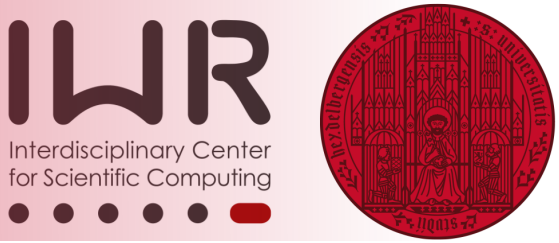


UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Towards the design of molecular materials

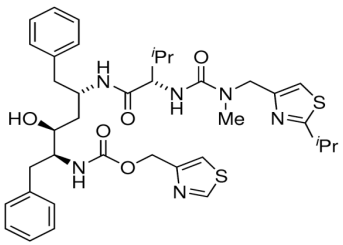
Gerit Brandenburg <j.g.brandenburg@gmx.de> | August 22, 2019

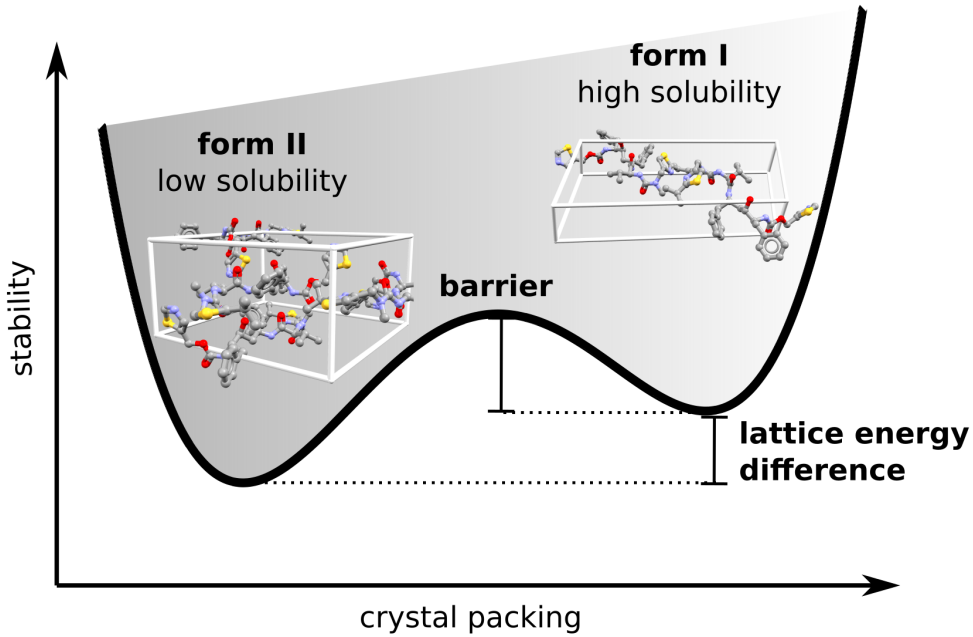
TALK AT 32ND EUROPEAN CRYSTALLOGRAPHIC MEETING, VIENNA, AUSTRIA

Interdisciplinary Center for Scientific Computing – Heidelberg University
Thomas Young Centre – University College London

Many properties depend on the polymorphic form of a crystal

API: ritonavir
- HIV antiviral drug
- industrially produced in form I
- form II found 2 years after market launch
- reformulation required [1]





- tools to predict possible polymorphs would be valuable[2]

  → Simulation methods aim at complementing experimental screening

[1] J. Bauer, et al., *J. Pharm. Res.* **18**, 859-866 (2001).

[2] S. L. Price, JGB, *Molecular Crystal Structure Prediction*; Elsevier Australia ISBN: 9780128098356 (2017).

Predict most stable crystal polymorphs based on the molecular diagram

The inability to predict something as simple as how a molecule would crystallize is one of the continuing scandals in the physical sciences.[3–5]

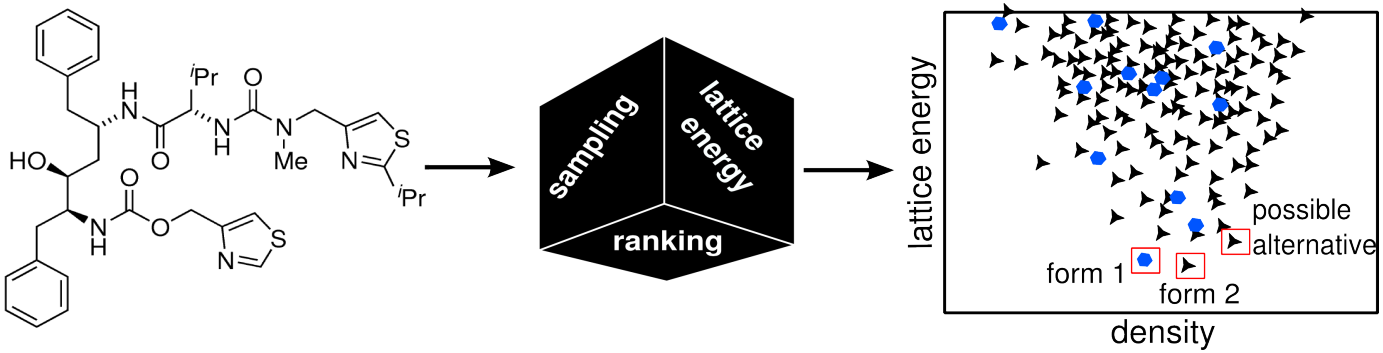
Task

- molecule is chosen due to its chemical/physical/biological properties
- based on the molecular diagram only, the most stable crystal structures should be predicted
- predict properties of interest for the most promising candidates

[3] A. Gavezzotti, *Acc. Chem. Res.* **27**, 309-314 (1994). [4] J. Maddox, *Nature* **335**, 201-201 (1988).

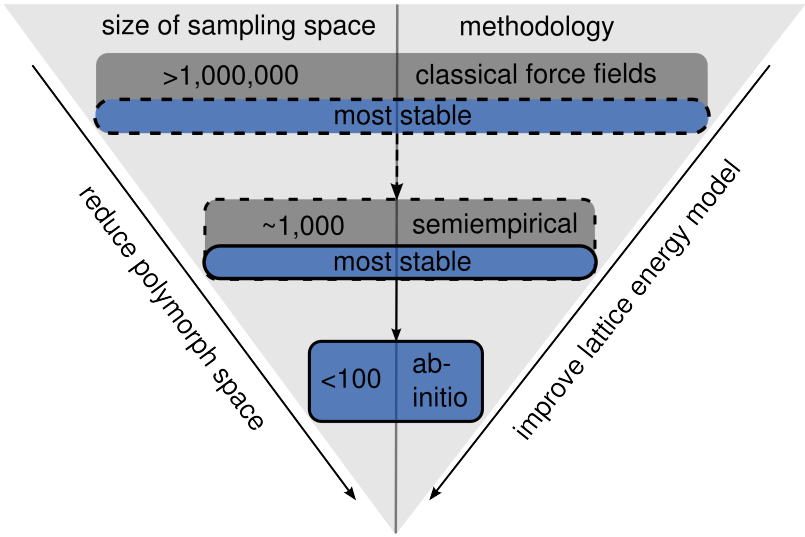
[5] K. N. Houk, F. Liu, *Acc. Chem. Res.* **50**, 539 (2017).

Sampling and energetic ranking for crystal structure prediction



Layers of complexity:

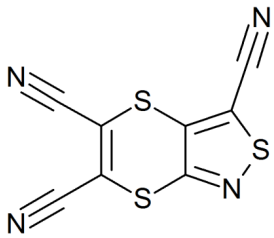
- molecular conformational space
- space groups (varying orientation, conformation, etc.)
- polymorph ranking according to free energy[6,7]



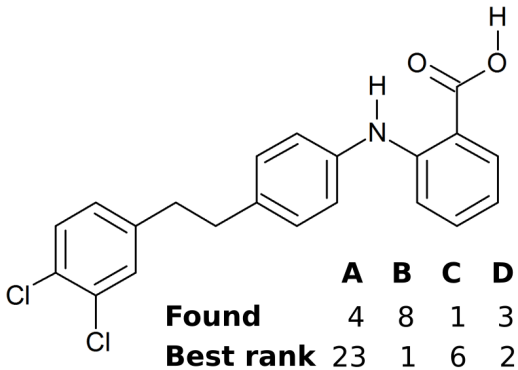
[6] S. Price, *Chem. Soc. Rev.* **43**, 2098 (2014)

[7] S. L. Price, JGB, *Molecular Crystal Structure Prediction*; Elsevier Australia, 336-363 (2017).

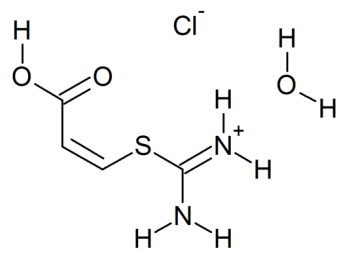
Results of the 6th blind test for organic crystal structure prediction



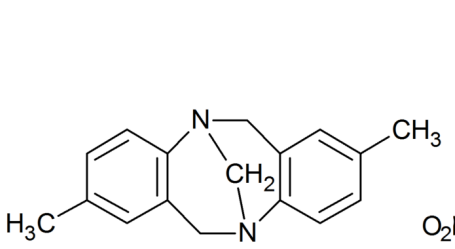
Found	12
Best rank	1



	A	B	C	D	E
Found	4	8	1	3	0
Best rank	23	1	6	2	

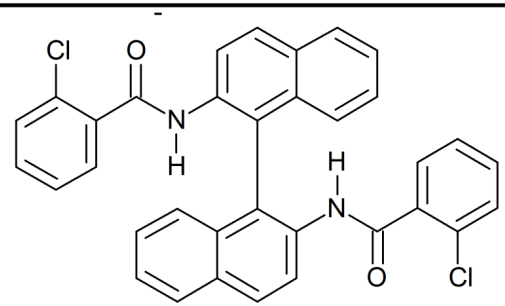


Found	1
Best rank	2



The Cambridge Crystallographic Data Centre

Found	5
Best rank	1



Found	3
Best rank	1

- No group had a perfect sampling and ranking of structures[8,9]

[8] A. Reilly, et al. *Acta Cryst. B*, **72**, 439 (2016). [9] J. G. Brandenburg, S. Grimme, *Acta Cryst. B*, **72**, 502 (2016).

CSP bind test highlights two remaining challenges

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

1. Ranking stage

Crystal polymorphs are often separated by just a few kJ/mol, exceeding the accuracy of standard density functional approximations (DFAs).
$\Longrightarrow$ **Are many-body methods feasible?**

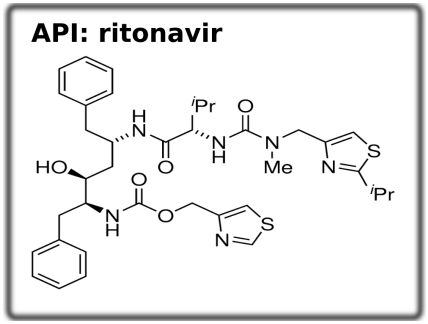
2. Sampling stage

Dealing with a vast search space, in particular for molecules with increased flexibility, one has to cover too many structures at DFT accuracy.
$\Longrightarrow$ **Can we improve modern DFAs & employ ML techniques?**

Outline of talk

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Challenge of treating the flexibility of large pharmaceutical-like molecules



- ritonavir is API with typical flexibility
- need of fast electronic structure method that can treat flexibility

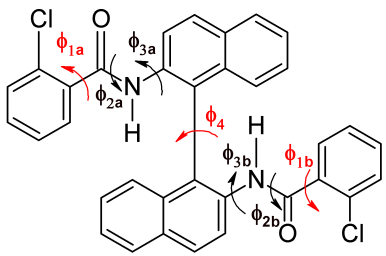
Density **F**unctional **T**ight **B**inding (DFTB3-D3):
3rd order, self consistent charges, 3OB atom-pair parameter, D3 dispersion

- huge speed-up compared to DFT ($\sim$ 4 orders of magnitudes)
- full treatment of all degrees of freedom
- good experience for proteins & energies of small molecular crystals

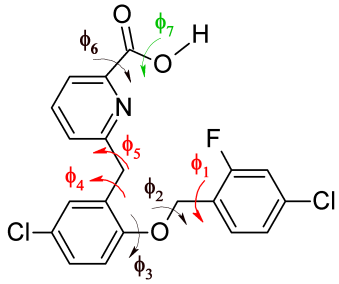
[10] A. S. Christensen, T. Kubař, Q. Cui, M. Elstner, *Chem. Rev.* **116**, 5301 (2016).

[11] JGB, S. Grimme, *J. Phys. Chem. Lett.* **5**, 1785 (2014).

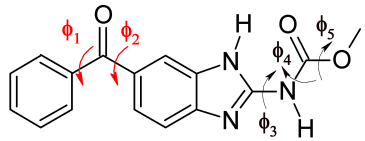
Focus on flexible molecules with previous CSP studies



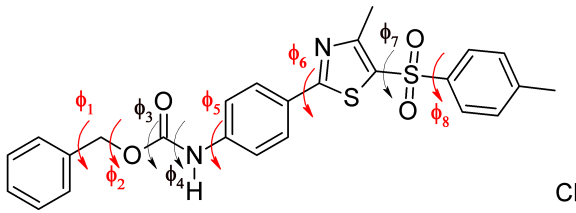
Molecule XXVI



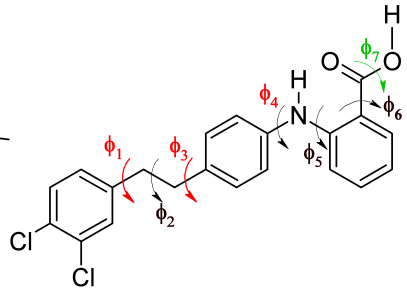
GSK269984B



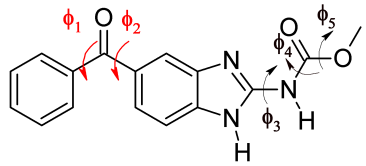
A-Mebendazole



Molecule XX



Molecule XXIII



C-Mebendazole

Possible advantages of DFTB3-D3 as an (intermediate) CSP step

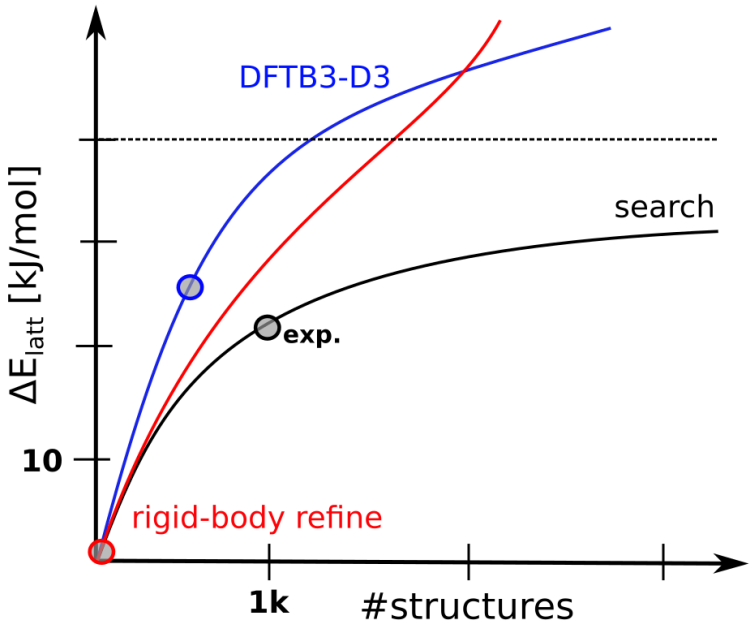
(1) improving geometries

(2) improving energy ranking

(3) reducing # minima

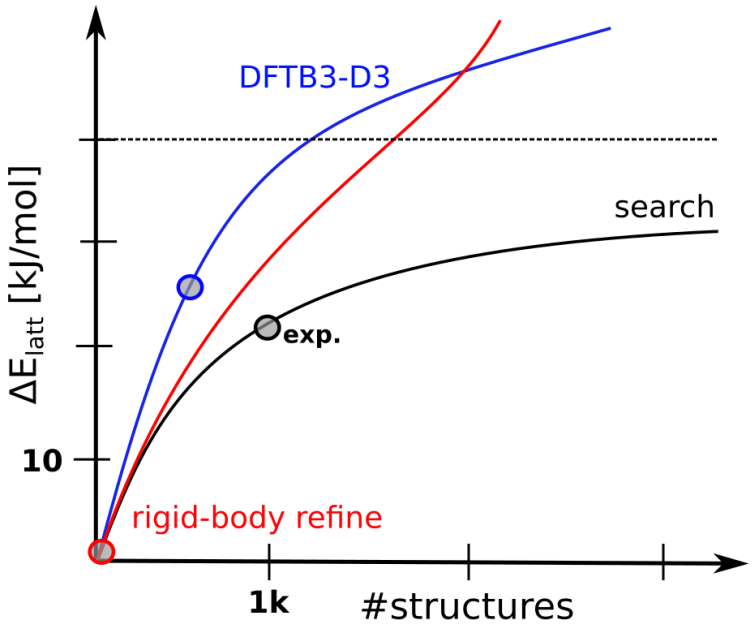
(4) free energy contributions

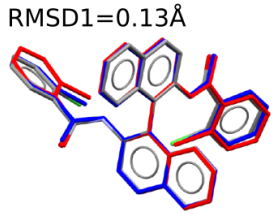
Though poor energetic ranking geometries improved



- 2.2 Mio structures generated for XXVI, 9 000 processed to reranking
- experimental polymorph is rather high in energy by DFTB3-D3
- successful reranking using rigid DFTB3-D3 structure

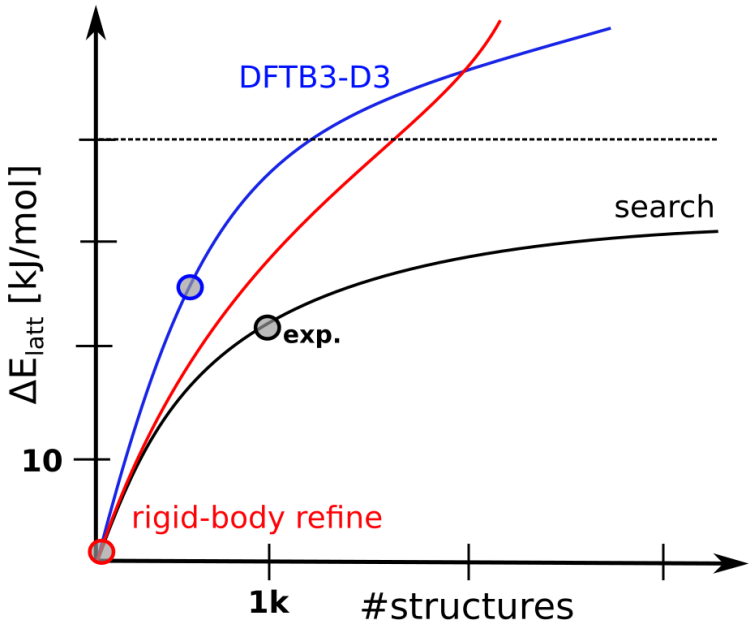
Though poor energetic ranking geometries improved

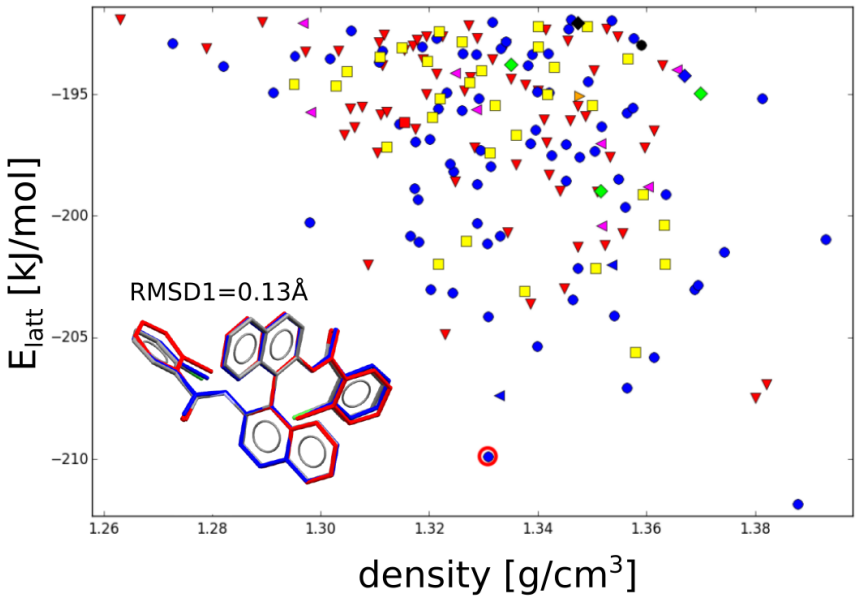




- 2.2 Mio structures generated for XXVI, 9 000 processed to reranking
- experimental polymorph is rather high in energy by DFTB3-D3
- successful reranking using rigid DFTB3-D3 structure

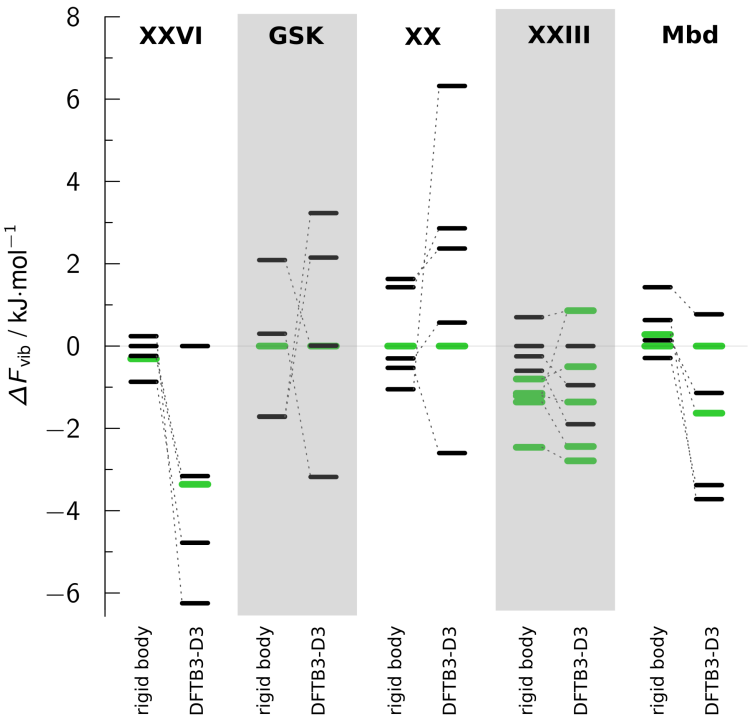
Though poor energetic ranking geometries improved





- 2.2 Mio structures generated for XXVI, 9 000 processed to reranking
- experimental polymorph is rather high in energy by DFTB3-D3
- successful reranking using rigid DFTB3-D3 structure
- final energy landscape: no important structure is lost

Impact of harmonic free energy contributions to relative stability

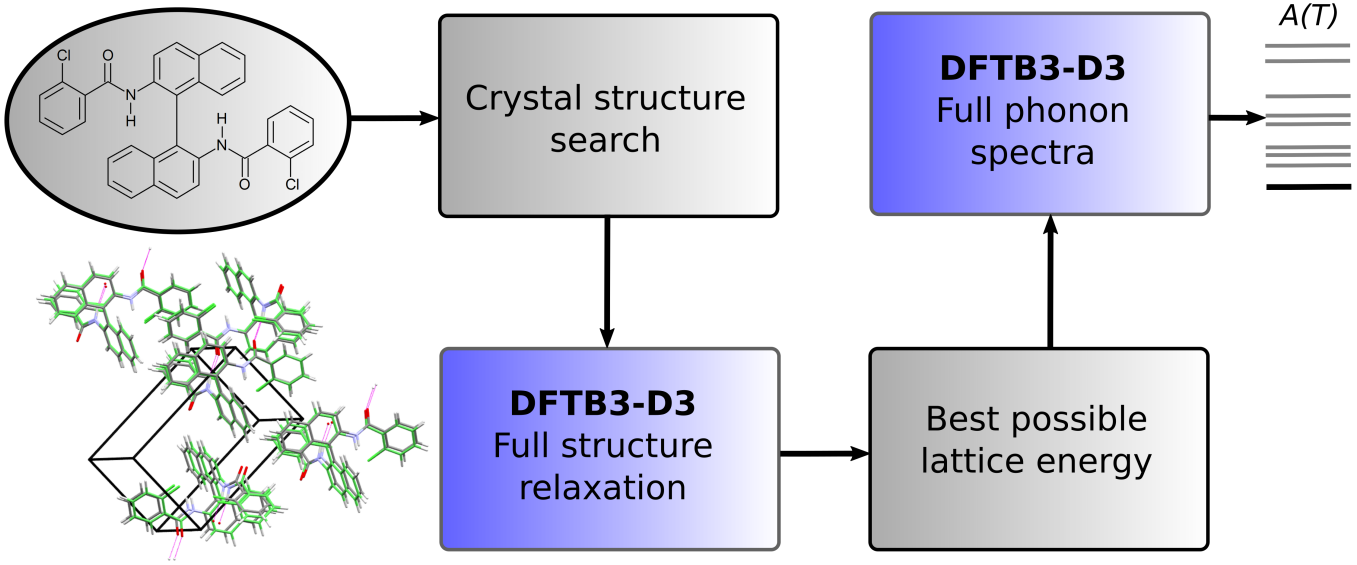


- significant impact of ZPE and thermal contribution on relative stability
- modified ranking of several putative polymorphs
- treating molecules as rigid underestimates the free energy
- rigid-body estimate: 9% of polymorph change stability upon including thermal contributions

[12] A.J. Cruz-Cabeza, J. Bernstein, *Chem. Rev.* **114** 2170 (2014).

[13] J. Nyman, G. M. Day, *CrystEngComm* **17**, 5154 (2015).

DFTB3-D3 intermediate optimization would be scaleable for larger molecules



Possible merits of DFTB3-D3

(1) improving geometries **yes**

(2) improving energy ranking **no**

(3) reducing # minima **partially**

(4) free energy contributions **yes**

[14] L. Iuzzolino, P. McCabe, S. L. Price, JGB, *Faraday Discuss.* **211**, 275 (2018).

Outline of talk

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Methods with 'benchmark quality' for noncovalent interactions

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

CCSD(T)

+ 'gold standard' of quantum chemistry
+ linear scaling variants
- pair-screening thresholds
- slow basis set convergence
- non-trivial to parallelize (shared memory: 100 GB/core)

DMC

+ exact projection to ground state
+ N^3 scaling, [-] huge prefactor
+ fast basis set convergence
+ scaleable to HPCs
- stochastic error
- uncertainties from FN and non-local pseudopotentials

No benchmark quality: MP2, RPA, CCSD, CCSD(T)/cc-pVTZ, VMC

Quantum Monte-Carlo in a nutshell: A scalable high-level method

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Fixed-node diffusion Monte-Carlo

1) Enforce nodal surface of Fermions

$$\Gamma = \{\mathbf{R}; |\Psi_T\rangle = 0\}$$

2) Walkers in configuration space

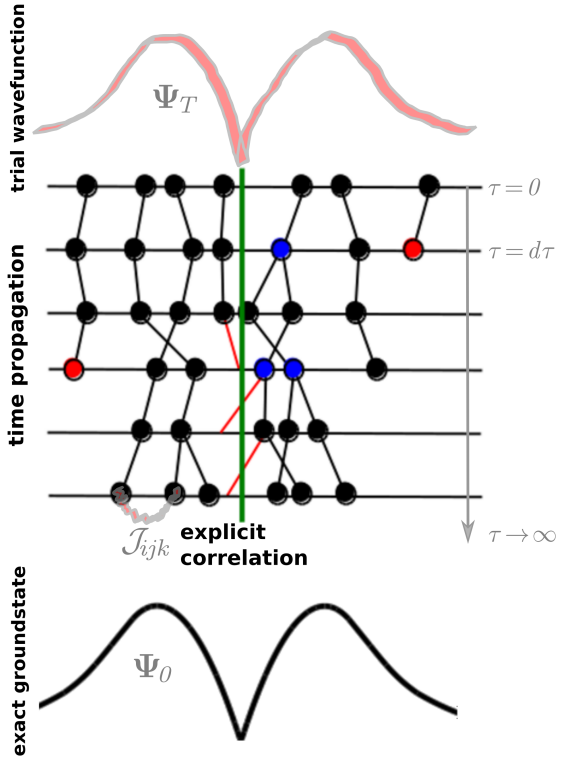
$$|\Psi_T(\mathbf{R},\tau)\rangle = \text{hist}\left[\sum \delta(\mathbf{R} - \mathbf{R_i}(\tau))\right]$$

3) Diffusion in imaginary time

$$\partial_\tau |\Psi_T(\mathbf{R},\tau)\rangle = \left[\frac{1}{2}\nabla^2_{\mathbf{R}} - (V - E_T)\right] |\Psi_T(\mathbf{R},\tau)\rangle$$

4) Projection to exact ground state

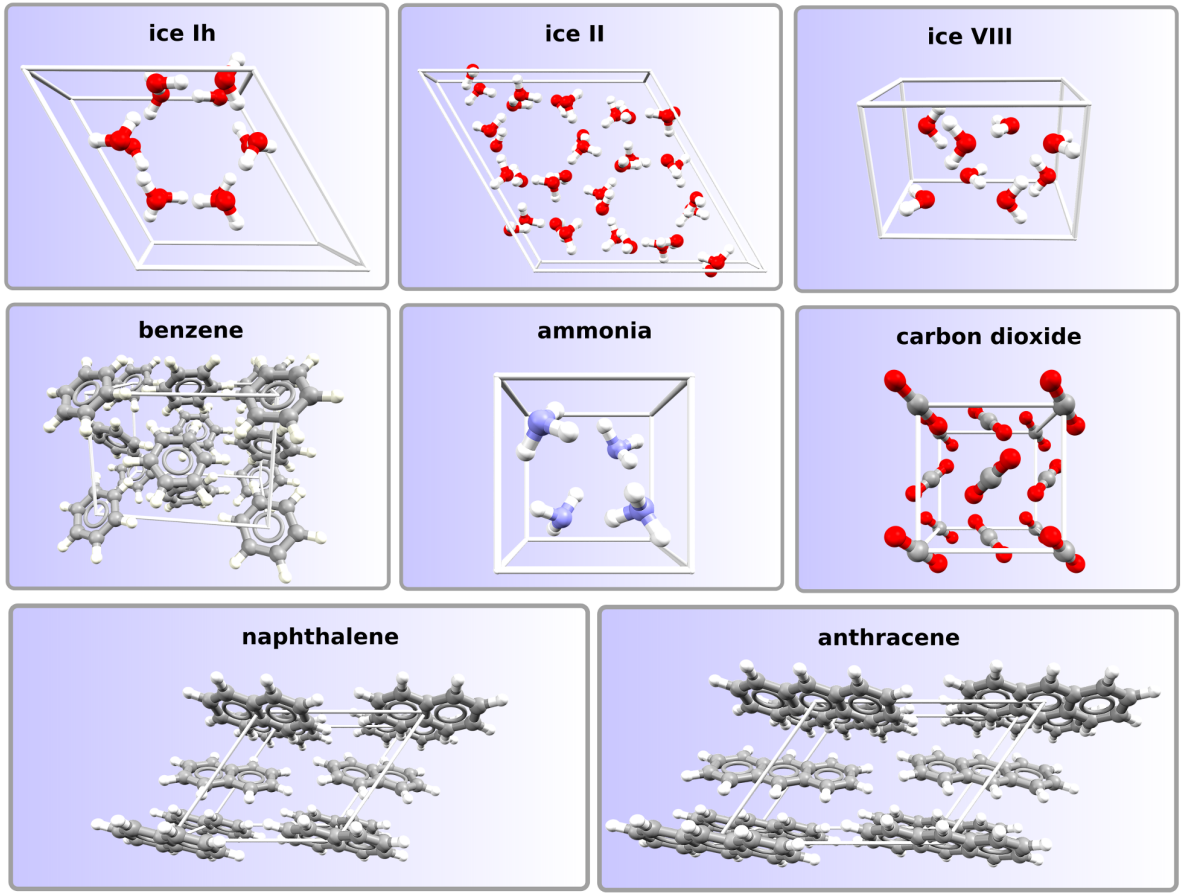
$$|\Psi_0(\mathbf{R})\rangle = \lim_{\tau\to\infty} \exp\left[-\tau(\hat{H} - E_T)\right] |\Psi_T(\mathbf{R},\tau)\rangle$$



- exact within Γ and $d\tau \to 0$ and $\tau \to \infty$[15]
- new DMC algorithms lead to substantial speed up[16]

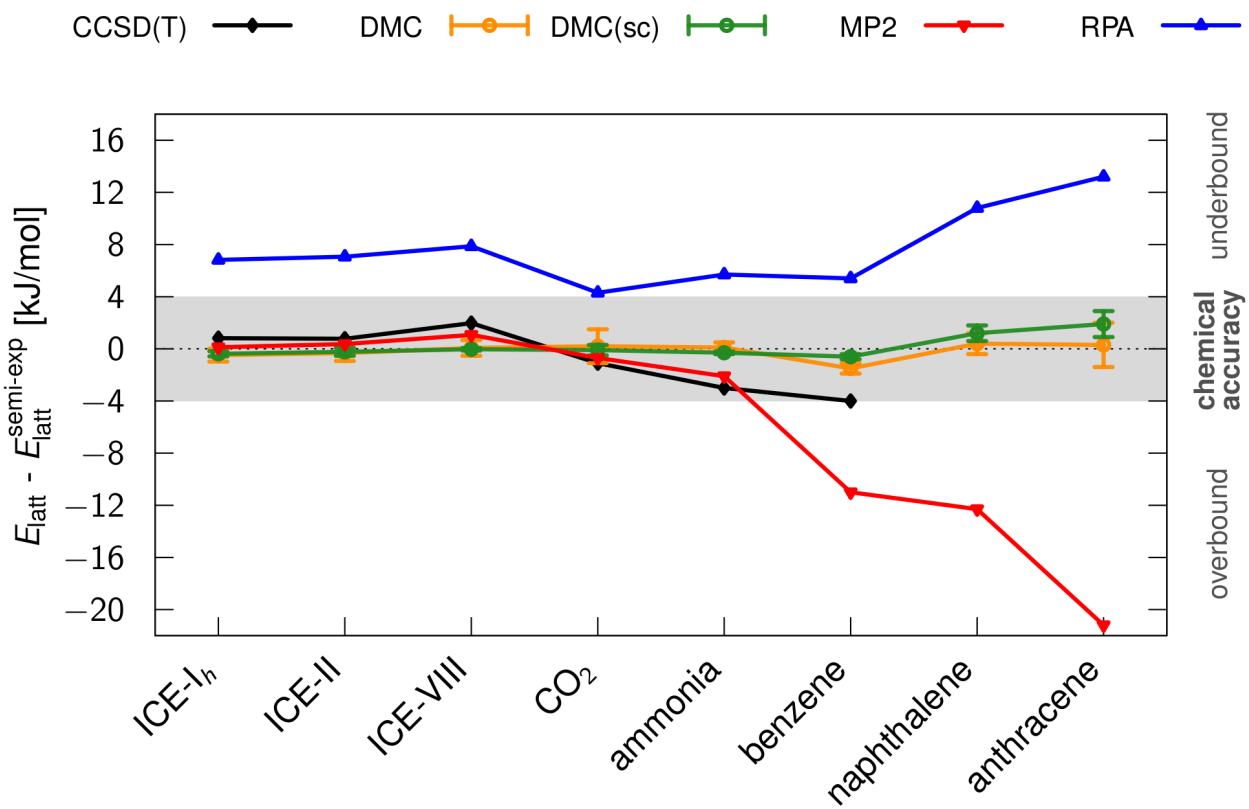
[15] *Chem. Rev.* **116**, 5188 (2016) [16] A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, D. Alfé, *Phys. Rev. B* **93**, 241118(R) (2016).

Molecular crystals as challenging test



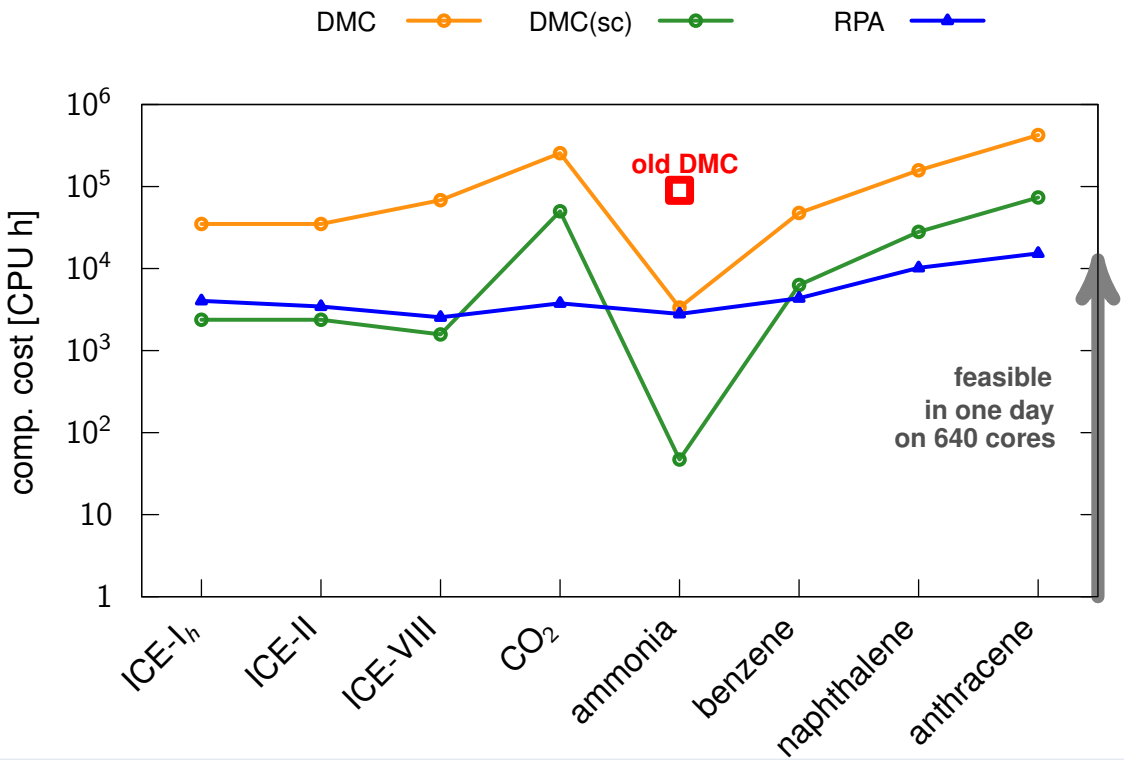
- strong H-bonds, vdW of saturated and unsaturated molecules
- problematic for all readily applicable methods (DFA-DISP, MP2)

QMC delivers (sub-) chemical accuracy for all tested systems



- excellent agreement with experiment and CCSD(T)
- uncertainty in H_{sub}^{exp} probably larger than DMC errors

QMC is feasible within one day on standard computer cluster



- up to three orders of magnitude speed-up compared to best DMC practice two years ago
- significance will extend to all classes of systems[17]

[17] A. Zen, JGB, J. Klimeš, A. Tkatchenko, D. Alfè, A. Michaelides, *Proc. Natl. Acad. Sci. U.S.A* , **115**, 1724 (2018).

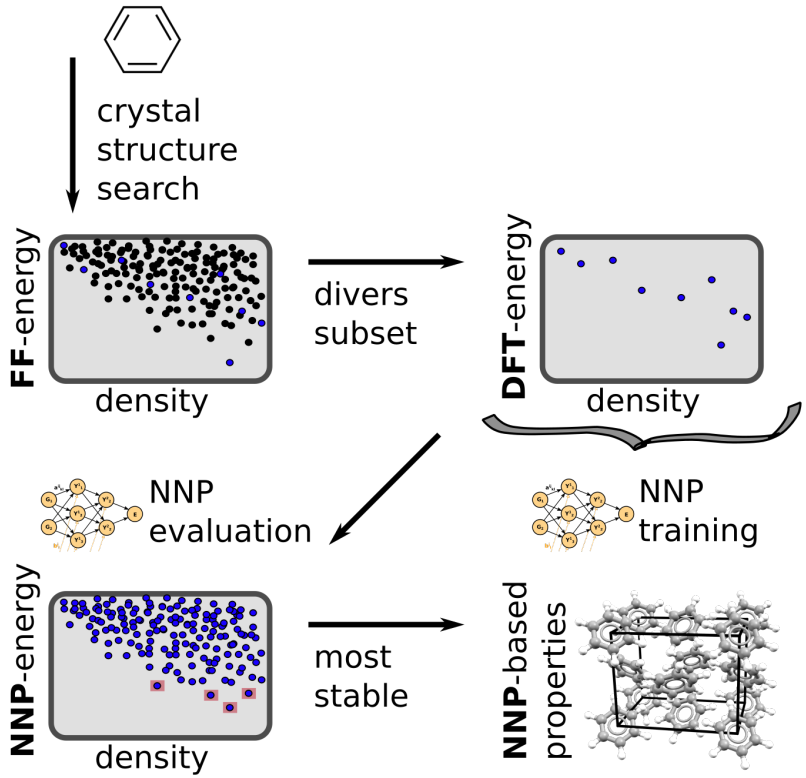
Outline of talk

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Employing ML to enhance CSP workflow

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

1. Generate putative structures with crude FF stability
2. Construct a divers subset of structures
3. Compute DFT stability on subset
4. Train a high-dimensional NNP on DFT references
5. Recompute stability of all structures employing the NNP



[18] V. L. Deringer, D. M. Proserpio, G. Csányi, and C. J. Pickard, *Faraday Discuss.* **211**, 45 (2018).

[19] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, *Phys. Rev. B* **99**, 064114 (2019).

[20] JGB and J. Behler, *in preparation*.

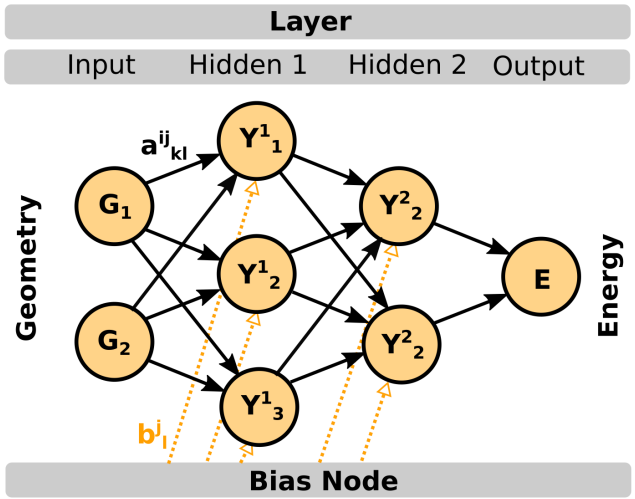
High-dimensional neural network for benzene crystal landscape

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Energy decomposed to atomic contributions

$$E_{\text{tot}} = \sum_{i}^{\text{atoms}} E_i \left[Z_i, G(\{R_j\}) \right]$$

- symmetry functions as local geometry descriptors G[21–23]
- element specific feed-forward neural network (2 hidden layers, 25 nodes each)
- training based on electronic energies and nuclear forces



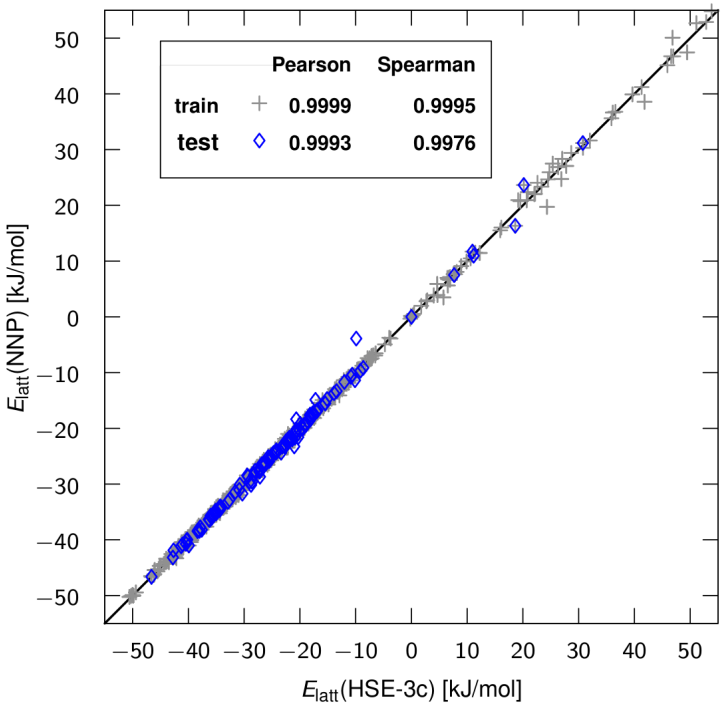
[21] J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).

[22] J. Behler, *J. Chem. Phys.* **134**, 074106 (2011).

[23] J. Behler, *Angew. Chem. Int. Ed.* **56**, 12828 (2017).

Stable training on 10% of generated putative polymorphs

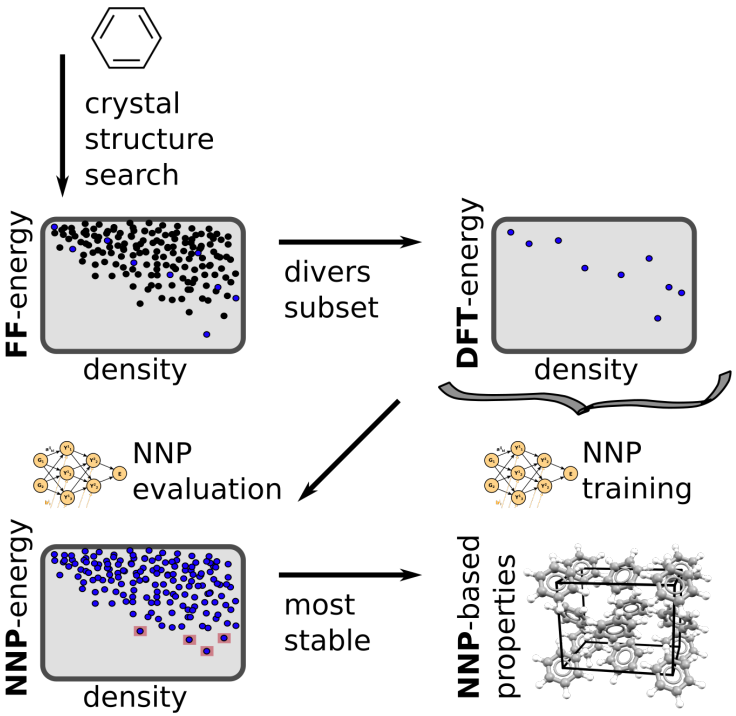
	training	testing
#structures	1 077	127
#data points	39 849	4 699
range E [kJ/mol]	[-50:147]	[-55:120]
range ρ [g/cm^3]	[0.1:1.1]	[0.1:1.1]
RMSD(E) [kJ/mol]	**0.53**	**0.93**
RMSD(F) [a.u.]	0.001	0.001

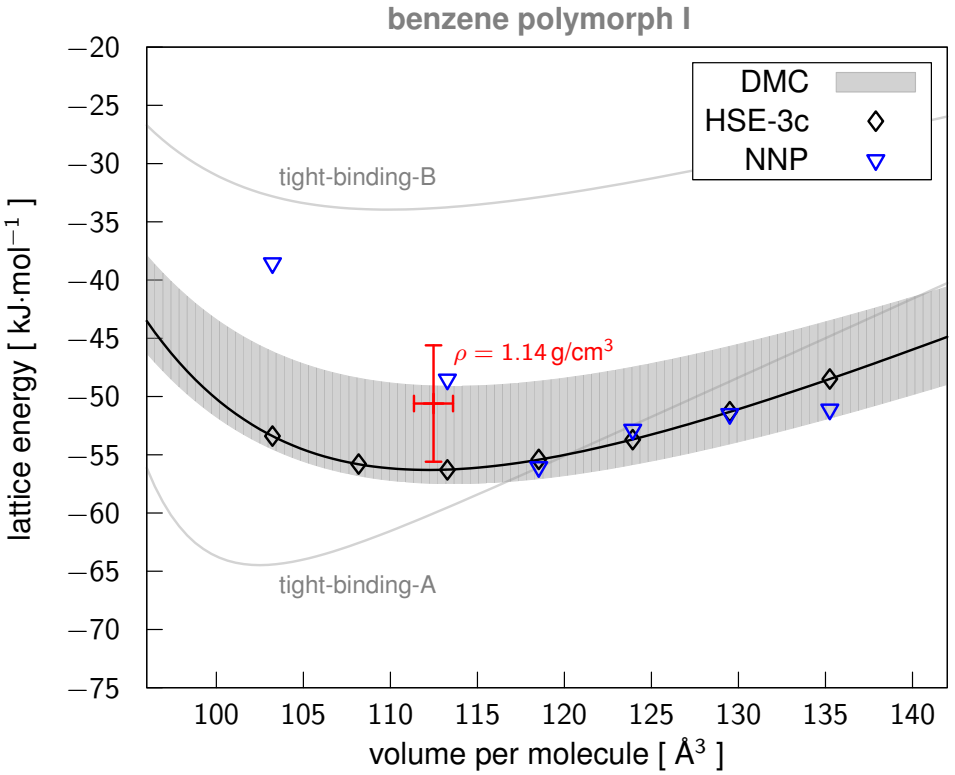


- neural network delivers sub-chemical accuracy
- CSP cost reduced by one order of magnitude [20]

[20] JGB and J. Behler, *in preparation*.

Employing ML to enhance CSP workflow





- EOS within 1 kJ/mol for ρ <1.1 g/cm^3, large errors for ρ >1.2 g/cm^3
- too few data points to generate globally smooth EOS

[20] JGB and J. Behler, *in preparation*.

Outline of talk

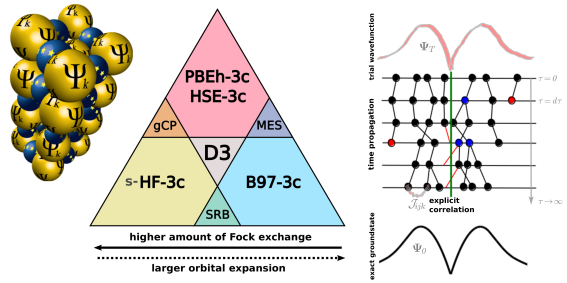
UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

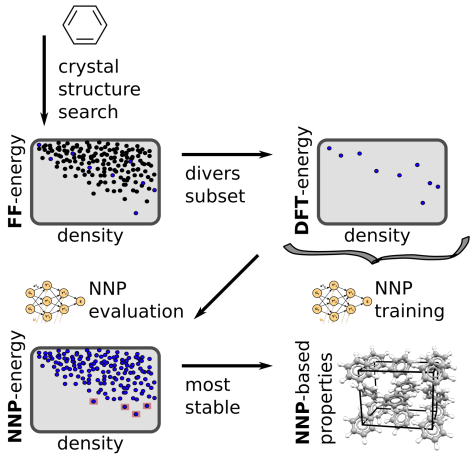
Remaining challenges in crystal structure prediction

- Thermodynamics of crystals, i.e. full PT phase diagram
- Kinetics of crystallization
- Reducing the human and CPU time for prediction

Summary

- Fast electronic structure methods to treat flexible molecules
- High-level solution of Schrödinger eq. for accurate energies
- Enhancing CSP by machine learning





Acknowledgements

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

PhD / Postdocs


Andrea Zen


Martin Fitzner


Michaelides (UCL)

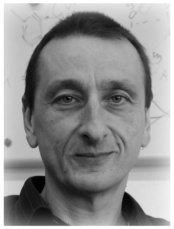

Alfè (UCL)

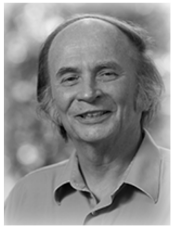

Tkatchenko (Luxembourg)


Klimes (Prague)


Benjamin Ramberger


Theodoros Tsatsoulis


Grimme (Bonn)


Perdew (Temple)


Kresse (Vienna)


Behler (Göttingen)


Rebecca Sure (BASF)


Eike Caldeweyher


Luca Iuzzolino


Price (UCL)


Grüneis (Vienna)


Civalleri (Turin)


Maschio (Turin)


Antreas Afantitis (NovaMechanics)

Principal investigator

Alexander von Humboldt Stiftung/Foundation

THE THOMAS YOUNG CENTRE TYC

erc

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

ATERIALS AND MOLECULAR MODELLING HUB

CSCS

Acknowledgements

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

PhD / Postdocs


Andrea Zen


Martin Fitzner


Michaelides (UCL)

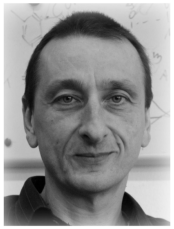

Alfè (UCL)

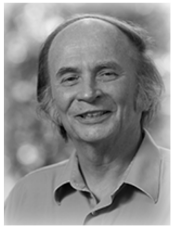

Tkatchenko (Luxembourg)


Klimes (Prague)


Benjamin Ramberger


Theodoros Tsatsoulis


Grimme (Bonn)


Perdew (Temple)


Kresse (Vienna)


Behler (Göttingen)


Rebecca Sure (BASF)


Eike Caldeweyher


Luca Iuzzolino


Price (UCL)


Grüneis (Vienna)


Civalleri (Turin)


Maschio (Turin)


Antreas Afantitis (NovaMechanics)

Principal investigator

Thanks

Alexander von Humboldt Stiftung/Foundation

THE THOMAS YOUNG CENTRE

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

MATERIALS AND MOLECULAR MODELLING HUB

CSCS

Key references

- Crystal structure prediction:

  S. L. Price, JGB, *Molecular Crystal Structure Prediction*, G. DiLabio, A. Otero-de-la-Roza, Eds., Elsevier Australia, 336-363 (2017).

  L. Iuzzolino, P. McCabeb, S. L. Price, JGB, *Faraday Discuss.* **211**, 275 (2018).

- QMC for molecular crystals:

  A. Zen, JGB, J. Klimeš, A. Tkatchenko, D. Alfè, A. Michaelides, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1724 (2018).

- DFT development

  S. Grimme, A. Hansen, JGB, C. Bannwarth, *Chem. Rev.* **116**, 5105 (2016).

  E. Caldeweyher, JGB, *J. Phys.: Condens. Matter* **30**, 213001 (2018).

- website: **gerit-brandenburg.de**