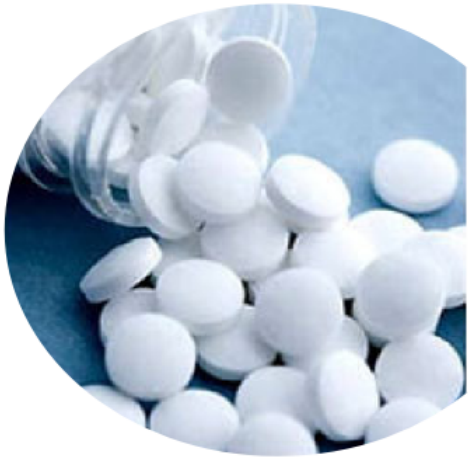
Molecular crystal polymorph prediction via multilevel strategies

Gerit Brandenburg <g.brandenburg@ucl.ac.uk> | 30th of May 2017

PHYSICAL CHEMISTRY INSTITUTE - UNIVERSITY OF GÖTTINGEN, GER

Outline of talk

Understanding phenomena in nature



How can a Gecko stick to a glass wall?

- cannot be explained by classical mechanics
- London dispersion (C_6)

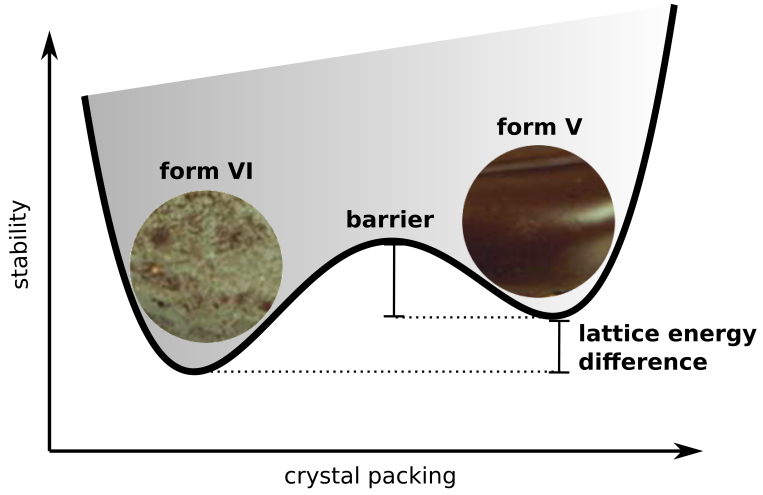
Why does a tablet change its properties?

- van der Waals interaction stabilizes tablet dependent on its form
- relevant for production process and dosage

Many properties depend on the polymorphic form of a crystal

Polymorphism

- ability of a molecule to crystallize in more than one structure
- properties change with crystal packing, e.g. solubility, color, etc.[1]



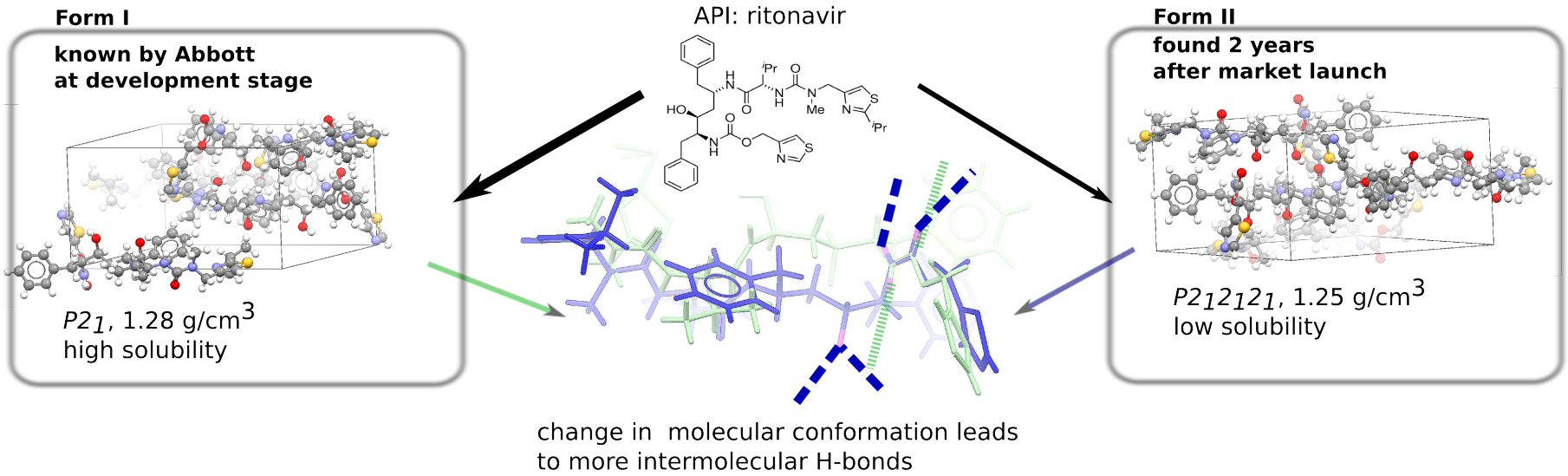
- cocoa butter (form VI) has a dull surface, soft texture, higher melting point
- metastable form V has glossy surface, crisp hardness, melts at 300 K[2]

[1] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, J. Bernstein, *Chem. Soc. Rev.* **44**, 8619-8635 (2015).

[2] S. T. Beckett, *Science of Chocolate*; RSC Paperbacks (2000).

[3] S. L. Price, JGB, *Molecular Crystal Structure Prediction; Non-covalent interactions in Quantum Chemistry and Physics*, G. DiLabio, A. Otero-de-la-Roza, Eds., Elsevier Australia, Melbourne, in press (2017).

Late appearing polymorph disrupted supply of antiviral drug ritonavir



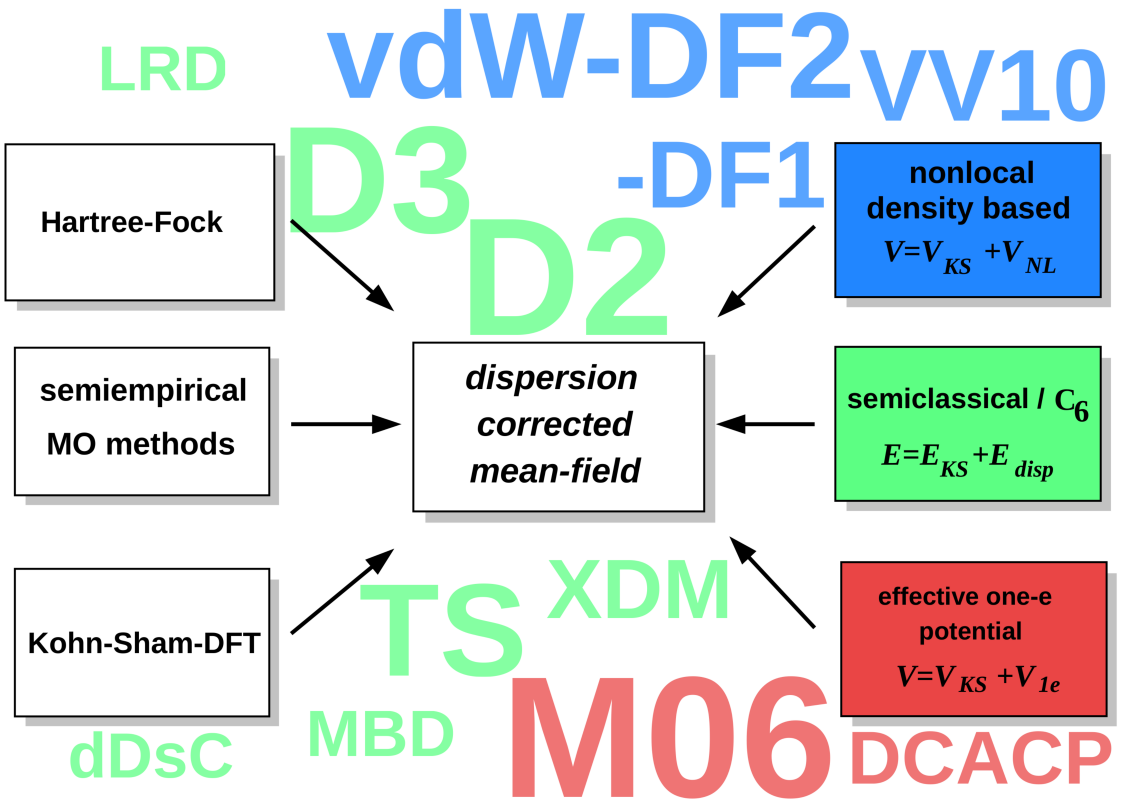
- ritonavir was industrially produced in form I for 2 years
- more stable form II suddenly crystallized
- much lower solubility made reformulation necessary[4]

→ Tools to predict possible polymorphs would be valuable

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

Outline of talk

Many good dispersion corrections exist



- focus on D3 scheme as it can be used at all computational levels
- own work started with extension of D3 type corrections

[5] A. Stone *The Theory of Intermolecular Forces, 2nd ed.*; Oxford University Press, Oxford (2013)

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

Exact expression for correlation energy cannot be solved for many systems

$$E_c = -\frac{1}{2\pi}\int_0^1 \mathrm{d}\lambda \int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \frac{1}{|\mathbf{r}-\mathbf{r}'|} \times \int_0^\infty \mathrm{d}\omega \left[\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega) - \chi_0(\mathbf{r},\mathbf{r}',i\omega)\right]$$

- from adiabatic connection fluctuation dissipation theorem[7]
- Coulomb interaction scaled by λ
- dynamical charge density susceptibility

$$\chi_0(\mathbf{r},\mathbf{r}',i\omega) = -4\sum_i\sum_a \frac{\omega_{ai}}{\omega_{ai}^2+\omega^2}\varphi_i(\mathbf{r})\varphi_a(\mathbf{r})\varphi_a(\mathbf{r}')\varphi_i(\mathbf{r}')\,,$$

$\rightarrow$ approximation needed

[7] A. Zangwill, P. Soven, *Phys. Rev. A* **21**, 1561 (1980)

Coarse-graining to atomic contributions leads to significant simplifications

- coarse-grain to atomic contributions

$$\alpha_{ij}(i\omega) = \int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\, \mathbf{r}_i \mathbf{r}'_j \chi(\mathbf{r}, \mathbf{r}', i\omega)$$

- expansion of Coulomb operator in multipoles[8]

$$E_{\text{disp}}^{AB} = -\underbrace{\frac{3}{\pi}\int \mathrm{d}\omega\, \alpha_A(i\omega)\alpha_B(i\omega)}_{C_6^{AB}} \times \frac{1}{R_{AB}^6}$$

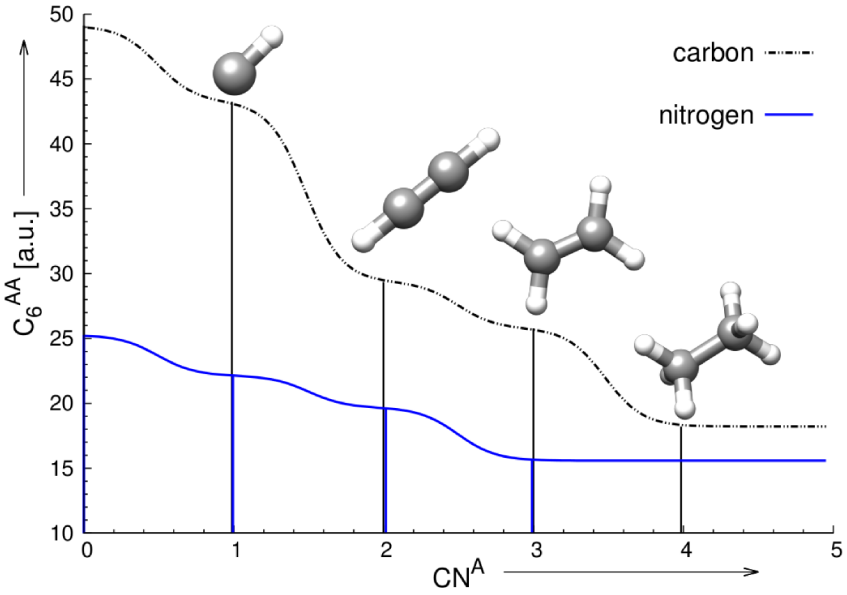
- C_8^{AB} and C_9^{AB} from recursion relations and averages[9]

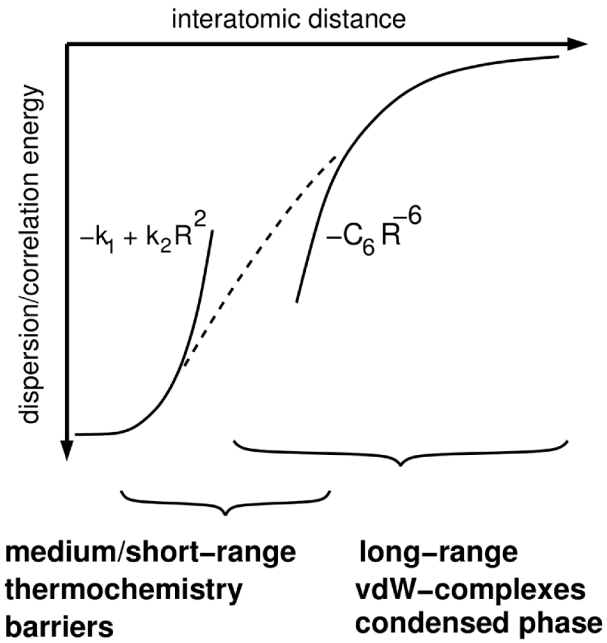
$$E_{\text{disp}}^{(\text{D3})} = -\frac{1}{2}\sum_{n=6,8}\sum_{A,B}^{\text{atoms}} \frac{C_n^{AB}}{R_{AB}^n} \cdot f_n^d(R_{AB}) - \frac{1}{6}\sum_{A,B,C}^{\text{atoms}} \frac{C_9^{ABC}}{R_{ABC}^9} \cdot f_9^d(R_{ABC}, \theta_{ABC})$$

[8] H. B. G. Casimir. D. Polder, *Phys. Rev.* **73**, 360 (1948)

[9] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010)

Geometric coordination number for interpolations of reference C_6

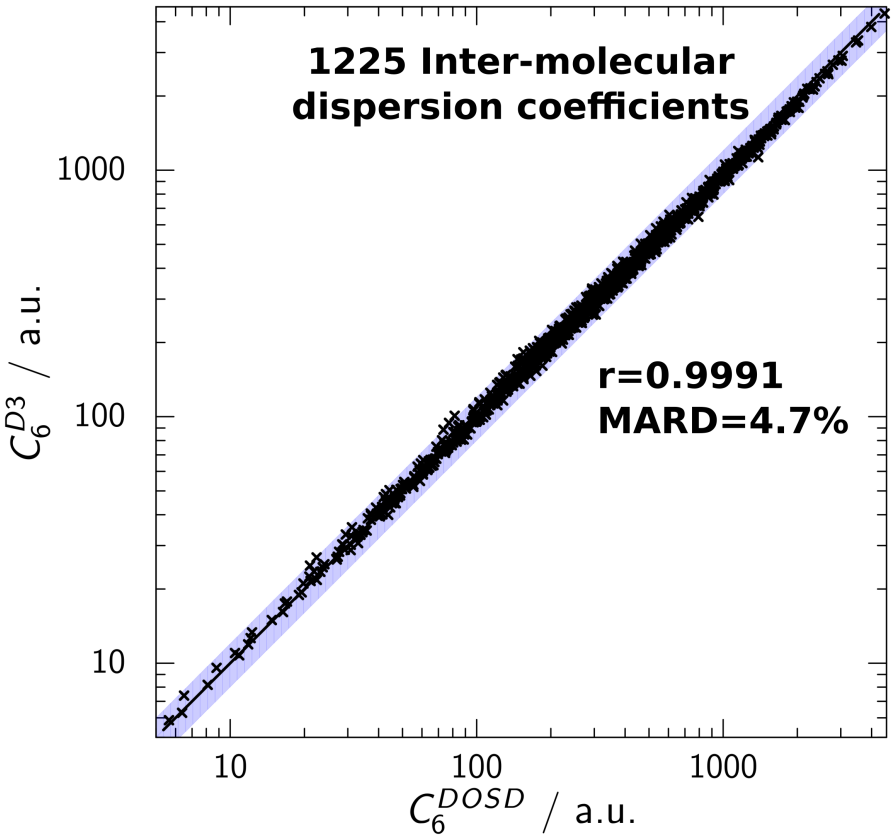




- atoms-in-molecules C_6 by TD-DFT of $\alpha(i\omega)$ on model hydrides
- reference C_6 are mapped to real system via geometrical coordination
- short-range damping avoids double counting[11]

[11] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.*, **32**, 1456 (2011)

Semi-classical scheme yields highly accurate dispersion coefficients



D3 correction

- dipole oscillator strength distribution (DOSD) yield $C_6^{\text{exptl.}}$ (compiled by A. Tkatchenko)
- residual long-range mean absolute relative deviation (MARD) of D3 $<$ 5% [12]
- deviations are close to intrinsic TD-DFT errors of $\alpha(i\omega)$

[12] S. Grimme, *WIREs Comput. Mol. Sci.* **1**, 211-228 (2011)

D3 in a nutshell: High accuracy at force field speed

Advantages of D3 model

- intermolecular C_6 coefficients are very accurate
- no electronic structure input needed
- computation is extremely fast D3(two-body) $\sim$ FF speed
- analytical first (and second) derivatives

Possible shortcomings

- no automatic adjustment to unusual electronic structures[13]
- missing anisotropy of dispersion interaction
- no many-body contributions beyond Axilrod-Teller-Muto term
- high empiricism in short-range damping

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

[13] work on extension by E. Caldeweyher, S. Grimme, C. Bannwarth, *J. Chem. Phys.* submitted (2017)

Outline of talk

Lessons learned in the past years

- realistic structures are the key to many important physical and chemical properties
- Kohn-Sham density functional theory is method of choice for structures; wavefunction methods may take over for energies
- configurational sampling, entropy, and solvation issues are as important as good convergence in electronic energy

Multilevel methodologies: finding the right compromise

	task/property	example method
accurate QM	single-point energy	DLPNO-CCSD(T) DMC, RPA+SE
cheap QM	optimization	metaGGA (SCAN-D3[14]) HSE-3c[15,16]
very cheap QM	optimization/Hessians conformations	semi-empirical HF-3c[17,18], DFTB3-D3[19]
force field	dynamics conformational sampling	transferable or molecule specific (QM derived) FF

[14] JGB, J. E. Bates, J. Sun, J. P. Perdew *Phys. Rev. B*, **94**, 115144 (2016)

[15] JGB, E. Caldeweyher, S. Grimme, *Phys. Chem. Chem. Phys.*, **18**, 15519 (2016)

[16] S. Grimme, JGB, C. Bannwarth, A. Hansen, *J. Chem. Phys.*, **143**, 054107 (2015)

[17] R. Sure, S. Grimme, *J. Comput. Chem.*, **34**, 1672 (2013) [18] JGB, S. Grimme, *Top. Curr. Chem*, **345**, 1 (2014)

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

HSE-3c: Small basis DFT with semi-classical corrections

UCL

Requirements

- $\sim$10 $\times$ faster vs. standard DFA $\rightarrow$ small atomic orbital expansion
- reduce self-interaction error $\rightarrow$ use Fock exchange
- numerically robust $\rightarrow$ long-range screening of exchange

$$E_{\text{tot}}^{\text{HSE-3c}} = E^{(\text{modHSE})} + E_{\text{DISP}}^{(\text{D3})} + E_{\text{BSSE}}^{\text{gCP}}$$

Technical details:[15]

- modified HSE[20] in small def2-mSVP[16] basis set
- D3 and gCP semi-classical corrections (7 global parameters)

[15] JGB, E. Caldeweyher, S. Grimme, *Phys. Chem. Chem. Phys.*, **18**, 15519 (2016)

[16] S. Grimme, JGB, C. Bannwarth, A. Hansen, *J. Chem. Phys.*, **143**, 054107 (2015)

[20] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **124**, 219906 (2006)

Compromise of known functionals for exchange correlation functional

$$E_{\mathrm{xc}}^{(\mathrm{modHSE})} = \mathbf{a_x}\, E_{\mathrm{x}}^{(\mathrm{HF,SR})}(\omega) + (1 - a_x)\, E_{\mathrm{x}}^{(\mathrm{HSE,SR})}(\omega) + E_{\mathrm{x}}^{(\mathrm{HSE,LR})}(\omega) + E_{\mathrm{c}}^{(\mathrm{modPBE})}$$

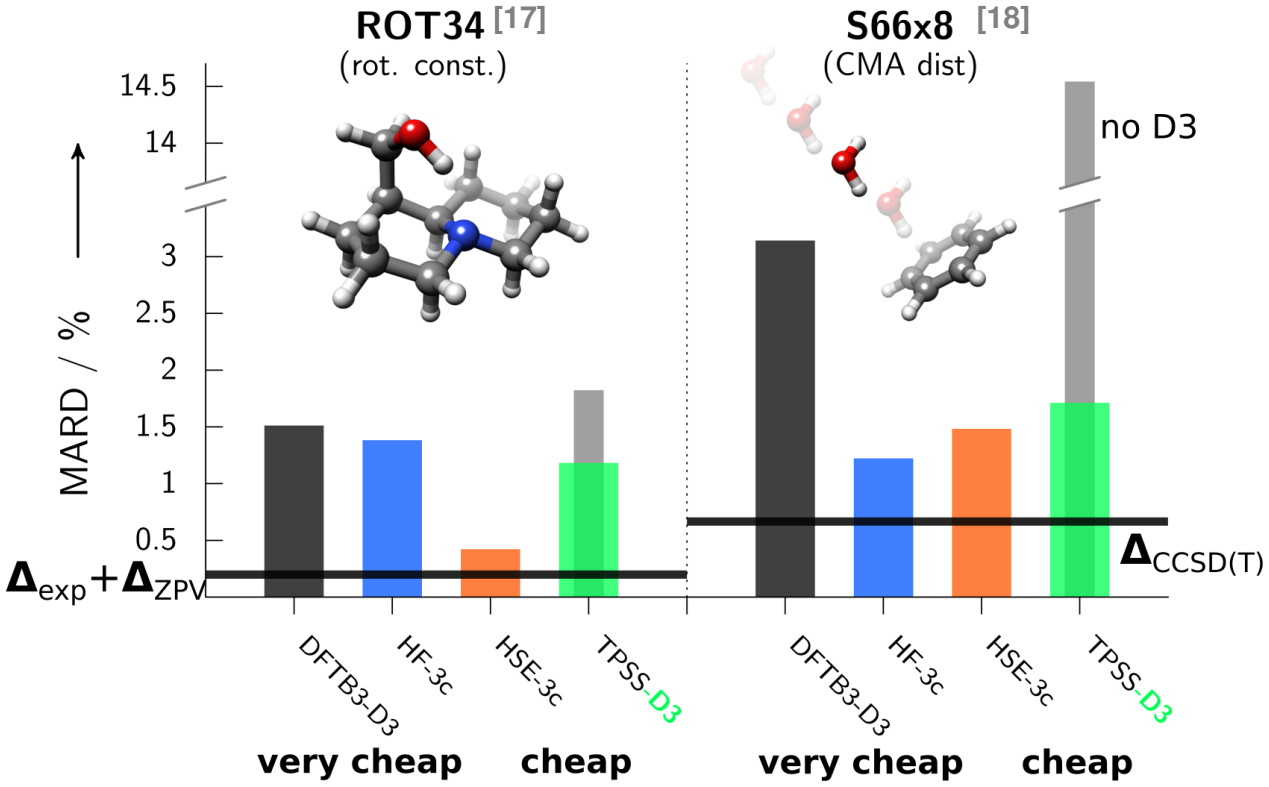
- modified HSE to reproduce modified PBE-XC

$$F_X^{\mathrm{PBE}} = 1 + \frac{\mu s}{1 + \frac{\mu s^2}{\kappa}}, \qquad s = |\nabla\rho/\rho^{4/3}|$$

- μ from PBEsol, κ averaged from PBE/revPBE
- $\beta = 0.03$ in F_C^{PBE} fitted to atomization energies
- $a_x = 0.42$: getting bond length right (standard range-separation $\omega = 0.11$)
- mSVP atomic orbitals fixed and available for whole PES

$\rightarrow$ only **seven** globally fitted parameters

Systematically improvable bond length and molecular structures

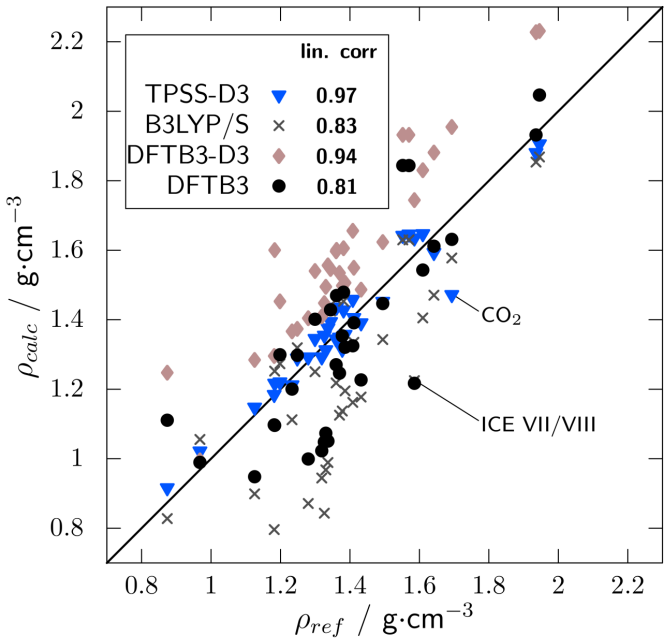


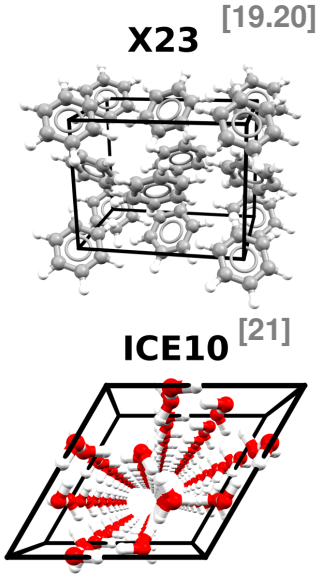
- error estimates from measurement and theoretical back-correction
- significant effect of dispersion interaction

[17] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.*, **15**, 16031 (2013)

[18] J. Rezáč, K. Riley, P. Hobza, *J. Chem. Theory Comput.*, **8**, 2427 (2011)

Good mass densities on diverse set of molecular crystals





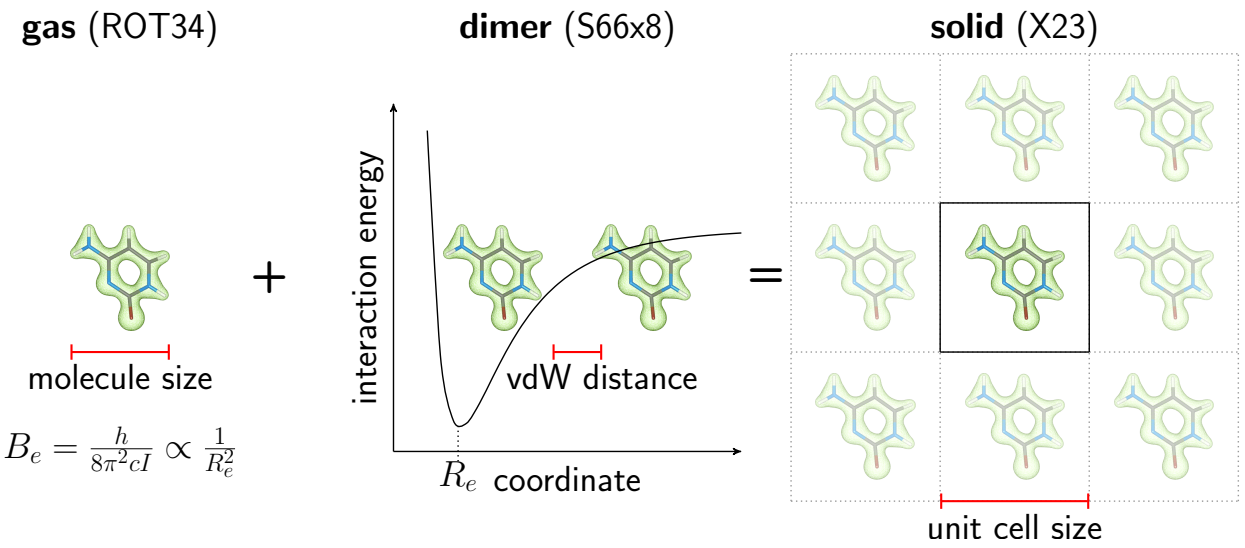
- impact of zero-point and thermal contribution important
- outliers due to problematic induction effects

[19] E. Johnson, A. Otero-de-la-Roza, *J. Chem. Phys.* **137**, 054103 (2012)

[20] A. Reilly, A. Tkatchenko, *J. Chem. Phys.* **139**, 024705 (2013)

[21] JGB, T. Maas, S. Grimme, *J. Chem. Phys.* **142**, 124104 (2015)

Consistency for structures achieved



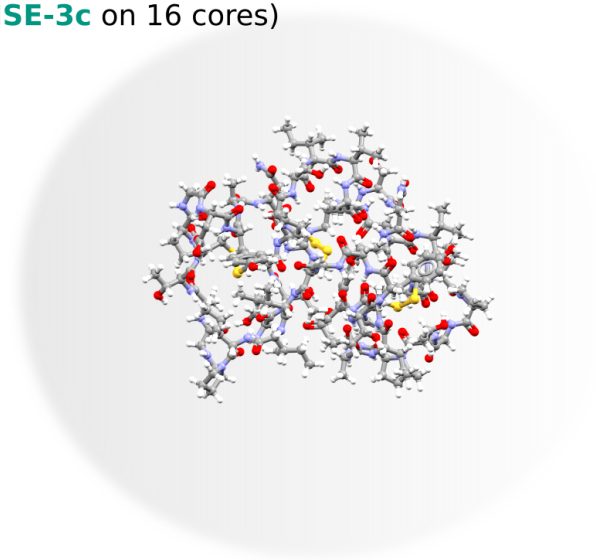
	ROT34 org. mol.	**S66x8** non-covalent	**X23** molecular crystal
TPSS/large basis	1.9	14.6	27.9
TPSS-D3/large basis	1.3	1.3	1.0
HF-3c (very cheap QM)	1.5	-1.2	-5.7
DFTB3-D3 (very cheap QM)	1.2	-2.9	-12.6
HSE-3c (cheap QM)	**0.2**	**1.3**	**0.7**

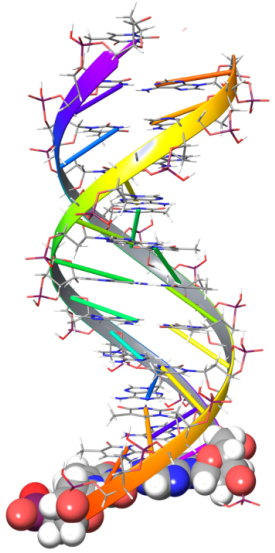
mean of relative deviation in %

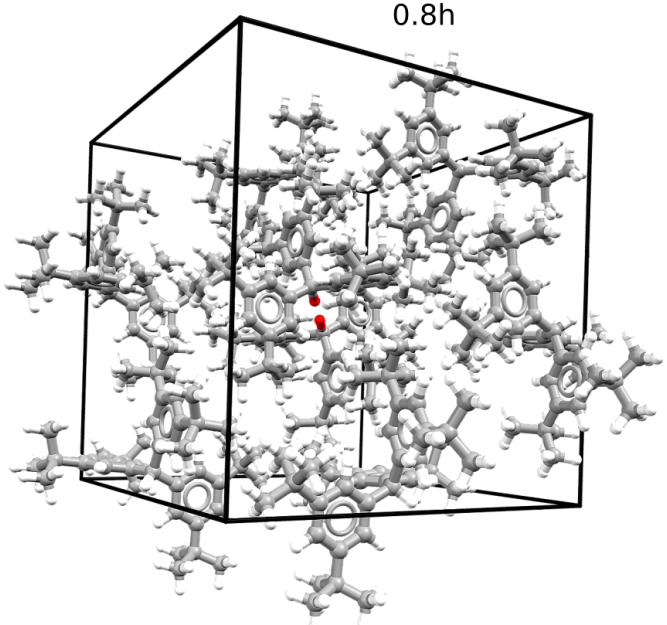
Fast electronic structure with ab-initio accuracy

	gas phase protein	DNA helix	molecular crystal
dim:	0D	1D	3D
#atoms/unit:	647 (647)	726 (66)	856 (37)
wall time:	5h	0.5h	0.8h

(**HSE-3c** on 16 cores)







- fast computer code CRYSTAL17[21] with cost-efficient methods
- enabling routine electronic structure calculation of large systems

[21] R. Dovesi, et al., *Int. J. Quantum Chem.*, **114**, 1287-1317 (2014) , new release in 2017

Outline of talk

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.[22,23]

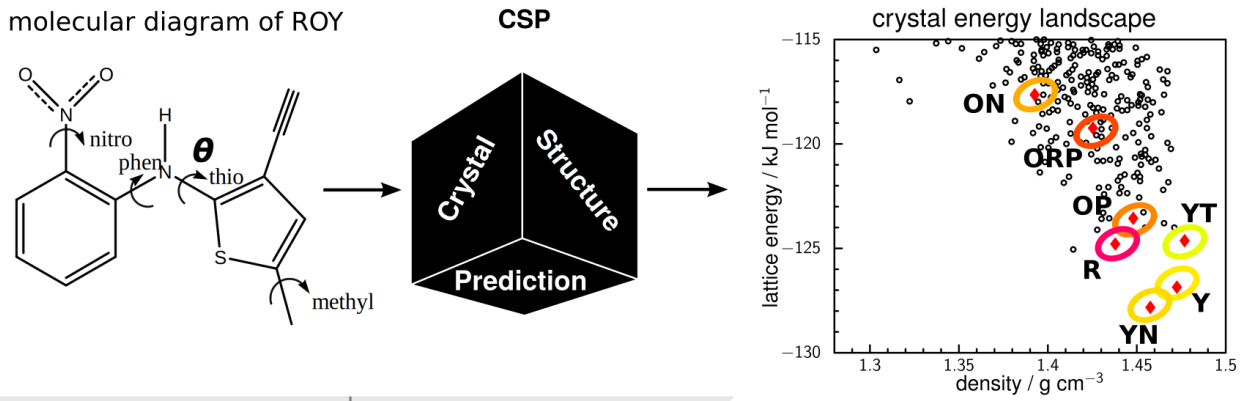
Task

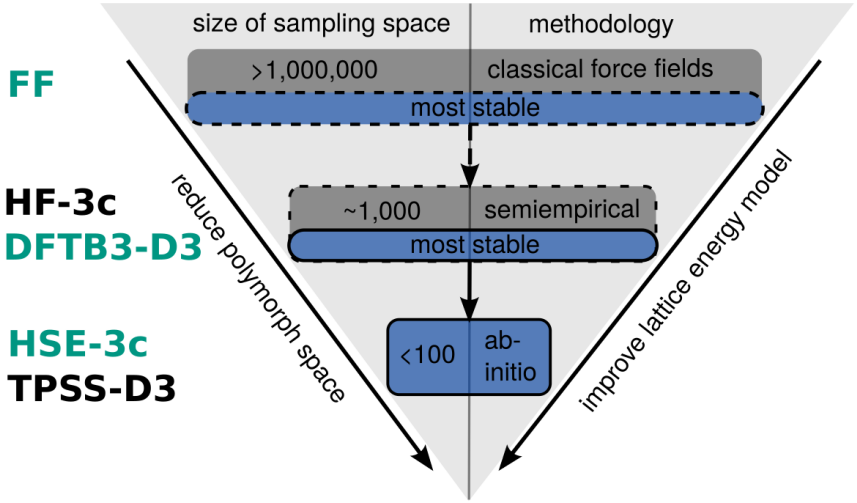
- molecule is chosen due to its chemical/physical/biological properties.
- based on the molecular diagram only, the most stable crystal structures should be predicted.
- predict properties of interest for the most promising candidates.

[22] A. Gavezzotti, *Acc. Chem. Res.* **27**, 309-314 (1994).

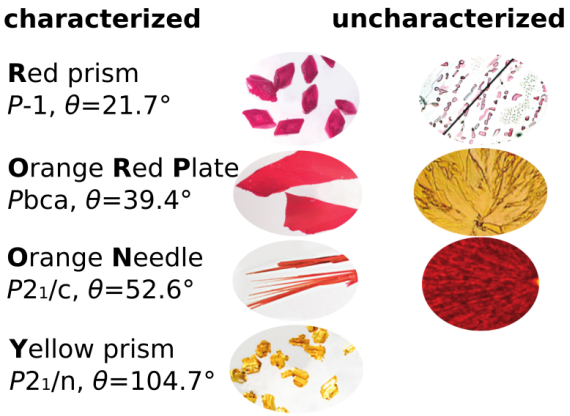
[23] J. Maddox, *Nature* **335**, 201-201 (1988).

Sampling and energetic ranking for crystal structure prediction





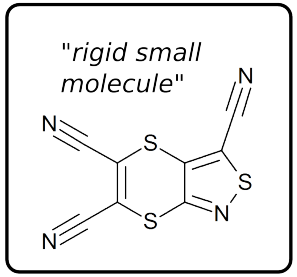
Properties
of ROY polymorphs

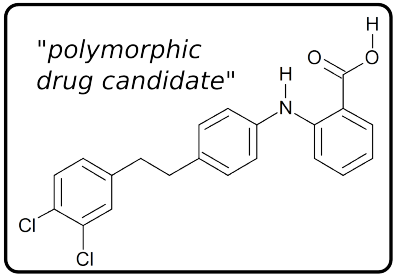


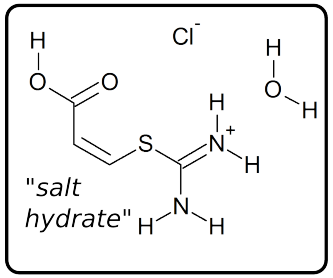
[22] S. Price, *Chem. Soc. Rev.* **43**, 2098 (2014)

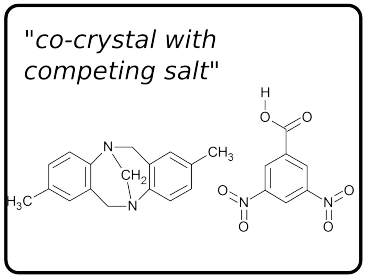
[23] M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, *Acta Cryst. B* **68**, 677 (2012)

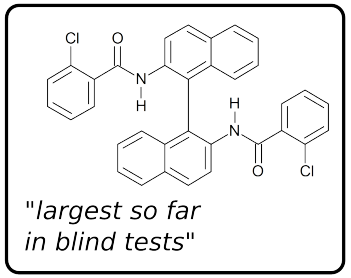
Promising results in the 6th blind test











	22	23	24	25	26
PBE	2	1-9	6	3	1
PBE-D3	**1**	**1-9**	**1**	**1**	**1**
PBE-MBD	1	1-7	1	2	1
vdW-DF2	1	4-8	1	3	2
M06L	1	4-13	1	1	7

lattice energy on fixed TPSS-D3 structures

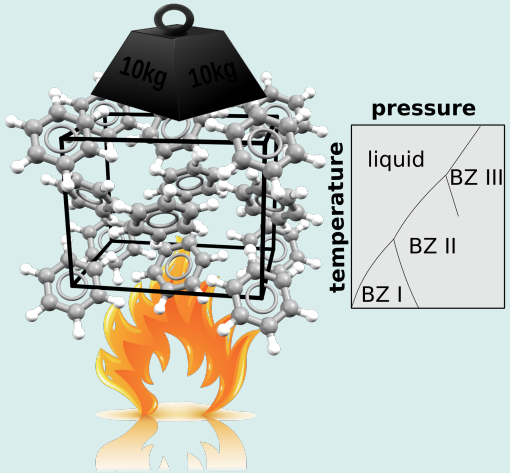
- good lattice energy based ranking of PBE-D3[24,25]
- some structures lost in FF $\longrightarrow$ DFT transition

[24] A. Reilly, et al. *Acta Cryst. B*, **72**, 439 (2016)

[25] JGB, S. Grimme *Acta Cryst. B*, **72**, 502 (2016)

Further improvement needed for reliable crystal structure prediction

1. Influence of thermodynamic



- *P*-*T*-dependent phase diagramm
- polymorph can change with thermodynamic conditions[26]

2. Accurate lattice energy

- improve DFT methods[27]
- many-body methods, like RPA and DMC are promising.[28]

3. Treatment of flexibility

- QM derived intramolecular FF
- machine learning potentials with DFT training set

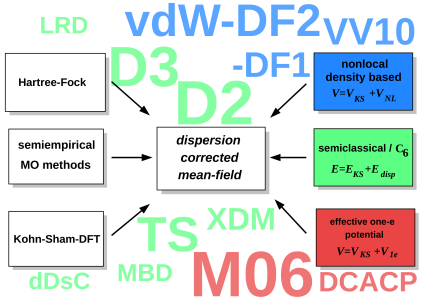
[26] QHA based work in progress [27] triple-ζ based composite method B97-3c under development

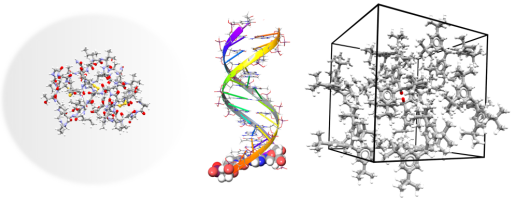
[28] A. Zen, JGB, J. Klimes, D. Alfè, A. Michaelides, *Phys. Rev. Lett.*, submitted (2017)

Summary

Conclusions

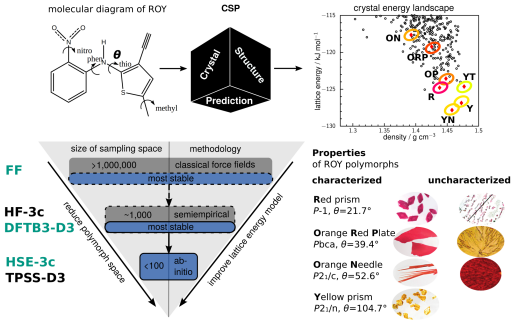
- D3 dispersion correction can be combined with various methods
- cheap QM methods (HSE-3c, DFTB3-D3) useful for fast electronic structures
- promising results of crystal energy rankings in CSP blind test





Outlook and possible improvements

- exploring the merit of DFT-D methods in context of CSP further
  $\longrightarrow$ generated DFT-D references for ML?
- analysis of phonon spectra, free energies



Acknowledgements

Collaborators

- Stefan Grimme (Bonn)
- Sally Price (London)
- Angelos Michaelides (London)
- Felix Fernandez-Alonso (Harwell Oxford)
- Eike Caldeweyher (Bonn)
- Bartolomeo Civalleri (Torino)
- Roberto Orlando† (Torino)
- Anthony Reilly (Cambridge)

Funding

Alexander von Humboldt Stiftung/Foundation

813 dfg·sfb Chemie an Spinzentren SFB 813 der Deutschen Forschungsgemeinschaft

THE THOMAS YOUNG CENTRE TYC

cposs

Acknowledgements

Collaborators

- Stefan Grimme (Bonn)
- Sally Price (London)
- Angelos Michaelides (London)
- Felix Fernandez-Alonso (Harwell Oxford)
- Eike Caldeweyher (Bonn)
- Bartolomeo Civalleri (Torino)
- Roberto Orlando† (Torino)
- Anthony Reilly (Cambridge)

Funding

Alexander von Humboldt Stiftung/Foundation

Chemie an Spinzentren SFB 813 der Deutschen Forschungsgemeinschaft

THE THOMAS YOUNG CENTRE

cposs

Thanks

Key references

- Dispersion corrections:

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

- DFT development

  JGB, E. Caldeweyher, S. Grimme, *Phys. Chem. Chem. Phys.*, **18**, 15519 (2016)

  S. Grimme, JGB, C. Bannwarth, A. Hansen, *J. Chem. Phys.*, **143**, 054107 (2015)

- DFA-DISP for crystal structure prediction:

  JGB, S. Grimme, *Top. Curr. Chem*, **345**, 1 (2014)

  S. L. Price, JGB, *Molecular Crystal Structure Prediction*, Elsevier, Melbourne, Australia, in press (2017).

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