Machine Learning for Material Science

(a biased introduction)

Stefano de Gironcoli
Scuola Internazionale Superiore di Studi Avanzati
Trieste-Italy

Credits:

Emine Kucukbenli, Harvard (MA, USA)

Ruggero Lot, Univ Toulouse (F)

Franco Pellegrini, ENS Paris (F)

Yusuf Shaidu, Berkeley (CA, USA)

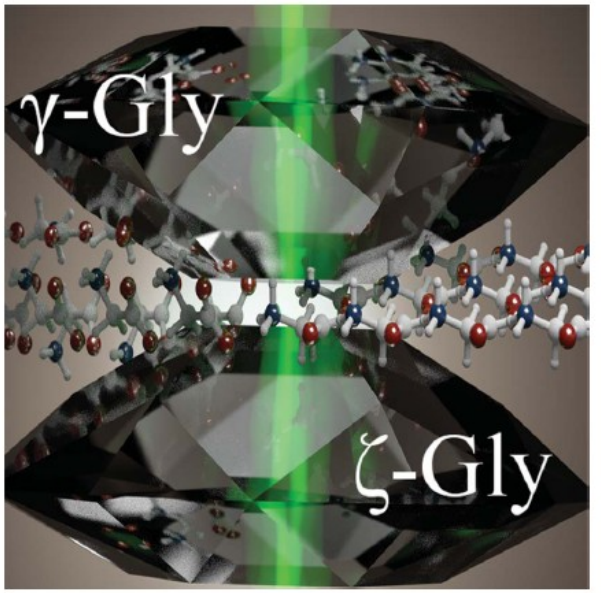
A few projects with Quantum ESPRESSO
a personal selection

Stefano de Gironcoli

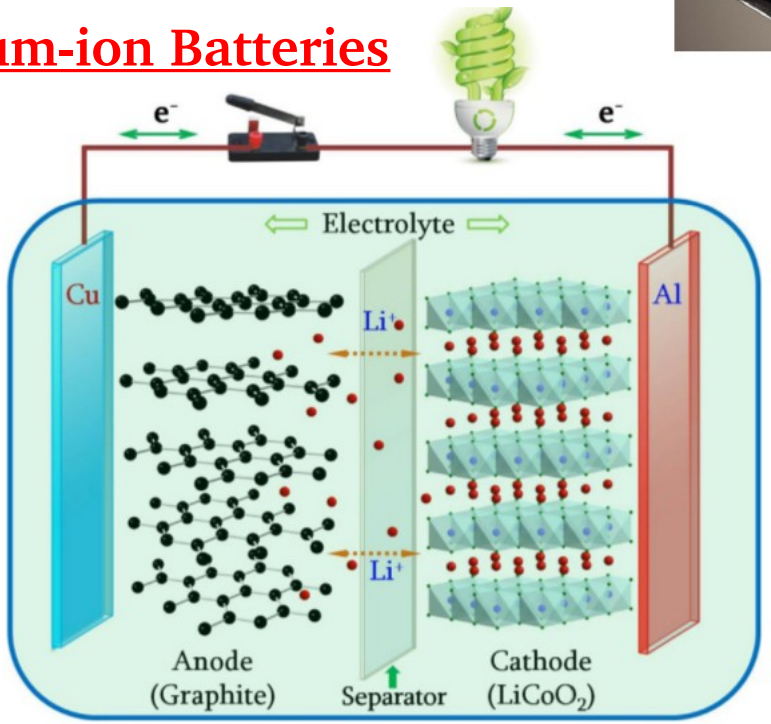
SISSA and CNR-IOM DEMOCRITOS

Complete $_{13}C$ Chemical Shift Assignment for Cholesterol Crystal

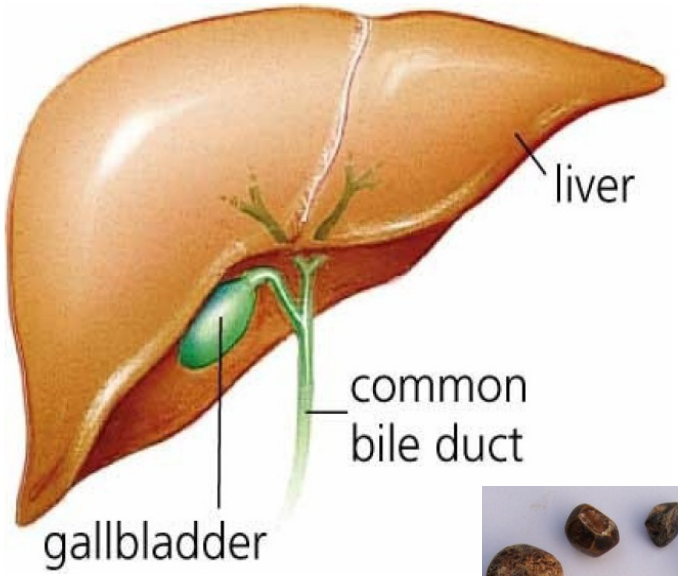
ζ-Glycine: Insight into the mechanism of a polymorphic phase transition

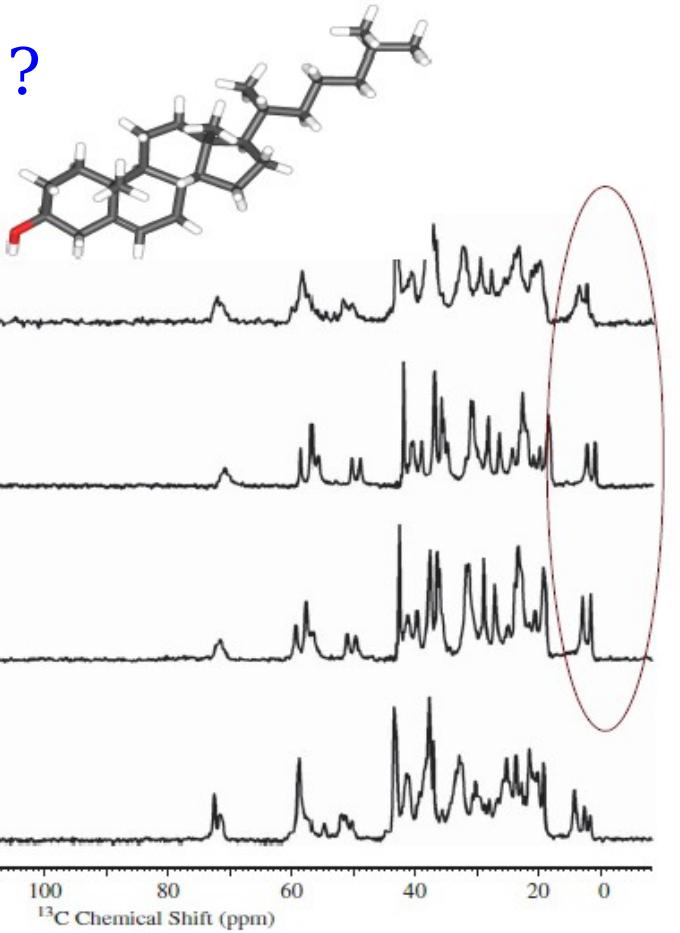


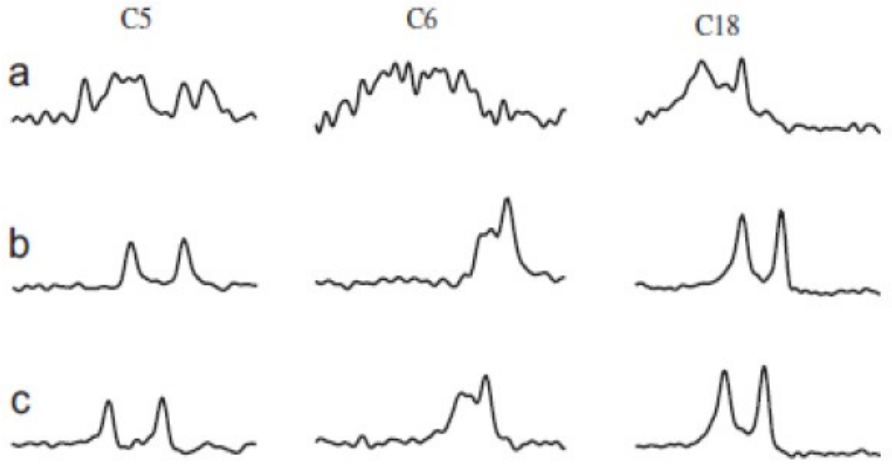
Lithium-ion Batteries



Why Cholesterol ?







a) Gallbladder Cancer

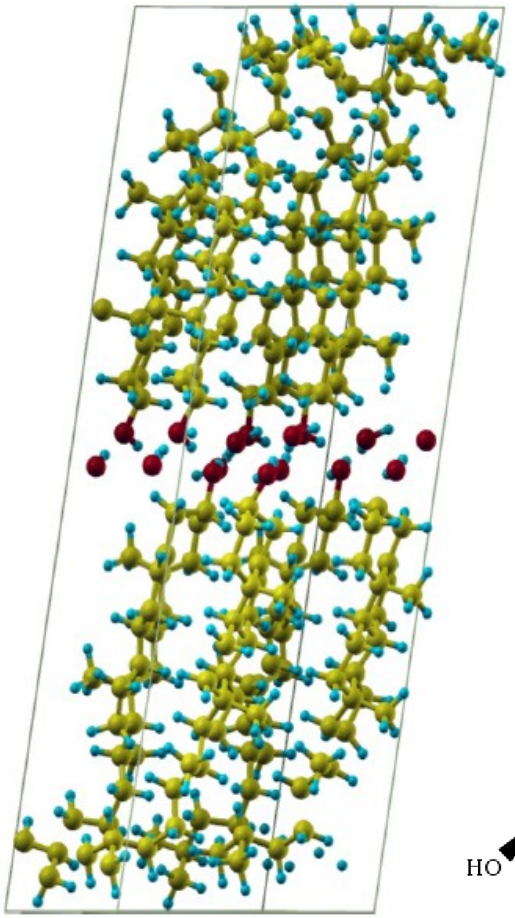
b) Chronic Cholecystitis

c) Xantho-Granulomatous Cholecystitis

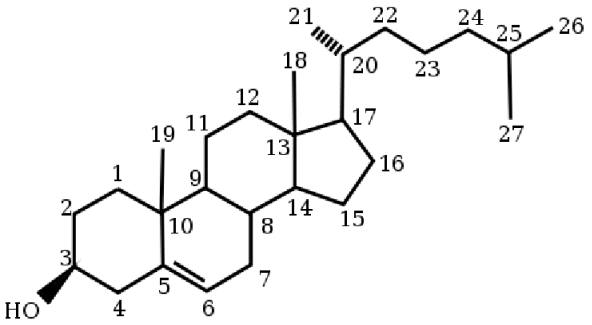
Gall stones associated to different pathologies have distinct NMR spectra.

It would be important to understand the structural differences between these stones and the underlying reasons.

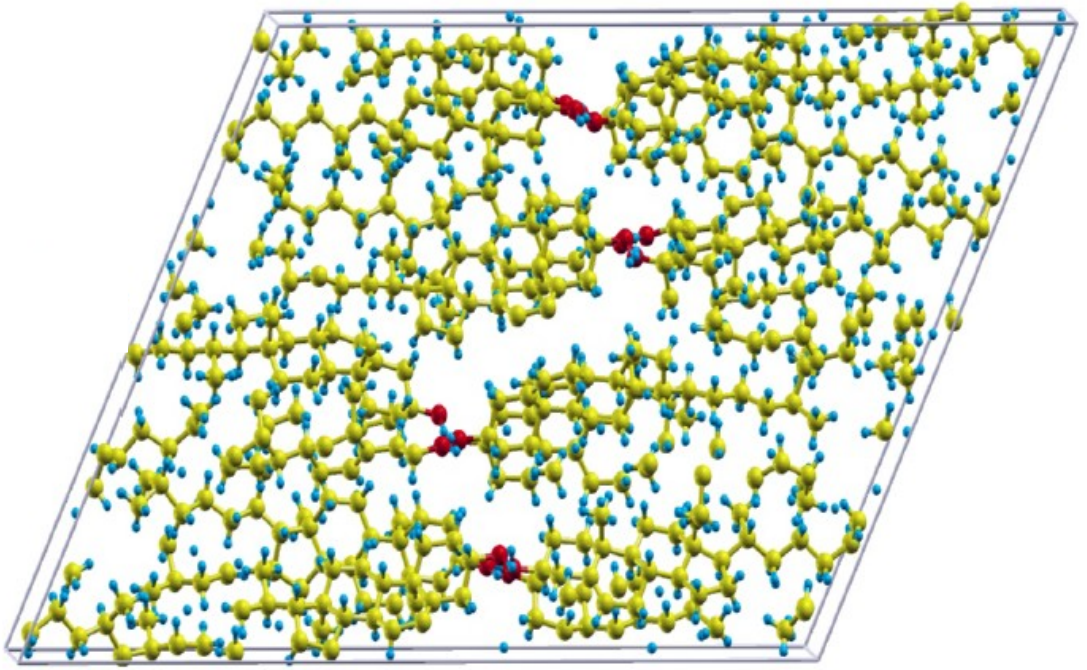
Can we distinguish the NMR spectra of different Cholesterol crystalline polymorphs ?



Monohydrade Cholesterol (ChM)
8 CLR +8 w molecules - 616 atoms



High temperature Anhydrous Cholesterol (ChAh)
16 CLR mol – 1184 atoms

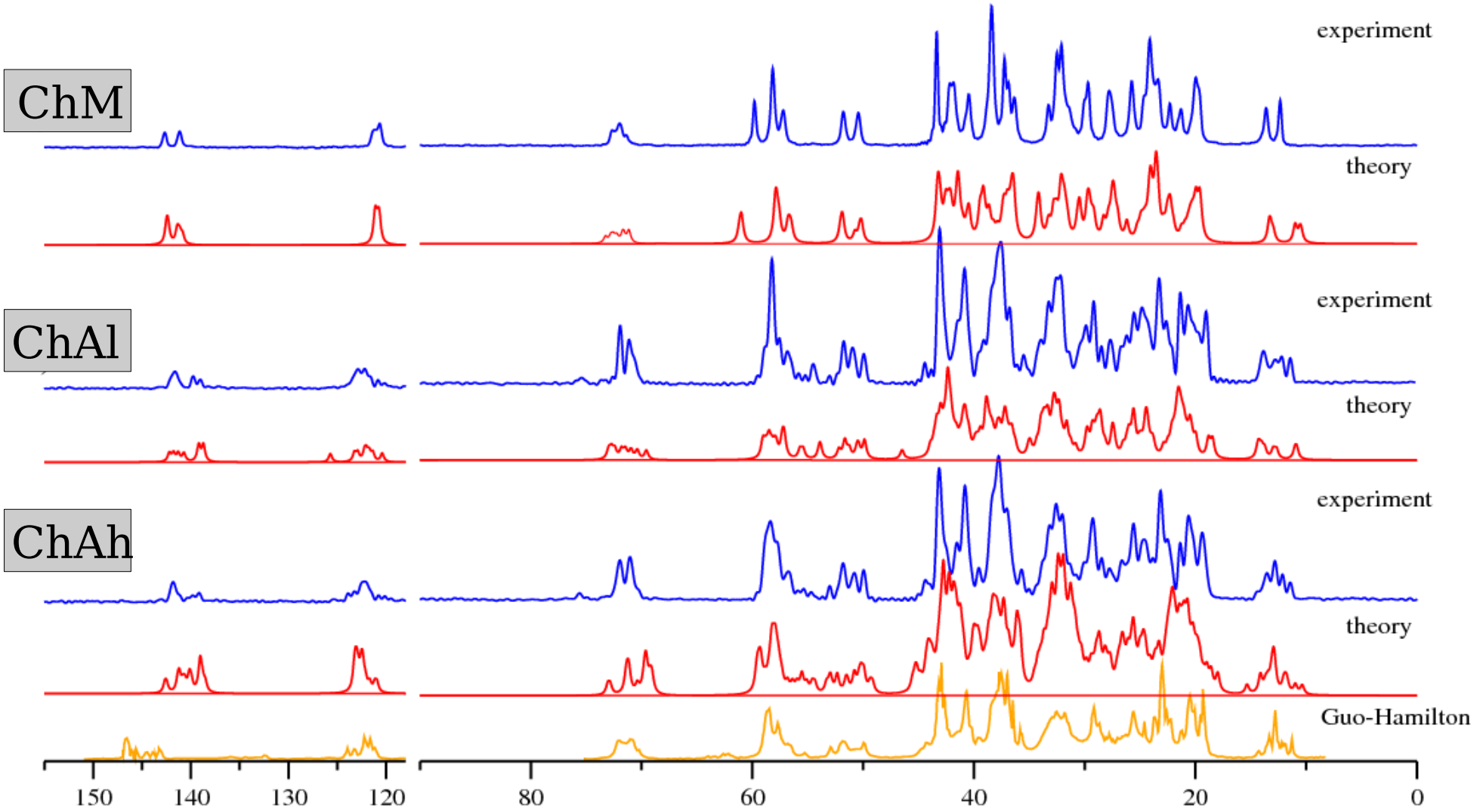


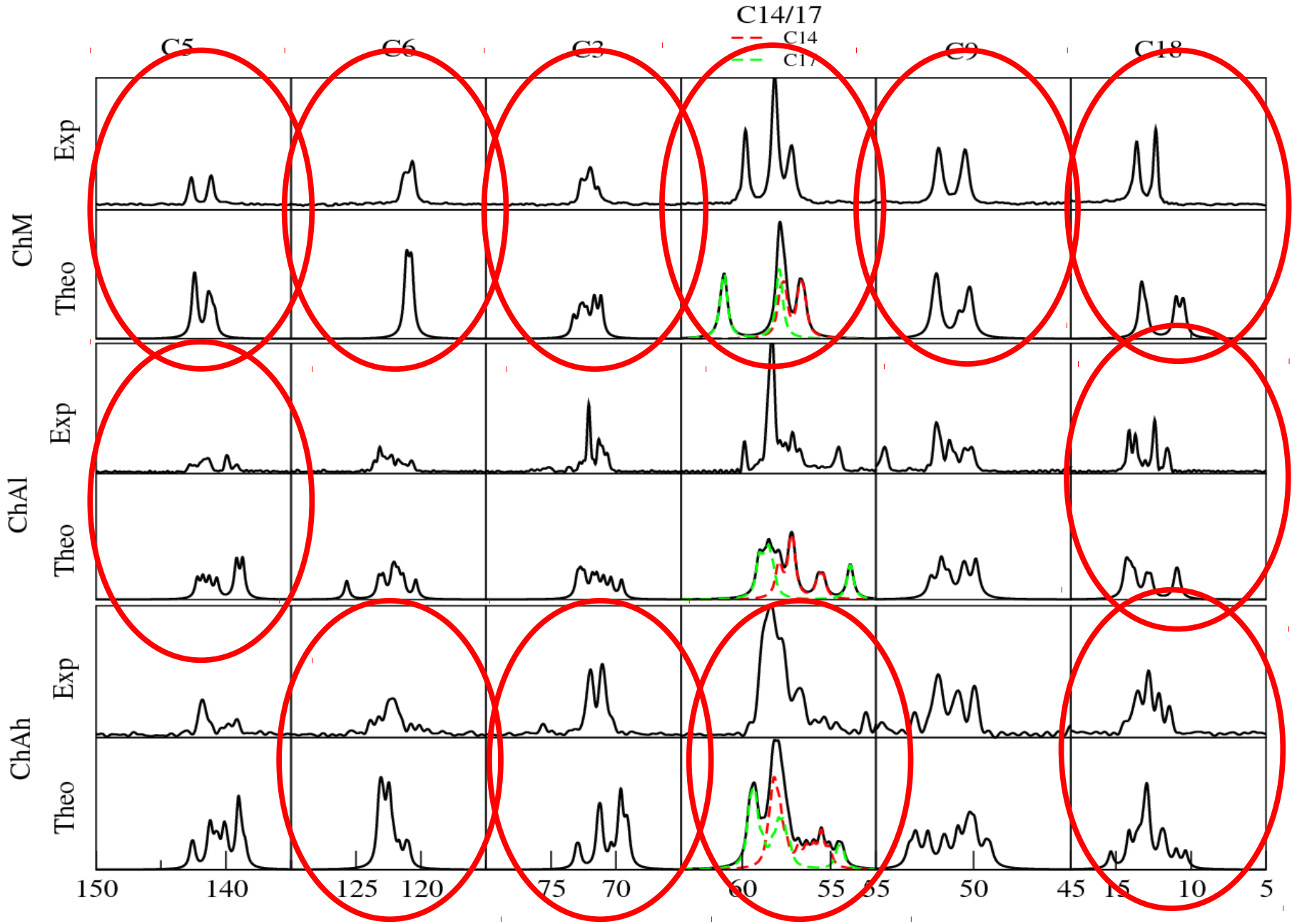
Low temperature Anhydrous Cholesterol (ChAl)
8 CLR mol – 592 atoms

(not shown)

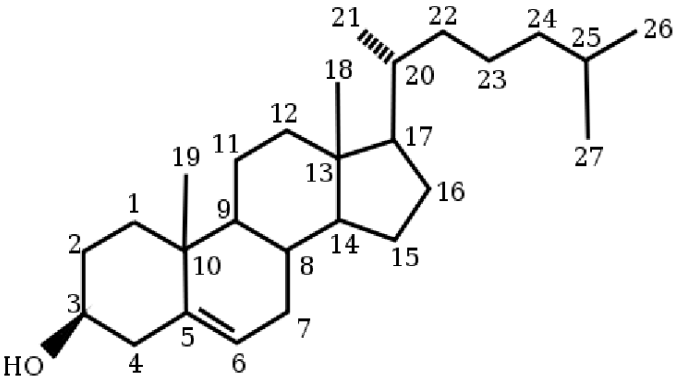
Comparison with Experimental Spectra

(E.Kucukbenli, K.Sonkar, N. Shina, SdG, JPCA 2012)




C14/17
C14
C17
C5
C6
C3
C9
C18
ChM
ChAl
ChAh
Exp
Theo
150
140
125
120
75
70
60
55
55
50
45
15
10
5

Peak assignement in ChM



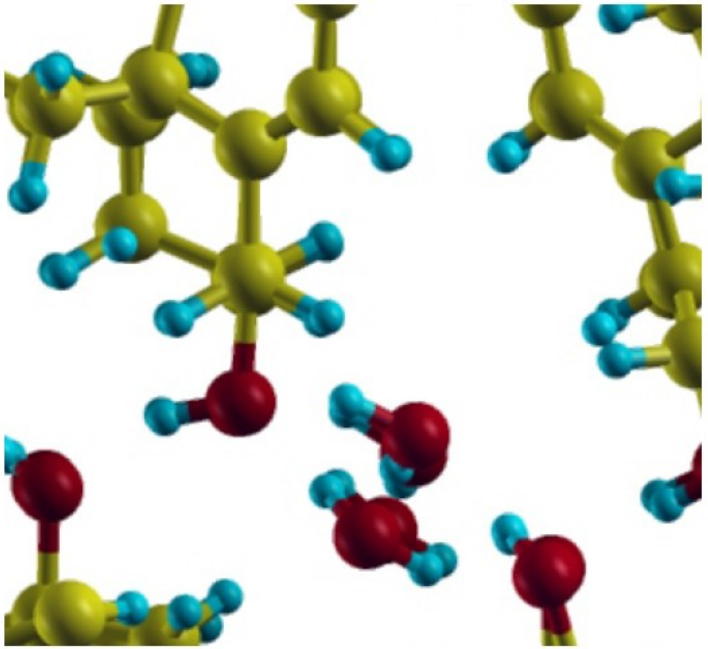
a) ChM-Exp

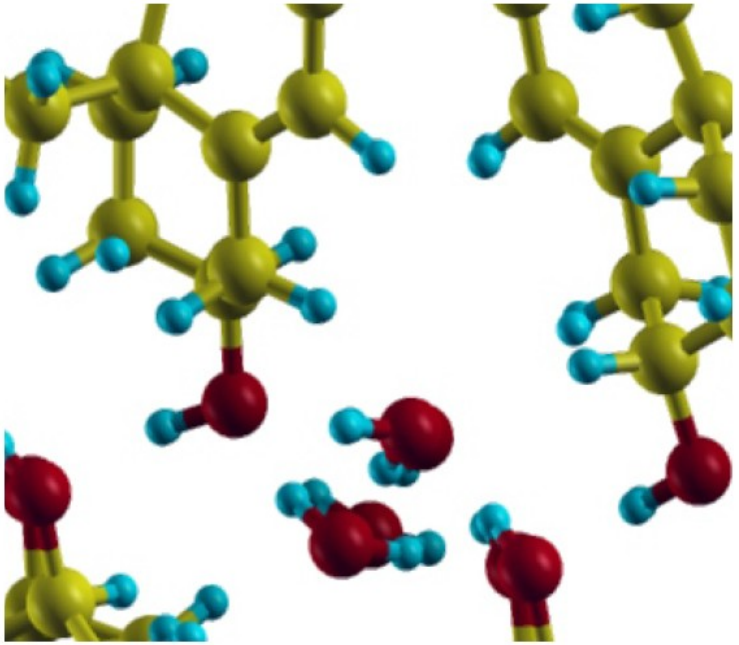
C17
C14
C17
C14
C9
C13
C12
C4
C20
C1
C10
C20
C7
C8
C15
C11
C16
C2
C27
C21
C19
C18
C22
C24
C25
C23
C26

b) ChM-Theory

60
50
40
30
20
10

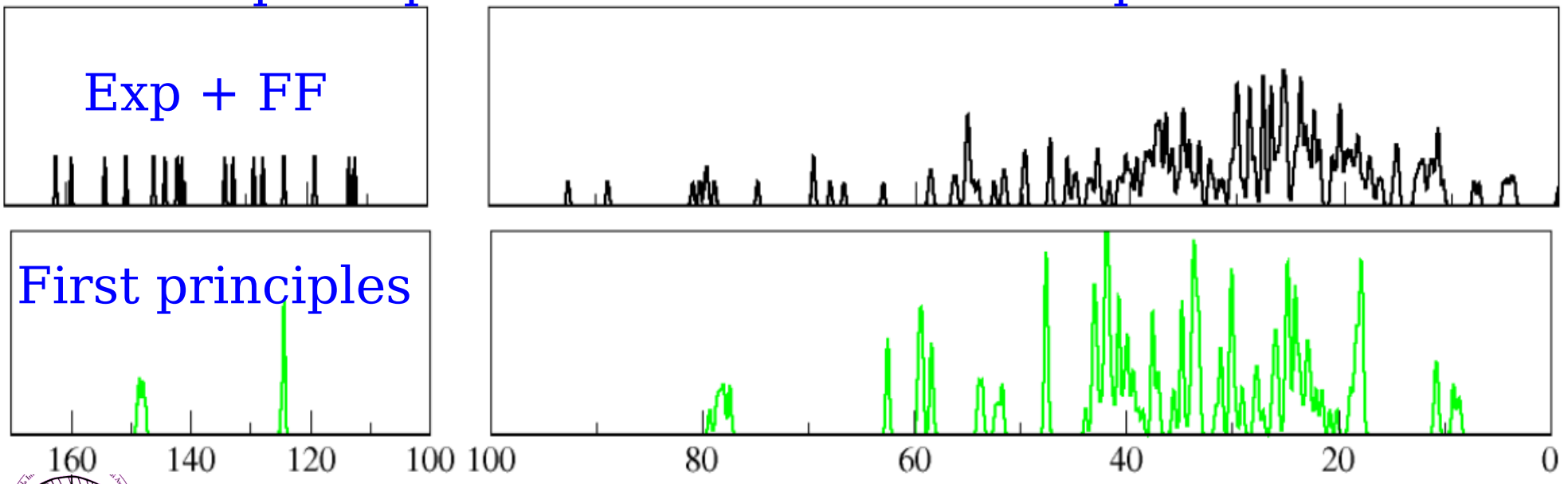
Importance of accurate structural relaxation (ChM)





First principles

Exp + FF



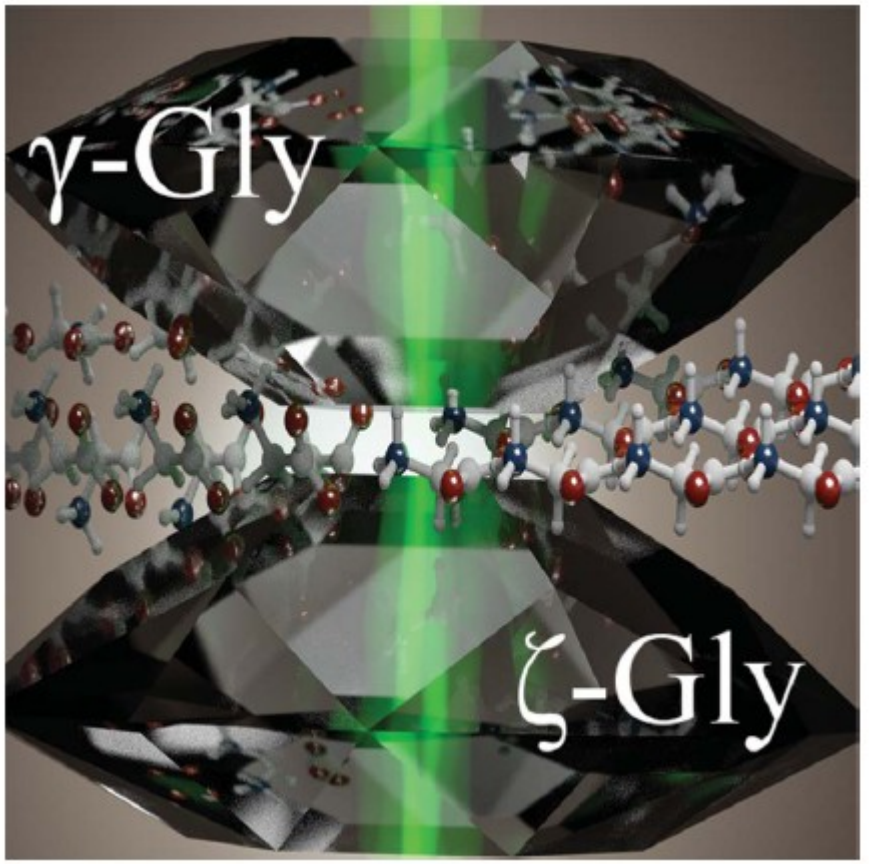
Summary

- Theory:
  GIPAW +PAW fully implemented

- ^{13}C NMR of Cholesterol Crystals
  Calculations reproduce the main features of the spectra and are able to distinguish different polymorphs
  It is possible to provide a complete peak assignment, confirmed with spectral editing exp.

- Search for unknown polymorphs would require a very accurate energy description at a fraction of the DFT cost

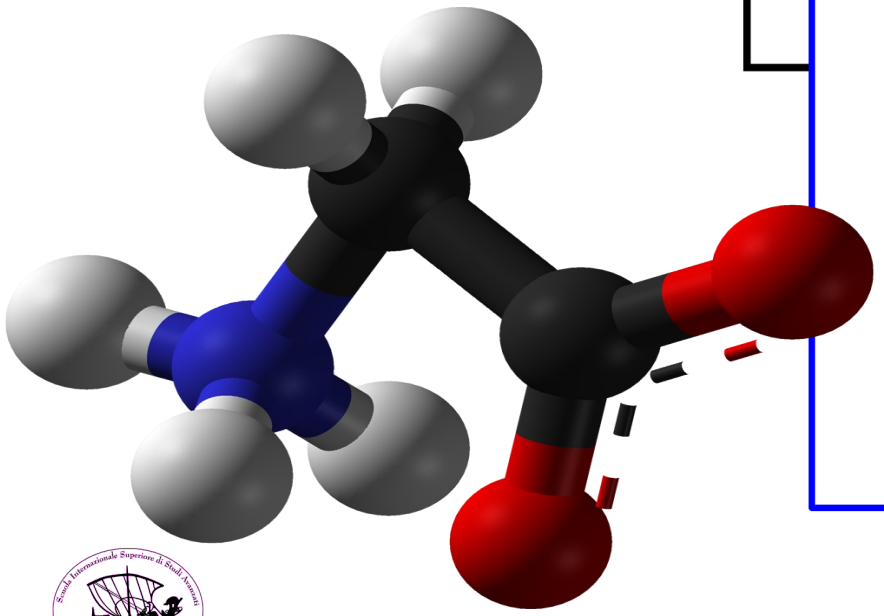
ζ-Glycine: Insight into the mechanism of a polymorphic phase transition



CSP is a formidable task

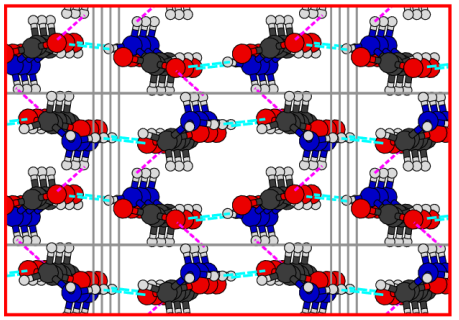
- CSP problem: Name a chemical or stoichiometric formula; find the (local) minima of the free energy landscape under given thermodynamic conditions (often at certain T,P)
- “What is the most stable structure of glycine at ambient conditions?” “What is the carbon structure that is stable at very high pressures”
- Challenges:
  - A very vast space of possibilities.
  - Free energy landscape is very expensive to obtain accurately

Glycine



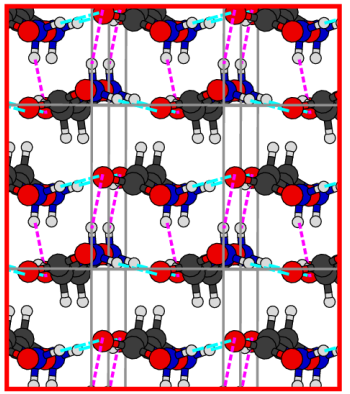
Pressure

δ(Z=4)



No phase transition

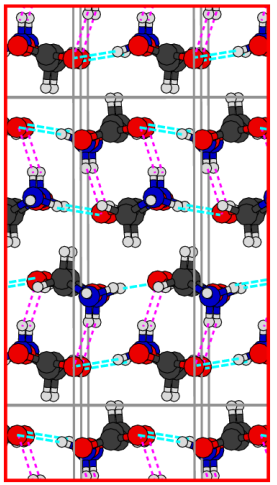
ε(Z=2)



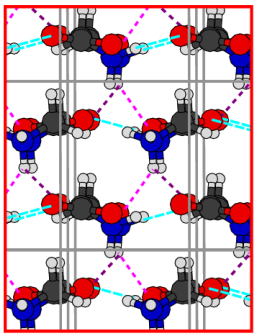
Unknown

ζ

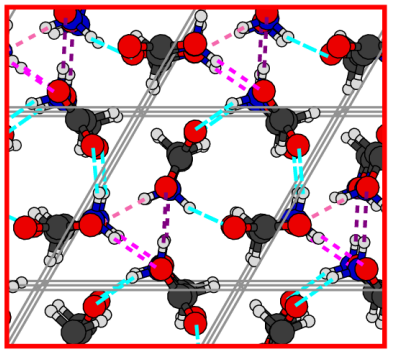
α(Z=4)



β(Z=2)



γ(Z=3)



Stability

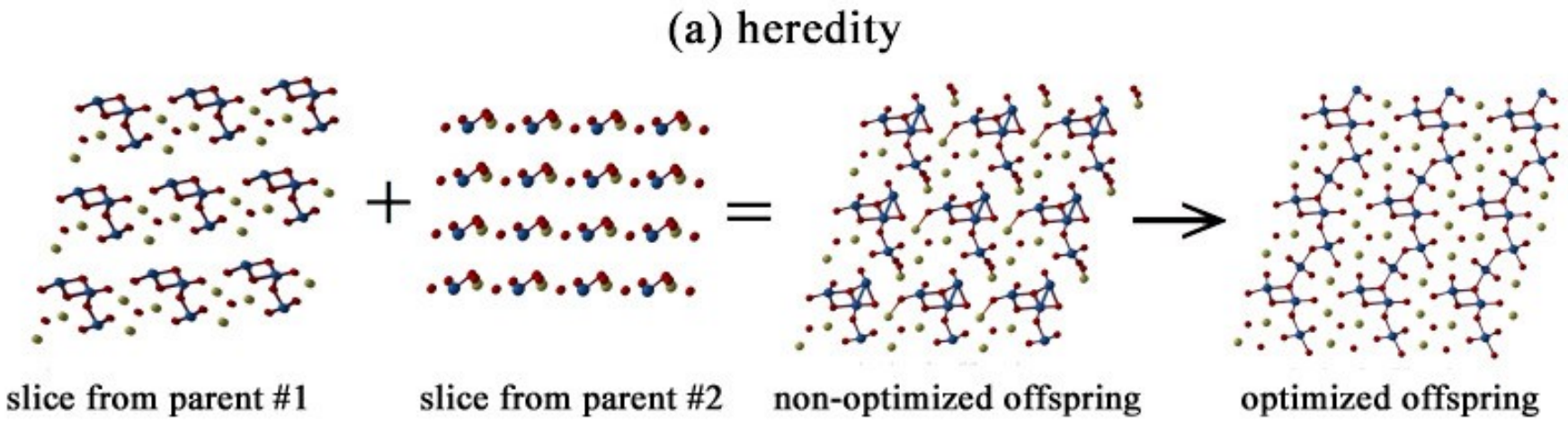
How to tackle CSP?

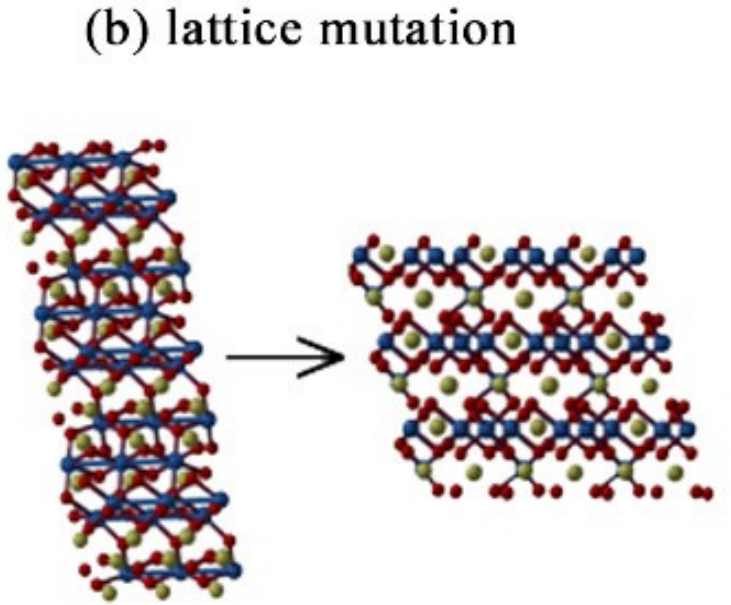
Explore: Use smart algorithms to explore as much of the landscape as possible

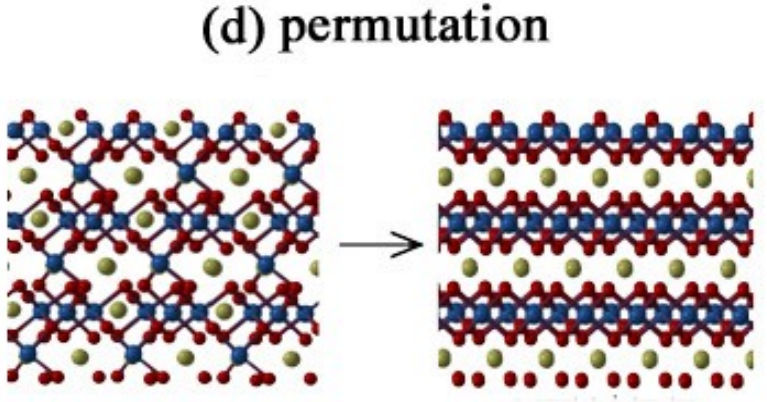
Molecular dynamics / Monte Carlo walkers

- Simulated annealing
- Metadynamics
- Basin hopping
- Minima hopping
- Genetic algorithm

Genetic algorithm



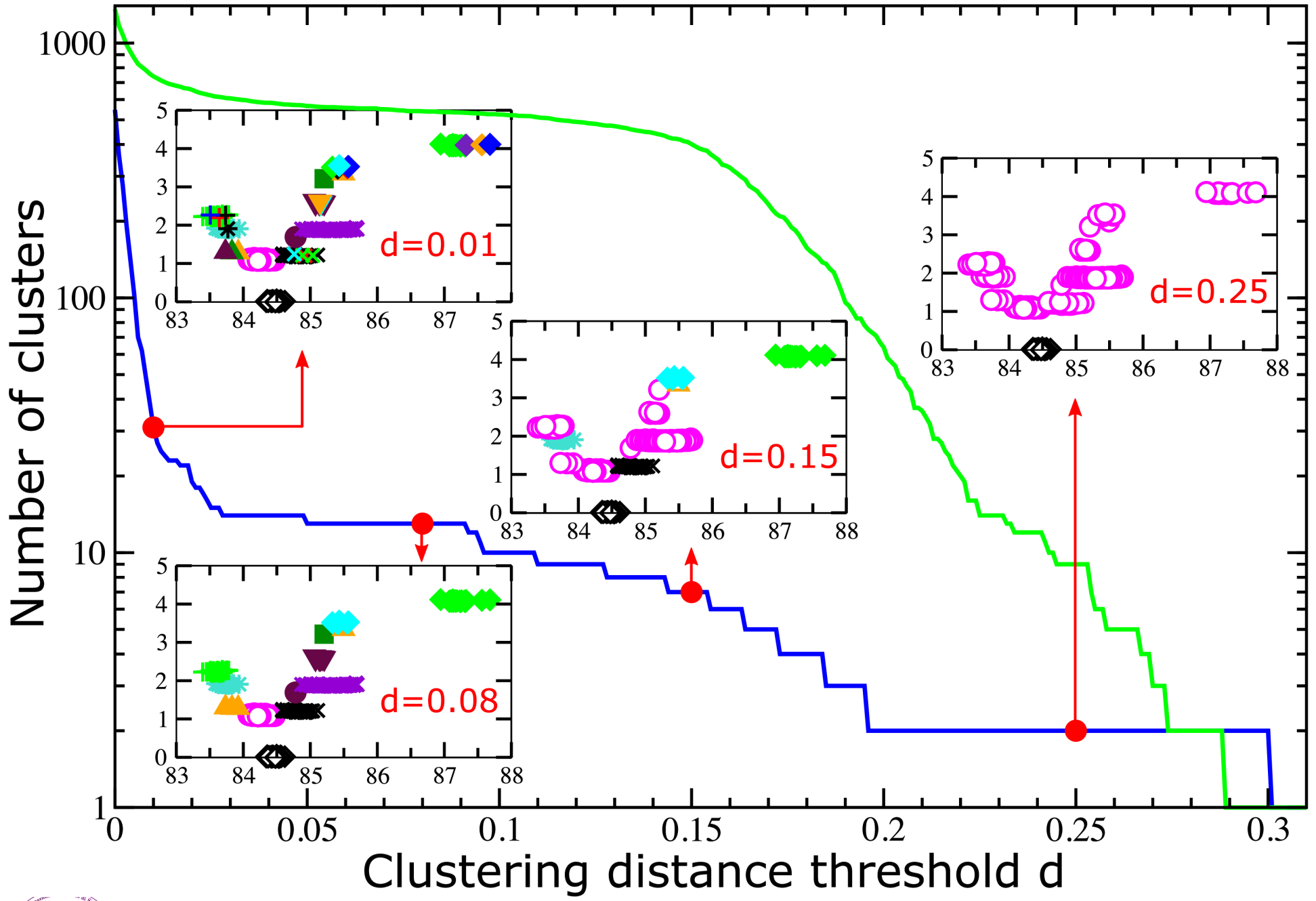


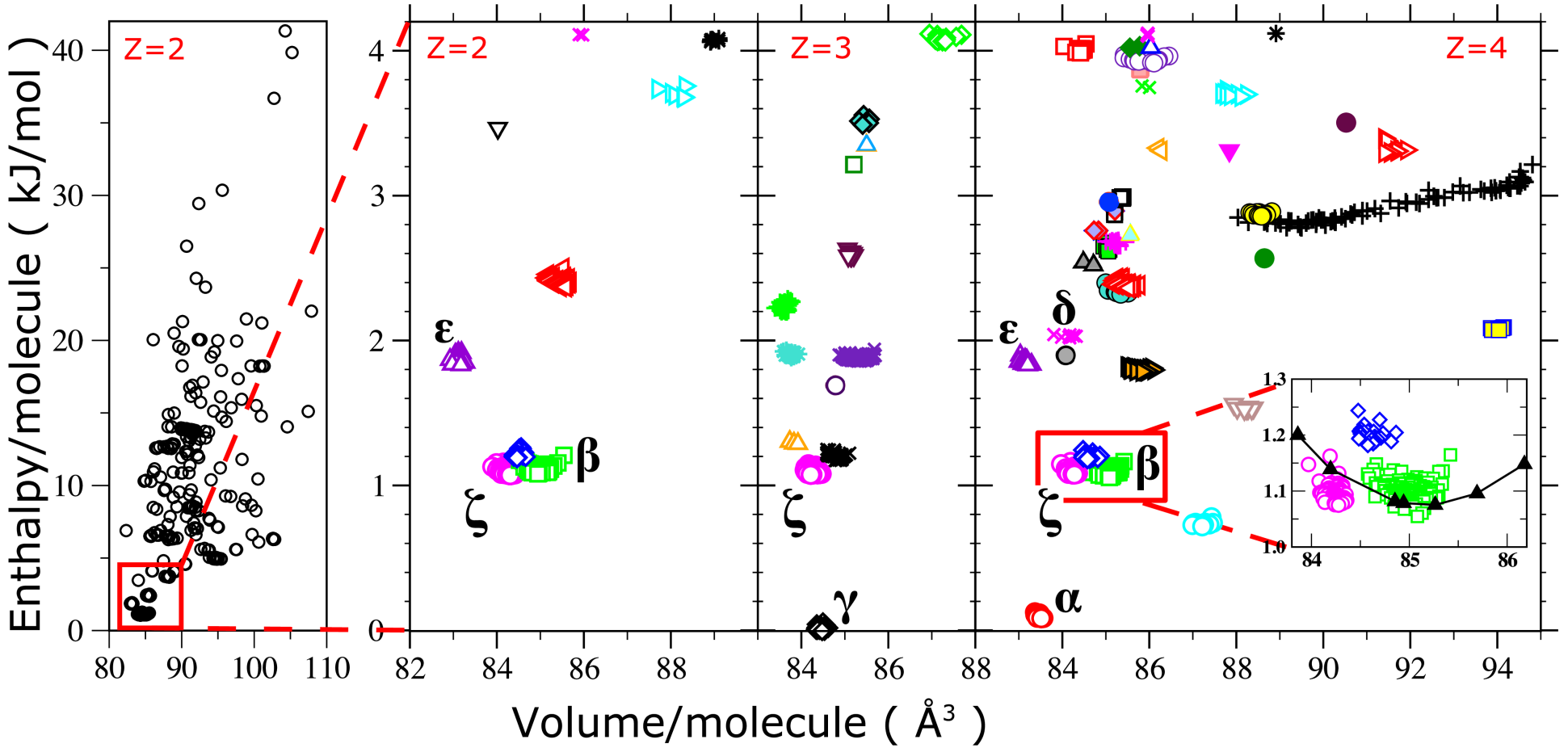


USPEX operations

+

+ vdWDF + clustering


1000
100
10
1
Number of clusters
0
0.05
0.1
0.15
0.2
0.25
0.3
Clustering distance threshold d
d=0.01
d=0.08
d=0.15
d=0.25


Z=2
Z=2
Z=3
Z=4
Enthalpy/molecule (kJ/mol)
Volume/molecule (Å³)
α
β
γ
δ
ε
ζ

ζ–phase

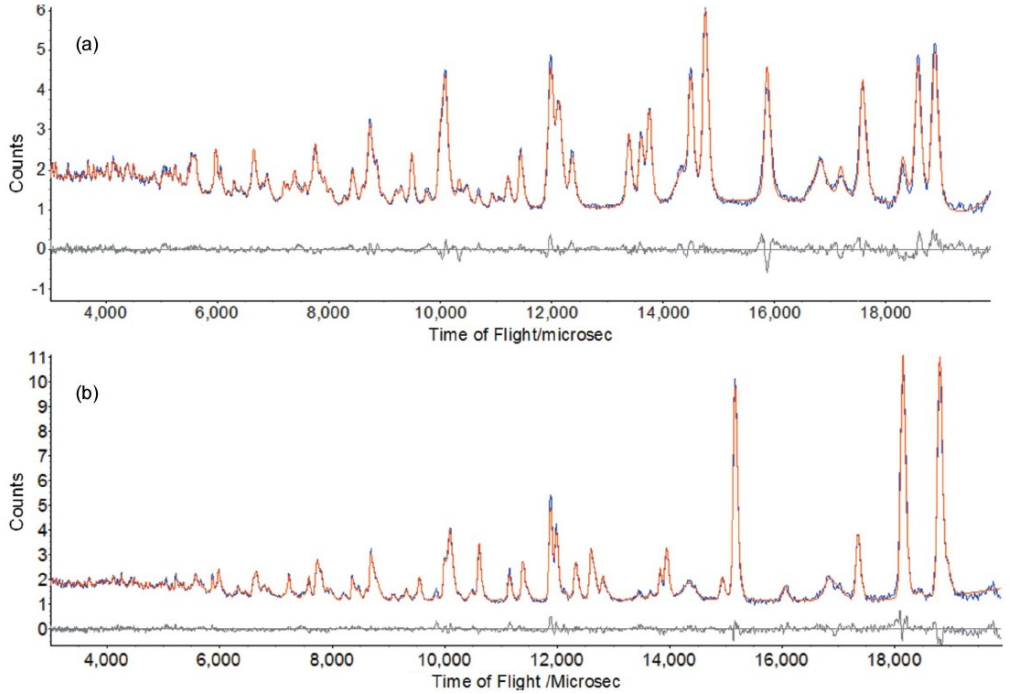


Figure 2
(*a*) Rietveld fit of the neutron powder diffraction pattern of ζ-glycine at 100 K (blue = observed, red = calculated). In addition to the peaks ζ-glycine, the pattern also shows the presence of residual ε- and a trace of γ-glycine. Other peaks arise from the sample environment, namely th pressure marker and the Al_2O_3 and ZrO_2 components of the anvils of the pressure cell. (*b*) Rietveld fit of the neutron powder diffraction patte β-glycine (contaminated with ζ- and a trace of γ-glycine) at 290 K. A 1 Å *d* spacing approximates to 4837 μs in time-of-flight.

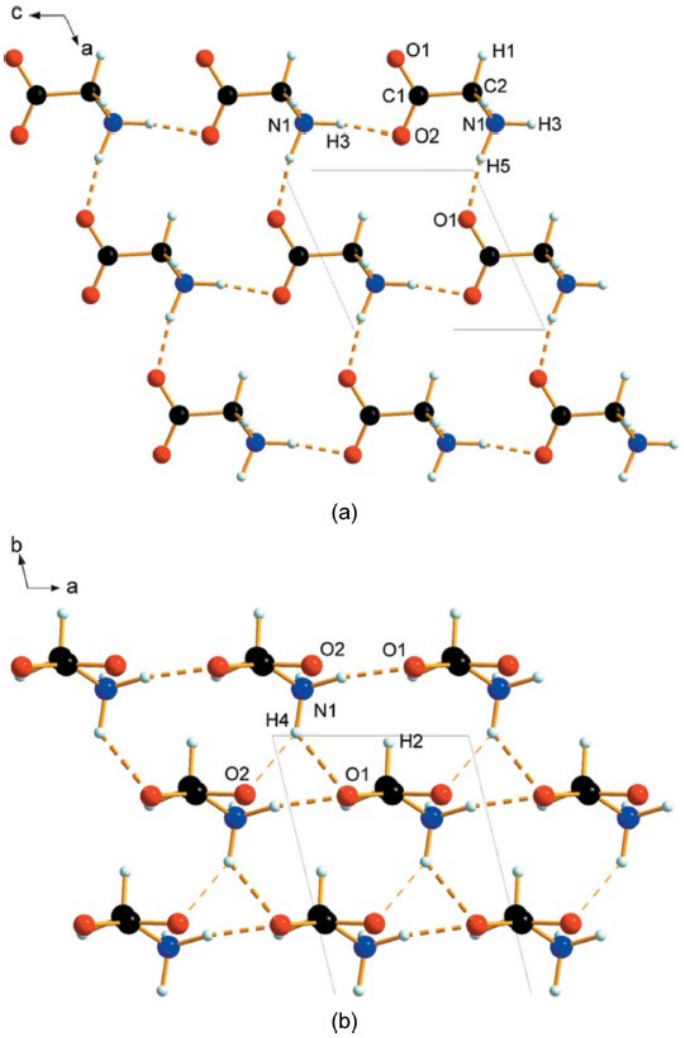


Figure 3
Intermolecular interactions in ζ-glycine. (*a*) Layers formed in the *ac* plane, viewed along **b**. (*b*) Stacking of the layers, viewed along **c**.

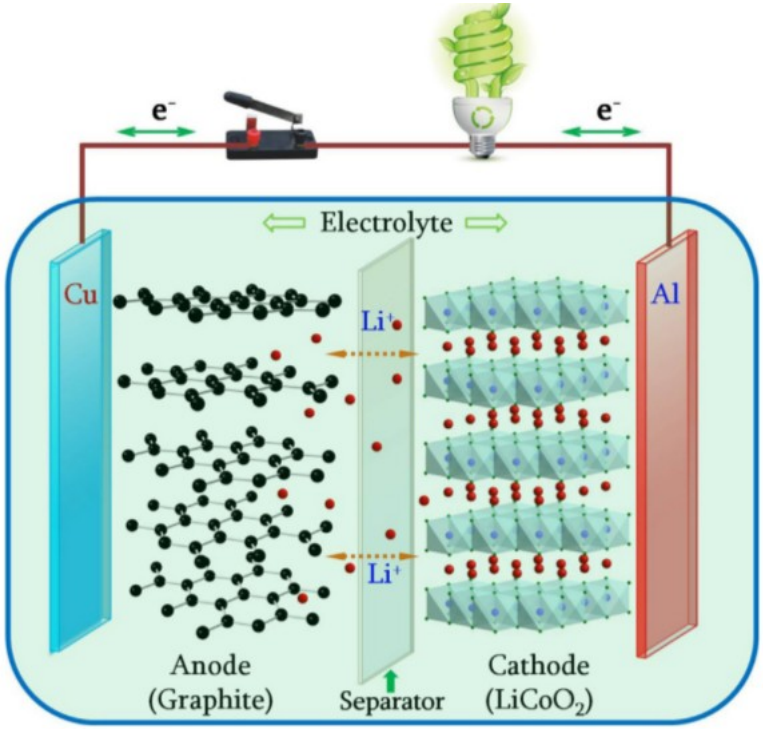
E Kucukbenli, CH Pham, SdG,
C Bull, G Flowitt-Hill, HY Playford, M Tucker, S Parsons
Int Union Crist J **4**, 569–574 (2017)

Exploring the phase space for larger molecules (ex. CLR) requires fast and accurate energetics

Lithium Interaction with Graphene Materials at Finite Temperature

~10Wh ~100Wh ~10,000Wh

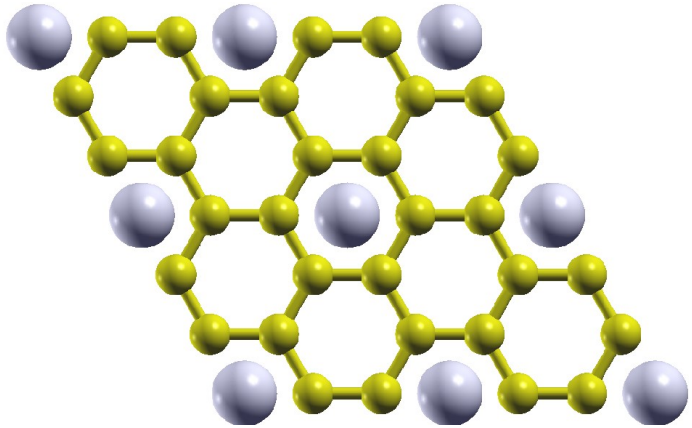
Lithium ion battery



- Cathode: Source of lithium
- Electrolyte: Ionic conductivity
- Anode: Lithium holder
- Current collectors

Capacity: The amount of Li absorbed by anode

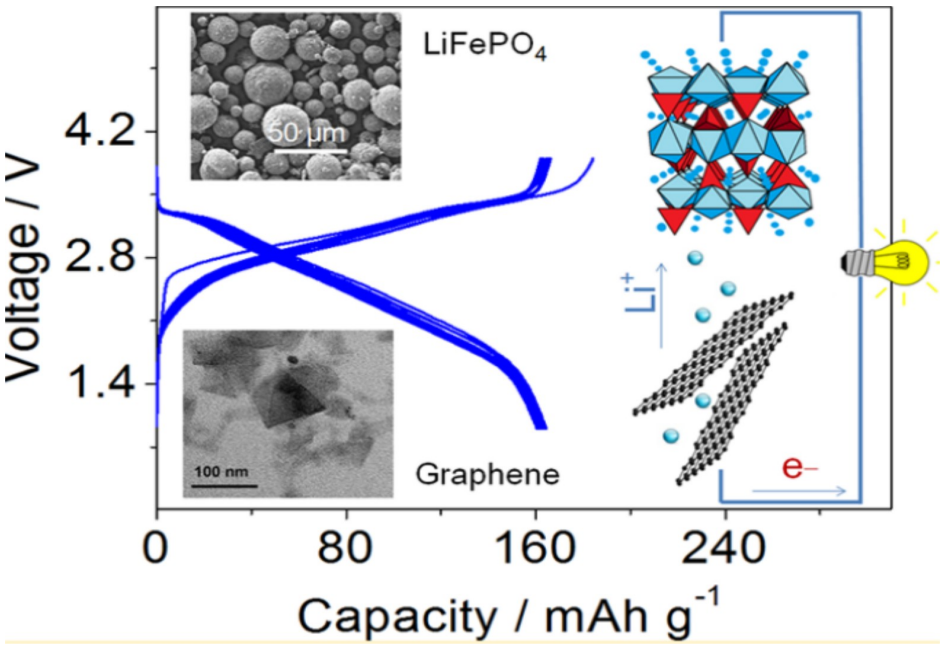
- Stoichiometry of Li adsorbed graphite is LiC_6



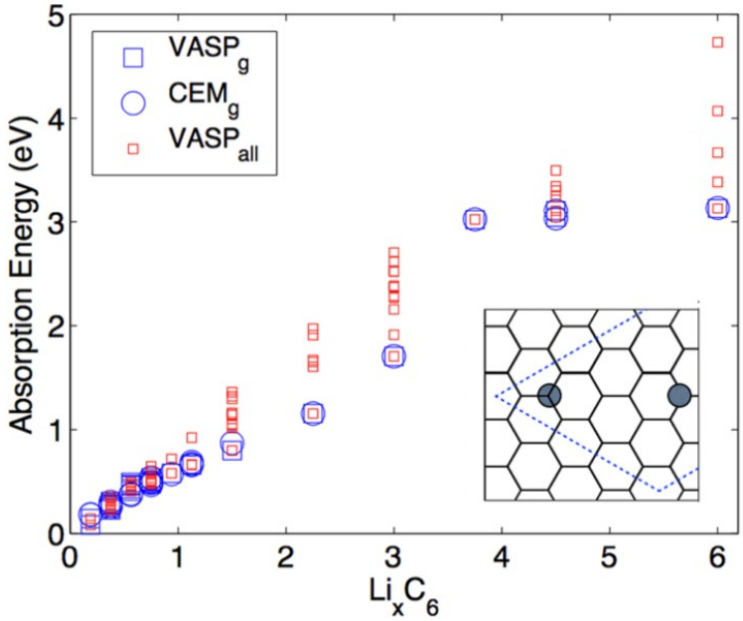
Alternative anode materials :
Graphene due to its large surface to mass ratio and good electrical conductivity.

Materials Today 19(2):109-123, 2016

Experiment



DFT



- Exploited graphene nanoflakes as alternative anode
- Flakes were ~30-100 nm lateral dimension
- **Very high Li uptake: LiC_2**

High Li-concentration is unstable with respect to Li-metal.

GOAL: Understand the nature of high Li uptake by graphene materials.

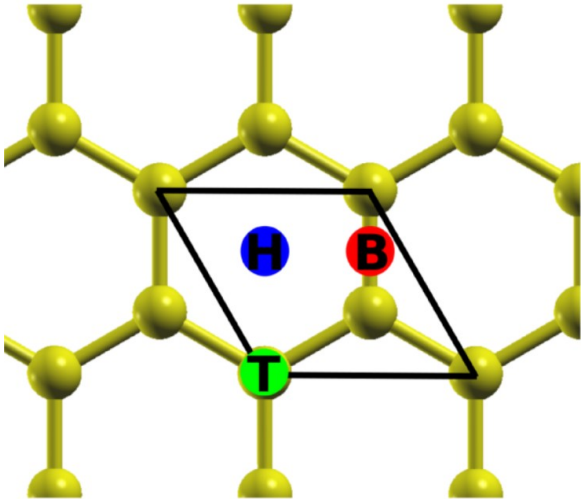
Hassoun et al. Nano Lett. 2014, 14, 4901-4906

Eunseok Lee and Kristin A. Persson, Nano Lett. 2012, 12, 4624−46

Li on graphene at 0K

- Density functional theory (DFT)
- rVV10 functional

Adsorption sites



Sabatini et al ,Phys. Rev. B 87 041108 , 2013

Adsorption energy

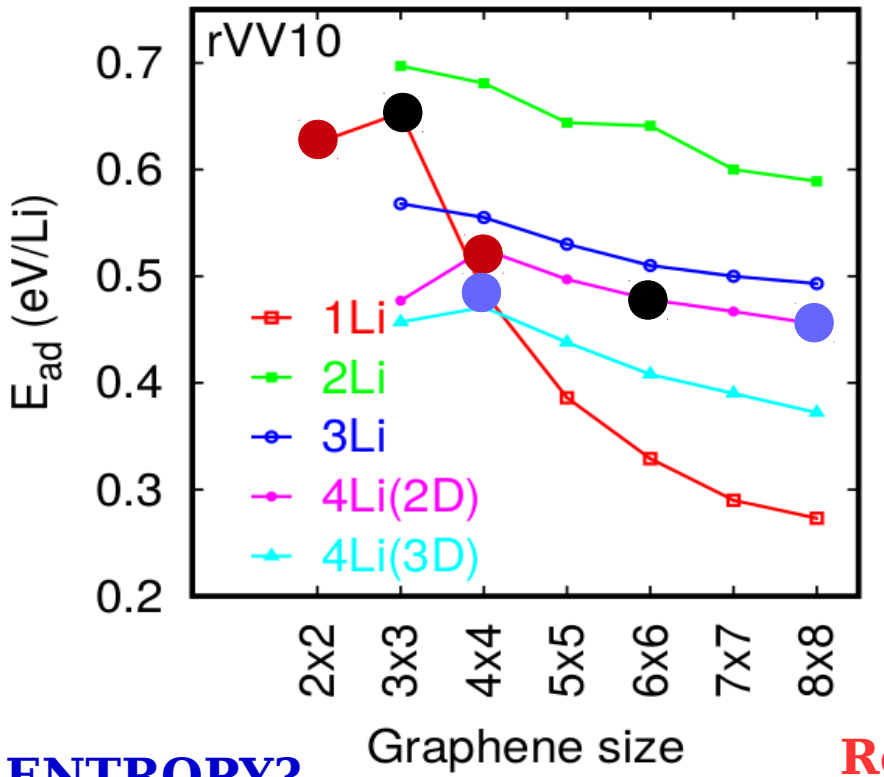
$$E_{ad}(n) = \frac{E(n) - E(n=0) - nE_{Libulk}}{n}$$

- Adsorption energy gives the stability of Li adsorbed graphene against phase separation into graphene and bulk Li.

System	E_H^{ad}(eV)	E_B^{ad}(eV)	E_T^{ad}(eV)
Li@Gr(2x2)	0.625	0.759	0.767

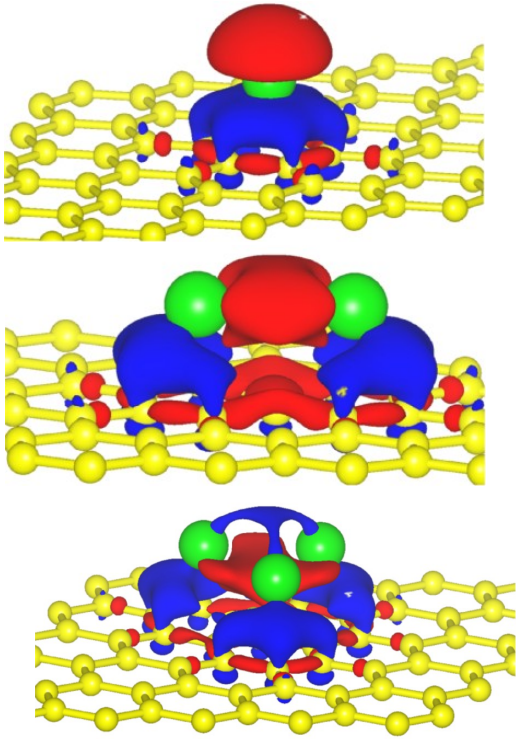
Results: Li on graphene at 0K

Energetics



ENTROPY?

Mechanisms



Li-adatom

1 Li

Li-clusters

2 Li

3 Li

Red (**Blue**) positive (negative) charge density change upon adsorption of Li on graphene.

$$\Delta\rho = \rho(gr + nLi) - \rho(gr) - \rho(nLi)$$

- **Adsorption energies are positive.**
- **Mechanism:** Mainly electrostatics
- **Charge distribution and energetics :** Li-adatom vs Li-clusters

Finite temperature

Grand Canonical Monte Carlo method (GCMC)

Compute thermodynamic properties at fixed μ and T

$$\langle A \rangle = \frac{\sum_N \sum_{\{c_i\}} exp(-\beta(U(\{c_i\})-\mu N)A(\{c_i\})}{\sum_N \sum_{\{c_i\}} exp(-\beta(U(\{c_i\})-\mu N))}$$

configuration space is sampled:

- Select site at random
- Add particle if empty or remove if occupied
- Accept or reject the move according to

$$t_{acc}(\text{old}_i \rightarrow \text{new}_j) = min(1, exp(-\beta(\Delta U - \mu \Delta N)))$$

System:

- 20x20 graphene supercell with 400 sites.
- Used periodic boundary conditions

We need an AFFORDABLE interatomic potential

Cluster expansion

Model:

- Map Li adsorption into a lattice model
- Assign Li occupation number to each site,

$c_i = 0$ if site i is empty

$c_i = 1$ if site i is occupied

- Configuration:

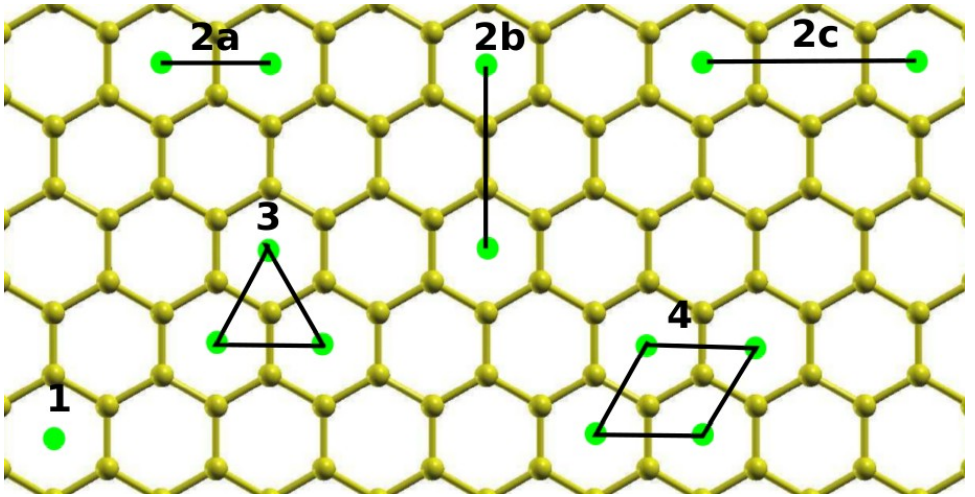
$$C_N = \{c_i\}, i = 1, ..., N,$$
for N lattice sites

$$E_{CE}(C_N) = \sum_f J_f \Pi_f(C_N)$$

$$= \sum_f J_f \prod_{i \in f} c_i$$

Approximations:

- Li adsorption geometry is restricted to 2D.
- **Expansion is truncated.**



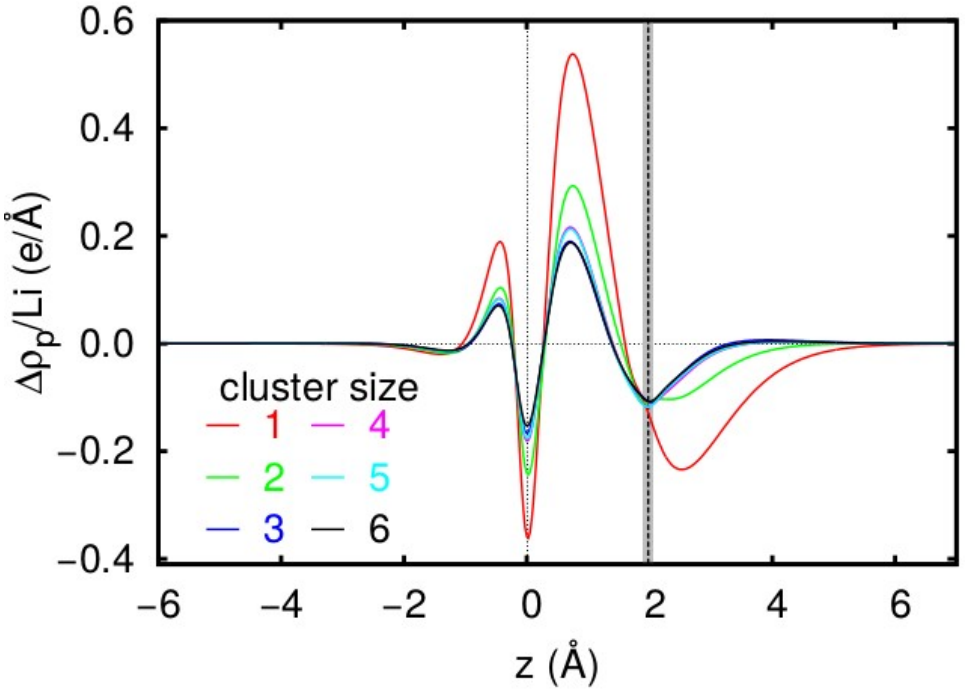
+

Long-range electrostatic contribution due to dipole interactions

Cluster expansion

$$\mathrm{E}_{CE}(C) = J_1 \sum_{i\in f_1} c_i + \frac{1}{2}\sum_{p=a,b,c} J_{2p} \sum_{\{ij\}\in f_{2p}} c_i c_j + \frac{J_3}{6}\sum_{\{ijk\}\in f_3} c_i c_j c_k + \frac{J_4}{24}\sum_{\{ijkl\}\in f_4} c_i c_j c_k c_l + \frac{J_{dd}}{2}\sum_{ij,|\mathbf{r}_i-\mathbf{r}_j|>2a_0} \frac{\alpha_i\alpha_j}{|\mathbf{r}_i-\mathbf{r}_j|^3} c_i c_j$$

Li-adatom vs Li-clusters



$$\alpha_i = \begin{cases} 1 & \text{if } NN_i = 0 \\ \gamma & \text{if } NN_i > 0 \end{cases}$$

J_{dd} is the same for all sites

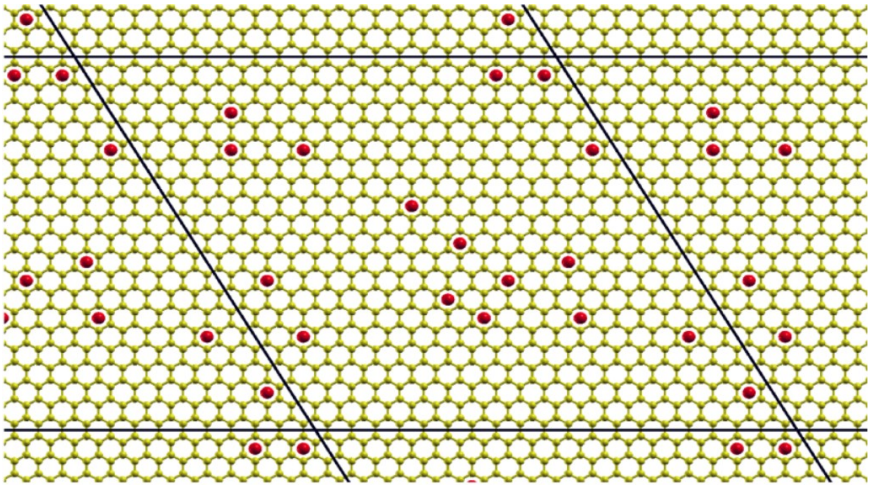
α distinguish the electrostatic interaction strength between isolated Li and clusters

Fit model to DFT energies by minimizing square deviation.

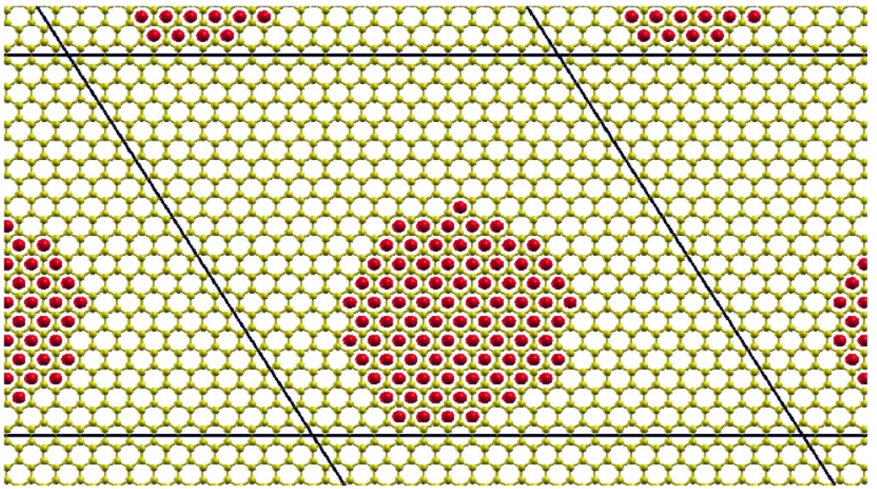
$$\sigma^2_{Li} = \frac{1}{N_{config}}\sum_{i=1}^{N_{config}} \left[\frac{E^i_{DFT}-E^i_{CE}}{N_{Li}}\right]^2 + t\sum_{i=1}^{N_J} J_i M_{ii} J_i$$

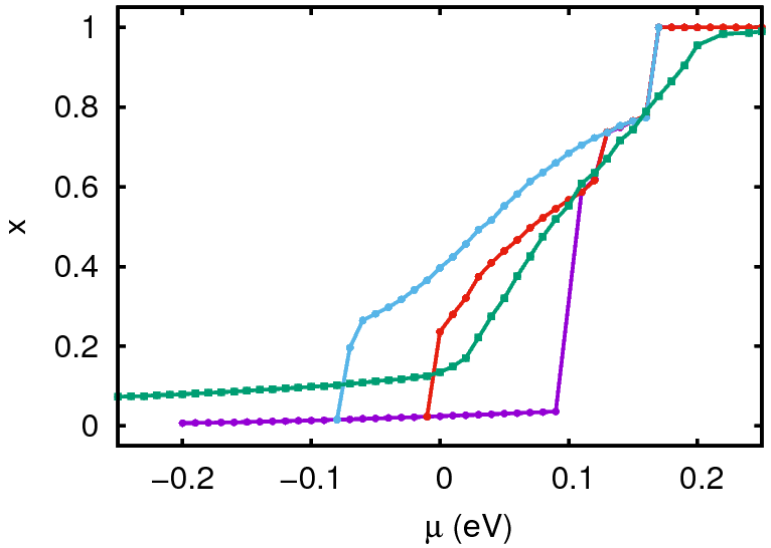
Phases

a) Li-gas

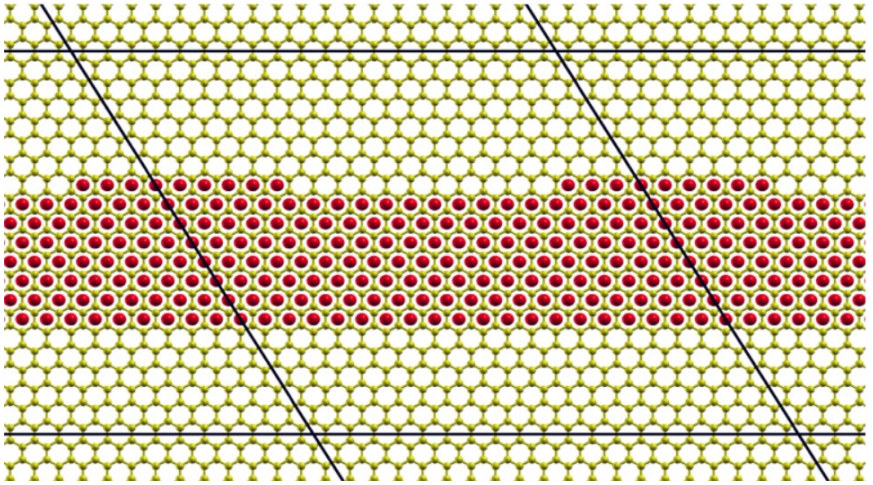


b) Li-island



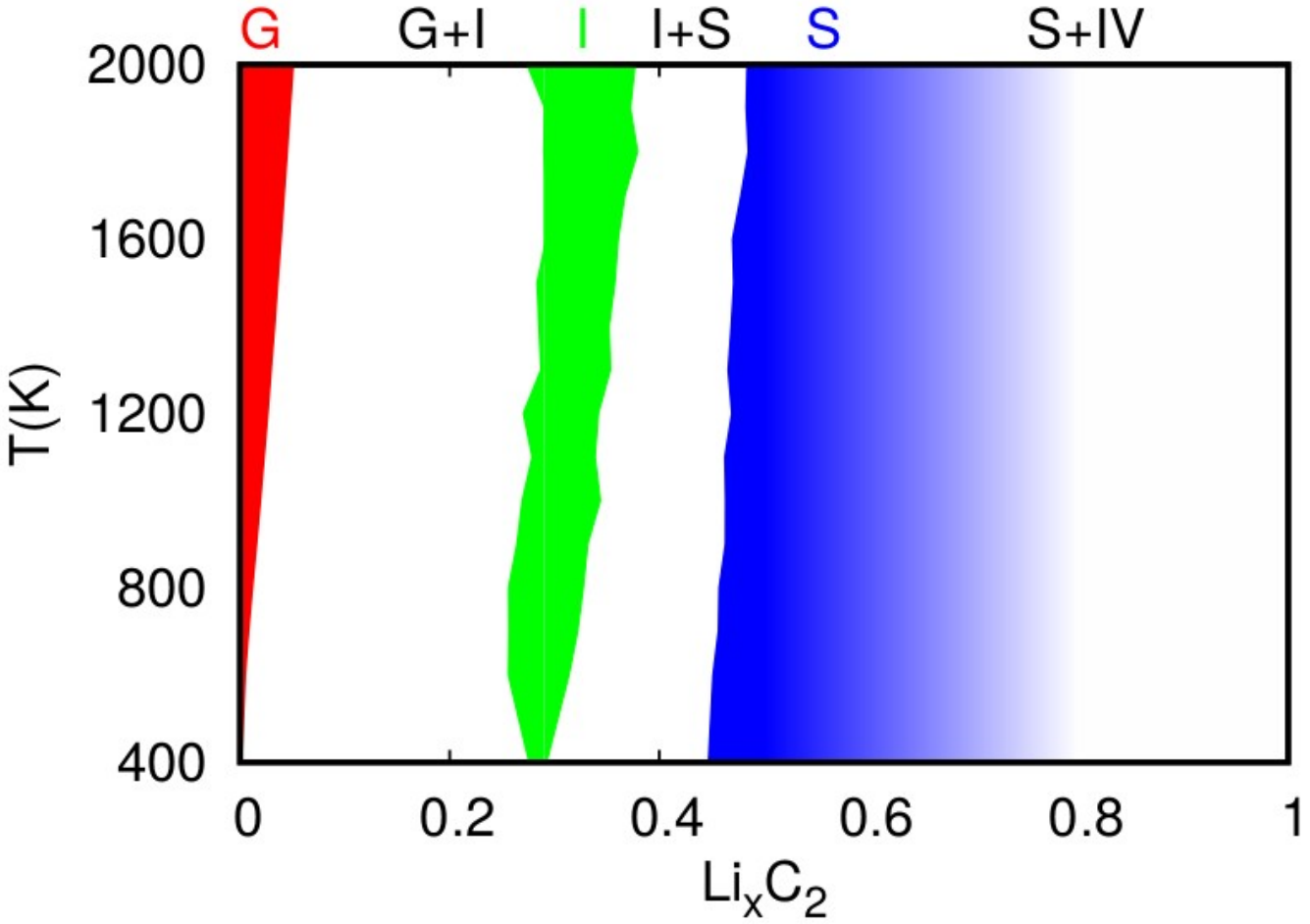


c) Li-stripe



Red dots: Lithium ions

Results: Phase diagram



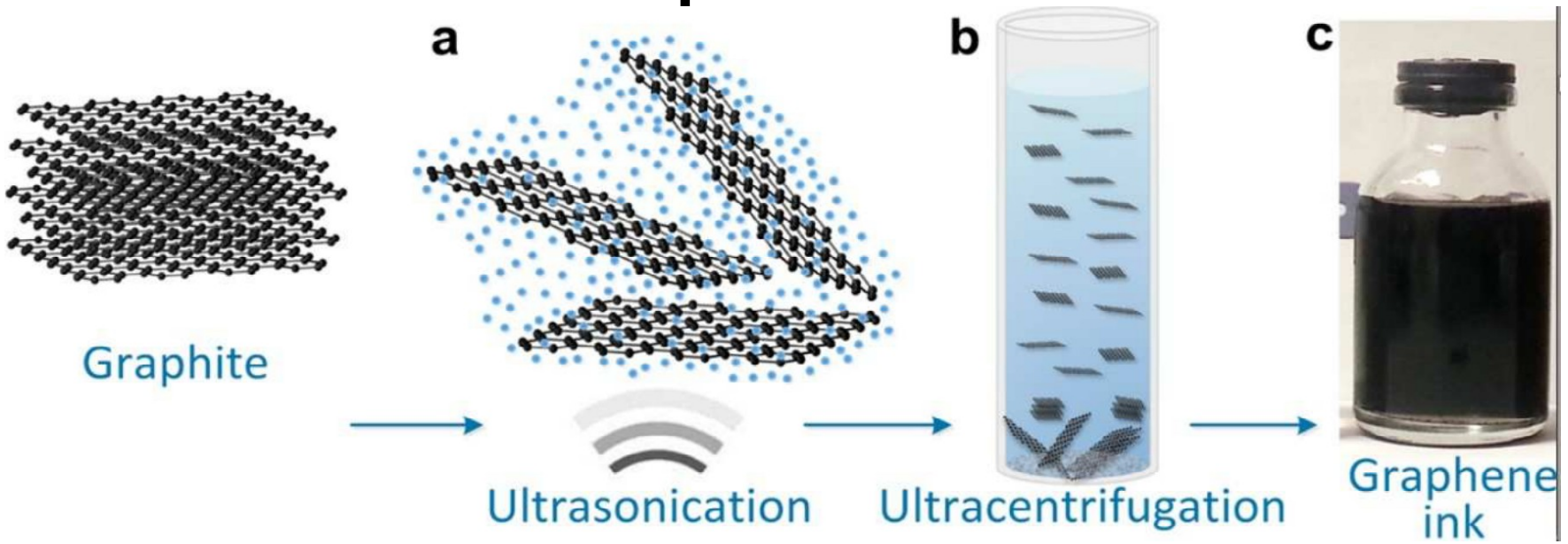
- **G: Li-gas**
- **I: Li-island**
- **S: Li-stripe**
- **IV: island of vacancies**

Two main phases:

1. **Random distribution of Li**
2. **2D Li-clusters**

Is Li adsorption in carbon materials really 2D?

Limitations of Cluster expansion



- Applied to only well defined underlying lattice.
- The dynamics of Li must be restricted to the lattice sites.
- **This calls for an accurate and yet affordable Li/C interaction potential.**

MACHINE LEARNING

Traditionally model potentials construction requires a lot of physical intuition and are strongly dependent on the available experimental information.

Not transferable to experimentally unexplored regions.

Limited accuracy due to rigid functional form.

DFT is a viable option to gather accurate information but requires a systematic approach to build a potential that can incorporate its features.


Artificial
Intelligence
is better than none

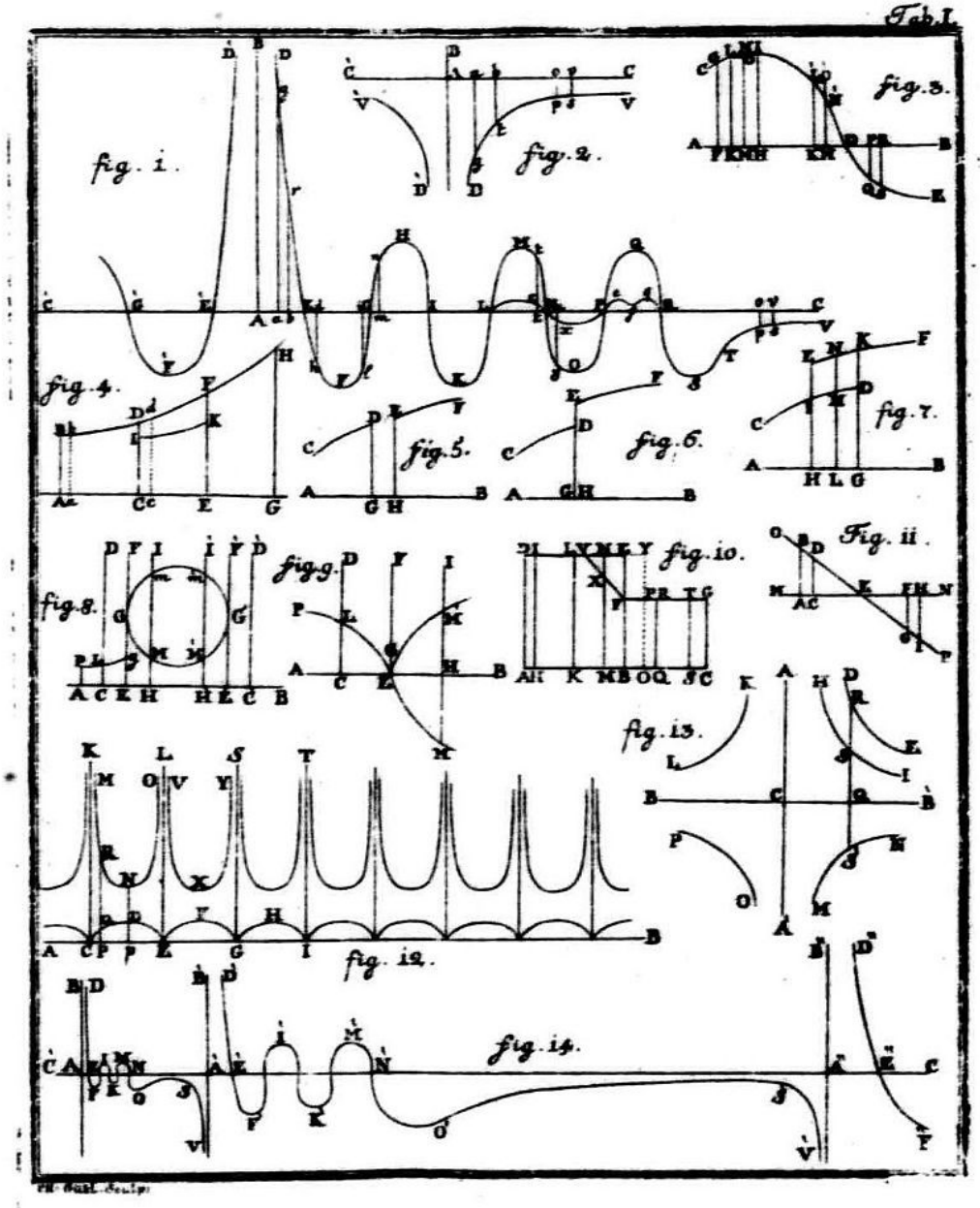
Replace the expensive DFT total energy calculations (or other accurate methods) with an interatomic potentials built to reproduce DFT data in a variety of environments

$$E(c) = \sum_{\alpha} \sum_{i \in \alpha} \varepsilon_{\alpha}(\mathbf{d}_i) + \text{ long range contrib}$$

- Kernel Ridge Regression (and Gaussian Processes)

- Neural Networks

- local environment descriptors



Ruđer Bošković, *Philosophiæ naturalis theoria redacta ad unicam legem virium in natura existentium,* (Vienna 1758)

Force Matching Method

Furio Ercolessi, JB Adams
"Interatomic potentials from first-principles calculations"
MRS Online Proceedings Library Archive, 291
arXiv:cond-mat/9306054

Peter Brommer, Franz Gähler
"Potfit: effective potentials from ab-initio data"
Modelling and Simulation in Materials Science and Engineering **15**, 295 (2007),
arXiv:0704.0185

$$Z = Z_{\mathrm{F}} + Z_{\mathrm{C}},$$

$$\text{with} \quad Z_{\mathrm{F}} = \sum_{j=1}^{N_A} \sum_{\alpha=x,y,z} W_j \frac{\left(f_{j_\alpha} - f_{0,j_\alpha}\right)^2}{\vec{f}_{0,j}^{\,2} + \varepsilon_j},$$

$$\text{and} \quad Z_{\mathrm{C}} = \sum_{k=1}^{N_c} W_k \frac{\left(A_k - A_{0,k}\right)^2}{A_{0,k}^2 + \varepsilon_k},$$

The high-throughput highway to computational materials design

S Curtarolo, GLW Hart, M Buongiorno Nardelli, N Mingo, S Sanvito & O Levy

Nature Materials **12**, 191 (2013)

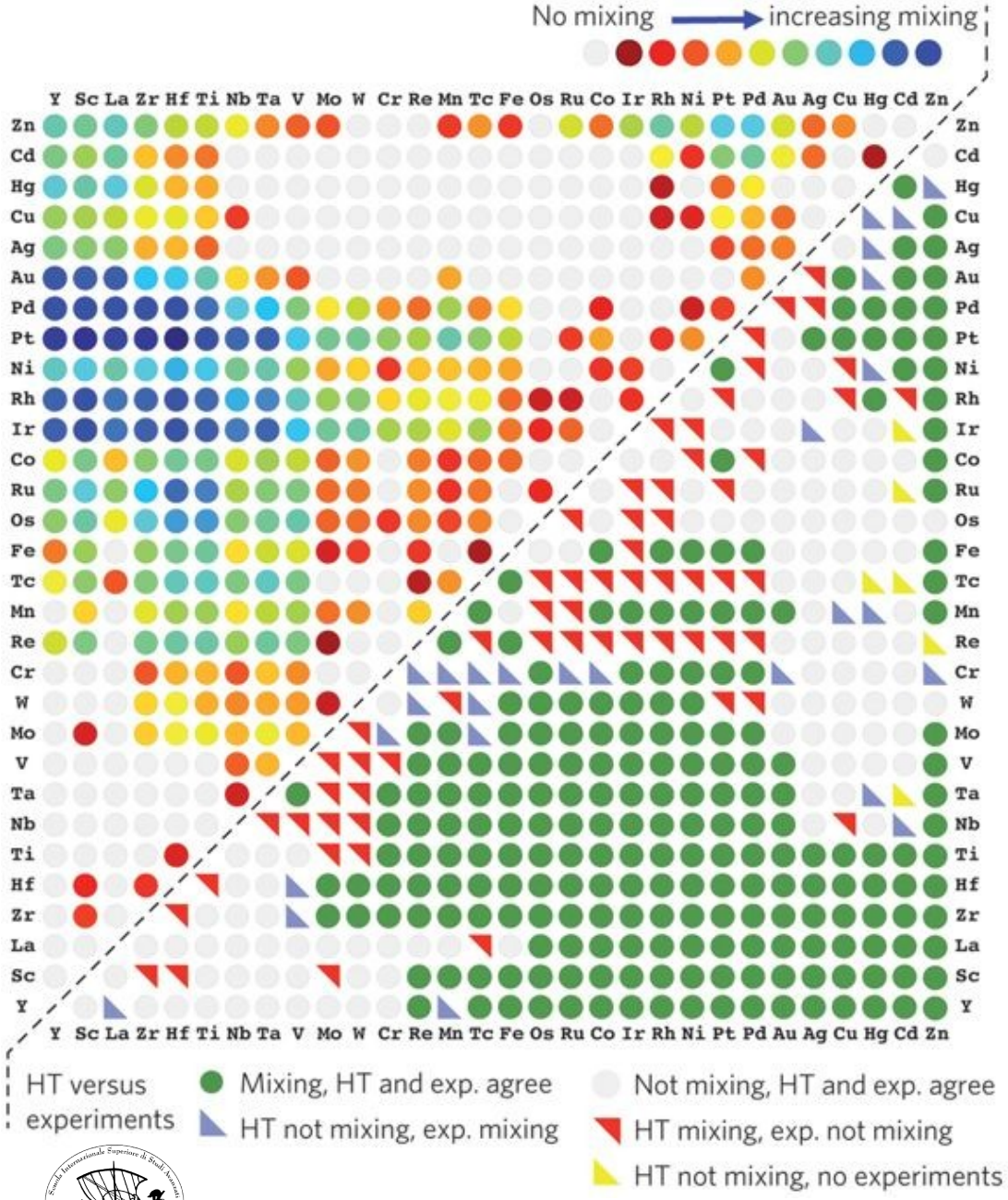
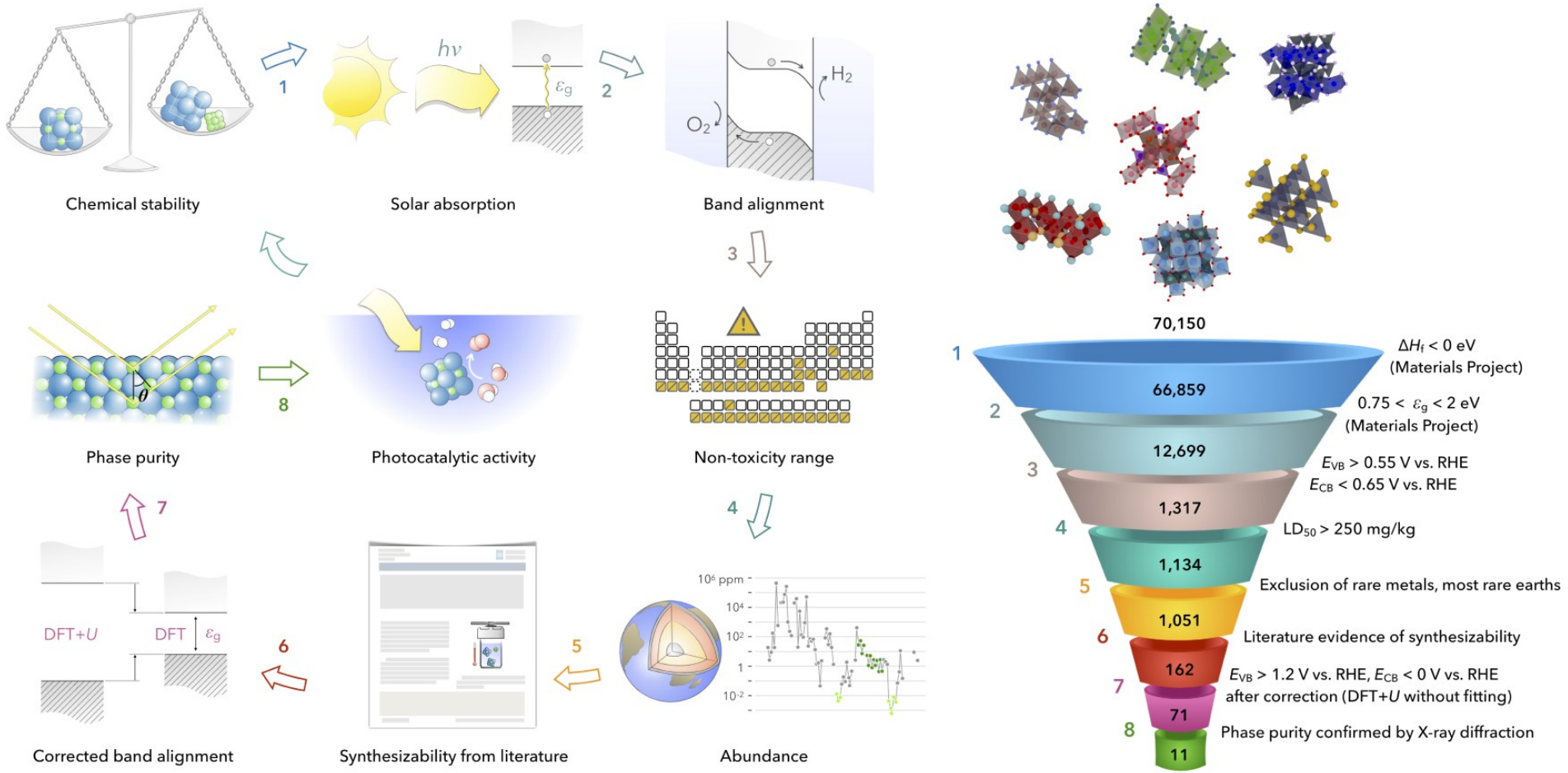


Figure 1 | High-throughput analysis of binary intermetallics[24]. Top left triangle: ordering tendency of the mixtures, as defined in the main text, for elements ordered by Pettifor's chemical scale[54]. Grey circles indicate no ordering, whereas darker blue circles indicate increasing capability to form stable compounds. Bottom right triangle: comparison of HT versus experimental results[24]. Green and grey circles denote agreement between calculation and experimental data on the existence (green) or absence (grey) of compounds. Purple (red) triangles indicate disagreement of HT predictions of compound absence (existence) versus experimental existence (absence). Yellow triangles indicate that data is unavailable for comparison.

Marco is a scientist, composer and flutist at the University of North Texas

SISSA PhD 1993

Optimizing accuracy and efficacy in data-driven materials discovery for the solar production of hydrogen”



Yihuang Xiong, Quinn T. Campbell, Julian Fanghanel, Catherine K. Badding, Huaiyu Wang, Nicole E. Kirchner-Hall, Monica J. Theibault, Iurii Timrov, Jared S. Mondschein, Kriti Seth, Rebecca Katz, Andres Molina Villarino, Betül Pamuk, Megan E. Penrod, Mohammed M. Khan, Tiffany Rivera, Nathan C. Smith, Xavier Quintana, Paul Orbe, Craig J. Fennie, Senorpe Asem-Hiablie, James L. Young, Todd G. Deutsch, Matteo Cococcioni, Venkatraman Gopalan, Hector D. Abruña, Raymond E. Schaak, Ismaila Dabo

arXiv:2102.01154

Thank you for your attention