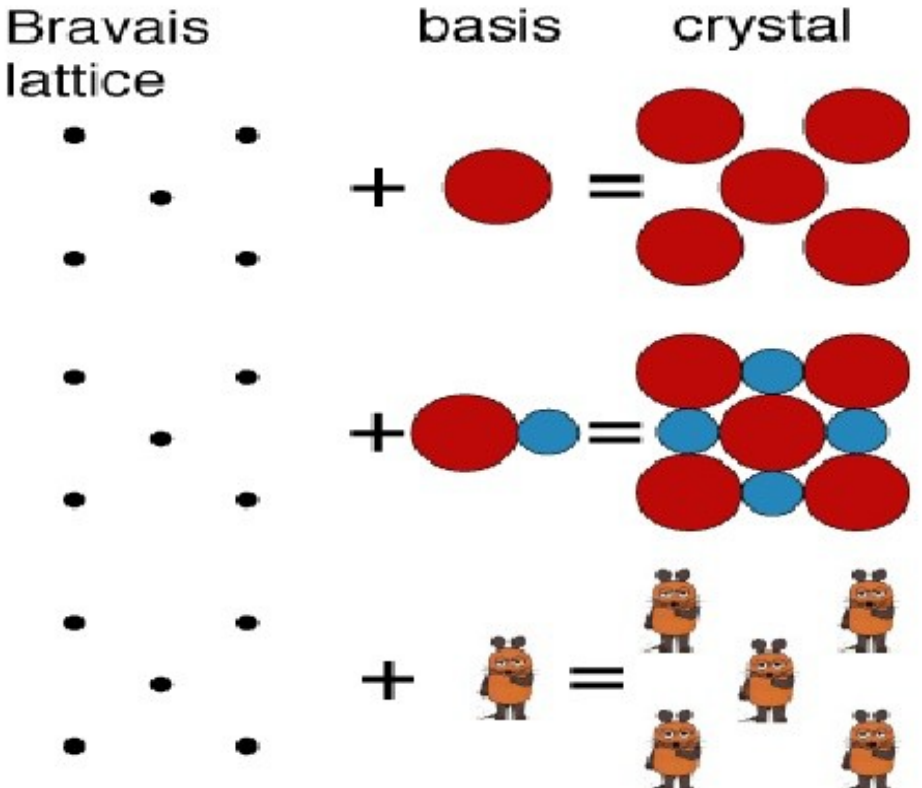
Crystal Structure Prediction From First Principles

Emine Kucukbenli – SISSA, Italy

Crystals

- *Condensed* Matter Physics
- *Solid* State Physics
- *Ordered* solid matter => Crystals
- Lattice + Basis

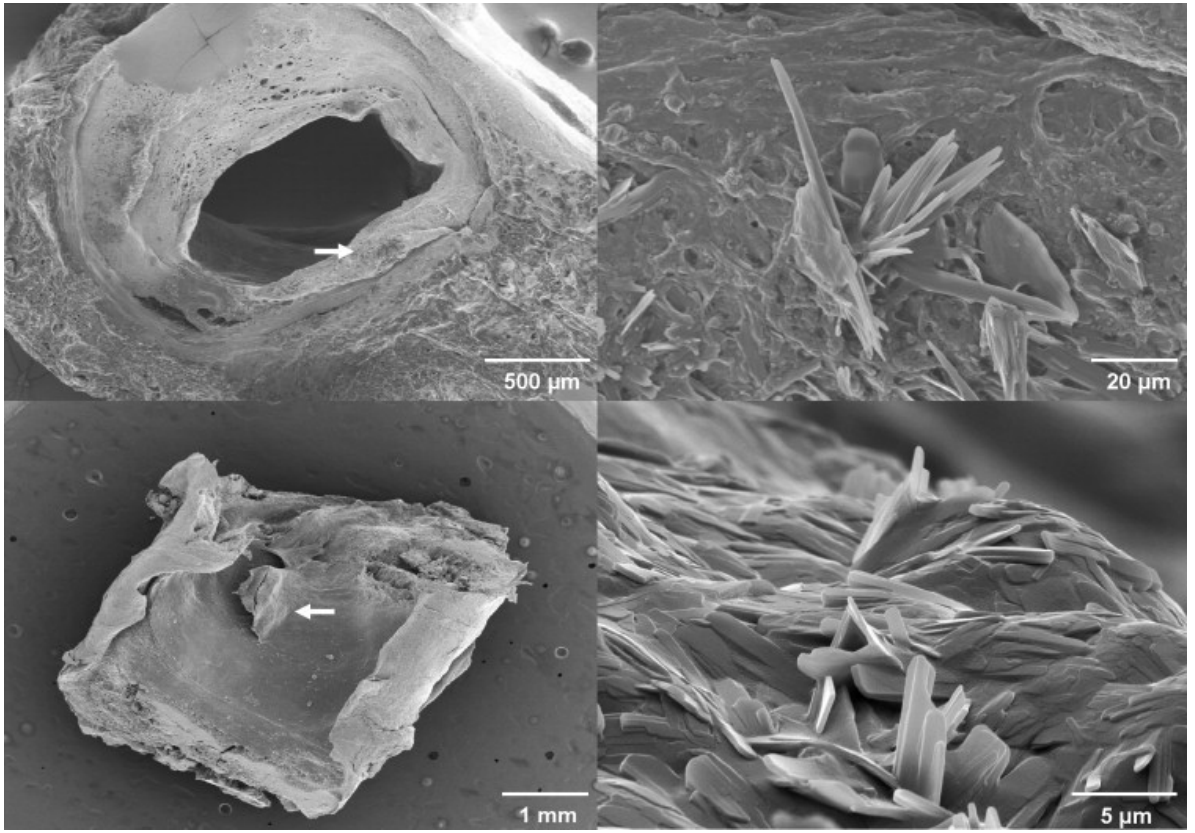
A two-dimensional Bravais lattice with different choices for the basis



Crystals

- Condensed Matter Physics
- Solid State Physics
- Ordered solid matter => Crystals
- Lattice + Basis

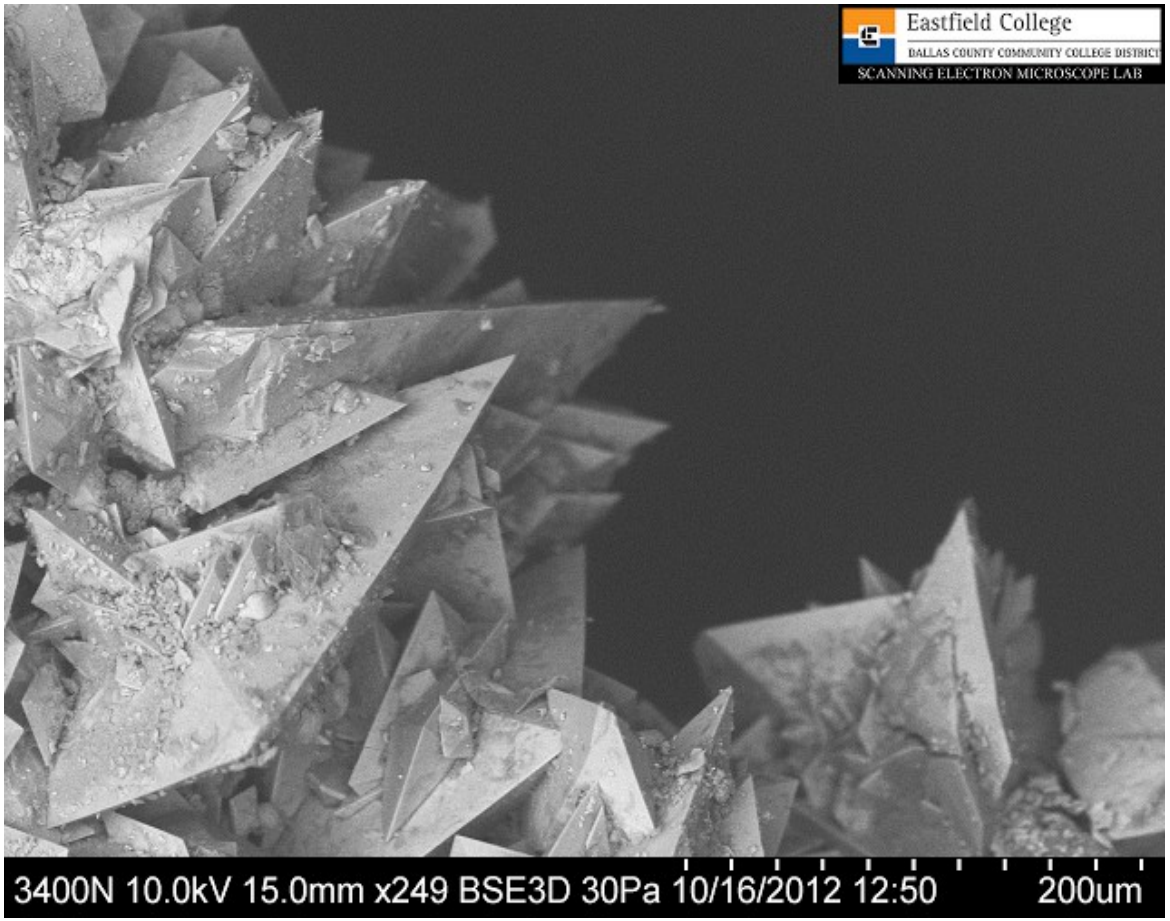
Cholesterol crystals



Verde et al. Atherosclerosis 203 (2009).

Crystals

- Condensed Matter Physics
- Solid State Physics
- Ordered solid matter => Crystals
- Lattice + Basis



Uric acid crystals

Not a perfect representation but a good one to start from

Eastfield College SEM Lab (2012)

Studying Crystals

- What keeps crystals together?

  An ordered system of N-many atoms bound by Coulomb interaction.

$$[T_I + T_e + W_{eI} + W_{ee} + W_{II}]\,\Psi_\alpha(\mathbf{r},\mathbf{R}) = E_\alpha \Psi_\alpha(\mathbf{r},\mathbf{R})$$

$$\Psi(\mathbf{r},\mathbf{R}) = \Phi(\mathbf{r}|\mathbf{R})\chi(\mathbf{R}), \qquad M_I >> m_e$$

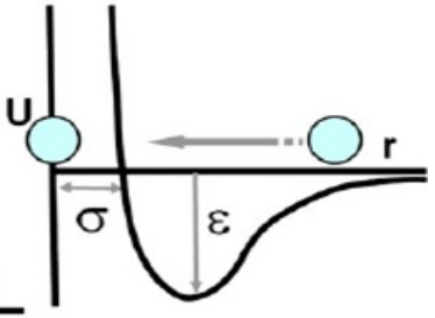
$$[T_e + W_{eI} + W_{ee} + W_{II}]\,\Phi_\nu(\mathbf{r}|\mathbf{R}) = E_\nu(\mathbf{R})\;\Phi_\nu(\mathbf{r}|\mathbf{R})$$

$$V(\mathbf{r},\mathbf{R}) = \frac{e^2}{2}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{e^2}{2}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

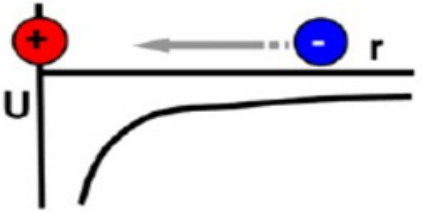
Studying Crystals

- What keeps crystals together?

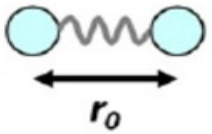
$$U = \sum_{i<j} \sum 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$



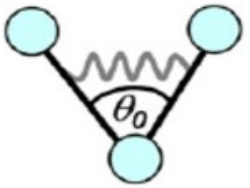
$$+ \sum_{i<j} \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



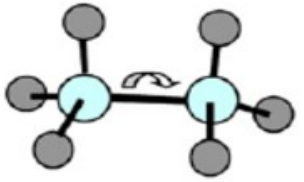
$$+ \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2$$



$$+ \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2$$



$$+ \sum_{torsions} k_\phi [1 + \cos(n\phi - \delta)]$$



Papadimitriou et al., Molecular Simulation 41, (2015)

Studying Crystals from First Principles

- No experimental parametrization of the potential; but physical assumptions are accepted.
  - A mapping of the QM-MB equation to a simpler one under certain assumptions:
    - Heavy nuclei
    - Ground state
    - Core-valance
    - Wavefunction in the form of Slater determinants or similar
    - For DFT, the functional form LDA/GGA etc.

Density Functional Theory

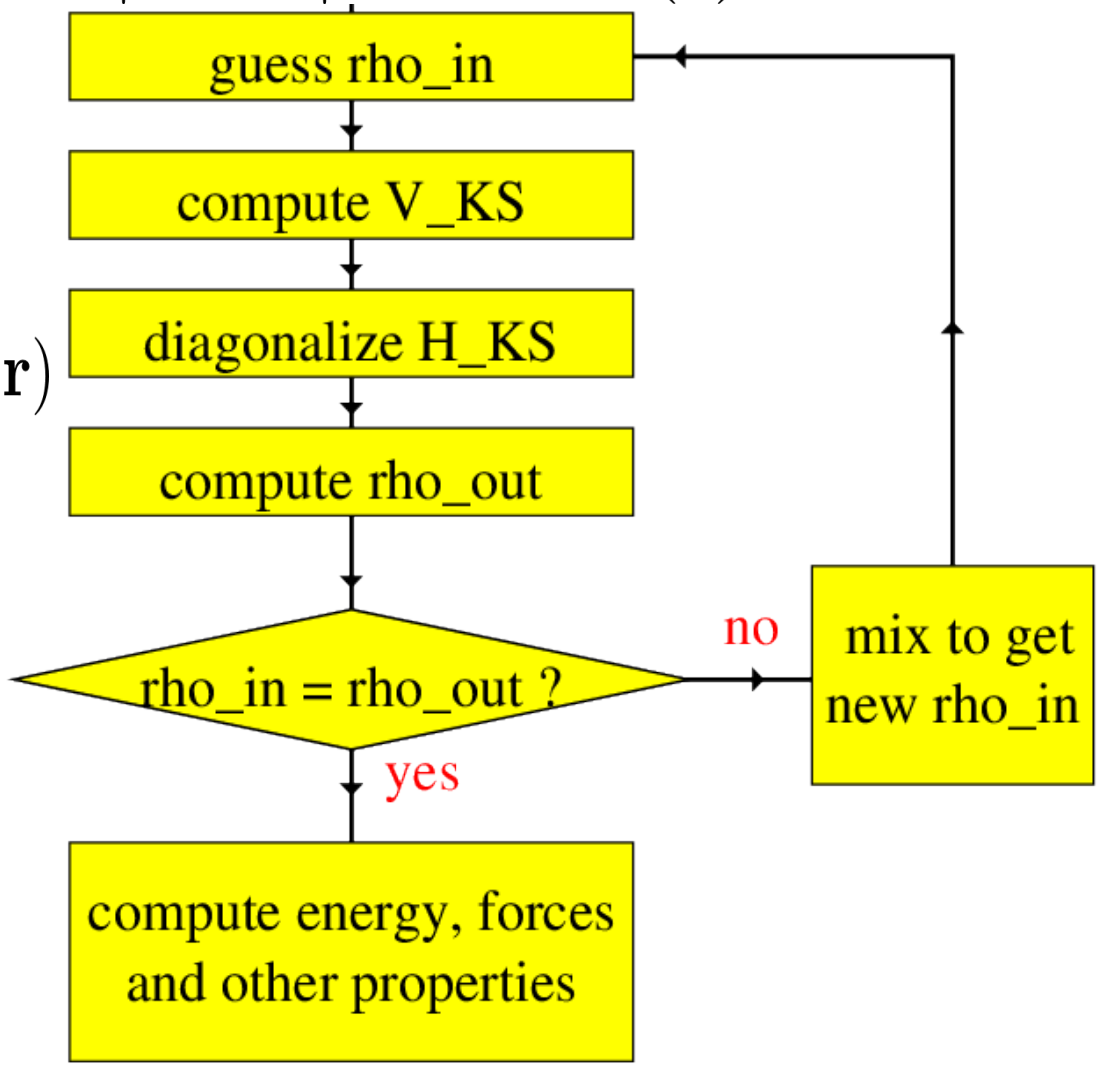
- The most popular first principles approach (2/10 most cited papers in all fields, all times)
- Affordable, Accurate, Available.
- Describing ground state density and potential as Legendre conjugate pairs E[V] <->F[n]
- Predictive for several ground state properties across the Periodic Table, as a first principles approach should be.

Density Functional Theory

$$V_{eff}(\mathbf{r},\mathbf{R}) = V_{ext}(\mathbf{r},\mathbf{R}) + e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}$$

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}(\mathbf{r},\mathbf{R}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r})$$

$$n(\mathbf{r}) = 2\sum_i |\phi_i(\mathbf{r})|^2$$

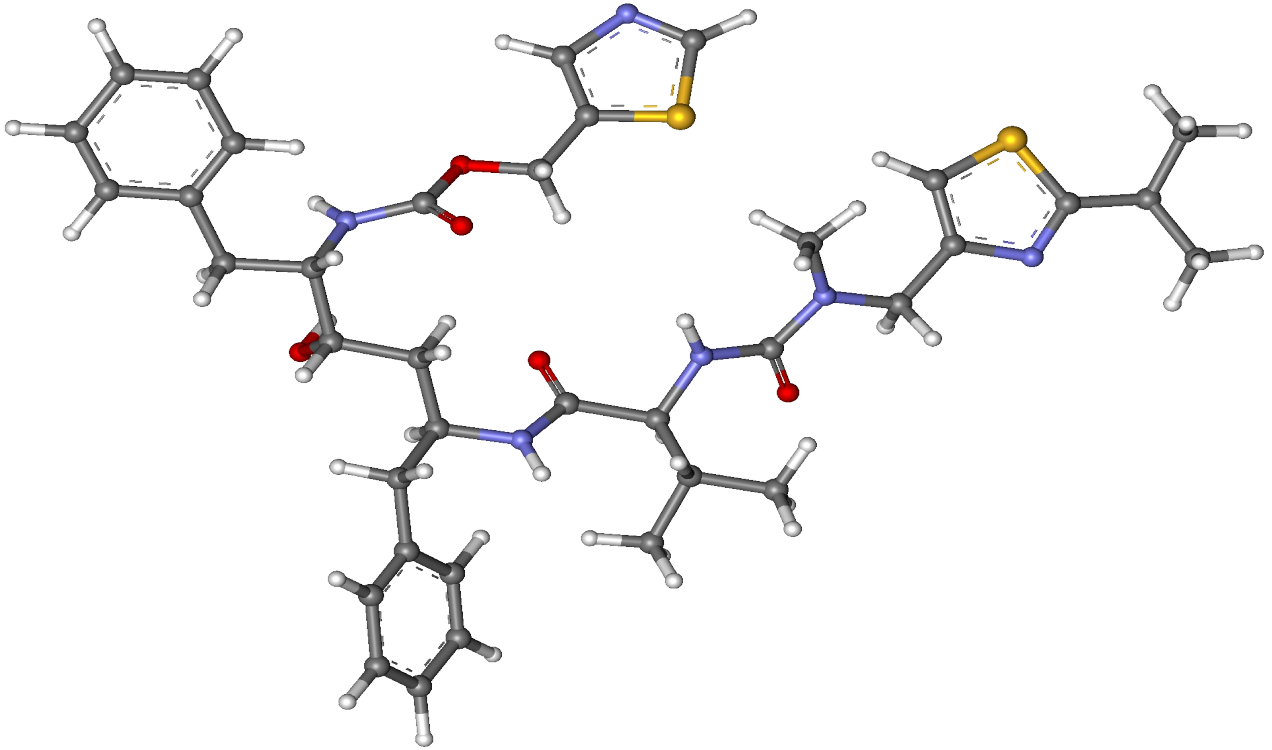


Studying Crystals with DFT

- Given a lattice + basis calculate
  - E : Ground state energy per unit cell
  - F_i : Force vector on each atom
  - S_ij : Stress on the unit cell and the pressure the cell feels
- What is the energetically most favorable lattice for this material?

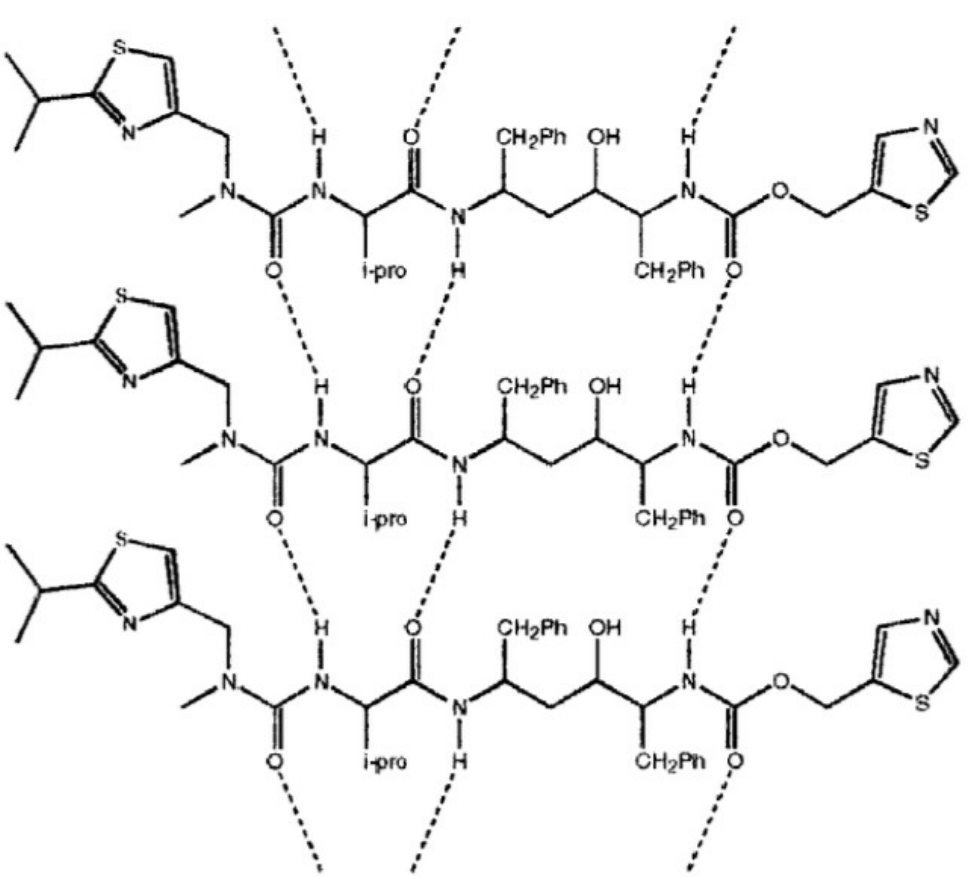
  => Crystal structure prediction
  - Structural properties,
  - Electronic&optical properties,
  - All these properties under pressure, at higher temperatures, in contact to humidity etc.

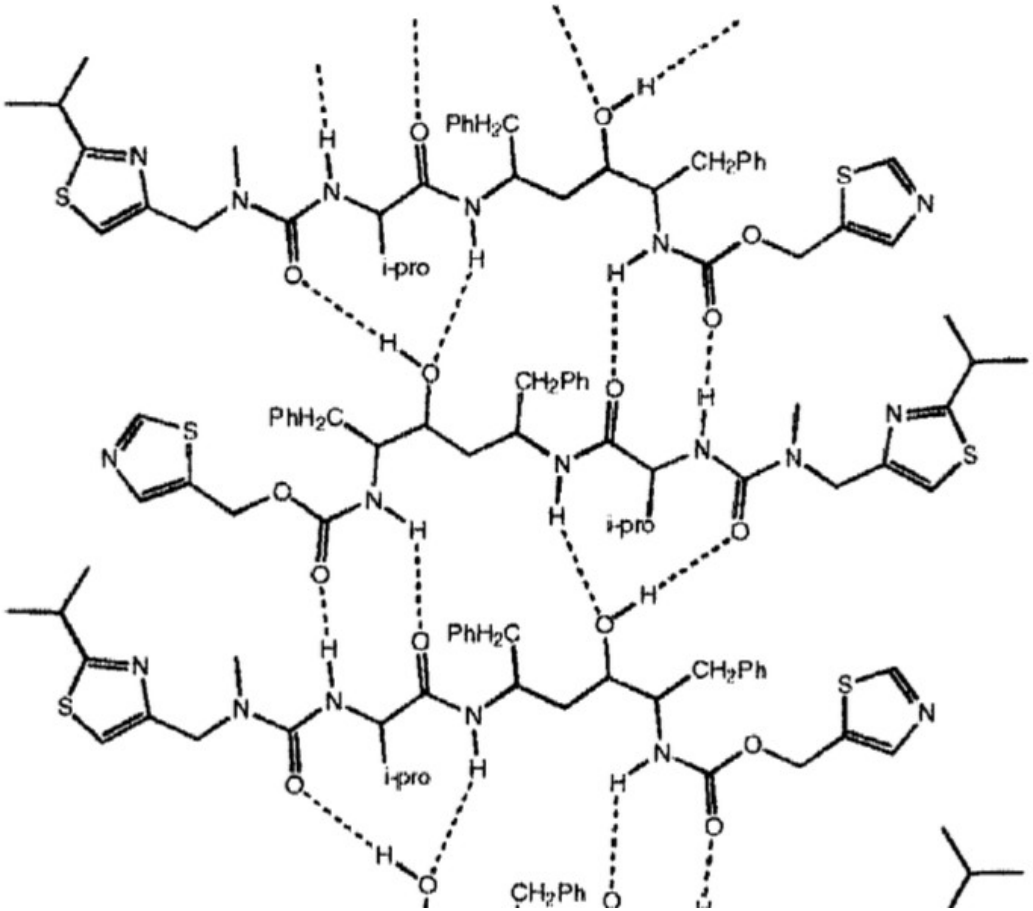
Story 1: Ritonavir



1996 - in the market – Non-refrigerated capsules
1998 – removed from the market in form II

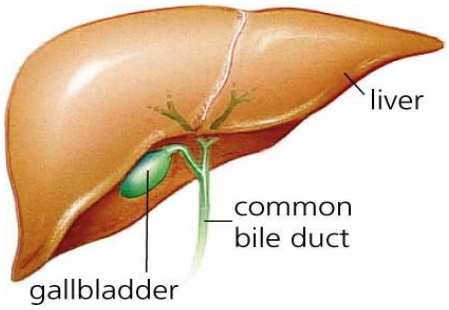
Story 1: Ritonavir





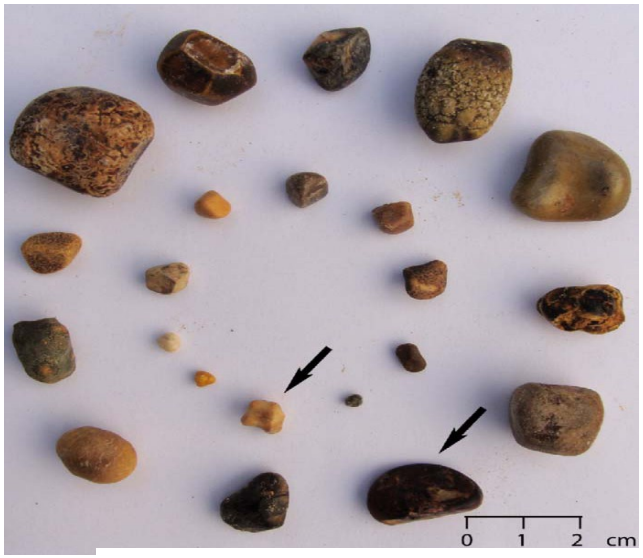
1996 - in the market – Non-refrigerated capsules
1998 – removed from the market in form II

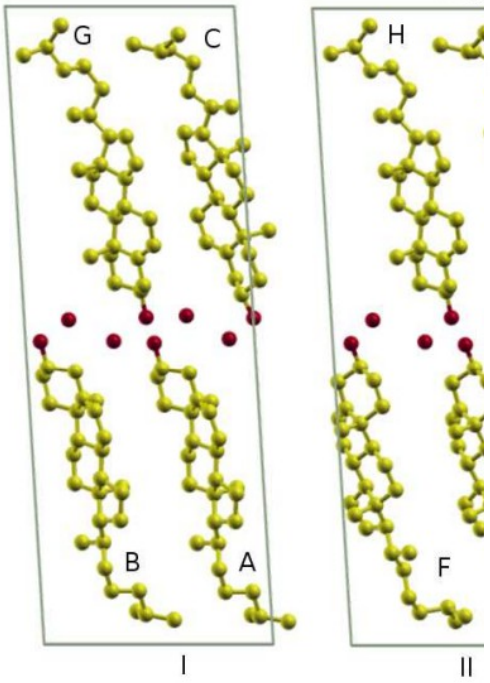
Bauer et al., Pharmaceutical Research 18 (2001)

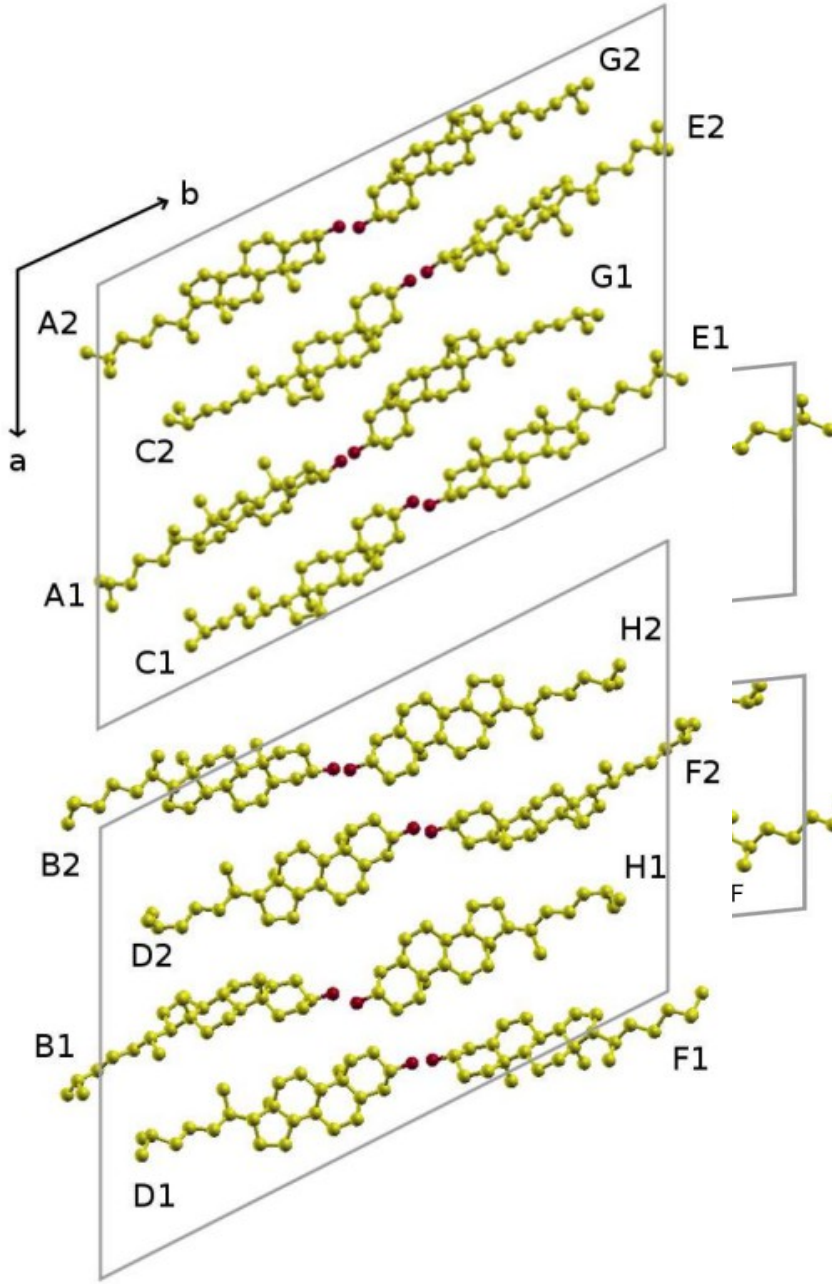


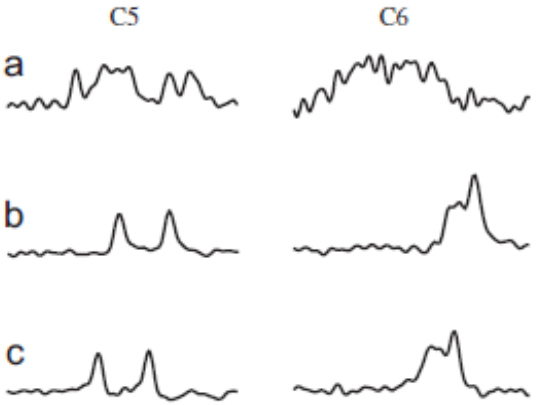
Story 2: Cholesterol

- Different crystal forms are as different diseases









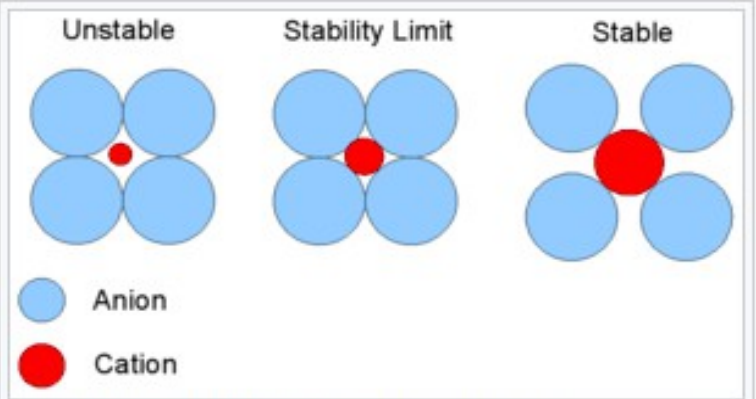
Only GBC patients

Is CSP a formidable problem?

- 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 too vast space of possibilities.
  - Free energy landscape is very expensive to obtain accurately

How to tackle CSP?

- Exploit vs Explore
- Pauling, 1929 CSP for ionic crystals



Critical Radius Ratio. This diagram is for coordination number six: 4 anions in the plane shown, 1 above the plane and 1 below. The stability limit is at $r_C/r_A = 0.414$

Polyhedron and minimum radius ratio for each coordination number

C.N.	Polyhedron	Radius ratio
3	triangular	0.155
4	tetrahedron	0.225
6	octahedron	0.414
7	capped octahedron	0.592
8	square antiprism (anticube)	0.645
8	cube	0.732
9	triaugmented triangular prism	0.732
12	cuboctahedron	1.00

How to tackle CSP?

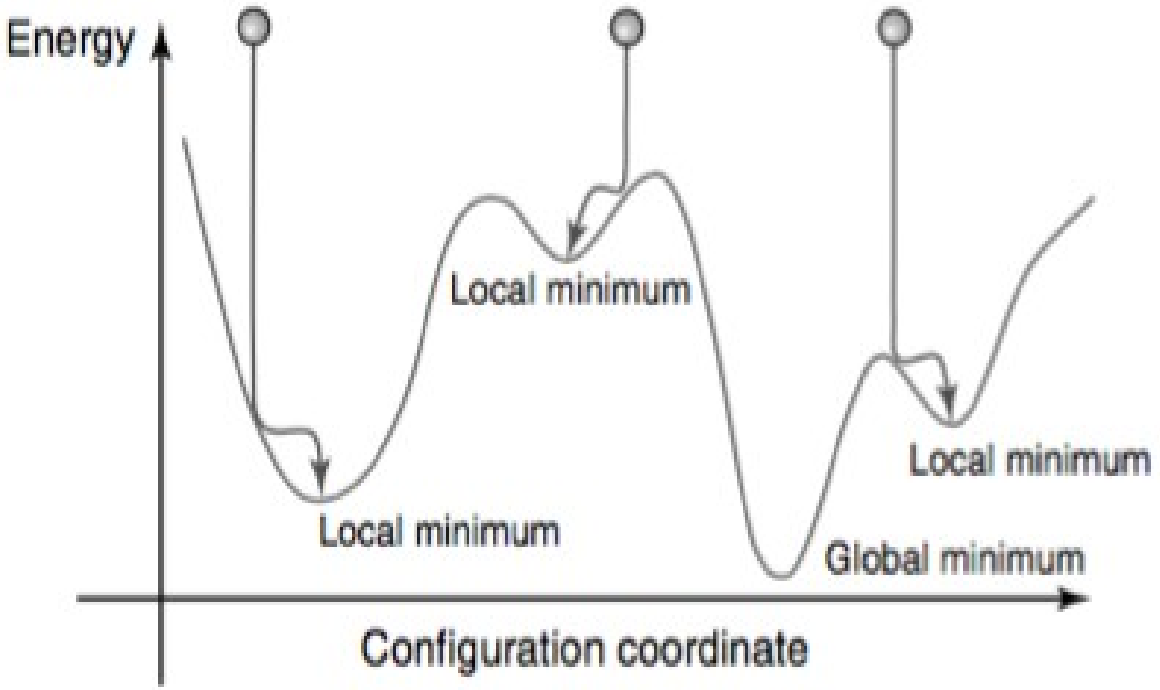
- Exploit: Data mining
- Binary alloys: choose the most common structures known + various compositions (114) of 55 alloys (55x114)
- Principle component analysis to project “Energy vs Structure” onto the smallest manifold (d=10-20)
- As few as possible first principles energies (d) the rest via linear combination

Curtarolo et al. PRL 91 (2003) Fischer et al. Nature mat. 5 (2006)

How to tackle CSP?

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

Well, not so smart but.. The random search



Random generation of initial guess str.

Meaningful constraints

Geometry optimization

How to tackle CSP?

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

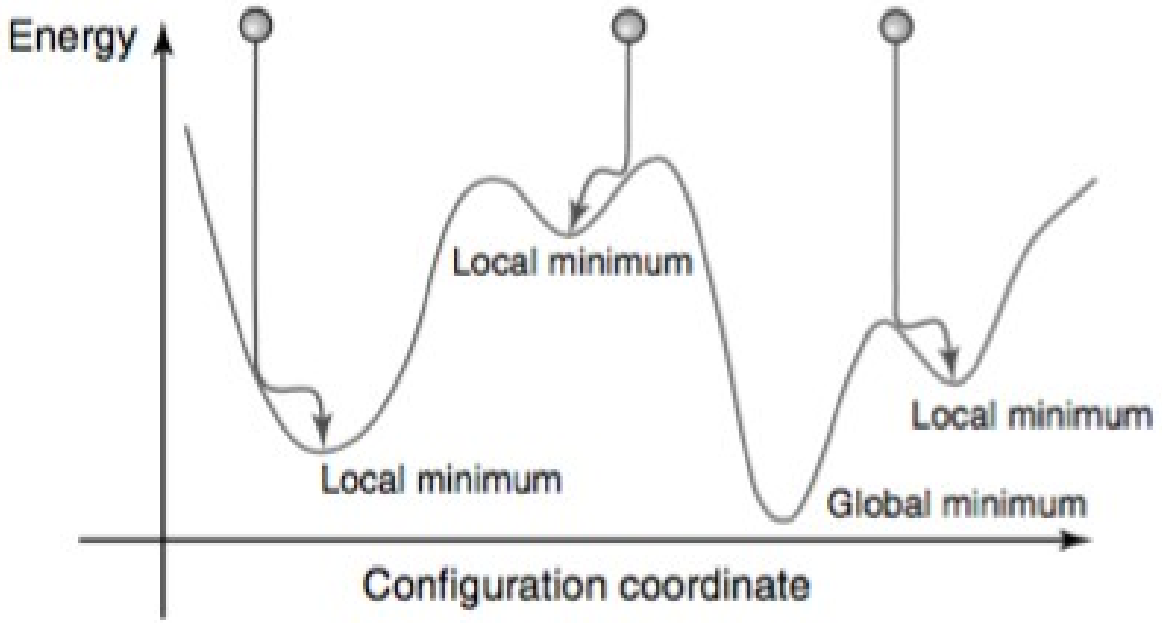
Molecular dynamics / Monte Carlo walkers

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

How to tackle CSP?

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

Molecular dynamics



How to tackle CSP?

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

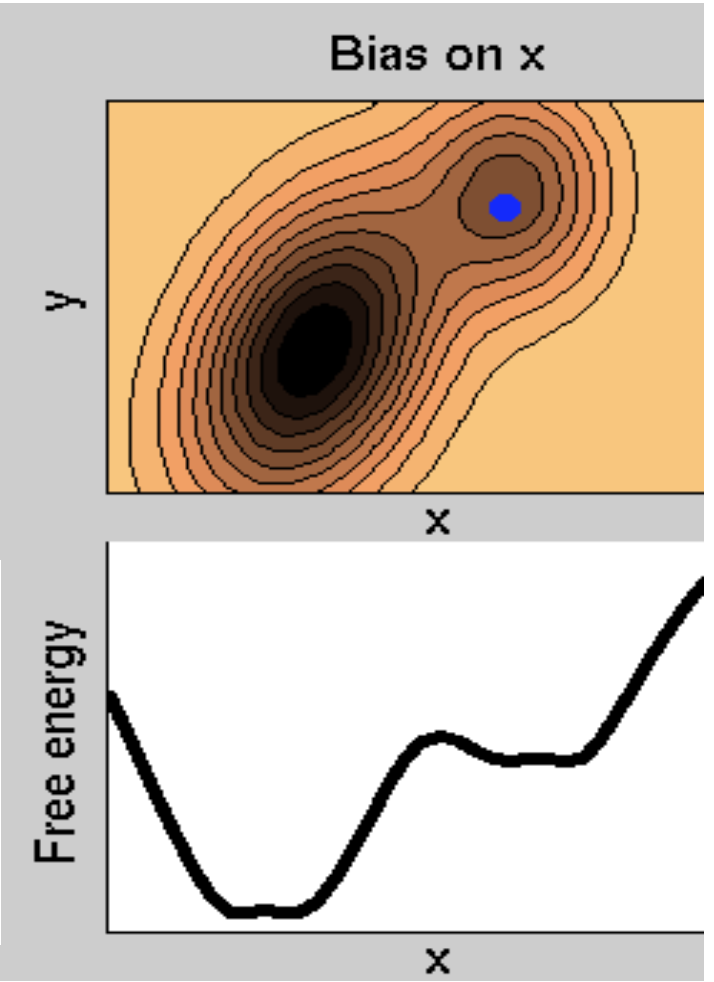
Molecular dynamics

- Metadynamics

H = T + V + V_bias

V_bias(s): s collective variable

$$V_{\mathrm{bias}}(\vec{s}) \approx \tau \sum_{j=0}^{\left\lfloor \frac{t_{\mathrm{sim}}}{\tau} \right\rfloor} \omega \exp\left(-\frac{1}{2} \left| \frac{\vec{s} - \vec{s}_j}{\vec{\sigma}} \right|^2 \right).$$



Laio & Parinello,PNAS 2002

How to tackle CSP?

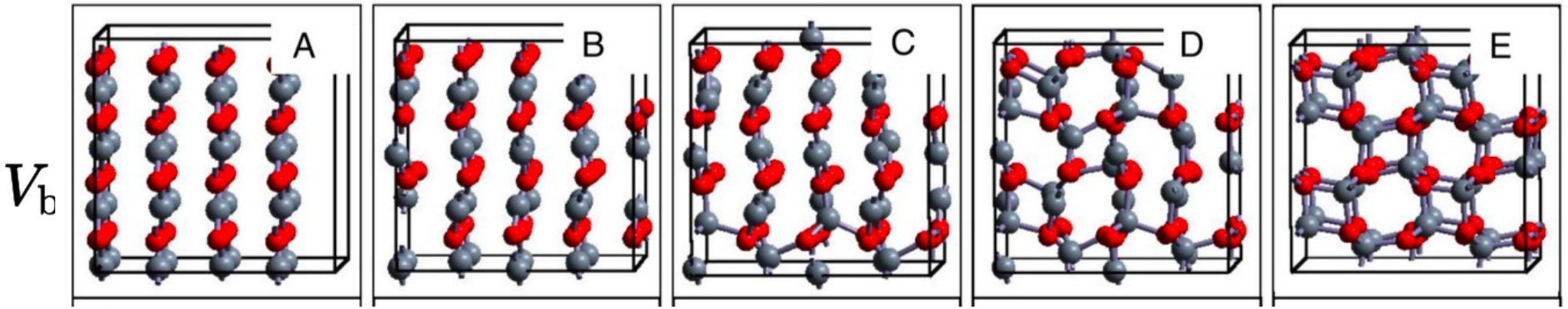
Explore: Use smart algorithms to explore as much of the landscape as possible

Molecular dynamics

- Metadynamics

H = T + V + V_bias

V_bias(s): s collective variable



Sun et al, PNAS 106 (2009)

How to tackle CSP?

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

Simulated annealing & Stochastic methods

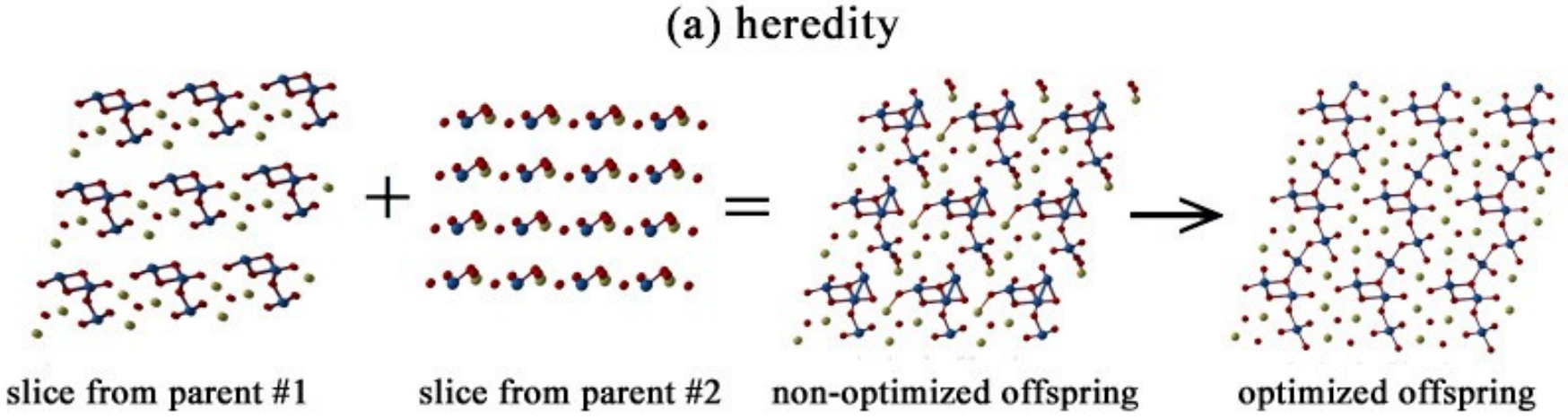
- Define a configuration space, cost function, move class.
- Minima hopping: Start at a local
- minimum; perform MD at temp T; Geometry optimization to find the closest minimum; Accept/Reject the new minimum based on its energy; increase or decrease the temperatue.

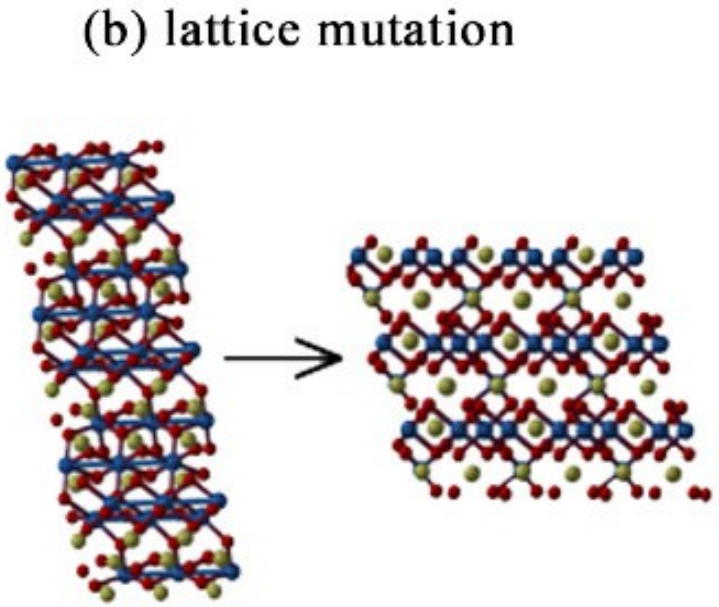
Genetic algorithm

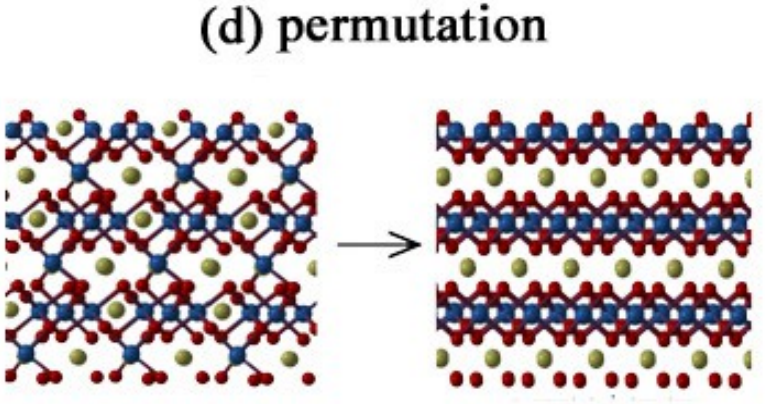
- Start with a set of random structures ->

  -> Parent population
- Geometry optimization (local minima)
- Heredity operations, mutation, strangers -> Children's population
- Cost function comparison to determine the next generation

Genetic algorithm







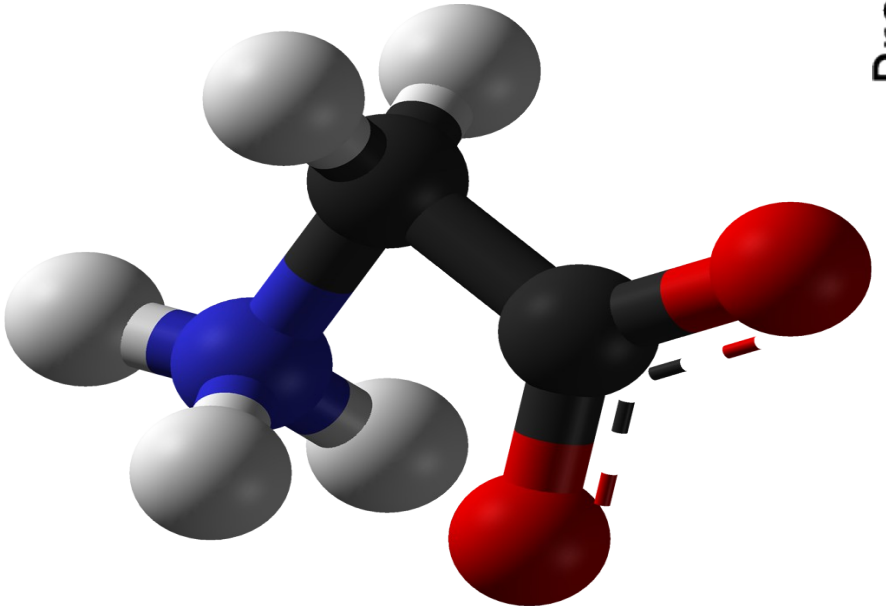
USPEX operations

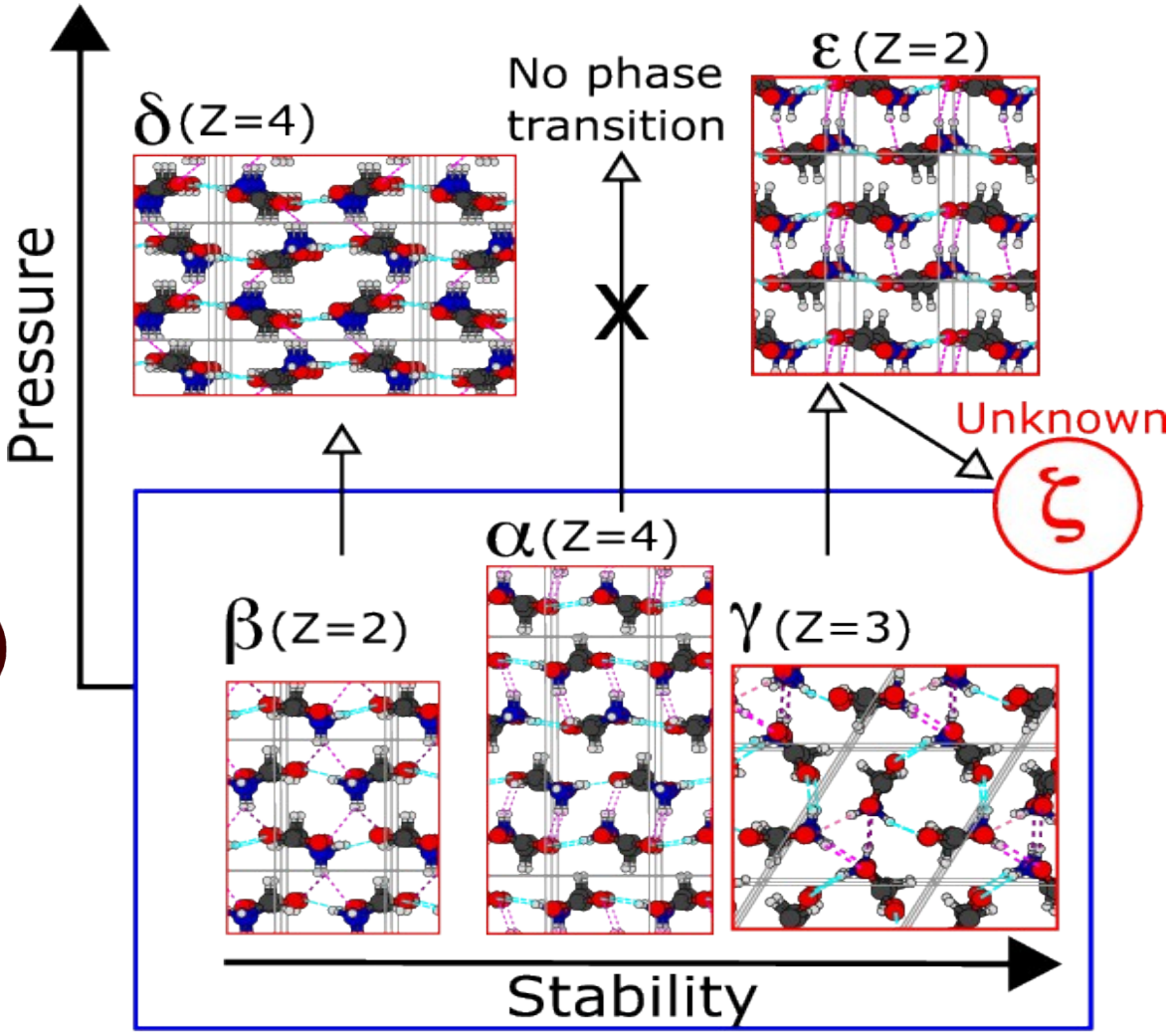
Success stories

They are many.

CSP problem has been “solved” for the last several years

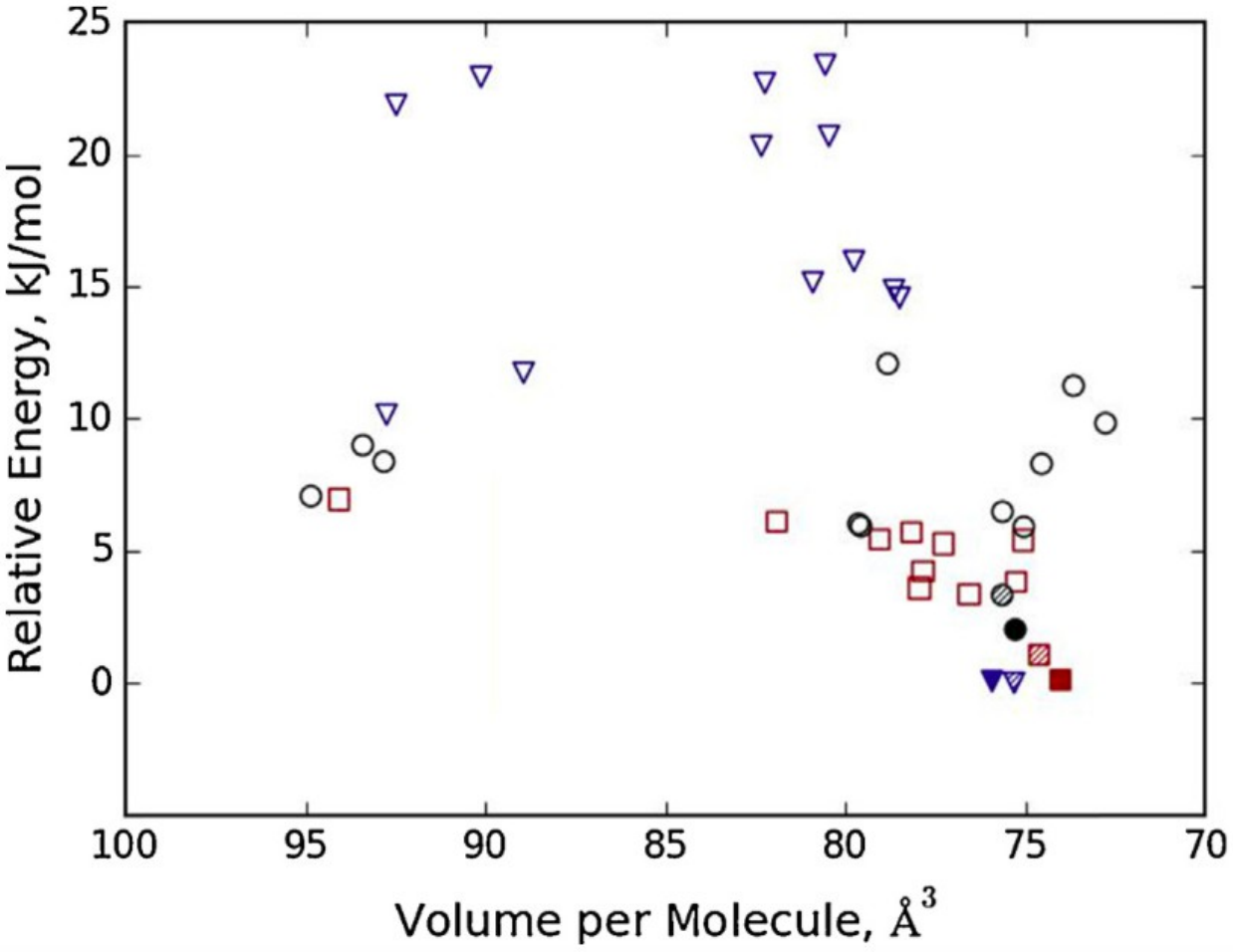
Case of Glycine

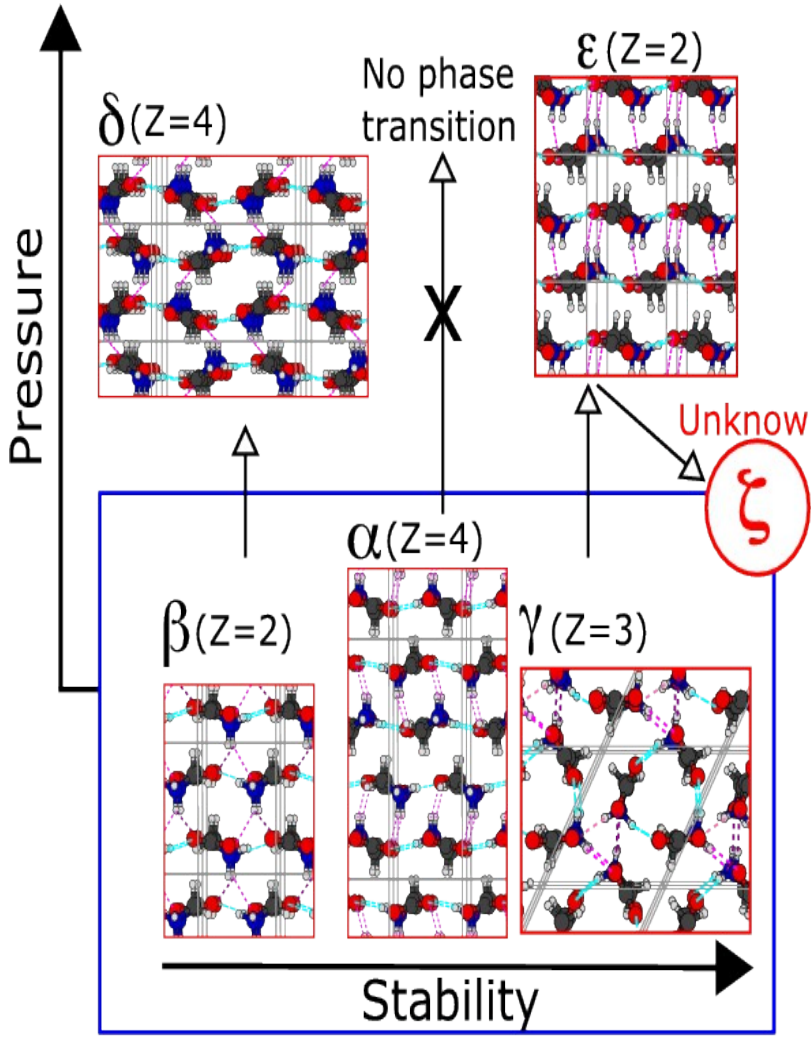




Success stories

Case of Glycine

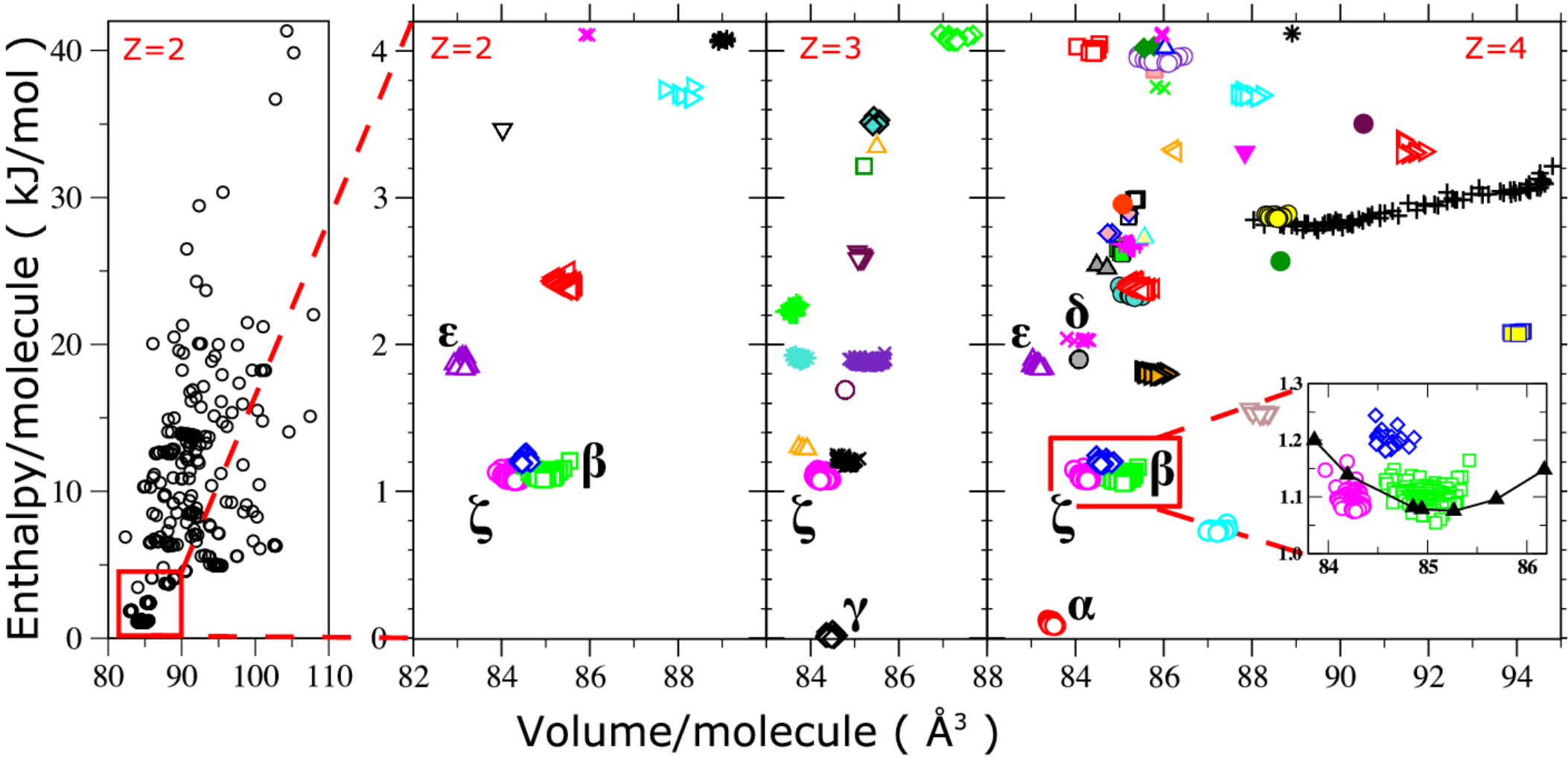




Lund et al, Chem Phys Let 262 (2015)

Success stories

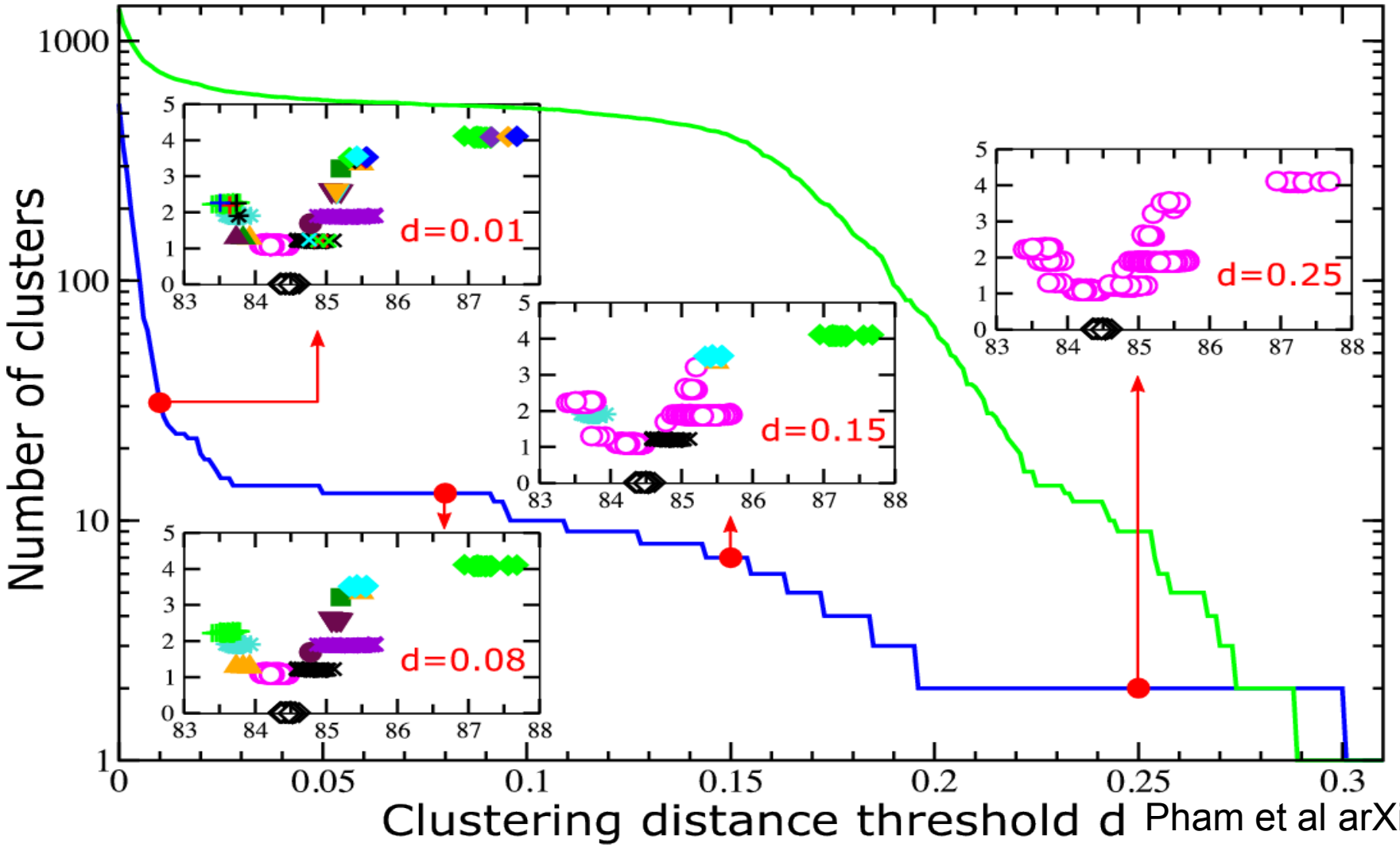
Case of Glycine



Pham et al arXiv:1605.00733

Success stories

- How to identify families of strtuctures "clustering"



Pham et al arXiv:1605.00733

Remaining Challenges for you

- Vast space to explore = Cost ->Machine learning algorithms, better exploitation of known landscapes
- Data Analysis & representation -> Dynamic variation of “fingerprint” tolerance, different fingerprinting
- Can we learn more Pauling rules? Extract insights from results