Stony Brook University

Structure Prediction for Molecular Crystals Using Evolutionary Algorithms: Methodology and Applications

Qiang Zhu

Department of Geosciences

Stony Brook University

Methodology

- *Constrained evolutionary algorithm*

Applications

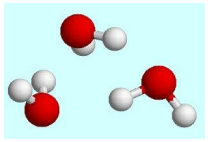
- *polyatomic molecules*
- *Pharmaceutical compounds*
- Organo metallics
- *Inorganic system*

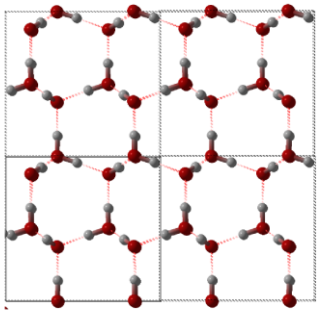
Outlook

- *Accuracy*
- *Force field*
- *User interface*

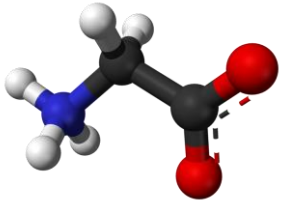
Molecular Crystals

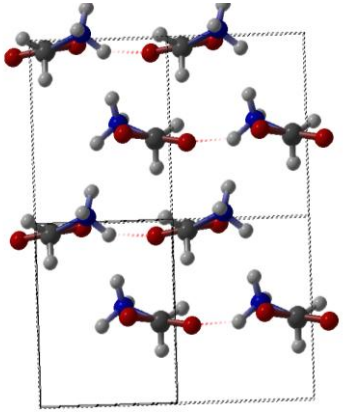
Ice



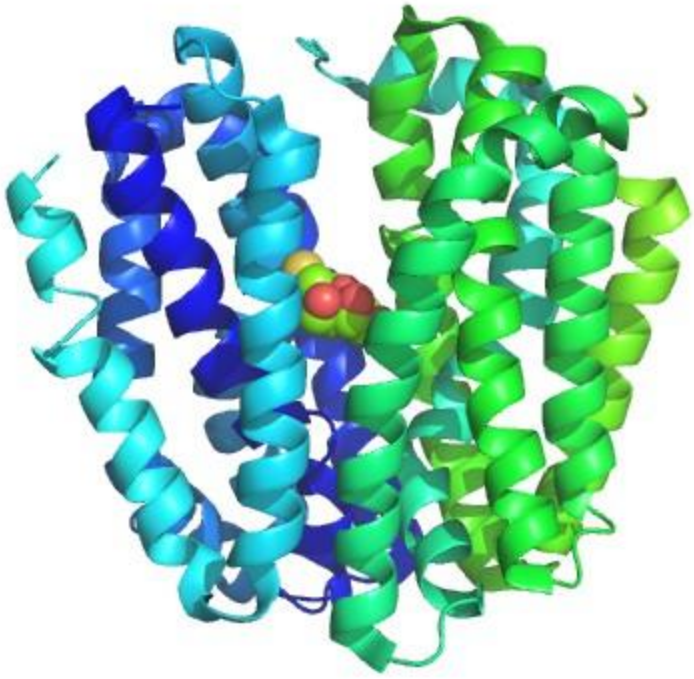


glycine





Lactose Permease



Organic Crystal Structure Prediction Blind Tests

Cambridge Crystallographic Data Centre

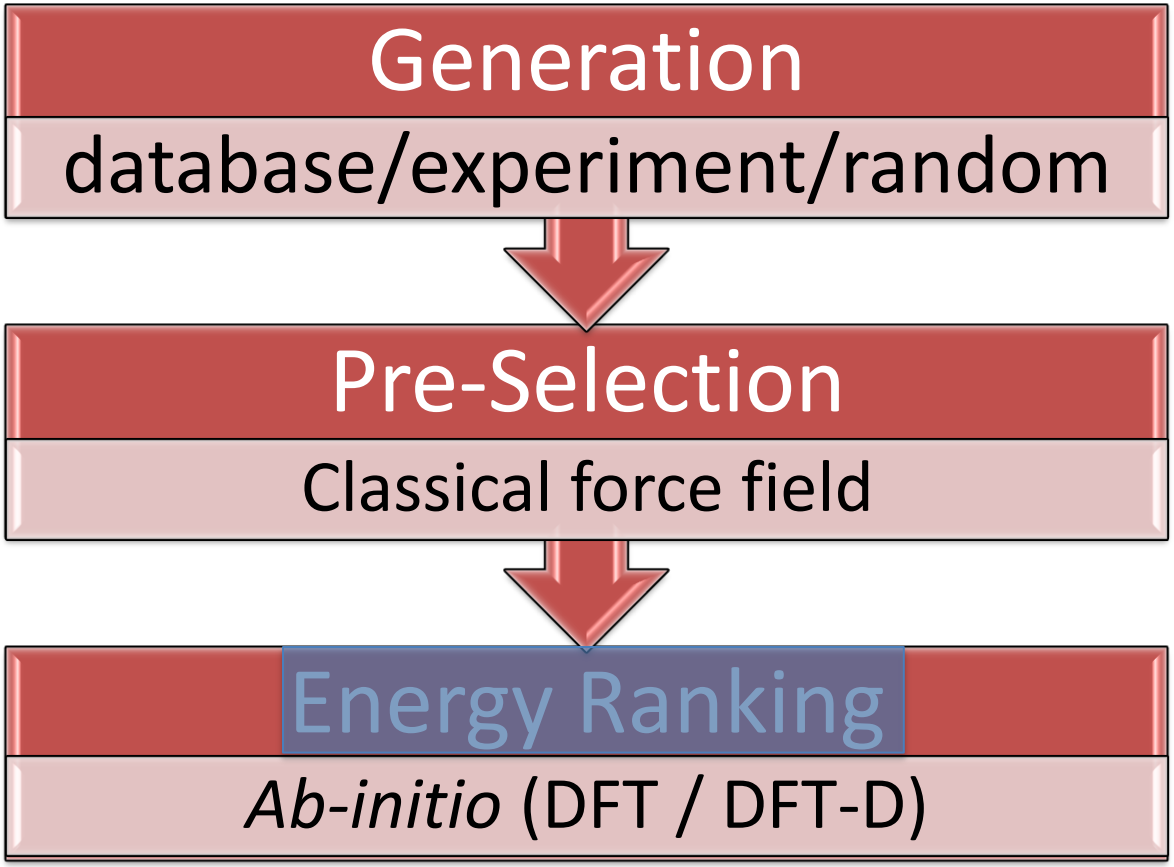
A test of crystal structure prediction of small organic molecules
J.P.M. Lommerse, W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, W.T.M. Mooij, S.L. Price, B. Schweizer, M.U. Schmidt, B.P. Van Eijck, P. Verwer, D.E. Williams, *Acta Crystallographica,Section B: Structural Science*, 56, 697-714, 2000
doi:10.1107/S0108768100004584

Crystal structure prediction of small organic molecules: a second blind test
W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooij, S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B P. van Eijck, P. Verwer, D.E. Williams, *Acta Crystallographica,Section B: Structural Science*, 58, 647-661, 2002
doi:10.1107/S0108768102005669

A third blind test of crystal structure prediction
G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, P. Verwer, *Acta Crystallographica,Section B: Structural Science*, 61, 511-527, 2005
doi:10.1107/S0108768105016563

Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test
G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf, B. Schweizer, *Acta Crystallographica,Section B: Structural Science*, 65, 107-125, 2009
doi:10.1107/S0108768109004066

Conventional Approach



Searching?

Molecular Crystals – Prediction?

Reliable

Ranking

Searching

Universal

Transferable

Economical

Computing time

Automatic

User friendly

(Parameter free)

New method: USPEX

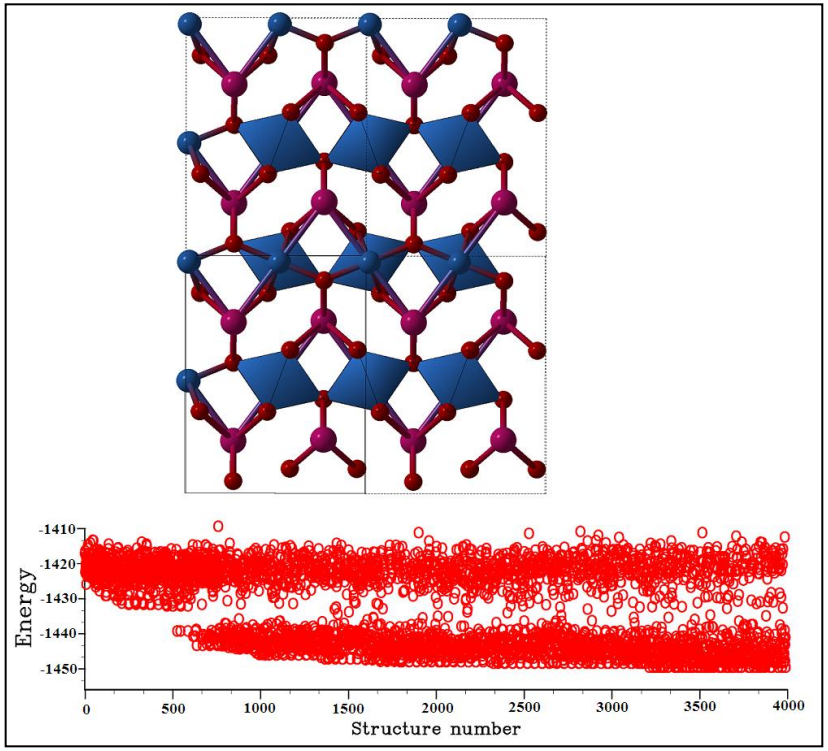
Ab initio determination

Given information as least as possible

Powerful Searching

Invariant with the system

Self-improved learning process



USPEX: from atom to molecule

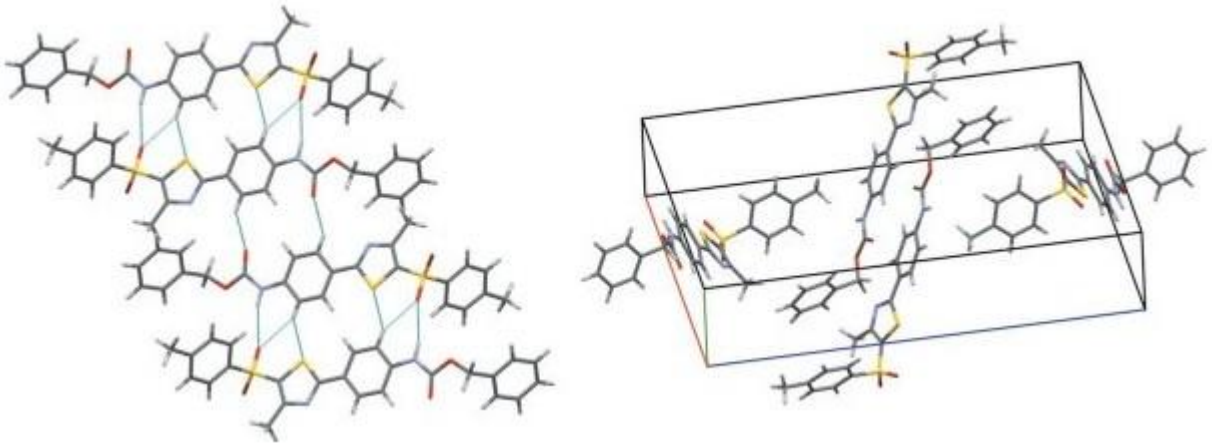
Universal Approach to predict organic molecular crystals

Main features of molecular crystals:

- Large amount of atoms
- Sparse space
- Metastable
- Symmetry preference

Molecular Crystals

1. Large amount of atoms



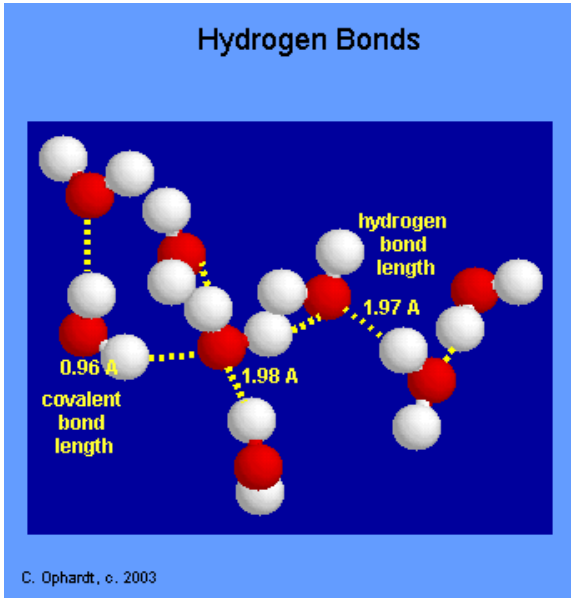
Benzyl-(4-(4-methyl-5-(p-tolylsulfonyl)-1,3-thiazol-2-yl)phenyl)carbamate

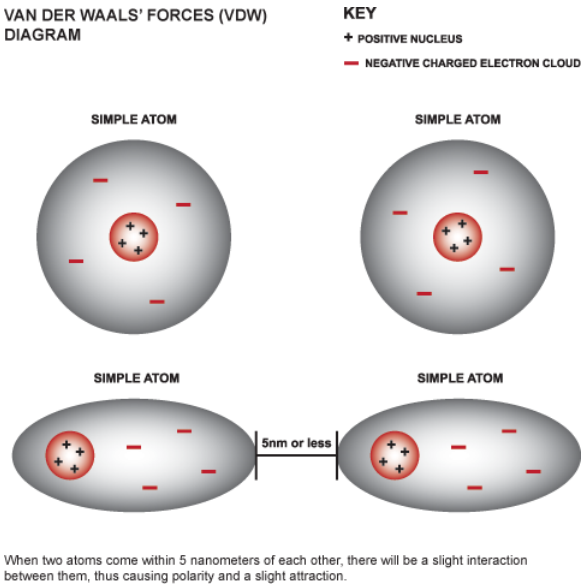
Molecular Crystals

2. Sparse space

H – bond

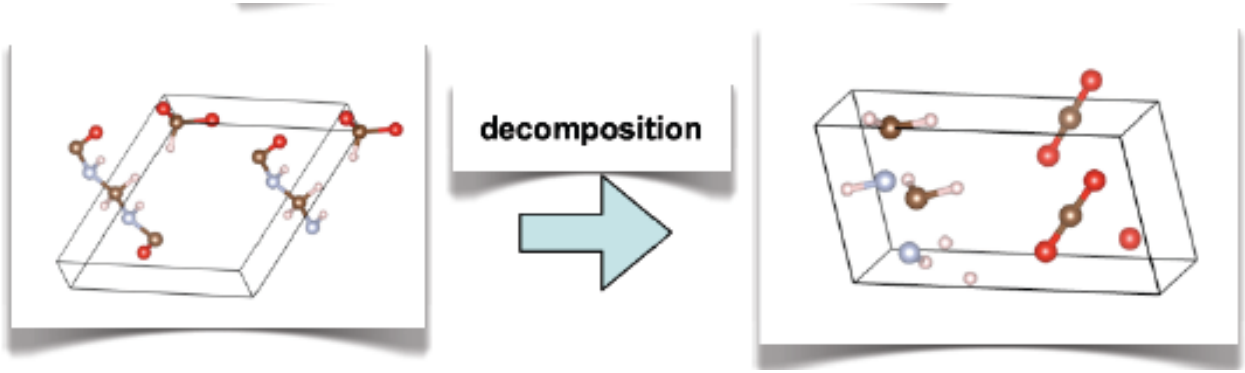
Van der waals bond





Molecular Crystals

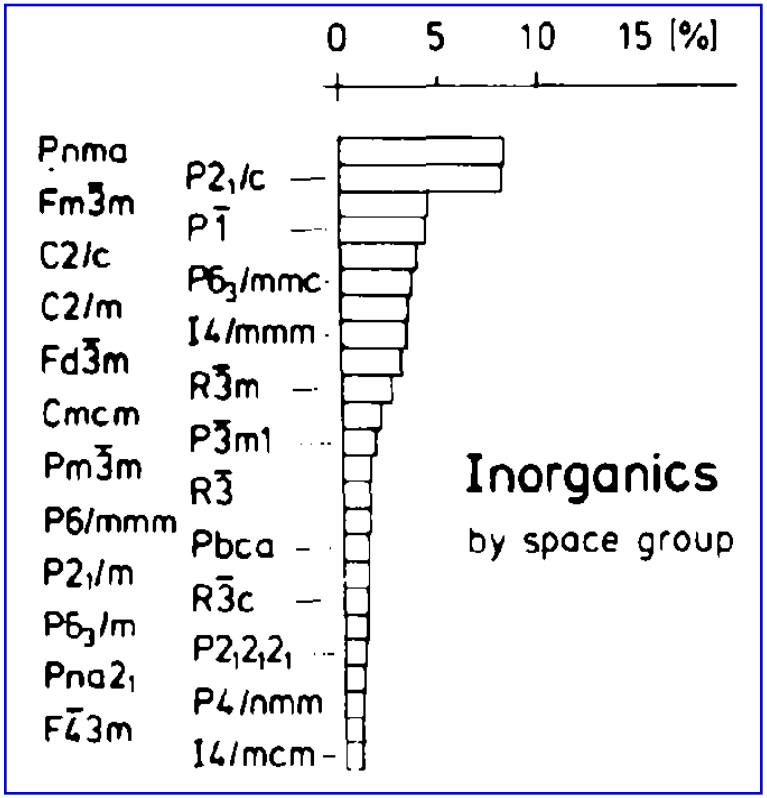
3. Metastability

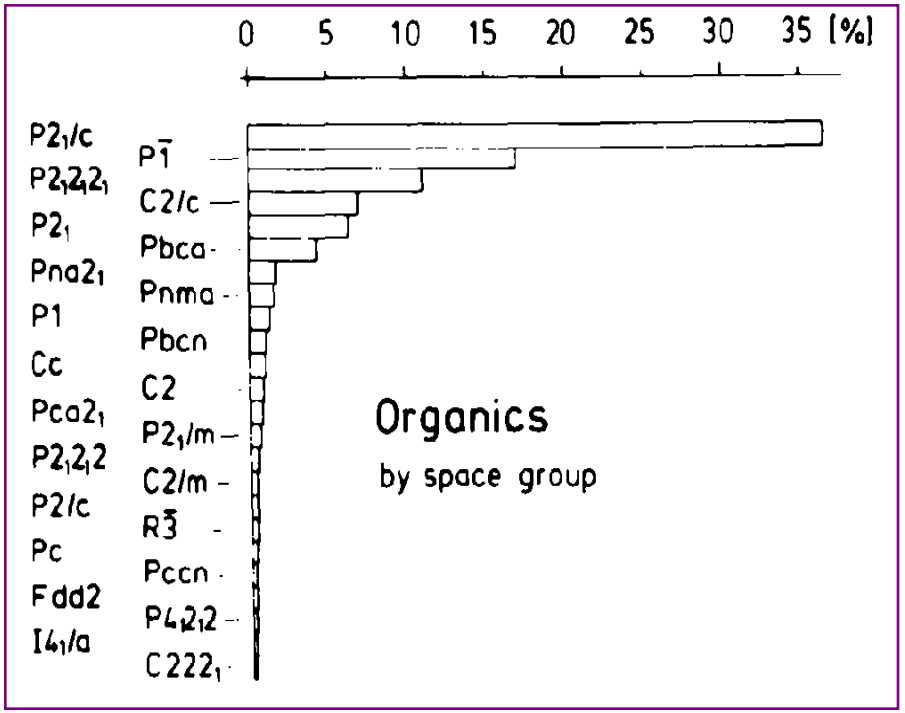


Molecular form is not always favored by energy !

Molecular Crystals

4. Symmetry Preference



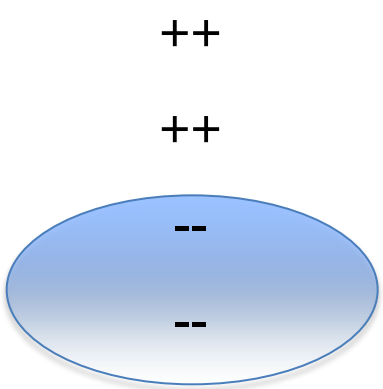


Molecular Crystals

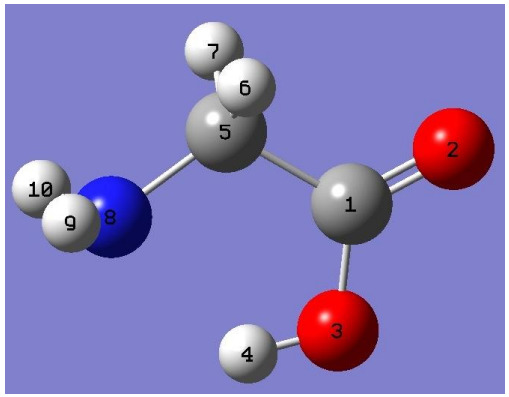
Main feature

- Large amount of atoms
- Sparse space
- Metastability
- Symmetry preference

Search space



Methodology



Apply Constraints

Advantage:

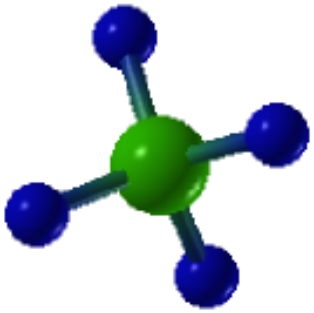
- avoid decomposition
- decrease search space

Type	Degree of freedom
Unconstrained:	3*10
Rigid:	6
Partly flexible:	6~3*10

Constrained Global Optimization

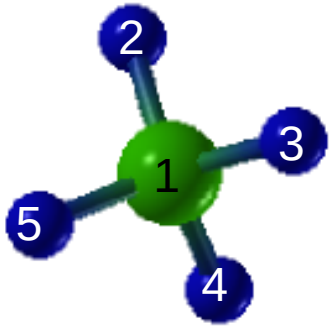
Z-matrix

description of each atom in a molecule in terms of its atomic number, **bond length, bond angle, and dihedral angle** (so-called **internal coordinates**).



Cartesian

```
C      0.000000      0.000000      0.000000
H      0.000000      0.000000      1.089000
H      1.026719      0.000000     -0.363000
H     -0.513360     -0.889165     -0.363000
H     -0.513360      0.889165     -0.363000
```

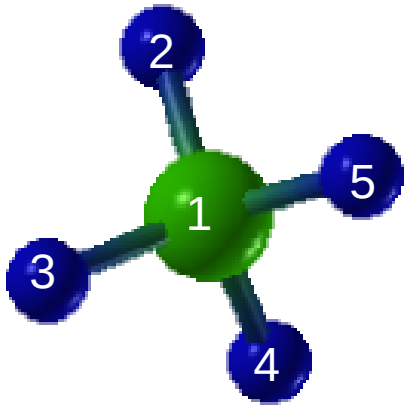


Z-matrix

```
C
H    1 1.089000
H    1 1.089000   2   109.4710
H    1 1.089000   2   109.4710   3   120.0000
H    1 1.089000   2   109.4710   3  -120.0000
```

Define molecule in USPEX

Species	X	Y	Z	i	j	k	ift
C	0.00000	0.00000	0.00000	0	0	0	1
H	0.63512	0.63512	-0.63512	1	0	0	1
H	-0.63512	-0.63512	-0.63512	1	2	0	1
H	0.63512	-0.63512	0.63512	1	2	3	0
H	-0.63512	0.63512	0.63512	1	2	3	0



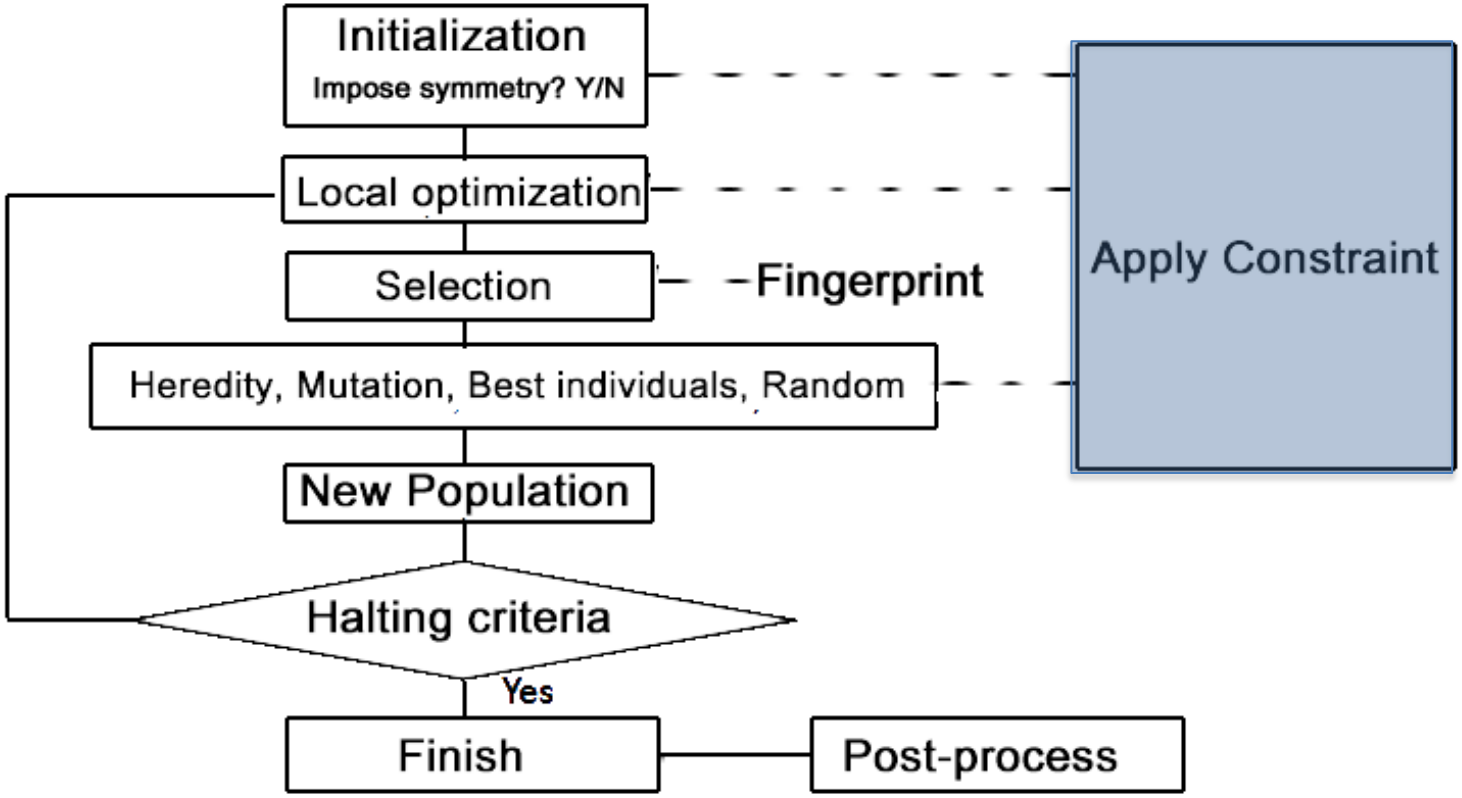
Species

X, Y, Z
Cartesian coordinate of the piece of molecule.

i, j, k
Topological relation in Zmatrix style.

ift
Optimization flag for torsion angle

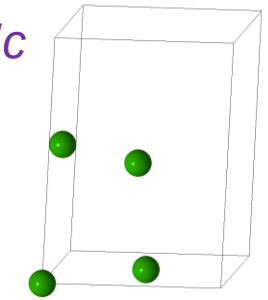
Constrained evolutionary algorithm



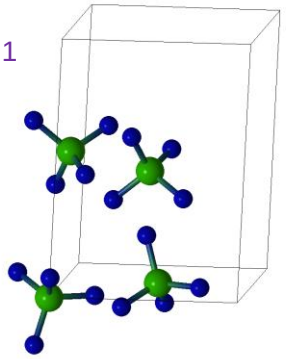
Symmetrization-
random structure with random space group

1. Build the wyckoff positions
2. Track the symmetry operations
   (Point group + Translation)
3. Construct the whole molecule

$P2_1/c$
2a
2d

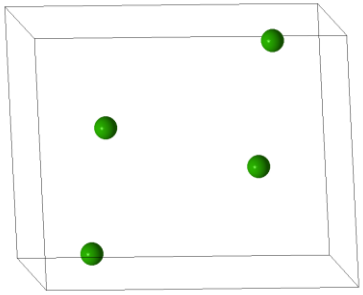


$P2_1$



14: $P2_1/c$, 4 molecules/cell

4e

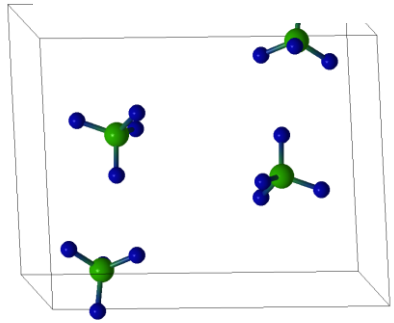


Point group
2/m

Translation

(x, y, z)	(0, 0, 0)
(-x, y,-z)	(0,1/2,1/2)
(-x,-y,-z)	(0, 0, 0)
(x,-y, z)	(0,1/2,1/2)

$P2_1/c$



Don't completely rely on symmetry !

Local optimizer

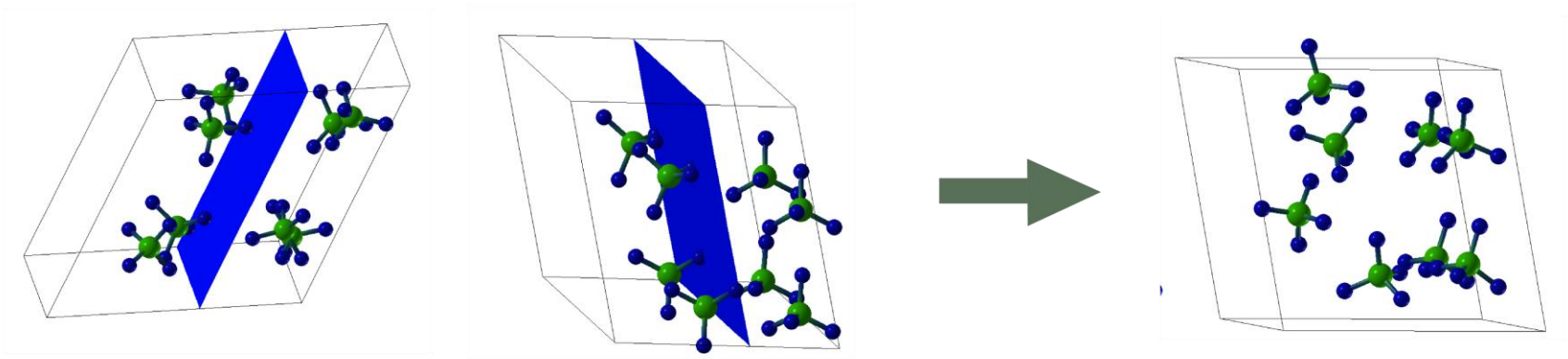
1. Constrained optimization: Zmatrix (DMACRYS, SIESTA)

2. Van der Waals correction: (VASP, SIESTA),

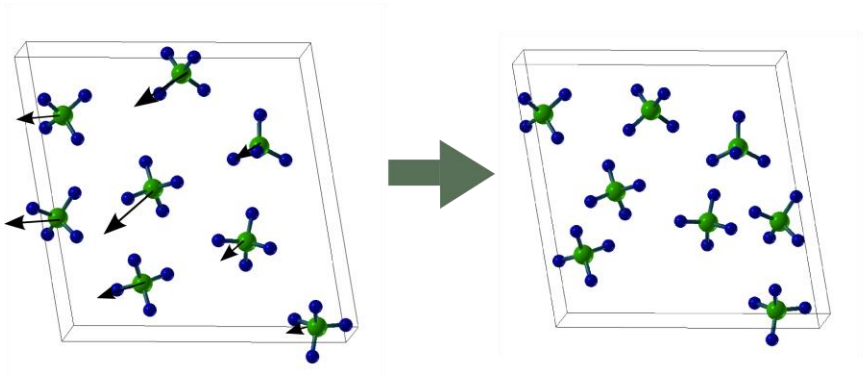
- DFT -D $E_{\mathrm{KS}} + E_{\mathrm{dispersion}}$ Latest version, DFT-D3, Grimme, JCP, 2010
- Random Phase Approximation, Li et al, JPC, 2010
- vdW-Functional, originally proposed by Dion et al, PRL, 2004; significantly speeded up by Kilimes, et al, PRB, 2011

Level	Skip it	Final stages (without Zmatrix)
Classical		DMACRYS+GULP
Ab-initio		VASP

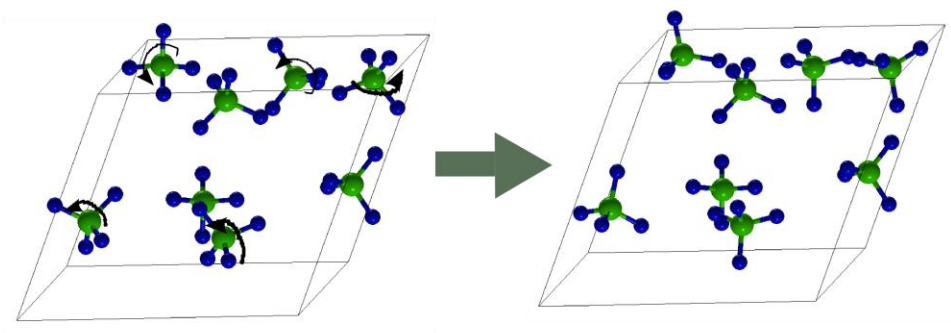
Variation Operators



(a) heredity

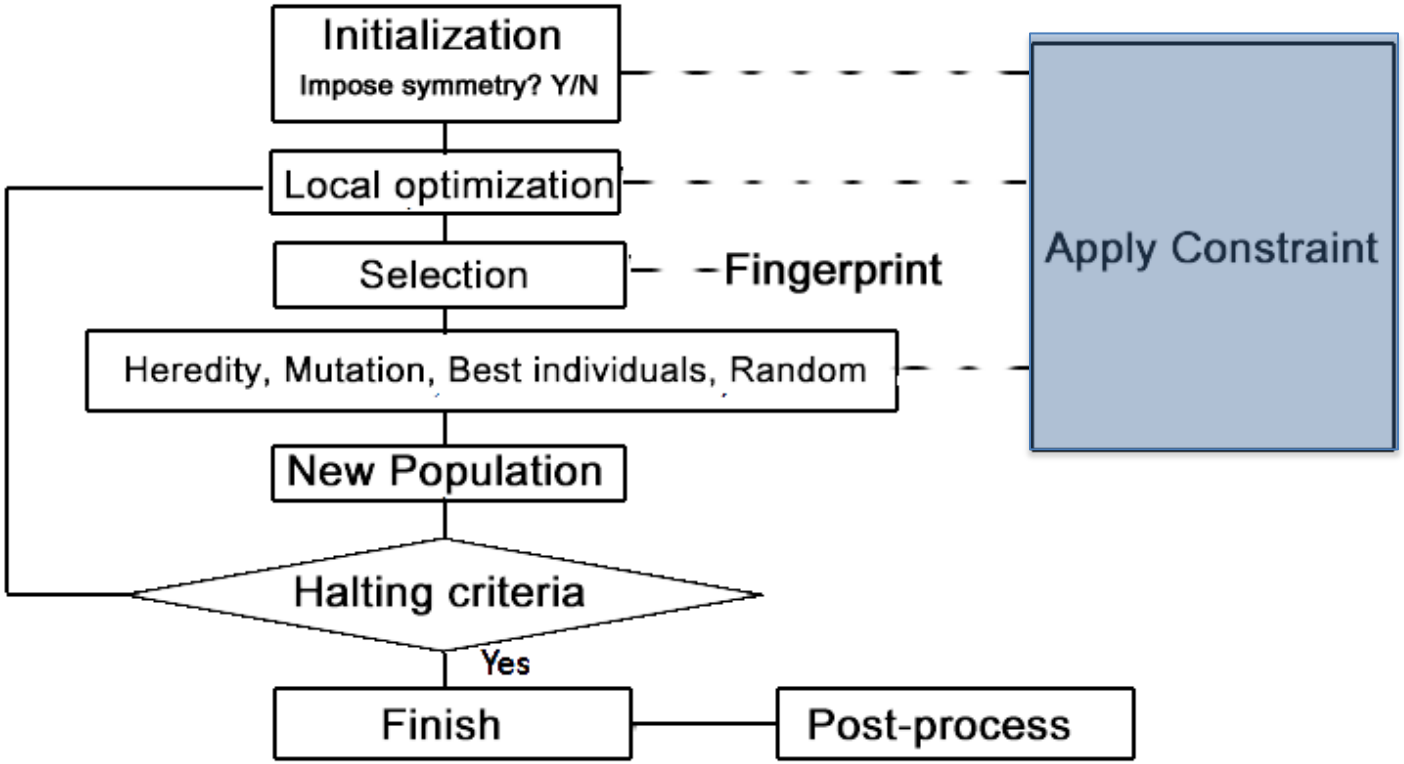


(b) coordinate mutation



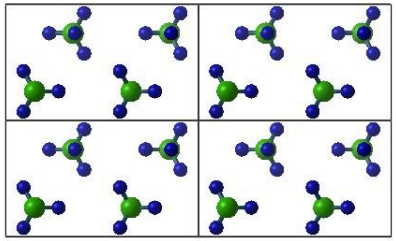
(c) rotational mutation

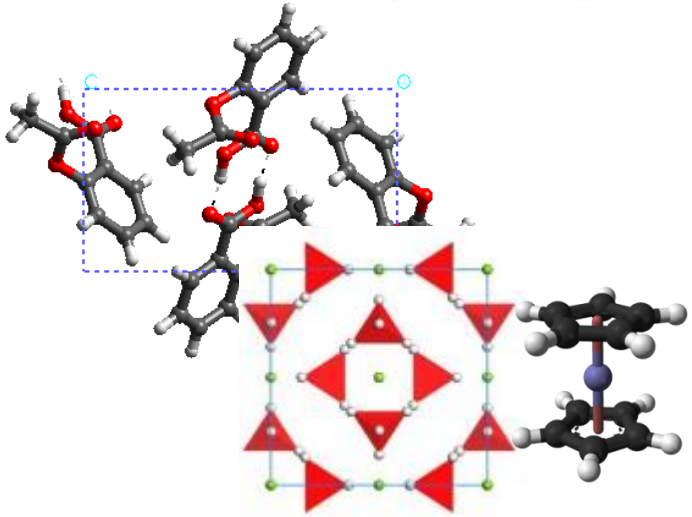
Constrained evolutionary algorithm



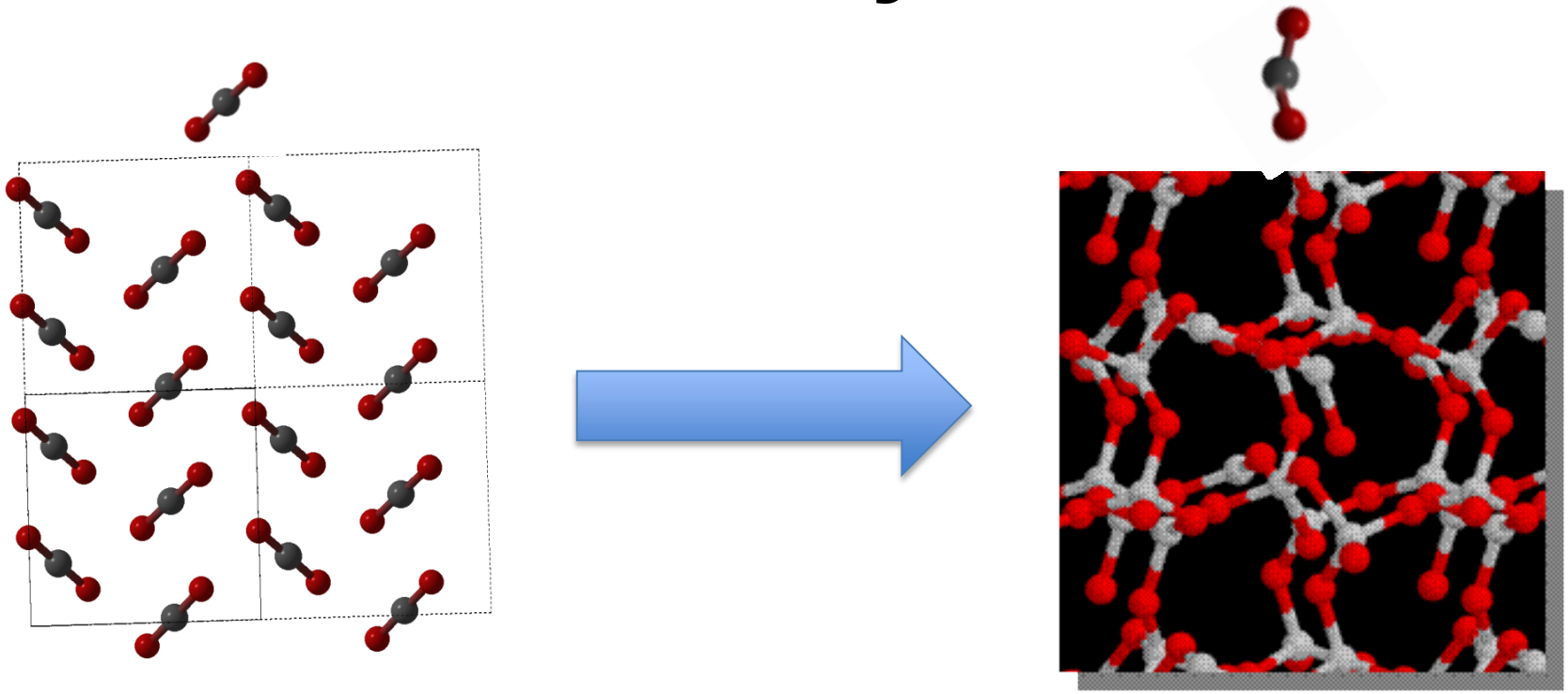
Application

- Simple molecules at extreme conditions
  - H_2O, NH_3, CH_4, CO_2, C_6H_6
- Pharmaceutical compounds
  - glycine, aspirin
- Organo-metallic complex
  - ferrocene
- Polymers
- Inorganic crystals
  - $M(BH_4)x$





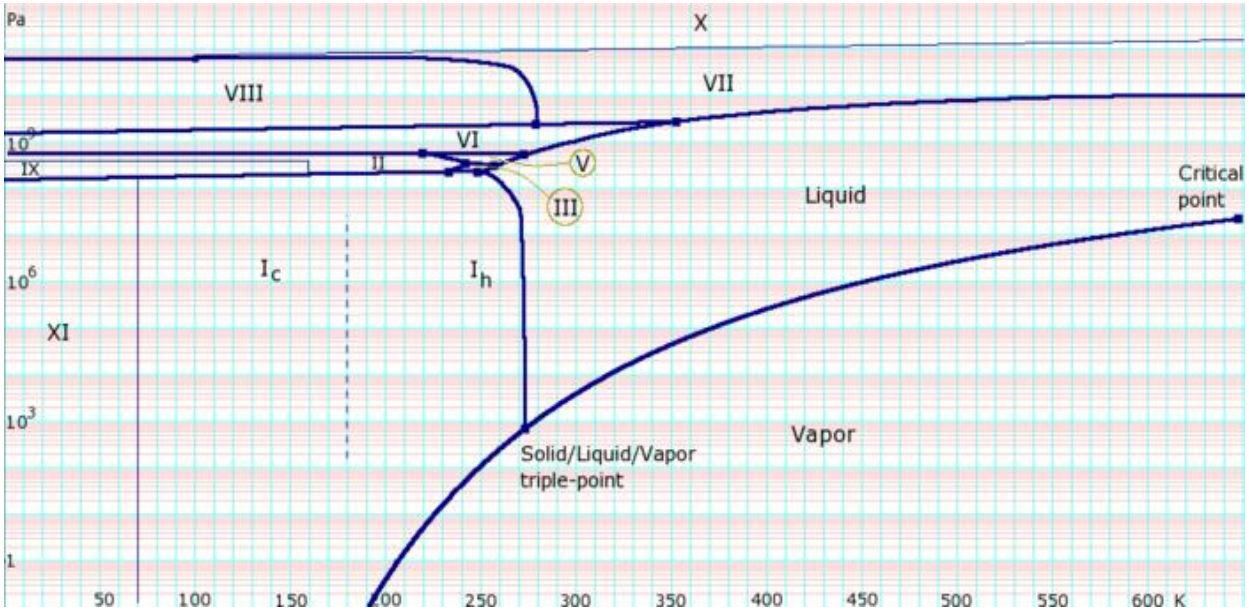
Extended Systems

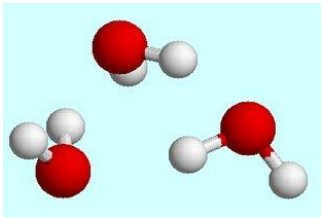


Molecular Packing

Polymerization

Ice





Strong H bonding

Bent molecule

USPEX input:

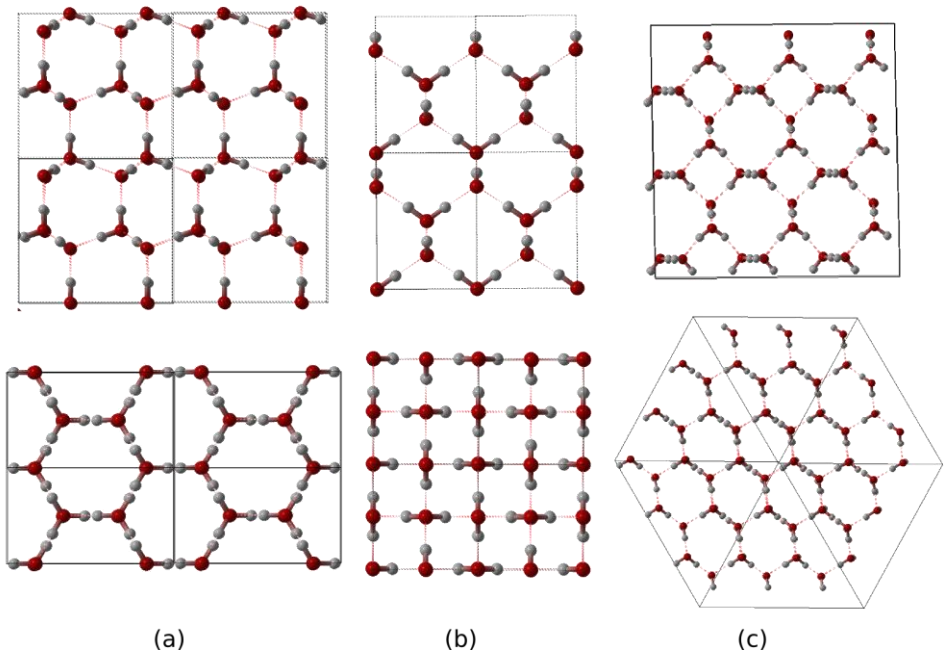
Z-matrix of H_2O

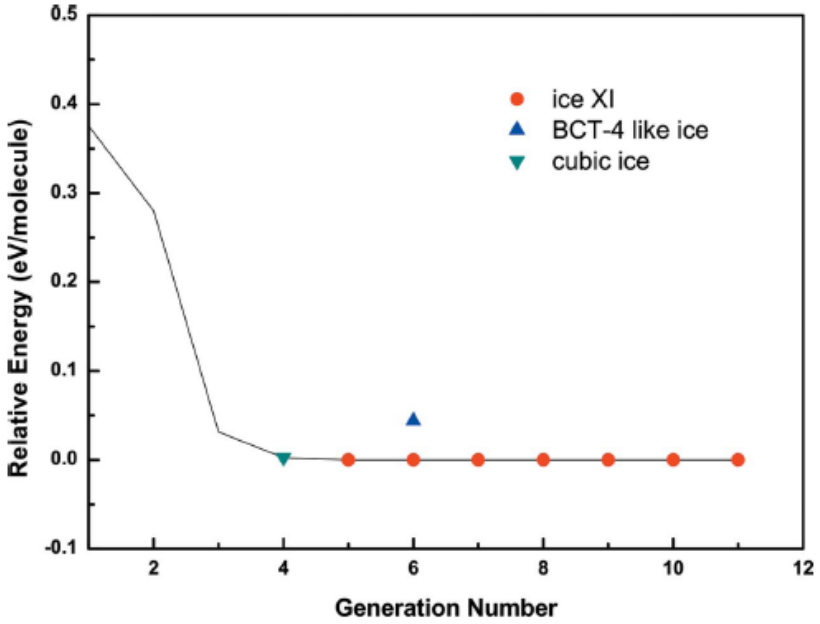
Siesta Input file

4 molecules/cell

Target Pressure: 0 GPa

Ice at ambient pressure





Ice:	Ice-XI	Cubic ice	Tetragonal	**Umemoto, PRL , 2010**
Carbon:	lonsdaleite	diamond	BCT-4	**Zamann, PCCP, 2011**

One single calculation: < 1 day in 8 core Dell XPS Desktop !!

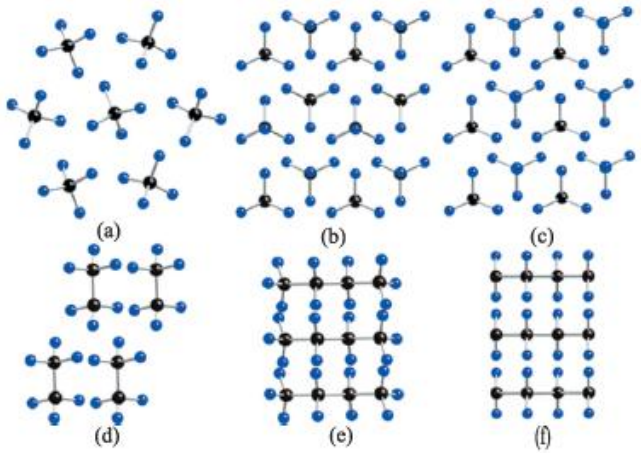
Methane

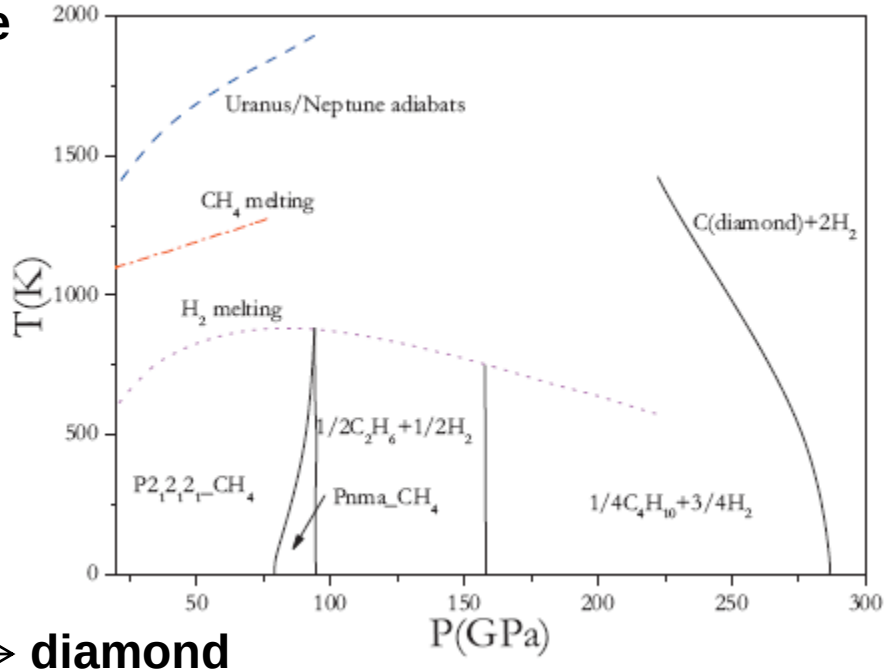
THE JOURNAL OF CHEMICAL PHYSICS 133, 144508 (2010)

Dissociation of methane under high pressure

Guoying Gao,[1,a)] Artem R. Oganov,[2,a)] Yanming Ma,[1,b)] Hui Wang,[1] Peifang Li,[1] Yinwei Li,[1] Toshiaki Iitaka,[3] and Guangtian Zou[1]

Predicted high pressure phases of methane





A subsequent dissociation:

Methane ⟶ ethane ⟶ butane ⟶ diamond

Methane

vdW bonding

tetrahedra molecule

USPEX input:

3~8 molecules/cell

Target Pressure: 20~200GPa

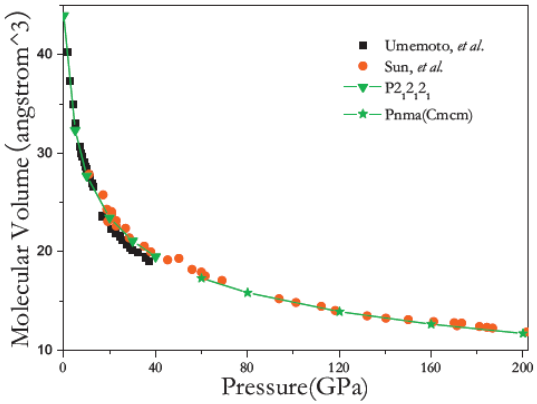
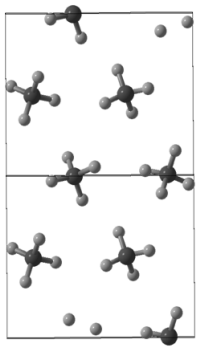
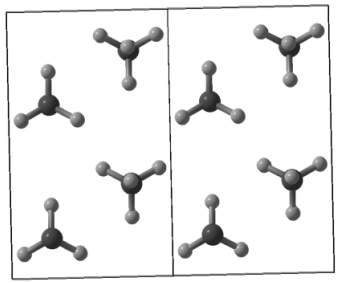


FIG. 1. The calculated equation of state compared with the experimental data (Refs. 4 and 12).

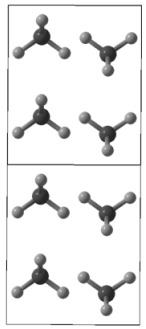
Gao,, J Chem. Phys, 2010



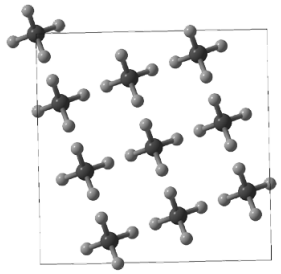
***P*2₁2₁2₁(20-58 GPa)**



***Pnma* (58~138 GPa)**

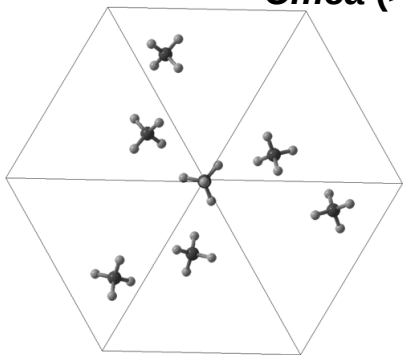


***Cmca* (> 138 GPa)**



***I*-4**

5 molecules/ primitive cell



R*-3*R

7 molecules/ primitive cell

Ammonia

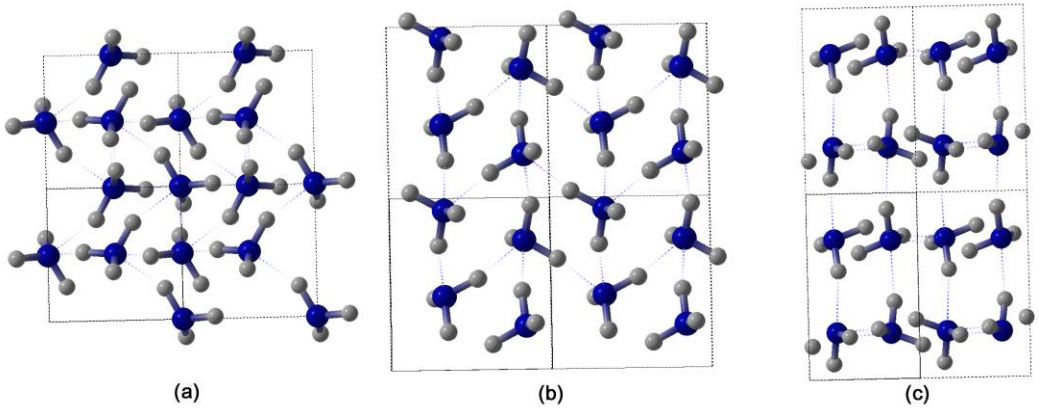
NH3, weak H bonding

Pseudo tetrahedral

USPEX input:

4 molecules/cell

Target Pressure: 0~100GPa



Ammonia polymorphs found by USPEX. a) *P*213 phase (1 - 4 GPa, Z=4); b) *P*21/c phase (4 - 7 GPa, Z=4); c) *P*212121 phase (7 - 60 GPa, Z=4)

Pickard, Nature Materials, 2008

CO_2

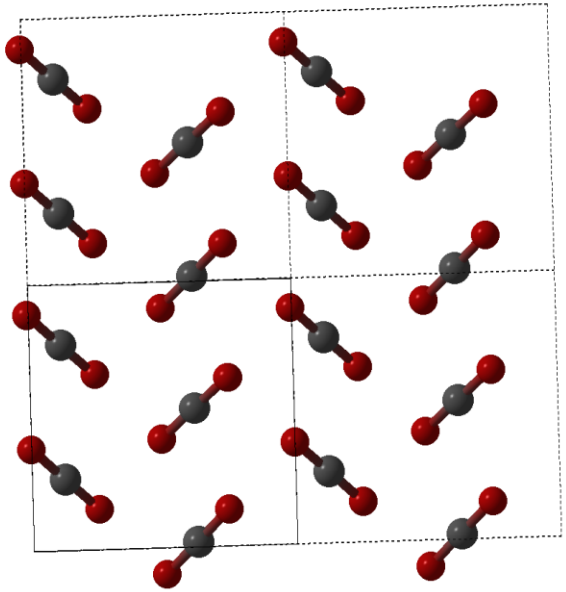
CO_2, vdW bonding

Linear molecule

USPEX input:

4 molecules/cell

Target Pressure: 20~80GPa



CO_2 III (*Cmca*, 4 molecules/cell)

Benzene

C_6H_6, π- π stacking planar molecule

USPEX input:

4 molecules/cell

Target Pressure: 5~20GPa

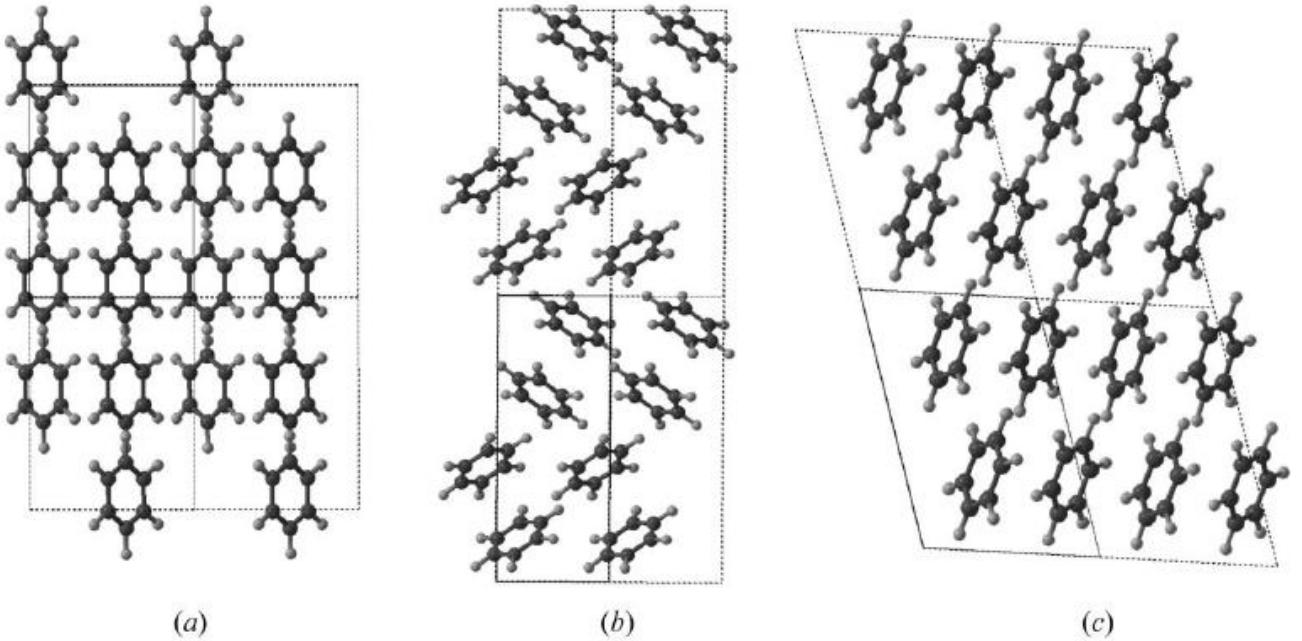
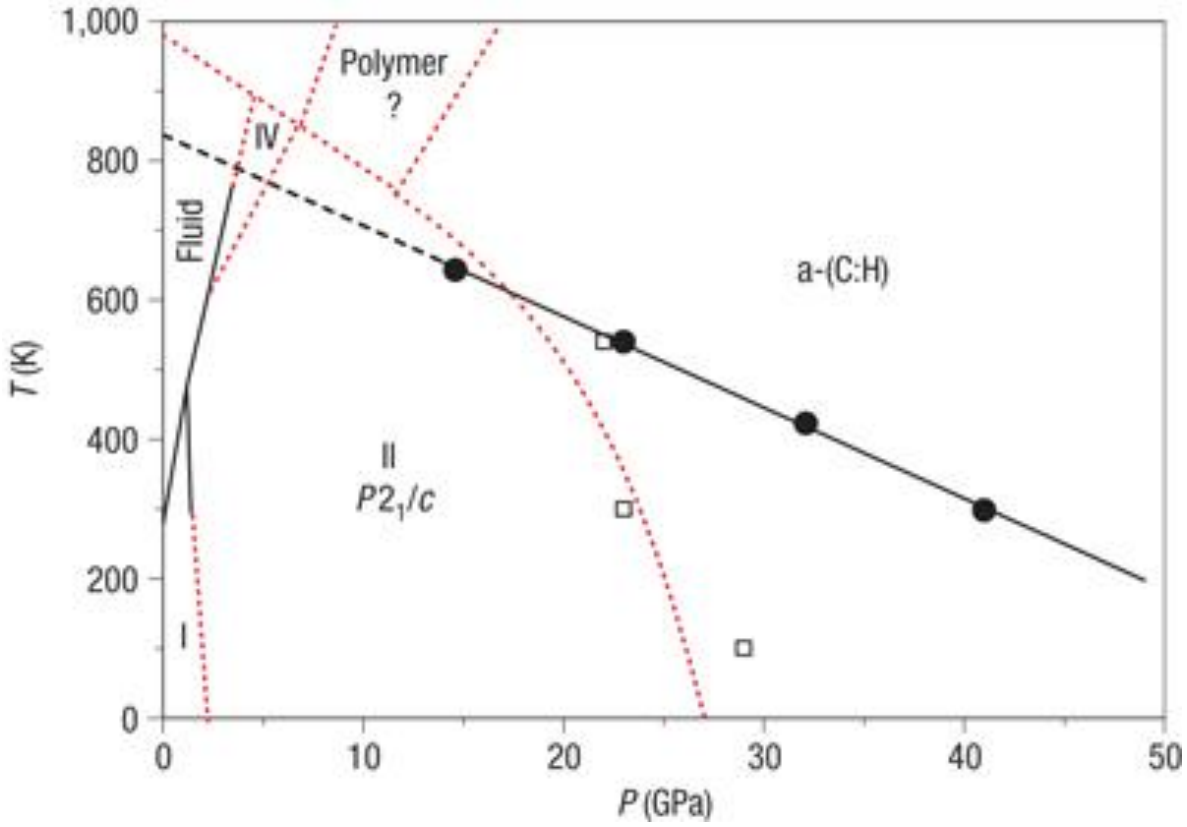


Figure 8
Crystal structures of benzene (*a*) orthorhombic phase (I) (*Pbca*, $Z = 4$); (*b*) tetragonal phase (II) ($P4_32_12$, $Z = 4$); (*c*) monoclinic phase ($P2_1/c$, $Z = 2$).

Benzene

C_6H_6, π- π stacking planar molecule

USPEX input:
4 molecules/cell
Target Pressure: 5~20GPa



Ciabini, et al, Nature Mat., 2006

Glycine

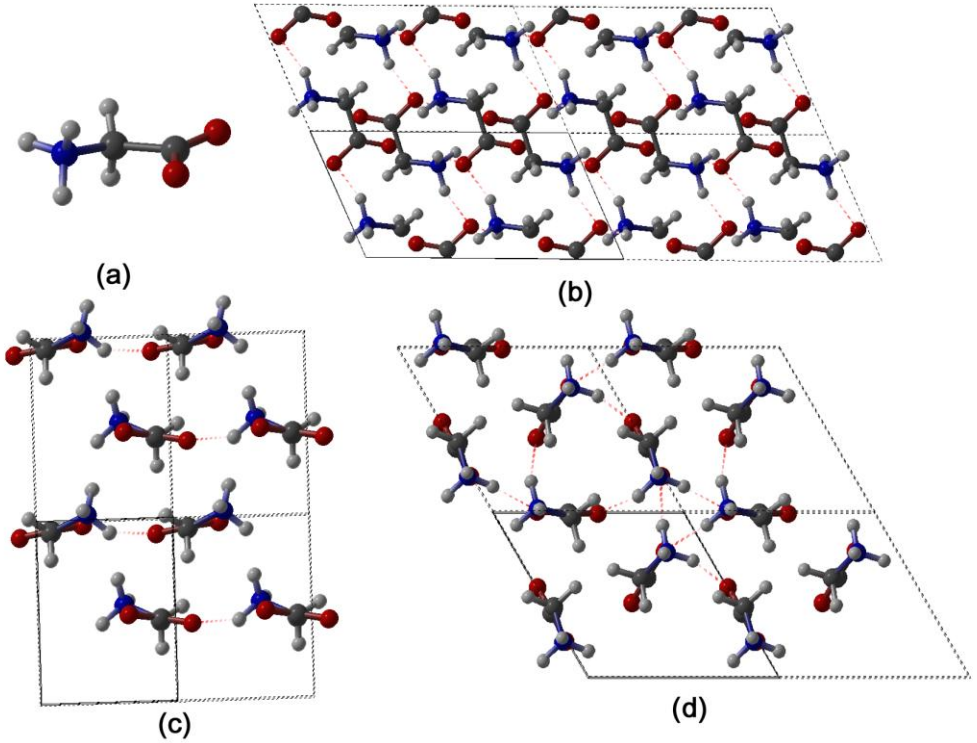
$C_2H_5NO_2$, vdW + H bonding

Amino acid molecule

USPEX input:

2, 3, 4 molecules/cell

Target Pressure: 0~2 GPa

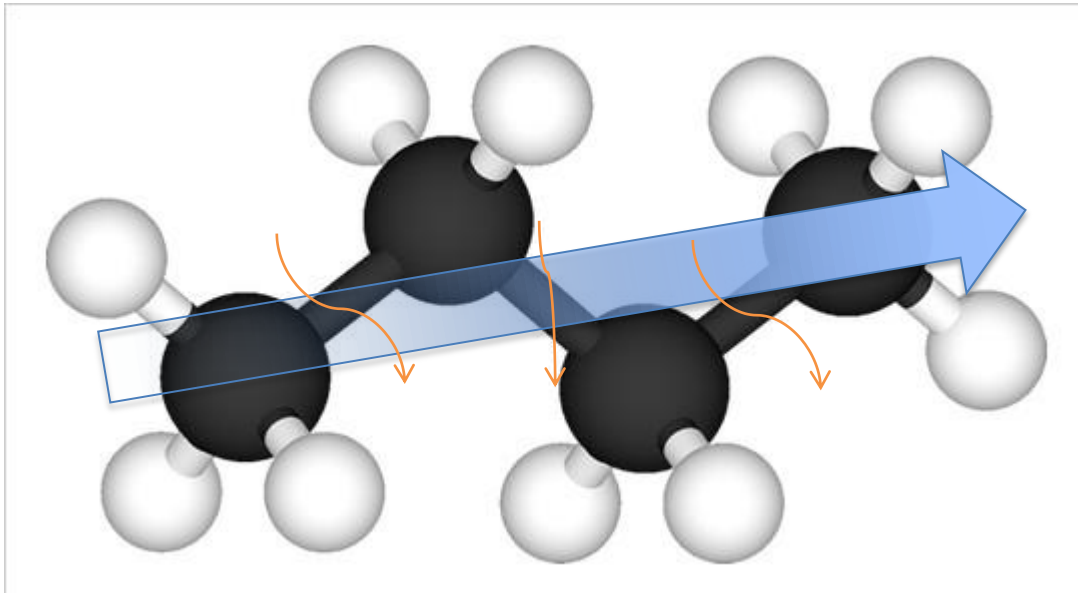


a) representation of glycine zwitterion; b) α glycine at 2 GPa; c) β-glycine at 0.4 GPa; d) γ-glycine at 1 GPa.

DFT+D: α < γ < β
Exp: γ < α < β
VdW-DF: γ < α < β

DFT+D could not give the right energy ranking for glycine polymorphs !!!!!!

Improved rotational mutation



1. Principle axis from inertia tensor

$$I = \sum_{i=1}^{N} m_i r_i^2 \ .$$

$$[I_C^B] = [Q][\Lambda][Q^T],$$

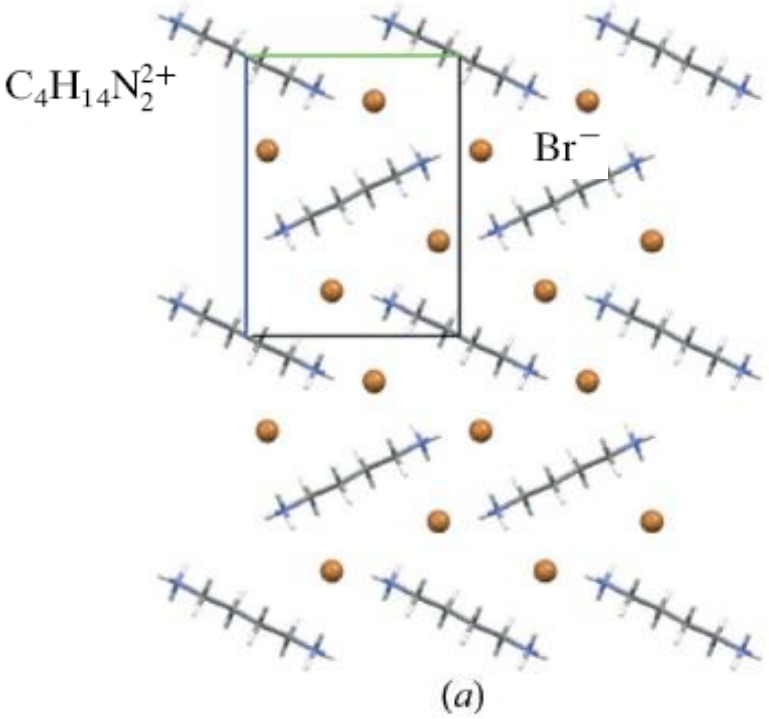
where

$$[\Lambda] = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix}.$$

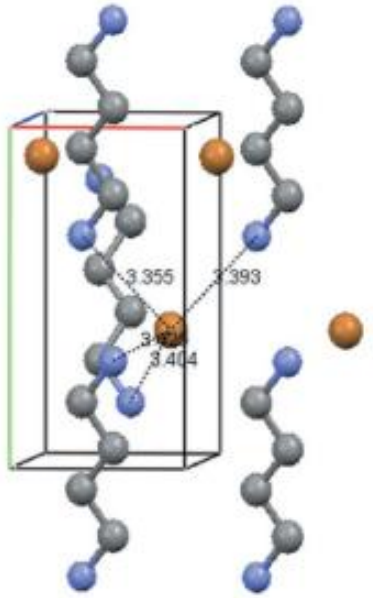
2, Flexible molecules,
allowing the variation of dihedral angles

Butane-1,4-diammonium dibromide

Flexible molecules are also applicable !



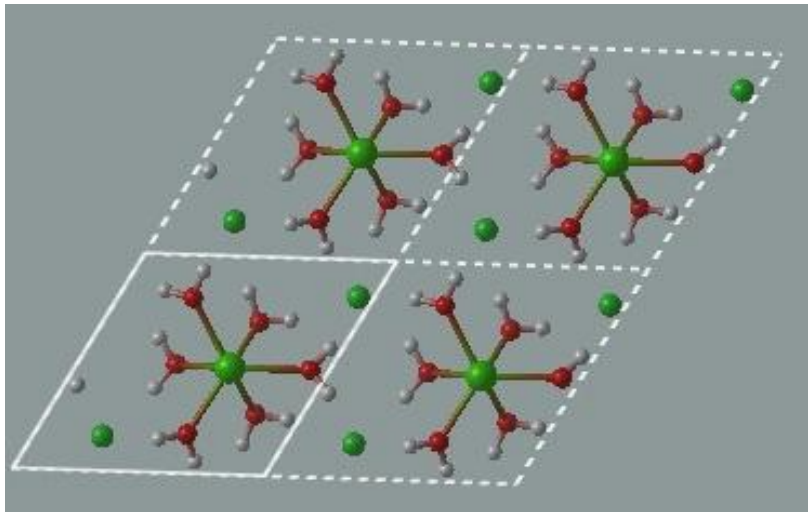
(a)



(b)

From organic to inorganic

Some organic crystals can also be viewed as the packing of well defined motifs



Ca(H2O)6 + Cl2

From organic to inorganic

$Mg(BH_4)_2$

USPEX input:

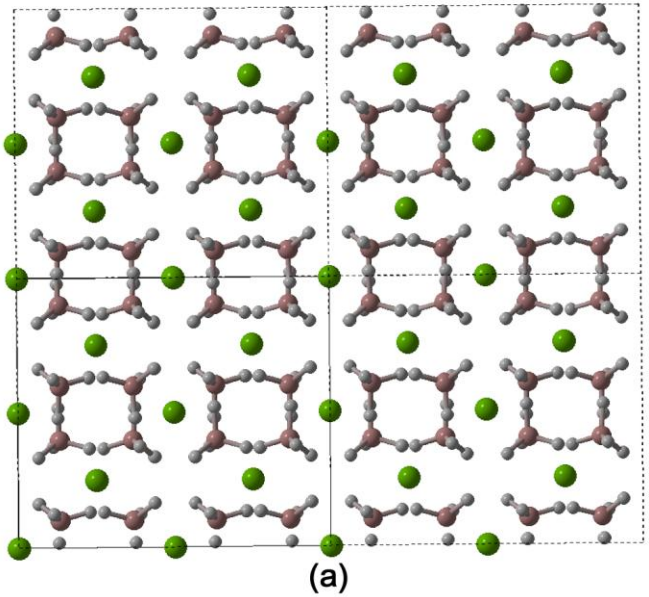
MOL 1: Mg

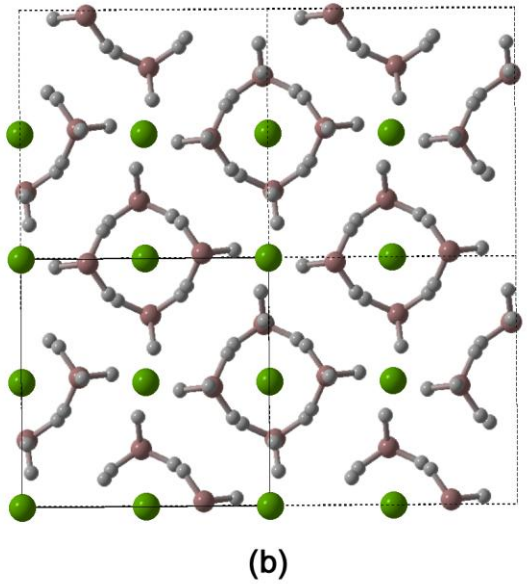
MOL 2: BH_4

4 + 4 molecules/cell

Target Pressure: 0 Gpa

~ 300 structural relaxations





$Mg(BH_4)_2$ polymorphs found by USPEX. a) *F*222 phase; b) *I*4122 phase.
Ozolins V., et al, PRL, 2008
Zhou X.F., et al, PRB, 2009
Pnma-LiBH4 was also trivially identified in 60 atoms /cells

Stony Brook University

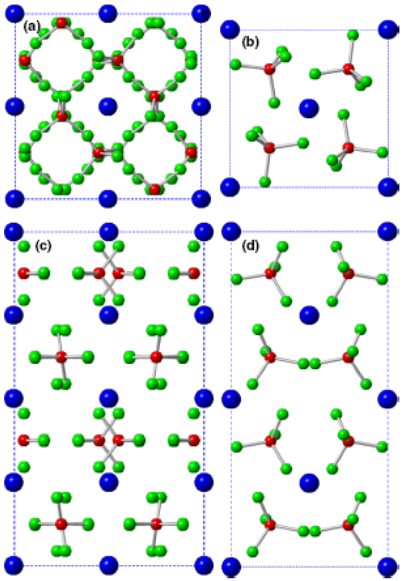
PRL 109, 245503 (2012) PHYSICAL REVIEW LETTERS week ending 14 DECEMBER 2012

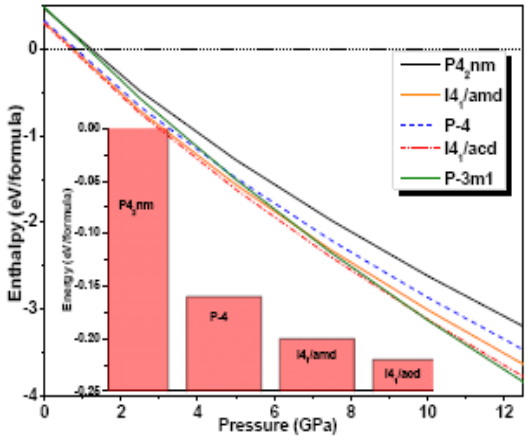
First-Principles Determination of the Structure of Magnesium Borohydride

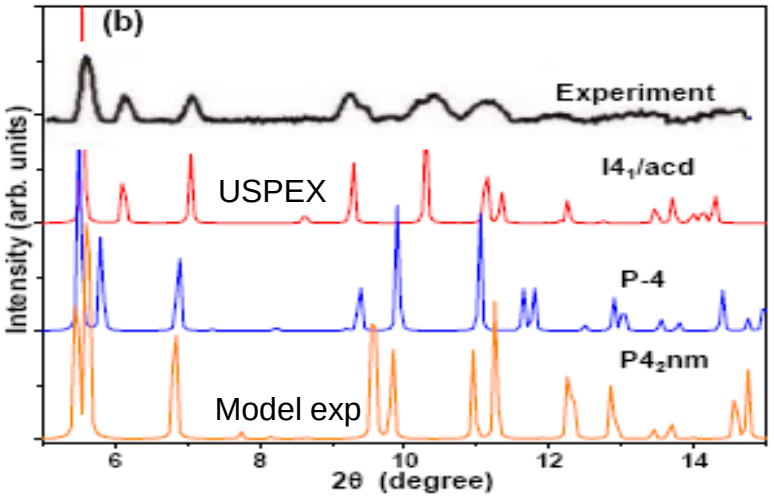
Xiang-Feng Zhou,[1,2,*] Artem R. Oganov,[2,3] Guang-Rui Qian,[2] and Qiang Zhu[2]

[1]*School of Physics and Key Laboratory of Weak-Light Nonlinear Photonics, Nankai University, Tianjin 300071, China*
[2]*Department of Geosciences, Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA*
[3]*Geology Department, Moscow State University, Moscow 119992, Russia*

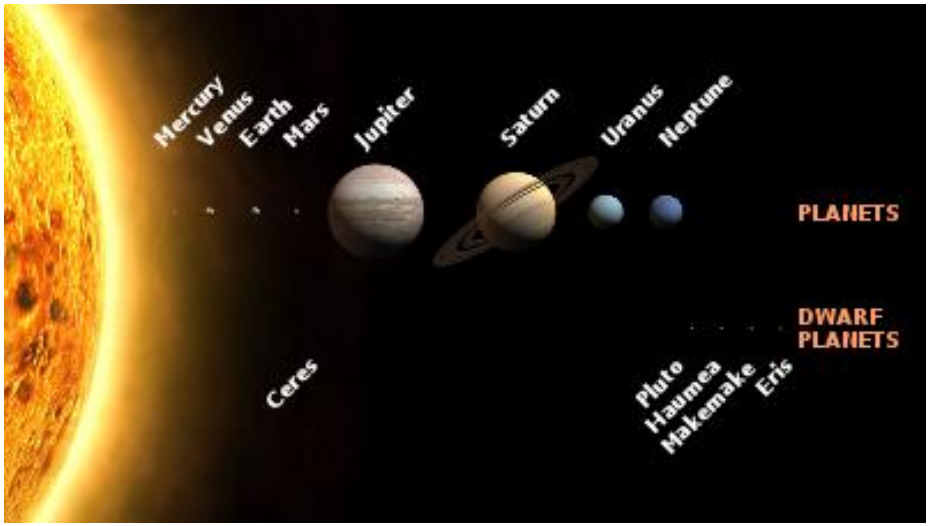
(Received 2 October 2012; published 10 December 2012)

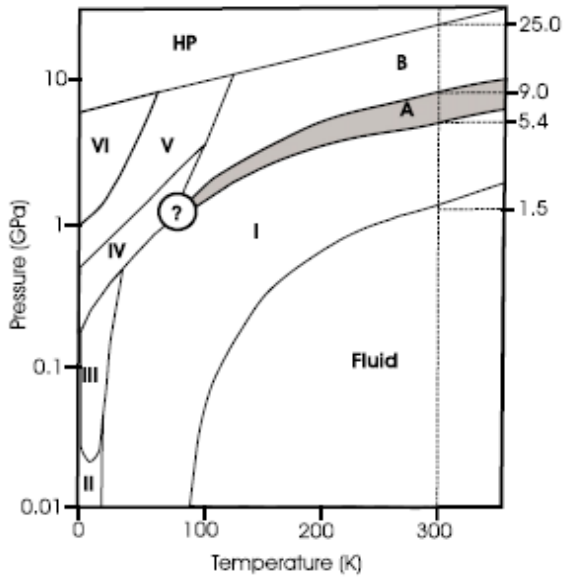






Methane under pressure

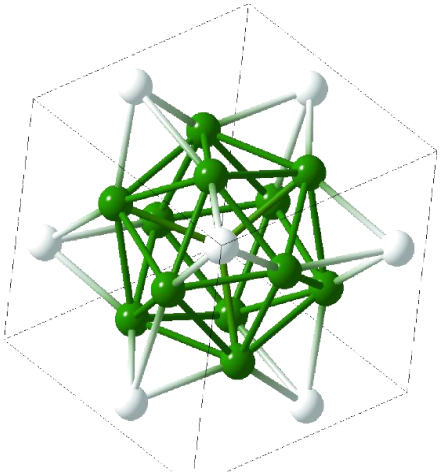




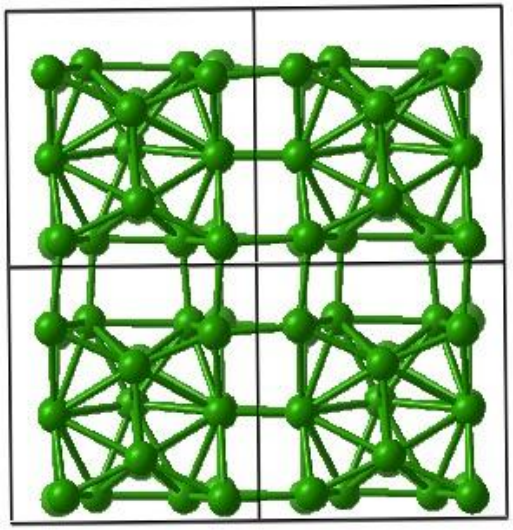
21 molecules (105 atoms)
Can it be predicable????

Nakahata, CPL, 1999

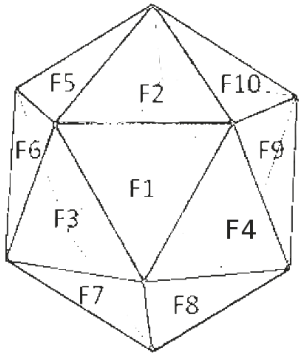
Methane under pressure



(111) view



2*2*2 (100) view

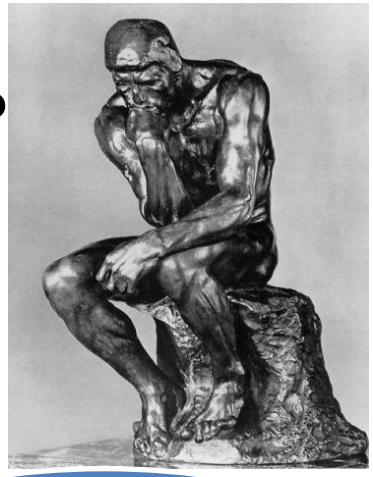


21 = 12 + 1 + 8

12: icosahedron site
1 : icosahedron center
8 : outside icosahedron

OUTLOOK

Molecular Crystals – Prediction?



Ab initio or empirical ??

Reliable

Ranking

Searching

Universal

Transferable

Economical

Computing time

Automatic

User friendly

vdW bonding
- DFT+D
- VDW -DF

EA

No bias introduced

Least given information

Reference

feature articles

Acta Crystallographica Section B
Structural
Science

ISSN 0108-7681

Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications

Qiang Zhu,[a]* Artem R. Oganov,[a,b] Colin W. Glass[c] and Harold T. Stokes[d]

[a]Department of Geosciences, Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York, USA, [b]Geology Department, Moscow State University, Moscow, Russia, [c]High Performance Computing Center Stuttgart (HLRS), Germany, and [d]Department of Physics and Astronomy, Brigham Young University, Provo, Utah, USA

Evolutionary crystal structure prediction proved to be a powerful approach for studying a wide range of materials. Here we present a specifically designed algorithm for the prediction of the structure of complex crystals consisting of well defined molecular units. The main feature of this new approach is that each unit is treated as a whole body, which drastically reduces the search space and improves the efficiency, but necessitates the introduction of new variation operators described here. To increase the diversity of the population of structures, the initial population and part (~ 20%) of the new generations are produced using space-group symmetry combined with random cell parameters, and

Received 6 February 2012
Accepted 19 April 2012

Conclusion

1. Based on USPEX , we succeeded in developing a new efficient approach to predict the structure of molecular crystals
2. It can find the stable for the systems with various geometry shape (tetrahedral, linear, bent, planar and regular molecules), and chemical bonding (vdW bonding, weak and strong hydrogen bonding, π-π stacking, etc)
3. We also extended this method to the prediction of inorganic complexes (if one wants to fix some bond connectivity).
4. Variable compositional prediction for molecular compounds (Monday)
5. 1D polymers crystal (Wednesday).

Acknowledgement

- A. R. Oganov
- A. O. Lyakhov
- X-F Zhou
- H. T. Stokes