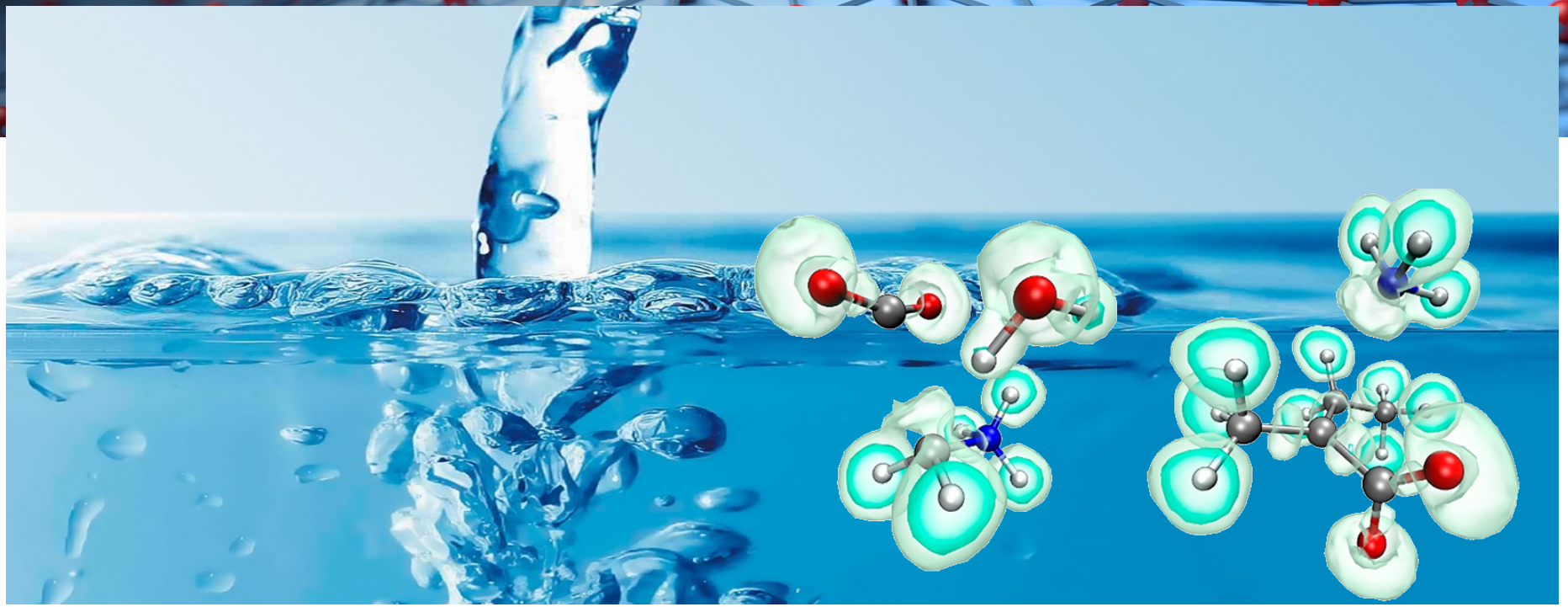
HPC SCHOOL ON QUANTUM COMPUTATIONAL MATERIALS SCIENCE

Nov 18, 2020

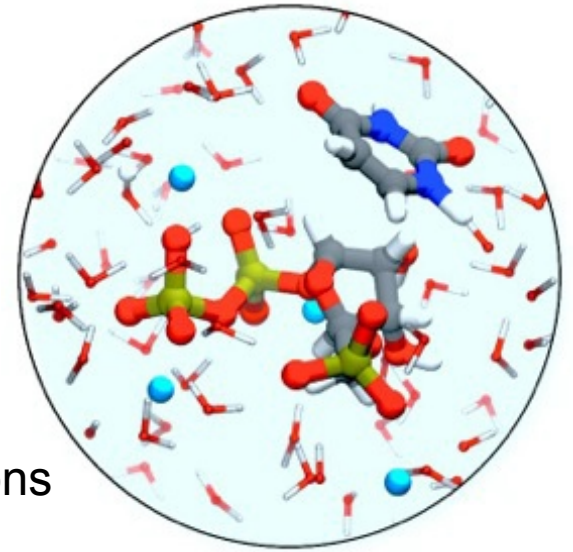


# Free-energy calculation of chemical reactions in solution

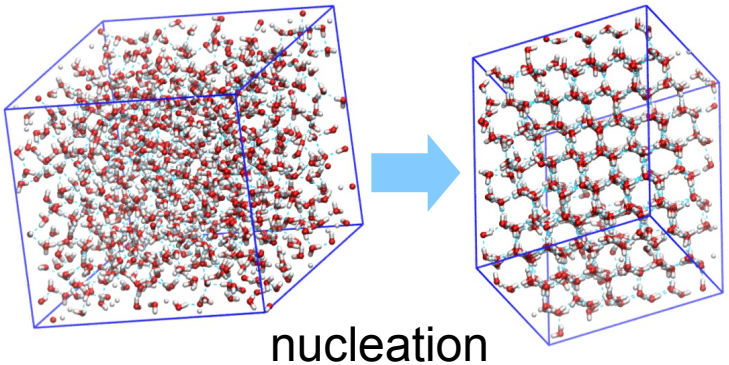
Fabio Pietrucci - Sorbonne Université, IMPMC

***objective:***

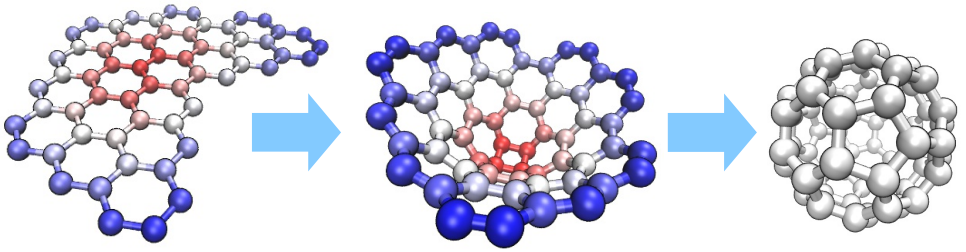
study the mechanism, thermodynamics and kinetics of the transformations of matter



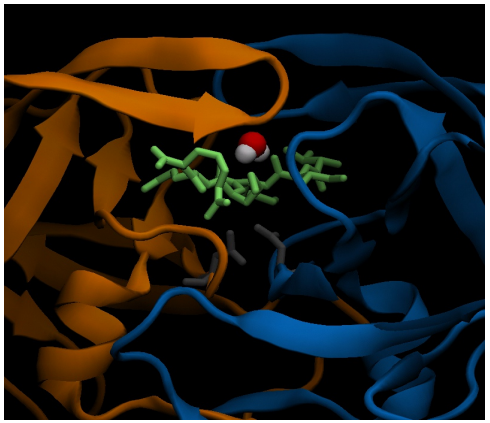
reactions



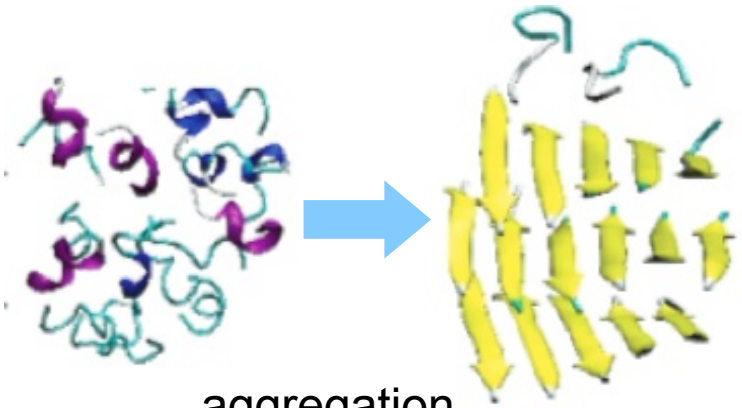
nucleation



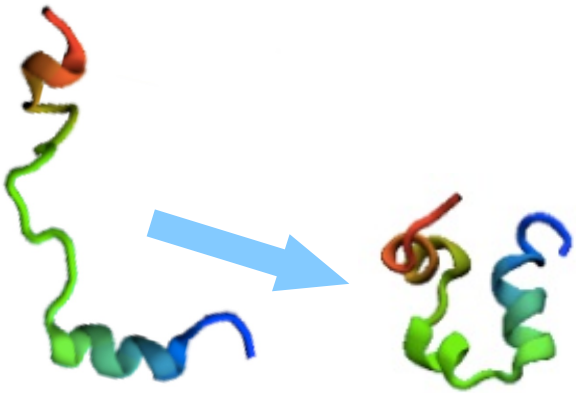
nano-transformations



drug binding



aggregation



folding

## *why is it important to simulate transformations?*

1) transition pathways are evanescent, hard to catch in experiments

2) access to free-energy landscapes and kinetic rates

3) gap between *in silico* predictions (e.g., crystal structures) and empirical “recipes” for synthesis

**the fundamental puzzle: thermodynamics vs kinetics**

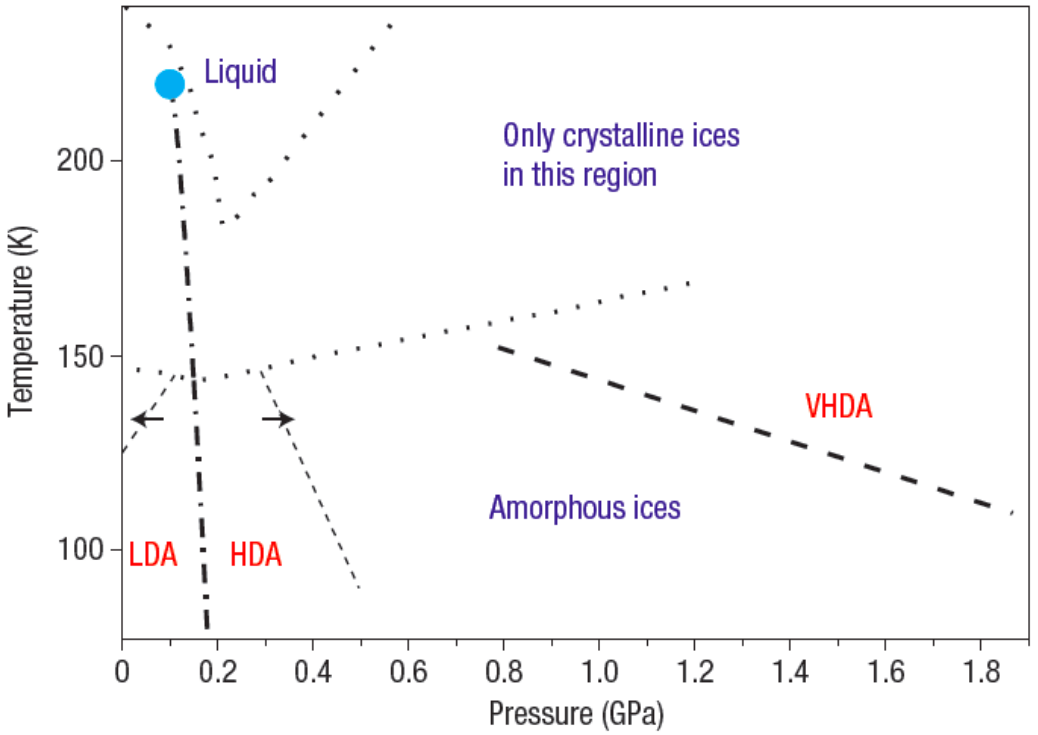
*Rev. Phys.* 2, 32 (2017)

nature physics 2, 414 (2006)

# Annealed high-density amorphous ice under pressure

RICHARD J. NELMES[1]*, JOHN S. LOVEDAY[1]*, THIERRY STRÄSSLE[2†], CRAIG L. BULL[1], MALCOLM GUTHRIE[1], GÉRARD HAMEL[2] AND STEFAN KLOTZ[2]

We find that HDA annealed at 0.2–0.3 GPa to 130 K is like the amorph obtained by compressing LDA at 130 K, and is not VHDA. We have recovered this annealed HDA to ambient pressure at low temperature, and warmed it up to and through the transition to LDA, and observe an abrupt transition at ~128 K rather than the quasi-continuous transformation at ~110 K obtained with unannealed HDA and with VHDA.



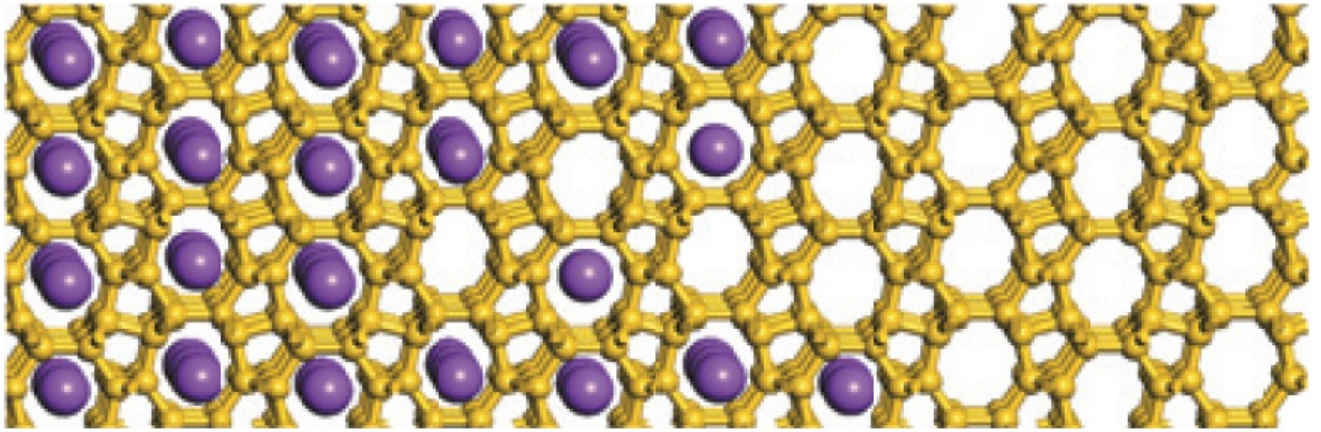
nature materials

LETTERS

PUBLISHED ONLINE: 17 NOVEMBER 2014 | DOI: 10.1038/NMAT4140

# Synthesis of an open-framework allotrope of silicon

Duck Young Kim[1†], Stevce Stefanoski[1†], Oleksandr O. Kurakevych[1,2†] and Timothy A. Strobel[1*†]



$Na_4Si_{24} \longrightarrow Na_xSi_{24} \longrightarrow Si_{24}$

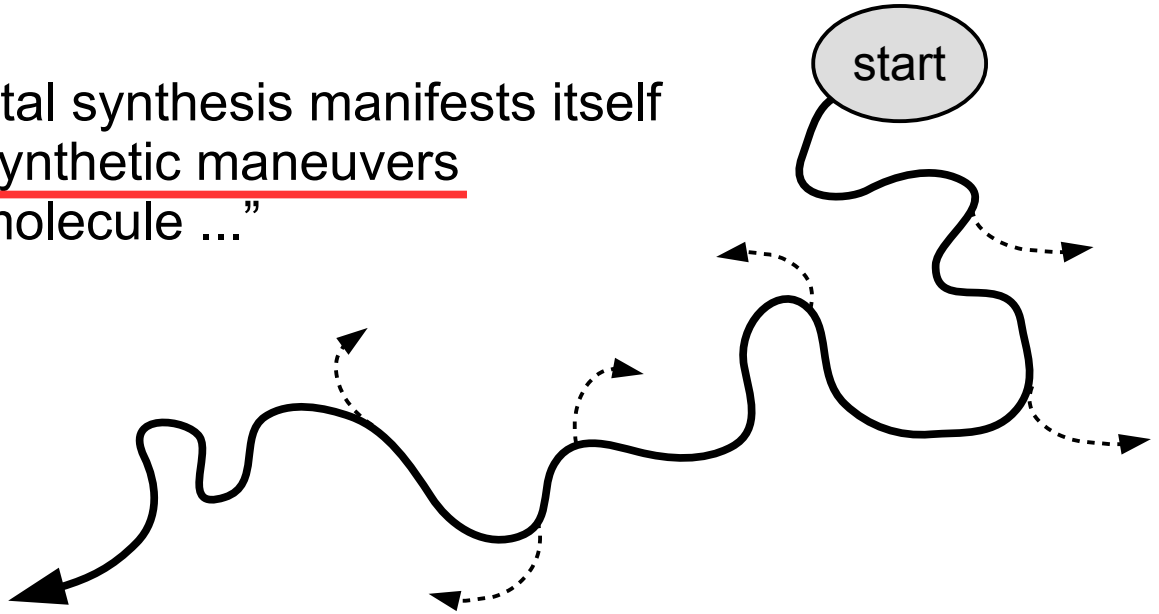
“ In the first step, $Na_4Si_{24}$ was synthesized from a Na/Si mixture with 15 mol% Na. The mixture was ground in a ceramic mortar for one hour ... The mixture was pressurized in a 1500 ton multianvil press at a rate of 10 bar $h^{-1}$ to a pressure of 10 GPa and reacted at 800 °C in two steps: preheating at 400 °C for 30 min ... and reaction at the final temperature for one hour, after which the sample was quenched ... In the second step, polycrystalline agglomerates of $Na_4Si_{24}$ were placed in a furnace under a dynamic vacuum of $10^{-5}$ torr and 'degassed' at 400 K for eight days to obtain the empty $Si_{24}$ structure. ”

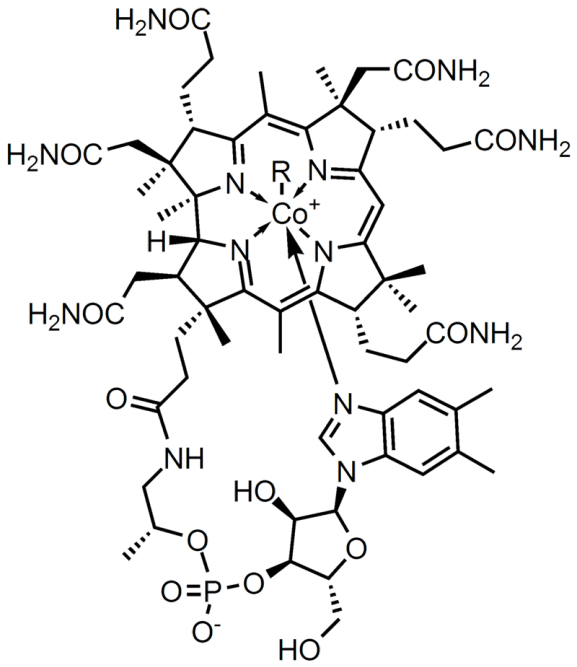
*Chem. Soc. Rev.*, 2009, **38**, 2993–3009

# The art of total synthesis through cascade reactions

**K. C. Nicolaou* and Jason S. Chen**

"The artistic nature of total synthesis manifests itself in the selection of the synthetic maneuvers that lead to the target molecule ..."

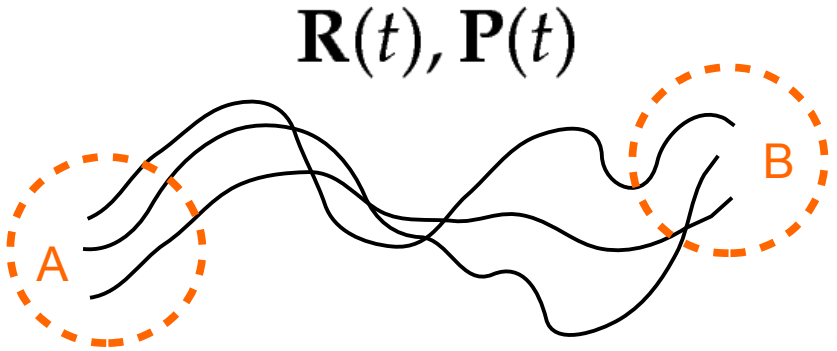



"The complete laboratory synthesis of B12 was achieved by R.B. Woodward and A. Eschenmoser in 1972, and remains one of the classic feats of organic synthesis, requiring the effort of 91 postdoctoral fellows (mostly at Harvard) and 12 PhD students (at ETH) from 19 nations."

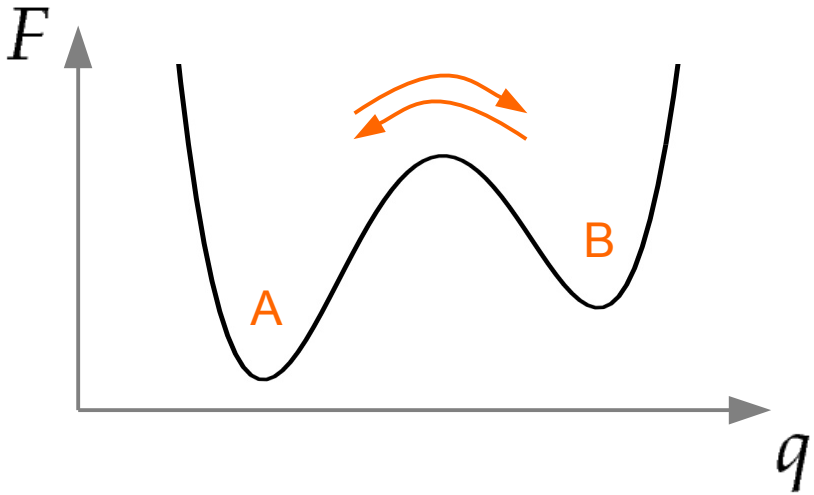
# studying rare events with molecular dynamics

1) generate transition pathways




2) sample the free energy landscape

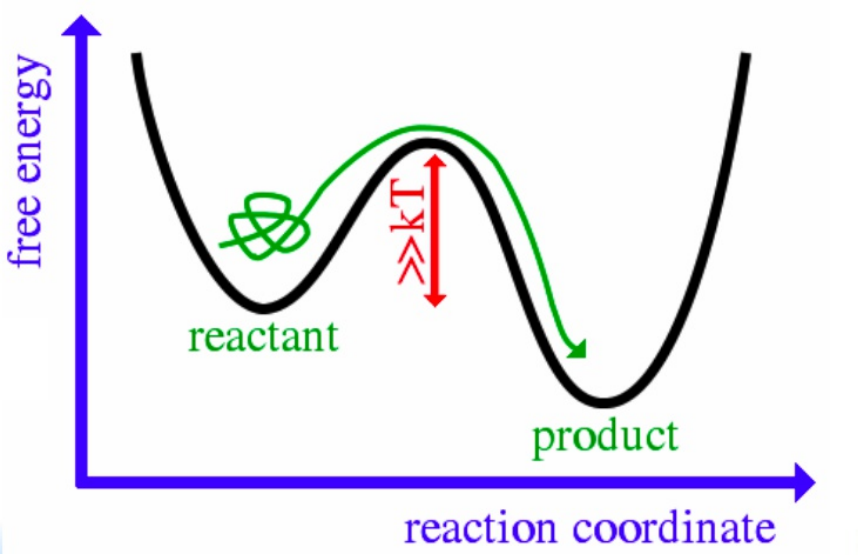
$$F(q(\mathbf{R})) = -kT \log P(q(\mathbf{R}))$$

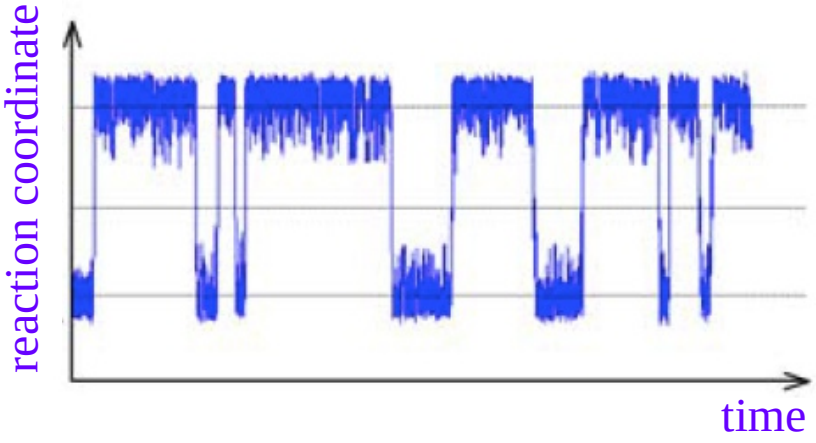


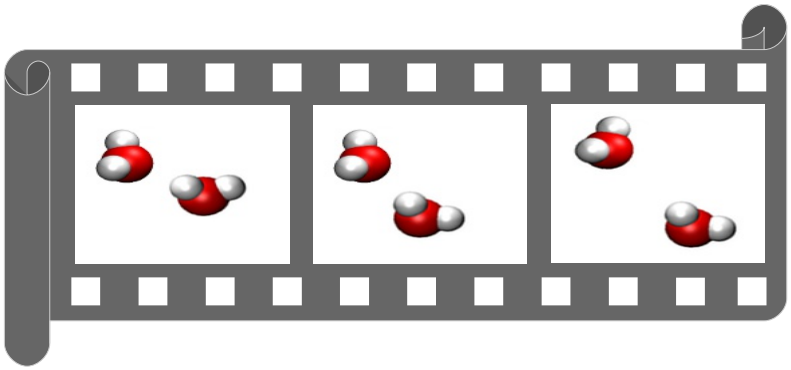

3) compute kinetic rates

$$m\ddot{q} = -\frac{dF}{dq} - \int_0^t dt' \Gamma(t') \dot{q}(t - t') + \eta(t)$$

# the challenge of rare events



in principle, molecular dynamics based on classical / *ab initio* forces is an ideal approach...

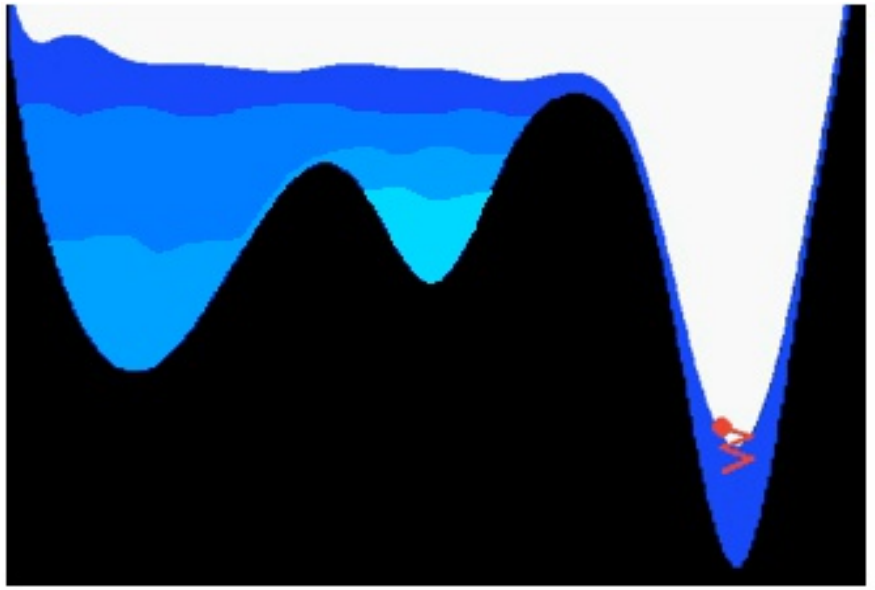
$$\text{rate} \propto e^{-\Delta G^* / kT}$$

**barriers vs rates at 300 K:**

| | |
|---|---|
| 20 kcal/mol | 1 / minute |
| 30 kcal/mol | 1 / human life |
| 40 kcal/mol | 1 / billion years |

## metadynamics

Laio & Parrinello, *PNAS* 2002
Bussi & Laio, *Nat Phys* 2020

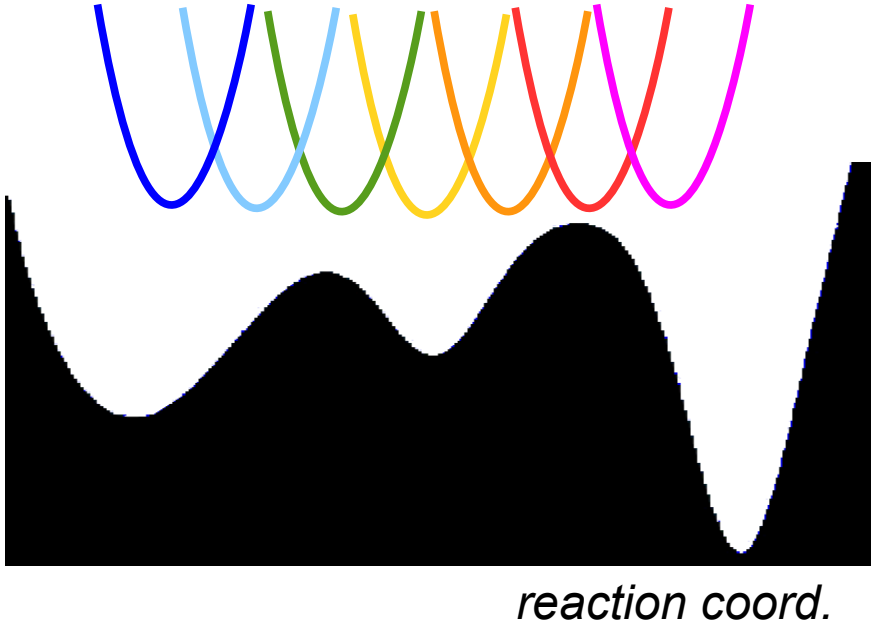



$$U^B(s,t) = \sum_{t_1,\ t_2,\ \ldots\ <t} \omega \exp\left(-\frac{(s-s(t_i))^2}{2\sigma^2}\right)$$

$$\lim_{t\to\infty} U^B(s,t) \approx -F(s)$$

## umbrella sampling

Torrie & Valleau, *J Chem Phys* 1977
Roux, *Comput Phys Commun* 1995



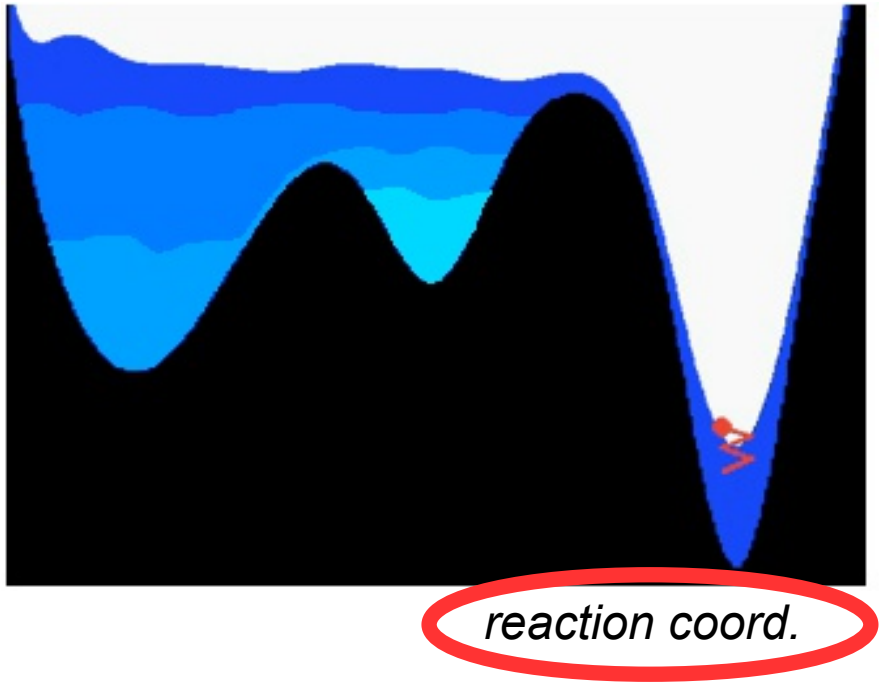

$$U_i^B(s) = \frac{1}{2}k(s-s_i)^2$$

$$P_i(s) = P_i^B(s)e^{\beta(U_i^B(s)-f_i)}$$

$$P(s) = C\sum_i \pi_i(s)P_i(s) \qquad \sum_i \pi_i(s) = 1$$

## metadynamics

Laio & Parrinello, *PNAS* 2002
Bussi & Laio, *Nat Phys* 2020



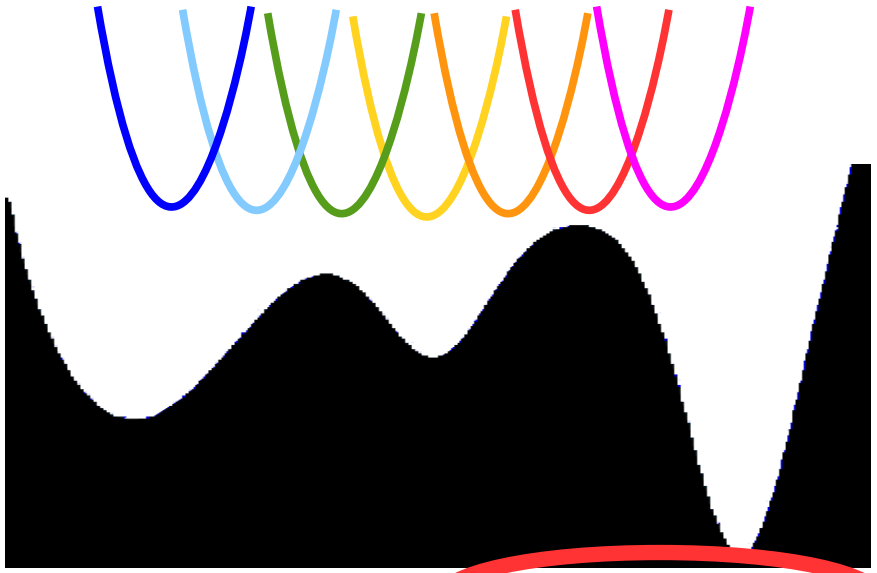

??

$$U^B(s,t) = \sum_{t_1,\ t_2,\ \ldots\ <t} \omega \exp\left(-\frac{(s-s(t_i))^2}{2\sigma^2}\right)$$

$$\lim_{t\to\infty} U^B(s,t) \approx -F(s)$$

## umbrella sampling

Torrie & Valleau, *J Chem Phys* 1977
Roux, *Comput Phys Commun* 1995




??

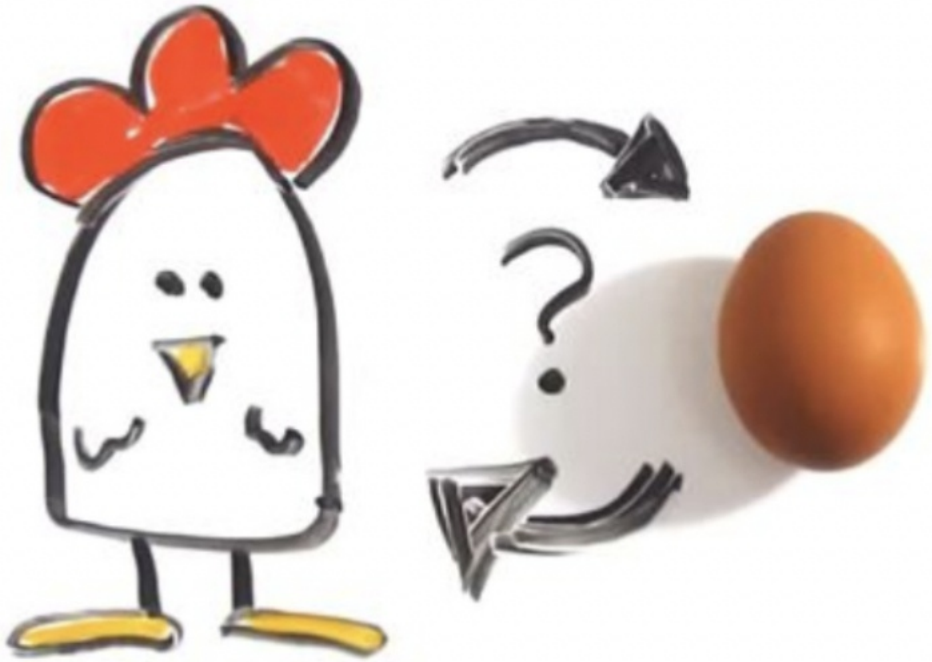
$$U_i^B(s) = \frac{1}{2}k(s-s_i)^2$$

$$P_i(s) = P_i^B(s)e^{\beta(U_i^B(s)-f_i)}$$

$$P(s) = C\sum_i \pi_i(s)P_i(s) \qquad \sum_i \pi_i(s) = 1$$

## how to find good reaction coordinates ?

**chicken & egg dilemma !**



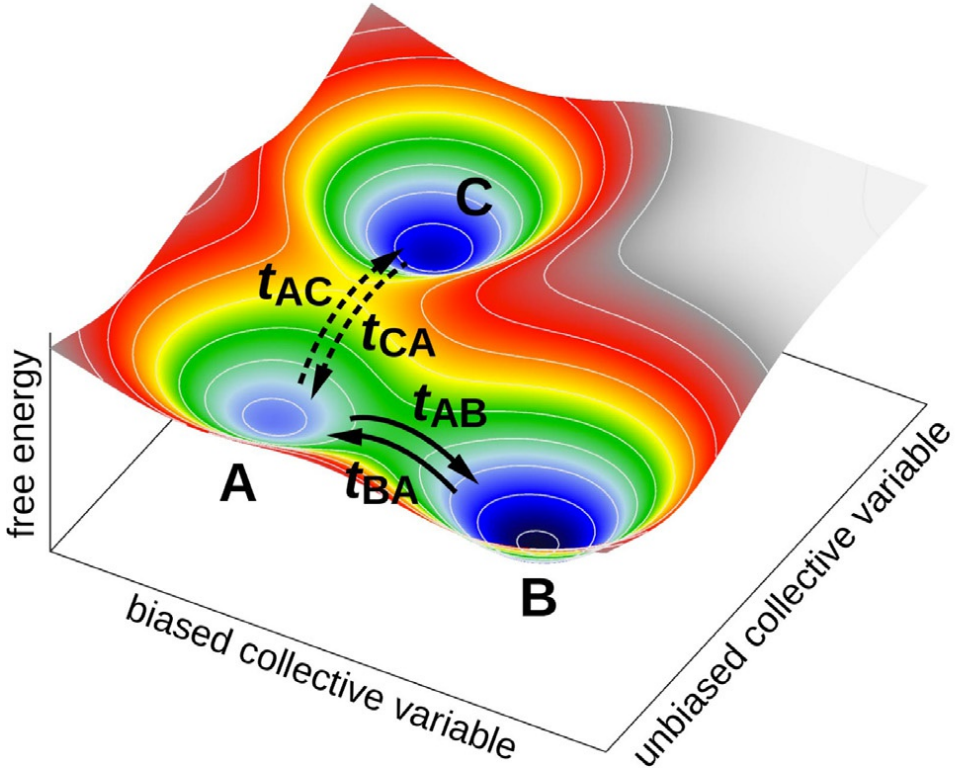
because you need to know
very well the phase space
of the transformation
to find the optimal coordinate,

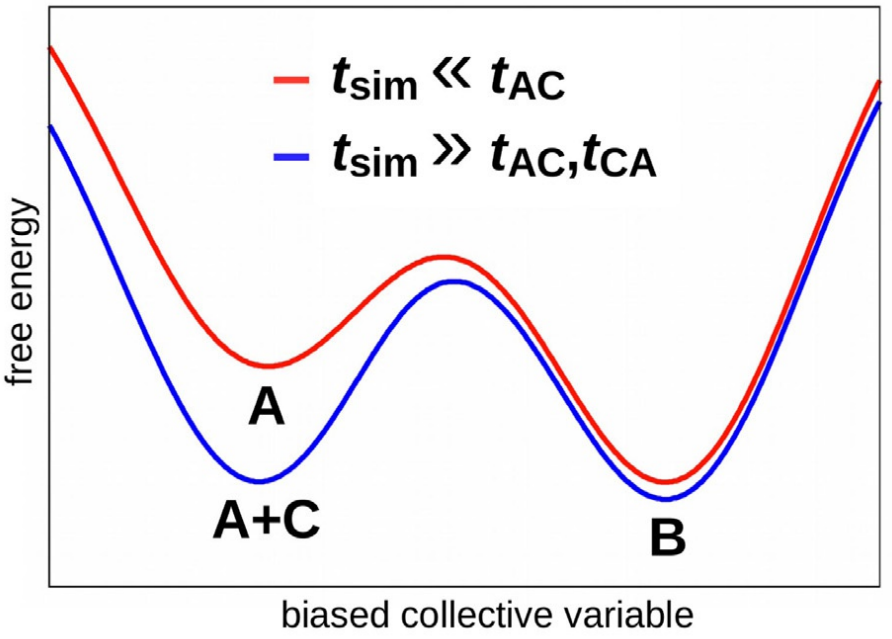
& you need the optimal coordinate
to explore the phase space
of the transformation...

Peters, *Annu Rev Phys Chem* 2016
Jungblut & Dellago, *Eur Phys J E* 2016

# sampling with bias on collective variables, in general

in all methods, the convergence of free energy profiles
depends, in a way difficult to predict,
on **choice of variables** & **length of simulations**



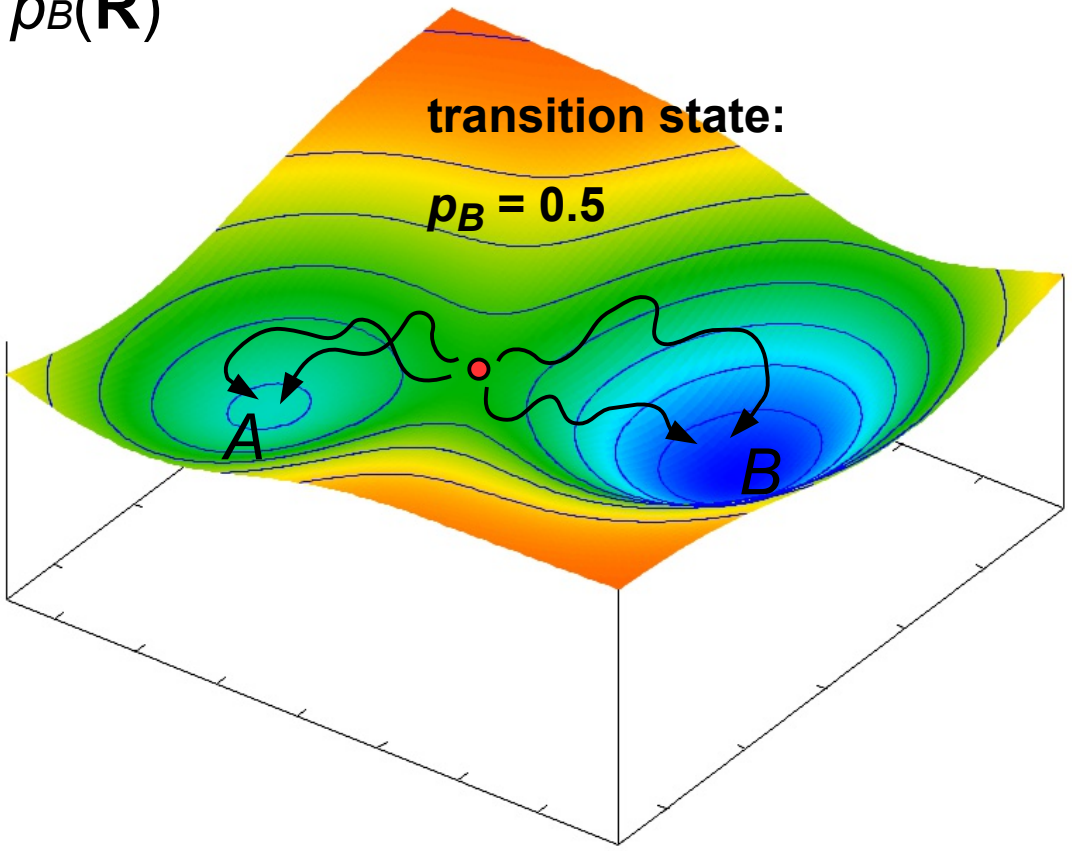

**no convergence if $t_{sim} \sim t_{AC}$**


FP, *Rev Phys* 2, 32 (2017)
Zhu & Hummer, *J Comput Chem* 33, 453 (2012)

Banushkina & Krivov, *WIREs* 2016
Peters, *Annu Rev Phys Chem* 2016
Jungblut & Dellago, *Eur Phys J E* 2016

**optimal reaction coordinate** =
probability that atomic configuration **R** will evolve to *B* before *A*

"**committor**" $p_B(\mathbf{R})$

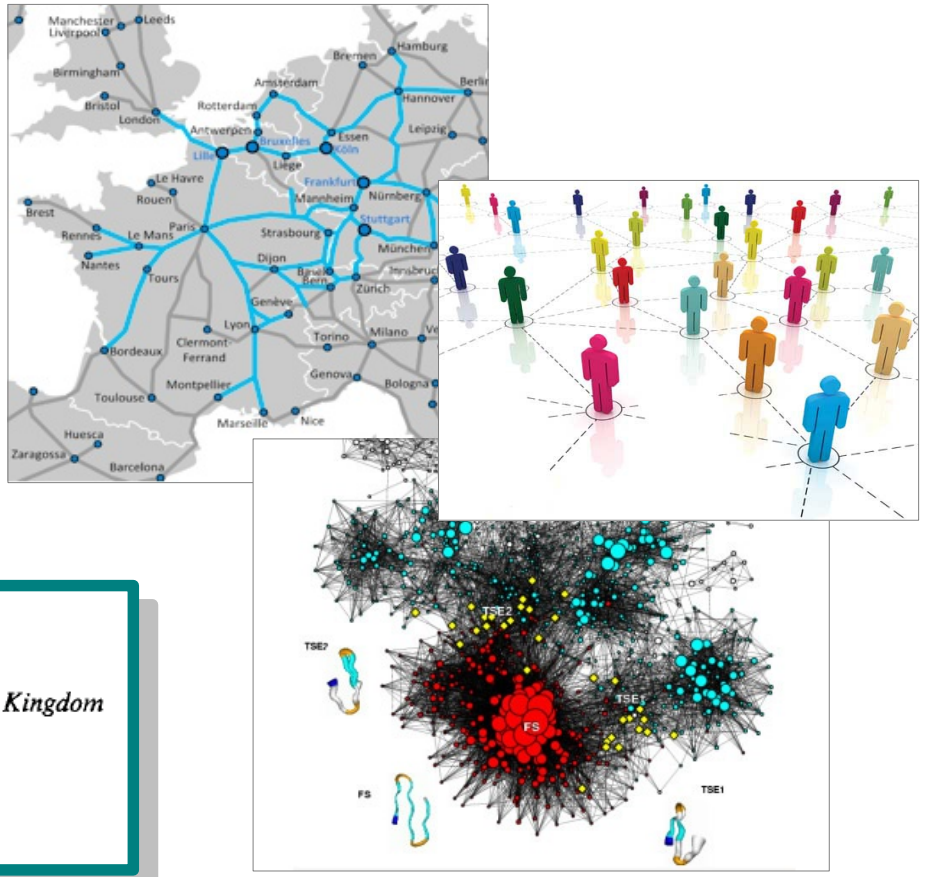



T=0, T>0, enthalpic/entropic barriers,
nucleation, chemical reactions, protein folding...

# can we describe in a unified way the transformations of matter ?

## we try graph theory

it provided powerful tools
in many domains:




SIAM REVIEW
Vol. 48, No. 3, pp. 569–581

Google Italia

**The $25,000,000,000 Eigenvector:**
**The Linear Algebra behind Google***


Kurt Bryan†
Tanya Leise‡


**Molecular graphs, point groups, and fullerenes**


David E. Manolopoulos
*Department of Chemistry, University of Nottingham, Nottingham NG7 2RD, United Kingdom*
Patrick W. Fowler
*Department of Chemistry, University of Exeter, Exeter EX4 4QD, United Kingdom*



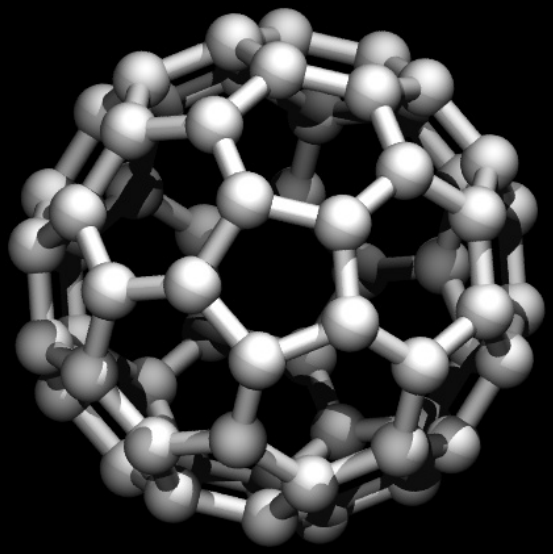
J. Chem. Phys. **96**, 7603 (1992)

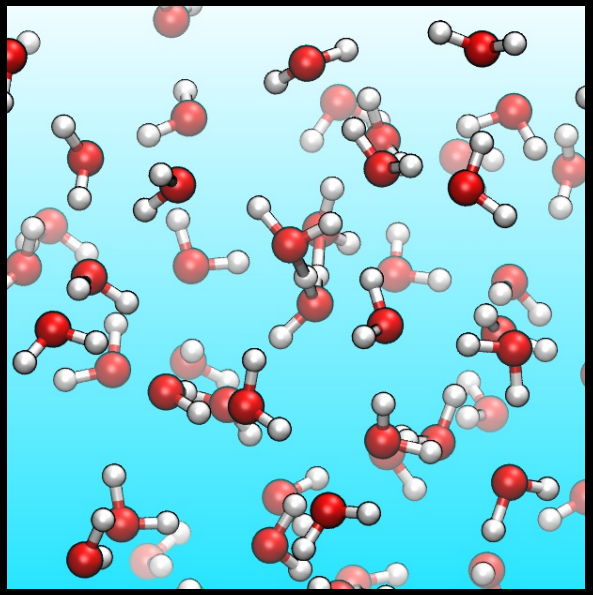


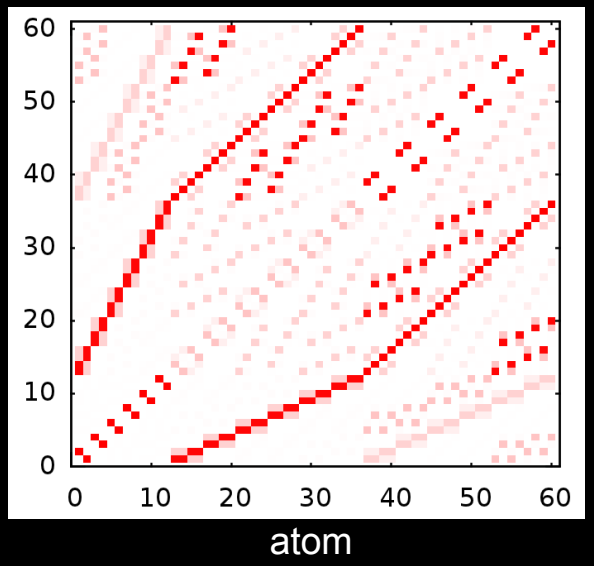
VOLUME 92, NUMBER 21 PHYSICAL REVIEW LETTERS week ending 28 MAY 2004

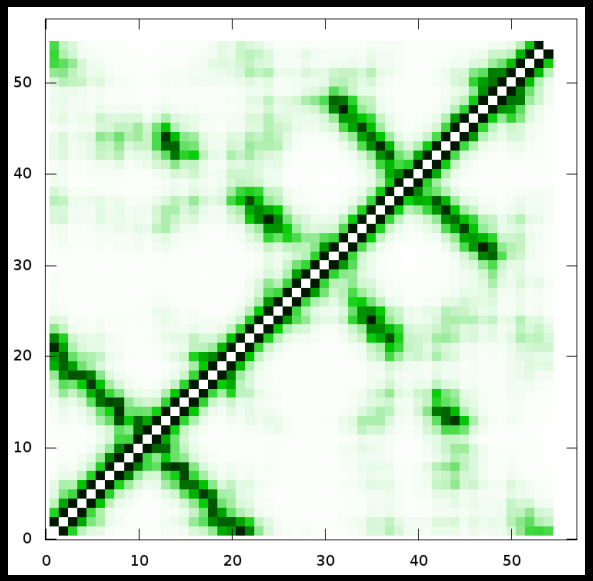

**Reconstruction of Protein Structures from a Vectorial Representation**

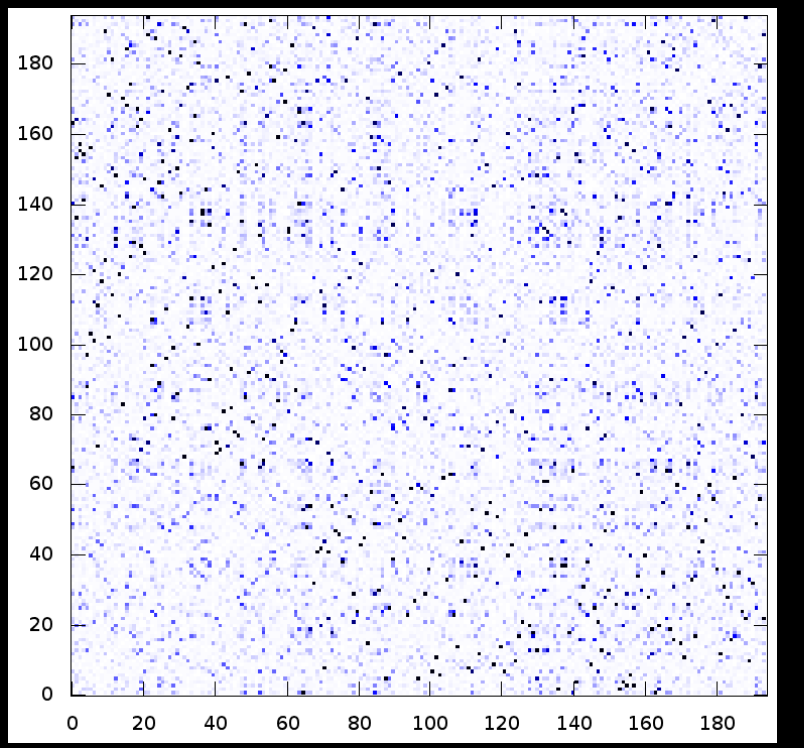

Markus Porto,[1,2] Ugo Bastolla,[3] H. Eduardo Roman,[4] and Michele Vendruscolo[5]



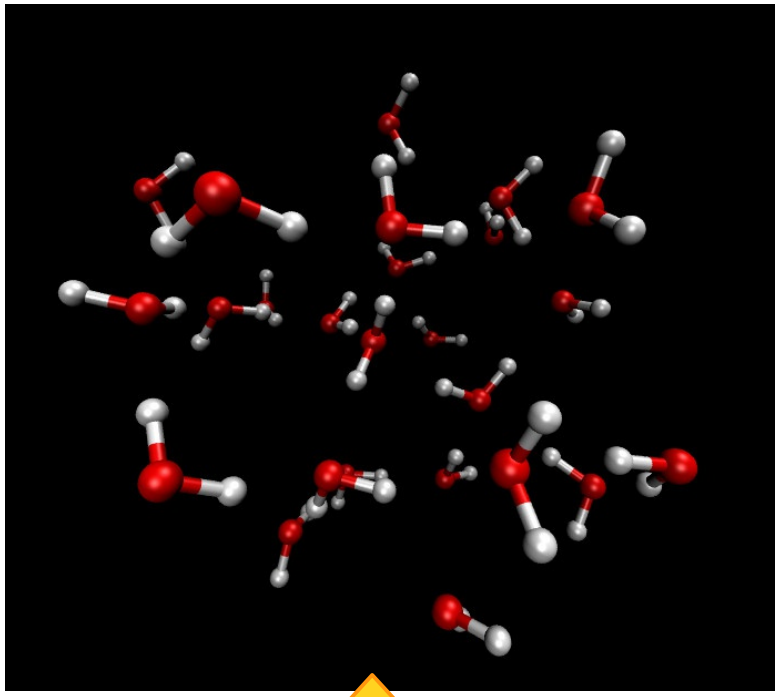


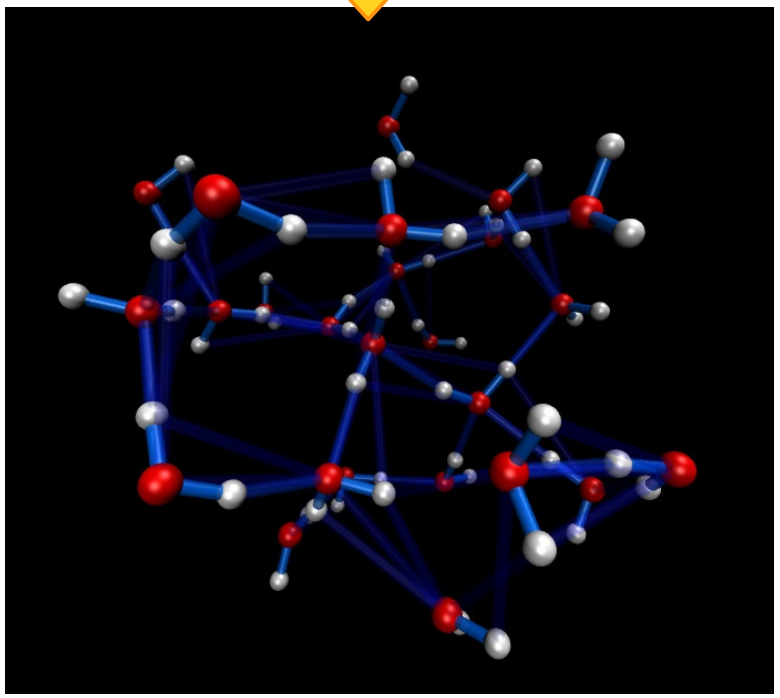



**adjacency matrices are**
**a very fruitful starting point !**

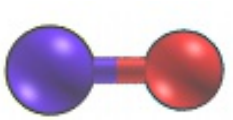




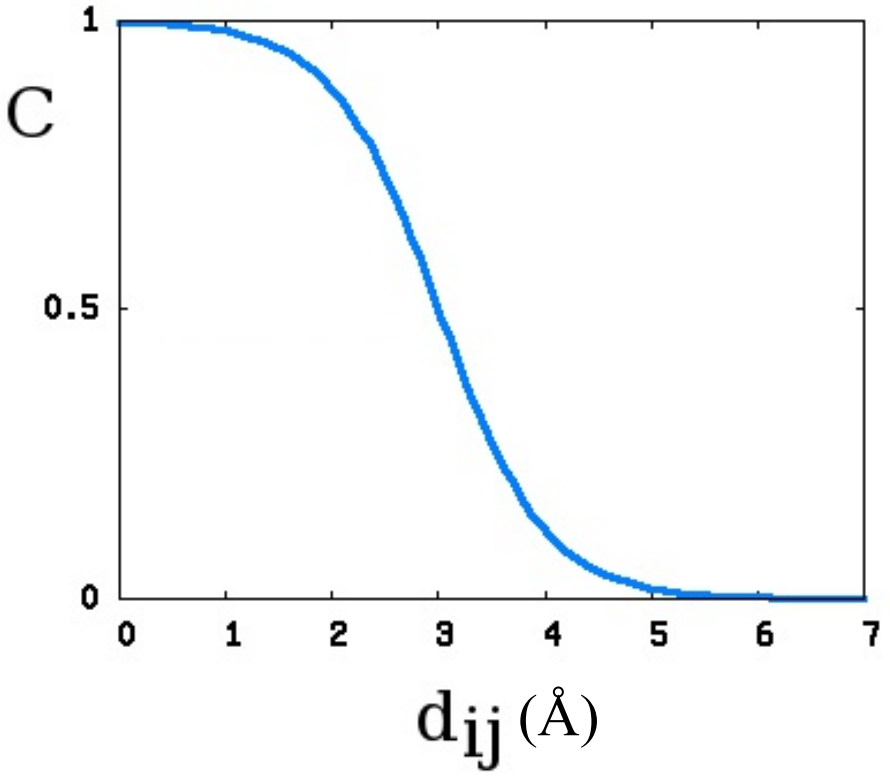
# 1st step: from an atomic system to a graph

filter distances through a switching function

C = 1

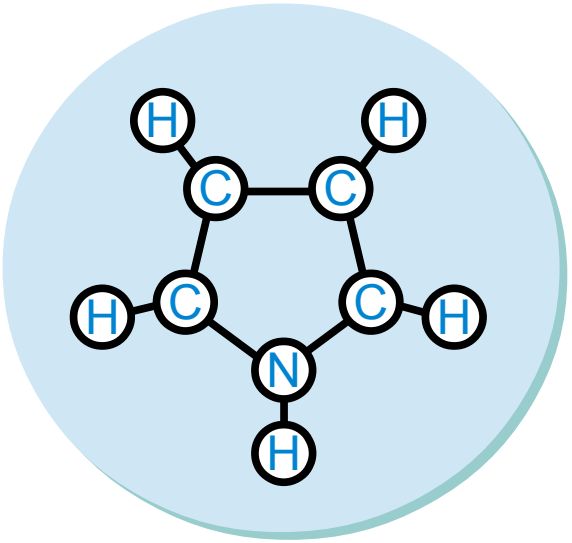


C = 0



**complete** description
of the topology

**approximations**
useful in simulations




adjacency matrix $a_{ij}$

$\lambda^{max}$ sort($v_i^{max}$)

principal eigenvalue/vector

gas-phase reactions, nanoclusters

Pietrucci & Andreoni, *PRL* 2011

**p** = sort($a_{ij}$)

permutation invariant vector

phase trans. in solids & liq., nanoclusters

Gallet & Pietrucci, *JCP* 2013
Pipolo et al, *PRL* 2017

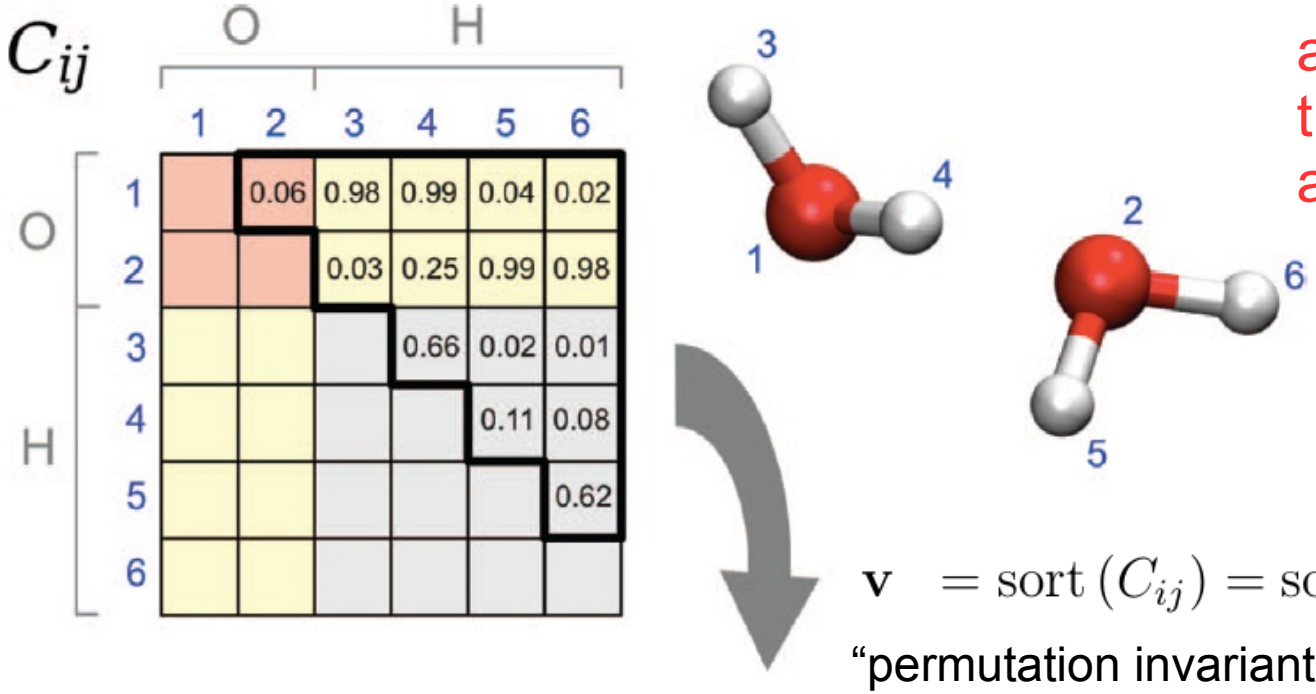
$\{a_{ij}\} \rightarrow$ path CV

coordination patterns

reactions in gas phase & solution

Pietrucci & Saitta, *PNAS* 2015

# from the adjacency matrix to useful coordinates




a general way to estimate the distance between atomic structures

$$\mathbf{v} = \mathrm{sort}\,(C_{ij}) = \mathrm{sort}\left(C(|\mathbf{R}_i - \mathbf{R}_j|)\right) \quad , \quad i > j$$

"permutation invariant vector" (PIV)

| 0.06 | 0.02 | 0.03 | 0.04 | 0.25 | 0.98 | 0.98 | 0.99 | 0.99 | 0.01 | 0.02 | 0.08 | 0.11 | 0.62 | 0.66 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $V_1$ | $V_2$ | $V_3$ | $V_4$ | $V_5$ | $V_6$ | $V_7$ | $V_8$ | $V_9$ | $V_{10}$ | $V_{11}$ | $V_{12}$ | $V_{13}$ | $V_{14}$ | $V_{15}$ |

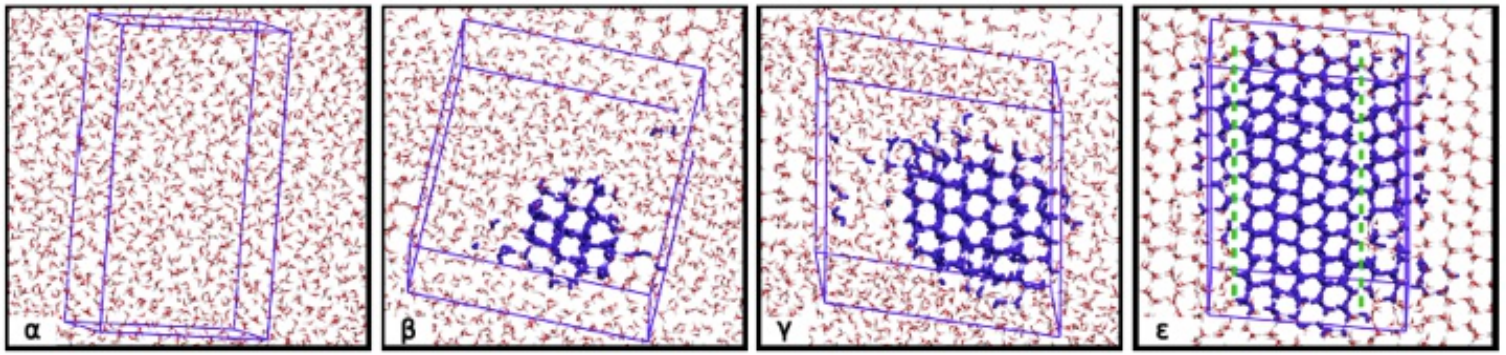
we analyzed > 80 crystalline / amorphous / liquid structures
of molecular / ionic / covalent / metallic materials: **all are well resolved**

C, Na, Si, P, S, Fe, $H_2O$, SiC, $SiO_2$, RbCl, $Fe_2O_3$, $B_2O_3$, $CO_2$ $MgSiO_3$, benzene, paracetamol

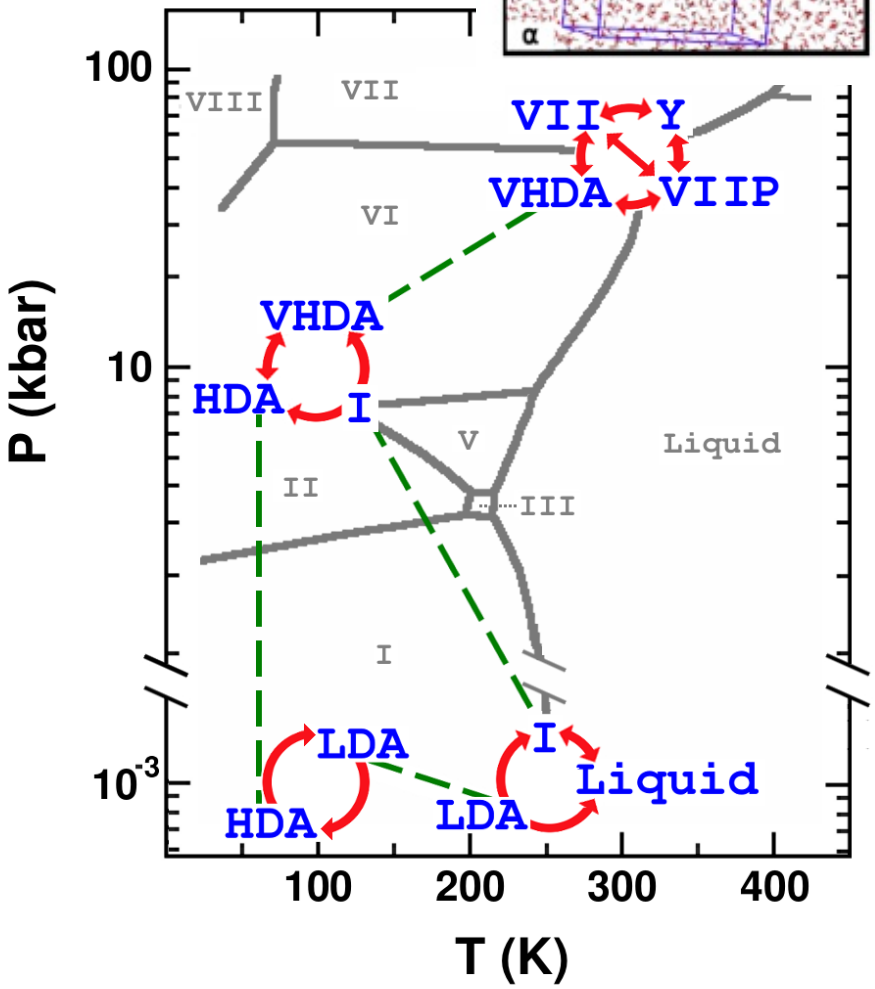
Gallet & Pietrucci, *J Chem Phys* 2013
Pipolo, Salanne, Ferlat, Klotz, Saitta, Pietrucci, *PRL* 2017

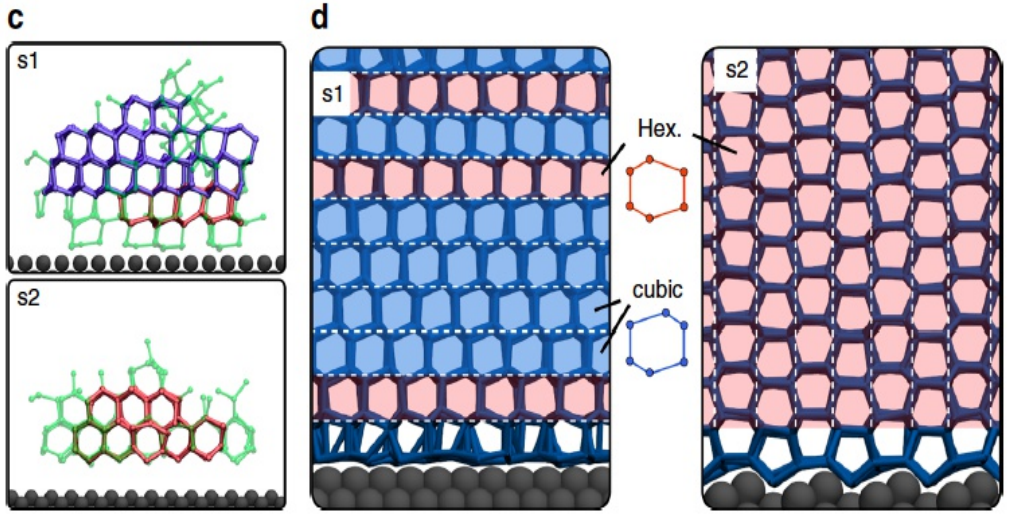
# from the adjacency matrix to useful coordinates

metadynamics,
committor analysis,
umbrella sampling




homogeneous ice nucleation



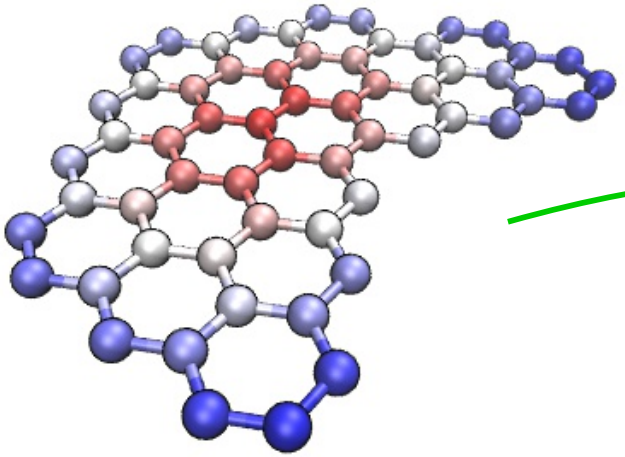




heterogeneous ice nucleation

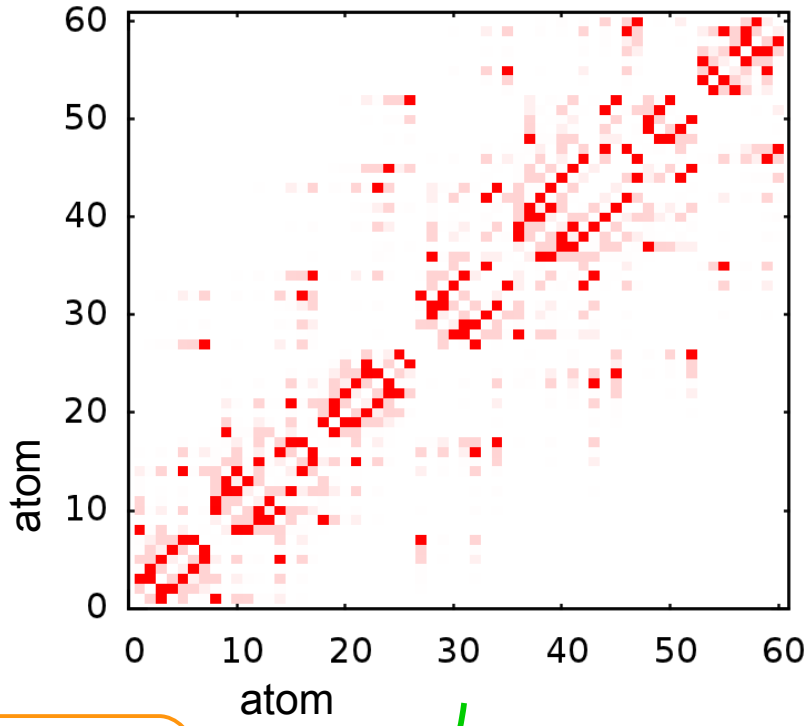
Pipolo, Salanne, Ferlat, Klotz, Saitta, Pietrucci, *PRL* 2017
Fitzner, Sosso, Pietrucci, Pipolo, Michaelides, *Nat. Commun.* 2017

# from the adjacency matrix to useful coordinates

atomic structure



adjacency matrix



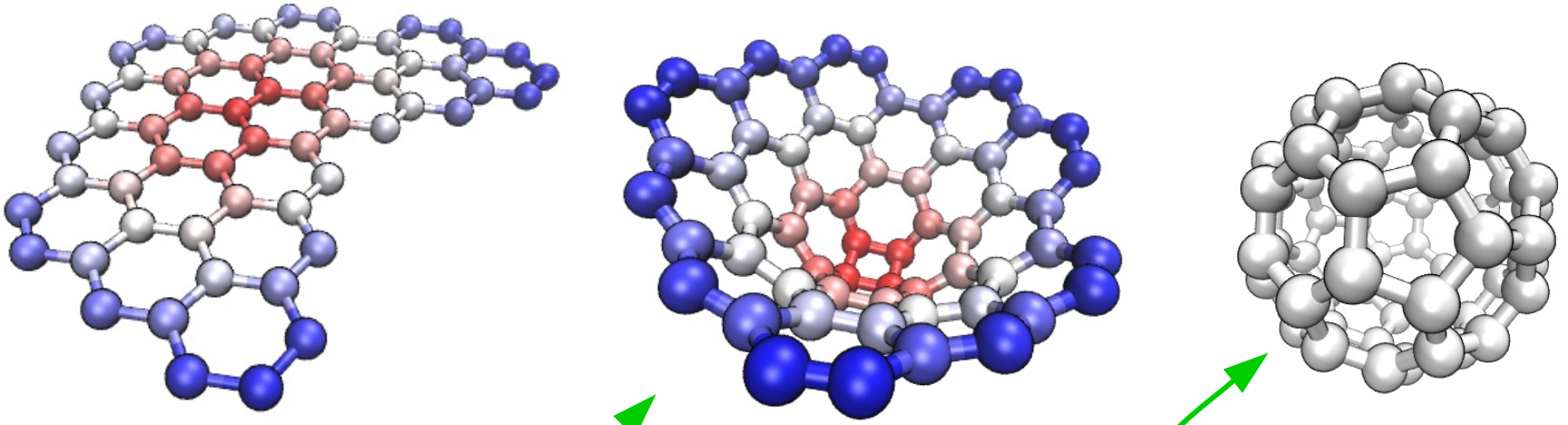

$$S_i = \sqrt{N}\ \lambda^{\max}\ v_i^{\max}$$

a “social” coordinate is assigned to each atom

Pietrucci & Andreoni, *PRL* 2011

# from the adjacency matrix to useful coordinates

atomic structure



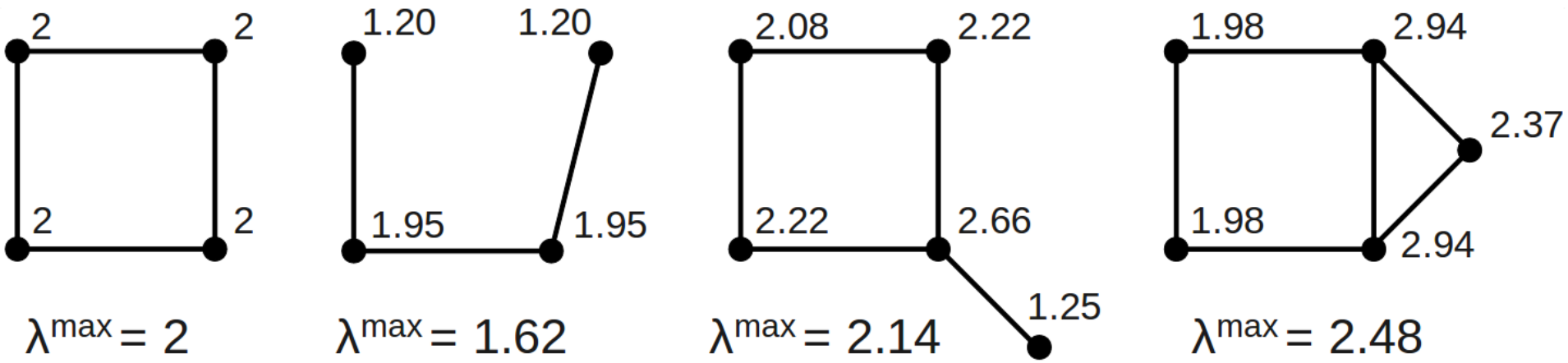
$$S_i = \sqrt{N}\ \lambda^{\max}\ v_i^{\max}$$

a “social” coordinate is assigned to each atom

Pietrucci & Andreoni, *PRL* 2011

# from the adjacency matrix to useful coordinates

sensitive to topology:




$$S_i = \sqrt{N}\ \lambda^{\max}\ v_i^{\max,\text{sorted}}$$

$$S_1 \leq S_2 \leq S_3 \leq \ldots \leq S_N$$

Social PeRmutation-INvarianT coordinates

Pietrucci & Andreoni, *PRL* 2011

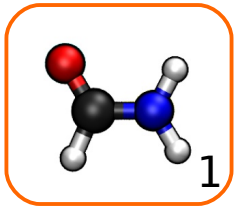
# general-purpose reaction coordinates for chemistry

a collaborative work @ SORBONNE UNIVERSITÉ

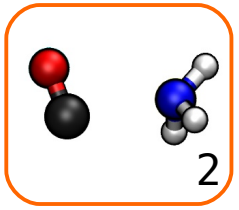
Marco Saitta

Théo Magrino

# general-purpose reaction coordinates for chemistry




only endpoints ?




$$s(t) = \frac{\sum_{k=1}^{N_f} k e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)}}{\sum_{k'=1}^{N_f} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_{k'})}}$$

$$z(t) = -\frac{1}{\lambda} \log \left( \sum_{k=1}^{N_f} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)} \right)$$

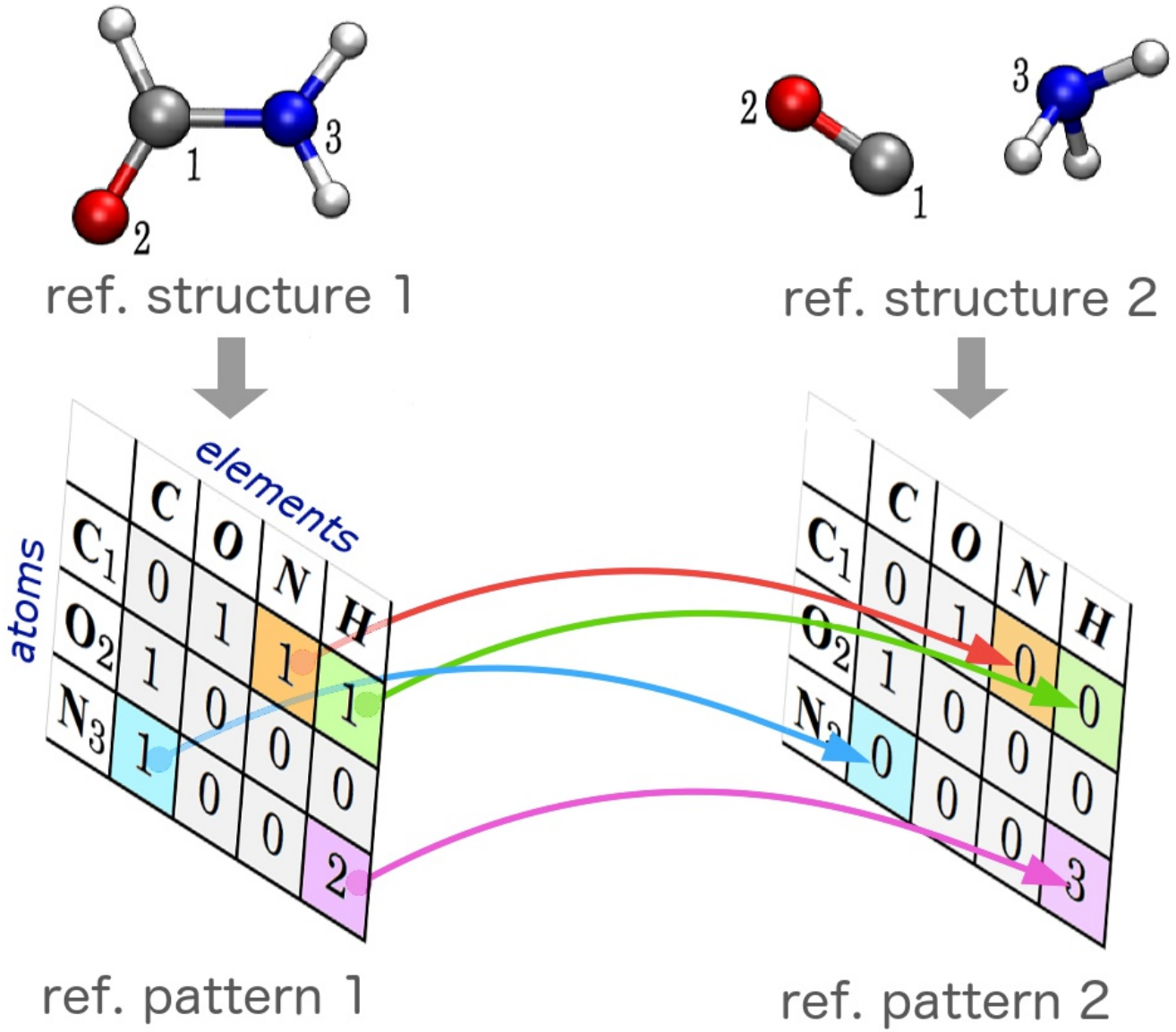
**idea:**
introduce a metric
flexible enough to include
the participation of solvent

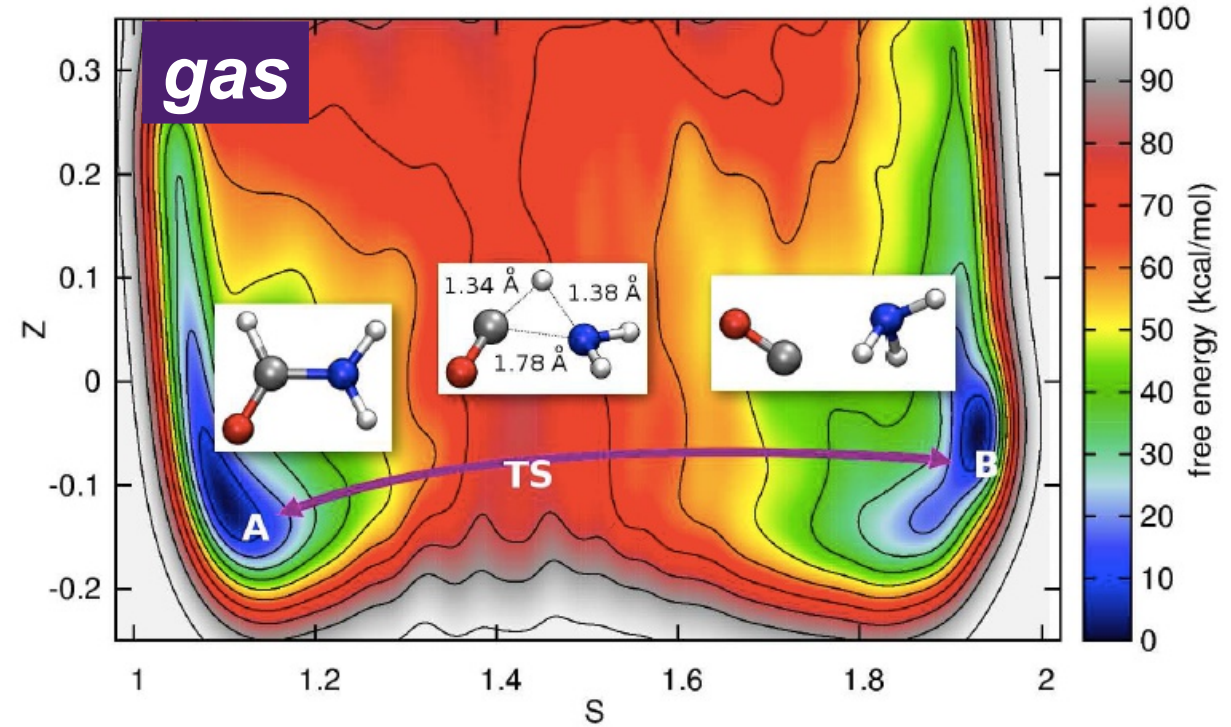
$$D\left(\mathbf{R}(t), \mathbf{R}_k\right) = \sum_{IS} \left( C_{IS}(t) - C_{IS}^k \right)^2$$

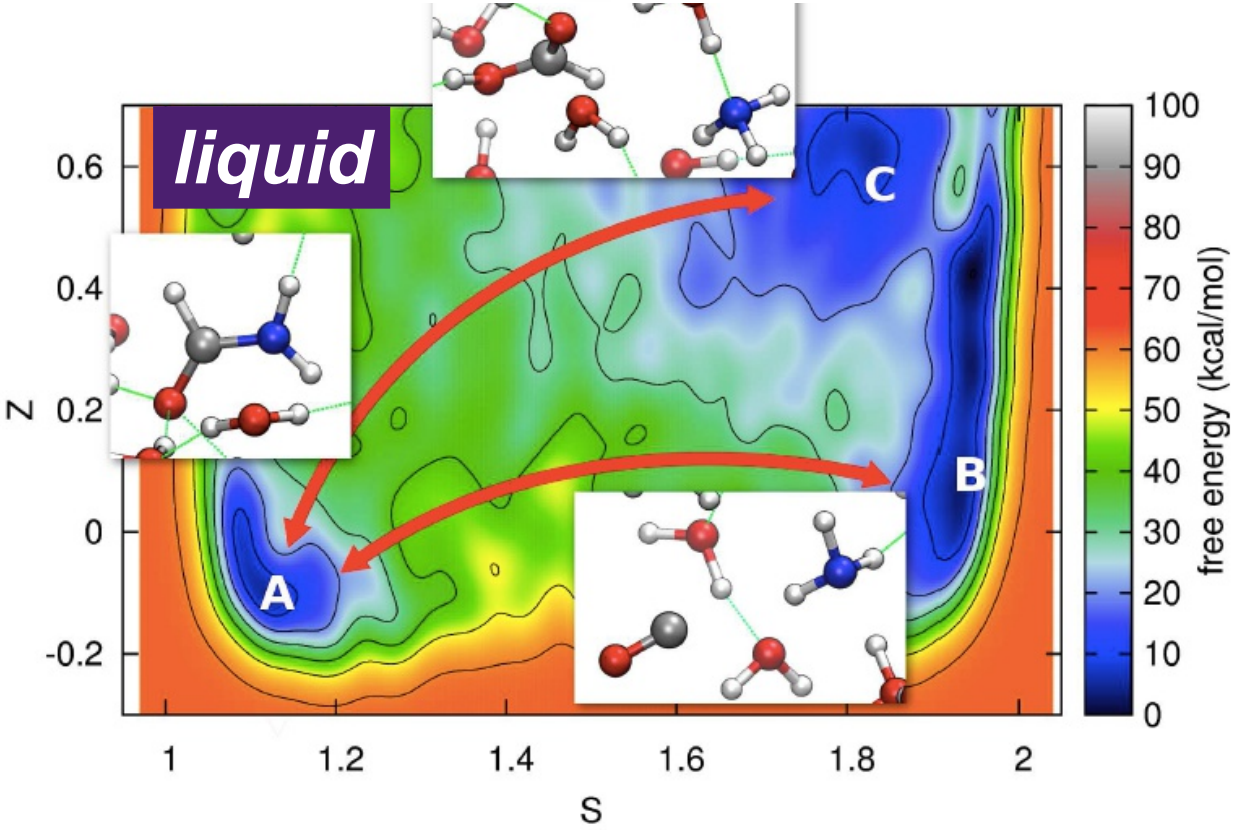
coordination number
of atom $I$ w.r.t. species $S$

Pietrucci & Saitta, *PNAS* 112, 15030 (2015)

# general-purpose reaction coordinates for chemistry





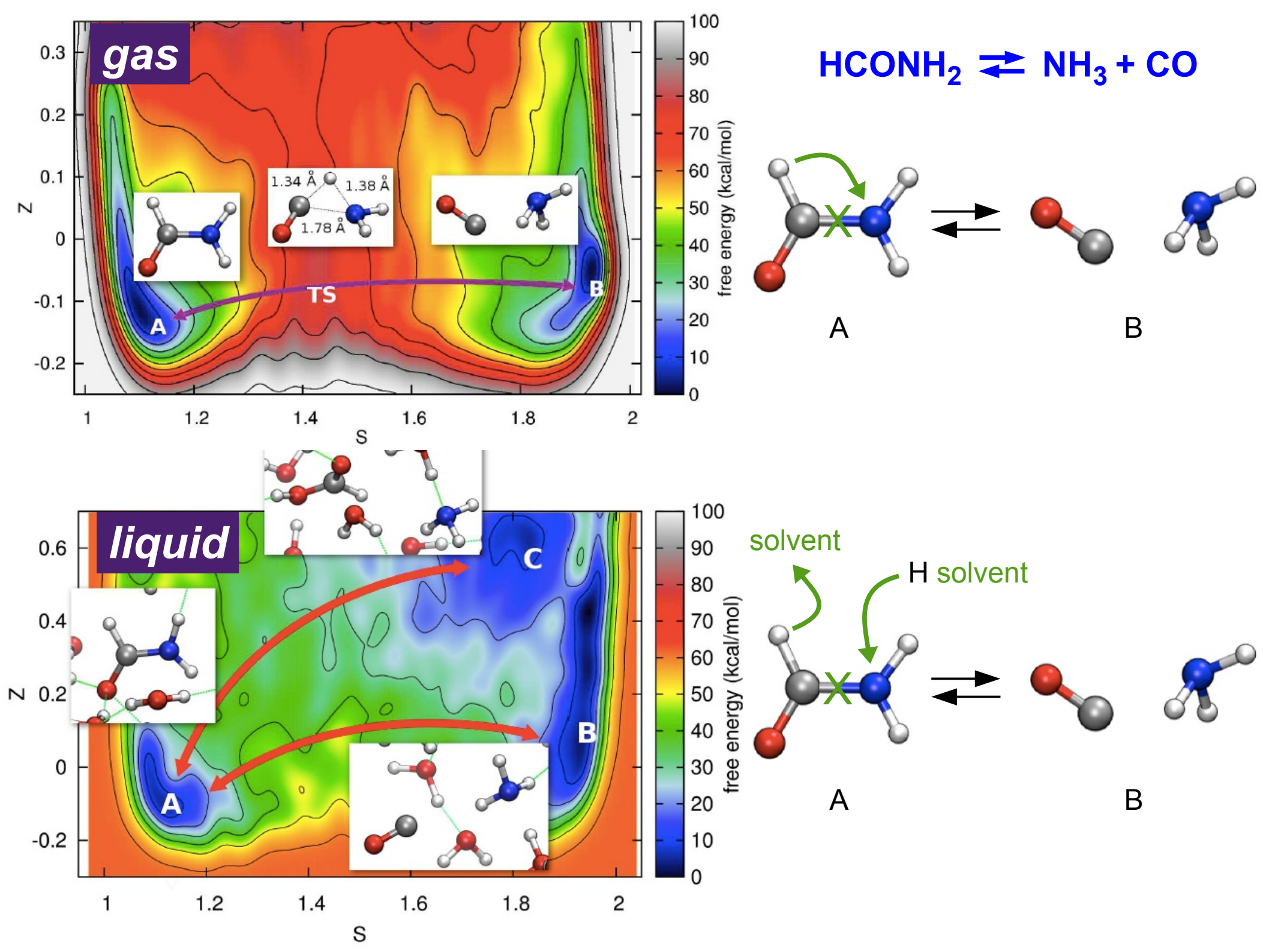




$HCONH_2 \rightleftharpoons NH_3 + CO$

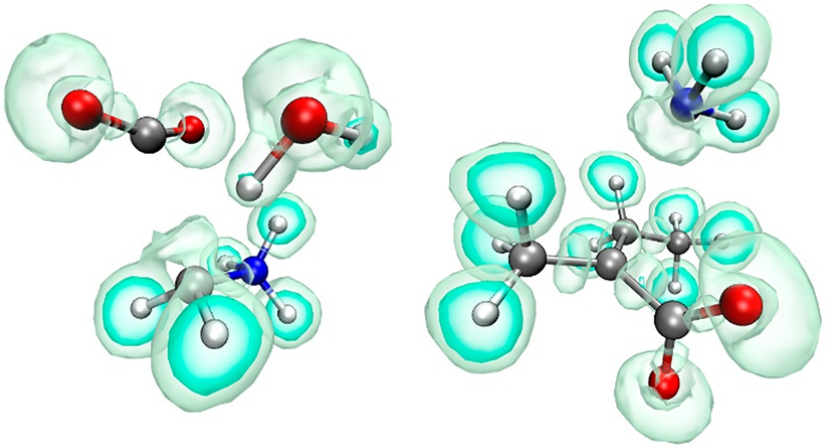
input = A & B

the simulation automatically discovers

- transition pathways
- intermediate states
- off-pathway states

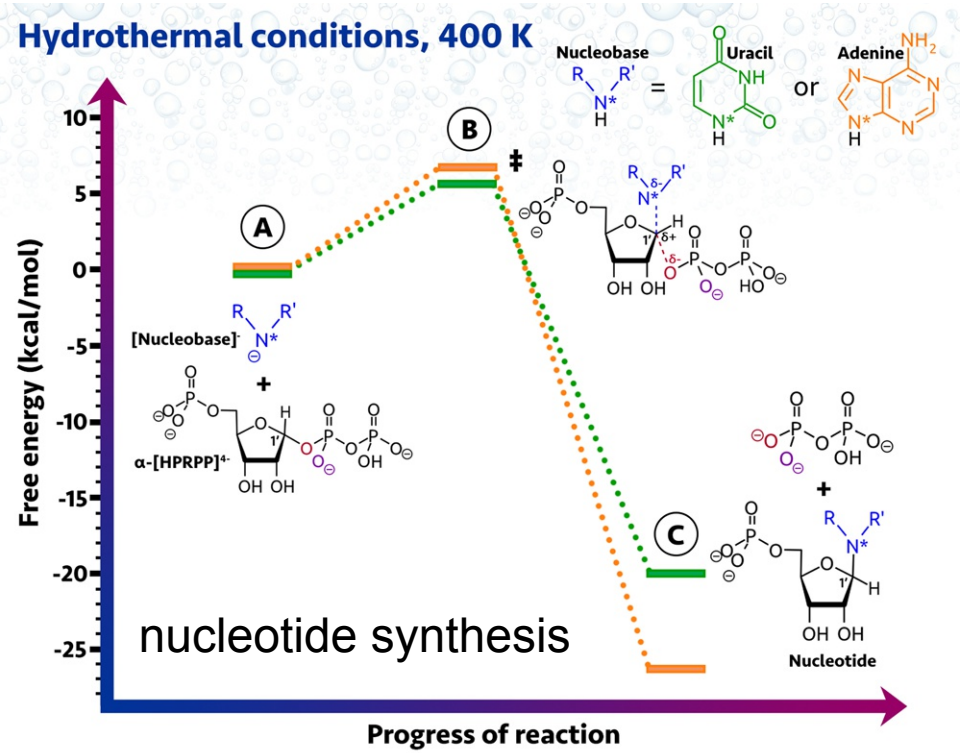

gas
liquid
HCONH2 ⇌ NH3 + CO
1.34 Å
1.38 Å
1.78 Å
TS
A
B
C
S
Z
free energy (kcal/mol)
solvent
H solvent

# several applications so far...

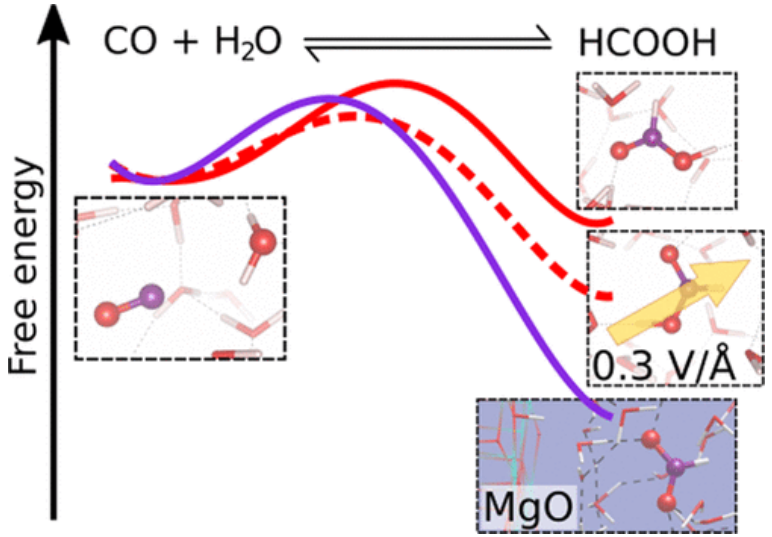


amino acid decomposition

*ACS Earth Space Chem* 2018

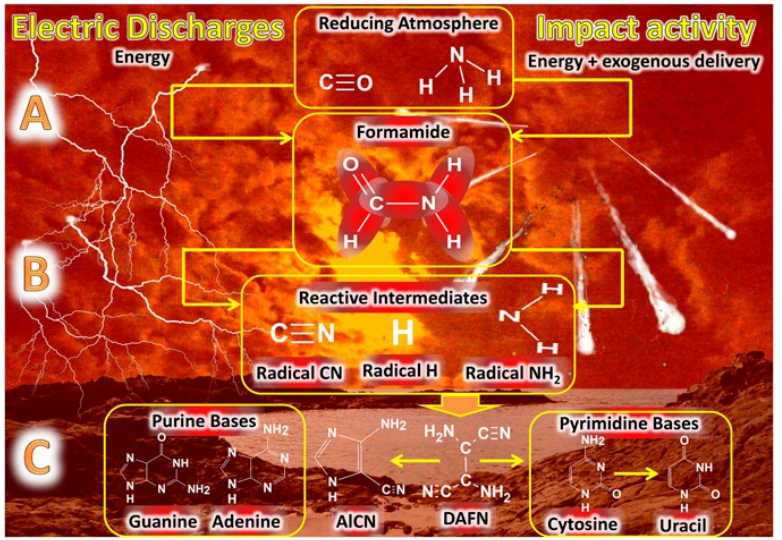



*J Phys Chem Lett* 2018




formic acid synthesis at water-mineral interface

*J Phys Chem C* 2020



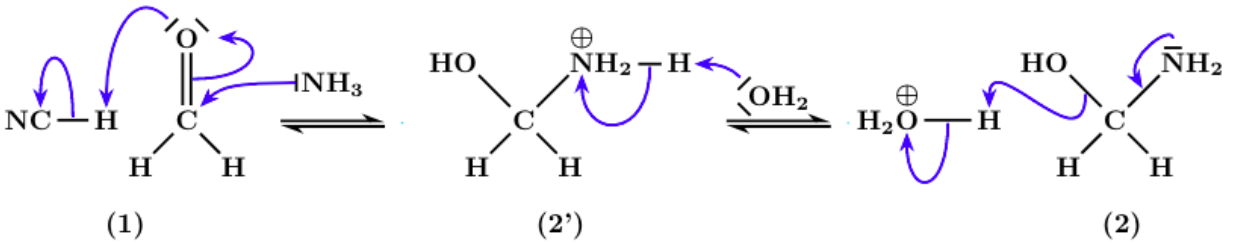

formation of nucleobases

*PNAS* 2017

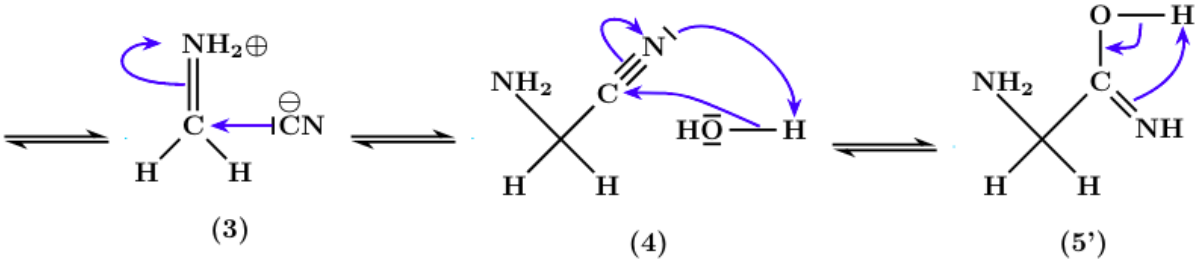
Théo Magrino

## simulating for the first time the complete Strecker amino acid synthesis

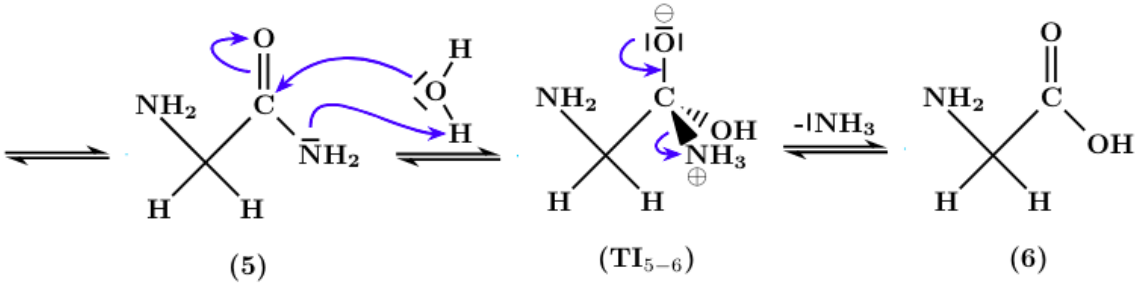
one century old, invoked by Miller (1952), no full study




DFT ab initio MD
T = 300K
81 water molecules
about 4 ns in total (!)




supercomputers:
CINES, TGCC, PRACE

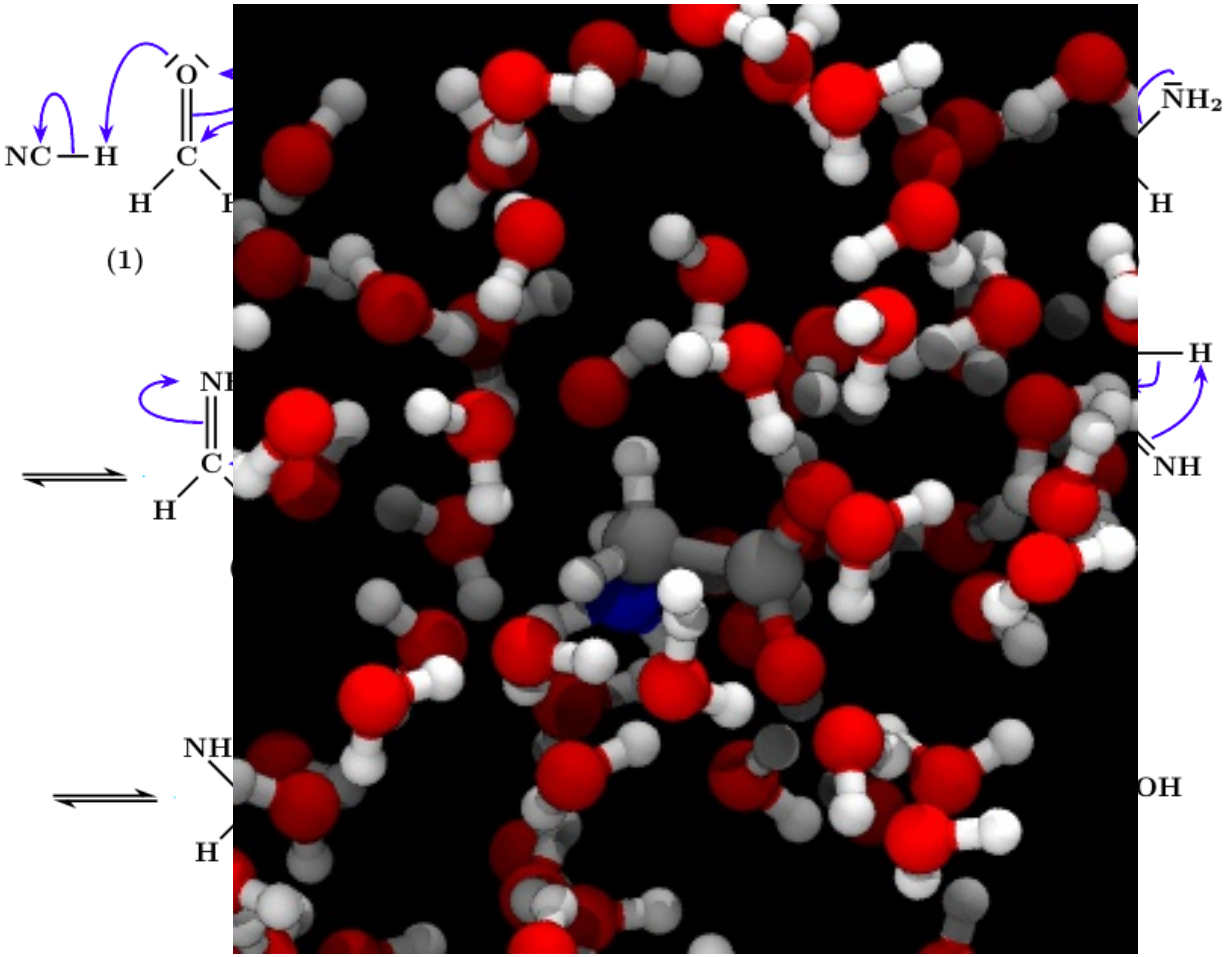



**a complex multi-step pathway:**
**only one form of reaction coordinates**

Théo Magrino

## simulating for the first time the complete Strecker amino acid synthesis

one century old, invoked by Miller (1952), no full study




DFT ab initio MD
T = 300K
81 water molecules
about 4 ns in total (!)

supercomputers:
CINES, TGCC, PRACE

**a complex multi-step pathway:**
**only one form of reaction coordinates**

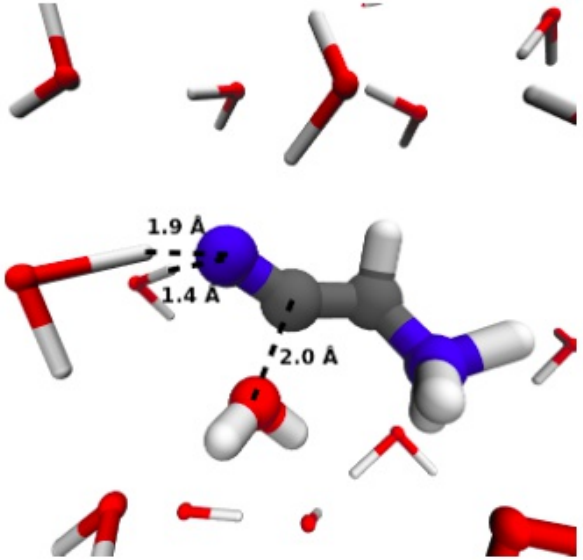
Théo Magrino

**simulating for the first time the complete Strecker amino acid synthesis**

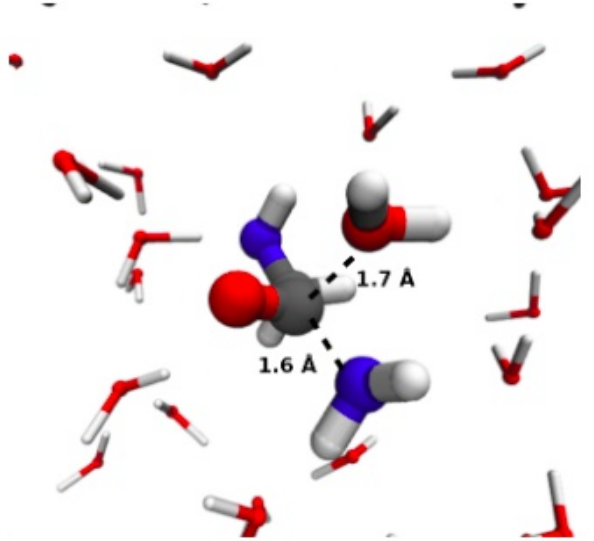
one century old, invoked by Miller (1952), no full study

DFT ab initio MD
T = 300K
81 water molecules
about 4 ns in total (!)

supercomputers:
CINES, TGCC, PRACE




(a) Nitrile hydration transition state, step **(4)**→**(5')**. $\Delta F^{\ddagger} = 22$ kcal/mol.




(b) Amide hydrolysis transition state, step **(5)**→**(6)**. $\Delta F^{\ddagger} = 29$ kcal/mol.

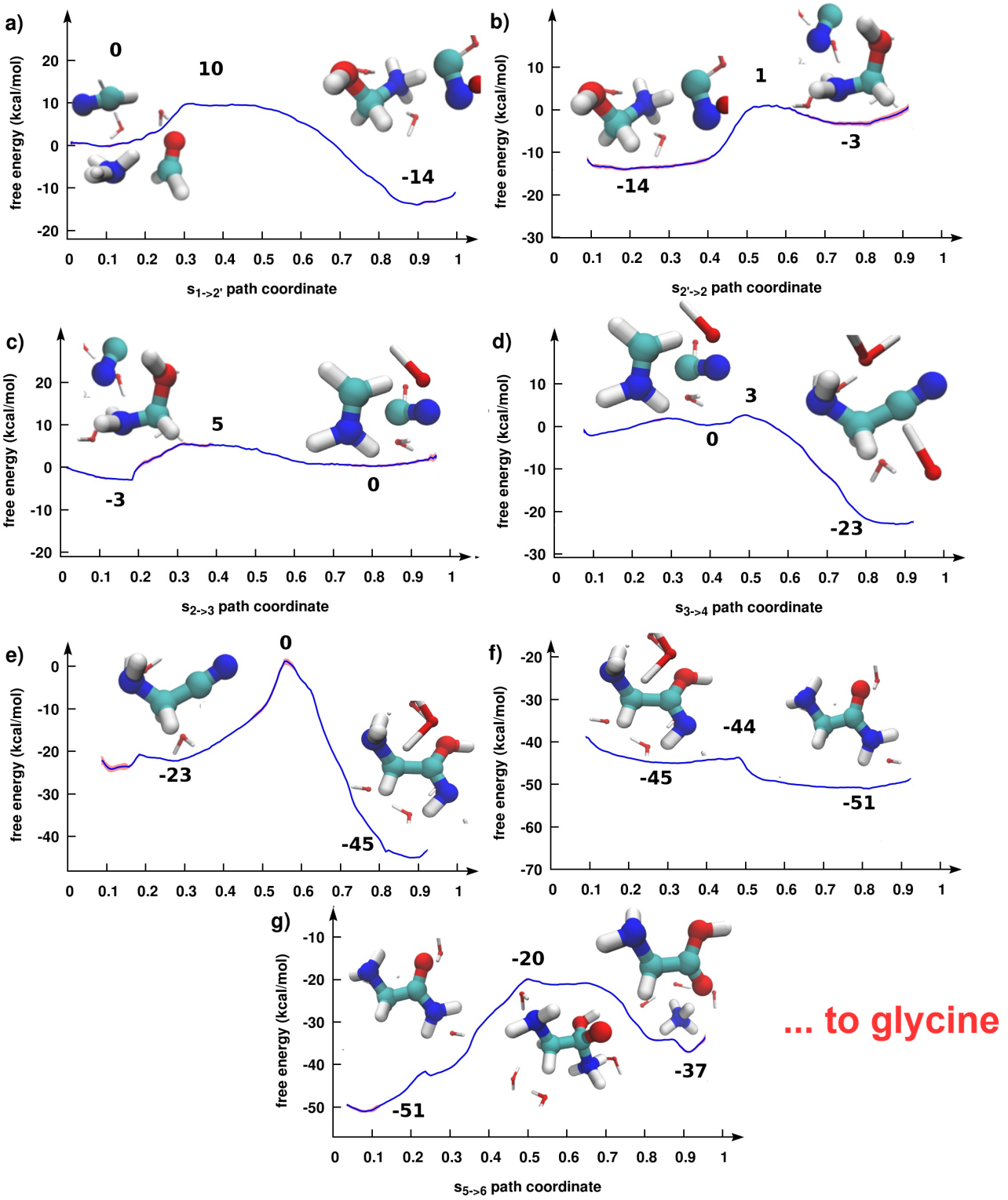
**a complex multi-step pathway:**
**only one form of reaction coordinates**

Théo Magrino

**from simple molecules ...**

DFT ab initio MD
T = 300K
81 water molecules
about 4 ns in total (!)

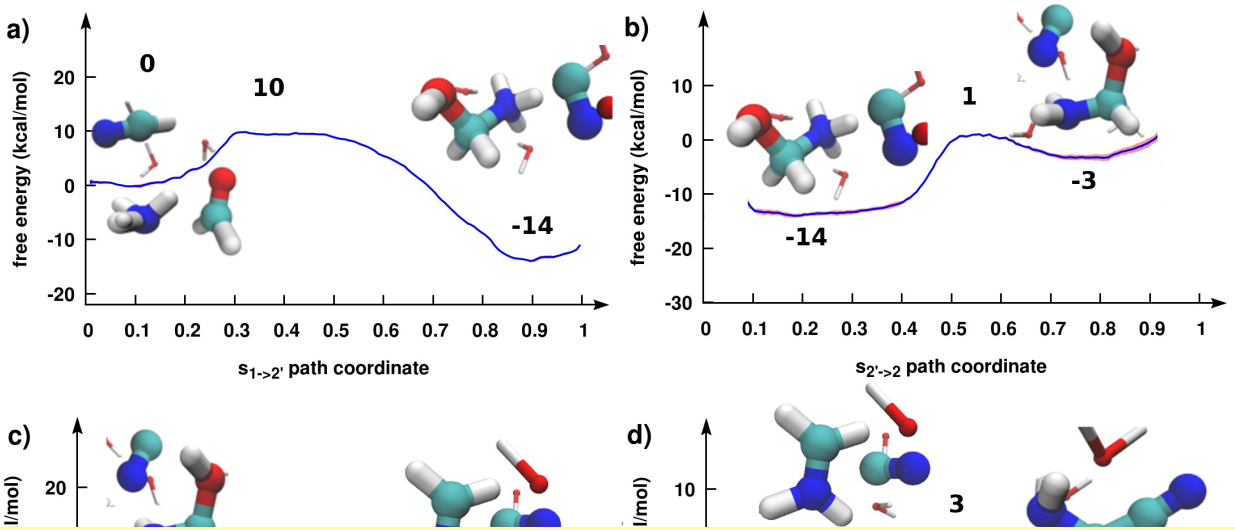
supercomputers:
CINES, TGCC, PRACE




**... to glycine**

Théo Magrino

**from simple molecules ...**



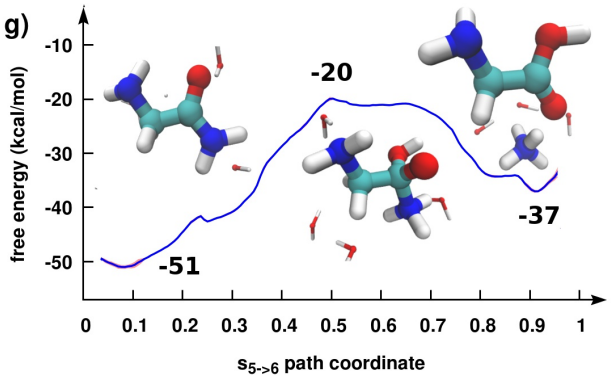

**conclusion:**

**good agreement with available exp. free energies (within DFT-PBE typical errors)**

**a reference for future prebiotic chemistry studies (effect of catalysts like rocks or meteorites?)**

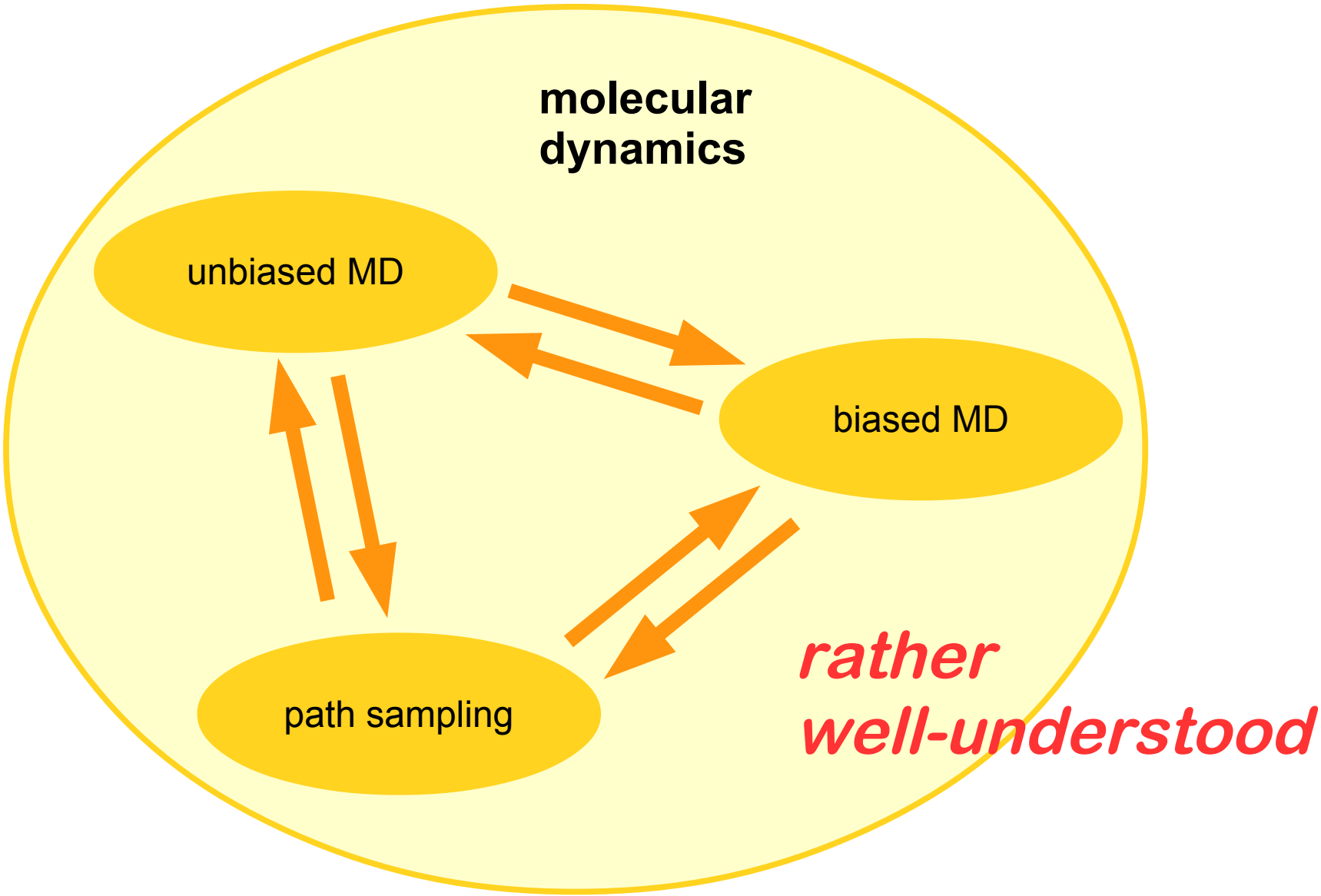
(submitted for publication)

PLUMED PLUMED-NEST




**... to glycine**

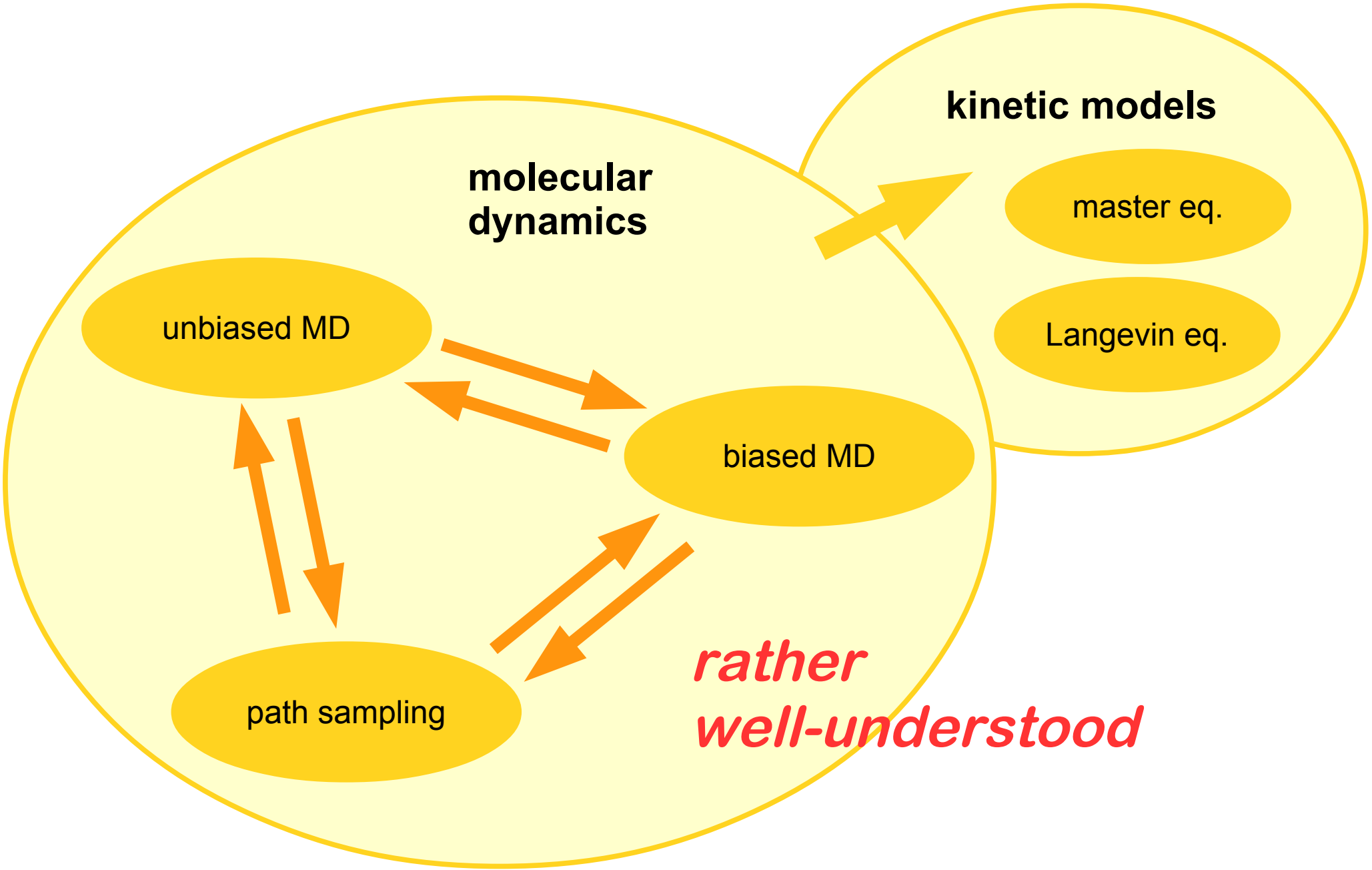
the larger picture:




*Adv Phys X 3, 1477531* (2018)

the larger picture:

*work in progress...*




*Adv Phys X 3, 1477531* (2018)