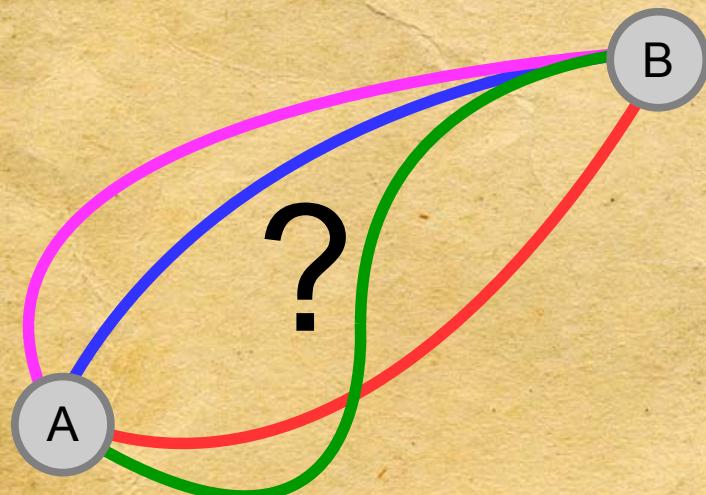
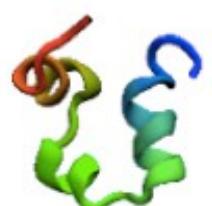
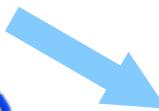
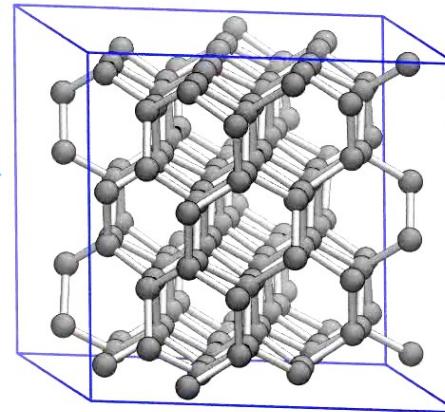
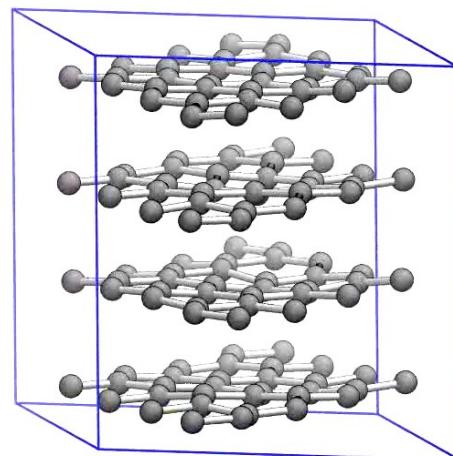
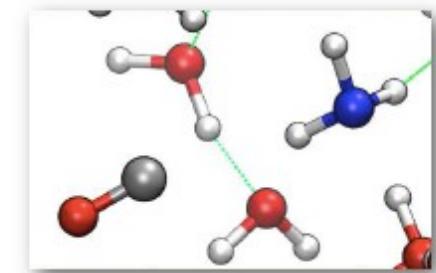
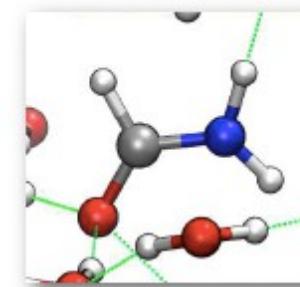
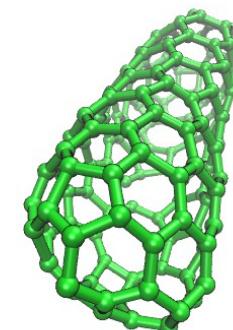
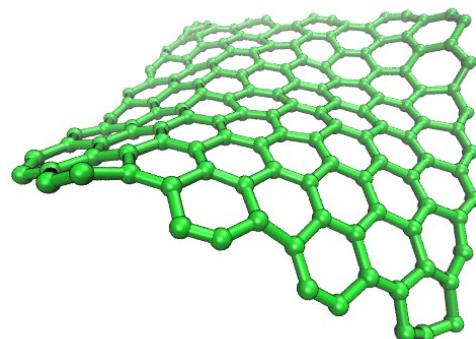


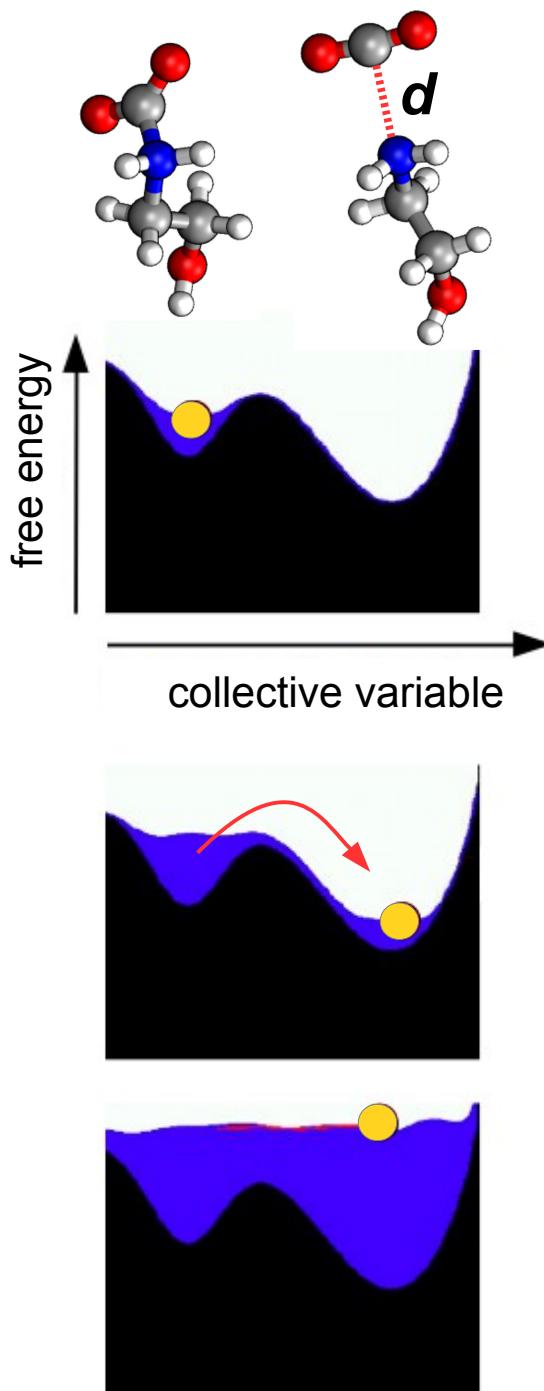
graph theory-inspired coordinates: from chemical reactions to phase transitions



how to characterize in atomic detail
the transformation mechanisms of matter?

is there a general way to describe
many different mechanisms?





molecular dynamics

+

history-dependent bias potential

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

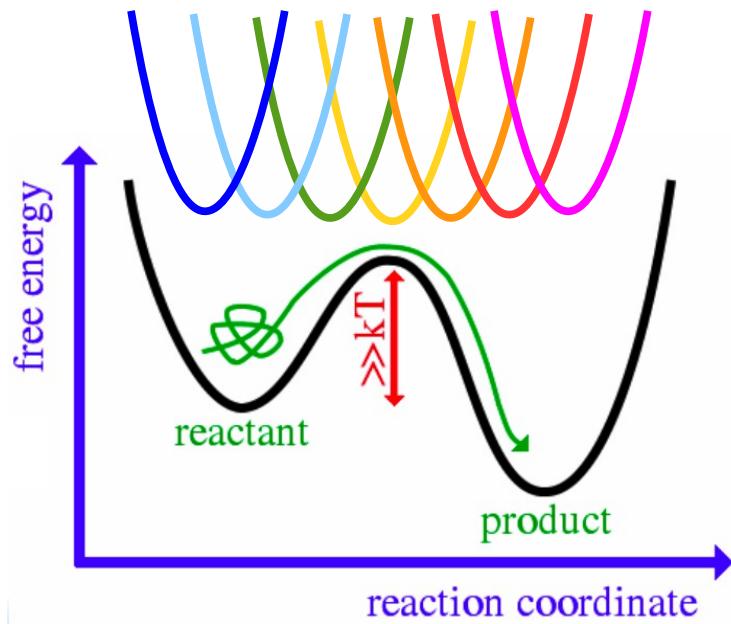
→ accelerated exploration

→ reconstruction of the equil.
free energy landscape

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

umbrella sampling

Torrie & Valleau, *J Chem Phys* **63**, 187 (1975)
Roux, *Comput Phys Commun* **91**, 275 (1995)



add bias potential (e.g., parabola)
to sample each region *separately*

- + control of sampling precision
- + perfectly parallel
- + put simulation where you need
- monitor trajectory carefully
- need to fill holes after first attempt

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

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

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

reweighting (WHAM) gives
equilibrium probability

how to choose a “good” collective variable?

- Distances
- Angles
- Dihedral angles
- Coordination numbers
- Density
- Energy
- ...



you must pull
the right strings!

how to choose a “good” collective variable?

- Distances
- Angles
- Dihedral angles
- Coordination numbers
- Density
- Energy
- ...



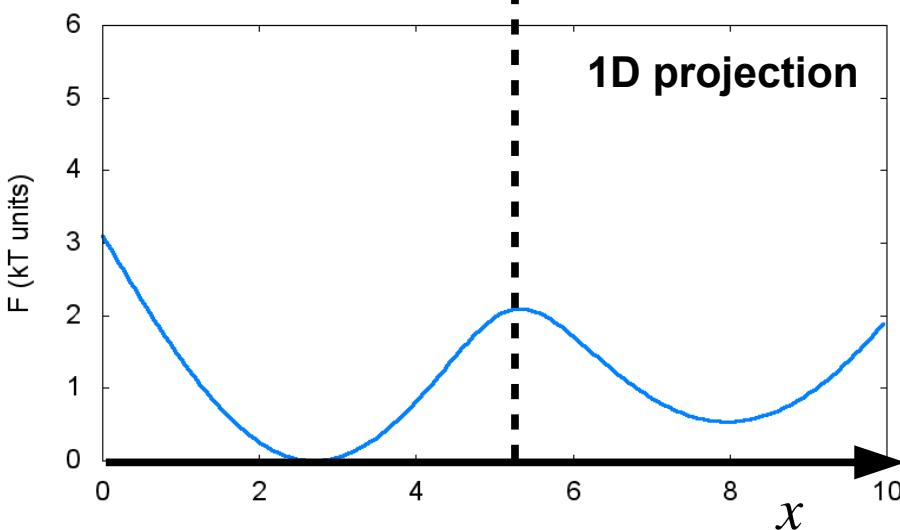
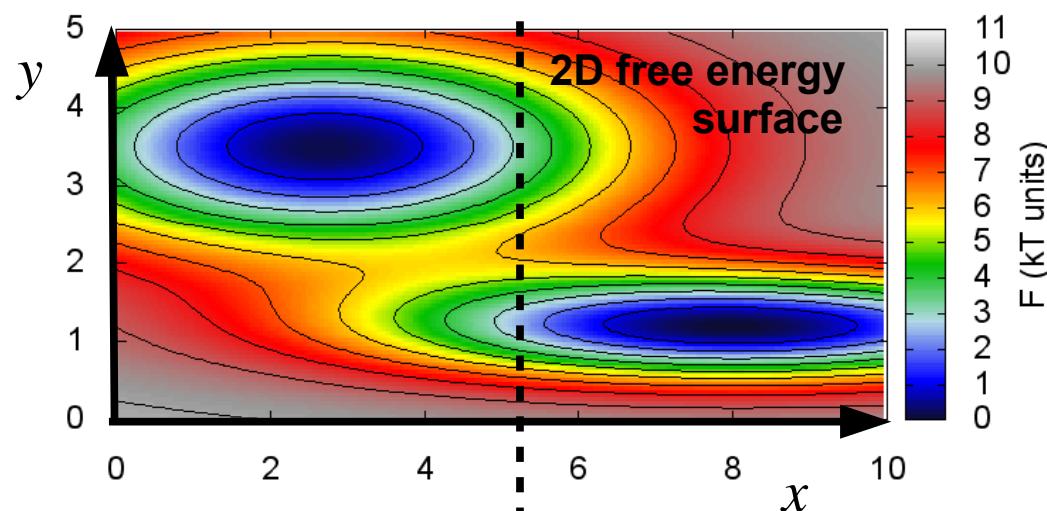
you must pull
the right strings!

M. Brando et al, 1972

order parameter:

a collective variable able to distinguish reactants from products

easy ...



reaction coordinate:

“a dynamically relevant measure for the progress of a reaction”

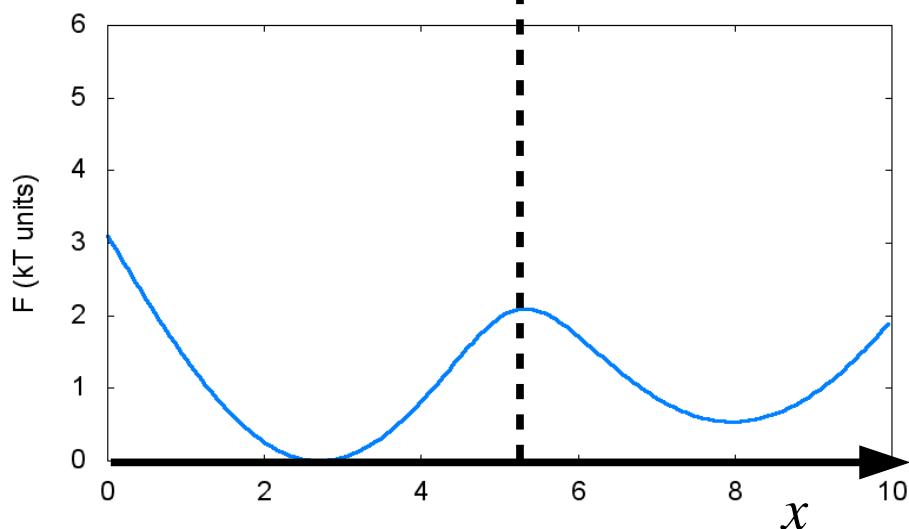
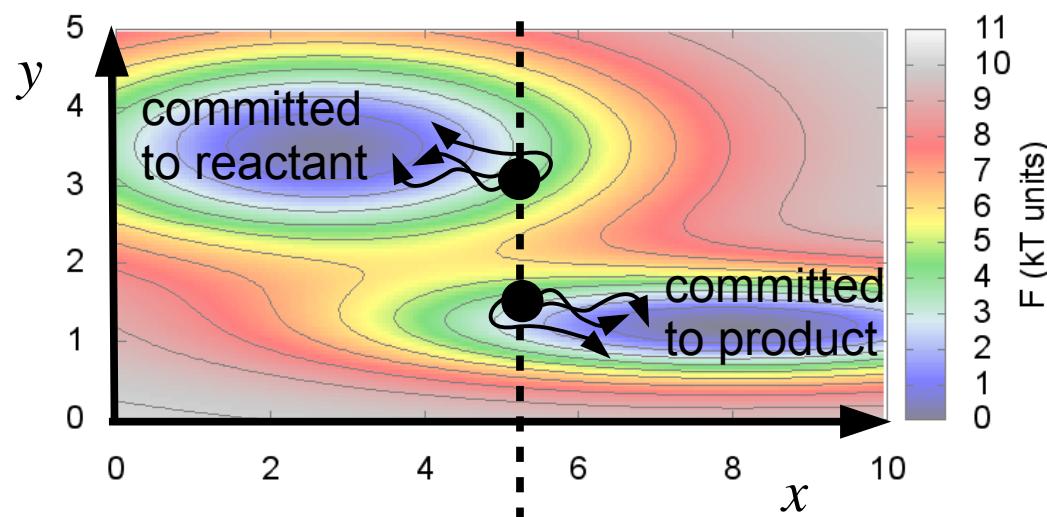
“a coordinate describing the dynamical mechanism of the transition”

difficult !

order parameter:

a collective variable able to distinguish reactants from products

easy ...



reaction coordinate:

“a dynamically relevant measure for the progress of a reaction”

“a coordinate describing the dynamical mechanism of the transition”

difficult !

**committor analysis:
a test for good RC**

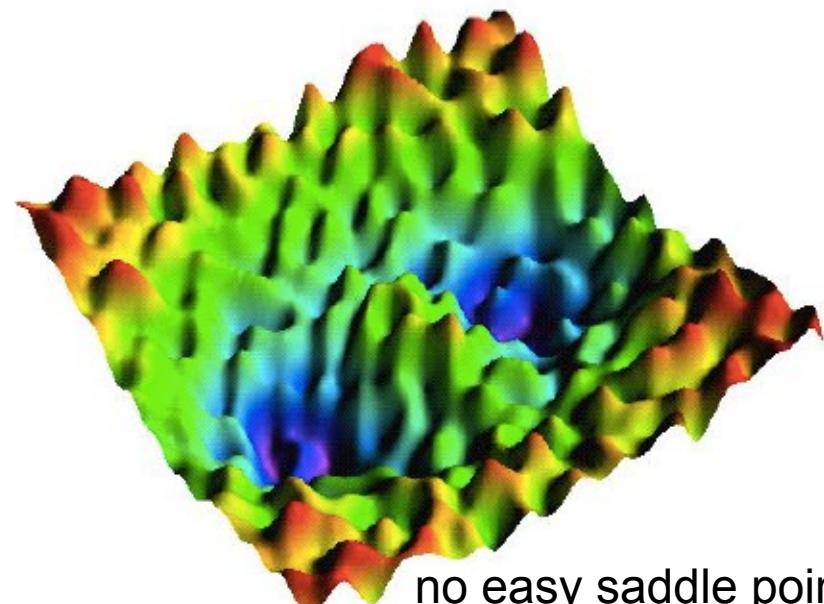
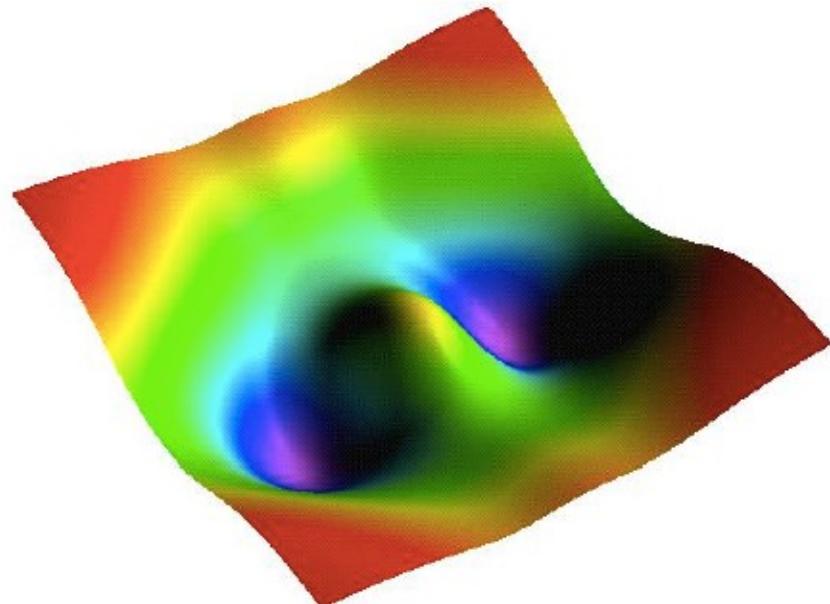
Dellago, Bolhuis, Geissler,
Adv Chem Phys 2002

x = good order parameter
bad reaction coordinate

general definition:

the reaction coordinate tells us the probability of each atomic configuration to evolve towards reactants or products

$T=0$, $T>0$, enthalpic and/or entropic barriers, chemical reactions, protein folding, diffusive processes...



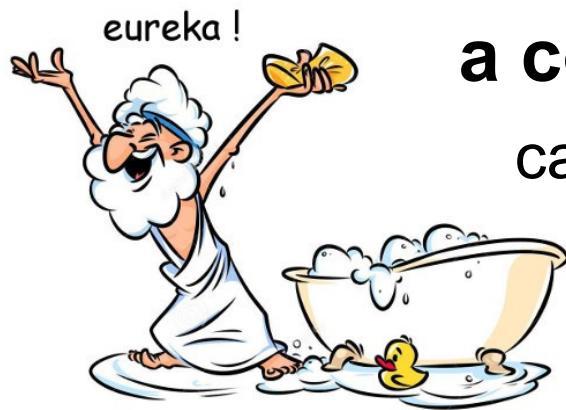
no easy saddle point here

the **reaction coordinate** has twofold importance:



a simulation tool

umbrella sampling, constrained dynamics,
metadynamics, ABF, steering, ...



a conceptual tool

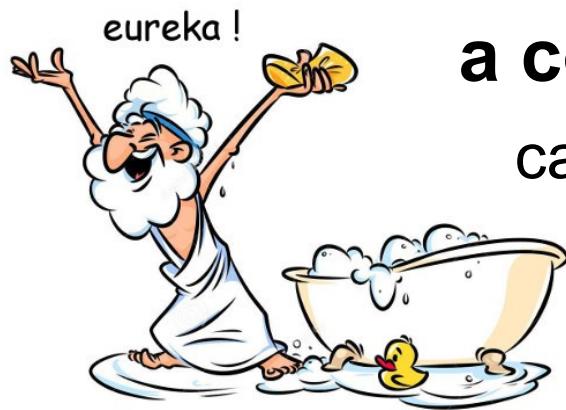
capturing all key features of the mechanism

the **reaction coordinate** has twofold importance:



a simulation tool

umbrella sampling, constrained dynamics,
metadynamics, ABF, steering, ...



a conceptual tool

capturing all key features of the mechanism

customary approach:

**system-tailored definitions based on
heuristics, intuition, trial and error ...**

can we do better ?

how to describe in a general way
the possible transformations of matter ?

graph theory provides powerful tools in a number of domains:

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



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

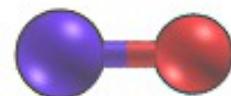
Reconstruction of Protein Structures from a Vectorial Representation

Markus Porto,^{1,2} Ugo Bastolla,³ H. Eduardo Roman,⁴ and Michele Vendruscolo⁵

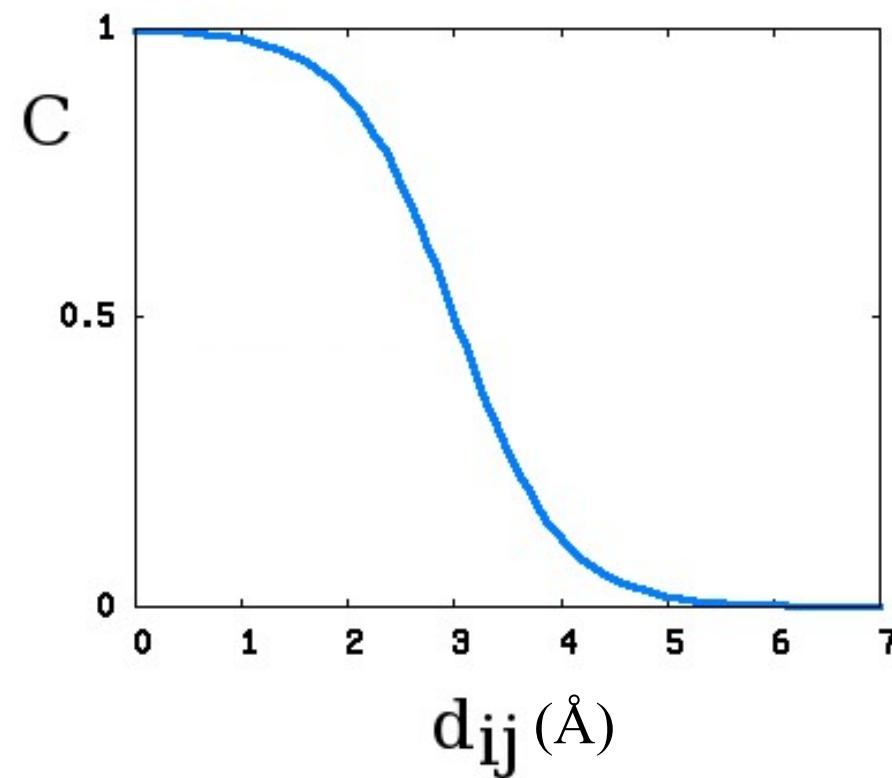
first step: build a graph from a system of atoms

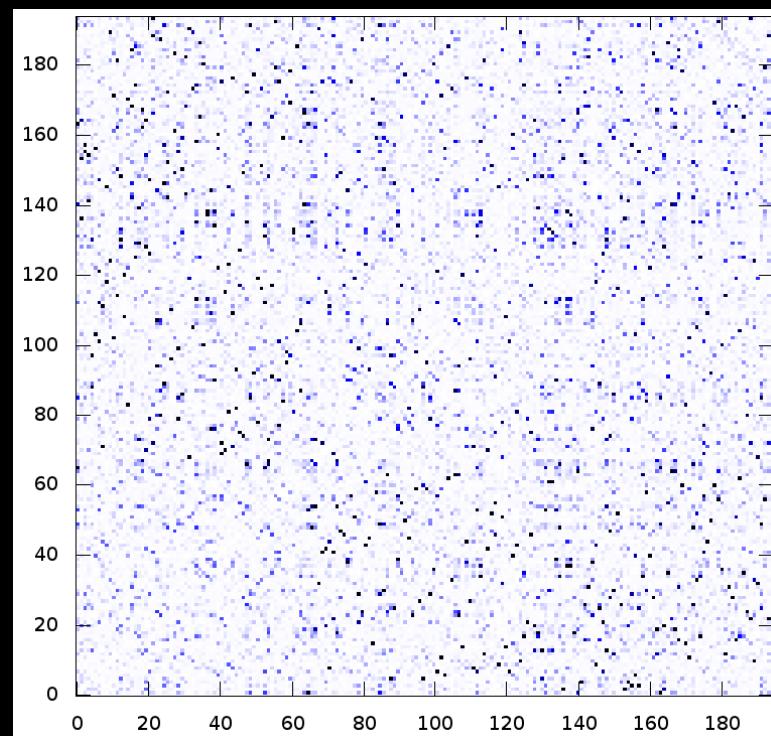
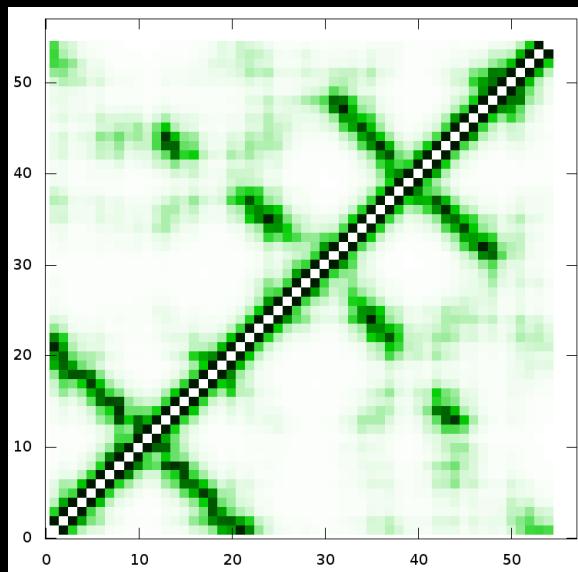
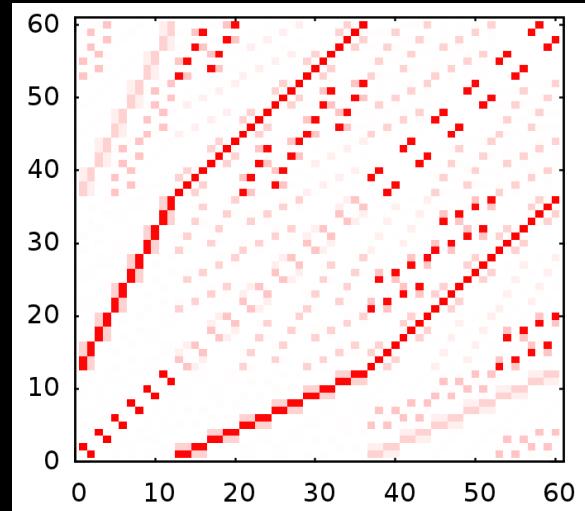
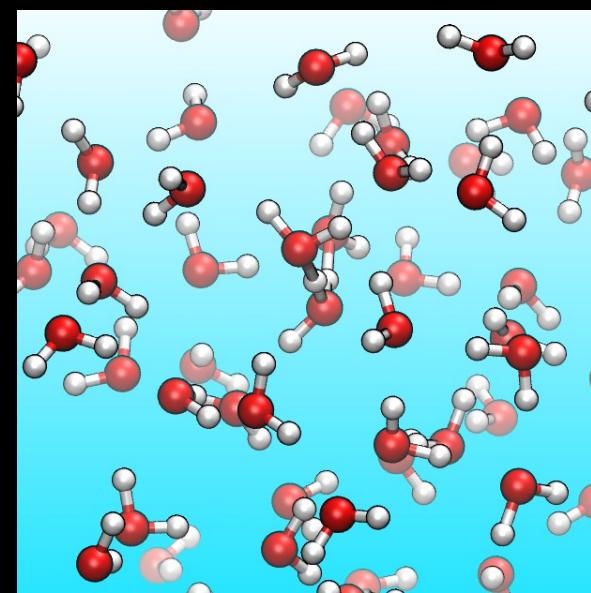
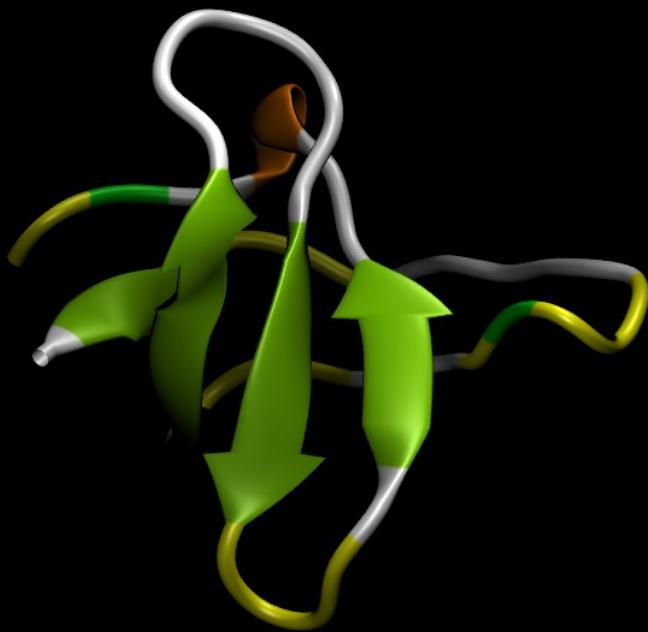
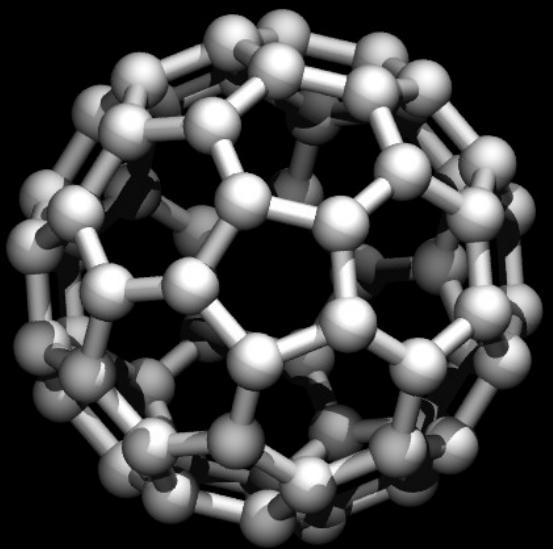
all you need is a switching-function $C(d_{ij})$

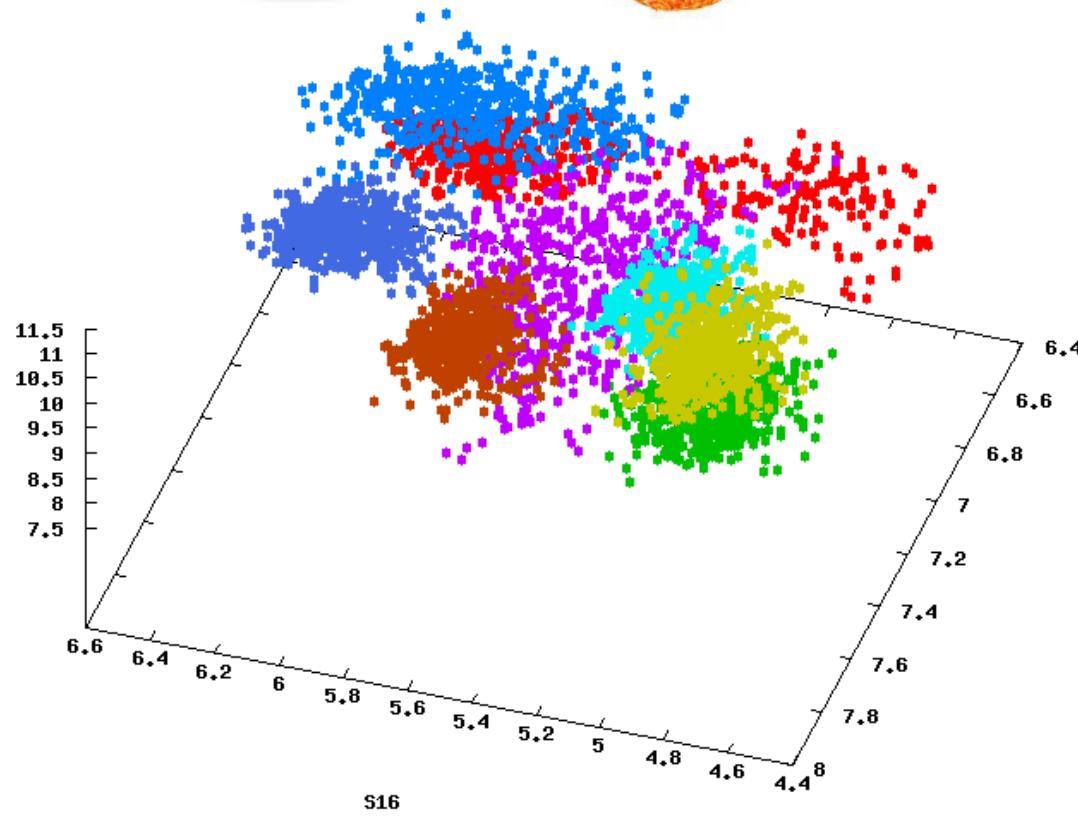
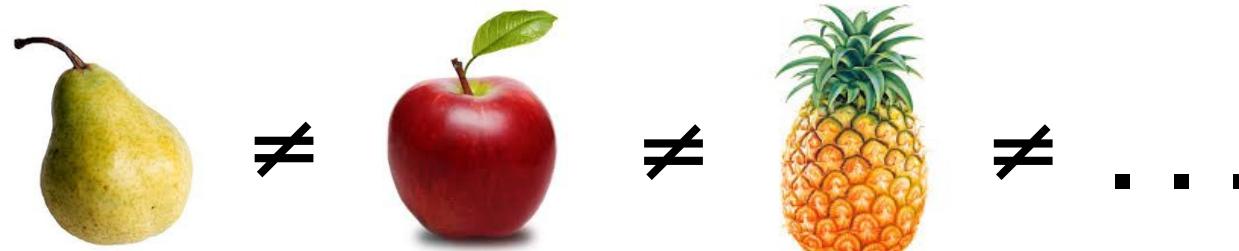
$$C = 1$$



$$C = 0$$



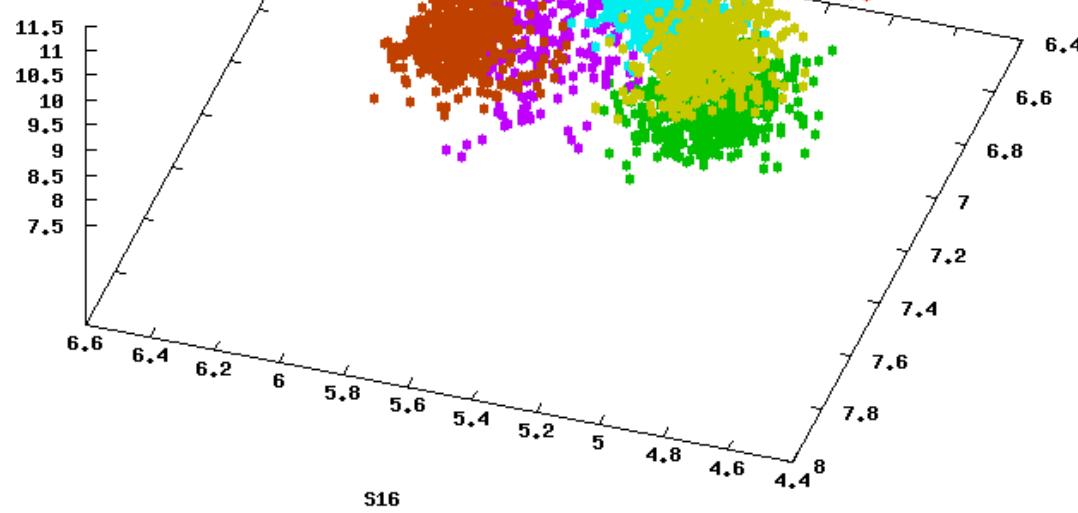




many possible definitions of topological coordinates, how to choose one?

two criteria:

- 1) coordinates must distinguish metastable states
- 2) coordinates must be invariant upon permutation of identical atoms



many possible definitions of topological coordinates, how to choose one?

two criteria:

- 1) coordinates must distinguish metastable states
- 2) coordinates must be invariant upon permutation of identical atoms

you can find many definitions of structural fingerprints:

Randic, J. Chem. Inf. Comput. Sci. 15, 105 (1975)

Balaban, Ciubotariu, and Medeleanu, J. Chem. Inf. Comput. Sci. 31, 517 (1991)

Bera et al., J. Phys. Chem. A 110, 4287 (2006)

Vasquez-Perez et al., J. Chem. Phys. 131, 124126 (2009)

Valle & Oganov, Acta Cryst. A66, 507 (2010)

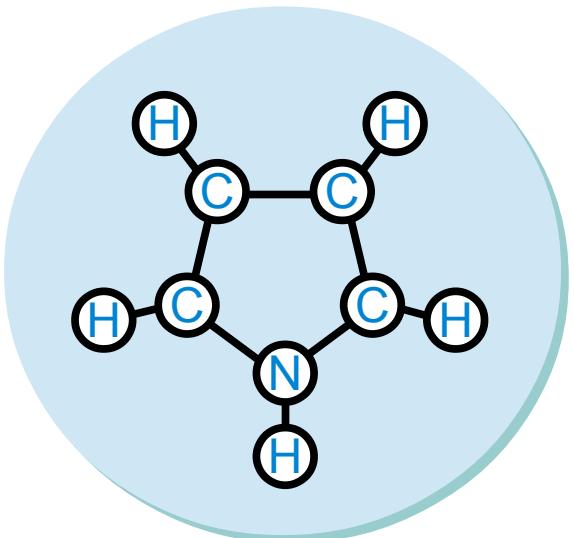
Sadeghi, Ghasemi, Schaefer, Mohr, Lill, Goedecker, JCP 139, 184118 (2013)

De, Bartok, Csanyi, Ceriotti, PCCP 18, 13754 (2016)

... a very incomplete list

theory

**complete description
of the topology**



adjacency matrix a_{ij}

practical simulations

approximations:
dimensional reduction,
coarse graining

$$\lambda^{max}, v_i^{max}$$

principal eigenvalue/vector

Pietrucci & Andreoni, PRL 2011

$$p = \text{sort}(a_{ij})$$

permutation invariant vector

Gallet & Pietrucci, JCP 2013
Pipolo et al, arXiv 2017

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

coordination patterns

Pietrucci & Saitta, PNAS 2015

gas-phase
reactions,
nanoclusters

phase trans.
in solids & liq.

reactions in
gas phase
& solution

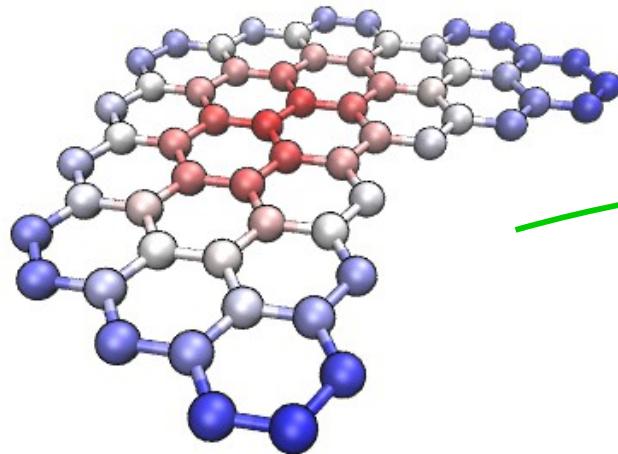
advantages of the new coordinates:

- robust and reversible simulations of chemical reactions, including the solvent
- systematic simulation of transitions across a phase diagram, including crystal nucleation from liquid or amorphous
- transformations among nanostructures, including changes of symmetry and/or shape

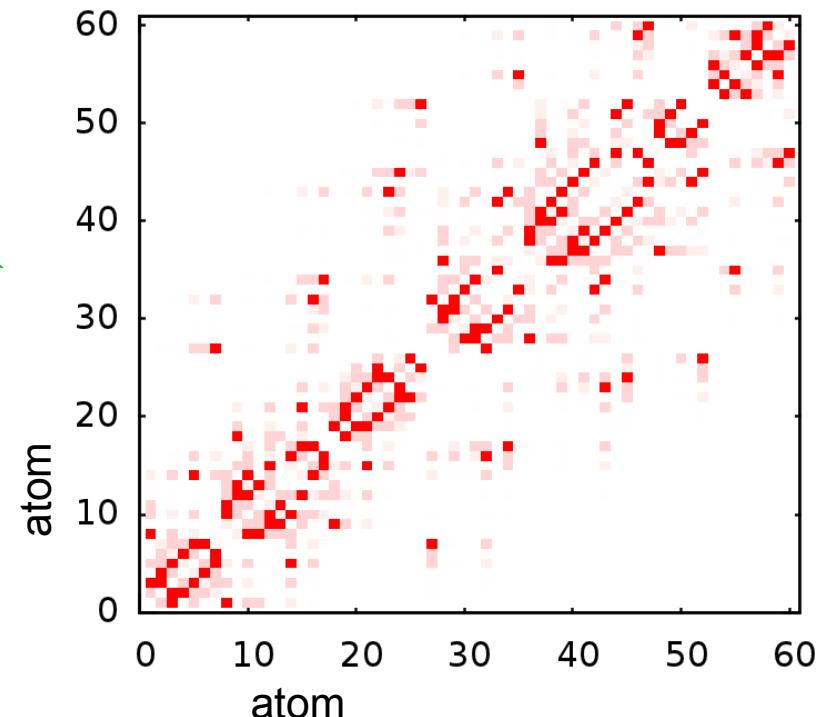
no need to guess the pathways in advance

SPRINT: a new class of reaction coordinates

atomic structure



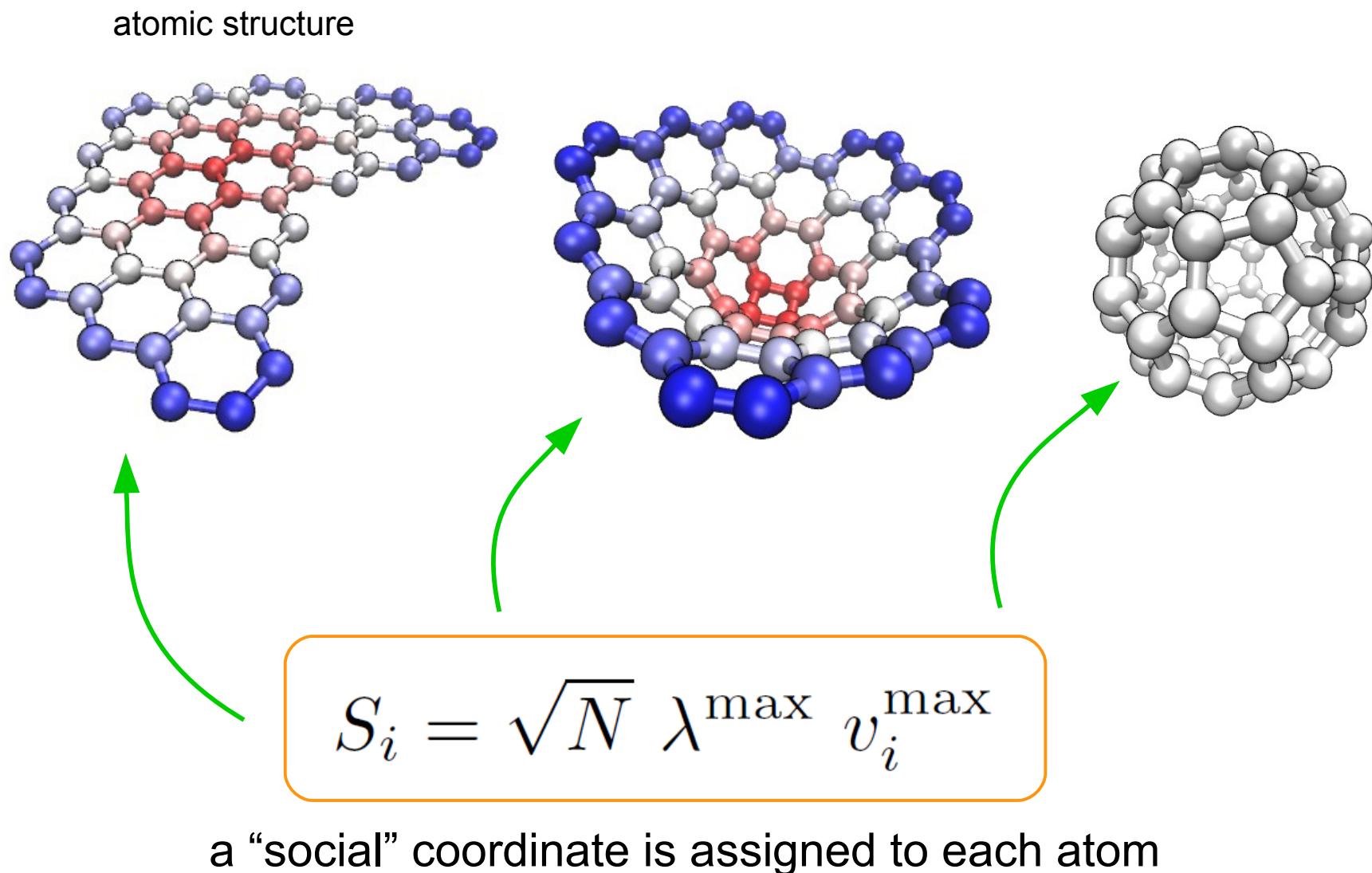
adjacency matrix



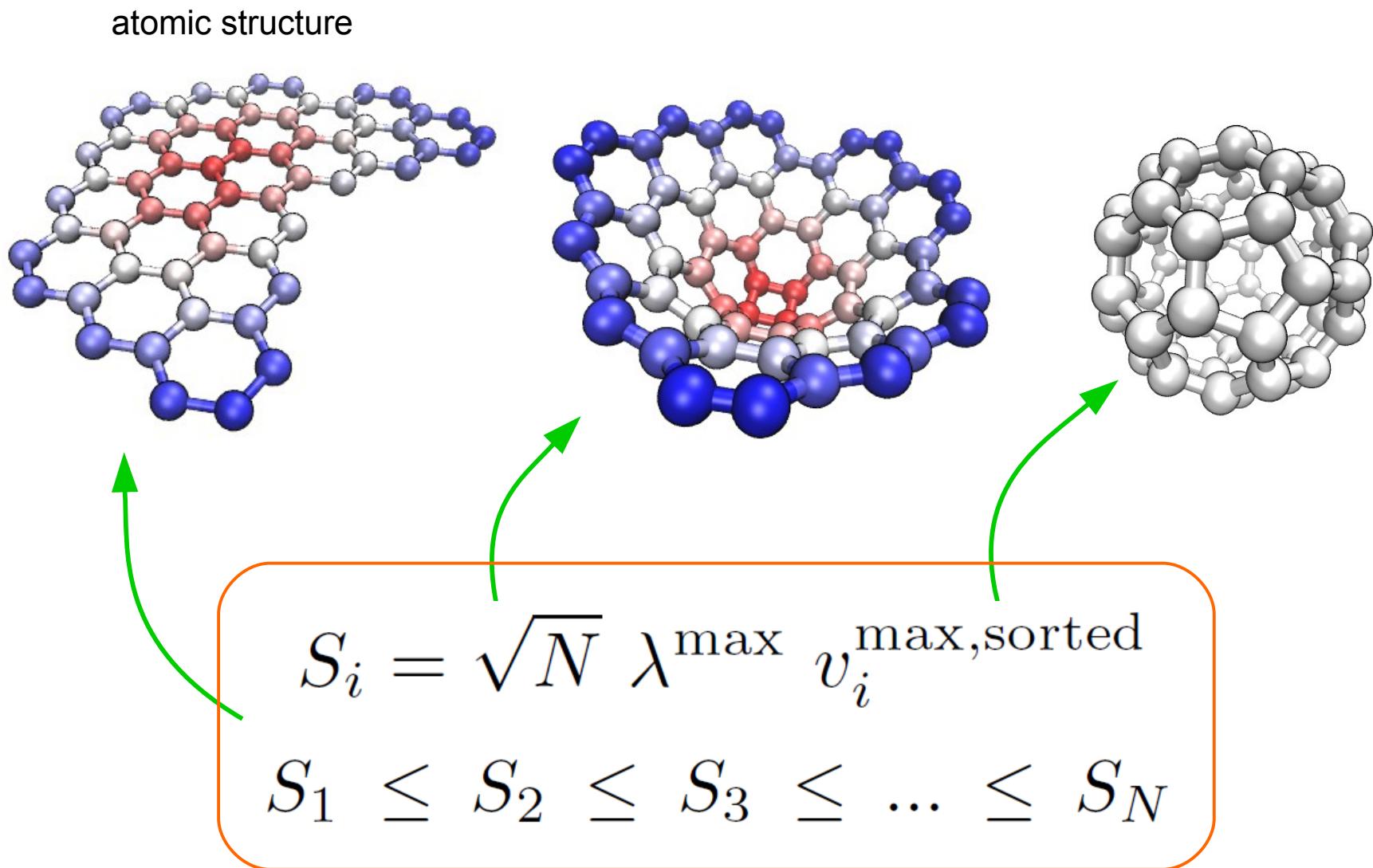
$$S_i = \sqrt{N} \lambda^{\max} v_i^{\max}$$

a “social” coordinate is assigned to each atom

SPRINT: a new class of reaction coordinates



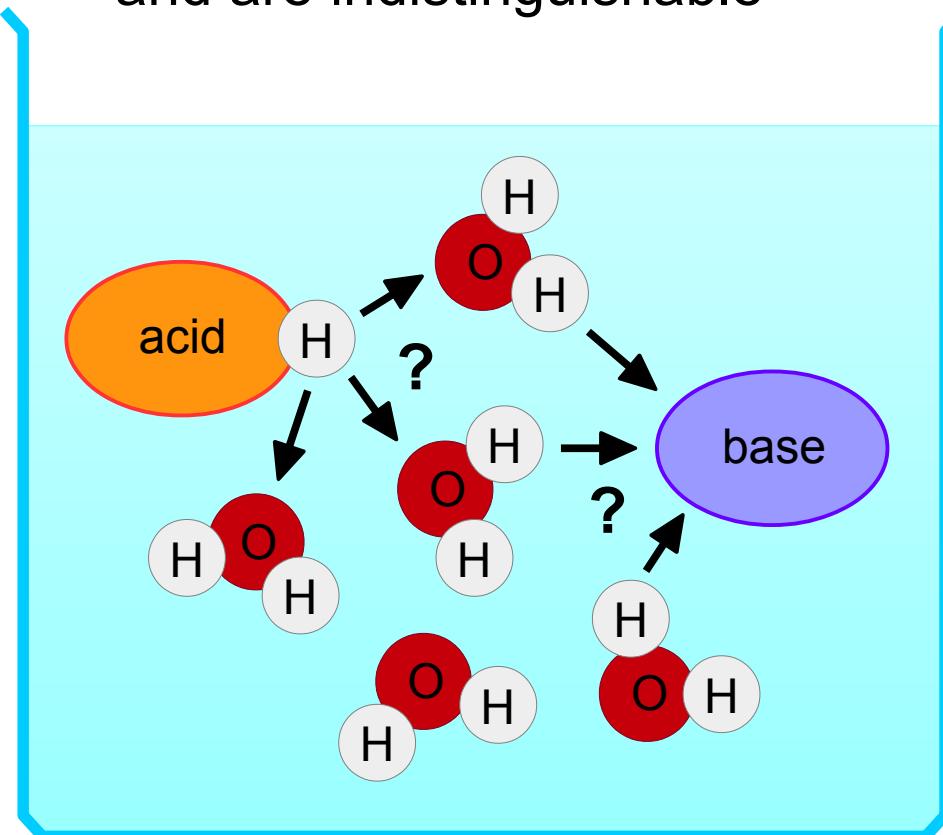
SPRINT: a new class of reaction coordinates



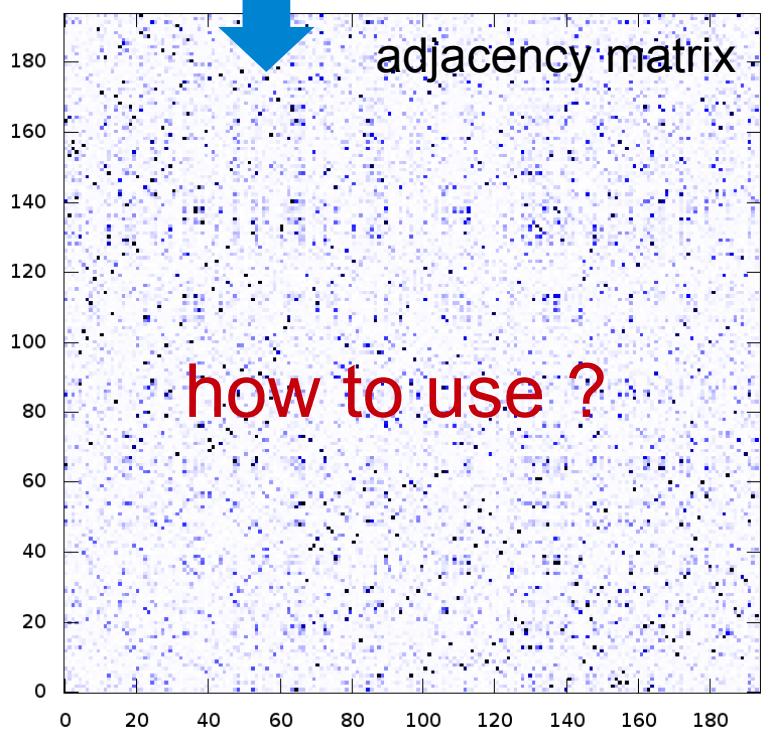
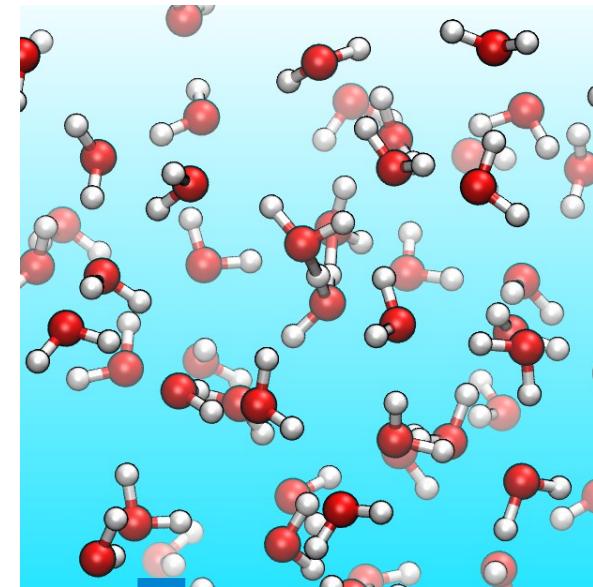
Social PeRmutation-INvarianT coordinates

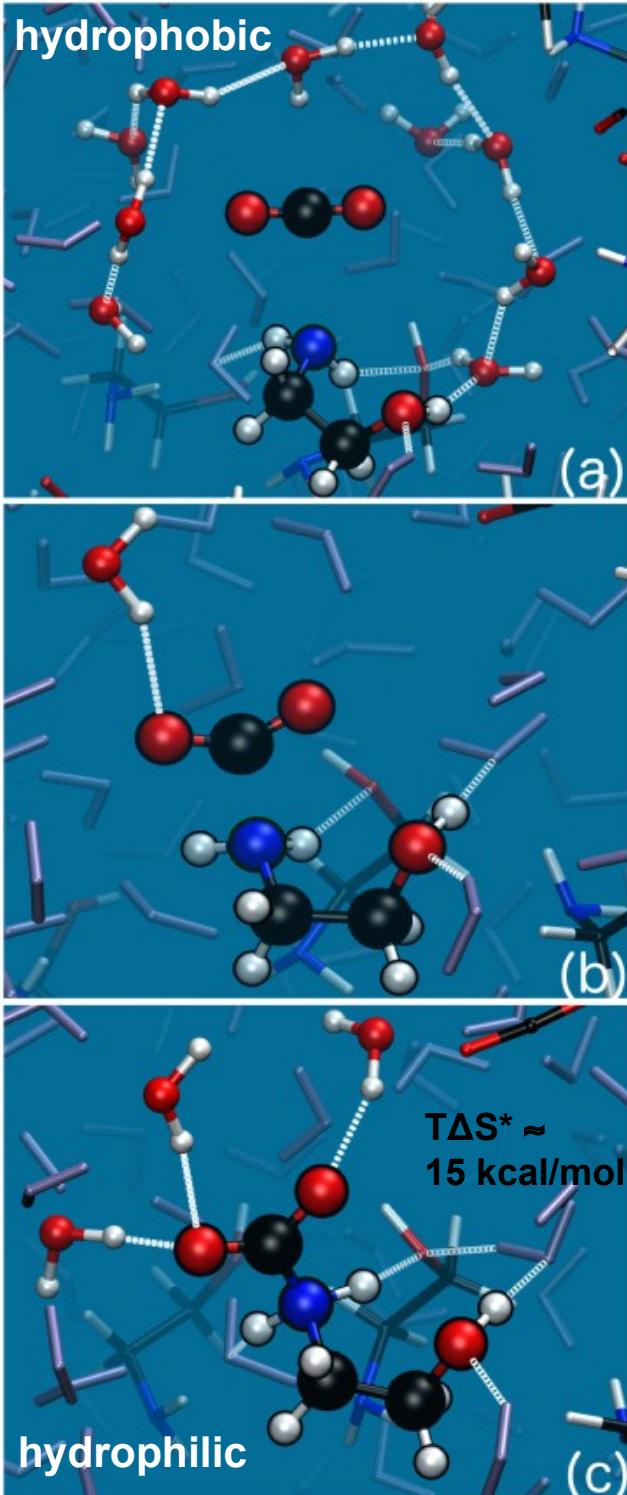
difficulties dealing with solutions:

- many degrees of freedom
- solvent molecules diffuse and are indistinguishable

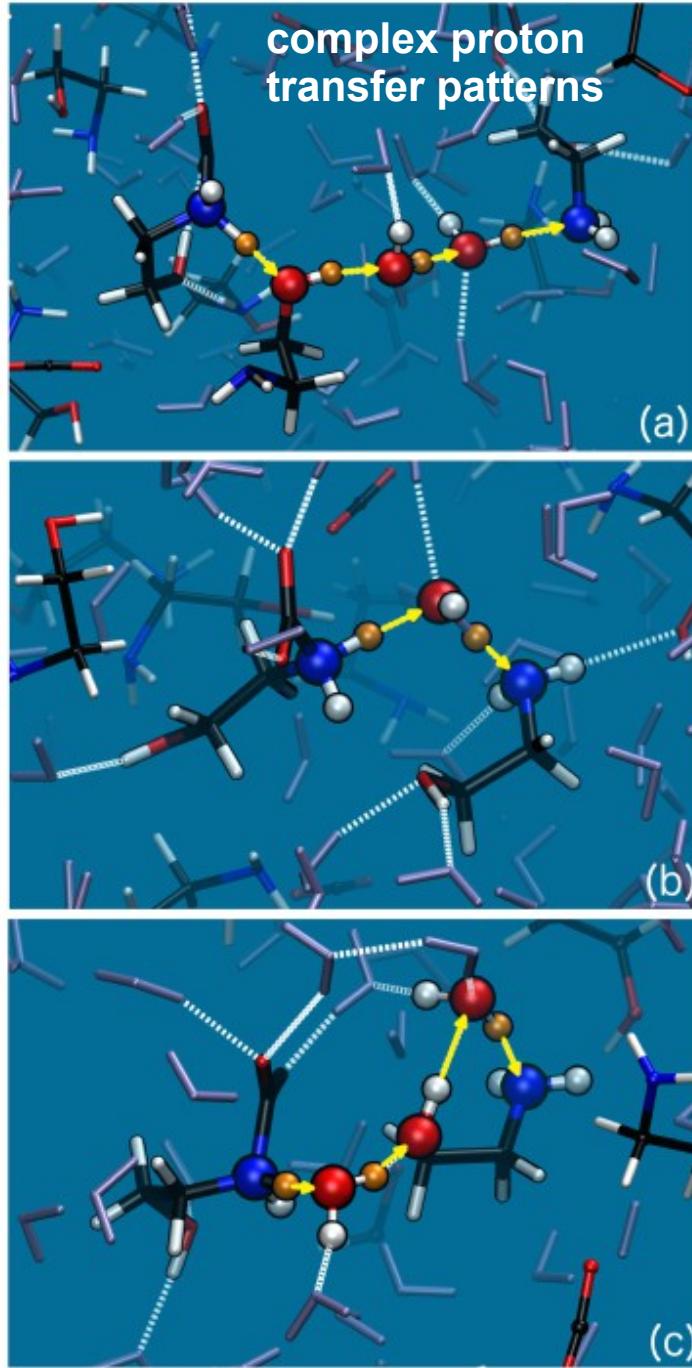


which solvent molecule reacts?
it depends on details of configuration at time t





e.g.: CO_2 capture in amine solutions



30 wt% aq. solution of monoethanolamine

Born-Oppenheimer MD

> 400 atoms

$T = 300\text{-}400 \text{ K}$

DFT-PBE + Grimme

metadynamics,
umbrella sampling, ...

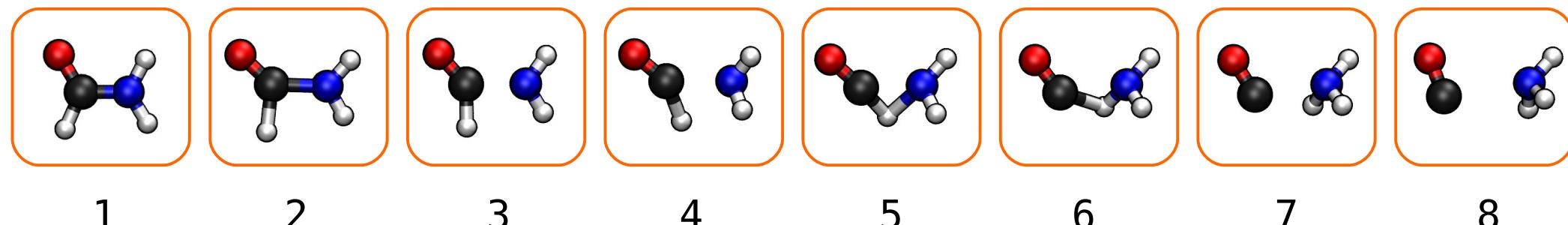
Guido, Pietrucci, Gallet,
Andreoni, JCTC 2013

Ma, Pietrucci, Andreoni
J Phys Chem Lett 2014
JCTC 2015

path collective variables

Branduardi, Gervasio, Parrinello, *J Chem Phys* **126**, 054103 (2007)

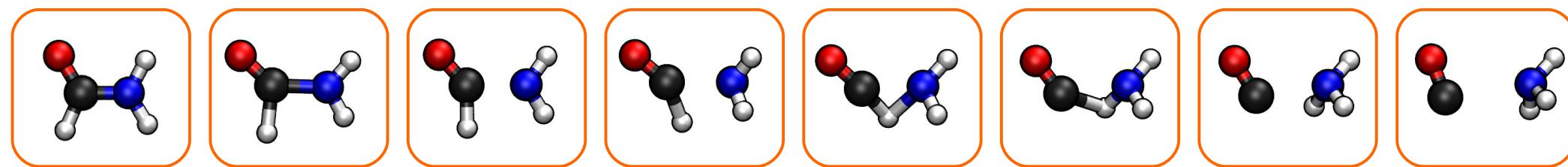
initial guess of path



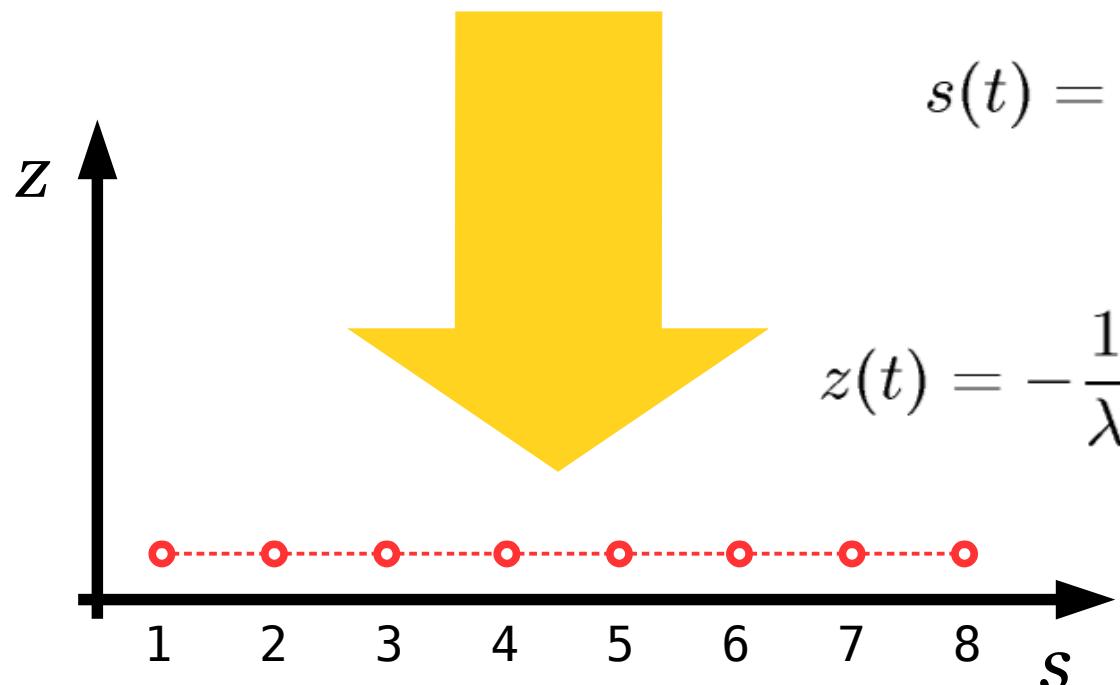
path collective variables

Branduardi, Gervasio, Parrinello, *J Chem Phys* **126**, 054103 (2007)

initial guess of path



1 2 3 4 5 6 7 8



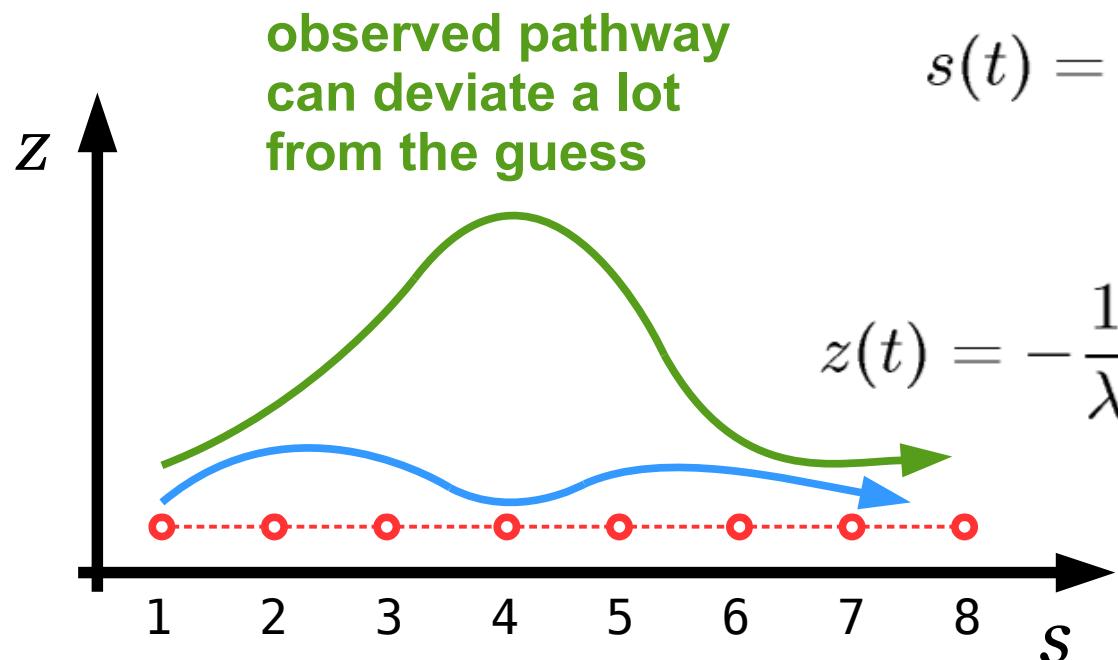
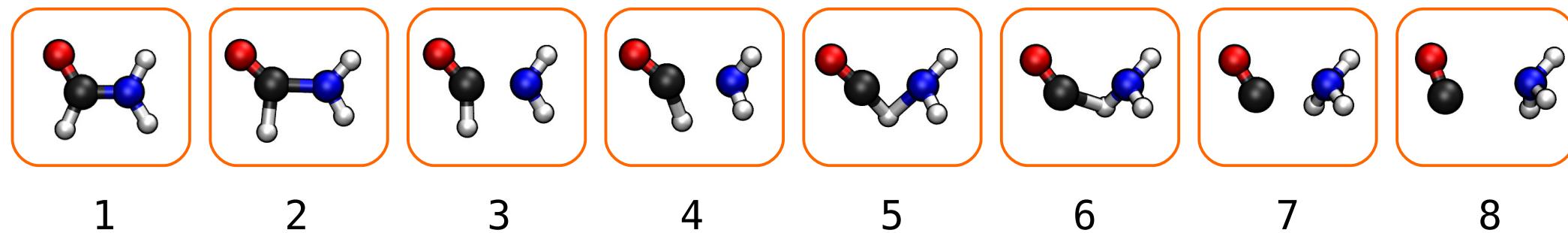
$$s(t) = \frac{\sum_{k=1}^{N_f} k e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)}}{\sum_{k'=1}^{N_f} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_{k'})}}$$

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

path collective variables

Branduardi, Gervasio, Parrinello, *J Chem Phys* **126**, 054103 (2007)

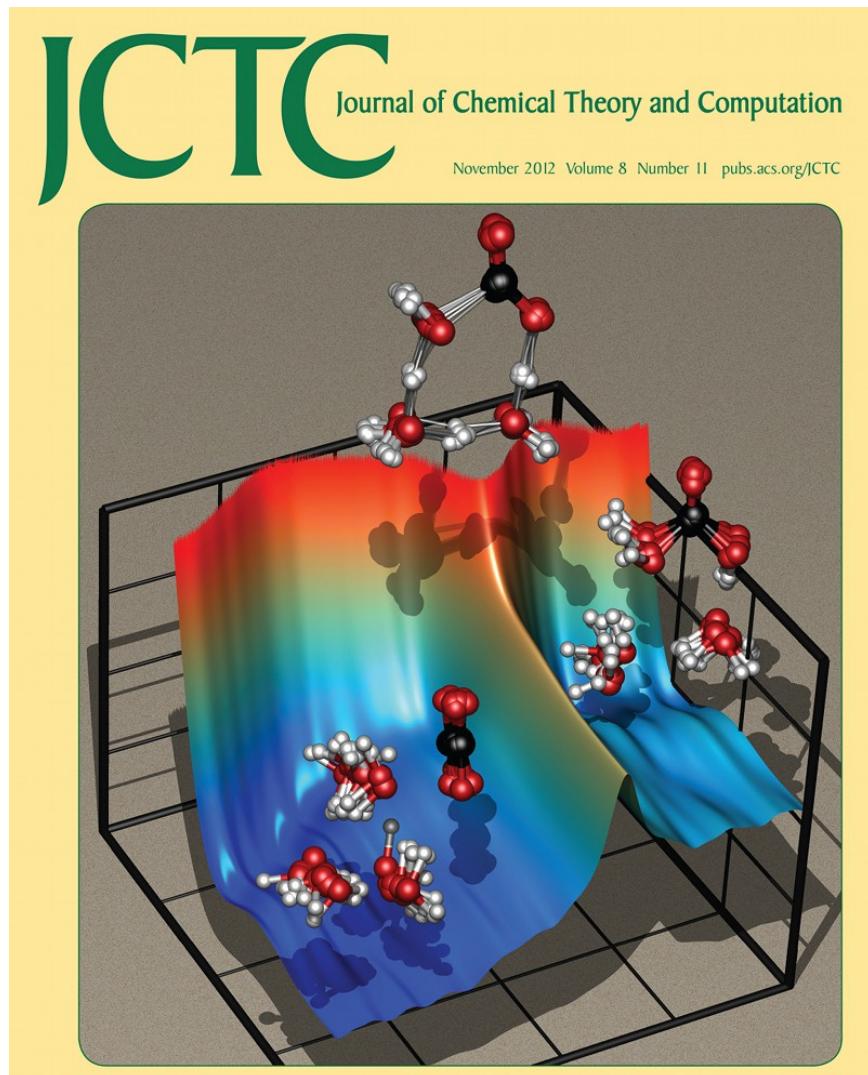
initial guess of path



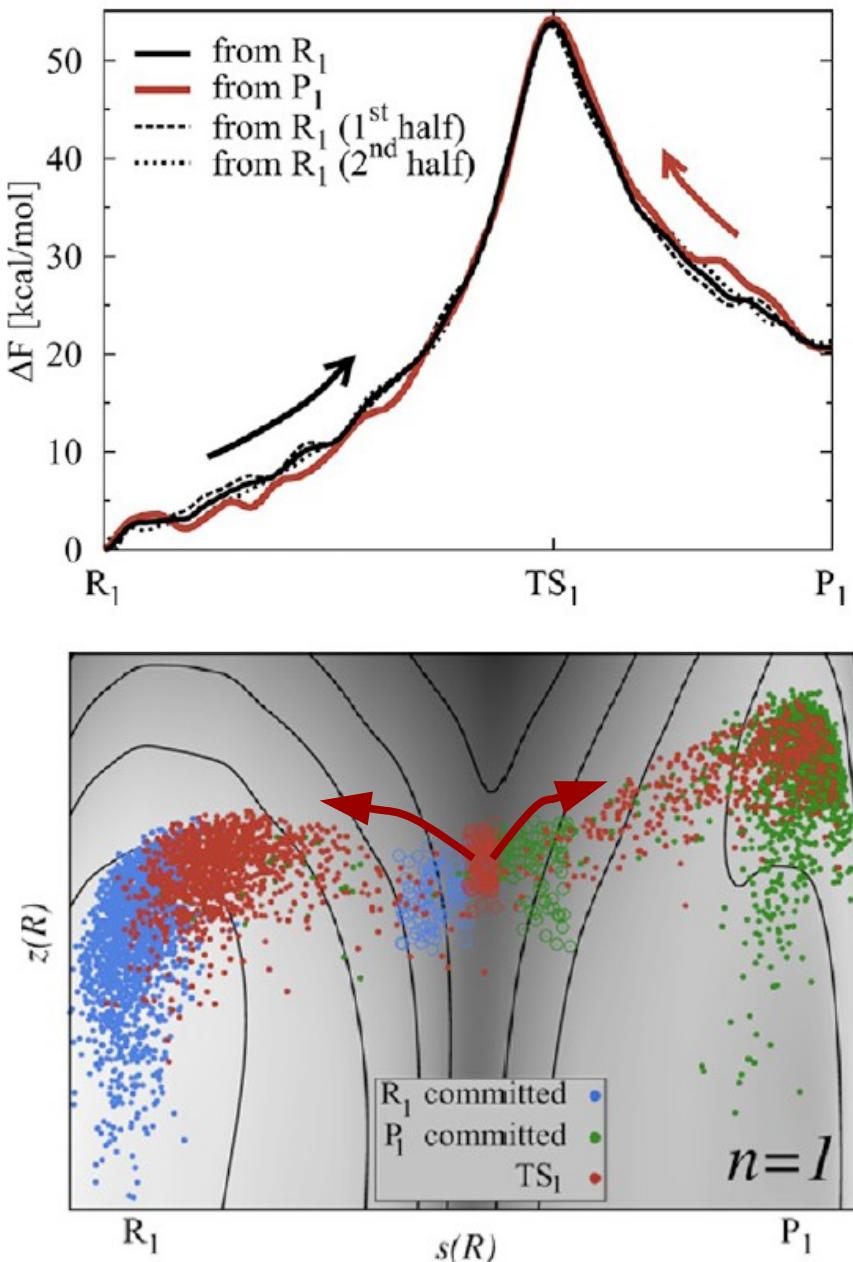
$$s(t) = \frac{\sum_{k=1}^{N_f} k e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)}}{\sum_{k'=1}^{N_f} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_{k'})}}$$

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

plain RMSD metric + path CVs: good for gas-phase

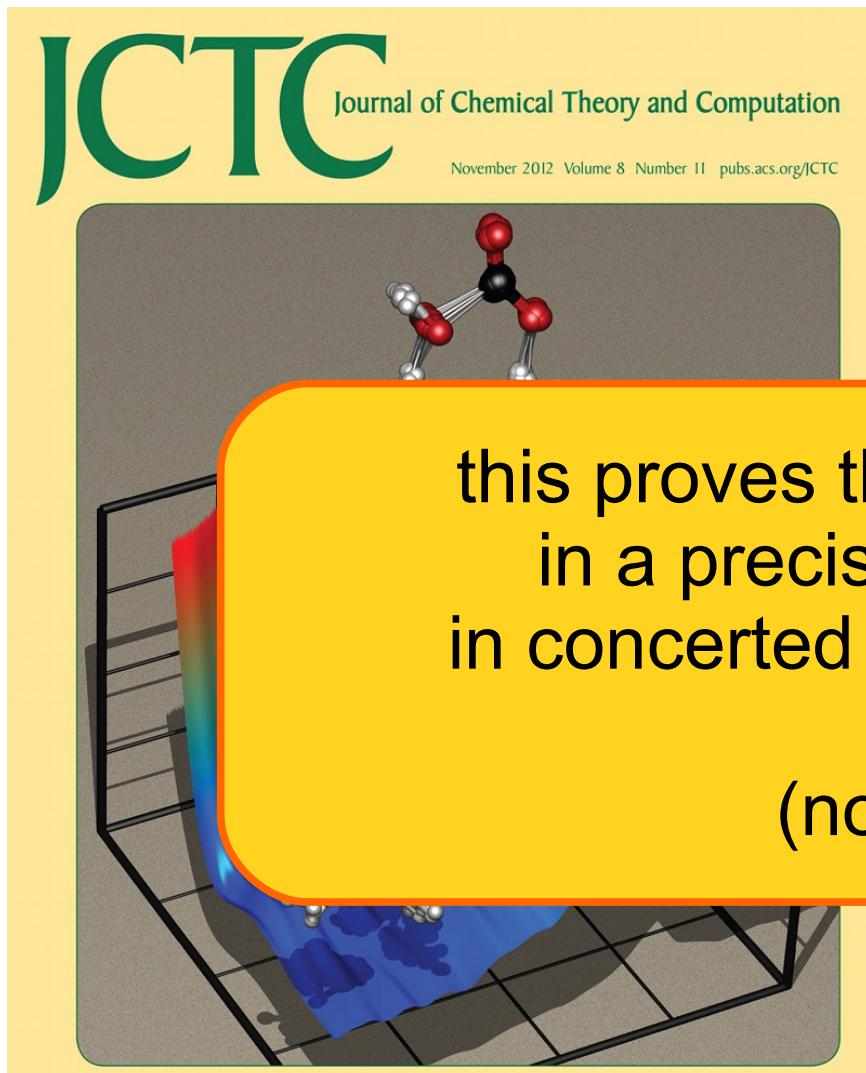


- fully reversible reactive simulations
- parallel growth of bias potential
- committer certifies reaction coordinates



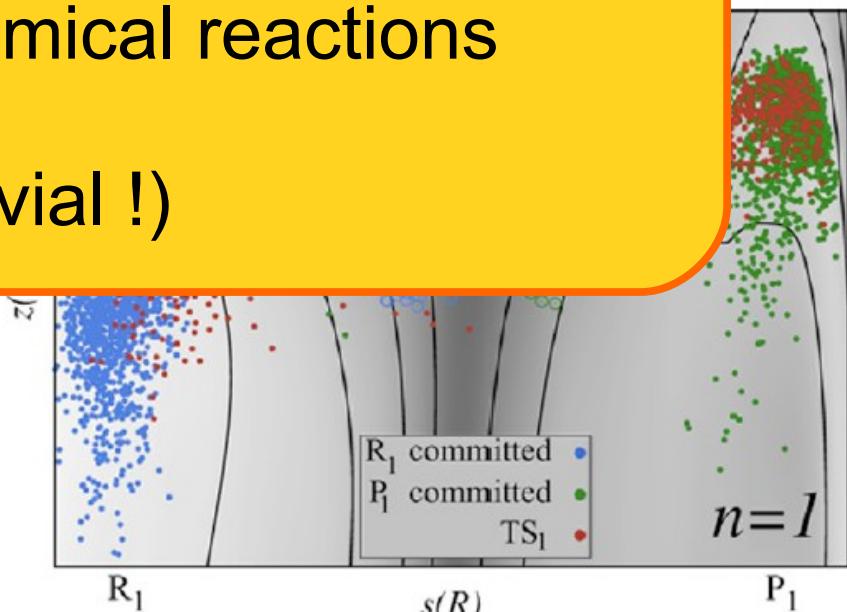
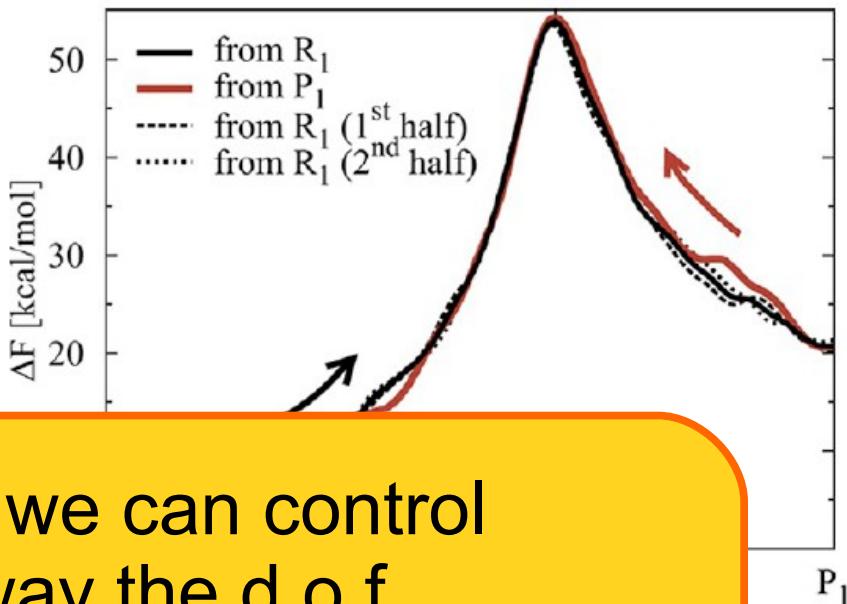
Branduardi et al, JCTC 2011
Gallet, Pietrucci, Andreoni, JCTC 2012

plain RMSD metric + path CVs: good for gas-phase



this proves that we can control
in a precise way the d.o.f.
in concerted chemical reactions

(not trivial !)



- fully reversible reactive simulations
- parallel growth of bias potential
- committer certifies reaction coordinates

a new approach to chemical reactions

with A.M. Saitta

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

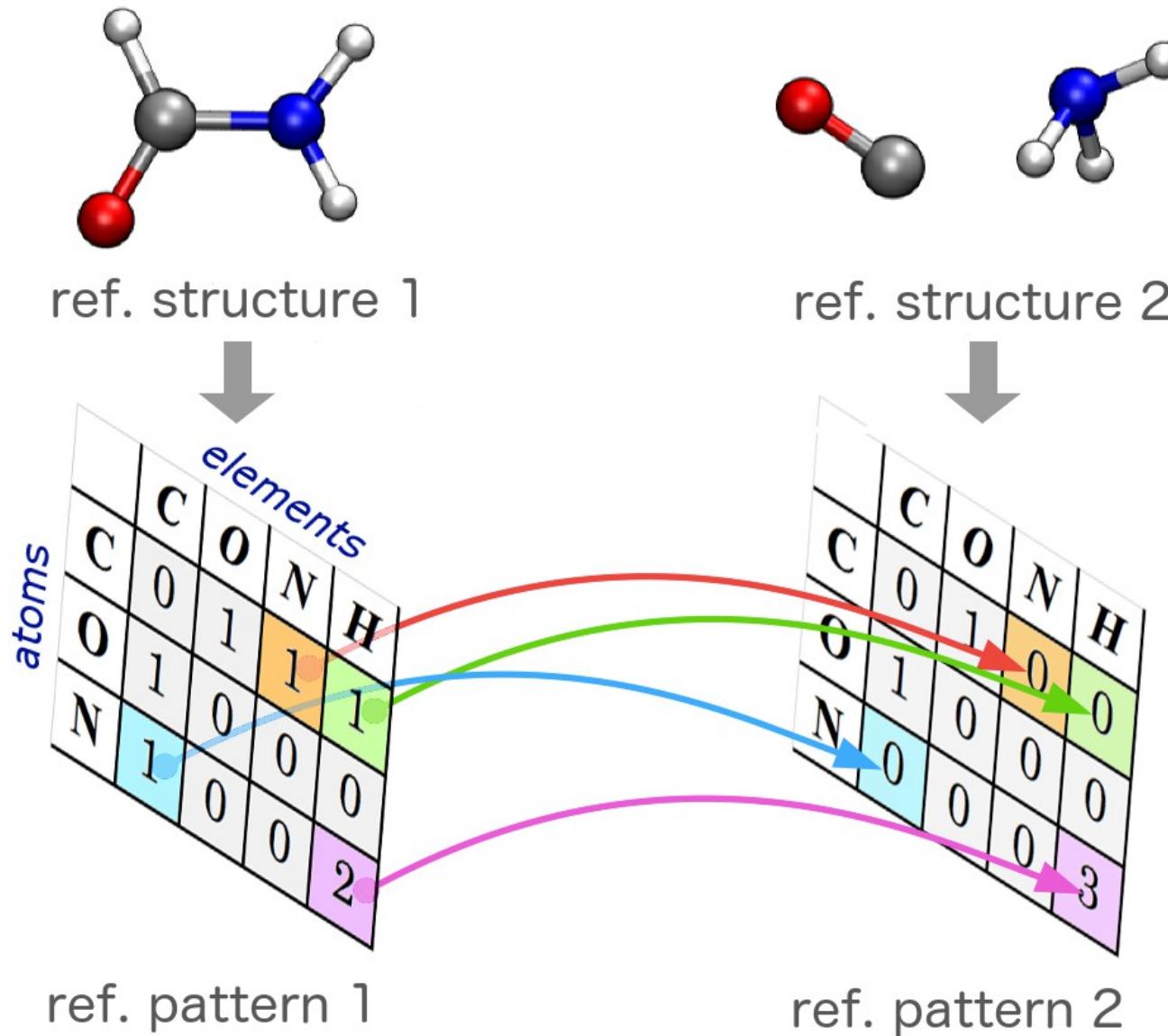
$$z(t) = -\frac{1}{\lambda} \log \left(\sum_{k=1}^{N_f} e^{-\lambda D(\mathbf{R}(t), \mathbf{R}_k)} \right)$$

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

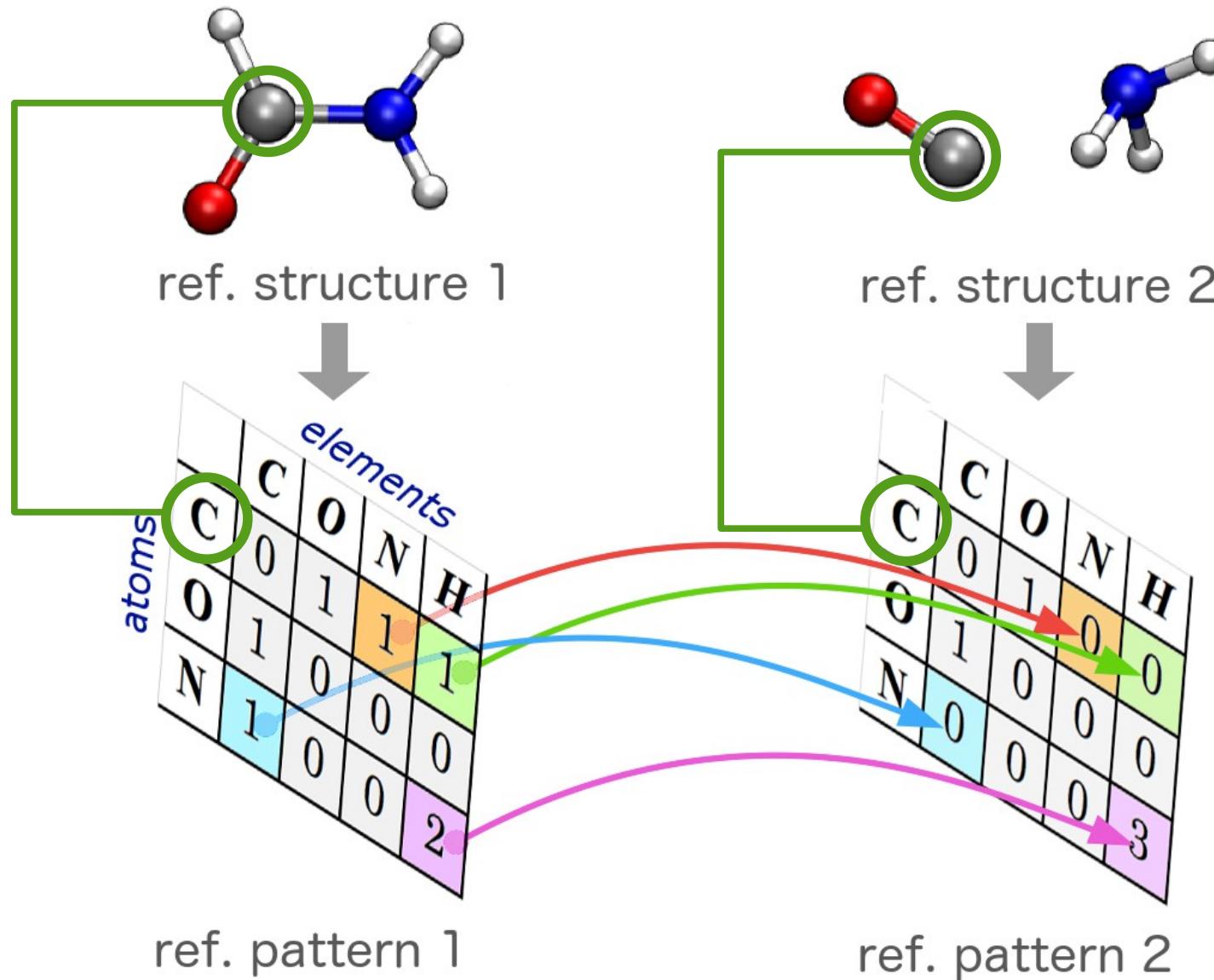
$$D(\mathbf{R}(t), \mathbf{R}_k) = \sum_{IS} \left(C_{IS}(t) - C_{IS}^k \right)^2$$

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

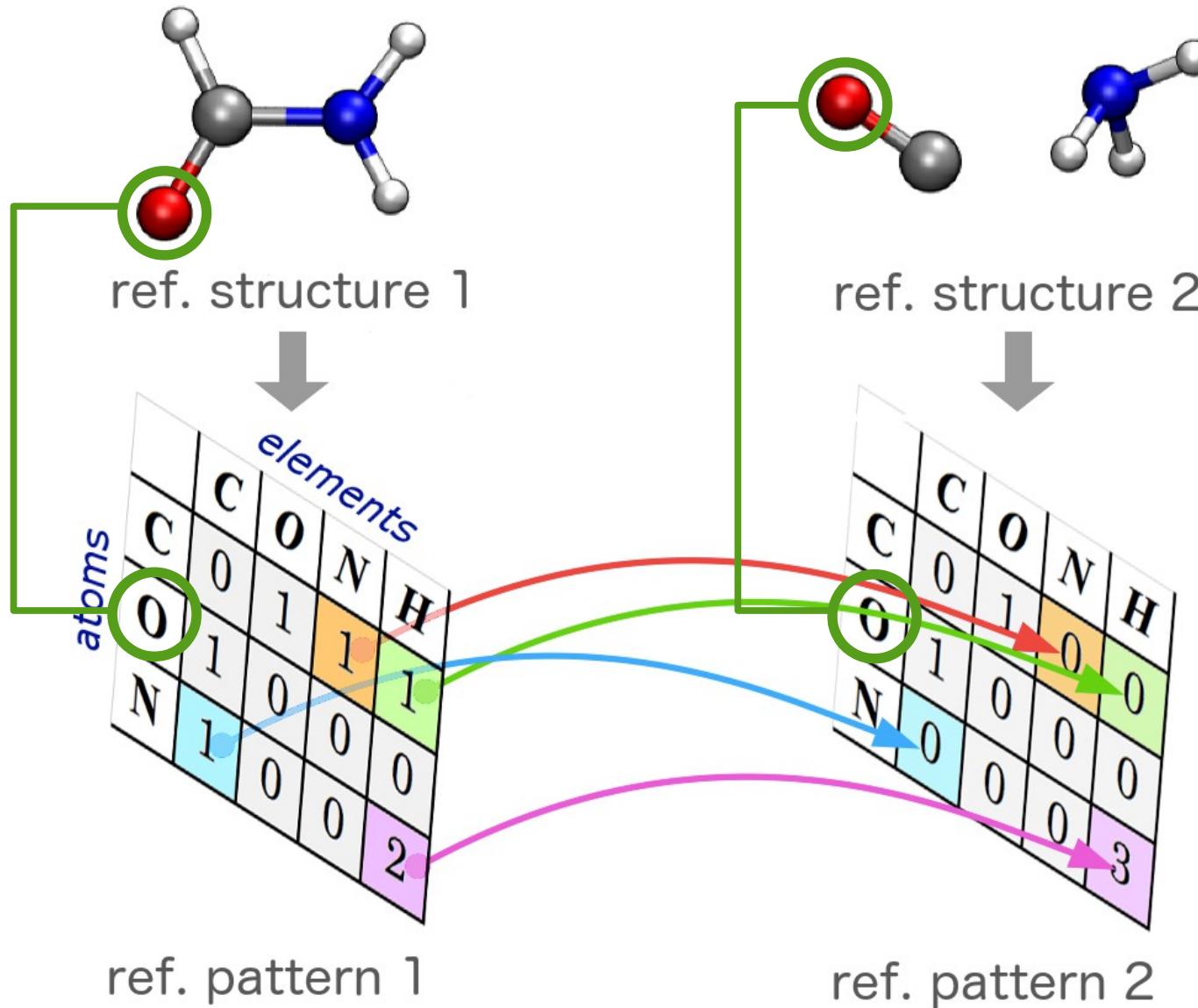
a new approach to chemical reactions



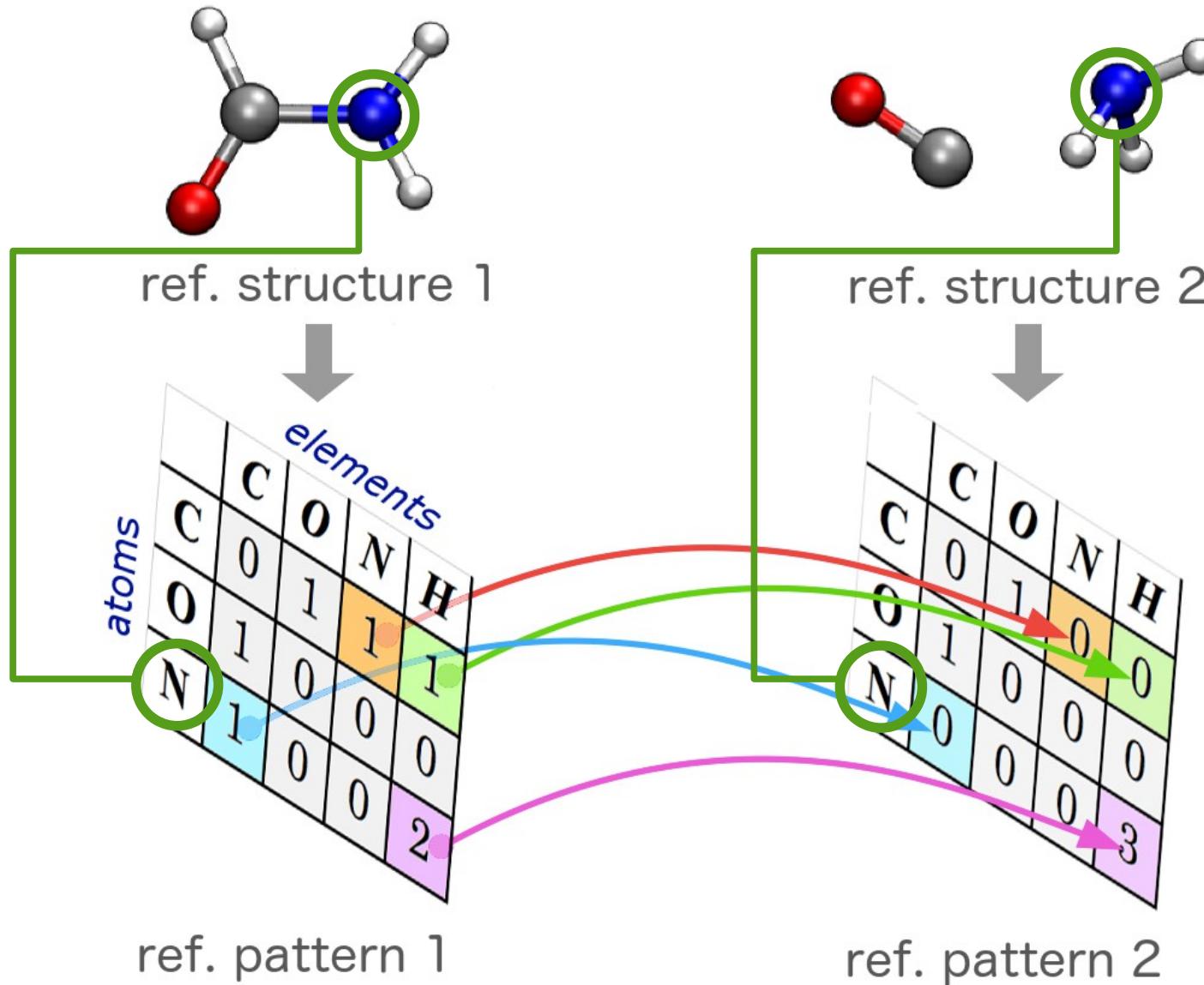
a new approach to chemical reactions



a new approach to chemical reactions



a new approach to chemical reactions



a new approach to chemical reactions

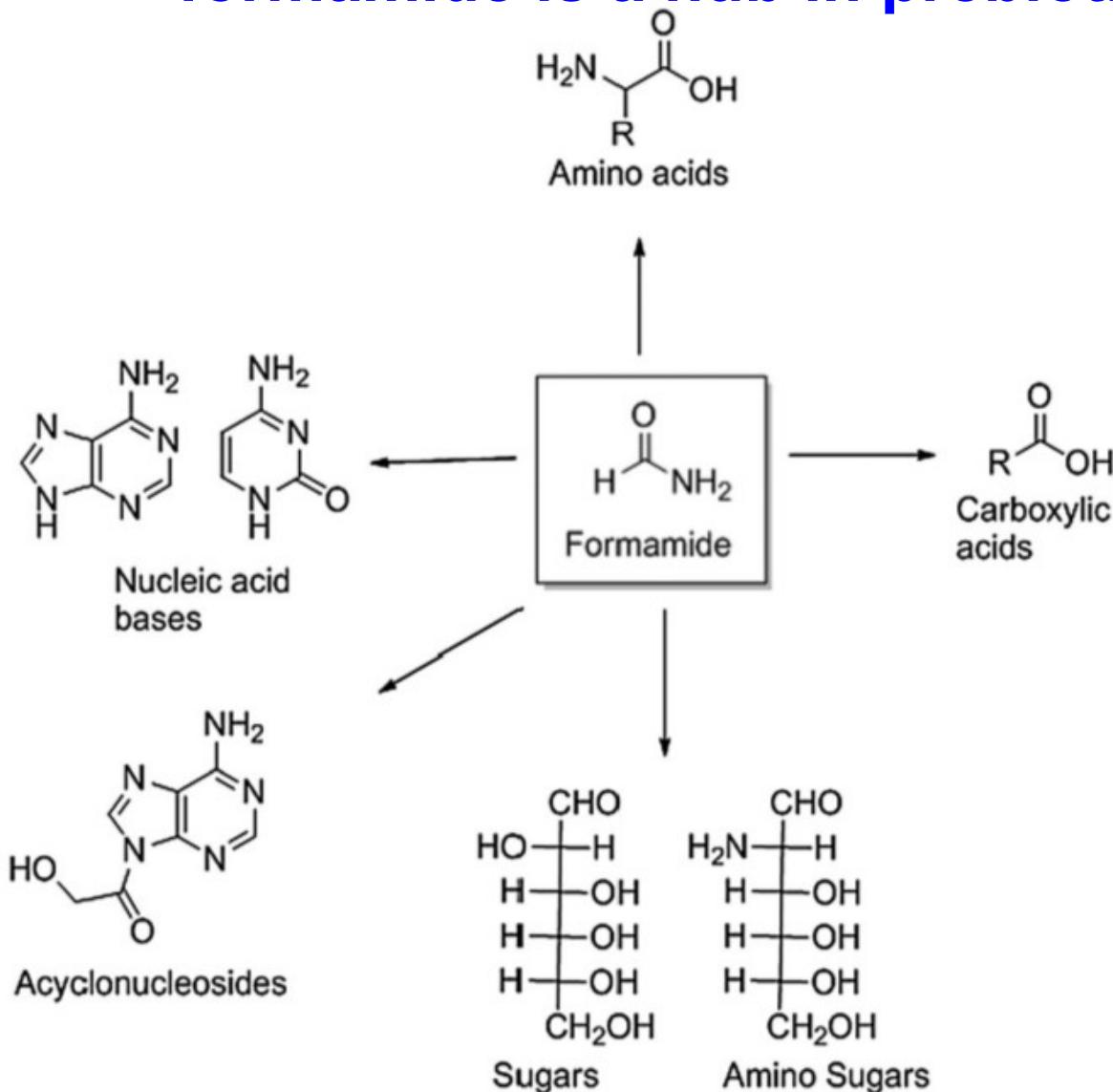
*input in **gas phase**:*

NAT	3	FRAMES	2
1			
1	1	0	1
2	2	1	0
3	3	0	1
2			
1	1	0	1
2	2	1	0
3	3	0	0

*input in **solution**:*

NAT	3	FRAMES	2
1			
1	1	1	0
2	2	2	1
3	3	3	0
2			
1	1	1	0
2	2	2	1
3	3	3	0

a good test case (many pathways): formamide is a hub in prebiotic chemistry



ubiquitous in the universe
(from star-forming regions
to comets and satellites)

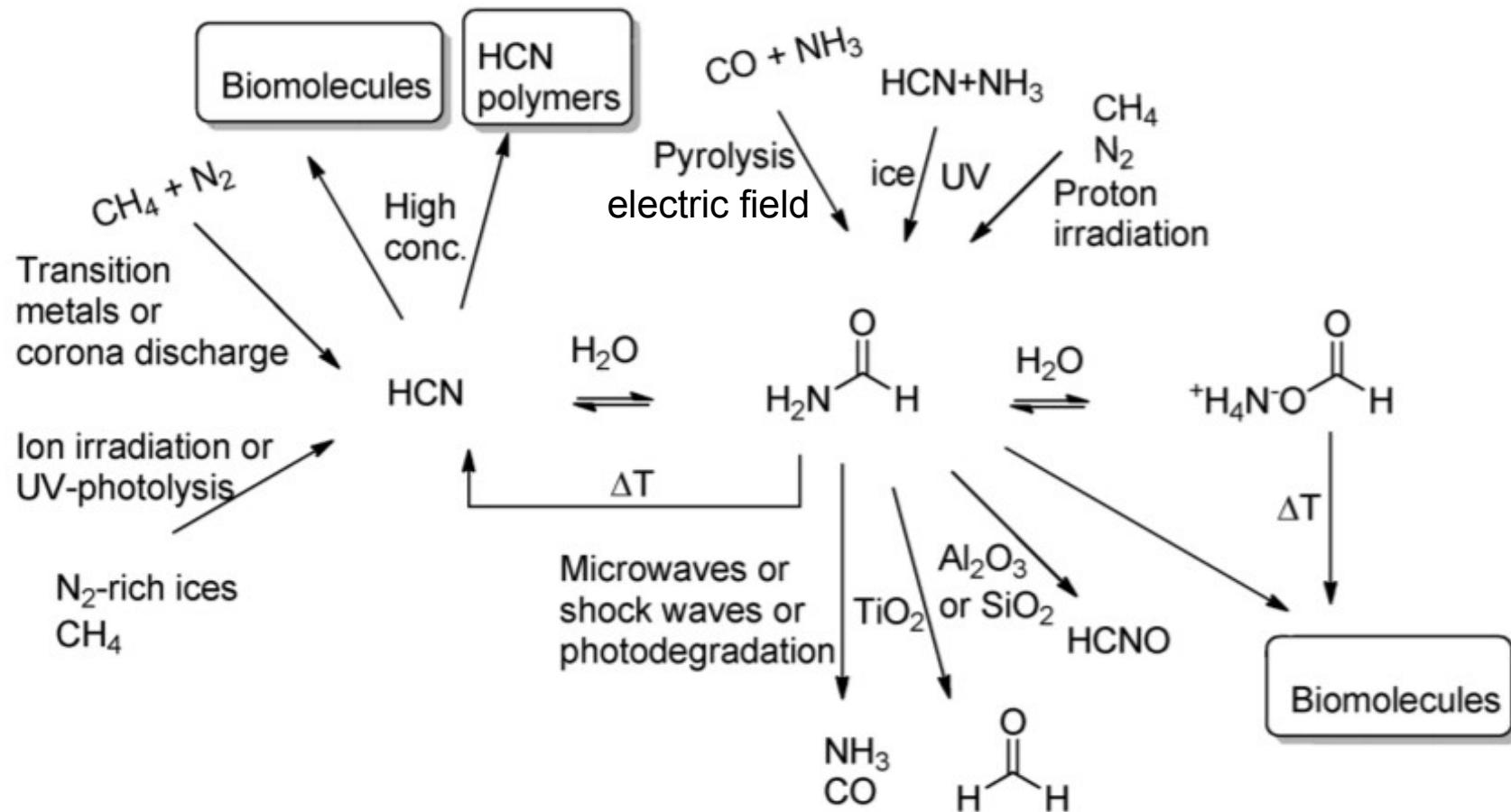
parent of biomolecular
units under diverse conditions:

- UV / laser spark / proton irradiation
- meteorite impact
- heat + meteorite as catalyst
- intense electric fields

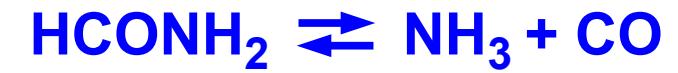
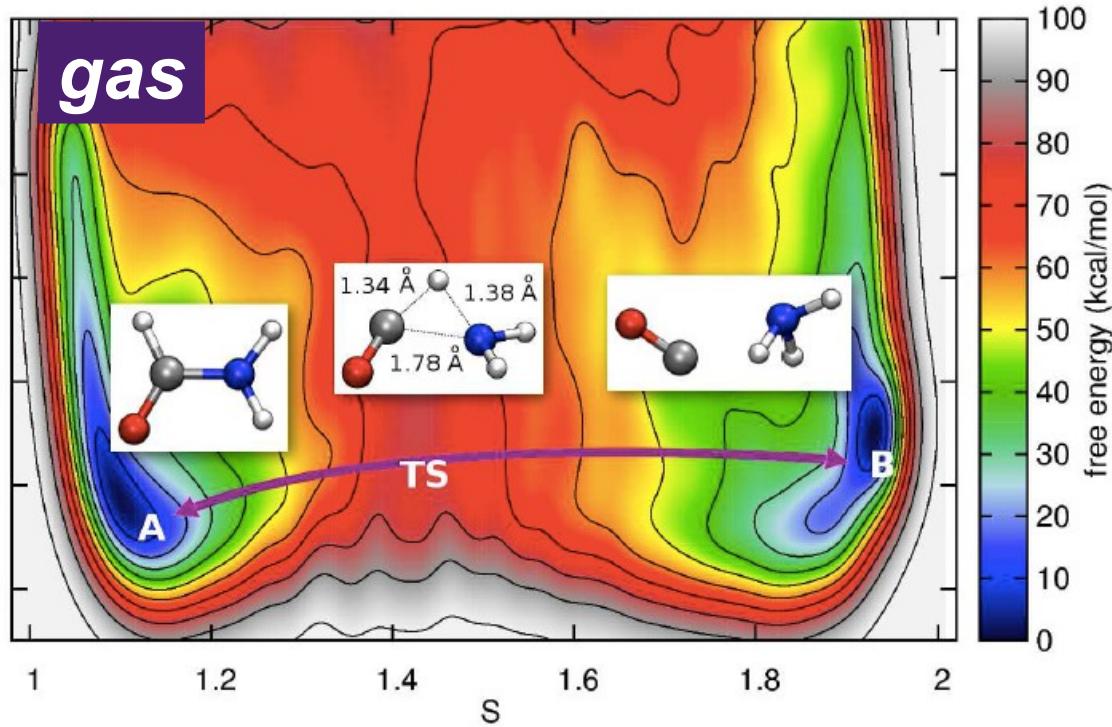
...

Saladino et al, *Phys Life Rev* 2012
Saitta & Saija, *PNAS* 2014
Ferus et al, *PNAS* 2015
Saladino et al, *PNAS* 2015

multiple routes lead to formamide synthesis



Saladino et al, *Phys Life Rev* 2012
Saitta & Saija, *PNAS* 2014
Ferus et al, *PNAS* 2015
Saladino et al, *PNAS* 2015



total time $\approx 50\text{-}100$ ps
 ~ 10 recrossings

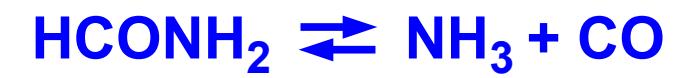
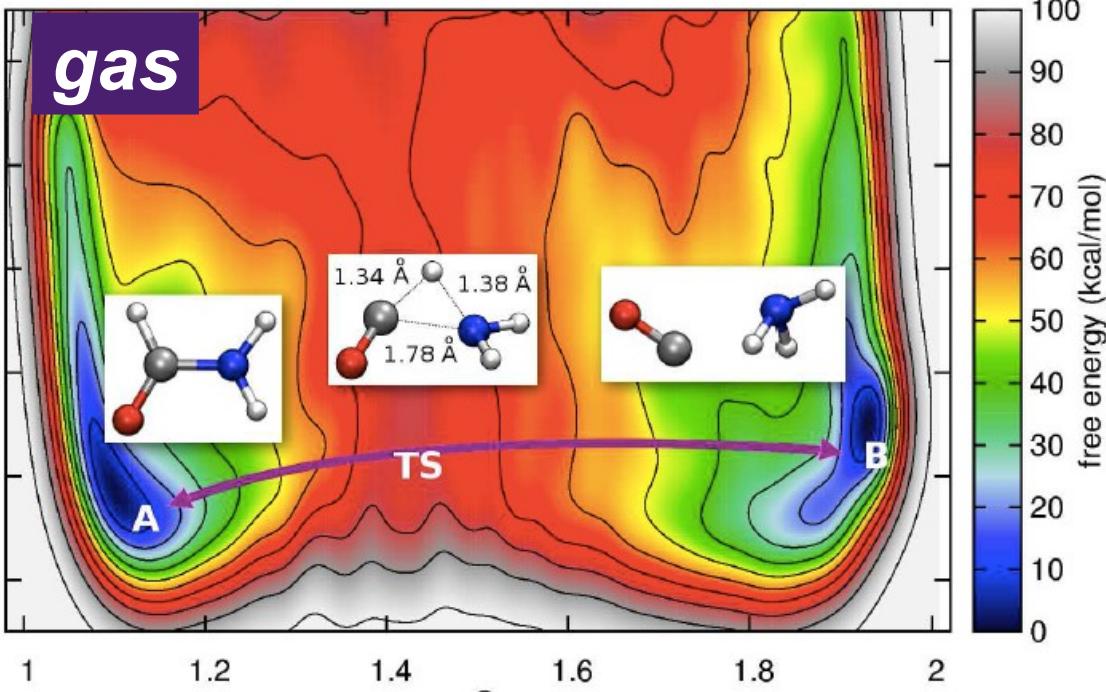
$T = 300$ K

input: A & B topologies
 \rightarrow **reversible** trajectories, free energy landscape

$\Delta F^* = 79$ kcal/mol (our model: PBE + Grimme, J Comput Chem 2006)

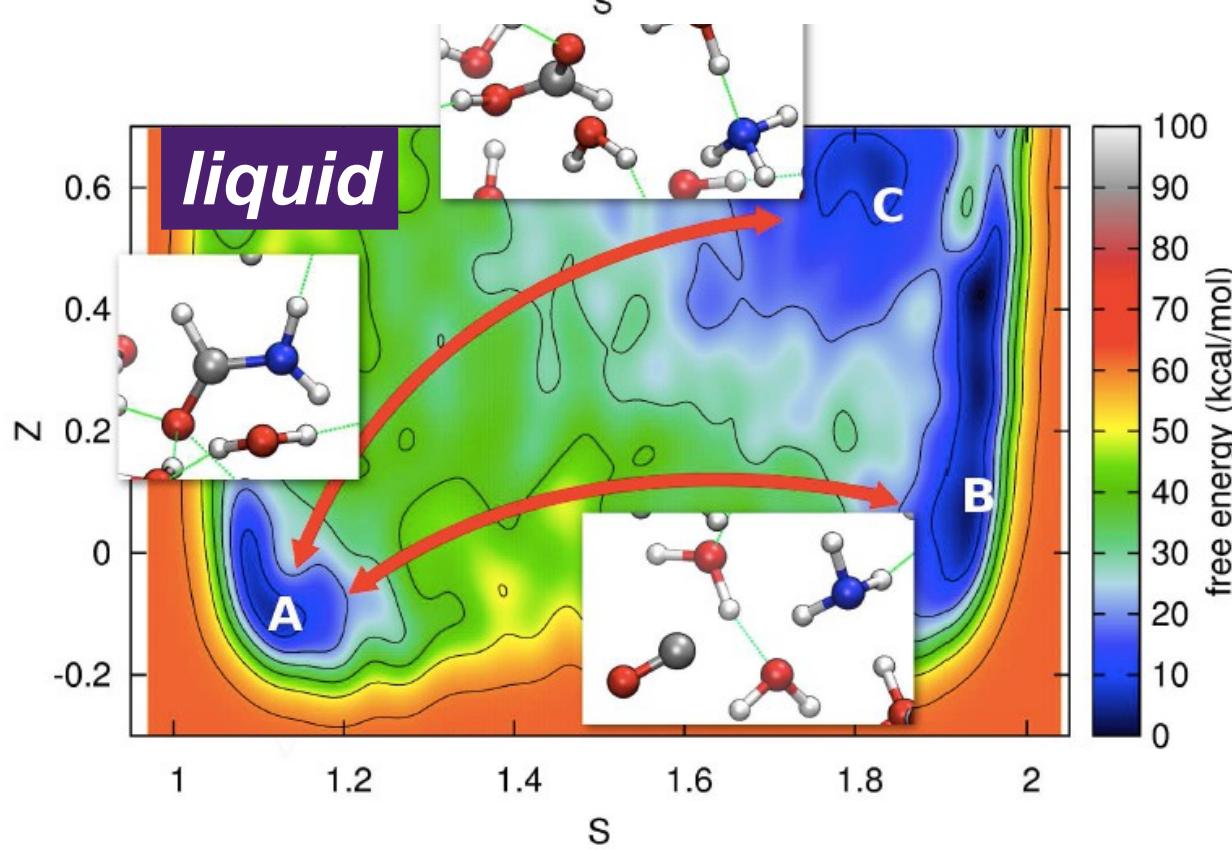
80.5 kcal/mol at CCSD(T)/CBS+ZPE level, T=0K

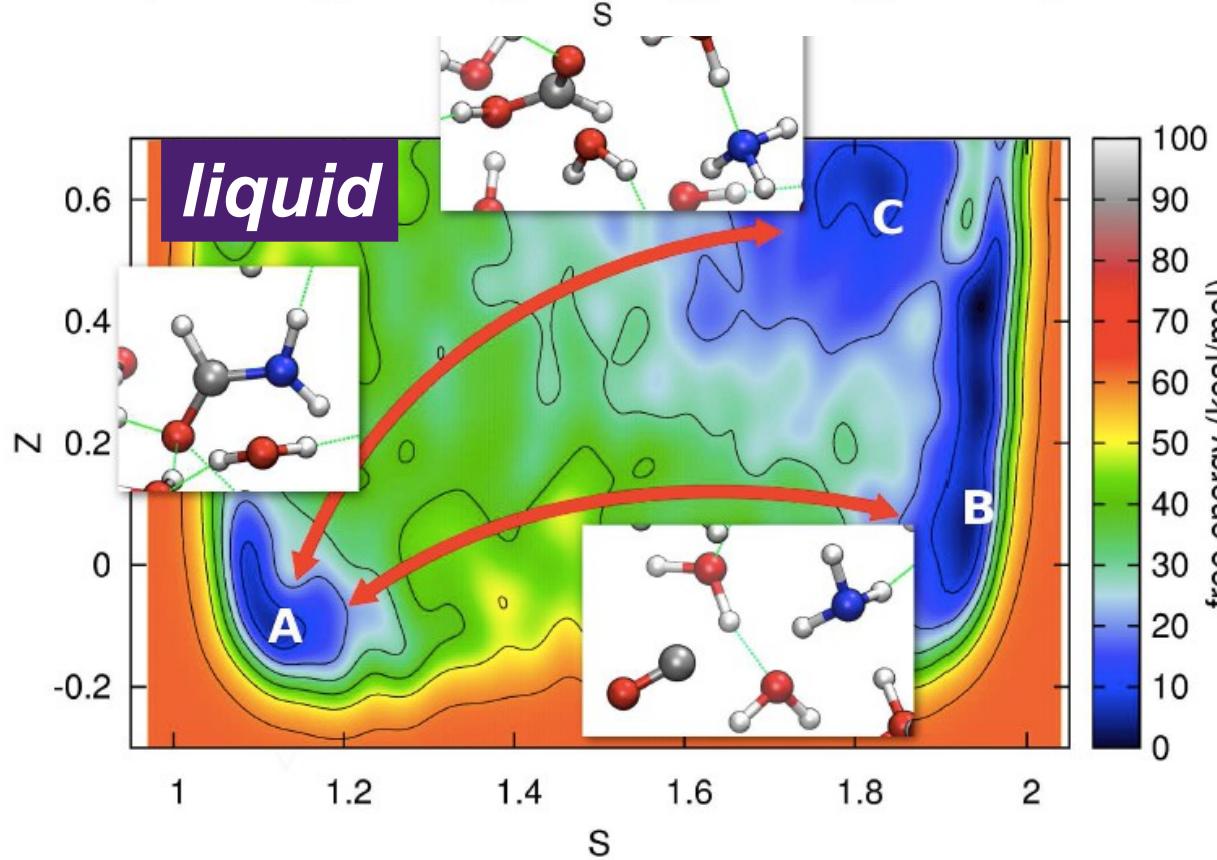
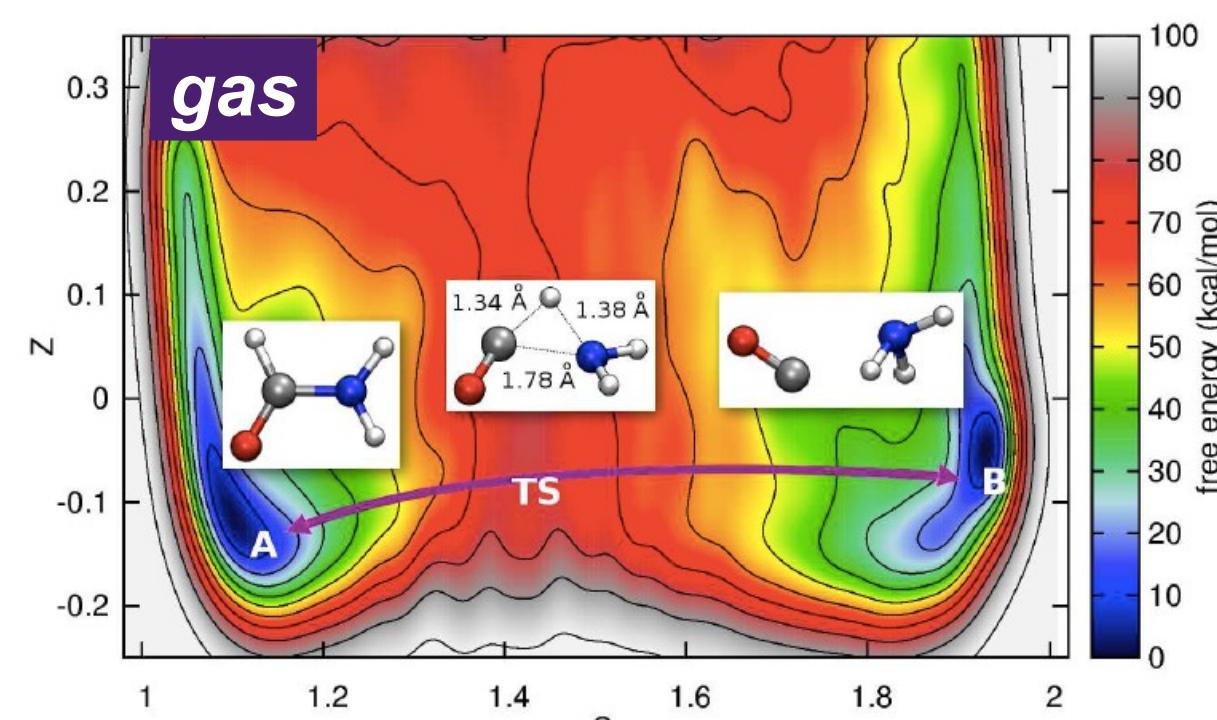
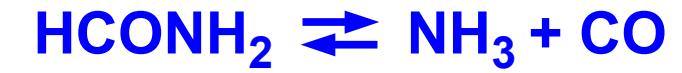
Nguyen et al, J. Phys. Chem. A 115, 841 (2011)



total time $\approx 50\text{-}100$ ps

same space of coords
in gas & in solution

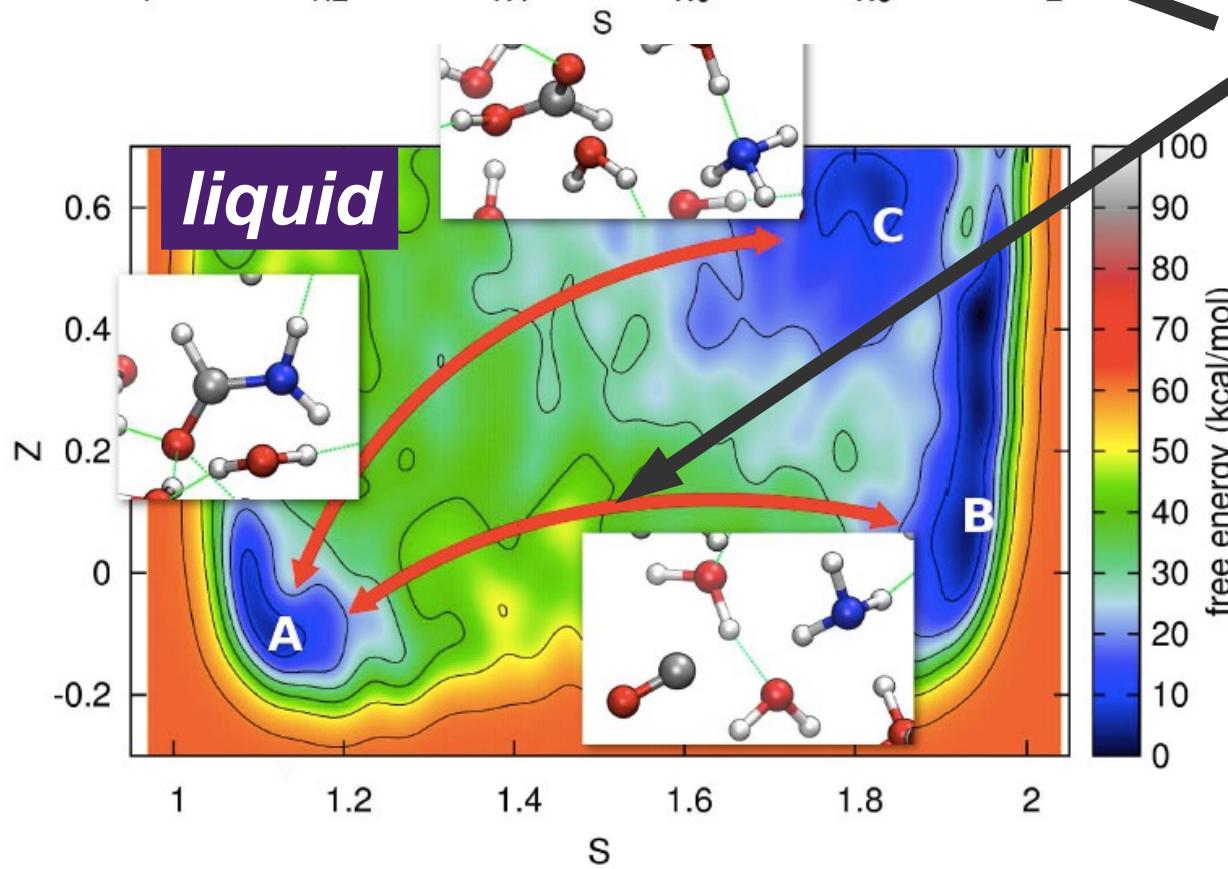
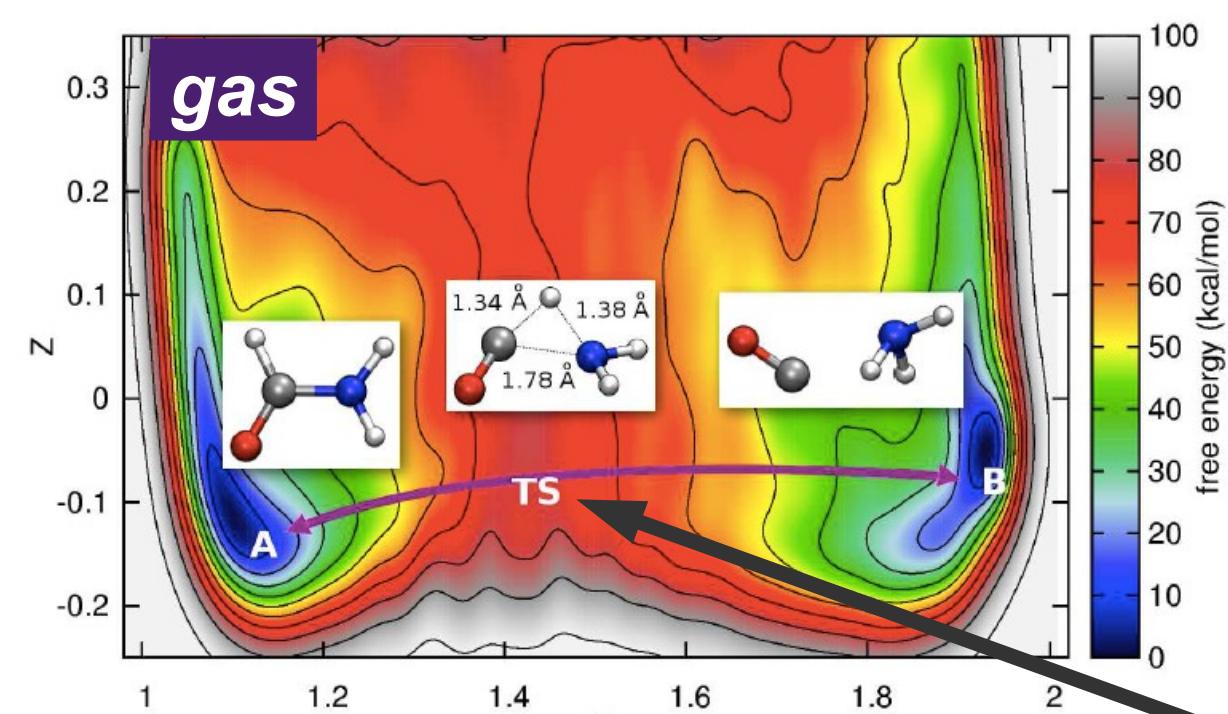
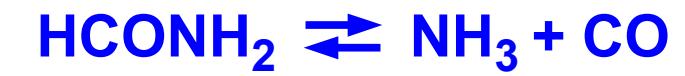




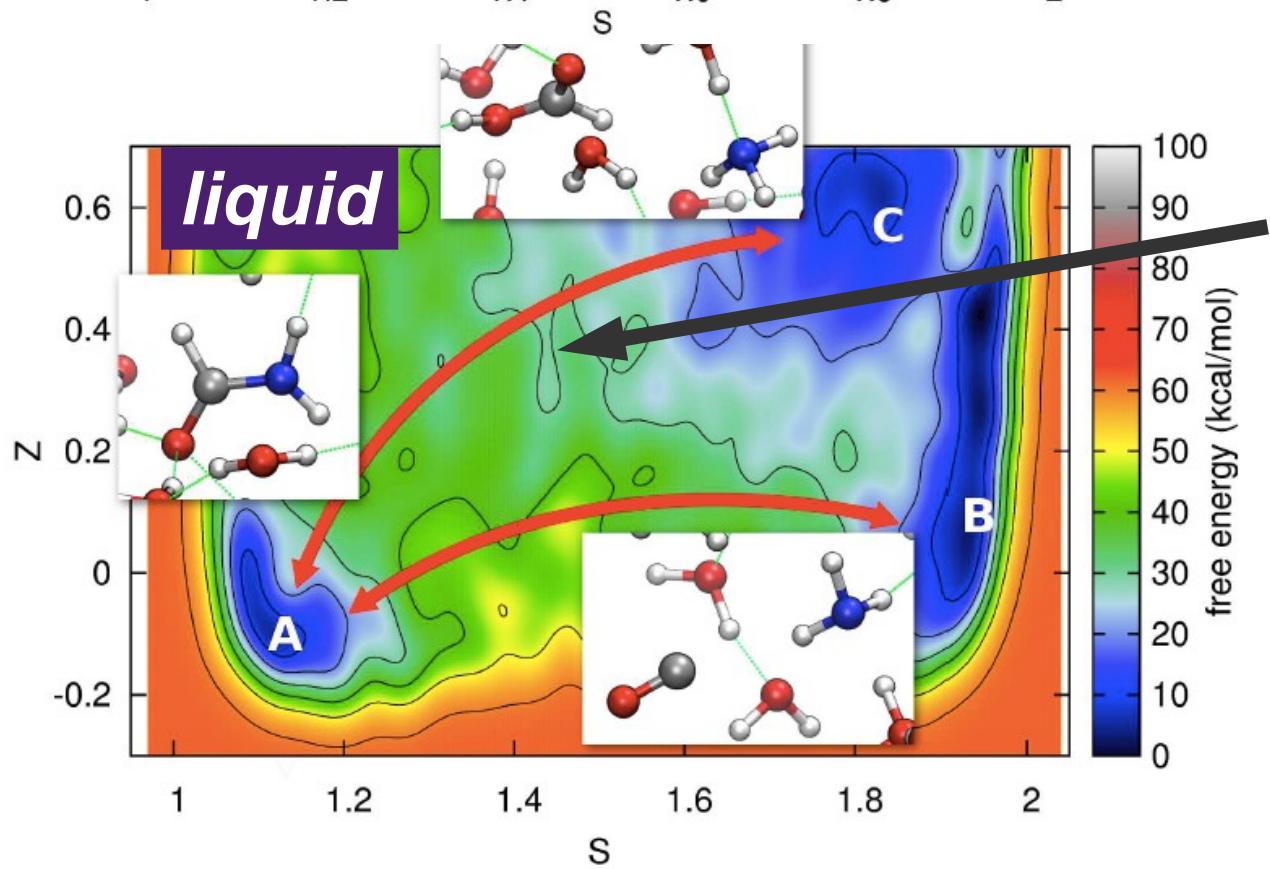
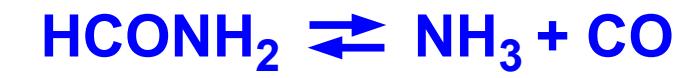
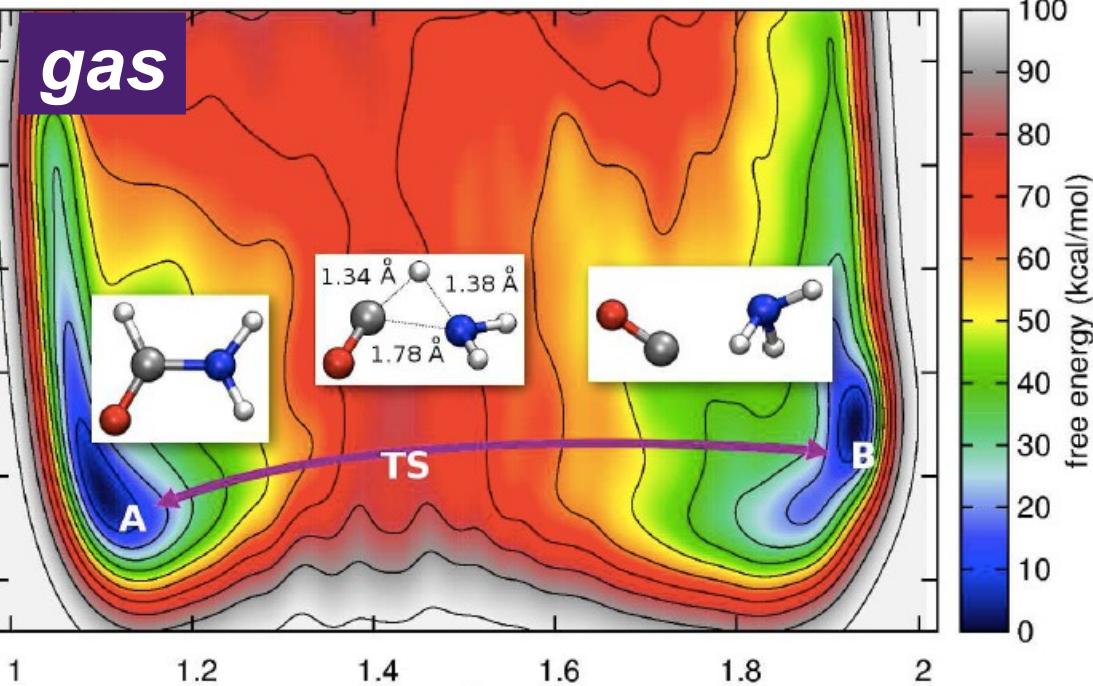
input = A & B

the simulation automatically discovers

- transition pathways
- intermediate states
- off-pathway states

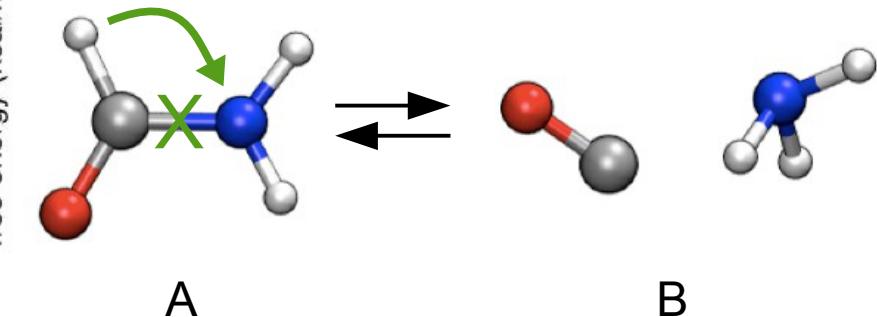
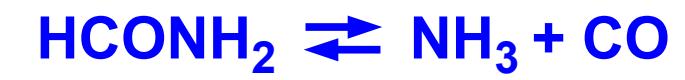
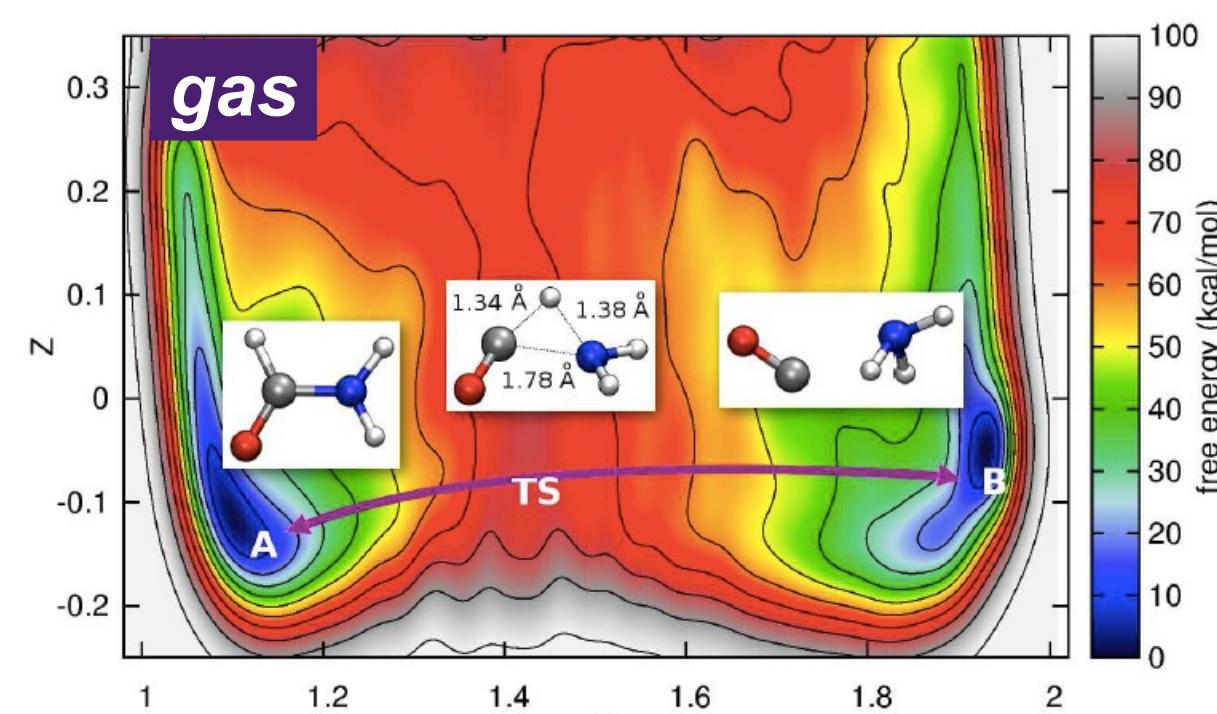


quantitative effect of solvent:
lower barrier

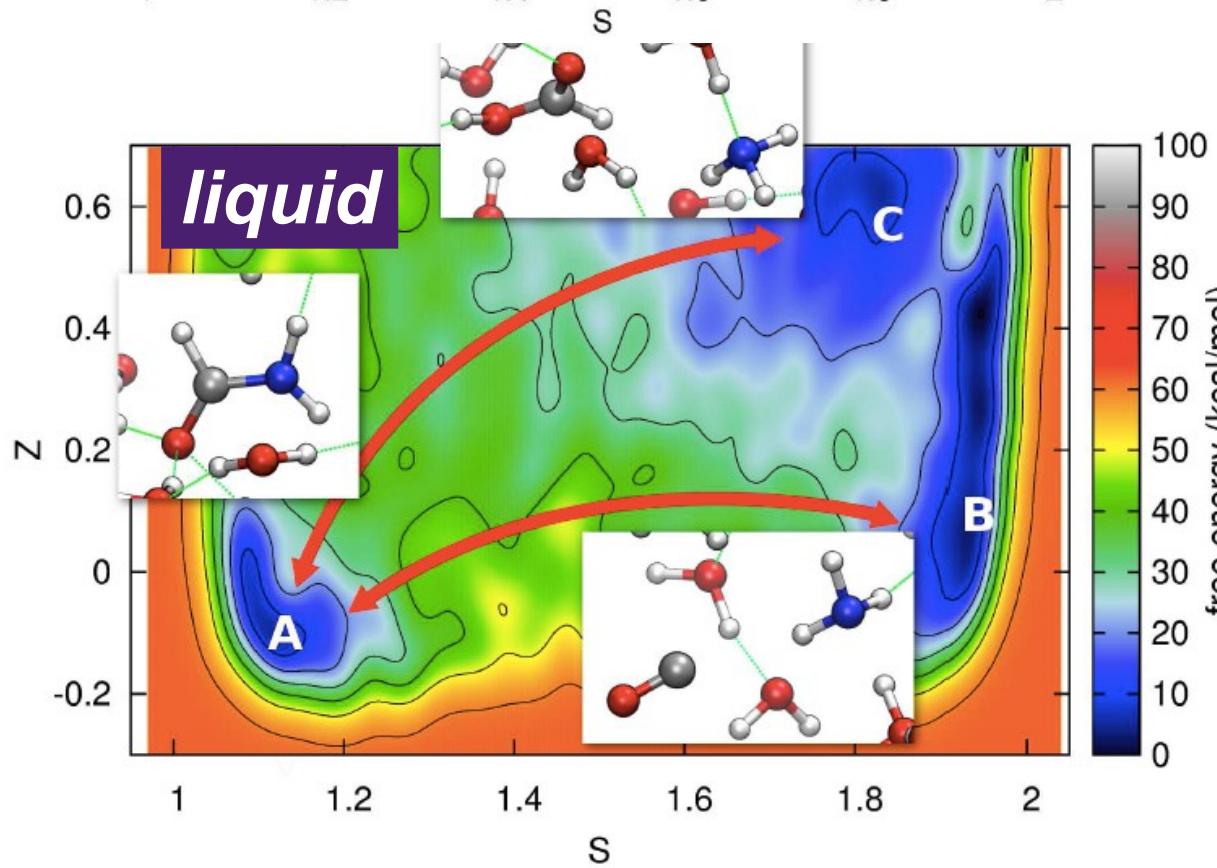


*qualitative effect of solvent:
new pathway to formic acid*

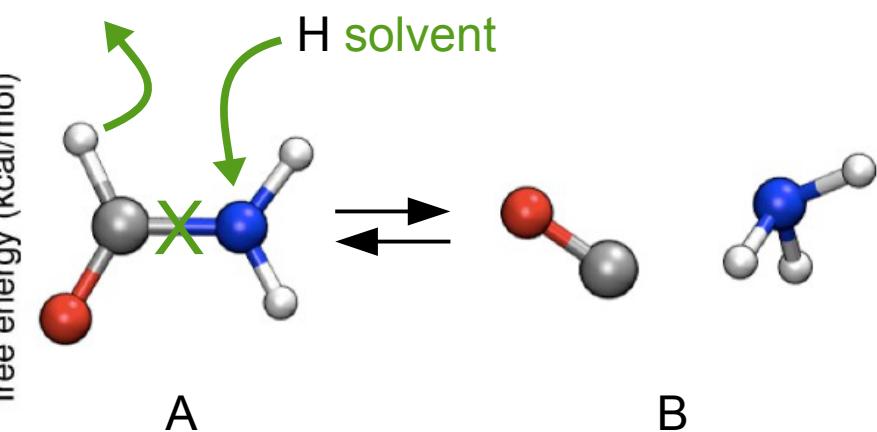
gas

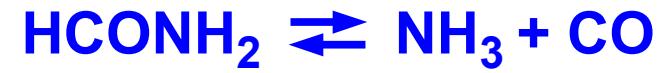
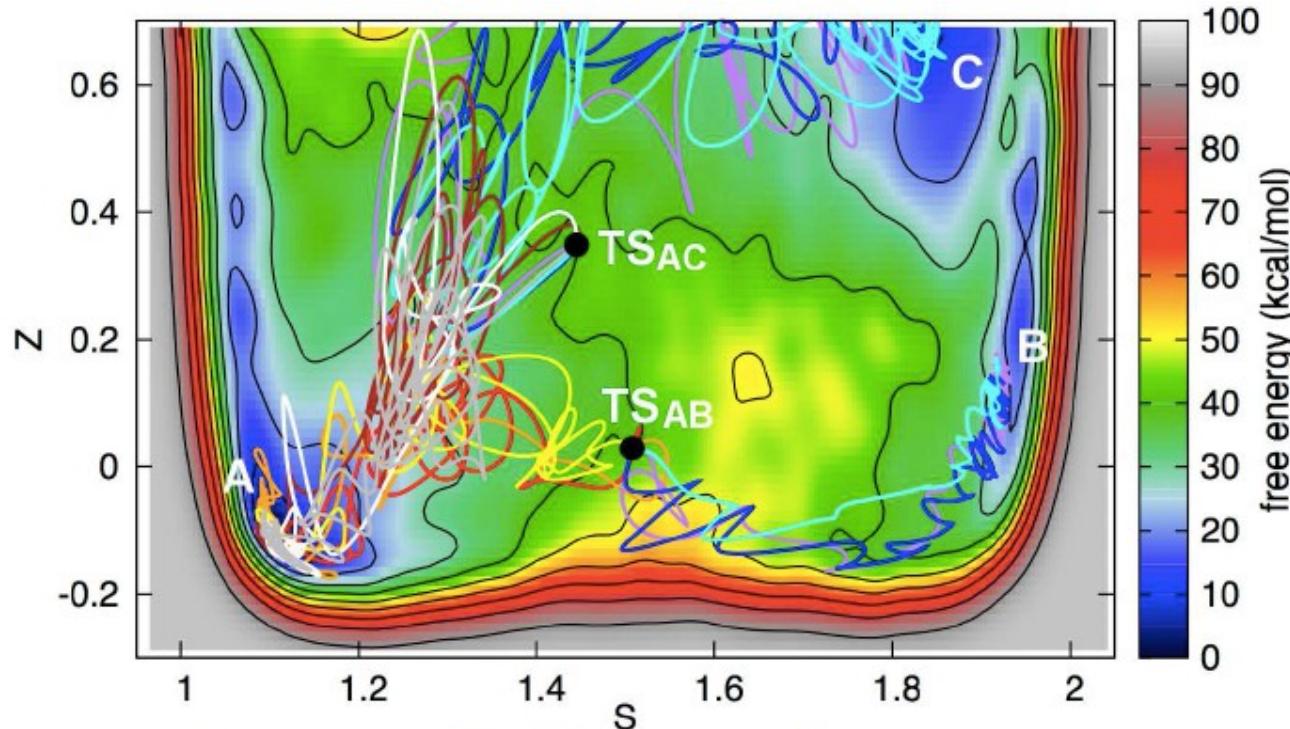


liquid

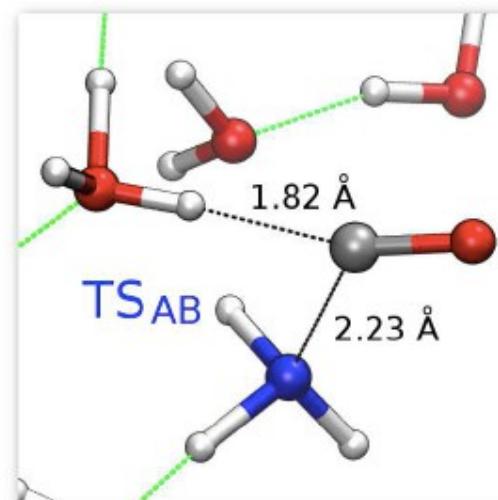
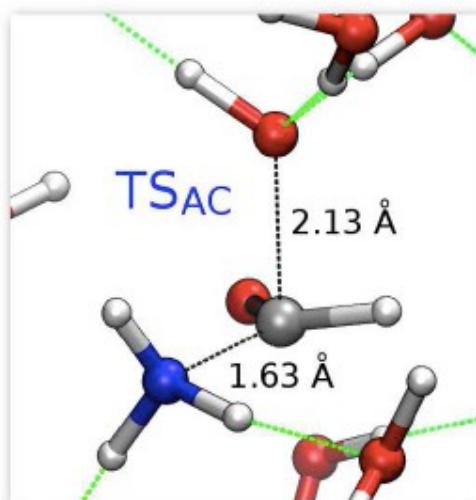


solvent





committor analysis
gives insight into kinetics
and transition states

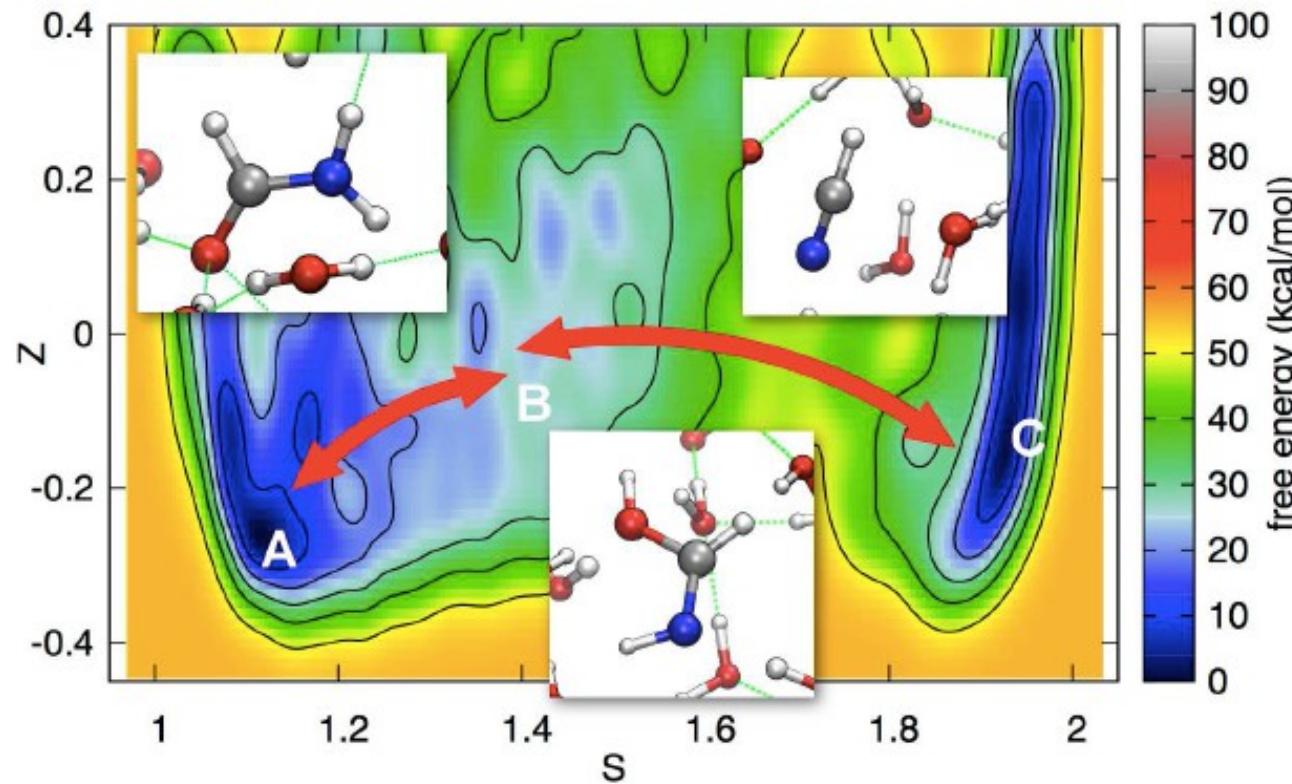


OH^- attacks the carbon

water deprotonates the carbon



formamide and HCN are central molecules in prebiotic chemistry

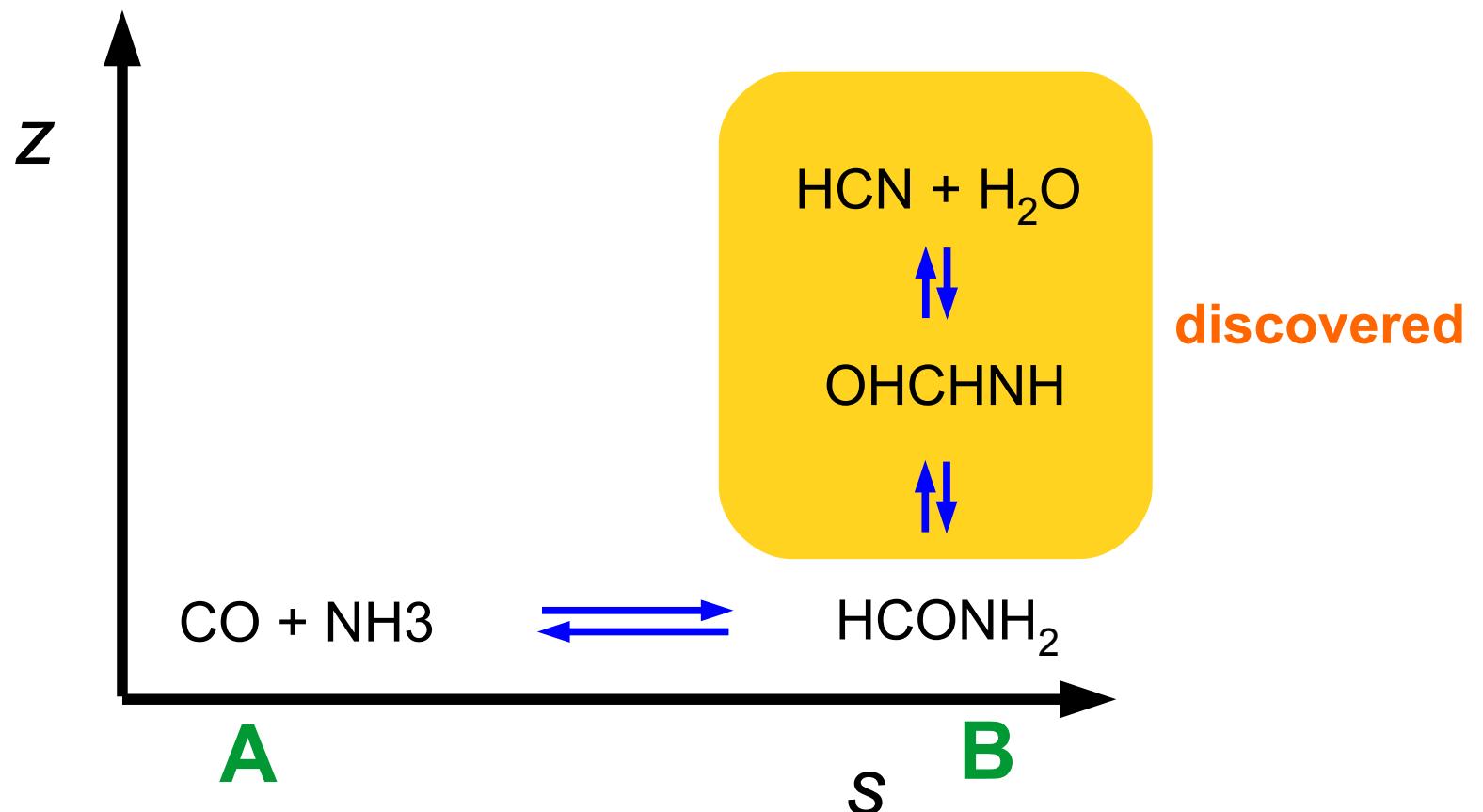
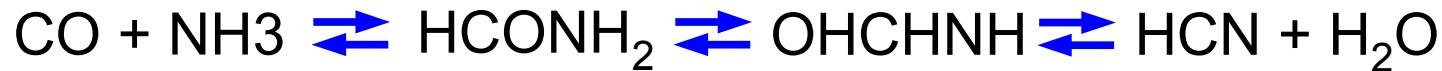


barriers

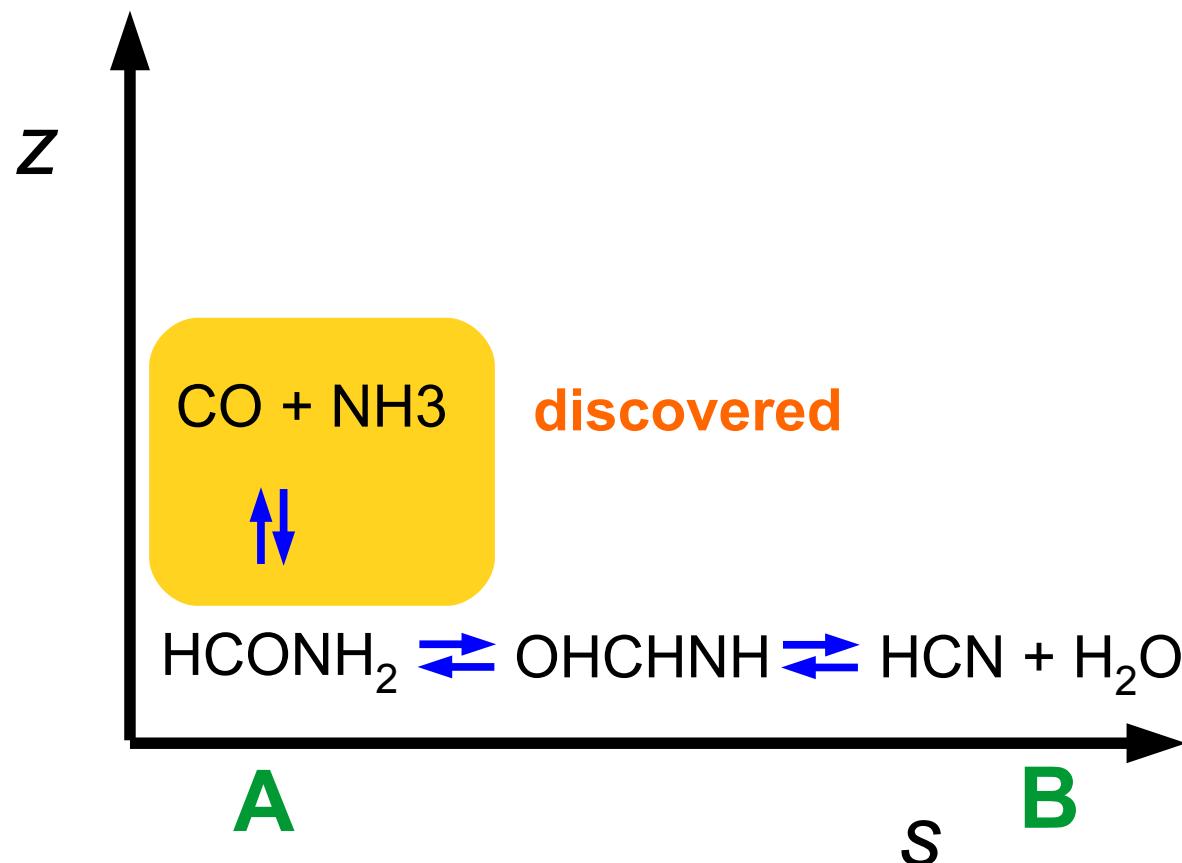
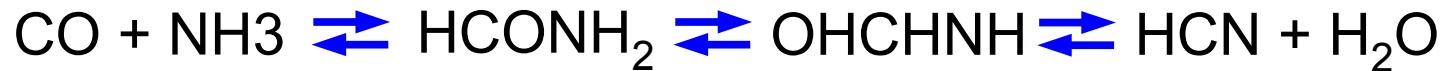
solution: $25 \text{ kcal/mol } \text{HCONH}_2 \rightarrow \text{OHCHNH}$
 $20 \text{ kcal/mol } \text{OHCHNH} \rightarrow \text{HCN} + \text{H}_2\text{O}$

gas phase: $45 \text{ kcal/mol } \text{HCONH}_2 \rightarrow \text{OHCHNH}$
 $55 \text{ kcal/mol } \text{OHCHNH} \rightarrow \text{HCN} + \text{H}_2\text{O}$

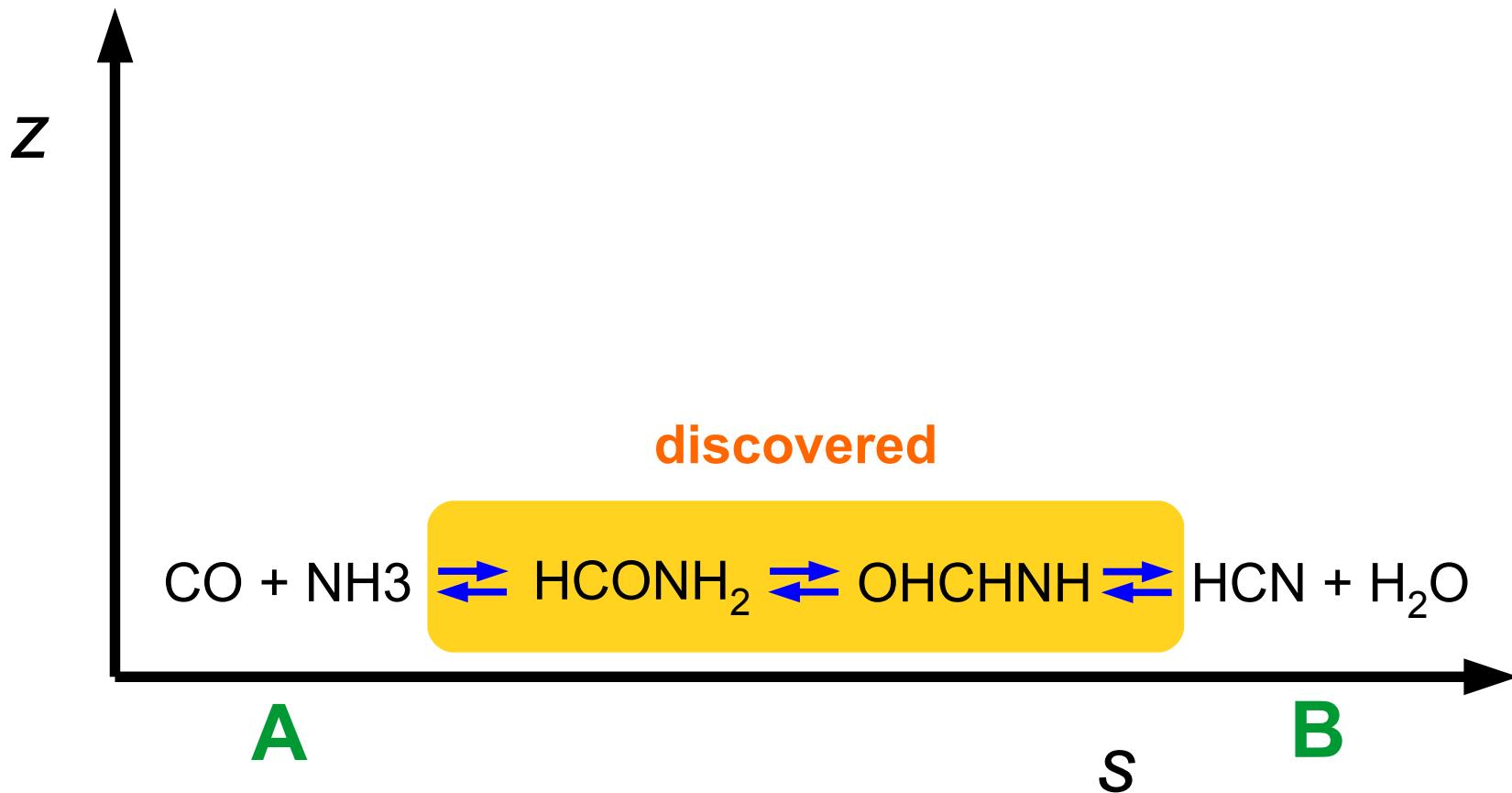
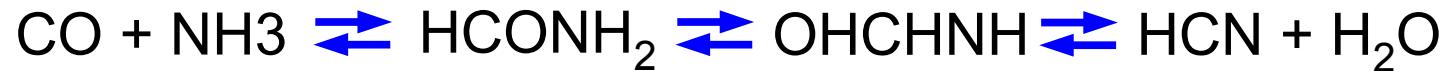
robustness: states on- and off-pathway are systematically discovered



robustness: states on- and off-pathway are systematically discovered



robustness: states on- and off-pathway are systematically discovered



how general is this approach?

- **synthesis of nucleobases in Miller exp. and in meteorite impact**

M. Ferus, F. Pietrucci, et al, *PNAS* 2017

- **new 1-step reaction from methanol to $\text{CH}_4 + \text{H}_2\text{CO}$ under E-field**

G. Cassone, F. Pietrucci, F. Saija, F. Guyot, A.M. Saitta, *Chem. Sci.* 2017

- **decomposition of amino acids in solution**

F. Pietrucci, A.M. Saitta, R. Starr, J. Aponte, J. Elsila, J. Dworkin (in preparation)

- **formation of HCOOH at water / mineral interfaces**

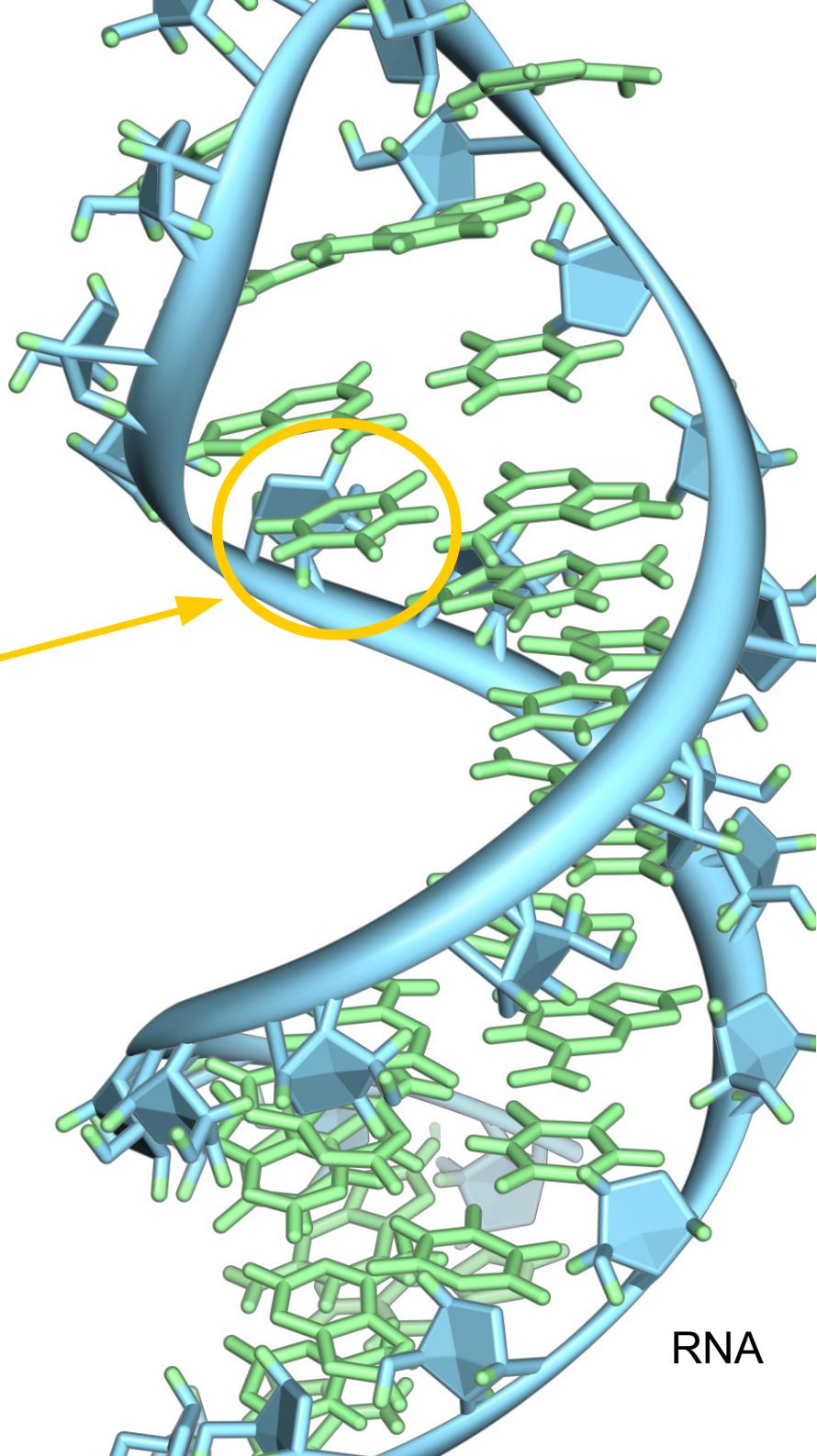
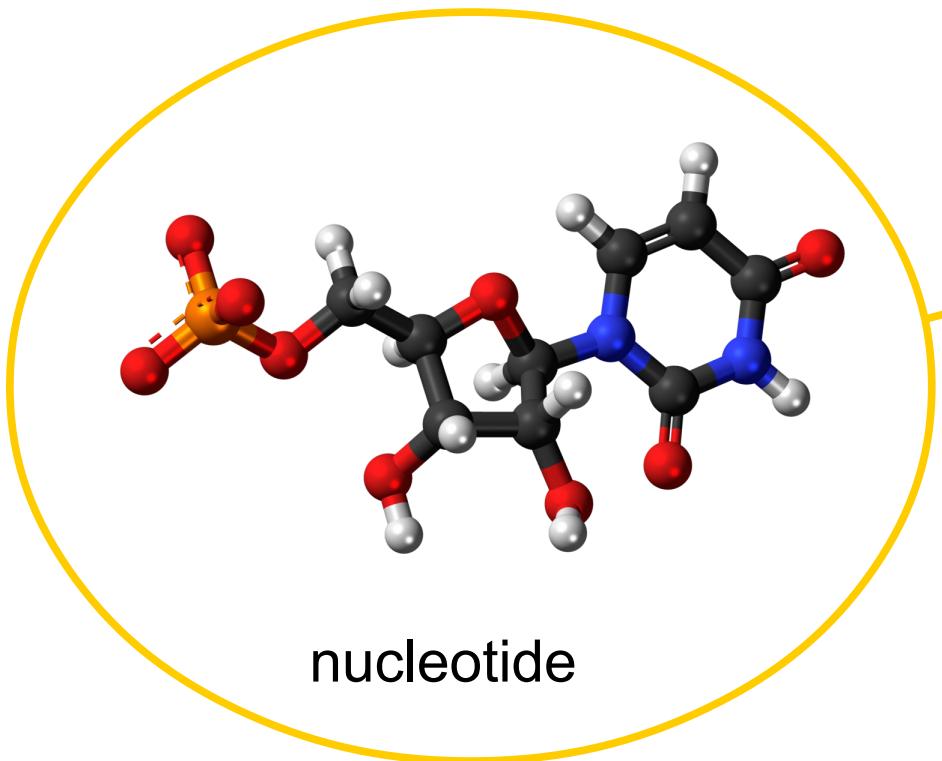
S. Laporte, F. Pietrucci, F. Guyot, A.M. Saitta (almost submitted)

- **hydrothermal prebiotic pathways to RNA nucleotides**

A. Perez-Villa et al (almost submitted)

...

RNA nucleotides synthesis & degradation

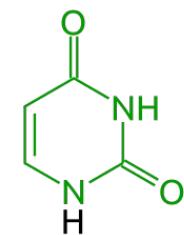
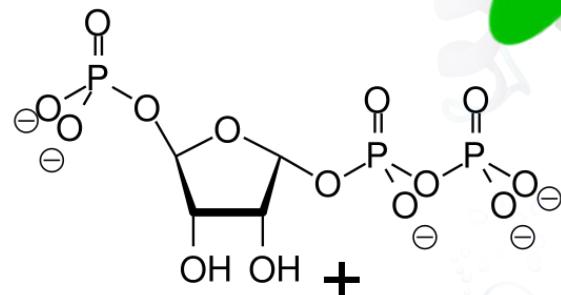


with Andrea Pérez-Villa,
A.M. Saitta, F. Guyot, M.C. Maurel,
T. Georgelin, J.F. Lambert

perhaps enzymes inherited a prebiotic reaction
PRPP + base -> nucleotide ?

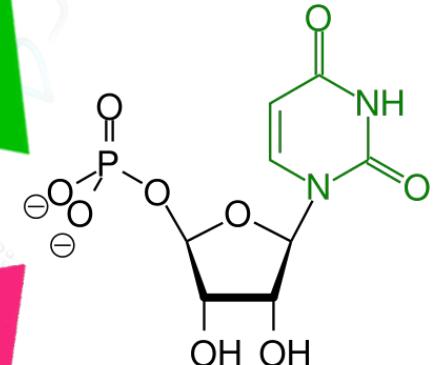
A. Pérez-Villa

phosphoribosyl
pyrophosphate



Biological

300 K



400 K

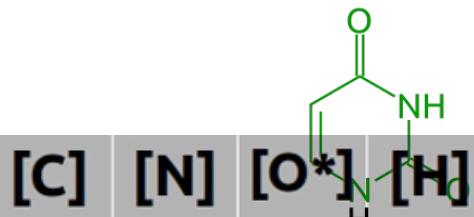
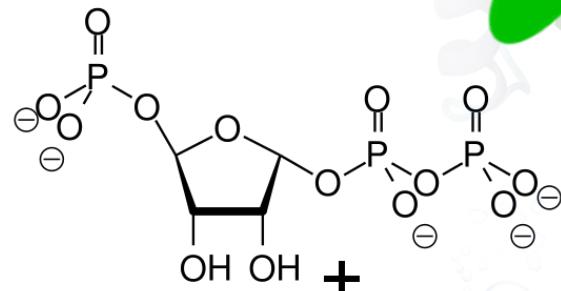
Hydrothermal
?

approach: ab initio MD + NMR experiments

perhaps enzymes inherited a prebiotic reaction
 PRPP + base -> nucleotide ?

A. Pérez-Villa

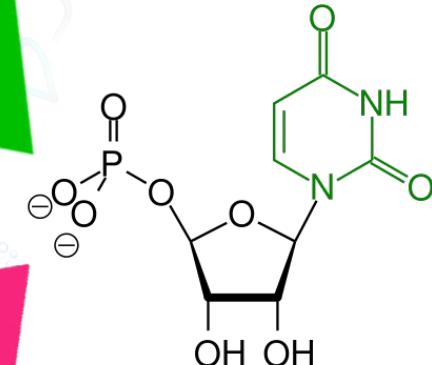
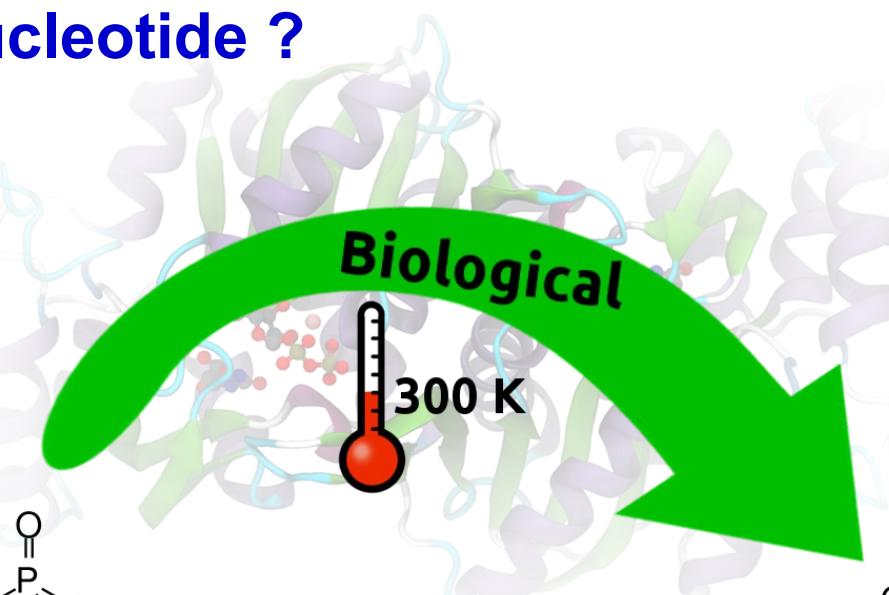
phosphoribosyl
 pyrophosphate



	[C]	[N]	[O*]	[H]
C ^{1'}	1	0	1	1
N ^{gly}	2	0	0	1
O ¹	1	0	0	0
O ²	0	0	0	0

Biological

300 K



400 K

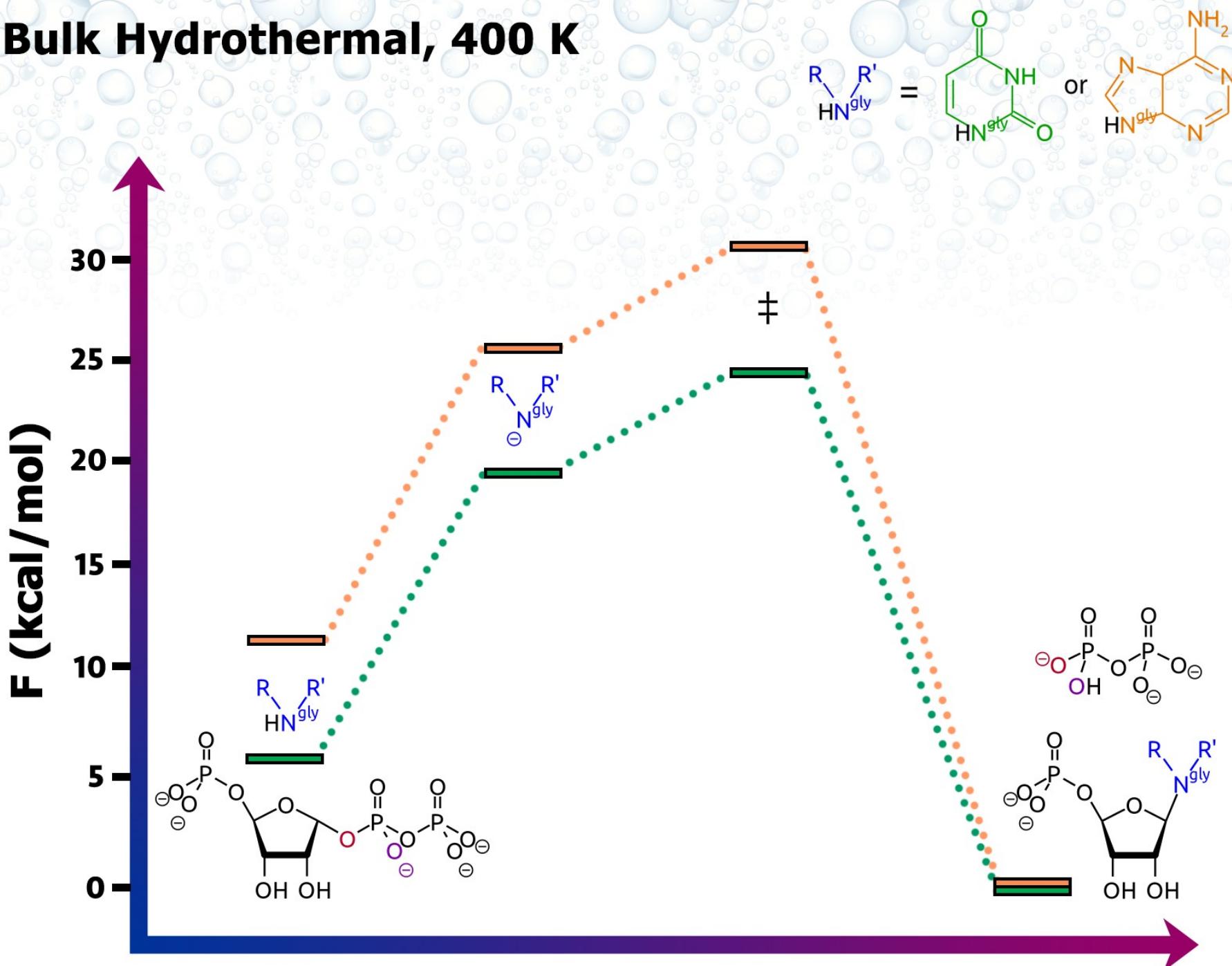
Hydrothermal

?

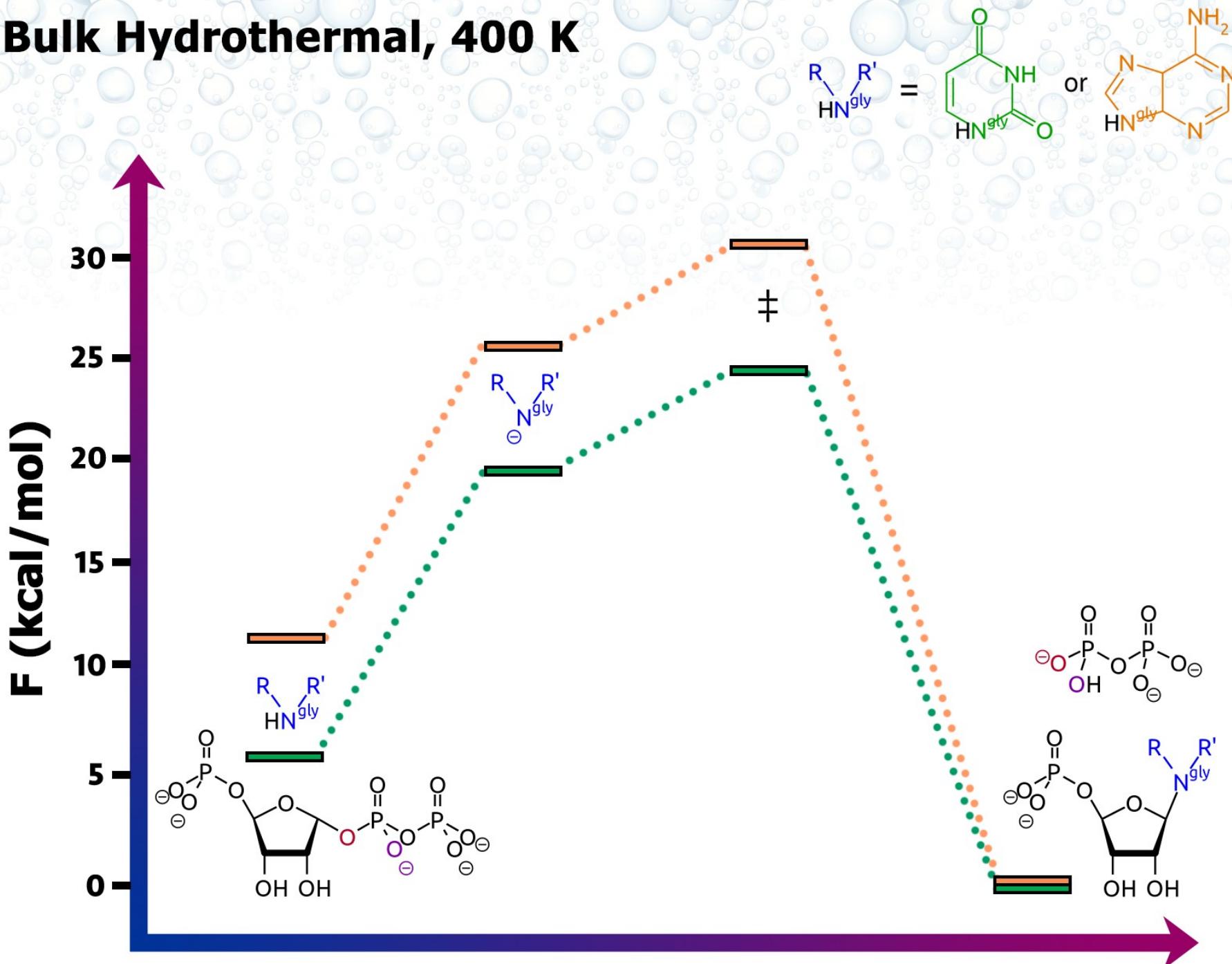
new path CVs

	[C]	[N]	[O*]	[H]
C ^{1'}	1	1	0	1
N ^{gly}	3	0	0	0
O ¹	0	0	0	0
O ²	0	0	0	1

Bulk Hydrothermal, 400 K



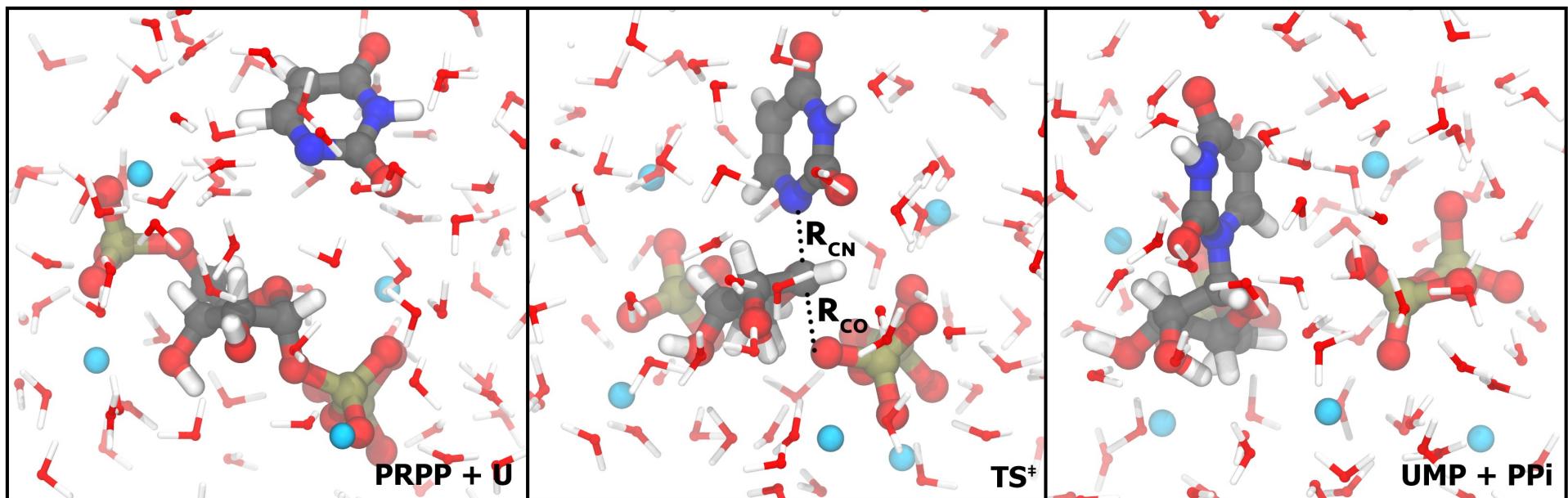
Bulk Hydrothermal, 400 K



we deduce: better to do exp. with excess of base, pH > 7

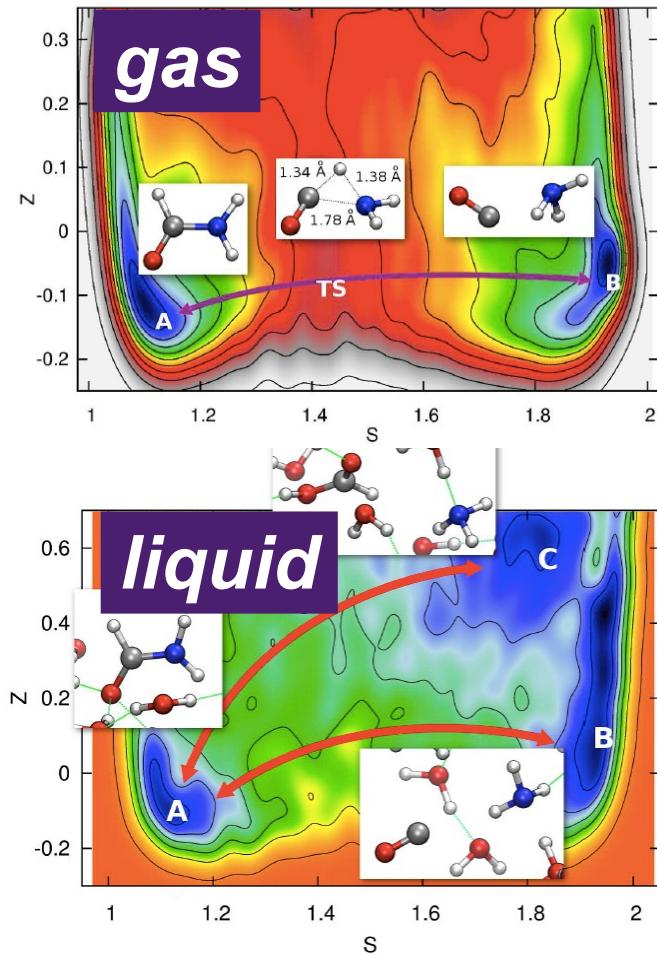
DFT-PBE + Grimme
explicit solvent (400 atoms)

T = 400 K
time ~ 100 ps



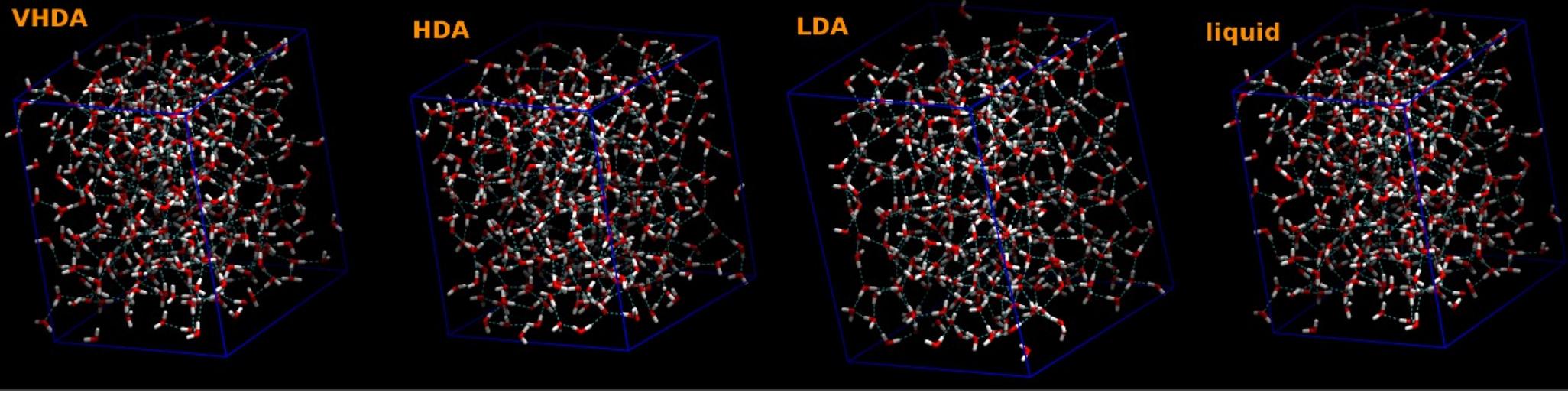
we found an S_N2 pathway
inverting chirality of the C1' atom: **confirmed by NMR**

an “amphibian” approach ...



- ✔ passing in a seamless way from gas-phase to solution
- ✔ A & B topologies as input: textbook formulas
- ✔ automatic discovery of intermediates and transition states
- ✗ need to define who is solute and who is solvent

amorphous ices and liquid water

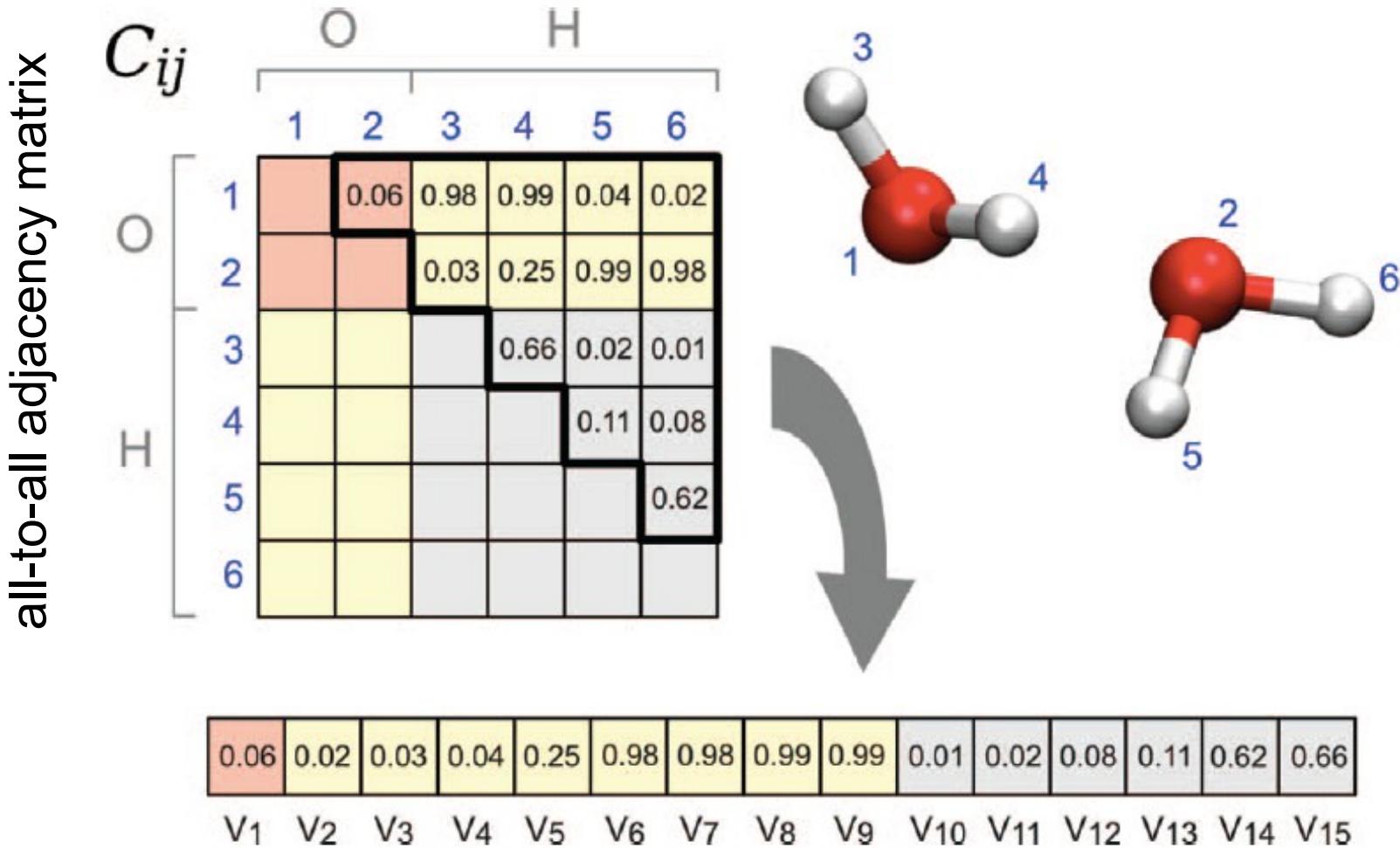


find the difference ...

water is challenging:

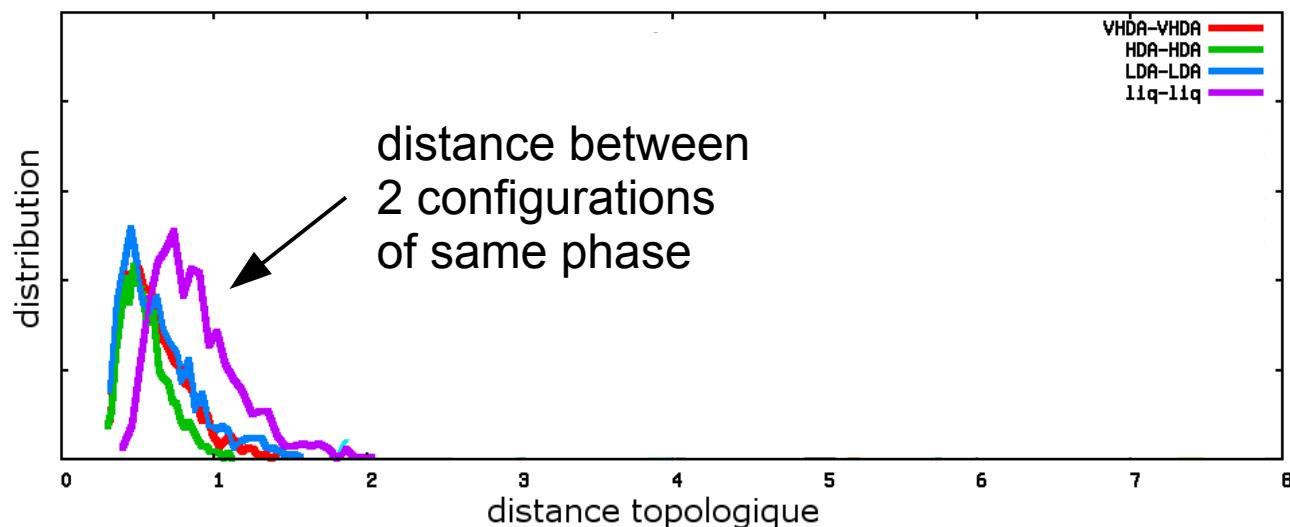
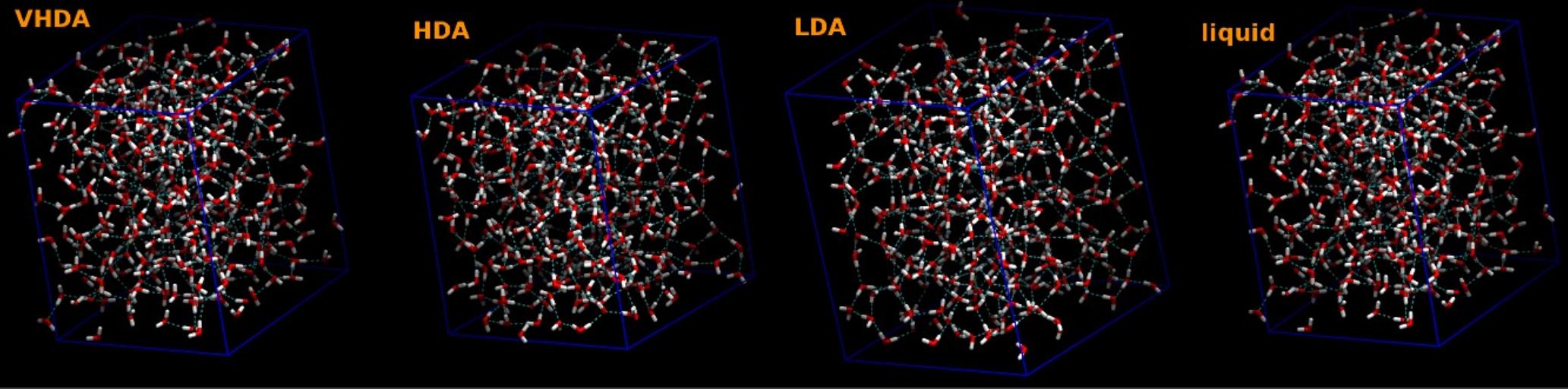
- a rich phase diagram, with many poly(a)morphs
- large sim. boxes with periodic boundary conditions (no RMSD...)
- no large differences in bonding

Permutation invariant vector (PIV)

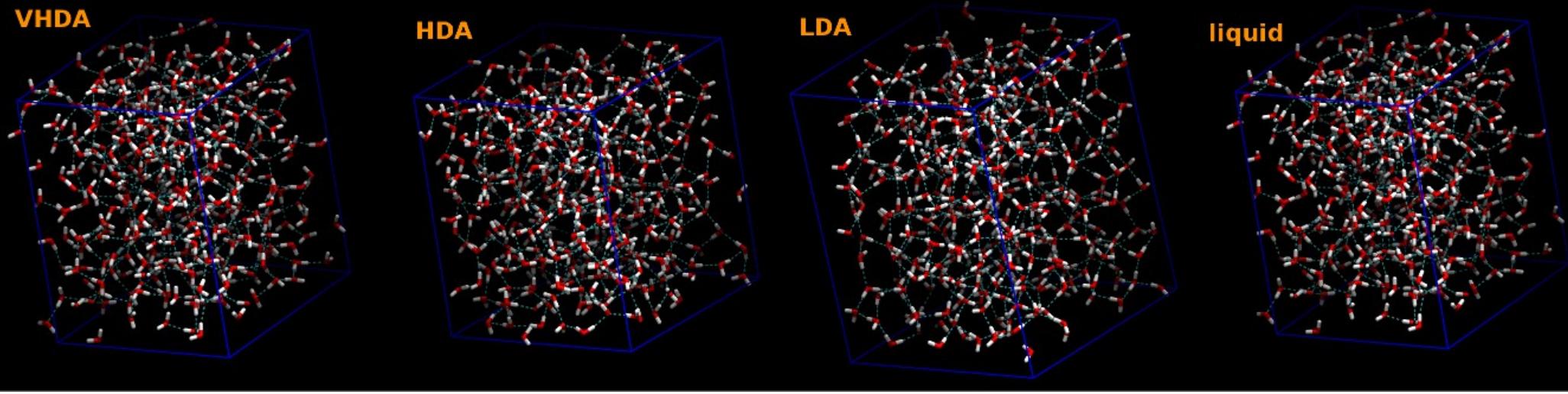


$$\mathbf{v} = \text{sort}(C_{ij}) = \text{sort}\left(C(|\mathbf{R}_i - \mathbf{R}_j|)\right), \quad i > j$$

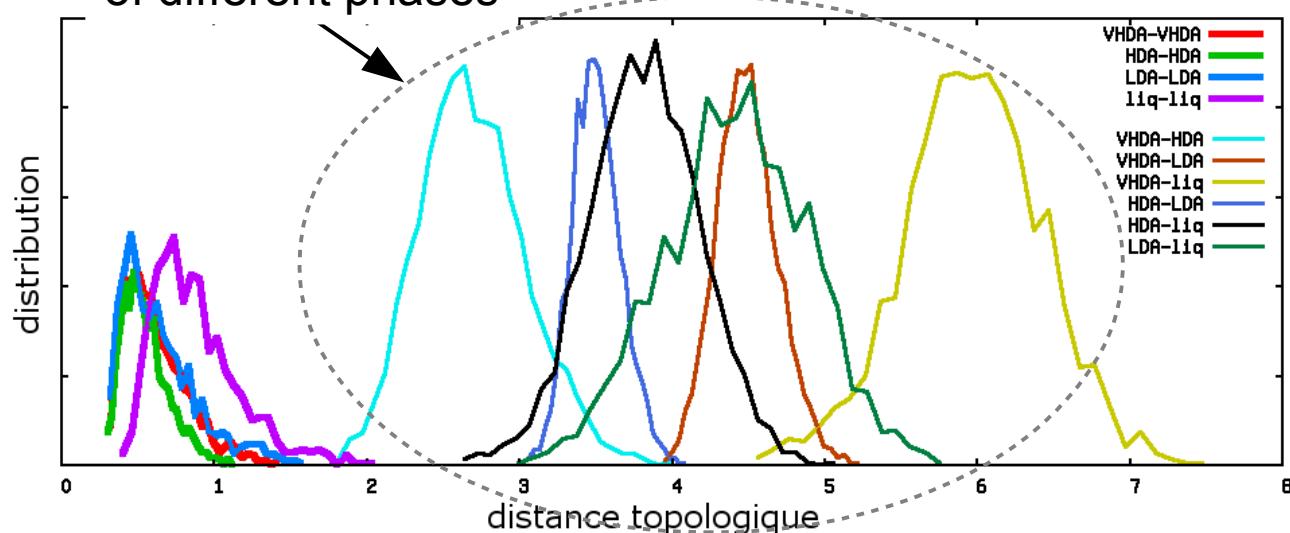
amorphous ices and liquid water



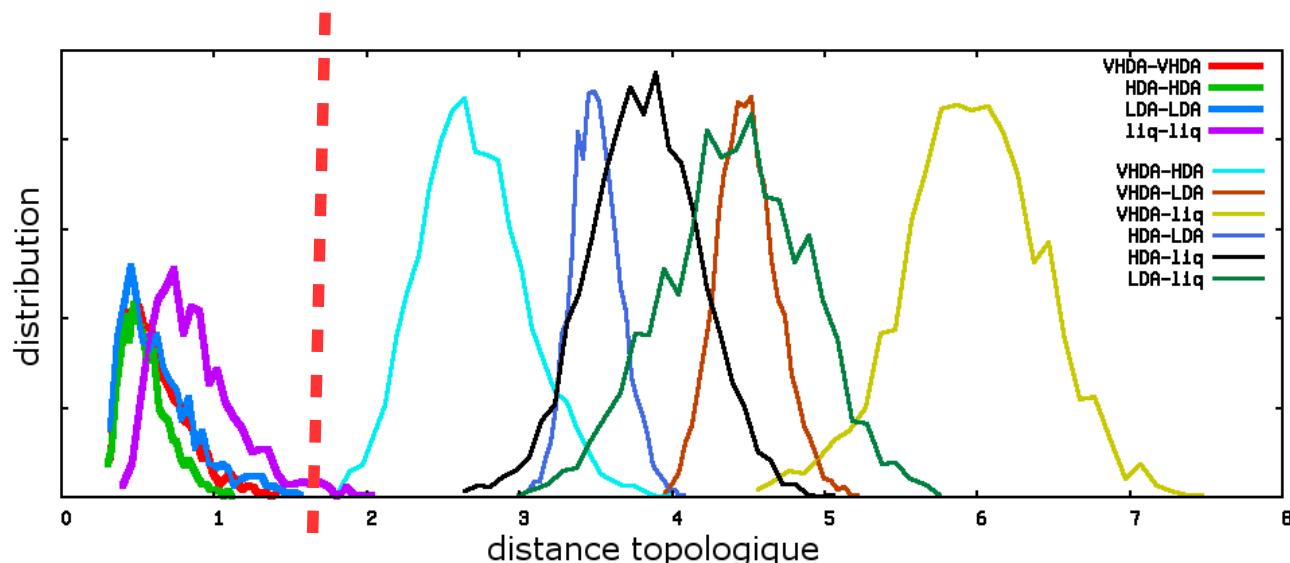
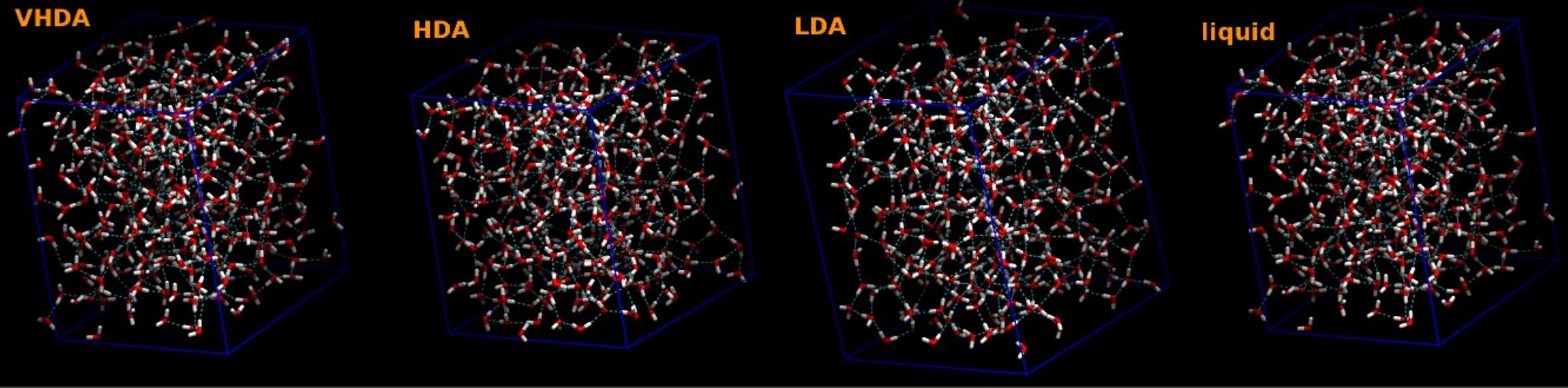
amorphous ices and liquid water



distance between
2 configurations
of different phases



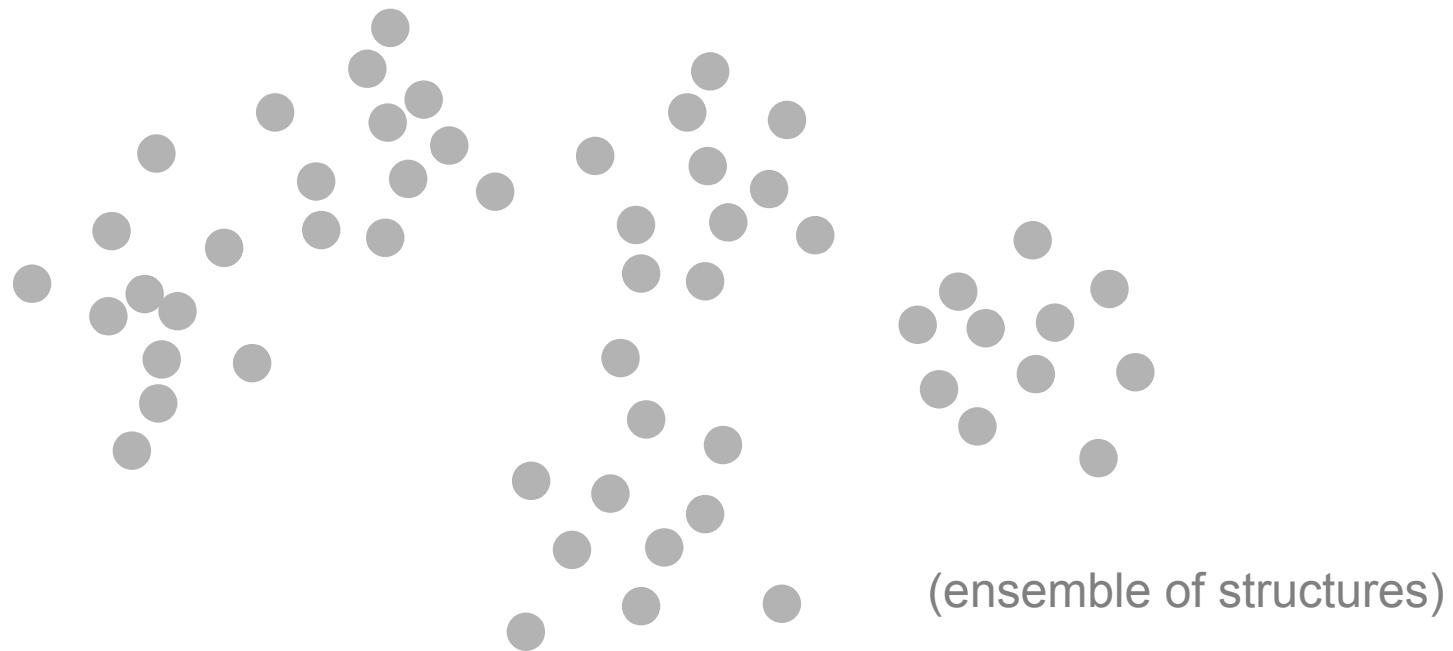
amorphous ices and liquid water



structures are
systematically assigned
to 4 distinct groups

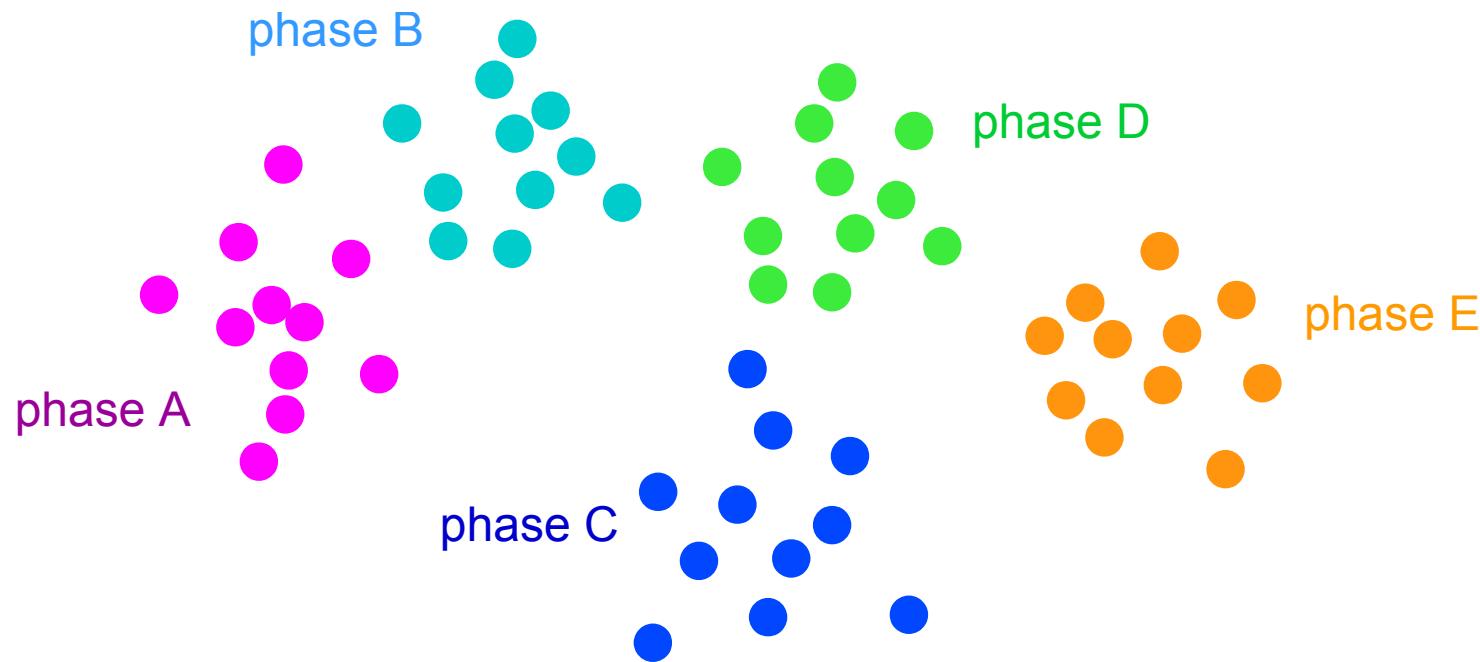
Permutation invariant vector (PIV)

$$\mathcal{D}_{AB} = |\mathbf{v}^A - \mathbf{v}^B| = \left[\sum_{k=1}^{\frac{1}{2}N(N-1)} (v_k^A - v_k^B)^2 \right]^{\frac{1}{2}}$$



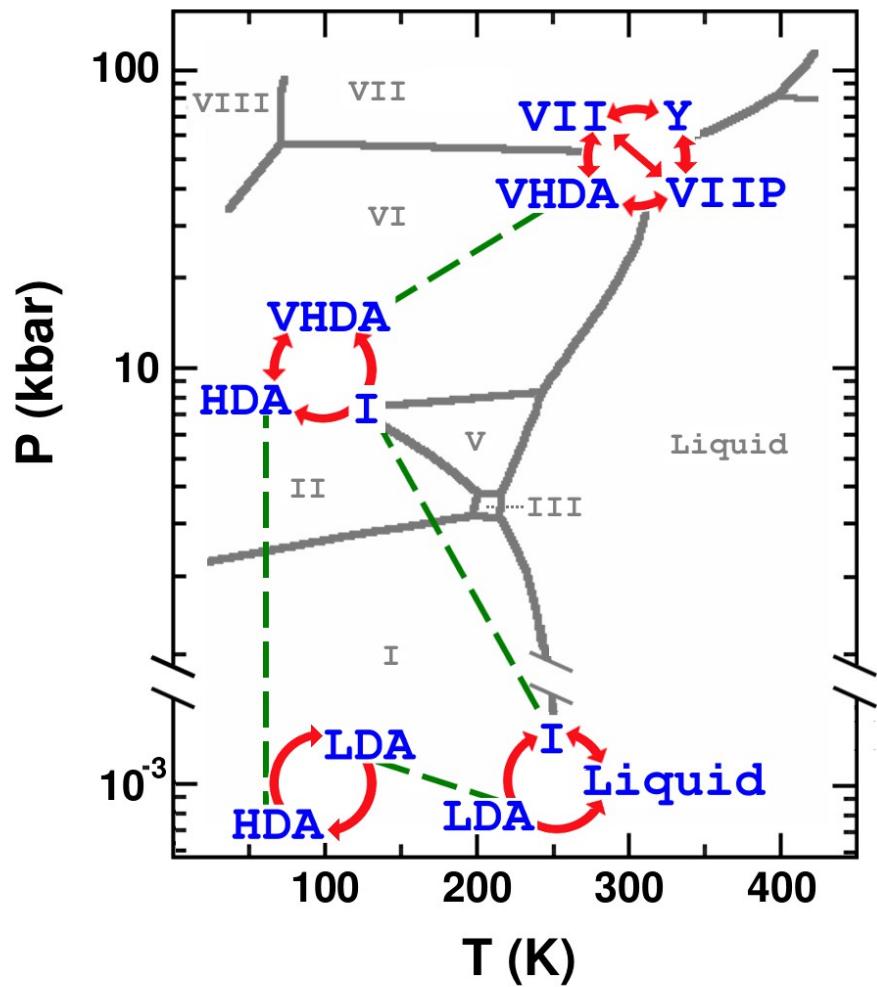
Permutation invariant vector (PIV)

$$\mathcal{D}_{AB} = |\mathbf{v}^A - \mathbf{v}^B| = \left[\sum_{k=1}^{\frac{1}{2}N(N-1)} (v_k^A - v_k^B)^2 \right]^{\frac{1}{2}}$$

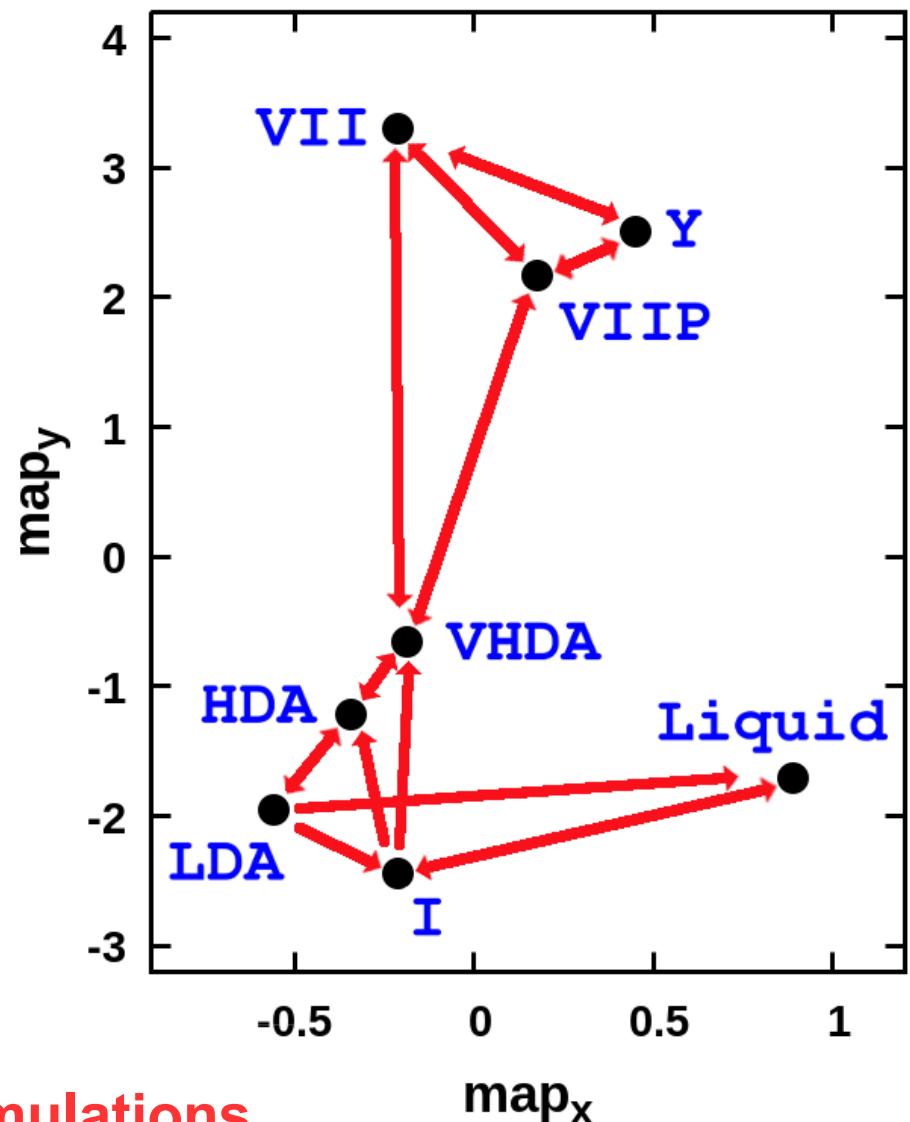


(many available clustering algorithms...)

NPT metadynamics / umbrella sampling
800 TIP4P/2005 water molecules



arrows = simulations



combining permutation invariant vectors with 2-state path CVs:

$$v_{kk'}^{\beta\beta'} = c_{kk'} \mathcal{S} \left(\sqrt[3]{\frac{\Omega_0}{\Omega}} |\mathbf{r}_{\beta k} - \mathbf{r}_{\beta' k'}| \right)$$

↑
switching function

element pair ↓
 ↑
atom

permutation invariant vector

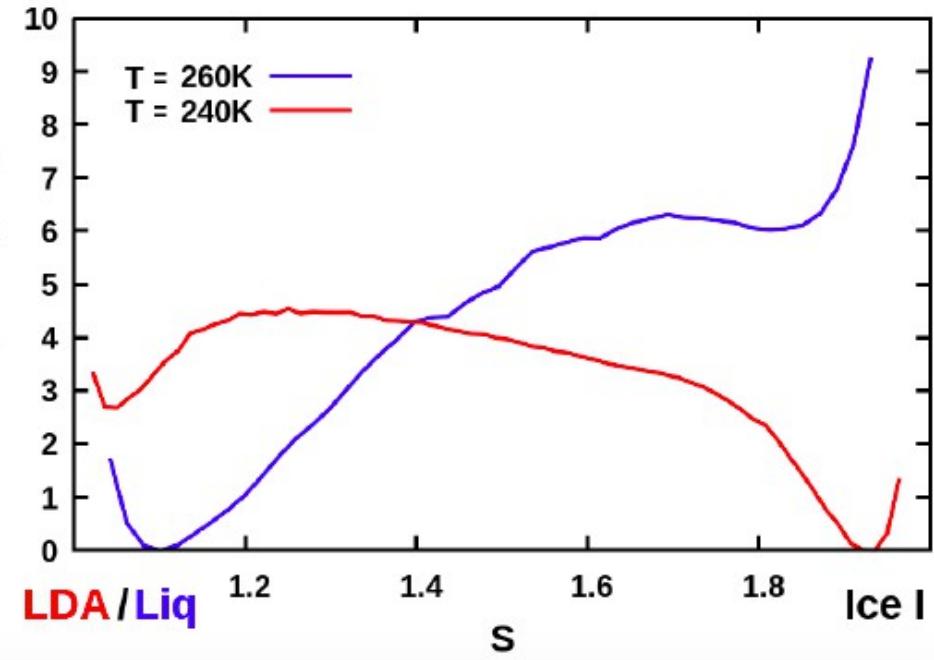
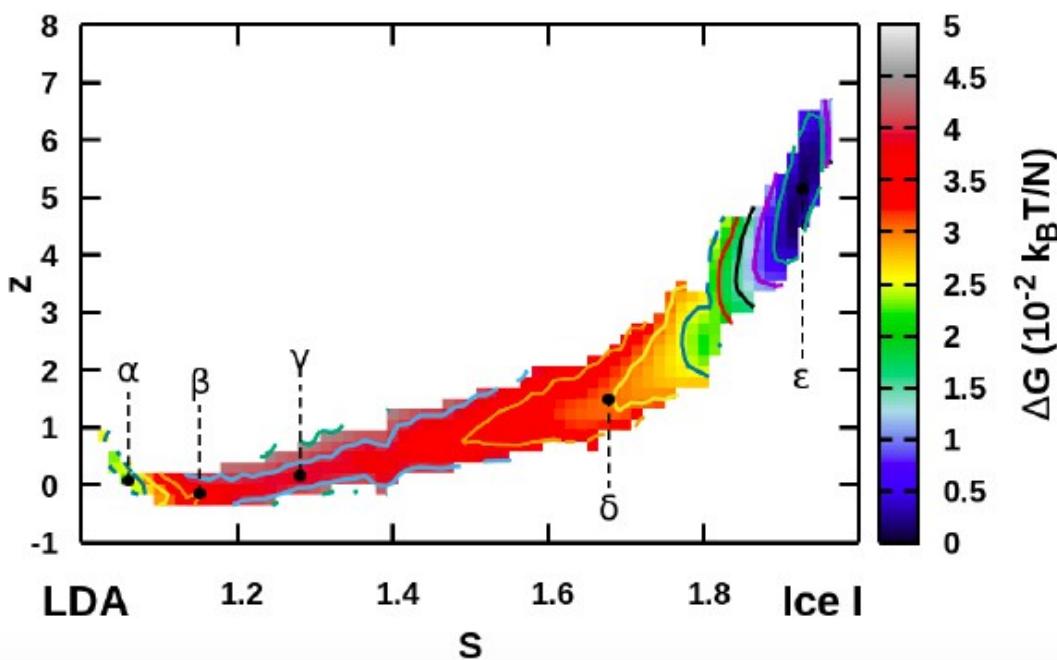
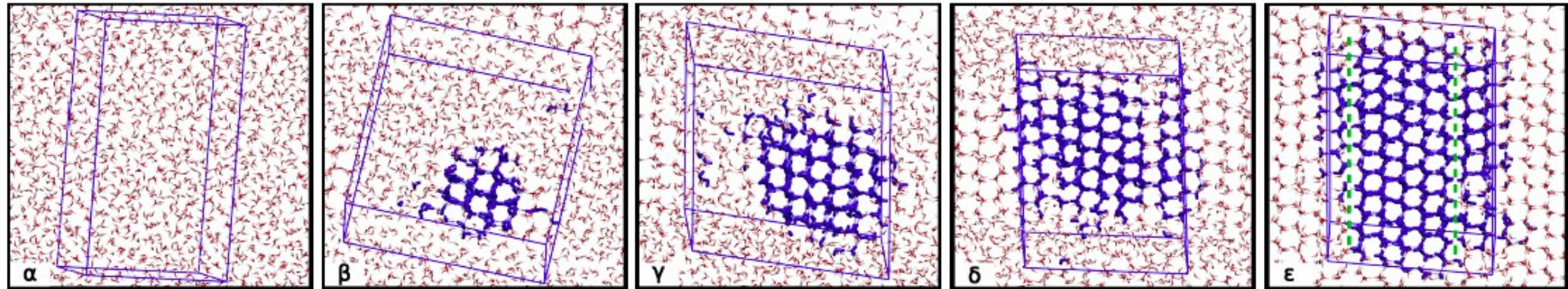
$$s_X = \frac{1 \cdot e^{-\lambda D_{AX}} + 2 \cdot e^{-\lambda D_{BX}}}{e^{-\lambda D_{AX}} + e^{-\lambda D_{BX}}}$$

$$z_X = -\lambda^{-1} \log (e^{-\lambda D_{AX}} + e^{-\lambda D_{BX}})$$

2-state path CVs:
from A to B

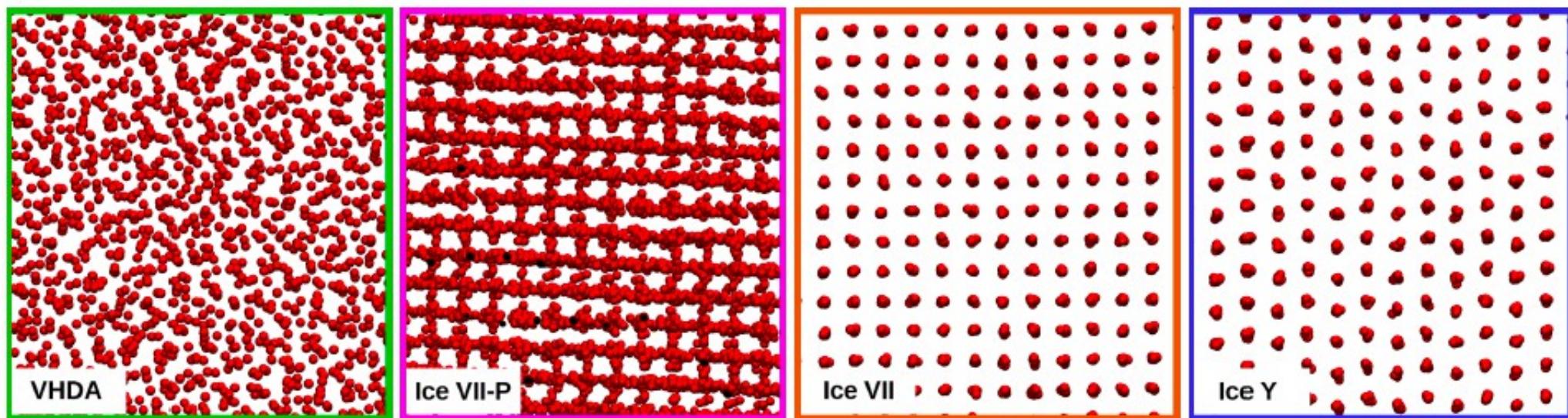
no guess about the mechanism: it is discovered

ice nucleation (above the melting point !)

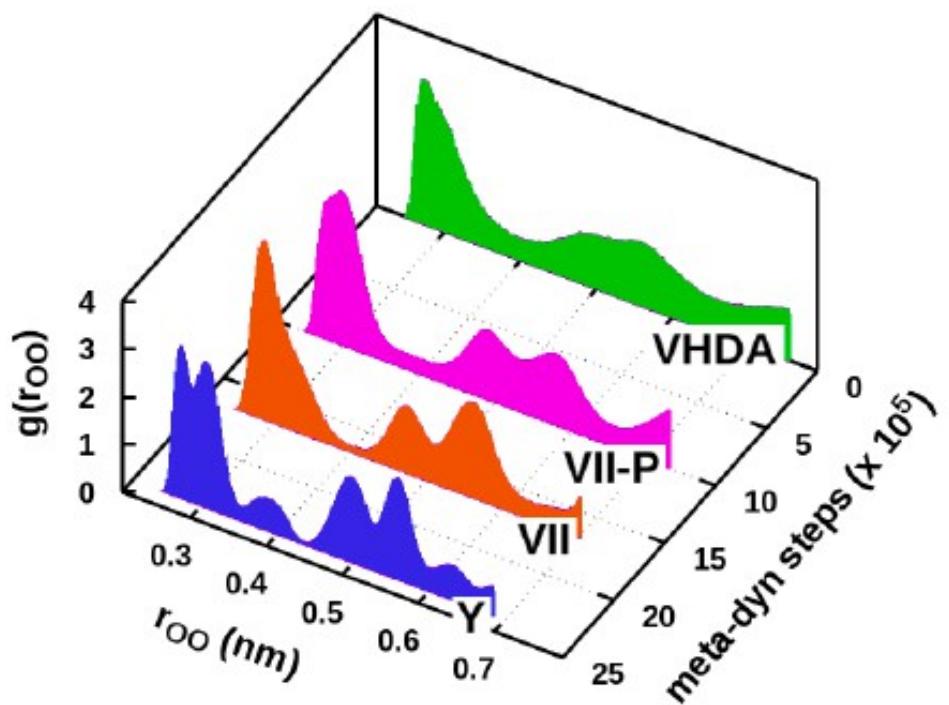


high pressure: disorder to order, and a new phase

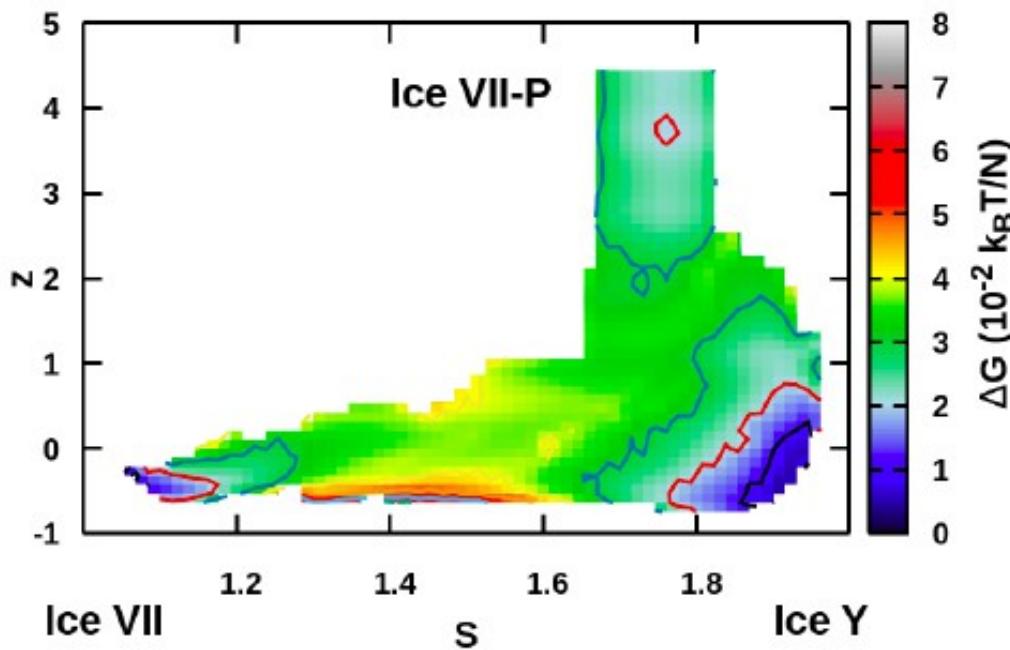
a



b



c



PIV – based path CVs

Pipolo, Salanne, Ferlat, Klotz, Saitta, Pietrucci, arXiv 2017

- very general approach: only 1 parameter (distance range)
- application to phase diagram of water, CO₂, SiO₂, B₂O₃ ...
- application to transitions of metal nanoclusters

implemented in



2

acknowledgements

Andrea Pérez-Villa

S. Pipolo

A. Marco Saitta

M. Salanne

G. Ferlat

S. Klotz

François Guyot

Sara Laporte

Giuseppe Cassone

M.-C. Maurel

UPMC Sorbonne

Wanda Andreoni

Grégoire A. Gallet

Changru Ma

EPFL



IDRIS

€ :



LABEX MATISSE



we recruit post-docs...