THE UNIVERSITY OF WARWICK

Centre for Scientific Computing

# Predictive modelling: a view from the atomic level

P.M. Rodger
Department of Chemistry
& Centre for Scientific Computing
University of Warwick

# Acknowledgements


- **David Quigley, Igor Khovanov**, **Matt Bano, Yuriy Bushuev**, **Aaron Finney, Ritchie-Mae Gamot, Yuanwei Xu, Salvatore Cosseddu** (Warwick)
- **John Harding, Colin Freeman** and **Riccardo Innocenti Malini** (Sheffield)
- **Paul Smeets, Heiner Friedrich, Jozua Laven, Nico Sommerdijk** (Eindhoven), **Wouter Habraken** (Postdam), **Fabio Nudelman** (Edinburgh)
- Mike Allen (Warwick), Dorothy Duffy (UCL), Julian Gale (Curtin)
- Warwick CSC
- HECToR & ARCHER
- MidPlus
- **EPSRC (mIb: Materials Interfacing Biology)**

# Philosophy

- Complex materials
  - complete model is not possible
- Mechanistic information
  - “predictions” not reliable without the correct underlying physics/chemistry
- Dimensionality
  - identify key degrees of freedom
  - target analysis
  - drive or simplify modelling

# Unexpected Mechanisms: old examples

- Decay of the Velocity Autocorrelation Function

  B. J. Alder and T. E. Wainwright
  Phys. Rev. A **1**, 18 – Published 1 January 1970

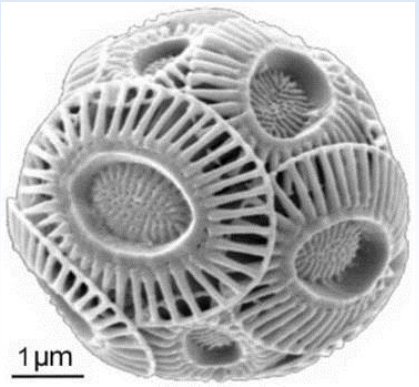
- Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events

  Arthur F. Voter
  Phys. Rev. Lett. **78**, 3908 – Published 19 May 1997

# Motivation: Biominerals

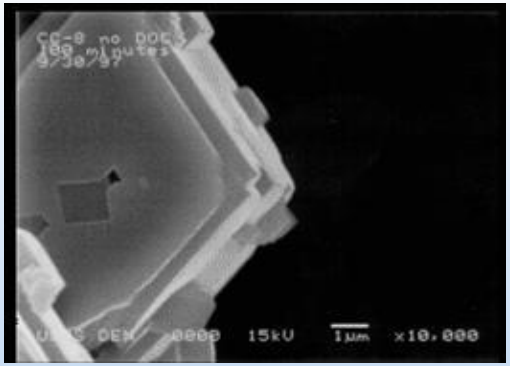
- Nature:
  - Exquisite control of crystal morphology & aggregation found in nature
  - Leads to very well "designed" functional materials
  - Ability to simulate implies ability to guide synthesis

Emiliania huxleyi coccoliths




Henriksen, K., S. L. S. Stipp, et al. *American Mineralogist* **89**, 1709-1716 (2004)

- Laboratory:




Reddy, M. M. and A. R. Hoch. *Journal of Colloid and Interface Science* **235**, 365-370 (2001)

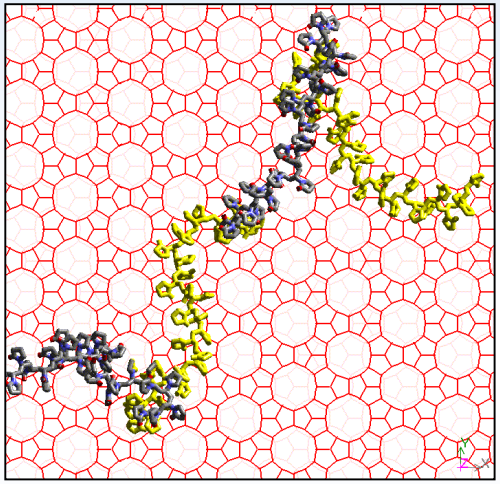
# Motivation: additives for inhibition

- Often need to suppress crystal growth
  - Scale, wax, hydrate
- "kinetic" inhibitors
  - *Delay* nucleation or *slow* growth
  - Active at low concentrations
  - Require ***molecular*** understanding of nucleation and growth

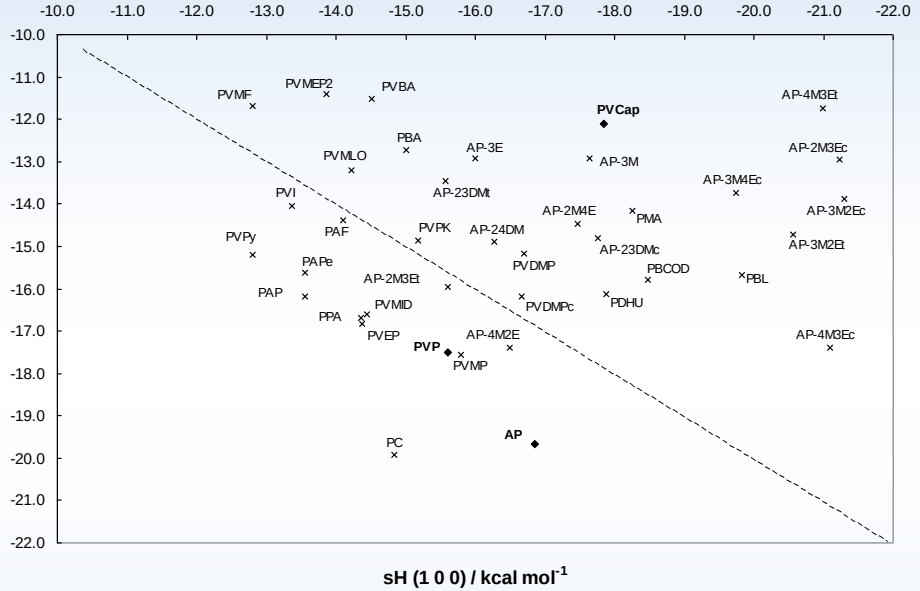
# Mechanistically inspired prediction?

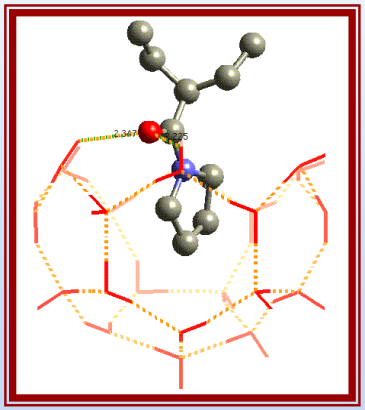
- Inhibition by surface adsorption

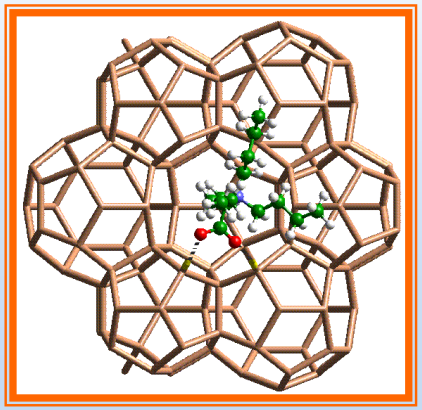
**validate**

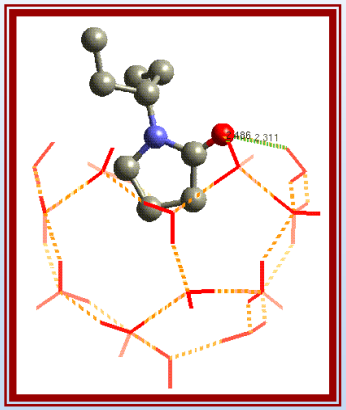


**screen**



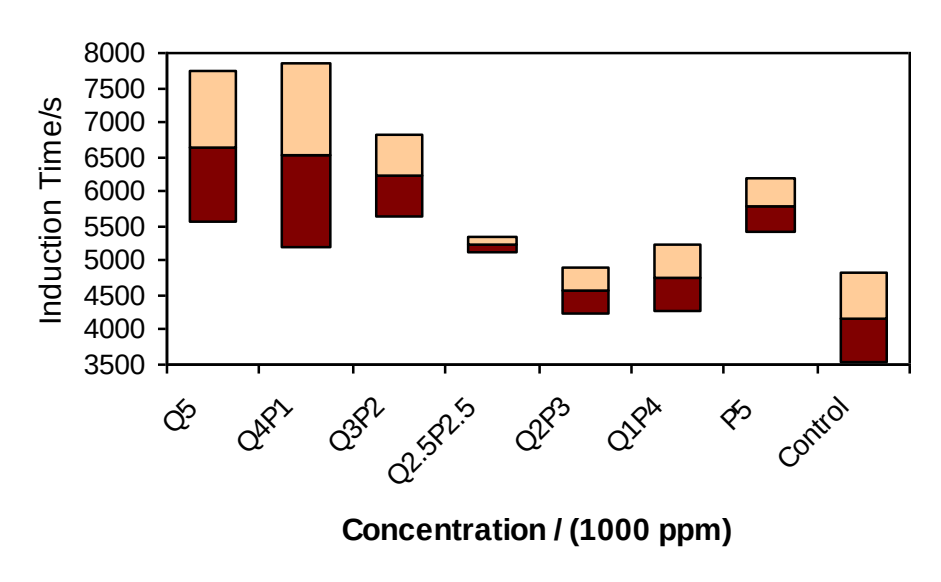







**new modes?**

# Synthesise and test




| | $t_{induction}$ / s | $R_{nucl}$ / (ml$^{-1}$ min$^{-1}$) | $R_{growth}$ / (μm min$^{-1}$) |
|---|---|---|---|
| Control | -90 (32) | 1439 (1031) | 1.56 (2.18) |
| tba3S (0.1 %) | 1705 (422) | 89 (–) | 0.61 (0.23) |
| tba3S (0.5 %) | 2678 (231) | 116 (13) | 0.86 (0.25) |

**THF Hydrate Tests**

induction times for 0.5% mixtures of tba3S (the “quat”, or Q) and PVP (P). The Control contains no inhibitor. Bars indicate ± one standard deviation

**Ethane Hydrate Tests**

Induction times, nucleation rates ($R_{nucl}$) and growth rates ($R_{growth}$). Standard deviations are given in parentheses

# 2nd generation

| Test No. | KI | Gas | Test P/T psia/C | Subcooling /C | Induction Time/ hrs |
|---|---|---|---|---|---|
| 1 | JI-C003:1 | NG | 1023 / 5.3 | 10.7 | 5 |
| 2 | JI-C003:1 | NG | 1023 / 5.3 | 10.7 | >67 |
| 3 | JI-C003:1 | NG | 1015 / 5 | 11 | >67 |
| 4 | JI-C003:1 | NG | 1023 / 5 | 11 | >150 |
| 5 | JI-C003:1 | Methane | 1700 / 4.5 | 10 | >150 |
| | Commercial | NG | | 10 | 16 |
| | Commercial | Methane | | 10 | 5 |
| 6 | JI-C002:1 | NG | 1095 / 4.5 | 12.3 | 3 |
| 7 | JI-C002:1 | NG | 1095 / 4 | 12.8 | 40 |
| 8 | JI-C002:1 | NG | 1095 / 5 | 11.8 | >42 |

- 2nd generation inhibitors perform about 4 times better than current commercial inhibitor

# But!

- Experiments not repeatable!
  - Oil companies can't repeat each other's screens
  - Heriot-Watt: activity changed with test-cell stirrer design
  - Scale of activity with new compound by test-site
    - Heriot-Watt > Toulouse > Halliburton > Heriot-Watt

- Similar story with wax inhibition (model oil using well characterised edible oil mixture)

- Experimentalists working with nanotoxicology want legislative standards based on computational tests because experiments are too irreproducible!

- Simplistic model: dynamic interface; nucleation; …

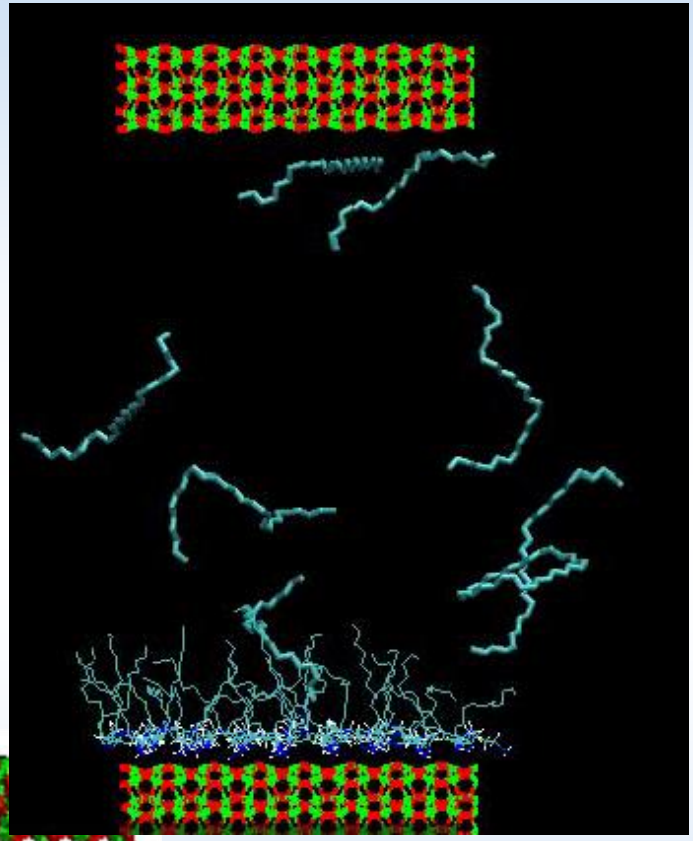
# Nucleation from MD?

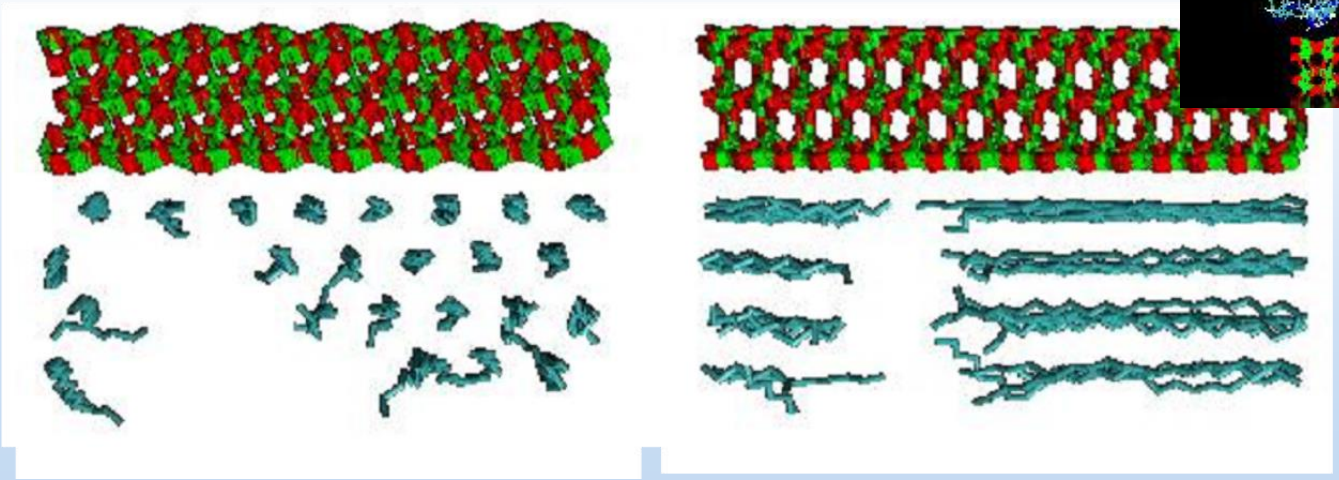
- Homogeneous nucleation?
  - Experimental nucleation rates < $O(10^6)$ nuclei per $cm^3$ per second
  - MD simulation 10,000 molecules:
    - 1 nucleation event every 32 millenia!

- Heterogeneous nucleation?
  - (some) success with seeding
  - (some) success with interfaces

# Standard MD: modelling surfaces

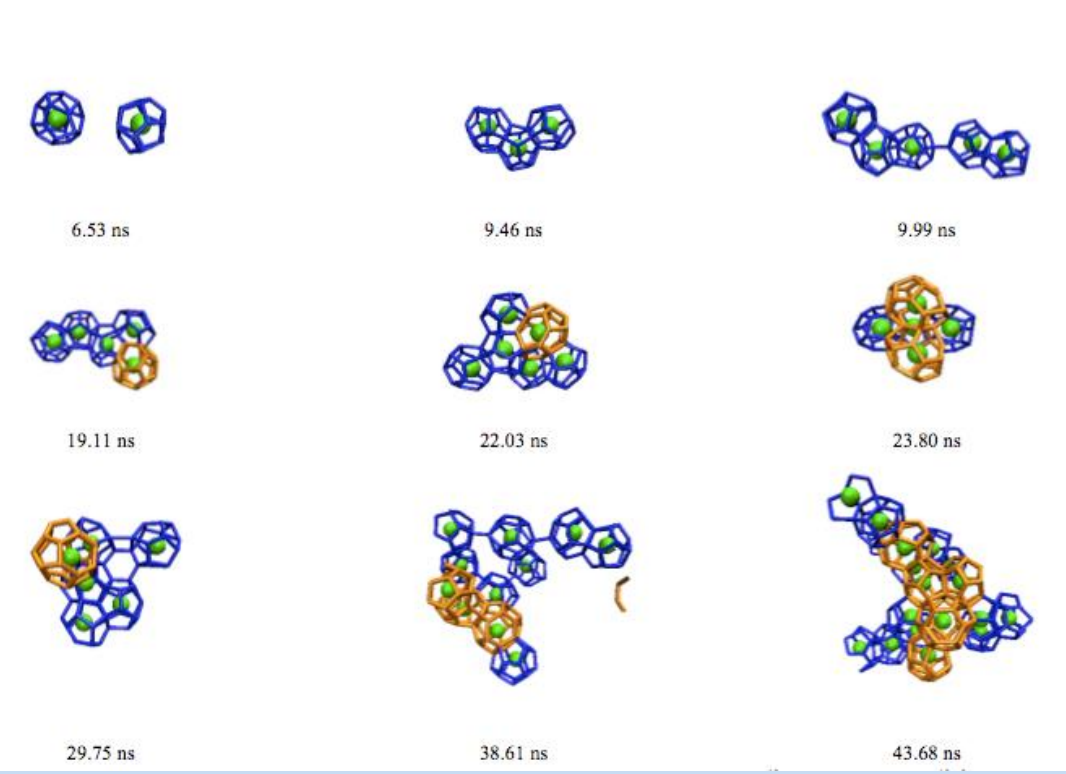
- Deposition of wax on $Fe_2O_3$
  - $C_{28}$ from $C_7$

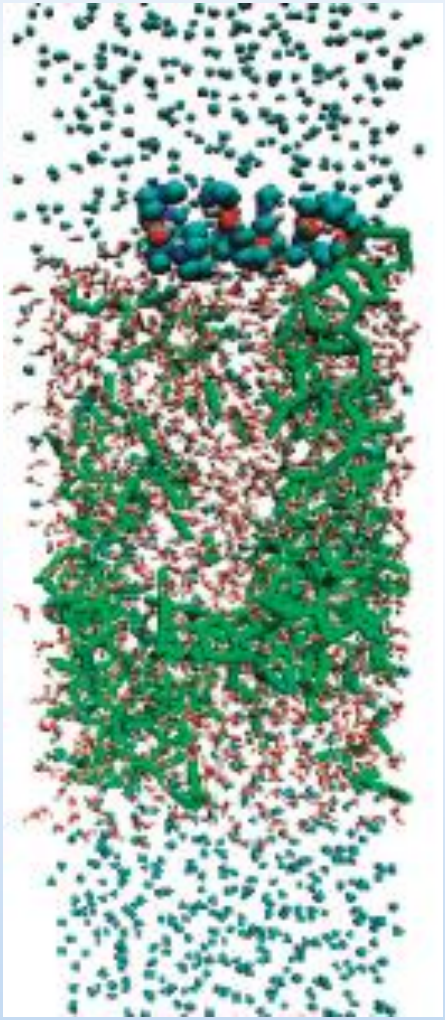




# Standard MD: crystal nucleation

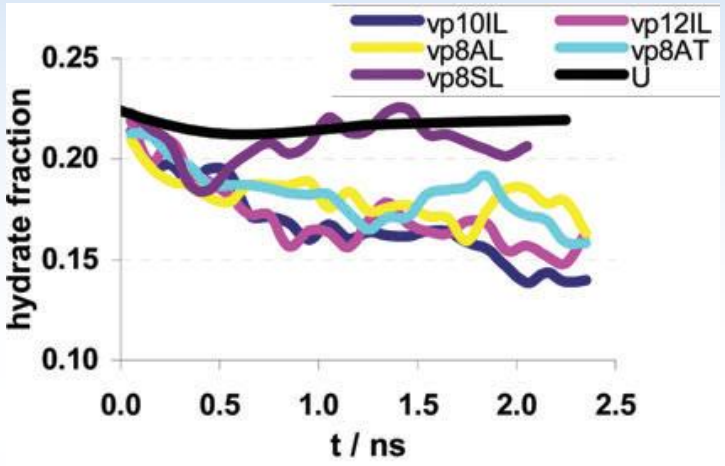
- heterogeneous and/or
  high supersaturation/subcooling



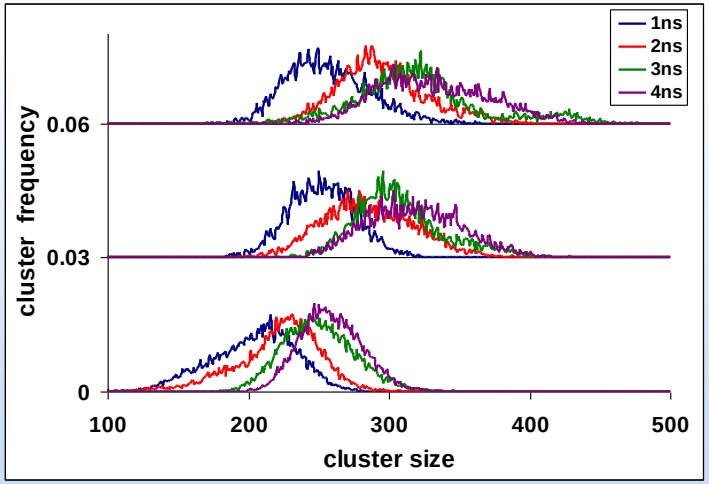



# Polydispersity in inhibitors

- molecular weight
  - ➢ amount of hydrate present
  - ➢ PVP




- molecular shape
  - ➢ growth of hydrate clusters
  - ➢ pDMAEMA

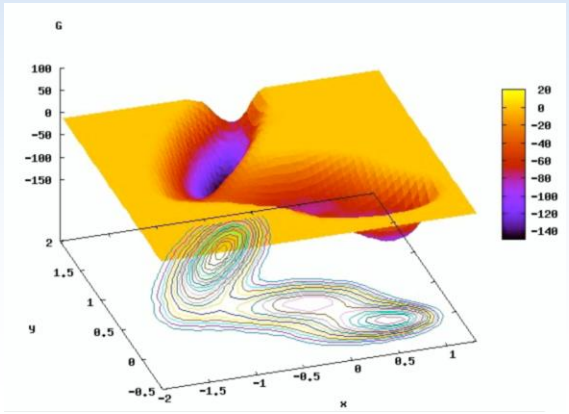


# Nucleation more generally

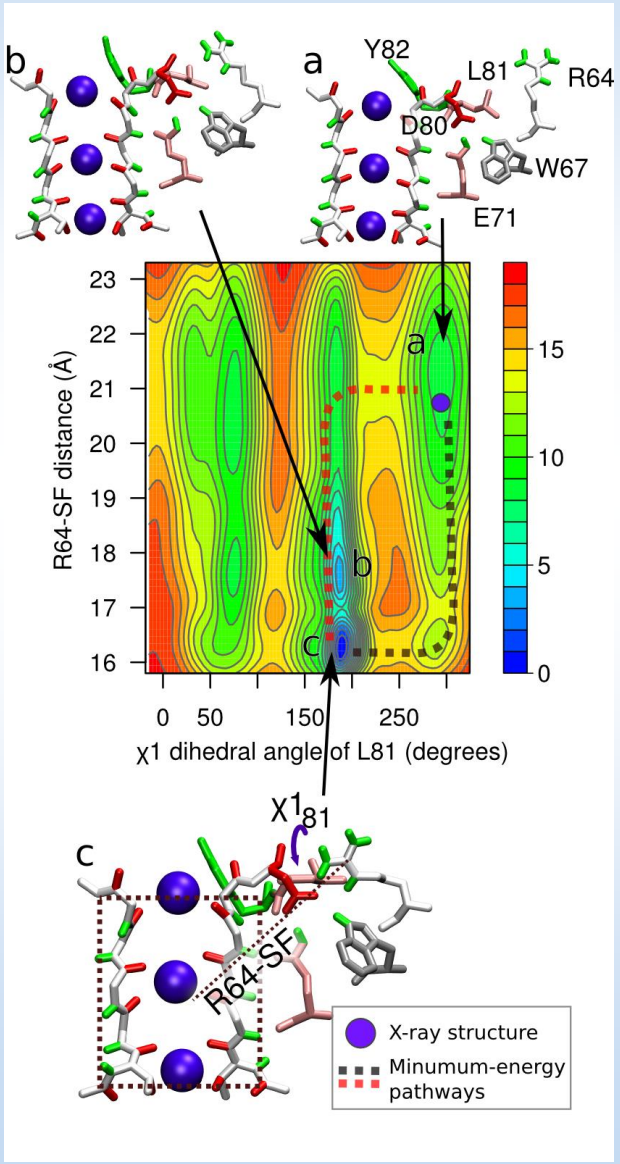
- Project onto “important” manifold
  - Define key dimensions (order parameters / reaction coordinates / collective variables)
  - Project analysis onto these dimensions
  - Bias simulations to explore these dimensions
    - constrain to portion of manifold (umbrella sampling)
    - trap fluctuations that explore the manifold (adaptive bias force)
    - disfavour current portion of manifold (Wang-Landau; metadynamics)

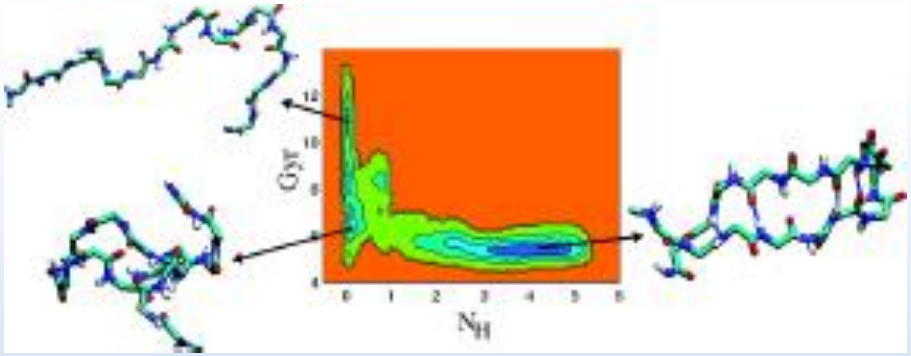
# ⇒ Free energy hypersurfaces

3-basin model
Quigley, PMR



KcsN ion channel
Cosseddu, PMR, Khovanov

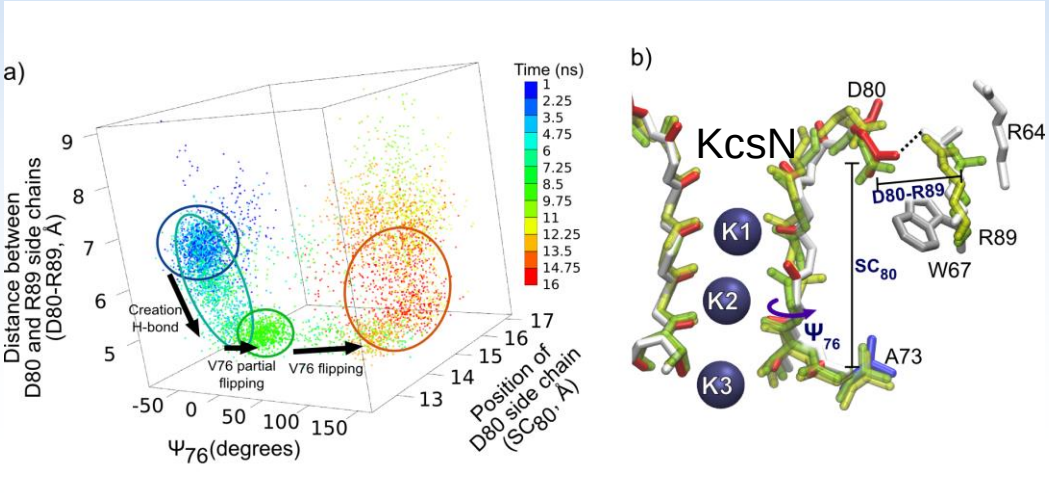


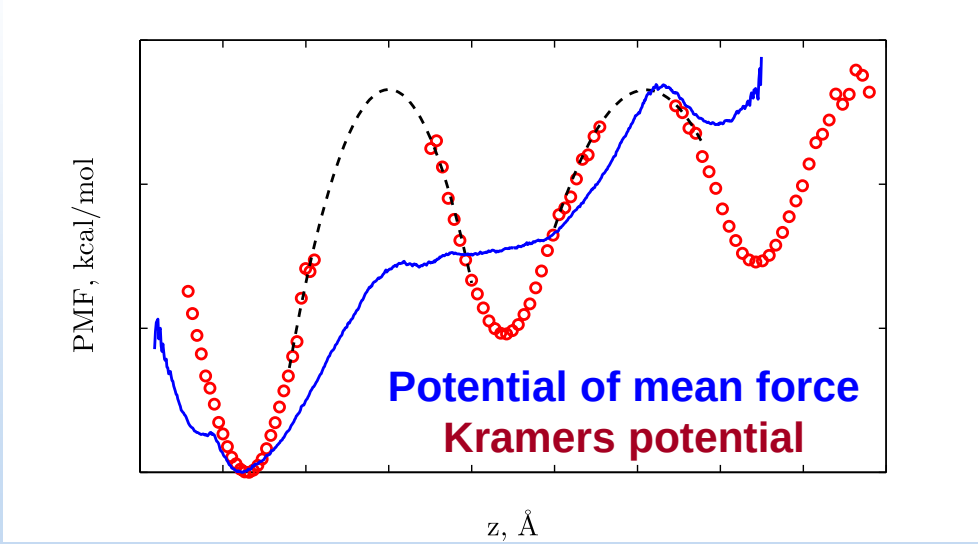
peptide folding
Bussi, …, Parrinello

# Limitations and Pitfalls: dimensionality!

- Equilibrium / comprehensive sampling
  - ➢ Typically 1–2D (≤ 6; can do ~100 for basin escape & rough sampling)
  - ➢ projection merges basins

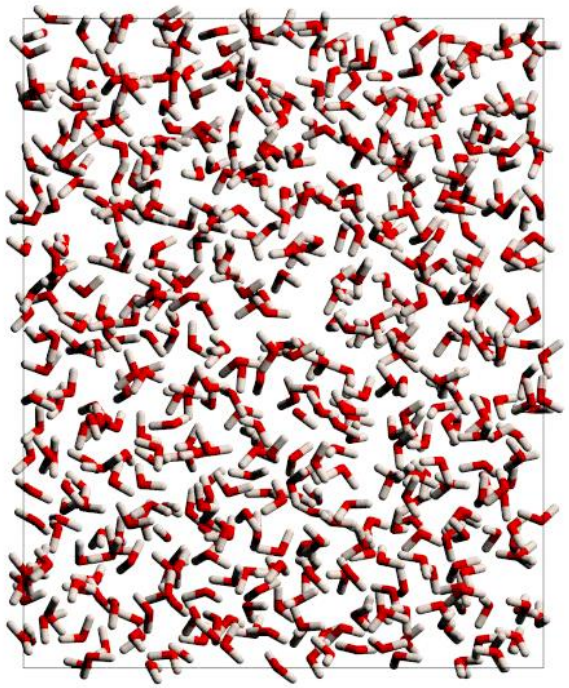



- Adiabatic Surfaces
  - ➢ dynamics defined by metastable substates?

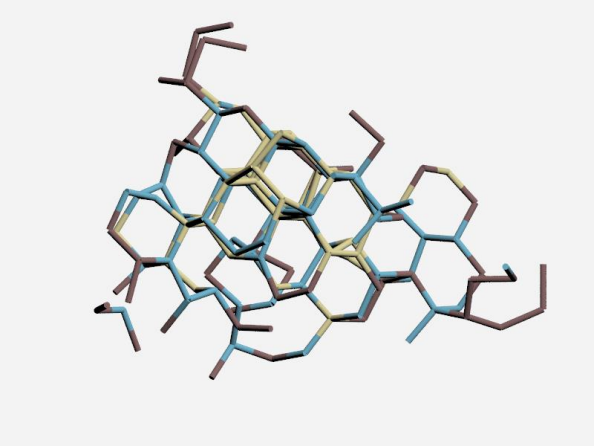


# Nucleation with metadynamics: ice

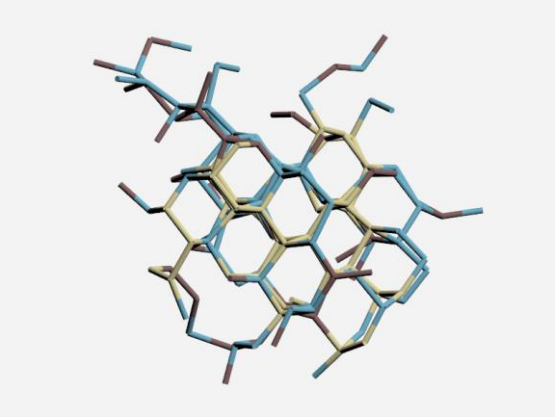
- 4D bias:
  - Q4, Q6, ç and potential energy (simaltaneous)
- *NPT* simulation
  - density change is spontaneous
- *T* = 180 K; *ca.* 11 “ns” simulation
- Results independent of periodic boundaries



# An *ensemble* of critical nuclei

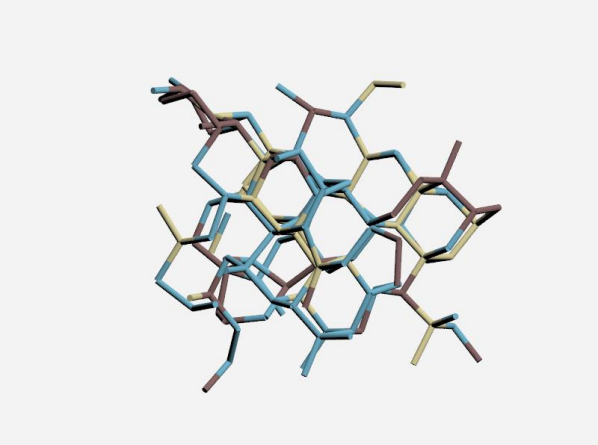


N = 147

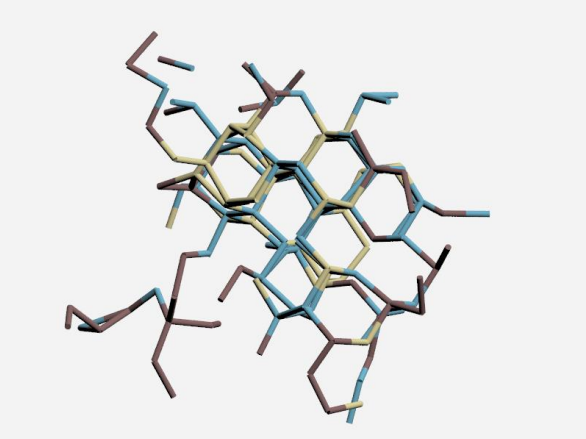


N = 157

ice $I_c$
ice $I_h$
both



N = 138

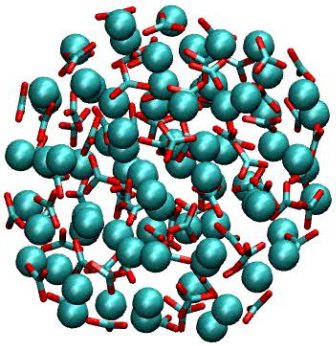


N = 161

# Nucleation with metadynamics: $CaCO_3$

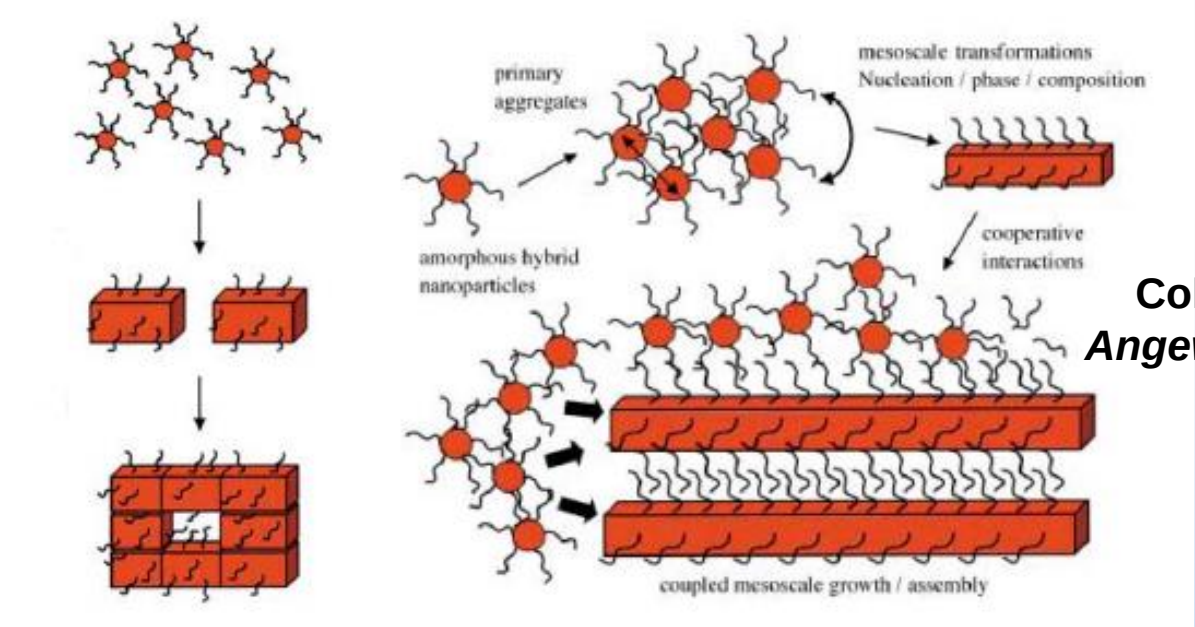
- 6 order parameters
  - $Q_4$ for: Ca-Ca, Ca-C, Ca-O, C-C, C-O
    - measures orientations of X about Y
  - energy associated with $CaCO_3$
- *ca.* 10 ns of MD

Exploring configurations for 75 units (water not shown)



# Calcium Carbonate Biominerals

- Multi-stage hierarchical formation
- ο ρ γ α ν ι χ  χ ο ν τ ρ ο λ




**Colfen & Mann**
***Angew. Chem*, 2003**

# Early stages of carbonate biomineralisation

ions & ion pairs $\rightarrow$ hACC $\rightarrow$ ACC $\rightarrow$ crystalline

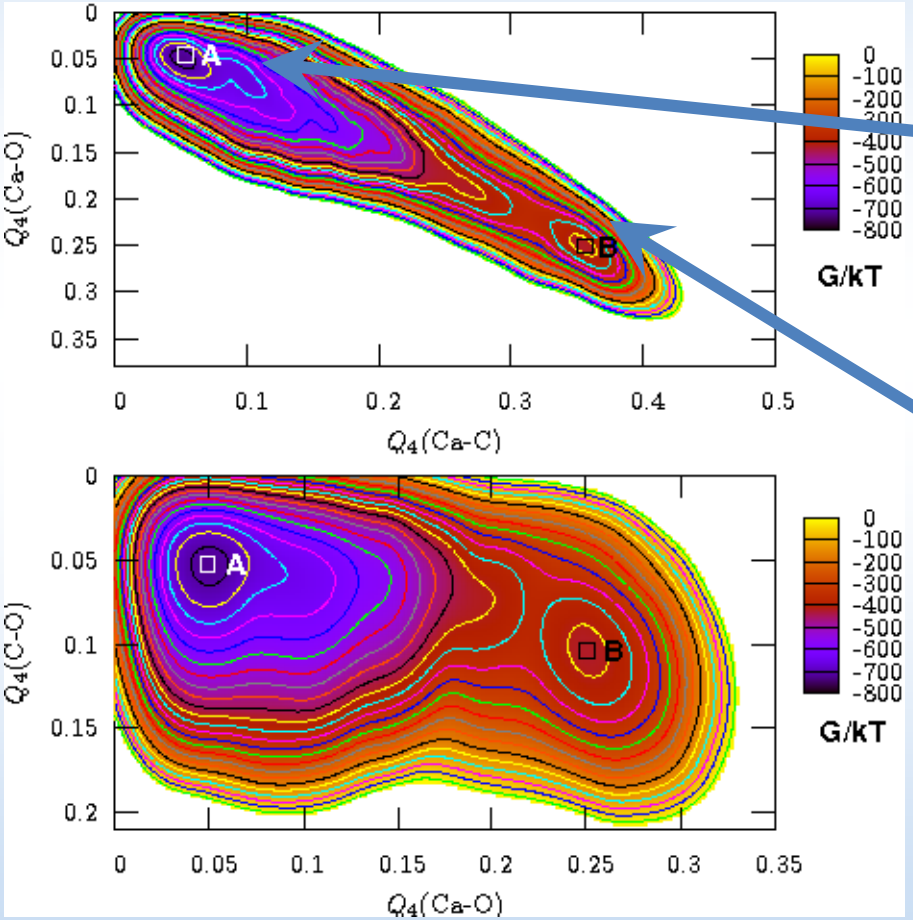
hACC: hydrated amorphous calcium cabonate
composition *ca.* $CaCO_3.H_2O$

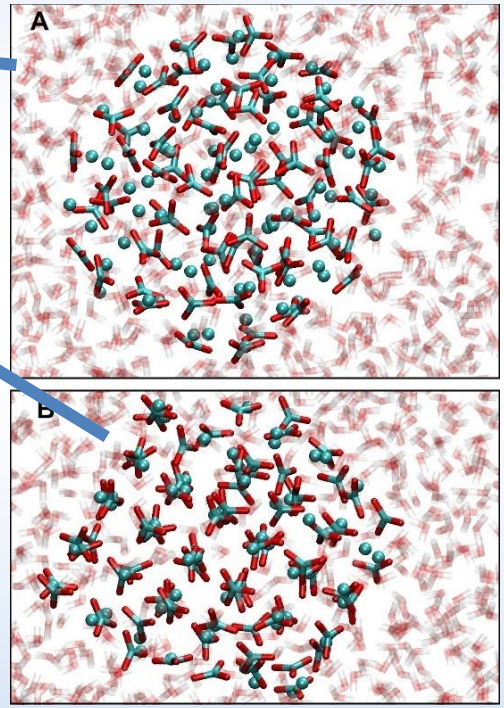
ACC: (anhydrous) amorphous calcium carbonate

Crystalline: calcite, aragonite, vaterite

# Free energy maps: 75 $CaCO_3$ units

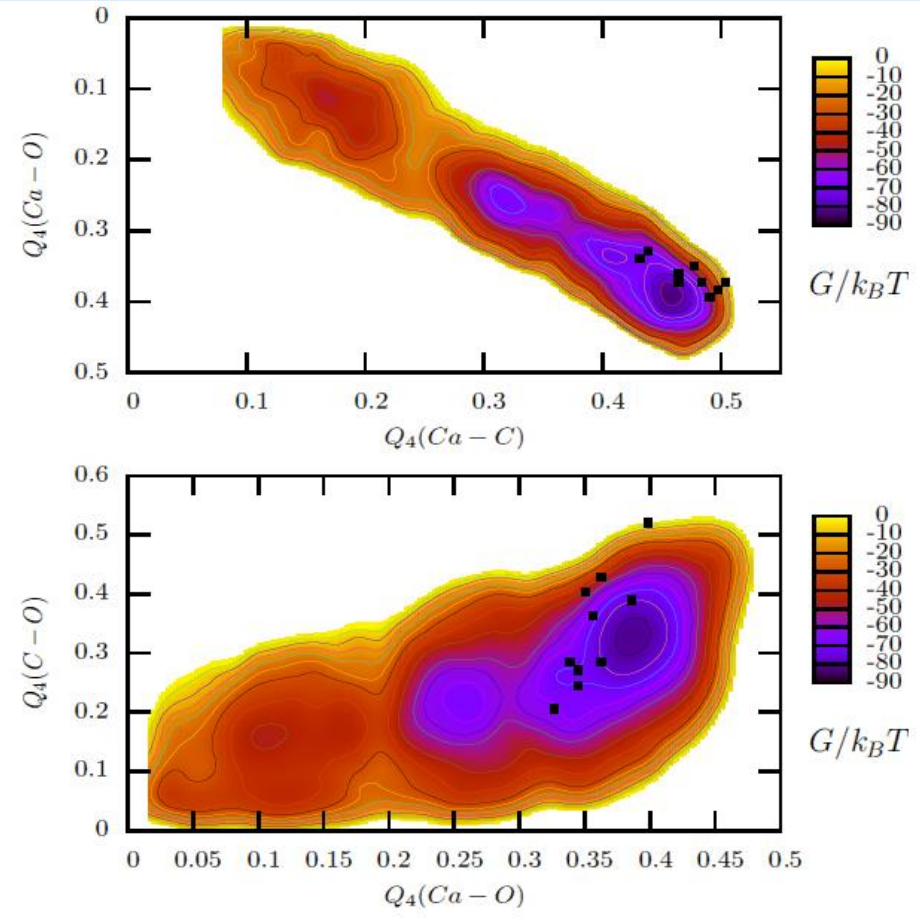
- *NVT*
- favours amorphous




DQ and PMR, *J. Chem. Phys.* **128** (2008) 221101

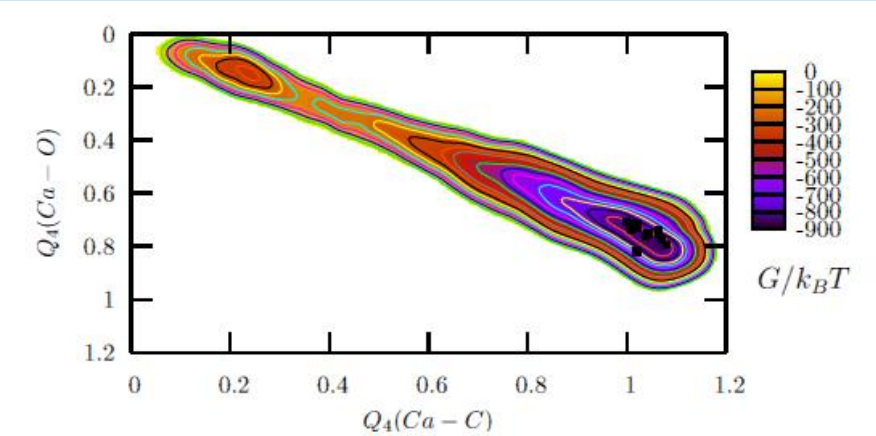
# Free energy maps: 75 $CaCO_3$ units

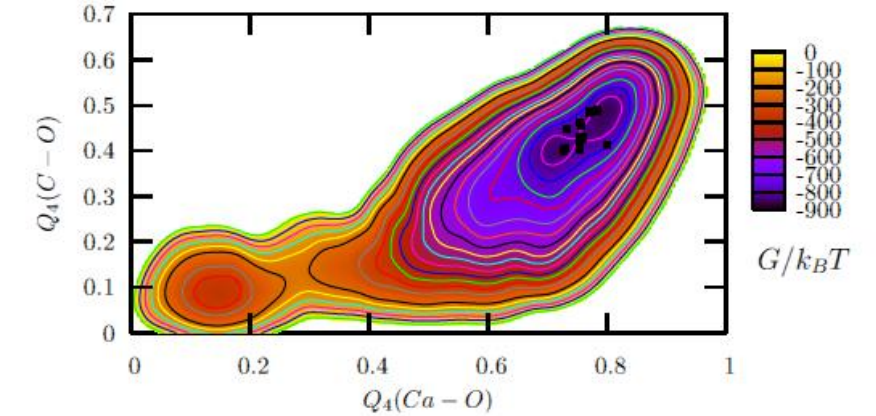


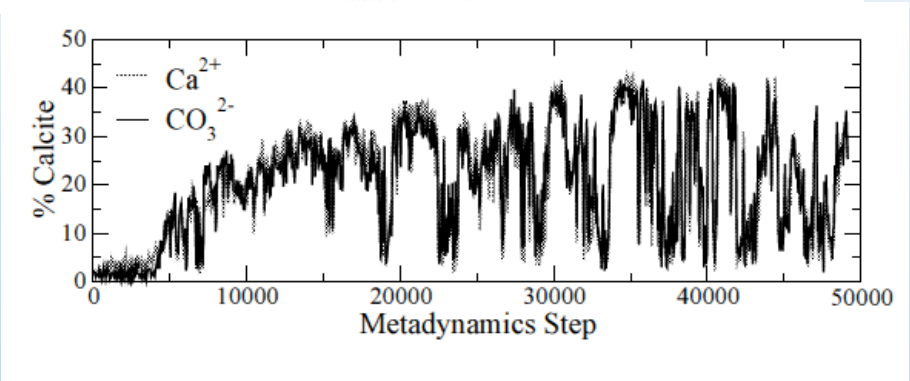

- *NPT*
- Dominated by calcite
  - ➢ very low barrier (10–20 *kT*)

DQ, CLF, JHH and PMR, *J. Chem. Phys.* **134** (2011) 044703

# 300 $CaCO_3$ Units in water

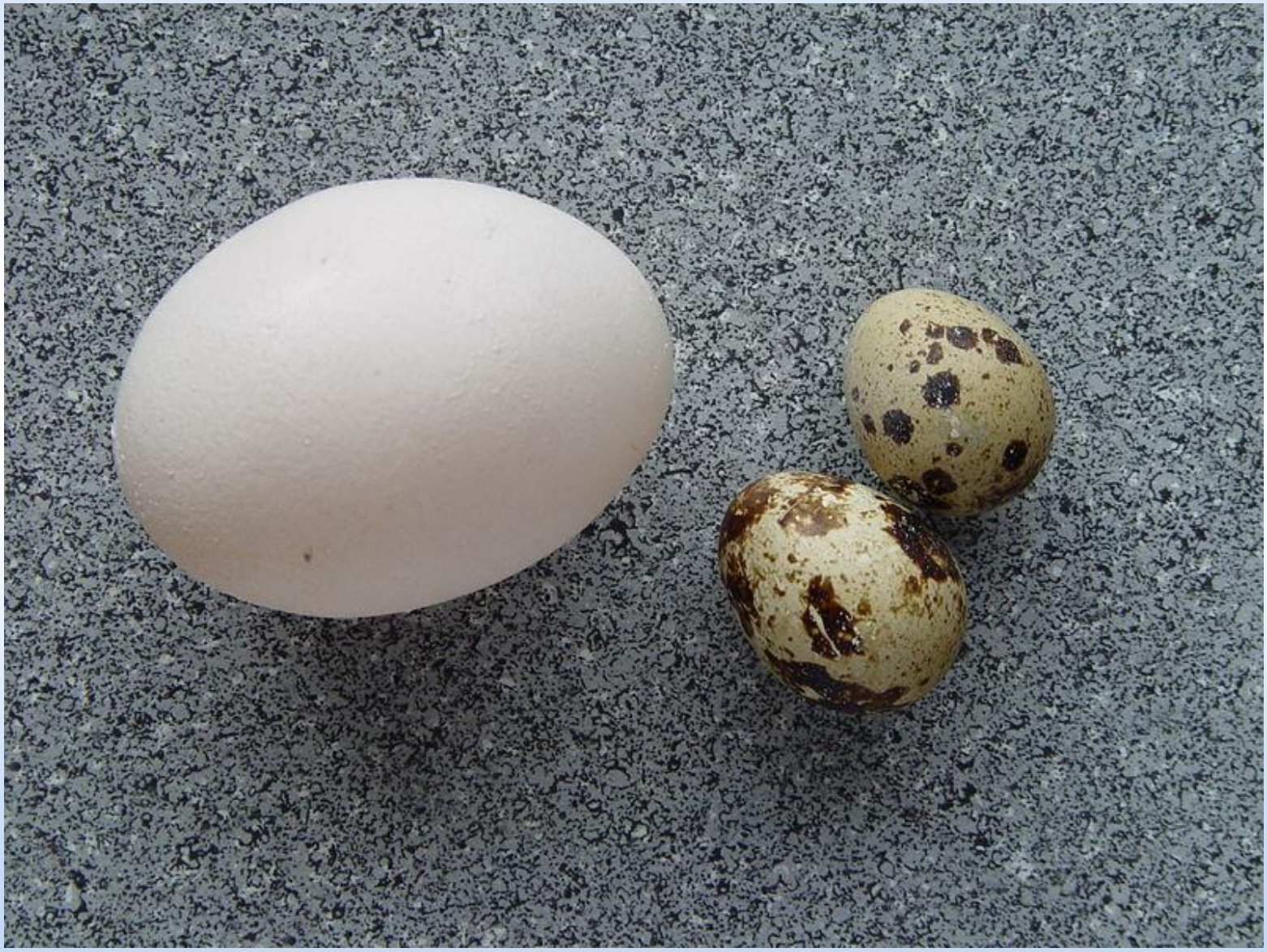


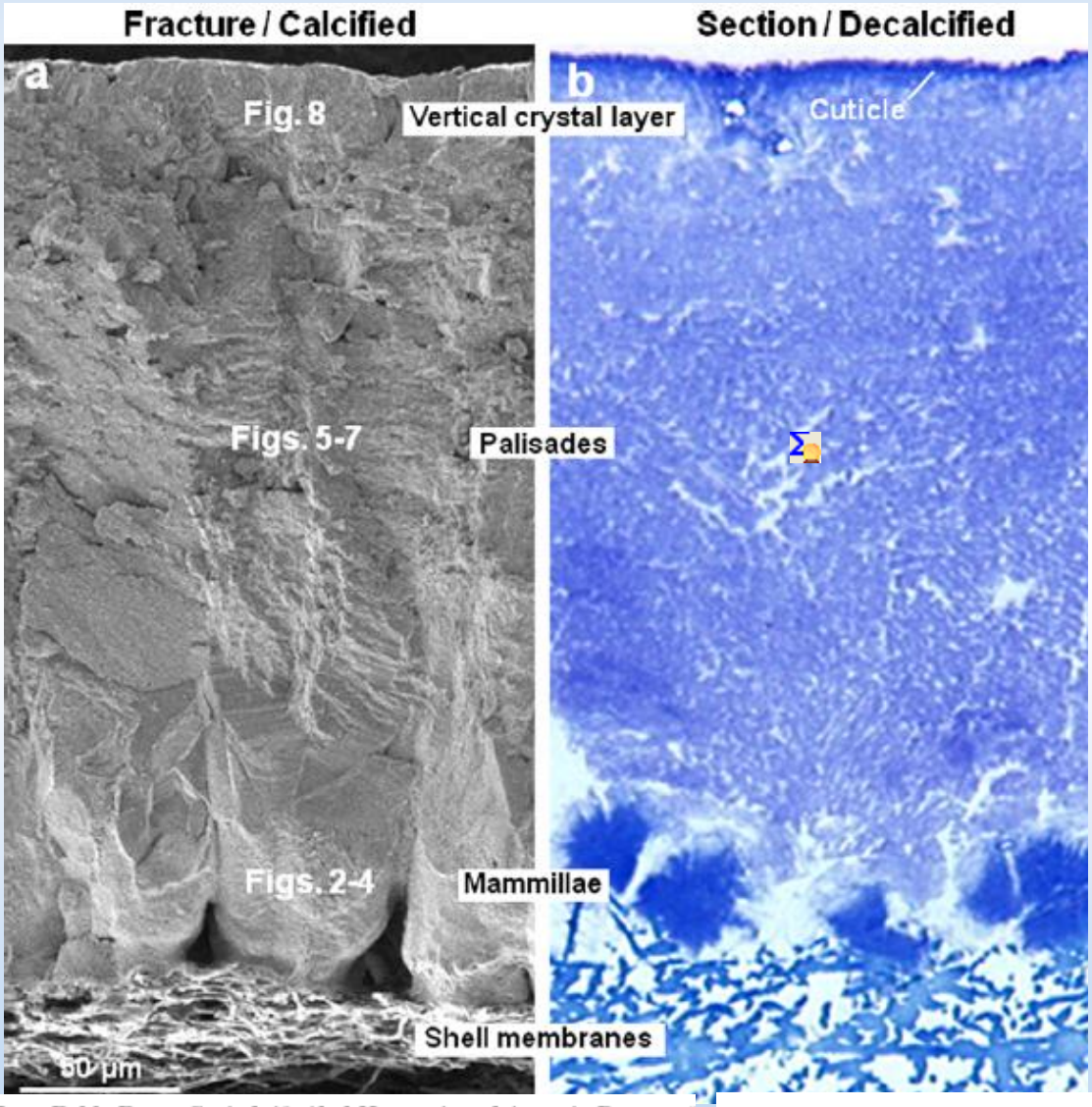




- two stable states
- calcite (dominant)
- amorphous
- energy barrier > $10^2$ *kT*
- multiple transitions

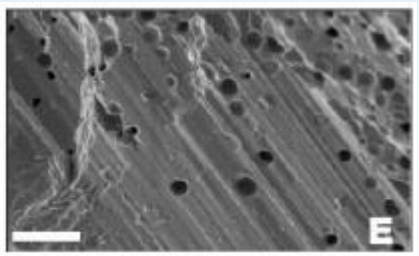
# Biomineralisation: egg shells



# Structure of an eggshell

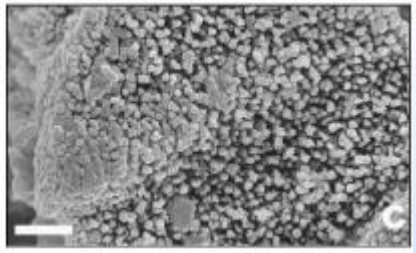



Juan Pablo Reyes-Grajeda‡§, Abel Moreno‡, and Antonio Romero¶‖ THE JOURNAL OF BIOLOGICAL CHEMISTRY



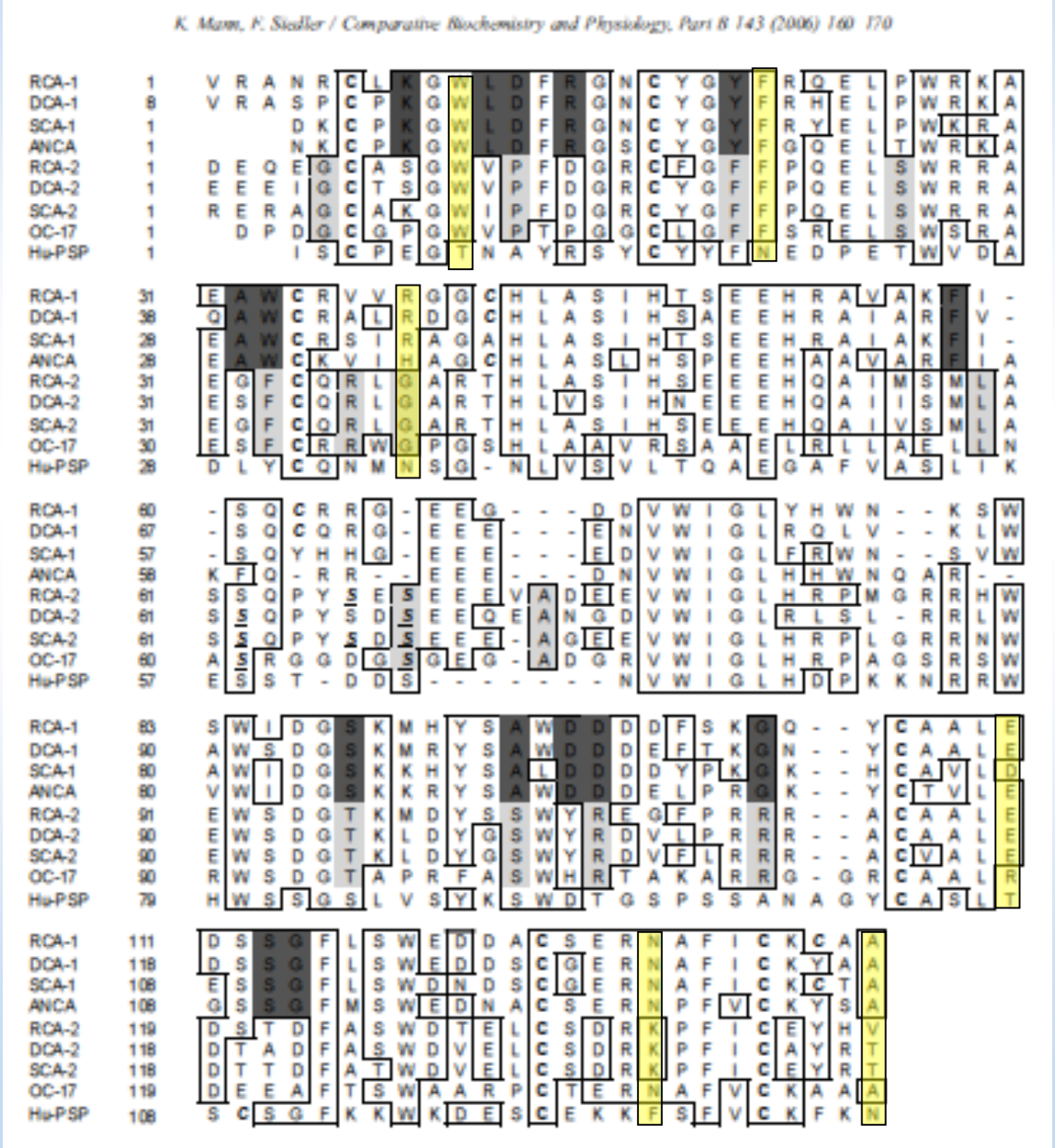

*Biomacromolecules, Vol. 7, No. 11, 2006* **3203**

Lakshminarayanan et al.

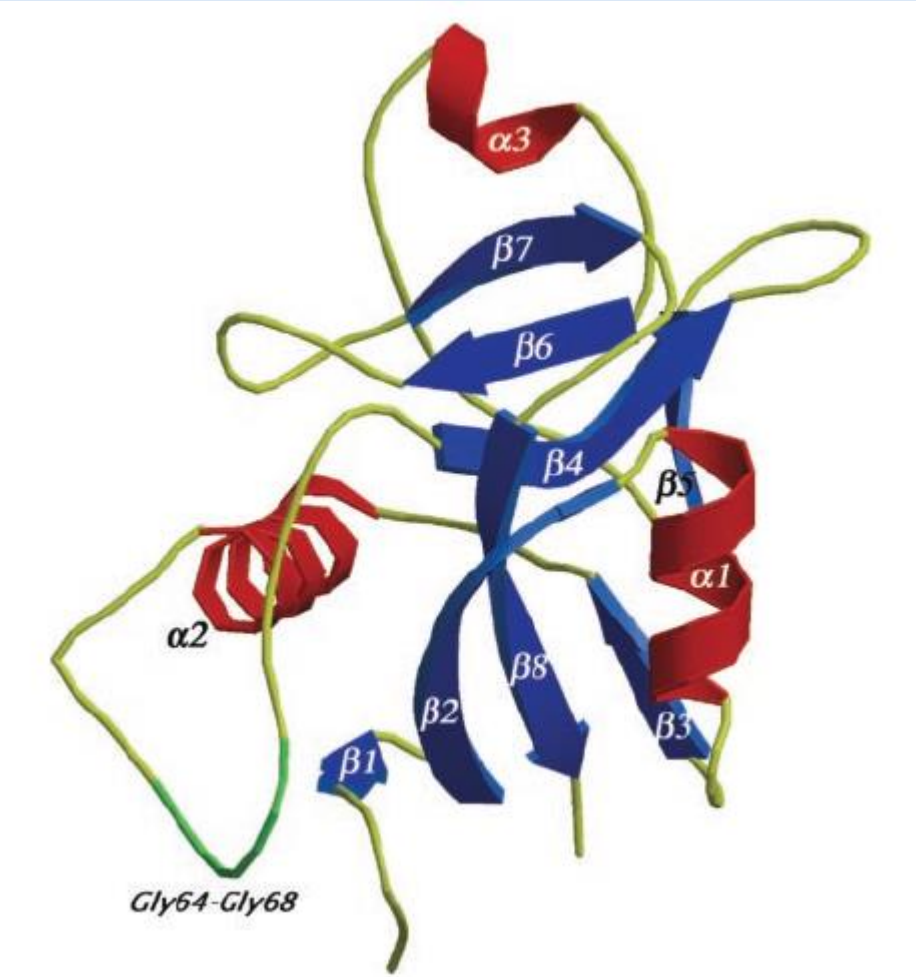


# G-type lectin-type proteins

- found in
  - Rhea
  - Duck
  - Chicken
  - Emu
  - Ostrich



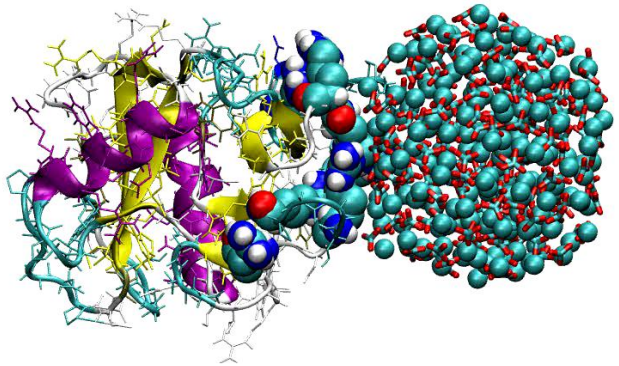
# Ovocleidin-17




**Reyes-Grajeda, Moreno, Romero;**
***J. Biol. Chem.*, 2004**

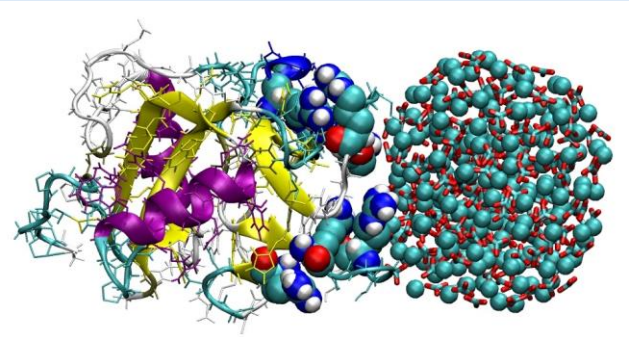
# metaDynamics of OC17 + $CaCO_3$

- 192 & 300 unit nanoparticles
  - 22,000 water molecules
- 20 different protein/nanoparticle orientations
- Select “best” 4 for long meta-dynamics
- Potentials: due to Freeman *et al.*
  - Pavese $CaCO_3$ ; Amber protein; Tip3p water;
  - cross terms derived from crystals with scaled charges



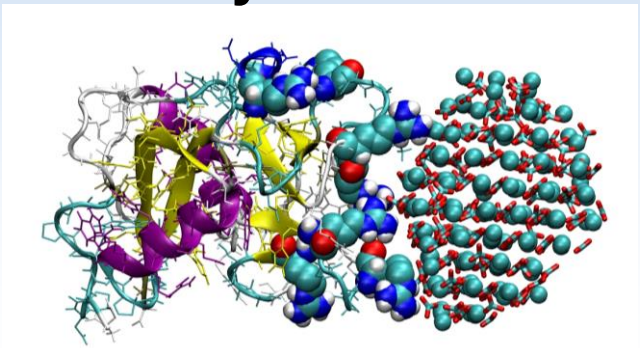
# Free energy landscapes

**Initial**



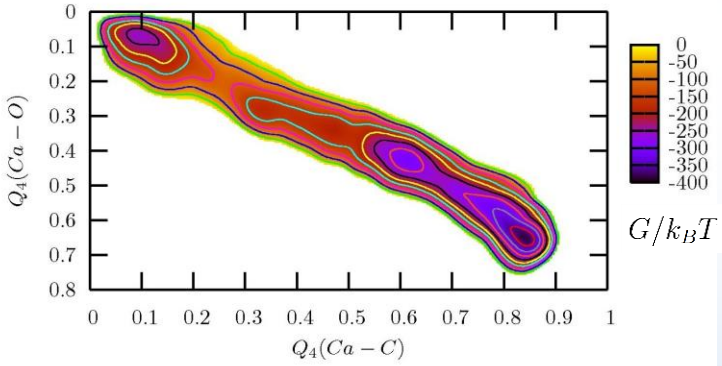
**metadynamics**

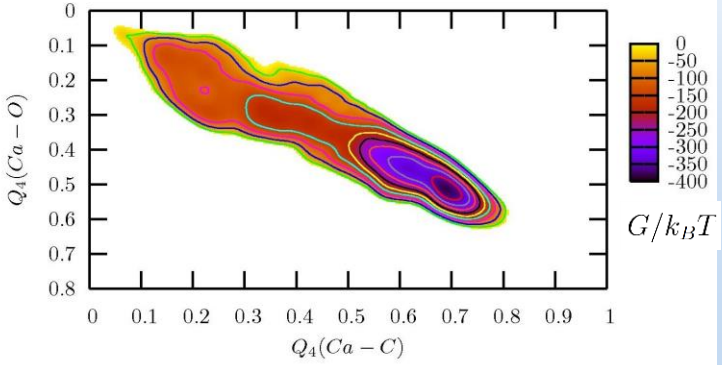
**Crystallised**



- 192 units of $CaCO_3$ in water

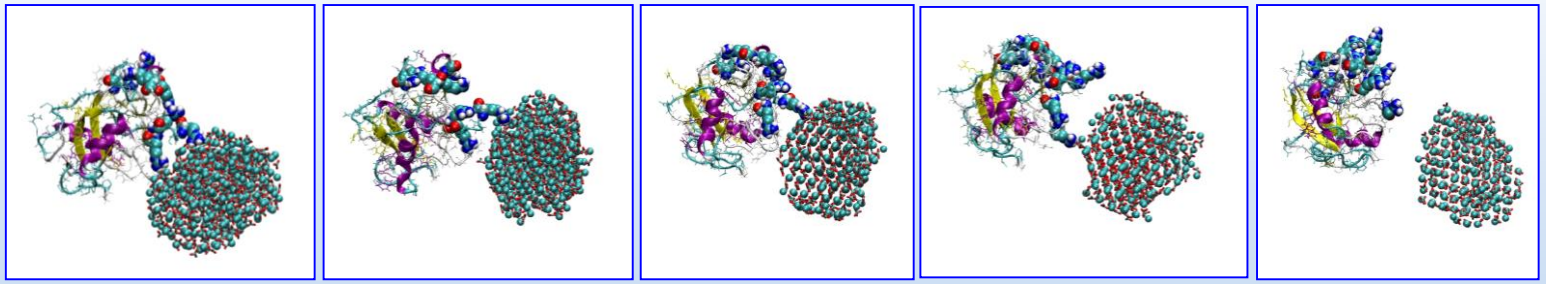
- 192 units of $CaCO_3$ bound to Ovocleidin-17 in water



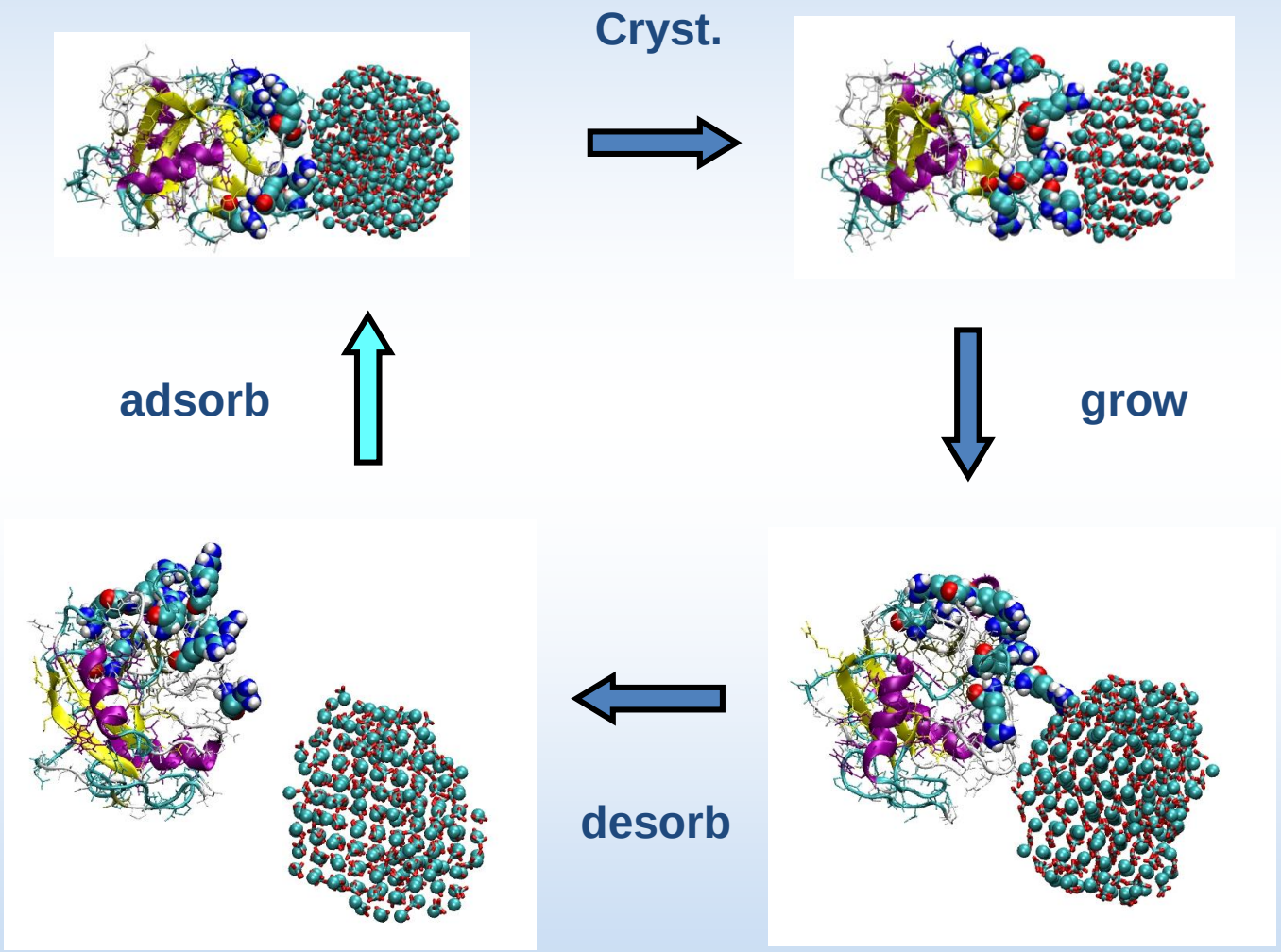



# Larger nonaparticles?

- 300 $CaCO_3$ unit particles don't stay bound on crystallisation
- Mechanism for proteins
  - Bind to small nanoparticles
  - Facilitate transformation to calcite
  - Desorb as crystal grows
- Gives catalytic cycle for polycrystalline mamillary layer

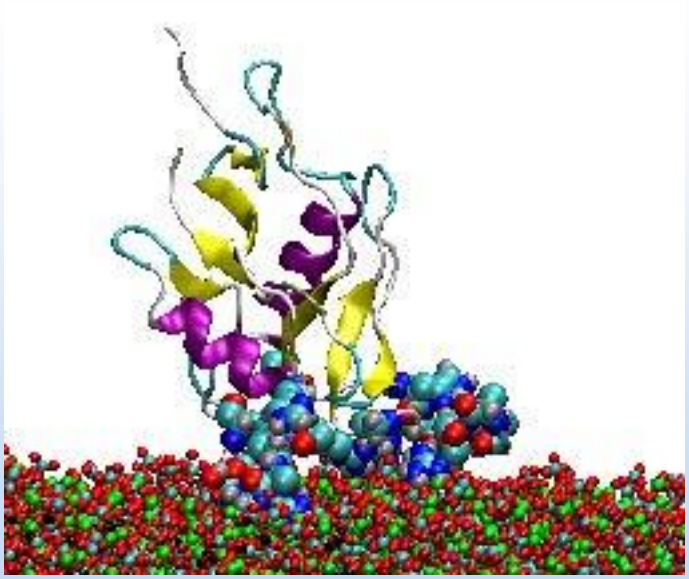


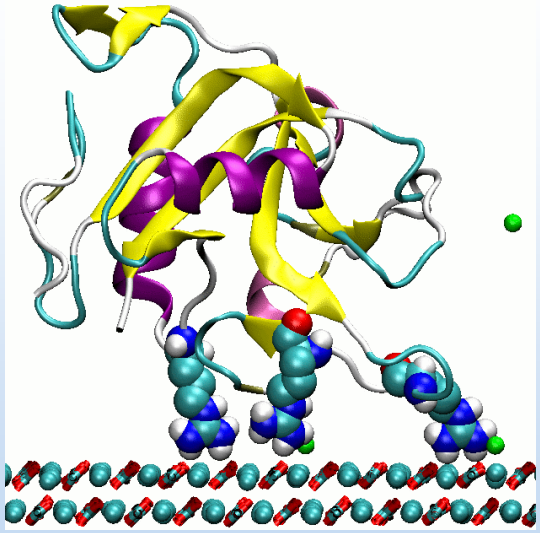
# OC17: catalytic cycle?



# Surface binding & structured water

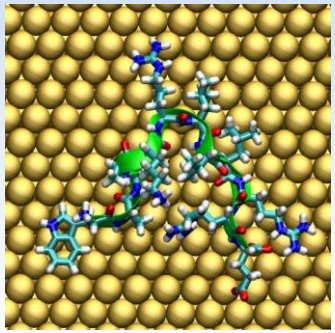
- Binding is seen to planar surfaces
  - Structured water dominates adsorption energy on large crystalline surfaces
  - Calcite nanoparticles and ACC do not give structured water layer; leads to “flat” adsorption

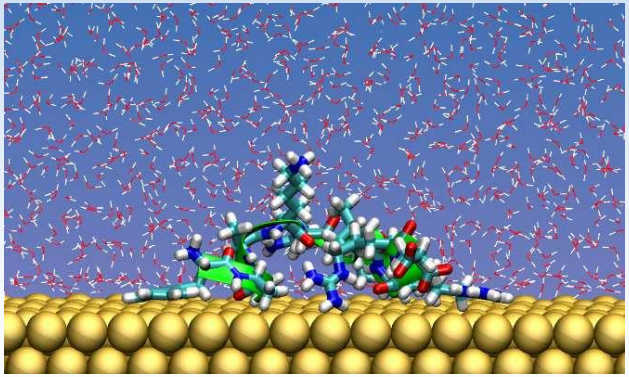




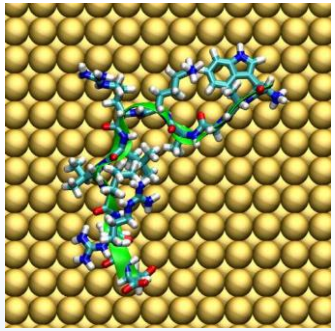
# Also seen with AuBP1 on gold

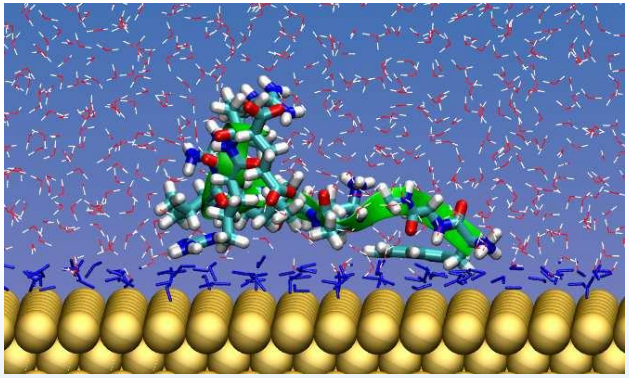
(100) 5x1



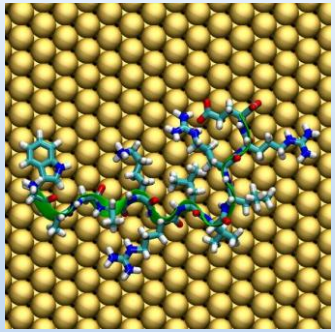


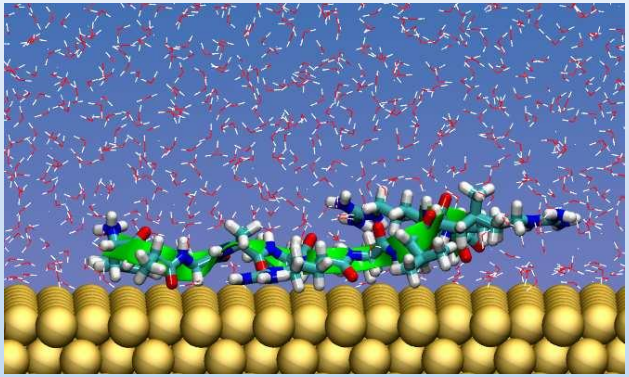
(100)





(111)





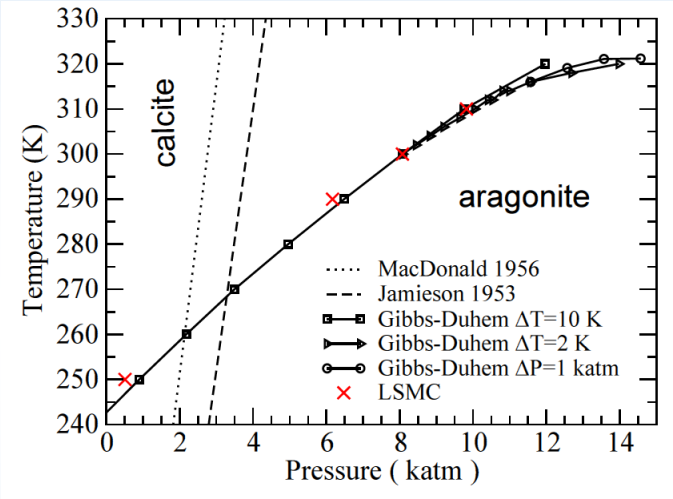
# hACC nulcleation

- Problem
  - intermolecular potentials

- Controversy

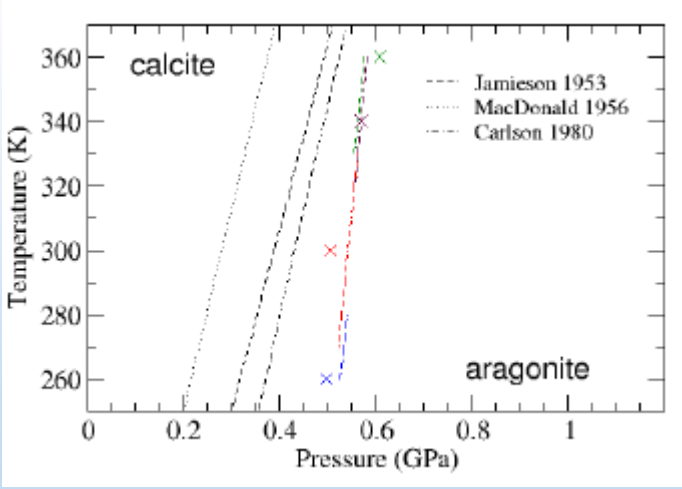
# Aragonite?

- needs a better potential (Gale & co workers)




**Pavese, Catti, Parker, Wall, *Phys Chem. Miner.* , 1996**

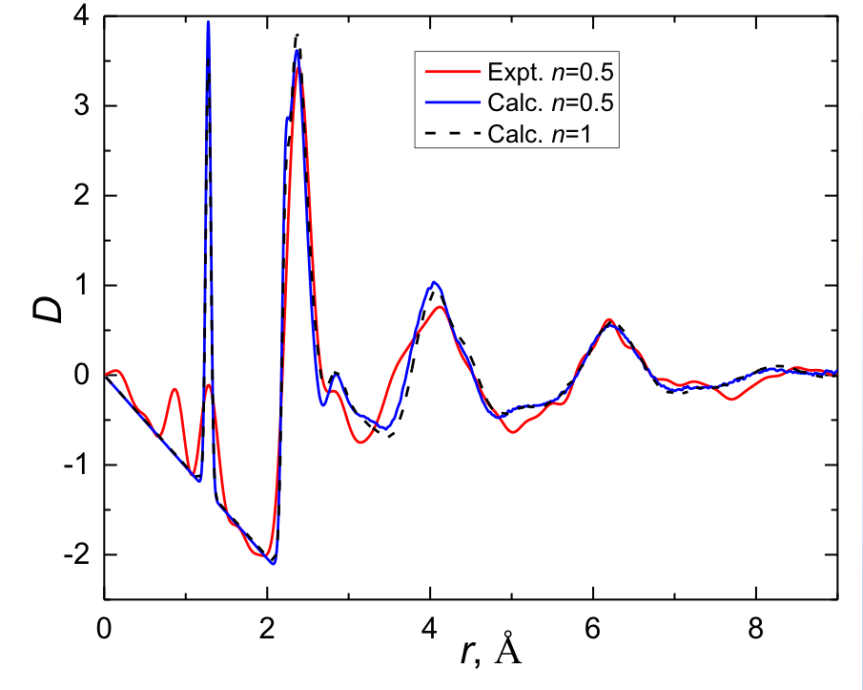
**Raiteri, Gale, Quigley, Rodger, *J. Phys. Chem. C*, 2010**



# Structure of hACC

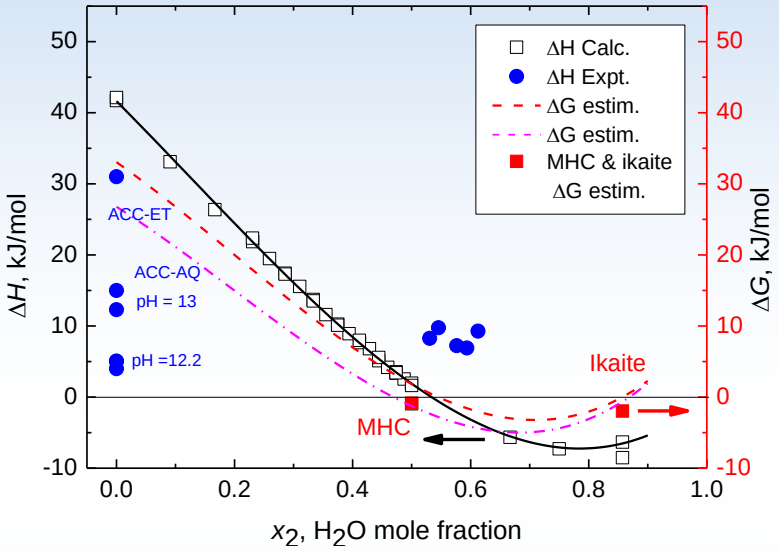
- create with several protocols
  - ➢ melt monohydrocalcite; melt and dehydrate ikaite; anneal random distribution …

- Good agreement with expt structure factor

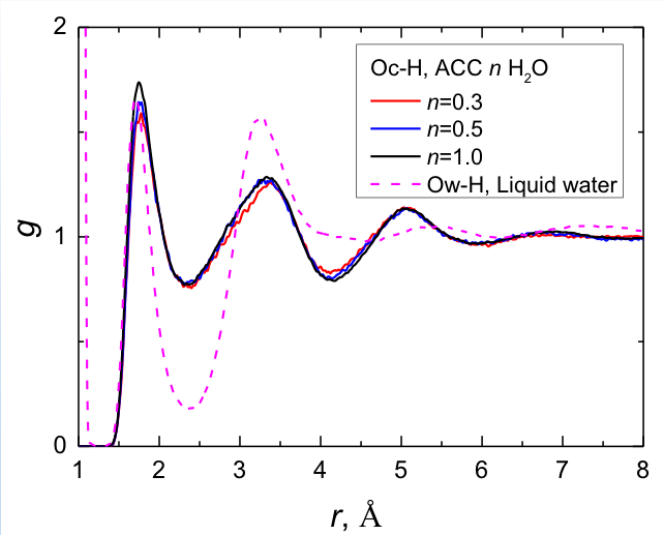


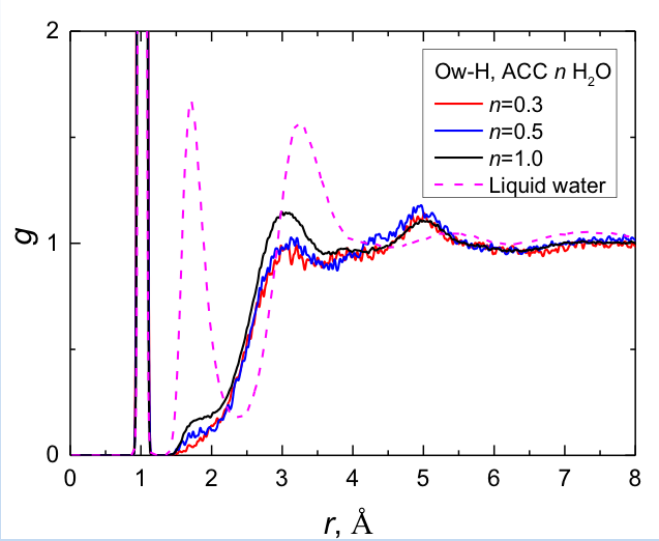
# Water in hACC

- zones of stability for hACC



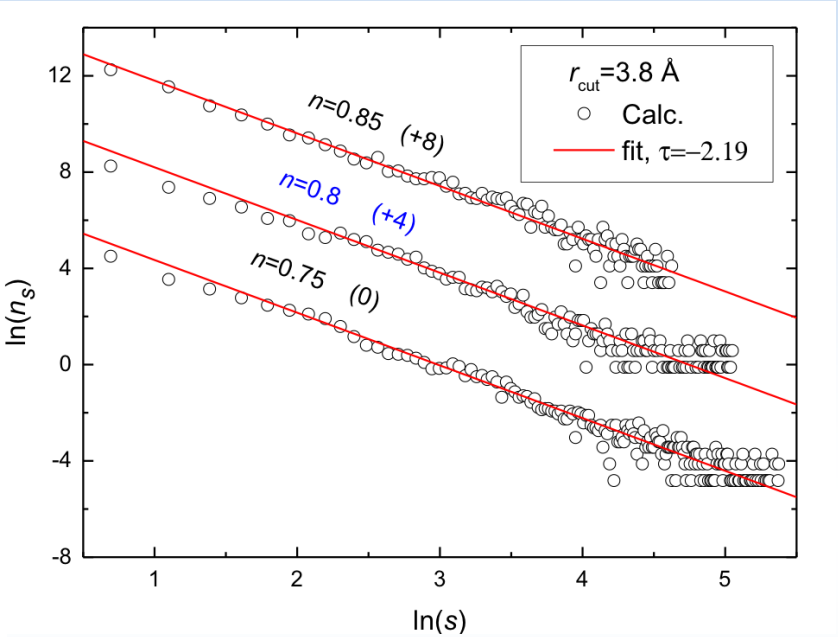

- carbonate-mediated H-bond networks

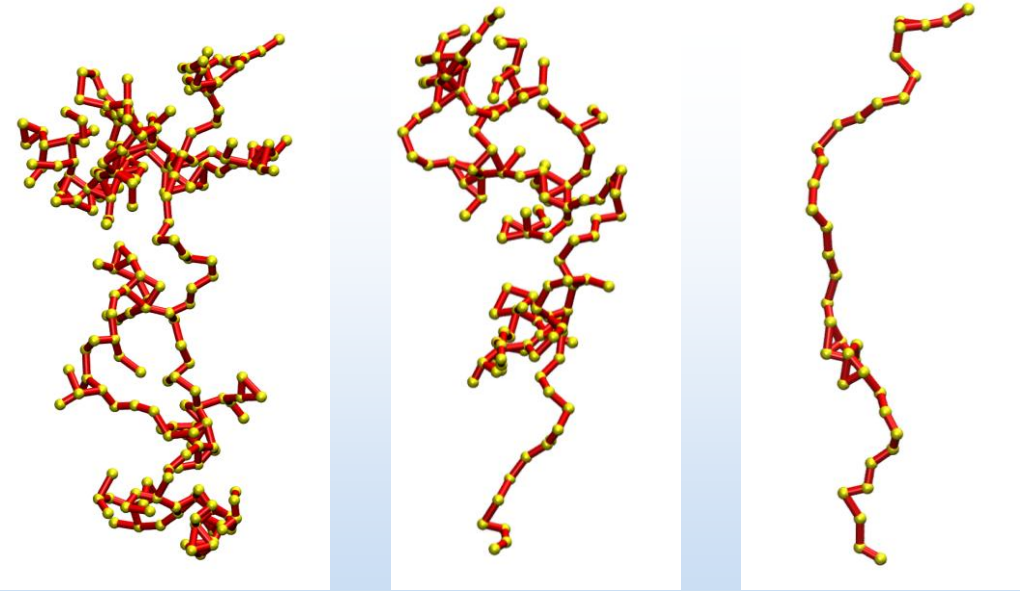


# Percolating water clusters

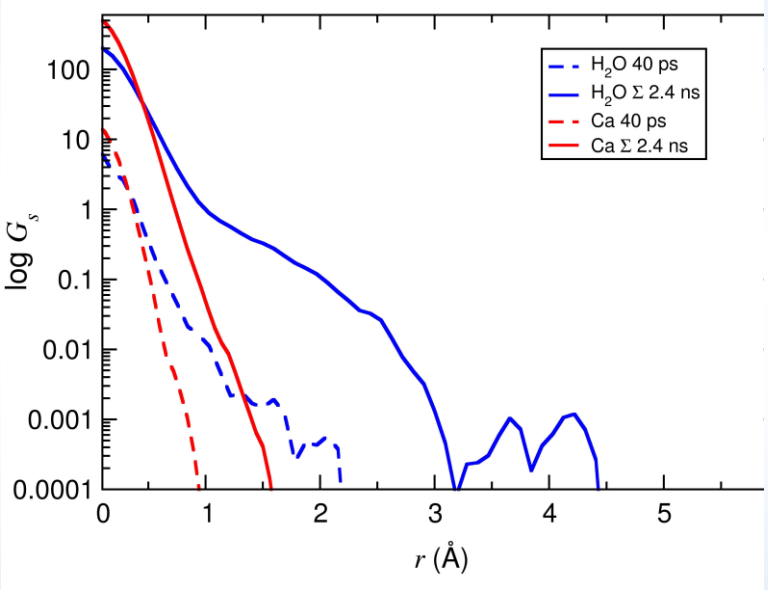
- Universal scaling behaviour
- percolation threshold ~ $n = 0.9$
- underlying hexagonal lattice

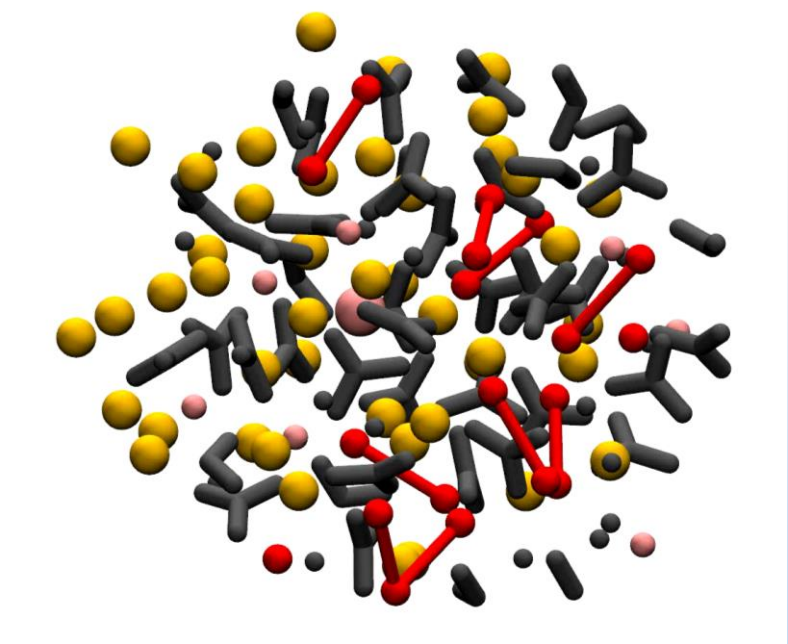


# correlated water dynamics

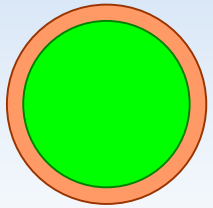
- correlated large hops along chains of water



# Classical nucleation

- Activated process:
  - Favourable "bulk" energy
  - Unfavourable interfacial energy
  - Critical "cluster" size

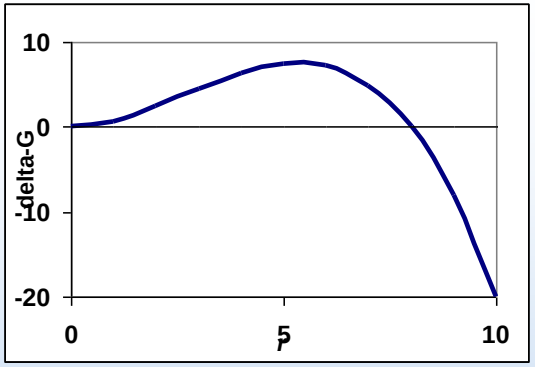


**Favourable**
**Unfavourable**

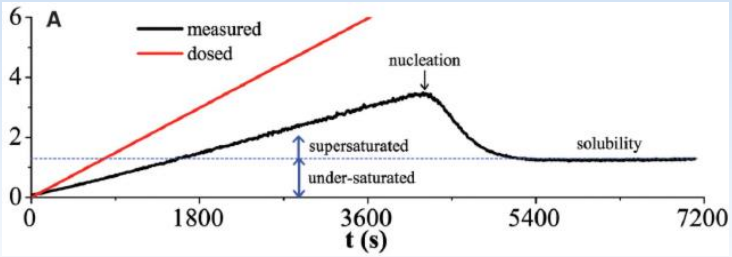
- Classical Nucleation Theory

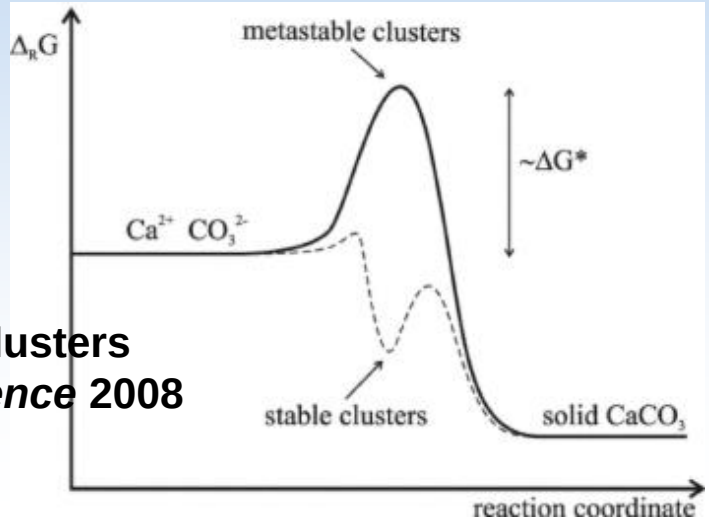
$$\Delta G = \frac{4}{3}\pi\rho\Delta\mu r^3 + 4\pi\gamma r^2$$

- Fundamentally Stochastic

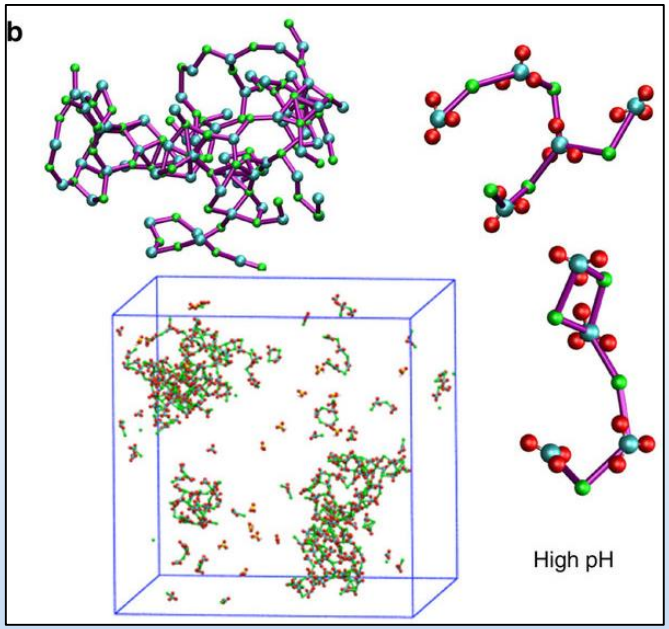


# hACC nucleation: classical?

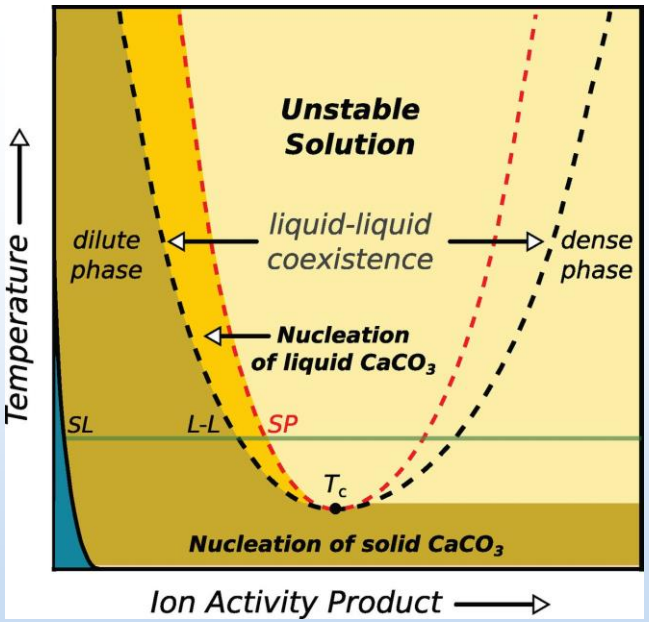



**pre-nucleation clusters**
**Gebauer *et al.*, *Science* 2008**




**DOLLOP, Raiteri *et al.*,**
***Nature Comms, 2011***




**dense liquid phase**
**Wallace *et al.*, *Science, 2014***

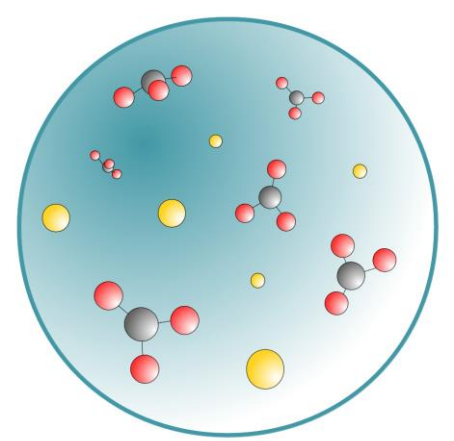
# Nature of the clusters?

- spontaneous clusters dominate?

- Random Structure searches
  - random arrangement of ions; minimise; repeat

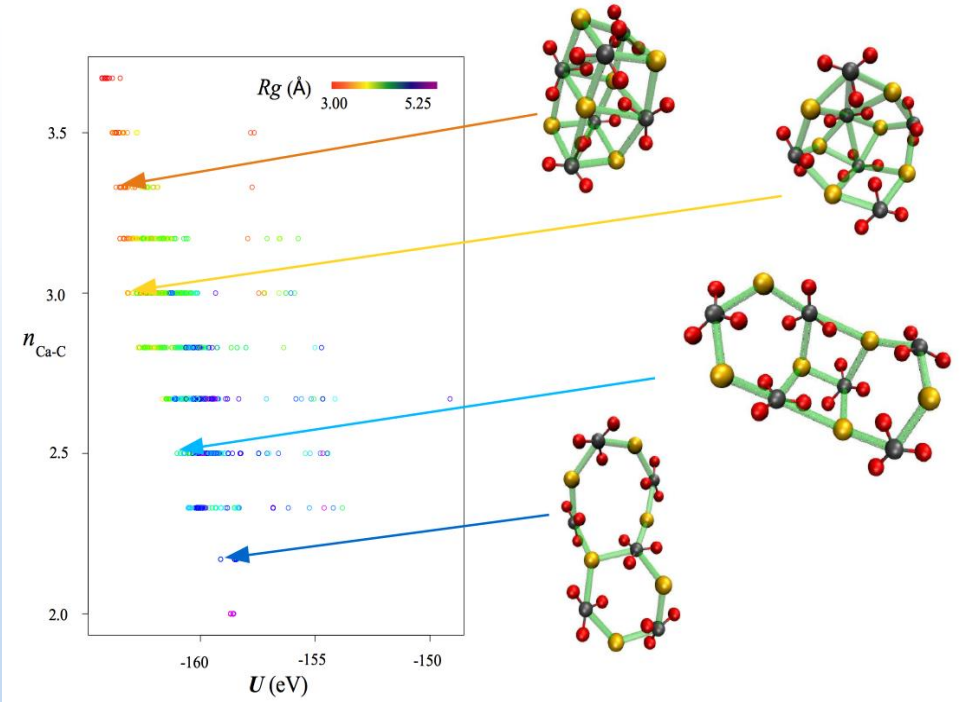
- Intelligent Water Drop algorithm
  - nature-inspired global optimsation
  - erosion of soil to define river valleys
  - applications to binary LJ mixtures and Janus particles
    - (not yet carbonates)

# Random Structure Searches

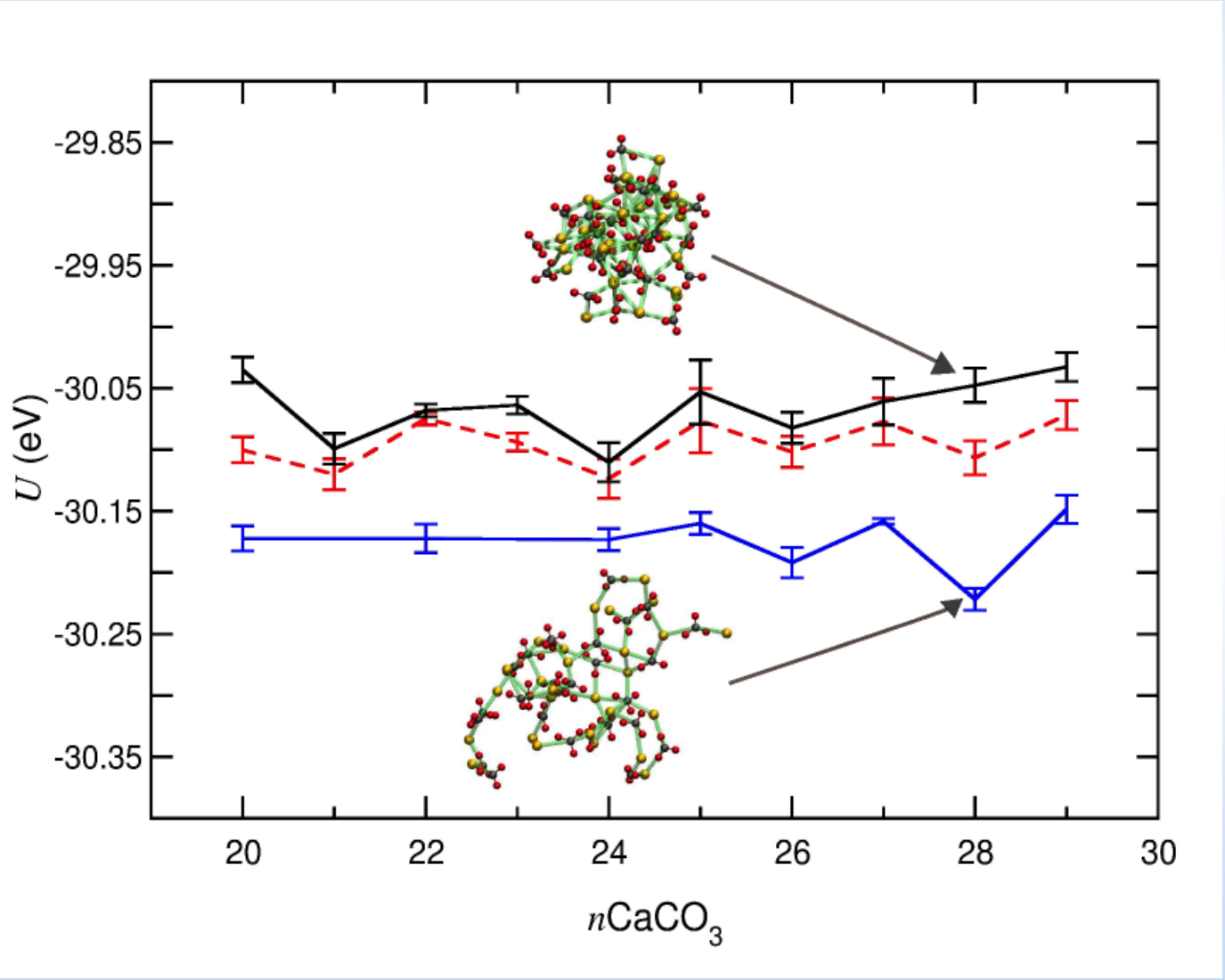


- random arrangement of ions within sphere
- Conjugate Gradient optimisation
- ~10,000 initial structures for each cluster size (1–40 formula units)

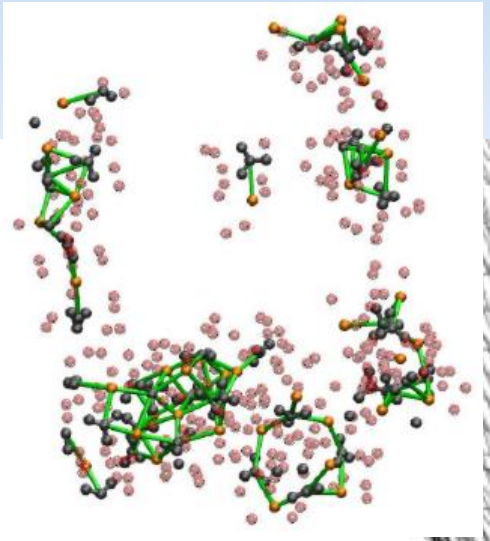
- variety of cluster properties
  - e.g. energy vs radius of gyration

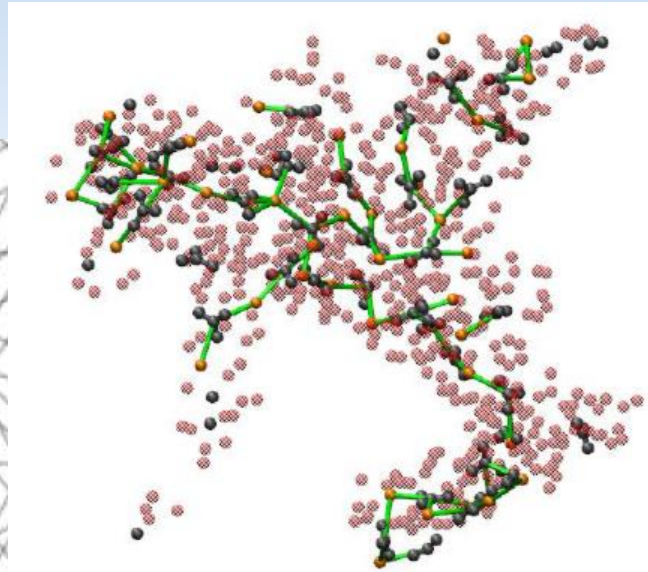


# Energetics of solvated clusters



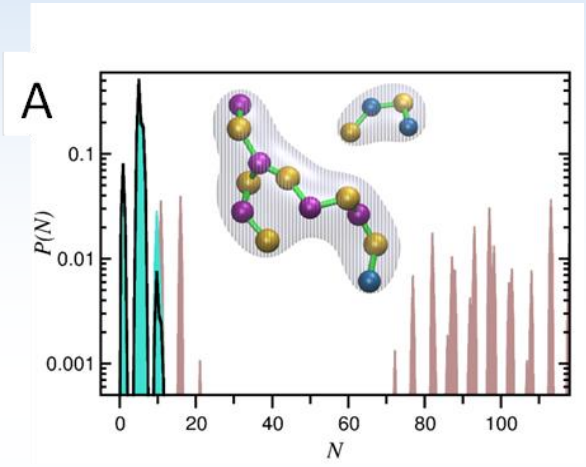
# Free energy Landscape

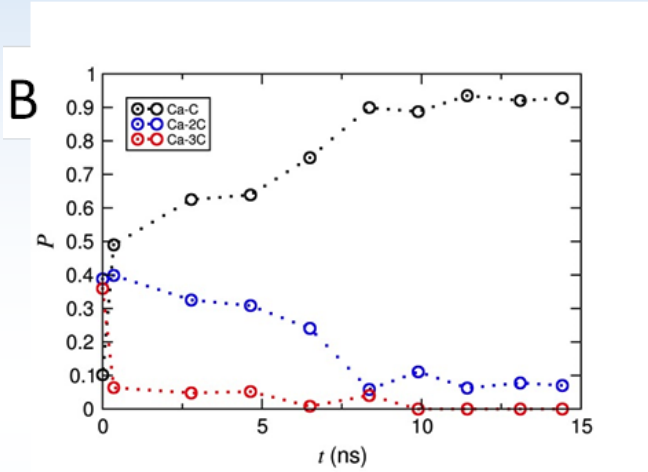




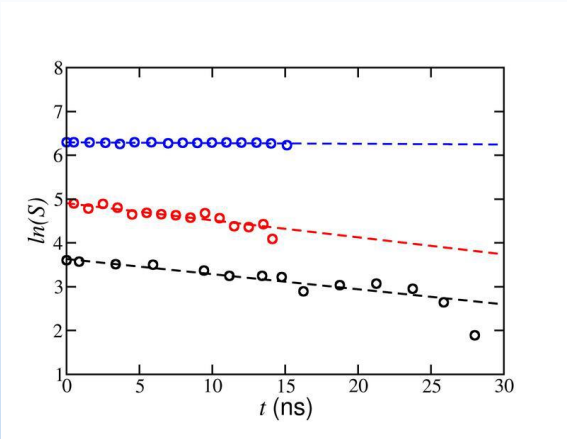
# Long timescale MD: 20–50 mM

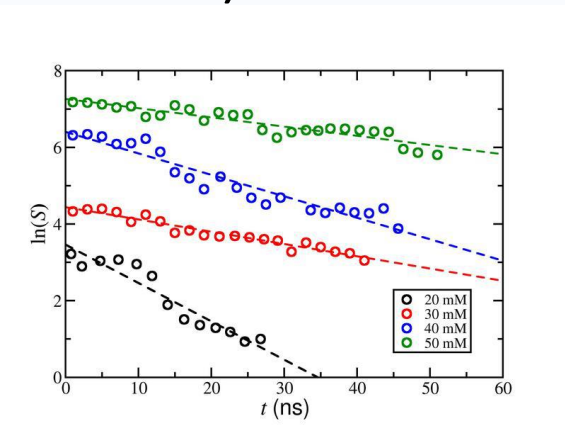
- “slow” dissolution: convergence from dispersion or initial clusters




- dissolution slows with cluster size (concentration)

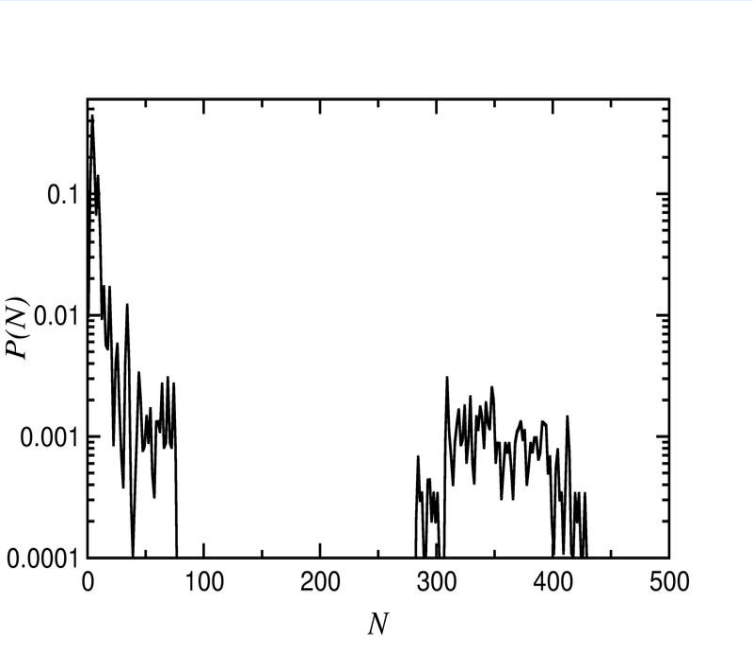


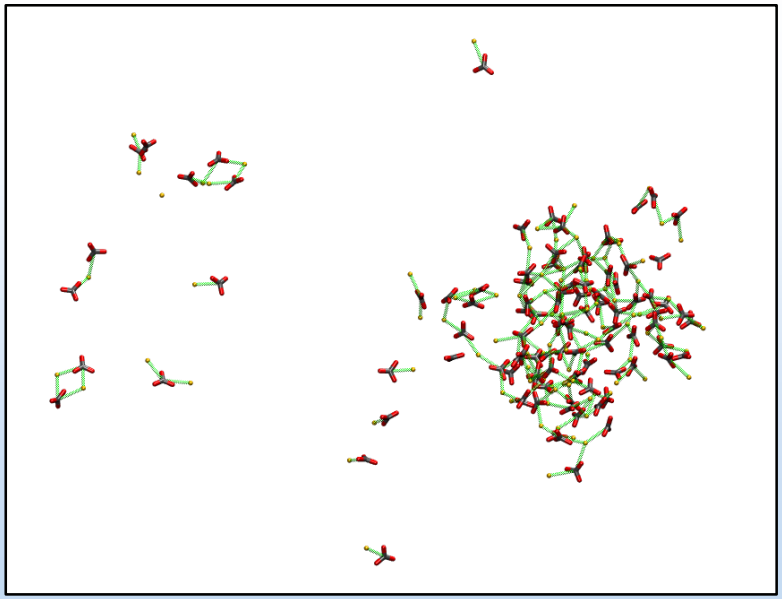




- (requires free energy barrier)

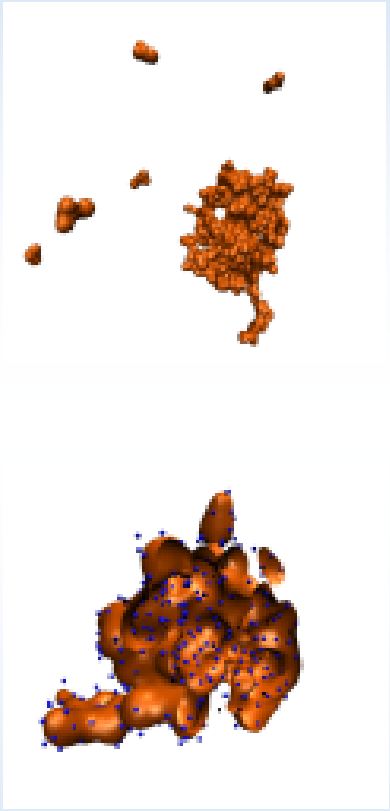
# Long timescale MD: *ca*. 0.5 M

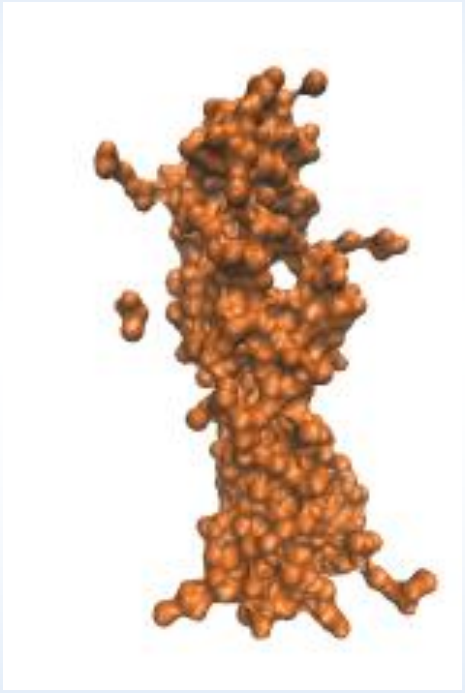
- stable “DOLLOP” behaviour; large clusters

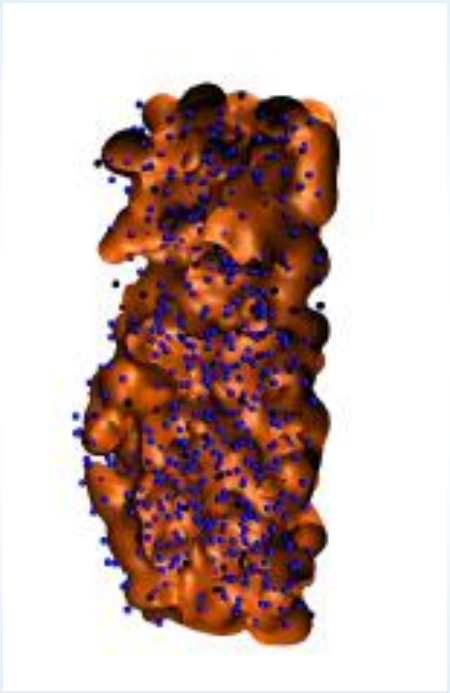


# And higher concentrations: 0.5–1.5 M

- 





**0.57 M**
**stable clusters**
**DOLLOP**

**1.1 M**
**infinite, liquid like**
**(phase transition?)**

# Conclusions

- Source of new mechanistic insight
  - kinetic inhibition; eggshell proteins
  - does your predictive model have the right features?
- Can capture some elements of intrinsic complexity
  - inhibitor polydispersity; nanoparticle structure
  - need for intelligent sampling
- Progress through data dimensionality reduction
  - excellent methods for a handful of dimensions
  - need to identify key dimensions
  - need to recognise and avoid pitfalls of too few (or wrong) dimensions