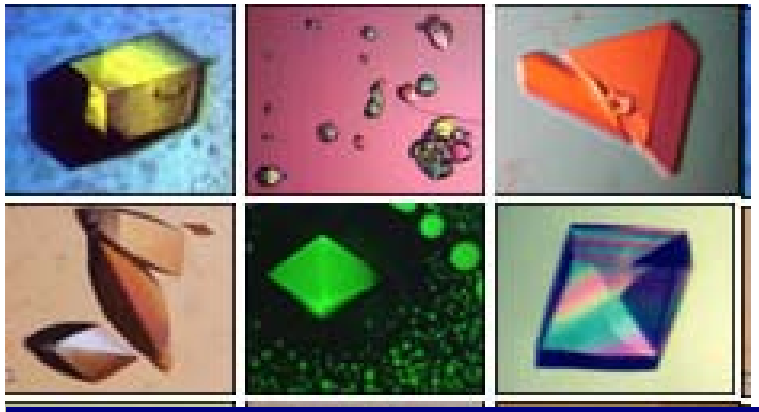
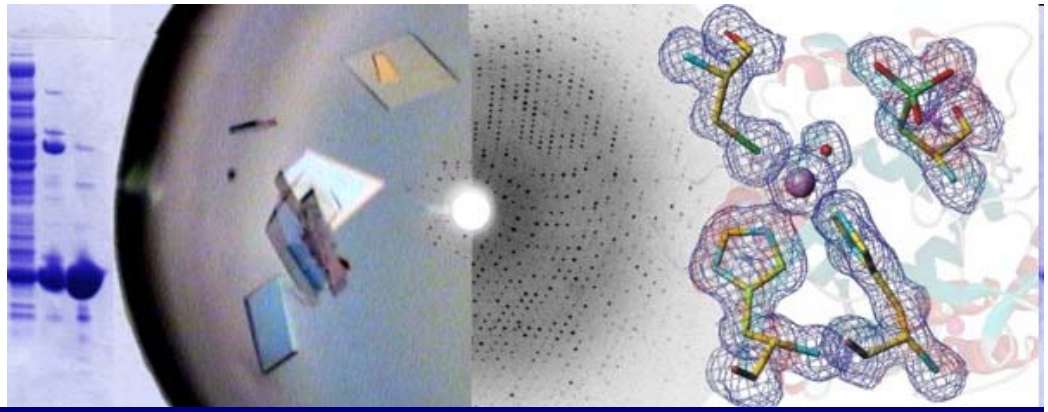


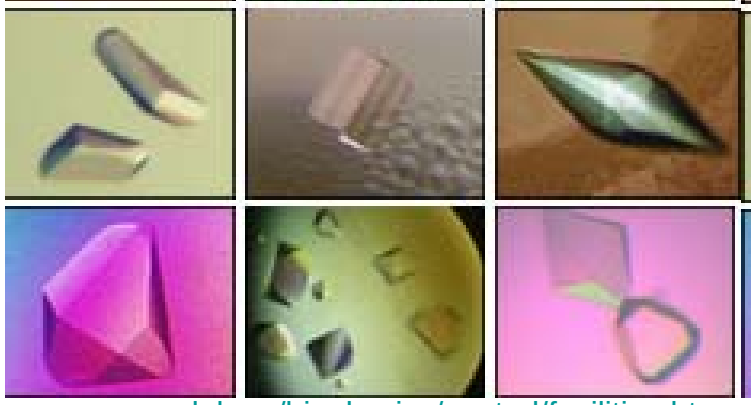


Protein Crystallization: Some Physical Insights

Fajun Zhang

Institut für Angewandte Physik

Universität Tübingen



www.uoguelph.ca/biophysics/central/facilities.htm

www.sheffield.ac.uk/mbb/research

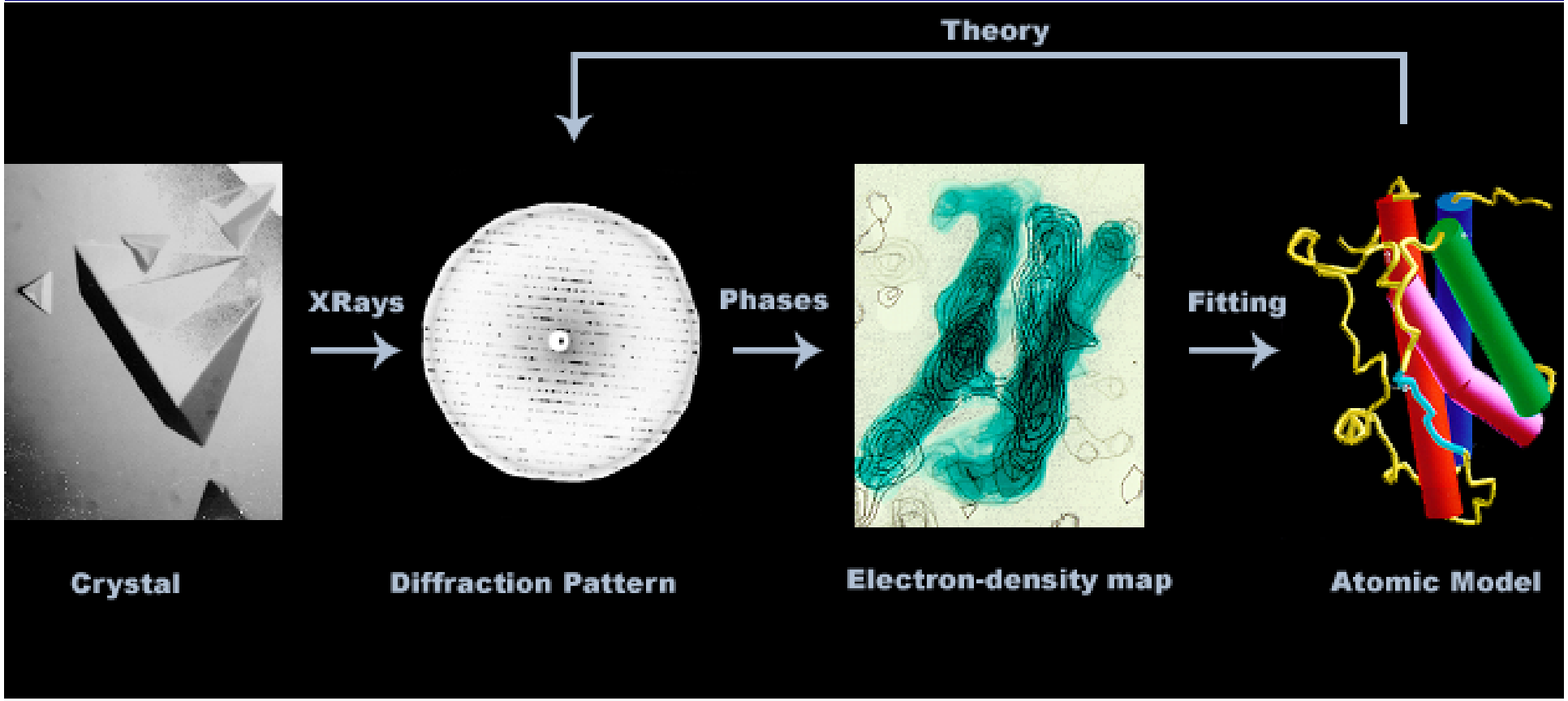
- Introduction
  - Why protein crystals?
  - Why protein crystallization (PC) difficult?
- Physical Chemistry of PC
  - Interactions in protein solutions;
  - Phase diagrams;
  - Crystal growth mechanism;
- Summary

References & Scope

1. R. F. Service, *Science* 2002, 298, 948
2. S. D. Durbin, G. Feher, *Annu. Rev. Phys. Chem*. 1996, 47, 171
3. V. J. Anderson, H. N. W. Lekkerkerker, *Nature* 2002, 416, 811
4. A. George, W. W. Wilson, *Acta Crystal*. 1994, D50, 361
5. D. Leckband, J. Israelachvili, Intermolecular forces in biology. *Q. Rev. Biophys*., 2001,34, 105-267
6. P. R. Ten Wolde, D. Frenkel, *Science* 1997, 277, 1975
7. D. Erdemir, A. Y. Lee, A. S. Myerson, *Acc. Chem. Research* 2009, 42, 621
8. F. Zhang, G. Zocher, A. Sauter, T. Sthele, F. Schreiber, J. Appl. Cryst. 2011,
9. P. G. Vekilov, Soft Matter, 2010, 6, 5254-5272

1. Crystallization of membrane proteins
2. Protein Crystallography
3. Crystallization techniques/methods (not in detail)

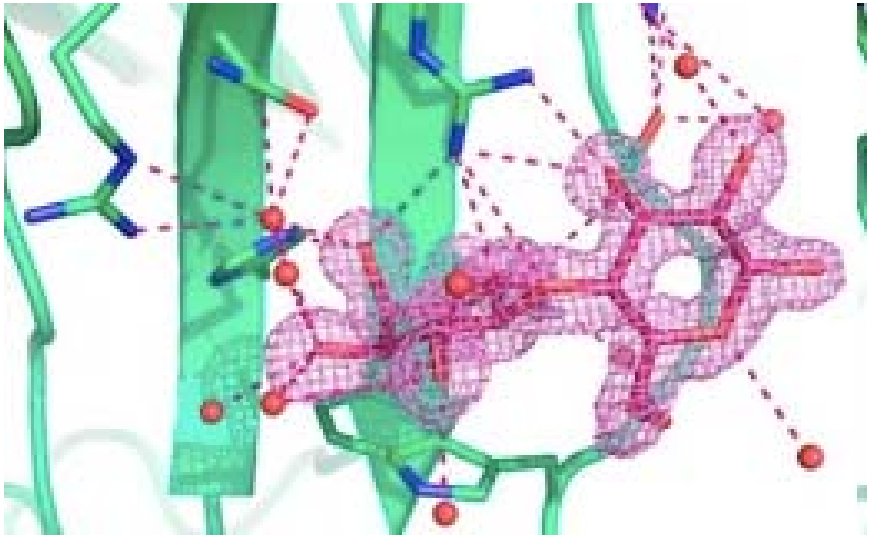
Why protein crystals?

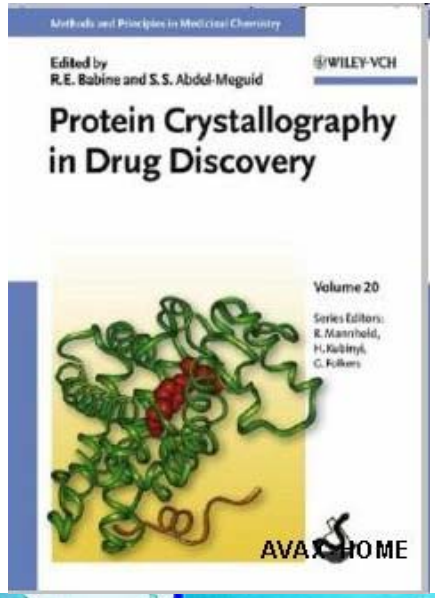


- Structural biology: biofunction, crystallography
- Evolution: scanning large numbers of structures, gain insights into how families of similar proteins evolved

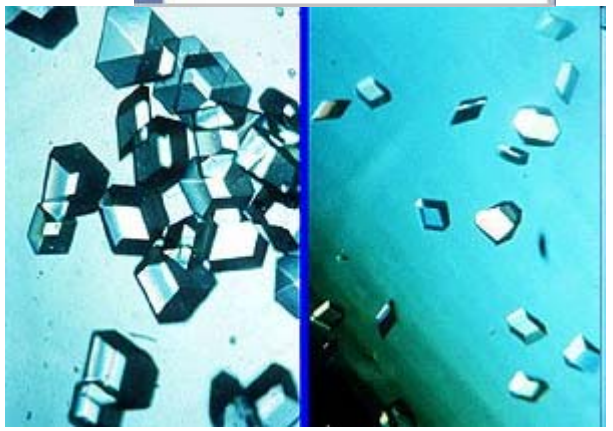
http://www.proxychem.com/macromolecular_crystallography.html

Why protein crystals?



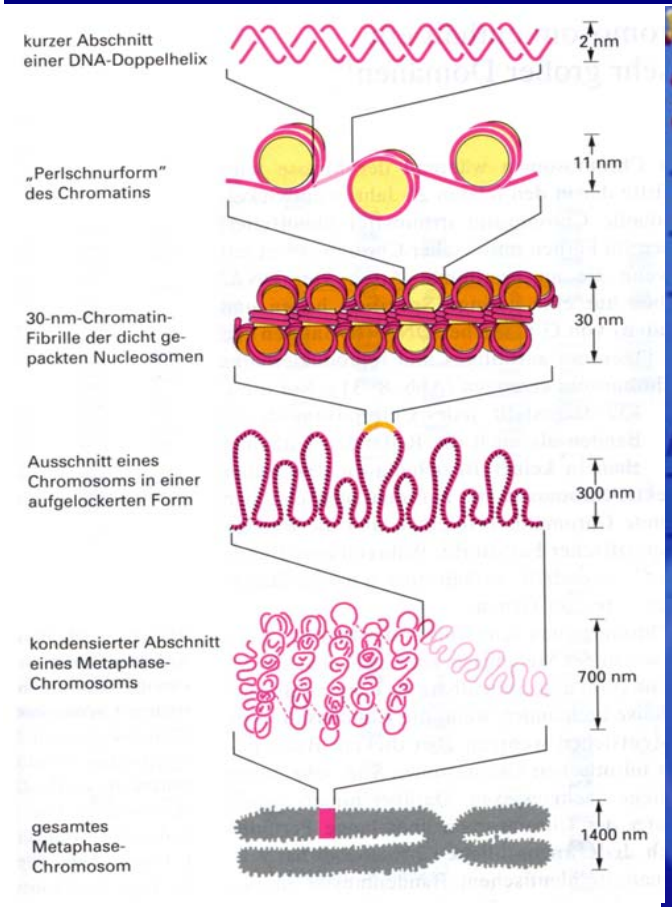


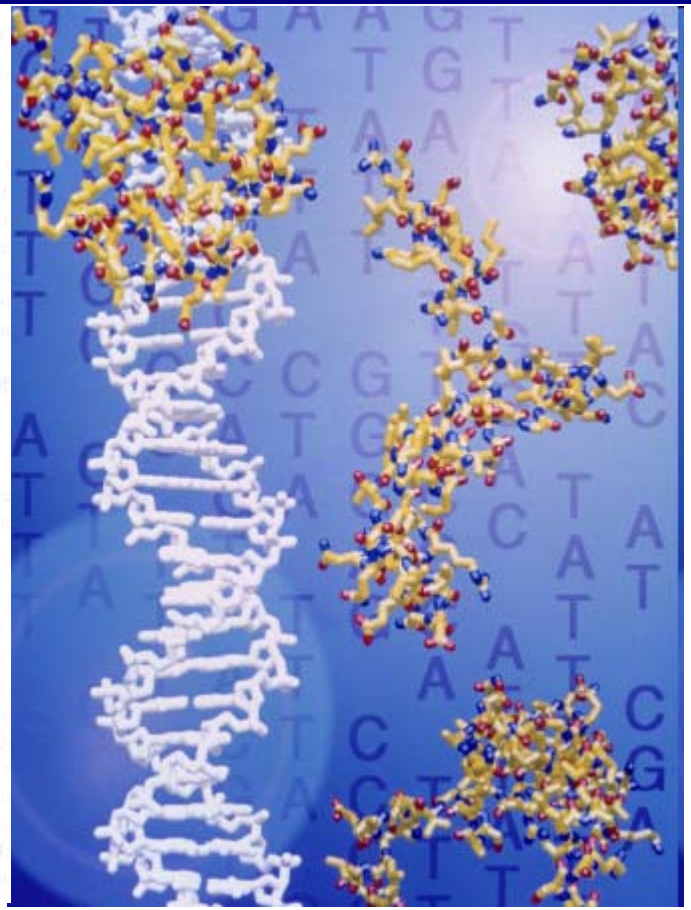
- Drug designers: help tailor pharmaceuticals to block or enhance a protein's chemical activity
- Drug release control: insulin crystal size and distribution



Insulin crystals from SpaceRef.com

Why protein crystals – Human Genome Project

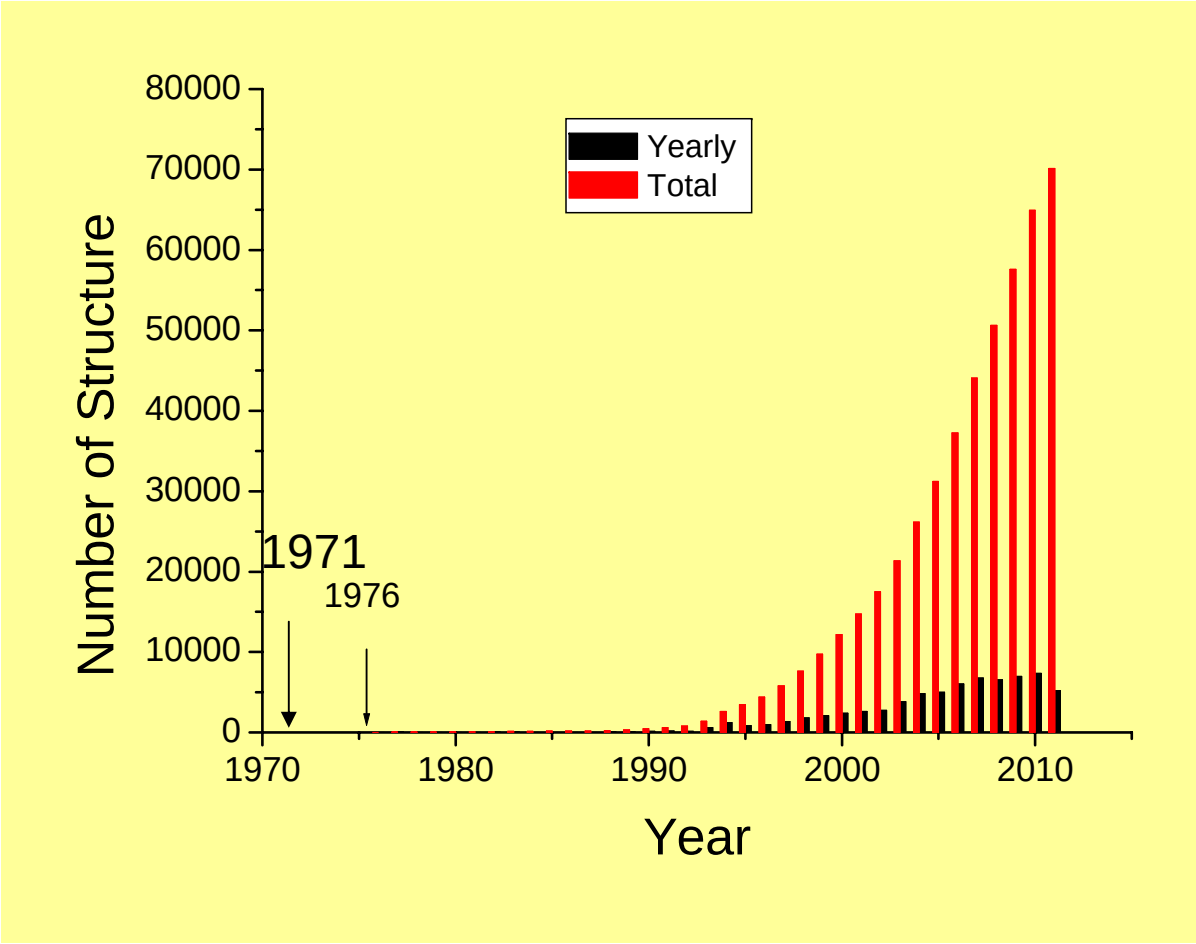




- Human Genome Project: Structral genomics: high-throughput
- Tradition: picking targets because they are biologically interesting.
- Scan genome databases for stretches of DNA encoding genes of completely unknown function, hunt down their proteins, study the results, and discover entirely new realms of biology in the process.

NIGMS: National Institute of General Medical Sciences,
$50 million/year
Eight Institutions for 5 years
1997-2002

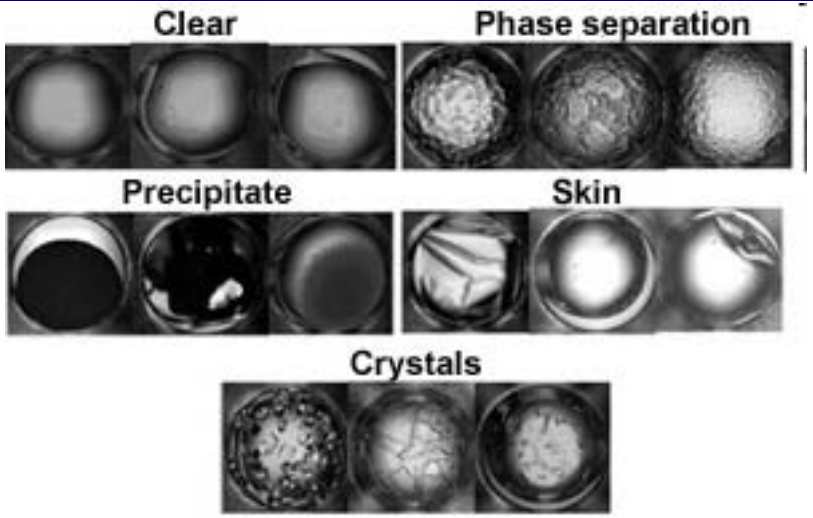
Protein Data Bank



Protein sequences so far: 2,867,124 only **2.64%** structures have been solved. It contributes to **0.23% / year**

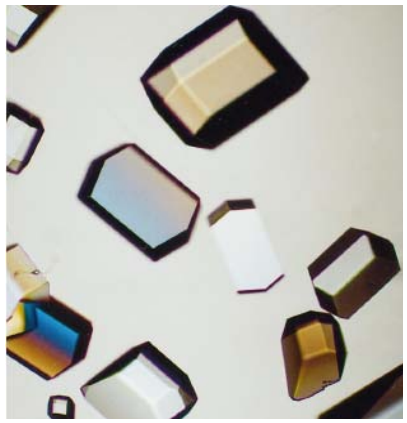
- **Sep 06, 2011 at 5 PM PDT there are 75694 Structures at www.pdb.org**
- **6573 structures / year in the recent 5 years**

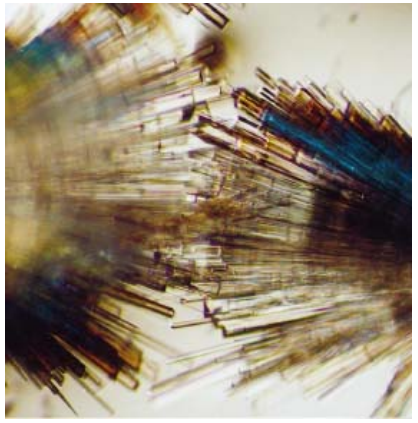
How do protein crystals grow?



Protein crystallization is:

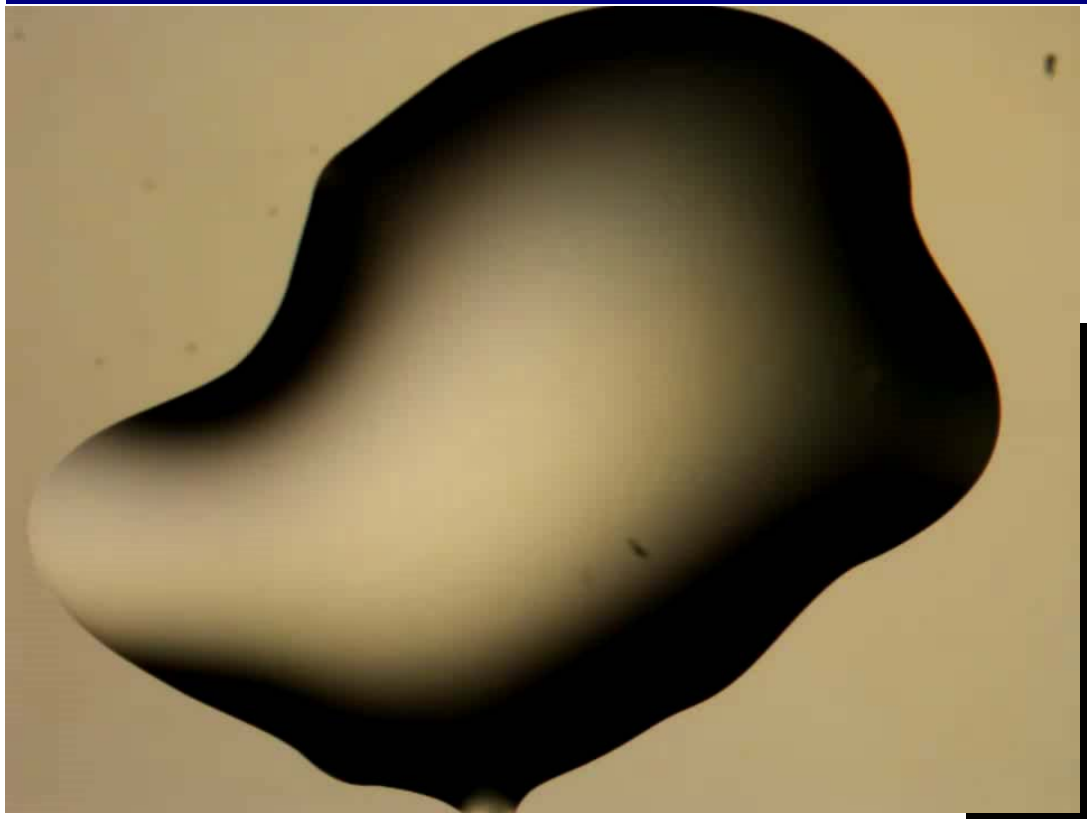
- un-predictable, trial-and-error method;
- bottle-neck of structural biology;
- less understanding of the interactions in the protein solutions.



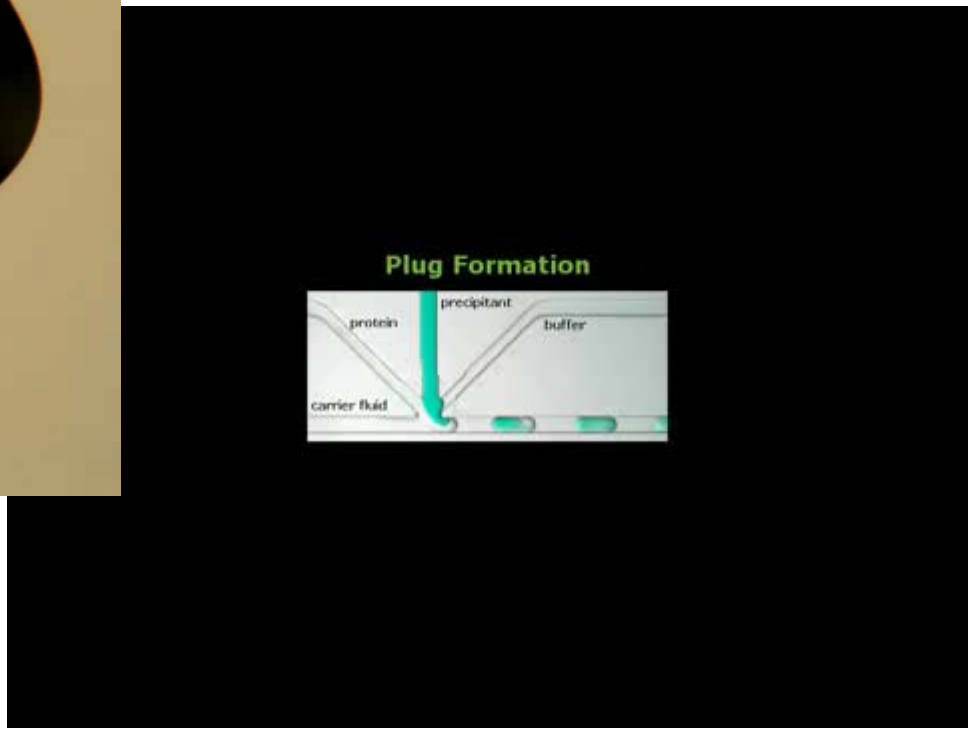


Velev, et al. *Biophys. J.* 1998, 75, 2682
http://www.chtsb.org/

How do protein crystals grow?

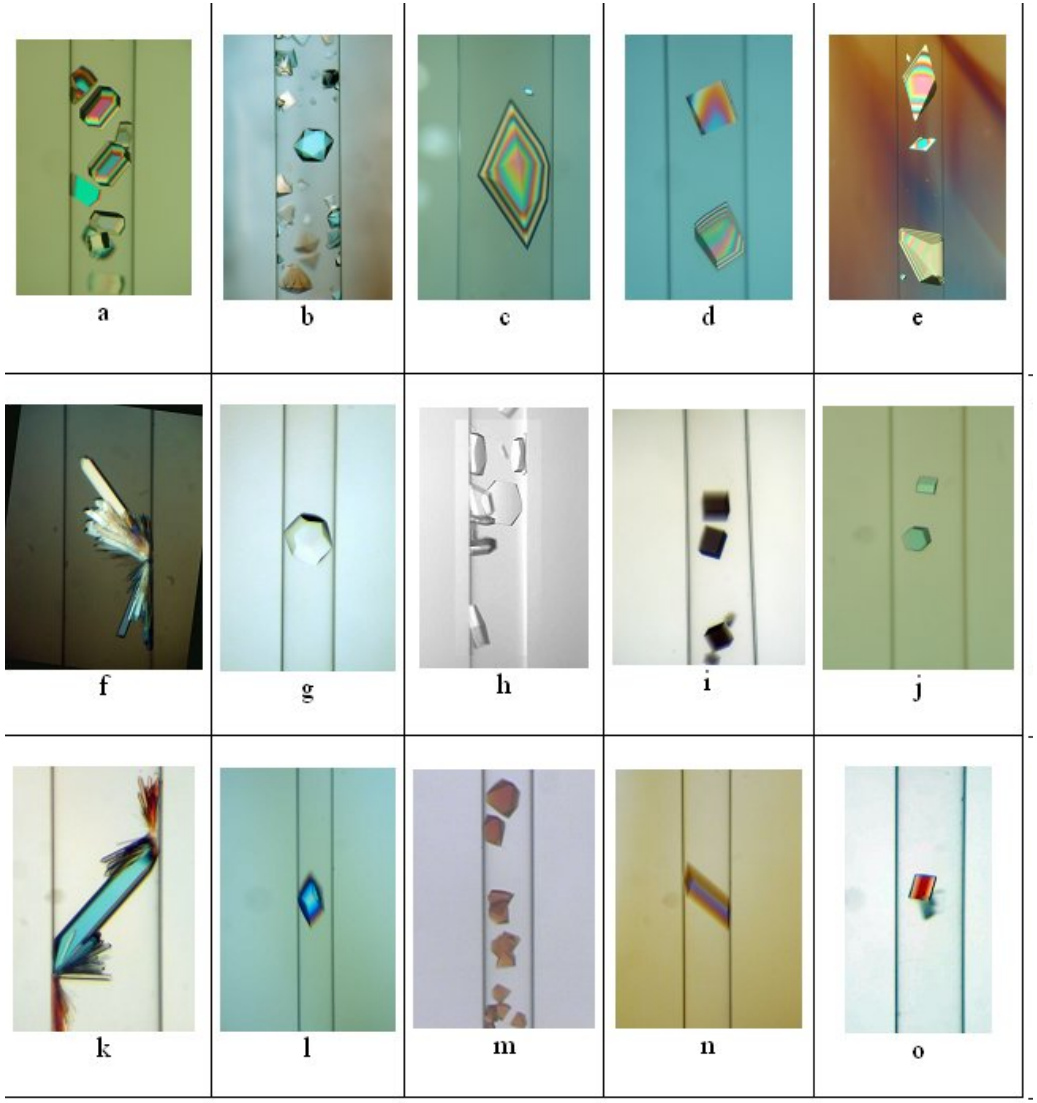


Lysozyme crystal growth



L. Liang, R. F. Ismagilov, Annu. Rev. Biophys. 2010, 39, 139

Examples of Protein Crystal



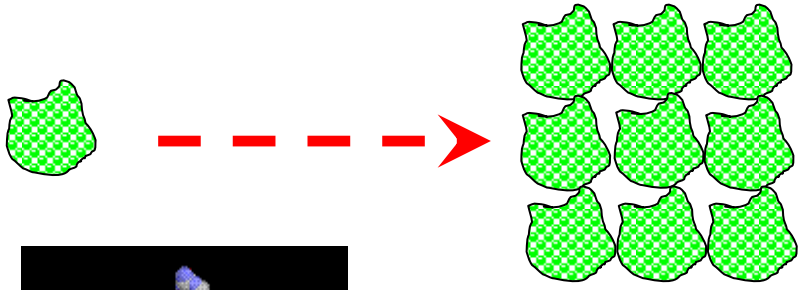
a) Lysozyme
b) Dehydroquinase
c) e) Thaumatin
d) Concanavalin A
f,k,n) Antilysozyme
g) Insulin
h) Lumazine synthase
i) catalase
j,o) Factor XII
l) Saicar synthase
m) Ferritin

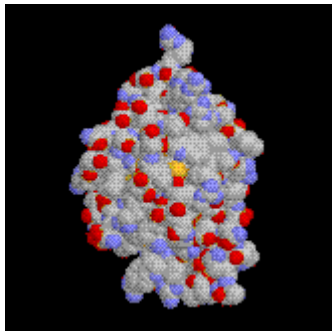
Granada Crystallisation Facility (GCF)

Why protein crystallization difficult?

Atom or small molecules

Globular proteins



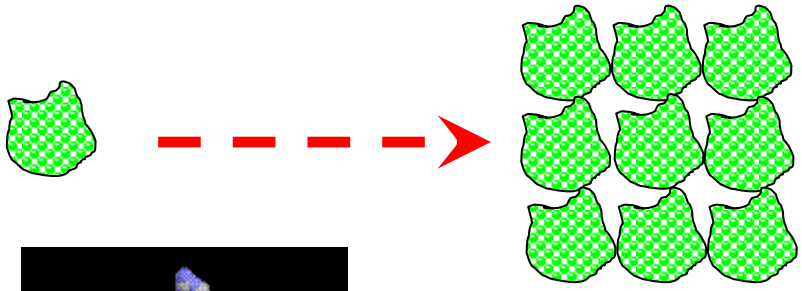


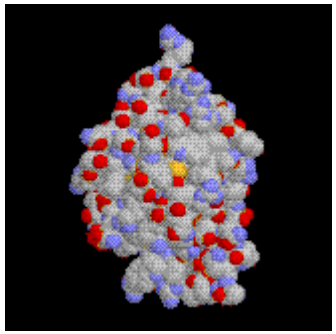
- Size – adhesion energy (E_{ad}) linearly increase with size, but bonding life times, dynamics, relaxations and equilibrion times as $\exp(E_{ad}/k_BT)$; this leads to energy landscape;
- Shape – contacts; the number of contacts limited the ability of formaing 3D structure,
  - reducing coherent energy, entropy;
  - packing density, solvent mediated contact
- ...

Why protein crystallization difficult?

Atom or small molecules

Globular proteins





- Rotation – low symmetry low entropy;
- Surface mobile side chains;
- Solvent involved interaction and structure formation...
- Negative design by Nature
- Unique: proteins do not share universal behavior (structure, geometry, interactions, etc.)

Interactions: orientation-dependent, short & long range, specific & non-specific;

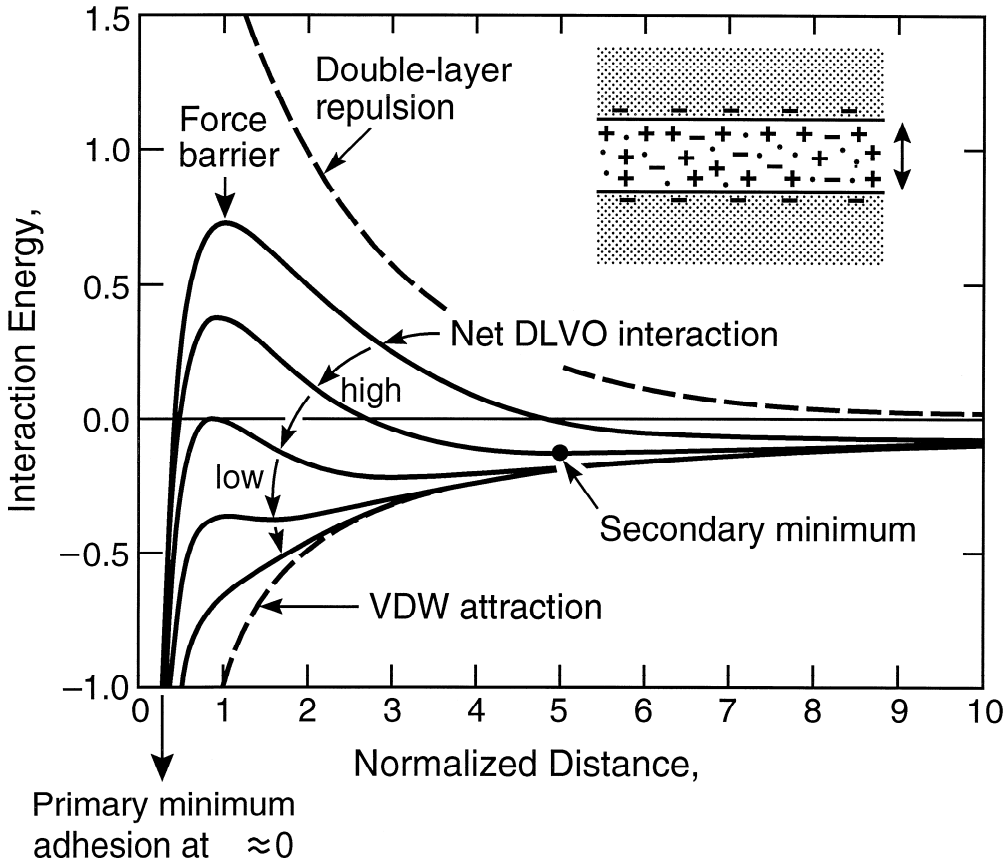
JPK, Doye, et al. Phys. Biol. 2004, 1, 9

Physical Chemistry of PC

- Interactions in protein solutions
  - Interactions for protein crystallization;
  - Predictor for protein crystallization (PC);
  - Second virial coefficient, A_2 & B_2;
- Phase diagram of protein solutions
  - New features of „big atom“
- Nucleation mechanism
  - Classical nucleation theory
  - Two-step nucleation procedure

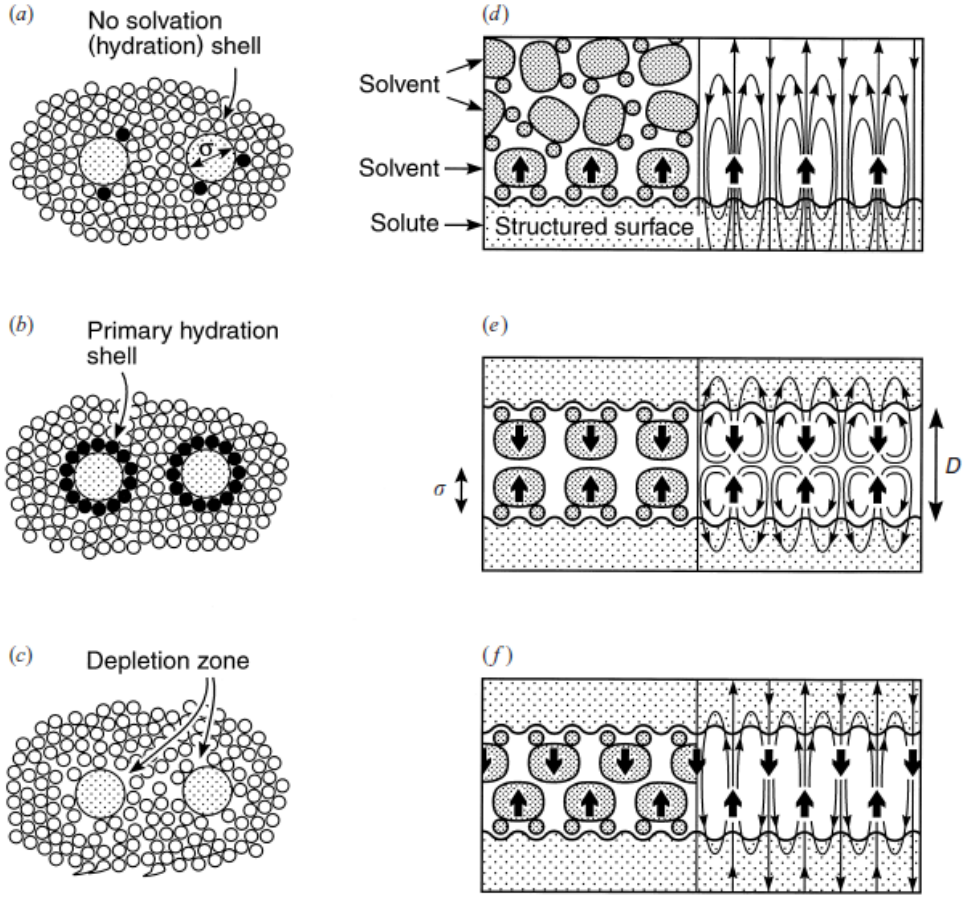
Interactions in biological system

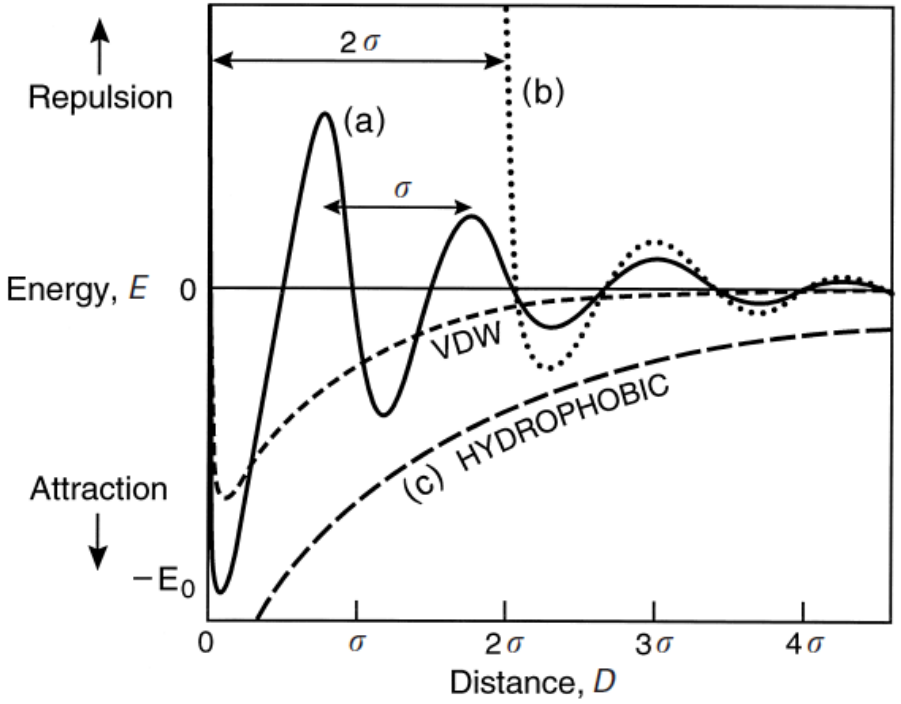
- Van der Waals interaction;
- Electrostatic interactions, double-layer;
- Break down of DLVO theory at small distances;
- Solvation, hydration;
- Hydrophobic interaction;
- Depletion effect
- Specific interactions
- Static & dynamic



D. Leckband, J. Israelachvili, Q. Rev. Biophys., 2001,34, 105-267
Lecture by Prof. Frank Schreiber

Interactions in biological system



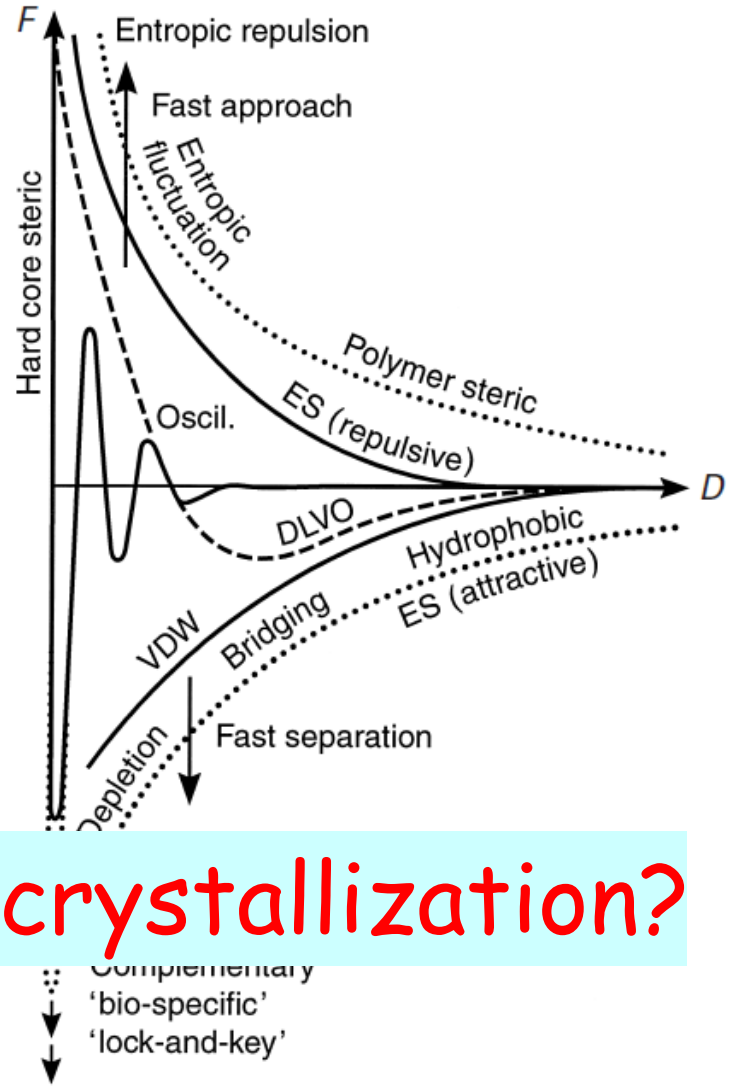


- Solvation, hydration;
- Hydrophobic interaction;
- Depletion effect

D. Leckband, J. Israelachvili, Q. Rev. Biophys., 2001,34, 105-267
Lecture by Prof. Frank Schreiber

Interactions in biological system

- Steric, bridging and Depletion
- Thermal fluctuation forces (repulsive)
- Specific interactions



How to predict protein crystallization?

Predictors for Protein Crystallization

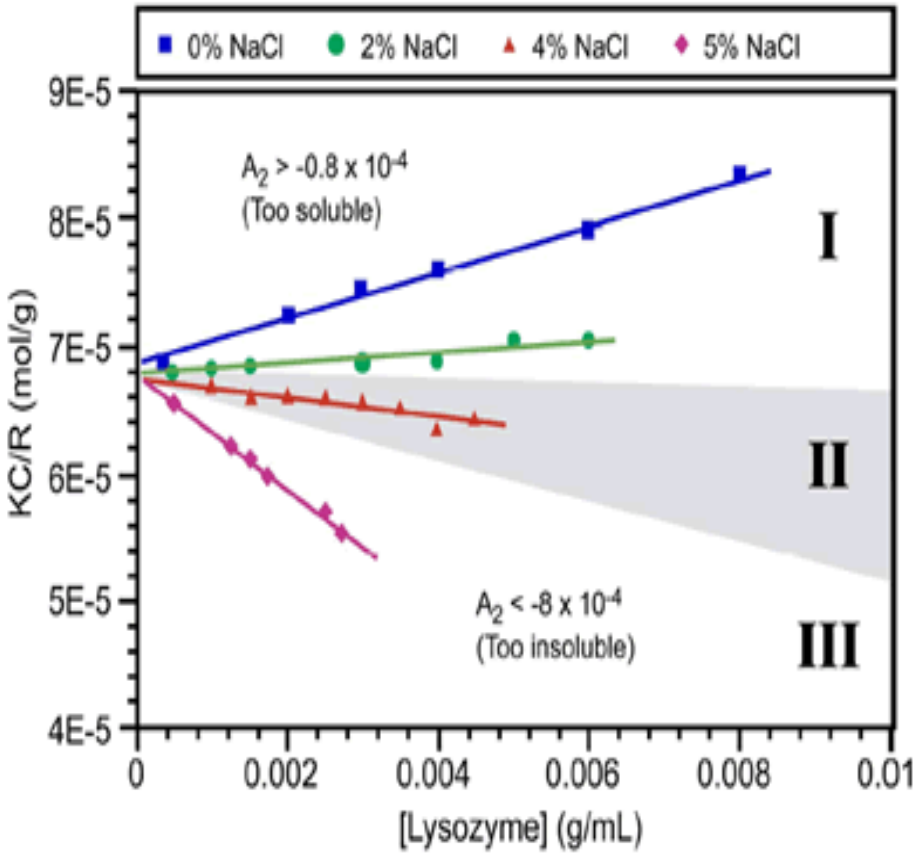
a) Test protein solutions for the probobality of crystallization;
b) Reduce the time and number of screening experiments;
c) Universal predictor for fine tuning existing conditions

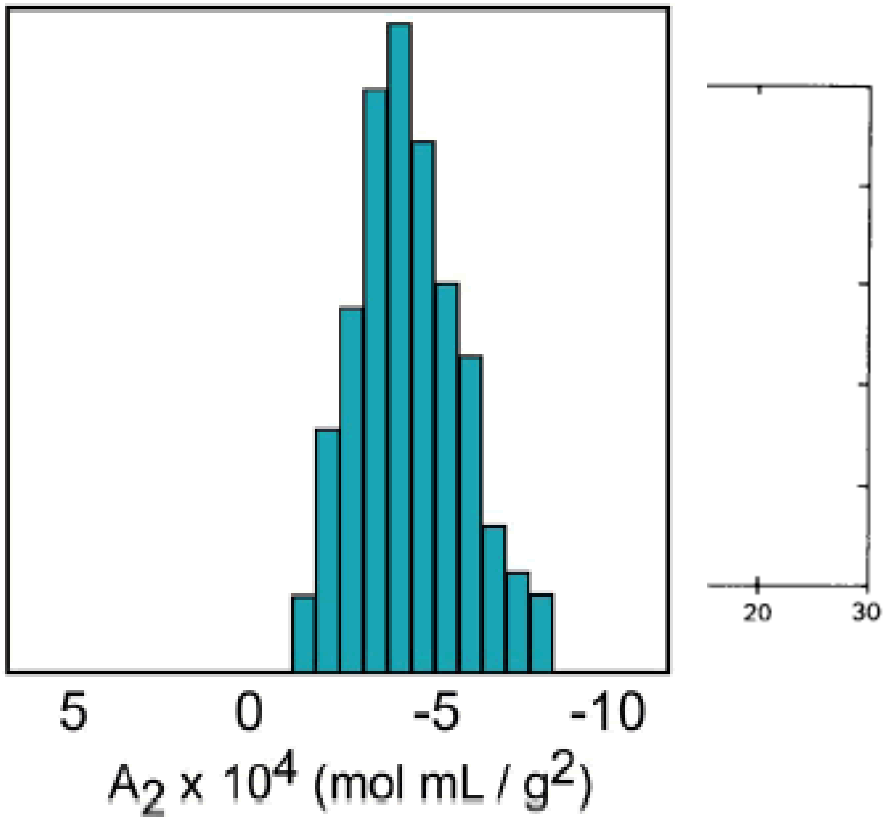
Requirement for the predictors:

a) Discriminating response;
b) Require minimum amount of protein and non-destructive;
c) Accommodating any solvent condition that may be required for protein crystallization;
d) Non-invasive: prevent protein interaction with foreign bodies;
e) Can be routinely performed by laboratory technicians;
f) ...

A. George, W. W. Wilson, Acta Crystal. 1994, D50, 361

Predictor of Protein Crystallization: A_2



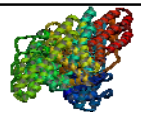


A. George, W. W. Wilson, Acta Crystal. 1994, D50, 361

Second Virial Coefficient, A_2

TABLE V. The second virial coefficient for a number of proteins, data taken from Ref. 43.

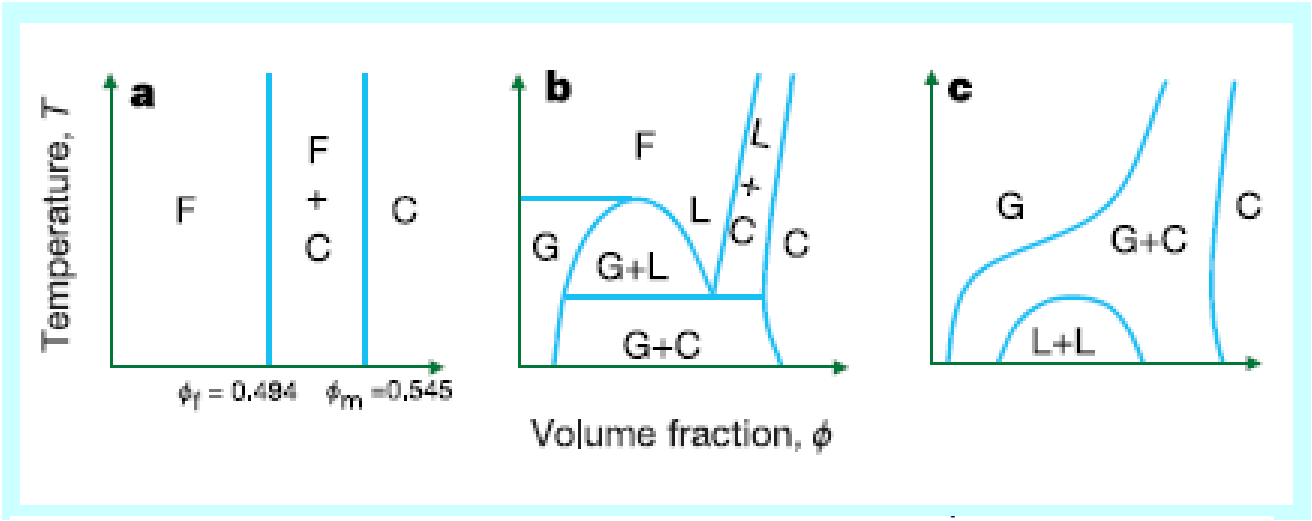
Protein	Mass (kD)	B_2 (10^{-4}mol cm^3 g^{-2})	B_2/v_0
BSA	65	−2.1	−18.6
Canavalin	141	−0.8	−15.3
Concanavalin A	102	−2.5	−34.7
Concanavalin A	102	−1.9	−26.4
α-chymotrypsin	25	−8.4	−28.6
α-lactalbumin	14	−7.3	−13.9
β-lactoglobulin A	36	−2.4	−11.8
β-lactoglobulin A	36	−6.2	−30.4
β-lactoglobulin B	36	−2.8	−13.7
β-lactoglobulin B	36	−6.2	−30.4
Lysozyme	14	−2.8	−5.3
Ovostatin	720	−7.1	−695
Ovalbumin	43	−6.1	−35.7
Pepsin	36	−7.8	−38.2
Pepsin	36	−2.8	−13.7
Pepsin	36	−0.8	−3.9
Ribonuclease A	14	−4.1	−7.8
STMV	1500	−1.8	−367
Thaumatin	22	−3.0	−9.0

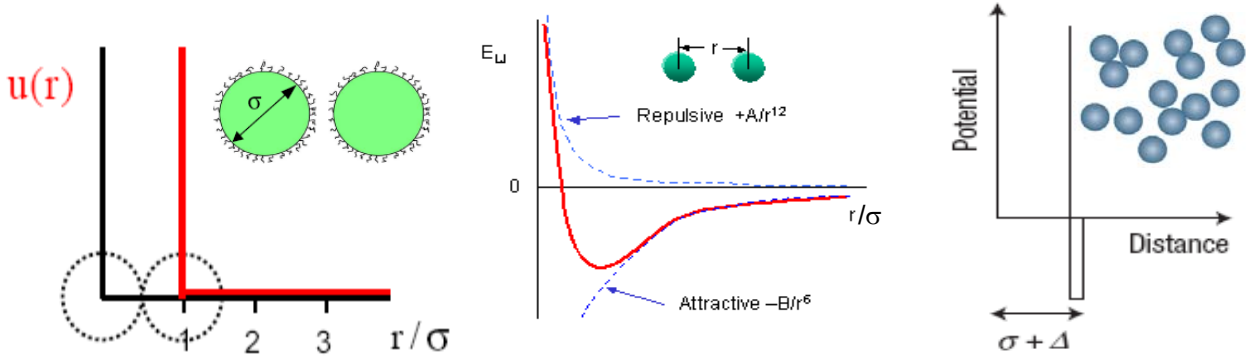


Physical Chemistry of PC

- Interactions in protein solutions
  - Interactions for protein crystallization;
  - Predictor for protein crystallization;
  - Second virial coefficient, A_2 & B_2;
- Phase diagram of protein solutions
  - New features of „big atom“
- Nucleation mechanism
  - Classical nucleation theory
  - Two-step nucleation theory

Typical phase diagrams

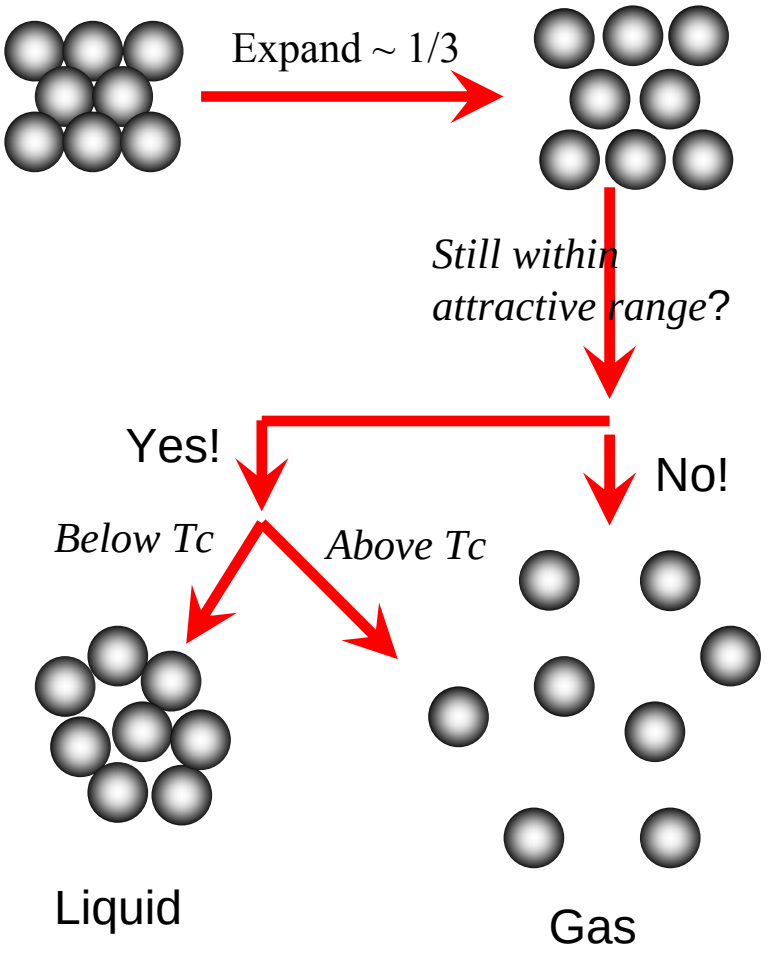




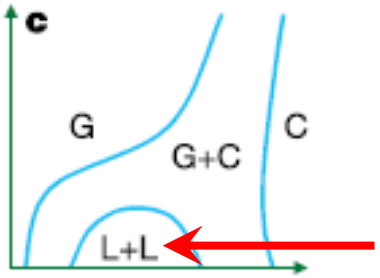
a. Hard sphere
b. Hard sphere with long-range attraction, $\Delta > 0.25\sigma$
c. Hard sphere with short-range attraction, $\Delta < 0.25\sigma$

Anderson, Nature 2002, 416, 811
Sciortino, F. Nature Mater.2003
Nägele 2005, Vekilov 2010

Short-range attraction: How short?



Expansion of the density by ~1/3 corresponds to a ~10% increase in the average interparticle spacing (cf the Lindemann criterion).



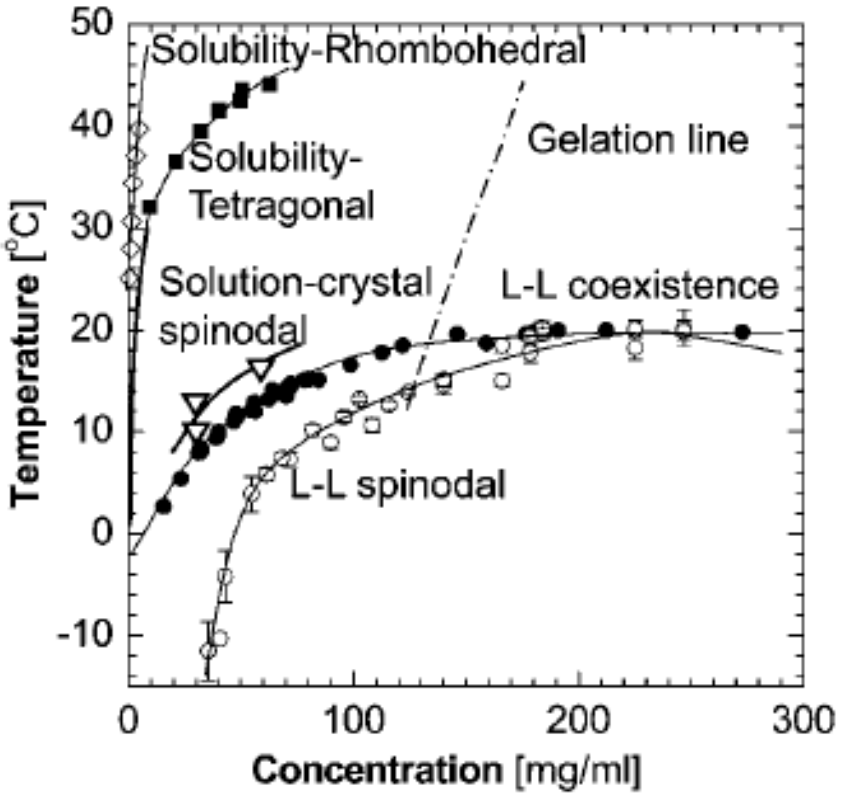
Meta-stable LLPS

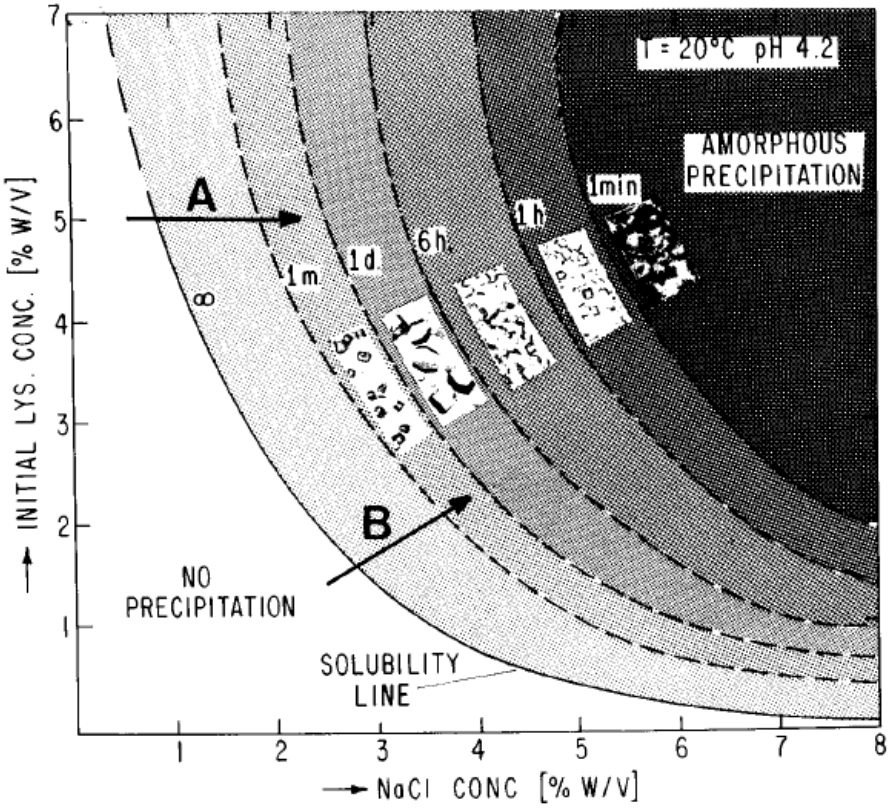
$\Delta < 0.25\sigma$

A schematic illustration of why an interparticle attraction of long enough range is needed for a thermodynamically stable liquid phase to occur.

Poon, WCK, J. Phys. Condens. Matter 2002, 14, R859

Phase diagrams in (c_p, T) (c_p, c_s) plane

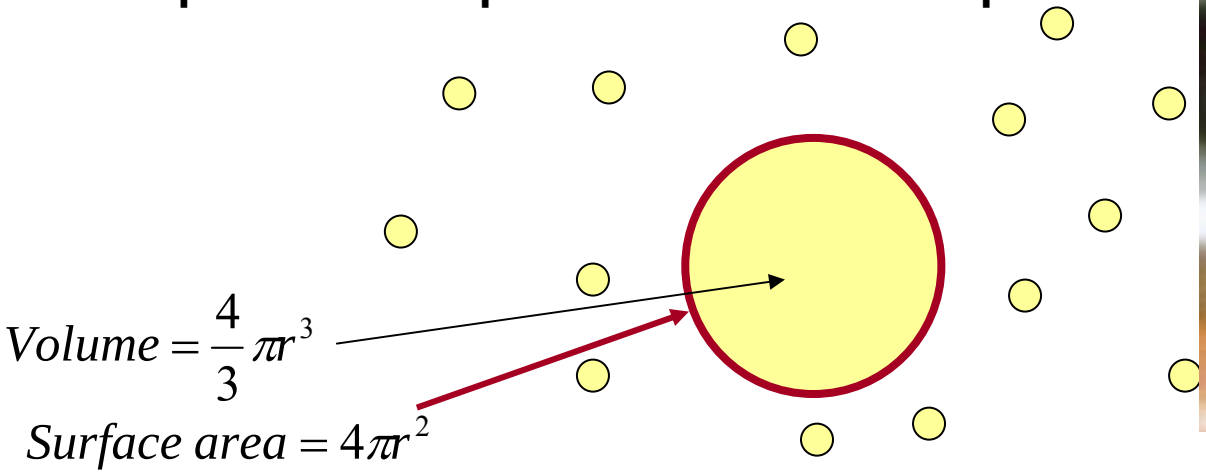


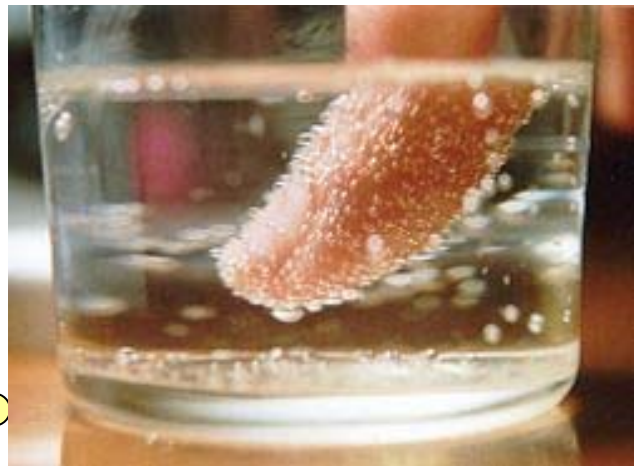


Vekilov 2010, Durbin 1996

Classical Nucleation Theory

Classical problem: Droplet in contact with vapor





$$\Delta G = \frac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \sigma$$

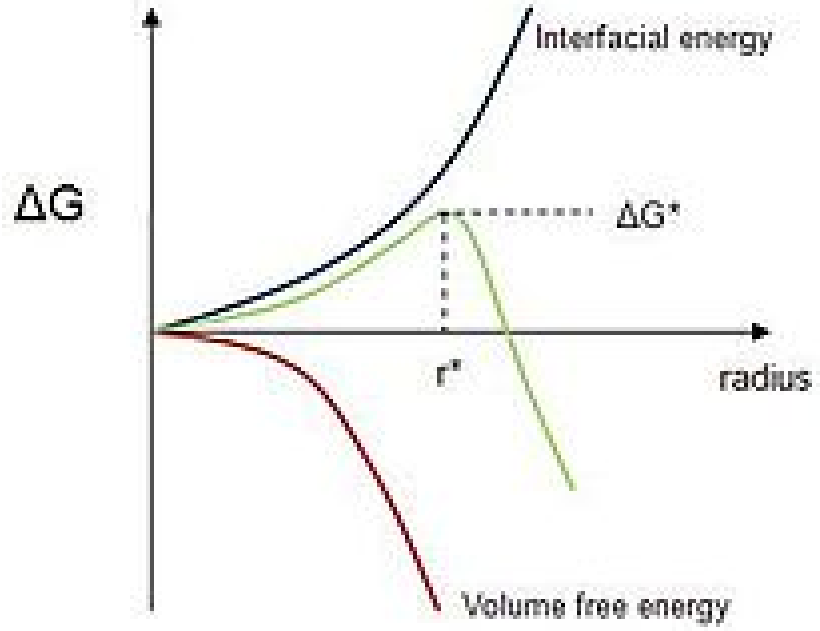
Surface tension
= Free Energy per unit area
of interface

Free energy change
per unit volume

Free energy change
per unit area of surface

http://en.wikipedia.org/wiki/Nucleation

Classical Nucleation Theory



- Critical nuclei: small size and short life time micoseconds
- Clusters r < r* shrink; r > r* grow

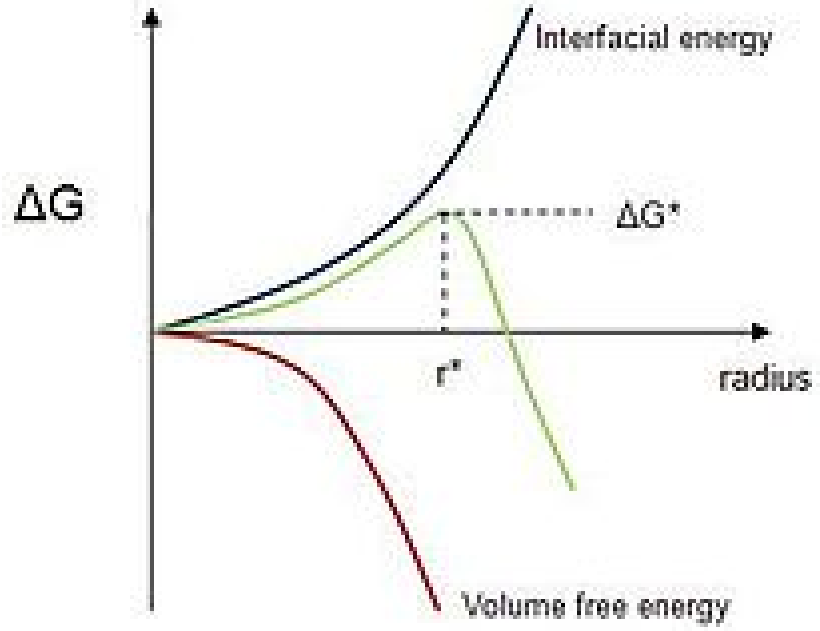
$$\Delta G = \frac{4}{3}\pi r^3 \Delta G_V + 4\pi r^2 \sigma$$

$$\frac{\partial \Delta G}{\partial r} = 4\pi r^2 \Delta G_V + 8\pi r\ \sigma$$

At critical nucleus size

$$\frac{\partial \Delta G}{\partial r} = 0, \text{so}\ \ r_{\text{critical}} = \frac{-2\sigma}{\Delta G_V}$$

Classical Nucleation Theory



$$\Delta G^* = \frac{16\pi\sigma^3}{3(G_v)^2} \quad \Delta S_v = \frac{\Delta H_v}{T_m}$$

$$\Delta G_v = \Delta H_v - T(\frac{\Delta H_v}{T_m})$$

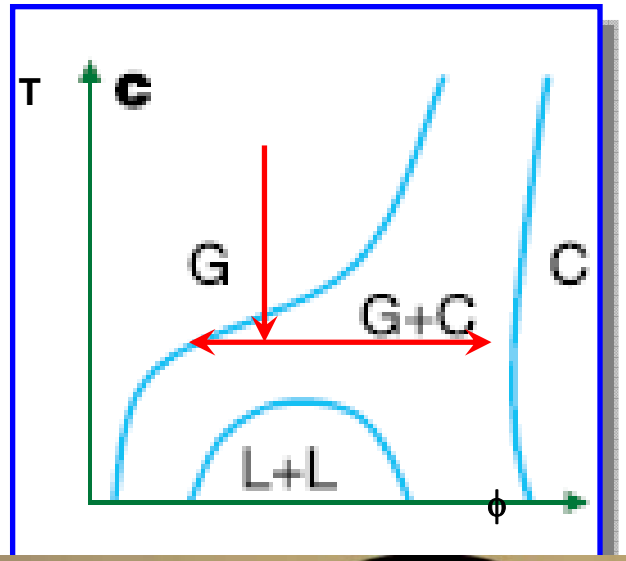
$$\Delta G_v = \frac{\Delta H_v}{T_m}\Delta T$$

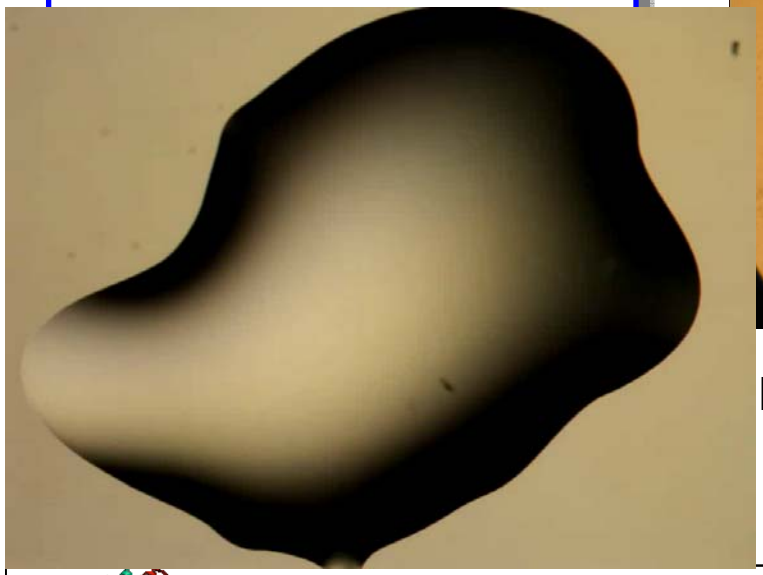
$$r^* = \frac{2\sigma T_m}{\Delta H_s}\frac{1}{\Delta T}$$

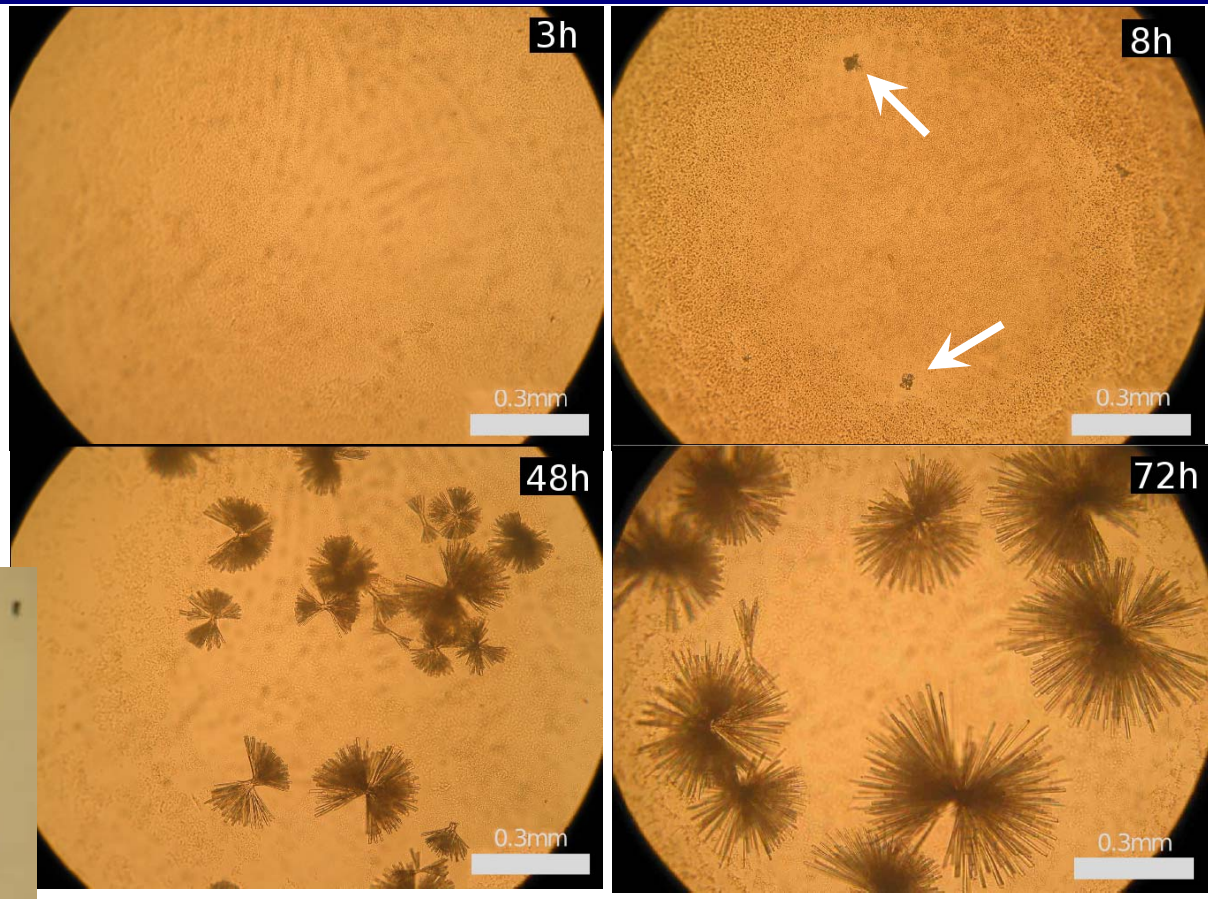
$$\Delta G^* = \frac{16\pi\sigma^3 T_m^2}{3\Delta H_s^2}\frac{1}{(\Delta T)^2}$$

- The greater the supercooling, the smaller the critical radius and the less energy needed to form it.

Classical Nucleation Theory



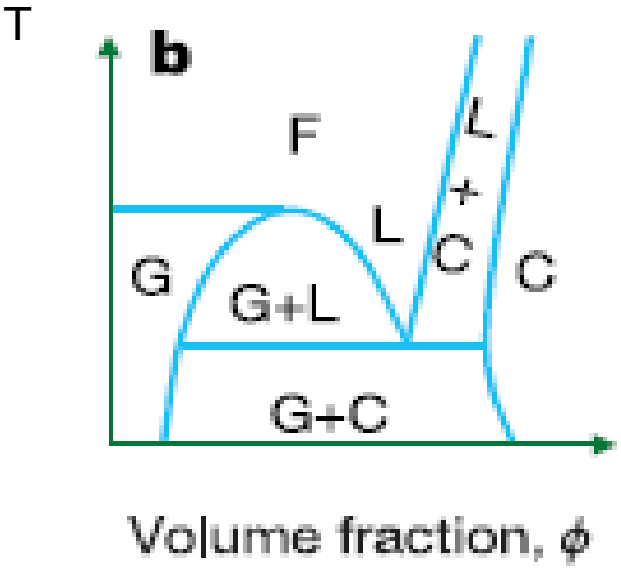




LG 10 mg/mL YCl_3 0.3 mM, by hanging-drop method

F. Zhang, et al. J. Appl. Cryst. 2011,

Classical Nucleation Theory



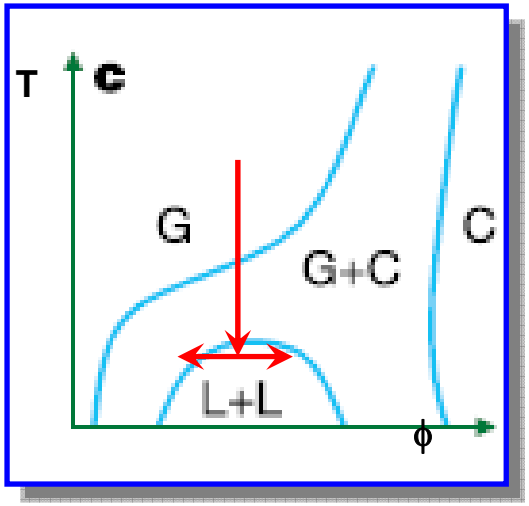
- The limitation of CNT

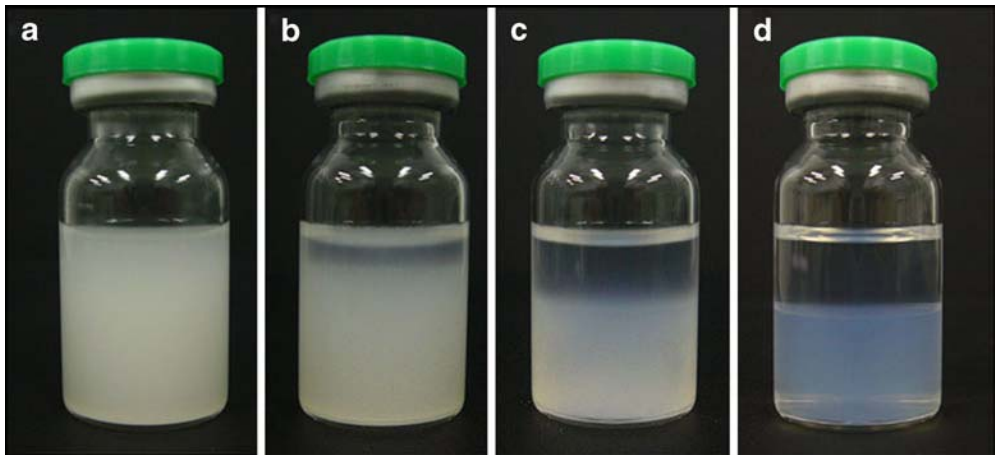
1. Homogeneous structure of dense droplets
2. Surface tension is equal to the final phase

G to L: density
L to C: density, mainly structure
Protein crystallization from solution, both density and structure changes

D. Erdemir, A. Y. Lee, A. S. Myerson, *Acc. Chem. Research* 2009, 42, 621

Meta-stable LLPS in protein solutions





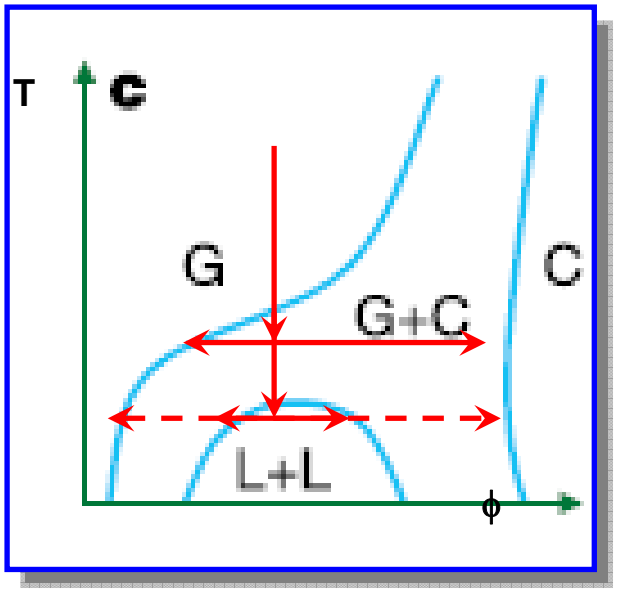
G: solution
LLPS gives two coexisting liquid phases with different density

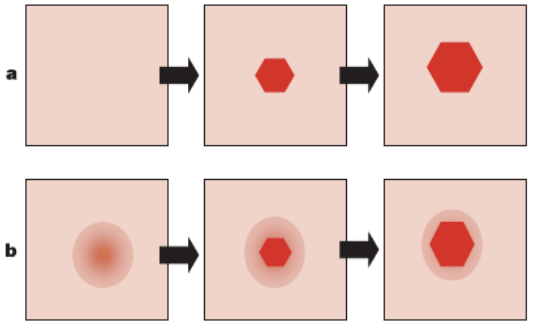
IgGI 107 mg/mL and HSA30.3mg/mL with 4mM YCl_3

H. Nishi et al. Pharm Res 2010, 27, 1348

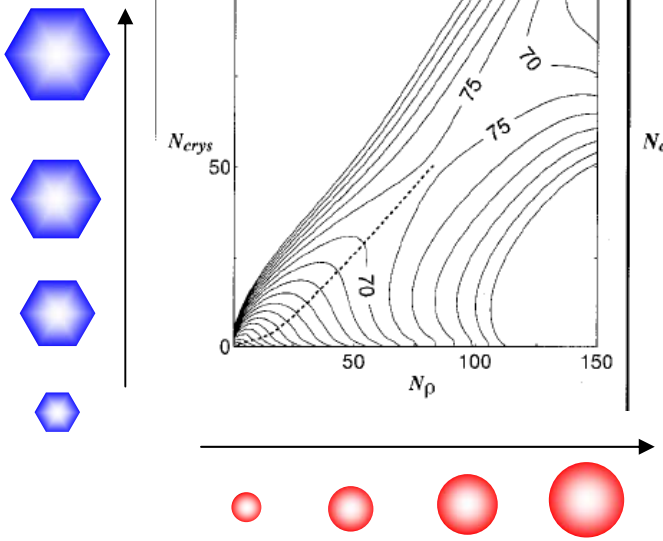
Y. Wang et al. PNAS 2011

Beyond CNT: Two-Step Mechanism



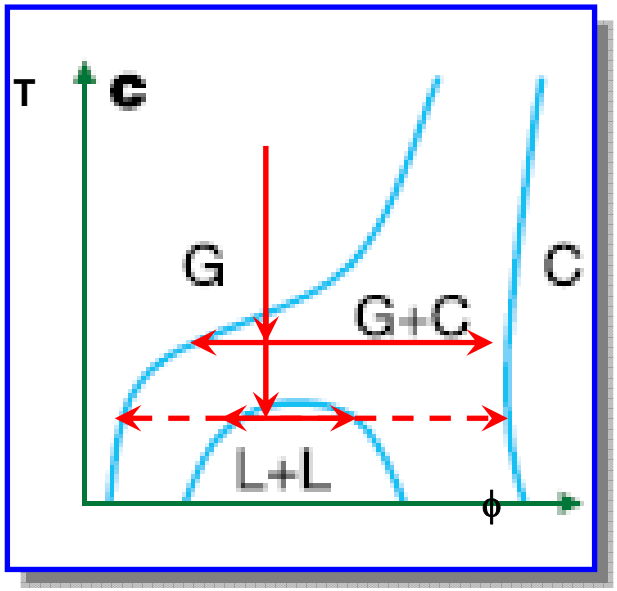


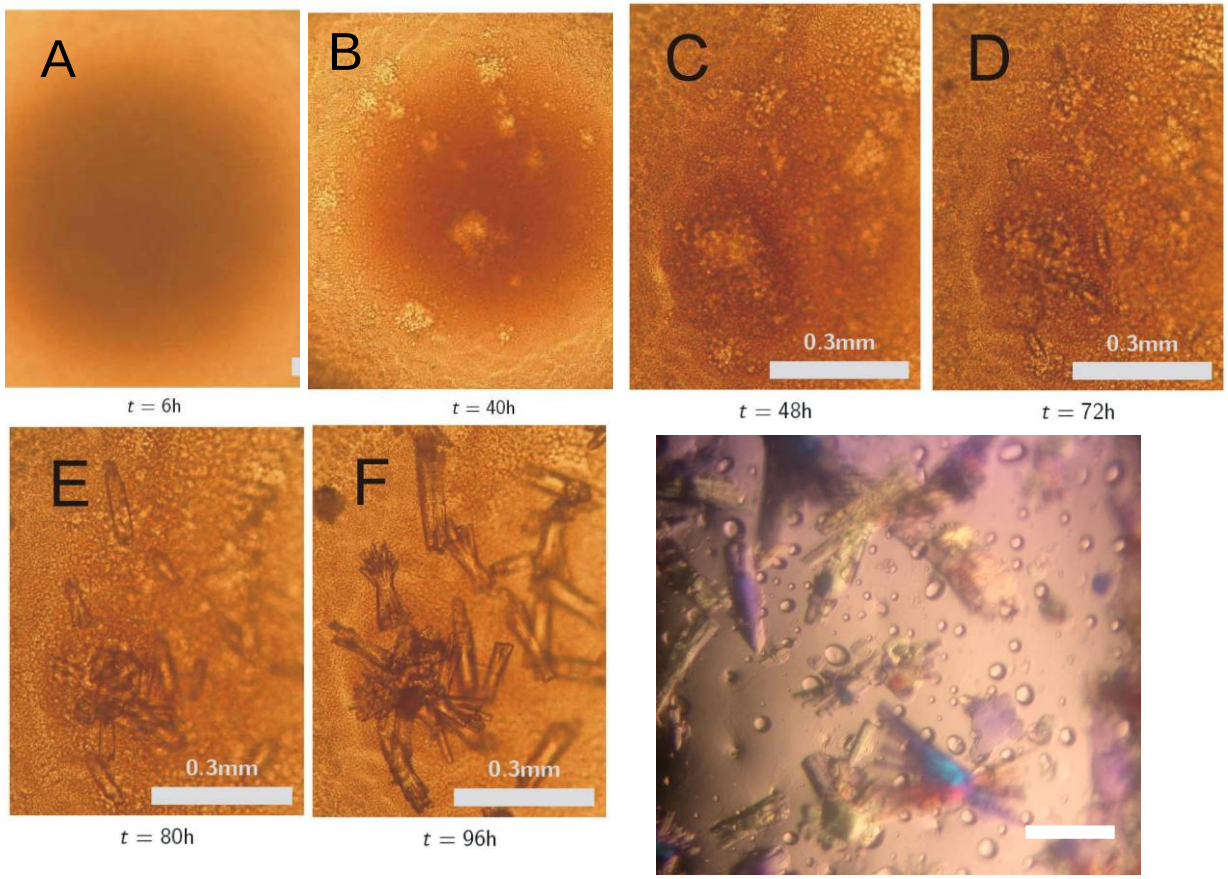
Anderson, Nature 2002, 416, 811



Order parameter separation at the critical point
P.R. ten Wolde, D. Frenkel. Science 1997

Beyond CNT: Two-Step Mechanism



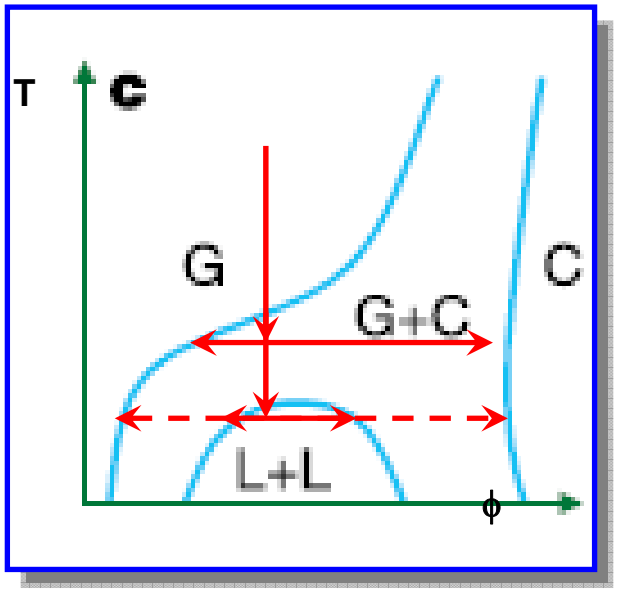


BLG 10 mg/mL YCl_3 3.0 mM

BLG 50 mg/mL YCl_3 8.0 mM

F. Zhang, et al. J. Appl. Cryst. 2011

Beyond CNT: Two-Step Mechanism



G: solution
LLPS gives two coexisting liquid phases with different density

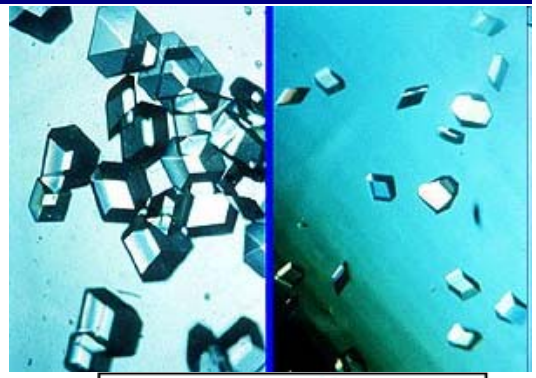
Normal Hemoglobins LLPS
Sickle cell Hb polymerization from LLPS

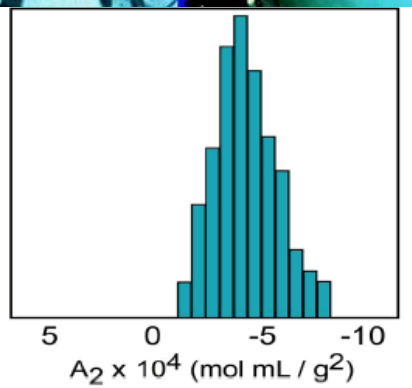
Anderson, Nature 2002, 416, 811

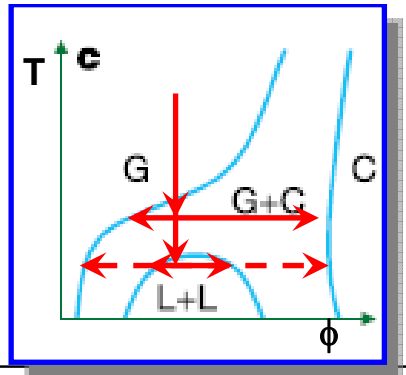
Galkin, O. et al. PNAS 2000, PNAS. 2002, 99, 8479

Summary

- Protein crystallization: important in many fields but difficult, un-predictable still;
- Physical chemistry of protein solutions;
  - A2 so far is the only predictor (not fast enough);
  - Metastable L-L coexistence
- Nucleation mechanisms: classical nucleation theory and beyond
  - CNT is not the only pathway for protein crystallization
  - Separation of order parameter (density, structure) leads to two-step nucleation procedure.







Second Virial Coefficient, A_2 & B_2

Ideal gas: pV=nRT

Van der Waals

$$\left(p + \frac{n^2 a}{V^2}\right)(V - nb) = nRT$$

Virial expansion

$$P = kT\frac{n}{V}\left[A_1 + A_2\left(\frac{n}{V}\right) + A_3\left(\frac{n}{V}\right)^2 + \cdots\right]$$

- Static Light Scattering
- A_2 and B_2

In solution, Osmotic pressure of ideal case:
$\Pi = MRT$

$$\Pi = RTc\left(\frac{1}{M_w} + B_2 c + \cdots\right)$$

B_2 is defined by the virial equation of state, which describes the nonideality of the osmotic pressure.

$$B_2 = -2\pi\int_0^\infty \left(e^{-u(r)/kT} - 1\right) r_{12}^2 dr_{12}$$

The Relationship between A_2 and B_2

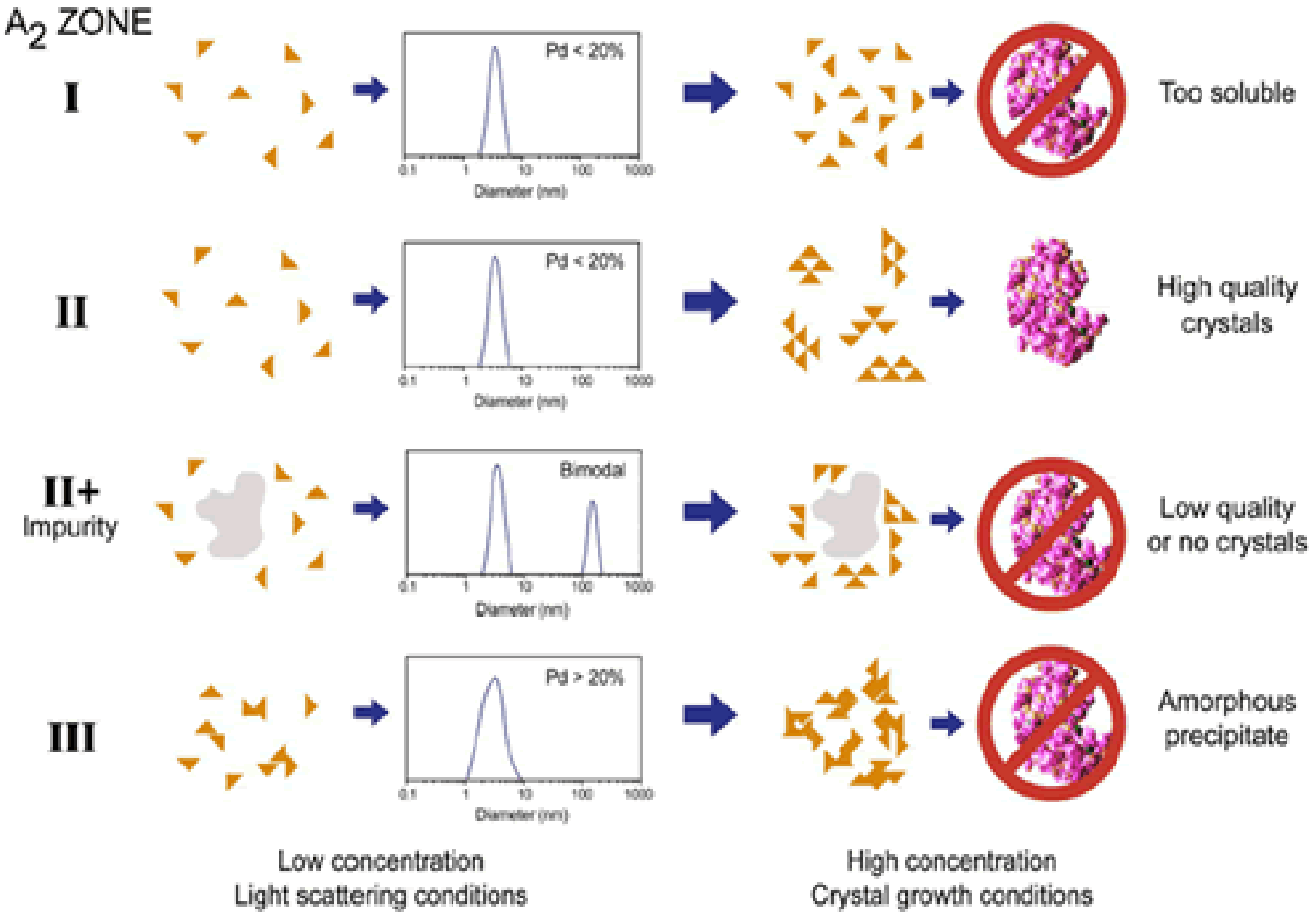
$$\frac{\Pi}{c_p \cdot R \cdot T} = \frac{1}{M_W} + A_2 c_p + \ldots$$

$$\frac{\Pi}{\rho \cdot k_B \cdot T} = 1 + B_2 \rho + \ldots$$

$$\rho = c_p N_A / M_W$$

$$[A_2 c_p] = \left[\frac{B_2 \rho}{M_W}\right] \Leftrightarrow [A_2] = \left[\frac{B_2 \rho}{c_p \cdot M_W}\right] = \left[\frac{B_2 c_p \cdot N_A}{M_W^2 \cdot c_p}\right] = \left[\frac{B_2 N_A}{M_W^2}\right]$$

Dynamic Light Scattering



- Pd ~ 20% gives high quality protein crystals;
- Emperical observation, physical meaning is not clear

(Malvern online info.)

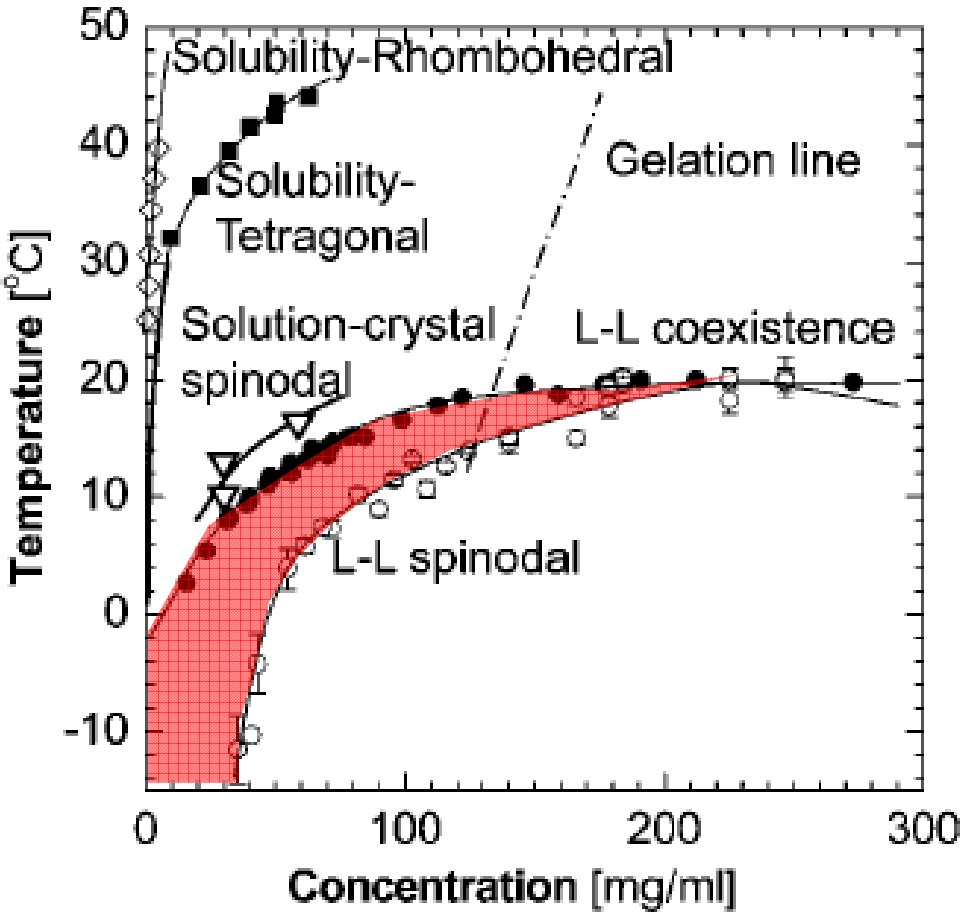
Summary I

- Protein crystallization: important in many fields but difficult, un-predictable still;
  - Length scale in general
  - Shape or contacts
  - Interactions
- Understanding of the Physical chemistry of protein solutions;
  - A2 so far is the only predictor (not fast enough);
  - DLS screening: Pd ~ 20% (no clear physical meaning)
  - Better predictor in the future?

Brief History

- 1895: W. C. Roentgen discovers X rays (Bragg, p. 1).
- 1912: Max von Laue discovers X-ray diffraction by crystals (Bragg, p. 7).
- 1913: W. L. Bragg reports the crystal structure of NaCl, providing the first experimental evidence for the absence of salt "molecules". (Bragg; Glusker p. 3).
- 1928: Kathleen Lonsdale reports the structure of benzene as having six equal sized bonds instead of alternating double and single bonds (Glusker, p. 3).
- 1935: J. M. Robertson *et al.* solve the structures of pthalocyanins, the first case of a complex organic molecule solved independently by crystallography (Bragg, p. 180).
- 1948: Bijvoet *et al.* solve strychnine, perhaps the first case in which crystallography decided between alternatives proposed by organic chemists (Bragg, p. 182).
- 1949-57: Dorothy Crowfoot Hodgkin *et al.* solved the structures of penicillin (1949) and vitamin B-12 (1957). She won the Nobel Prize in Chemistry in 1964. (Bragg, p. 189)
- 1958: **Myoglobin**, sperm whale, solution reported by Kendrew *et al.*. No PDB entry (nor obsolete PDB entry) represents the original myoglobin structure, since the PDB was not established until 1971. In 1973, Watson and Kendrew deposited 1MBN, a 2.0 Å structure refined with then-current procedures.
- 1962: Max Ferdinand Perutz and Sir John Cowdery Kendrew win the Nobel Prize in Chemistry for their studies on the structures of globlular proteins. See interviews in the yellow box above.
- 1965: **Lysozyme**, hen egg white, solution reported by Phillips *et al.*.

Typical phase diagrams

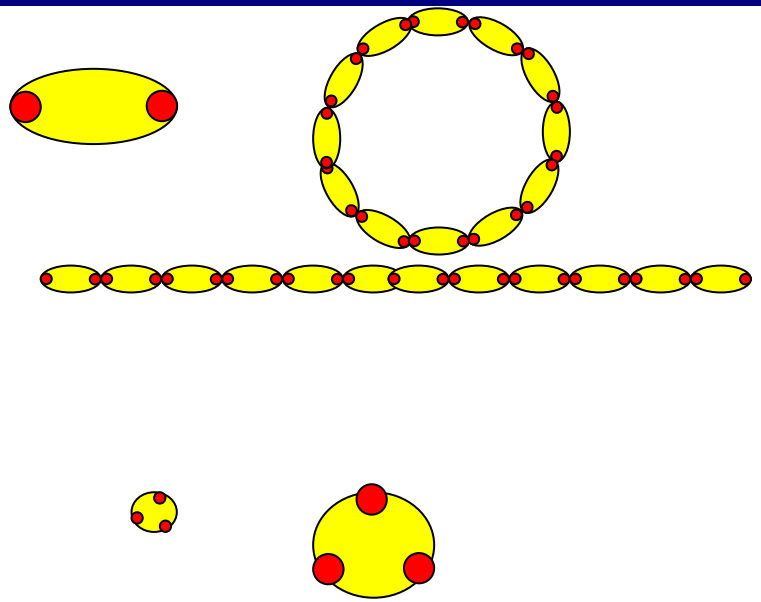


- Lysozyme
- γ-crystallin
- Limited results of protein systems

Vekilov 2010

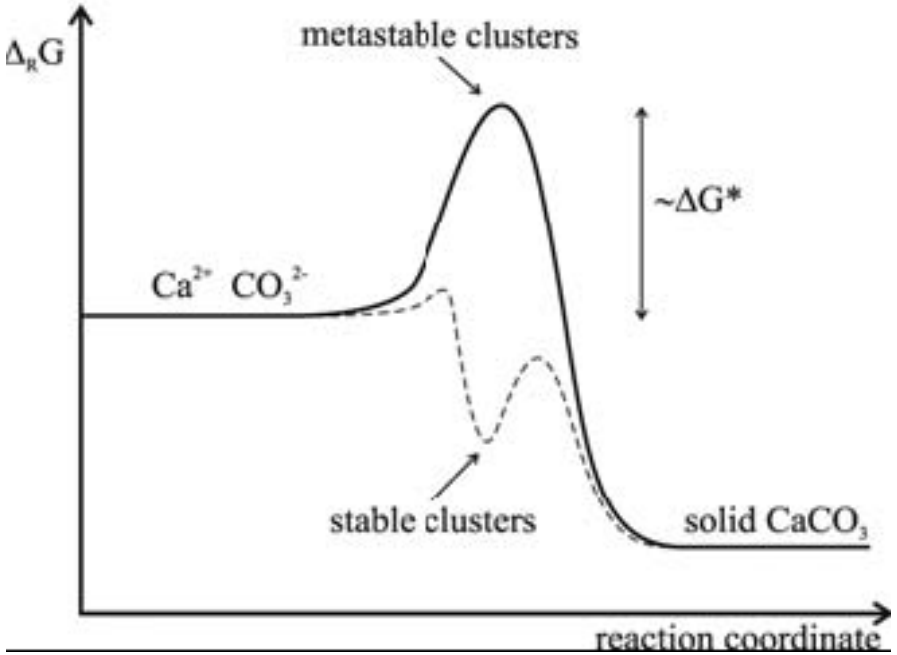
Why protein crystallization difficult?

- More reasons:
- Negative design by Nature
- Unique: proteins do not share universal behavior (structure, geometry, interactions, etc.)
- …



J. P. K. Doye, et al. Phys. Biol. 2004, 1, 9-13

Now you see it



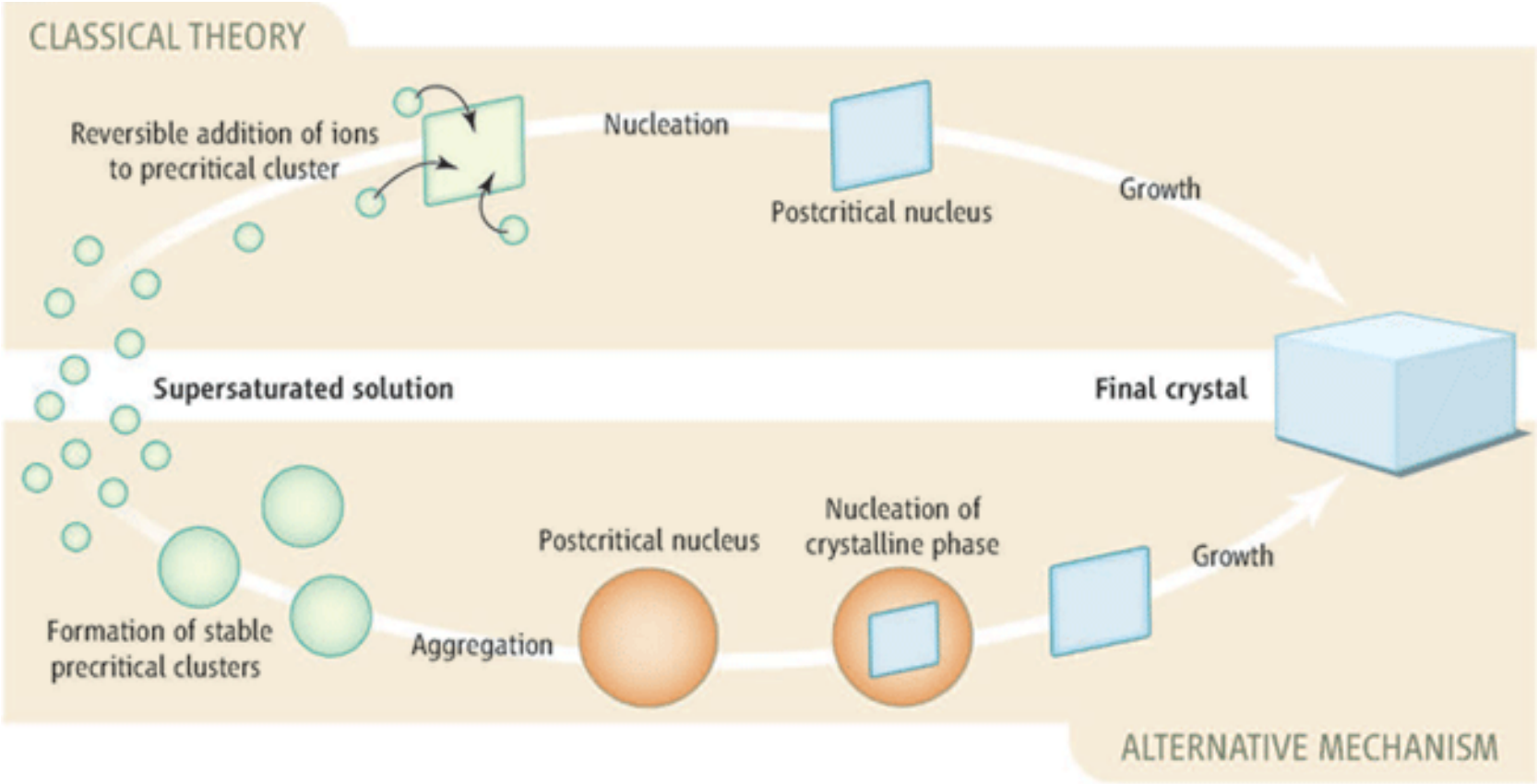
Gebauer et al. Observe long-lived precritical clusters, about 2 nm in diameter, and suggest that they grow by colliding and coalescing.

Why do they neither dissolve nor grow?

Structure of cluster

D. Gebauer, A. Völkel, H. Gölfen, *Science* 2008, 322, 1819

Can you really see it?



F.C. Meldrum, R. P. Sear, Science 2008,322,1802

You may see it *indirectly*

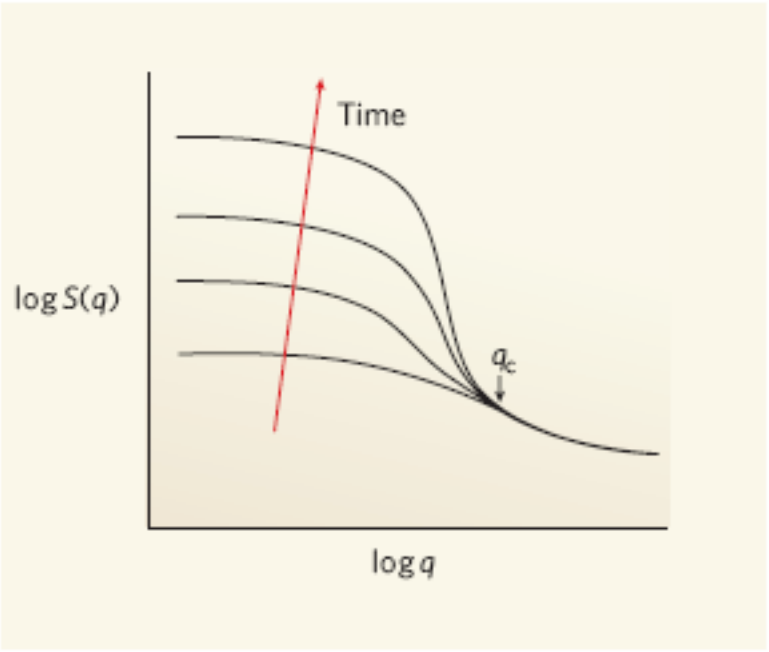
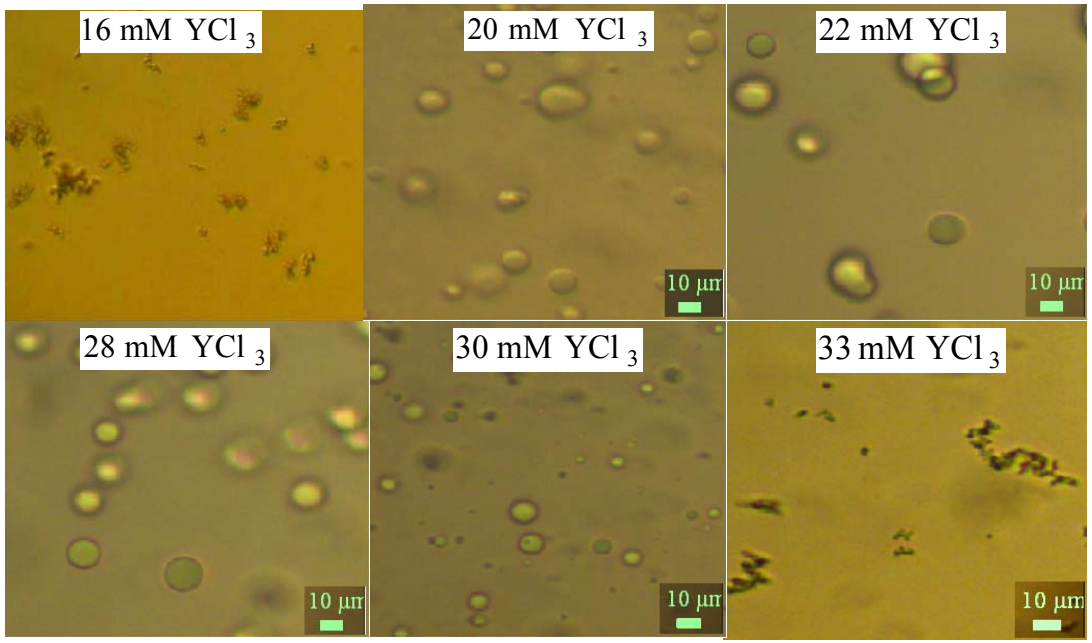


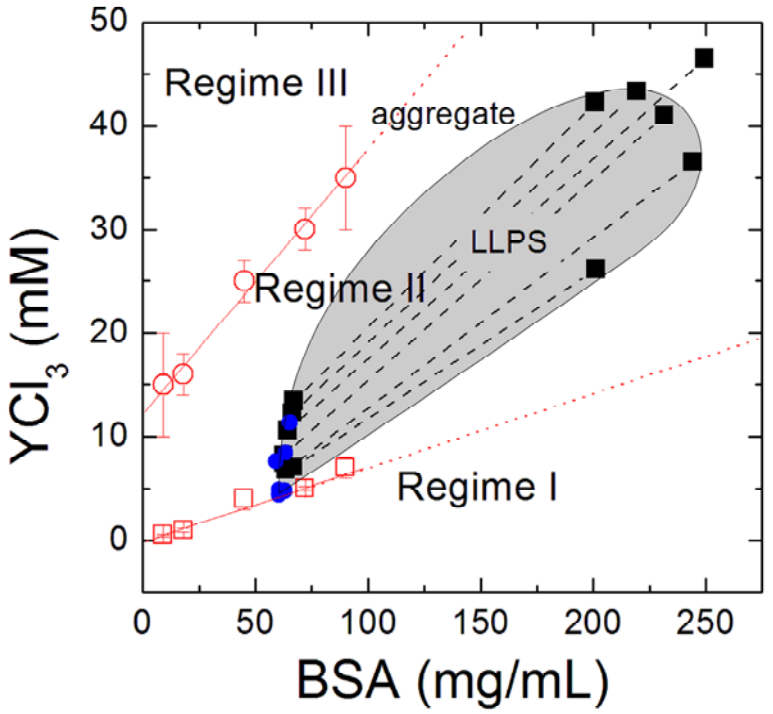
Figure 1 | The search for the critical nucleus. Scattering intensity, or structure factor $S(q)$, as measured by Pan and colleagues[1] for nucleation in a phase-separating mixture. Curves obtained at different times during a phase transition merge at a critical scattering vector q_c, implying that the length $1/q_c$ is a signature of the critical nucleus from which the new phase forms.

- When a phase is born
- A possible way to determine the size of critical nuclei
- A chellange to CNT since this method indicate the desity fluctuation like nucleation??

Science 2006, 441, 168

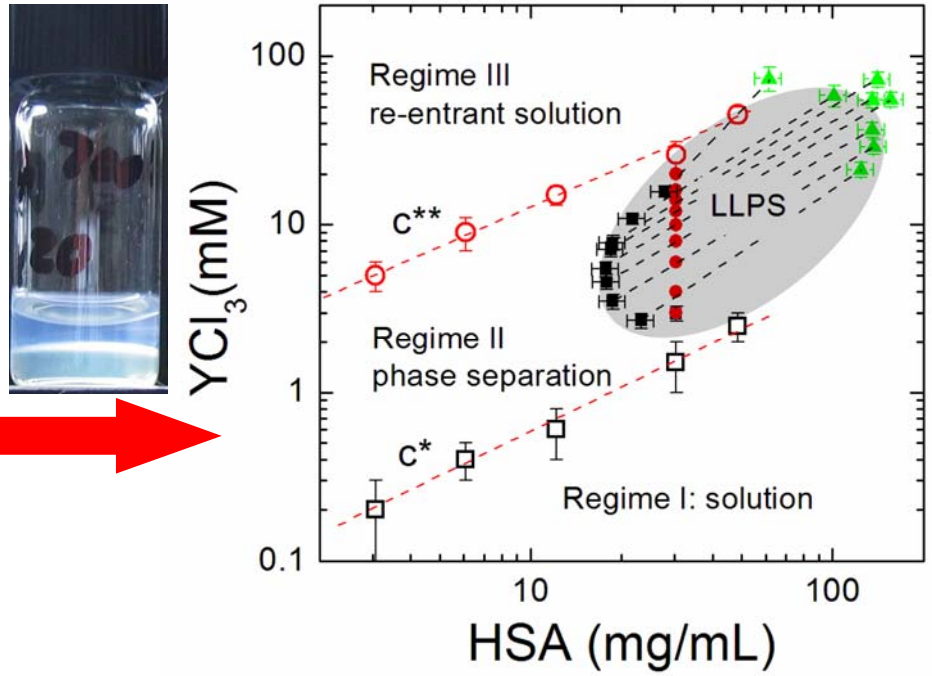
Meta-stable LLPS: BSA





Liquid-liquid phase separation for BSA200mg/mL with YCl3 at 22°C

Meta-stable LLPS: HSA



HSA30.3mg/mL with 4mM YCl_3

c_s^1 and c_p^1 was determined by X-ray and UV absorption, respectively and c_s^2 and c_p^2 was calculated from the volume of each phase and the initial c_s and c_p.

F. Zhang et al. *submitted* **2011a**

3D Phase Diagram