(Some) Physical Chemistry of Crystallization

Bernhard Rupp
Prosecutor
Societas Eruditorum
q.e.d. life science discoveries
Livermore, CA 94551, USA
University of California, Irvine

- Thermodynamics vs. kinetics
- Solubility of proteins
- Phase diagrams and what they (not) mean
- Roaming the phase space with different techniques

I thought all we need is more protein ....
Crystal: periodically stacked molecules

Unit lattice + Motif = Unit cell
Periodic network of sparse and weak intermolecular interactions

Highly specific and LOCAL phenomenon!
What is protein crystallization?

Protein crystallization is a special case of phase separation from thermodynamically metastable solutions under the control of kinetic parameters.

- Mix cocktail and protein on glass slide
- Turn slide and seal well
- Watch and wait/see/pray/hope
  - poke/seed/spike
- Vapor diffuses into well, concentrations in drop increase
- Harvest and mount crystals

Well with crystallization cocktail (precipitants, additives, detergents etc - unlimited combinations possible!)

Movie courtesy G. Sheldrick’s group
What is protein crystallization?

Protein crystallization is a special case of phase separation from thermodynamically metastable solutions under the control of kinetic parameters.

A

The solution must be in or move into a composition region where phase separation is thermodynamically possible and a crystalline phase is stable: macroscopic phase equilibria and protein solubility.

B

Kinetic parameters, such as nucleation rates, growth kinetics, convection (gravity) must be conducive to crystal formation - microscopic foundations.

Only if (A and B) then $\text{prob}(C)$
Crystallization Techniques

Various techniques may have different virtues for screening and optimization (production).

- The Classic: HDVD
- The Variant: SDVD
- Micro-batch under oil
- Micro-dialysis
- Free interface diffusion

Sample amount
Miniaturizability
Ease of sealing
Ease of harvesting
Coverage and pathway through phase space
Cost
Reproducibility
Mapping the thermodynamically possible - the phase diagram

Disclaimer: the 'phase diagrams' used in protein crystallization are not true equilibrium phase diagrams, but rather playgrounds for imagination and may contain kinetic and even spatial information

1) Conditions for phase stability

2) Conditions for phase equilibrium

3) Reaction thereof to compositional change
Simple ideal systems and their reaction to change

This is boring, because the system is IDEAL.

The fun starts when things become non-ideal and excessive.

\[ P = n \frac{RT}{V} \]

\[ \left( \frac{\partial P}{\partial V} \right)_{T,n} < 0 \]
Non-ideal systems and their reaction to change

Non-ideality:

A - attraction between molecules (how fun starts)

B - molecules have finite volume (also necessary for fun)

\[ P = \frac{nRT}{V_0 - n \cdot b} - a \cdot \left( \frac{n}{V_0} \right)^2 \]

\[ P_v = \frac{nRT}{V} (1 + B'P + C'P^2 + D'P^3 \ldots) \]

Contains nonideality
Non-ideal systems and their reaction to change

In red part:

Pressure increase leads to increase in volume!

Not possible.

Stability requirement:

\[
\left( \frac{\partial P}{\partial V} \right)_{T,n} < 0
\]

\[
P = \frac{nRT}{V_0 - n \cdot b} - a \cdot \left( \frac{n}{V_0} \right)^2
\]
Fundamental Equations of state

\[ \Phi = \sum_{i=1}^{n} x_i P_i \]

\[ d\Phi(x_i) = \sum_{i=1}^{n} P_i \, dx_i \]

\[ d\Phi(x_i) = \sum_{i=1}^{n} \left( \frac{\partial \Phi}{\partial x_i} \right)_{x_j \neq i} \, dx_i \]

\[ U_{(S,V)} = PV + TS \]
Reaction to change in multicomponent systems

\[ dG(P,T,x_i) = VdP - SdT + \sum_{i=1}^{n} \left( \frac{\partial \bar{G}}{\partial x_i} \right)_{Xj \neq i} dx_i \]

\[ . \]

\[ dG(P,T,x_i)_{T,P} = \sum_{i=1}^{n} \left( \frac{\partial \bar{G}}{\partial x_i} \right)_{Xj \neq i} dx_i \]

\[ z_m = \bar{Z} - \sum_{i=1}^{m-1} \frac{\partial \bar{Z}}{\partial x_i} x_i \]
Non-ideal multicomponent systems and their reaction to compositional change

In red part:

Curvature becomes negative

Gaht nüt, gibts net, nil, nada, zilch, meschugge.

Spontaneous (spinodal) decomposition

\[
\left( \frac{\partial^2 G}{\partial x_1^2} \right)_{P,T} > 0
\]
From $G/x$ diagrams $T/x$ diagrams can be constructed....
...from which finally isothermal sections of the water rich corner our ternary phase diagram can be derived.
There is our basic ‘phase diagram’ with the solubility and instability curve.

- **Precipitate + protein solution**
- **Two-phase region**
- **Unstable, spontaneous decomposition**
- **Solubility line**
- **Metastable solution**
- Will eventually separate into protein rich phase (maybe crystals) plus saturated solution
- **Clear protein solution**
- **Single phase**
- **Stable**
- **Pure water**
- **Precipitant concentration**
- **Protein concentration**

---

PEPC-V Course, EMBL Hamburg, August 2006

Copyrighted material not for public distribution © Bernhard Rupp 2006
Solubility is different for each precipitant
Example: PEGs
Solubility is different for each precipitant

Example: Salts

Opposite correlations for anions and cations with their degree of strong hydration

Ionic strength

log S (protein solubility)

single phase, stable

metastable supersaturated solution

salting in

salting out

slope \(-k\)
pH and solubility

Isoelectric point (pI)

pI is NOT the most common pH of crystallization
What about pH of crystallization?

- pH from PDB
- pI from PDB
Some net charge is favorable - think locally!

KK and BR, Bioinformatics 2004
Summary thermodynamics

A) The solution must be in or move into a composition region where phase separation is thermodynamically possible.

Necessary requirement: supersaturation

B) For a crystal to form, the enthalpic gain through formation of contacts must outweigh the entropic loss due to increased order in the system:

$$\Delta G_c = \Delta H_S - T\Delta S_S$$

Necessary requirement: Phase stability

However, B) does not have to happen – Activation and nucleation
Activation and nucleation

Activation Energy ($\Delta G^\#$)

- Metastable super-saturated solution, single phase
- 2-phase equilibrium, crystal + saturated solution
Crystallization Techniques

Various techniques may have different virtues for screening and optimization (production)

The Classic: HDVD
The Variant: SDVD
Micro-batch under oil
Micro-dialysis
Free interface diffusion

Sample amount
Miniaturizability
Ease of sealing
Ease of harvesting
Coverage and pathway through phase space
Cost
Reproducibility
Pathway of Vapor Diffusion experiment I

stable → metastable → unstable

growth
spontaneous nucleation

Protein concentration

Pure water
Precipitant concentration

p
p/2

P/EPC-V Course, EMBL Hamburg, August 2006
Copyrighted material not for public distribution © Bernhard Rupp 2006
Pathway of Vapor Diffusion experiment II

stable → metastable → unstable

growth  spontaneous nucleation

MOVIE

PEPC-V Course, EMBL Hamburg, August 2006

Copyrighted material not for public distribution © Bernhard Rupp 2006
Pathway of Vapor Diffusion experiment III

Stable $\rightarrow$ metastable $\rightarrow$ unstable

Protein concentration

growth

spontaneous nucleation

Pure water $\rightarrow$ Precipitant concentration

$\text{Pathway of Vapor Diffusion experiment III}$
Pathway of batch crystallization experiments

stable → metastable → unstable

Protein concentration

growth
spontaneous nucleation

Pure water
Precipitant concentration

1. stable region
2. metastable region
3. unstable region
4. spontaneous nucleation
5. growth

123
Pathway of dialysis experiments

Protein concentration

spontaneous nucleation

growth

Pure water

Precipitant concentration
Free interface diffusion with micro-fluidics

Hansen, Berger, Quake

Uses very little material
Different traversing of phase space
Difficult harvesting
Scale-up and harvesting difficult and overlap not yet determined
Free-interface diffusion: spatial and temporal concentration profiles

- $t = 0$
- $t = t_{1/3}$
- $t = t_{2/3}$
- $t = t_E$
Pseudo-Pathways of Free Interface Diffusion experiment

\[ t = 0 \]

\[ t = \frac{1}{3} t_E \]

\[ t = \frac{2}{3} t_E \]

Protein concentration

Spontaneous nucleation

Pure water

Precipitant concentration
Pseudo-Pathways of Free Interface Diffusion experiment

Reagent (precipitant) concentration

\[ \frac{c}{(\text{Protein})} \]

[L+S]

[
L
]

['Water corner']

1:1

3:1

1:3

PEPC-V Course, EMBL Hamburg, August 2006

Copyrighted material not for public distribution © Bernhard Rupp 2006
Change of method affects outcome—different pathways

A

sitting drop VD

hanging drop VD

B

microbatch

microdiffusion
Summary kinetics

A) The supersaturated solution must have critical nucleation events – not too many but at least some. Rationale for Seeding

B) Different methods travel through the crystallization phase diagram through different paths and with varying coverage. The same reagents do not necessarily give the same results!

Irrespective of all above - efficiency is main concern: most successful experiments with the least material and cost.
Choice of effective reagents
Easy proteins come in multiple forms

again ~ 30% of proteins ...
Static light scattering sheds light on the interactions between protein molecules in a given medium (system) – TD = necessary but not sufficient for crystallization.

\[ P_{(v)} = \frac{nRT}{V}(1 + B'P + C'P^2 + D'P^3) \ldots \]

\[ \frac{K \cdot c}{R_{90}} = \frac{1}{M} + 2B_{22} \cdot c \]
**Dynamic** light scattering sheds light on the hydrodynamic properties of protein molecules in a given medium (system) - not a reliable predictor for absence of crystallization.

- **Narrow size distribution** indicates low conformational microheterogeneity.
- **Monomodal distribution** indicates lack of aggregates.
Key strategies

Accept that your chance of obtaining diffracting crystals of your protein without any additional procedural adjustments or protein modifications is only 10-20%.

Whether your crystallization will succeed or not is largely predetermined by the protein construct itself. If it cannot crystallize, it will not, no matter what you do. Consider your protein and its preferences and dislikes in your screening design.

Accept the probabilistic nature of the crystallization game. You can win only by increasing your odds, not by seeking certainty.

In other words, do nothing stupid, but sample everything else efficiently.

Try to gain a rapid assessment of your protein’s crystallization propensity by using a two-tiered approach. Start with a pH-PEG screen or Index™-type screen and expand the sample space in the next round.
Key strategies

Do not oversample - if no promising results are obtained after about 300 trials, it is likely that your protein construct is a hard case for crystallization. Consider other constructs or orthologs.

Use robotics for repetitive and tedious tasks and for miniaturization, but do not expect your robots to think for you - or to save you much time.

The choice of screening kit or reagent set is probably the least significant variable. Most kits share the same basic reagents. Proteins that are going to crystallize at all will usually do so under multiple conditions.

Do not believe any tips or claims that lack a clear rationale. Causality rules, even for statistically infrequent events.

Good luck!
More 'tips' for when things go really wrong

- Stop screening - diminishing return.
- Start thinking
- Don’t see a tip psychic and don’t continue wasting material
- Read McPherson’s book (2nd Edition)
- Try seeding
- If material available - post mortem analysis
  - DLS - aggregation, microheterogeneity, CD for sec struc
  - Alternatives - protein engineering - natural or rational
- Consider cofactors and complex partners, allosteric affectors
- If it is important, it will crystallize
- If it crystallizes, it is important
- Try NMR