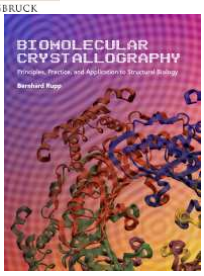
Lecture 1: From protein solution to protein crystals



www.ruppweb.org

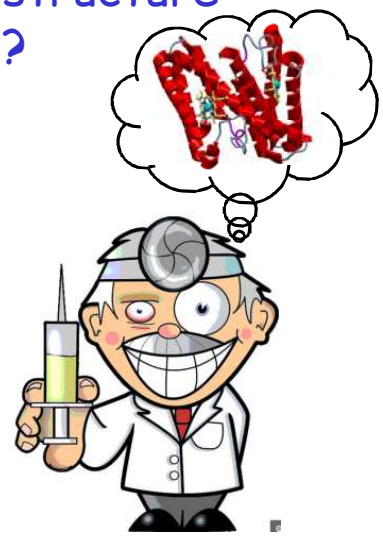
Bernhard Rupp
Dept. of Forensic Crystallography
k.-k. Hofkristallamt
Vista, CA 92084, USA
Innsbruck, A 6020 Austria

- Nature and properties of crystals (and proteins)
- Thermodynamics vs. kinetics
- Solubility of proteins

- Phase diagrams and what they (not) mean
- Roaming the phase space with different techniques
- HTPX and post-mortem analysis (second lecture)

Why do we care about macromolecular structure models ?

Structure models are useful because:
Molecular structure defines biological function - from basic research to understanding of disease, molecular medicine, therapeutic drug design



Molecular form defines function: huge variety in shapes and size

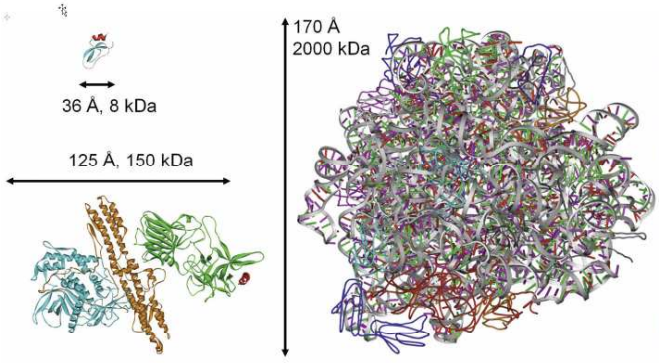


Figure 2-1 Protein structures determined by X-ray crystallography. The small protein (top left) is the bovine pancreatic trypsin inhibitor (58 residues, ~8 kDa), one of the first proteins whose structure was refined at atomic resolution.[4] Its two distinct secondary structure elements are shown in ribbon presentation (a red α-helix and two cyan anti-parallel β-strands).The larger molecule (bottom left, ~1300 residues, ~150 kDa) is the neurotoxin secreted by the bacterium *Clostridium botulinum*. Botulinum neurotoxins are the most toxic substances known to man. They consist of three functionally distinct domains: a Zn-protease (cyan ribbon), a translocation domain (orange), and the ganglioside-binding domain (green), which is shown bound to a recognition peptide from the neuronal cell surface (red helix).[5] The right panel shows the large (50S) ribosomal subunit of the bacterium *Haloarcula marismortui*. The huge, nearly 2 MDa large structure[6] contains 27 different proteins (~4000 residues) and the ribosomal 5S and 23S RNA, together having 2833 nucleotides. PDB entries 1bpi,[4] 3bta,[5] and 1ffk.[6]

X-ray crystallography provides these structures without size limitation and often at atomic detail level and relevant beyond solid state – with some caveats

Structure determination involves a large array of different techniques

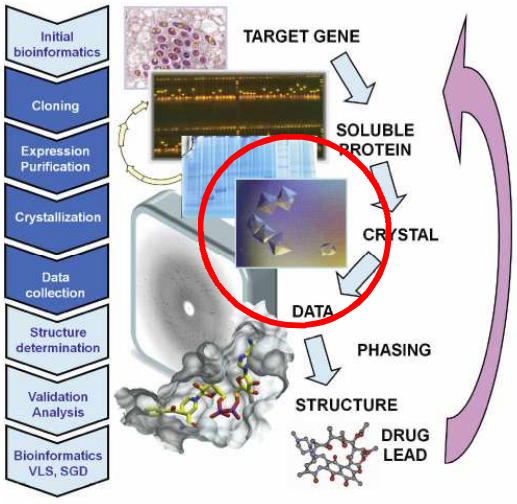
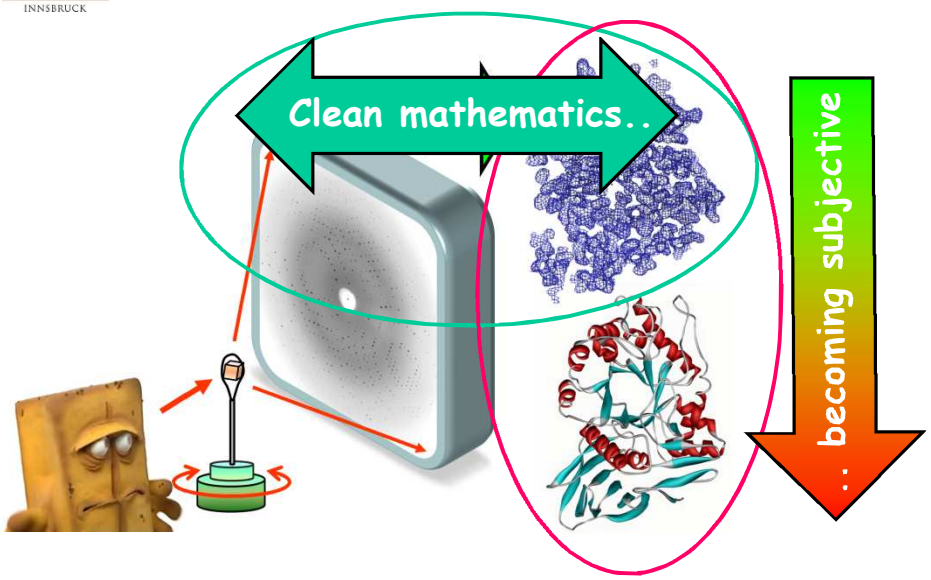


Figure 1-4 Overview of protein structure determination. The bar on the left side of the figure lists major stages of a crystal structure determinaton project. The dark blue shading indicates experimental procedures while the light shading indicates work performed in-silico on computers. The results of the structure analysis frequently feed back into the design of a refined study, particularly in structure guided drug discovery. VLS: virtual ligand screening; SGD: structure guided drug discovery. Consult Figure 1-8 for a more detailed diagram of key steps in structure determination and the corresponding Chapters in this book.

In 2 hrs we can cover only a few selected topics

The fundamental concept behind X-ray crystallography is deceptively simple



Higher resolution = more data = more information

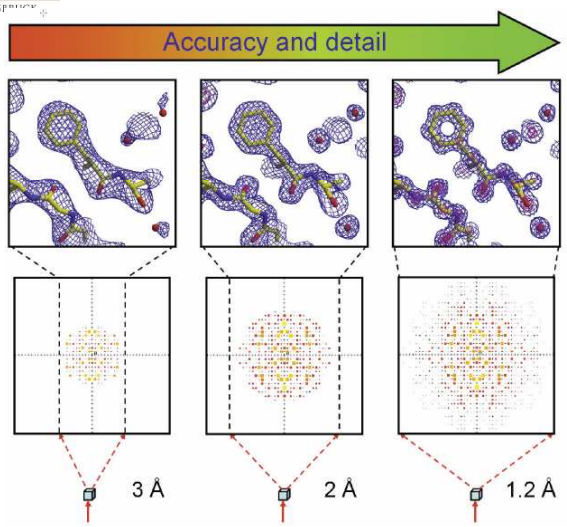


Figure 1-6 Data quality determines structural detail and accuracy. The qualitative relation between the extent of X-ray diffraction, the resulting amount of available diffraction data, and the quality and detail of the electron density reconstruction and protein structure model are evident from this figure: The crystals are labeled with the nominal resolution d_{min} given in Å (Ångström) and determined by the highest diffraction angle (corresponding to the closest sampling distance in the crystal, thus termed d_{min}) at which X-ray reflections are observed. Above each crystal is a sketch of the corresponding diffraction pattern, which contains significantly more data at higher resolution, corresponding to a smaller distance between discernible objects of approximately d_{min}. As a consequence, both the reconstruction of the electron density (blue grid) and the resulting structure model (stick model) are much more detailed and accurate. The non-SI unit Å (10^{-8} cm or 0.1 nm = 10^{-10} m) is frequently used in the crystallographic literature, simply because it is of the same order of magnitude as atomic radii (~0.77 Å for carbon) or bond lengths (~1.54 Å for the C–C single bond).

Holton Movie

Higher resolution = more data = more information

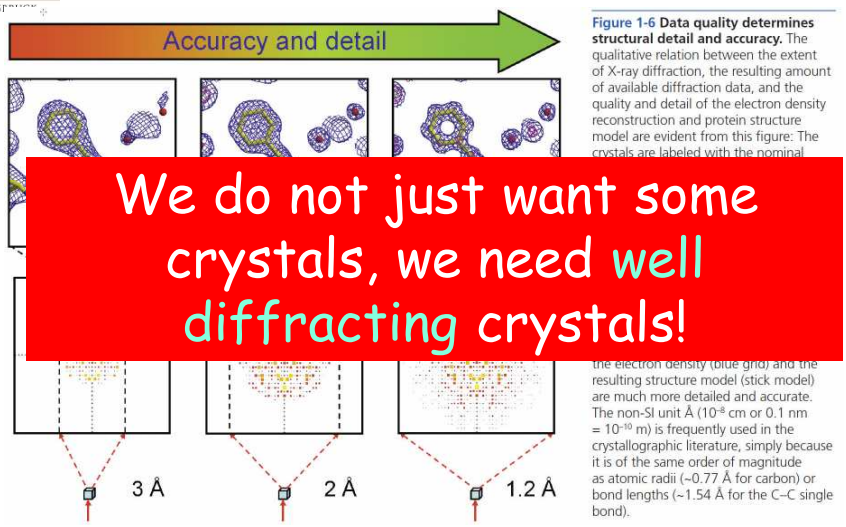


Figure 1-6 Data quality determines structural detail and accuracy. The qualitative relation between the extent of X-ray diffraction, the resulting amount of available diffraction data, and the quality and detail of the electron density reconstruction and protein structure model are evident from this figure: The crystals are labeled with the nominal

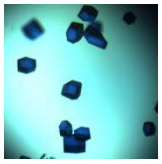
the electron density (blue grid) and the resulting structure model (stick model) are much more detailed and accurate. The non-SI unit Å (10^{-8} cm or 0.1 nm = 10^{-10} m) is frequently used in the crystallographic literature, simply because it is of the same order of magnitude as atomic radii (~0.77 Å for carbon) or bond lengths (~1.54 Å for the C–C single bond).

We do not just want some crystals, we need well diffracting crystals!

Holton Movie

Workflow of a protein crystallization project

Figure 3-3 Workflow of a crystallization project. The flow diagram shows the basic tasks and their relation and feedback in a crystallization project, starting at the protein level. Color indicates basic liquid handling (magenta); crystallization plate setup and handling (light blue); and mounting and data collation tasks (green).



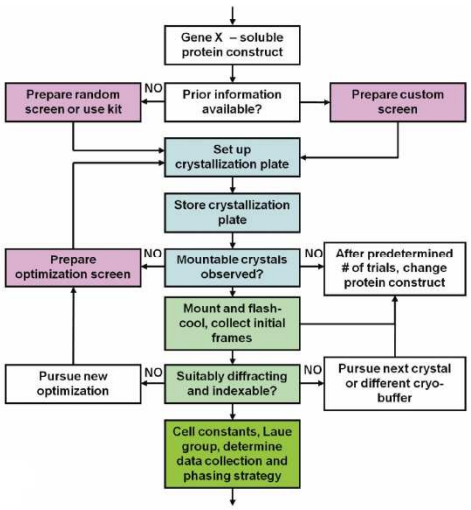


Figure 3-4 Crystals of tetragonal lysozyme. The figure shows tetragonal lysozyme crystals of varying quality, imaged from a hanging drop with a digital microscope camera (Figure 3-36). Hen egg white lysozyme is a hardy perennial of protein crystallization and widely used in practical demonstrations as well as in systematic crystallization studies. It is readily available and can be easily crystallized with common reagents (Sidebar 3-4). Note that some crystals are very well developed single crystals with sharp edges, while others are twinned and inter-grown in clusters. External appearance does not always correlate with diffraction quality. Perfectly formed crystals may not diffract, while unsightly fragments dissected from crystal clusters can produce excellent diffraction patterns.

TEST (no-one said it is going to be easy)

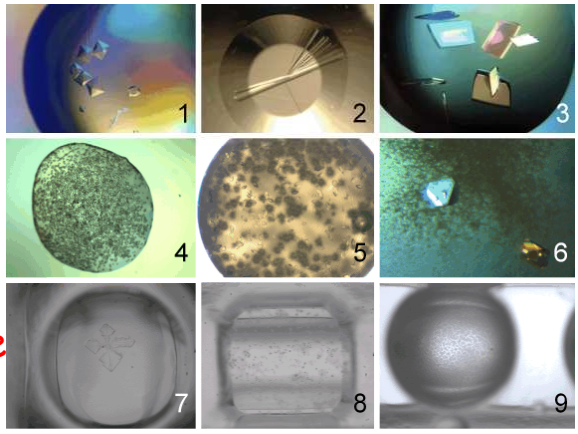
Name a few properties of crystals:

- Beautiful
- Hard
- Durable
- Precious

TEST (no-one said it is going to be easy)

Name a few properties of protein crystals:

- Beautiful
- Soft
- Fragile
- Sensitive
- Deceptive
- Unpredictable
- Even more precious…



What is a crystal – formal view

Definition: A 3-dimensional translationally periodic stacking of unit cells

A very useful mathematical concept for computation but it ignores the minutiae of protein self assembly and details of crystal contacts.

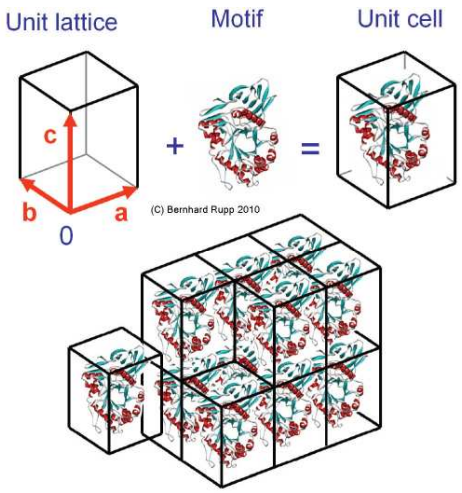


Figure 5-24 Assembly of a primitive triclinic 3-dimensional crystal from unit cells. In analogy to the 2-dimensional case, the unit lattice is filled with a motif, and the crystal is built from translationally stacked unit cells. The basis vectors form a right-handed system [0, **a**, **b**, **c**].

What is a protein crystal – biocrystallization view

Protein-protein contacts are mediated by weak and sparse, non-covalent intermolecular interactions

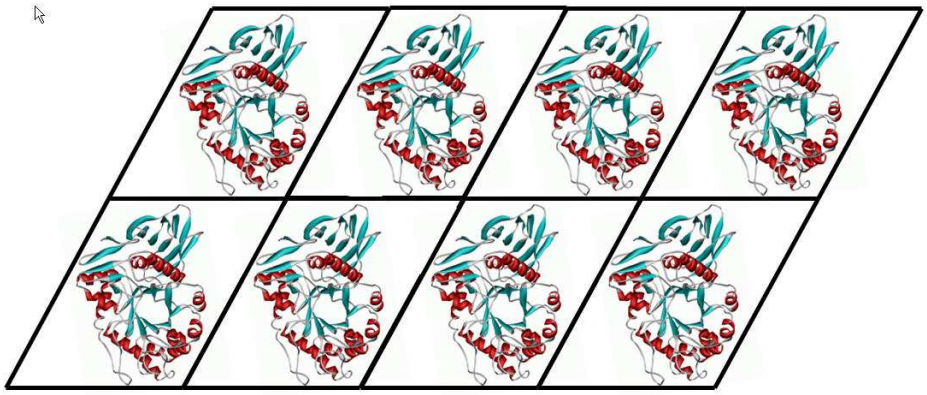
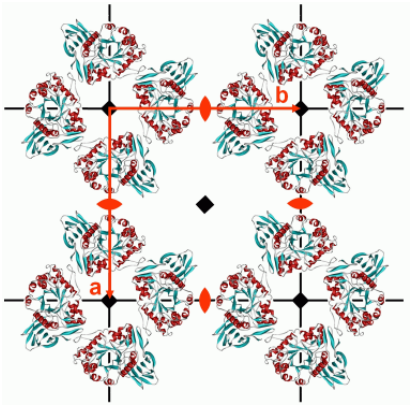


Figure 3-5 Protein crystals are formed by a sparse network of weak intermolecular interactions. The example shows protein molecules assembled into a primitive 2-dimensional lattice, connected by three different types (red, green, blue) of periodically repeating intermolecular interactions. The interactions are both sparse and weak, and as a consequence protein crystals are fragile and sensitive to mechanical stress and environmental changes.

Consequences of a crystal being a network of sparse, weak, and specific interactions

- Sensitive to mechanical stress
- Sensitive to environmental changes - ΔT, ΔpH, Δμ
- Contain large fraction of solvent
- Contain solvent channels important for ligand soaking



Ligand soaking and co-crystallization

Figure 3-38 Soaking of blue dye into lysozyme crystals. The crystals were dyed by adding 0.5 µl of methylene blue solution to a 2 µl drop, and imaged immediately after dye addition, after 1 h, and after 10 h. The solution becomes successively lighter while the crystals absorb nearly all of the dye. It is also important to realize that it takes quite a while even for small molecules to diffuse into a crystal. It is not possible to soak large ligands such as peptides into crystals in seconds and expect to observe any electron density.[27]

Note: diffusion is a slow process ! (movie)

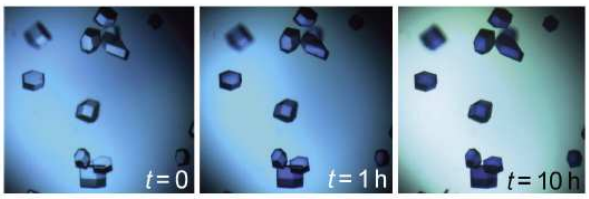
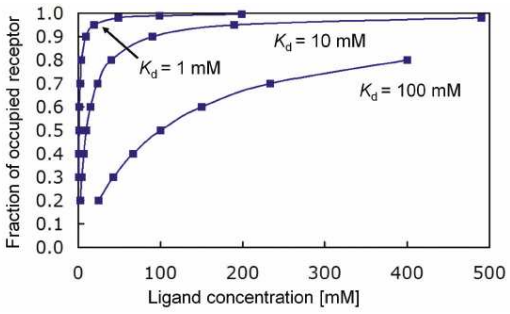


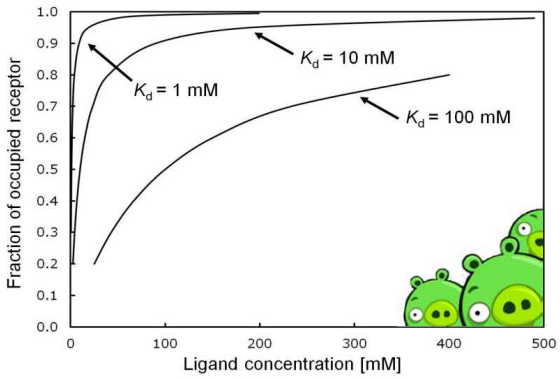
Figure 3-40 Fraction of occupied receptor sites against ligand equilibrium concentration for three different binding constants. Note that while at mM and lower K_d range small concentrations of ligand suffice to achieve good binding site occupation (between 70–90%), quite impractical concentrations of ligand in the crystallization drop are required for poor binders.



Binding sites will pick up everything they can from the cocktail!

Bindings sites suck (up stuff)

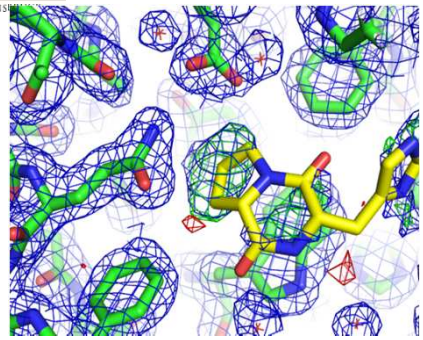
Out of principle, binding sites are never fully occupied:

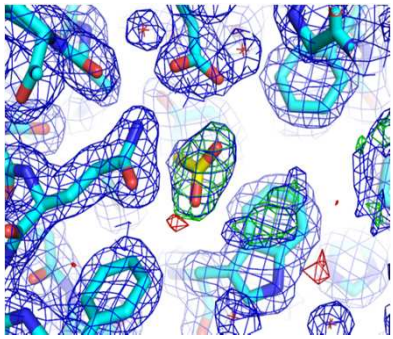


Fraction of occupied receptor sites plotted against ligand equilibrium concentration for three different binding constants. While at mM and lower K_d range small concentrations of ligand suffice to achieve reasonable binding site occupancy (between 70-90%), quite impractical concentrations of ligand in the crystallization drop are required for poor binders. On the other hand, given sufficiently high concentration, even weakly binding and non-native ligands can be forced into a binding site.

-> There is almost always some obscure density in sites that beckons to be filled with a ligand of desire

Ligands that are cocktail components

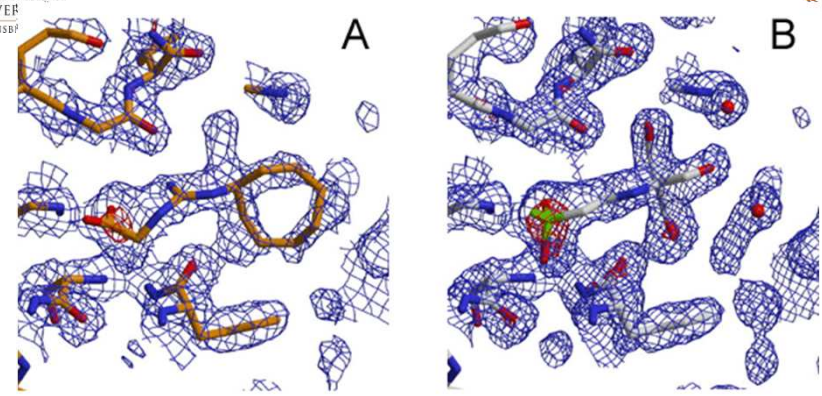




Ligands placed into mother liquor density, ligand omit maps. A: In the structure of the *B. cereus* chitinase, PDB entry 3n1a, (Hsieh *et al.*, 2010), the cyclo-(L-His-L_Pro) molecule (CHQ A1514) is placed into low level electron density that is difficult to interpret, and which may be plausibly interpreted as an acetate molecule present in crystallization cocktail at 200 mM.

Very tempting and very common – check your imagination!

Binding sites want to bind – anything they can



TES buffer in ligand binding site. 2.1 Å maps contoured at 1σ (blue) and 5σ (red). (A) presumed ligand built into *CNS* ML $2mF_o$-DF_c map; (B) *Shake&wARP* map, with TES buffer built into density. Map has less noise and cleaner connectivity and reveals the true nature of the ligand. A questionable VdW contact is also obvious between 'ligand' and protein in the left panel (A).

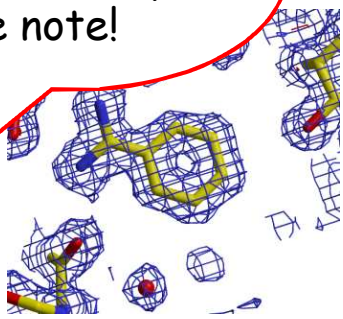
Lack of supervision and training may often be responsible!

Why ligand-models may be dangerous to your career

1. Global indicators of (reciprocal space) data fit like R-values are completely insensitive. Ligand scattering mass is often only 1/1000 of th… …ine this with high B-factors and p… becomes even worse. Dit… …nil.

2. Therefore we need loca… space) indicators that show th… fit between model and electron density. The electron density – preferably minimally biased positive omit difference density - is the primary evidence !

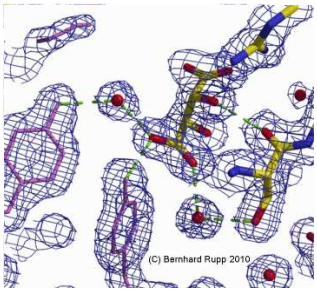
(Poster) presenters please take note!

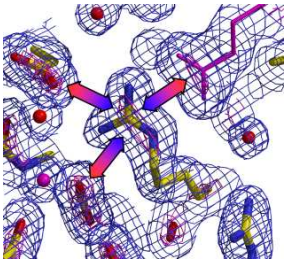


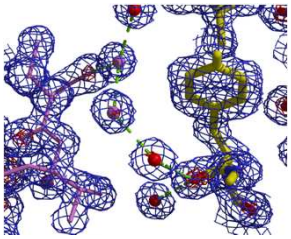
Interactions between protein molecules

Preciously few weak interactions that must be in the right place for self-assembly into in a fragile protein crystal

- Hydrogen bonds
- Salt bridges (charged interactions)
- Polar & hydrophobic interactions
- VdW contacts
- Solvent mediated
- Often combinations of above!

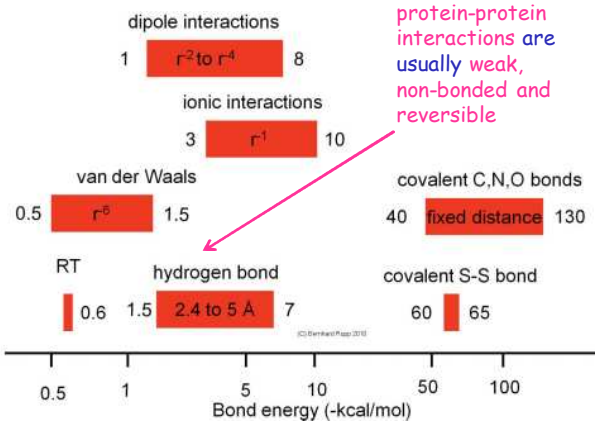






Energy range of non-bonded interactions

Figure 2-29 Typical ranges for bond energies of side chain interactions. RT denotes the thermal energy at room temperature (293 K). Note the logarithmic energy scale. The numbers in the red boxes give the approximate radial distance dependence for charged and polar interactions, and an approximate interaction range for the directionally dependent hydrogen bonds. The numbers left and right of the red boxes flank the approximate bond energy range. In contrast to the weak non-covalent interactions, covalent bonds have specific and discrete bond distances and bond angles.



protein-protein interactions are usually weak, non-bonded and reversible

Conclusion: Small changes in environment can have significant impact on protein crystallization !

Extreme sensitivity to pH is no surprise

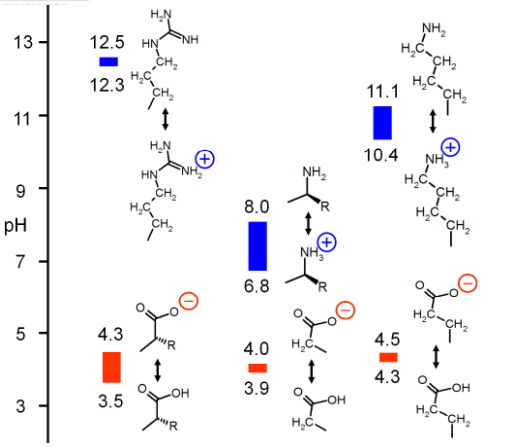


Figure 2-28 Logarithmic acid dissociation constant (pK_a) of charged amino acids. The figure provides a quick overview of the normal range of pK_a values for charged amino acid side chains and of amino- and carboxy-terminal groups. Unusual and extreme local environments can further shift these normal ranges to atypical values.

note the highly variable protonation state of His at physiological pH

Change of pH shifts the surface charge distribution and thus the intermolecular interactions – pH is therefore a most significant factor for protein crystallization

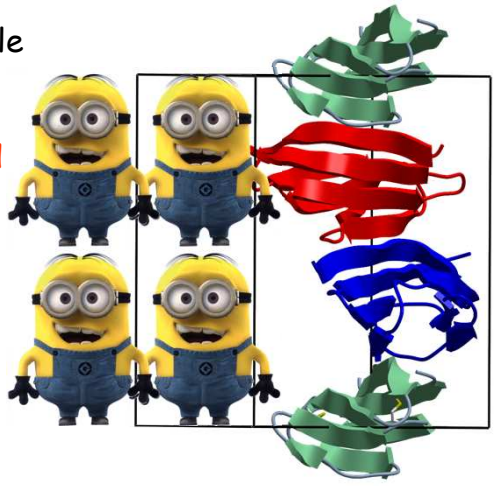
First invariable conclusion :

If there are no suitable contacts at the **right surface locations** to form an **adequate 3-d network**, then

NO MATTER WHAT

that protein will **not** crystallize.

Amen.

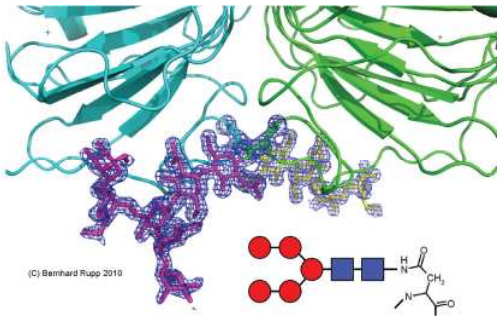


Crystallization is not the first challenge

It's the protein – (i) getting a soluble protein in the first place – and (ii) one that is inherently crystallizable

- Expression problems with eukaryotic proteins
- Disulfide links – host selection (extracellular or secreted)
- PTMs and decorations – conformationally inhomogeneous – but functionally and sometimes structurally important (TCR, QC)

Figure 4-20 Glycosylated residues stabilizing homodimer. In quercetin 2,3-dioxygenase, a copper-containing enzyme that catalyzes the insertion of molecular oxygen into polyphenolic flavonols, a biantennary (Man_3)-Man-GlcNAc-GlcNAc moiety (red circles mannose, blue squares N-acetylglucosamine) is covalently linked to Asn191 and forms a stable dimer contact. Enzymatic deglycosylation with EndoH was necessary to obtain diffraction quality crystals. Although EndoH truncates exposed oligosaccharides after the first GlcNAc, the biantennary sugar at Asn191 remained intact in contrast to the remaining four other oligosaccharide decorations. Image from PDB entry 1juh[112] courtesy of Roberto Steiner, Kings College London.



Recombinant DNA techniques in protein crystallography

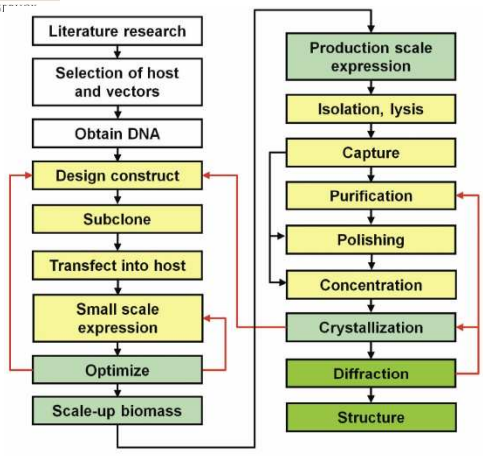


Figure 4-1 Recombinant DNA technologies allow efficient protein expression, purification, and crystallization. One of the most significant advantages of recombinant DNA techniques is the ability to add affinity capture tags which allow rapid separation and concentration of the tagged protein in a single step. In fortunate cases further steps such as purification, polishing, and concentration can be bypassed and the protein directly screened for crystallization. More often than not, further modification to the protocol and ultimately the protein construct will be necessary (red arrows). In summary: • Study all available literature • select host(s) and expression vector(s); • obtain DNA with the open reading frame (ORF) of interest; • design the initial expression constructs; • subclone the cDNA if needed; • introduce the recombinant DNA into the chosen host(s); • perform small-scale expression trials; • analyze the small-scale trials and select optimum expression parameters; • scale up biomass production; • extract the protein of interest from lysed cells or from the expression media; • purify the protein; • condition the protein for crystallization; • perform crystallization and cycle back (red arrows) to purification, expression, or cloning with redesigned construct as needed.

Even very rare proteins with a defined sequence can be overexpressed in chemically (but not necessarily conformationally) pure form

Recombinant DNA techniques in protein crystallography

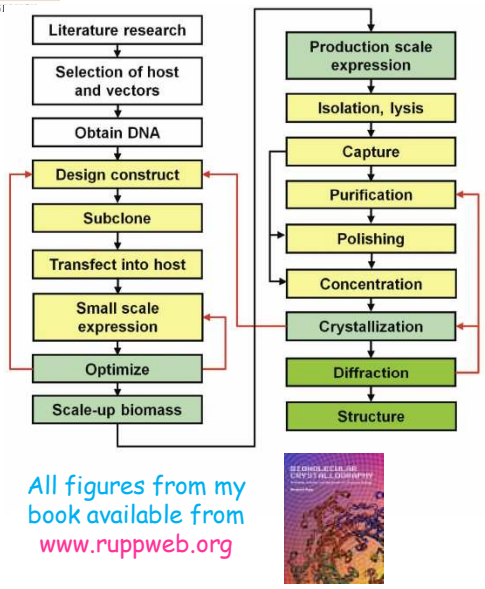
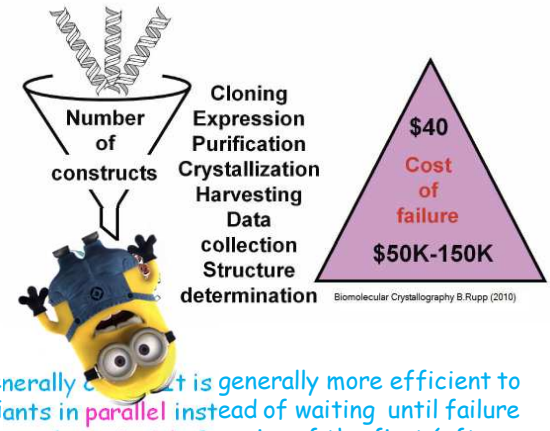


Figure 4-1 Recombinant DNA technologies allow efficient protein expression, purification, and crystallization. One of the most significant advantages of recombinant DNA techniques is the ability to add affinity capture tags which allow rapid separation and concentration of the tagged protein in a single step. In fortunate cases further steps such as purification, polishing, and concentration can be bypassed and the protein directly screened for crystallization. More often than not, further modification to the protocol and ultimately the protein construct will be necessary (red arrows). In summary: • Study all available literature • select host(s) and expression vector(s); • obtain DNA with the open reading frame (ORF) of interest; • design the initial expression constructs; • subclone the cDNA if needed; • introduce the recombinant DNA into the chosen host(s); • perform small-scale expression trials; • analyze the small-scale trials and select optimum expression parameters; • scale up biomass production; • extract the protein of interest from lysed cells or from the expression media; • purify the protein; • condition the protein for crystallization; • perform crystallization and cycle back (red arrows) to purification, expression, or cloning with redesigned construct as needed.

All figures from my book available from www.ruppweb.org

Parallel vs. serial approaches

Figure 1-5 Keeping the funnel full. Given the high attrition rate in early experimental steps of a protein structure determination project, a parallel approach pursuing multiple variants of the protein target significantly increases the chances of success. Parallel approaches are also less susceptible to late stage failures, because the pressure to pursue one single protein construct that already shows warning signs to the bitter end is diminished. The most expensive failure is a perfectly good diffraction data set that cannot be phased, because all the costly and time-consuming laboratory experiments have been conducted already. Developing a suitable phasing strategy early on is thus part of good study planning.



Early stage failures are generally c t is generally more efficient to pursue multiple target variants in parallel instead of waiting until failure and then repeating the procedure. Avoids focusing of the first (often marginal) success until ultimate failure.

Methods for protein modification

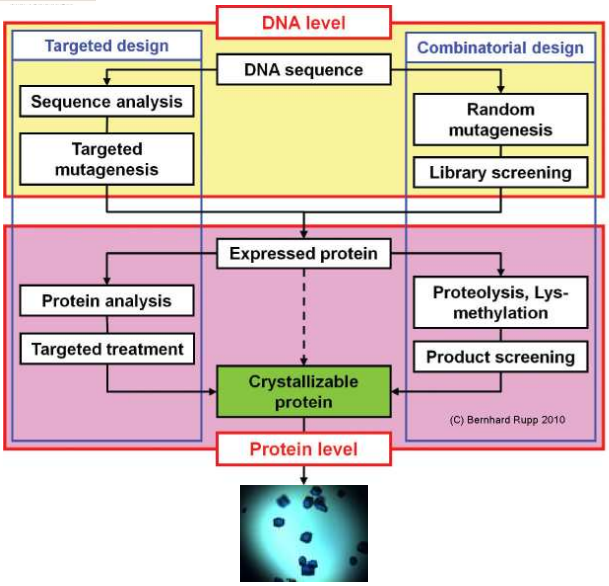


Figure 4-2 Conceptual overview of design strategies toward protein engineering for crystallization. Protein engineering can occur at the DNA level and the protein level. Targeted experimental designs as well as combinatorial designs can be applied to either level. Often several approaches are combined and pursued in a parallel strategy.

Sidebar 4-1 Common protein modification techniques.

Biochemical/enzymatic:
- Proteolysis
- Deglycosylation
- Phosphorylation/ dephosphorylation

Chemical:
- Methylation of primary amines
- Reduction/oxidation of disulfides or thiols
- Modification of exposed thiols (e.g., with iodoacetamide)
- Selective derivatization with heavy atom reagents (e.g., iodination)

Domain truncation

Independently folding domains of proteins can often be separately expressed and crystallized – example bitopic membrane protein EGF-R

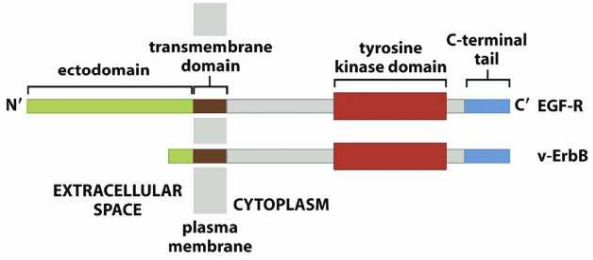


Figure 2-34 Domain structure of the epidermal growth factor receptor (EGF-R) and the related mutant oncoprotein v-ErbB. Both the growth factor binding ectodomain of the EGF-R as well as the tyrosine kinase (TK) domains have been separately expressed and the X-ray structures determined. Normal signal transduction occurs via dimerization of receptors upon ligand binding and subsequent transphosphorylation, where one receptor TK domain phosphorylates the other dimer partner. The oncogenic v-ErbB receptor lacks a complete ectodomain but causes persistent mitogenic signaling independent of ligand binding. See also Sidebar 5-2 discussing the role of dimerization. With permission from reference 29 (Garland Science).

Predicting where to cut domains

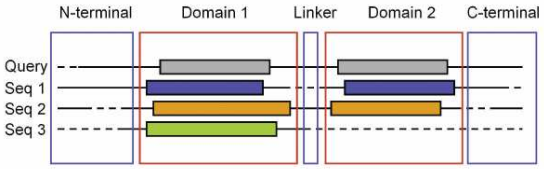


Figure 4-3 Schematic of alignment suggesting possible strategies for protein engineering. A multiple sequence alignment often reveals the domain organization of a protein of unknown 3-dimensional structure. Query represents the target protein, Seq the similar sequences from the data base. Domains generally show a higher degree of sequence conservation, while loops and termini show more variation in sequence as well as in length. The alignment with a shorter, presumably single domain protein as shown in Seq 3 further supports the proposed domain structure (the dashed lines represent alignment gaps or missing residues). Secondary structure predictions can further delineate the domain boundaries. Expression levels and stability can be affected by as much as a single residue truncation or retention, and multiple constructs are thus generally pursued.

A single residue more or less can make all the difference

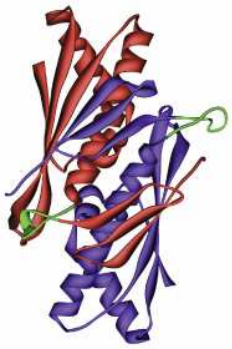


Figure 4-4 Domain swapped dimer. The domain swapped dimer of *Bacillus subtilis* organic hydroperoxide-resistance protein (OhrB) serves as an example where domain truncation would not be effective. Truncating the protein sequence at the disordered loops (green) would likely lead to loss of structural integrity because a part of the β sheet of the domain structure would be missing. These loops, however, remain viable targets for site directed mutagenesis (see Figure 4-6). PDB entry 2bjo.[28]

Fusion proteins and fusion tags

Most frequently used tags are His-6 tags for IMAC (Immobilized metal-ion affinity) capture

Figure 4-12 Generic design of a tagged fusion construct. The protein construct contains an affinity tag for easy affinity capture with a shorter linker to a solubility enhancing fusion protein finally linked to the target sequence. The cleavage site is located right before the N-terminus of the target sequence to leave as few residues as possible behind. The Gateway p-DEST-HisMBP vector is a typical example,[81] with the TEV cleavage site providing for simple IMAC purification when a His-tagged TEV protease is used (Figure 4-13).

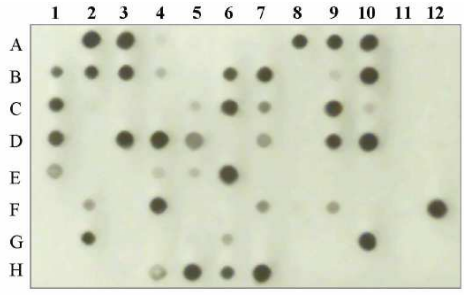


Figure 4-13 His-tag antibody Western blot of C-terminally His_6-tagged proteins. A stained immuno-blot of lysates from a small scale, parallel expression experiment in 96-well format allows rapid identification of colonies that have overexpressed a large amount of full length constructs (intense dark spots). Figure reproduced[46] with permission from Springer Verlag.

Fusion proteins and fusion tags

Most frequently used tags are His-6 tags for IMAC (Immobilized metal-ion affinity) capture

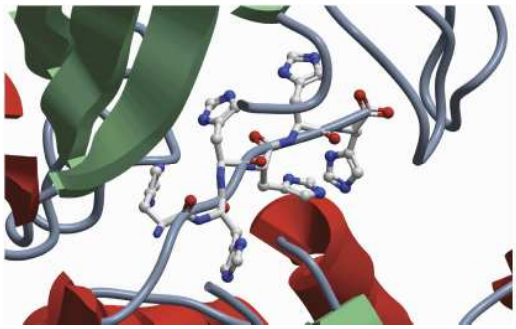
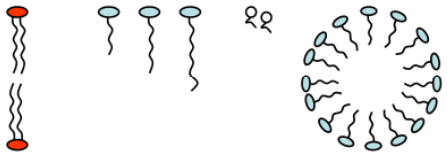


Figure 4-12 Generic design of a tagged fusion construct. The protein construct contains an affinity tag for easy affinity capture with a shorter linker to a solubility enhancing fusion protein finally linked to the target sequence. The cleavage site is located right before the N-terminus of the target sequence to leave as few residues as possible behind. The Gateway p-DEST-HisMBP vector is a typical example,[81] with the TEV cleavage site providing for simple IMAC purification when a His-tagged TEV protease is used (Figure 4-13).

Figure 3-6 C-terminal histidine affinity purification tag participating in crystal contacts. The C-terminal His_6-tag of the molecule in the lower part of the figure interacts with the symmetry related copy shown in the top part of the figure. Not shown are additional solvent molecules that participate in an intricate network of intermolecular contacts. PDB entry 2bv9 and untagged molecule 2bvn.[83]

Membrane protein expression

Figure 4-17 Phospholipid, detergents, amphiphiles, and detergent micelle. Phospholipids comprise the majority of the fluid bilayer cellular membrane. A glycerol molecule is linked to two C16 to C18 fatty acids, forming the hydrophobic tail. The third position of the core glycerol is linked to a phosphate group and choline or other polar groups symbolized by the red oval. Detergents similarly have a polar group (blue oval) and shorter hydrophobic tails of varying length. Mild ionic, zwitterionic, and non-ionic detergents are used in membrane protein crystallization. Amphiphiles (white head groups) are small detergent-like molecules that can be further used to adjust the size of the detergent collar around the transmembrane stem of a membrane protein. With increasing concentration, detergents form spherical micelles (right panel). The concentration at which micelles and free detergent are in equilibrium is called the critical micelle concentration or CMC.



Hydrophobic trans-membrane stem needs protection

Detergents are added for solubilization and crystallization

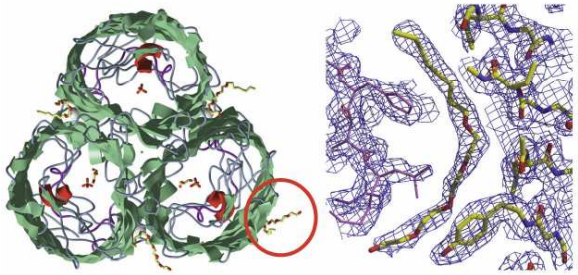


Figure 4-18 The trimeric bacterial outer membrane protein OprP. Published in 2007, the OprP structure revealed a new structural motif in bacterial membrane proteins, a 9-residue arginine ladder. OprP controls the transport of essential phosphate anions into the pathogenic bacterium *Pseudomonas aeruginosa*. The Arg ladder extends from the extracellular surface down to a constriction zone where a phosphate anion is bound (visible in the center of the channels). Note that the crystal structure again contains detergent molecules along the transmembrane barrel, in this case hydroxyethyloxytri(ethyloxy)octane or C8E. The right panel shows the region around the circled detergent molecule in electron density. In this position, the detergent in addition mediates a crystal contact to a symmetry related molecule (magenta). PDB entry 2o4v.[105]

Fusions for transport and membrane insertion

Sidebar 4-8 Mystic Mistic. The recent discovery of Mistic (Membrane Integrating Sequence for Translation of Integral membrane protein Constructs) has opened another promising avenue for membrane protein expression. Mistic is an unusual *Bacillus subtilis* dual-topology protein that folds and inserts autonomously into the membrane forming an integral four-helix bundle membrane protein (Figure 4-16).[100] In addition, Mistic lacks any signal sequence and bypasses the cellular translocon complex, and therefore provides a highly suitable fusion partner for expression of eukaryotic membrane proteins.

The auto-inserting ability of Mistic together with the bypassing of the cellular translocon has raised the probability that more structures of crucial membrane proteins of therapeutic interest, such as ion channels and 7-TM G-protein coupled receptors (GPCRs), will become available soon. More than 1000 GPCRs exist, which are involved in regulation of a wide variety of biological functions, and many of them are drug targets. Another avenue to expression of GPCRs via antibody scaffolding[101] was outlined in Chapter 3, and the scaffolding via the insertion of a T4-lysozyme mutant into a flexible loop of the GPCR structure.[102]

Similar to the idea of using Mistic as a transport and integration fusion, members of the bacterial Tol proteins[103] export a wide range of molecules across the periplasmic space and outer membrane of Gram-negative bacteria, and might provide a further avenue for membrane protein expression in bacteria. An engineered version of the anti-apoptotic Bcl-2 family protein Bcl-XL has also been used in a fusion system for membrane proteins.[104]

A good resource for membrane protein production and structure is Steven White's web page: http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html

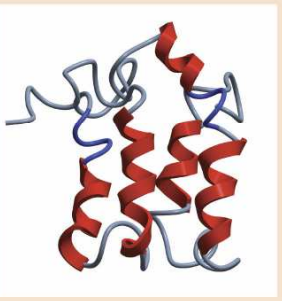
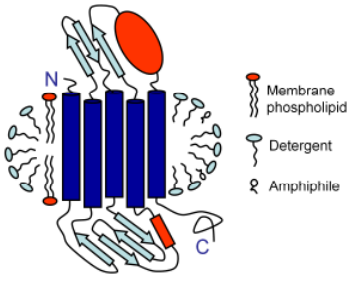


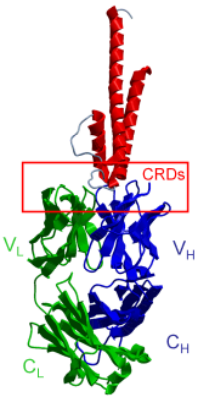
Figure 4-16 NMR solution structure of Mistic. The ribbon diagram shows the lowest energy conformer of the irregular four-helix bundle of Mistic. The lipid facing surfaces are unusually polar, and the presence of a concentric ring of solubilizing LDAO detergent molecules was deduced from NMR NOE interactions. PDB entry 1ygm.[100]

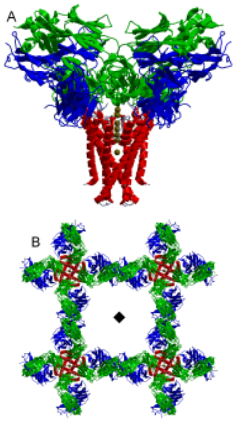
Scaffolding as general method

* It's the protein – (i) getting a protein in the first place – and (ii) one that is inherently crystallizable

- Membrane proteins, receptors - scaffolding







140. Zhou Y, Morais-Cabral JH, Kaufman A, et al. (2001) Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 A resolution. *Nature* 414(6859), 43-48.

Protein-DNA complex crystallization

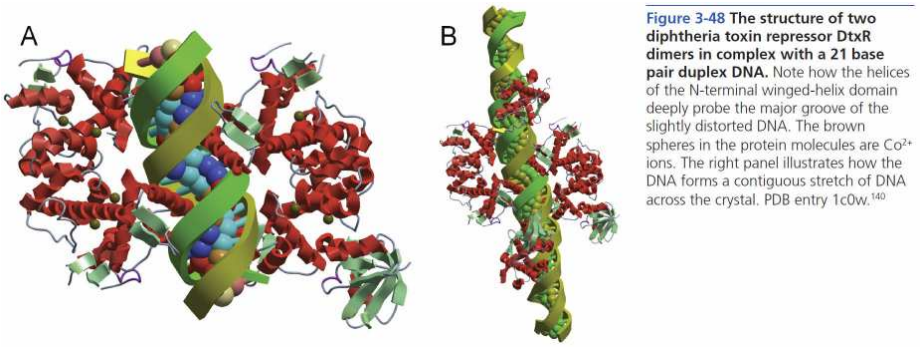


Figure 3-48 The structure of two diphtheria toxin repressor DtxR dimers in complex with a 21 base pair duplex DNA. Note how the helices of the N-terminal winged-helix domain deeply probe the major groove of the slightly distorted DNA. The brown spheres in the protein molecules are Co^{2+} ions. The right panel illustrates how the DNA forms a contiguous stretch of DNA across the crystal. PDB entry 1c0w.[140]

Almost always the length of the DNA oligomer is a decisive factor for protein-DNA complex crystallization.

What defines a stable protein crystal ?

Protein-protein contacts are mediated by weak and sparse, non-bonded intermolecular interactions

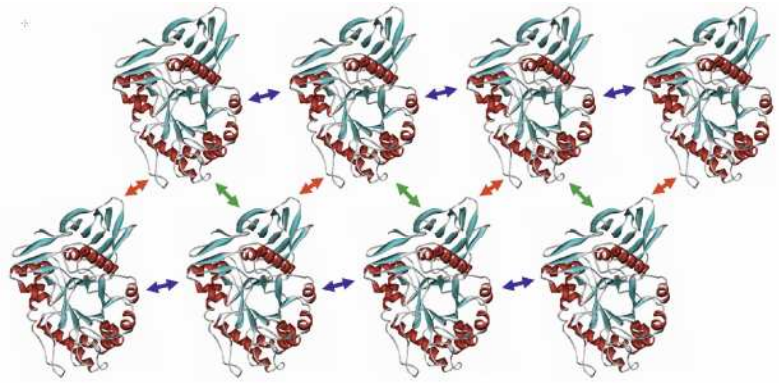
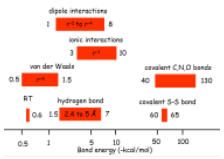


Figure 3-5 Protein crystals are formed by a sparse network of weak intermolecular interactions. The example shows protein molecules assembled into a primitive 2-dimensional lattice, connected by three different types (red, green, blue) of periodically repeating intermolecular interactions. The interactions are both sparse and weak, and as a consequence protein crystals are fragile and sensitive to mechanical stress and environmental changes.

Determining factors for stability:

A) How many contacts and B) how strong ?

A) ~ 15 contacts/molecule
B) contact surface area 100-500 Å^2
obligate dimers: ~ 800 Å^2 and up
in between gray area



Free energy of crystallization

$$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$$

Not much — Decisive term

Crystallization is strongly entropy driven !
rationale for surface (entropy) engineering

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 protein solution must be in or move into a metastable, supersaturated 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**

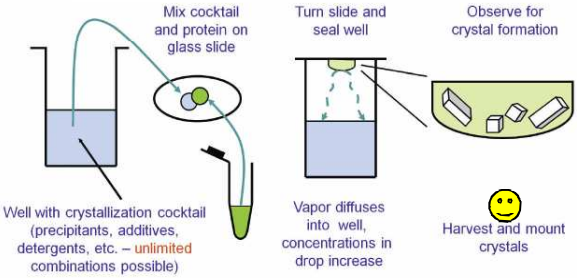
Only if (A and B) then *prob* (C)

What is protein crystallization and how-to?

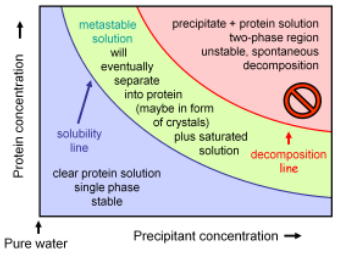
Protein crystallization is a special case of phase separation forming a protein rich phase from thermodynamically metastable (supersaturated) solution under the control of kinetic parameters

Figure 3-1 Basic hanging-drop vapor diffusion. Hanging-drop vapor diffusion has been in use for over 30 years for the manual setup of protein crystallization. The reservoir (generally one well of a multi-well assay plate) is partially filled with several hundred µl of crystallization cocktail. A small drop (a few µl or less) of this cocktail is set in the center of a siliconized cover slide, and mixed there with an equal volume of protein stock solution (green). The cover slide is then turned over and placed on the greased rim of the reservoir well. The mixing with protein has reduced the precipitant cocktail concentration to half of the original value, and the sealed system thus equilibrates by water vapor diffusion from the drop into the reservoir solution, thus effectively increasing the concentration of all constituents (protein and precipitation cocktail reagents) in the crystallization drop. During this process the drop becomes supersaturated, nucleation can occur, and protein crystals may grow from the supersaturated solution.

Figure 3-2 Crystallization plates. A 24-well Linbro plate used for a manual hanging-drop crystallization setup is shown to the left. Originally designed for cell culture work, the Linbro plates made from polystyrene are cheap and commercially available with pre-greased rims. The wells are covered with siliconized 22 mm circular cover slides. Although not suitable for automated work, the hanging-drop method using Linbro plates is still useful in small laboratories. The crystallization plate to the right is a typical 96-well sitting-drop plate in the smaller, standardized high-throughput SBS (Society for Biomolecular Sciences) format (Intelliplate, Art Robbins Instruments). Crystallization robotics most frequently use sitting-drop plates with a SBS footprint.



How to: crystallization techniques



How do different crystallization methods traverse the crystallization phase space?

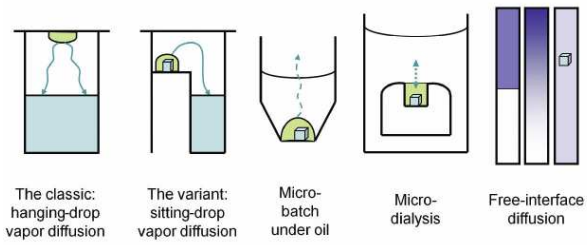
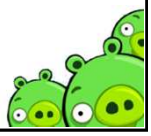


Figure 3-19 Schematic sketches of popular crystallization techniques. Hanging-drop vapor diffusion is a common method used in small-scale manual setups while sitting-drop vapor diffusion is preferred with robotic setups. The absence of additional sealing requirements and the ease of miniaturization favors automated microbatch screening under oil, although harvesting tends to be more difficult. Use of silicone oils in microbatch wells allows partial exchange of solvent vapor (indicated by the broken arrow). Microdialysis is hard to miniaturize but can be used to grow very large crystals. Miniaturized free-interface diffusion screening chips are gaining popularity, but automation and harvesting issues remain to be resolved. Each method traverses the crystallization phase space in a different path and the same chemical screening conditions can produce widely varying results.

Basic crystallization (phase) diagram

'pseudobinary phase diagram'

Playground for Imagination



Concentration of Protein

'Water corner'

Reagent (precipitant) concentration

Only few things in the crystallization (phase) diagram we do know:

unstable 2-phase region

$$\left(\frac{\partial^2 G}{\partial x_i^{\,2}}\right)_{P,T} < 0$$

meta-stable region

spontaneous (homogeneous) nucleation 'zone'

heterogeneous nucleation 'zone'

single phase protein solution

$$\left(\frac{\partial^2 G}{\partial x_i^{\,2}}\right)_{P,T} > 0$$

Concentration of Protein

'Water corner'

Reagent (precipitant) concentration

Reducing solubility with precipitants

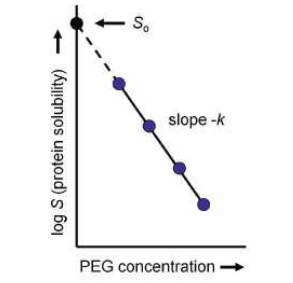


Figure 3-14 Solubility of proteins in PEG. The solubility of proteins in polyethylene glycols decreases exponentially with increasing PEG concentration. A plot of the logarithmic solubility log (*S*) against PEG concentration is therefore a linear function. S_0 is the extrapolated (but not necessarily achievable) protein solubility at zero PEG concentration.

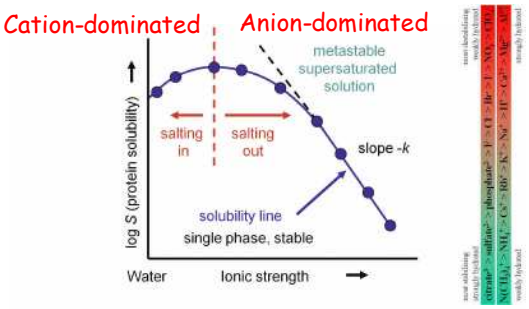


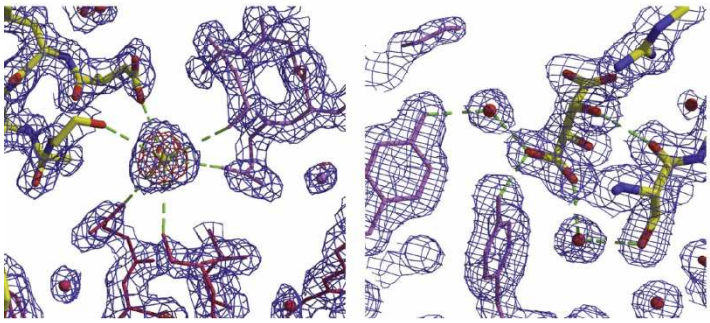
Figure 3-13 Solubility diagram typical for a protein in an ionic solvent. The high ionic strength (concentration) range can be described with a linearized logarithmic law, while the behavior in the low salt or low ionic strength region is more complex. We can distinguish a region where the solubility increases with small salt additions (salting in) and the salting-out region at higher concentrations.

Polyethylene glycols (PEGs) and salts are the primary precipitants

Using additives to modify or mediate crystal contacts

Metal ions, polydentate ligands, detergents, salts, PEGS, all can act as additives and modify or mediate crystal contacts.

Figure 3-16 Additives mediating intermolecular contacts. The left panel shows a Cd^{2+} ion in ferritin bridging three symmetry related molecules. Note that the Cd^{2+} ion is located on a threefold crystallographic axis. The right structure shows a tartrate molecule connecting two molecules of the sweet-tasting protein thaumatin from the African berry *Thaumatococcus daniellii*. Symmetry related molecules are shown in magenta and red. PDB entries 1aew[49] and 1thw.[50]



pH and protein solubility

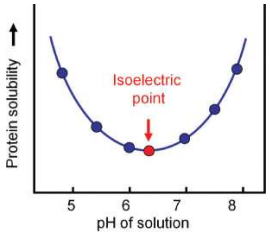


Figure 3-8 Protein solubility versus pH of protein solution. The protein shown in this example has its solubility minimum at its isoelectric point of ~6.3, where the sum of positive and negative charges (the net charge of the protein) is zero. Even at the isoelectric point, there are still numerous (but net compensating) local charges present on the surface of the protein.

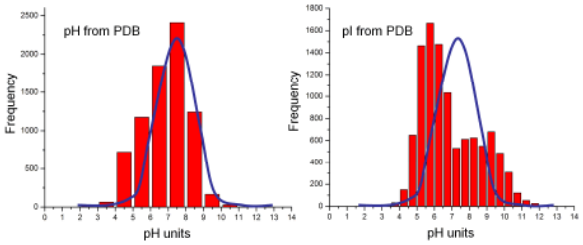


Figure 3-17 Distribution of crystallization pH and distribution of isoelectric point of proteins. The blue curve shows that the peak of the pH distribution around 7.4 falls directly into the gap between the modes of the bimodal pI distribution (right panel). Detailed pair-wise analysis has shown that acidic proteins prefer to crystallize above their pI and basic ones below their pI. The sum of these binned pH distribution gives the resulting overall distribution shown in the left panel.

pI is not the most common pH of crystallization. Some net charge seems favorable – still a local distribution of contacts

Protein solubility and nucleation regimes

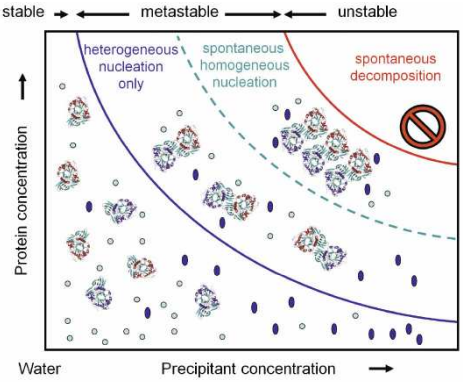


Figure 3-10 The location of nucleation zones in a protein crystallization diagram. Precipitant molecules are represented as dark blue ovals and water molecules as light blue circles. As a rule of thumb, higher supersaturation is necessary for spontaneous formation of stable crystallization nuclei (homogeneous nucleation), while at low supersaturation nucleation requires external seeds in the form of microcrystals or other particulate matter (heterogeneous nucleation). The zone of homogeneous (spontaneous) nucleation is occasionally referred to as the "labile" zone, but this term should be avoided because it leads to confusion with the different concept of labile or neutral equilibrium used in physical chemistry and thermodynamics.

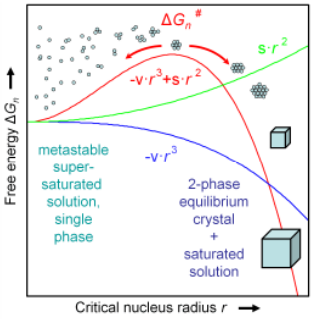


Figure 3-9 Nucleation energy. To achieve crystallization, nucleation must overcome the kinetic barrier that exists for phase separation (crystallization) from the metastable solution (left). At the critical size, it is equally likely for the nucleus (symbolized by the small aggregate) to fall apart again (left red arrow) or to grow into a crystal (right red arrow). Once a nucleus above a critical size defined by the critical free energy of nucleation $\Delta G_n^{\#}$ is formed, additional gain of binding enthalpy overcomes entropic loss during crystal growth, and the system can proceed towards its 2-phase equilibrium state (right side of image).

Improving nucleation: seeding

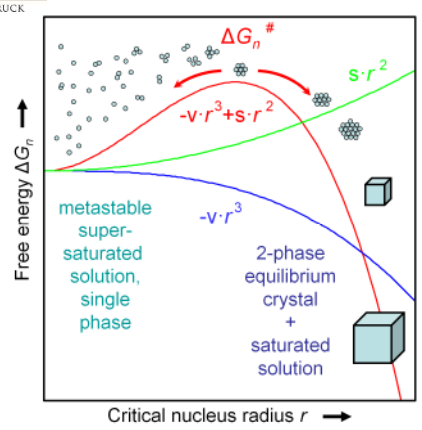


Figure 3-9 Nucleation energy. To achieve crystallization, nucleation must overcome the kinetic barrier that exists for phase separation (crystallization) from the metastable solution (left). At the critical size, it is equally likely for the nucleus (symbolized by the small aggregate) to fall apart again (left red arrow) or to grow into a crystal (right red arrow). Once a nucleus above a critical size defined by the critical free energy of nucleation $\Delta G_n^{\#}$ is formed, additional gain of binding enthalpy overcomes entropic loss during crystal growth, and the system can proceed towards its 2-phase equilibrium state (right side of image).

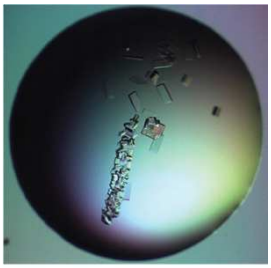
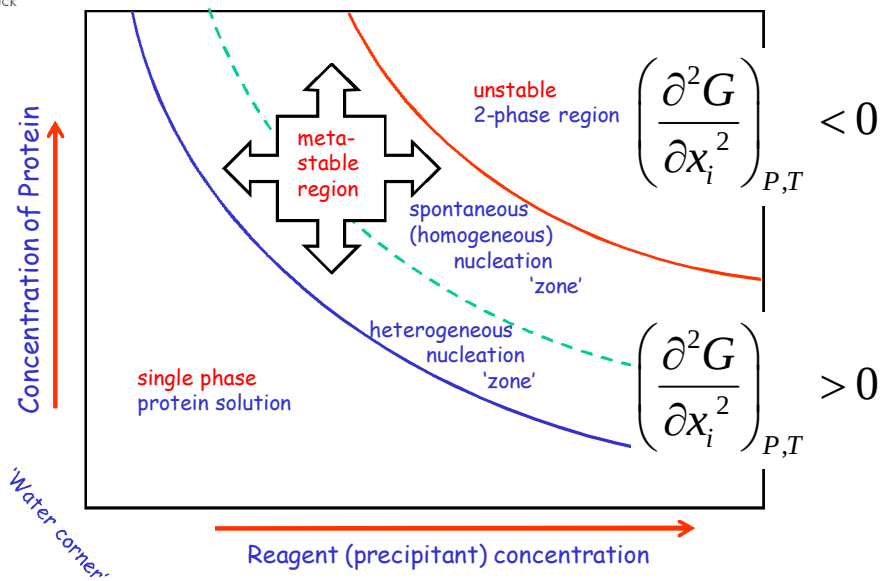
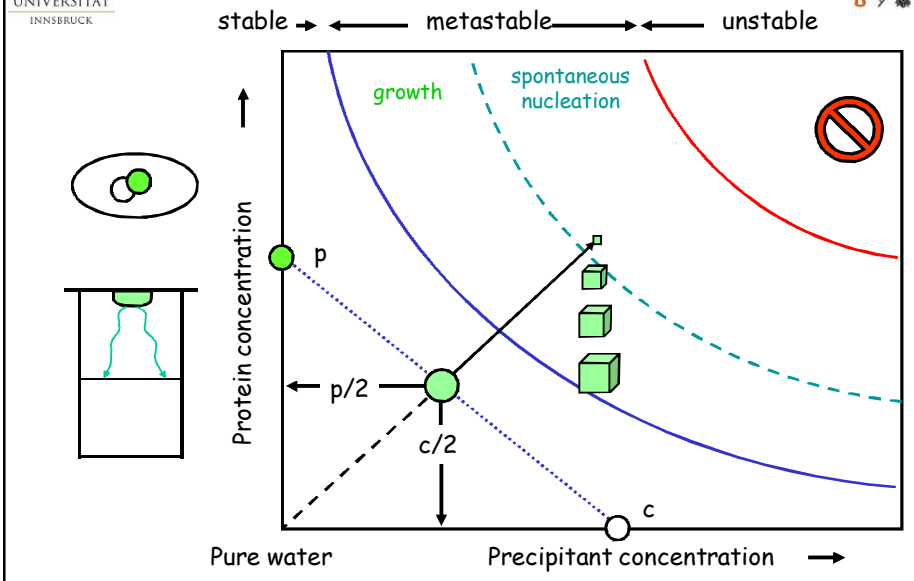


Figure 3-32 Crystals grown along the line of a seed streak. In microseeding, tiny microcrystals, crushed crystals, or other fine particulate matter are used to induce nucleation. The microcrystals or seeds are easily introduced using a whisker that is streaked or swiped through the crystallization drop, giving the name *streak seeding* to this technique. Image courtesy of Andrea Schmidt, EMBL Outstation Hamburg.

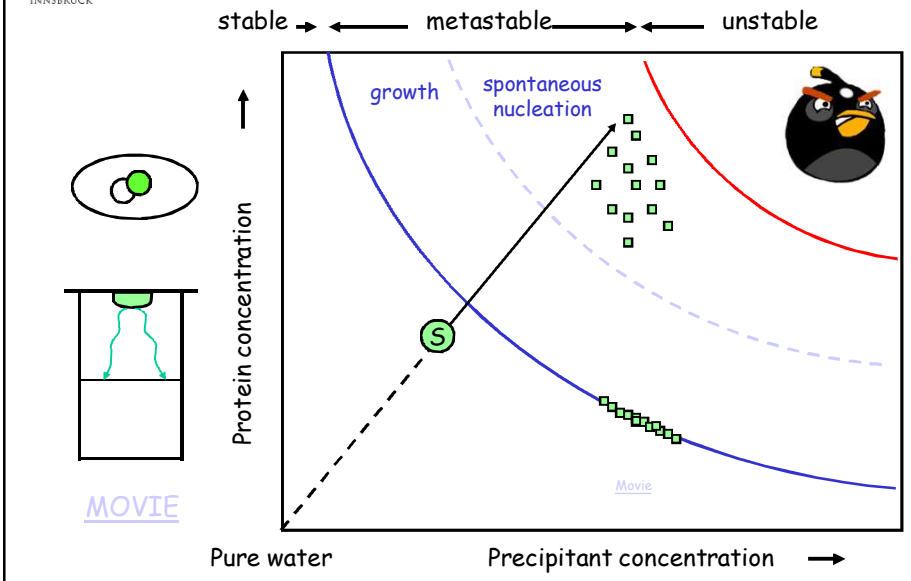
Only few things in the crystallization (phase) diagram we do know:



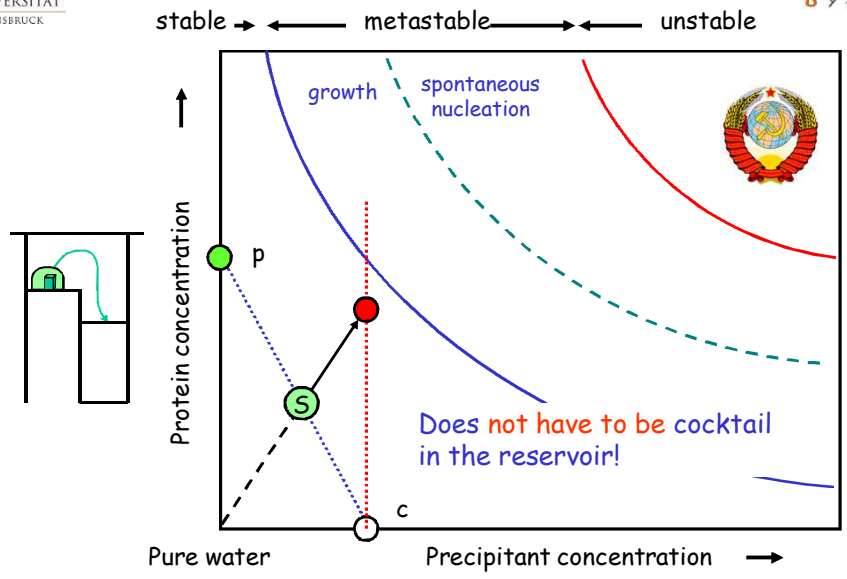
Pathway of Vapor Diffusion experiment I



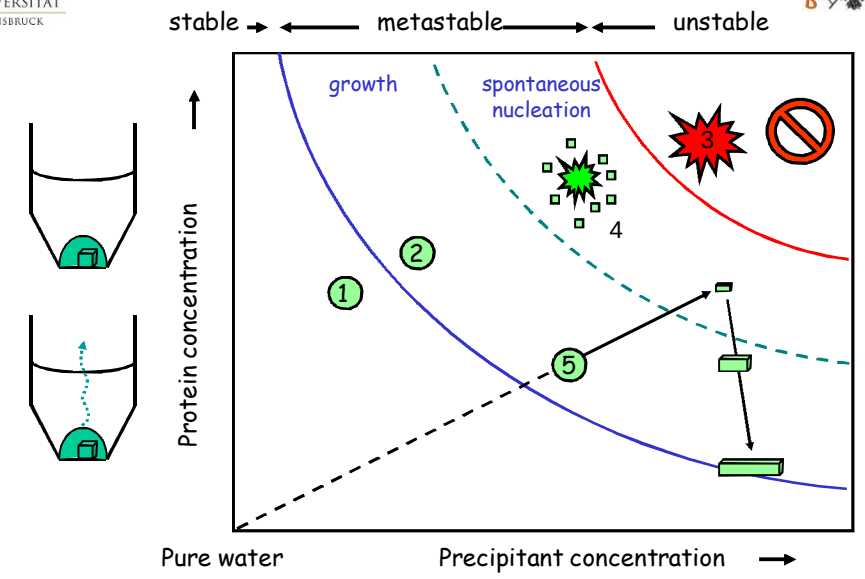
Pathway of Vapor Diffusion experiment II



Pathway of Vapor Diffusion experiment III



Pathway of batch crystallization experiments



Free interface diffusion

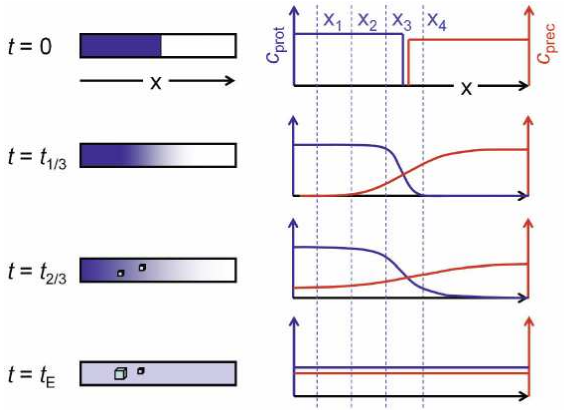
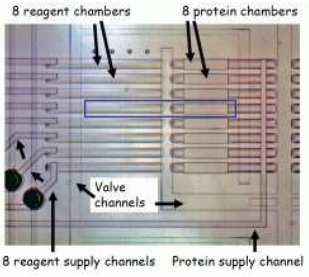
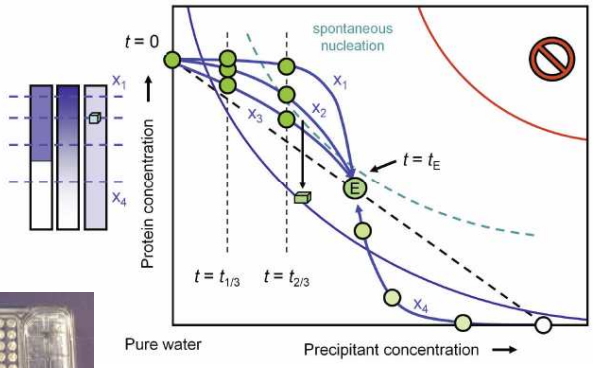


Figure 3-25 The time-development of the concentration profiles in a free-interface diffusion experiment. The left panel depicts the experimental setup and possible outcomes. The right panel shows the concentration profiles for a protein (blue) and a small molecule or salt precipitant (red) at different locations (x_1–x_4). Note the slow diffusion of the protein compared with the small molecules in the precipitation cocktail.



Free interface diffusion pathway

Figure 3-26 Pathway of free-interface diffusion through a crystallization phase diagram. The diagram shows the pathways of each of the four different locations x_1 to x_4 corresponding to the sections of Figure 3-25 and at different times during the process ($t = 0$, start; t_E = end, time when diffusion equilibrium is reached). Note that the pathway at location x_4 starting at pure precipitant never sufficiently enters the nucleation region. In this scenario crystals can only form in the protein-rich part of the diffusion chamber.



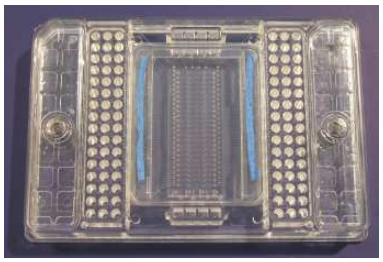


Figure 3-29 A microfluidic free-interface screening chip. The actual silicone microfluidic chip (~20 × 50 mm) is mounted in the center of an SBS footprint compliant plate that houses the sample and precipitant receptacles, and the control valve connections feeding into the chip. This chip model can screen eight proteins (~1 µl each) against 96 reagent conditions (10 µl each), executing a total of 768 simultaneous crystallization screening experiments.

Uses very little material
Different traversing of phase space
Scale-up and harvesting difficult and overlap with VD not yet determined

Free interface diffusion in gels

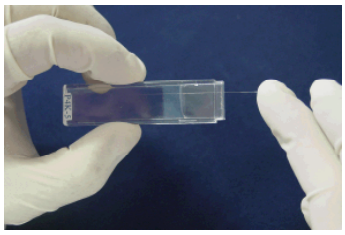
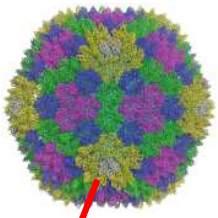
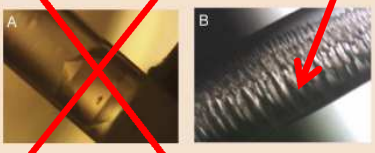


Figure 3-27 Counter-diffusion in thin capillary. The protein is wicked up in the capillary, one end sealed, and the capillary is inserted into a gel saturated with precipitant solution. The gel blocks diffusion of the protein, but allows a wave of precipitant to propagate through the capillary, effectively probing the crystallization space. Image courtesy Juan-Maria Garcia-Ruiz.



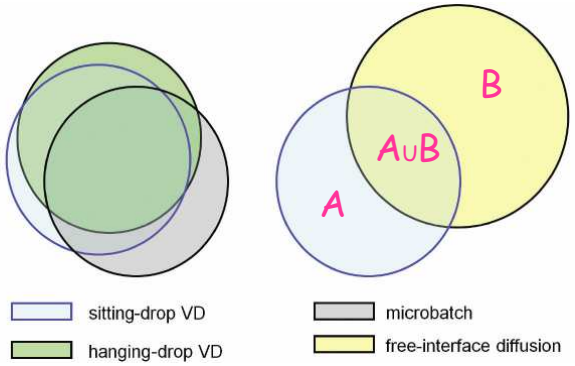
Sidebar 3-11 Interface diffusion in virus crystallography. The spectacular 66 MDa structure of the double-stranded DNA bacteriophage PRD1 (Figure 2-2) has been determined from capillary-grown crystals[70] diffracting to 4 Å (Figure 3-28). These crystals were too sensitive to be grown by the sitting-drop method requiring separate handling steps during harvesting and mounting. A combination of methods including glutaraldehyde cross-linking of the virus particles and vapor diffusion combined with counter-diffusion in capillaries mounted in a vapor-diffusion cell finally led to diffracting crystals.[71]

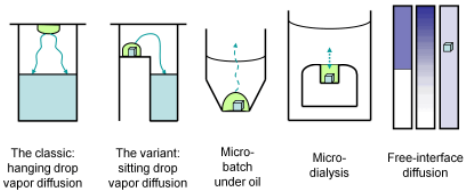
Figure 3-28 Crystals of the 66 MDa bacteriophage PRD1 grown in a capillary. Crystals of these huge molecular particles (640 Å diameter) are so sensitive that they cannot be grown using other techniques such as vapor diffusion or batch crystallization which require harvesting and manipulation. Interestingly, the pretty and well-formed crystals (A) did not diffract, while the "fronds" of the crystals growing in a fern-like structure (B) diffracted to 4 Å—a reminder that looks of crystals can be deceptive. Images courtesy of Nicola Abrescia, David Stewart, and Jonathan Grimes, Oxford University, UK.



Changes in methods will change results

Figure 3-31 Venn diagrams representing different crystallization scenarios. The circles represent different techniques, and circles of same diameter indicate equal overall success rate for each method. The left panel shows that overlap in outcomes given the same reagents and drop sizes between hanging- and sitting-drop vapor-diffusion (VD) techniques is generally quite large, whereas microbatch experiments may have fewer conditions in common with either of the vapor diffusion techniques.[76] The panel on the right side visualizes a hypothetical scenario representing free-interface diffusion with potentially higher success rates in initial screening (larger circle) but limited overlap with sitting-drop VD technique.





A change in methods can reveal new conditions but can make it difficult to reproduce previously successful conditions

Summary of physical chemistry of crystallization

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

Necessary requirement: supersaturation

B) The free energy of formation is driven by competition between entropic terms and not the enthalpic terms:

$$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$$

Necessary requirement: Phase stability

c) The kinetics must allow that the thermodynamically stable state – hopefully a crystal – can be reached.

Necessary requirement: nucleation

A few points for review (I)

- The protein is the key factor for success in crystallization. Whether your crystallization can succeed or not is already 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.
- Be ready to use multiple constructs and protein engineering to increase your chances for success

Next lecture:

Irrespective of all we have learned so far efficiency is our main concern: How to set up a maximum of successful experiments with the least amount of material and cost?

Handling - Automation -> Statistics -> Analysis -> Prediction -> Modification -> Success!

Complete summary, Q&A

Key concepts of Chapter 3

Box 3-1 Protein crystallization basics. Protein crystals are periodic self-assemblies of large and often flexible macromolecules, held together by weak intermolecular interactions. Protein crystals are generally fragile and sensitive to environmental changes. In order to form crystals, the protein solution must become supersaturated. In the supersaturated, thermodynamically metastable state, nucleation can occur and crystals may form while the solution equilibrates. The most common technique for protein crystal growth is by vapor diffusion, where water vapor equilibrates from a drop containing protein and a precipitant into a larger reservoir with higher precipitant concentration. Given the large size and inherent flexibility of most protein molecules combined with the complex nature of their intermolecular interactions, crystal formation is an inherently unlikely process, and many trials may be necessary to obtain well-diffracting crystals.

Box 3-2 The nature of protein crystals. The protein molecules in a crystal are connected through a network of few and specific intermolecular interactions. Between the packed protein molecules, large voids remain that are filled with solvent. The voids between molecules allow the crystal to exchange liquid with the environment, and small molecules such as ligands or drug molecules can be soaked into crystals. For experimental phase determination, heavy metal ions can be soaked into the crystal, where they may specifically bind to certain residues and form marker atoms for phase determination.

Intermolecular packing interactions can change the conformation of surface residues or flexible loops of a molecule, but the core of a protein maintains its native conformation in the crystalline state; enzymes generally remain active in the crystalline state.

Box 3-3 The protein is the most crucial factor in determining crystallization success. Given that a crystal can only form if specific interactions between molecules can occur in an orderly fashion, the inherent properties of the protein itself are the primary factors determining whether crystallization can occur. A single-residue mutation can make all the difference between successful crystallization and complete failure. Important factors related to the protein that influence crystallization are its purity, the homogeneity of its conformational state, the freshness of the protein, and the additional components that are invariably present—but often unknown or unspecified—in the protein stock solution.

Box 3-4 Fundamental properties of protein solutions. Crystallization is a special form of phase separation from a homogeneous solution, where the protein-rich phase in equilibrium with the protein solution is an ordered crystal. For phase separation to occur, the solution must become supersaturated, where it is thermodynamically metastable and will upon nucleation equilibrate into a protein-rich phase and protein solution. Supersaturation is a thermodynamic necessity to achieve phase separation. If the nucleation process and other kinetic parameters such as growth kinetics are favorable, the protein-rich phase may form as a protein crystal. Other possible protein-rich phases that can form are various liquid phases (protein "oils") and solid precipitates.

The phase relations in a protein solution can be represented in a pseudo-binary phase diagram with protein concentration and the precipitant concentration as parameters. The pH of a protein solution has a strong effect on the solubility, but crystallization does not preferentially occur at the isoelectric point, where protein solubility is at its minimum. Temperature also affects protein solubility and crystallization, but no general preferences can be predicted. Protein crystallization is an entropy-driven process; the release of water molecules across both hydrophobic and polar surface residues during the formation of a crystal contributes to the entropy gains of the solvent, exceeding the entropy loss largely caused by the loss of motional degrees of freedom upon crystallization. Self-assembly into crystals can be assisted by trial of additives that stabilize the protein, mediate crystal contacts, fine-tune intermolecular interactions, or otherwise modify protein solubility.

Box 3-5 Kinetics determine crystallization events. Once the protein solution has reached thermodynamically metastable supersaturation, nucleation determines how the phase separation into protein-rich phase and saturated protein solution occurs. At high supersaturation, spontaneous homogeneous nucleation of the protein rich phases occurs, while at low supersaturation heterogeneous nucleation must be induced by seeding. Real single crystals are not perfect; they consist of multiple slightly misaligned domains forming a mosaic crystal.

Key concepts of Chapter 3

Box 3-6 Composition of a crystallization cocktail. The purpose of a crystallization cocktail is to act as a precipitant reducing the solubility of the protein and to introduce other reagents that are potentially beneficial to crystal formation. Salts and PEGs are the major precipitants, and additives are selected to either facilitate crystal contact formation or otherwise stabilize or improve crystal formation. The pH of the cocktail strongly affects the distribution of charges on the protein surface, and therefore is a major determinant for crystallization. The effects of the reagents in the cocktail are generally synergistic and difficult to predict. At high concentrations, the precipitants (PEGs and salts) can also act as cryoprotectants.

Box 3-7 Crystallization techniques. The inability to predict *ab initio* any conditions favoring protein crystallization means that, in general, several hundred crystallization trials must be set up in a suitable format and design. Crystallization screening experiments are commonly set up manually or robotically in multi-well format crystallization plates. The most common procedure for achieving supersaturation is the vapor-diffusion technique, performed in sitting-drop or hanging-drop format. In vapor-diffusion setups, protein is mixed with a precipitant cocktail, and the system is closed over a reservoir into which water vapor diffuses from the protein solution. During vapor diffusion, both precipitant and protein concentration increase in the crystallization drop and supersaturation is achieved.

Other protein crystallization methods include batch crystallization under oil, dialysis methods, and free-interface diffusion techniques. Microfluidic chips or thin-walled capillaries are used for free-interface diffusion. The advantages of free-interface diffusion methods are a comprehensive coverage of the crystallization phase space, and that very little material is required in the case of microfluidic chip methods. The pathway through crystallization phase space can be visualized with the help of crystallization phase diagrams. Crystallization diagrams combine information about thermodynamically defined phase relations with a tentative assignment of kinetic nucleation regions.

As a rule of thumb, low supersaturation favors controlled crystal growth, while high supersaturation is required for spontaneous nucleation of crystallization nuclei. Seeding is a method to induce heterogeneous nucleation at low supersaturation, which is more conducive to controlled crystal growth.

Box 3-8 Crystallization strategies. Finding suitable conditions for protein crystallization—provided the protein is inherently crystallizable—requires sampling of a nearly unlimited combination of parameters such as reagent combinations, pH, and temperature. Searching for successful crystallization conditions involves sampling of a multidimensional, sparsely populated, and ill-defined sampling space. Efficient sampling requires proper design of experiments.

Random screening experiments with minimum bias have confirmed a number of empirical general rules for crystallization screening and provided in addition a number of important insights affecting screening strategies. Rapid assessment of a protein's crystallization propensity can be gained by using a 2-tiered approach, starting with pH–PEG or index screens and expanding the sample space in the next round.

Random sampling and other large-scale trials have shown that if no promising results are obtained after about 300 trials, it is likely that the protein is a difficult case for crystallization: consider other protein constructs, orthologs, or protein engineering. Accept that the chance of obtaining diffracting crystals of a protein without any additional procedural adjustments or protein modifications is only 10–20%. Whether crystallization will succeed or not is already predetermined by the protein construct itself. If the protein cannot crystallize, it will not, no matter how many crystallization trials are performed. In contrast, proteins that crystallize frequently under multiple conditions also tend to diffract well. This fact lends additional weight to the case for protein engineering rather than expending excessive effort in crystallization of a poorly crystallizing protein that will in all likelihood never yield well-diffracting crystals.

Box 3-9 Ligand and heavy atom soaking. The prevalence of large solvent channels in protein crystals permits small molecules and ions to be readily soaked into crystals. A small drop of concentrated ligand or heavy atom solution is added to the mother liquor of the crystallization drop harboring the crystal. Ligand complexes usually take hours to weeks to form. In small molecule ligand soaking, the limited solubility of the substances in aqueous solutions can be problematic. Specifically bound heavy metal ions are required for isomorphous replacement phasing and are also valuable for anomalous phasing. Native gel shift assays show whether heavy atoms have bound to the protein or not. Successful ligand or heavy atom binding is validated by data collection through analysis of isomorphous difference data.

Review protein engineering strategies

Box 4-1 Protein production for crystallography. Proteins for crystallographic studies are, with few exceptions, produced by heterologous overexpression in cellular hosts. Heterologous expression requires design and cloning of a recombinant DNA molecule encoding the target protein sequence and its transfer into the expression host. The advantage of recombinant DNA methods is the great flexibility in modifying the protein sequence and the overexpression of proteins with otherwise very low natural abundance.

Box 4-2 Protein engineering strategies and levels. Protein engineering is possible at the DNA level by modifying the sequence of the protein construct and at the protein level by modifying the expressed protein. In both cases targeted, rational design strategies or random (combinatorial) design strategies can be applied. Frequently, many different protein constructs or expression conditions are screened in parallel for expression and solubility levels. The most fundamental property of a protein suitable for crystallographic studies is sufficient solubility.

Box 4-3 Targeted design strategies at the DNA level. Design strategies at the DNA level are based on the analysis of the protein sequence and the derived protein properties. Multiple sequence alignments supported by available structural information can identify domain boundaries and terminal regions. Additional tools identifying disordered regions, transmembrane helices, signal peptides, and binding sites augment the analysis and help identify potential problem regions. In eukaryotic sequences mRNA splicing sites may delineate domain boundaries or regions dispensable for stable protein structure.

Box 4-4 Predicting crystallization success. Inclusion of statistical data relating crystallization success to specific protein properties allows the establishment of conditional probabilities for crystallization success for a given protein. However, these probabilities are largely based on predicted protein properties and cannot pinpoint the specific and local intermolecular interactions that actually determine crystallization. Crystallization predictions are generally more reliable in clear cut cases (very easy or very difficult to crystallize) but provide less guidance in intermediate cases. Crystallization trials remain the only authoritative (and in fact, fast and simple) method of determining crystallization propensity of a specific protein construct.

Box 4-5 Surface entropy reduction and site specific mutations. Surface entropy reduction (SER) exemplifies a targeted design strategy at the DNA level. Mutation of high entropy surface residues such as exposed Lys, Glu, and Gln residues to Ala, Thr, or Val can lead to significant improvements in crystallization. However, various other properties such as overall protein charge, local charge distribution, and conformation are also varied at the same time and the precise causality between residue mutations and crystallization success cannot be easily established. A number of targeted protein design studies have shown that the rationale leading to various construct modifications is often not causal to the crystallization success—the readiness to modify the protein is important.

Box 4-6 Combinatorial designs at the DNA level. Combinatorial DNA libraries are based on the generation of random mutants of the target gene *in vitro*. Truncation mutant libraries and domain fragment libraries are examples of basic DNA libraries used to generate soluble and crystallizable constructs. Directed evolution is a refined combinatorial technique in which a library of random mutants is exposed to selective pressure such as an expression and solubility screen at the phenotype level. DNA shuffling allows recombining mutants with desirable traits to further improve solubility and ultimately, crystallizability and diffraction quality. Combinatorial DNA library generation requires extensive colony picking, selection, and sequencing and is generally conducted using robotic equipment.

