# Large volume protein crystal growth for Neutron Macromolecular Crystallography

by

Joseph D. Ng[1,2*], James K. Baird[3], Leighton Coates[4], Juan M. Garcia-Ruiz[5], Teresa A. Hodge[3] and Sijay Huang[3]

[1]Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL 35899

[2]iXpressGenes Inc., Hudson Alpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL 35806, USA

[3]Department of Chemistry, University of Alabama in Huntsville, Huntsville, AL 35899

[4]Oak Ridge National Laboratory, Biology and Soft matter Division, PO Box 2008, MS6475, Oak Ridge, TN 37831-6475

[5]Laboratorio de Estudios Cristalográficos (IACT), CSIC-Universidad de Granada, Av. de la Innovación s/n, Armilla (Granada), Spain

[*]Corresponding author: Phone 1.256.824.6166

E-mail: ngj@uah.edu

## Abstract

Neutron Macromolecular Crystallography (NMC) is the prevailing method that accurately determines the positions of hydrogen atoms in macromolecules. In over 100,000 macromolecular structures reported to-date in the RCSB Protein Data Base, NMC has been associated with only less than 0.1% of the crystallographic structures. The major limiting factor

in determining protein structures with neutron diffraction is obtaining large volume crystals. As neutron sources are becoming more available to general users, finding means to optimize growing protein crystals to sizes suitable for NMC is extremely important. Historically, much has been learned about growing crystals for X-ray diffraction. But due to new generation synchrotron X-ray facilities and sensitive detectors, protein crystal sizes as small as in the nano-range have been adequate for structure determination, lessening the necessity to grow large crystals. Here, we revisit some of the approaches, techniques and considerations to grow crystals to significant dimensions that are now relevant to NMC. These include experimental strategies utilizing solubility diagrams, ripening effects, classical crystallization techniques, microgravity and theoretical considerations.

## 1. Introduction

Neutron Macromolecular Crystallography (NMC) is a powerful method to obtain accurate positioning of hydrogen atoms in biological structures of 50—175kDa molecular weight with neutron data collected near atomic resolution (1.5 Å -2.5 Å) (Blakeley *et al*., 2008). Macromolecules will primarily refer to proteins unless stated otherwise. Hydrogen constitutes 50% of the protein atoms but they scatter X-ray weakly such that their positions can only be accurately identified at a resolution higher than 1.0Å. Most three-dimensional macromolecular structures have been determined by X-ray crystallography within the resolution range of 1.5Å – 2.0Å (RCSB Protein Data Bank). As a result, there is very little knowledge of the location of hydrogen atoms and water molecules in 99% of the structures determined to-date. Knowing the

hydrogen positions in macromolecules and in interacting water molecules is necessary to completely understand the reaction mechanisms, pathway and structure-function relationships among proteins with their important binding partners. Therefore, it is invaluable to obtain macromolecular structures that can provide unique and complimentary insights into hydrogen-bonding interactions, protonation states, catalytic mechanisms and hydrations states of biological structures that are not available from X-ray analysis alone.

The primary limiting factor in obtaining a neutron crystallographic macromolecular structure rests on the success of growing adequate size and quality protein crystals that can diffract to atomic resolution. Large volume protein crystals are required to overcome the weak flux associated with neutron sources (Coates *et al*., 2001, Teixeira *et al*., 2008). NMC is dependent on scattering by the atomic nuclei providing information on the location of hydrogen atoms from the protein, its substrates and surrounding water. The present protein crystal volume prerequisite for neutron diffraction is at least 0.1 $mm^3$ (Blakeley *et al*., 2008, Howard *et al*., 2011, Weber *et al*., 2013). In most cases a larger protein crystal (~1 $mm^3$ for non-perdeuterated proteins) is needed for successful neutron analysis. These substantial crystal sizes are not necessary for X-ray structure determination because of the accessibility of intense synchrotron focused beam sources and sensitive X-ray detectors.

Globally, there are only a handful of neutron facilities available for single macromolecular crystal diffraction. LADI-III (Blakeley *et al*., 2010) at the Institut Laue-Langevin, Grenoble, France is a quasi-Laue diffractometer able to collect high-resolution data between 1.5Å-2.5Å. It has been repositioned closer to the cold neutron source, thereby increasing the incident neutron flux considerably. The ability to collect data at cryogenic temperatures of around 100K has also just been added to the instrument. Several newly

constructed neutron beamlines have also recently come online such as the BIODIFF diffractometer at the FRM2 reactor in Garching, Germany. BIODIFF is designed to analyze crystals with large unit cell sizes (> 200Å) with a monochromatic neutron source. In the United States, the MaNDi (Coates *et al*., 2010) and IMAGINE (Meilleur *et al*., 2013) instruments are available at ORNL using a Spallation Neutron Source equipped with capabilities to conduct cryo-data collection and with large unit cell sizes. In Japan the IBIX instrument at J-PARC has been operational since December 2008. All of these facilities have significantly increased the availability of neutron diffraction beamtime to the global user community. Even though high power neutron sources with state-of-the-art facilities are becoming more available with faster data collection capabilities, the large crystal size requirement still persists.

Extended protein crystal growth to acquire large volumes has been very challenging because macromolecules are usually dynamic and flexible and as a result contains inherent heterogeneity (Malkin & Thorne, 2004) despite the extent of their purity. In addition, the presence of partially denatured protein most often poisons the growing surface (Ng, Kuznetsov, *et al*., 1997). Nonetheless, there is a small number of protein neutron crystallographic structures determined to-date, compared to those deciphered by X-ray. Less than 75 non-redundant neutron structures have been deposited in the RCSB Protein Data Bank at present compared to over 90,000 of those deciphered by X-ray. All the macromolecular structures that have been obtained by neutron crystallography are listed in Table 1 showing their PDB ID numbers, space groups and determined resolution.

Obtaining large volume crystals begins with the same common approach for crystallizing any macromolecules out of solution. The general strategy for protein crystallization is to reduce the solubility of the macromolecule from the solvent such that an extreme supersaturation state is

achieved in respect to the protein. There are a number of devices and procedures to bring a protein solution to a supersaturated state (for review see (McPherson, 1999, Ducruix & Giege, 1999, Ng & Garcia-Ruiz, 2006, McPherson & Cudney, 2014, Loll, 2014) and ultimately lead to a large crystal after appropriate optimization. The most commonly used in obtaining protein crystals are the batch and vapor diffusion methods (Cudney & Patel, 1994, Bergfors, 2007). Batch crystallization entails the direct mixing of an undersaturated protein solution with a precipitant solution (Rayment, 2002, Chayen *et al*., 1990). In vapor diffusion, undersaturated protein solution is mixed with precipitants in a droplet and the resulting mixture is allowed to evaporate in a closed system against a reservoir solution containing the precipitant (Benvenuti & Mangani, 2007). At this point there is a high probability that critical nuclei will spontaneously form in solution. Counter-diffusion crystallization in restricted geometry is also a well-known method to screen and optimize conditions for protein crystallization (García-Ruiz, 2003). In counter-diffusion equilibration, the samples of the precipitating agent and protein solution are arranged in juxtaposition to one another inside a volume-constrained capillary. The two solutions are set to diffuse against each other in a convection-minimized environment resulting in a spatial-temporal gradient of supersaturation along the length of the capillary (García-Ruiz, 2003, Otalora *et al*., 2009). This technique has been successfully used for *in situ* X-ray diffraction analyses for structure determination (Gavira *et al*., 2002, Ng, Stevens, *et al*., 2008, Ng *et al*., 2003). Regardless of the initial approach to find early crystallization conditions, strategic efforts and consideration must be made to optimize on-going crystallization processes to attain maximum crystal volume. Here we will review some established experimental tools, approaches and observations associated with large crystal growth for neutron crystallography. These include the usage of phase-diagram information, counter-diffusion crystallization, Ostwald Ripening,

microgravity and essential manipulations once protein crystals suitable for neutron crystallization are obtained.

## 2. Phase-diagram-guided crystallization

The knowledge of protein solubility from phase-diagram information can be most useful in knowing how to find optimal nucleation and growth conditions. While obtaining phase diagrams may not be necessary to obtain crystals for conventional X-ray crystallographic determination, knowing the solubility barriers of a protein is important in cultivating crystals to maximum size for neutron crystallography. Lysozyme and a DNA decamer are examples of macromolecules in which their phase diagrams were essential in obtaining crystals suitable for neutron crystallographic determination (Iwai *et al*., 2008, Niimura, 2010). A protein phase diagram describes the liquid and solid states of the macromolecules as a function of chemical or physical variables that affect the protein's solubility (for review see Rupp 2015). Most commonly used are the precipitant or additive concentration, pH, temperature, or ionic strength. As shown in Fig. 1A, the solubility and super-solubility curve divide the plot into three regions, namely the undersaturated zone for protein concentration below the solubility barrier and the supersaturated zone for protein concentration larger than solubility. This supersaturated region is itself divided into the labile zone at larger supersaturation values, where nucleation occurs instantaneously, and the metastable zone where nucleation probability varies between 0 (in the solubility curve) and 1 (in the supersolubility curve) (Ng *et al*., 1996).

A phase diagram for conventional crystallization (Fig 1A) highlights the extreme supersaturation region often referred to as the labile region (Miers & Isaac, 1907, Feigelson,

1988). The lesser value of supersaturation is the metastable region where nucleation events are low and crystal growth is favorable. In traditional protein crystallization screens, chemical reagents in buffered solution can be randomly, systematically or selectively chosen to mix and equilibrate with varying protein concentrations. In this process, a supersaturated condition targeting a single labile region for spontaneous nuclei and metastable region for the greatest crystal growth can be acquired. Upon evaluation of a phase diagram for a targeted protein, one can strategically maneuver between the nucleation and crystal growth conditions (Fig 1A) as a strategic crystallization approach to obtain large volume crystals suitable for neutron diffraction.

There are many procedures to obtain information for a protein phase diagram (Cacioppo *et al*., 1991, Asherie, 2004, Dunuwila & Berglund, 1997). The most commonly applied phase diagrams are two-dimensional having only one independent variable even though multidimensional diagrams exist (Ewing *et al*., 1994, Sauter *et al*., 1999, Rakel *et al*., 2014). Here, we will only discuss the single independent variable diagrams. To establish a phase diagram, one must first determine the solubility of the protein as a function of ambient conditions- that is the concentration of proteins in the solution at equilibrium under a certain chemical condition. The most classical method to determine solubility for a phase diagram is measuring protein concentrations at equilibrium under saturated and undersaturated conditions (Sazaki *et al*., 1996, Haire & Blow, 2001, Gray *et al*., 2001, Ewing *et al*., 1994). One may obtain solubility data points by placing some protein crystals in a protein-free solution of interest while gently stirring the solution. As a function of time, the soluble protein concentration is measured by absorbance (A280) or bicinchoninic acid assay (BCA) (Smith *et al*., 1985) to determine how much of the crystals have dissolved back into the solution. Concurrently, the targeted protein is set up against a supersaturated solution (Asherie, 2004). During the equilibration process, the

protein concentration as a function of time is measured as described for the undersaturated equilibration process. The point where the protein concentrations converge for the super- and undersaturated equilibration is the solubility of the protein for the selected condition (Ceq). The equilibrium method can provide valuable information on protein solubility and phase transitions; however, the technique requires significant amount of protein materials (up to gram quantities) and time.

The use of Michaelson interferometry has been an effective tool in obtaining protein solubility data (Sazaki *et al*., 1996, Gray *et al*., 2001, Nakazato, 2008) requiring less time and materials than that of the equilibrium technique. By measuring the changes of surface and concentration gradient solution interference fringes in the presence of small crystals, accurate estimates of protein solubility are achievable. Lysozyme has been the most well reported protein used in interferometry measurements in which surface fringe movements were correlated with crystal growth or dissolution as a function of a variable condition (i.e. temperature). Consequently, informative solubility curves can be derived for selected conditions and used to adjust crystal growth conditions to obtain optimized crystal volumes.

Thermal and chemically stable proteins are convenient and often preferred candidates for large volume crystallization. These include proteins derived from thermophilic organisms or mesophilic proteins that are functionally and structurally tolerant to wide temperature changes (i.e. plant storage proteins, proteases, nucleases etc.). In this case, temperature can be exploited as a useful variable in keeping the crystallization solution metastable during the growth process. One successful approach in using temperature controlled ripening processes to grow protein crystals suitable for NMC is the use of a semi-automated protein crystal-growth arrangement coupled to a custom made quartz crystal growth cell with a microscope temperature-controlled

system (Budayova-Spano *et al*., 2007). This approach includes seeding a protein solution in a predetermined crystallization condition and observing *in situ* growth or dissolution as a function of temperature under batch equilibration . Upon fine temperature changes, the areas of the phase diagram can be qualitatively identified. Large-volume crystals of perdeuterated urate oxidase (Uox) were successfully obtained by this method. By carefully controlling and optimizing the crystallization conditions for Uox around its metastable zone (with no occurrence of spontaneous nucleation), the neutron crystallographic structure of Uox was determined from the crystals grown by temperature adjustments (Budayova-Spano *et al*., 2007, Oksanen *et al*., 2009).

## 3. Counter-diffusion crystallization

Protein crystallization by counter-diffusion in restricted geometry is an effective technique to obtain macromolecular crystals suitable for neutron crystallography. Contrary to conventional techniques such as vapor diffusion, varying supersaturation conditions leading to protein precipitation, nucleation and crystal growth can be obtained simultaneously in a capillary channel as exemplified in Figure 1B using the model protein thaumatin (Ng, Clark*, et al*., 2008). Crystals obtained by this technique are a consequence of selective growth conditions from different supersaturation environments (García-Ruiz, 2003, Ng *et al*., 2003, Garcia-Ruiz & Ng, 2006). The crystals can grow out to the diameter of the capillary and be directly analyzed by X-ray radiation *in situ* without ever touching the crystals (Ng, Clark*, et al*., 2008, Ng *et al*., 2003, Gavira *et al*., 2002, Ng & Garcia-Ruiz, 2006, Ng, Stevens*, et al*., 2008).

The counter-diffusion crystallization procedure is able to compose a continuous supersaturation gradient. During the diffusion process a supersaturation state with respect to the

protein can be attained in the labile region while crystal growth can occur in a continuous range in the metastable area of the phase diagram (Fig 1B). Hence, crystal formation is caused by the progression of a nucleation front resulting from the non-linear interplay among mass transport, protein crystal nucleation and growth (Garcıa-Ruiz & C., 2001, Carotenuto *et al*., 2002).

The implementation of counter-diffusion crystallization relies on the maximum reduction of convection allowing diffusion to govern the mass transport process along the length of the protein chamber. In practice, the optimal supersaturation gradient achieved is dependent on minimizing the ratio between buoyant and viscous forces. The quality of this diffusive scenario can be described by the non-dimensional Grashof number ($N_G$) used in fluid dynamic studies with the following relationship:

$$N_G = \frac{\text{buoyancy forces}}{\text{viscous drag forces}} = d^3 \times \rho(\delta\rho/\delta c) \times \Delta c \times g \times u^{-2} \qquad (1)$$

where $d$ is the characteristic dimensions of the system, ρ is the density of the fluid, $\delta\rho/\delta C$ is the density gradient, ΔC is the change in concentration, $g$ is the gravity value and $u$ is the viscosity of the solution (García-Ruiz, 2003). We have shown proteins with different molecular weights and isoelectric points can be crystallized by counter-diffusion crystallization under a wide range of chemical and physical conditions (Ng *et al*., 2003).

There are some advantages to consider in this approach. Single nucleation and maximum crystals growth can be experimentally revealed along the length of the capillary since the equilibration process passes through the labile and metastable conditions along a solubility gradient. Therefore, optimization for growing large crystal volumes filling up the entire diameter of the capillary can be realized without solubility data. Second, diffraction analysis can be performed *in situ* without physical handling of the crystals (Ng, Clark, *et al*., 2008, Gavira *et*

*al*., 2002). Crystals that grows across the diameter of the capillary are attached to the chamber walls and are immobilized. Therefore X-ray or neutron diffraction can be performed without the need to separately mount the crystals minimizing damage from physical manipulation, particularly for large fragile crystals. Third, crystals grown in the capillary can undergo soaking of ligands, metals, cryogenic preservatives (Gavira *et al*., 2002) and exchanging $H_2O$ for $D_2O$ for preparation of neutron diffraction (discussed in the next section).

When crystals must be grown in large diameters (> 0.3mm), high viscosity precipitants are preferred and, in some cases, accompanied by strategic placement of agarose gel plugs. The initial arrangement is prepared similar to a Free Interface Diffusion (FID) crystallization technique where the same geometry is used. Typical FID preparations can be considered as batch experiments conducted with very slow mixing (Otálora *et al*, 1997, García-Ruiz *et al*., 1997). As shown in Fig 2, a supersaturation gradient can be observed in the large diameter channel similar to those seen in narrow ones.

**4. Ostwald Ripening**

Ostwald Ripening is a process in which small stable particles grow at the expense of smaller ones such that a more stable thermodynamic state is established (Veesler *et al*., 1994, Ostwald, 1897, Mullin, 1993). While this process was discovered and observed in small molecule crystallization, it is not well documented for proteins. However, characteristic features of Ostwald Ripening have been observed for macromolecules in which depletion zones were present as large crystals grow at the expense of the smaller ones. These proteins include alpha-amylase, TBSV, thermostable AspRS (Ng *et al*., 1996) and lysozyme (Streets & Quake, 2010). In these cases, large volume crystals can be achieved when conditions for Ostwald Ripening are

permitted. The thermostable AspRS has been particularly well observed for its ripening behavior and subsequently crystals measuring up to 1mm in the longest dimensions were achieved (Ng *et al*., 1996).

Ostwald Ripening is often observed with large volume crystal growth in counter-diffusion equilibration. When the diameter of the crystallization capillary is large (> 0.3mm), convective mixing may follow the formation of a supersaturated gradient along the length of the crystallization chamber. This was the case for the large volume protein crystal growth of the hyperthermophilic Inorganic Pyrophosphatase (IPPase) (Hughes *et al*., 2009). Below describes a scenario where large proteins of IPPase crystals were achievable (Fig 2).

Conditions for initial crystallization were already known where small crystals of IPPase were obtained from vapor diffusion sparse-matrix screens. Concentrated protein solution was loaded into quartz capillary by syringe aspiration with diameters ranging from 1-2mm. It is worth noting that one must use quartz capillaries for neutron diffraction because boron-silicate glass strongly scatters neutrons. In addition, the protein concentration required for counterdiffusion crystallization is approximately twice that needed in vapor diffusion or batch processes. Fig 2 illustrates the progress of equilibration accompanied by the observation of the actual protein crystals obtained. Several significant results were produced. The precipitant contained a minimal concentration of 30% 2-methyl-2-4-pentanediol (MPD) (Anand *et al*., 2002) and was sufficient to achieve a ratio between buoyancy and viscous forces (see equation *(1)*) allowing diffusion to govern the mass transport process despite the large tube volume. This was evident by the supersaturation gradient obtained giving rise to a precipitation front closest to the precipitant interface and the progression of high to low nucleation of crystal growth along the length of the capillary. When the capillary diameters are large, the presence of additives that can

increase the viscosity of the solution is important to sustain a diffusion limiting process. Reagents such as glycerol, PEG and general cryo-protectants have been shown to be effective.

A single crystal filling the entire diameter of the capillary can be obtained within 10 days. During the initial course of the equilibration, crystals are observed to grow along the length of the capillary. However, slow mixing shortly occurs and the existing crystals undergo Ostwald-like ripening where depletion zones form around the growing crystals that eventually fill up the capillary. To avoid osmotic shock, agarose plugs are implemented dividing the protein and precipitating solutions prior to initializing the crystallization process. Consequently, a large single crystal usually grows on and sometimes into the agarose plug.

**5. Deuteration**

For NMC, the challenge does not end with growing the protein crystals to a suitable size. Once a large volume crystal is obtained, one must prepare the crystal for diffraction with $D_2O$ exchanges without compromising the quality of the crystals. This is because large incoherent scattering background of hydrogen attributes to the low gain in signal-to-noise ratios of neutron diffraction images. Approaches to remedy this issue include the complete replacement of hydrogen atoms with deuterium (Shu *et al*., 2000), $D_2O$ exchange with $H_2O$ by vapor exchanges or by soaking procedures (Niimura & Bau, 2008). Here we will discuss exchanging $H_2O$ with $D_2O$ in the capillary geometry without any physical manipulations on the protein crystal.

When crystals are grown in capillaries, it is easy to remove the mother liquor and carefully equilibrate the targeted protein with an equivalent solution containing $D_2O$. This is done by carefully excising the ends of the capillaries, withdrawing all the solutions and smaller crystals leaving the targeted crystal immobilized with some residual mother liquor. Thereafter,

the capillary is then filled from both sides with deuteriated solution leaving an adequate gap between the solution and the crystals to allow vapor diffusion (Fig 3A). The capillaries are sealed again and this process can be repeated several times during the course of the $D_2O$ exchanges.

Crystals obtained in the presence of an agarose plug are preferred because slow exchanges of new solutions through a porous barrier minimizes physical and chemical stress. The deuteriated solutions are replaced as described before and directly placed against the agarose plug allowing free liquid diffusion (Fig 3B). A final alternative for effective $D_2O$ exchanges is simply using Free Interface Diffusion against a solution containing a similar concentration of the protein and precipitant as the mother liquor. In a typical protein crystal, 66% of all the hydrogen atoms in the unit cell can exchange for deuterium. The percentage that can exchange within the protein is much lower at approximately 26%. Highly stable non-flexible domains resist deuterium exchange giving data on protein dynamics. It is thought that the bulk solvent in the crystal rapidly exchanges with $D_2O$. However, many users undergo deuterium exchanges for days with several exchanges to assure adequate replacement of dissociable hydrogen atoms with the deuterium.

**6. Cryogenic neutron protein crystallography**

In X-ray crystallography, collecting diffraction data at the synchrotron and home X-ray sources under cryogenic conditions (near liquid nitrogen temperature) (Hope, 1988) have become routine. Protein crystals that are cryo-cooled have reduced radiation damage by X-ray and consequently, significant improvement of resolution in data collection can be obtained with a single crystal (Garman & McSweeney, 2007). In contrast to X-ray radiation, protein crystals are

not damaged by thermal neutron diffraction. Presently, most neutron diffraction analyses are conducted in ambient temperatures with minimal or no radiation damage as long as the crystals are chemically and thermally stable during the data collection. However, collecting neutron diffraction data requires a very extended exposure time, ranging from 6-18 hours per image. This means data acquisition time can be on the order of days to weeks. In comparison, capturing X-ray diffraction images usually requires only seconds per image and even much faster with the recent development of shutterless cameras. In the interest of performing NMC with protein crystals that may not be chemically or thermally robust, trapping short-lived intermediates or reducing atomic displacement parameters, cryogenic neutron protein crystallography is necessary. In addition, other important crystallographic rationales for neutron cryogenic data collection include the desire to compare X-ray and neutron structures at the same temperature (Juers & Weik, 2011), measure the impact of radiation damage and determine their influence on X-ray/neutron joint refinement. The study of crycrystallography for neutron-sized crystals is still in its nascent stage and there is no systematic study to-date that examines the ability to cryo-cool a large crystal for neutron studies as a function of its size, solvent content, cooling kinetics and cryoprotecting agents as those crystals used for X-ray. In contrast, there are extensive reports on cryo-cooling methods and approaches for crystals appropriate for X-ray crystallography and the principles of these techniques may be considered for large volume crystals (Juers & Matthews, 2004b, a, Garman & Owen, 2007, Garman & Doublie, 2003, Borgstahl *et al*., 2000). We will focus our discussion here on reporting approaches that have been recently performed on neutron beamlines that are amendable to neutron cryocrystallography.

To-date only two protein cryogenic neutron structures have been reported. The first was concanavalin A (Blakeley *et al*., 2004) and the most recent was Toho-1 b-lactamase (Coates *et*

*al*., 2014). Cryo-treating large volume crystals have been thought to be very difficult, if not impossible. The difficulty lies in selecting proper cryo solutions, soaking protocols and rate of supercooling crystals without disrupting the crystal quality (i.e. increasing mosaicity). Even though there are many techniques and approaches on cryogenic preparation of protein crystals for X-ray studies (Hope, 1988, Kitago *et al*., 2005), detailed studies on the cryo-cooling of large volume crystals have not been well documented and there are no standard procedures to prepare protein crystals for cryogeneic neutron data collection.

Cryogenic neutron data collection for protein crystals has recently become available at the new Biological Diffractometer (BIODIFF) at the FRM II and the Macromolecular Diffractometer (MaNDi) at the Spallation Neutron Source, Oak Ridge National Laboratory (ORNL). The Toho-1 b-lactamase is the first perdeuterated cryogenic protein ligand complex captured at 100C using NMC (Coates *et al*., 2014). In this case, the crystals measuring about 1mm$^3$ were grown by batch method and exposed to cryogenic solution (in this case it was 30% perdeuterated trehalose) very briefly (30 sec). The crystal was then flash cooled immediately with a liquid nitrogen open-flow Cryostream. Short soaking times (in seconds) of large protein crystals provided better results than those that have undergone long sequential soaks in cryoprotective solutions. This was also the case with hypertheromophilic IPPase where crystals greater than 1mm$^3$ were analyzed with X-rays under cryogenic conditions as a function of soaking time. Quick soaks (<30 seconds) followed by flash cooling provided the best results as determined by mosaicity and resolution limit measurements. In addition to Toho-1 β-lactamase and IPPase, we have observed in other large volume protein crystals (>1mm$^3$ ) that regardless of solvent content, space group and cryo- preservative solutions, quick cryo-soaks followed by flash freezing retained crystal integrity much better than those that have gone through extensive cryo-

soaks followed by storage in liquid nitrogen (results not published). BIODIFF uses a large cylindrical neutron-sensitive area detector that completely surrounds the crystal in the presence of a cryostream. This system proves cryogenic neutron crystallography can be very feasible using standard equipment and procedure. Other neutron facilities are now being equipped with cryogenic capabilities as well (see introduction section). Neutron cryocrystallography will be necessary in the future as many biologically relevant proteins are temperature and chemically labile. As neutron facilities are still limiting, extended storage and transportation periods will be required while waiting for data collection opportunities. More studies development in cryo-perservation of large volume crystals will be needed in the future.

## 7. Microgravity

Protein crystal growth under microgravity has been shown to produce larger and higher quality protein crystals compared to those obtained on Earth (Day & McPherson, 1992, Ng *et al*., 2002, Ng, 2002, DeLucas *et al*., 1999, Lorber *et al*., 2002, Ng, Lorber, *et al*., 1997). While microgravity protein crystal growth experiments have been curtailed in the U.S. and Europe in the past 10 years, the International Space Station (ISS) has been recently used as a platform, managed by the Center for the Advancement of Science in Space (CASIS), to grow large volume protein crystals for X-ray and Neutron crystallography. Space X, an American space transport services company, can now provide lift vehicles to the ISS delivering protein crystallization experiments. While obtaining large protein crystals may not be as important now as it was in the past for X-ray crystallography, it is absolutely necessary for NMC today. Microgravity may play a critical role in obtaining protein crystals suitable for NMC where it would not be otherwise possible on Earth. When proteins grow to large volumes in the presence of gravity, lattice

defects or irregularities accumulate due to convective mixing and sedimentation during the equilibration process. In microgravity, molecules will traverse the crystallization chamber at a very slow rate providing more time for crystal lattice alignment thus producing a higher quality crystal.

The protein IPPase was recently crystallized by counter-diffusion equilibration in large diameter (2mm) quartz capillaries on the ISS for 6 months, the longest duration of time for any protein crystal growth experiment using the counter-diffusion geometry as described earlier. Large crystal volumes were obtained growing across the length of the capillary as a result of a progressive nucleation front.

Reflecting the values of the viscosity normally used in protein crystallization, a pure diffusive mass transport can be only assured for capillary diameters of 20 microns considering the diffusive scenario determined by Grashof number (eq 1). For a diameter of 2000 microns used in neutron diffusion experiments there is a strong contribution of convection to mass transport processes, in particular when a solid crystalline phase form, unless high viscous protein solutions or gels are used. Alternatively, one can use microgravity. Clearly this is shown in Fig 4 showing the well separated large volume crystals grown from microgravity compared to that on Earth. The detail of this work will be reported elsewhere. This method of growing large volume crystals holds promise for obtaining protein crystals suitable for NMC when it would otherwise not be possible on Earth.

**8. Theoretical considerations on protein crystal size**

In the sections above, we have focused on laboratory methods for controlling parameters that would be conducive to large crystal growth. Theory can also serve as a guide. It can be

shown, for example, that some of the difficulties encountered in obtaining large protein crystals are due to the peculiar dependence of the free energy of dissolution of a protein crystal on its size. The free energy is a function of the charge on the protein molecule, however, which suggests that proper adjustment of pH can be an effective method for controlling the free energy.

Consider the size, $a$, of a microscopic crystal of a substance, such as glucose, whose molecules are uncharged in aqueous solution. The free energy of dissolution, $\Delta G(a)$, of the crystal is given by (Rowe & Baird, 2007),

$$\Delta G(a) = k_B T \ln\left(\frac{m(a)}{m_s}\right) - \frac{2v\gamma_o}{a} \qquad (2)$$

In Eq. (2), $m(a)$ is the solubility in mg/mL of a crystal of radius, $a$, $m_s$, is the solubility of the crystal when it has reached macroscopic size, $v$ is the molecular volume of glucose, $\gamma_o$ is the interfacial energy at the crystal/solution boundary, $T$ is the Kelvin temperature, and $k_B$ is Boltzmann's constant. The quantity, $m_s$, can be identified with the experimentally observed "bulk solubility" or the "solubility limit" of the protein. The condition for a crystal of microscopic size, $a$, to be in equilibrium with the growth solution is $\Delta G(a) = 0$. According to Eq.(2), this can occur for finite values of $a$ only if the growth solution is supersaturated in glucose, i.e. $m(a) > m_s$. As the crystal grows in size, it takes up more glucose molecules from the solution, so the ambient concentration, $m(a)$, falls, and the supersaturation defined by $m(a)/m_s$, diminishes. As the crystal reaches its final macroscopic size, regarded mathematically as $a \to \infty$, its solubility, $m(a)$, approaches $m_s$.

In contrast to molecules of glucose, the molecules of a water soluble protein are charged when they are dissolved in a pH buffered aqueous growth solution. Depending upon the pH, the

functional groups responsible for the charges on protein molecules are (Baird & Kim, 2002): (1) the COOH and $NH_3^+$ amino acid groups at the ends of the polypeptide chain, and (2) the acid/base groups, which are in the side chains of certain of the amino acid residues, specifically lysozyme, arginine, histidine, tyrosine, aspartic acid, glutamic acid, and also cysteine, when cysteine is not forming a disulfide bond with another cysteine. Not only are the protein molecules in solution charged, but so also are the crystals formed from these molecules (Lee *et al.*, 2001). The charges on the crystals are most likely on the surface and occur whenever a protein molecule in the crystal is oriented so that one or more of its ionizable groups comes in contact with the growth solution.

The crystal surface thus contains a number of acidic sites, HA, and a number of basic sites, B. The acid sites ionize according to

$$HA \rightarrow H^+ + A^- \qquad (3)$$

The fraction , $\theta$ , of the HA sites which are in the charged form, $A^-$, depends upon the pH and is given by

$$\theta = \frac{1}{1+10^{-(pH-pK_a)}} \qquad (4)$$

where $pK_a = -\log K_a$, and $K_a$ is the acid ionization constant of HA. By contrast, the basic sites, B, when in acid form, $BH^+$, ionize according to

$$BH^+ \rightarrow B + H^+ \qquad (5)$$

The fraction, $\theta$ , of the basic sites, B, which are in the charged form, $BH^+$, is given by

$$\theta = \frac{1}{1+10^{-(pK_a-pH)}} \qquad (6)$$

where $K_a$ is the acid ionization constant of $BH^+$.

The presence of ionizable sites on the crystal surface contributes to the free energy of the crystal (Rowe & Baird, 2007). These contributions augment the interfacial tension, $\gamma_o$, which appears in Eq.(2). The contributions, which are both energetic and entropic are:

(1) The surface charges on the crystal attract ions of opposite sign in the growth solution and repel ions of like sign. This displacement of ions the away from their original random positions in the growth solution has the effect of decreasing the entropy at the crystal/solution interface. The decrease in entropy increases the free energy.

(2) By attracting oppositely charged ions and repelling like charged ions, the crystal surface builds up electrostatic energy. This stored electrostatic energy makes a negative contribution to the free energy.

(3) According to Eqs. (4) and (6), when the value of the pH equals the value of the $pK_a$ of an ionizable group on the surface, half of the sites of this particular type will be in the charged form and half will be in the uncharged form. In this situation, $H^+$ ions can hop between charged sites and uncharged sites. This mixing of charged and uncharged sites increases the entropy of the crystal surface and makes a negative contribution to the free energy.

When the above three contributions are taken into account, the free energy of dissolution of the crystal assumes the form (Hodge, 2014)

$$\Delta G(a) = (\omega + 1)k_B T \ln\left(\frac{m(a)}{m_s}\right) - \frac{2v}{a}\left[\bar{\gamma} - \frac{a\sigma^2(3+2\kappa a)}{4\varepsilon\varepsilon_o(1+\kappa a)^2}\right] \qquad (7)$$

where the overall interfacial energy, $\bar{\gamma}$, is given by (Rowe & Baird, 2007)

$$\bar{\gamma} = \gamma_o + k_B T \Gamma \ln(1-\theta) \qquad (8)$$

In Eq.(7), $\omega$ is the number of charged groups on a protein molecule, and $\sigma$ is the charge density on the surface of the crystal. Both $\omega$ and $\sigma$ are evaluated at the prevailing pH. The other

electrostatic parameters are $\varepsilon$, the dielectric constant of water; $\varepsilon_o$, the dielectric constant of vacuum; and $\kappa$, the reciprocal of the Debye length. In the case of an aqueous solution at $T$ = 298 K with NaCl serving as the electrolyte,

$$\kappa = 3.29x10^{9}\ \mathrm{m^{-1}\ L^{1/2}\ mol^{-1/2}}\ \sqrt{c} \qquad (9)$$

where $c$ is the molar concentration of NaCl (Rowe & Baird, 2007). In Eq. (8), we have assumed that the growth solution pH is close to the $pK_a$ value of one of the ionizable groups on the surface of the crystal. The surface density of this group is $\Gamma$. All the other groups with $pK_a$ values less than this value of the pH will be present in their basic forms, either B or $A^-$. All ionizable groups with $pK_a$ values greater than this pH will be present in their acid forms, either $BH^+$ of HA. The ionizations of groups with $pK_a$ values distant from the growth solution pH are saturated and essentially independent of the pH. The contributions of these groups to the free energy are included in the surface charge density, $\sigma$, and also in the background interfacial tension, $\gamma_o$.

Eq.(7) was derived by solving the Poisson-Boltzmann equation in the Debye-Huckel approximation (Rowe & Baird, 2007), which limits its quantitative application to electrolytes with concentrations, $c < 0.1$ M A numerical integration of the Poisson-Boltzmann equation would be required to extend the range of applicability of the theory to concentrations in excess of 0.1 M. Although $c$ = 0.1 M is about a factor of ten below that ordinarily employed in crystal growth solutions, we can still draw some qualitative conclusions by comparing Eqs. (2) and (7).

Our discussion of Eq.(2) indicates that an uncharged crystal of microscopic size (i.e. $a << \infty$) cannot be in equilibrium ($\Delta G(a) = 0$) with a growth solution where the ambient concentration, $m(a)$, equals the bulk solubility, $m_s$. Such a crystal will dissolve. The molecules released will diffuse through the growth solution and precipitate onto the surfaces of any larger crystals which are present. As explained above, this process is called Ostwald ripening. By

contrast, when $\sigma \neq 0$, Eq.(7) predicts that a crystal of microscopic size, $a << \infty$, can be at equilibrium ($\Delta G(a) = 0$) with an ambient concentration, $m(a)$, equal to the bulk solubility, $m_s$. To find this condition, we set the left hand side of Eq.(7) to zero and let $m(a) = m_s$ on the right. These restrictions force the bracket on the right hand side of Eq.(7) to be zero. The value of $a$ which makes the bracket zero is (Hodge, 2014),

$$a = \frac{1}{\kappa}\left[\frac{(2-3\beta)+\sqrt{9\beta^2-4\beta}}{2(2\beta-1)}\right] \tag{10}$$

where

$$\beta = \frac{\sigma^2}{4\varepsilon\varepsilon_o\overline{\gamma}\kappa} \tag{11}$$

Eq.(10) predicts the size, $a$, of a protein crystal, which is in equilibrium with a growth solution where the protein concentration is equal to the bulk solubility, $m_s$. Since the ambient concentration equals the bulk solubility, such a solution cannot support any *net* exchange of mass with crystals of this size. *In the absence of a net exchange of mass, the process of Ostwald ripening is frustrated, which may explain why Ostwald ripening is not well documented in protein crystallization*. Until one or more of the three variables, pH, temperature, and salt concentration are changed, the crystal remains stuck in equilibrium with the growth solution.

It is apparent from Eq.(10) that $a > 0$ for $\beta > 1/2$. Experimentally, the values of $a$ predicted by Eq.(10) make sense, however, only within the range $(3v/4\pi)^{1/3} < a < a_{\max}$, where $(3v/4\pi)^{1/3}$ is an estimate of the size of a protein molecule, and $a_{\max}$ is the size of a single crystal consisting of all of the protein which at the beginning of the experiment was in excess of $m_s V$, where $V$ is the volume of the growth solution. At $\beta = 1/2$, Eq.(10) has a singular point where

$a$ gets very large. For values of $\beta$ only slightly greater than ½, $a$ diminishes rapidly, numerically reaching the molecular limit, $a = (3v/4\pi)^{1/3}$ as $\beta$ approaches unity. Hence, crystals of macroscopic size are to be expected only for those values of $\beta$ which are in close proximity to $\beta = 1/2$. *This tight restriction may explain why large protein crystals are so rare.*

At temperature $T = 298$ K, the dielectric constant of water is $\varepsilon = 78.54$ (Rowe & Baird, 2007). If we consider lysozyme as a model protein, then for a NaCl concentration of $c = 0.05$ M, the value of $\sigma$ will be no greater than 2 $\mu C/cm^2$ (Kim *et al*., 2003). When Eq.(9) is used to calculate $\kappa$, and the values of $\kappa$, $T$, $\varepsilon$, and $\sigma$ are substituted into Eq.(11), we find $\beta = 1/2$ when $\overline{\gamma} = 0.4$ mJ/m$^2$. This estimate of $\overline{\gamma}$ is close to the values of the interfacial energy calculated from the results of experiments where the rate of advance of a macroscopic crystal facet as determined optically (Malkin et al. 1999). According to Eq.(8), this moderate value of $\overline{\gamma}$ can be expected to occur for values of $\theta$ between the limits $\theta = 0$, where $\overline{\gamma}$ has its maximum value, $\overline{\gamma} = \gamma_o$, and $\theta = 1$, where $\overline{\gamma} = -\infty$. The latter is an unphysical value of $\overline{\gamma}$, corresponding to an ionizable group with $pK_a$ value far from the prevailing pH. As pointed out, the contributions of such groups are not represented by Eq.(8) but are included within the parameters $\sigma$ to the $\gamma_o$. In order to achieve $\beta = 1/2$, a value of $\theta$ near ½ would seem to be a good compromise. According to Eqs.(4) and (6), the value $\theta = 1/2$ occurs when the pH equals the $pK_a$ of one of the ionizable groups on the protein molecule. *Hence, in the absence of complications, such as the effect of pH on nucleation, Eq.(10) makes the simple prediction that to obtain large protein crystals, the pH of the growth solution should be adjusted to a value which is close to the $pK_a$ of one the ionizable groups on the protein molecular chain.*

The theory takes into account explicitly only the effects of protein molecular charge on crystal size. The effects of the short range van der Waals forces acting between the protein molecules and the crystal are subsumed within the interfacial energy parameter, $\gamma_o$. This simple approximation is forced on the theory because the details of the distance dependence of the van der Waals forces cannot be treated analytically within the Debye-Huckel approximation that gives rise to Eq.(7). Nevertheless, for all distances short of molecular contact, the electrostatic forces can be expected to dominate the van der Waals interactions. For this reason, the electrostatic theory can be expected to have some predictive power albeit perhaps only qualitative. Finally, for the case where $m(a)$ is not too much greater than $m_s$, the theory predicts additional values of $a$ besides the one given by Eq.(10) which make $\Delta G(a) = 0$. The conditions appropriate for this additional case must be determined numerically and are currently under investigation.

**Conclusion**

Funding agencies from different countries have invested considerable resources in building structural repositories of three-dimensional protein structures (over 90,000 to-date), yet 50% of the atoms in these structures are comprised of hydrogen and they are invisible in most X-ray structures. Therefore the role of hydrogen atoms in protein structures is severely understudied. The precise location of hydrogen atoms cannot be experimentally determined by biochemical methods or by X-ray crystallography in the average resolution range but can be definitively located using neutron diffraction. Despite the development and renovation of powerful neutron sources, protein crystal size remains a bottleneck for NMC. The knowledge of current

techniques to grow large crystals and understanding how do they grow will be critical to future neutron crystallographic studies.

**Acknowledgements**

Part of the information acquired here was from research supported by the Center for the Advancement of Science in Space (CASIS), *GA-2013-100* and sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle LLC for the US Department of Energy under Contract No. DE-AC05-00OR22725. We thank Mr. Jorge Barcena for his assistance in the preparation of this manuscript.

**References**

Anand, K., Pal, D. & Hilgenfeld, R. (2002). *An overview on 2-methyl-2,4-pentanediol in crystallization and in crystals of biological macromolecules*. *Acta Crystallographica Section D* **58**, 1722-1728.

Asherie, N. (2004). *Protein crystallization and phase diagrams*. *Methods* **34**, 266-272.

Baird, J. K. & Kim, Y. W. (2002). *Theory of the nucleation of protein macromolecular ions*. *Mol. Phys.* **100**, 1855-1866.

Benvenuti, M. & Mangani, S. (2007). *Crystallization of soluble proteins in vapor diffusion for x-ray crystallography*. *Nat Protoc* **2**, 1633-1651.

Bergfors, T. (2007). *Screening and optimization methods for nonautomated crystallization laboratories*. *Methods Mol. Biol* **363**, 131-151.

Blakeley, M. P., Kalb, A. J., Helliwell, J. R. & Myles, D. A. (2004). *The 15-K neutron structure of saccharide-free concanavalin A*. *Proc Natl Acad Sci U S A* **101**, 16405-16410.

Blakeley, M. P., Langan, P., Niimura, N. & Podjarny, A. (2008). *Neutron crystallography: opportunities, challenges, and limitations*. *Current Opinion in Structural Biology* **18**, 593-600.

Borgstahl, G. E., Pokross, M., Chehab, R., Sekher, A. & Snell, E. H. (2000). *Cryo-trapping the six-coordinate, distorted-octahedral active site of manganese superoxide dismutase*. *Journal of Molecular Biology* **296**, 951-959.

Budayova-Spano, M., Dauvergne, F., Audiffren, M., Bactivelane, T. & Cusack, S. (2007). *A methodology and an instrument for the temperature-controlled optimization of crystal growth*. *Acta Crystallographica Section D* **63**, 339-347.

Cacioppo, E., Munson, S. & Pusey, M. L. (1991). *Protein solubilities determined by a rapid technique and modification of that technique to a micro-method*. *Journal of Crystal Growth* **114**, 286-292.

Carotenuto, L., Piccolo, C., Castagnolo, D., Lappa, M., Tortora, A. & Garcia-Ruiz, J. M. (2002). *Experimental observations and numerical modelling of diffusion-driven crystallisation processes*. *Acta Crystallographica Section D* **58**, 1628-1632.

Chayen, N. E., Shaw-Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *An automated system for micro-batch protein crystallization and screening*. *J. Appl. Cryst.* **23**, 297-302.

Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. & Cooper, J. B. (2001). *A neutron Laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism*. *Biochemistry* **40**, 13149-13157.

Coates, L., Stoica, A. D., Hoffmann, C., Richards, J. & Cooper, R. (2010). *The macromolecular neutron diffractometer (MaNDi) at the Spallation Neutron Source, Oak Ridge: enhanced optics design, high-resolution neutron detectors and simulated diffraction*. *J. Appl. Cryst.* **43**, 570-577.

Coates, L., Tomanicek, S. J., Schrader, T. E., Weiss, K. L., Ng, J. D. & Ostermann, A. (2014). *Cryogenic neutron protein crystallography: Routine methods and potential benefits*. *J. Appl. Cryst.* **47**, 1431-1434.

Cudney, B. & Patel, S. (1994). *Crystallization as a tool for bioseparation. Am Biotechnol Lab* **12**, 42.

Day, J. & McPherson, A. (1992). *Macromolecular crystal growth experiments on International Microgravity Laboratory-1*. *Protein Science* **1**, 1254-1268.

DeLucas, L. J., Moore, K. M. & Long, M. M. (1999). *Protein crystal growth and the International Space Station*. *Gravitational and space biology bulletin : publication of the American Society for Gravitational and Space Biology* **12**, 39-45.

Ducruix, A. & Giege, R. (1999). *Crystallization of nucleic acids and proteins:A practical approach*. *Oxford University Press*.

Dunuwila, D. D. & Berglund, K. A. (1997). *ATR FTIR spectroscopy for in situ measurement of supersaturation*. *Journal of Crystal Growth* **179**, 185-193.

Ewing, F., Fortsythe, E. & Pusey, M. L. (1994). *Orthorhombic lysozyme solubility*. *Acta Crystallographica Section D* **50**, 424-428.

Feigelson, R. S. (1988). *The relevance of small molecule crystal growth theories and techniques to the growth of biological macromolecules*. *J. Cryst. Growth* **90**, 1-13.

García-Ruiz, J.M. and Otálora, F. (1997). Crystal growth studies in microgravity with the APCF. II: Image analysis study. *Journal of Crystal Growth* **182,** 155-167.

García-Ruiz, J. M. (2003). *Counterdiffusion methods for macromolecular crystallization*. *Methods in Enzymology* **368**, 130-154.

García-Ruiz, J. M. & Ng, J. D. (2006). *Protein crystallization strategies for structural genomics*. edited by N. E. Chayen, p. Chapter 5: International University Line.

García-Ruiz, J. M., Otalora, F., Novella, M.L., Gavira, J.A., Sauter & Vidal, O. (2001). *A supersaturation wave of protein crystallization. J. Cryst. Growth* **232**, 149-155.

Garman, E. & Owen, R. L. (2007). *Cryocrystallography of macromolecules: practice and optimization*. *Methods Mol Biol* **364**, 1-18.

Garman, E. F. & Doublie, S. (2003). *Cryocooling of macromolecular crystals: optimization methods*. *Methods in Enzymology* **368**, 188-216.

Garman, E. F. & McSweeney, S. M. (2007). *Progress in research into radiation damage in cryo-cooled macromolecular crystals*. *Journal of Synchrotron Radiation* **14**, 1-3.

Gavira, J. A., Toh, D., Lopez-Jaramillo, J., García-Ruiz, J. M. & Ng, J. D. (2002). *Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling*. *Acta Crystallographica Section D* **58**, 1147-1154.

Gray, R. J., Hou, W. B., Kudryavtsev, A. B. & DeLucas, L. J. (2001). *A new approach to the measurement of protein solubility by Michaelson interferometry*. *Jouranal of Crystal Growth* **232**, 10-16.

Haire, L. F. & Blow, D. M. (2001). *A novel spin filter method for the measurement of solubility*. *Journal of Crystal Growth* **232**, 17-20.

Hope, H. (1988). *Cryocrystallography of biological macromolecules: a generally applicable method*. *Acta Crystallographica Section B* **44 ( Pt 1)**, 22-26.

Howard, E. I., Blakeley, M. P., Haertlein, M., Petit-Haertlein, I., Mitschler, A., Fisher, S. J., Cousido-Siah, A., Salvay, A. G., Popov, A., Muller-Dieckmann, C., Petrova, T. & Podjarny, A. (2011). *Neutron structure of type-III antifreeze protein allows the reconstruction of AFP-ice interface*. *Journal of Molecular Recognition* **24**: 724-732.

Hughes, R. C., Coates, L., Byrne-Steele, M. L., Tomanicek, S., Kovalevsky, A., Langan, P. & Ng, J. D. (2009). *Inorganic pyrophosphatase crystals from Thermococcus thioreducens for X-ray and neutron diffraction*. *Acta Crystallographica Section F* **68,** 1482-1487

Iwai, W., Yagi, D., Ishikawa, T., Ohnishi, Y., Tanaka, I. & Niimura, N. (2008). *Crystallization and evaluation of hen egg-white lysozyme crystals for protein pH titration in the crystalline state*. *Journal of Synchrotron Radiation* **15**, 312-315.

Juers, D. H. & Matthews, B. W. (2004a). *The role of solvent transport in cryo-annealing of macromolecular crystals*. *Acta crystallographica. Section D* **60**, 412-421.

Juers, D. H. & Matthews, B. W. (2004b). *Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization*. *Quarterly Reviews of Biophysics* **37**, 105-119.

Juers, D. H. & Weik, M. (2011). *Similarities and differences in radiation damage at 100 K versus 160 K in a crystal of thermolysin*. *Journal of Synchrotron Radiation* **18**, 329-337.

Kim, Y. W., Barlow, D. A., Caraballo, K. G. & Baird, J. K. (2003). *Kinetics of supersaturation decay in the crystallization of lysozyme*. *Molecular Physics* **101**, 2677-2686.

Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Structure determination of a novel protein by sulfur SAD using chromium radiation in combination with a new crystal-mounting method*. *Acta Crystallographica Section D* **61**, 1013-1021.

Lee, H.-M., Kim, Y. W. & Baird, J. K. (2001). *Electrophoretic mobility and zeta-potential of lysozyme crystals in aqueous solutions of some 1 : 1 electrolytes*. *Journal of Crystal Growth* **232**, 294-300.

Loll, P. J. (2014). *Membrane proteins, detergents and crystals: what is the state of the art? Acta crystallographica. Acta Crystallographica Section F* **70**, 1576-1583.

Lorber, B., Theobald-Dietrich, A., Charron, C., Sauter, C., Ng, J. D., Zhu, D. W. & Giege, R. (2002). *From conventional crystallization to better crystals from space: a review on pilot crystallogenesis studies with aspartyl-tRNA synthetases*. *Acta Crystallographica Section D* **58**, 1674-1680.

Malkin, A. J. & Thorne, R. E. (2004). *Growth and disorder of macromolecular crystals: insights from atomic force microscopy and X-ray diffraction studies*. *Methods* **34**, 273-299.

McPherson, A. (1999). *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press.

McPherson, A. & Cudney, B. (2014). *Optimization of crystallization conditions for biological macromolecules*. *Acta Crystallographica Section F* **70**, 1445-1467.

Meilleur, F., Munshi, P., Robertson, L., Stoica, A. D., Crow, L., Kovalevsky, A., Koritsanszky, T., Chakoumakos, B. C., Blessing, R. & Myles, D. A. (2013). *The IMAGINE instrument: first neutron protein structure and new capabilities for neutron macromolecular crystallography*. *Acta Crystallographica Section D* **69**, 2157-2160.

Miers, H. A. & Isaac, F. (1907). *The spontaneous crystallization of binary mixtures: experiments on salol and betol. Proc. R. Soc. London A* **79**, 322.

Mullin, J. W. (1993). *Crystallization. 3rd ed Butterworth-Heinemann, Oxford*.

Nakazato, K. (2008). *Method of measuring protein solubility, process for producing crystal and apparatus therefor. U.S. Patent* **No. 7,416,708 B2**.

Ng, J. D. (2002). *Space-grown protein crystals are more useful for structure determination. Ann NY Acad Sci* **974**, 598-609.

Ng, J. D., Clark, P. J., Stevens, R. C. & Kuhn, P. (2008). *In situ X-ray analysis of protein crystals in low-birefringent and X-ray transmissive plastic microchannels. Acta Crystallographica Section D* **64**, 189-197.

Ng, J. D. & Garcia-Ruiz, J. M. (2006). *Counter-diffusion capillary crystallization for structural genomics. Trends in Drug Discovery*.

Ng, J. D., Gavira, J. A. & García-Ruiz, J. M. (2003). *Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. Journal of Structural Biology* **142**, 218-231.

Ng, J. D., Kuznetsov, Y. G., Malkin, A. J., Keith, G., Giege, R. & McPherson, A. (1997). *Visualization of RNA crystal growth by atomic force microscopy. Nucleic Acids Research* **25**, 2582-2588.

Ng, J. D., Lorber, B., Giege, R., Koszelak, S., Day, J., Greenwood, A. & McPherson, A. (1997). *Comparative analysis of thaumatin crystals grown on earth and in microgravity. Acta Crystallographica Section D* **53**, 724-733.

Ng, J. D., Lorber, B., Witz, J., Théobald-Dietrich, A., Kern, D. & Giegé, R. (1996). *The crystallization of biological macromolecules from precipitates: evidence for Ostwald ripening. Jouranal of Crystal Growth* **168**, 50-62.

Ng, J. D., Sauter, C., Lorber, B., Kirkland, N., Arnez, J. & Giege, R. (2002). *Comparative analysis of space-grown and earth-grown crystals of an aminoacyl-tRNA synthetase: space-grown crystals are more useful for structural determination. Acta Crystallographica Section D* **58**, 645-652.

Ng, J. D., Stevens, R. C. & Kuhn, P. (2008). *Protein crystallization in restricted geometry: advancing old ideas for modern times in structural proteomics. Methods in Molecular Biology* **426**, 363-376.

Niimura, N. (2010). *Neutron protein crystallography: Hydrogen and hydration in proteins*. Springer.

Niimura, N. & Bau, R. (2008). *Neutron protein crystallography: beyond the folding structure of biological macromolecules. Acta Crystallographica Section A* **64**, 12-22.

Oksanen, E., Blakeley, M. P., Bonnete, F., Dauvergne, M. T., Dauvergne, F. & Budayova-Spano, M. (2009). *Large crystal growth by thermal control allows combined X-ray and neutron crystallographic studies to elucidate the protonation states in Aspergillus flavus urate oxidase*. *J R Soc Interface* **6 Suppl 5**, S599-610.

Ostwald, W. (1897). *Besprechung der Arbeit von Liesenganga A-Linien. Z. Physical Chemistry* **23**, 365.

Otálora, F. and García-Ruiz, J.M. (1997). Crystal growth studies in microgravity with the APCF. I: Computer simulation of transport dynamics. *Journal of Crystal Growth* **182** 141-154.

Otalora, F., Gavira, J. A., Ng, J. D. & García-Ruiz, J. M. (2009). *Counterdiffusion methods applied to protein crystallization*. *Progress in Biophysics and Molecular biology* **101**, 26-37.

Rakel, N., Baum, M. & Hubbuch, J. (2014). *Moving through three-dimensional phase diagrams of monoclonal antibodies*. *Biotechnol Prog* **30**, 1103-1113.

Rayment, I. (2002). *Small-scale batch crystallization of proteins revisited: an underutilized way to grow large protein crystals*. *Structure* **10**, 147-151.

Rowe, J. D. & Baird, J. K. (2007). *Reduced capillary length scale in the application of Ostwald Ripening theory to the coarsening of charged colloidal crystals in electrolyte solutions*. *International Journal of Thermophysics* **28**, 855-864.

Rupp, B. (2015). *Origin and use of crystallization phase diagrams*. *Acta Crystallographica Section F* **71**, 247-260.

Sauter, C., Lorber, B., Kern, D., Cavarelli, J., Moras, D. & Giege, R. (1999). *Crystallogenesis studies on yeast aspartyl-tRNA synthetase: use of phase diagram to improve crystal quality*. *Acta Crystallographica Section D* **55**, 149-156.

Sazaki, G., Kurihara, K., Nakada, T., Miyashita, S. & Komatsu, H. (1996). *A novel approach to the solubility measurement of protein crystals by two-beam interferometry*. *Journal of Crystal Growth* **169**, 355-360.

Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000). *Enhanced visibility of hydrogen atoms by neutron crystallography on fully deuterated myoglobin*. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3872-3877.

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985). *Measurement of protein using bicinchoninic acid*. *Analytical Biochemistry* **150**, 76-85.

Streets, A. M. & Quake, S. R. (2010). *Ostwald ripening of clusters during protein crystallization*. *Phys Rev Lett* **104**, 178102.

Teixeira, S. C., Ankner, J., Bellissent-Funel, M. C., Bewley, R., Blakeley, M. P., Coates, L., Dahint, R., Dalgliesh, R., Dencher, N., Dhont, J., Fischer, P., Forsyth, V. T., Fragneto, G., Frick, B., Geue, T., Gilles, R., Gutberlet, T., Haertlein, M., Hauss, T., Haussler, W., Heller, W. T., Herwig, K., Holderer, O., Juranyi, F., Kampmann, R., Knott, R., Kohlbrecher, J., Kreuger, S., Langan, P., Lechner, R., Lynn, G., Majkrzak, C., May, R., Meilleur, F., Mo, Y., Mortensen, K., Myles, D. A., Natali, F., Neylon, C., Niimura, N., Ollivier, J., Ostermann, A., Peters, J., Pieper, J., Ruhm, A., Schwahn, D., Shibata, K., Soper, A. K., Straessle, T., Suzuki, U. I., Tanaka, I., Tehei, M., Timmins, P., Torikai, N., Unruh, T., Urban, V., Vavrin, R., Weiss, K. & Zaccai, G. (2008). *New sources and instrumentation for neutrons in biology*. *Chem Phys* **345**, 133-151.

Veesler, S., Marcq, S., Lafont, S., Astier, J. P. & Boistelle, R. (1994). *Influence of polydispersity on protein crystallization: a quasi-elastic light-scattering study applied to alpha-amylase*. *Acta Crystallographica Section D* **50**, 355-360.

Weber, I. T., Waltman, M. J., Mustyakimov, M., Blakeley, M. P., Keen, D. A., Ghosh, A. K., Langan, P. & Kovalevsky, A. Y. (2013). *Joint X-ray/neutron crystallographic study of HIV-1 protease with clinical inhibitor amprenavir: insights for drug design*. *Journal of Medicinal Chemistry* **56**, 5631-5635.

**TABLE AND FIGURE LEGENDS**

**Table 1. Crystallographic structures determined to-date by neutron crystallography**

**Figure 1. Schematic phase diagram and solubility measurements.** Panel A displays the solubility diagram of a protein versus the concentration of precipitant. The solubility curve (solid line) where the waiting time for nucleation is infinite and the supersolubility curve (dashed line) where the waiting time for nucleation is one, are the boundaries of the metastable region. The arrows illustrate the pathway of a solution that crystallize by typical vapor diffusion experiments. The blue dot marks the condition at the beginning of the equilibration process. Note that in vapor diffusion experiment the solution moves out of equilibrium until supersaturation is reached. The labile zone is where nucleation occurs and the metastable zone is defined by the absence of nucleation but the crystal will continue to grow. Beyond the labile zone is the precipitation area where proteins will amorphously come out of solution. The quality and size of the crystals depends on how

fast the solution moves across the metastable zone, i.e. depends on the rate of change of supersaturation, a variable that it is difficult to control. Panel B describes a counterdiffusion experiment in which the precipitant solution diffuses across a long capillary containing the protein solution. This creates a pattern of supersaturation values across the capillary that varies from very high (thus, a high flow of nucleation of amorphous particles or small crystals) to very low (thus forming very few crystal of large size) as shown by the subsequent pathways at different time shown in the solubility diagram. Note that in this case the solution moves towards equilibrium, towards low supersaturation located in the metastable region, thus creating large crystals at the end of experiment. The case is exemplified with the growth of thaumatin crystals grow by counterdiffusion inside microfluidic cell (Ng et al., 2008).

**Figure 2. Schematic diagram of capillary configuration for large volume crystal growth.** The capillary, used as growth cell, is filled with protein solution (pink) to about half the volume followed by an equal volume of the solution of the precipitating agent (PA). Both ends of the capillary are sealed with soft wax (W) and enamel varnish (red). $T_0$ to $T_6$ indicate time progression within a span of about 30 days of equilibration. Actual photograph of the entire tube and crystals of the IPP crystals grown out of a supersaturation gradient are shown in black and white after two days and ten days of the start of the experiment. The left panel shows the case with no initial barrier between the protein and precipitant solutions. The right panel shows the equilibration with a transparent agarose plug (APlg) separating the protein and the precipitant solution.

**Figure 3. Procedure for deuterium exchange in capillary.** Large volume crystals grown or mounted in a large diameter (> 0.3mm) quartz capillary can undergo $D_2O$ exchange quite readily. In the case where the crystals are grown *in situ*, the capillary ends can be excised and the solutions withdrawn one end at a time and replaced with precipitant agent (PA) in $D_2O$ (A). The exchanging $D_2O$ solution is placed close to the targeted crystal leaving an air gap for vapor diffusion. When there is an agarose plug separating the precipitant and protein solution, the precipitant agent is simply replaced with the $D_2O$ solution filled against the agarose plug while the other end is left empty (B). $D_2O$ exchanges can also be directly diffuse into the existing precipitant solution with a solution containing deuteriated precipitating agent and protein (C).

**Figure 4. Space- and Earth-grown IPPase crystals within quartz capillaries.** Capillaries of quartz containing crystals grown under microgravity (A) and on Earth (B). All capillaries are 50mm in length and have an inner diameter of 2mm. In both cases the starting precipitating solution is located at the left. Both space and Earth crystallization experiments were prepared with the same proteins, concentration, loading protocol, location, laboratory and environmental conditions. For a canonical counterdiffusion experiment, the protein solution should diffuse from left to right creating a supersaturation wave that moves across the capillary with increasing width and decreasing amplitude (Garcıa-Ruiz & C., 2001). Thus, the number of crystals should decrease and the size of the crystals should increase along the length of the capillary (left to right). This spatial-temporal gradient of supersaturation is clearly observed in the capillaries of the microgravity grown crystals (A) while it is not observed in the capillaries with Earth-grown crystals. The reason is that the capillaries with 2mm in diameter are too large

to reduce convective flow, as expected from simple fluid dynamics considerations. Typical crystals grown in space are shown under polarized light(C).

**Table 1. Crystallographic structures determined to-date by neutron crystallography**

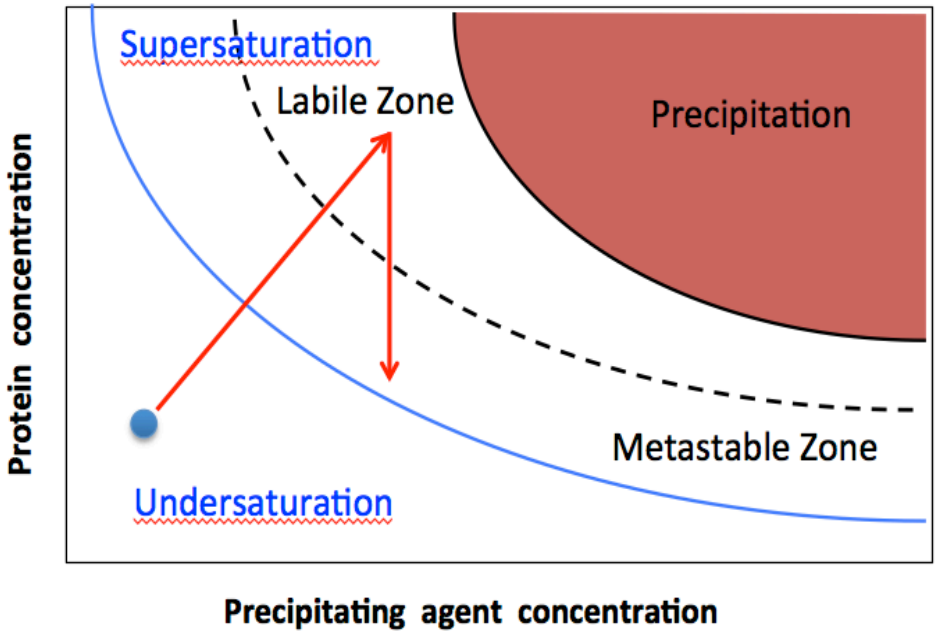
| | Macromolecule | PDB ID | Space Group | Resolution (Å) |
|---|---|---|---|---|
| 1 | Achromobacter protease I | 4GPG | P 1 | 1.90 |
| 2 | Antifreeze protein (Type-III) | 3QF6 | P $2_1$ $2_1$ $2_1$ | 1.85 |
| 3 | Apo-D-Xylose Isomerase | 3KCJ | I 2 2 2 | 1.80 |
| 4 | Aspartic proteinase | 2JJI | P$2_1$ | 1.57 |
| 5 | Aspartic proteinase | 2JJJ | P$2_1$ | 1.00 |
| 6 | B-DNA decamer d(CCATTAATGG)2 | 1WQZ | P$2_1$ | 3.00 |
| 7 | Beta-lactamase -Toho-1 R274N R276N double mutant (Perdeuterated) | 2XQZ | P $3_2$ 2 1 | 2.10 |
| 8 | Beta-lactamase Toho-1 E166A R274N R276N triple mutant | 2WYX | P $3_2$ 2 1 | 2.10 |
| 9 | Beta-lactamase Toho-1 R274N R276N double mutant (Perdeuterated) in complex with a fully deuterated boronic acid (BZB) | 4C3Q | P $3_2$ 2 1 | 2.20 |
| 10 | Bovine pancreatic ribonuclease A | 3A1R | P $2_1$ | 1.70 |
| 11 | Bovine pancreatic trypsin inhibitor | 5PTI | P $2_1$ $2_1$ $2_1$ | 1.00 |
| 12 | Carbonmonoxymyoglobin | 2MB5 | P $2_1$ | 1.80 |
| 13 | Concanavalin A | 1C57 | I 2 2 2 | 2.40 |
| 14 | Concanavalin A | 2YZ4 | I 2 2 2 | 2.20 |
| 15 | Crambin | 4FC1 | P $2_1$ | 1.10 |
| 16 | Crambin (H/D exchanged) | 3U7T | P $2_1$ | 0.85 |
| 17 | Cytochrome c peroxidase | 3R98 | P $2_1$ $2_1$ $2_1$ | 2.40 |
| 18 | Cytochrome c Peroxidase (Compound I intermediate) | 4CVJ | P $2_1$ $2_1$ $2_1$ | 2.50 |
| 19 | D-Xylose isomerase | 2GVE | I 2 2 2 | 2.20 |

| | | | | |
|---|---|---|---|---|
| 20 | D-xylose isomerase in complex with 2Ni2+ and per-deuterated D-sorbitol | 4DVO | I 2 2 2 | 2.00 |
| 21 | D-Xylose Isomerase in complex with two Cd2+ cations and d12-D-alpha-glucose | 3KCL | I 2 2 2 | 2.00 |
| 22 | D-Xylose Isomerase in complex with two Ni2+ cations and d12-D-glucose | 3KCO | I 2 2 2 | 1.80 |
| 23 | Deoxy human hemoglobin | 2DXM | P $2_1$ | 2.10 |
| 24 | Dihydrofolate Reductase (*E. coli)* Bound to the Anti-cancer drug, Methotrexate | 2INQ | P $6_1$ | 2.20 |
| 25 | Diisopropyl fluorophosphatase | 3BYC | P $2_1$ $2_1$ $2_1$ | 2.20 |
| 26 | Diisopropyl fluorophosphatase (DFPase) (Perdeuterated) | 3KGG | P $2_1$ $2_1$ $2_1$ | 2.10 |
| 27 | Dissimilatory sulfite reductase D (DsrD) | 1WQ2 | P $2_1$ $2_1$ $2_1$ | 2.40 |
| 28 | Endothiapepsin complexed with transition state analogue inhibitor H261 | 1GKT | P $2_1$ | 2.10 |
| 29 | Endothiapepsin in complex with a gem- diol inhibitor | 2VS2 | P $2_1$ | 2.00 |
| 30 | Ferric Cytochrome c peroxidase | 4CVI | P $2_1 2_1 2_1$ | 2.41 |
| 31 | Get5 ubiquitin-like domain | 4GOC | P 21 21 21 | 2.40 |
| 32 | Hen egg-white lysozyme | 1LZN | P1 | 1.70 |
| 33 | HIV-1 protease in complex with clinical inhibitor amprenavir (Perdeuterated) | 4JEC | P $2_1$ $2_1$ 2 | 2.00 |
| 34 | HIV-1 Protease in Complex with Potent Inhibitor KNI-272 | 2ZYE | P $2_1$ $2_1$ 2 | 1.90 |
| 35 | Human Carbonic anhydrase II | 3KKX | P $2_1$ | 2.00 |
| 36 | Human transthyretin | 3U2J | P $2_1 2_1 2$ | 2.00 |
| 37 | Human transthyretin | 3U2I | P $2_1 2_1 2$ | 1.70 |
| 38 | Human transthyretin (TTR) at room temperature | 4PVM | P $2_1$ $2_1$ 2 | 2.00 |
| 39 | Inorganic pyrophosphatase | 3Q3L | C 2 | 2.50 |
| 40 | Insulin | 2ZPP | I $2_1$ 3 | 2.50 |
| 41 | Insulin | 3INS | H 3 | 1.50 |
| 42 | Insulin (Porcine 2$Zn^{2+}$) | 3FHP | H 3 | 2.00 |

| | | | | |
|---|---|---|---|---|
| 43 | Leucine and valine methyl protonated type III antifreeze | 4NY6 | P $2_12_12_1$ | 1.85 |
| 44 | Lysozyme in $D_2O$ (Wild type human) | 2ZWB | P $2_12_12_1$ | 1.80 |
| 45 | Magnesium activated Inorganic pyrophosphatase | 3Q46 | H 3 2 | 0.99 |
| 46 | Major Urinary Protein I | 1I05 | P $4_3$ $2_1$ 2 | 2.00 |
| 47 | Myoglobin (Fully perdeuterated) | 1CQ2 | P $2_1$ | 2.00 |
| 48 | Oxidized Amicyanin | 3L45 | P $2_1$ | 1.80 |
| 49 | Oxymyoglobin | 1MBD | P $2_1$ | 1.40 |
| 50 | Photoactive yellow protein (PYP) | 2ZOI | P $6_3$ | 1.50 |
| 51 | Photoactive yellow protein (PYP) | 2QWS | P $6_3$ | 2.50 |
| 52 | PKGIbeta in complex with cGMP | 4QXK | P $4_1$ $2_1$ 2 | 2.20 |
| 53 | Porcine pancreatic elastase complexed with a potent peptidyl inhibitor FR130180 | 3HGN | P $2_1$ $2_1$ $2_1$ | 1.65 |
| 54 | Ribonuclease A with uridine vanadate | 6RSA | P $2_1$ | 2.00 |
| 55 | Rubredoxin (Fully perdeuterated) | 3KYX | P $2_1$ $2_1$ $2_1$ | 1.68 |
| 56 | Rubredoxin (Fully perdeuterated) | 3KYY | P $2_1$ $2_1$ $2_1$ | 1.10 |
| 57 | Rubredoxin (Oxidized form perdeuterated from *Pyrococcus furiosus*) | 4AR3 | P $2_12_12_1$ | 1.05 |
| 58 | Rubredoxin (Perdeuterated) | 4K9F | P $2_12_12_1$ | 1.75 |
| 59 | Rubredoxin (Wild Type) from *Pyrococcus Furiosus* | 1VCX | P $2_1$ $2_1$ $2_1$ | 1.50 |
| 60 | Rubredoxin from *Pyrococcus furiosus* (Perdeuterated) | 4AR4 | P $2_1$ $2_1$ $2_1$ | 1.38 |
| 61 | Rubredoxin mutant | 1IU5 | P $2_1$ $2_1$ $2_1$ | 1.50 |
| 62 | Rubredoxin mutant from *Pyrococcus Furiosus* | 1IU6 | P $2_1$ $2_1$ $2_1$ | 1.60 |
| 63 | Saccharide-free concanavalin A | 1XQN | I 2 2 2 | 2.50 |
| 64 | Sperm whale met-myoglobin | 1L2K | P $2_1$ | 1.50 |
| 65 | Streptomyces rubiginosus D-xylose isomerase in complex with two $Cd^{2+}$ ions and cyclic beta-L-arabinose | 4QDP | I 2 2 2 | 2.00 |
| 66 | Streptomyces rubiginosus D-xylose isomerase in complex with two $Ni^{2+}$ ions and linear L-arabinose | 4QDW | I 2 2 2 | 1.80 |

| | | | | |
|---|---|---|---|---|
| 67 | Trypsin | 1NTP | P $2_1$ $2_1$ $2_1$ | 1.80 |
| 68 | Trypsin complexed with BPTI (Bovine Pancreatic Trypsin Inhibitor) | 3OTJ | I 2 2 2 | 2.15 |
| 69 | Urate oxidase in complex with 8-azaxanthine | 4N3M | I 2 2 2 | 1.90 |
| 70 | Urate oxidase in complex with 8-hydroxyxanthine | 4N9M | I 2 2 2 | 2.30 |
| 71 | Urate oxidase in complex with 8-hydroxyxanthine | 4N9S | I 2 2 2 | 1.06 |
| 72 | Xylose isomerase E186Q mutant (Cyclic glucose bound) | 4LNC | I 2 2 2 | 2.19 |
| 73 | Z-DNA | 3QBA | P $2_1$ $2_1$ $2_1$ | 1.53 |
| 74 | Z-DNA hexamer CGCGCG | 1V9G | P $2_1$ $2_1$ $2_1$ | 1.80 |

A




B

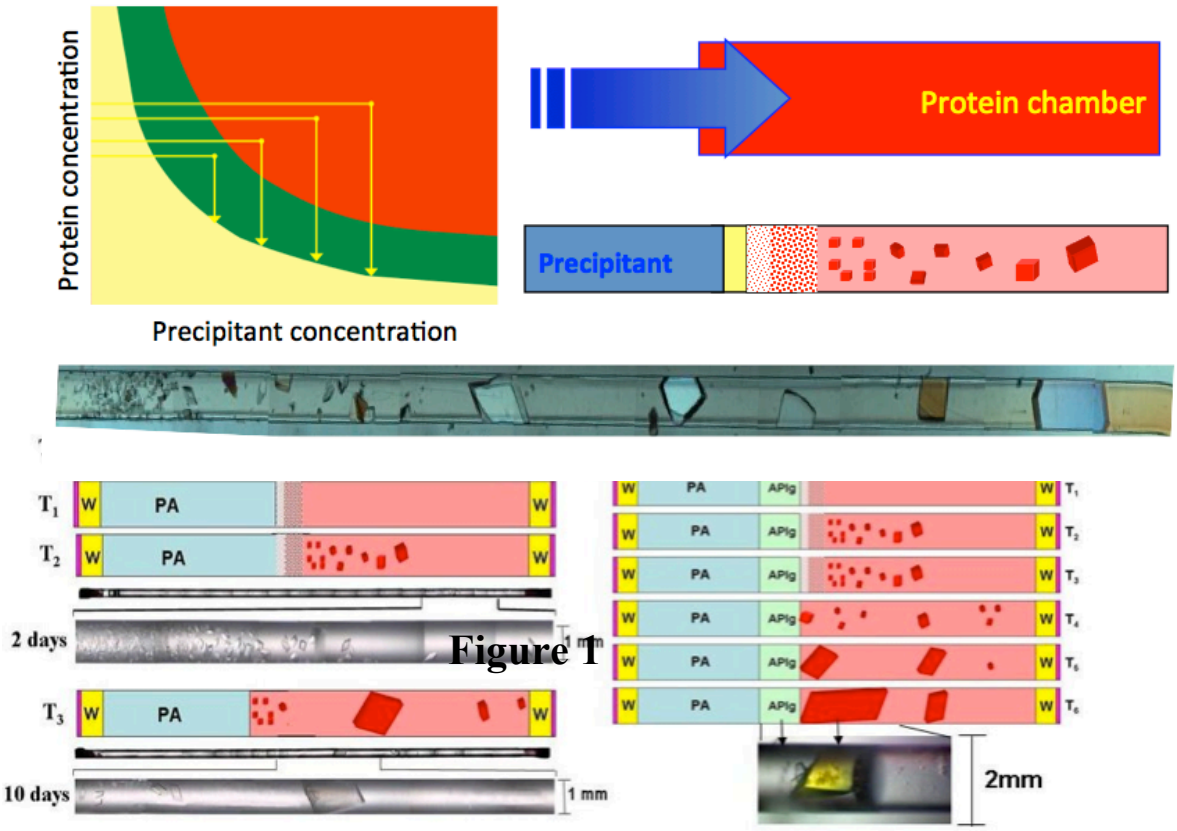
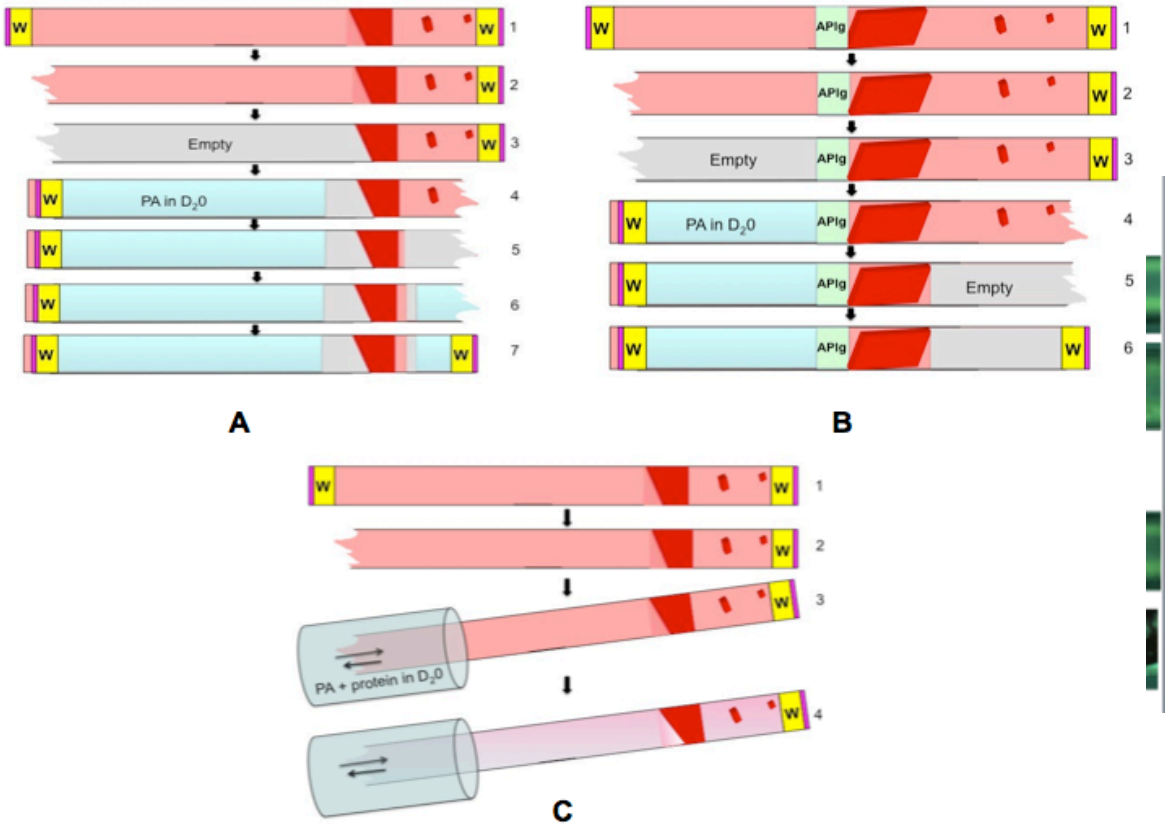



Figure 1

Figure 2


W
1
2
Empty
3
PA in D20
4
5
6
7
A
APIg
Empty
PA in D20
B
PA + protein in D20
C

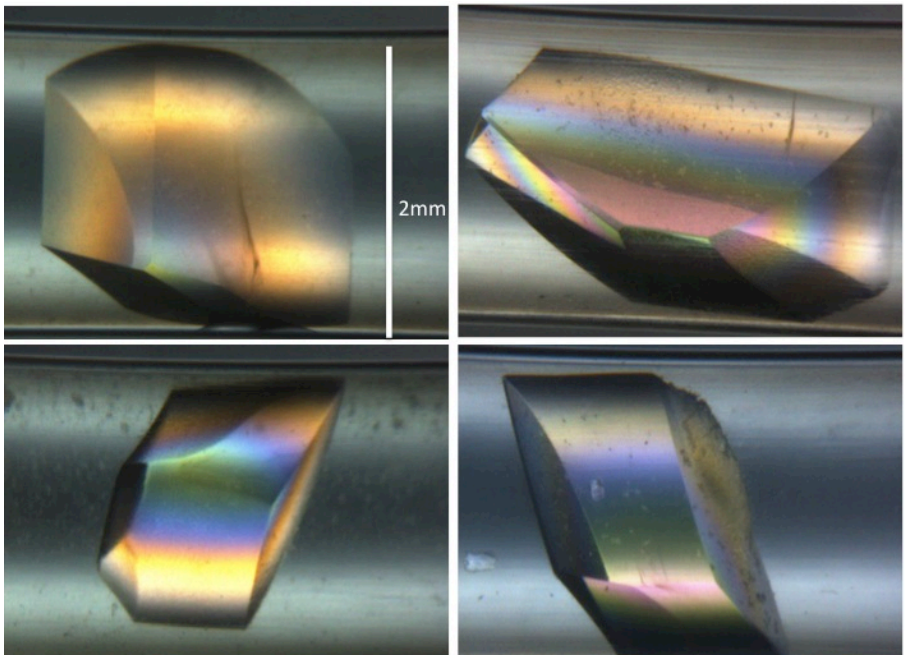

2mm

**Figure 4**