ELSEVIER

Journal of Crystal Growth 232 (2001) 520–535

JOURNAL OF CRYSTAL GROWTH

www.elsevier.com/locate/jcrysgro

Mosaic spread analysis of Canadian advanced protein crystallization experiment on the Russian space station, Mir

T.-S. Yoon[a], S. Tetreault[a], H.E. Bosshard[b], R.M. Sweet[b], J. Sygusch[a,*]

[a] *Department of Biochemistry, University of Montreal, Biochemistry, CP 6128, Station Centre Ville, Montréal, Quebec, Canada, H3C 3J7*
[b] *Brookhaven National Laboratory, Biology, Upton, NY 11973, USA*

Abstract

Protein crystallization experiments were performed on the Russian space station, Mir, using liquid–liquid interface diffusion. The technique was activated in orbit by the sliding together of two half-wells containing protein and precipitant fluids, respectively. Imperfections in protein crystals were analyzed from rocking curve measurements of the diffracted intensities using synchrotron radiation. Data were collected on microgravity and earth-grown crystals, and 10 different protein pairs were compared. To avoid bias, a double-blind protocol was used throughout the data analysis. Rocking curves for individual reflections were analyzed in terms of crystal domains, each fitted by a three-dimensional Gaussian profile. The results of Gaussian analysis were consistent with domain segregation corresponding to spatially different regions of the protein crystal exhibiting distinct mosaic spreads. When crystals were grown in microgravity the domain mosaic spreads were consistent with five of 10 different proteins exhibiting fewer imperfections, three other proteins showed no significant difference while a remaining two proteins displayed a greater number of apparent imperfections. Ground (earth-grown) controls were also conducted on protein samples flown to assess protein stability as a function of solution storage time prior to protein crystal growth (PCG) activation in microgravity. Protein samples were stored in ground controls at concentrations used to initiate crystallization, and aliquots were analyzed after a 30-day period by dynamic light scattering. Polydispersity estimates indicated that prolonged storage induced heterogeneity in all protein samples. Stable aggregates were present, and they were concentration independent, as shown by resistance to protein sample dilution. A PCG growth model is proposed that takes into account large scale aggregation or self-impurities present during crystal growth and predicts domain segregation. Trapping or rejection of self-impurities using this model can qualitatively explain differences in domain mosaic spreads observed as a function of gravitational environment. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: A1. Growth models; A1. Impurities; A1. Segregation; A1. X-ray diffraction; A2. Microgravity conditions; B1. Proteins

1. Introduction

Macromolecular crystals have become a basis of molecular biology, biochemistry, and biotechnology [1–3]. Understanding how proteins express their biological function depends on knowledge of the macromolecular architecture at the atomic

*Corresponding author. Tel.: +1-514-343-2389; fax: +1-514-343-6463.
E-mail address: jurgen.sygusch@umontreal.ca (J. Sygusch).

0022-0248/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0022-0248(01)01165-4

level. The accuracy of structures determined by X-ray crystallography is limited by the disorder in the protein crystals. Focus of microgravity research in protein crystal growth (PCG) has been based on the observation that PCG in a microgravity environment yields protein crystals that in a number of cases have reduced disorder [4–8]. Reduction in lattice disorder by protein crystals grown in microgravity compared to ground controls offers enhanced resolution of the diffracted intensities [4,8–11] and translates at the atomic level into more precise knowledge of the protein architecture [12].

Rocking-curve widths of the diffracted intensities is one of a number of techniques that have been used to assess quality of protein crystals [7,13–17]. A mosaic-block model has been used to analyze diffraction profiles. The model takes into account orientation of the blocks, their size distribution, and variation in unit-cell dimensions among blocks. Mosaic-block analysis, somewhat simpler than yet complementary to analysis of resolution limits, affords a parameterized description of crystal imperfection [18]. The departure from crystal perfection is characterized by a mosaicity parameter derived from full-width of the diffracted intensity at half-maximum (FWHM) and obtained from rocking-curve measurements. Profile broadening is a convolution of distributions describing the mosaic spread, the intrinsic profile width of a perfect crystal as well as the instrumental factors. FWHM characterizing the broadening is taken as a square root of the sum of the second moments of the individual distributions. Theoretical rocking widths for perfect protein crystals are typically of the order of $1''$ ($<0.5\times10^{-3\circ}$) [13] while instrumental factors can be derived from primary beam divergence, bandwidth, and source size. By using synchrotron radiation, instrumental effects contribute less than several millidegrees to FWHM [19,20]. The magnitude of these factors suggests that, when using synchrotron radiation, mosaic disorder is the dominant source of profile broadening in protein crystals..

Rocking-curve profiles have been analyzed in terms of Lorentzian or Gaussian shape functions [15,30] and, in case of complicated curve shapes, diffraction profiles are analyzed as linear sums of such functions centered at different peak positions [17]. Topographical analysis of lysozyme crystals suggests that individual diffraction profiles can be associated with a specific scattering region or domain in a protein crystal [21,22]. Mosaic spread for each scattering domain is extracted in the same manner from the physical and instrumental broadening factors as would be in analysis of the rocking curve profile from a single crystal.

Various crystal imperfections, or macroscopic disorder in the protein crystal, affect the rocking width differentially and can be discriminated using the mosaic model [18]. The origin of crystal imperfections has been linked to growth mechanisms and impurities incorporated during the growth process [23], environmental factors that affect growth process, as well as physical manipulation of protein crystals [17,29].

Microgravity can result in a disturbance-free environment by suppressing fluid convection. This can influence crystal growth significantly by affecting movement and distribution of proteins in the bulk fluid and controlling introduction of fresh solute and attachment of solute to crystal surfaces. Most proteins aggregate at high concentrations used in PCG, forming more-or-less ordered structures that may be transiently or kinetically stable. The latter are readily detected using light scattering techniques or gel filtration. These entities or self-impurities may be a major source of contaminants that influence growth processes. Because of their large size and low diffusivities, the movement and distribution of such entities can be significantly different in convection-free environments. The long-term storage of a purified protein in soluble form, prior to experiment activation in microgravity, may even increase the concentration of such entities. Protein instability or unfolding promotes production of irreversible aggregates in a protein solution at high concentration, especially if it has been maintained in soluble form for several weeks. Given the high supersaturation conditions required for nucleation, protein crystal nuclei may contain significant concentrations of amorphous aggregates. Protein crystals also grow by direct

addition of three-dimensional nuclei or volume elements [24], and their transport by virtue of their size will be significantly modified in a diffusive rather than convective environment. The quantitative description of crystal imperfection obtained from mosaic spread analysis of rocking curves should thus afford insight into factors influencing crystal quality.

Mosaic spread analysis was performed using synchrotron radiation on protein crystals obtained from the Canadian advanced protein crystallization experiment (CAPE) flown on the Russian Space Station, Mir and compared with an identical analysis on earth-grown protein crystals. The mission, a high-density PCG experiment, provided microgravity access to the protein crystallographic community in Canada. In total, 670 different crystallization experiments were performed using several different crystallization methods and were intended to provide a quantitative benchmark for protein crystal growth in microgravity. The protein-crystallization experiments were performed using a sliding block technology that superposes, following block activation, two fluid half-wells that were initially separated [25]. This technology is best performed in a microgravity environment since it is based on free liquid–liquid interface diffusion of a precipitant solution against a protein solution. Superposition allows diffusion to occur, inducing nucleation and crystal growth. Ground control protein crystals were supplied by participating laboratories and exemplified protein crystal growth under the best laboratory conditions.

2. Experimental procedure

The CAPE mission was launched September 1997 and stayed on Mir until end of January 1998 when it was returned to earth.

2.1. Loading of CAPE samples

Loading of the CAPE flight hardware commenced beginning of September 1997. Preparation and loading occurred over a 3-week period and involved coordinated receipt of more than 1400 different protein crystallization samples. Physical loading of the CAPE flight hardware took place over a 5-day period. Quality assurance personnel closely monitored all aspects of loading to ensure that each sample was pipetted into its assigned well. The CAPE experiment was then flown by leased jet from integration site at St. Hubert, Quebec, Canada to Cape Kennedy, Florida launch site 3 days prior to launch.

2.2. Crystal harvesting

The CAPE payload was retrieved shortly upon shuttle landing at Cape Kennedy and returned by leased jet to St. Hubert, Quebec, Canada where sample unloading took place. The contents of each sample well were first drawn into 100 μl capillaries and then each well was rinsed twice with freshly prepared precipitant solution to retrieve all residual material. Over 3500 capillaries were harvested, documented by microscope examination and a written analysis, and digital microscope images were recorded. The capillaries were then returned to the participating CAPE laboratories by end of week three. Crystals of selected proteins were retained for mosaic-spread analysis at the National synchrotron light source (NSLS) at Brookhaven National Laboratory. Selection was made on basis of the crystal morphology, size — at least one dimension exceeding 100 μm, and a crystal habit suitable for mounting in capillaries.

2.3. Data-collection protocol

Crystals were mounted using their mother liquor in thin-walled capillaries and photographed before being carefully transported to NSLS. Rocking-curve measurements were performed at beam line X26-C using a Brandeis 1024×1024 pixel CCD-based detector. Profile scans were taken at 0.02° intervals using normal-beam geometry such that the spindle axis (X_1-direction) was in the horizontal plane and perpendicular to the incident beam (X_3-direction). Crystals were initially exposed for 10 s per 0.5° step for five steps in order to adjust the crystal-to-detector distance and

to optimize counting times. Exposure times for fine slicing was 5–10 s for a total of 125 frames, at the end of which the crystal was rotated 90° about the spindle axis and exposed for another 125 frames provided radiation damage was not significant. The starting orientation was always chosen such that the crystal in each capillary was on the side closest to the detector. Data were collected first on the ground controls for a given protein crystal and then followed by collection on the microgravity-grown crystals. Rocking-curve measurements were obtained from 109 crystals: 64 space grown crystals and 45 ground controls representing 20 different proteins. Most protein crystal data sets were measured at least twice. In addition, data sets were collected on several silicon test crystals to analyze the instrumental resolution function of the X26-C beam line. X26-C beam line parameters used for instrumental resolution function calculations were the following: horizontal and vertical slit sizes of 0.2 mm, horizontal and vertical beam divergences of 2 and 0.2 mrad, respectively; source to instrument distance of 20 m; Si(1 1 1) beam monochromatization and relative wavelength resolution of 4.5×10^{-4}.

2.4. Double-blind experiment

To avoid possible experimenter bias in the 3-D profile analysis, a double-blinded experimental protocol was employed. For each of the diffraction images from the 109 protein crystal data sets, the diffraction data set identification was removed in the Montreal laboratory, renamed with random ID numbers, and sent to Dr. Sweet's laboratory at the Brookhaven National Laboratory. The data processing used the data reduction program d*trek [26] for indexing of the diffracted spots, determination of the crystal orientation matrix, and collection of successive 2-D image slices belonging to the same reflection into a contiguous record or "shoebox". Before returning the shoeboxes to Montreal, Dr. Sweet's group introduced new random ID numbers. In this way, the data set identities were kept hidden to each group to ensure objectivity throughout the analysis. Certain data sets could not be reduced by d*trek due to crystal splitting and inadequate resolution and were not used in the analysis. Data reduction resulted in both ground and space data sets for 10 different protein crystal pairs that were returned to the Montreal laboratory for rocking-curve analysis. Proteins for which paired data sets suitable for analysis were obtained are shown in the first column of Table 1.

2.5. Modeling of diffraction profiles

For each shoebox containing a 3-D profile of a Bragg reflection HKL, Eq. (1) was used to fit the observed pixel values from the CCD detector,

$$P^{HKL}(X) = B^{HKL} + I^{HKL} \sum_{i=0}^{N} p_i \phi\{M_i^{HKL}, \mu_i^{HKL}\}(X), \quad (1)$$

where $P^{HKL}(X)$ is the intensity recorded for each pixel and which has Cartesian coordinates $X(X_1, X_2, X_3)$, B^{HKL} is the background, I^{HKL} is the net intensity, p_i is the population for the ith mosaic domain, the Gaussian basis function, φ, is described below, M_i^{HKL} is a matrix describing the anisotropic intensity broadening effects due to the ith scattering domain, and μ_i^{HKL} is the vector representing the ith domain peak position. This representation attributes to each crystallite domain, as has been done in 1-D analysis [17], a peak position, μ_i^{HKL}, and specific mosaic spread parameters, M_i^{HKL}. The linear sum of population-weighted functions, $\varphi\ \{\cdot\cdot\}$, then accounts for the presence of more than one mosaic domain. For a given crystal, the population parameters should be similar for all Bragg reflections HKL because they reflect the total volume ratios of the scattering domains in the crystal.

Although the analysis allows potential use of any functional description, Gaussian functions are good representations not only of spatial interference but also of data that have been subject to instrumental effects. In the limit of an infinite number of instrumental effects, any diffraction profile has a Gaussian shape. Gaussian functions are also easy to use in making corrections for instrumental broadening. The 3-D Gaussian basis

Table 1
Data analysis and crystallization condition references of protein samples flown on CAPE

Protein	Environment	Data points	Gaussian functions	Goodness of fit	Crystallization reference
		⟨pixels⟩	⟨n⟩	⟨$\chi 2$⟩	
Cholesterol oxidase	Ground	4718	2.7	0.88	Croteau, Vrielink [37]
	Space	4792	3.6	1.27	
Duck I crystallin	Ground	1621	2.9	0.77	Abu-Abed et al. [38]
	Space	3214	3.4	0.82	
E.coli aldolase	Ground	5130	5.2	1.16	Blom et al. [39]
	Space	6371	5.4	1.04	
Fru-1,6 pase	Ground	12 705	7.1	1.60	Zhu et al. [40]
	Space	5085	2.9	0.67	
H162N Duck II crystallin	Ground	9450	4.3	0.70	Vallee et al. [41]
	Space	4932	4.1	1.03	
Lysozyme	Ground	4193	3.7	1.45	Pusey et al. [42]
	Space	4095	2.7	0.99	
Plasminogen activator inhibitor-1	Ground	16 469	8.9	1.27	Courtesy of Dr. R. Read
	Space	13337	6.9	0.91	
Rabbit muscle aldolase	Ground	5500	4.6	1.32	Blom, Sygusch [43]
	Space	8609	2.3	0.69	
Thaumatin	Ground	3085	4.0	1.02	Ng et al. [10]
	Space	3545	2.6	0.95	
Xylanase	Ground	5705	4.6	1.12	Bedarkar et al. [44]
	Space	4371	3.4	1.41	

function has the following form:

$$\phi\{M_I, \mu_I\}(X_1, X_2, X_3) = \frac{(\det M^{-1})^{1/2}}{(2\pi)^{3/2}} \times \exp\left[-\frac{1}{2}\sum_{I=1}^{3}\sum_{J=1}^{3}(X_I - \mu_I)M_{IJ}^{-1}(X_J - \mu_J)\right]. \quad (2)$$

Given the non-linear behavior of the basis functions in terms of refinable parameters, least-squares regression analysis requires reasonable initial estimates for one to refine the peak positions, μ_I, and anisotropic mosaic spread parameters, M_{IJ}. The second moment about the mean μ_I for the Ith Gaussian function defines the half-width of the Gaussian function in the X_I direction. $\mathbf{M}_{33}$ then represents the mosaic spread uncorrected for instrumental factors of the ith domain. Multiple peak recognition and estimates of the refinable model parameters is not readily automated in 3-D diffraction profile fitting and often requires manual intervention. An interactive graphical interface was constructed that provides initial estimates of the parameters characterizing each domain and then refines these estimates by non-linear iterative least squares analysis.

2.6. Computer program

The 3-D graphical interface program was written using the MATLAB language to

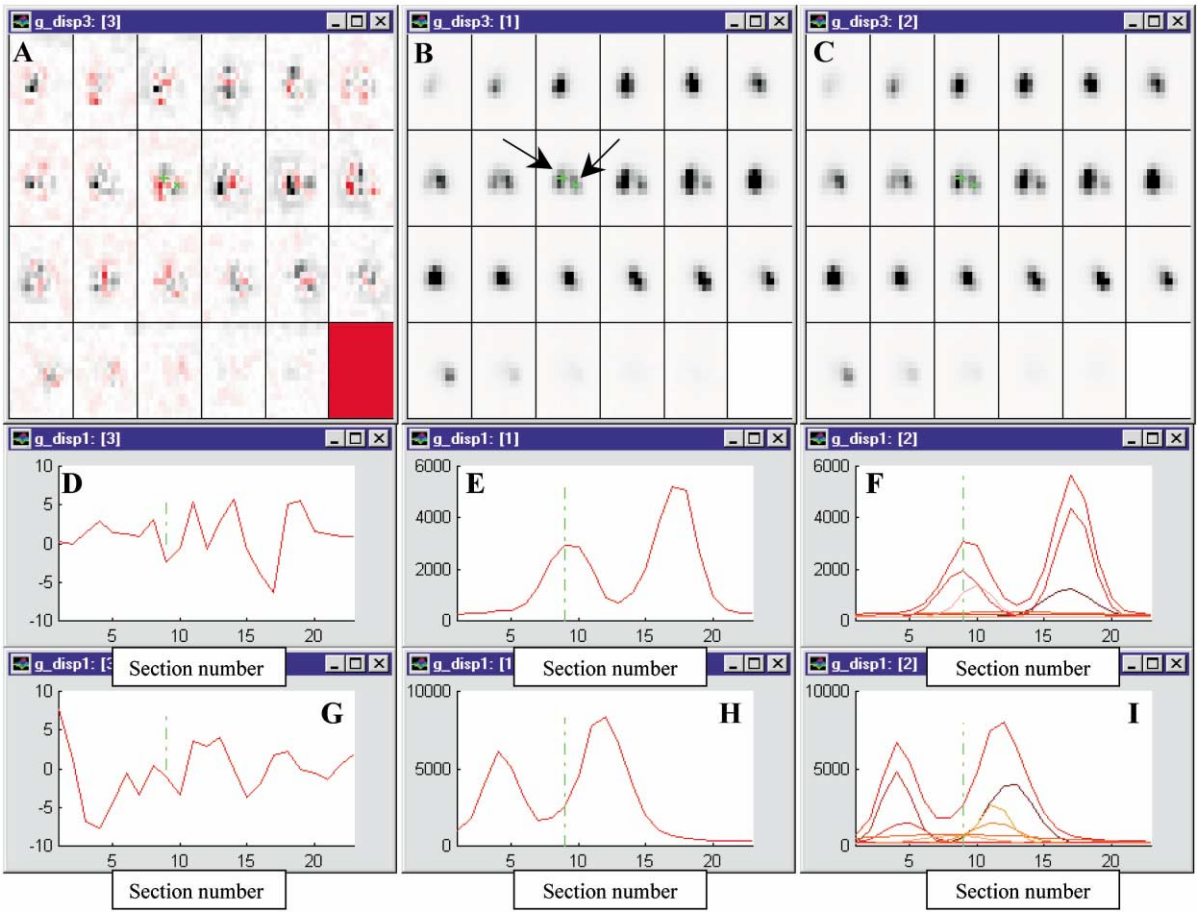

Fig. 1. Shoebox display of a rocking-curve profile for rabbit muscle aldolase comprising $15 \times 15 \times 23$ pixels. Panels on left show differences between observed and calculated values, middle panels correspond to observed data, while right panels show calculated values from the least squares analysis. Scales in panels A, D and G are in sigma units. Each pixel value was normalized by residual standard deviation of the difference between observed and calculated values. Positive deviations are shown in shades of light gray while negative deviations are in graduations of red. Panels B and C are in pixel values, shown as a gray-scale display, for each 2-D image along the rocking curve. Panels D–F show linear profiles obtained by plotting the intensity value of the same pixel, indicated by the left arrow in panel B, for each section. Panels G–I show a linear profile taken through pixel indicated by the right arrow in panel B. Panels E and H clearly show that the particular crystal contained at least four different resolvable domains, several of which are spatially distinct. Goodness of fit attained was 1.38.

implement the above formalism and will be presented elsewhere. The program uses MATLAB version 5.3 (obtained from the MathWorks Inc, Natick, MA) and can run on both SGI /IRIX and Windows NT operating systems. In Fig. 1, a shoebox representing an actual diffraction profile containing distinct peaks is analyzed. The program estimates peak positions, μ_I, and mosaic spread parameters, M_{IJ}, for each of the three peaks. As can be seen, compared to the 3-D analysis, a line or 1-D representation would not resolve the intensity peaks. The refinement results for each fine slice in a given shoebox are displayed in Fig. 1 showing observed and calculated values as well as differences between the observed and calculated values.

2.7. Data analysis

The great majority of the intensities chosen for analysis lay within a 30° wedge of the plane perpendicular to the rotation axis. This selection strategy minimizes profile broadening due to both Lorentz factor and beam divergence. For each ground control and space-grown crystal, the shoeboxes were sorted as a function of resolution and at least 10 reflections were chosen uniformly sampled and of highest signal intensity within the given resolution shell. Whenever possible, protein data sets having comparatively weaker or an insufficient number of strong intensities were not used in the analysis. Noise in weak reflections $\{I < 10\sigma(I)\}$ did not allow for the same level of discrimination of scattering domains as for strong intensities. To prevent data overfitting, the weighted goodness of fit was used as an objective statistical criterion to terminate function fitting [27]. Analysis of the diffracted profiles thus consisted of adding recognizable scattering domains via the graphical interface until the goodness-of-fit value approached 1 for a given profile. Residual fluctuations in the data less than 5 standard deviations were not considered significant. Weights used for the analysis were based on Poissonian counting statistics. For each profile, convergence was attained when all refinable parameters had shifts less than their standard deviation. Standard deviations were estimated from the normal equation matrix. Depending on the number of fine slices per shoebox, the amount of observations used in the refinement per shoebox was generally between 3000 and 15 000 pixels. The graphical interface was very useful for those cases requiring manual intervention due to slow convergence or divergence. Residual profiles were identified on the display, and Gaussian profiles were added or removed and refinement was continued. All data analyzed satisfied the criterion of convergence and corresponded to a stable minimum. Scattering domains having relative populations of less than 0.10 were not refined fully anisotropically due to poor convergence and refinement instability and the $\mathbf{M}_{IJ}$ matrix was set diagonal with $\mathbf{M}_{11} = \mathbf{M}_{22}$. In most cases this amounted to less than 15% of the total population.

2.8. Mosaic spreads

The mosaic spreads for a given scattering domain were calculated from the second moment $\mathbf{M}_{33}$ obtained from each scattering domain and corrected for its instrumental broadening and Lorentz profile broadening according to Colapietro et al. [19] and Ferrer et al. [20], respectively. Relative population-weighted averages were calculated for the mosaic spread of each scattering domain for each protein. Domains having populations less than 0.10 were not used in data compilation.

Second moments $\mathbf{M}_{33}$ obtained for two silicon test crystals corresponded to values of 0.0066° and 0.0052°, and correction for instrumental and Lorentz broadening yielded mosaic spreads of 0.0043° and 0.0042°, respectively. The mosaic spreads were, however, quite sensitive to energy resolution – an increase of 2 eV reduced the mosaic spreads by 1–2 millidegrees. For the majority of the protein data, the contribution of the instrumental broadening factor was not significant and only influenced mosaic spreads of the very sharp Gaussian domains by less than one millidegree.

2.9. Protein storage

Protein stability was assessed in terms of protein aggregation. Irreversible protein aggregation is protein concentration independent and can be distinguished by dilution from reversible aggregation that is concentration dependent. Dynamic light scattering (DLS) experiments on dilute protein solutions were used to detect aggregation. Measurements were performed using a DynaPro-801 TC (Protein Solution, Charlotteville, VA) apparatus. The laser wavelength was 853.4 nm, and measurements were taken at a scattering angle of 90°. DLS software yielded estimates of sample polydispersity, allowing assessment of protein sample heterogeneity. Sample polydispersity is defined by the standard deviation of the spread of particle sizes about the average hydrodynamic radius. Relative polydispersity is defined by the ratio of the standard deviation to the mean value. A protein solution having a value of $<25\%$ is considered to be essentially monodisperse

according to the manufacturer. The protein samples used for DLS analysis were residual protein samples from the CAPE mission, and measurements were taken on earth to coincide within days of on-orbit activation on Mir. For comparison, DLS measurements also have been included from a simulation where samples of purified protein were stored in a soluble state at concentrations for crystallization experiments and immediately shipped to the CAPE mission integration center. To ensure adequate and non-artifactual DLS signal measurement, protein samples were centrifuged and filtered prior to measurement through an inorganic alumina matrix membrane filter (0.02 μm) recommended by the manufacturer. Each sample was diluted such that the protein-concentration-dependent DLS signal exceeded 3 times the signal from sample buffer alone. Protein samples were diluted to lowest possible protein concentrations ($\sim$0.2 mg/ml final protein concentration) to promote a non-aggregated state for each protein sample. A double-exponential (bimodal) analysis was performed of the DLS signal to estimate gross populations of protein species at the time of CAPE activation.

3. Results

3.1. Domain analysis

Each intensity rocking curve measured was satisfactorily fitted by a linear sum of 3-D Gaussian profiles. A breakdown into gravitational environment and quality of the Gaussian decomposition of the rocking intensity measurements for the paired protein crystals is summarized in Table 1. The number of Gaussian functions used to interpret the rocking-curve data were between four and five on average, and increased with the complexity of the rocking-curve profile. Refinement against individual pixel data in each fine slice in a shoebox afforded in many cases unambiguous assignment of the scattering domain centroid, which would not have been evident from one-dimensional data analysis. An example of this is shown in Fig. 1.

Individual scattering domains discerned were fitted by a single 3-D anisotropic Gaussian function, and additional Gaussian functions were added if necessary to improve the fit to the data. Gaussian functions characterizing populations of <0.10 were omitted from the subsequent data analysis. The enhanced resolution afforded by the 3-D data resulted in good convergence and more stable refinement, and the number of pixels used to fit the data afforded a very large data-to-parameter ratio. The fit to the rocking-curve data by the Gaussian model was satisfactory with agreement between observed and calculated values on the most part to within the theoretical goodness-of-fit value of 1. The weighting used in the refinement did not take into account incident beam fluctuations and thus data fit for strong intensity profiles resulted in goodness-of-fit values somewhat greater than 1.

3.2. Segregated mosaic model

Scattering domains were distinguished in terms of their positions and moment matrices. Well resolved scattering domains, characterized by small moment matrices, hence possessing well separated centroids, is consistent with spatial separation of the scattering domains, hence diffraction occurring from distinct regions of the protein crystal (see peaks in Fig. 2). Scattering domains whose centroids overlap, seen in Fig. 1, are consistent with the description of a protein crystal containing embedded yet segregated scattering domains as shown in Fig. 3. In the case of a few scattering domains, that is 3–4 Gaussians with significant populations, the spatial organization of the domains was such that adjacent domains frequently had least differences in mosaic spread.

3.3. Paired protein crystal comparisons

Diffraction profiles were compared on a domain basis. When minor populations were excluded at least two to four significant domain populations remained in most instances. The analysis of differences between crystals then consisted of comparing domains having the largest and smallest mosaic spread as well as the domain with the

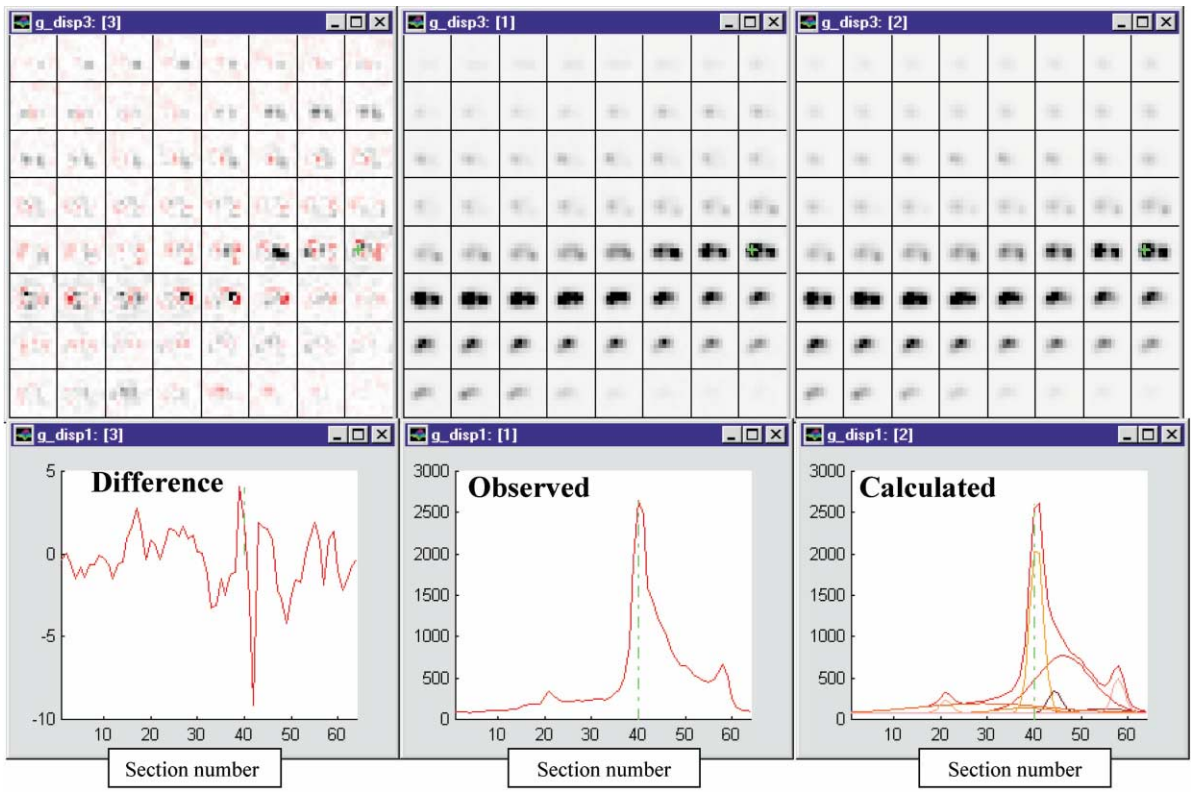

Fig. 2. Shoebox display of a rocking-curve profile for ground crystal of plasminogen activator inhibitor-1 comprising $11 \times 11 \times 64$ pixels. Panels are the same as in Fig. 1. Panels on the left shows the difference between observed and calculated values and are scaled in residual sigma units, the middle panels correspond to observed data, while the right panels show the calculated values from the least squares analysis. Lower panels are linear profiles taken through the pixel indicated by the green + symbol. Lower right-hand side panel clearly shows that the particular crystal contained at least six different domains, the majority of which are spatially resolved. Goodness of fit obtained corresponded to 1.06.

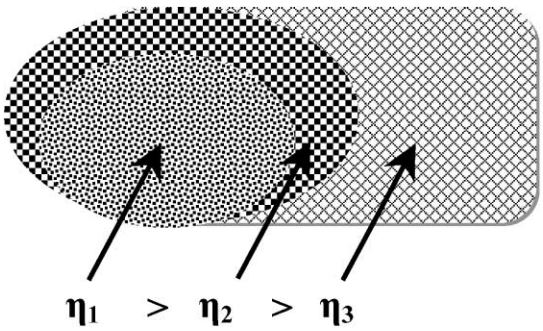

Fig. 3. Segregated mosaic model describing the domain distribution in a scattering crystal. The domains are discrete and correspond to distinct regions in the protein crystal. Each domain has a particular mosaic spread, and the spatial organization of the domains is such that adjacent domains have minimal differences in mosaic spread.

largest population. As the same domain description applies to each reflection of a given crystal, the pertinent mosaic spreads for each intensity profile of a protein crystal were averaged, using as weights the refined domain population. This weighting scheme compensated for minor populations that varied among diffraction profiles. The average population per protein crystal was also computed for each tabulated domain. The averaged mosaic spreads and populations are shown for each protein crystal pair in Table 2.

The sums of the population averages calculated for the lowest and largest mosaic spreads, shown in Table 2, comprise nearly half of the total scattering domains and in most instances amount to more than two thirds. The dispersion or standard

Table 2
Mosaic spreads analyzed from protein samples flown on CAPE mission[a]

Protein	Environment	Population having largest mosaic spread, averaged		Population having smallest mosaic spread, averaged		Mosaic spread of largest population, averaged	
		⟨p⟩	⟨η⟩[b]	⟨p⟩	⟨η⟩[b]	⟨p⟩	⟨η⟩[b]
Cholesterol oxidase[c]	Ground	*0.59*	76.2	*0.36*	26.7	*0.63*	69.3
	Space	*0.34*	32.1	*0.38*	16.8	*0.53*	22.1
Duck I crystallin[c]	Ground	*0.39*	24.1	*0.48*	11.7	*0.61*	16.2
	Space	*0.34*	15.3	*0.41*	7.2	*0.48*	8.5
E. coli aldolase[d]	Ground	*0.27*	41.6	*0.25*	14.6	*0.35*	26.5
	Space	*0.38*	91.9	*0.15*	29.7	*0.48*	73.2
Fru-1,6 pase[c]	Ground	*0.26*	159.6	*0.20*	69.9	*0.34*	131.9
	Space	*0.52*	55.4	*0.20*	16.3	*0.52*	50.6
H162N Duck II crystallin[c]	Ground	*0.51*	106.0	*0.25*	49.9	*0.55*	99.3
	Space	*0.30*	46.8	*0.32*	28.2	*0.48*	34.2
Lysozyme	Ground	*0.28*	36.8	*0.39*	16.8	*0.48*	21.7
	Space	*0.39*	36.4	*0.60*	20.0	*0.66*	24.8
Plasminogen activator inhibitor-1[c]	Ground	*0.24*	224.9	*0.16*	47.4	*0.29*	155.8
	Space	*0.31*	171.1	*0.15*	25.9	*0.33*	150.8
Rabbit muscle aldolase[d]	Ground	*0.31*	20.5	*0.22*	8.7	*0.39*	13.3
	Space	*0.46*	52.9	*0.53*	47.8	*0.58*	47.4
Thaumatin	Ground	*0.41*	11.6	*0.37*	9.8	*0.54*	10.9
	Space	*0.44*	41.1	*0.41*	10.1	*0.52*	31.0
Xylanase	Ground	*0.35*	68.0	*0.33*	14.9	*0.47*	40.4
	Space	*0.44*	68.0	*0.34*	17.1	*0.52*	56.0

[a] Mosaic spread determination was performed using a double-blind protocol, described, to ensure unbiased analysis.
[b] η units are in millidegrees.
[c] Microgravity reduced mosaic spread.
[d] Microgravity was not beneficial.

deviations calculated with respect to the average populations (not shown) were less than 50% of the population value in all proteins and with least dispersion obtained invariably where rocking-curve analysis contained only few minor populations. The results reflect that a given scattering domain is present approximately to the same extent in each of the diffraction profiles for a protein crystal.

The mosaic spreads $\langle\eta\rangle$ for both smallest and largest scattering domains shown in Table 2, clearly show that for five out of the 10 protein pairs (denoted by superscript c), a shift to smaller mosaic domain spreads is observed for protein crystals grown in microgravity. These domains in many cases can represent together more than 2/3 of the scattering matter making up the protein crystal. Only for two protein pairs shown in Table 2 (denoted by superscript d) does microgravity apparently have a negative influence on PCG. The remaining three protein pairs are neutral with regard to microgravity influence. These conclusions also apply to the mosaic

Table 3
Dynamic light scattering experiments on CAPE flight samples coincident with hardware activation on Mir

Protein	Storage concentration (mg/ml)	Relative polydispersity		Bimodal analysis 30 days
		Initial[a]	30 days	
Cholesterol oxidase	21.8	0.55	0.38	Monomeric and aggregates[b]
Duck I crystallin	9	0.30	0.36	Monomeric and aggregates[b]
E. coli aldolase	7.5	0.16	0.39	Monomeric and aggregates[b]
Fru-1,6 pase	5	ND	—	Signal unstable[c]
H162N Duck II crystallin	9.3	0.51	0.65	Aggregation[d]
Lysozyme	10	<0.1	0.15	Monomeric and few aggregates[b]
Plasminogen activator inhibitor-1	2	ND	—	Signal to noise insufficient[e]
Rabbit muscle aldolase	3	0.17	0.39	Monomeric and aggregates[b]
Thaumatin	30	<0.1	0.30	Monomeric and aggregates[b]
Xylanase	3.75	0.55	—	Signal to noise insufficient[e]

[a] Initial refers to measurements made within less than 5 days of purification.
[b] Aggregates detected based on molecular weights obtained from bimodal analysis.
[c] DLS signal displayed significant fluctuations unsuitable for molecular weight estimation.
[d] Monomeric population not detected
[e] Protein signal was less than 3X of sample buffer. ND—Not determined

spread $\langle\eta\rangle$ of the domain having the largest population.

Analysis of the domain mosaic spreads of each reflection profile for a given protein crystal revealed no systematic variation in $\langle\eta\rangle$ values as a function of resolution.

3.4. Mission problems

The success of the CAPE mission was affected by fluid leakage out of some of the sliding block half-wells and temperature excursions on the space station Mir from the nominal set point of 22°C. In some cases, the latter could have caused the former. Thus, not all protein solutions flown on the mission yielded crystals; however, due to the large number of protein samples flown on the mission, numerous protein crystals were obtained.

3.5. Protein aggregation

Polydispersity of protein samples, summarized in Table 3, is significant even when the protein solution was diluted at the end of 30 days. For some proteins, however, relative polydispersity was considerable even prior to experiment activation. The bimodal molecular weight analysis indicated the presence of higher molecular weight species than native protein, consistent with protein aggregation in the protein samples. The presence of aggregates or aggregation is consistent with signal losses and instability noted following sample filtration. Without sample filtration, DLS measurements were highly erratic. Retention of aggregates in the sub-micron filter may indicate the presence of micron-sized particles in solution prior to CAPE activation.

4. Discussion

All protein crystals grown, be it in a microgravity environment or on earth, exhibit imperfections. The mosaic spread values obtained from our analysis when compared to the mosaic spread calculated for a perfect protein crystal, 200 μm in size [13], are at least an order of magnitude greater. From Table 2, protein crystals of duck I crystallin, grown in microgravity, had the lowest mosaic spread $\langle\eta\rangle$ of 0.0072°. Rabbit muscle aldolase crystals grown on earth had the next lowest value for $\langle\eta\rangle$ of 0.0087°. Both lysozyme and thaumatin crystals showed no clear distinction between growth on earth or microgravity although their domain mosaic spreads had values of 0.017° and 0.010°, respectively, among the lowest of all

protein pairs. Rocking-curve data reported for lysozyme [15] and thaumatin [28], also derived from microgravity studies, have $\langle\eta\rangle$ values of 0.0002° and 0.002°, respectively, and are an order of magnitude smaller compared to the values in this study. The $\langle\eta\rangle$ values shown in Table 2 are, however, similar to $\langle\eta\rangle$ values reported for apocrustcyanin C_1 crystals grown in microgravity [29]. That several crystals on CAPE had $\langle\eta\rangle$ values smaller than the smallest $\langle\eta\rangle$ value of 0.0118° reported for apocrustcyanin C_1 crystals suggests that CAPE mission problems did not adversely impact all proteins. Depending on its direction, crystal cracking could produce an apparent trend with resolution based on a 1-D rocking curve analysis; however, the use of 3-D Gaussian analysis minimizes such a potential artifact. In cases where diffraction profiles from protein crystals showed multiple peaks, such as those seen in Figs. 1 and 2, mosaic spreads derived from scattering domains of the different peaks did not show systematic differences. If mechanical handling contributed strain, this was not apparent from the analysis nor was it anisotropic in nature.

The spatial resolution of individual scattering domains from the rocking-curve analysis by anisotropic three-dimensional Gaussian functions is consistent with each scattering domain physically belonging to a distinct region in the protein crystal. The segregated mosaic model proposed in Fig. 3 embodies such a spatially discrete distribution of scattering domains having distinct mosaicities. The model implies that individual mosaic blocks are not randomly distributed throughout the crystal but segregated into domains that possess characteristic mosaic spreads and giving the protein crystal texture. Topographical studies on tetragonal lysozyme crystals at different rocking curve settings indicate texture by showing considerable variation in scattered intensity across the crystal [30]. The segregated model, shown in Fig. 3, also entails spatial ordering of domains in terms of their mosaic spreads implying concentration gradients with respect to perfection within the protein crystal. Synchrotron studies on a tabular hydrogenase crystal indeed observed progressive changes in resolution of diffraction patterns along the crystal length [31].

The influence of microgravity does not entail conceptual modification of the segregated mosaic model. Rather, protein crystal growth in microgravity merely produces a global shift to domains having smaller mosaic spreads in 5 of 10 protein crystals and in two cases to larger mosaicity. Except for the substantial greater value in mosaicity of the largest domain of thaumatin in microgravity, the values of the mosaic spreads shown for lysozyme, xylanase and thaumatin in Table 2 indicate negligible influence by microgravity on these proteins.

All protein samples showed significant protein aggregation after prolonged storage. Furthermore, from Table 3, some protein samples became heterogeneous within days following purification. Protein aggregation does not necessarily compromise protein crystal growth if it is reversible on a time scale capable of sustaining protein crystal growth. Losses upon protein solution filtering, even upon prior dilution, as well as sample heterogeneity at low dilution, argue in favor of concentration independent or irreversible aggregates present in the protein sample. Filtering losses imply these aggregates may be micron-sized in some cases. The aggregates are in effect self-impurities, and their origin in purified protein can be a consequence of protein instability. When protein is maintained for prolonged periods in solution prior to activation in microgravity or has been subjected to aggressive purification or both, protein is susceptible to spontaneous denaturation and unfolding. Exposure of hydrophobic surfaces upon partial denaturation then leads to formation of irreversible aggregates. Evidence for differential colligative behavior has been shown for apoferritin molecules under solution conditions where interactions among monomers are strongly repulsive while those of apoferritin dimers and higher oligomers, species resulting from partial denaturation of apoferritin monomers, were shown to be attractive [32]. Given the high protein concentrations required to induce nucleation at supersaturation, protein solutions may contain significant concentrations of irreversible aggregates. Furthermore, for a crystal growth time scale that is comparable to protein aggregation phenomena, irreversible aggregates become a significant con-

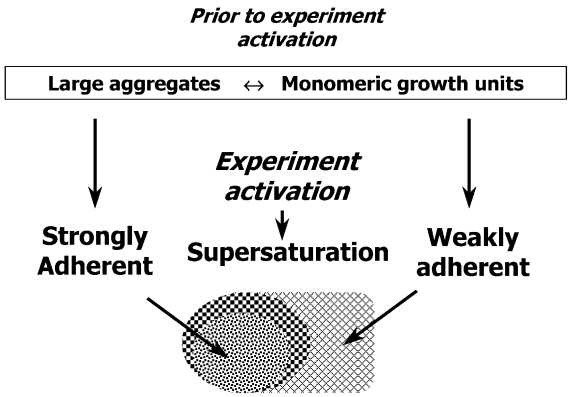

Fig. 4. Model describing origin of segregated domains. Large aggregates composed of convalently or conformationally modified protein are in equilibrium with smaller aggregates including monomeric growth units at sub-saturating concentrations and in conditions approaching those used for crystallization. The larger aggregates, containing partially unfolded protein, can possess greater hydrophobic surface area than the monomeric growth units, making them more susceptible to aggregation under supersaturation or PCG activation, hence more adherent. The larger aggregates, which can be micron in size, will assemble preferentially forming the initial domain. The subsequent domains are composed of less adherent aggregates and the monomeric growth units forming the last domain.

taminant of the crystal growth process. It was therefore surprising that half of the proteins flown on CAPE showed improvement when crystallized in a microgravity environment were those exhibiting sample heterogeneity.

Self-impurities that are aggregates in equilibrium in the protein solution with smaller monomeric protein growth units can affect the protein crystal growth process. A PCG model is proposed in Fig. 4 that considers the presence of aggregates during crystal growth. The model proposes that supersaturation conditions induce a graded interaction among aggregates and monomeric growth units with strongest interaction among aggregates, weakest interaction among monomers and intermediate interaction between monomers and aggregates. This tenant is reasonable given that the aggregates may originate from modified and/or partially denatured protein material that have greater hydrophobic surface area and which provides the driving force for association as water molecules are displaced from the aggregate surface in presence of precipitant. Large aggregates, by providing a larger hydrophobic contact surface, would thus be enriched in initial domains thereby effectively trapping the self-impurities. Preferential aggregate association promotes segregation of weaker interacting protein material and enriching subsequent domains in the less adherent aggregates. If monomeric growth units are weakest interacting then these will be enriched in the terminal domain.

The growth model equally predicts domain formation when self-impurities are rejected from the crystal surface under growth conditions consistent with aggregates being less adherent than monomeric growth units. Supersaturation conditions would promote domain formation composed of monomeric growth units initially and least adherent aggregates enriching the terminal domain. In this case, concentration of aggregates excluded by the initial domain growth would increase on the domain surface potentially hindering further monomeric transport and eventually promoting formation of a domain composed of less adherent aggregates. Domain dimension would be a function of the colligative properties of the aggregates or self-impurities as well as on their transport kinetics in relation to the monomeric growth units. Independent of which growth scenario applies, domains having higher concentrations of self-impurities have greater number of imperfections and thus larger mosaic spreads. Studies on the incorporation of either turkey egg-white lysozyme or ovotransferrin into hen egg-white lysozyme crystals created stresses in lysozyme crystals producing defects that were manifested in mosaic broadening [33].

Convection and sedimentation can influence the PCG model proposed for domain formation and is outlined in Fig. 5. Self-impurities, because of their larger hydrodynamic size, have lower diffusivities than monomeric growth units. If self-impurities are preferentially trapped, as shown in Fig. 5a, a concentration gradient will be transiently produced proximal to the growing crystal surface, which will be more depleted in protein species of larger hydrodynamic size compared to lower molecular weight species. Modeling of crystal growth process wherein the kinetic coefficient for impurity trapping is larger than the kinetic

coefficient describing incorporation of the basic protein predicts an impurity depletion zone around the growing crystal face [34,35]. As the smaller molecular weight species are incorporated into the terminal domain these in turn will exhibit a depletion zone as has been observed in lysozyme [36]. Convection will be deleterious to this process by disturbing the depletion zone, thereby reducing the concentration of monomeric growth units vicinal to the crystal face and enriching, by convective transport, the concentration of self-impurities at the crystal surface that would otherwise be kinetically excluded because of their low diffusion coefficient. Disturbances due to convection, sedimentation out of the depletion zone, or g-jitter processes would promote growth of domains, composed of larger aggregates, which have large mosaic spreads. A diffusion-dominated transport regime such as that present in a microgravity environment could thus promote growth of domains with smaller mosaicities. This growth model is consistent with the observation that the group of five proteins flown on CAPE were among those exhibiting greatest sample heterogeneity nevertheless showed significant improvement in mosaic spreads of their constituent domains compared to ground controls. Protein crystal growth experiments conducted in presence of foreign proteins have indeed shown that there are differential distributions between terrestrial and space environments and lower impurity incorporation in microgravity was consistent with an impurity depletion zone arising around growing crystals in absence of buoyancy driven convection or stirring [34,35].

If self-impurities are rejected, then as shown in Fig. 5b, self-impurities concentrate at the growing domain surface. Convective disturbances or other gravity-induced processes will favor domain growth by monomeric growth units firstly by reducing concentration of self-impurities at the domain surface and secondly by enriching with

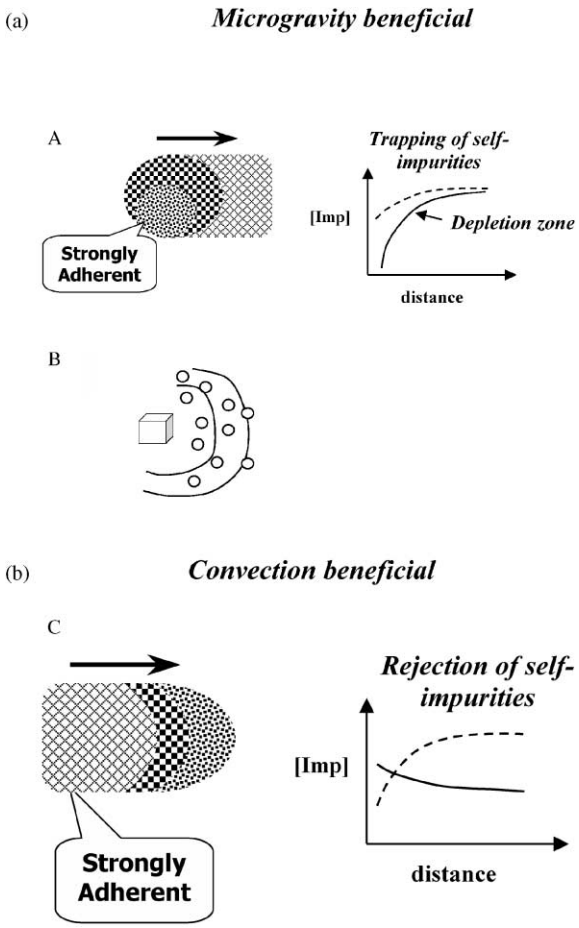

Fig. 5. (a) Growth of a protein crystal, shown by arrow in panel A, wherein larger aggregates are adherent resulting in segregation of weakly adhering aggregates and monomeric growth units to the domain surface, as suggested in panel B. As the crystal increases in size, aggregate concentration diminishes in front of the growing crystal face, shown by the solid line in panel A graph, thereby creating an impurity concentration gradient or depletion zone. Comparatively, monomeric growth units are less depleted in front of the crystal face (dotted line in graph A), and are capable of sustaining outgrowth of the terminal domain. Convection would interfere with aggreate trapping by continuously replenishing the crystal surface with self-impurities. (b) Growth of a protein crystal, shown by arrow in panel C, where the larger aggregates become less adherent upon supersaturation than monomeric growth units. In this case, the domain composed of monomeric growth units will form first followed by the domain composed of weaker interacting aggregates. The aggregates accumulate on the surface of the sticky domain, creating a concentration gradient around it (solid line in graph C), while the region is depleted in monomeric growth units (dotted line). The aggregate concentrations eventually interfere with monomeric transport and growth of the monomeric domain ceases. Convection would continuously cleanse the surface of the aggregate particles promoting growth of the monomeric domain.

monomeric growth units. In a microgravity environment, a growth mechanism dominated by rejection of self-impurities would result in domains having larger mosaicities as was observed for the aldolases. The large mosaic spread value observed for thaumatin when grown in space might be a consequence of the sample polydispersity observed at 30 days. Trapping of self-impurities would account for the larger mosaic spread. The remaining two proteins, lysozyme and xylanase, did not show significant differences in mosaic spreads of domains in space compared to earth. Trapping of self-impurities by rapid growth of monomeric growth units that virtually exhausts all available protein before significant concentrations of new self-impurities could build up would be consistent with these findings.

Acknowledgements

Dr. Tae-Sung Yoon was a recipient of a postdoctoral fellowship from the Korea Science and Engineering Foundation (KOSEF). Drs. A. Berinstain and P. Gregory of the Canadian Space Agency provided critical support and many helpful discussions. Logistic support by H. Moniz was greatly appreciated. Work was funded by the Canadian Space Agency. Work performed by the Brookhaven Laboratory staff was supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences, and the National Science Foundation.

References

[1] R. Giege, B. Lorber, A. Theobald-Dietrich, Acta Crystallogr. D50 (1994) 339.
[2] A. McPherson, A.J. Malkin, Y.G. Kuznetsov, Structure 3 (1996) 759.
[3] A. McPherson, Trends Biotechnol. TIBTECH 15 (1997) 197.
[4] L.J. DeLucas, C.D. Smith, H.W Smith, S. Vijay-Kumar, S.E. Senadhi, S.E. Ealick, D.C. Carter, R.S. Snyder, P.C. Weber, F.R. Salemme et al., Science, 246 (1989) 651. Published erratum appears in Science 249 (1990) 1359.
[5] J. Day, A. McPherson, Protein Sci. 1 (1992) 1254.
[6] L.J. DeLucas, K.A. Moore, T.L. Bray, W.M. Rosenblum, H.M. Einspahr, L.L. Clancy, G.S.J. Rao, B.G. Harris, S.H. Munson, B.C. Finzel, C.E. Bugg, J. Phys D: Appl. Phys 26 (1993) B100.
[7] J.R. Helliwell, E. Snell, S. Weisberger. in: L. Ratke, H. Walter, B. Feuerbacher (Eds.), Proceedings of the 9th European Symposium on Gravity Dependent Phenomena in Physical Sciences, Springer, Berlin, 1995, pp. 155–170.
[8] S. Koszelak, J. Day, C. Leja, R. Cudney, A. McPherson, Biophys, J. 69 (1995) 13.
[9] M.R. Wardell, R.D. Skinner, C. Carter, P.D. Twigg, J.P. Abrahams, Acta Crystallogr. D 53 (1997) 622.
[10] J.D. Ng, B. Lorber, R. Giege, S.J. Day Koszelak, A.J. Greenwood, A. McPherson, Acta Crystallogr. D 53 (1997) 724.
[11] G.D. Smith, E. Ciszak, W. Pangborn, Protein Sci. 5 (1996) 1502.
[12] R. Skinner, J.P. Abrahams, J.C. Whisstock, A.M. Lesk, R.W. Carrell, M.R. Wardell, J. Mol. Biol. 266 (1997) 601.
[13] J.R. Helliwell, J. Crystal Growth 90 (1988) 259.
[14] R. Fourme, A. Ducruix, M. Ries-Krautt, B. Capelle, J. Synchrotron Rad. 2 (1995) 136.
[15] E.H. Snell, S. Weisburger, J.R. Helliwell, Acta Crystallogr. D 51 (1995) 1099.
[16] J.L. Ferrer, J. Hirschler, M. Roth, J.C. Fontecilla-Camps, ESRF Newslett. 27 (1996).
[17] F. Otalora, B. Capelle, A. Ducruix, J.M. Garcia-Ruiz, Acta Crystallogr. D 55 (1999) 644.
[18] C. Nave, Acta Crystallogr. D 54 (1998) 848.
[19] M. Colapietro, G. Cappucio, C. Marciante, A. Pifferi, R. Spagna, J.R. Helliwell, J. Appl. Cryst. 25 (1992) 192.
[20] J.-L. Ferrer, M. Roth. J. Appl. Cryst. 31 (1998) 523.
[21] O. Vidal, M.C. Robert, B. Arnoux, B. Capelle, J. Crystal Growth 196 (1999) 559.
[22] R. Fourme, A. Ducruix, M. Ries-Kautt, B. Capelle, J. Crystal Growth 196 (1999) 535.
[23] A.A. Chernov, Acta Crystallogr. A 54 (1998) 859.
[24] A.J. Malkin, Y.G. Kuznetsov, T.A. Land, J.J. DeYoreo, A. McPherson, Nat. Struct. Biol. 2 (1995) 956.
[25] J. Sygusch, R. Coulombe, J.M. Cassanto, M.G. Sportiello, P. Todd, J. Crystal Growth 162 (1996) 167.
[26] J.W. Pflugrath, Acta Crystallogr. D 55 (1999) 1718.
[27] W. Hamilton, Statistics in Physical Sciences. Ronald Press, New York, 1964, p. 128.
[28] B. Lorber, C. Sauter, M.C. Robert, B. Capelle, R. Giege, Acta Crystallogr. D 55 (1999) 1491.
[29] E.H. Snell, A. Casetta, J.R. Helliwell, T.J. Boggon, N.E. Chayen, E. Weckert, K. Holzer, K. Schroer, E.J. Zagasky, P.F. Gordon, Acta Crystallogr. D 53 (1997) 231.
[30] F. Otalora, J.M. Garcia-Ruiz, J.A. Gavira, B. Capelle, J. Crystal Growth 196 (1999) 546.
[31] Y. Higuchi, T. Okamoto, N. Yasuoka, J. Crystal Growth 168 (1996) 99.

[32] N.P. Petsev, B.R. Thomas, S.-T. Yau, P.G. Vekilov, Biophys. J. 78 (2000) 2060.
[33] C.L. Caylor, I. Dobrianov, S.G. Lemay, C. Kimmer, S. Kriminski, K.D. Finkelstein, W. Zipfel, W.W. Webb, B.R. Thomas, A.A. Chernov, R.E. Thorne, Proteins, : Struct. Funct. Gen. 36 (1999) 270.
[34] D.C. Carter, K. Lim, J.X. Ho, B.S. Wright, P.D. Twigg, T.Y. Miller, J. Chapman, K. Keeling, J. Ruble, P.G. Vekilov, B.R. Thomas, F. Rosenberger, A.A. Chernov, J. Crystal Growth 196 (1999) 623.
[35] B.R. Thomas, A.A. Chernov, P.G. Vekilov, D.C. Carter, J. Crystal Growth 211 (2000) 149.
[36] A. McPherson, A.J. Malkin, Y.G. Kuznetsov, S. Koszelak, M. Wells, G. Jenkins, J. Howard, G. Lawson, J. Crystal Growth 196 (1999) 572.
[37] N. Croteau, A. Vrielink, J. Struct Biol. 116 (1996) 317.
[38] M. Abu-Abed, M.A. Turner, J. Atkinson, K. Dole, P.L. Howell, J. Mol. Biol. 243 (1994) 944.
[39] N.S. Blom, S. Tetreault, R. Coulombe, J. Sygusch, Nat. Struct. Biol. 3 (1996) 856.
[40] D.W. Zhu, G.J. Xu, P.H. Rehse, A. Azzi, F.K. Zhao, S.X. Lin, Acta Crystallogr. D 55 (1999) 1342.
[41] F. Vallee, M.A. Turner, P.L. Lindley, P.L. Howell, Biochemistry 38 (1999) 2425.
[42] M.L. Pusey, R.S. Snyder, R. Naumann, J. Biol. Chem. 261 (1986) 6524.
[43] N. Blom, J. Sygusch, Nat. Struct. Biol. 4 (1997) 36.
[44] S. Bedarkar, N.R. Gilkes, D.G. Kilburn, E. Kwan, D.R. Rose, R.C. Miller Jr., R.A. Warren, S.G. Withers, J. Mol. Biol. 228 (1992) 693.