Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A thermal stability assay can help to estimate the crystallization likelihood of biological samples

Florine Dupeux,[a] Martin Röwer,[a] Gael Seroul,[a] Delphine Blot[b] and José A. Márquez[a]*

[a]European Molecular Biology Laboratory, Grenoble Outstation and Unit of Virus Host-Cell Interactions, UJF–EMBL–CNRS, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble CEDEX 9, France, and [b]Institute de Biologie Structurale Jean-Pierre Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France

Correspondence e-mail: marquez@embl.fr

Received 13 May 2011
Accepted 5 September 2011

The identification of crystallization conditions for biological molecules largely relies on a trial-and-error process in which a number of parameters are explored in large screening experiments. Currently, construct design and sample formulation are recognized as critical variables in this process and often a number of protein variants are assayed for crystallization either sequentially or in parallel, which adds complexity to the screening process. Significant effort is dedicated to sample characterization and quality-control experiments in order to identify at an early stage and prioritize those samples which would be more likely to crystallize. However, large-scale studies relating crystallization success to sample properties are generally lacking. In this study, the thermal stability of 657 samples was estimated using a simplified Thermofluor assay. These samples were also subjected to automated vapour-diffusion crystallization screening under a constant protocol. Analysis of the data shows that samples with an apparent melting temperature (T_m) of 318 K or higher crystallized in 49% of cases, while the crystallization success rate decreased rapidly for samples with lower T_m. Only 23% of samples with a T_m below 316 K produced crystals. Based on this analysis, a simple method for estimation of the crystallization likelihood of biological samples is proposed. This method is easy, rapid and consumes very small amounts of sample. The results of this assay can be used to determine optimal incubation temperatures for crystallization experiments or to prioritize certain constructs. More generally, this work provides an objective test that can contribute to making decisions in both focused and structural genomics crystallography projects.

1. Introduction

The identification of initial crystallization conditions for biological molecules relies largely on a trial-and-error process in which a number of parameters, such as the composition of the crystallization solution or the incubation temperature, are explored through screening experiments (McPherson, 1999). This task has been notably advanced by the introduction of high-throughput automated screening methods. These techniques not only speed up the process of crystallization screening but also increase the likelihood of success by allowing the screening of a larger number of different conditions while consuming limited amounts of sample (Banci *et al.*, 2006). One additional advantage of automated crystallization is that of producing large numbers of experiments under very

constant experimental conditions. This situation, which did not exist before the generalization of automated crystallization, represents a significant advantage for the correlation of crystallization results with experimental parameters, since the incidence of methodological and operational variability is minimized.

Currently, construct design and sample formulation are recognized as critical parameters and often a number of protein variants are assayed for crystallization either sequentially or in parallel. This strategy has proven successful (Banci *et al.*, 2006). However, it adds complexity to the process and requires significant additional effort. Nowadays, a substantial amount of work is dedicated to sample characterization and quality-control experiments in many crystallization projects, with the aim of identifying at an early stage and prioritizing those samples which would be more likely to crystallize (Geerlof *et al.*, 2006). This approach assumes that samples that are stable and monodisperse and that lack unfolded regions show a higher tendency to crystallize (Zulauf & D'Arcy, 1992) and typically involves experiments such as gel filtration, mass spectrometry, dynamic light scattering, ultracentrifugation or monodimensional NMR, for example. However, only very few studies have been carried out in order to correlate the results of such sample-characterization experiments with the likelihood of crystallization. Zulauf & D'Arcy (1992) analyzed 15 proteins by dynamic light scattering that were subsequently subjected to crystallization screening and found that proteins with a tendency to form aggregates in dilute solutions and in the absence of precipitants did not crystallize in the majority of cases. Nordlund and coworkers described a method to search for buffer components and small molecules that increase the thermal stability of the sample (Ericsson *et al.*, 2006). They found that additives that increased the apparent melting temperature (T_m) of the sample also produced an increase in the number of crystallization hits, while buffer components that produced a decrease in T_m led to a decrease in the number of crystallization hits. The Thermofluor assay (Pantoliano *et al.*, 2001) relies on the use of an environmentally sensitive fluorescent probe, *i.e.* a fluorophore with low quantum yields in water but that is highly fluorescent in organic solvents. This fluorophore is added to a protein solution, which is then subjected to stepwise increasing temperatures. As the protein denatures, its hydrophobic core is exposed and the fluorophore partitions into it, producing an increase in fluorescence (Pantoliano *et al.*, 2001). In this way, protein unfolding can be monitored as a function of temperature through the increase in fluorescence signal, the T_m being approximated as the midpoint in the protein-unfolding curve.

In this study, we have analyzed the thermal stability of 657 samples using a Thermofluor assay that were also subjected to automated crystallization screening at the High Throughput Crystallization Laboratory (HTX Lab) of the EMBL Grenoble Outstation under a constant protocol. Analysis of the results identifies a critical threshold for the value of T_m, as determined using this method, below which samples show a decreased crystallization success rate.

2. Methods

The samples used for this analysis were those submitted by users to the HTX Lab for automated crystallization screening over a period of 18 months. No selection criteria were applied other than that sufficient sample was available to carry out the Thermofluor assay. Samples were diluted to a final protein concentration of 10 μM in 0.040 ml of a solution consisting of 20 mM HEPES pH 7.5, 150 mM NaCl and 5× SYPRO Orange (a 5000× SYPRO Orange preparation from Molecular Probes, Invitrogen was used as stock solution; catalogue No. S6650). The samples were then subjected to thermal denaturation in a Real Time PCR machine (Stratagene Mx3005P) with a temperature gradient from 298 to 368 K in steps of 1 K and 1 min. Protein unfolding was monitored by the increase in the fluorescence of the SYPRO Orange probe, which was recorded every minute using excitation and emission wavelengths of 492 and 516 nm, respectively. The relative fluorescence emission intensity (R) was plotted as a function of the temperature and the T_m for each individual sample was estimated as the temperature corresponding to the midpoint between the baseline and the point with maximum fluorescence intensity. Nanovolume crystallization experiments were performed using a Cartesian PixSys 4200 robot (Genomic Solutions). Sitting-drop vapour-diffusion experiments were set up in Greiner CrystalQuick plates using 0.0001 ml of sample and 0.0001 ml of crystallization solution equilibrated against 0.088 ml of reservoir solution and incubated at 278 or 293 K (Dimasi *et al.*, 2007). The choice of incubation temperature was made by the user. All of the experimental parameters and results, including images of crystallization experiments, were automatically recorded and stored using the Crystallization Information Management System (CRIMS), an electronic laboratory information system developed by the HTX Lab. Three months after the initial setup, users were requested to provide feedback on the results of the crystallization experiments through the CRIMS interface. This time was judged to be sufficient to allow the user to perform X-ray diffraction tests with the crystals identified. Users were requested to indicate the conditions that produced crystals and whether these crystals were formed by protein or salt. The information concerning the conditions producing crystals was verified by staff at the HTX Lab. When discrepancies were noted between the information provided by the user and that recorded in CRIMS, the user was contacted for clarification. The crystallization screen was considered to be successful for a particular sample when at least one condition produced protein crystals.

3. Results and discussion

The HTX Lab (https://embl.fr/htxlab) is an open-access facility offering automated crystallization screening services to the members of the Partnership for Structural Biology (PSB) and to scientists working in European research institutions. This external access program is supported through the EC-funded Protein Production Platform (P-CUBE; http://www.p-cube.eu/).

The samples included in this analysis were those received at the HTX Lab over several months and no selection criteria were applied other than that sufficient sample was available to carry out the Thermofluor assay as described in §2. A total of 1534 Thermofluor experiments (including replicates) were carried out, corresponding to 657 unique samples from 173 different users. The majority of these samples originated from focused rather than structural genomics projects. 254 proteins were of bacterial origin, 203 were of human origin, 42 were from other mammals, 32 were from plants, 62 were from viruses and 64 were of other origins. A total of 45 different species are represented. Feedback on crystallization experiments was obtained for 93% of these samples (616 samples).

According to the results of the Thermofluor experiments, samples were classified into three groups (see Fig. 1*a*). In the majority of cases (66%) the samples produced typical denaturation curves with a clear and sharp temperature transition, allowing a straightforward estimation of the T_m. In 30% of cases (196 samples) no clear temperature transition was observed, precluding the calculation of a value for T_m. In a few cases (24 samples) complex temperature transitions were observed. Although the latter curves were often very informative they were not included in this analysis, since it was not possible to assign a single unambiguous T_m to these samples.

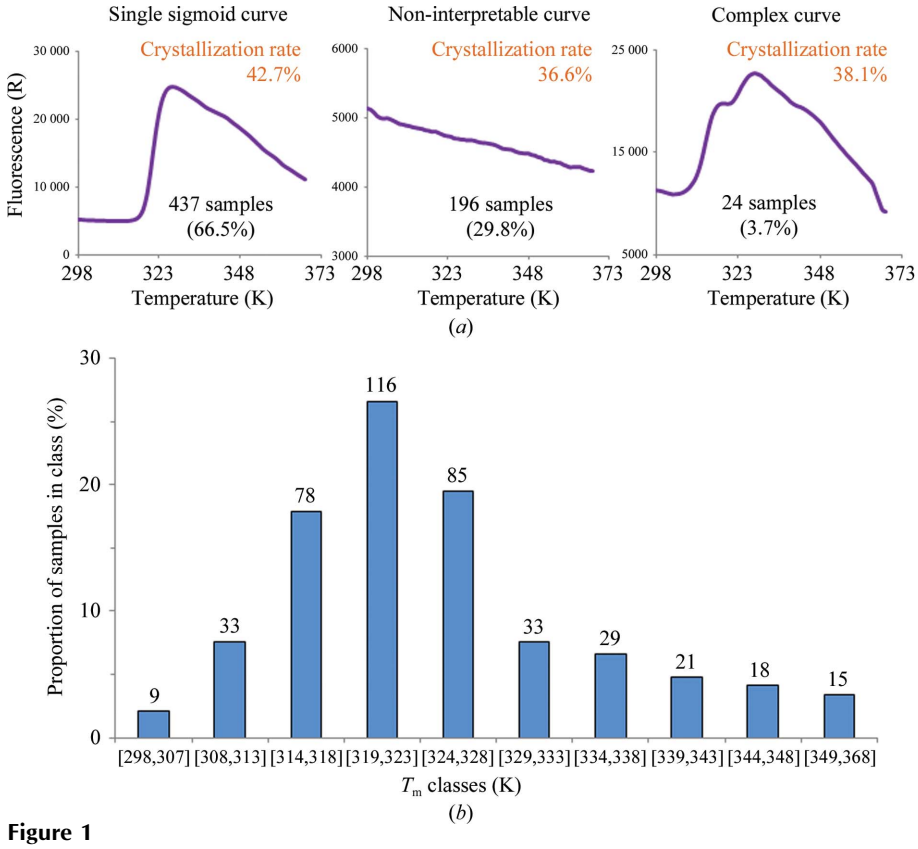

Figure 1
Thermofluor assay and T_m histogram. (*a*) According to the results of the Thermofluor experiment samples can be classified into three groups: those producing a sharp sigmoidal denaturation curve (left panel), those with no clear temperature transition (centre) and those with complex denaturation curves. The crystallization success rates for each of these groups are indicated. (*b*) Histogram of T_m for the 437 samples producing single sigmoidal curves. The numbers over the bars indicate the number of samples in the class.

Of the samples from the first group with single and readily interpretable temperature transitions 42.7% were crystallized, while only 36.6% of the samples that showed no temperature transition produced crystals. This suggests that although samples producing a negative result in the Thermofluor assay show a slightly lower success rate, lack of a clear temperature transition in this assay is not a strong discriminating factor.

Fig. 1(*b*) shows the distribution of T_m for the 437 samples from group 1. The average T_m for the ensemble of samples was 324.5 K, with the most populated groups being those with T_m in the range 319–323 K. The T_m histogram has a bell shape but is not exactly symmetric, showing a slight bias towards higher temperatures. This might reflect the use of traditional thermal stabilization screens by users before the samples were submitted to the HTX Lab or the preferential use of samples from extremophiles in some projects.

Differences in crystallization success rate are observed as a function of T_m. Fig. 2 shows the crystallization success rate as a function of T_m for samples assayed for crystallization at 293 K. As can be observed, samples with T_m values of 318 K or higher have a greater tendency to produce crystals (49.1% on average), while the crystallization success rate decreases rapidly for samples with lower values of T_m. Collectively, samples with T_m in the range 298–317 K produced crystals in 26.8% of cases and samples with T_m below 316 K produced crystals in 23.2% of cases. Notably, for samples with T_m between 318 and 368 K the crystallization rate remained fairly constant. In contrast, 12 samples originating from extremophile organisms, with T_m values between 343 and 368 K, showed a significantly higher success rate (69.2%) compared with samples from mesophilic organisms with similar T_m (see Fig. 2).

Our data suggest that there is a threshold value for the T_m of around 318 K above which the crystallization likelihood is maximized and remains constant above this value (except for samples of thermophilic origin). Taking into account that for the samples in Fig. 2 crystallization experiments were incubated at 293 K, the results indicate that crystallization likelihood is maximized if the T_m of the sample, as determined using our simplified Thermofluor method, is at least 25 K higher than the incubation temperature of the crystallization experiments and remains constant for higher values of T_m. In our view, this observation is in agreement with the basic principles determining protein stability. Protein folding is a highly cooperative process and protein-denaturation experiments typically lead to narrow sigmoidal curves (see Fig. 1*a*,

left panel). While small temperature changes in a range close to the T_m are expected to have a strong influence on the proportion of folded and unfolded protein for a particular sample, and hence its ability to crystallize, this influence would be very low at temperatures that are well above or below the T_m.

Following the same reasoning, it would be expected that for samples with T_m between 316 and 298 K the crystallization likelihood would be maximized if the incubation temperature for the crystallization experiments was 278 K instead of 293 K. Unfortunately, the number of samples in this study with T_m in this range and that were incubated at 278 K was too low to estimate a crystallization rate with confidence. However, this criterion has been used to systematically recommend incubation at low temperature to users of the HTX Lab, with positive results. As an example, Fig. 3 shows two proteins with a T_m of 313 K that did not produce crystals in the initial screening experiments at the incubation temperature selected by the user (293 K) but that produced crystals when the screening experiments were incubated at 278 K following our recommendations.

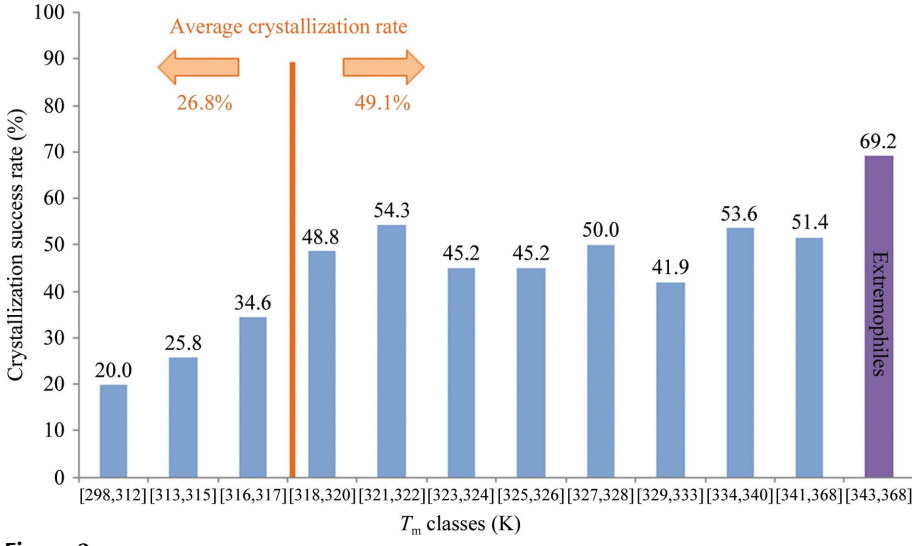

Figure 2
T_m and crystallization success rate: crystallization success rates as a function of T_m for samples assayed for crystallization at 293 K (excluding samples originating from extremophilic organisms). The crystallization success rate for 12 proteins originating from extremophilic organisms (with T_m between 343 and 368 K) is also shown. The numbers over the bars indicate the crystallization success rates for each class. The crystallization success rate decreases rapidly for samples with T_m below 318 K.

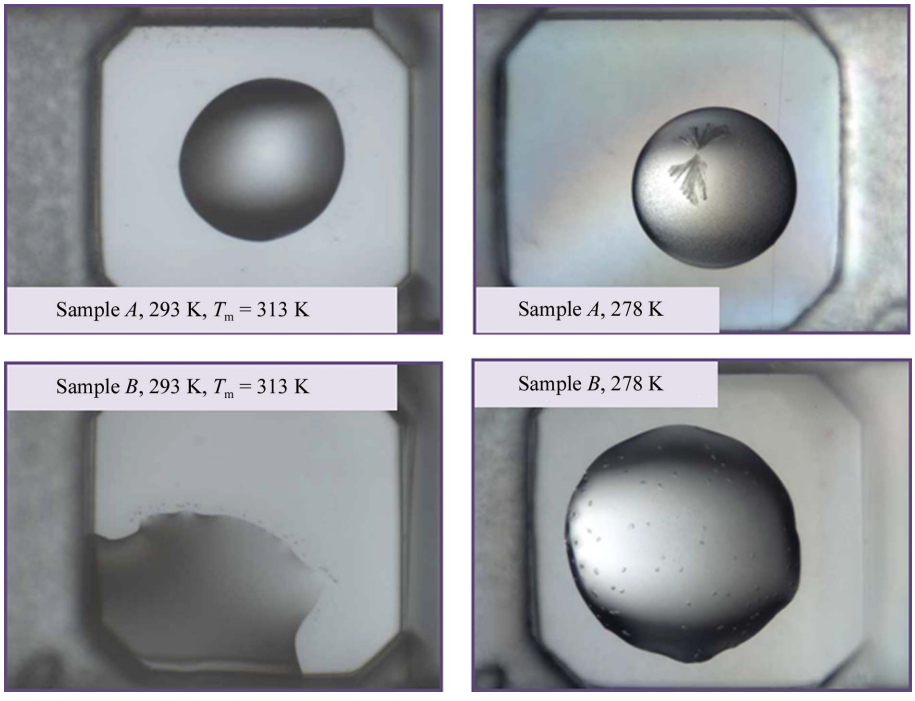

Figure 3
Samples with low T_m. The HTX Lab recommends systematic incubation of crystallization experiments at low temperature when the Thermofluor assay indicates a T_m of 317 K or lower. The figure shows two examples of proteins with a T_m of 313 K that did not produce crystals in the initial screening experiments (incubation at 293 K) but that produced crystals when the sample was re-screened at 278 K.

Our study confirms that proteins of thermophilic origin show a higher tendency to crystallize. Interestingly, the crystallization success rate of thermophilic proteins is notably higher than that of proteins from mesophilic organisms with a similar T_m. This suggests that the increased crystallization rate of thermophilic proteins is not only a consequence of increased folding stability. Thermophilic proteins have been shown to contain a lower proportion of unstructured regions (Szilágyi & Závodszky, 2000), which tend to inhibit crystallization. It may be that this characteristic or potentially other distinctive properties could explain their increased crystallization success rate.

It is well known that samples may show a different T_m in buffers of different compositions. However, the purpose of this work was not to find the optimal buffer for a given protein (methods for this have already been proposed by Ericsson *et al.*, 2006), but to compare the T_m values of a large number of proteins against a standard buffer of reference and to investigate whether they correlate with the likelihood of crystallization. We believe that the data that we provide here and our simplified Thermofluor assay can help

to estimate the crystallization likelihood of a biological sample. This method could be used to rationalize the decisions made with regard to prioritization of particular samples, to help to decide the incubation temperature for crystallization experiments or when to undertake other types of experiments oriented to modify sample properties, such as, for example, classical thermal stabilization screening experiments. The advantage of this approach over other sample-characterization techniques is that it consumes very low amounts of sample (0.4 nmol), is easy to carry out, is inexpensive and can be performed in 96-well format, making it compatible with the workflow at high-throughput platforms. We believe that this approach can be generally applied as an objective test to help the making of decisions in focused and structural genomics crystallography projects.

We would like to thank the users of the HTX Lab for providing the feedback on the results of crystallization experiments that has made this work possible and Stephen Cusack (EMBL) and Christoph Müller (EMBL) for their continued support of the crystallization facility. We would also like to thank Dr Timmins (ESRF) and Dr Gaboriaud (IBS) for the pictures shown in Fig. 3. Access to the High Throughput Crystallization Facility of the Partnership for Structural Biology in Grenoble (https://embl.fr/htxlab) for European researchers is funded by the P-CUBE project of the European Community (FP7/2007-2013; grant No. 227764).

References

Banci, L. *et al.* (2006). *Acta Cryst.* D**62**, 1208–1217.
Dimasi, N., Flot, D., Dupeux, F. & Márquez, J. A. (2007). *Acta Cryst.* F**63**, 204–208.
Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
Geerlof, A. *et al.* (2006). *Acta Cryst.* D**62**, 1125–1136.
McPherson, A. (1999). *Crystallization of Biological Macromolecules.* New York: Cold Spring Harbor Laboratory Press.
Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J. Biomol. Screen.* **6**, 429–440.
Szilágyi, A. & Závodszky, P. (2000). *Structure*, **8**, 493–504.
Zulauf, M. & D'Arcy, A. (1992). *J. Cryst. Growth*, **122**, 102–106.