**Chen-Yan Zhang,[‡] Zi-Qing Wu,[‡]
Da-Chuan Yin,^{*} Bo-Ru Zhou,
Yun-Zhu Guo, Hui-Meng Lu,
Ren-Bin Zhou and Peng Shang**

Institute for Special Environmental Biophysics,
Key Laboratory for Space Bioscience and
Biotechnology, School of Life Sciences,
Northwestern Polytechnical University, Xi'an,
Shaanxi 710072, People's Republic of China

[‡] These authors contributed equally to this work.

Correspondence e-mail: yindc@nwpu.edu.cn

Received 17 December 2012

Accepted 17 May 2013

A strategy for selecting the pH of protein solutions to enhance crystallization

The pH of a solution is an important parameter in crystallization that needs to be controlled in order to ensure success. The actual pH of the crystallization droplet is determined by the combined contribution of the buffers in the screening and protein solutions, although the contribution of the latter to the pH is often ignored. In this study, the effects of the buffer and protein solution pH values on the results of screening are systematically investigated. It was found that these parameters significantly affected the results and thus the following strategy for the selection of appropriate pH values is proposed: (i) when screening with only one protein solution, the pH should be as low, as high or as divergent from the pI as possible for a basic, acidic or neutral protein, respectively, within its stable pH range; (ii) when screening with two protein solutions, the pH values should be well separated from one another; and (iii) when multiple pH values are utilized, an even distribution of pH values is the best approach to increase the success rate of crystallization.

1. Introduction

The pH value of a solution is of great importance in protein crystallization, as it can modulate the charges on the protein molecules (Bernardo *et al.*, 2004), thereby affecting the electrostatic interactions between the molecules and further influencing molecular packing during crystallization (Neal *et al.*, 1999). Many reports have demonstrated that the pH of a solution affects crystal nucleation and growth processes (Lee *et al.*, 2001; Iwai *et al.*, 2008; Aldabaibeh *et al.*, 2009). For example, it has been reported that the area of the nucleation zone enlarges when the pH value diverges from the pI (isoelectric point) (Iwai *et al.*, 2008). Crystals have also been found to grow faster as the pH-governed charge on the protein increases (Schmit & Dill, 2012). Using data deposited in the PDB, Kantardjieff and Rupp analyzed the correlation between the pI and crystallization and found that the probability of crystallization increases at 0–2.5 pH units above the pI value for acidic proteins and at 0.5–3 pH units below the pI value for basic proteins (Kantardjieff & Rupp, 2004). Clearly, controlling the pH of a protein solution is an essential factor in ensuring successful protein crystallization (Neal *et al.*, 1999; Judge *et al.*, 1999).

The actual pH of the crystallization droplets is determined by the pH values of both the crystallization reagent and the protein solution. However, the pH of the protein solution is not intentionally controlled and the protein typically remains in solution throughout the purification process. In reality, there has been little consideration of whether the pH of the protein solution is suitable for crystallization. Some efforts (Jancarik *et al.*, 2004; Collins *et al.*, 2005) using initial tests to select a suitable protein buffer (and pH) prior to crystallization screening have been reported; however, the effect of the protein solution pH is often still ignored. Practically speaking, choosing the pH of a protein solution may require a trial-and-error strategy that can be directly applied without preliminary testing.

In this paper, we systematically investigated the effects of the protein solution pH on crystallization and demonstrated that the pH value can significantly affect the success rate of protein crystallization. Based on this investigation, we propose a strategy for selecting proper protein solution pH value(s) to promote protein crystallization.

Table 1

The proteins utilized in this investigation.

Protein	Supplier	Source	Concentration (mg ml ⁻¹)	pI value	Protein stable pH range
Lysozyme	Seikagaku	Chicken egg white	20	11.3	5–10 (Hayashi <i>et al.</i> , 1968; Sun <i>et al.</i> , 2011)
Concanavalin A	Sigma–Aldrich	<i>Canavalia ensiformis</i>	15	5.5	2–10 (Khan <i>et al.</i> , 2005)
α -Chymotrypsinogen A	Sigma–Aldrich	Bovine pancreas	15	9	3–11 (Blow, 1976)
Catalase	Sigma–Aldrich	Bovine liver	15	5	4–8.5 (Chance, 1952)
Glucose isomerase	Hampton Research	<i>Streptomyces rubiginosus</i>	7	3	6–8 (Schafhagser & Storey, 1992)
Myoglobin	Sigma–Aldrich	Horse heart	30	7	6–13 (Björck, 1949)

Table 2

The pH and the major chemical components of the two buffer systems.

pH value	Single-component buffer system	Multiple-component buffer system† (Newman, 2004)
3	25 mM sodium citrate	25 mM SA, SDP and GLY
4	25 mM sodium acetate	25 mM SA, SDP and GLY
5	25 mM sodium acetate	25 mM SA, SDP and GLY
6	25 mM sodium acetate	25 mM SA, SDP and GLY
7	25 mM sodium HEPES	25 mM SA, SDP and GLY
8	25 mM sodium HEPES	25 mM SA, SDP and GLY
9	25 mM Tris	25 mM SA, SDP and GLY
10	25 mM Tris	25 mM SA, SDP and GLY

† Succinic acid, sodium dihydrogen phosphate and glycine are abbreviated SA, SDP and GLY, respectively; they were mixed in a 2:7:7 ratio in the multiple-component buffer.

2. Materials and methods

2.1. Materials

Six commercial proteins (Table 1) were utilized without further treatment. They were chosen as model proteins in this study because their pI values are evenly distributed over a broad pH range (3–11.3). The chemical reagents sodium chloride and sodium citrate were purchased from the Chemical Reagent Co. Ltd (Beijing, People's Republic of China). Sodium acetate, succinic acid and sodium dihydrogen phosphate monohydrate were purchased from Beijing Chemical Factory (Beijing, People's Republic of China). Sodium HEPES and glycine were obtained from Amresco (Solon, USA) and Tris was purchased from Fluka (St Louis, USA). The Index crystallization screening kit was obtained from Hampton Research (Aliso Viejo, USA).

2.2. Experiments

2.2.1. Preparation of buffers and protein solutions. Proteins were dissolved in two sets of buffer systems (single-component buffer system and multiple-component buffer system) with pH values of 3, 4, 5, 6, 7, 8, 9 and 10 (Table 2). The single-component buffers (buffer system No. 1) were prepared using the following procedure. Sodium citrate, sodium acetate, sodium HEPES and Tris were each dissolved in pure deionized water ($R = 18.3 \text{ M}\Omega$) to a concentration of 25 mM and adjusted to the desired pH by adding 1 M HCl or 1 M NaOH (PB-10.23991772 pH meter, Sartorius, Germany; Stoll & Blanchard, 1990).

The multiple-component buffer system (Newman, 2004; buffer system No. 2) can be used to achieve a wide pH range without changing the chemical components of the buffer. Succinic acid, sodium dihydrogen phosphate monohydrate and glycine were mixed in a 2:7:7 ratio to obtain a 1 M stock solution. These buffers were prepared following the protocol provided in the literature to obtain the desired pH value (Newman, 2004) and were then diluted using pure deionized water to a final concentration of 25 mM. The protein solutions were prepared by formulating the proteins into these buffers at eight pH values. Table 2 lists the buffer information.

2.2.2. Crystallization screening experiments. The sitting-drop method was utilized in the crystallization screening experiments. The crystallization trials were set up by mixing the protein solution with the crystallization reagents from the Index screening kit in a volume ratio of 1 μ l:1 μ l. The volume of each reservoir was 80 μ l. An automated protein-crystallization robot (Screenmaker 96+8, Innovadyne Technologies Inc., USA) was utilized to prepare the trials. The crystallization plates used were 96-well sitting-drop Intelli-Plates (Hampton Research, USA). After preparing the crystallization trials, the crystallization plates were placed in a sealed chamber (inner dimensions 28 $\times$ 23 $\times$ 11 cm) that was connected to a programmable refrigerated circulator (Polyscience 9712 refrigerated circulator, Polyscience Inc., USA) to control the temperature inside the chamber to within ± 0.1 K. The temperature utilized was 293 K and the incubation time was 96 h. After the incubation period, images of the crystallization droplets were captured using an automated crystal image reader (XtalFinder; XtalQuest Inc., People's Republic of China) equipped with a UV light source (CRYSTALIGHT 100 UV source; Molecular Dimensions, USA). The crystallization hits ('hits' were defined as the number of crystallization conditions that yielded observable protein crystals under the microscope) were then obtained from the images.

3. Results

3.1. Effects of protein solution pH on crystallization screening

Crystallization screening experiments were conducted using several commercial proteins at different protein solution pH values (Fig. 1). The crystallization screening hits varied with the different protein solution pH values for the two buffer systems, indicating that a different choice of protein solution pH can result in different screening hits. An unsuitable protein solution pH can lead to a low crystallization success rate (for example, pH 3 or 10 for lysozyme in A1, pH 3 for concanavalin in B1, pH 9 for chymotrypsinogen A in C1 and pH 3 and 4 for glucose isomerase in E1 in Fig. 1) or even failure (for example, zero screening hits when the pH is 4, 5, 6 or 7 for myoglobin in F1 in Fig. 1). In contrast, an appropriate choice of protein solution pH can provide satisfactory results (for example, pH 5 for lysozyme in A1, pH 10 for concanavalin in B1, pH 3 for chymotrypsinogen A in C1, pH 8 for glucose isomerase in E1 and pH 10 for myoglobin in F1 in Fig. 1). This difference can be very large at different protein solution pH values. For example, in the case of glucose isomerase there was one screening hit at pH 3 or 4, but there were 37 screening hits at pH 8.

In further examining the data shown in Fig. 1, we found that the screening hits increased when the protein solution pH moved away from the pI value (within its stable pH range). This fact may provide a guide for selecting a pH value for protein solutions.

Fig. 1 also shows that utilization of the multiple-component buffer system (No. 2) exhibited a slightly higher success rate than the single-component buffer system (No. 1) in some cases. These results

demonstrate that the crystallization success rate was also dependent on the specific buffer system, with a multiple-component buffer system possibly providing a better performance in achieving a higher crystallization success rate than a single-component buffer system.

3.2. Effects of protein solution pH on the final pH of the crystallization solution

The above studies demonstrate that the pH value of the protein solution can significantly affect the crystallization process. A likely explanation is that the actual pH of the final crystallization solution was affected by the protein solution pH. To verify this assumption, we mixed 1 ml of crystallization reagents (selected from Index screening kits) with 1 ml of buffer system No. 1 and then measured the final pH of the mixture (Fig. 2). It is clear that the pH levels of both the crystallization reagent and the protein solution significantly influence the final pH value of the crystallization solution. For droplets of the same crystallization reagent mixed with a protein solution at different pH levels, the final pH of the droplets changed. This explains why varying the protein solution pH can yield new crystallization conditions, as shown in Fig. 1. In other words, using only one protein solution pH may ignore other potentially useful crystallization conditions.

4. Discussion

Considering the effects of the protein solution pH on crystallization, we here propose a strategy for selecting suitable protein solution pH value(s) to promote crystallization.

4.1. If only one protein solution pH is allowed

At a specific protein concentration, the lower the solubility is, the higher the supersaturation will be (Chen *et al.*, 2012; Yin *et al.*, 2002; Zhang *et al.*, 2008). Therefore, it was previously believed that the optimal pH for protein crystallization was the pI value of the protein owing to the pronounced solubility minimum (Rupp, 2009). However, an analysis based on data from the PDB revealed that there is no clear correlation between pI and pH for protein crystallization (Kantardjieff & Rupp, 2004). In fact, proteins tend to aggregate rapidly at their pI, often leading to the formation of amorphous precipitates (Rupp, 2009). Therefore, the pI value of a protein should be avoided when choosing the pH for a protein solution.

Another important consideration is the pH range that is suitable for formulating the protein. Proteins are only stable within a certain pH range (McPherson, 1995) and such stability is favourable for crystallization. Therefore, the selected protein solution pH must be within its stable pH range, which can be found in the literature for proteins that have been investigated extensively. For other proteins, such information can be obtained from certain biochemical experiments, such as thermal melt analysis (Ericsson *et al.*, 2006), the isotope method (Everley *et al.*, 2007) and the light absorption method (Landsman *et al.*, 1976).

The remaining question is which pH value is recommended. It has been suggested from the data accumulated in the PDB that protein crystallization is facilitated at 0–2.5 pH units above the pI for acidic proteins and 0.5–3 pH units below the pI for basic proteins, respectively (Kantardjieff & Rupp, 2004); however, the preferred pH value for crystallization remains unknown. Nevertheless, we may infer from

Figure 1

Number of crystallization screening hits for different protein-solution pH values formulated with two buffer systems. A, B, C, D, E, and F represent lysozyme, concanavalin, chymotrypsinogen A, catalase, glucose isomerase and myoglobin, respectively. 1 and 2 represent buffer systems Nos. 1 and 2, respectively. The concentration of the protein before mixing is shown in Table 1. The screening kit used was the Index screening kit from Hampton Research. The error bars indicate the standard deviation ($n = 3$).

Figure 2

The final pH of the crystallization reagent after mixing with protein buffer system No. 1, which was used for protein formulation (1:1 ratio). The open squares represent the pH of the crystallization reagent, the open triangles represent the pH of buffer system No. 1 and the solid dots represent the final pH values after mixing. Each mixture was labelled with a group number for easy identification. Group Nos. 1, 2, ..., 14 correspond to the serial numbers (Nos. 1, 1, 1, 2, 65, 70, 10, 75, 38, 56, 6, 6, 31 and 6, respectively) of crystallization reagents in the Index screening kit from Hampton Research.

a previous report that when the pH deviates from the protein pI the area of the nucleation zone increases (Iwai *et al.*, 2008), increasing the probability of obtaining crystals. Our results (Fig. 1) are in good agreement with this postulation. Thus, in accordance with the results obtained from our study and from previous studies, we recommend choosing a protein solution pH within the stable pH range that is as low, as high or as divergent from the pI as possible for basic, acidic or neutral proteins, respectively.

Although the protein solution pH values suggested in our strategy appear to be different for many proteins, they are actually not very different from those of the cumulative data (Kantardjieff & Rupp, 2004). For example, according to the rules that we suggest the protein solution pH for crystallizing myoglobin is pH 10. Because the majority of the pH values of the Index screening kit are in the range 5.5–8.5 (note that the pH values mentioned here are those of the buffers used in the crystallization reagents and are not exactly the actual pH values of the crystallization reagents; Newman *et al.*, 2010; Newman, 2011), the actual pH of the solution after mixing with the crystallization reagents will be in the range of approximately 8.5–9.4 (derived from our measurements), which is approximately 1.5–2.4 pH units from the pI of myoglobin. For proteins with a pI higher than 9 or lower than 5.5, the recommended pH may not be suitable for achieving crystallization at a pH that is approximately 0–3 units from the pI. Nevertheless, using this strategy for these proteins, we

Figure 3

The number of nonredundant crystallization hits plotted against the number of protein solution pH values for the two buffer systems. A, B, C, D, E and F represent lysozyme, concanavalin, chymotrypsinogen A, catalase, glucose isomerase and myoglobin, respectively. 1 and 2 represent buffer systems Nos. 1 and 2, respectively. Nonredundant crystallization hits was defined as the number of crystallization conditions that yielded protein crystals, in which repeated crystallization conditions were considered as one hit. The upper and lower limits are not error bars but rather represent the maximum and minimum number of crystallization hits, respectively. The black squares between the upper and lower limits are the average number of hits for the specific number of protein solution pH combinations and the red triangles represent the number of hits using the protein solution pH selection strategy proposed here.

obtained results (Fig. 1) that were found to be in good agreement with the literature (Iwai *et al.*, 2008), indicating that this strategy was still effective.

4.2. If multiple protein solution pH values are allowed

It has been reported that a change of 0.1 pH units can result in significant differences in protein solubility (Rupp, 2009; Chayen, 2005), which would ultimately affect the crystallization of the target protein. Therefore, a combination of multiple protein solution pH values is preferred to increase the likelihood of obtaining crystals.

In our investigation, we used eight pH values (3, 4, 5, 6, 7, 8, 9 and 10). If two protein solution pH values are used, there will be C_8^2 (= 28) possible combinations of pH values. Similarly, there will be 56, 70, 56, 28, eight and one possible pH combinations for three, four, five, six, seven and eight protein solution pH values, respectively. We counted the number of non-redundant crystallization hits for all possible pH combination groups and summarized the results in Fig. 3. As shown in this figure, a clear trend is observed as the average number of hits increases with more combined pH conditions. However, we also note that if the combination of protein solution pH values is not suitable, few crystallization hits will be obtained. For example, in the case of lysozyme (see A1 in Fig. 3), when protein solution pH values 9 and 10 were utilized 18 hits were obtained; significantly fewer than 40 hits would be obtained for the combination using protein solution pH values 5 and 10. Clearly, arbitrarily chosen pH combinations may result in a low probability of crystallization; therefore, a rational strategy to achieve higher crystallization hits is necessary.

Furthermore, if the selected protein solution pH values are too close to each other, many repetitions of the crystallization conditions may be encountered. To avoid this situation, we suggest choosing protein solution pH values that are sufficiently far apart. That is, for two pH combination groups, the second pH should be well separated from the first, which was chosen as proposed above. For example, if the first buffer pH has already been determined to be pH 5, then a second choice of pH 10 is better than pH 6 as it can provide a broader final pH range (approximately 5.1–6.7 and 7.7–9.2 as opposed to approximately 5.1–7.2). If more protein solution pH values are allowed, we suggest choosing pH values that are distributed as evenly as possible within the pH range.

According to the above strategy, we selected combinations of protein solution pH values and then examined the resulting crystallization hits. The detailed pH-selection results and numbers of nonredundant crystallization hits are given in the Supplementary Material¹ (Supplementary Tables S1-1 to S2-6). As shown in Fig. 3, rationally chosen protein solution pH values (shown as red triangles) can provide a higher number of hits compared with randomly chosen protein solution pH values, indicating that the pH selection strategy proposed above is effective. For example, in the case of lysozyme crystallization and buffer system No. 1, when the number of pH combinations of protein solution was two, three, four, five, six, seven and eight, the hit improvement (*i.e.* the ratio of the average number of hits to that for one pH condition) for lysozyme is 69.8, 86.3, 112.3, 128.8, 135.8, 137.3 and 144.7%, respectively. It can thus be concluded that rationally chosen protein solution pH values can help to significantly increase the likelihood of obtaining crystals.

4.3. Other considerations

In our current investigation, we chose pH values of integral units; however, finer grades of pH values can just as easily be utilized. This

would allow a further improvement of crystallization results with a more even distribution of pH values.

Finally, it is necessary to note that none of the proteins were desalted in this study; thus, the samples were likely to contain some chemical species that may have affected crystallization (Riès-Kautt *et al.*, 1994). Therefore, if the same protein from a different source is tested, a different number of hits could be obtained. However, the trend that the number of crystallization hits increases when the protein solution pH deviates from the pI value is expected to remain the same. Thus, the strategy proposed in this study is still applicable.

5. Conclusions

In this paper, we have demonstrated that varying the protein solution pH can significantly affect the crystallization results and that randomly chosen protein solution pH values can result in a low crystallization success rate. To rationally utilize the pH of protein solutions for improved crystallization, we propose the following strategy for selecting appropriate protein solution pH values without any initial testing: (i) when only one pH value is allowed, we recommend using a pH that is as low, as high or as divergent from the pI as possible for basic, acidic or neutral proteins, respectively, within their stable pH range; (ii) when two pH values are allowed, the values should be well separated within the stable pH range; and (iii) when multiple pH values are allowed, the chosen pH values should be distributed as evenly as possible (also applicable for biochemically uncharacterized proteins). Finally, it is necessary to emphasize that the buffer components could also affect the crystallization results, with multiple-component buffer systems exhibiting a better performance in increasing crystallization hits than single-component buffer systems.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31170816 and 11202167), the National Basic Research Program of China (973 Program; Grant No. 2011CB710905), the Natural Science Basic Research Plan of the Shaanxi Province of China (Grant No. 2012JQ3009), the China Postdoctoral Science Foundation (Grant No. 2012M512029) and the Fundamental Research Foundation of NPU in China (Grant No. JC20110285).

References

- Aldabaibeh, N., Jones, M. J., Myerson, A. S. & Ulrich, J. (2009). *Cryst. Growth Des.* **9**, 3313–3317.
- Bernardo, A., Calmanovici, C. E. & Miranda, E. A. (2004). *Cryst. Growth Des.* **4**, 799–805.
- Biörck, G. (1949). *Acta Med. Scand.* **133**(S226), 33–38.
- Blow, D. M. (1976). *Acc. Chem. Res.* **9**, 145–152.
- Chance, B. (1952). *J. Biol. Chem.* **194**, 471–481.
- Chayen, N. (2005). *Biophys. Mol. Biol.* **88**, 329–337.
- Chen, R.-Q., Yin, D.-C., Lu, Q.-Q., Shi, J.-Y. & Ma, X.-L. (2012). *Acta Cryst. D* **68**, 584–591.
- Collins, B., Stevens, R. C. & Page, R. (2005). *Acta Cryst. F* **61**, 1035–1038.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.
- Everley, P. A., Gartner, C. A., Haas, W., Saghatelian, A., Elias, J. E., Cravatt, B. F., Zetter, B. R. & Gygi, S. P. (2007). *Mol. Cell. Proteomics*, **6**, 1771–1777.
- Hayashi, K., Kugimiya, M. & Funatsu, M. (1968). *J. Biochem.* **64**, 93–97.
- Iwai, W., Yagi, D., Ishikawa, T., Ohnishi, Y., Tanaka, I. & Niimura, N. (2008). *J. Synchrotron Rad.* **15**, 312–315.
- Jancarik, J., Pufan, R., Hong, C., Kim, S.-H. & Kim, R. (2004). *Acta Cryst. D* **60**, 1670–1673.
- Judge, R. A., Jacobs, R. S., Frazier, T., Snell, E. H. & Pusey, M. L. (1999). *Biophys. J.* **77**, 1585–1593.
- Kantardjieff, K. A. & Rupp, B. (2004). *Bioinformatics*, **20**, 2162–2168.
- Khan, R. H., Naeem, A. & Baig, M. A. (2005). *Cell. Mol. Biol. Lett.* **10**, 61–72.

¹ Supplementary material has been deposited in the IUCr electronic archive (Reference: NJ5145).

- Landsman, M. L., Kwant, G., Mook, G. A. & Zijlstra, W. G. (1976). *J. Appl. Physiol.* **40**, 575–583.
- Lee, H. M., Kim, Y. W. & Baird, J. K. (2001). *J. Cryst. Growth*, **232**, 294–300.
- McPherson, A. (1995). *J. Appl. Cryst.* **28**, 362–365.
- Neal, B. L., Asthagiri, D., Velez, O. D., Lenhoff, A. M. & Kaler, E. W. (1999). *J. Cryst. Growth*, **196**, 377–387.
- Newman, J. (2004). *Acta Cryst. D* **60**, 610–612.
- Newman, J. (2011). *Methods*, **55**, 73–80.
- Newman, J., Fazio, V. J., Lawson, B. & Peat, T. S. (2010). *Cryst. Growth Des.* **10**, 2785–2792.
- Riès-Kautt, M., Ducruix, A. & Van Dorselaer, A. (1994). *Acta Cryst. D* **50**, 366–369.
- Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice and Application to Structural Biology*, pp. 89–90, 99–101. New York: Garland Science.
- Schafhagser, D. Y. & Storey, K. B. (1992). *Appl. Biochem. Biotechnol.* **36**, 63–74.
- Schmit, J. D. & Dill, K. (2012). *J. Am. Chem. Soc.* **134**, 3934–3937.
- Stoll, V. S. & Blanchard, J. S. (1990). *Methods Enzymol.* **182**, 24–38.
- Sun, J., Xu, R. & Yang, Y. (2011). *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **879**, 3053–3058.
- Yin, D. C., Inatomi, Y., Wakayama, N. I., Huang, W. D. & Kuribayashi, K. (2002). *Acta Cryst. D* **58**, 2024–2030.
- Zhang, C.-Y., Yin, D.-C., Lu, Q.-Q., Guo, Y.-Z., Guo, W.-H., Wang, X.-K., Li, H.-S., Lu, H.-M. & Ye, Y.-J. (2008). *Cryst. Growth Des.* **8**, 4227–4232.