used, particularly by biochemists. However, many crystallographers routinely use dynamic light scattering to check concentrated protein preparations for aggregation (Ferré-D'Amaré & Burley, 1994). The method is relatively simple, very sensitive to small amounts of aggregation and has the additional advantage that it does not consume the sample. After testing, the sample (which is often precious) can still be used for crystallization trials.

If sample heterogeneity is detected, one is faced with the issue of whether it will adversely affect crystallization, and if so, how to remove it. Unfortunately, there do not seem to be general rules. Heterogeneity at the termini of proteins is a common occurrence. In many crystal structures, the termini are disordered and heterogeneity at these unstructured ends would not be expected to be a significant problem. Indeed, in a number of instances, N-terminal sequence analysis of proteins obtained by dissolving crystals has indicated substantial heterogeneity. However, in other cases, properly defined domain boundaries are thought to have been a critical factor in obtaining useful crystals. Domain boundaries can be determined by a combination of limited proteolysis, followed by identification of the fragments using mass spectrometry (Cohen *et al.*, 1995; Hubbard, 1998). Subsequent re-engineering of expression constructs with modified termini is a relatively easy task. Similar engineering can also be used to alter internal sequences, such as removal of sites of post-translational modification or introduction of mutations that improve solubility (Chapter 4.3).

3.1.6.2. Protein storage

Even when the efforts of those engaged in crystallization and those engaged in producing the desired recombinant protein are well coordinated, it is not usually appropriate or desirable to use all the available protein for crystallization at the same time. This means that some of the material must be stored for later use. Even under the best of circumstances, protein solutions are subject to a number of unwanted events that can include, but are not limited to, oxidation, racemization, deamination, denaturation, proteolysis and aggregation. As a general rule, it is better to store proteins as highly purified concentrated solutions. This reduces problems of proteolysis (since the proteases have been removed), and, in general, proteins are better behaved if they are relatively concentrated (greater than 1 mg ml^{-1}). This is not an absolute rule, however; if there are problems with aggregation, these can sometimes be minimized by storage of proteins in dilute solutions, followed by concentration of the samples immediately prior to crystallization. If the protein contains oxidizable sulfurs (free cysteines are a particular problem), reducing agents can be added (and should be refreshed as necessary), and the solutions held in a non-reducing (N_2) atmosphere. In some cases, it is easier to mutate surface cysteines to produce a more stable protein (see Chapter 4.3).

In general, proteins behave best under conditions of pH and ionic strength similar to those they would experience in the normal host. Usually this means a pH near, or slightly above, neutral and intermediate ionic strength. These conditions are often not the ideal conditions for crystallization, and dialysis or other forms of buffer exchange may be required before beginning crystallization trials. In general, protein solutions are stored either at 4 °C in a cold room or refrigerator, or at 0 °C on ice. It is essential that the protein be stored in a manner that will not allow microbial growth, usually achieved by sterilization of the protein solution by filtration through 0.2 micron filters and/or addition of antimicrobial agents, such as NaN_3. For long-term storage (periods longer than a few weeks), protein solutions are often precipitated in ammonium sulfate or frozen at either −20 or −70 °C. Repeated freezing and thawing is not recommended; if a protein sample is to be frozen, it should be divided into aliquots small enough so that each will be thawed only once. Whenever a protein sample is frozen and thawed, some loss of quality and/or activity can be expected. Freezing samples of intermediate concentration (1–3 mg ml^{-1}) usually works better than freezing either extremely dilute or concentrated samples. Cryoprotective agents can be added to protein samples destined to be frozen; however, it should be remembered that the same reagents that are helpful when freezing a protein sample may be distinctly unhelpful when that sample is thawed and used for crystallography. Most biochemists willingly add glycerol to their protein samples before freezing; crystallographers are not usually happy to find that their protein sample is dissolved in 50% glycerol. Both pH and ionic strength can affect a protein's tolerance to freezing and thawing. In many cases, buffer exchange and concentration procedures need to be performed to convert stored protein solutions to ones suitable for crystallization.

As is so often true in science, decisions about whether to freeze a particular protein sample and, if it is to be frozen, exactly how the freezing should be done, depend on experience. If the protein in question is an enzyme, it is often useful to set up a series of trials in which small aliquots of the protein are stored under a variety of conditions. If the aliquots are tested on a fairly regular basis, how stable the protein is in solution can usually be determined, as well as how well it will tolerate a cycle of freezing and thawing, with or without an added cryoprotectant. If enzyme assays are not available, other methods of characterization, such as gel electrophoresis, mass spectrometry and light scattering, can be used to check for degradation, oxidation of cysteines and aggregation. Armed with this information, and with a plan for how the protein will be used for crystallization, it is usually a fairly simple matter to decide whether or not to freeze a particular sample, and, if the sample is to be frozen, how best to do it. It is a good idea to make such tests early in a major crystallization effort. This will avoid the awkward dilemma that occurs when a large amount of a highly purified protein is available, and the knowledge of how best to store it is not.

3.1.7. Reprise

We have reached a point where it is possible to use recombinant DNA techniques to produce most proteins in quantities sufficient for crystallography. Both high-level expression systems and methods for making defined modifications of recombinant proteins vastly simplify the process of purification. This has played a direct and critical role in the ability of crystallographers to produce an astonishing array of new and exciting protein structures. We are beginning to come to grips with the next level of the problem: using the ability to modify the sequence of proteins to improve their crystallization properties. This is a difficult problem; however, there are already notable, if hard won, successes. It would appear that the marriage of genetic engineering and crystallography – clearly a case in which opposites attract – has been a happy union. This is entirely for the good. Collaborations between specialists in these disciplines have led to the solution of problems too difficult for any individual armed only with the skills of one or the other partner. It is important that genetic engineering be fully integrated into future crystallographic efforts, either directly within the crystallography

laboratory or through close collaborations. There yet remain formidable problems in protein structure and function that will require all the combined talents of the most skilled practitioners of these arcane arts.

References

Abelson, J. N. & Simon, M. I. (1990). *Guide to protein purification. Methods Enzymol.* **182**, 1–894.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1995). *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 3rd ed. New York: Greene Publishing Associates and Wiley.

Beggs, J. D. (1978). *Transformation of yeast by a replicating hybrid plasmid. Nature (London)*, **275**, 104–109.

Bhandari, P. & Gowrishankar, J. (1997). *An Escherichia coli host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. J. Bacteriol.* **179**, 4403–4406.

Biswas, E. E., Fricke, W. M., Chen, P. H. & Biswas, S. B. (1997). *Yeast DNA helicase A: cloning, expression, purification, and enzymatic characterization. Biochemistry*, **36**, 13277–13284.

Bollag, D. M., Rozycki, M. D. & Edelstein, S. J. (1996). *Protein Methods.* New York: Wiley-Liss.

Boyer, P. L. & Hughes, S. H. (1996). *Site-directed mutagenic analysis of viral polymerases and related proteins. Methods Enzymol.* **275**, 538–555.

Brinkmann, U., Mattes, R. E. & Buckel, P. (1989). *High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. Gene*, **85**, 109–114.

Broach, J. R. (1983). *Construction of high copy number yeast vectors using 2 μm circle sequences. Methods Enzymol.* **101**, 307–325.

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H. & Xu, M. Q. (1997). *Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene*, **192**, 271–281.

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler, F. B. & Xu, M. Q. (1996). *Protein splicing involving the Saccharomyces cerevisiae VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J. Biol. Chem.* **271**, 22159–22168.

Cohen, S. L., Ferre-D'Amare, A. R., Burley, S. K. & Chait, B. T. (1995). *Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. Protein Sci.* **46**, 1088–1099.

Cole, P. A. (1996). *Chaperone-assisted protein expression. Structure*, **4**, 239–242.

Cregg, J. M., Vedick, T. S. & Raschke, W. C. (1993). *Recent advances in the expression of foreign genes in Pichia pastoris. Biotechnology*, **11**, 905–910.

De Bernardez Clark, E. (1998). *Refolding of recombinant proteins. Curr. Opin. Biotechnol.* **9**, 157–163.

De Boer, H. A. & Kastelein, R. A. (1986). *Biased codon usage: an exploration of its role in optimization of translation.* In *Maximizing Gene Expression*, edited by W. S. Reznikioff & L. Gold, pp. 225–285. Boston: Butterworths.

Del Tito, B. J. Jr, Ward, J. M., Hodgson, J., Gershater, C. J. L., Edwards, H., Wysocki, L. A., Watson, F. A., Sathe, G. & Kane, J. F. (1995). *Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli. J. Bacteriol.* **177**, 7086–7091.

Enfors, S.-O. (1992). *Control of in vivo proteolysis in the production of recombinant proteins. Trends Biotechnol.* **10**, 310–315.

Ernst, J. F. & Kawashima, E. (1988). *Variations in codon usage are not correlated with heterologous gene expression in Saccharomyces cerevisiae and Escherichia coli. J. Biotechnol.* **7**, 1–9.

Ferré-D'Amaré, A. R. & Burley, S. K. (1994). *Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. Structure*, **2**, 357–359.

Fischer, B., Sumner, I. & Goodenough, P. (1993). *Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies. Biotechnol. Bioeng.* **41**, 3–13.

Georgiou, G. & Valax, P. (1996). *Expression of correctly folded proteins in Escherichia coli. Curr. Opin. Biotechnol.* **7**, 190–197.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981). *Translational initiation in prokaryotes. Annu. Rev. Microbiol.* **35**, 365–403.

Goldman, E., Rosenberg, A. H., Zubay, G. & Studier, F. W. (1995). *Consecutive low-usage leucine codons block translation only when near the 5′ end of a message in Escherichia coli. J. Mol. Biol.* **245**, 467–473.

Gottesman, S. (1990). *Minimizing proteolysis in Escherichia coli: genetic solutions. Methods Enzymol.* **185**, 119–129.

Grinna, L. S. & Tschopp, J. F. (1989). *Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, Pichia pastoris. Yeast*, **5**, 107–115.

Guise, A. D., West, S. M. & Chaudhuri, J. B. (1996). *Protein folding in vivo and renaturation of recombinant proteins from inclusion bodies. Mol. Biotechnol.* **6**, 53–64.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J.* **9**, 1665–1672.

Hernan, R. A., Hui, H. L., Andracki, M. E., Noble, R. W., Sligar, S. G., Walder, J. A. & Walder, R. Y. (1992). *Human hemoglobin expression in Escherichia coli: importance of optimal codon usage. Biochemistry*, **31**, 8619–8628.

Higgins, D. R. & Cregg, J. (1998). *Methods in Molecular Biology*, Vol. 103. *Pichia protocols.* Totowa: Humana Press.

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. (1989). *Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl Acad. Sci. USA*, **86**, 8247–8251.

Hockney, R. C. (1994). *Recent developments in heterologous protein production in Escherichia coli. Trends Biotechnol.* **12**, 456–463.

Hofmann, A., Tai, M., Wong, W. & Glabe, C. G. (1995). *A sparse matrix screen to establish initial conditions for protein renaturation. Anal. Biochem.* **230**, 8–15.

Hollenberg, C. P. & Gellissen, G. (1997). *Production of recombinant proteins by methylotrophic yeasts. Curr. Opin. Biotechnol.* **8**, 554–560.

Hubbard, S. J. (1998). *The structural aspects of limited proteolysis of native proteins. Biochim. Biophys. Acta*, **1382**, 191–206.

Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (1990). *PCR Protocols: A Guide to Methods and Applications.* San Diego: Academic Press.

Jarvis, D. L., Kawar, Z. S. & Hollister, J. R. (1998). *Engineering N-glycosylation pathways in the baculovirus-insect cell system. Curr. Opin. Biotechnol.* **9**, 528–533.

Jones, I. & Morikawa, Y. (1996). *Baculovirus vectors for expression in insect cells. Curr. Opin. Biotechnol.* **7**, 512–516.

Kane, J. F. (1995). *Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr. Opin. Biotechnol.* **6**, 494–500.

Kaufman, R. J. (1990). *Selection and coamplification of heterologous genes in mammalian cells. Methods Enzymol.* **185**, 537–566.

Kim, R., Sandler, S. J., Goldman, S., Yokota, H., Clark, A. J. & Kim, S.-H. (1998). *Overexpression of archaeal proteins in Escherichia coli. Biotechnol. Lett.* **20**, 207–210.

Krueger, J. K., Kulke, M. H., Schutt, C. & Stock, J. (1989). *Protein inclusion body formation and purification. BioPharm*, March issue, 40–45.

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature (London)*, **393**, 648–659.

LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. & McCoy, J. M. (1993). *A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology*, **11**, 187–193.

LaVallie, E. R. & McCoy, J. M. (1995). *Gene fusion expression systems in Esherichia coli. Curr. Opin. Biotechnol.* **6**, 501–506.

Lee, H. W., Joo, J.-H., Kang, S., Song, L.-S., Kwon, J.-B., Han, M. H. & Na, D. S. (1992). *Expression of human interleukin-2 from native and synthetic genes in E. coli: no correlation between major codon bias and high level expression. Biotechnol. Lett.* **14**, 653–658.

Lu, A. & Miller, L. K. (1996). *Generation of recombinant baculoviruses by direct cloning. Biotechniques*, **21**, 63–68.