[2] Optimization of Protein Solubility and Stability for Protein Nuclear Magnetic Resonance

By STEFAN BAGBY, KIT I. TONG, and MITSUHIKO IKURA

Introduction

Nuclear magnetic resonance (NMR) spectroscopic techniques and hardware for the study of biomacromolecular structure and function have developed to the point where we can envisage obtaining high quality spectra of biomacromolecules and biomacromolecular assemblies of greater than 100 kDa molecular mass.[1] This will permit structure determinations of larger proteins that cannot be crystallized and allow studies of many intermolecular interactions in solution.[2,3]

For those NMR laboratories focusing on a particular target or type of target for structural analysis and to a lesser extent those pursuing a structural genomics approach,[4] there remains, however, the prosaic but fundamental and often difficult problem of generating suitable samples for detailed NMR study: one of the major bottlenecks in the analysis of protein structure and function in solution by high resolution NMR methods is generating protein samples that are stable and soluble. NMR studies require the protein to be stable in the magnet for several weeks (unless the researchers have the time, energy, and funds to prepare numerous batches of sample) at high concentrations (ideally 1 m*M* or higher). This problem has been exacerbated by the move toward study of larger proteins by NMR with their greater tendency to aggregate.

Here we review methods that have been developed to optimize the polypeptide construct, facilitate initial screening of structural integrity, and assess aggregation state. We consider additives that may be used to improve protein stability and solubility at high concentrations without affecting the structure of the protein and protocols that have been developed to allow screening of a wide range of solution conditions for protein NMR studies using small amounts of protein.

Polypeptide Constructs: Defining Domain Boundaries and Segmental Isotope Labeling

Large proteins typically comprise several smaller domains, with most domains falling in the range of 100–250 amino acids. To date technical limitations have

[1] G. Wider and K. Wüthrich, *Curr. Opin. Struct. Biol.* **9,** 594 (1999).
[2] H. Takahashi, T. Nakanishi, K. Kami, Y. Arata, and I. Shimada, *Nat. Struct. Biol.* **7,** 220 (2000).
[3] K. Wüthrich, *Nat. Struct. Biol.* **7,** 188 (2000).
[4] S. K. Burley, S. C. Almo, J. B. Bonanno, M. Capel, M. R. Chance, T. Gaasterland, D. Lin, A. Sali, F. W. Studier, and S. Swaminathan, *Nat. Genet.* **23,** 151 (1999).

0076-6879/00 $35.00

forced structural biologists using NMR to tackle such multidomain proteins by a "divide and conquer" strategy whereby single domains are studied in isolation. Because most full-length proteins are only marginally stable at physiological temperature, and selection of the start and end points for subcloning of domains has often been carried out in the absence of concrete information on the domain boundaries, it is not surprising that isolated fragments are frequently partially unfolded and/or prone to aggregation. Subcloning sites may typically be selected using secondary structure prediction and alignment of multiple sequences.

Structural information that permits identification of domain boundaries and therefore assists in selection of suitable sites for subcloning can alternatively be obtained from limited proteolysis, N-terminal sequencing, and matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI) mass spectrometry.[5,6] The principle of this method for designing constructs is that amino acid residues within a folded domain are protected from proteolysis whereas solvent-exposed, flexible amino acid residues are susceptible to rapid cleavage. The fragments generated by limited proteolysis are separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) or high-performance liquid chromatography (HPLC) and characterized by N-terminal sequencing and MALDI or ESI mass spectrometry. This method has been used to define the domain boundaries of a number of proteins, including Max[5] and an NAD^+-dependent DNA ligase.[6]

An alternative to the divide and conquer approach for studies of multidomain proteins is to generate multidomain polypeptides in which only one of the domains is labeled with NMR-active isotope(s) such as ^{15}N and/or ^{13}C.[7–10] Techniques for joining together protein segments,[7–9,11] based on protein splicing, permit such domain-selective labeling and potentially allow structure determination by NMR of a single domain within the context of the full-length protein. Yamazaki and coworkers[7] have developed a trans-splicing approach to segmental labeling of proteins for NMR studies which involves a denaturation step. A mild chemical ligation procedure for joining together folded recombinant domains which does not require a denaturation step has been demonstrated by chemical ligation of the SH3 and SH2 domains of the Abelson tyrosine kinase with the SH2 domain ^{15}N-labeled.[9] The peaks in the ^1H–^{15}N heteronuclear single quantum coherence (HSQC) spectrum of the SH2 domain in this chemically produced fusion coincided almost exactly with those of the recombinant SH3–SH2 construct in which both domains

[5] S. L. Cohen, A. R. Ferré-d'Amaré, S. K. Burley, and B. T. Chait, *Prot. Sci.* **4,** 1088 (1995).

[6] D. J. Timson and D. B. Wigley, *J. Mol. Biol.* **285,** 73 (1999).

[7] T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino, and H. Nakamura, *J. Am. Chem. Soc.* **120,** 5591 (1998).

[8] T. Otomo, N. Ito, Y. Kyogoku, and T. Yamazaki, *Biochemistry* **38,** 16040 (1999).

[9] R. Xu, B. Ayers, D. Cowburn, and T. W. Muir, *Proc. Natl. Acad. Sci. U.S.A.* **96,** 388 (1999).

[10] D. Cowburn and T. Muir, *Methods Enzymol.* **339,** [3] 2001 (this volume).

[11] T. W. Muir, D. Sondhi, and P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **95,** 6705 (1998).

were ^{15}N-labeled. Both trans-splicing and chemical ligation approaches can be extended to allow three recombinant protein segments to be regioselectively linked together: the feasibility of joining three segments by chemical ligation has been demonstrated in a model synthetic peptide system[12]; *trans*-splicing has been used for selective isotope labeling of a central segment of maltose binding protein[8] and can be used to label selectively any segment between structurally flexible residues.

Polypeptide Folding

Once the polypeptide construct has been decided, the usual sequence of events would involve protein expression, protein purification, and then qualitative assessment of the structural integrity of the pure protein by recording a fingerprint spectrum such as ^{1}H–^{15}N HSQC; the backbone amide cross peaks in such a spectrum will cluster around 8 ppm if the protein is denatured. This lengthy and labor-intensive process is often fruitless, particularly if the construct boundaries are selected using alignment of multiple sequences or data from secondary structure prediction rather than the more rigorous limited proteolysis/mass spectrometry method discussed above. Structural integrity of the protein can instead be assessed rapidly by expression of the protein in ^{15}N-labeled minimal medium, removal of the cell debris, and acquisition of a ^{1}H–^{15}N HSQC spectrum on the crude cell lysate. This was illustrated for two proteins, interleukin-1β and a double mutant of the B1 immunoglobulin (Ig) binding domain of streptococcal protein G, both of which comprised 15–25% of total expressed cellular protein.[13] In these cases, $^{15}NH_4Cl$ was used as the sole nitrogen source throughout the growth of the cells. In cases where the protein of interest is expressed at levels corresponding to 5–10% of total cellular protein, $^{14}NH_4Cl$ can be used as the nitrogen source until just prior to induction when the medium is changed to one that contains $^{15}NH_4Cl$ as the sole nitrogen source. If the peak dispersion observed in the ^{1}H–^{15}N HSQC spectrum indicates that the protein or protein fragment is folded, then it is obviously worth proceeding with further purification and spectral analysis.

This fast and simple method to assess the structural integrity of overexpressed proteins and domains may not be applicable to proteins that are very sensitive to solution conditions. It is also only applicable to proteins that are expressed in a soluble form, i.e., not packaged into inclusion bodies. These limitations have been tackled by the design of expression vectors specifically for the purpose of rapid screening by NMR. The vectors reported by Huth *et al.*,[14] for example, encode the immunoglobulin-binding domain of streptococcal protein G (GB1 domain) fused to the N terminus of the relevant protein or protein fragment. The presence

[12] J. A. Camarero, G. J. Cotton, A. Adeva, and T. W. Muir, *J. Pept. Res.* **51,** 303 (1998).

[13] A. M. Gronenborn and G. M. Clore, *Protein Sci.* **5,** 174 (1996).

[14] J. R. Huth, C. A. Bewley, B. M. Jackson, A. G. Hinnebusch, G. M. Clore, and A. M. Gronenborn, *Prot. Sci.* **6,** 2359 (1997).

of the GB1 domain enhances expression and improves the chances of expression in a soluble form, and its small size (56 amino acid residues) means that NMR spectra can be acquired without separating the GB1 domain from the protein of interest. This last point represents a considerable advantage for rapid screening over expression systems that encode fusions with larger proteins such as glutathione transferase and maltose binding protein, where the fusion must be cleaved before structural integrity can readily be assessed.

Typically, 0.1–1.0 liter cultures are required for screening and the GB1 fusions offer the choice of recording a ^{1}H–^{15}N HSQC spectrum on the crude cell lysate or purifying the fusion protein using Ni^{2+} or IgG Sepharose affinity chromatography prior to acquisition of the ^{1}H–^{15}N HSQC spectrum.[14] Proteins greater than 30 kDa molecular weight may require use of ^{1}H–^{15}N transverse relaxation-optimized spectroscopy (TROSY)[15,16] instead of a standard ^{1}H–^{15}N HSQC for screening of structural integrity since the TROSY technique provides superior spectral resolution and improved effective sensitivity for larger proteins.

As an alternative to NMR spectroscopy, circular dichroism (CD) spectroscopy can be used to assess the structural integrity of the polypeptide. Many secondary structure motifs in proteins, such as the α helix, β sheet, and β turn, give rise to characteristic CD spectra, and CD spectroscopy can be used to estimate the percentage secondary structure composition of polypeptides.[17] The sample for CD spectroscopy must be free of contaminating proteins and other optically active impurities such as nucleotides and also free of optically active buffer material or additives. One advantage of CD spectroscopy for preliminary characterization is that spectra can be recorded with relatively small amounts of protein. CD spectroscopy therefore permits the researcher to assess the likely value of further work to optimize expression and purification in cases where initial protein preparations provide low yields. For example, we have recorded CD spectra of the N-terminal adhesion domain of epithelial cadherin using protein concentrations as low as 8 μM (0.12 mg/ml),[18] although higher concentrations (50–100 μM) would typically be used. Details of instrumentation, sample preparation, and theory and applications of CD spectroscopy have been reviewed.[19]

Aggregation State of Polypeptide

Having confirmed that the polypeptide is folded, it is equally important to assess the aggregation state of the polypeptide. Polypeptide solutions used for structural

[15] K. Pervushin, R. Riek, G. Wider, and K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.* **94,** 12366 (1997).
[16] L. E. Kay, *Methods Enzymol.* **339,** [9] 2001 (this volume).
[17] N. J. Greenfield, *Anal. Biochem.* **235,** 1 (1996).
[18] K. I. Tong, P. Yau, M. Overduin, S. Bagby, T. Porumb, M. Takeichi, and M. Ikura, *FEBS Lett.* **352,** 318 (1994).
[19] R. W. Woody, *Methods Enzymol.* **246,** 34 (1995).

studies by NMR (and for protein crystallization) should exhibit homogeneous association behavior, i.e., the solution should consist uniformly of monomers (usually most convenient) or dimers or trimers or other multimers rather than a mixture of different types of aggregate. Such a uniform solution is said to be monodisperse.

Light scattering[20] can provide information on a molecule's hydrodynamic radius, particle size and dimensions, mono/polydispersity, and molecular size distribution. Light scattering has been more widely exploited in protein crystallography to assess the crystallizability[21] and other properties[22] of macromolecules, but should be equally useful in combination with NMR studies. Analytical ultracentrifugation can be used to provide similar information on the properties of solutions of biomacromolecules.[23,24] When used to quantify the degree of protein self-association as a function of solution conditions and protein concentration, light scattering and analytical ultracentrifugation can be a powerful adjunct to the microdialysis button test and microdrop screen (described below) that are used for rapid assessment of optimal solution conditions for NMR studies.

Folded Polypeptides: Optimization of Solubility and Stability

Solvent Additives and Protocols for Screening Solution Conditions

Details of two protocols for rapid and efficient screening of solution conditions, the microdialysis button test[25] and microdrop screen,[26] have been published. Both rely on tools and techniques previously used for protein crystallization trials, but obviously with the opposite aim of maximizing protein solubility rather than precipitation. Starting with a single solution condition, small volumes (1–5 μl) of protein are mixed with solutions in which pH, buffer type, salt concentration, and additives are varied systematically. Both methods rely on visual inspection of the sample to determine the extent, if any, of protein precipitation. Many proteins can be concentrated to levels suitable for high resolution NMR studies but subsequently precipitate over a period of hours or days. Such proteins are particularly suitable for screening by the microdialysis button test or microdrop screen to determine solution conditions for optimum solubility. In both cases, the protein is first concentrated to its solubility limit.

In either screen, it is helpful to adopt a progression[26] in which first the optimum pH and buffer type are established. The screens are best started with buffers that

[20] P. J. Wyatt, *Anal. Chim. Acta* **272,** 1 (1993).
[21] A. R. Ferré d'Amaré and S. K. Burley, *Methods Enzymol.* **276,** 157 (1997).
[22] H. Wu, P. D. Kwong, and W. A. Hendrickson, *Nature* **387,** 527 (1997).
[23] P. Hensley, *Structure* **4,** 367 (1996).
[24] T. M. Laue and W. F. Stafford III, *Annu. Rev. Biophys. Biomolec. Struct.* **28,** 75 (1999).
[25] S. Bagby, K. I. Tong, D. Liu, J. R. Alattia, and M. Ikura, *J. Biomol. NMR* **10,** 279 (1997).
[26] C. A. Lepre and J. M. Moore, *J. Biomol NMR* **12,** 493 (1998).

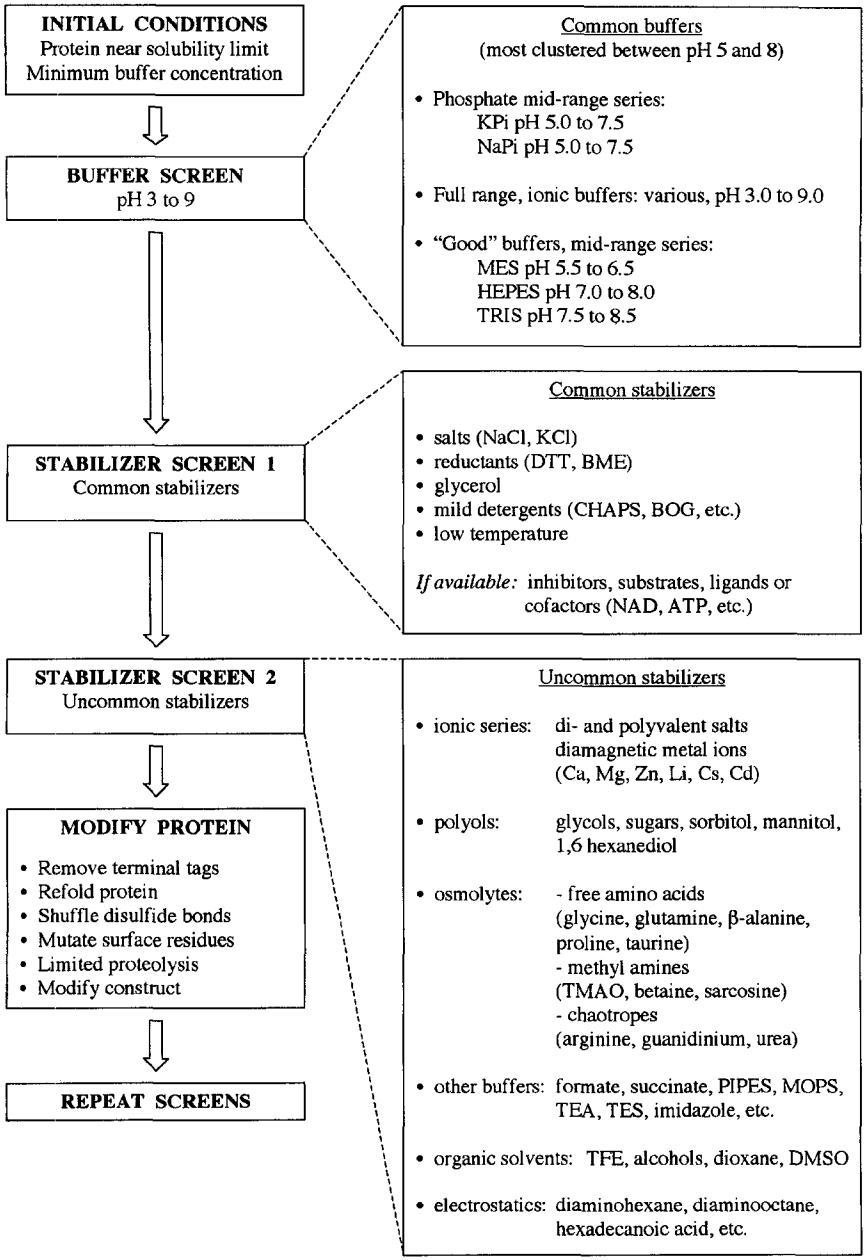

FIG. 1. Flow chart depicting a typical progression for screening solution conditions using either the microdialysis button test[25] or microdrop screen.[26] (Reproduced with permission from Lepre and Moore.[26])

TABLE I
PROTEIN COSOLUTES

Common additives	Suggested amount	References[a]
Ionic compounds and salts		
Na_2SO_4	0.25–1.0 *M*	(1, 2, 3)
$MgSO_4$	0.60–1.0 *M*	(1, 4, 5, 6)
Guanidine sulfate	0.50–2.0 *M*	(6, 7)
KCl, NaCl	0.02–1.0 *M*	(4, 8)
CH_3COONa	0.10–1.0 *M*	(4)
NaSCN	0.02–0.4 *M*	(6, 8, 9)
$CaCl_2$, $CaSO_4$	1–20 m*M*	(10)
$MgCl_2$	0.10–1.0 *M*	(4, 6, 11)
Osmolytes		
Polyols and sugars		
Glycerol	5–40%	(3, 6, 12)
Sucrose	0.1–1.0 *M*	(13, 14)
Trehalose	10–40%	(15, 16, 17)
Glucose	1.0–3.0 *M*	(6, 18)
Lactose	0.4–1.0 *M*	(6, 18)
Inositol	5–10%	(6, 19)
Amino acids and derivatives		
Glycine	0.1–2 *M*	(6, 13, 20, 21)
Alanine (α-, β-)	0.1–2 *M*	(6)
(Sodium) glutamate	0.50–1.5 *M*	(6, 22, 23)
(Potassium) aspartate	0.20–0.5 *M*	(6, 23, 24)
Arginyl glutamate	0.20–0.77 *M*	(6, 24)
Arginine hydrochloride	0.40–1.5 *M*	(6, 24)
Betaine	20–500 m*M*	(8, 25, 26)
Sarcosine	0.20–1.0 *M*	(26)
Trimethylamine *N*-oxide (TAMO)	0.25–1.0 *M*	(6, 27)
γ-Aminobutyric acid	20–500 m*M*	(8)
Organic molecules		
2-Methyl-2,4-pentanediol (MPD)	10–60%	(6, 28)
Other common additives		
Dithiothreitol (DTT)	1–10 m*M*	
Tris(2-carboxyethyl)phosphine, HCl (TCEP-HCl)	1–10 m*M*	
2-Mercaptoethanol (BME)	1–10 m*M*	
Bis(2-mercaptoethyl)sulfone (BMS)	1–10 m*M*	
Pefabloc	0.1–1 m*M*	
Benzamidine	0.1–1 m*M*	
EDTA	0.01–0.1 m*M*	
EGTA	0.01–0.1 m*M*	
NaN_3	50–100 μM	

[a] *Key to references*: (1) T. Arakawa and S. N. Timasheff, *Biochemistry* **21,** 6545 (1982); (2) O. Zhang and J. D. Forman-Kay, *Biochemistry* **34,** 6784 (1995); (3) D. Liu, R. Ishima,

have been routinely used for NMR studies of proteins, including ones that are aprotic or commercially available in deuterated form (for example, potassium and sodium phosphates, sodium acetate, and Tris), covering a fairly broad pH range (e.g., from 4.0 to 9.0). Attempts may then be made to improve solubility by screening common additives such as salt that may stabilize a protein, followed by screening of more exotic additives. This progression is represented in Fig. 1. Before describing the button test and microdrop screen, we outline below the characteristics of some of the more common additives.[27,28] Some of these additives together with relevant references are listed in Tables I and II.

Protein Structure Stabilizers and Destabilizers

Certain compounds affect the stability of protein structures.[28–30] Some compounds stabilize protein structure independent of their own concentration and the pH of the solution, others destabilize protein structure, and a third class of

[27] C. H. Schein, *Biotechnology* **8,** 308 (1990).
[28] S. N. Timasheff, *Adv. Protein Chem.* **51,** 355 (1998).
[29] S. N. Timasheff, *Annu. Rev. Biophys. Biomol. Struct.* **22,** 67 (1993).
[30] S. N. Timasheff, *in* "Methods in Molecular Biology, Vol 40: Protein Stability and Folding: Theory and Practice" (B. A. Shirley, ed.), Ch. 11, Humana Press Inc., Totowa, NJ, 1995.

K. I. Tong, S. Bagby, T. Kokubo, D. R. Muhandiram, L. E. Kay, Y. Nakatani, and M. Ikura M, *Cell* **94,** 573 (1998); (4) T. Arakawa and S. N. Timasheff, *Methods Enzymol.* **114,** 49 (1985); (5) T. Arakawa, R. Bhat, and S. N. Timasheff, *Biochemistry* **29,** 1924 (1990); (6) S. N. Timasheff, *Adv. Protein Chem.* **51,** 355 (1998); (7) T. Arakawa and S. N. Timasheff, *Biochemistry* **23,** 5924 (1984); (8) C. H. Schein, *Biotechnology* **8,** 308 (1990); (9) P. H. von Hippel and T. Schleich, *in* "Structure and Stability of Biological Macromolecules" (S. N. Timasheff and G. D. Fasman, eds.), p. 417. Dekker, New York, 1969; (10) J. J. Robinson, *Biochem. J.* **256,** 225 (1988); (11) T. Arakawa, R. Bhat and S. N. Timasheff, *Biochemistry* **29,** 1914 (1990); (12) K. Gekko and S. N. Timasheff, *Biochemistry* **20,** 4667 (1981); (13) S. N. Timasheff, *in* "Methods in Molecular Biology, Vol. 40: Protein Stability and Folding: Theory and Practice" (B. A. Shirley, ed.), p. 253. Humana Press Inc., Totowa, NJ, 1995; (14) A. Wang, A. D. Robertson, and D. W. Bolen, *Biochemistry* **34,** 15096 (1995); (15) T. Hottiger, C. De Virgilio, M. N. Hall, T. Boller, and A. Wiemken, *Eur. J. Biochem.* **219,** 187 (1994); (16) M. A. Singer and S. Lindquist, *Trends Biotech.* **16,** 460 (1998); (17) M. A. Singer and S. Lindquist, *Mol. Cell.* **1,** 639 (1998); (18) T. Arakawa and S. N. Timasheff, *Biochemistry* **21,** 6536 (1982); (19) K. Gekko and T. Morikawa, *J. Biochem. (Tokyo)* **90,** 51 (1981); (20) S. J. Matthews and R. J. Leatherbarrow, *J. Biomol. NMR* **3,** 597 (1993); (21) R. L. Foord and R. J. Leatherbarrow, *Biochemistry* **37,** 2969 (1998); (22) L. Wilson, *Biochemistry* **9,** 4999 (1970); (23) T. Arakawa and S. N. Timasheff, *J. Biol. Chem.* **259,** 4979 (1984); (24) Y. Kita, T. Arakawa, T. Y. Lin, and S. N. Timasheff, *Biochemistry* **33,** 15178 (1994); (25) T. Arakawa and S. N. Timasheff, *Arch. Biochem. Biophys.* **224,** 169 (1983); (26) T. Arakawa and S. N. Timasheff, *Biophys. J.* **47,** 411 (1985); (27) T. Y. Lin and S. N. Timasheff, *Biochemistry* **33,** 12695 (1994); (28) E. P. Pittz and S. N. Timasheff, *Biochemistry* **17,** 615 (1978).

TABLE II
POTENTIALLY USEFUL DETERGENTS

Detergents	CMC[a](m*M*)	References[b]
Nonionic		
Dodecyl-β-D-maltoside	0.1–0.6(0.12)	(1)
Decyl-β-D-maltoside	1.6(1.5)	(1)
Octyl-β-D-thioglucopyranoside[c]	9(4)	(1)
Heptyl-β-D-thioglucopyranoside[c]	30(23)	(1)
Decyl-β-D-glucopyranoside[c]	2–3(2)	(1)
Nonyl-β-D-glucopyranoside[c]	6.5(6.2)	(1)
Octyl-β-D-glucopyranoside[c]	20–25(19–25)	(1, 2, 3)
Heptyl-β-D-glucopyranoside[c]	79(65)	(1)
Hexyl-β-D-glucopyranoside[c]	250(195)	(1)
MEGA-8 (octanoyl-*N*-methylglucamide)	58(48)	(1)
MEGA-9 (nonanoyl-*N*-methylglucamide)	19–25 (18)	(1)
MEGA-10 (decanoyl-*N*-methylglucamide)	6–7(6.2)	(1)
Triton X-100	0.2–0.9(0.29)	(1)
Tween 20	0.059	(1)
Nonidet P-40	0.05–0.3(0.11)	(1)
Ionic		
Deoxycholic acid (sodium salts)	2–6(1–4)	(1)
CTAB (Celyltrimethylammonium bromide)	1	(1)
Zwitterionic		
CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propane sulfonate)	6–10(3–5)	(1, 4)
LDAO (lauryldimethylamine oxide)	1–2(0.14)	(1, 5)
ZWITTERGENT 3–08 (*N*-octylsulfobetaine)	330(140)	(1)
ZWITTERGENT 3–10 (*N*-decylsulfobetaine)	25–40(14)	(1)
ZWITTERGENT 3–12 (*N*-dodecylsulfobetaine)	2–4(1.4)	(1)
ZWITTERGENT 3–14 (*N*-tetradecylsulfobetaine)	0.1–0.4(0.14)	(1)
ZWITTERGENT 3–16 (*N*-hexadecylsulfobetaine)	0.01–0.06(0.014)	(1)

[a] Temperature, 20–25°; 0–0.05 *M* Na^+ (CMC values in the presence of 0.1–0.2 *M* Na^+ are given in parentheses).

[b] Key to references: (1) J. Neugebauer, "A Guide to the Properties and Uses of Detergents in Biology and Biochemistry," Clarkson University, Potsdam, NY, 1988; (2) L. Stryer, *J. Biol. Chem.* **266,** 10711 (1991); (3) J. B. Ames, A. M. Dizhoor, M. Ikura, K. Palczewski, and L. Stryer, *J. Biol. Chem.* **274,** 19329 (1999); (4) J. Anglister, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax, *J. Biomol. NMR* **3,** 121 (1993); (5) J. Ames, unpublished data (2000).

[c] Glucopyranoside and glucoside are chemical synonyms.

compounds can be stabilizers or destabilizers depending on their concentration and the solution pH. The stabilizers include most osmolytes (see below) and some salting out salts, the destabilizers are the strong denaturants such as urea and guanidine hydrochloride, and the third class consists of weakly acting agents such as $MgCl_2$, some amino acid salts, and dimethyl sulfoxide (DMSO).[28,30] Protein structure destabilizers and stabilizers operate by unrelated mechanisms: denaturants

bind directly to proteins and interact favorably with the unfolded state, whereas stabilizers are preferentially excluded from the protein surface. Hence, the protein structure stabilizers are sometimes referred to as preferentially excluded agents and in their presence proteins are said to be preferentially hydrated. Stabilizers and denaturants can be considered as a single class of compounds that form a continuum from strong stabilizers to strong denaturants. A similar continuum exists for the effect of compounds on protein solubility. In both cases, the action of a compound on stability or solubility depends on the balance between the affinities of the protein for water and the compound.[29]

Ionic Compounds and Salts. It is advisable to use salts at low concentrations initially, for several reasons: most bacterial and mammalian enzymes function under low salt conditions and are inhibited by high salt; at high concentrations, salts can lead to precipitation by competing for water molecules such that the hydration shell around the protein cannot be maintained (salting out—commonly used in protein isolation and in protein crystallization).[31,32] At low salt concentrations, Debye–Hückel screening can lead to an increase in protein solubility (salting in). An NMR-specific reason for using a low salt concentration in initial solubility screens is that salt concentrations significantly above 100 m*M* lead to longer pulse widths and sample heating.

It is not usually possible to predict the effect of a particular salt on protein solubility because of the variety of polar groups on a protein surface and the variation of their distribution. In general, large ions such as citrate, sulfates, acetate, and phosphates are better at stabilizing proteins than small ions such as chloride and nitrates. Hence, $MgSO_4$ is a good structure stabilizer whereas $MgCl_2$ is generally regarded as a salting-in agent and structure destabilizer but shows complex variation in behavior with its own concentration.[28] In aqueous solution conditions close to physiological, the isolated N-terminal SH3 domain of the *Drosophila* protein drk existed in slow exchange on the NMR time scale between folded and unfolded states such that the ^{1}H–^{15}N HSQC spectrum comprised two sets of peaks of roughly equal intensity.[33,34] Na_2SO_4 stabilized the folded form of the SH3 domain: addition of Na_2SO_4 caused the progressive disappearance of the resonances of the unfolded state. In the presence of 0.4 *M* Na_2SO_4, the SH3 domain gave rise to one set of peaks in the ^{1}H–^{15}N HSQC spectrum with dispersion characteristic of a folded polypeptide.[34]

Small changes in salt concentration can have a dramatic effect on protein solubility, so it is advisable to employ small increments in salt concentration

[31] A. McPherson, *Eur. J. Biochem.* **189,** 1 (1990).

[32] A. Ducruix and R. Giegé, "Crystallization of Nucleic Acids and Proteins: A Practical Approach." Oxford University Press, New York, 1992.

[33] O. Zhang, L. E. Kay, J. P. Olivier, and J. D. Forman-Kay, *J. Biomol. NMR* **4,** 845 (1994).

[34] O. Zhang and J. D. Forman-Kay, *Biochemistry* **34,** 6784 (1995).

such as 25 m*M* (for example, a change of 50 m*M* KCl concentration produced a 20-fold change in solubility of T7 RNA polymerase[27]). Divalent cations can exert a powerful effect on protein solubility. Minute levels of metals such as Cu, Zn, and Mn can induce aggregation. In general, $CaCl_2$ is seen as a good protein solubilizer.

Detergents. In cases where protein aggregation is presumed to arise from hydrophobic interactions, one option for disruption of these interactions is a nondenaturing detergent (Table II).[35,36] In general, nonionic detergents are useful for breaking lipid–lipid and lipid–protein interactions. Ionic detergents are better suited to breaking protein–protein interactions. Zwitterionic detergents are useful for breaking protein–protein interactions, and they do not bind to ion-exchange resins. It has been recommended that detergent : protein (w : w) ratios covering the range 10 : 1 to 0.1 : 1 be used for initial solubility trials.[35] An alternative guideline is that 6–11 micelles for each protein molecule may be used.[37] It is necessary to calculate the concentration of micelles in a detergent solution in order to prepare a solution that provides a certain number of micelles per protein molecule. The formula[36] for micelle concentration in moles per liter is:

$$[\text{micelles}] = ([C_B] - \text{CMC})/N$$

where $[C_B]$ is the bulk molar concentration of detergent, N is the mean aggregation number, and CMC is the critical micelle concentration in moles per liter. CMC values for micelle formation can be found in Table II and in Neugebauer.[36]

The calcium-stimulated phosphatase calcineurin exhibited backbone amide proton T_2 values that were 3- to 4-fold shorter than expected for a protein in the 16–20 kDa size range, indicating that calcineurin is aggregated. Variation of temperature, pH, and salt concentration produced small changes in T_2, but the protein could not be made to behave as a monomer. Addition of 10 m*M* CHAPS (Table II), a nondenaturing detergent, increased the average amide T_2 from 8.3 to 17.5 ms and produced a dramatic increase in sensitivity and resolution.[38] In a similar case, addition of 25 m*M* [$^2H_{28}$]octyl-β-glucoside appeared to prevent dimerization of guanylyl cyclase activating protein-2 (GCAP-2) and consequently resulted in significantly sharper NMR signals.[39] The original purification of GCAP-2 from the retina indicated that octyl-β-glucoside does not appear to denature or inactivate GCAP.[40]

Osmolytes. All organisms except halobacteria have evolved a response to denaturing stresses such as high temperature, desiccation, the presence of denaturants,

[35] L. M. Hjelmeland and A. Chrambach, *Methods Enzymol.* **104,** 305 (1984).

[36] J. Neugebauer, "A Guide to the Properties and Uses of Detergents in Biology and Biochemistry." Clarkson University, Potsdam, NY, 1988.

[37] J. A. Reynolds, *Receptors and Recognition, Series B* **11,** 34 (1981).

[38] J. Anglister, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax, *J. Biomol. NMR* **3,** 121 (1993).

[39] J. B. Ames, A. M. Dizhoor, M. Ikura, K. Palczewski, and L. Stryer, *J. Biol. Chem.* **274,** 19329 (1999).

[40] L. Stryer, *J. Biol. Chem.* **266,** 10711 (1991).

and high osmotic pressure that involves intracellular production and accumulation of high levels of low molecular weight organic compounds called osmolytes.[29,41] These can be polyols such as glycerol, sugars (sucrose, trehalose), polysaccharides, neutral polymers, amino acids and their derivatives, and large dipolar molecules such as trimethylamine *N*-oxide (TMAO).[27] Osmolytes stabilize macromolecules and so conserve biological activity. The stabilizing effect is exerted indirectly: osmolytes do not interact with the biomacromolecule directly but alter the solvent properties of the surrounding water and therefore affect macromolecule–solvent interactions. One proposal is that stabilization arises from preferential hydration of the protein, whereby protein molecules are surrounded by water with exclusion of additive molecules from the protein's vicinity. The native protein conformation is favored because unfolded protein is less compact and leads to further additive exclusion, which is thermodynamically unfavorable.[42] Many osmolytes can stabilize enzymes with little apparent change in structure and function[43,44] even when present at high concentrations and hence are known as "compatible solutes." For the purposes of protein solubility screening, it should be noted that these stabilizers can reduce protein solubility and promote salting out, although the fact that NMR studies of lysozyme, chymotrypsin inhibitor 2, and horse heart cytochrome *c* could be performed in the presence of 2 *M* glycine (see below) suggests that it is worthwhile including osmolytes in the later stages of solution condition screening.

Sugars. Preferential interaction studies have been carried out on sucrose, trehalose, lactose, and glucose. The first three showed total preferential exclusion, i.e., they did not interact with the protein surface except at thermodynamically indifferent loci.[28] In contrast, glucose gradually formed contacts with the protein surface as the glucose concentration was increased. Of the sugar osmolytes, trehalose has perhaps been of most interest recently.[45] Trehalose is produced by a wide variety of organisms and is best known for its role in protecting certain organisms from desiccation. Work in yeast has indicated that trehalose also promotes survival under conditions of extreme heat by efficiently protecting enzymes against heat inactivation[46] and by suppressing the aggregation of denatured proteins.[47]

Polyols. In a study of the preferential interactions of bovine serum albumin (BSA) with several polyols (ethylene glycol, glycerol, xylitol, mannitol, sorbitol,

[41] P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus, and G. N. Somero, *Science* **217,** 1214 (1982).

[42] S. N. Timasheff, *in* "Stability of Protein Pharmaceuticals, Part B: In Vivo Pathways for Degradation and Strategies for Protein Stabilization" (T. J. Ahern and M. C. Manning, eds.), p. 265. Plenum Press, New York, 1992.

[43] M. M. Santoro, Y. Liu, S. M. Khan, L. X. Hou, and D. W. Bolen, *Biochemistry* **31,** 5278 (1992).

[44] G. N. Somero and P. H. Yancey, *in* "Handbook of Physiology, Section 14: Cell Physiology" (J. F. Hoffman and J. D. Jamieson, eds.), p. 441. Oxford University Press, New York, 1997.

[45] M. A. Singer and S. Lindquist, *Trends Biotech.* **16,** 460 (1998).

[46] T. Hottiger, C. De Virgilio, M. N. Hall, T. Boller, and A. Wiemken, *Eur. J. Biochem.* **219,** 187 (1994).

[47] M. A. Singer and S. Lindquist, *Mol. Cell.* **1,** 639 (1998).

and inositol), all except inositol gave low preferential hydration.[28] The strong preferential hydration displayed by inositol was ascribed to its strongly hydrophilic character and its high degree of hydration.

Amino acids, amino acid salts, and methylamines. Small neutral amino acids such as glycine, α-alanine, and β-alanine exhibit a concentration-independent degree of preferential hydration and therefore belong to the first class of protein structure stabilizers.[28–30] Of the amino acid salts, sodium glutamate (NaGlu) and potassium aspartate (KAsp) are strongly preferentially excluded. Lysine hydrochloride displayed opposite behavior toward BSA and lysozyme compared with these anionic amino acids: preferential hydration in the NaGlu–BSA and KAsp–BSA systems was greater than with positively charged lysozyme, whereas preferential hydration was greater in the LysHCl–lysozyme system than in LysHCl–BSA.[28]

Arginyl glutamate provides an interesting example of compensation between binding (destabilizing) and exclusion (stabilizing). Arg^+ should interact favorably with amide and peptide groups whereas glutamate is a protein structure stabilizing agent. The strong preferential exclusion of glutamate from the protein surface compensates for the binding tendency of Arg^+, giving a concentration-independent preferential hydration with values smaller than those observed with NaGlu.[28]

Methylamines such as sarcosine, betaine, and TMAO stabilize protein structure. Betaine induced strong preferential hydration with a small concentration dependence[48] and sarcosine was strongly excluded from lysozyme.[49] In renal medullas, it is thought that methylamines stabilize macromolecules to counteract the effects of high levels of urea.[50] Betaine may act as a thermoprotectant in *Escherichia coli.*[51]

Gopal and Ahluwalia[52] have used differential scanning calorimetry to compare the stabilizing effects of methylamine and amino acid osmolytes. These compounds were found to increase the thermal stability of ribonuclease A and lysozyme in the order: trimethylamine *N*-oxide > glycine > β-alanine > γ-aminobutyric acid > sarcosine > serine > α-alanine > betaine > proline.

Osmolytes and NMR. Since concentrations up to several molar may be required, osmolytes can introduce intense NMR signals and lead to sample heating. The presence of 2 *M* glycine (perdeuterated form) allowed acquisition of a 1D NMR spectrum of lysozyme characteristic of fully folded protein at 348 K.[53] The 1D NMR spectrum of lysozyme at 348 K in the absence of added glycine indicated that the protein was at least partially denatured. Glycine stabilization of

[48] T. Arakawa and S. N. Timasheff, *Arch. Biochem. Biophys.* **224,** 169 (1983).
[49] T. Arakawa and S. N. Timasheff, *Biophys J.* **47,** 411 (1985).
[50] M. B. Burg and E. M. Peters, *Am. J. Physiol.* **274,** F762 (1998).
[51] T. Caldas, N. Demont-Caulet, A. Ghazi, and G. Richarme, *Microbiol.* **145,** 2543 (1999).
[52] S. Gopal and J. C. Ahluwalia, *J. Chem. Soc. Faraday Trans.* **89,** 2769 (1993).
[53] S. J. Matthews and R. J. Leatherbarrow, *J. Biomol. NMR* **3,** 597 (1993).

chymotrypsin inhibitor 2 and horse heart cytochrome *c* has also been investigated,[54] through analysis of the effect of glycine on hydrogen exchange rates that provide information on behavior at discrete locations throughout the protein structure. The addition of 2 *M* glycine significantly stabilized both proteins and reduced the exchange rates of most slowly exchanging protons. The effect of glycine on exchange rates varied across almost three orders of magnitude for different protons, with no apparent correlation with primary structure, amino acid type, static solvent accessibility, or temperature factors from X-ray crystal structures. Similar conclusions were drawn from examination of the effects of 1 *M* sucrose on hydrogen exchange rates of ribonuclease A[55]: slowly exchanging amide protons exchanged more slowly in the presence of sucrose, but the exchange rates of intermediate-exchanging protons were not affected. The authors concluded that fast and intermediate exchange occurs mainly from the native state ensemble of the protein, which is not significantly affected by osmolytes. In contrast, slow amide proton exchange occurs mainly from the compact unfolded state ensemble of the protein. Slow exchange requires exposure of large protein surface areas, as occurs in unfolding. Sucrose opposes this exposure of surface area and so reduces the rate of slow exchange and also affects protein stability.[55]

Miscellaneous Common Additives for Protein NMR Samples. It is general practice to add some or all of the following to protein NMR samples: reductant such as dithiothreitol (DTT) to protect free sulfhydryls from oxidation and prevent intermolecular sulfhydryl cross-linking; protease inhibitors such as Pefabloc, a serine protease inhibitor, and EDTA, a metalloprotease inhibitor; EDTA or EGTA to chelate divalent metal ions which may otherwise react with proteins; and sodium azide, a bacteriocide.

Microdialysis Button Test

Microdialysis buttons are machined from transparent Perspex and are the size of a small button (hence their name). The buttons have a convex top surface in the middle of which is situated a well for the protein solution (Fig. 2). Dialysis buttons can be obtained with a range of well volumes (5–350 μl); we typically use buttons with a 5 μl well. Once the protein sample has been placed in this well, the button is covered with a piece of dialysis membrane which is held in place with a rubber O ring. The dialysis membrane permits passage of small molecules such as ions, buffer materials, and additives but prevents passage of molecules larger than the molecular weight cutoff of the membrane. Microdialysis buttons can be obtained from Cambridge Repetition Engineers Ltd., Greens Road, Cambridge CB4 3EQ, UK, or from Hampton Research (www.hamptonresearch.com).

[54] R. L. Foord and R. J. Leatherbarrow, *Biochemistry* **37,** 2969 (1998).
[55] A. Wang, A. D. Robertson, and D. W. Bolen, *Biochemistry* **34,** 15096 (1995).

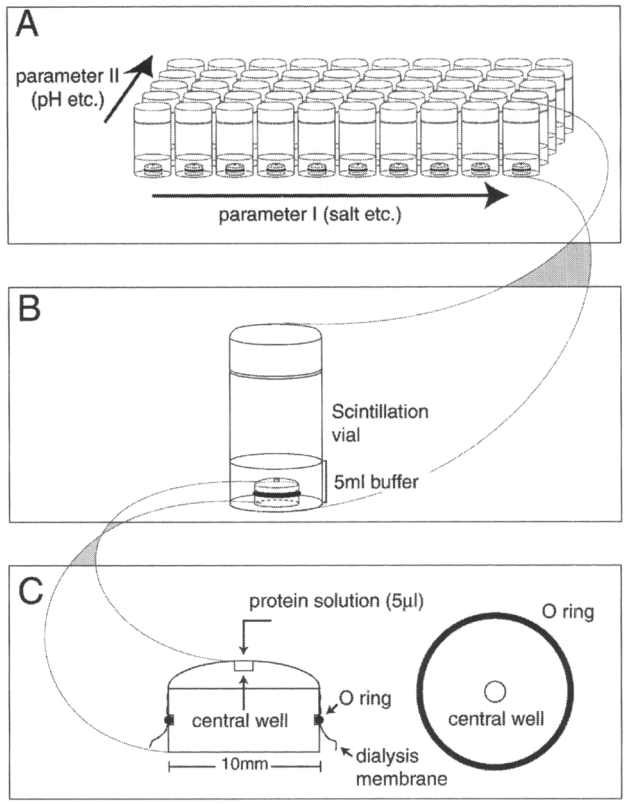

FIG. 2. A typical setup for the microdialysis button test. (A) An array of microdialysis buttons or cells contained within scintillation vials. A solution parameter such as salt concentration might be varied along one dimension of the array and another parameter such as pH might be varied along the other dimension, in this case producing 50 different solution conditions. (B) One of the scintillation vials from the array, showing the volume of solution required to submerge the microdialysis cell. (C) A microdialysis cell shown in vertical cross section (left) and viewed from directly above (right). The various parts of the microdialysis cell are labeled, and its approximate diameter is indicated.

In a screen of conditions for NMR studies of TFIIB core domain,[25] the protein is first exchanged into distilled deionized water containing 7.5 m*M* dithiothreitol and then concentrated to 1 m*M* (a typical protein concentration for high resolution NMR studies). A lower initial concentration of protein might be used to screen conditions for proteins with particularly low solubility. The protocol for setting up a microdialysis button sample is as follows:

1. A piece of standard dialysis membrane, for example Spectra/Por molecular porous membrane from Spectrum (www.spectrapor.com), is prepared according

to the manufacturer's instructions and then cut into squares of roughly 2.5 cm. The pieces of dialysis membrane can be kept moist and ready for use by placing them between wet Kimwipes.

2. The protein solution is pipetted into the central well of a microdialysis button, employing a circular motion with the pipette as the solution is pipetted in order to avoid air bubble formation. Narrow tips, such as those designed for SDS–PAGE gel loading, can be used.

3. Once the protein solution has been transferred to the well, a piece of dialysis membrane and rubber O-ring are applied over the top of the button using an applicator (again available from Cambridge Repetition Engineers). The O-ring fits in a groove running around the circumference of the microdialysis cell (Fig. 2). A smooth, quick action is required to apply the dialysis membrane without inducing air bubbles in the protein solution.

4. The dialysis button is then submerged membrane side up in a small volume of test solution. It is probably most convenient for subsequent microscope observation if each button is submerged within a Linbro plate (Hampton Research) reservoir and the reservoir sealed using a cover slide. If Linbro plates are unavailable, the buttons can be submerged in 5 ml solution contained within a scintillation vial (Fig. 2).

5. In cases of particularly low protein solubility or stability, the microdialysis buttons can be submerged in precooled test solutions and left in a 4° environment. This initial gentle handling may allow identification of promising conditions for further exploration at higher temperature. Otherwise the test samples are placed in a temperature-controlled environment at a temperature desirable for NMR experiments, for example 25° or even 37° in cases where the protein is known to be soluble and stable at the higher temperature. The effect of temperature on solubility might be investigated by placing microdialysis samples in the same test solution in environments at different temperatures.

6. The test samples can be monitored for precipitation both by naked eye and inspection using a standard dissecting microscope. As suggested,[26] a scale can be used for scoring the extent of precipitation in each test sample. In the screen for TFIIBc, the samples were checked for precipitation at least once a day over a period of 7 days, although the general trends of protein solubility (e.g., whether solubility is increased by high or low salt concentration) were apparent after 24–36 hr.

Microdrop Screen

Like the microdialysis method, microdrop screening has been employed previously to screen conditions for protein crystallization.[31,32] The microdrop screen for protein crystallization relies on the phenomenon of vapour diffusion whereby a drop of protein solution, typically combined with a precipitating agent, and the test solution are sealed within a chamber so that solvent is gradually drawn from the protein drop until its solution conditions change to those of the test solution. This

leads to supersaturation, protein precipitation, and ideally formation of protein crystals.

A group from Vertex Pharmaceuticals (Cambridge, MA) has adapted the hanging drop technique for vapor diffusion to screen solution conditions rapidly and efficiently for maximum protein solubility rather than protein precipitation.[26] The protein [in the reported example human recombinant glia maturation factor-β (GMF-β) with an N-terminal hexahistidine (His_6) tag] was exchanged into 10 m*M* potassium phosphate buffer. (Volumes, concentrations, and buffer types indicated in parentheses in the following are those used by Lepre and Moore[26] unless otherwise stated). The screen employs 24-well tissue culture (Linbro) plates and siliconized glass coverslips, available as kits designed for protein crystallization trials from Hampton Research (Web address above). The microdrop screen is set up as follows (Fig. 3):

1. Each test buffer solution (1 ml) is pipetted into a well in the tissue culture plate. The test buffer (1 μl of 100 m*M*) is added to the protein solution (2 μl) on each glass coverslip and mixed by carefully pipetting up and down.
2. Each coverslip is then inverted and sealed onto the appropriate well using vacuum grease.
3. The Linbro plates are placed at a particular temperature (room temperature) and the extent of precipitation is scored every 12–24 hr using a microscope to assess the fraction of the drop covered by precipitate. As in the microdialysis screen of TFIIBc solubility,[25] the pattern of relative stabilities for GMF-β was clearest after 24 hr.
4. Once optimum buffer conditions have been identified (i.e., the type of buffer and the pH which produce least protein precipitation), additives can be screened using those buffer conditions (sodium phosphate at pH 7.5 and HEPES at pH 7.0). For GMF-β, the effect of sodium chloride (25, 50, and 100 m*M*), 2-mercaptoethanol, glycerol, and CHAPS on solubility was tested, in addition to varying the concentration of sodium phosphate. Of these, only the addition of the nonionic detergent CHAPS to 2 m*M* resulted in improved solubility of GMF-β.

In general, two approaches can be followed in the microdrop screen, the concentration method (increasing the concentration of protein in the drop by diffusing solvent out of the drop) or the dilution method (reducing protein concentration by diffusing solvent into the drop). Selection between these two is made simply by using a lower or higher concentration of solute in the drop than in the buffer reservoir. In the concentration method, the final protein concentration in the drop can be increased approximately fourfold by reducing the amount of reservoir buffer added to the drop when setting up the screen. Alternatively, the protein concentration in the drop can be incremented by larger factors *in situ* by progressively stepping up the concentration of buffer in the reservoir. If desired, different concentrations of

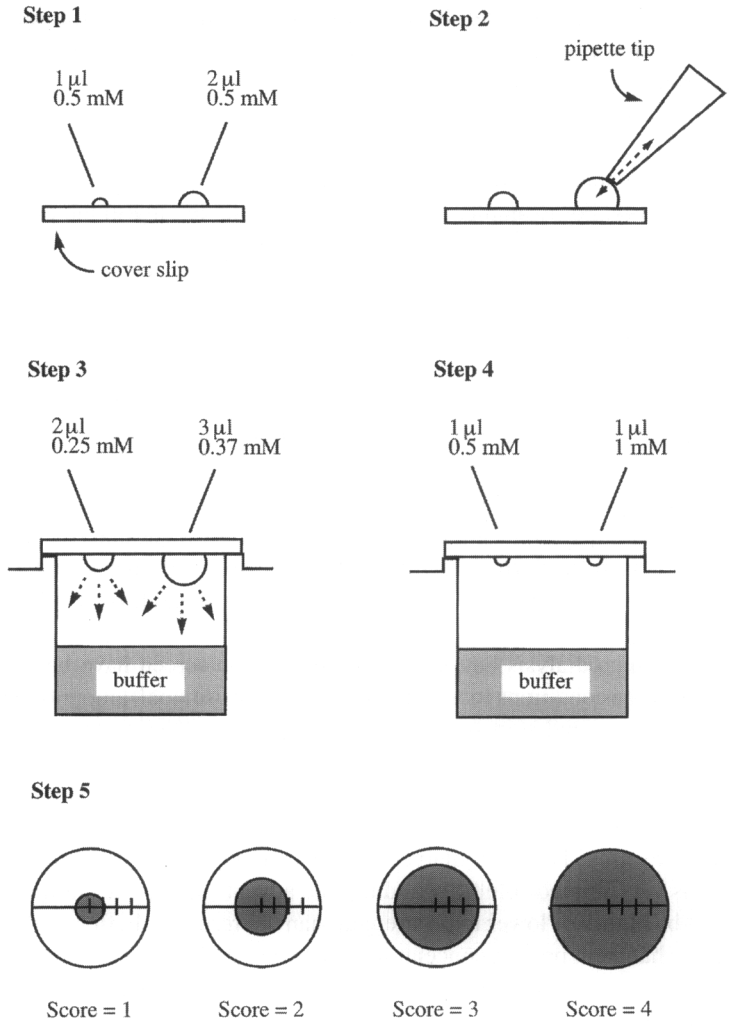

FIG. 3. Procedure for setting up a microdrop screen. In step 1, the protein solution in starting buffer is pipetted onto a coverslip for a 24-well Linbro plate. The drop volume and protein concentration are indicated. Using two drops per coverslip allows two protein concentrations to be tested in one well. In step 2, the protein solution is mixed with 1 μl of test solution from the well. The coverslip is then inverted and sealed over the well (step 3). The test samples are then left undisturbed to allow vapor equilibration (step 4) and finally the degree of precipitation in each drop is assessed by estimating the fraction of the drop covered by precipitate (step 5). The drop boundary is represented by the outer circle and precipitate by the inner circle. (Reproduced with permission from Lepre and Moore.[26])

protein may be tested per well by pipetting multiple drops onto the same coverslip, each with a different ratio of initial protein solution to reservoir buffer. This flexibility in protein concentration and the capacity to test multiple concentrations per well is an advantage of the microdrop screen over the microdialysis method.

The dilution approach to the microdrop screen is useful for proteins with very limited solubility. For example, a protein may be soluble only in the presence of high levels of a stabilizing compound or if unfolded by a chaotrope. The microdrop screen would then be set up by concentrating the protein in the presence of stabilizer or chaotrope and mixing the protein solution with a solution containing a different stabilizer or combination of stabilizers. The protein drop is then equilibrated with a dilute reservoir such that solvent diffuses into the drop and gradually reduces the concentration of the original stabilizer or chaotrope to a level below which the protein was previously insoluble. Only those drops containing an additive that stabilizes the protein will then remain clear. A drawback of this approach is that solvent diffusion into the drop reduces the protein concentration at the same time as reducing the stabilizer concentration. In this respect, the microdialysis button test may be more useful for proteins with very limited solubility because the protein concentration remains constant as the concentration of additive in the protein solution increases.

Both the microdialysis button test and microdrop screen have been used to optimize solution conditions for several proteins. Both use small amounts of protein (well under 5 mg) and start from a uniform condition. These screens therefore avoid the need to prepare several batches of protein, for example for making NMR samples with different solvent conditions. The microdialysis button test and microdrop screen represent a significant advance over trial-and-error methods of condition screening. The microdialysis button is more suitable when the protein of interest has very limited solubility and has the advantage that the starting protein buffer can be higher or lower ionic strength than the reservoir buffer, both with no significant change in protein concentration between initial and equilibrium states. The microdrop screen has more flexibility in protein concentration and capacity to test multiple concentrations per well.

Polypeptides: Not Folded; Folded, but Not Soluble

In the case where its CD or ^{1}H–^{15}N HSQC spectrum indicates that a protein or protein fragment is not folded, or if it appears to be folded but solution conditions cannot be found in which it is soluble over the long time periods required for NMR data acquisition, it may be necessary to express an alternative fragment that is trimmed or extended at the N or C terminus, to carry out some posttranslational modification or site-directed mutagenesis, or to mix or refold with a binding partner (e.g., peptide, polypeptide, or nucleic acid oligomer), cofactor, or metal ion.

Wild-type human Mad2 spindle assembly checkpoint protein formed oligomers, but when the N-terminal 10 amino acid residues were omitted from the construct, most (about 75%) of the truncated and still active Mad2 was monomeric.[56] Truncation of Mad2 permitted detailed NMR study.[57] Even single amino acid mutations can affect protein stability and solubility: the aim of such mutations is obviously to manipulate particular unfavorable interactions to make them favorable or at least neutral. In this section we discuss possible mutation, chemical modification,[58] and evolutionary approaches to protein stabilization.

Cysteine is a common target for mutation, particularly if the cysteine is predicted to be surface exposed and is therefore potentially capable of forming intermolecular disulfide bonds. Histidine ammonia-lyase, for example, gave rise to undefined aggregates in the absence of reducing agents. Even in the presence of reducing agents that reverse aggregation, crystals suitable for structure determination could not be produced. Of the seven cysteines, one was predicted to be solvent-exposed. Mutation of this surface cysteine to alanine resulted in monodisperse solutions of histidine ammonia-lyase that yielded high-quality crystals.[59]

Insights into the potential effects on stability of other amino acid mutations have come from a detailed investigation of the effects of various substitutions on the stability of T4 lysozyme.[60] Mutations that sought to introduce a salt bridge in four different locations or to reduce repulsion between like charges (T4 lysozyme has a net positive charge of 9 at neutral pH) had little effect on T4 lysozyme stability. A more successful mutation strategy involved introduction of a negative charge close to the N terminus of an α helix, with the aim of complementing the partial positive charge due to the helix dipole. This appeared to increase stability on a consistent basis. These results indicated that mutations seeking to manipulate electrostatic interactions are best targeted toward rigid portions of the protein such as α helices. Probably more important for protein stability is tight packing of the hydrophobic core; it has been noted that mutations that introduce cavities destabilize proteins.

Other clues as to potential routes to protein stabilization and, ideally, greater solubility arise from analysis of the structures of proteins from thermophilic organisms. These structures show small, subtle variations from their mesophilic counterparts: for example, thermophile protein structures tend to contain smaller loops; thermophile enzymes have additional salt bridges; thermophile enzymes have a much lower content of asparagine and glutamine, perhaps because these amino acids are prone to destructive deamidation; analyses of thermophilic *Bacillus*

[56] G. Fang, H. Yu, and M. W. Kirschner, *Genes Dev.* **12,** 1871 (1998).

[57] X. Luo, G. Fang, M. Coldiron, Y. Lin, H. Yu, M. W. Kirschner, and G. Wagner, *Nat. Struct. Biol.* **7,** 224 (2000).

[58] C. Ó Fágáin, *Biochim. Biophys. Acta* **1252,** 1 (1995).

[59] T. F. Schwede, M. Bädeker, M. Langer, J. Rétey, and G. E. Schulz, *Prot. Eng.* **12,** 151 (1999).

[60] B. W. Matthews, *Annu. Rev. Biochem.* **62,** 139 (1993).

lactate dehydrogenases (LDH) indicated that phenylalanine and valine tend to replace mesophilic isoleucine and leucine[61]; proline content increases linearly with increasing thermostability in *Bacillus* oligo-1,6-glucosidases[62] (the proline five-membered ring is rigid and reduces flexibility of the polypeptide chain, reducing the tendency to unfold); arginine tends to replace lysine in thermophilic proteins, probably because of its higher pK_a (12 vs 9.5 and therefore arginine remains charged in more alkaline conditions and at higher temperatures), larger positively charged surface available for stabilizing interactions, and shorter side-chain hydrophobic component (three $-CH_2$ moieties rather than four $-CH_2$) to reduce unfavorable interactions with solvent.

Another possible approach to improving protein stability and solubility involves chemical modification of surface residues in order to make the protein more hydrophilic and improve its interaction with solvent. Conversion of surface tyrosines of trypsin and chymotrypsin to aminotyrosines, for example, increased stability; trypsin with four modified tyrosines was more than 100-fold more resistant to heat inactivation than wild-type trypsin.[63]

In view of the fact that the principles of protein stability remain relatively poorly understood, evolutionary approaches to protein stabilization offer an intriguing alternative to site-directed mutagenesis and chemical modification. One approach, termed Proside ("protein stability increased by directed evolution") uses a rapid method for the selection of stabilized variants of a protein.[64] Proside depends on the relationship between the stability of a protein and its resistance to proteolysis. Variants of the protein of interest are inserted between the domains of a phage protein, the phage is subsequently subjected to *in vitro* proteolysis, and the phage infectivity is then lost if the engineered phage protein has been cleaved as a result of an unstable insert. The phage containing the most stable variants of the protein under investigation are enriched through rounds of proteolysis, infection, and propagation.

Proteins that bind a ligand such as a metal ion, a small organic compound, a peptide, or a polypeptide are generally stabilized by binding to that ligand. Ca^{2+}-binding proteins such as calmodulin and troponin C, for example, are greatly stabilized by Ca^{2+} binding. TATA binding protein (TBP) precipitates readily in the absence of a binding partner but produced stable NMR samples when bound to an inhibitory domain from a TBP-associated factor.[65] In our studies of the 1,4,5-trisphosphate (IP_3) receptor, NMR spectra of a 45 kDa domain responsible for IP_3 binding were dramatically improved upon addition of the ligand.[66]

[61] H. Zuber, *Biophys. Chem.* **29,** 171 (1988).
[62] Y. Suzuki, *Proc. Jpn. Acad. Sci.* **65 ser B,** 146 (1990).
[63] V. V. Mozhaev, V. A. Siksnis, N. S. Melik-Nubarov, N. Z. Galkantaite, G. J. Denis, E. P. Butkus, B. Yu. Zaslavsky, N. M. Mestechkina, and K. Martinek, *Eur. J. Biochem.* **173,** 147 (1988).
[64] V. Sieber, A. Pluckthun, and F. X. Schmid, *Nature Biotech.* **16,** 955 (1998).
[65] D. Liu, R. Ishima, K. I. Tong, S. Bagby, T. Kokubo, D. R. Muhandiram, L. E. Kay, Y. Nakatani, and M. Ikura, *Cell* **94,** 573 (1998).
[66] I. Bosanac, unpublished data (2000).

Perspective

A major limitation of the microdialysis and microdrop methods described above for efficient screening of a wide range of solution conditions is that they can be used to assess only the macroscopic property of whether or not a protein precipitates. Separate analysis is needed to determine the aggregation state, thermodynamic stability, folding state, and activity of the protein. This issue was mentioned in the microdrop screening paper, with the proposal that where important further analyses require larger volumes of protein solution, vapor diffusion using the sitting drop method[31,32] would permit drops of several hundred microliters to be used.[26] It would be very useful to develop a high-throughput system that carries out numerous analytical procedures automatically on samples used for solution condition screening contained in a Linbro or 96-well microtiter plate. Such an analytical system would indeed appeal to anyone who has wrestled with the problems of protein solubility and stability and might have application in the numerous ongoing structural genomics projects.[4]

[3] Segmental Isotopic Labeling Using Expressed Protein Ligation

By DAVID COWBURN and TOM W. MUIR

Recent advances in using nuclear magnetic resonance (NMR) for structural investigation of proteins have heavily depended on the incorporation of stable isotopes of C, N, and H to achieve: (1) increased occurrence of an NMR-active isotope, (2) dilution of an NMR-active isotope (usually H), and/or (3) stereospecific incorporation of isotope.[1,2] For biologically expressed proteins, these incorporations may be done uniformly, or by amino acid class. In the latter case, some amino acids can be directly incorporated bacterially when they are in favorable synthetic pathways; others require auxotrophs,[3] and others can be achieved by cell-free synthesis.[4] Total chemical synthesis of isotopically labeled proteins, although achievable,[5] is generally economically impractical because of the cost of precursors, but it is an attractive objective because of the ability to control completely

[1] G. M. Clore and A. M. Gronenborn, *Nat. Struct. Biol.* **4**(Suppl.), 849 (1997).
[2] G. Wider and K. Wüthrich, *Curr. Opin. Struct. Biol.* **9**, 594 (1999).
[3] M. Kainosho, *Nat. Struct. Biol.* **4**(Suppl.), 858 (1997).
[4] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, and S. Yokoyama, *FEBS Lett.* **442**, 15 (1999).
[5] T. W. Muir, P. E. Dawson, and S. B. H. Kent, *Methods Enzymol.* **289**, 266 (1997).

0076-6879/00 $35.00