J|A|C|S
ARTICLES
Published on Web 07/01/2004

A Simple Method for Improving Protein Solubility and Long-Term Stability

Alexander P. Golovanov,*,† Guillaume M. Hautbergue,‡ Stuart A. Wilson,‡ and Lu-Yun Lian*,†

Contibution from the Department of Biomolecular Sciences, University of Manchester Institute of Science and Technology (UMIST), P. O. Box 88, Manchester M60 1QD, U.K., and Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, U.K.

Received February 9, 2004; E-mail: lu-yun.lian@umist.ac.uk; a.golovanov@umist.ac.uk

Abstract: Increasing a protein concentration in solution to the required level, without causing aggregation and precipitation is often a challenging but important task, especially in the field of structural biology; as little as 20% of nonmembrane proteins have been found to be suitable candidates for structural studies predominantly due to poor protein solubility. We demonstrate here that simultaneous addition of charged amino acids L-Arg and L-Glu at 50 mM to the buffer can dramatically increase the maximum achievable concentration of soluble protein (up to 8.7 times). These amino acids are effective in preventing protein aggregation and precipitation, and they dramatically increase the long-term stability of the sample; additionally, they protect protein samples from proteolytic degradation. Specific protein−protein and protein−RNA interactions are not adversely affected by the presence of these amino acids. These additives are particularly suitable for situations where high protein concentration and long-term stability are required, including solution-state studies of isotopically labeled proteins by NMR.

Introduction

High protein concentration and long-term stability are universally required in a wide range of applications, from the preparation of liquids in pharmaceuticals[1] to general biochemical studies,[2] and are particularly necessary in the field of structural biology.[3−6] Preparation of a concentrated and stable protein sample is a prosaic but often difficult task, as proteins frequently aggregate or precipitate at higher concentrations, and are sometimes subject to spontaneous proteolytic degradation. The first published results from high-throughput structural genomics programs[4,7] identified poor protein solubility as one of the main bottlenecks. The estimates show[8−10] that ca. 33−50% of all expressed nonmembrane proteins are not soluble, and ca. 25−57% of remaining soluble proteins aggregate or precipitate during concentration. That means that simply because of the low solubility, a significant proportion of proteins cannot be studied at all using some of the most powerful and informative methods (such as those used for the structural studies) requiring high protein concentration. Solving the problem of protein solubility may increase up to two-fold the number of proteins amenable to structural studies, which in turn will proportionally increase the number of targets for such applications as rational drug design.

A number of strategies have been suggested to improve the yield of soluble proteins; they are generally aimed either at the modification of the protein itself (which is not always possible or desirable) or at optimization of the protocols for expression, purification, and solubilization.[3,5,11−13] In many cases, proteins can be produced as inclusion bodies and then denatured, purified, and refolded,[14,15] thus significantly improving the yield of soluble material.[2,6] However, even if the first step is successful and a pure protein is produced in soluble form, it is often

† UMIST.
‡ University of Sheffield.

(1) Wang, W. *Int. J. Pharm.* **1999**, *185*, 129−188.
(2) Bondos, S. E.; Bicknell, A. *Anal. Biochem.* **2003**, *316*, 223−231.
(3) Bagby, S.; Tong, K. I.; Ikura, M. *Methods Enzymol., Pt. B* **2001**, *339*, 20−41.
(4) Schmid, M. B. *Trends Microbiol.* **2002**, *10*, S27−S31.
(5) Pedelacq, J. D.; Piltch, E.; Liong, E. C.; Berendzen, J.; Kim, C. Y.; Rho, B. S.; Park, M. S.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2002**, *20*, 927−932.
(6) Maxwell, K. L.; Bona, D.; Liu, C. S.; Arrowsmith, C. H.; Edwards, A. M. *Protein Sci.* **2003**, *12*, 2073−2080.
(7) Montelione, G. T.; Zheng, D. Y.; Huang, Y. P. J.; Gunsalus, K. C.; Szyperski, T. *Nat. Struct. Biol.* **2000**, *7*, 982−985.
(8) Christendat, D.; Yee, A.; Dharamsi, A.; Kluger, Y.; Savchenko, A.; Cort, J. R.; Booth, V.; Mackereth, C. D.; Saridakis, V.; Ekiel, I.; Kozlov, G.; Maxwell, K. L.; Wu, N.; McIntosh, L. P.; Gehring, K.; Kennedy, M. A.; Davidson, A. R.; Pai, E. F.; Gerstein, M.; Edwards, A. M.; Arrowsmith, C. H. *Nat. Struct. Biol.* **2000**, *7*, 903−909.
(9) Yee, A.; Chang, X. Q.; Pineda-Lucena, A.; Wu, B.; Semesi, A.; Le, B.; Ramelot, T.; Lee, G. M.; Bhattacharyya, S.; Gutierrez, P.; Denisov, A.; Lee, C. H.; Cort, J. R.; Kozlov, G.; Liao, J.; Finak, G.; Chen, L.; Wishart, D.; Lee, W.; McIntosh, L. P.; Gehring, K.; Kennedy, M. A.; Edwards, A. M.; Arrowsmith, C. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1825−1830.
(10) Yee, A.; Pardee, K.; Christendat, D.; Savchenko, A.; Edwards, A. M.; Arrowsmith, C. H. *Acc. Chem. Res.* **2003**, *36*, 183−189.
(11) Zhou, P.; Lugovskoy, A. A.; Wagner, G. *J. Biomol. NMR* **2001**, *20*, 11−14.
(12) Edwards, A. M.; Arrowsmith, C. H.; Christendat, D.; Dharamsi, A.; Friesen, J. D.; Greenblatt, J. F.; Vedadi, M. *Nat. Struct. Biol.* **2000**, *7*, 970−972.
(13) Savchenko, A.; Yee, A.; Khachatryan, A.; Skarina, T.; Evdokimova, E.; Pavlova, M.; Semesi, A.; Northey, J.; Beasley, S.; Lan, N.; Das, R.; Gerstein, M.; Arrowmith, C. H.; Edwards, A. M. *Proteins* **2003**, *50*, 392−399.
(14) Misawa, S.; Kumagai, I. *Biopolymers* **1999**, *51*, 297−307.
(15) De Bernardez Clark, E. *Curr. Opin. Biotechnol.* **2001**, *12*, 202−207.

difficult to increase its concentration up to the necessary level (e.g., typically 0.5−1 mM for structural studies) without causing precipitation or aggregation. It is widely accepted that the solubility and stability of proteins can be increased by the use of additives in buffers (e.g., ionic compounds, salts, detergents, osmolytes, etc). However, the lists of possible additives are quite extensive,[2,3] and it is difficult to guess ab initio which of these additives, and at what concentration, will be successful for a particular protein. Optimization of buffer conditions for the best solubility is often achieved by screening,[2,16,17] where pH, salt concentration, buffer type, and additives are varied systematically. For proteins with very low initial solubility, several iterations of buffer optimization might be needed, which can restrict the usage of such screens.

Here, we describe an alternative way of increasing protein solubility and long-term stability, using a single set of additives for different protein targets. We show that simultaneous addition of the charged L-amino acids Arg and Glu (Arg+Glu) to the dilute protein solution significantly reduces aggregation during the process of concentration. The presence of 50 mM Arg+Glu in the buffer significantly increased the solubility limit of several proteins; it did not alter the structure of the proteins, nor prevent specific protein−protein and protein−RNA interactions. Addition of Arg+Glu to the buffer dramatically increased the lifetime and stability of the protein samples and prevented protein degradation and precipitation over time. The simplicity and effectiveness of the protocol make it highly attractive for high-throughput protein preparation for structural genomics studies, but may be used in many other areas where high protein concentration and long-term sample stability are required.

Materials and Methods

Protein Expression and Purification. DNA sequences encoding amino acids 1−153 and full-length murine REF2-1, human MAGOH, Y14, UAP56, TAP, and amino acids 8−120 of Herpesvirus saimiri ORF57 were cloned into various pET vectors (Novagen) allowing fusions to a 6xHis tag sequence. *GST*-tagged *REF2-1* and *Y14* constructs were built into pGEX vectors (Amersham Biosciences). Proteins were expressed into *E. coli* BL2I(DE3) cells (Novagen) harboring RP or RIL pUBS vectors (Statagene) in 750 mL to 8.5 L of Terrific Broth medium using shaking flasks or a fermentor, after induction of the protein synthesis with 200 μM isopropyl-1-thio-β-D-galactopyranoside (IPTG). 6xHis-tagged proteins were purified using a standard metal affinity protocol on Fast-Flow TALON/Cobalt beads (Clontech) packed into columns of various sizes (Amersham Biosciences). Target proteins were eluted by step in 50 mM Tris-HCl pH 8.0, 1 M NaCl, 200 mM imidazole. Pooled fractions were diluted to 1.0−1.5 mg/mL in the elution buffer to avoid protein precipitation and then dialyzed against different buffers. TAP was further purified on a 5 mL Hi-Trap Heparin column (Amersham Biosciences) and was eluted in 50 mM Tris pH 8.0, 700 mM NaCl, 1 mM DTT, whereas UAP56 was loaded onto a 120 mL Sephacryl-S200 gel filtration column (Amersham Biosciences) in 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT. At the end of purification, the dilute proteins were in the following buffers: Ref2NM and (8−120)ORF57, 20 mM Na phosphate, pH 6.3, 100 mM NaCl, 50 mM β-mercaptoethanol (β-ME), 10 mM DTT; Y14, 50 mM Tris pH 8.0, 150 mM NaCl, 10 mM β-ME; TAP, 50 mM Tris pH 8.0, 700 mM NaCl, 1 mM DTT; WW34, 5 mM Na phosphate, pH 7.0, 50 mM NaCl, 2 mM KCl.

Table 1. Solubility of Proteins Concentrated in the Absence (−) or in the Presence (+) of 50 mM Arg+Glu

				protein solubility, mM		
protein	MW,[a] kDa	pI[a]	pH[b]	−	+	solubilization factor[c]
Ref2NM	19.6	10.0	6.3	0.15	1.3	8.7
MAGOH	18.0	5.8	6.3	0.11	0.9	8.2
WW34	10.3	5.3	5.3	0.42[d]	1.5	3.6
Y14	20.7	6.3	6.3	0.34	0.8	2.4
ORF57	14.9	9.5	6.3	0.17	0.7	4.1
TAP	71.0	9.0	6.3	0.04	0.2	5.0

[a] Molecular weight (MW) and pI were calculated for each protein taking into account tag sequences if used for protein purifications. [b] pH of the buffer used for concentrating experiments. [c] Calculated as the ratio between the protein solubility in the presence and absence of Arg+Glu. [d] Determined by dialysis of concentrated 1.5 mM protein solution against the buffer without Arg+Glu.

Protein Solubility Studies. For dialysis experiments, purified protein solutions (in the buffers described above) were supplemented with Arg and Glu (Sigma) to a final concentration of 100 mM and were then concentrated up to their solubility limits using a 50 mL stirring concentrator (Amicon, Inc.) in combination with solvent evaporation. Concentrated protein samples of 30−40 μL were then microdialyzed against buffers (20 mM Na phosphate buffer pH 6.3, 100 mM NaCl, 10 mM DTT) containing decreasing concentrations of Arg and/or Glu.

The ab initio concentration experiments of the dilute proteins were done once in the presence and once in the absence of Arg+Glu. For one set of these experiments, Arg+Glu (final concentration 50 mM) was added to the purified dilute protein solutions in the original buffers (see above) before dialysis against 20 mM Na phosphate buffer (pH 5.3 or 6.3, see Table 1), 100 mM NaCl, 50 mM Arg, 50 mM Glu, 5 mM EDTA, and 10 mM DTT. In another set of ab initio concentration experiments, no Arg+Glu was added to the protein samples and buffer solutions (which were otherwise the same). Concentration up to the solubility limit was performed on a 50 mL stirring concentrator. Soluble protein concentrations were measured using Bradford assays (Biorad).

GST Pull-Down Assays. Soluble protein extracts from 0.4 g of induced cells were added to 25 μL of GSH beads (Amersham Biosciences) for 20 min at 4 °C to allow the binding of ~25 μg of proteins to the beads. An excess of purified human UAP56 or MAGOH (100 μg) was added to the washed beads in RB100 buffer (25 mM Hepes-KOH pH 7.5, 100 mM KOAc, 10 mM $MgCl_2$, 1 mM DTT, 10% glycerol) for 20 min at 4 °C. Aliquots of eluted proteins in 50 μL of buffer containing 40 mM reduced glutathione were loaded on 12% SDS-PAGE and stained with Coomassie blue.

RNA Electrophoretic Mobility Shift Assays (EMSA). *S. pombe* Ras and REF2-1 proteins purified by metal affinity chromatography on cobalt columns were dialyzed against EMSA buffer, before RNA binding reactions were allowed in the absence or in the presence of 50 mM Arg+Glu using various concentrations of Ras or REF2-1, and 0.5 ng of a ^{32}P-continuously labeled 58 mer RNA probe (25 000 cpm) (Promega), essentially as described in ref 18.

NMR Spectroscopy. NMR experiments were performed on a Bruker Avance DRX600 spectrometer equipped with either a TXI or a CryoProbe probehead, using standard pulse sequences. The receiver gain (40−128) was lower for homonuclear spectra because of the presence of strong signals from 50 mM Arg+Glu; in heteronuclear spectra, these signals were mainly filtered out, and the receiver gain was similar to that when no Arg+Glu were present and mainly limited by the presence of residual water signal. NMR spectra were collected at 298 K in the 20 mM phosphate buffer described previously. Either pure ^2H_2O or a 50 mM solution of Arg+Glu in ^2H_2O was added to the protein samples (at 8%) for the deuterium lock. The 2D NOESY spectrum of Ref2NM in ^2H_2O was acquired with a mixing time 150 ms at 303 K using the TXI probe.

(16) Bagby, S.; Tong, K. I.; Liu, D. J.; Alattia, J. R.; Ikura, M. *J. Biomol. NMR* **1997**, *10*, 279−282.
(17) Lepre, C. A.; Moore, M. *J. Biomol. NMR* **1998**, *12*, 493−499.
(18) Wilson, S. A.; Brown, E. C.; Kingsman, A. J.; Kingsman, S. M. *Nucleic Acids Res.* **1998**, *26*, 3460−3467.

Results

Protein Test Sample Set. To assess a solubilization protocol, we chose a set of six proteins with unrelated sequences and poor solubility properties. This set consists of fragment 1−153 of murine REF2-1 (Ref2NM),[19] human MAGOH,[20] Y14,[21] and TAP,[22,23] WW domains 3 and 4 from *Drosophila* Su(dx) protein (WW34),[24] and fragment 8−120 of the Herpesvirus saimiri ORF57 protein.[25,26]

Addition of Equimolar Amounts of Arg and Glu Improves the Solubility and Long-Term Stability of Proteins. To test the concentration effects of Arg and/or Glu on protein solubility, three proteins were used: Ref2NM, MAGOH, and WW34. The purified protein solutions were concentrated (for buffer content, see Materials and Methods) in the presence of 100 mM Arg+Glu to their maximum solubility limits, which were in all cases significantly higher than in the buffer without Arg and/or Glu (data not shown). Microdialysis reactions were then set up against the buffers containing various decreasing amounts of Arg and/or Glu. After dialysis, the concentrations of the remaining soluble protein were measured and normalized to concentrations obtained in a dialysis reaction performed without Arg or Glu. Precipitation was observed after dialysis under all buffer conditions, implying that solubility limits have been reached in all cases. The normalized concentrations (solubilization factors) directly reflect the solubilization effects of different amounts of Arg and Glu (Figure 1A). Either Arg or Glu alone does not improve significantly the solubility of the proteins, whereas the presence of equimolar amounts of both Arg and Glu clearly increases the solubility of all three proteins (Figure 1A). The solubilization factors for each of the proteins in 50 mM Arg+Glu are approximately double those at 25 mM, suggesting that the protein solubility in this range of concentrations is roughly proportional to the concentration of Arg+Glu (Figure 1A). These results are reflected qualitatively by differences in the intensity of the Coomassie stained protein bands on SDS-PAGE, where equal aliquots of soluble fractions were loaded (Figure 1B). The presence of Arg and/or Glu in the samples loaded on an SDS-PAGE gel does not affect the electrophoretic mobility of the proteins and does not interfere with this method routinely used for protein characterization.

Soluble protein supernatants obtained after dialysis were stored for 5 days at room temperature to assess the precipitation and degree of proteolytic degradation over time in buffers containing different amounts of Arg and/or Glu. The same amount of total protein stored in each buffer was then loaded on SDS-PAGE to check for degradation. All three proteins were significantly degraded in the buffer containing neither Arg nor Glu (Figure 1C, lane 1), whereas in buffers containing Arg and/or Glu, proteolytic degradation was essentially suppressed (compare lanes 2−7 with 1). No precipitation was noticed after 5 days for protein samples in 50 mM Arg+Glu.

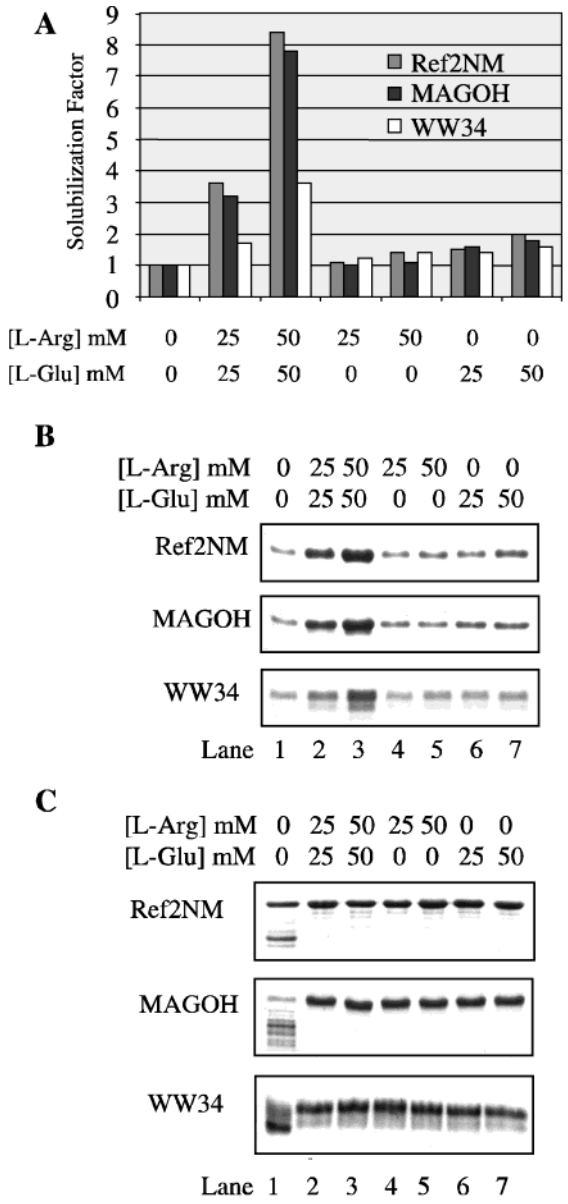

Figure 1. Arg and Glu effects on protein solubility and stability. Concentrated proteins Ref2NM, MAGOH, and WW34 were dialyzed against buffers containing various amounts of Arg and Glu. (A) Soluble protein concentrations normalized to concentrations measured without Arg or Glu. (B) Same volumes of protein solutions recovered after buffer exchanges were loaded on 13.5% SDS-PAGE and stained with Coomassie blue. (C) After incubation for 5 days at 25 °C, the same amounts of total protein extracts were loaded on 13.5% SDS-PAGE and stained with Coomassie blue.

Concentrating Proteins in the Presence of 50 mM Arg and Glu. The performance of the new protocol for obtaining proteins at high concentrations was tested on five proteins (Ref2NM, MAGOH, Y14, ORF57, and TAP), by ultrafiltration of the dilute protein solutions and comparing the maximum achievable concentrations of soluble proteins in the presence or absence of 50 mM Arg+Glu (Table 1). For completeness, data for WW34 are added to Table 1; however, this protein had only been concentrated in the presence of Arg+Glu, and the data on its solubility without these amino acids were taken

(19) Stutz, F.; Bachi, A.; Doerks, T.; Braun, I. C.; Seraphin, B.; Wilm, M.; Bork, P.; Izaurralde, E. *RNA* **2000**, *6*, 638−650.
(20) Kataoka, N.; Diem, M. D.; Kim, V. N.; Yong, J.; Dreyfuss, G. *EMBO J.* **2001**, *20*, 6424−6433.
(21) Kataoka, N.; Yong, J.; Kim, V. N.; Velazquez, F.; Perkinson, R. A.; Wang, F.; Dreyfuss, G. *Mol. Cell* **2000**, *6*, 673−682.
(22) Kang, Y.; Cullen, B. R. *Genes Dev.* **1999**, *13*, 1126−1139.
(23) Katahira, J.; Strasser, K.; Podtelejnikov, A.; Mann, M.; Jung, J. U.; Hurt, E. *EMBO J.* **1999**, *18*, 2593−2609.
(24) Cornell, M.; Evans, D. A. P.; Mann, R.; Fostier, M.; Flasza, M.; Monthatong, M.; Artavanis-Tsakonas, S.; Baron, M. *Genetics* **1999**, *152*, 567−576.
(25) Whitehouse, A.; Cooper, M.; Meredith, D. M. *J. Virol.* **1998**, *72*, 857−861.
(26) Goodwin, D. J.; Hall, K. T.; Stevenson, A. J.; Markham, A. F.; Whitehouse, A. *J. Virol.* **1999**, *73*, 10519−10524.

from the dialysis experiment. For all proteins concentrated in the presence of 50 mM Arg+Glu, solubilities were significantly improved (up to 8.7 times). A concentration of at least 0.7 mM was obtained in five out of six cases. A relatively low absolute concentration (0.2 mM) was achieved for TAP, a 71 kDa protein which displays extremely poor solubility in the absence of Arg+Glu (0.04 mM). Nevertheless, for this protein, a 5-fold improvement of solubility was accomplished. Remarkably, the solubility of WW34 and Y14, which have a pI equal to the pH of the solution, a situation where proteins are expected to have low solubility, was also significantly increased, thus allowing us to obtain concentrations of soluble proteins of 1.5 and 0.8 mM, respectively. The solubilization factors for Ref2NM and MAGOH achieved in this "ab initio concentrating" experiment (Table 1) were higher than those obtained in microdialysis experiments (Figure 1A), suggesting that in the microdialysis procedure some Arg and Glu were sticking to the proteins, thus improving their solubility even in the absence of charged amino acids in the exchange buffer. Therefore, dialysis experiments provide underestimated values for true solubilization factors. The ab initio concentration of Ref2NM in the presence of Arg+Glu was performed on seven different occasions (although not to the solubility limit); concentrations in the range of 0.96−1.14 mM were consistently achieved without significant precipitation (<10%, data not shown).

We noticed that the protein samples concentrated up to their solubility limits in the presence of Arg+Glu are very sensitive to further changes in salt concentration or buffer content. Often at some stage of the sample preparation protocol it is necessary to reduce the amount of salt present in the buffer (e.g., from 0.4 to 1.0 M of NaCl typically present in fractions collected after ion-exchange column to 100 mM for NMR studies). We found (data not shown) that the greatest solubilization effect is achieved if buffer and salt conditions are manipulated in the presence of Arg+Glu when the protein is still in the dilute form. Dialysis against a low salt buffer containing 50 mM Arg+Glu prior to finally concentrating the sample (by ultrafiltration) without further buffer change significantly reduces aggregation and precipitation. We also observed that for proteins concentrated to their solubility limit in the presence of 50 mM Arg+Glu, even addition of a small aliquot of ^2H_2O (to make up 5−10% of solvent required for NMR lock) caused sample turbidity and precipitation, probably due to the local decrease in Arg+Glu concentration. When 50 mM Arg+Glu solution in ^2H_2O was added, no precipitation or turbidity occurred. The presence of Arg+Glu in solution may, however, interfere with protein binding to the ion-exchange columns used for the protein purification (data not shown), in which case these amino acids should be added only after an ion-exchange chromatography purification step, but before final concentration.

Addition of 50 mM Arg and Glu Does Not Disrupt Specific Protein−Protein or Protein−RNA Interactions. Although protein aggregation at high concentrations is expected to be mainly nonspecific, due care must be taken to ensure that the agents disrupting the protein aggregation do not also affect or disrupt specific protein−ligand interactions. To test if Arg+Glu affects the specific protein−ligand interactions, a series of assays were performed for known protein−protein complexes, REF2-1·UAP56 [27] and Y14·MAGOH.[20] GST pull-down assays were performed using purified 6xHis-tagged UAP56 or MAGOH. No protein binding is detected in pull-downs using GST by itself (Figure 2, panels A and B, lane 3), this providing a negative control to check for the specificity of REF2-1 or Y14 interactions. In contrast, nearly stoichiometric binding of UAP56 to GST-REF2-1 and of MAGOH to GST-Y14 was observed (panels A and B, compare lane 4 to 3). The presence of 30 or 50 mM Arg+Glu in the buffer does not noticeably affect interactions between GST-REF2-1 and UAP56 or GST-Y14 and MAGOH (panels A and B, respectively, compare lanes 5 and 6 to 4). Moreover, the presence of Arg+Glu significantly decreases proteolytic degradation of GST-REF2-1 during the course of the reactions (Figure 2A, compare lane 4 with 5 and 6).

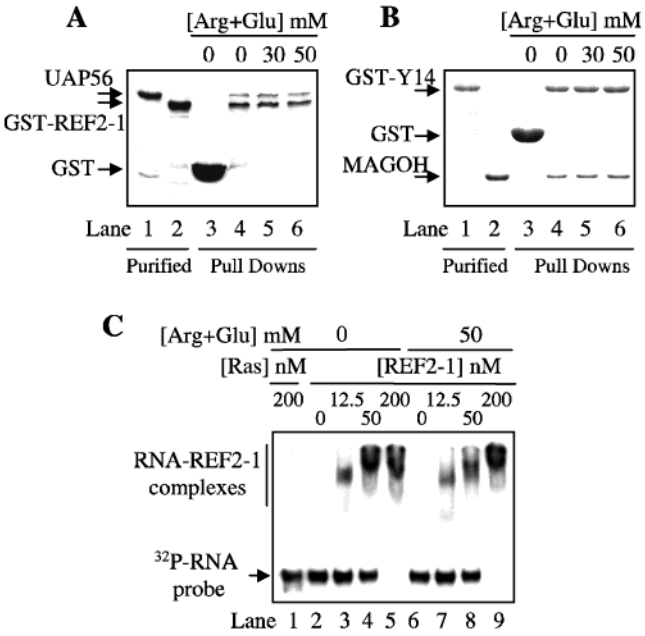

Figure 2. Protein and RNA interactions in the presence of Arg+Glu. UAP56/Ref2 (A) and MAGOH/Y14 (B) GST pull-downs were performed using various Arg plus Glu concentrations. Lanes 1 and 2 correspond to the purified proteins. GST negative control (lane 3), GST-REF2 (lanes 4−6, panel A), or GST-Y14 (lanes 4−6, panel B) were first bound to glutathione-sepharose beads. Purified UAP56 (A) or MAGOH (B) was then added to the beads into buffers containing increasing amounts of Arg+Glu. Eluted proteins were loaded on 12% SDS-PAGE and stained with Coomassie blue. (C) RNA band-shift assays using a ^{32}P-labeled 50mer RNA probe and purified proteins Ras or REF2 were performed without (lanes 1−5) or with 50 mM Arg+Glu (lanes 6−9). Negative control (lane 1) is a Ras protein that does not exhibit RNA binding activity. Binding reactions were then loaded on 5% native acrylamide gel before autoradiography.

Interactions between protein and nucleic acids are often sensitive to the salt concentration, because they are mainly driven by electrostatic interactions. Furthermore, we anticipated that free Arg can directly bind RNA and perhaps block the interaction with protein. For these reasons, we analyzed RNA−protein interactions in the presence of 50 mM Arg+Glu by electrophoretic mobility shift assays (EMSA) using REF2-1. A ^{32}P-labeled 58-mer RNA probe was added to various concentrations of REF2-1 in the presence or absence of 50 mM Arg+Glu to allow binding reactions (Figure 2C). No band-shift is detected with Ras, a *S. pombe* control protein, which does not exhibit any RNA binding activity (Figure 2C, lane 1). Addition of 50 mM Arg+Glu does not disrupt the formation of the RNA-

(27) Luo, M. J.; Zhou, Z. L.; Magni, K.; Christoforides, C.; Rappsilber, J.; Mann, M.; Reed, R. *Nature* **2001**, *413*, 644−647.

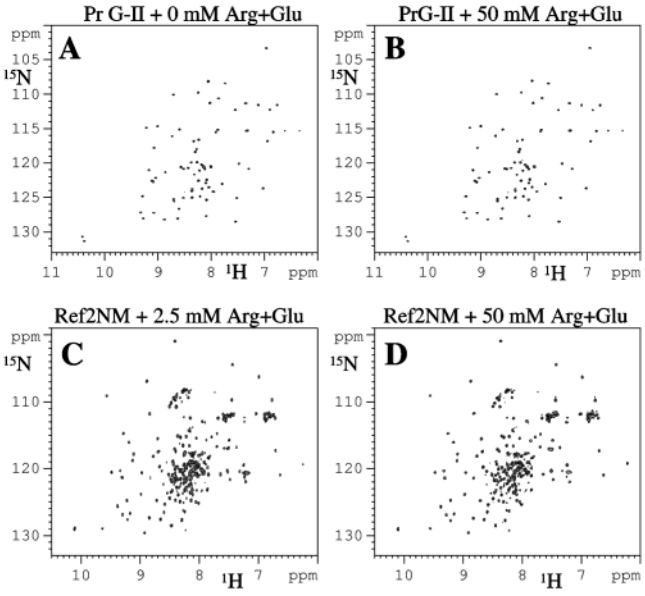

Figure 3. 2D HSQC spectra of uniformly ^{15}N-labeled PrG-II in the absence (A) and presence (B) of 50 mM Arg+Glu (protein concentration ca. 0.2 mM) and spectra of ^{15}N,^{2}H-labeled Ref2NM in the presence of 2.5 mM (C) and 50 mM (D) Arg+Glu (protein concentrations 0.03 and 0.6 mM, respectively).

REF2-1 complexes (compare lanes 7, 8, 9 to lanes 3, 4, 5), although the affinity of the RNA:REF2-1 is slightly decreased in the presence of Arg+Glu (Figure 2C, compare lane 8 with 4).

The results of pull-down and band-shift assays suggest that the specific protein−protein and protein−RNA binding properties are not adversely affected by the addition of 50 mM Arg+Glu, even at the very low protein concentrations used in these experiments.

NMR Spectra of the Proteins Are Not Affected by 50 mM Arg+Glu. An analysis of the ^{1}H−^{15}N heteronuclear single quantum correlation (HSQC) NMR spectra is an easy way to identify potential structural changes caused by the additives. Due to low solubility and strong aggregation, it was not possible to obtain a satisfactory HSQC spectrum of Ref2NM prepared without Arg+Glu in the buffer. Therefore, we used a "typical" soluble protein, domain II of the protein G (PrG-II),[28] which gives a good reference spectrum at high concentration in the absence of Arg+Glu, to check if the addition of 50 mM Arg+Glu changes the protein NMR spectrum. The 2D ^{1}H−^{15}N-HSQC spectra of uniformly ^{15}N-labeled PrG-II recorded in the absence and in the presence of 50 mM Arg+Glu (Figure 3A,B) are very similar. Similarly, the spectrum of Ref2NM at 600 μM (in the presence of 50 mM Arg+Glu) is identical to the spectrum of diluted sample at 30 μM (in the presence of 2.5 mM Arg+Glu) (Figure 3D and C, respectively). These data indicate that protein amide chemical shifts and protein structures are not affected by the presence of 50 mM Arg+Glu in the sample buffer, and a higher protein concentration achieved in the presence of Arg+Glu does not cause any noticeable line broadening as compared to the spectrum of diluted protein.

Experimental Practicalities of Using Arg and Glu for NMR Studies. The addition of free nondeuterated amino acids Arg and Glu at 50 mM gives rise to signals in ^{1}H NMR spectra which are at least 50 times more intense than the signals originating from a protein itself (with a typical concentration of 0.2−1 mM). However, the modern NMR experiments for structural studies routinely use ^{15}N- and ^{15}N,^{13}C-labeled proteins in combination with multidimensional pulse sequences which filter out signals originating from nonlabeled material (e.g., buffer). In our 3D and triple-resonance NMR experiments used for signal assignment and structure calculation of Ref2NM, no significant interference was observed from nondeuterated Arg+Glu signals (Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L.-Y., manuscript in preparation). The value of receiver gain used in such experiments was similar to that used in the absence of Arg+Glu, as it was determined mainly by the intensity of the residual water signal. The signals from ^{13}C natural abundance Arg+Glu (effective concentration ca. 0.55 mM) were comparable to ^{13}C signals of ^{13}C,^{15}N-enriched protein originating from flexible amino acid residues and, in our experience, did not pose considerable problems. A significant increase in long-term sample stability due to the presence of 50 mM Arg+Glu in the buffer allowed Ref2NM protein samples at 1 mM to last for several weeks at 30 °C without noticeable proteolytic degradation and without any precipitation (instead of just a few days and at 8.7 times lower concentration in the absence of Arg+Glu). WW34 sample was also stable for several weeks during NMR experiments in the presence of Arg+Glu.

In the case of homonuclear ^{1}H NMR spectroscopy, the quality of protein samples and the extent of protein folding can still be assessed despite the interference of the strong signals from free nondeuterated Arg and Glu (this may be useful for preliminary experiments with nonlabeled protein). The amide groups of free Arg and Glu are in fast exchange with the solvent; hence, only protein signals are visible in the amide region of the spectra. The guanidyl group of Arg at neutral pH gives a very broad signal that does not significantly interfere with protein amide signals. This means that the most informative parts of ^{1}H NMR spectra (amide region and high-field methyl signals around 0 ppm) are not obscured by the strong signals originating from these free amino acids, even if they are not perdeuterated. The dynamic range of modern spectrometers is sufficient to observe simultaneously the relatively weak protein signals and the strong signals in the aliphatic region originating from the buffer (although at slightly increased noise level). Figure 4 shows homonuclear 2D NOESY spectra of Ref2NM collected in D_2O, where the NOE signals from aromatic protons are still clearly visible. The aliphatic part of the spectrum is largely obscured by t_1 noise originating from the strong signals from the buffer. Using perdeuterated free amino acids Arg and Glu should eliminate these strong signals and reduce t_1 noise in homonuclear NMR spectra. In summary, our experience shows that in many cases (e.g., heteronuclear experiments for signal assignment and structure calculation using ^{15}N- or ^{15}N,^{13}C-labeled proteins, and some ^{1}H homonuclear applications using nonlabeled proteins) the addition of the nondeuterated 50 mM Arg+Glu does not significantly interfere with the NMR experimental setup and is not detrimental to spectral quality.

Discussion

The main area of applicability of the technique proposed here is a situation when a protein can be purified and is soluble at low concentration, but aggregates or precipitates during the

(28) Lian, L. Y.; Derrick, J. P.; Sutcliffe, M. J.; Yang, J. C.; Roberts, G. C. K. *J. Mol. Biol.* **1992**, *228*, 1219−1234.

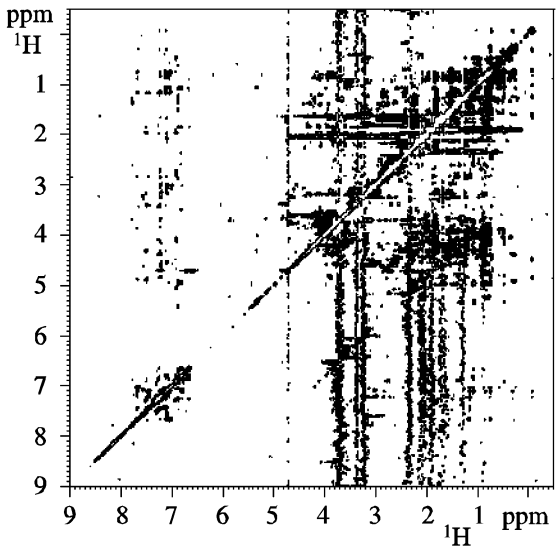

Figure 4. 2D NOESY spectra (mixing time 0.15 s) of 1 mM Ref2NM in the presence of nondeuterated 50 mM Arg+Glu collected at 303 K in ^2H_2O overnight.

process of concentrating up to a higher level (e.g., such as 0.5−1 mM typically required for the NMR structural analysis or other biochemical/biophysical studies). Here, we demonstrated that the addition of a 50 mM 1:1 mixture of Arg:Glu to the dilute protein solution can significantly increase the maximum achievable concentration of soluble protein and reduce its aggregation over time. Interestingly, the additives additionally protect the proteins from proteolytic degradation, which by itself often presents a significant challenge.

The effects of the charged amino acids on the stability and aggregation properties of proteins have been studied before, and these effects were found to vary significantly depending on the protein charge and amino acid concentration. Positively charged Arg at concentrations greater than 0.5 M is, in fact, one of the most common additives known to assist refolding of recombinant proteins from the inclusion bodies.[2,14,15,29−31] Above 0.1 M, it prevents heat-induced aggregation[32,33] and aggregation following protein dilution from 8 M urea.[32] At lower concentrations (0.1−0.4 M), Arg neither stabilizes nor destabilizes proteins against thermal stress; however, at higher concentrations (0.5−2 M), it decreases the melting temperature and destabilizes protein.[33,34] The mechanism by which Arg reduces heat- and denaturation-induced aggregation is still unclear; potentially, it may influence both the solubility and the stability of the native, intermediate, denatured, and aggregated states in a complex manner, or it can affect the relative rates of folding and/or aggregation.[15,32] Negatively charged amino acids (Asp and Glu) were found to promote lysozyme aggregation after protein dilution from 8 M urea, but to suppress its aggregation after thermal stress.[32] Either Arg or Glu separately, and at a concentration of 50 mM, suppressed to various extents the heat-induced aggregation for the majority of the proteins in the test set used in ref 32. On the basis of the surface tension measurements, it was previously concluded that at lower concentrations, ArgHCl penetrates the solvation layer of negatively charged BSA; at high concentrations (above 0.5 M), addition of amino acid salts ArgHCl and NaGlu causes preferential hydration of proteins (Arg^+ and Glu^- are excluded from the protein surface), thus stabilizing them.[35] Glu^- ions were found to bind to the positively charged lysozyme, thus offsetting the contribution from the increased surface tension effect.[35] The same study revealed that the dipeptide ArgGlu (0.2−0.77 M) also caused preferential hydration both for lysozyme and for BSA and stabilized these proteins.[35] Addition of 1 M NaGlu increased the transition temperature of both lysozyme and BSA.[36]

Several general conclusions appear from all of these published studies. First, at relatively high concentrations of Arg or Glu (≫0.1 M), the preferential hydration of proteins, which is the main mechanism of action of osmolytes,[37] provides the major influence on protein properties in solution. Second, at lower concentrations (<0.1 M), the effects of specific binding of Arg^+ or Glu^- ions to a protein (depending on its overall charge) dominate over the effect of preferential hydration; protein thermodynamic stability is not significantly affected.

To our knowledge, the current work is the first observation that simultaneous addition of free amino acids Arg and Glu is required to achieve the maximal suppression of aggregation and increase in protein solubility. A quantitative evaluation is also presented. As seen from Figure 1, addition of either Arg or Glu at a concentration up to 50 mM does not increase significantly the limiting concentration of proteins in our test set, whereas simultaneous addition of Arg and Glu increases protein solubility dramatically (up to 8.7 times, Table 1). Another prominent feature of the Arg+Glu mixture is that whereas the solubility of Glu itself is relatively low (50−60 mM in water), it becomes highly soluble in combination with Arg. This enables the preparation of a 1 M Arg+Glu aqueous stock solution (pH ∼6.5) which is convenient to add to the samples and to prepare the buffers.

A rigorous description of the mechanism by which the Arg+Glu mixture suppresses aggregation is likely to be complicated, as oppositely charged Arg and Glu interact with each other and with the protein, water shell, and the different ions in the buffer. In the crystal of Arg•Glu salt,[38] molecules of Arg and Glu form separate alternating layers, stabilized by ionic interactions and hydrogen bonds involving amino, guanidyl, and carboxylate groups and water molecules. It is likely that in solution Arg+Glu will also form highly structured associates. We assume that the protein aggregation and precipitation during the concentration process from dilute solutions is driven mainly by the surface interactions in the folded state, as opposed to aggregation of exposed hydrophobic cores of unfolded proteins.[8] We speculate that at the neutral or slightly acidic pH, when both Arg and Glu are charged, they interact with (and mask) oppositely charged groups on the surface of the protein, while the aliphatic hydrophobic parts of the side chains of Arg and Glu interact with and cover the adjacent exposed hydrophobic

(29) Suenaga, M.; Ohmae, H.; Tsuji, S.; Itoh, T.; Nishimura, O. *Biotechnol. Appl. Biochem.* **1998**, *28*, 119−124.
(30) De Bernardez Clark, E.; Schwarz, E.; Rudolph, R. *Methods Enzymol.* **1999**, *309*, 217−236.
(31) Lu, H.; Zhang, H. H.; Wang, Q.; Yuan, H. Y.; He, W.; Zhao, Z.; Li, Y. Y. *Curr. Microbiol.* **2001**, *42*, 211−216.
(32) Shiraki, K.; Kudou, M.; Fujiwara, S.; Imanaka, T.; Takagi, M. *J. Biochem.* **2002**, *132*, 591−595.
(33) Arakawa, T.; Tsumoto, K. *Biochem. Biophys. Res. Commun.* **2003**, *304*, 148−152.
(34) Taneja, S.; Ahmad, F. *Biochem. J.* **1994**, *303*, 147−153.
(35) Kita, Y.; Arakawa, T.; Lin, T. Y.; Timasheff, S. N. *Biochemistry* **1994**, *33*, 15178−15189.
(36) Arakawa, T.; Timasheff, S. N. *J. Biol. Chem.* **1984**, *259*, 4979−4986.
(37) Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721−9726.
(38) Bhat, T. N.; Vijayan, M. *Acta Crystallogr.* **1977**, *B33*, 1754−1759.

parts of the protein surface, thus reducing its general "stickiness" which causes protein aggregation. Combining hydrophobic and electrostatic energetic components should provide more efficient counterion binding than that by the monovalent ions (e.g., Na^+, K^+, Cl^-), which are usually added to protein samples to improve solubility. Taking into account that protein surfaces are made of the same types of amino acid residues, and that the nature of the noncovalent interactions (i.e., electrostatic and hydrophobic) driving self-aggregation of different proteins is the same, it can be expected that the Arg+Glu mixture will be as universal as monovalent salts (e.g., NaCl or KCl) for suppressing aggregation for a wide variety of proteins (although more efficient). In the present paper, this effect was demonstrated experimentally for six unrelated proteins with known solubility problems.

Here, we have observed that protein samples concentrated to their solubility limit in the presence of Arg+Glu are very sensitive to the further changes in buffer conditions. For example, removing an excess of NaCl (performed by dialysis against same buffer, but with less NaCl) from such concentrated samples causes significant precipitation. This suggests that Na^+ and Cl^- ions (which initially are present in excess) compete with the binding of Arg^+ and Glu^- to the protein surface. If the NaCl concentration is decreased and bound monovalent counterions are released, the protein self-association occurs before Arg^+ and Glu^- are able to cover the surface of the protein. Performing buffer exchanges at low protein concentrations, when self-association is reduced, maximizes the yield of soluble protein.

The amount of Arg+Glu required to cover the charged patches (and, hence, to prevent aggregation) is likely to depend on the number of charged exposed residues (i.e, size of the protein and its amino acid composition) and the number of protein molecules (i.e., protein concentration). Proteins with more charges may require more Arg+Glu to achieve the same molar protein solubility level. Protein TAP, which has the lowest absolute solubility (Table 1), has approximately 3 or 4 times more charges then the other proteins in the test set, suggesting that there was not enough Arg+Glu in solution to cover its charged patches at higher protein concentration. Our experiments showed that an increase of Arg+Glu concentration from 25 to 50 mM increases protein solubility approximately twice (consistent with a 2-fold increase in Arg^+ and Glu^+ ions able to bind with the protein surface). However, the level of protein concentration achieved for a majority of proteins in the test set was sufficient for typical structural NMR work; therefore, concentrations of Arg+Glu higher than 50 mM were not quantitatively studied here. The concentrations of Arg+Glu much higher than 100 mM could potentially start to affect the stabilities of the proteins due to an increased osmolytic effect and may weaken protein−ligand interactions. In NMR applications, higher Arg+Glu concentrations may necessitate the use of perdeuterated Arg and Glu to reduce their proton signals.

Using nondeuterated 50 mM Arg+Glu in our experience did not cause any significant problems for signal assignment and structure calculation of ^{15}N- or ^{15}N,^{13}C-labeled proteins, although using perdeuterated Arg and Glu will be advantageous (although at additional cost) for some applications where the presence of strong signals will obscure the signals of interest (e.g., for nonlabeled protein samples). The current cost of deuterated L-Arg and L-Glu is comparable and only slightly higher than that of other deuterated compounds used as additives or buffers for NMR (like DTT, MES, HEPES, etc). The necessity to add Arg and Glu to a relatively large volume of diluted protein solution before concentrating it (see Results) may significantly affect the cost of the sample preparation in the presence of perdeuterated amino acids. To reduce the costs, it is recommended that a final buffer exchange step be introduced following protein concentration during which the nondeuterated Arg+Glu is exchanged for the perdeuterated versions.

It is widely accepted that working at a pH close to the isoelectric point (pI) of a protein is risky because protein molecules tend to aggregate more when electrostatic repulsions between them vanish. Here, we demonstrated that the addition of Arg+Glu significantly increased the solubility of WW34 and Y14 even at a pH equal to pI. The ability to suppress the aggregation at a pH close to the pI can be advantageous, for example, when working with multiprotein complexes where it may be difficult to choose a sample pH discrete from the pI values of all the components.

The mechanism by which Arg and Glu inhibit sample degradation is not entirely clear. We assume that this degradation is caused by the traces of proteases that are present in the protein samples. As we discussed above, protein thermodynamic stability is not affected by the low (50 mM) concentration of Arg or Glu; hence, protein stabilization is unlikely to be responsible for greater protease resistance. We speculate that the free charged amino acids, due to their significant excess, act as competitive inhibitors of proteases. An alternative explanation is that binding of Arg and Glu to the surface of the protein "masks" recognition sites for the proteases and thus interferes with protease binding.

The protocol for improving protein solubility proposed in this paper has been successfully tested for several unrelated protein targets with poor solubility. This method is simple, does not modify proteins or their specific binding properties, and, importantly, simultaneously protects against proteolytic degradation. It has made possible structural and biochemical studies that have otherwise been precluded and hampered by poor solubility. Both Arg and Glu are very mild chemical compounds with inherent affinities to the proteins; they are "natural", as they are present in a cell and surround proteins in vivo. Whether the presence of Arg+Glu affects protein crystallization requires additional work. The charged amino acids, Arg and Glu, can be routinely incorporated in sample preparation procedures in NMR and any other biochemical applications, including high-throughput ones, where high protein concentrations and long-term sample stability are required.

Acknowledgment. We are grateful to Drs. J. Avis and O. Fedoroff who supplied WW34 protein, to Dr. D. Hughes for the gift of pET14b-Ras plasmid, and to Marina Golovanova for supplying ^{15}N-labeled PrG-II and technical support. The work was supported by a grant from the BBSRC (S.A.W.) and a NESBiC Pump-Priming Grant (S.A.W, L.-Y.L., and A.P.G.). The Wellcome Trust is acknowledged for funding the 600 MHz spectrometer at UMIST (Grant No: 057806 to L.-Y.L.).

JA049297H