Journal of Magnetic Resonance **148,** 142–146 (2001)
doi:10.1006/jmre.2000.2213, available online at http://www.idealibrary.com on IDEAL®

Selectively Labeling the Heterologous Protein in *Escherichia coli* for NMR Studies: A Strategy to Speed Up NMR Spectroscopy

F. C. L. Almeida,*[,1] G. C. Amorim,* V. H. Moreau,* V. O. Sousa,* A. T. Creazola,* T. A. Américo,* A. P. N. Pais,* A. Leite,† L. E. S. Netto,‡ R. J. Giordano,§ and A. P. Valente*[,1]

**Centro Nacional de Ressonância Magnética Nuclear, Departamento de Bioquímica Médica, ICB, CCS, Universidade Federal do Rio de Janeiro–UFRJ, Cidade Universitária, Rio de Janeiro, RJ 21941-590, Brazil; †Universidade de Campinas UNICAMP, Campinas, SP, Brazil; ‡Instituto de Biologia, Universidade de São Paulo–USP, São Paulo, SP, Brazil; and §M.D. Anderson Cancer Center, GU Medical Oncology Department, University of Texas, Houston, Texas*

Received June 27, 2000

Nuclear magnetic resonance is an important tool for high-resolution structural studies of proteins. It demands high protein concentration and high purity; however, the expression of proteins at high levels often leads to protein aggregation and the protein purification step can correspond to a high percentage of the overall time in the structural determination process. In the present article we show that the step of sample optimization can be simplified by selective labeling the heterologous protein expressed in *Escherichia coli* by the use of rifampicin. Yeast thioredoxin and a coix transcription factor Opaque 2 leucine zipper (LZ) were used to show the effectiveness of the protocol. The $^{1}H/^{15}N$ heteronuclear correlation two-dimensional NMR spectrum (HMQC) of the selective ^{15}N-labeled thioredoxin without any purification is remarkably similar to the spectrum of the purified protein. The method has high yields and a good $^{1}H/^{15}N$ HMQC spectrum can be obtained with 50 ml of M9 growth medium. Opaque 2 LZ, a difficult protein due to the lower expression level and high hydrophobicity, was also probed. The ^{15}N-edited spectrum of Opaque 2 LZ showed only the resonances of the protein of heterologous expression (Opaque 2 LZ) while the ^{1}H spectrum shows several other resonances from other proteins of the cell lysate. The demand for a fast methodology for structural determination is increasing with the advent of genome/proteome projects. Selective labeling the heterologous protein can speed up NMR structural studies as well as NMR-based drug screening. This methodology is especially effective for difficult proteins such as hydrophobic transcription factors, membrane proteins, and others. © 2001 Academic Press

Key Words: **NMR; Opaque 2; rifampicin; selective labeling; thioredoxin.**

INTRODUCTION

Protein NMR[2] studies need high protein concentration and purity. Purification of proteins that behave well in solution is straightforward. On the other hand, some proteins can aggregate and be very difficult to handle in solution. The majority of proteins in the cell can be difficult and the methodology used to deal with them needs improvement. For those proteins, purification can be the most time-consuming step for structure determination by NMR. Poor yields during purification and protein aggregation are problems that often occur and can be avoided by working with salt, pH, and temperature and sometimes by the addition of cosolvents or even by making protein complexes.

Protein expression in the presence of rifampicin enables labeling of only the heterologous proteins. This occurs since cells self protein synthesis is inhibited but not heterologous expression. Rifampicin binds to the bacterial β-subunit of RNA polymerase, inhibiting transcription initiation (*1*). On the other hand, rifampicin does not inhibit T7 RNA polymerase, which is used for expression of the heterologous protein. This is a well-known methodology for labeling with specific amino acids (*2*, *3*).

The present work shows a simple and cheap methodology using rifampicin to simplify NMR sample optimization for structural determination. The growth is done in nonlabeled M9 medium. Rifampicin is added, inhibiting the cell self protein synthesis. The cells are then pelleted and resuspended in ^{15}N-labeled M9 medium with isopropyl-β-D-thiogalactopyranoside (IPTG) and rifampicin. The use of rifampicin in the M9 medium with IPTG permits the selective labeling of protein with heterologous expression. The labeling was monitored by the comparison of the proton spectrum and the $^{1}H/^{15}N$ HMQC spectrum. Yeast thioredoxin type 2 (*4*) (TRX, 12 kDa) and Opaque 2 leucine zipper (LZ, 9 kDa), a transcription factor from coix (*5*), were used to test the rifampicin protocol. The nonpurified NMR samples showed a HMQC spectrum that was remarkably similar to those of the purified protein and consistent with the expected properties.

[1] To whom correspondence should be addressed at Av. Brigadeiro Trompowiski s/n, CCS-ICB, Bloco E sala 10, Cidade Universitaria, Rio de Janeiro, RJ 21941-590, Brasil. Fax: (5521) 2708647. E-mail: valente@cnrmn.bioqmed.ufrj.br; fameida@cnrmn.bioqmed.ufrj.br.

[2] Abbreviations used: HMQC, heteronuclear multiple quantum coherence; NMR, nuclear magnetic resonance; TPPI, time proportional phase increment; TRX, thioredoxin.

1090-7807/01 $35.00

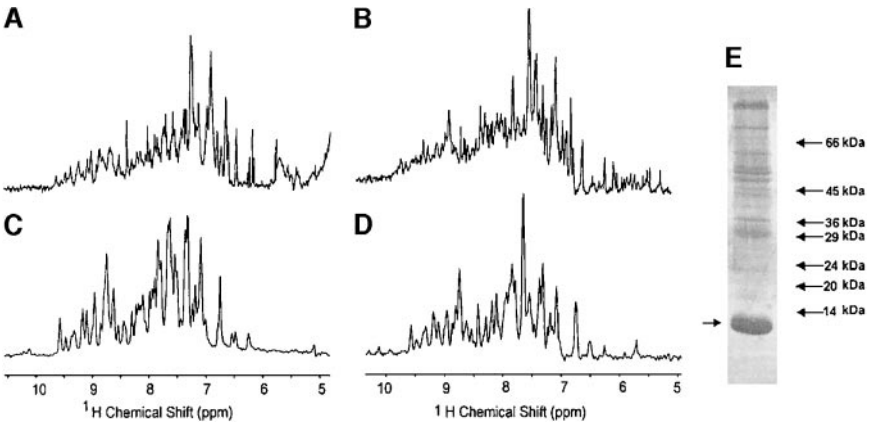

FIG. 1. ^{1}H spectra (A, B) and ^{15}N-edited ^{1}H spectra (C, D) of TRX. Spectra of purified TRX are shown in A and C. The samples obtained using rifampicin protocol samples are shown in B and D. E shows the SDS–PAGE of the rifampicin NMR sample, the same as spectra B and D.

RESULTS

The effectiveness of the rifampicin protocol was tested by comparing the spectra of purified and nonpurified TRX, prepared using the rifampicin protocol. The comparison of the ^{1}H spectrum and the one edited for ^{15}N enables the evaluation of the selectiveness of the labeling, since ^{1}H spectra will show the resonances from all proteins in the cell and the edited spectrum will show resonances only from the labeled protein. In Fig. 1 we show the ^{1}H spectra of the purified TRX and the sample obtained with the rifampicin protocol. Several proton resonances show up in the sample prepared with rifampicin (Fig. 1B) when compared with the purified TRX (Fig. 1A), since resonances from cell self proteins also appear in this spectrum. Different results are obtained for the ^{15}N-edited spectra (Figs. 1C and 1D). The spectra in Figs. 1C and 1D are remarkably similar, showing that the use of rifampicin methodology stops cell self protein synthesis and enables the ^{15}N labeling of, exclusively, TRX.

Figure 2 confirms the observation in Fig. 1 but with more details. The two-dimensional HMQC of the purified TRX (Fig. 2A) is almost the same as the nonpurified TRX (Fig. 2B). The threshold was set low in spectrum B to show the absence of weak extra peaks. The spectra of the purified TRX were obtained from 1 L of growth medium and the one with rifampicin was obtained with 50 mL (see the legend to Fig. 2 for experimental details).

As shown in the SDS–PAGE in Fig. 1E, TRX has a high expression level. It may represent more than 50% of all proteins in the cell. Since in an NMR spectrum the low-molecular-weight proteins give rise to the most intense peaks, it is expected that the spectra of the TRX grown always in labeled media has the great majority of peaks from TRX. The spectrum in Fig. 2C shows this important control. The supernatant of the lysed cells grown all the time in ^{15}N-labeled M9 shows that several proteins other than TRX are also labeled (Fig. 2C). Of course, TRX resonances are clearly prominent. However, several extra peaks also appear in the amide region. This control shows the importance of using rifampicin to get selectivity in the labeling of the heterologous protein even if the protein is well behaved in solution and the clone has very high expression levels.

It is important to show the effectiveness of the rifampicin protocol for a difficult protein. Opaque 2 LZ is a transcription factor from coix (*5*). This protein has an unknown structure. It is a leucine zipper with a probable coiled coil structure. It is highly hydrophobic and the expression yield is low. The rifampicin protocol was performed and a step of ultrafiltration was included for sample concentration and semi-purification. The Opaque 2 LZ is 9 kDa and passes through the 30-kDa membrane and is retained in the 3.5 kDa membrane (Fig. 3C).

The material retained in the 3.5-kDa membrane (where Opaque 2 LZ is present) was checked for ^{15}N labeling. The differences between the ^{1}H and ^{1}H/^{15}N HMQC spectra clearly show that the heterologous protein is the one that is being labeled (Fig. 3). The SDS–PAGE in Fig. 3C shows that even after ultrafiltration steps, the sample has several proteins other than Opaque 2 LZ.

The two-dimensional ^{1}H/^{15}N HMQC spectrum (Fig. 4) is consistent with a leucine zipper, with low dispersion similar to GCN4 (*6*) although the protein has a sign of partial aggregation. At this moment, we are improving sample conditions. This protein has 86 amino acids and we can count approximately 70 cross peaks (Fig. 4). It is important to mention that there are no cross peaks from free amino acids or small peptides, since the sample was ultrafiltrated in a membrane of 3.5-kDa molecular weight cutoff.

Another way to control the selectiveness of the labeling is to check the presence of labeled proteins with higher molecular weights (retained on the 30-kDa cutoff ultrafiltration membrane). When the induction is performed in the presence of rifampicin, no ^{15}N signal could be measured in the sample retained in the 30-kDa membrane (Fig. 5B). In contrast, in the

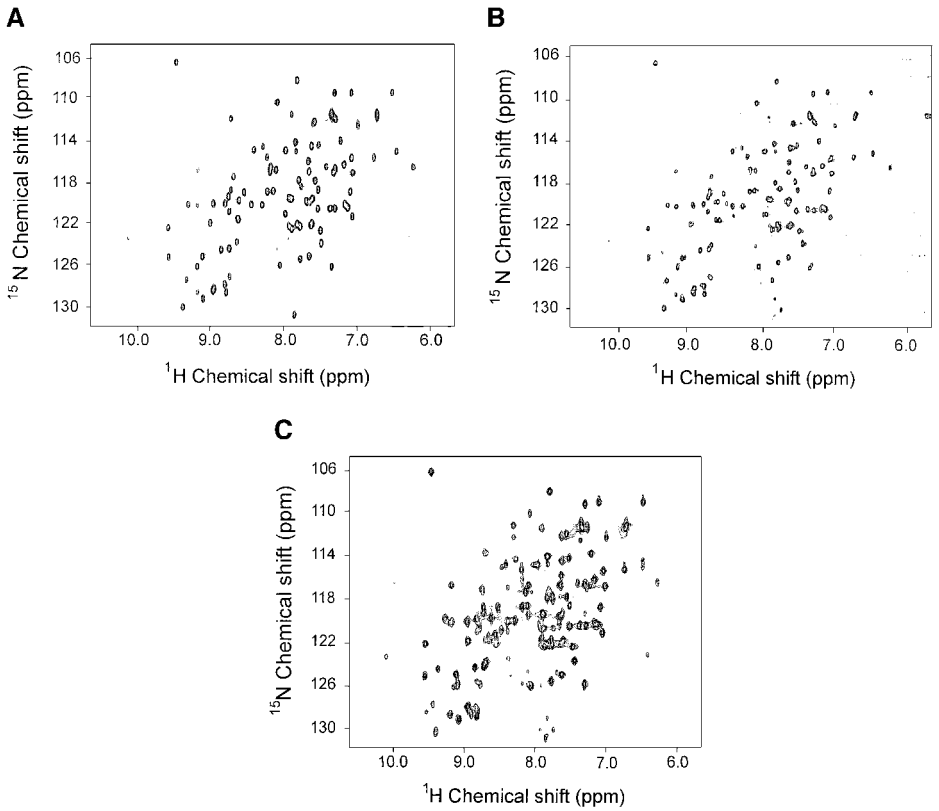

FIG. 2. Two-dimensional ^{15}N, ^{1}H-HMQC spectra of TRX. Purified TRX (A), rifampicin protocol sample (B), and totally labeled control (C) HMQC spectra were analyzed. The spectra were obtained at 303 K. The threshold was set low in spectrum B to show the absence of extra weak peaks. All spectra were acquired with 1024 × 200 points, 256 scans, and a recycle delay of 1.2 s. The processing was performed with zero filling and a square sine multiplication shifted by 90° in both dimensions. The States–TPPI method was used for quadrature detection in the indirect dimension (*12*). GARP was used to decouple ^{15}N during acquisition (*13*) and WATERGATE for water supression (*14*).

sample obtained when the cells grew all the time in labeled medium, a strong ^{15}N signal is present. These results show that rifampicin drives the labeling to the heterologous protein (Fig. 5A).

DISCUSSION AND CONCLUSION

The rifampicin protocol simplifies the protein labeling for NMR studies. The first steps of NMR sample optimization are facilitated. The first NMR spectrum can be obtained after growing, cleaving the cells and acquiring a ^{15}N-edited spectrum.

The two-dimensional ^{15}N HMQC of the rifampicin protocol sample spectrum is very similar to those of the purified protein. ^{13}C labeling and ^{15}N/^{13}C double labeling also give good results (data not shown). Gronenborn and Clore proposed a similar procedure by using an overexpressed protein (*7*). The strategy presented here is also simple, offers high yields and, most importantly, the rifampicin protocol can also be used for difficult low-expressing proteins. The rifampicin protocol can make the protein preparation cheaper by reducing the amounts of ^{13}C-glucose and $^{15}NH_4^+$ to be used. A 250-ml growth medium culture is sufficient to perform three-dimensional experiments. Lee *et al.* used a similar protocol to specifically label proteins with selected amino acids for NMR studies (*2*). In the present work we show that such a methodology has broader uses.

For the Opaque 2 LZ we have performed a fast prepurification step by using ultrafiltration membranes. This procedure is fast, with almost no loss of protein. It is important because, by selecting only the low-molecular-weight proteins, it enables solubilization at the concentrations necessary for NMR purposes.

The rifampicin protocol can also be used to follow proteins during purification by screening the ^{15}N signal on column fractions or in the supernatant of protein precipitation. This procedure can be easily done through a one-dimensional HMQC spectrum.

Several genome projects are being carried out nowadays. These projects consist of the identification and isolation of many novel genes. The next step will be the proteome projects. These include the elucidation of expression control, structure, and function of proteins coded by the large number of gene

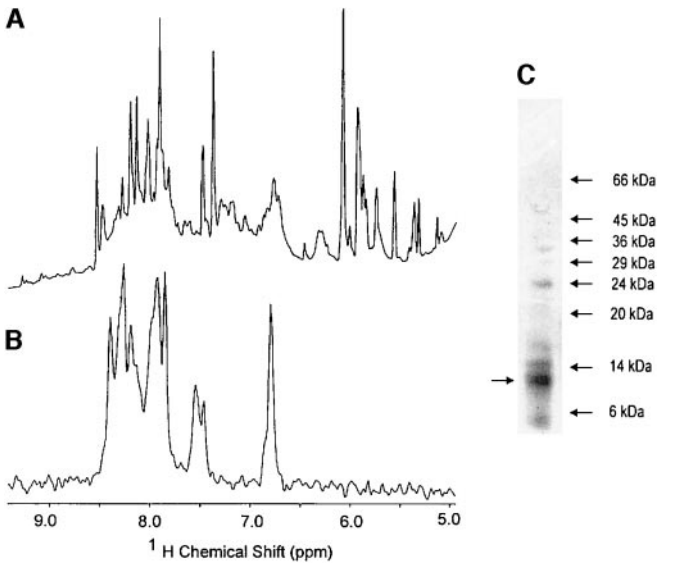

FIG. 3. ^{1}H spectrum (A) and ^{15}N-edited ^{1}H spectrum (B) of Opaque 2 LZ. The ^{15}N-edited ^{1}H spectrum shows low dispersion and is consistent with a leucine zipper (*6*). The proton spectrum shows several different peaks with higher dispersion. This spectrum was obtained with 250 ml of growth medium. The sample was prepared using the Amicon by passing 250 ml through a membrane of 30-kDa molecular cutoff and then concentrating it to 500 μl using a membrane of 3.5 kDa. C shows the SDS–PAGE of the rifampicin NMR sample used to obtain spectra A and B.

sequences. The investigation of these new proteins demands fast methodology for structural studies. Different methodologies have been proposed (*8, 9*). Selective labeling of the heterologous protein can speed up the screening of clones that leads to folded proteins for NMR structural studies.

HMQC spectrum analysis can also be used to monitor changes in protein structure upon drug addition. This will make rational drug design faster as the screening can be performed quickly.

We strongly believe that there might be cases where protein purification in high amounts is difficult and the structure determination could be elucidated by using the rifampicin protocol.

EXPERIMENTAL

NMR experiments. NMR spectra were made at 303 K in a Bruker Avance DRX 600 MHz, using 5-mm triple-resonance probes. Bruker XWINNMR software and NMRView (*10*) were used for data acquisition and processing and for data analysis, respectively. The sample preparation includes the addition of 10% D_2O for lock. All spectra were recorded at pH 7.0.

All reagents were of analytical grade purity.

Selective protein labeling: Rifampicin protocol samples. Escherichia coli BL21(DE3) with plasmids (pET) containing TRX or Opaque 2 LZ were grown in 100 ml of nonlabeled M9 growth medium (*11*) at 37°C and were shaken at 200 rpm up to mid-log phase (OD about 0.7 at 600 nm). At this point 1 mM of IPTG was added. This enabled the induction of T7 RNA polymerase. A small amount of unlabeled TRX or Opaque 2 LZ will be obtained. After 5 min, 200 μg/ml of rifampicin was added. The cells grew for 15 more min and then were centrifuged to pellet. This time interval is necessary in order to let rifampicin diffuse into the cell, inhibiting self synthesis while still in the unlabeled medium. The cells were then resuspended in 100 ml of ^{15}N-labeled M9 growth medium containing 1 mM IPTG and 200 μg/ml rifampicin. The induced cells grew for 2 to 3 h more and were cleft by freezing and thawing in cleavage buffer (1 ml TRX and 15 ml Opaque2 LZ per 100 ml of growth medium). The samples were then centrifuged and the superna-

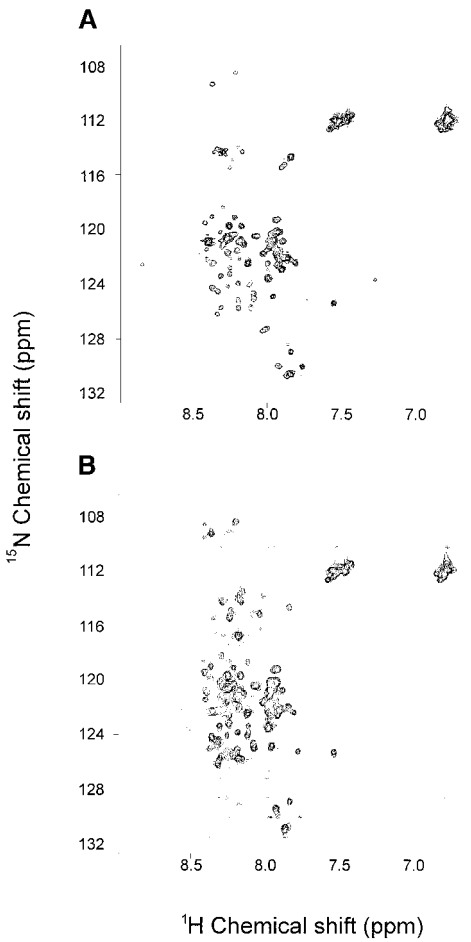

FIG. 4. Two-dimensional ^{15}N, ^{1}H-HMQC spectra of Opaque 2 LZ obtained with (A) and without (B) rifampicin. The 2D spectrum (A) shows 70 peaks while the primary sequence has 86 amino acids. The dispersion is very similar to the one obtained for GCN4 (*6*). The sample obtained without rifampicin shows several extra peaks. The sample was prepared as described in Fig. 3. All spectra were acquired with 1024 × 512 points, 64 scans, and a recycle delay of 1.3 s. The processing was performed with zero filling and a square sine multiplication shifted by 90° in both dimensions. The States–TPPI method was used for quadrature detection in the indirect dimension (*12*). GARP was used to decouple ^{15}N during acquisition (*13*) and WATERGATE for water supression (*14*).

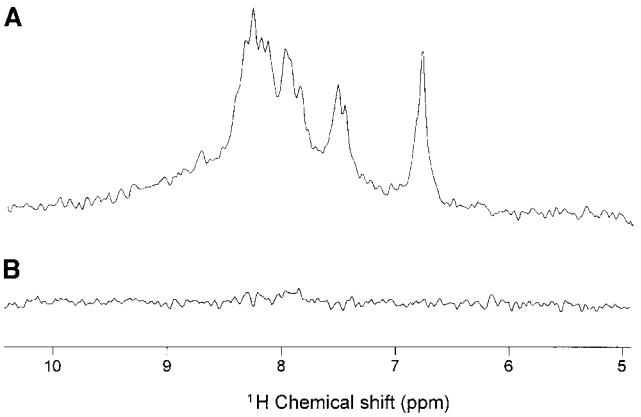

FIG. 5. ^{15}N-Edited ^{1}H spectrum of the sample retained in the 30-kDa cutoff ultrafiltration membrane prepared with (B) and without rifampicin (A). The spectra were obtained with the same number of scans of the material that is retained in the Amicon of the 30-kDa cutoff membrane (for details see Experimental). As expected, the sample prepared with rifampicin shows a ^{15}N signal only in the fraction where the Opaque 2 LZ is present (low molecular weight). The ^{1}H signal has the same intensity in the two samples (not shown).

tant containing the specifically labeled protein was used to perform NMR experiments. In the case of TRX, the sample was dialyzed in a membrane of molecular weight cutoff of 3.5 kDa to remove low-molecular-weight compounds (free amino acids). In the case of Opaque 2 LZ, a step for sample concentration was included and is described below.

Controls were made to probe the efficiency of the rifampicin protocol: bacterium cells grew always in labeled M9 medium. Briefly, the cells grew up to an OD (600 nm) of 0.7 when protein expression is induced with 1 mM IPTG, in the absence of rifampicin. The cells grew for 2 to 3 h more. Sample preparation is exactly the same as for samples made with the rifampicin protocol. These samples were called "totally labeled control."

The cleavage buffer used was 20 mM phosphate buffer, pH 7.0; 10 mM β-mercaptoethanol; 40 μg/ml PMSF.

Opaque 2 LZ sample preparation. The protein was expressed as described above. Cells grown in 250 ml of growth medium were cleaved with 50 ml of cleavage buffer (described above). The supernatant was ultrafiltrated at a molecular weight cutoff of 30 kDa. The flow-through was collected and applied to another ultrafiltration step, but now with a molecular weight cutoff of 3.5 kDa, in order to concentrate the sample and free it of low-molecular-weight compounds (such as peptides and free amino acids). The sample was concentrated to 500 μL and 100 mM of KCl was added.

Thioredoxin purification. The purified thioredoxin was obtained by cleaving the cells as described above, the supernatant was applied to a Toyopearl-650M column (ion exchange) at pH 8, and the fraction eluted with 50 mM NaCl was collected, dialyzed, and applied to the same column. TRX is eluted with 50 mM NaCl. Sample purity and expression were confirmed by SDS–PAGE.

ACKNOWLEDGMENTS

The authors thank PRONEX, CAPES, CNPq, FAPERJ, and ICGEB–International Center for Genetic Engineering and Biotechnology. We also thank Dr. C. Griesinger for helpful discussions.

REFERENCES

1. A. Travers, Control of transcription in bacteria, *Nature (London) N. Biol.* **229,** 69 (1971).
2. K. M. Lee, E. J. Androphy, and J. D. Baleja, A. novel method for selective isotope labeling of bacterially expressed proteins, *J. Biomol. NMR* **5,** 93–96 (1995).
3. W. W. Gao and E. Goldman, Use of sodium dodecyl sulfate–polyacrylamide gel eletrophoresis to resolve mRNA and its protein product in one gel, *FASEB J.* **11,** 1153–1156 (1997).
4. A. Holmgren, Thioredoxin, *Annu. Rev. Biochem.* **54,** 237–271 (1985).
5. A. L. Vettore, J. A. Yunes, G. C. Neto, M. J. Silva, P. Arruda, and A. Leite, The molecular and functional characterization of a Opaque2 homologue gene from Coix and a new classification of plant bZIP proteins, *Plant Mol. Biol.* **36,** 249–263 (1998).
6. C. Bracken, P. Carr, J. Cavanagh, and A. G. Palmer, Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA, *J. Mol. Biol.* **285,** 2133–2146 (1999), doi: 10.1006/jmbi.1998.2429.
7. A. M. Gronenborn and G. M. Clore, Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy, *Protein Sci.* **5,** 174–177 (1996).
8. D. T. Jones, GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences, *J. Mol. Biol.* **287,** 797–815 (1999), doi: 10.1006/jmbi.1999.2583.
9. C. C. Wang and C. L. Tsou, Post-genome study—Proteomics, *Acta Biochim. Biophys. Sin.* **30,** 533–539 (1998).
10. B. A. Johnson and R. A. Blevins, NMR View—A computer-program for the visualization and analysis of NMR data, *J. Biomol. NMR* **4,** 603–614 (1994).
11. J. Sambrook, E. F. Fritsch, and T. Maniatis, "Molecular Cloning—A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
12. D. Marion, M. Ikura, R. Tschudin, and A. Bax, Rapid recording of 2D NMR spectra without phase cycling-application to the study of hydrogen exchange in proteins, *J. Magn. Reson.* **85,** 393–399 (1989).
13. A. J. Shaka, P. B. Barker, and R. Freeman, Computer-optimized decoupling scheme for wideband applications and low-level operation, *J. Magn. Reson.* **64,** 547–552 (1985).
14. M. Piotto, V. Saudek, and V. Sklenar, Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous solutions, *J. Biomol. NMR* **2,** 661–665 (1992).