Protein Science (1997), 6:2359–2364. Cambridge University Press. Printed in the USA.

Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR

JEFFREY R. HUTH,[1] CAROLE A. BEWLEY,[1] BELINDA M. JACKSON,[2] ALAN G. HINNEBUSCH,[2] G. MARIUS CLORE,[1] AND ANGELA M. GRONENBORN[1]

[1]Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520
[2]Laboratory of Eukaryotic Gene Regulation, Building 6B, National Institute for Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892

(RECEIVED June 12, 1997; ACCEPTED July 25, 1997)

Abstract

Two protein expression vectors have been designed for the preparation of NMR samples. The vectors encode the immunoglobulin-binding domain of streptococcal protein G (GB1 domain) linked to the N-terminus of the desired proteins. This fusion strategy takes advantage of the small size, stable fold, and high bacterial expression capability of the GB1 domain to allow direct NMR spectroscopic analysis of the fusion protein by ^{1}H-^{15}N correlation spectroscopy. Using this system accelerates the initial assessment of protein NMR projects such that, in a matter of days, the solubility and stability of a protein can be determined. In addition, ^{15}N-labeling of peptides and their testing for DNA binding are facilitated. Several examples are presented that demonstrate the usefulness of this technique for screening protein/DNA complexes, as well as for probing ligand–receptor interactions, using ^{15}N-labeled GB1-peptide fusions and unlabeled target.

Keywords: domain structure; GBI-fusion protein; intermolecular interactions; NMR

When investigating whether a protein or protein–DNA complex is suitable for NMR structural analysis, several technical goals must be met. First, the size, solubility, and stability of the protein domain must be optimized. Second, the protein should be expressed in a host (usually bacteria), allowing for easy and inexpensive labeling with ^{13}C and ^{15}N. A third challenge pertains to the preparation of protein–DNA complexes where both the protein and DNA sequences must be optimized to achieve a complex that is most amenable to structural investigation; that is, the system should not exhibit intermediate exchange on the NMR time scale. The practical size limit for structure determination by current solution NMR methodology is approximately 40 kDa, so that, for many proteins, subcloning of a smaller domain that retains functional qualities of the full-length protein is required. All too frequently, however, engineering of protein domains results in limited solubility and/or low stability, often to the point where many engineered proteins are partially or predominately unfolded. In practice, many variations of a protein sequence may be evaluated by NMR before one emerges that exhibits the desired spectral qualities. Current methods of NMR sample screening require a significant experimental commitment and are both labor and time intensive, specifically, subcloning, optimization of bacterial expression, protein purification, and complex formation have to be performed (Clubb et al., 1994; Werner et al., 1995; Huth et al., 1997; Omichinski et al., 1997). Frequently, a refolding procedure may also be required (Huth et al., 1994; Werner et al., 1994) and, to obtain concentrated samples (1 mM), several buffer conditions have to be evaluated. With these factors in mind, we have developed two T7 RNA polymerase-based (Studier & Moffatt, 1986) expression vectors that are tailored specifically for NMR sample preparation and screening. In the strategy described here, the engineered protein is synthesized as a fusion with the B1 immunoglobulin binding domain of streptococcal protein G (GB1; 56 residues) (Gronenborn et al., 1991). As is true for other N-terminal fusion proteins, such as those containing domains from glutathione S-transferase (Smith & Johnson, 1988), maltose binding protein (di Guan et al., 1988), protein A (Nilsson et al., 1985; Jansson et al., 1996), or thioredoxin (LaVallie et al., 1993), the presence of the GB1 domain results in very high expression levels of the fusion protein and limits the need to optimize expression conditions. For most fusion protein-based expression strategies, the N-terminal domain is rather large and is removed by proteolysis prior to NMR analysis. This com-

Reprint requests to: G. Marius Clore or Angela M. Gronenborn, Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520; e-mail: clore@vger.niddk.nih.gov or gronenborn@vger.niddk.nih.gov.

plicates the screening procedure because protocols for digestion and purification of the cleaved products can vary depending on the sample. In the case of the GB1 expression system, we show that, owing to the small size of GB1, it is readily possible to perform the spectroscopic analysis on the GB1 fusion proteins. Using this approach, we demonstrate that one can overexpress, purify, and acquire a ^{1}H-^{15}N correlation spectrum of a labeled protein in several days, and thus rapidly evaluate its solubility, stability, and binding characteristics.

Results and discussion

The design of the GB1 fusion protein expression vectors is outlined in Figure 1. The first vector (GEV1) encodes a noncleavable linker between the GB1 domain and the desired protein, whereas the second vector (GEV2) contains a thrombin-cleavable linker as well as several more 3′ cloning sites. Both vectors encode an optional poly His-tag at the C terminus of the fusion protein to allow purification using Ni^{2+} affinity chromatography under nondenaturing conditions.

On its own, the GB1 domain expresses at very high levels in *Escherichia coli* and ~3 μmol of ^{15}N-labeled protein can be purified from a 100-mL culture in minimal medium (A.M. Gronenborn, unpubl. results). We have found that GB1 fusion proteins also express at high levels such that 0.1–1-L cultures are typically grown for sample screening. Figure 2 shows an SDS-PAGE analysis of GB1 fusions with either the activation domain of yeast GCN4 (residues 39–139) (Jackson et al., 1996) or the death domain of FAS (residues 210–304) (Suda et al., 1993). The GB1-GCN4 fusion protein was purified by Ni^{2+} affinity chromatography, whereas the GB1-Fas fusion protein was purified by affinity chromatography using IgG-sepharose. In both cases, a single-step procedure yielded 80–90% pure protein, sufficient for analysis by 2D ^{1}H-^{15}N correlation spectroscopy. In fact, purification may not be required for the initial screening because direct analysis of a ^{1}H-^{15}N correlation spectrum recorded on whole-cell extracts, after removal of the cell debris, is possible for small- to medium-sized (up to ~20 kDa) soluble proteins (Gronenborn & Clore, 1996).

The ^{1}H-^{15}N correlation spectra of the GB1-GCN4 and GB1-Fas fusion proteins are presented in Figures 3 and 4, respectively, and demonstrate the simplicity of this approach for evaluating the stability and solubility of these proteins. A qualitative assessment of the spectra permits the presence of regular secondary structure elements to be evaluated on the basis of the chemical-shift dispersion of backbone and side-chain amide groups. Random coil amide protons typically resonate between 7.8 and 8.4 ppm, whereas the chemical shifts of hydrogen bonded amides in β-sheets and α-helices frequently lie outside of this region. Because of its small size, the GB1 fusion domain does not interfere significantly with the detection of amide chemical shifts of the expressed partner protein, but rather, acts as an internal control for sample preparation and instrument operation. Note that, with the exception of the crowded,

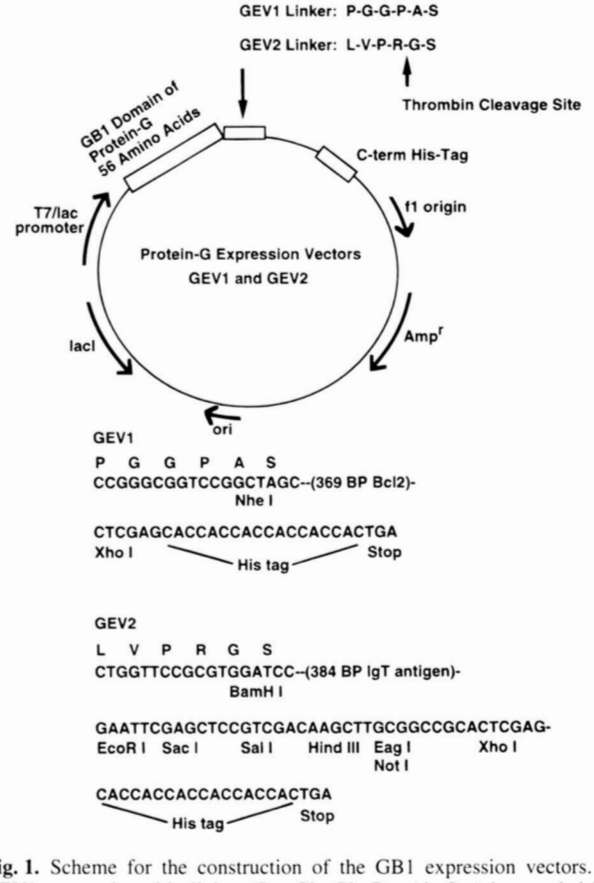

Fig. 1. Scheme for the construction of the GB1 expression vectors. In GEV1, a noncleavable linker (Pro-Gly-Gly-Pro-Ala-Ser) is encoded between the GB1 domain and the N-terminus of the protein of interest. Genes can be inserted into this vector using *Nhe* I and *Xho* I cloning sites, and, if no stop codons are introduced at the 5′ end of an insert, the vector encodes a C-terminal poly His-tag. The GEV2 vector contains a linker with a thrombin cleavage site. Genes are inserted between a *Bam*H I site at the 5′ end and several possible restriction sites at the 3′ end. As for GEV1, cloning into the *Xho* I site allows incorporation of a poly His-tag.

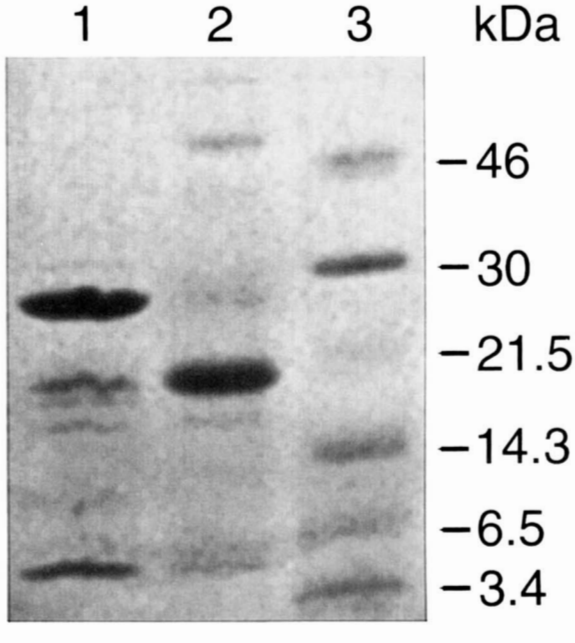

Fig. 2. SDS-PAGE separation of GB1-GCN4 (39–139) following purification by Ni^{2+} affinity chromatography (lane 1) and GB1-Fas (210–304) purified by IgG affinity chromatography (lane 2). Molecular weight markers (lane 3) include ovalbumin (46 kDa), carbonic anhydrase (30 kDa), trypsin inhibitor (21.5 kDa), lysozyme (14.3 kDa), aprotinin (6.5 kDa), and insulin B chain (3 kDa). Note that the GB1-GCN4 fusion runs at an anomalous high molecular weight; mass spectroscopic analysis, however, confirmed the molecular weight to be 19.3 kDa.

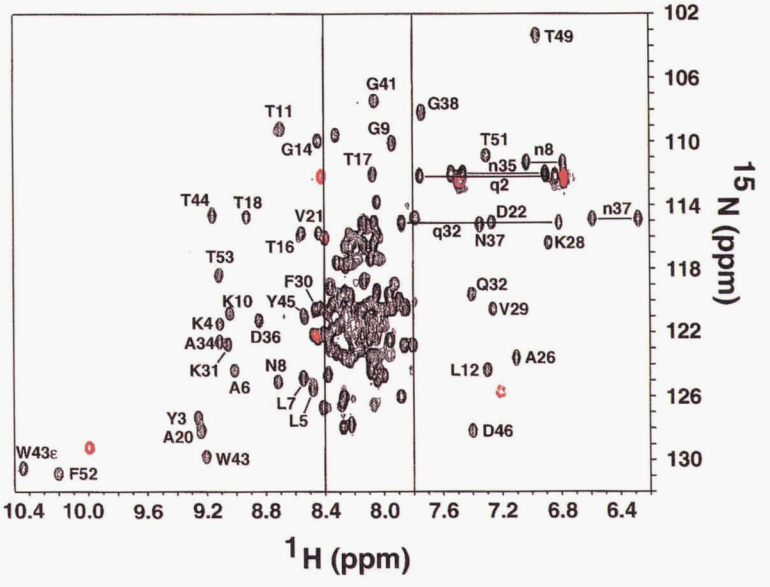

Fig. 3. ^{1}H-^{15}N HSQC spectrum of GB1-GCN4 (39–139) recorded at 600 MHz with total acquisition times of 64 and 102 ms in the acquisition and indirect dimensions, respectively. Eight scans were recorded per t_1 increment. GB1 resonances that lie outside of the random coil region (7.8–8.4 ppm) are labeled and resonances originating from GCN4 amino acids are colored in red. Side-chain amino groups of GB1 are indicated by lowercase letters.

random-coil region between 7.8 and 8.4 ppm, all the GB1 signals in the GB1-GCN4 fusion protein are easily assigned in the 2D ^{1}H-^{15}N correlation spectrum by comparison with assignments reported previously (Barchi et al., 1994). A 3D ^{15}N-edited NOE spectrum of a GB1-HMG-I fusion supports these assignments (data not shown). Simple inspection of the remainder of the GB1-GCN4 spectrum (Fig. 3) indicates that the activation domain of GCN4 contains little, if any, stable secondary structure, and, therefore, is in all likelihood unfolded. Apart from the sole tryptophan indole NH at 10.0 ppm and the side-chain amide protons (shown in red), only two resonances from GCN4 lie outside of the random coil region of the spectrum, albeit very close to the border. Further, the dispersion of side-chain amide protons from the three glutamine and four asparagine residues is very poor, also suggesting a random coil conformation. Thus, this simple screen indicates that the GCN4 activation domain requires more amino acids or, more likely,

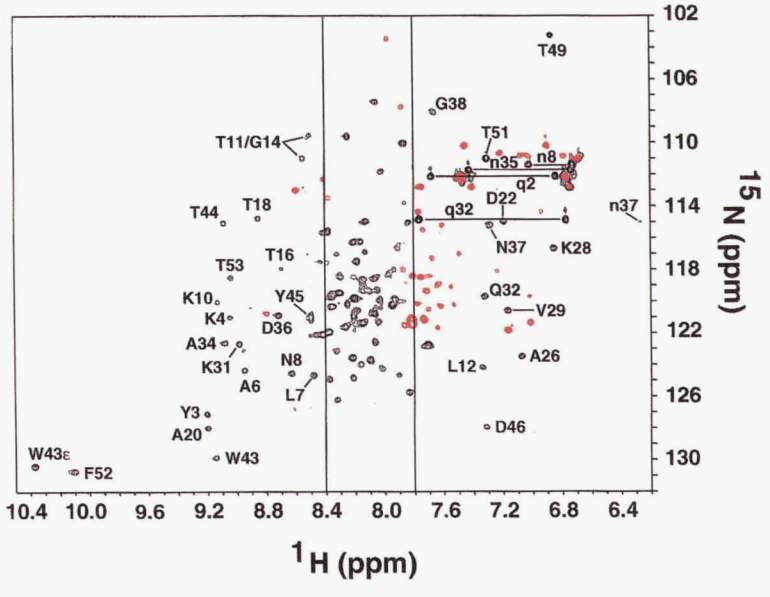

Fig. 4. ^{1}H-^{15}N HSQC spectrum of GB1-Fas (210–304) recorded at 600 MHz with total acquisition times of 64 and 56 ms in the acquisition and indirect dimensions, respectively. Eight scans were recorded per t_1 increment. Of the signals that lie outside of the random coil region, those from the GB1 domain are black and those from Fas are colored in red. Side-chain amino groups of GB1 are indicated by lowercase letters.

its binding partner for stabilization of the folded structure. On the other hand, the ^{1}H-^{15}N correlation spectrum of the GB1-Fas death domain fusion protein (Fig. 4) indicates that the domain comprising amino acids 210–304 of Fas is indeed folded. Eighteen signals between 6.9 and 7.8 ppm and three signals downfield of 8.4 ppm were observed, and the side-chain amide proton resonances are dispersed. The difficulty with this sample is due to aggregation, as evidenced by line broadening of both Fas and GB1 amide resonances (cf. the GB1 signals in Figs. 3 and 4).

From the two examples presented above, the screening process immediately suggests the appropriate course of action for optimizing the NMR samples. In the case of the activation domain of GCN4, this requires the difficult task of obtaining a complex consisting of the fusion protein and the interacting domain from a target or identifying the limits of a domain that maintains a stable fold. In the case of Fas, either solubility conditions would need to be explored for an already folded domain or the boundaries of the domain would have to be extended slightly to enhance solubility. The recent publication of the solution structure of the Fas death domain (Huang et al., 1996) confirms our results presented above, namely that residues 210–305 do indeed delineate the folded domain. However, the construct employed by Huang et al. (1996) was slightly longer, comprising residues 202–319, which presumably increased its solubility. In addition, the structure of the Fas death domain was determined at pH 4.0 in 50 mM sodium acetate and 50 mM NaCl, which may also have enhanced the solubility, whereas our spectrum was recorded in the absence of any salt at pH 4.3.

The application of the GB1-fusion strategy for exploring conditions for the formation of a peptide–DNA complex demonstrates another strength of this approach. Figure 5 illustrates an experiment to evaluate the binding of a 25-residue peptide to a target DNA site. In this case, the second DNA-binding domain (DBD2, residues 50–75) of HMG-I as the GB1 fusion was ^{15}N-labeled and purified by IgG-sepharose affinity chromatography (~2 μmol/L culture). As is evident from the spectrum of the GB1-HMG-I fusion protein alone, only GB1 resonances are detected outside of the random coil region of the ^{1}H-^{15}N correlation spectrum (Fig. 5A). Upon addition, however, of the DNA dodecamer comprising the PRDII region of the interferon-β gene enhancer (Thanos & Maniatis, 1992; Fig. 5B), HMG-I amide resonances appear outside of the random coil region at positions previously assigned for a larger HMG-I/DNA complex (Huth et al., 1997). Interestingly, GB1 resonances are not changed significantly, indicating that the GB1 portion of the fusion protein does not interact with the DNA. From these data, it is evident that the HMG-I DBD2 binds on its own to the PRDII target site, that the complex is not in intermediate exchange, and that the structure of HMG-I DBD2 bound to this DNA dodecamer could be solved by NMR.

In summary, the GB1 fusion approach presented here represents a valuable tool for evaluating potential structural projects by NMR spectroscopy. Because the GB1 fusion domain does not, in any significant manner, interfere with substrate binding, this approach may be particularly useful for initial evaluations of protein–DNA, protein–protein, or protein–drug complexes. The advantages of the GB1 expression system, as well as the ability to label peptides with ^{15}N as GB1 fusions, facilitates the application of NMR in mapping macromolecular interactions. For example, a series of ^{15}N-labeled GB1-peptides spanning the sequence of a protein ligand could be prepared using the GEV expression vectors. Analysis of the ^{1}H-^{15}N correlation spectra of these labeled fusion peptides in the presence and absence of unlabeled receptor would detect those amino acids of the ligand that contact the receptor without prior knowledge of the ligand three-dimensional structure.

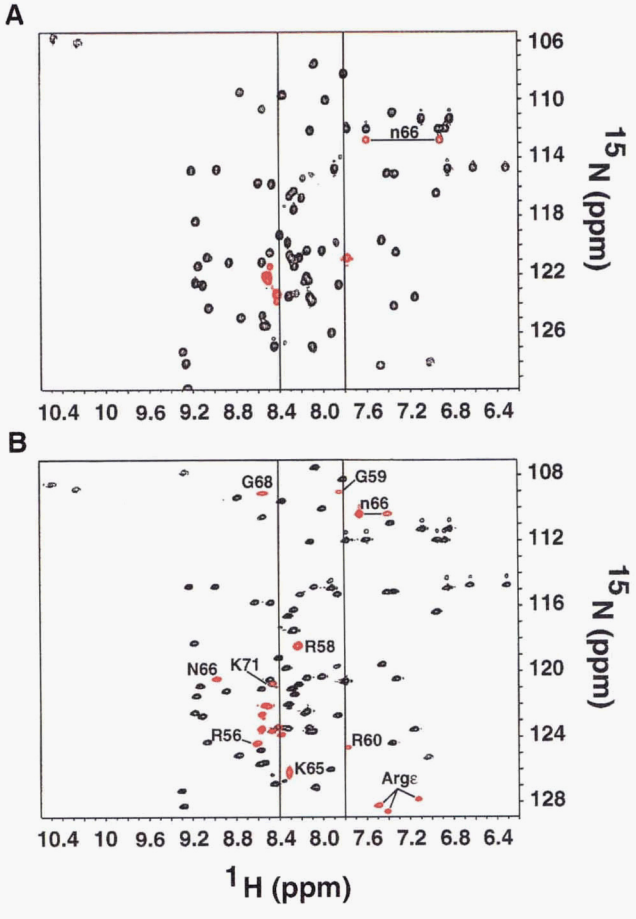

Fig. 5. ^{1}H-^{15}N HSQC spectra of GB1-HMG-I (50–75) in the absence (**A**) and presence (**B**) of DNA. Spectrum A was recorded at 600 MHz with total acquisition times of 64 and 51 ms in the acquisition and indirect dimensions, and eight scans per t_1 increment. Spectrum B was recorded at 750 MHz with total acquisition times of 43 and 77 ms in the acquisition and indirect dimensions, and eight scans per t_1 increment. The latter contained an equimolar amount of the double-stranded dodecamer 5′d(GGG AAATTCCTC)·5′d(GAGGAATTTCCC). Signals from HMG-I (50–75) are colored in red and in B are assigned based on previous assignments of an HMG-I(50–91)/DNA complex using the same DNA dodecamer (Huth et al., 1997). Resonances corresponding to the side-chain amino protons of Asn 66 for both DNA-free and bound GB1-HMG-I (50–75) are labeled by lowercase letters. Note that the chemical-shift changes for the Asn 66 side-chain amino protons are indicative of the interaction of this functional group with the DNA. Note that the two cross-peaks at 10.2–10.4 ppm in the ^{1}H dimension are folded in the ^{15}N dimension and arise from W43(NG) and F52(NH) of GB1.

Materials and methods

Preparation of the protein G fusion expression vectors (GEV and GEV2)

Coding sequences for the IgG binding domain of protein G with the amino acid changes T2Q, which eliminates methionine heterogeneity at the N-terminus from bacterial expression, and I6A (Smith et al., 1994) was PCR amplified using the primers 5′-ATAATT

AAACATATGCAGTACAAGCTTGCT (primer 1) and 5′-GCTAG CCGGACCGCCCGGTTCGGTTACCGTGAAGGT (primer 2) to create sequences encoding GB1 fused to the protease-resistant linker Pro-Gly-Gly-Pro-Ala-Ser. In a second PCR reaction, this linker was encoded on the 5′ end of the sequence for huBCl2(82–204) using the primers 5′-CCGGGCGGTCCGGCTAGCGCGCTCAG CCCTGTGCCA (primer 3) and 5′-AATAAACTCGAGTTACTT TTTTCGCATGCTGGGGCCATATAG (primer 4). These PCR products were mixed and used as a template to amplify the sequence for the GB1-linker-BCl2(82–204) using primers 1 and 4 in a third PCR reaction. The final PCR product was then cloned into pET21a (Novagen) using *Nde* I and *Xho* I cloning sites. The same strategy was used to prepare the GEV2 vector. Here the sequence for the GB1 domain was fused to the thrombin-cleavable linker Leu-Val-Pro-Arg-Gly-Ser using primer 1 and 5′-GGATCCACGC GGAACCAGTTCGGTTACCGTGAAGGT (primer 5). The initial insert for placement between the peptide linker and the C-terminal His-tag consisted of the coding sequence for residues 131–259 of the SV40 large T-antigen (IgT) and was amplified using 5′-CTG GTTCCGCGTGGATCCAAGGTAGAAGACCCCAAGGACTTT (primer 6) and 5′-CGTCA-GGAATTCTTAAGGATTAAAATCAT GCTCCTTTAA (primer 7). The GB1-linker2-IgT(131–259) fragment was amplified with primers 1 and 7 and cloned into pET21a using *Nde* I and *Eco*R I restriction sites.

Preparation of GCN4, Fas, and HMG-I expression vectors

DNA encoding residues 39–139 of GCN4 from yeast *Saccharomyces cerevisiae* was amplified by PCR using the primers 5′-TCTACTGCCAAACCAAGATCTATGGTTGGCCAATTG and 5′-GGATTCAATTGCCTTCTCGAGATCAGCCAATGAAAC ATC and cloned into the *Bam*H I and *Xho* I sites of GEV2 using *Bgl* II to create a compatible end for the *Bam*H I site. DNA encoding residues 210–304 of human Fas (the death domain) was PCR amplified using the primers 5′-GCGGGATCCGATGTTGA CTTGAGTAAA and 5′-CGTCCTCTCGAGTTAACTAGTAATG TCCTTGAG and cloned into GEV2 (Fig. 1) using 5′ *Bam*H I and 3′ *Xho* I sites, removing the IgT insert. A stop codon was introduced to prevent the addition of the poly-His sequence at the C-terminus. DNA for amino acids 50–91 of human HMG-I and a stop codon was amplified using the primers 5′-GCTCGTGGAT CCGAAGTGCCAACACCTAAG and 5′-CTGGCTGAATTCTT AGGTTTTCCGGGTCTTGGC and cloned into GEV2 (Fig. 1) using 5′ *Bam*H I and 3′ *Eco*R I sites.

Expression and ^{15}N-labeling of fusion proteins

BL21(DE3) cells (Novagen) were transformed with the respective vectors and grown to saturation ($OD_{600} \sim 4$) in LB medium that contained 50 μg/mL ampicillin. Cultures were diluted fourfold with medium, shaken for 10 min at 37 °C, and then induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) for 4 h. To determine the extent of soluble expression, 100-mL cultures were pelleted, opened in a French pressure cell, cleared by centrifugation, and analyzed by SDS-PAGE using the Phast-Kast electrophoresis system (Pharmacia). For ^{15}N labeling, glycerol stocks of cells in LB medium were grown in minimal medium with 1.1 g ^{15}N-ammonium chloride as the nitrogen source, 5 g glucose, and 25 μL of a 10% solution of yeast extract per liter of culture. Glycerol stocks were prepared from this growth and used to inoculate cultures that were grown in minimal medium for 14 h to OD_{600} 3.5–4. These cultures were diluted fourfold with minimal medium that lacked glucose and yeast extract, shaken at 37 °C for 30 min, supplemented with glucose, and induced with 1 mM IPTG, and grown for 4 h more at 37 °C prior to harvesting.

Purification of fusion proteins

BL21(DE3) cells that expressed GB1-Fas or GB1-HMG-I were broken in a French pressure cell in 50 mM Tris, pH 7.5, 5 mM EDTA, and 5 mM benzamidine. After brief sonication and centrifugation, the supernatant was filtered through a 0.45-μm membrane prior to loading onto a 20-mL IgG sepharose-fast flow column (Pharmacia), equilibrated in 50 mM Tris, pH 7.5, 150 mM NaCl. The column was washed with 50 mM Tris, pH 7.5, followed by 5 mM sodium acetate buffer, pH 5.0, and proteins were eluted in 0.5 M sodium acetate, pH 3.5. Proteins were concentrated by Amicon filtration, dialyzed against water, and further concentrated. NMR samples of GB1-HMG-I were concentrated to 1.2 mM in 10 mM sodium phosphate, pH 6.1, 90% H_2O/10% D_2O. The GB1-Fas sample was concentrated to ~1 mM in 90% H_2O/10% D_2O, pH 4.3. GB1-GCN4 was purified by Ni^{2+} affinity chromatography (Pharmacia) according to the manufacturer's instructions, dialyzed against water, and concentrated to ~1 mM using centricon-3 microconcentrators (Amicon) in 90% H_2O/10% D_2O, pH 6.2.

NMR spectroscopy

NMR experiments were performed on Bruker DMX600 and DMX750 spectrometers equipped with x,y,z-shielded gradient triple resonance probes. Spectra were processed with the NmrPipe package (Delaglio et al., 1995) and analyzed using the program PIPP (Garrett et al., 1991). The ^{1}H-^{15}N HSQC spectra (Bax et al., 1990; Piotto et al., 1992) were recorded with 128* × 512* complex points in the indirect (^{15}N) and acquisition (^{1}H) dimensions, respectively. Details of the spectroscopy are provided in the figure legends.

Acknowledgments

We thank Charles Zacharchuk, Lynne Regan, Raymond Reeves, and Ellen Fanning for providing clones used in this work. This work was supported in part by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (G.M.C. and A.M.G.).

References

Barchi JJ, Grasberger B, Gronenborn AM, Clore GM. 1994. Investigation of the backbone dynamics of the IgG-binding domain of streptococcal protein G by heteronuclear two-dimensional ^{1}H-^{15}N nuclear magnetic resonance spectroscopy. *Protein Sci 3*:15–21.

Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R. 1990. Comparison of different modes of two-dimensional reverse correlation NMR for the study of proteins. *J Magn Reson 86*:304–318.

Clubb RT, Omichinski JG, Clore GM, Gronenborn AM. 1994. Mapping the binding surface of interleukin-8 complexes with an N-terminal fragment of the type 1 human interleukin-8 receptor. *FEBS Lett 338*:93–97.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR 6*:277–293.

di Guan C, Li P, Riggs PD, Inouye H. 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene 67*:21–30.

Garrett DS, Powers R, Gronenborn AM, Clore GM. 1991. A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson 95*:214–220.

Gronenborn AM, Clore GM. 1996. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci 5*:174–177.

Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science 253*:657–661.

Jackson BM, Drysdale CM, Natarajan K, Hinnebusch AG. 1996. Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. *Mol Cell Biol 16*:5557–5571.

Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. 1996. High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli*. *J Biomol NMR 7*:131–141.

Huang B, Eberstadt M, Olejniczak ET, Meadows RP, Fesik SW. 1996. NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain. *Nature 384*:638–641.

Huth JR, Bewley CA, Nissen MS, Evans JNS, Reeves R, Gronenborn AM, Clore GM. 1997. The solution structure of an HMG-I(Y)/DNA complex defines a new architectural minor groove binding motif. *Nature Struct Biol 4*:657–665.

Huth JR, Norton SE, Lockridge O, Shikone T, Hsueh AJW, Ruddon RW. 1994. Bacterial expression and in vitro folding of the β-subunit of human chorionic gonadotropin (hCGβ) and functional assembly of recombinant hCGβ with hCGα. *Endocrinology 135*:911–918.

LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/Technology 11*:187–193.

Nilsson B, Abrahmsén L, Uhlén M. 1985. Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors. *EMBO J 4*:1075–1080.

Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM, Clore GM. 1997. The solution structure of a specific GAGA factor–DNA complex reveals a modular binding mode. *Nature Struct Biol 4*:122–132.

Piotto M, Saudek V, Sklenar V. 1992. Gradient-tailored excitation for single quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR 2*:661–665.

Smith DB, Johnson KS. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene 67*:31–40.

Smith CK, Withka JM, Regan L. 1994. A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry 33*:5510–5517.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol 189*:113–130.

Suda T, Takahashi T, Golstein P, Nagata S. 1993. Molecular cloning and expression of the Fas ligand, a novel member of the tumor necrosis factor family. *Cell 75*:1169–1178.

Thanos D, Maniatis T. 1992. The high mobility group protein HMG I(Y) is required for NF-β-dependent virus induction of the human IFN-β gene. *Cell 71*:777–789.

Werner MH, Clore GM, Gronenborn AM, Kondoh A, Fisher RJ. 1994. Refolding proteins by gel filtration chromatography. *FEBS Lett 345*:125–130.

Werner MH, Huth JR, Gronenborn AM, Clore GM. 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell 81*:705–714.