Protein Science (1996), 5:174–177. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy

ANGELA M. GRONENBORN AND G. MARIUS CLORE

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

(RECEIVED October 17, 1995; ACCEPTED November 2, 1995)

Abstract: A simple and rapid method based on ^{15}N labeling and ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy is presented to directly assess the structural integrity of overexpressed proteins in crude *Escherichia coli* extracts without the need for any purification. The method is demonstrated using two different expression systems and two different proteins, the B1 immunoglobulin-binding domain of streptococcal protein G (56 residues) and human interleukin-1β (153 residues). It is shown that high quality ^{1}H-^{15}N correlation spectra, recorded in as little as 15 min and displaying only cross-peaks arising from the overexpressed protein of interest, can be obtained from crude *E. coli* extracts.

Keywords: ^{1}H-^{15}N HSQC spectroscopy; interleukin-1β; protein GB1 domain; structural screening

Rapid developments in recombinant DNA techniques have resulted in the identification and isolation of many novel genes. This invariably results in the need to express the gene product or parts and mutants thereof for structure–function studies, in vivo testing, or in vitro reagent production. A large variety of systems have been created for expression purposes and it is frequently possible to proceed from gene sequence to expression within days or weeks. Traditionally, the second step in any study involves the purification of the protein in question or its variants to assess their structural integrity prior to advancing to functional tests. The only two methods that permit structural assessment at the atomic level are X-ray crystallography and NMR spectroscopy, both of which require purified protein and often a considerable commitment in time and effort. In this report, we present a novel and rapid procedure for the identification of folded, structurally intact overexpressed protein directly out of crude *Escherichia coli* extracts using 2D ^{1}H-^{15}N heteronuclear single quantum coherence NMR spectroscopy.

Results and discussion: Figure 1 shows an SDS-polyacrylamide gel of crude extracts of *E. coli* cells harboring either an expression plasmid for the T2Q, I6A double mutant of the B1 immunoglobulin binding domain of streptococcal protein G (GB1; 56 residues), or an expression plasmid for human interleukin-1β (IL-1β; 153 residues). No purification of the two proteins was carried out, although, as can be appreciated from the gel, both proteins are expressed at a level corresponding to approximately 15–25% of the total cellular protein. Cells were harvested by centrifugation and disrupted by passage through a French pressure cell. The cell debris was removed by centrifugation and the supernatant was concentrated, with a concomitant buffer change, and directly subjected to NMR spectroscopy. Figure 2 presents the resulting 2D ^{1}H-^{15}N HSQC correlation spectra for the two cell extracts. The spectrum of the GB1-containing cell extract was recorded in 15 min, whereas that for the IL-1β-containing extract was recorded in 1 h. It is strikingly obvious that only signals from the expressed proteins are visible. None of the other cellular proteins that are present in the extract (cf. Fig. 1) result in visible cross-peaks. Thus, in the case of the GB1-containing extract (Fig. 2B), there are a total of 65 resolved cross-peaks, 55 arising from backbone amides, and 10 from the side-chain amino groups of three asparagine and two glutamine residues. In addition, the samples contain large amounts of nucleic acids, none of which yield any signals (despite the fact that the UV spectra exhibit their maxima at 260 nm).

The present samples were prepared with ^{15}N labeling of proteins in minimal medium using $^{15}NH_4Cl$ as the sole nitrogen source throughout the growth of the cells. For proteins that are expressed at lower levels (5–10% of the total cellular protein), it is possible to use a pulse-labeling approach in which $^{14}NH_4Cl$ is used as the nitrogen source in the initial growth of the cells, followed by a change in growth medium to one that contains

Reprint requests to: A.M. Gronenborn or G.M. Clore, 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: gronenborn@vger.niddk.nih.gov or clore@vger.niddk.nih.gov.

Abbreviations: GB1, B1 immunoglobulin binding domain of streptococcal protein G; IL1-β, interleukin-1β; HSQC, heteronuclear single quantum coherence; IPTG, isopropyl β-D-thiogalactopyranoside.

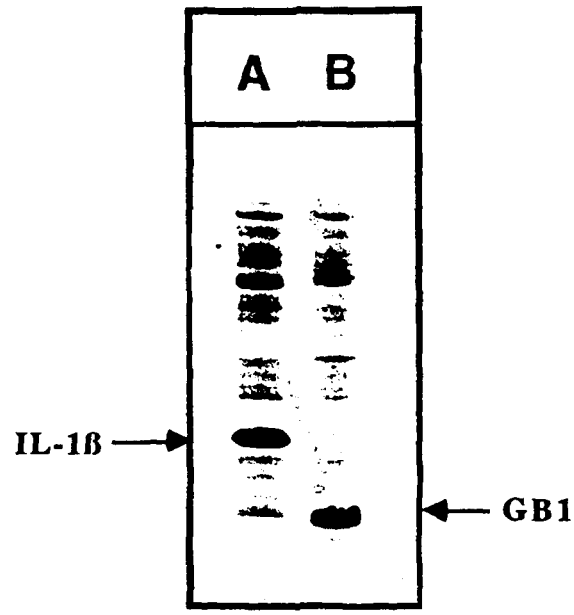

Fig. 1. SDS-polyacrylamide gel of the crude cell extracts used for NMR of **(A)** IL-1β and **(B)** GB1 (T2Q; I6A) expressing *E. coli* cells. The ^1H-^{15}N HSQC spectra of these extracts are shown in Figure 2.

$^{15}NH_4Cl$ as the sole nitrogen source just prior to the induction step. This has been verified on another mutant of GB1, which only expresses at the 5–10% level (data not shown). In this manner, all cellular proteins will contain the NMR-inactive ^{14}N isotope, and the desired, expressed protein will have incorporated the NMR active ^{15}N isotope.

The ^1H-^{15}N HSQC spectra shown in Figure 2 are essentially indistinguishable from those recorded for the corresponding purified proteins published previously for wild-type GB1, with the exception of a few shifts arising from the presence of the double T2Q, I6A mutation (Gronenborn & Clore, 1993), and for IL-1β (Driscoll et al., 1990), demonstrating that these two protein structures are not perturbed by other cellular proteins or the nucleic acids in the crude extract. This is a notable observation given the fact that chemical shifts are exquisitely sensitive to small structural or electronic changes. In addition, in the case of the GB1-containing extract, the concentration of GB1 is sufficient to permit 3D ^{15}N-separated spectra to be recorded with ease for the purposes of resonance assignment and secondary structure determination (Clore & Gronenborn, 1991).

To address the question of generality of the above approach, two different expression systems were used for GB1 and IL-1β. GB1 was expressed using the T7 system with IPTG induction (Studier et al., 1990), whereas IL-1β was expressed using the λ P_l promoter (Remaut et al., 1981) from a plasmid carrying the temperature-sensitive CI 857 repressor with heat induction at 42 °C. No difference was found for both systems, indicating that the choice of expression system is irrelevant for the final outcome of the method. It is likely that the methodology outlined above will also be applicable to the expression of proteins using a secretion system.

Why does this simple approach work so well? Two factors come into play. First, proteins above ~40 kDa or smaller proteins that are part of multiprotein complexes larger than ~40 kDa will be extensively line broadened owing to their slow rotational correlational time and hence will not be detectable in the ^1H-^{15}N HSQC spectrum. Second, the signal-to-noise ratio is directly proportional to concentration. Hence, proteins in the crude extract that are a factor of 10 lower in concentration than the overexpressed protein will not be detectable either. Thus, for example, it is readily appreciated from the SDS polyacrylamide gels in Figure 1 that the concentration of individual proteins below about 40 kDa is at least 20-fold less than that of the two overexpressed proteins.

Naturally, some limitations may be encountered regarding the general applicability of the technique. First, it is only suitable for proteins that are expressed in soluble form. Second, it may not be suitable in those cases where the structural integrity of the protein under consideration is very sensitive to solution conditions, because such a protein could potentially be unfolded in the unpurified extract, but folded in conditions only determined by screening purified protein in defined buffers.

We believe the power of the NMR screening approach presented in this report lies in its speed and simplicity. In particular, it is possible to assess the structural integrity of mutants within a day. For example, we were able to screen for the overall fold of a substantial number of core GB1 mutants within a day of the mutagenesis (A.M. Gronenborn & G.M. Clore, in prep.).

Experimental: *Sample preparation:* GB1 (T2Q, I6A double mutant) (Smith et al., 1994) was expressed from the plasmid pET11 (Novagen) in strain HMS174. A 50-mL culture of transformed cells was grown in minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source to an OD_{600} of approximately 1.6, induced with 1 mM IPTG, and grown for a further 3 h at 37 °C. Cells were harvested by centrifugation, taken up in 20 mL phosphate-buffered saline, pH 7.4, and passed twice through a French pressure cell. Cell debris was removed by centrifugation and supernatant was concentrated in a centriprep 3 device (Amicon) at 4 °C to a volume of about 0.8 mL. At the same time, the buffer was changed to 50 mM sodium phosphate, pH 5.4; 0.45 mL of the concentrated extract was used for NMR after the addition of 50 μL of D_2O.

IL-1β was expressed from a plasmid carrying the λ P_l promoter (Remaut et al., 1981) and the temperature-sensitive CI 857 repressor in strain RB791. A 50-mL culture of transformed cells was grown in minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source at 30 °C to an OD_{600} of approximately 0.5, induced by raising the temperature to 42 °C, and grown for a further 2.5 h at 42 °C. Cells were treated exactly as described above for the GB1-expressing cells, except that a centriprep 10 concentrator was used and the final buffer comprised 100 mM sodium acetate-d_3, pH 5.4, in a final volume of approximately 1.6 mL.

NMR spectroscopy: The ^1H-^{15}N HSQC spectra (Bax et al., 1990; Piotto et al., 1992) were recorded at 25 °C (GB1) and 35 °C (IL-1β) on a Bruker AMX600 spectrometer equipped with a triple resonance z-shielded gradient probe. Spectra were recorded with 128* × 512* complex points in the indirect (^{15}N) and acquisition (^1H) dimensions, respectively, and with total acquisition times of 64 ms in both dimensions. Spectra of the GB1- and IL-1β-containing crude *E. coli* extracts were acquired with 4 and 16 scans per t_1 increment, respectively.

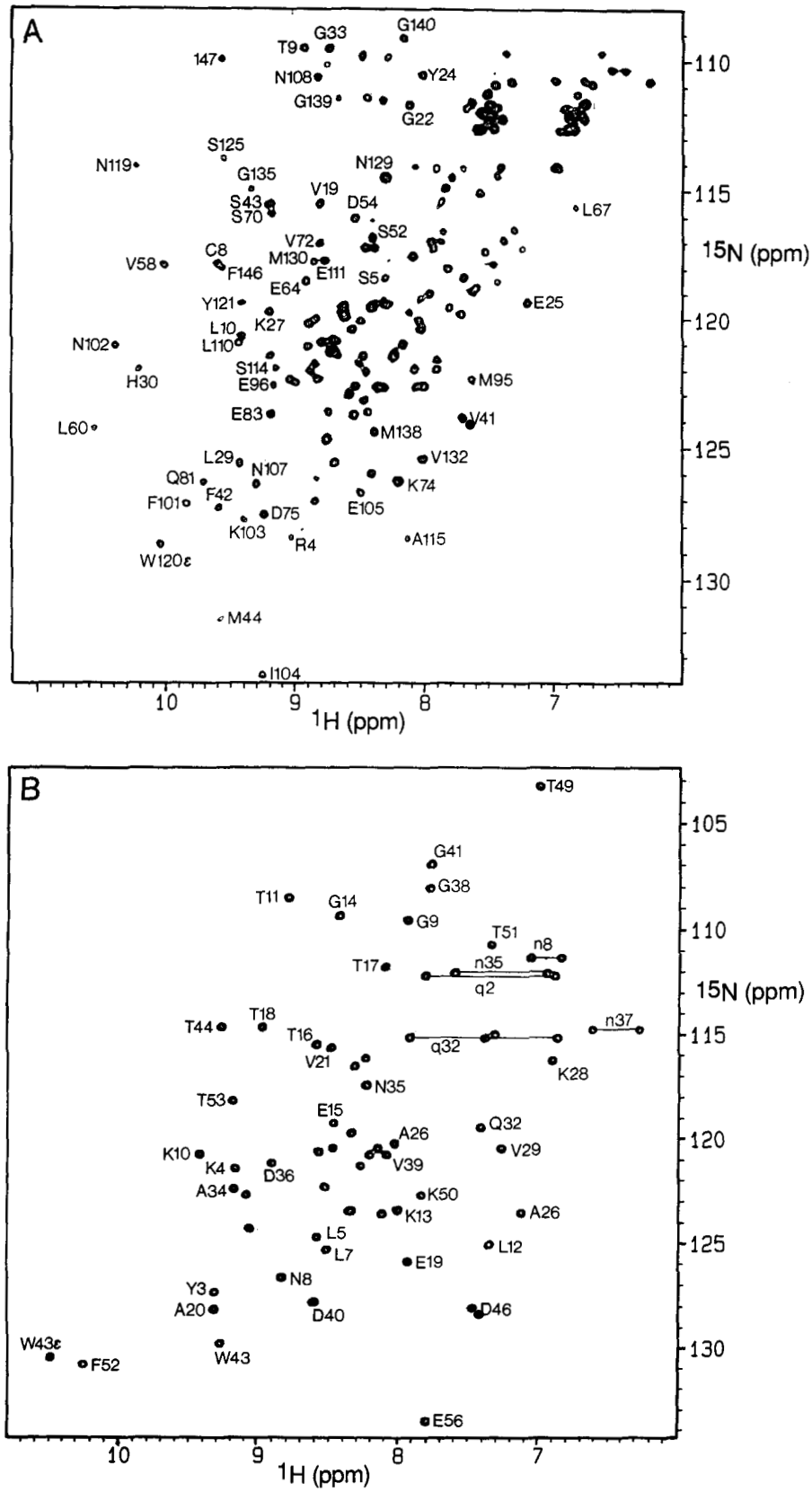

Fig. 2. ^{1}H-^{15}N HSQC spectra of crude *E. coli* cell extracts containing expressed **(A)** IL-1β and **(B)** GB1 (T2Q, I6A double mutant). The spectrum in A was recorded at 35 °C in 1 h, and that in B was recorded at 25 °C in 15 min. The buffer in A comprised 100 mM sodium acetate-d_3, pH 5.4, and that in B, 50 mM sodium phosphate, pH 5.4. Several cross-peaks are labeled and side-chain amino groups are indicated by lower case letters.

Acknowledgments: This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health. Special thanks go to Man-Yee Wong for carrying out initial trial experiments.

References

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.

Clore GM, Gronenborn AM. 1991. Structures of larger proteins in solution: Three- and four-dimensional heteronuclear NMR spectroscopy. *Science 252*:1390–1399.

Driscoll PC, Clore GM, Marion D, Wingfield PT, Gronenborn AM. 1990. Complete resonance assignment for the polypeptide backbone of interleukin-1β using three-dimensional heteronuclear NMR spectroscopy. *Biochemistry 29*:3542–5356.

Gronenborn AM, Clore GM. 1993. Identification of the contact surface of a streptococcal protein G domain complexed with a human Fc fragment. *J Mol Biol 233*:331–335.

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

Remaut E, Stanssens P, Fiers W. 1981. High-efficiency-expression plasmid vectors controlled by the p_L promoter of bacteriophage lambda. *Gene 15*:81–93.

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, Rosenberg AH, Dunn JJ, Dubendorff JW. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol 185*:60–89.