J|A|C|S
ARTICLES
Published on Web 08/01/2002

Four-Dimensional NMR Spectroscopy of a 723-Residue Protein: Chemical Shift Assignments and Secondary Structure of Malate Synthase G

Vitali Tugarinov,[†] Ranjith Muhandiram,[†] Ayeda Ayed,[‡] and Lewis E. Kay*[,†]

Contribution from the Protein Engineering Network Centres of Excellence and the Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8, and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario, Canada M5G 2M9

Received April 22, 2002. Revised Manuscript Received June 11, 2002

Abstract: A four-dimensional (4-D) NMR study of *Escherichia coli* malate synthase G (MSG), a 723-residue monomeric enzyme (81.4 kDa), is described. Virtually complete backbone ^{1}HN, ^{15}N, ^{13}C, and $^{13}C^{\beta}$ chemical shift assignments of this largely α-helical protein are reported. The assignment strategy follows from our previously described approach based on TROSY triple resonance 4-D NMR spectroscopy [Yang, D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2571−2575. Konrat, R; Yang, D; Kay, L. E. *J. Biomol. NMR* **1999**, *15*, 309−313] with a number of modifications necessitated by the large size of the protein. A protocol for refolding deuterated MSG in vitro was developed to protonate the amides deeply buried in the protein core. Of interest, during the course of the assignment, an isoaspartyl linkage in the protein sequence was unambiguously identified. Chemical shift assignments of this system are a first step in the study of how the domains of the protein change in response to ligand binding and for characterizing the dynamical properties of the enzyme that are likely important for function.

Introduction

The assignment of chemical shifts to specific sites in a macromolecule forms the basis for NMR studies of its structure and dynamics.[1] In the past decade, triple resonance ^{15}N, ^{13}C spectroscopy has evolved into a powerful methodology for resonance assignments of proteins,[2,3] and complete assignments of molecules up to approximately 25 kDa in molecular weight are now becoming increasingly routine. Applications to larger proteins generally require the use of deuteration in concert with ^{15}N and ^{13}C labeling,[4−8] and a number of studies of systems on the order of 300−400 amino acids have now appeared in the literature.[7,8] The development of TROSY spectroscopy in which the relaxation interference between ^{1}H−X dipolar and X chemical shift anisotropy interactions is exploited offers significant improvements in resolution and sensitivity in ^{1}HN−^{15}N and ^{1}H−$^{13}C^{aromatic}$ correlation experiments.[9,10] In combination with ^{2}H-based triple resonance schemes, TROSY experiments have significantly increased the size of molecules amenable to detailed NMR studies.[11] In this regard, Wüthrich and co-workers have recently reported near complete backbone assignments of a 110 kDa homooctameric protein, 7,8-dihydroneopterin aldolase.[12] Additional important applications include studies of membrane proteins, such as OmpX[13,14] and OmpA[15] in micelles, with effective molecular masses on the order of 50−60 kDa. Although the size of these systems is large, the number of residues, and hence correlations in spectra, is relatively small, and the resolution afforded by three-dimensional spectroscopy is sufficient.

Over the past several years, our laboratory has developed a family of four-dimensional TROSY-based triple resonance NMR experiments for application to high molecular weight single polypeptide chains in which both the size and the number of residues become an issue.[16,17] The development of tools for the

* To whom correspondence should be addressed. E-mail: kay@pound.med.utoronto.ca.

[†] Protein Engineering Network Centres of Excellence and the Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto.

[‡] Department of Medical Biophysics, University of Toronto.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.
(2) Bax, A. *Curr. Opin. Struct. Biol.* **1994**, *4*, 738−744.
(3) Clore, G. M.; Gronenborn, A. M. *Science* **1991**, *252*, 1390−9.
(4) Grzesiek, S.; Anglister, J.; Ren, H.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 4369−4370.
(5) Yamazaki, T.; Lee, W.; Revington, M.; Mattiello, D. L.; Dahlquist, F. W.; Arrowsmith, C. H.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 6464−6465.
(6) Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994**, *116*, 11655−11666.
(7) Farmer, B. T.; Venters, R. A. In *NMR of Perdeuterated Large Proteins*; Krishna, N. R., Berliner, L. J., Eds.; Kluwer Academic/Plenum Publishers: New York, 1998; Vol. 16, pp 75−120.
(8) Gardner, K. H.; Kay, L. E. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 357−406.
(9) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(10) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. *J. Am. Chem. Soc.* **1998**, *120*, 6394−6400.
(11) Wider, G.; Wüthrich, K. *Curr. Opin. Struct. Biol.* **1999**, *9*, 594−601.
(12) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *J. Am. Chem. Soc.* **2000**, *122*, 7543−7548.
(13) Fernandez, C.; Hilty, C.; Bonjour, S.; Adeishvili, K.; Pervushin, K.; Wüthrich, K. *FEBS Lett.* **2001**, *504*, 173−8.
(14) Fernandez, C.; Adeishvili, K.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2358−63.
(15) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 334−8.
(16) Yang, D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 2571−2575.

assignment of such systems is critical when one considers that the average human protein length is 460 residues,[18] with many important biological molecules significantly larger. As a preliminary test of the methodology, NMR studies of maltose-binding protein (MBP, 370 residues) at 5 °C were carried out.[16] At this temperature, the overall molecular tumbling time is 46 ns, equivalent to a molecule on the order of 100 kDa at 37 °C. The data sets obtained were of high sensitivity and, as expected, of very high resolution, allowing near complete backbone assignment. More recently, assignments of dimers of the human tumor suppressor protein p53[19] (67 kDa, 279 residues) have allowed us to rule out a regulatory mechanism which involves interactions between the C-terminal domain and other domains of the protein.[20] These two studies clearly indicate the utility of 4-D-TROSY spectroscopy for the investigation of proteins (or protein complexes) with moderate numbers of residues. It is important now to establish that the methodology is also of use for application to proteins of increasing complexity.

Malate synthase G (MSG) from *Escherichia coli*, a 723-residue monomeric enzyme[21] (81.4 kDa), catalyzes the condensation of glyoxylate with an acetyl group of acetyl-CoA, producing malate, an intermediate in the citric acid cycle. The crystal structure of MSG complexed with magnesium and glyoxalate was solved recently at 2.0 Å resolution.[22] Malate synthase G is a globular multidomain protein; its structure is based on the $\beta 8/\alpha 8$ (TIM barrel) fold, with insertions within the fold of the barrel and extensions at both ends forming additional domains and regions of secondary structure.[22] To obtain a complete description of the catalytic mechanism, it is of considerable importance to establish how the domains in the protein change orientation and what intradomain structural perturbations occur in response to ligand binding (only the glyoxalate-bound structure is available). Our recent studies of domain orientation in both maltose binding protein[23−25] and T4 lysozyme[26] suggest that interdomain structures in solution and in the crystal can be quite different. With this in mind, we have begun a detailed NMR investigation of MSG, starting with the apo-form of the protein. Herein, we describe a four-dimensional NMR study that has allowed near complete (>95%) backbone ^{1}HN, ^{15}N, $^{13}C^{\alpha}$, ^{13}CO, and side-chain $^{13}C^{\beta}$ chemical shift assignments of the protein.

Materials and Methods

MSG Overexpression and Initial Purification. ^{15}N, ^{13}C, ^{2}H-labeled samples of MSG were obtained by overexpression from cultures of *E. coli* BL21(DE3)pLysS cells transformed with the plasmid pMSG-B[22] encoding all the residues of MSG with (i) the substitution of Ser 2 with an alanine, (ii) the addition of an N-terminal methionine, and (iii) the addition of an eight-residue C-terminal hexa-histidine tag (LEHHHHHH). The transformed cells were grown in 5 mL of LB medium at 37 °C for 3 h, then transferred to 25 mL of unlabeled minimal M9/H_2O medium, and grown until an A_{600} of ∼0.6. The cells were then separated from the medium and transferred to 100 mL of M9/D_2O culture containing 1 g/L of $^{15}NH_4Cl$ as the sole nitrogen source and 2 g/L of ^{13}C, ^{2}H-glucose as the sole carbon source. At $A_{600} \approx 0.6$, the culture was diluted to 400 mL with M9/D_2O, grown until an A_{600} of ∼ 0.6, and then diluted to a final volume of 2 L. The expression of the histidine-tagged enzyme was induced at $A_{600} \approx 0.3$ with 1 mM IPTG, and the culture was allowed to grow for an additional 7−8 h (final $A_{600} \approx 1.0$). The cells were disrupted by sonication, and the cellular debris was pelleted by centrifugation. The lysate was passed through a nickel-affinity column (Ni-NTA Agarose, Qiagen), and the bound malate synthase was washed and subsequently eluted with 20 mM Tris buffer containing 300 mM NaCl, 10 mM β-mercaptoethanol, and 500 mM imidazole, pH 7.8. The fractions containing pure enzyme as judged by SDS−PAGE were combined and concentrated. Yields of MSG were estimated to be ∼40−50 mg per L of M9/D_2O medium prior to refolding (see below).

MSG Unfolding and Refolding. To fully protonate the amide positions of those residues buried in the core of MSG after growths in D_2O-based media, the protein was completely unfolded after initial purification in 20 mM Tris buffer containing 100 mM NaCl, 10 mM β-mercaptoethanol, and 6 M guanidinium hydrochloride (GuHCl), pH 7.8 (protein concentration ≈ 1.5 mg/mL), and incubated for 1 h at room temperature. The protein was then refolded by rapid dilution into GuHCl-free buffer containing 20 mM Tris, 300 mM NaCl, 10 mM β-mercaptoethanol, 5 mM magnesium sulfate, 2 mM benzamidine, and 10% sucrose, pH 7.8 (final protein concentration ≈ 35 μg/mL). The dilution was carried out at 4 °C and accompanied by vigorous stirring. After a 2 h incubation period at room temperature, the solution was filtered from insoluble protein aggregates and loaded onto a Ni-NTA column. The refolded malate synthase was then eluted as described above, and the purest fractions were combined and concentrated. The refolded MSG was further purified on a G-200 Sephadex size-exclusion column equilibrated with 20 mM sodium phosphate buffer (pH 7.1) containing 5 mM dithiothreitol (DTT) and 0.05% NaN_3.

NMR Samples. NMR samples were between 0.7 and 0.8 mM in protein dissolved in 20 mM sodium phosphate buffer (pH 7.1), 5 mM DTT, 20 mM $MgCl_2$, 0.05% NaN_3, 0.1 mg/mL Pefabloc, and 10% D_2O. Even at these concentrations, a small amount of protein precipitated over time; increased protein concentrations did not lead to improved spectral quality, presumably because of increases in sample viscosity and/or protein aggregation. All of the experiments used in the course of the assignment were recorded on a single MSG sample produced using ^{13}C, ^{2}H-glucose, D_2O-based media to ensure the highest possible level of deuteration. In the course of data acquisition, a small number of correlations in ^{15}N−^{1}HN HSQC spectra were found to move and weaken. Therefore, the NMR sample was periodically passed through a G-200 Sephadex size-exclusion column and redissolved in fresh buffer. The sample concentration decreased somewhat by this procedure and was approximately 0.5 mM in protein by the end. Dynamic light scattering (Protein Solutions, NJ) was used to verify the absence of any significant degree of aggregation of MSG at concentrations close to those used in NMR samples.

A U-[^{15}N, ^{2}H]-labeled MSG sample with Ile, Leu, and Val selectively labeled with ^{13}C was produced as described above except for the use of ^{12}C, ^{1}H-glucose as the sole carbon source and the addition of 50 mg each of U-[^{15}N, ^{13}C]-labeled Ile, Leu, and Val per 1 L of M9 medium 1 h prior to induction. In this case, the growth was continued for only 5 h after induction to minimize isotopic dilution. U-[^{15}N, ^{13}C]-Leu and -Val were obtained from Cambridge Isotope Laboratories Inc. (Andover, MS) and used without further purification, while U-[^{15}N, ^{13}C]-Ile was purchased from Isotec (Miamisburgh, PA).

(17) Konrat, R.; Yang, D.; Kay, L. E. *J. Biomol. NMR* **1999**, *15*, 309−313.
(18) Apweiler, R.; Biswas, M.; Fleischmann, W.; Kanapin, A.; Karavidopoulou, Y.; Kersey, P.; Kriventseva, E. V.; Mittard, V.; Mulder, N.; Phan, I.; Zdobnov, E. *Nucleic Acids Res.* **2001**, *29*, 44−8.
(19) Mulder, F. A.; Ayed, A.; Yang, D.; Arrowsmith, C. H.; Kay, L. E. *J. Biomol. NMR* **2000**, *18*, 173−6.
(20) Ayed, A.; Mulder, F.; Yi, G. S.; Kay, L. E.; Arrowsmith, C. H. *Nat. Struct. Biol.* **2001**, *8*, 750−760.
(21) Molina, I.; Pellicer, M. T.; Badia, J.; Aguilar, J.; Baldoma, L. *Eur. J. Biochem.* **1994**, *224*, 541−8.
(22) Howard, B. R.; Endrizzi, J. A.; Remington, S. J. *Biochemistry* **2000**, *39*, 3156−68.
(23) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W. Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295*, 1265−1273.
(24) Evenas, J.; Tugarinov, V.; Skrynnikov, N. R.; Goto, N. K.; Muhandiram, R.; Kay, L. E. *J. Mol. Biol.* **2001**, *309*, 961−74.
(25) Hwang, P. M.; Skrynnikov, N. R.; Kay, L. E. *J. Biomol. NMR* **2001**, *20*, 83−8.
(26) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745−64.

NMR Spectroscopy. All NMR experiments were performed on either 800 or 600 MHz four-channel Varian Inova spectrometers operating at 37 °C and equipped with pulsed-field gradient triple resonance probes. 4-D TROSY−HNCACO[16] (correlations of the form [$\omega_{C\alpha}(i)$, $\omega_{CO}(i)$, $\omega_N(i)$, $\omega_{HN}(i)$] and, in some cases, [$\omega_{C\alpha}(i-1)$, $\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]), 4-D TROSY−HNCO$_{i-1}$CA$_i$[17] ([$\omega_{C\alpha}(i)$, $\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$] and, in some cases, [$\omega_{C\alpha}(i-1)$, $\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]), 4-D ^{15}N, ^{15}N−NOESY[27,28] (mixing time of 90 ms), 3-D TROSY−HNCO[29,30] ([$\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]), 3-D TROSY−HN(CA)CO[31] ([$\omega_{CO}(i)$, $\omega_N(i)$, $\omega_{HN}(i)$] and, in some cases, [$\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]), and 3-D TROSY−HNCACB[16,30] spectra ([$\omega_{C\alpha}(i)/\omega_{C\beta}(i)$, $\omega_N(i)$, $\omega_{HN}(i)$] and, to a lesser extent, [$\omega_{C\alpha}(i-1)/\omega_{C\beta}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]) were recorded at 800 MHz using pulse sequences described previously, with the parameters indicated in Table 1 of the Supporting Information. Selection of the TROSY component in all experiments except the 4-D NOESY was achieved passively by relying exclusively on the efficient relaxation of the anti-TROSY component during the relatively long transfer delays in the pulse schemes.[16,29] Only in six cases was a cross-peak derived from the anti-TROSY signal observed. The 4-D ^{15}N, ^{15}N−NOESY spectrum was recorded with the selection of the TROSY component[9] in F_3 (the second ^{15}N dimension)/F_4. In total, 1.5 months of spectrometer time was required to record all of the data sets listed above.

During the analysis of the 4-D TROSY−HNCACO, we noticed that a number of the correlations were either weaker than expected or missing altogether. This experiment employs a pair of $^{13}C^{\alpha}$-selective refocusing pulses during the $^{13}C^{\alpha} \rightarrow {}^{13}CO$ transfer steps to refocus evolution resulting from the one-bond $^{13}C^{\alpha}-{}^{13}C^{\beta}$ scalar coupling and hence increase the sensitivity of the majority of cross-peaks in the resulting spectrum (see ref 16). Specifically, attenuation of signal was noted for cross-peaks derived from residues where the selective pulses affect both $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ spins. Serines were affected most significantly, and intraresidue correlations for the majority of these residues (22 of the 28 ultimately assigned) were missing. The origin of the sensitivity loss and a slight modification to the original pulse scheme which improves signal-to-noise for these residues is presented in Results and Discussion. We chose not to re-record the 4-D HNCACO (with the modification) but rather opted to measure a 3-D TROSY−HN(CA)CO spectrum with the selective pulses replaced by broadbanded (for the aliphatic region) RE-BURP pulses[32] and the delays used to transfer magnetization from $^{13}C^{\alpha}$ to ^{13}CO reduced (from 8.2 to 5.4 ms). In this manner, sensitivity losses for correlations involving Ser are minimized, and 19 of the 22 cross-peaks that could not be observed in the 4-D spectrum could now be visualized.

A number of experiments (4-D TROSY−HNCOCA,[16] providing correlations of the form [$\omega_{C\alpha}(i-1)$, $\omega_{CO}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$], and 3-D TROSY−HN(CO)CACB,[16,30] [$\omega_{C\alpha}(i-1)/\omega_{C\beta}(i-1)$, $\omega_N(i)$, $\omega_{HN}(i)$]) transfer magnetization from the ^{1}HN, ^{15}N pairs of residue i to the $^{13}C^{\alpha}/^{13}C^{\beta}$ spins of the preceding residue. Two of the transfer steps ($^{13}CO \rightarrow {}^{13}C^{\alpha}$ and back) place transverse magnetization on the ^{13}CO spin for periods of $\sim 1/(2J_{C\alpha CO})$, which is not insignificant in relation to the transverse relaxation time of the carbonyl spin. For example, for a protein tumbling isotropically in solution with a correlation time of 37 ns (see below), the contribution to the relaxation rate of the carbonyl spin from chemical shift anisotropy at 800 MHz is 93 s^{-1} (T_2 of 10.8 ms), assuming values for (σ_x, σ_y, σ_z) of (73.3, 7.3, −80.6) ppm.[33] HNCOCA type experiments recorded at 800 MHz are predicted, therefore, to be approximately a factor of 2 less sensitive than the corresponding experiments at 600 MHz, and this difference has been observed in test two-dimensional spectra recorded at both fields. Thus, we have chosen to record these experiments at 600 MHz.

Both 3-D HN(CO)CACB and 3-D HNCACB experiments were recorded with nonconstant-time $^{13}C^{\beta}$ evolution periods and the $^{13}C^{\alpha}-{}^{13}C^{\beta}$ transfer times optimized for maximum sensitivity of $^{13}C^{\beta}$ peaks using delays that were somewhat less than $1/(2J_{C\alpha C\beta})$. This led to the appearance of typically weak $^{13}C^{\alpha}$ correlations in these spectra, in addition to cross-peaks involving the $^{13}C^{\beta}$.

NMR experiments for the measurement of ^{15}N T_1, ^{15}N $T_{1\rho}$, ^{15}N $T_{1\rho}$ of the TROSY component and HN T_1 and T_2 values were recorded at 800 MHz ^{1}H frequency using TROSY versions of pulse sequences described elsewhere.[34−36] The spin lock power used in the ^{15}N $T_{1\rho}$ experiments was 1.6 kHz. T_1 and $T_{1\rho}$ (T_2) values were determined by nonlinear least-squares fitting of the experimental data to monoexponential decay equations ($A \exp(-t/T_i)$) using the program ModelXY.[37] Uncertainties in the derived relaxation parameters were estimated with 120 Monte Carlo simulations.[38]

All NMR spectra were processed using the suite of programs provided in NMRPipe/NMRDraw software.[37] Briefly, the residual water signal was minimized by time domain deconvolution.[39] The ^{15}N time domain of all the spectra was doubled using mirror image linear prediction,[40] before apodization with a cosine-bell window function and Fourier transformation. The ^{13}C time domains of all of the spectra were doubled using either mirror-image[40] (for constant-time dimensions) or forward−backward linear prediction[41] and apodized with cosine-squared window functions. Linear prediction in a given dimension was performed only after all of the other spectral dimensions were transformed as described by Kay et al.[42] The frequency domain spectra acquired at 600 MHz were recalibrated in ^{1}H and ^{15}N dimensions for consistency with the 800 MHz data to account for differences in chemical shifts of the TROSY component (in ppm) at different spectrometer fields. The transformed data sets were reduced to include only the regions of interest and analyzed using the NMRView program.[43] Chemical shift indices of $^{13}C^{\alpha}$ were calculated using the program CSI.[44,45]

Results and Discussion

In Vitro Refolding of MSG. In the initial stages of the study, we dissolved a fully protonated MSG sample in D_2O, 25 °C, and periodically recorded ^{15}N−^{1}HN TROSY−HSQC spectra over a period of 2 months. Approximately 120 backbone amides remained protonated after 1 week, while ~70 remained unexchanged even after 6 weeks. Thus, it was necessary to unfold/refold MSG so that all of the amide groups could be protonated at a high level after protein preparation in D_2O-based media.

In vitro refolding of large multidomain proteins is known to be difficult without the use of molecular chaperones.[46] However,

(27) Venters, R. A.; Metzler, W. J.; Spicer, L. D.; Mueller, L.; Farmer, B. T. *J. Am. Chem. Soc.* **1995**, *117*, 9592−9593.
(28) Grzesiek, S.; Wingfield, P.; Stahl, S.; Kaufman, J.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 9594−9595.
(29) Yang, D.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 3−10.
(30) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wüthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13585−13590.
(31) Clubb, R. T.; Thanabal, V.; Wagner, G. *J. Magn. Reson.* **1992**, *97*, 213−217.
(32) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93−141.
(33) Teng, Q.; Iqbal, M.; Cross, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 5312−5321.
(34) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984−6003.
(35) Akke, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1996**, *118*, 911−912.
(36) Donaldson, L. W.; Skrynnikov, N. R.; Choy, W. Y.; Muhandiram, D. R.; Sarkar, B.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **2001**, *123*, 9843−9847.
(37) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(38) Kamith, U.; Shriver, J. W. *J. Biol. Chem.* **1989**, *264*, 5586−5592.
(39) Marion, D.; Ikura, M.; Bax, A. *J. Magn. Reson.* **1989**, *84*, 425−430.
(40) Zhu, G.; Bax, A. *J. Magn. Reson.* **1990**, *90*, 405−410.
(41) Zhu, G.; Bax, A. *J. Magn. Reson.* **1992**, *98*, 192−199.
(42) Kay, L. E.; Ikura, M.; Zhu, G.; Bax, A. *J. Magn. Reson.* **1991**, *91*, 422−428.
(43) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603−614.
(44) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647−1651.
(45) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171−180.
(46) Lilie, H.; Schwarz, E.; Rudolph, R. *Curr. Opin. Biotechnol.* **1998**, *9*, 497−501.

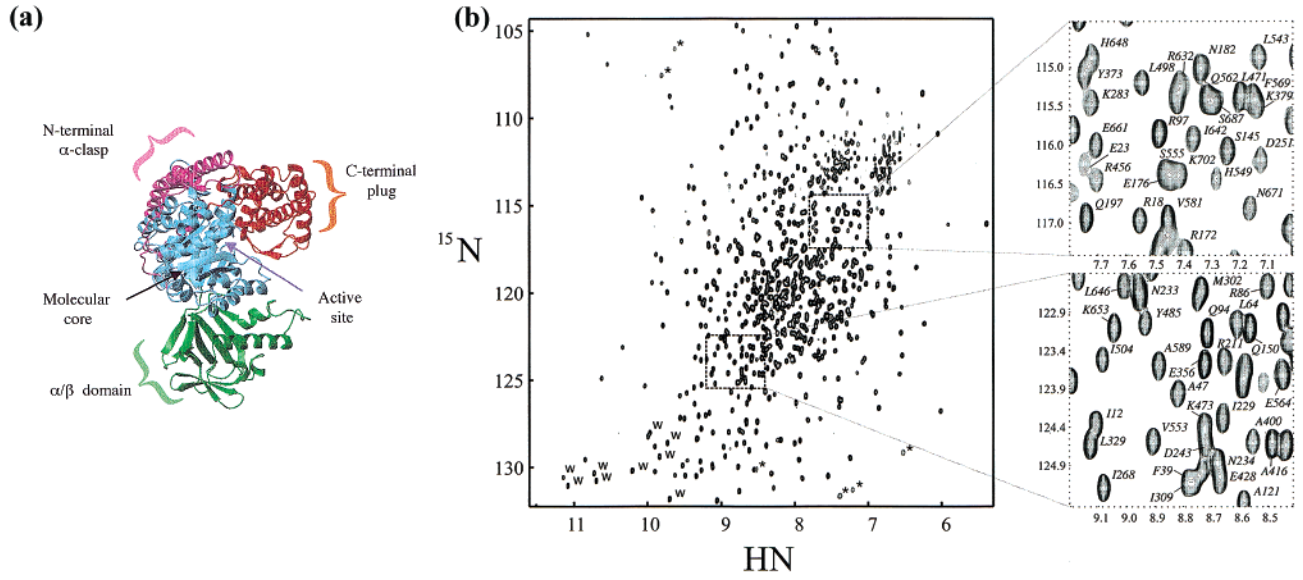

Figure 1. (a) Ribbon diagram of MSG illustrating the four main domains of the molecule. The MSG coordinates are taken from the Protein Data Bank, PDB code 1d8c.[22] (b) 800 MHz ^{15}N−^{1}HN TROSY−HSQC spectrum of malate synthase G recorded at 37 °C. Two regions of the spectrum are enlarged and labeled with residue numbers. Tryptophan side-chain resonances are labeled with "W", while peaks aliased in the ^{15}N dimension are marked with asterisks.

this step is essential for any detailed NMR study of a large deuterated protein using currently available strategies in which magnetization both originates and is detected on amide protons.[8] In the case of MSG, several of the commonly used methods,[46] including slowly dialyzing out the denaturant (6 M GuHCl or urea), refolding on a Ni metal-affinity column, as well as partial protein denaturation with 2.5 M GuHCl followed by fast dilution refolding used previously for NMR studies of deuterated maltose binding protein,[47] all failed to provide substantial yields of the refolded molecule. In contrast, complete denaturation of MSG with 6 M GuHCl with subsequent fast dilution refolding as described in Materials and Methods commonly provided refolding yields of 60 ± 15%. Rapid dilution from a fully unfolded state is presumably necessary to avoid the aggregation and ensuing precipitation that result if a partially denatured state is allowed to accumulate. No significant improvements in the refolding yield were observed with the addition of the substrate (sodium glyoxalate) to the refolding buffer. For other multidomain proteins with less favorable refolding properties, it may be necessary to resort to refolding chromatography with immobilized minichaperones[48] or to a fast dilution method with addition of nondetergent sulfobetaines ("chemical chaperones").[49] The refolded enzyme was structurally identical to native MSG which was not subject to unfolding/refolding as judged by peak positions in ^{15}N−^{1}HN TROSY−HSQC spectra recorded before and after the refolding step.

Sequential Assignments. Malate synthase G is a multidomain enzyme,[22] comprised of four main domains illustrated in Figure 1a. The centrally located core of the molecule is based on a highly stable β8/α8 barrel fold (residues 117−134; 263−550). It is supported on one side by an N-terminal α-helical clasp (residues 3−88) which is linked to the first strand of the barrel by a long extended loop (residues 89−116). On the opposite side of the barrel is an α/β domain comprising residues 135−262. The C-terminal end of the enzyme (589−723) consists of a five-helix plug connected to the barrel by a flexible loop-helix-turn-helix-loop motif (residues 551−588). In addition, several smaller loops protrude from the globular core of the molecule. The active site of the enzyme is located in a cleft at the interface between the C-terminal plug and loops at the C-terminal ends of the β-strands of the core barrel. It was suggested on the basis of the crystallographic data[22] that the C-terminal plug may be mobile and its position relative to the rest of the enzyme may differ in the apo (open) and substrate-bound (closed) states of the molecule. Such mobility could allow an opening of the active site cleft for substrate binding and product release. We show below, however, that the domains of the molecule are all rigid (at least on a time scale faster than the overall tumbling) with a correlation time for each ranging from 35 to 37 ns (37 °C) based on ^{15}N spin relaxation measurements. It is noteworthy that a correlation time of 36 ns for MSG (81 kDa) is double what was obtained for maltose binding protein (18.6 ns)[25] (42 kDa) at a temperature of 37 °C.

Figure 1b shows a resolution-enhanced ^{15}N−^{1}HN TROSY−HSQC spectrum of refolded MSG recorded at 800 MHz. A significant degree of resonance overlap is visible in the central part of the correlation map partially because of the high helical content in the protein, with 67% of regions of secondary structure and 47% of the entire protein comprised of helices.

The strategy that has been employed to assign MSG makes use of 4-D NMR spectroscopy and follows from what has been described earlier.[16,17,19] Because of the large number of residues in the protein, we have found it useful to simplify the ^{15}N−^{1}HN correlation map by recording a modified HN(CACB) experiment,[50] described in the Appendix, in which only Ala and those residues immediately following Ala are selected. Figure 2a shows such a spectrum in which correlations from 67 of the 71 Ala residues in the protein are observed, along with several cross-peaks from residues following Ala. Our choice of design-

(47) Gardner, K. H.; Zhang, X.; Gehring, K.; Kay, L. E. *J. Am. Chem. Soc.* **1998**, *120*, 11738−11748.
(48) Altamirano, M. M.; Golbik, R.; Zahn, R.; Buckle, A. M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3576−8.
(49) Goldberg, M. E.; Expert-Bezançon, N.; Vuillard, L.; Rabilloud, T. *Folding Des.* **1995**, *1*, 21−7.
(50) Wittekind, M.; Mueller, L. *J. Magn. Reson., Ser. B* **1993**, *101*, 201−205.

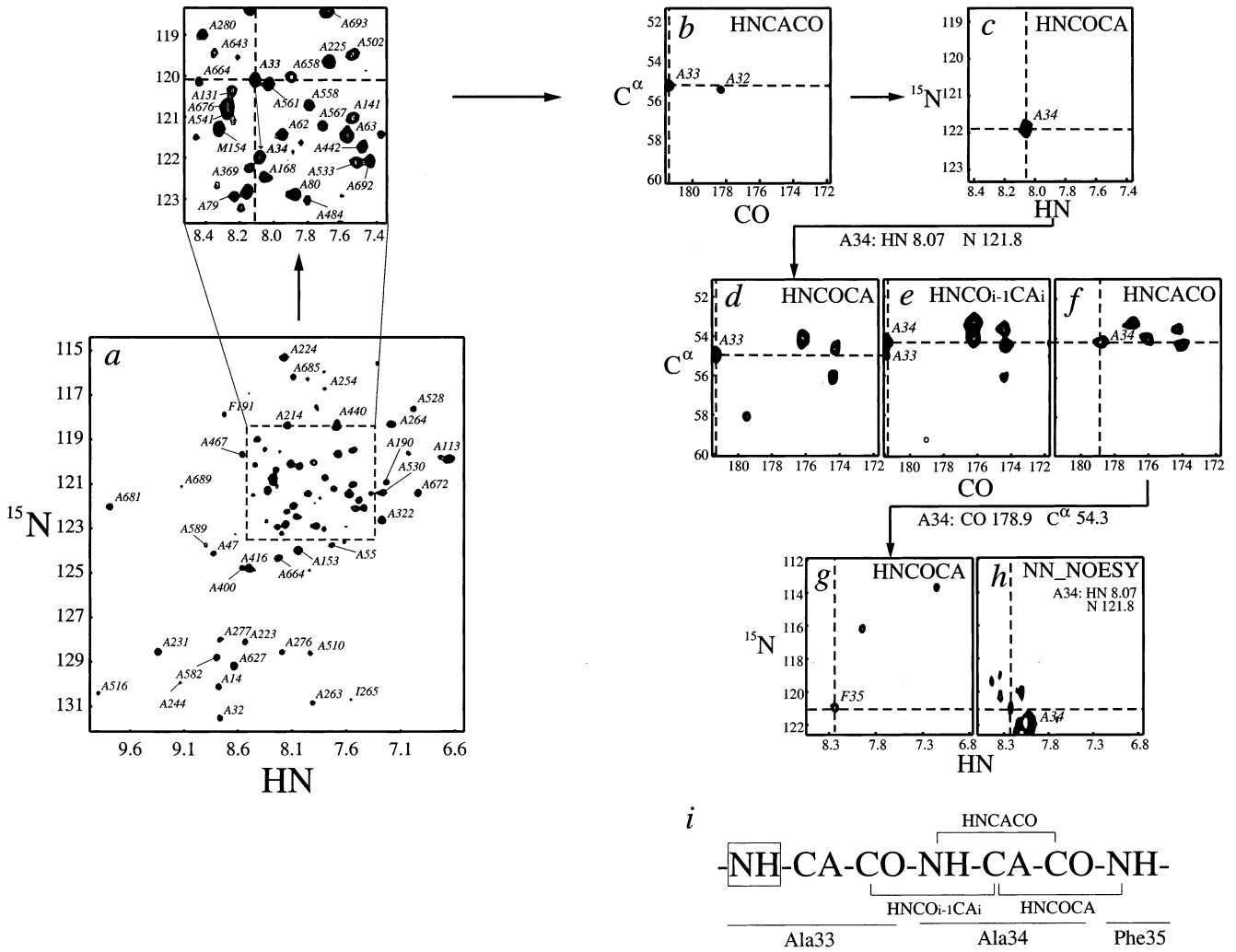

Figure 2. Sequential backbone assignments for residues A33−F35 of MSG starting from the amide correlation of A33. (a) Ala-selective ^{15}N−^{1}HN correlation spectrum recorded with the modified HNCACB pulse sequence described in the Appendix (see Figure 8). Ala residues and those following Ala are observed (and labeled in the spectrum). (b−h) ^{13}C$^{\alpha}$/^{13}CO and ^{15}N/^{1}HN planes of the 4-D spectra used in the assignment process. Arrows indicate the sequence of the assignment steps described in the text. (i) Diagram of the connectivities established from different 4-D NMR spectra for the HN A33−HN F35 stretch. The assignment process illustrated starts with the ^{15}N/^{1}HN of Ala 33.

ing an Ala-selective HSQC experiment is based on the fact that (i) Ala is the most abundant amino acid in MSG (10.1%) and that (ii) its distinctive ^{13}C$^{\beta}$ chemical shift makes selection particularly straightforward. The ^{15}N−^{1}HN correlations from the Ala-selective HSQC, along with correlations from Gly which are easy to identify in the 4-D data sets, were used as starting points in the assignment process and later to ensure the correct and unambiguous positioning of a connected stretch of residues in the 723-residue amino acid sequence of the protein.

Figure 2b−h illustrates how the family of 4-D experiments that we have developed is used for sequential assignment. A slice of the 4-D HNCACO at the amide ^{1}HN and ^{15}N chemical shifts of A33 (Figure 2b,i) allows assignment of the ^{13}C$^{\alpha}$ (55.0 ppm) and ^{13}CO (181.4 ppm) shifts of this residue because the (^{15}N, ^{1}HN) correlation for this amino acid is resolved in the ^{15}N−^{1}HN HSQC spectrum (Figure 2b). It is possible to immediately identify the (^{15}N, ^{1}HN) shifts of A34 from a slice of the 4-D HNCOCA at ^{13}C$^{\alpha}$ = 55.0 ppm, ^{13}CO = 181.4 ppm (Figure 2c) because these carbon shifts are also separated from all others. Figure 2d−f shows slices from 4-D data sets at the (^{15}N, ^{1}HN) shifts of A34, allowing the straightforward identification of the ^{13}C$^{\alpha}$ and ^{13}CO shifts of this residue. The 4-D HNCO$_{i-1}$CA$_i$ data set is critical in this regard because there are five residues in MSG with the same pairs of backbone amide shifts (Figure 2f), and it would thus not be possible to unambiguously assign the intraresidue carbon shifts of A34 with only the 4-D HNCACO. However, because the ^{13}CO chemical shift of A33 and the ^{15}N, ^{1}HN shifts of A34 are known, the ^{13}C$^{\alpha}$ shift of A34 can be obtained from the 4-D HNCO$_{i-1}$CA$_i$, facilitating the assignment of the A34 carbonyl from the 4-D HNCACO. The ^{13}C$^{\alpha}$/^{13}CO chemical shifts of A34 can, in principle, be used to assign the amide shifts of F35 from the 4-D HNCOCA. However, because of the 3-fold degeneracy of this pair of carbon shifts (Figure 2g), the correct assignment can only be made using additional correlations, in this case provided by ^{1}HN−^{1}HN NOEs from a 4-D resolved ^{15}N-edited NOESY data set (Figure 2h). Generally, a high degree of degeneracy among ^{13}C$^{\alpha}$/^{13}CO pairs of chemical shifts was observed, especially in helical regions of the protein, and the 4-D NOESY data set was extremely useful in these cases. Typically, for the mixing time used in the NOESY experiment, 90 ms, sequential ^{1}HN−^{1}HN NOE cross-peaks could not be detected in β-sheets and loops and in cases of chemical shift degeneracy, therefore, decisions were based on the connectivities from 3-D HN(CO)CACB and HNCACB

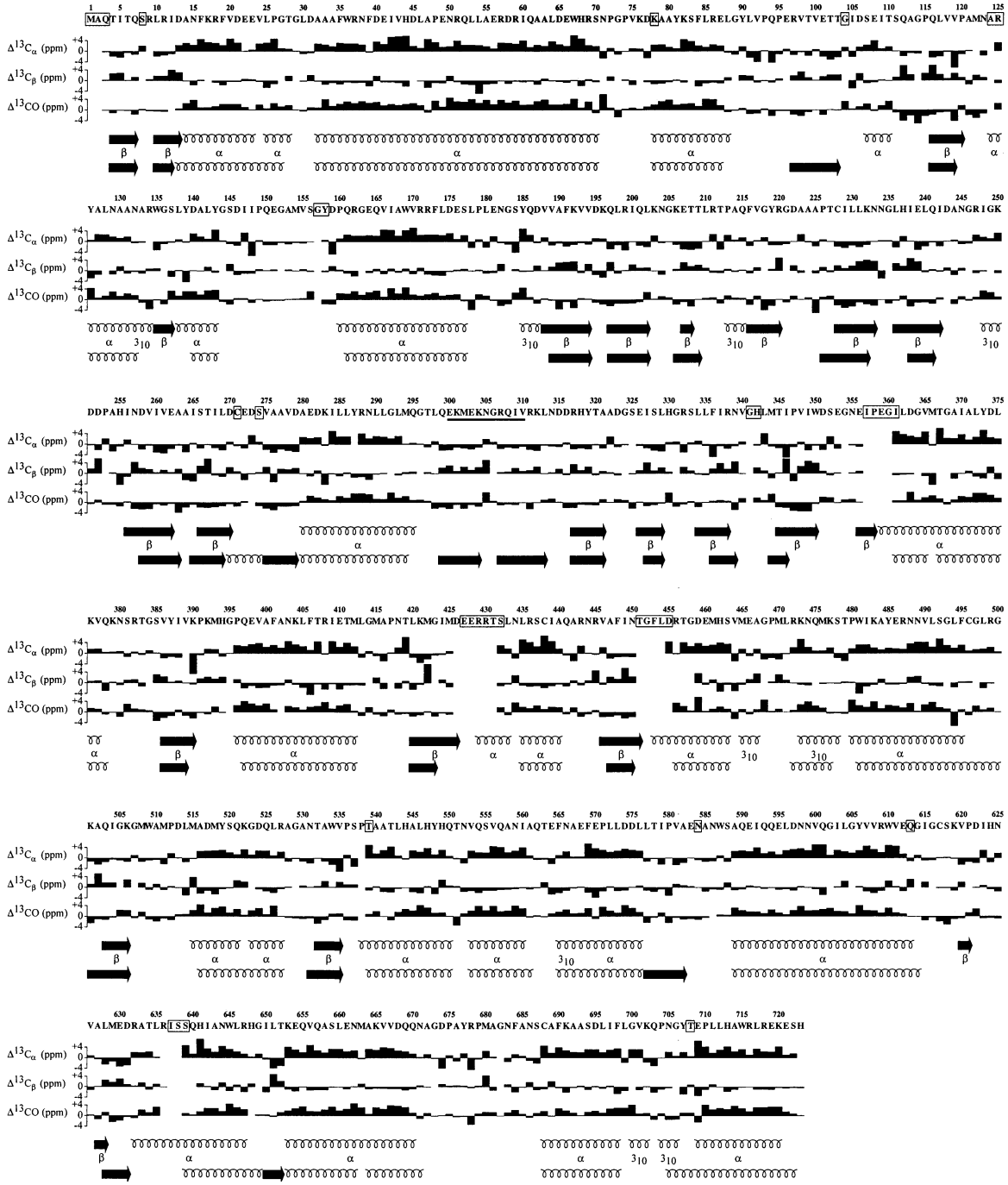

Figure 3. Deviations of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts of MSG from random coil values. The secondary structure elements shown in the upper row are based on the X-ray structure, PDB code 1d8c.[22] Random coil chemical shifts were taken from the reduced database of protein chemical shifts in the BioMagResBank (http://www.bmrb.wisc.edu) and corrected for one-, two-, and three-bond ^{2}H isotope effects.[64,65] The secondary structure elements shown in the lower row are predicted from the chemical shift index analysis[45] (CSI) of $^{13}C^{\alpha}$ shifts. Amides of MSG remaining unassigned are enclosed in rectangular boxes. The coordinates of the underlined stretch of residues (E300−V310) are missing from the X-ray structure.

spectra. Of interest, we have found that the helical regions of MSG were easier to assign than other secondary structural elements because ambiguities in $^{13}C^{\alpha}/^{13}CO$ chemical shift pairs could be easily resolved by NOE connectivities.

In general, starting from either Ala or Gly, residues were connected sequentially until another residue of an unambiguous type (one of Ala, Gly, Ser, or Thr) was reached. The connected stretch, for example, Ala(Gly)−$(X)_n$−Ala, was then positioned in the MSG sequence taking into account residue-type information of all the intervening residues X from 3-D HNCACB and HN(CO)CACB spectra. After one month of data analysis, 95% of the amide and 97% of the carbon ($^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO) chemical shifts of MSG were assigned.

Figure 3 shows deviations of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts from mean random coil values that have been corrected for deuterium isotope shifts as well as the secondary structure of MSG predicted on the basis of the chemical shift index of Wishart et al.[45] using $^{13}C^{\alpha}$ shifts exclusively. Inclusion of $^{13}C^{\beta}$ and ^{13}CO chemical shifts in the CSI analysis introduced only marginal changes to the predicted secondary structure. The NMR-based secondary structure prediction (bottom line) is in good agreement with that derived from the crystallographic

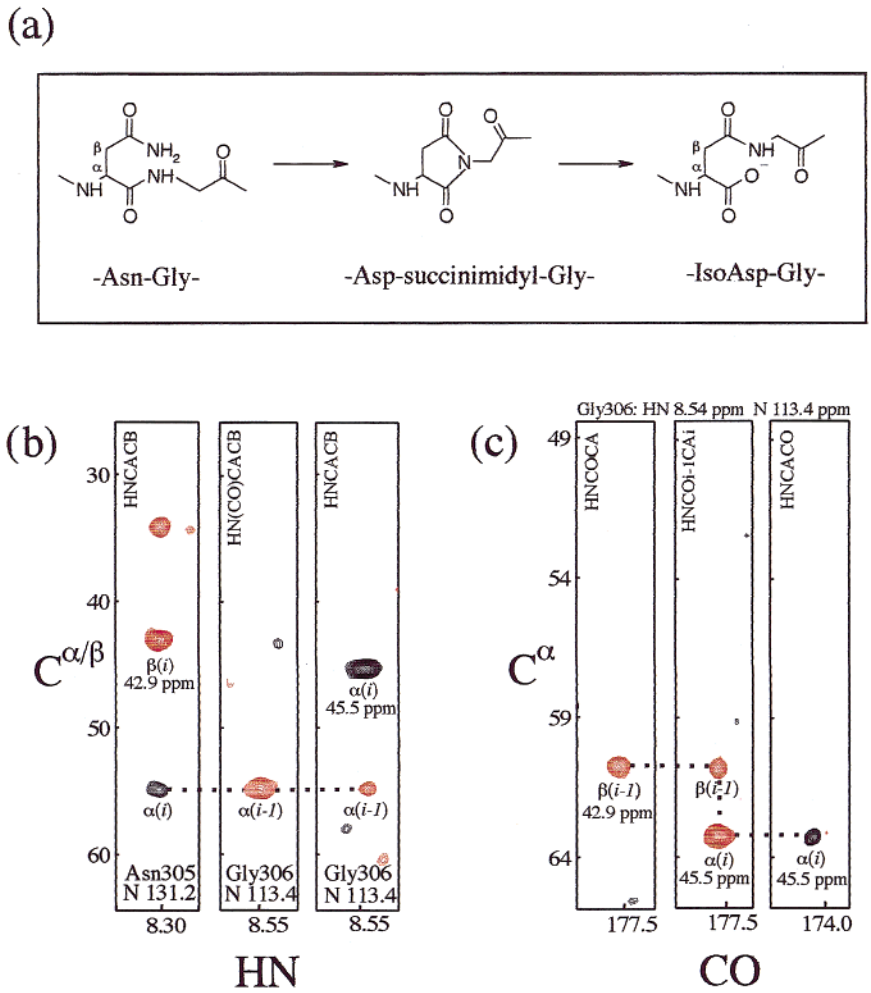

Figure 4. (a) The deamidation pathway of the N305 side chain resulting in the isoaspartyl linkage between the N305-G306 pair. (b) Panels from 3-D HN(CO)CACB[16,30] and HNCACB[16,30] spectra drawn at ^{1}HN and ^{15}N chemical shifts of N305/G306 are consistent with an −IsoAsp−Gly− linkage and not an −Asn−Gly− connectivity. Negative peaks are shown in red. (c) $^{13}C^{\alpha}/^{13}CO$ panels from 4-D HNCOCA,[16] 4-D HNCO$_{i-1}$CA$_i$,[17] and 4-D HNCACO[16] spectra drawn at the ^{1}HN, ^{15}N chemical shifts of G306. Negative peaks are shown in red and correspond to correlations aliased in the $^{13}C^{\alpha}$ dimension. Chemical shift values corrected for aliasing are shown below the appropriate cross-peaks. The data confirm the −IsoAsp−Gly− structure (see text).

coordinates (second line from bottom). Of interest, carbonyl chemical shift deviations agree better in many cases with the secondary structure than those of $^{13}C^{\beta}$. For example, in the helical stretch between P160 and S177, the positive deviations of ^{13}CO chemical shifts closely follow those of $^{13}C^{\alpha}$, while $^{13}C^{\beta}$ deviations do not exhibit any regular pattern. Generally, helical regions are predicted from $^{13}C^{\alpha}$ chemical shifts better than β-strands. However, several short α- and 3_{10}-type helices remain unidentified. Many of the unassigned amides shown in Figure 3 (residues indicated with rectangular boxes) belong either to turns or loops connecting regular secondary structure elements of MSG or to the first two or three (non-hydrogen-bonded) residues in a helix and may be unobservable because of fast exchange with solvent. Among the unassigned residues, E427, G452, F453, L454, and D455 belong to the glyoxalate binding site and are likely to be conformationally disordered and hence broadened in the absence of ligand.

Unambiguous Identification of N305−G306 Isoaspartyl Linkage. In the course of the assignment process, an unusual connectivity was observed between N305 and G306 of MSG and identified as an isoaspartyl linkage resulting from the deamidation of the side chain of N305 via a cyclic imide intermediate[51] as illustrated in Figure 4a. In this linkage, the polypeptide chain includes an additional CH_2 group, derived from the side-chain C^{β} of asparagine. Isolinkages occur predominantly in Asn−Gly pairs and have been characterized extensively.[52−57] Several years ago, Chazin, Forsen, and coworkers used ^{1}H NMR spectroscopy to confirm one such linkage in calbindin D_{9k}.[58] In our study, the isoaspartyl linkage was identified unambiguously on the basis of peak sign and aliasing information from 3-D and 4-D heteronuclear data. In Figure 4b, 3-D HNCACB and HN(CO)CACB data sets have been used to distinguish between the expected −Asn(305)−Gly(306)− connectivity and the observed −IsoAsp(305)−Gly(306)− connection. It is well known that cross-peaks involving $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei are of opposite sign in these spectra,[50,59] as illustrated

(51) Creighton, T. E. *Proteins: Structures and Molecular Properties*; W. H. Freeman and Company: New York, 1993.
(52) Robinson, A. B.; Scotchler, J. W.; McKerrow, J. H. *J. Am. Chem. Soc.* **1973**, *95*, 8156−9.
(53) Geiger, T.; Clarke, S. *J. Biol. Chem.* **1987**, *262*, 785−94.
(54) Stephenson, R. C.; Clarke, S. *J. Biol. Chem.* **1989**, *264*, 6164−70.
(55) Tyler-Cross, R.; Schirch, V. *J. Biol. Chem.* **1991**, *266*, 22549−56.
(56) Mine, S.; Ueda, T.; Hashimoto, Y.; Tanaka, Y.; Imoto, T. *FEBS Lett.* **1999**, *448*, 33−7.
(57) Robinson, N. E.; Robinson, A. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4367−4372.
(58) Chazin, W. J.; Kordel, J.; Thulin, E.; Hofmann, T.; Drakenberg, T.; Forsen, S. *Biochemistry* **1989**, *28*, 8646−53.
(59) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 6291−6293.

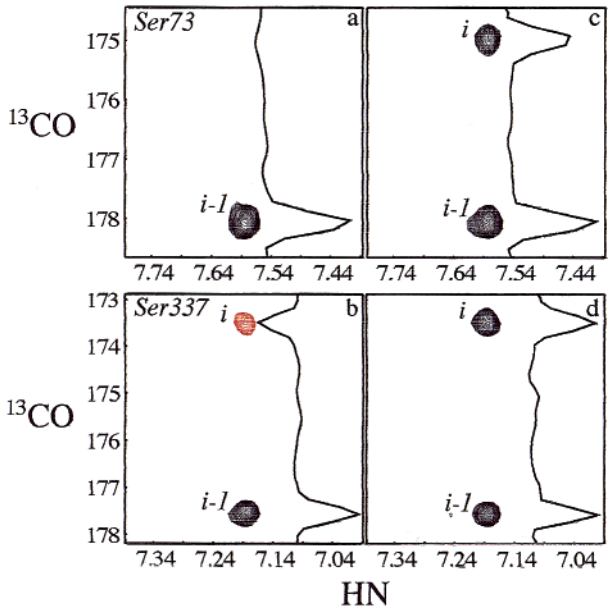

Figure 5. $^{13}CO-^{1}HN$ planes (and traces) from an HN(CA)CO spectrum of maltotriose loaded maltose binding protein[24] recorded at 800 MHz using the 4-D-HNCACO pulse scheme of Yang and Kay[16] without modification (a,b) and with insertion of an additional delay (c,d), as discussed in the text. Note that in the scheme of Figure 1b, ref 16, this additional delay is inserted between the $^{13}C^{\alpha}$ pulse following period D and the ensuing ^{13}CO pulse of phase $\phi 3$ so that the net duration between these elements is $1/(2J_{C\alpha CO})$. Negative peaks are shown in red.

in the HNCACB strip plot at the ^{15}N, ^{1}HN shifts of Asn305. Note that the $^{13}C^{\beta}$ correlation is more intense than the corresponding $^{13}C^{\alpha}$ cross-peak, as expected (for both HNCACB and HN(CO)CACB data sets) on the basis of the delays used in the experiments. The HN(CO)CACB slice at the ^{15}N, ^{1}HN shifts of Gly306 shows only a single correlation at the $^{13}C^{\alpha}$ chemical shift of Asn 305. If the structure shown on the left of Figure 4a were correct, the correlation to the $^{13}C^{\beta}$ carbon (not observed) would be stronger than that to the $^{13}C^{\alpha}$. More telling evidence that the −Asn−Gly− linkage is incorrect comes from the phase of the cross-peak observed in the HN(CO)CACB panel which indicates that the correlation must be derived from a carbon that is not directly attached to the carbonyl. In contrast, the data of Figure 4b are consistent with the −IsoAsp−Gly− structure (Figure 4a, right). Further evidence of the −IsoAsp−Gly− linkage comes from 4-D data sets. A slice from the 4-D HNCOCA at the ^{15}N, ^{1}HN shifts of Gly 306 (Figure 4c) shows a negative correlation (because of aliasing) at an aliphatic ^{13}C chemical shift of $\omega_C^{\alpha} - SW(C^{\alpha}) = 42.9$ ppm, also observed in the 4-D $HNCO_{i-1}CA_i$, corresponding to the chemical shift of the $^{13}C^{\beta}$ carbon of Asp 305. Because both 4-D experiments provide chemical shift information exclusively from backbone carbons, only the −IsoAsp−Gly− structure is consistent.

It is interesting that an isoaspartyl linkage was identified exactly in the middle of a flexible stretch (residues 300−310 of MSG, Figure 3) whose coordinates could not be determined by X-ray crystallography.[22] Two β-strands are predicted in this region by the CSI calculation (Figure 3), with the N305−G306 pair possibly forming a flexible connection between them. Because five other Asn−Gly pairs also exist in the MSG sequence with four of them located in loops and turns, and none of them form isoaspartyl linkages, this unusual flexibility (see Figure 6 below) may be a prerequisite for the formation of the cyclic imide intermediate. The deamidation of asparagines has been shown to be accelerated in phosphate buffers,[55,56] and formation of the isoaspartyl linkage might result from the specific buffer conditions (sodium phosphate, pH 7.1) chosen for NMR studies. In principle, hydrolysis of the succinimide intermediate can also lead to a conventional Asp−Gly linkage as a second reaction product. Indeed, a second set of weaker cross-peaks was identified in 3-D and 4-D spectra for residues R307−R311. However, cross-peaks from N305 and G306 of the second conformer have not been observed. It is unlikely that the isoaspartyl linkage could lead to any significant perturbation of the MSG structure, because the only additional cross-peaks that have been noted are those from the four-residue stretch (307−311).

An Improved 4-D HNCACO Pulse Scheme. As described in Materials and Methods, correlations involving residues with proximal $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ shifts were attenuated (or missing) from the 4-D HNCACO spectrum. The origin of the problem can be appreciated by examination of the 4-D HNCACO pulse scheme, Figure 1b of Yang and Kay.[16] Briefly, the flow of magnetization in this experiment can be summarized as

$$^{1}HN \xrightarrow{1} {}^{15}N \xrightarrow{2} {}^{13}C^{\alpha}(CT\text{-}t_1) \xrightarrow{3} {}^{13}CO(t_2) \xrightarrow{4} {}^{13}C^{\alpha} \xrightarrow{5} {}^{15}N(CT\text{-}t_3) \xrightarrow{6} {}^{1}HN(t_4) \quad (1)$$

where each of the magnetization transfers denoted by arrows is based on scalar couplings, the variable t_i, $i = 1-4$, corresponds to an acquisition time, and CT-t_i indicates a constant-time acquisition period. During the CT-t_1 period of duration, T_c evolution due to chemical shift and the $^{13}C^{\alpha}-^{13}CO$ one-bond coupling occurs, in preparation for the subsequent transfer of magnetization to the ^{13}CO spin (step 3 above). A significant sensitivity loss can occur from evolution due to the one-bond $^{13}C^{\alpha}-^{13}C^{\beta}$ coupling, $J_{C\alpha C\beta}$, during this interval and during a subsequent constant-time period also of duration T_c in which $^{13}C^{\alpha}$ magnetization anti-phase with respect to ^{13}CO becomes in-phase (between steps 4 and 5 above). In our experiment, $J_{C\alpha C\beta}$ evolution is refocused during both of these intervals through the use of $^{13}C^{\alpha}$ selective RE-BURP pulses of duration 1.14 ms, centered at 55 ppm (800 MHz). Simulations show that magnetization is refocused/inverted over a bandwidth of ±1.7 kHz, leaving nuclei resonating at frequencies greater than ±2.9 kHz from the center of excitation largely unaffected. As discussed previously,[16] for the majority of amino acids, evolution resulting from $^{13}C^{\alpha}-^{13}C^{\beta}$ couplings is refocused, leading to significant sensitivity gains. However, in the case of Ser, in particular, the small difference between $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts precludes refocusing of the scalar coupled evolution, and intensities of Ser correlations can be severely attenuated.

It can be shown that (i) for residues where the $^{13}C^{\alpha}-^{13}C^{\beta}$ couplings are not refocused, (ii) if $|\nu_{C\alpha} - \nu_{C\beta}| > 150$ Hz (weak coupling limit), and (iii) neglecting relaxation, the resultant signal intensity in this experiment is to good approximation proportional to $\cos^2(\pi J_{C\alpha C\beta} T_c) - {}^1/_2 \sin^2(\pi J_{C\alpha C\beta} T_c)$. The first term in this expression derives from $^{13}C^{\alpha}$ magnetization that effectively remains in-phase with respect to the coupled $^{13}C^{\beta}$ spin during both constant-time periods in the experiment (between 2 and 3 and 4 and 5 in eq 1). The second term originates from $^{13}C^{\alpha}$ magnetization that becomes anti-phase with respect to $^{13}C^{\beta}$ during the first T_c interval, $2C_X^{\alpha}C_Z^{\beta}$, which is

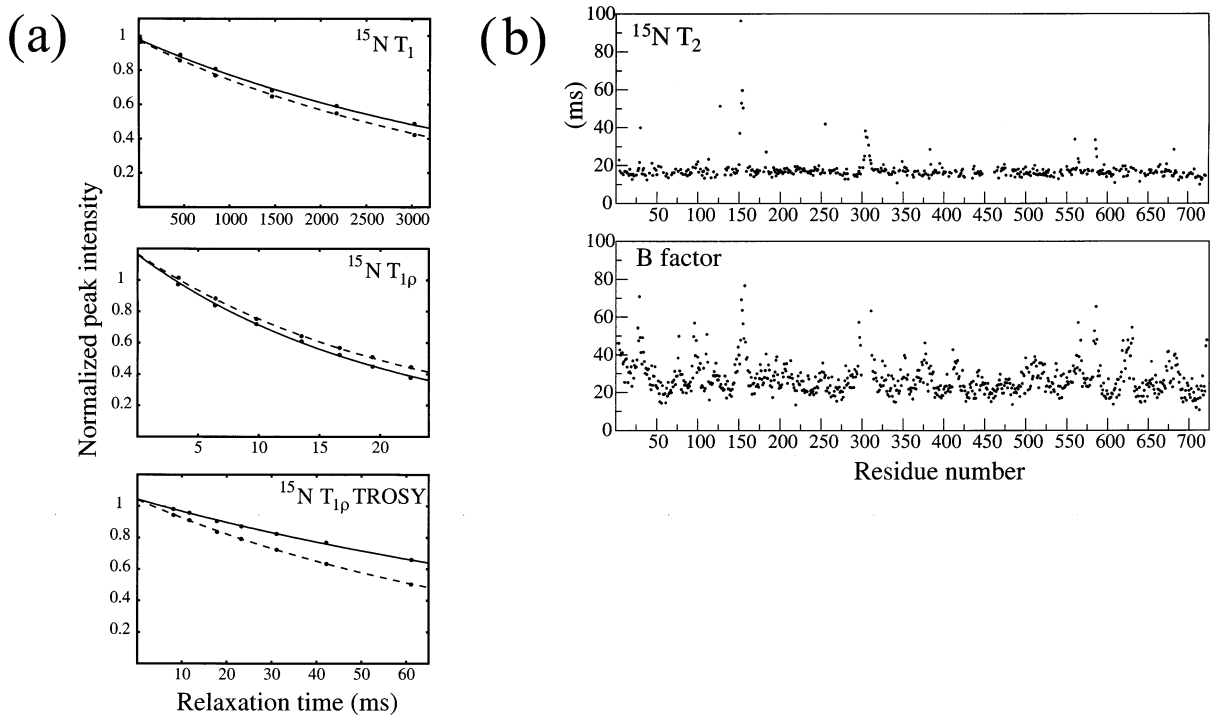

Figure 6. (a) Typical ^{15}N T_1, $T_{1\rho}$, and $T_{1\rho}$ (TROSY component) relaxation curves for residues V194 and A582 of MSG shown with solid and dashed lines, respectively. (b) ^{15}N T_2 values (upper panel) and crystallographic B-factors of backbone nitrogen atoms from the X-ray structure (PDB code 1d8c) of MSG[22] are plotted versus residue number. ^{15}N T_2 values of MSG were derived from $T_{1\rho}$ values corrected for resonance offset effects.[66]

subsequently converted into zero-quantum coherence by the action of pulses and gradients that follow this period and then transferred back to in-phase $^{13}C^{\alpha}$ magnetization during the second T_c period. Because the sign of the second term is opposite that of the first, the signals derived from the two pathways subtract, resulting in very poor sensitivity. For example, for T_c = 9.2 ms, typically used in this experiment (which includes the duration of the selective pulse), $\cos^2(\pi J_{C\alpha C\beta} T_c) = 0.28$, while $-(1/2)\sin^2(\pi J_{C\alpha C\beta} T_c) = -0.36$.

The sensitivity of the 4-D HNCACO can be improved by eliminating the pathway that contributes to the second term, $-(1/2)\sin^2(\pi J_{C\alpha C\beta} T_c)$. This can be achieved by noting that the signal of interest (leading to the $\cos^2$ term) is proportional to 2 $C_Z^{\alpha}CO_Z$ immediately prior to the transfer of magnetization to ^{13}CO, while magnetization that "traverses the zero-quantum pathway" is proportional to $4(C_Y^{\alpha}C_X^{\beta} - C_X^{\alpha}C_Y^{\beta})CO_Z$. Thus, inserting a delay of $1/(2J_{C\alpha CO})$ does not affect $2C_Z^{\alpha}CO_Z$ while $4(C_Y^{\alpha}C_X^{\beta} - C_X^{\alpha}C_Y^{\beta})CO_Z$ is converted to $-2C_X^{\alpha}C_X^{\beta} - 2C_Y^{\alpha}C_Y^{\beta}$ (neglecting chemical shift evolution), which is not refocused to observable magnetization by the subsequent action of the pulse scheme.

Figure 5 shows regions from ^{13}CO−^{1}HN planes of an HN(CA)CO spectrum recorded on maltotriose-bound maltose binding protein[24] using the pulse scheme of ref 16 without (panels a and b) and with (c and d) the delay described above. In the case of Ser 73, the intraresidue correlation is missing, indicating that the intensities of signal from the competing pathways are very similar. In contrast, for Ser 337 the signal from the zero-quantum pathway is stronger, giving rise to a negative intraresidue cross-peak. A more complete description of the magnetization transfer pathways than what is given here shows that the "zero-quantum signal" is relaxed more efficiently than the signal derived from $2C_Z^{\alpha}CO_Z$ (see above), so that depending on the details of relaxation, intraresidue cross-peaks from Ser can be negative, missing, or even positive. In contrast, the interresidue correlations for both of these residues are intense because neither are part of Ser−Ser pairs. Figure 5c and d shows that insertion of the $1/(2J_{C\alpha CO})$ delay does indeed improve the sensitivity of cross-peaks from Ser, although these correlations are still relatively weak because they are attenuated by the factor $\cos^2(\pi J_{C\alpha C\beta} T_c)$. In this regard, a 3-D "test" spectrum recorded on MSG using the improved pulse scheme resulted in an additional 10 intraresidue Ser correlations out of the 22 that were originally missing in the 4-D HNCACO data set. A more complete set of cross-peaks (19 of the missing 22) was obtained in the 3-D HN(CA)CO data set recorded with a shorter T_c value (5.4 ms, see Materials and Methods) and without selective pulses, albeit at the expense of decreased intensity for connectivities involving other residues.

MSG Tumbles as a Single Unit in Solution. ^{15}N T_1 and $T_{1\rho}$ values have been measured for MSG at 37 °C and used to estimate correlation times for each of the domains in the protein. In addition, the ^{15}N relaxation rates of TROSY multiplet components have been quantified as a point of interest, because the experiments used for assignment all rely on the long lifetime of the TROSY signal. Figure 6a shows typical relaxation curves (for V194 and A582), from which ^{15}N T_1, ^{15}N $T_{1\rho}$, and ^{15}N $T_{1\rho}$ (TROSY component) values are extracted. Average values of 3.48 ± 0.74 s, 18.8 ± 4.8 ms, 65.7 ± 16.1 ms, 2.03 ± 0.84 s, and 12.1 ± 3.3 ms are obtained for ^{15}N T_1, ^{15}N $T_{1\rho}$, ^{15}N $T_{1\rho}$ (TROSY component), ^{1}HN T_1, and ^{1}HN T_2 values, respectively. The relaxation times of both ^{15}N and ^{1}HN nuclei were found to be homogeneous throughout the MSG sequence, except for several flexible loops, as illustrated in Figure 6b, where ^{15}N T_2 values are plotted as a function of residue and compared with crystallographic B factors of backbone nitrogen atoms. It is

interesting that regions that are flexible in solution (on the picosecond to nanosecond time scale) also show disorder in the crystal, especially considering that the crystallographic data were collected on frozen samples at −173 °C.

T_1/T_2 ratios have been analyzed using the program R2R1[60] along with the X-ray structure of the protein.[22] Notably, the data for each domain were well fit with an isotropic model, and more complex models were not justified on the basis of F-test statistics. Correlation times of 36.8 ± 2.1, 36.3 ± 2.3, 35.3 ± 1.5, and 36.0 ± 2.1 ns were obtained for the core, N-terminal, α/β, and C-terminal domains, respectively. The X-ray structure shows that all of the domains in the glyoxylate bound form (an apo-structure is not available) make extensive surface contacts with the core of the molecule, with contact surface areas of ∼2300, ∼1330, and ∼2000 Å^2 between the core and the N-terminal α-clasp, α/β domain, and the C-terminal plug, respectively. These close contacts contribute to the overall globular shape of the enzyme (Figure 1a) and also explain why the molecule tumbles as a single entity despite the fact that the domains are connected to the core of the protein through flexible linkers.

Verification of Backbone Assignments. Chemical shift assignments of MSG were based exclusively on spectra recorded with a single ^{15}N, ^{13}C, ^{2}H-labeled sample. The redundancy of information available from the combination of 4-D and 3-D data sets led to a high degree of confidence in the resulting assignments. Nevertheless, because MSG is the largest single polypeptide chain that has been assigned to date by a substantial margin, we felt it important to prepare an additional sample of selectively labeled protein that could be used to verify a large subset of the assignments that had been made. To this end, a uniformly ^{15}N-labeled, highly deuterated sample was prepared that was selectively ^{13}C-labeled at Ile, Val, and Leu residues (159 ^{13}C labeled amino acids). Selective labeling of Ile, Leu, and Val residues has been used in the past in the context of obtaining proton−proton NOE constraints for structure determination of highly deuterated proteins.[8,61−63] In the present case, our goal was to record a series of triple resonance data sets and confirm assignments on the basis of the appearance of expected cross-peaks.

Figure 7 shows a 2-D ^{15}N−^{1}HN plane of the 3-D TROSY−HNCO spectrum recorded using the Ile, Leu, Val-labeled sample. Both this plane and the full 3-D spectrum feature exclusively residues following either Ile, Leu, or Val in the MSG

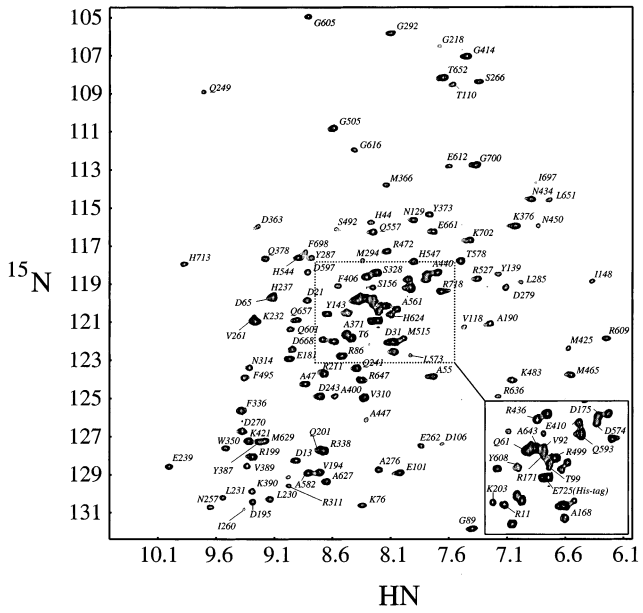

Figure 7. An ^{15}N−^{1}HN 2-D correlation map of malate synthase G selectively labeled with ^{13}C at Ile, Leu, and Val positions recorded using a 3-D TROSY−HNCO pulse scheme.[29,30] Peaks are labeled with residue numbers. Only cross-peaks from those amino acids immediately following Ile, Leu, or Val are observed.

sequence, and it is noteworthy that all of the cross-peaks in the spectrum could indeed be assigned to such residues using the compiled assignment list of the protein. Subsequently, both 3-D HN(CO)CA ($[\omega_{C\alpha}(i-1), \omega_N(i), \omega_{HN}(i)]$) and 3-D HN($CO_{i-1}$)$CA_i$ ($[\omega_{C\alpha}(i-1)/\omega_{C\alpha}(i), \omega_N(i), \omega_{HN}(i)]$) spectra were recorded. The HN(CO)CA data set provides correlations for any residue following Ile, Leu, or Val, while a comparison of both data sets identifies Ile, Leu, or Val residues that are immediately C-terminal to Ile, Leu, or Val. Finally, a modified 3-D TROSY−HNCA spectrum[16,30] was recorded which provides correlations ($[\omega_{C\alpha}(i), \omega_N(i), \omega_{HN}(i)]$) for any Ile, Leu, or Val that follows a residue that is not part of this three amino acid subset. In all of the cases, no contradictions with our assignments were found, and we were able to verify the positions of 218 previously assigned amides corresponding to either Ile, Leu, or Val or to residues immediately following these.

In summary, we have presented near complete backbone ^{15}N, ^{13}C, ^{1}HN, and side-chain $^{13}C^{\beta}$ chemical shift assignments for malate synthase G, an 81.4 kDa, 723-residue single polypeptide chain. The high sensitivity and resolution afforded by four-dimensional TROSY-based triple resonance spectroscopy have been essential for the completion of the task. Chemical shift assignments are a prerequisite for any further studies by NMR, including probing structure, investigating the role of dynamics in protein function, and studying ligand binding. In the specific case of MSG, future work will focus on characterizing how the domain structure of the enzyme changes in response to ligand binding, using residual dipolar couplings as a probe of structure. These investigations will facilitate an understanding of the structural changes that occur along the reaction coordinate. It is clear that NMR can make valuable contributions to detailed structural studies of high molecular weight proteins.

Acknowledgment. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K. We thank Prof. S. J. Remington (University of Oregon) for the

(60) Lee, L. K.; Rance, M.; Chazin, W. J.; Palmer, A. G. *J. Biomol. NMR* **1997**, *9*, 287−298.
(61) Metzler, W. J.; Wittekind, M.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *J. Am. Chem. Soc.* **1996**, *118*, 6800−6801.
(62) Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 7599−7600.
(63) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212.
(64) Gardner, K. H.; Rosen, M. K.; Kay, L. E. *Biochemistry* **1997**, *36*, 1389−1401.
(65) Venters, R. A.; Farmer, B. T.; Fierke, C. A.; Spicer, L. D. *J. Mol. Biol.* **1996**, *264*, 1101−1116.
(66) Peng, J. W.; Wagner, G. *J. Magn. Reson.* **1992**, *98*, 308−332.
(67) Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1990**, *89*, 496−514.
(68) Boyd, J.; Soffe, N. *J. Magn. Reson.* **1989**, *85*, 406−413.
(69) Patt, S. L. *J. Magn. Reson.* **1992**, *96*, 94−102.
(70) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.
(71) Schleucher, J.; Sattler, M.; Griesinger, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1489−1491.
(72) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393−399.
(73) Kay, L. E. *J. Am. Chem. Soc.* **1993**, *115*, 2055−2057.

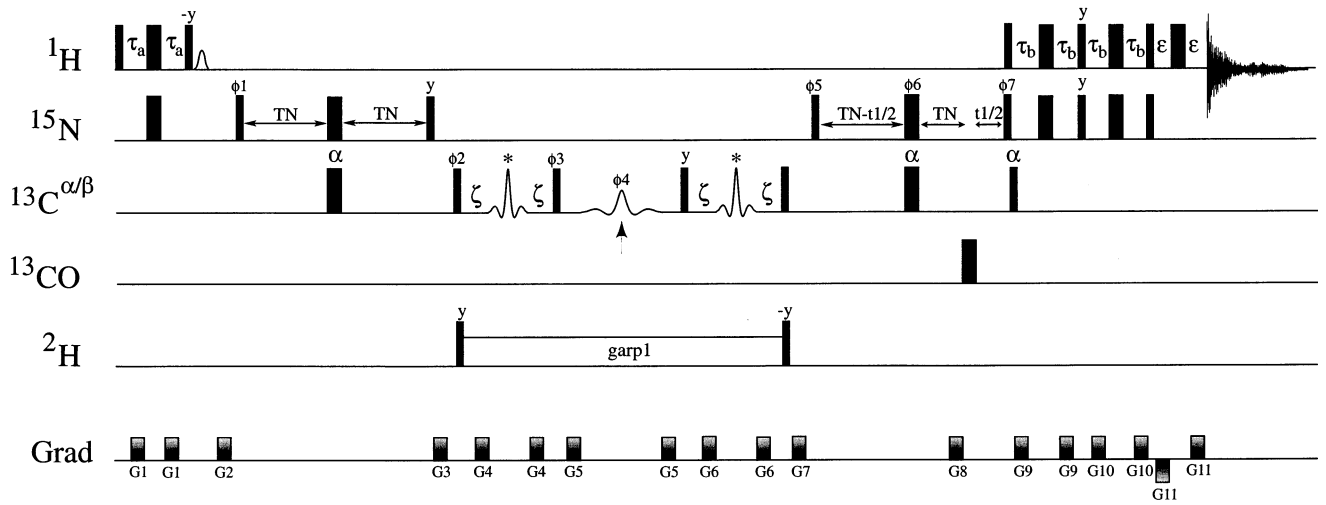

Figure 8. Modified TROSY−HNCACB pulse sequence[16,30] for the selective observation of Ala amide correlations or cross-peaks from residues following Ala. Many of the details pertaining to the experimental setup can be found in Yang and Kay[16] (see Supporting Information of ref 16). Briefly, all narrow (wide) rectangular bars correspond to pulses of flip angle 90° (180°). ^{15}N, $^{13}C^{\alpha/\beta}$, ^{13}CO, and ^{2}H carriers are placed at 119, 45, 176, and 4.7 ppm, respectively. All 90° rectangular ^{13}C pulses are applied at the highest field possible, unless indicated otherwise. 90° (180°) pulses labeled with α are applied with a field strength of $\Delta/\sqrt{15}$ ($\Delta/\sqrt{3}$)[67], where Δ is the difference (in Hz) between $^{13}C^{\alpha}$ (57 ppm) and ^{13}CO (176 ppm) chemical shifts and with the center of excitation at 57 ppm by phase modulation of the carrier.[68,69] Similarly, the 180° ^{13}CO pulse was applied with a field strength of $\Delta/\sqrt{3}$ with the center of excitation at 176 ppm. ^{13}C pulses marked with asterisks have the RE-BURP[32] profile (268 μs at 800 MHz) with the center of excitation at 42 ppm. The phases of these pulses are carefully adjusted for optimum sensitivity. The pulse marked by an arrow is a 2.5 ms (800 MHz) RE-BURP pulse centered at 18.3 ppm and selective for $^{13}C^{\beta}$ of Ala (refocusing bandwidth of ±750 Hz). Delays used are $\tau_a = \tau_b = 2.3$ ms; $T_N = 12.4$ ms; $\zeta = 5.0$ ms; $\epsilon = 250$ μs. All pulses were applied with phase x unless stated otherwise. The phase-cycle employed is $\phi1 = x, -x$; $\phi2 = 2(x), 2(-x)$; $\phi3 = 4(y), 4(-y)$; $\phi4 = 8(x), 8(y), 8(-x), 8(-y)$; $\phi5 = y$; $\phi6 = 4(x), 4(-x)$; $\phi7 = x$; rec = $2(x, -x, -x, x), 2(-x, x, x, -x)$. Quadrature detection in t_1 is achieved by the enhanced sensitivity pulsed field gradient method,[70,71] where for each t_1 value separate data sets are recorded for (g8, $\phi7$) and (−g8, $\phi7$ + 180°). For each successive t_1 value, $\phi5$ and the receiver phase are incremented by 180°.[72] The durations and strengths of the gradients are g1 = (0.5 ms, 5 G/cm); g2 = (0.8 ms, 15 G/cm); g3 = (0.9 ms, 10 G/cm); g4 = (0.1 ms, 15 G/cm); g5 = (0.15 ms, 10 G/cm); g6 = (0.1 ms, 15 G/cm); g7 = (0.4 ms, 15 G/cm); g8 = (1.25 ms, −30 G/cm); g9 = (0.4 ms, 3.1 G/cm); g10 = (0.4 ms, 5.35 G/cm); g11 = (62.5 μs, 28.75 G/cm). Decoupling was interrupted prior to application of the gradients.[73]

gift of the plasmid used for MSG expression and Prof. Cheryl Arrowsmith (Ontario Cancer Institute) for kindly allowing us to carry out the protein expression and purification in her laboratory. We also thank Dr. Frans Mulder (Lund University, Sweden) for NMRView tcl scripts for analysis of 4-D NMR data, Prof. J. Forman-Kay (Hospital for the Sick Children, Toronto) for useful advice on refolding in vitro, Mr. Peter Hwang (University of Toronto) for help with identifying this protein, and Dr. Oscar Millet (University of Toronto) for many helpful discussions. V.T. is a recipient of a Human Frontiers Postdoctoral Fellowship. A.A. acknowledges a fellowship from the National Cancer Institute of Canada. L.E.K. holds a Canada Research Chair.

Appendix

Figure 8 shows the Ala-HSQC pulse scheme that provides ^{15}N, ^{1}HN correlations of Ala or residues immediately following Ala. The pulse sequence is derived from the TROSY−HNCACB sequence[16,30] with magnetization transfer extending from ^{1}HN → ^{15}N → $^{13}C^{\alpha}$ → $^{13}C^{\beta}$ and then subsequently back via the reverse of the pathway. When the magnetization resides on $^{13}C^{\beta}$, a refocusing pulse which is selective for methyl carbons of Ala is applied (2.5 ms RE-BURP pulse,[32] 800 MHz, centered at 18.3 ppm). Magnetization from all other residues is effectively eliminated by phase cycling this pulse and by the action of the flanking gradients.

Supporting Information Available: One table listing the experimental parameters used for the NMR experiments described in the text and a table listing the ^{1}HN, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts of malate synthase G at pH 7.1, 37 °C (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0205636