Tollin, G., & Edmondson, D. E. (1980) *Methods Enzymol. 69*, 392-406.
Vervoort, J., Müller, F., Mayhew, S. G., van den Berg, W. A. M., Moonen, C. T. W., & Bacher, A. (1986) *Biochemistry 25*, 6789-6799.
Watenpaugh, K. D., Sieker, L. C., Jensen, L. H., Legall, J., & Dubourdieu, M. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3185-3188.
Witanowski, M., Stefaniak, L., & Webb, G. A. (1981) *Ann. Rep. NMR Spectrosc. 11B*, 1-502.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, pp 1-292, Wiley, New York.

NMR Sequential Assignment of *Escherichia coli* Thioredoxin Utilizing Random Fractional Deuteriation†

David M. LeMaster* and Frederic M. Richards
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511
Received June 29, 1987

ABSTRACT: All non-proline residues except for the N-terminal dipeptide have been assigned in the 108-residue protein *Escherichia coli* thioredoxin. Central to these experiments has been the use of protein samples in which all carbon-bound hydrogen positions are substituted to 75% with deuterium by bacterial growth on partially deuteriated carbon sources and media. The dilution of the local proton density gives rise to narrower line widths with little loss in sensitivity. In addition, passive or secondary coupling to protons not directly involved in the coherence transfer process of correlation experiments is largely suppressed, thus significantly improving the resolution for side-chain couplings. Simultaneous multiresidue-type assignments have been obtained by incorporation of several amino acids with differing selective α- and/or β-deuteriation into a fractionally deuteriated background. Combined with several single residue type labeling experiments, these selective labelings have yielded direct residue type assignments for two-thirds of the protein. In addition to improved resolution, the amide to carbon-bound proton NOESY spectra offered equivalent sensitivity while the amide to amide NOESY spectra offered superior sensitivity to that observed for natural abundance samples. The resultant sequential assignment has an average number of nearest-neighbor NOE connectivities of 2.35 out of the possible 3 α-amide, β-amide, and amide-amide connectivities.

In the 5 years since the first protein NMR sequential assignment was reported (Wagner & Wüthrich, 1982), application of this technique to significantly larger proteins has proceeded rather slowly. As anticipated a number of years ago (Jardetzky et al., 1978), 10 kdaltons has remained an effective upper boundary for assignment in the absence of a high-resolution X-ray structure. Even in cases where an X-ray determination is available, progress on larger proteins has been slow although recent multiple quantum experiments are encouraging (Dalvit & Wright, 1987). The difficulty is easily rationalized in light of the fact that both resolution and sensitivity of 2D NMR analysis decrease as approximately the cube of the molecular weight. The most straightforward view of resolution in the contour plots standardly used in 2D analysis is that of total occupied area. The number of cross-peaks is generally proportional to the molecule weight. The average cross-section of individual cross-peaks is roughly proportional to the square of the molecular weight because the correlation time of the protein, and therefore the T_2 relaxation rates of most protons, increases nearly linearly with molecular weight, and hence, the effective line widths are broadened in both dimensions of the 2D experiment. The same factors cause the decrease in sensitivity since broadening of the peak implies a decrease in the peak height which determines sensitivity in the contour plot. Furthermore, the practical millimolar concentrations tend to decrease with increasing molecular weight.

Various aspects of particular experiments will modify the details of this first-order analysis. For instance, sensitivity of a NOESY experiment will be somewhat less dependent on molecular weight while that of a phase-sensitive COSY will have a somewhat higher dependence. Nevertheless, in the absence of a major increase in available field strength, extension of detailed assignment experiments to significantly larger proteins will require direct attack on the resonance number and resonance line width problems.

Two approaches to spectral editing are available: (1) pulse sequences which select subsets of all the spin coupling patterns present and (2) selective isotope labeling. Pulse sequences, such as the spin topological filtration technique (Levitt & Ernst, 1985), can in principle edit out the various different side-chain coupling systems in proteins. However, sensitivity problems resulting from the shorter T_2 relaxation times seen in moderate-sized proteins are presently a significant limitation on these techniques. Isotope labeling has proven better suited for spectral editing of these proteins. A number of protein NMR experiments using ^2H, ^{13}C and ^{15}N labeling have been reported [e.g., Crespi et al. (1968), Markley et al. (1968), Chan and Markley (1982), Griffey et al. (1985), LeMaster and Richards (1985), and McIntosh et al. (1987)]. However, these isotopic labeling experiments have not played a significant role in any of the sequential assignments previously published.

The resonance line width problem requires reduction of the proton T_2 relaxation rates. For carbon-bound protein protons (and to a lesser extent nitrogen-bound protons) relaxation is predominantly due to ^1H-^1H dipolar interaction. Isolation of individual protons by selective deuteriation has been seen to give rise to narrower resonances in a number of protein experiments. However, besides a brief 1D NMR study of 90% ^2H-labeled elongation factor TU (Kalbitzer et al., 1985), little

† This investigation was supported by Program Grant GM-22778 from the National Institute of General Medical Sciences and by Project Grant PCM-8305203 from the National Science Foundation.

0006-2960/88/0427-0142$01.50/0

attention has been given to the idea of uniform isotopic dilution with deuterium as a means of narrowing all proton resonances. In part, this is due to concerns about the effect on sensitivity particularly in a 2D experiment where resonance intensity would be expected to decrease by roughly the square of the isotopic dilution factor. However, as noted above, it is peak height rather than integrated intensity which determines sensitivity, and the resultant line narrowing largely compensates for the loss in integrated intensity.

We have chosen to carry out NMR experiments on *Escherichia coli* thioredoxin since it offers a number of benefits as a model system for protein dynamics studies. An X-ray structure determination is available on the oxidized form (Holmgren et al., 1975). Two independent reconstitution systems have been established on the basis of cleavage of the sole methionine (Holmgren & Slaby, 1979) and sole arginine (Slaby & Holmgren, 1979). A number of other solution studies have been carried out including stopped-flow folding kinetics measurements, which indicate at least three kinetic phases (Kelley & Stellwagen, 1984). Site-directed mutagenesis has served to assign the slowest phase to the trans to cis isomerization of proline-76 (Kelley & Richards, 1987). Additional structural mutants and overproduction plasmids have been constructed suitable for the isotopic labeling experiments desired.

To obtain a suitable balance between reduced line widths and reduced sensitivity, we have prepared 75% ^{2}H-labeled *E. coli* thioredoxin by growth of an overproducing strain on partially deuteriated alanine and succinate in an H_2O/D_2O-containing medium. The resolution obtained is significantly better than that observed for a natural abundance sample while the sensitivity for amide (~95% ^{1}H)–amide NOESY is better than, and for amide–carbon-bound proton 2D data is equal to, that of the natural abundance sample. Even for cross-peaks between carbon-bound proton sites, the sensitivity is only reduced by roughly a factor of 3. The resultant data when combined with selective isotopic labeling have proven adequate to obtain the sequential assignment of this 108-residue protein.

Materials and Methods

Deuteriated Amino Acid Synthesis. For the single residue type enrichment experiments, uniformly 99% ^{2}H-labeled aspartic acid, isoleucine, leucine, lysine, and threonine were prepared in collaboration with the Stable Isotope Resource of Los Alamos by growth of the methanol-utilizing strain *Methylophilus methylotrophus* in a deuteriated medium (manuscript in preparation). The crude protein fraction was isolated and hydrolyzed, and the constituent amino acids were separated according to published procedures (LeMaster & Richards, 1982a). Pyridoxal-catalyzed exchange at pH 5 was used to produce α,β-deuteriated phenylalanine, isoleucine, and leucine (Tenenbaum et al., 1974; LeMaster & Richards, 1982b). Acetic anhydride catalyzed exchange was used to produce α-deuteriated tyrosine and valine as well as to back-exchange the α,β-deuteriated isoleucine to the β-deuteriated form (Upson & Hruby, 1977). Pyridoxal-catalyzed exchange at pH 10 and 50 °C for 2 days was used to prepare α-deuteriated serine. Acid-catalyzed exchange was used to prepare β-deuteriated aspartic acid (LeMaster & Richards, 1982b) as well as to exchange the aromatic ring positions to synthesize perdeuteriated phenylalanine and [α,ϵ,ϵ-$^2H_3$]tyrosine (Matthews et al., 1977).

Construction of Transaminase-Deficient Auxotrophic E. coli Strain DL39. Our earlier ^{15}N labeling experiments (LeMaster & Richards, 1985) utilized DL30, a polyauxotrophic derivative of DG30 (Gelfand & Steinberg, 1977), which carries mutations in the three glutamate-dependent general amino acid transaminase genes ilvE, tyrB, and aspC. The poor growth characteristics of DL30 combined with the demanding growth conditions needed for these labeling experiments prompted us to transduce the three transaminase lesions from DG30 into the prototrophic strain MG1655 (Guyer et al., 1981). The ilvE lesion was moved in by use of a selectable metE::Tn10 marker (from RK4349-CGSC strain 6403) and then transduced to met$^+$. The tyrB lesion was moved in by use of a malB::Tn10 marker (from TST1-CGSC strain 6137) and then transduced to mal$^+$. Finally, aspC was transduced in by use of AroA::Tn10 (from LCB273-CGSC strain 6542), and then the strain was transduced to aro$^+$. These lesions give rise to requirements for phenylalanine, tyrosine, aspartic acid, leucine, isoleucine, and valine (leaky). The additional auxotrophies required were obtained by transduction into the DL39 background amino acid biosynthesis loci mutated by Tn10 or Tn5 transposon insertions.

Construction of an E. coli and Phage T4 Thioredoxin Overproduction Plasmid. In order to obtain a high-expression plasmid compatible with the various bacterial strains anticipated in the labeling experiments, the temperature-sensitive λ repressor C_{1857} was cloned into the pUC18 plasmid along with the leftward promoter P_L and antiterminator N gene so as to make the expression system genetically mobile. The *Bgl*II–*Taq*I restriction fragment containing the C_I gene was blunt end ligated into the *Eco*RI site in a clockwise orientation while the *Sau*3A fragment containing the P_L promoter and N gene was cloned into the *Bam*H1 site in a counterclockwise orientation. The trxA (*E. coli* thioredoxin) containing *Hin*cII fragment from pBHK8 (Lunn et al., 1984) was cloned into the *Hin*cII site, and the nrdC (phage T4 thioredoxin) containing *Hae*III fragment (LeMaster, 1986) was blunt end ligated into the *Xba*I site, both in counterclockwise orientation, to yield the plasmid pDL59.

Growth Medium. For bacterial growth, 82%-labeled [^{2}H]succinic acid and 85%-labeled [^{2}H]-DL-alanine were used as carbon sources. The labeled succinic acid was prepared via two exchanges of succinic anhydride in deuteriated acetic acid (Stella, 1973; LeMaster & Richards, 1982a). The alanine was prepared by pyridoxal exchange as for the other amino acids discussed above with the variation that a 4 molal solution of alanine was exchanged at reflux for 2 days under nitrogen. M63 salts (Miller, 1972) were dissolved in 8 L of 85% ^2H_2O (kindly supplied by P. Moore) supplemented with 32 g of 82%-labeled sodium [^{2}H]succinate and 20 g of 85%-labeled [^{2}H]-DL-alanine. Five to ten percent isotopic selection against deuterium was observed to occur during incorporation into protein.

Purification of E. coli and Phage T4 Thioredoxins. Bacterial strains MG1655/pDL59 and DL39/pDL59 were grown at 30 °C in the deuteriated medium until an OD_{600} of 1.0 at which time the temperature was shifted to 42 °C and incubated for 3 h. For the single residue type labeling experiments, a highly supplemented protonated medium was used (LeMaster & Richards, 1985), and growth was carried out at 33 °C followed by a 3-h induction at 43 °C.

The cells were pelleted, resuspended in 3 volumes of 50 mM sodium phosphate, pH 7.0, and lyzed by sonication. After centrifugation, streptomycin sulfate was added to the supernatant to a final concentration of 1% (Holmgren & Reichard, 1967). The next supernatant then was treated with ammonium sulfate to isolate the 35%–75% fraction. The ammonium sulfate pellet was resuspended in 50 mM NH_4HCO_3 to a final volume of 40 mL and loaded on a 2.5 × 100 cm Ultragel

AcA54 column equilibrated in the same buffer (Lunn et al., 1984). Material from no more than roughly 20 g wet weight of cells is loaded on the sizing column at one time. The *E. coli* and phage T4 thioredoxin containing fractions were pooled and loaded onto a DE52 column equilibrated in 50 mM NH_4HCO_3 and eluted in a gradient to 400 mM NH_4HCO_3. T4 thioredoxin elutes at roughly 125 mM while *E. coli* thioredoxin elutes at 250 mM NH_4HCO_3. Approximately 25 mg and 15 mg per liter of *E. coli* and T4 thioredoxins, respectively, are obtained from growth on protonated media and roughly half as much during growth on deuteriated media.

NMR Data Collection and Analysis. Five to eight millimolar samples of *E. coli* thioredoxin in 150 mM NaCl and 20 mM phosphate, pH 5.7 and pH 2.8, were used in data collection on a Bruker WM500 at the Yale Chemical Instrumentation Center. Data collection times ranged from 15 h for the single residue type deuteriation experiments to 60 h for the NOESY experiments on the random fractionally deuteriated samples. Data were standardly collected to 4.2 Hz in the T_2 dimension and 6.0 Hz in the T_1 dimension. The data were multiplied by an unshifted skewed sine bell squared in the T_2 dimension and unshifted skewed sine bell in the T_1 dimension before transformation with the software written by D. Hare. Relaxation delay + acquisition times of 1.3 and 1.8 s were used for the protonated and fractionally deuteriated samples in ^1H_2O. Solvent saturation was applied continuously except during acquisition. In ^2H_2O, 1.8 and 3.3 s were used for these samples. A mixing time of 250 ms was used in the NOESY experiments for the random fractionally deuteriated sample.

Results and Discussion

Intraresidue Assignments. In the standard sequential assignment procedure, spin coupling mediated experiments such as COSY, coherence relay, and multiple quantum filtration are used to establish the intraresidue spin connectivities. In addition to identifying which resonances belong to the same side chain, the coupling patterns serve to provide partial residue-type assignments since several amino acids give rise to characteristic coupling patterns. On the other hand an AMX system is obtained for serine, cysteine, asparagine, and aspartic acid as well as all four aromatic amino acids since coupling is not observed between the β proton and ring protons in protein spectra. Generally these residue types cannot be unambiguously assigned with these experiments.

Isotopic labeling provides a more flexible approach to obtaining residue-type assignments since essentially all amino acids have been demonstrated to be selectively incorporated into protein with appropriate auxotrophs and growth conditions. In order to avoid label dilution or scrambling, particularly of the α position, all of the specific enrichment experiments reported here have been carried out with derivatives of the glutamate-dependent general amino acid transaminase deficient strain DL39 into which the *E. coli* thioredoxin overproducing plasmid pDL59 has been transformed. One of the simplest forms of selective labeling is obtained by incorporating one residue type labeled with deuterium for which the appropriate amino acid can readily be purchased or produced by chemical or biosynthetic means. When compared to natural abundance spectra, the resultant 2D spectra lack cross-peaks whenever either of the two interacting protons of that cross-peak are replaced with deuterium. In Figure 1 are illustrated phase-sensitive COSY spectra of the amide–α region of *E. coli* thioredoxin in which either [U-^{2}H]aspartic acid or [α-^{2}H]serine has been incorporated. The crosses mark positions where cross-peaks have been eliminated as can be verified

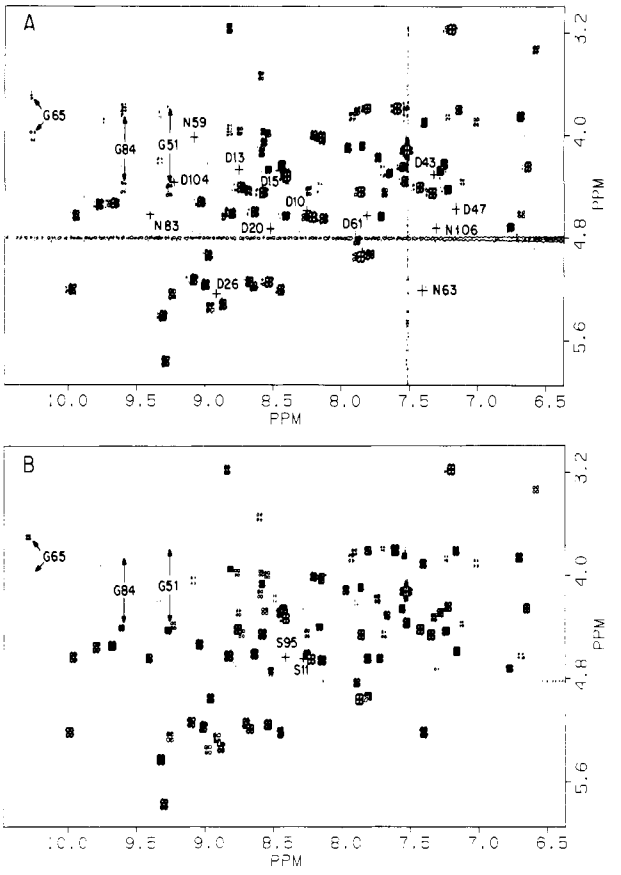

FIGURE 1: COSY analysis of *E. coli* thioredoxin labeled with [α,-β,β-$^2H_3$]aspartic acid and [α-^{2}H]serine. In the phase-sensitive spectra of the amide–α region crosses mark positions in which cross-peaks have been eliminated by selective deuteriation of aspartic acid in panel A and serine in panel B. The peaks missing in panel A can be seen in panel B and vice versa. The glycine biosynthetically derived from [α-^{2}H]serine is stereospecifically deuteriated in the *R* position resulting in elimination of passive coupling and hence signal enhancement of the *S* proton resonance. This effect is demonstrated for the downfield resonances associated with glycines-51, -65, and -84. The arrows point to the Nα and Nα' cross-peaks in panel A; the altered intensities in the equivalent positions can be seen in panel B.

by reference to the accompanying spectrum. Also marked are the cross-peaks for three of the downfield glycine residues. [α-^{2}H]Serine gives rise biosynthetically to glycine deuteriated in the *R* position (Jordan & Akhtar, 1970) thus resulting in loss of that resonance. In addition, the *S* proton resonance is intensified in the [α-^{2}H]serine-labeled spectrum due to the removal of the passive coupling to the geminal hydrogen.

Although such experiments are straightforward to execute and interpret, it would clearly be more desirable to obtain residue specific assignment information on several residue types at one time. The amide–α COSY, α–β COSY, and intraresidue amide–β NOESY cross-peaks can be observed for essentially all appropriate residues in a protein. Therefore, differential labeling at the amide, α and/or β positions permits several labeled residue types to be assigned simultaneously.

By incorporating several amino acids with various α,β-deuteriation patterns into a 75% random deuteriated background, we can observe both peak diminution and augmentation. Figure 2B gives the magnitude COSY spectra of a portion of the amide–α region for such a sample as compared to a reference sample uniformly labeled to 75% ^{2}H throughout, illustrated in Figure 2A. Leu-107 and Tyr-70 are strongly reduced due to incorporation of the α-deuteriated samples. Isoleucine (I4 and I38) and aspartic acid (D43) are increased in intensity relative to the 75% ^{2}H-labeled sample as a result

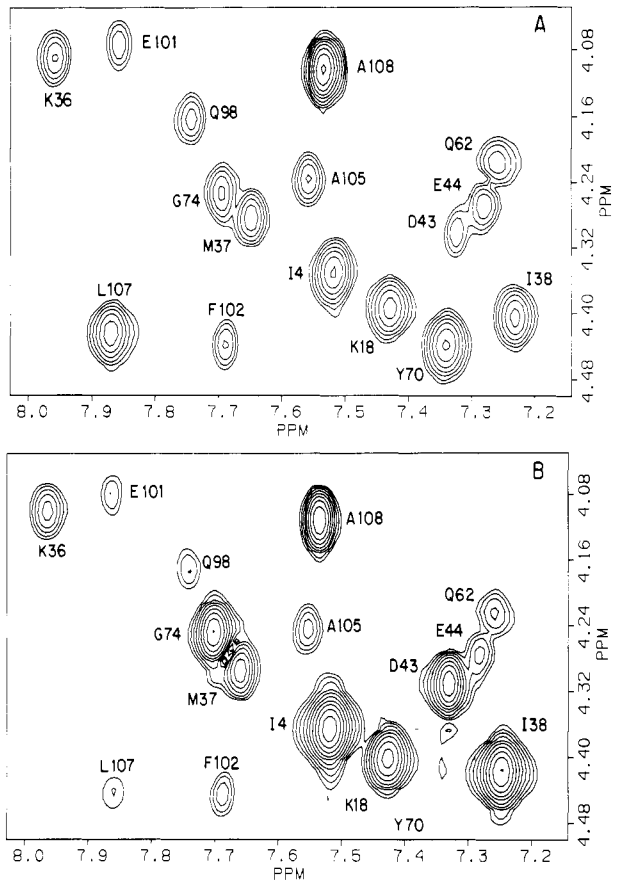

FIGURE 2: Multiresidue-type assignment by selective $\alpha\beta$-deuteriation. In panel A a portion of the amide–α region is displayed in magnitude mode for a random fractionally deuteriated *E. coli* thioredoxin sample. A second sample was prepared in which selectively α,β-labeled amino acids are incorporated in a 75% ^2H background. [α-^2H]Tyrosine and [α,β,β-$^2H_3$]leucine result in the strong reduction of the corresponding cross-peaks in panel B (Y70, L107) while [α-$^1H,\beta,\beta$-$^2H_2$]aspartic acid and -isoleucine, and [α,β,β-$^1H_3$]serine (hence glycine) yield significantly stronger peaks (D43, I4, I38, G74) than are observed in the reference spectrum. The resonances of all other residue types in the region are the same within one contour level (1.4 scale factor) for the two spectra.

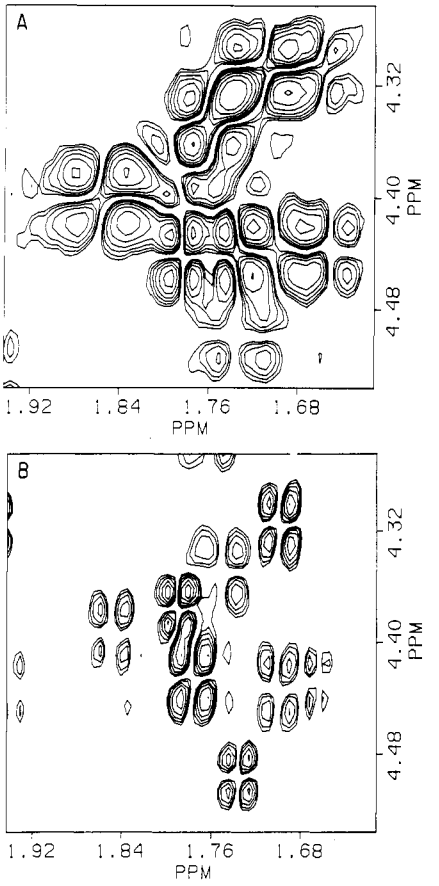

FIGURE 3: Elimination of passive spin coupling by random fractional deuteriation. In panel A is shown a section of the $\alpha\beta$ COSY region presented in phase-sensitive mode for a natural abundance *E. coli* thioredoxin sample. The corresponding spectrum for the 75% uniformly ^2H-labeled sample is displayed in panel B under identical acquisition and analysis parameters. The improved resolution observed in the fractionally deuteriated sample is partly due to the longer T_2 relaxation time and hence narrower line width, but mainly it is a result of the fact that coupling from other protons to the two protons involved in the coherence transfer is largely eliminated due to the isotopic dilution.

of incorporating α-$^1H,\beta$-^2H-labeled amino acids. Glycine-74 is likewise enhance due to incorporation of natural abundance serine. Note the chirality of labeling of the glycine is opposite to that discussed above.

The α-$^1H,\beta$-^2H labeling of aspartic acid gives rise to similarly labeled threonine, which served to resolve a standard ambiguity in resonance assignments. The $\beta\gamma$ COSY cross-peaks of threonine are often hard to distinguish from the $\alpha\beta$ COSY cross-peaks of alanine. By β-deuteriation of threonine the $\beta\gamma$ COSY cross-peaks are eliminated, thus allowing for facile assignment of the 12 alanine residues.

On the other hand, Figure 2 illustrates that α exchange during biosynthetic incorporation is not completely suppressed in these experiments. Perdeuteriated phenylalanine was also used in this experiment, yet the F102 cross-peak is virtually undiminished. Lysine and methionine are biosynthetically derived from aspartic acid; however, isotopic exchange of the α-hydrogen has clearly occurred during synthesis since the corresponding intensities of K18, K36, and M37 are nearly the same. Rather than being a nuisance, this selective biosynthetic back-exchange can be exploited to further broaden the potential for residue-type assignment. Clearly the most important extension of the multiresidue-type assignment experiment is the use of ^{15}N labeling (Senn et al., 1987). It is likely that ^{15}N-labeled aspartate would serve to enrich lysine, methionine, and threonine as well while the α-hydrogen back-exchange would serve to distinguish between these residue types.

The benefits of random fractional deuteriation become more evident in the COSY spectra as one looks out along the side chains. For proteins larger than ten kilodaltons the resolution in the aliphatic region is generally poor. This is largely a result of resonance broadening due to passive or secondary coupling to protons not directly involved in the coherence transfer process, analogous to the splitting one sees in 1D spectra of highly coupled systems. The secondary coupling is largely suppressed by partial deuteriation. In Figure 3 is shown a portion of the $\alpha\beta$ COSY region of a natural abundance spectrum and the corresponding region in a 75% ^2H-labeled sample. The spectral overlap in the natural abundance spectrum makes deconvolution of the individual cross-peaks problematical while in the fractionally deuteriated sample the analysis is straightforward. The cross-peaks between carbon-bound hydrogens represent the worst loss of sensitivity seen in this work. But although the peak heights are down by a factor of roughly 3, resolved $\alpha\beta$ coupling cross-peaks are observed for all but four residues (T8, W28, T66, and V86), and only for Trp-28 does this appear to reflect a problem of sensitivity.

Interresidue Assignments. The second half of the sequential assignment technique involves establishment of NOE con-

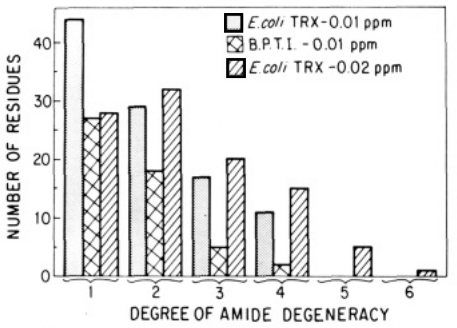

FIGURE 4: Amide degeneracy as a function of molecular weight and resolution. Illustrated are the number of cases for which multiple amide resonances of *E. coli* thioredoxin (101 assigned amide protons) occur within ±0.01-ppm or ±0.02-ppm intervals, which range from the nondegenerate case up to six amides with chemical shifts within ±0.02 ppm. As reference for the size dependence, the corresponding analysis for 52 amide protons of bovine pancreatic trypsin inhibitor is given.

nectivities between the amide, α and/or β protons of one residue and the amide proton of the following residue. This process is repeated all along the protein backbone. Establishment of a complete self-consistent set of sequential NOE connectivities is the most difficult part of the assignment process. This is mainly due to the problem of amide chemical shift degeneracy which becomes quite severe for proteins of this size. When two amide protons are degenerate, it is not possible to make specific interresidue connectivity assignments without appealing to the constraints of the residue-type assignment information discussed above, which is generally incomplete.

The problem of amide degeneracy increases with molecular weight due to both increased resonance number and decreased resolution. For reference, the degree of degeneracy for the 52 reported amide resonance chemical shifts of bovine pancreatic trypsin inhibitor at pH 4.6, 68 °C (Wagner & Wüthrich, 1982), are given in the histograms of Figure 4, assuming a 0.01 ppm window. Over half of the amide resonances are nondegenerate while over one-third have only one other amide resonance within ±0.01 ppm. In contrast for the 101 amide assigned protons of *E. coli* thioredoxin at 30 °C, pH 5.7, assuming a 0.02 ppm window, only one-fourth of the amide protons are nondegenerate while up to six amide protons are found to lie within a ±0.02 ppm region. The effect of improved resolution obtained by random fractional deuteriation is illustrated by considering the same *E. coli* thioredoxin data with a 0.01 ppm window. Although clearly still more than twice as severe as the BPTI example, the improvement in the amide degeneracy problem compared to the 0.02 ppm case is appreciable.

Random fractional deuteriation offers an additional benefit for NOE experiments since longer mixing times can be used to allow for NOE buildup. If the delay between the second and third pulse of the NOESY sequence is too long, magnetization will pass not only to spatially adjacent protons as desired but will pass to the next layer of protons as well, the so-called spin–diffusion effect. In the random fractionally deuteriated sample this effect is strongly curtailed since the probability of any such tertiary interaction will be reduced by the degree of deuteriation. The benefits of the longer mixing times and narrower line widths is particularly evident in the case of amide–amide NOE's. Here both sites are essentially fully protonated, yet the resonances still benefit from the reduced ^1H–^1H dipolar interactions. Sensitivity is significantly improved over the natural abundance sample, which has proven to be of considerable value in the sequential assignment.

The amide degeneracy problem is not quite as severe as the histogram of Figure 4 indicates since, for proteins which can tolerate the conditions, amide protons are more sensitive to pH and temperature variations than carbon-bound protons. Thus some amide degeneracies can be overcome without large shifts along the second axis, which otherwise complicates interpretation. In addition at low pH, data can be collected in D_2O so as to only observe connectivities involving slowly exchanging amides.

Nevertheless since no such technique has been able to overcome all amide degeneracies, sequential analysis involves following numerous short hypothetical connectivity pathways through the NOE data and then checking against the residue-type assignment to see if any of the paths are consistent with the protein sequence. The consistent paths are then extended and linked together so as to eventually cover the entire sequence. Given the level of degeneracy seen for this protein, it becomes impractical to carry out an exhaustive path search manually. We have developed software to assist in this analysis, which will be described in more detail elsewhere; however, the basic idea is straightforward. Amide NOE cross-peak chemical shifts and intraresidue cross-peak assignment information for pH 5.7 at 30 and 50 °C were used as input. With the intraresidue assignment information, all possible N_i–N_j, α_i–N_j, and β_i–N_j NOE's were examined. If the corresponding position in the NOE data set has a cross-peak within a defined error limit (usually 0.01 or 0.02 ppm), this is considered a possible sequential connectivity. These pairwise connectivities were then concatenated into strings, and the residues previously assigned by residue type are matched against the protein sequence.

The level of confidence that any particular string of connectivities is correctly assigned is determined primarily by the number of NOE connectivities seen for each peptide bond and the pattern of intensities seen along the chain. Secondary structural elements give rise to characteristic patterns of intensities. α-Helical regions are characterized primarily by sequential strong amide–amide NOE connectivities, while β sheets give rise to sequential strong α–amide NOE peaks. As with other proteins, this study of *E. coli* thioredoxin progressed by establishing confident assignments in several secondary structural regions and then expanding the assignments at the ends of these segments until these regions meet.

In Figure 5 is given the data at pH 5.7 and 30 °C used in the sequential assignment in which the intraresidue and interresidue connectivities for the sequence from aspartate-15 to aspartate-26 are indicated. This sequence covers the second half of the first α helix, a turn region, and half of the second β strand. The secondary structure is clearly reflected in the N–N NOESY data in which the first few residues of the D15 to D26 sequence have strong cross-peaks while the β region gives weak cross-peaks. On the other hand, the more downfield β sheet resonances have strong α–amide NOESY cross-peaks.

In Table I are listed the chemical shifts of all non-proline positions except for the first two residues. Isotopic labeling experiments directed toward proline sequential assignments will be considered in a later paper although it should be noted that three of the five prolines are traversed by either N_i–N_{i+2} or α_i–N_{i+3} connectivities. Sixty-eight residues were assigned by type via selective deuteriation, which provides unambiguous identification of the amide. α, and usually β positions, some β cross-peaks being obscured by overlap in samples with selective deuteriation in natural abundance backgrounds. All β protons giving rise to observable $\alpha\beta$ COSY cross-peaks also were found to give observable intraresidue Nβ NOESY

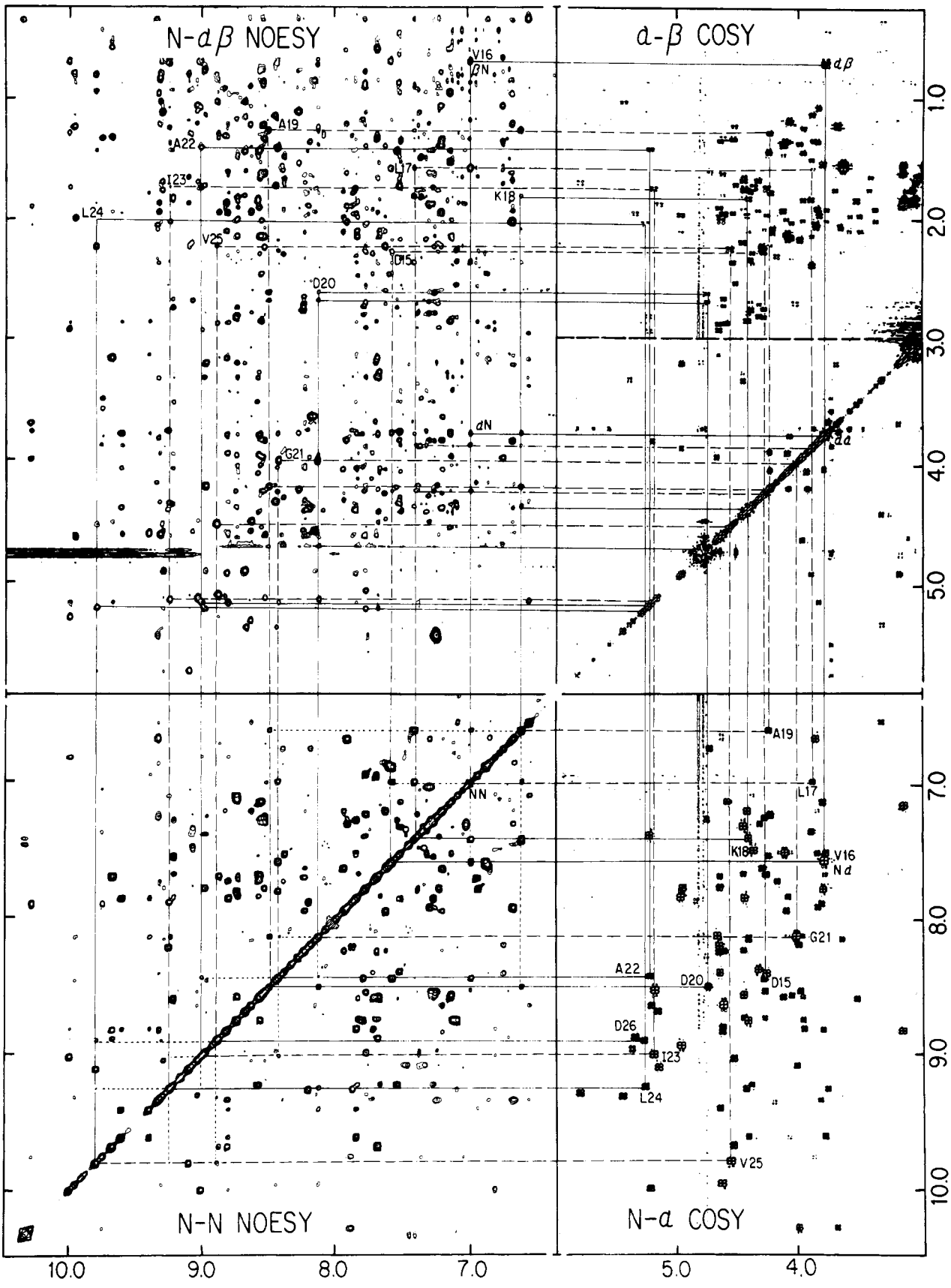

FIGURE 5: Sequential assignment of *E. coli* thioredoxin. NOESY and COSY data for the random fractionally (75%) deuteriated sample at pH 5.7, 30 °C, are shown with an 11-residue sequential pathway explicitly illustrated. The lower right-hand panel is the N_i–α_i COSY data. In the upper right-hand panel is α_i–β_i COSY data. The part below the dashed line is substituted by a symmetrized double quantum filtered COSY spectrum collected in ^2H_2O for ease of following the connectivity pathways near the diagonal. Both the α_i–N_{i+1} and β_i–n_{i+1} NOESY connectivities are shown in the upper left panel. The N–N NOESY data in the lower left serves to connect the upper left and lower right panels, and the N_i–N_{i+1} cross-peaks are shown by the dotted lines. For the set of lines leading from valine-16 to leucine-17, the atom pairs contributing to each cross-peak are labeled. Of the 32 potential pairwise interresidue NN, αN, and βN NOESY connectivities in this 11-residue sequence (the one glycine has no βN connectivity), only the N_{I23}–N_{L24} and β_{L24}–N_{V25} connectivities are not observed. For the residue types possessing β-methylene protons one often can observe αβ COSY and βN NOESY cross-peaks for both β protons as illustrated for Asp-20.

Table I: Resonance Assignments of *E. coli* Thioredoxin at 30 °C and pH 5.7

		chemical shift (ppm)					chemical shift (ppm)		
		amide	α	β			amide	α	β
1	Ser				55	Val[a]	9.96	4.63	2.00
2	Asp				56	Ala	9.32	5.41	1.03
3	Lys[a]	8.41	4.32	1.68, 1.92	57	Lys[a]	8.70	5.14	1.60
4	Ile[a]	7.55	4.38	1.74	58	Leu[a]	8.83	4.62	1.32
5	Ile[a]	8.46	4.23	1.74	59	Asn	9.09	4.00	2.21, 2.69
6	His	8.99	4.93	3.16, 3.21	60	Ile[a]	8.56	4.27	2.24
7	Leu[a]	8.70	4.43	1.81	61	Asp[a]	7.80	4.64	2.71, 2.92
8	Thr	8.12	4.82	4.52	62	Gln	7.26	4.23	1.78, 2.22
9	Asp	8.26	4.74	2.75, 2.80	63	Asn[a]	7.41	5.19	2.72, 2.80
10	Asp[a]	8.26	4.59	2.65, 2.71	64	Pro			
11	Ser[a]	8.26	4.63	4.36	65	Gly[a]	10.28	3.70, 4.00	
12	Phe	7.73	3.69	3.20, 3.33	66	Thr[a]	7.90	3.80	3.78
13	Asp[a]	8.76	4.27	2.74, 2.82	67	Ala	9.76	3.89	1.34
14	Thr[a]	7.82	3.80	4.08	68	Pro			
15	Asp[a]	8.46	4.27	2.27	69	Lys[a]	7.40	3.90	1.49, 1.83
16	Val[a]	7.61	3.79	0.70	70	Tyr[a]	7.35	4.44	3.44
17	Leu[a]	7.02	3.89	1.58	71	Gly[a]	7.55	3.78, 3.85	
18	Lys[a]	7.44	4.40	1.82	72	Ile[a]	7.17	3.79	1.53
19	Ala	6.64	4.23	1.27	73	Arg	8.60	4.44	1.63, 1.67
20	Asp[a]	8.52	4.74	2.62, 2.69	74	Gly[a]	7.71	3.79, 4.25	
21	Gly[a]	8.16	4.02		75	Ile	8.30	4.74	1.92
22	Ala	8.45	5.20	1.41	76	Pro			
23	Ile[a]	9.02	5.16	1.74	77	Thr[a]	7.80	4.93	3.90
24	Leu[a]	9.26	5.24	2.03	78	Leu[a]	9.30	5.77	1.70, 1.77
25	Val[a]	9.80	4.55	2.24	79	Leu[a]	9.12	5.14	1.87
26	Asp[a]	8.91	5.24	2.87	80	Leu[a]	8.89	5.32	1.96
27	Phe[a]	9.00	5.33	2.79, 3.33	81	Phe[a]	10.00	5.20	2.87, 2.93
28	Trp	8.66	5.17	3.16, 3.08	82	Lys[a]	8.82	4.62	1.62, 1.77
29	Ala	6.59	3.33	0.35	83	Asn[a]	9.41	4.62	2.86, 3.13
30	Glu	9.36	4.20	2.06, 2.10	84	Gly[a]	9.62	3.79, 4.40	
31	Trp	6.69	4.61	3.22	85	Glu	7.88	4.95	1.96, 2.11
32	Cys	6.88	4.74	2.78, 2.91	86	Val[a]	8.84	3.16	1.87
33	Gly[a]	9.61	4.07, 4.34		87	Ala	9.68	4.53	1.33
34	Pro				88	Ala	7.72	4.64	1.34
35	Cys	8.25	4.43	3.40	89	Thr[a]	8.55	5.16	3.84
36	Lys[a]	7.97	4.09	2.00, 2.12	90	Lys[a]	9.05	4.52	1.22
37	Met	7.66	4.29	2.22, 2.24	91	Val[a]	8.66	4.60	1.95
38	Ile[a]	7.23	4.41	1.96	92	Gly[a]	8.19	3.65, 4.40	
39	Ala	7.23	3.85	1.34	93	Ala	8.22	3.99	1.37
40	Pro				94	Lys	6.78	4.71	1.56, 1.61
41	Ile[a]	6.71	3.85	2.03	95	Ser[a]	8.42	4.64	3.96
42	Leu[a]	7.94	3.85	1.90	96	Lys[a]	9.34	3.83	1.88, 1.96
43	Asp[a]	7.33	4.31	2.69, 2.80	97	Gly[a]	8.83	3.80, 3.96	
44	Glu	7.29	4.27	2.07, 2.24	98	Gln	7.76	4.16	2.30, 2.47
45	Ile[a]	8.56	3.99	2.16	99	Leu[a]	8.60	4.12	1.48, 1.91
46	Ala	8.60	3.97	1.44	100	Lys[a]	8.76	3.97	1.83, 2.02
47	Asp[a]	7.17	4.57	2.85	101	Glu	7.86	4.08	2.11, 2.16
48	Glu	8.58	4.06	2.02, 2.16	102	Phe[a]	7.70	4.44	3.28, 3.33
49	Tyr[a]	8.76	4.40	2.82	103	Leu[a]	8.61	3.53	1.85
50	Gln	7.12	4.37	2.27, 2.36	104	Asp[a]	9.23	4.38	2.68, 2.74
51	Gly[a]	9.27	3.76, 4.43		105	Ala	7.57	4.24	1.44
52	Lys[a]	8.22	4.65	1.90	106	Asn[a]	7.29	4.73	2.63
53	Leu[a]	7.89	4.82	1.79	107	Leu[a]	7.88	4.42	1.75, 1.88
54	Thr[a]	8.17	4.66	3.97	108	Ala	7.54	4.10	1.35

[a] Residues assigned by type via selective deuteriation. Only one β chemical shift given does not imply degeneracy.

cross-peaks. In cases where the αβ COSY cross-peaks are too weak to be observed in our data, αβ NOESY data were used to verify assignments since the intensity of intraresidue αβ NOESY cross-peaks varies oppositely to that of the corresponding αβ COSY cross-peaks as a function of the χ_1 dihedral angle.

Since the confidence levels for the amide and α assignments are higher, only connectivities between these protons were used in the early stage of assignments. Here the superior quality of the amide–amide NOESY data proved valuable. Seventy-five percent of the anticipated N_i–N_{i+1} cross-peaks were observed, and nearly half of the remaining cross-peaks were obscured by the large diagonal peaks and might prove observable with the less-sensitive diagonal suppression sequences (Bodenhausen & Ernst, 1982; Harbison et al., 1985). Particularly noteworthy is that half of the N_i–N_{i+1} connectivities in β strands are observed, which is considerably more than what is generally obtained in natural abundance samples. With use of only amide and α connectivities in the path analysis program mentioned above, roughly 80% of the sequential assignment was rapidly obtained. In Figure 6 is given the NOE connectivity pattern for the assigned residues. With the exception of the α_2 helix all the secondary structural regions were rapidly assigned. For the second helix the anticipated pattern of strong amide–amide NOESY cross-peaks was interrupted by an internal proline and three cases of sequentially adjacent amides which, being nearly degenerate, were lost in the diagonal. In two of these cases the adjacent positions were degenerate at pH 5.7 for 30 and 50 °C as well as at pH 2.8 and 30 °C. Hence, interresidue αN and βN connectivities

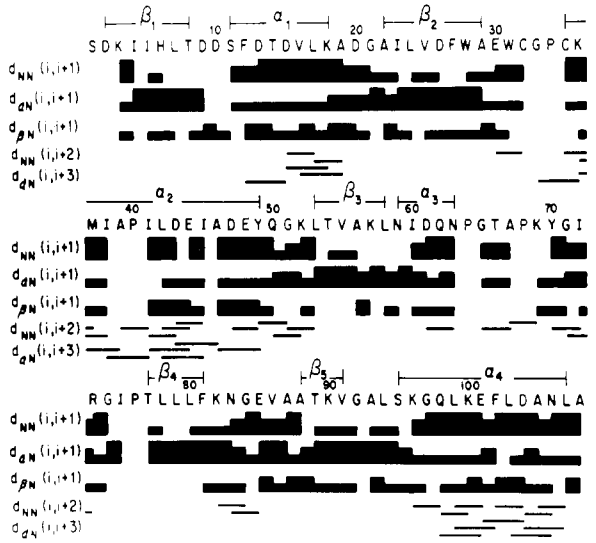

FIGURE 6: Summary of the sequential NOE connectivities in *E. coli* thioredoxin. For the NN, αN, and βN nearest-neighbor connectivities the weak, medium, and strong NOE's are denoted by the height of the darkened bar. The observed (i, i+2) and (i, i+3) connectivities are indicated only as bars. The secondary structural elements based on the X-ray analysis (Holmgren et al., 1975) are indicated above the sequence. Assignment of proline-containing peptides will be considered elsewhere.

could not be distinguished from the intraresidue NOE cross-peaks. The assignments were verified with the α_i–N_{i+3} connectivities characteristic of α helices. In addition, numerous N_i–N_{i+2} connectivities were observed here and in other helical and turn regions of the protein. Furthermore alanines-38 and -46 were placed by virtue of the previous assignment of the other 10 alanine residues.

Another assignment problem arose around the Gly-33-Pro-34 dipeptide. At pH 5.7 the Gly-33 amide proton is degenerate with the Gly-84 amide, and furthermore, the two Nα COSY cross-peaks for Gly-33 were quite weak as expected for a turn region. At pH 2.8 the two glycine amides were resolved and by virtue of the placement of the other glycines Gly-33 could be assigned. In addition an α_i–N_{i+3} connectivity to Lys-36 was observed. Unfortunately, no sequential connectivity was observed between Cys-32 and Gly-33. This is in part due to the location of the Cys-32 α resonance near the water peak.

The final difficulty in the assignment arose in the reverse turn region of residues 9–11. As can be seen from Table I, all three of these sequential amides have identical chemical shifts at 30 °C. At 50 °C the three amides are resolved but not enough to have the anticipated strong amide–amide NOE cross-peaks resolved from the diagonal. In addition the amide–α cross-peaks occur in a highly congested region so that the interresidue connectivities cannot be confidently assigned. However, the βN connectivities are well resolved and when combined with the selective aspartate and serine labeling experiments yield the appropriate assignment.

In addition to use in resonance assignment, protein NOE measurements have been used to obtain structural information in a number of small proteins [e.g., Braun et al. (1982), Kaptein et al. (1985), and Kline et al. (1986)]. Besides the previously discussed i to i+3 connectivities seen in α helices, the most commonly useful nonsequential connectivities are those connecting β strands. *E. coli* thioredoxin has a five-strand β sheet as schematized in Figure 7. Arrows are drawn to indicate the observed interstrand connectivities among the amide and α protons. The observed NOE connectivities are consistent with the earlier published X-ray structure except

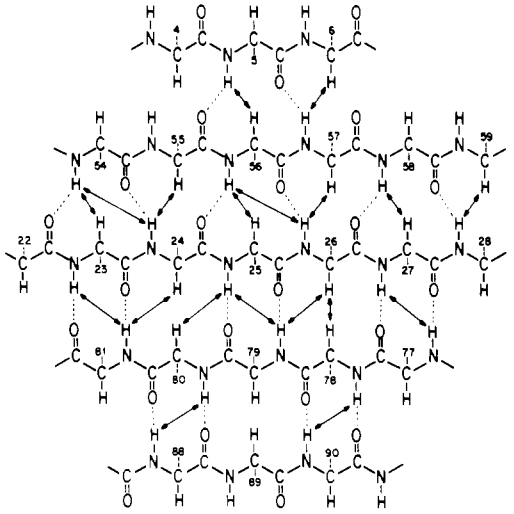

FIGURE 7: NMR analysis of the β sheet of *E. coli* thioredoxin. The observed interstrand NOE connectivities among the amide and α protons of the five β strands are indicated by arrows. Dotted lines mark the hydrogen bonds.

for the first strand, which is shifted one residue as compared to the reported structure. In our determination isoleucine-5 rather than isoleucine-4 is involved in two hydrogen bonds to the β_3 strand. In addition to the amide to α connectivities, parallel β sheets have close approaches between the α proton and the β protons of the opposite strand. As further verification of this assignment we observed NOE connectivities between Ile-4$_\alpha$ and Val-55$_\beta$, Ala-56$_\alpha$ and Ile-5$_\beta$, and His-6$_\alpha$ and Lys-57$_\beta$. This placement of the first β strand is consistent with that found in more recent analysis of the X-ray structure (H, Eklund, personal communication).

Conclusions

The most significant measure of confidence for a sequential assignment is the average number of observed nearest-neighbor NOE connectivities (αN, βN, and NN) per peptide bond. For several of the best determined assignments of small proteins, values of around 2.0 have been obtained although as the molecular weight of the proteins studied have approached 10 kdaltons, this value has generally significantly decreased. For the assigned peptide linkages of *E. coli* thioredoxin we have obtained value of 2.35. This value is slightly inflated due to the absence of proline-containing peptides, which generally show fewer connectivities. In addition several of the best determined small proteins have high β sheet content for which often only the dominant αN connectivity is observed. Nevertheless, the quality of the amide–amide NOE spectrum obtained in the random fractionally deuteriated sample has served to provide much better definition to the β strand regions for which the distance between adjacent amides is greater than 4 Å. Overall, the decreased line widths of the partially deuteriated samples combined with isotopic residue type assignments have proven suitable for obtaining sequential assignment of comparable quality to those of natural abundance samples for proteins of half the size.

Acknowledgments

We thank the Stable Isotopes Resource of Los Alamos (NIH RR02231) for their collaboration on the production of the perdeuteriated amino acids as well as P. Moore for the use of the D_2O used in the bacterial growth medium. The use of the instruments at the Yale Chemical Instrumentation Center is appreciated. We also thank B. Bachmann of the Yale Coli Genetic Stock Center for several of the strains used

in this study. The assistance of M. Bannon and J. Mouning in preparing the manuscript is also appreciated.

References

Bodenhausen, G., & Ernst, R. R. (1982) *Mol. Phys. 47*, 319.
Braun, W., Bösch, C., Brown, L. R., Go, N., & Wüthrich, K. (1981) *Biochim. Biophys. Acta 667*, 377.
Chan, T. M., & Markley, J. L. (1982) *J. Am. Chem. Soc. 104*, 4010.
Crepsi, H. L., Rosenberg, R. M., & Katz, J. J. (1968) *Science (Washington, D.C.) 161*, 795.
Dalvit, C., & Wright, P. E. (1987) *J. Mol. Biol. 194*, 313.
Gelfand, D. H., & Steinberg, R. A. (1977) *J. Bacteriol. 130*, 429.
Griffey, R. H., Redfield, A. G., Loomis, R. E., & Dahlquist, F. W. (1985) *Biochemistry 24*, 817.
Guyer, M., Reed, R. R., Steitz, J. A., & Low, K. B. (1981) *Cold Spring Harbor Symp. Quant. Biol. 45*, 135.
Harbison, G. S., Feigon, J., Ruben, D. J., Herzfeld, J., & Griffin, R. G. (1985) *J. Am. Chem. Soc. 107*, 5567.
Holmgren, A., & Reichard, P. (1967) *Eur. J. Biochem. 2*, 187.
Holmgren, A., & Slaby, I. (1979) *Biochemistry 18*, 5591.
Holmgren, A., Söderberg, B. O., Eklund, H., & Bränden, C. I. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2305.
Jardetzky, O., Conover, W. W., Sullivan, G. R., & Basus, V. J. (1978) *Magnetic Resonance and Related Phenomena, Proceedings of the Congress AMPERE, 20th* (Kundla, E., Ed.) p 19, Springer, West Berlin.
Jordan, P. M., & Akhtar, M. (1970) *Biochem. J. 116*, 277.
Kalbitzer, H. R., Leberman, R., & Wittinghofer, A. (1985) *FEBS Lett. 180*, 40.
Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179.
Kelley, R. F., & Stellwagen, E. (1984) *Biochemistry 23*, 5095.
Kelley, R. F., & Richards, F. M. (1987) *Biochemistry 26*, 6765–6774.
Kline, A. D., Braun, W., & Wüthrich, K. (1986) *J. Mol. Biol. 189*, 377.
LeMaster, D. M. (1986) *J. Virol. 59*, 759.
LeMaster, D. M., & Richards, F. M. (1982a) *Anal. Biochem. 122*, 238.
LeMaster, D. M., & Richards, F. M. (1982b) *J. Labelled Compd. Radiopharm. 19*, 639.
LeMaster, D. M., & Richards, F. M. (1985) *Biochemistry 24*, 7263.
Levitt, M. H., & Ernst, R. R. (1985) *J. Chem. Phys. 83*, 3297.
Lunn, C. A., Kathju, S., Wallace, B. J., Kushner, S. R., & Pigiet, V. (1984) *J. Biol. Chem. 259*, 10469.
Markley, J. L., Putter, I., & Jardetzky, O. (1968) *Science (Washington, D.C.) 161*, 1249.
Matthews, H. R., Matthews, K. S., & Opella, S. J. (1977) *Biochim. Biophys. Acta 497*, 1.
McIntosh, L. P., Griffey, R. H., Muchmore, D. C., Nielson, C. P., Redfield, A. G., & Dahlquist, F. W. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1244.
Miller, J. H. (1972) *Experiments in Molecular Genetics*, p 431, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Senn, H., Otting, G., & Wüthrich, K. (1987) *J. Am. Chem. Soc. 109*, 1090.
Slaby, I., & Holmgren, A. (1979) *Biochemistry 18*, 5584.
Stella, V. J. (1973) *J. Pharm. Sci. 62*, 634.
Tenenbaum, S. W., Witherup, T. H., & Abbott, E. H. (1974) *Biochim. Biophys. Acta 362*, 308.
Upson, D. A., & Hruby, V. J. (1977) *J. Org. Chem. 42*, 2329.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347.

Processive Action of the Two Peptide Binding Sites of Prolyl 4-Hydroxylase in the Hydroxylation of Procollagen

Anthony de Waal and Luitzen de Jong*
Laboratory of Biochemistry, University of Amsterdam, P.O. Box 20151, 1000 HD Amsterdam, The Netherlands
Received May 29, 1987; Revised Manuscript Received September 14, 1987

ABSTRACT: The number of peptide binding sites of prolyl 4-hydroxylase was manipulated with the peptide photoaffinity label *N*-(4-azido-2-nitrophenyl)glycyl-(Pro-Pro-Gly)$_5$, and the effect on hydroxylation of the relatively short peptide substrate (Pro-Pro-Gly)$_5$ and of the long natural substrate procollagen was studied. With (Pro-Pro-Gly)$_5$ as a substrate, a linear relation was found between enzyme activity and the amount of covalently bound photoaffinity label, approximately 50% inactivation being reached at 1 mol of label/mol of enzyme. No difference in K_m value for (Pro-Pro-Gly)$_5$ was detected between unlabeled and partially labeled enzyme preparations. These results indicate that enzyme molecules with only one free active site hydroxylated the synthetic substrate (Pro-Pro-Gly)$_5$ with the same K_m and at half the rate of native enzyme. In contrast, with procollagen as a substrate a 5–10-fold increase in K_m was found with the fraction of enzyme containing only one free active site, as compared to the K_m for procollagen with nonlabeled enzyme. This finding is explained by an enzyme–kinetic model based on a processive action of the two peptide substrate binding sites of prolyl 4-hydroxylase, preventing dissociation of the enzyme–substrate complex between successive hydroxylations of a long peptide with multiple substrate sites. Such a mechanism leads to a low K_m for a long peptide by overcoming the diffusional constraints on the rate of association between the enzyme and the individual substrate sites.

Prolyl 4-hydroxylase (prolyl-glycyl-peptide, 2-oxoglutarate:oxygen oxidoreductase, 4-hydroxylating, EC 1.14.11.2) catalyzes the 4-hydroxylation of proline residues located N-terminally of glycine in nascent pro-α chains of procollagen. This posttranslational modification is pivotal for the formation of the triple helix structure in the various col-

0006-2960/88/0427-0150$01.50/0