Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data[†]

John L. Battiste[§] and Gerhard Wagner*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue,
Boston, Massachusetts 02115

ABSTRACT: To test whether distances derived from paramagnetic broadening of ^{15}N heteronuclear single quantum coherence (HSQC) resonances could be used to determine the global fold of a large, perdeuterated protein, we used site-directed spin-labeling of 5 amino acids on the surface of ^{15}N-labeled eukaryotic translation initiation factor 4E (eIF4E). eIF4E is a 25 kDa translation initiation protein, whose solution structure was previously solved in a 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS) micelle of total molecular mass ~45–50 kDa. Distance-dependent line broadening consistent with the three-dimensional structure of eIF4E was observed for all spin-label substitutions. The paramagnetic broadening effects (PBEs) were converted into distances for modeling by a simple method comparing peak heights in ^{15}N-HSQC spectra before and after reduction of the nitroxide spin label with ascorbic acid. The PBEs, in combination with HN–HN nuclear Overhauser effects (NOEs) and chemical shift index (CSI) angle restraints, correctly determined the global fold of eIF4E with a backbone precision of 2.3 Å (1.7 Å for secondary structure elements). The global fold was not correctly determined with the HN–HN NOEs and CSI angles alone. The combination of PBEs with simulated restraints from another nuclear magnetic resonance (NMR) method for global fold determination of large proteins (methyl-protonated, highly deuterated samples) improved the quality of calculated structures. In addition, the combination of the two methods simulated from a crystal structure of an all α-helical protein (40 kDa farnesyl diphoshphate synthase) correctly determined the global fold where neither method individually was successful. These results show the potential feasibility of obtaining medium-resolution structures for proteins in the 40–100 kDa range via NMR.

There has been an increase in research aimed at calculating global folds of proteins with the limited NOE[1] data obtained from deuterated samples (*1*). The motivation is the molecular weight size gap between what can be sequentially assigned and for what a sufficient number of NOEs can be determined for structure calculations. Backbone resonances in macromolecular systems as large as 65 kDa have been assigned with deuterated samples (*2*) and recent advances with transverse relaxation-optimized spectroscopy (TROSY) experiments promise to extend the theoretical limit for backbone assignment to beyond 100 kDa (*3, 4*). However, the high levels of deuteration required to increase the relaxation times of the aliphatic ^{13}C-carbon resonances remove most of the side-chain protons for NOE analysis. The most simple methodology for global fold determination is to use the remaining NOEs between nitrogen-attached protons (HN) that can be obtained from 3D or 4D ^{15}N-NOESY–HSQC experiments of a deuterated protein (*5*). This limited subset of NOEs results in very poor quality structures that in many cases cannot correctly determine the global fold of the protein. Use of HN NOEs is particularly problematic for proteins with high α-helical content, since there are few long-range NOEs to amide protons in helices. Recently, methods for selectively protonating methyl or aromatic hydrogens in otherwise highly deuterated proteins have been used to obtain NOEs between hydrophobic side chains that are often in the core of proteins (*6–9*). With these augmented NOE data sets the global folds of small model proteins were correctly determined with varying success. Measuring residual dipolar couplings in partially oriented systems (*10, 11*) is also a promising technique for studying large proteins. Most applications of residual dipolar couplings to date, however, have been toward refinement of protein structure in combination with NOE data, and it has not been demonstrated how

* To whom correspondence should be addressed: Phone (617) 432-3213; Fax (617) 432-4383; E-mail wagner@hms.harvard.edu.
[§] Present address: Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, 244 Gortner Labs, St. Paul, MN 55108.

[1] Abbreviations: CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate; eIF4E, eukaryotic translation initiation factor 4E; EPR, electron paramagnetic resonance; FPS, farnesyl diphosphate synthase; HN, hydrogen bonded to nitrogen; HSQC, heteronuclear single quantum coherence; INEPT, insensitive nucleic enhanced by polarization transfer; TROSY, transverse relaxation-optimized spectroscopy; MTSL, (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiosulfonate; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PBE, paramagnetic broadening effect; PCR, polymerase chain reaction; SDSL, site-directed spin-labeling.

well the orientation information alone would determine protein structures. Nevertheless, new or complementary methods are needed for obtaining restraints for modeling calculations in the absence of extensive NOE data.

It has long been recognized that distance-dependent line broadening of nuclear magnetic resonances can be observed in protein samples containing paramagnetic electrons (either from metals or nitroxide spin labels; reviewed in refs *12* and *13*). However, few attempts to incorporate the broadening effect into distances for modeling calculations have been reported (*14, 15*). In most cases, such distances would be unnecessary or redundant given the quality of protein structures obtained from NOE-derived distances. However, they may be useful in situations where structural information from NOEs is limited, such as for large proteins. The usefulness of nitroxide spin labels has already been demonstrated in another situation where NOEs are limited, which is the unfolded state of staphylococcal nuclease (*14, 16*). Hydrogens attached to ^{15}N remain protonated in deuterated proteins after exchange with H_2O and can be used as probes for distance-dependent broadening via site-directed spin labeling (SDSL) (*17*). We wished to determine whether distances derived from paramagnetic broadening of ^{15}N−HSQC resonances via SDSL could be used to calculate the global fold of a protein with limited NOE data.

The 25 kDa yeast translation initiation protein eIF4E was used as a model system for several reasons. A high-resolution structure of eIF4E has been determined (*18*) and can be used as a reference for all spin-label calculations (Figure 1C). In addition, to prevent aggregation yeast eIF4E requires a CHAPS micelle of a total molecular weight of approximately 45 kDa (*18*). Therefore, the resonances of eIF4E have relaxation properties similar to the size protein to which the technique will ultimately be applied. Last, wild-type eIF4E contains no cysteines, so introduction of single-cysteine mutations for attachment of the spin label was simplified. Five samples containing spin labels attached at different residues were used to obtain ∼500 semiquantitative restraints for modeling calculations. These spin-label restraints, along with HN−HN NOEs obtained from ^{15}N-edited NOESY spectra and loose backbone angle restraints for secondary structure elements, correctly determined the global fold of eIF4E. Comparisons are also made to the methyl-protonation method using simulated restraints for eIF4E and the protein farnesyl diphosphate synthase.

MATERIALS AND METHODS

Sample Preparation. Single-cysteine mutations of yeast eIF4E (which contains no wild-type cysteines) were made by the Quick Change mutagenesis protocol utilizing whole-plasmid PCR (Stratagene). The five mutants were arginine 120 (R120C), arginine 132 (R132C), serine 169 (S169C), aspartate 190 (D190C), and serine 200 (S200C). All mutations were confirmed by DNA sequencing of the entire open reading frame. Protein mutants were overexpressed from a pGEM.2 vector on minimal medium enriched with ^{15}N-ammonium chloride and purified as described (*18*). After elution of the protein from an m^{7}GDP affinity column (with excess m^{7}GDP), the protein was modified with the spin-label reagent MTSL that attaches the nitroxide spin label via a disulfide bond to the single cysteine (Figure 1B) (*19*). MTSL
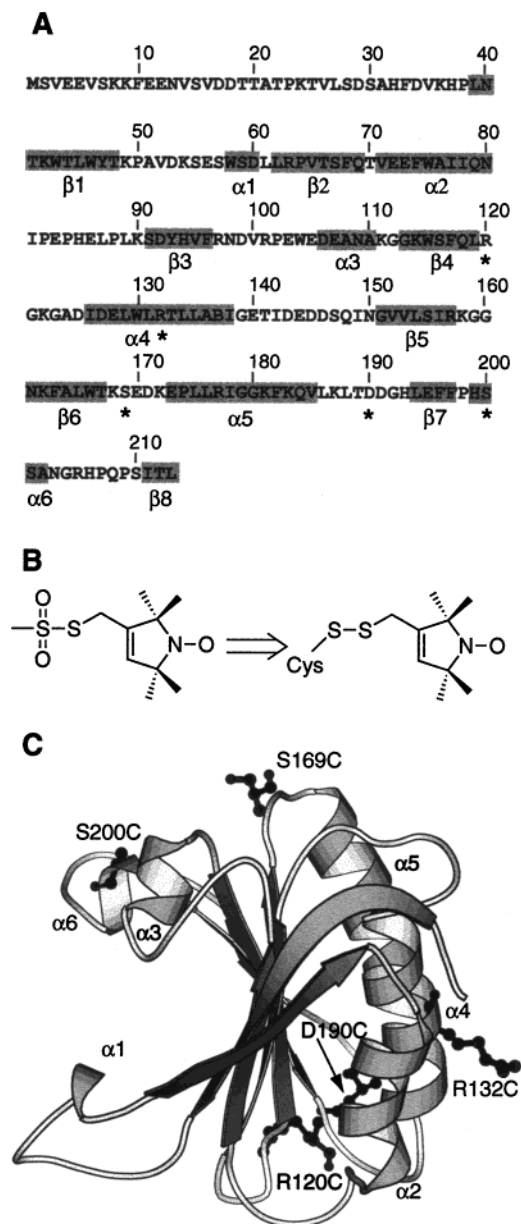

FIGURE 1: Sequence and location of spin label modifications in eIF4E. (A) Protein sequence of yeast eIF4E (single-letter code). Secondary structure elements are shaded and labeled. Amino acids where spin labels are attached are marked with an asterisk. (B) Chemical structure of the spin label reagent MTSL and the disulfide bond adduct formed with a single cysteine. (C) Ribbon diagram of the structure of eIF4E bound to the cap analogue m^{7}GDP (determined by NMR). The positions of the side chains used for attachment of the spin label (after mutation to cysteine) are shown in black.

was added from a concentrated stock in acetonitrile at a molar ratio of 3−5:1 MTSL:protein and incubated at room temperature for 12−16 h. Unreacted protein (with free sulfhydryls) was removed by passage over an organomercurial SH-affinity column (Bio-Rad). Excess MTSL reagent was removed by dialysis at 4 °C for 24 h into NMR buffer (50 mM NaPO$_4$, pH 6.5, and 50 mM KCl). The modified protein was concentrated to 250−500 μL for NMR. The nondenaturing detergent 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate (CHAPS) was added to a final concentration of 25 mM to stabilize the protein. Final protein concentrations ranged from 0.2 to 1.0 mM.

NMR Spectroscopy. ^{15}N-HSQC experiments were acquired at 25 °C at either 500 or 750 MHz. In many cases, ^{15}N-TROSY experiments (*3*) were also acquired for comparison. Total experiment times were typically 12−18 h. All experiments were repeated after the spin label was reduced with 2−3-fold excess ascorbic acid, adding 2−5 μL from a concentrated stock (dilution <1%). Samples were placed in the magnet at 25 °C for at least 1 h to ensure complete reduction of the spin label.

NMR Analysis. Two-dimensional spectra were processed and analyzed with the programs nmrPipe and nmrDraw (*20*). For peak fitting analysis (see below), spectra were processed with only 5−10 Hz exponential broadening in each dimension, linear predicted 2× points in the ^{15}N dimension, and zero-filled to 4096 × 1024 data points in t_2 and t_1, respectively. Spectra were fit with the program nlinLS, provided as a part of nmrDraw. Lorentzian models were used for the fitting in each dimension, starting from values obtained from the peak-peaking routine in nmrDraw. Intensity (peak height) ratios of cross-peaks in oxidized (paramagnetic) versus reduced (diamagnetic) spectra were extracted from the peak fitting. Intensity ratios were normalized by averaging the values for amides >30 Å away from the γ-carbon at the site of the spin label, as determined from the known three-dimensional structure. For amides that were broadened beyond detection in the oxidized spectra, an upper limit for the intensity ratio was estimated as the noise in the oxidized spectrum divided by the intensity in the reduced spectra. Intensity ratios were converted into a paramagnetic relaxation rate enhancements (R2^{sp}) by estimating the additional relaxation needed to reduce the intensity of the diamagnetic sample by the calculated ratio (equations below). Relaxation during all periods of the HSQC where ^{1}H is in the transverse plane was considered, including the INEPT and acquisition periods. Potential relaxation differences between in-phase and anti-phase ^{1}H magnetization (I_y and I_xN_z) were ignored. Peak intensity reduction can also occur through R1 relaxation of ^{1}H during the t_1 evolution time (I_zN_y); however, paramagnetic-induced R1 relaxation is typically insignificant compared to the R2 effects (*16*). Relaxation effects (R2 or R1) on ^{15}N nuclei from the spin label were considered to be negligible compared to ^{1}H due to the much lower gyromagnetic ratio (eq 6; see below).

During the INEPT delays the intensity of the magnetization is proportional to (ignoring *J*-evolution terms, which are constant and cancel)

$$I_{ox} \approx \exp(-R2^*t) \quad \text{and} \quad I_{red} \approx \exp(-R2t) \quad (1)$$

where I_{ox} and I_{red} are the peak intensities (heights) of oxidized and reduced resonances, respectively, *t* is the total INEPT evolution time of the HSQC (∼9 ms), and R2* and R2 are the transverse relaxation rates for oxidized and reduced amide spins, respectively. Similarly, the intensity (peak height for Lorentzian line at $\omega = \omega_0$) in the directly detected proton dimension of the HSQC is proportional to

$$I_{ox} \approx \frac{1}{R2^*} \quad \text{and} \quad I_{red} \approx \frac{1}{R2} \quad (2)$$

The total electron spin-enhanced relaxation rate R2* is a sum

of the intrinsic (R2) and spin contribution (R2^{sp})

$$R2^* = R2 + R2^{sp} \quad (3)$$

Substituting eq 3 into eqs 1 and 2 and then combining, the intensity ratio for a particular amide proton is equal to

$$\frac{I_{ox}}{I_{red}} = \frac{R2 \exp(-R2^{sp}t)}{R2 + R2^{sp}} \quad (4)$$

The intrinsic R2 relaxation rate for each amide was estimated from the reduced spectra and along with the measured intensity ratios was used to linearly fit for R2^{sp}.

Paramagnetic rate enhancements (R2^{sp}) were converted into distances by use of the following equation for the effect of paramagnetic spins on nuclear magnetic relaxation (*13, 21*):

$$r = \left[\frac{K}{R2^{sp}} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_h^2 \tau_c^2} \right) \right]^{1/6} \quad (5)$$

where *r* is the distance between the electron and nuclear spins, τ_c is the correlation time for this electron−nuclear interaction, ω_h is the Larmor frequency of the nuclear spin (proton), and *K* is 1.23×10^{-32} cm^6 s^{-2} (*12*) composed of the following physical constants:

$$K = \frac{1}{15} S(S+1)\gamma^2 g^2 \beta^2 \quad (6)$$

where γ is the nuclear gyromagnetic ratio, *g* is the electronic *g* factor, and β is the Bohr magneton. For calculating distances, the approximation was made that τ_c was equal to the global correlation time of the protein.

Three classes of distance restraints were utilized for structure calculations. Peaks with an intensity ratio <0.85 and detectable in the oxidized spectra were restrained as the calculated distance ± 4 Å bounds (± 3 Å bounds were also tested). Severely broadened peaks not detectable in the oxidized spectra were restrained with no lower bound and a target distance estimated from the noise in the spectrum plus an upper bound of 4 Å. Peaks with an intensity ratio >0.90 were restrained with no upper bound, a target distance calculated from an intensity ratio of 0.90, and a lower bound of 4 Å. The distances in modeling calculations were restrained from the nitrogen atom of the MTSL ring to the hydrogen atom detected in the HSQC. For side-chain NH$_2$ groups of asparagine and glutamine, the two distances were averaged, a correction of 1 Å was added, and the distance was restrained to a psuedoatom positioned between the two amide protons.

Simulated Restraints. One NMR method for global fold determination of large proteins is to selectively protonate methyl groups in an otherwise highly deuterated protein (*9*). We wished to compare this method to the spin label method on the same protein. Therefore, methyl−methyl and HN−methyl NOE distance restraints were calculated for eIF4E as described (*9*). These simulated NOE data sets are intended to approximate NOEs that would be observed for a highly deuterated, methyl-protonated protein sample (*9*). Actual methyl NOEs determined from previous NMR analysis were not used for these data sets.

To hypothetically test the spin-label method on another protein, simulated distance restraints were generated for farnesyl diphosphate synthase (FPS; PDB accession code 1fps) with the eIF4E results as a guide for the number and type of restraints that would be expected. Spin labels were introduced into the crystal structure coordinates approximately every 30 residues at positions on the surface of loops or helices (29, 64, 92, 122, 151, 176, 203, 240, 269, 298, 334, 360). The spin-label side chain was placed in an extended conformation (dihedrals $\chi^1-\chi^5$ all $\sim$180°). The only exception was residue 269, which required a χ^1 angle of $-60°$ to avoid van der Waals overlap with adjacent side chains. Distances were then measured from the nitrogen atom of each nitroxide to the hydrogen of all amides (including side chains). Distances were also measured to the carbon atoms of methyl groups of Val, Leu, Ala, and Ile ($\gamma2$ only). Fifty percent of all distances were randomly removed to approximate spectral overlap. Distances were separated into three classes, as described above, with <15 Å upper bound only, 15–23 Å upper and lower bound, and >23 Å lower bound only. Methyl–methyl, HN–methyl, and HN–HN NOE restraints were also generated as previously described (9).

Structure Calculations. Structures of eIF4E were calculated in two stages by the program XPLOR 3.851 on R10000 IndigoII Silicon Graphics workstations. First the "apo" structure of the protein was calculated, followed by introduction of the m7GDP cap in a random orientation and reannealing (18). Parameters for the apo calculation were 48 000 steps of 5 fs at 3000 K, followed by cooling to 300 K in 58 000 steps of 5 fs. For the calculations including the m7GDP "cap", annealing was 50 000 steps of 3 fs followed by cooling to 300 K in 100 000 steps of 2 fs. Force constants in apo calculations were 25 and 10 kcal/(mol Å^2) for NOE and PBE distances, respectively. Force constants in cap calculations were 45, 35, and 15 kcal/(mol Å^2) for intermolecular NOE, intramolecular NOE, and PBE distances, respectively.

Parameters for simulated annealing of FPS were 50 000 steps of 5 fs at 3000 K followed by cooling to 300 K in 100 000 steps of 5 fs. Force constants were 25 and 10 kcal/(mol Å^2) for NOE and PBE distances, respectively.

Converged structures were selected on the basis of lowest total and restraint violation energies, as well as no restraint violations greater than 0.5 Å for NOEs and 0.8 Å for PBEs.

RESULTS AND DISCUSSION

Methodology of Spin-Label Substitutions. Five single-cysteine mutants of eIF4E were constructed for introduction of the nitroxide spin label MTSL at defined positions in the protein. These mutations are in loops or on the solvent-exposed surface of helices (Figure 1). The positions were designed to be in regions that would not disrupt the three-dimensional structure of eIF4E and cover a sufficient portion of the protein surface for modeling calculations (range of each spin label $\sim$20–25 Å). While the known three-dimensional structure was used to guide the spin-label positions, it should be possible to successfully introduce spin labels in a de novo system. The process of backbone assignment of a double-labeled (^{13}C/^{15}N) and perdeuterated protein would determine the secondary structure. The

secondary structure and hydrophilicity patterns should be sufficient for introduction of spin labels with a high level of success. EPR studies have shown that the nitroxide spin label MTSL is readily introduced into all secondary structure elements (particularly helices) and in general does not significantly perturb protein structure (22).

Nevertheless, a concern for high-resolution work is whether the spin-label modification significantly disrupts the protein structure. Two criteria were applied to ascertain whether the three-dimensional fold of eIF4E was perturbed by mutation/modification. The purification procedure contains an affinity chromatography step with an m7GDP column. Binding of m7GDP is a major function of eIF4E in vivo, and successful purification of the single cysteine variants indicates that the global fold is not disrupted by mutation (affinity column is before spin-label modification). The effect of the spin-label modification on the protein structure was ascertained from chemical shift analysis of ^{15}N–HSQC spectra. Amide proton and nitrogen chemical shifts are very sensitive to three-dimensional structure and should not be affected if the structure of the protein is unaltered by spin-label modification. Of course, introduction of the spin label will have some effects on the chemical shift of amides in the proximity of the spin-label side chain, but these effects should be minimal and localized to a very small region around the mutation site. Generally, some large chemical shift changes were observed $\pm$2–3 residues (in sequence) around the spin label site, as would be expected from primary and secondary effects of side-chain mutation, while the remainder of the protein had small chemical shift changes <0.1/0.5 ppm in ^{1}H/^{15}N (data not shown). Shifts were small enough that the spin-labeled eIF4E resonances were assigned by comparison to wild-type spectra. At regions remote from the spin label in three-dimensional space, chemical shifts were virtually identical in all modified proteins, while occasional moderate changes were observed in regions that would be in close proximity to the spin-label side chains. For all spin-label modifications, we believe that the protein structure is not significantly altered and therefore will not adversely affect the distances calculated. Almost definitely, the protein structure is not perturbed at the level of the backbone or global fold. Furthermore, the lack of chemical shift changes implies that the nucleotide binding function of eIF4E remains intact after modification.

Distance-Dependent Line Broadening of ^{15}N-HSQC Spectra. ^{15}N-HSQC was chosen to analyze paramagnetic broadening, since ^{15}N-attached proton resonances are the only ones observable in a perdeuterated protein. In addition, the two-dimensional experiment affords sufficient resolution for analysis of up to several hundred amides per spin-label sample. Figure 2 shows sections of HSQC spectra of ^{15}N-labeled eIF4E with spin labels at two separate positions. Different patterns of specific broadening, resulting in lower intensity cross-peaks, are observed for each spin-label position, indicative of the three-dimensional structural information that can be extracted from the spectra. For the 200 non-proline amino acids in eIF4E, about 120 have sufficiently resolved cross-peaks to be used as quantitative probes of paramagnetic broadening by the spin label. After removal of the N-terminal 37 residues that are unstructured and severely overlapped, about 120 out of 165 residues are
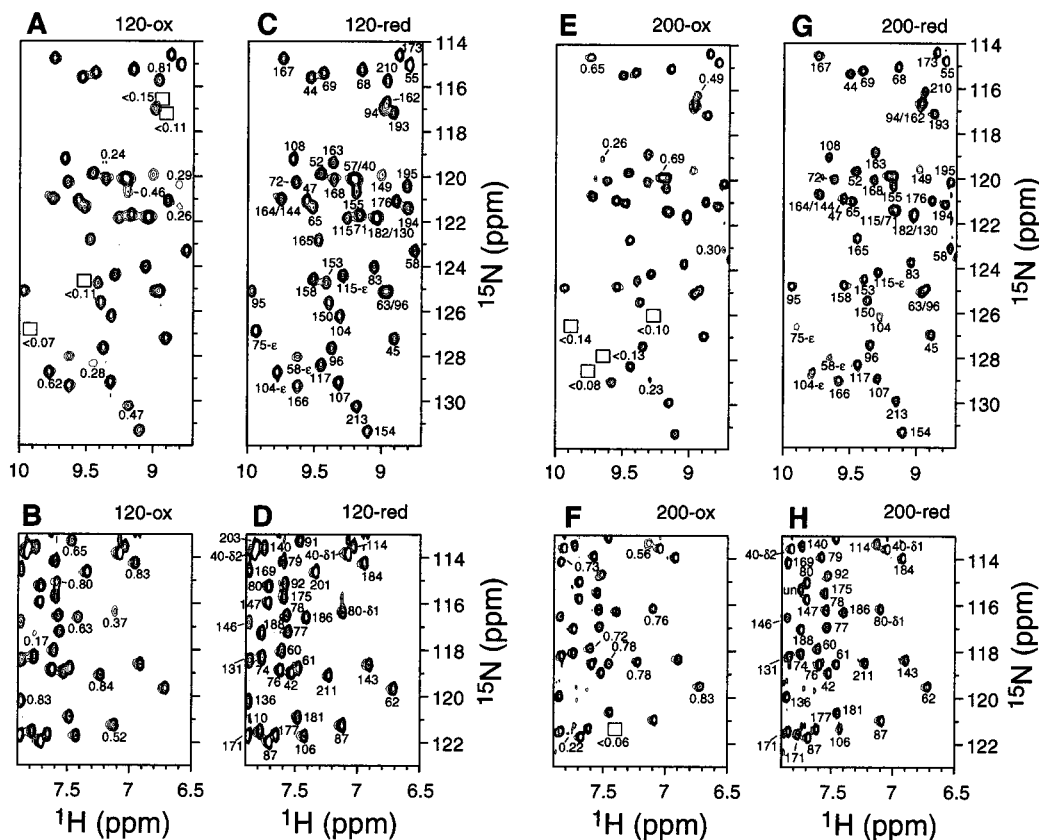
FIGURE 2: Regions of ^{15}N-HSQC spectra of spin-labeled eIF4E at 750 MHz. (A−D) Spin label at residue 120; (E−H) spin label at residue 200. Panels A, B, E, and F are oxidized or paramagnetic samples with a free radical on the nitroxide group. Panels C, D, G, and H are reduced or diamagnetic samples taken after reduction of the spin label with ascorbic acid. Resonance assignments are given in the reduced spectra. The suffixes ϵ and $\delta1/\delta2$ indicate side-chain HN groups of tryptophan and asparagine residues, respectively. Numbers in oxidized spectra are values for intensity ratios (I_{ox}/I_{red}) that are less than 0.85. For cross-peaks not detectable in the oxidized spectra, the values indicate an upper limit for the ratio calculated by using the noise in the oxidized spectrum as the numerator.

available as probes. Of the 120 resolved resonances, on average ∼30 were close enough to the spin-label probe to produce a detectable intensity reduction (I_{ox}/I_{red} < 0.85) in HSQC spectra. Similar results were obtained for ∼20 resolved side-chain HN resonances from tryptophan, asparagine, glutamine, and arginine.

A simple method for quantitating distances from intensities of cross-peaks in ^{15}N−HSQC spectra was used. This methodology was adapted from previous papers analyzing two-dimensional homonuclear (15, 23) or heteronuclear (16) spectra of spin-labeled proteins. The general procedure is to acquire spectra of the paramagnetic protein ("oxidized"), reduce the nitroxide to a diamagnetic species with ascorbic acid ("reduced"), and repeat the experiment under identical conditions. The difference in relaxation properties of the HN spins between oxidized and reduced spectra should be directly attributable to effects of the electron spin. Rather than directly measuring relaxation rates from a series of two-dimensional experiments, which is tedious and time-consuming, relaxation rates were indirectly extracted from ratios of cross-peak heights between oxidized and reduced spectra (I_{ox}/I_{red}) according to equations given under Materials and Methods. This procedure is nearly identical to the heteronuclear work done on unfolded and ^{15}N-labeled staphyloccocal nuclease, which showed that the intensity ratio procedure gave similar results to directly measured relaxation rates (16). For this work, two-dimensional peak fitting with the program nlimNS was used to expedite analysis, rather

than fitting of one-dimensional slices. The two-dimensional peak fitting was also important to obtain as many distance restraints as possible for modeling calculations by deconvoluting intensity changes for partially overlapping peaks with resolvable peak centers. For instance, in the absence of deconvolution, the tail of a cross-peak completely broadened may reduce the height of a neighboring peak 10−15%, giving a false positive for broadening effects due to the spin label. Deconvolution allowed an approximately 10% increase in the number of resonances that could be used for distance restraints and will be particularly important for larger systems that will have increased spectral overlap. In the future experiments, it should also be possible to extend the analysis to three-dimensional experiments, such as HNCO, which would increase the resolution.

Incorporation of PBE Distances into Restraints for Molecular Modeling. PBE distances were incorporated as semiquantitative restraints with large bounds for modeling calculations, similar to what is typically done for NOE analysis. There are enough sources of error or assumptions that may not be universally valid to prevent attempting more quantitative use of the distances (i.e., tighter bounds). Possible sources of systematic error include estimation of intrinsic $R2$ relaxation rates from reduced spectra, use of a global correlation time for all nitroxide−amide dipolar coupling interactions, and potential r^{-6} averaging due to highly flexible motions of the nitroxide side chain. Due to the r^{-6} dependence (eq 5), the calculated distance is relatively
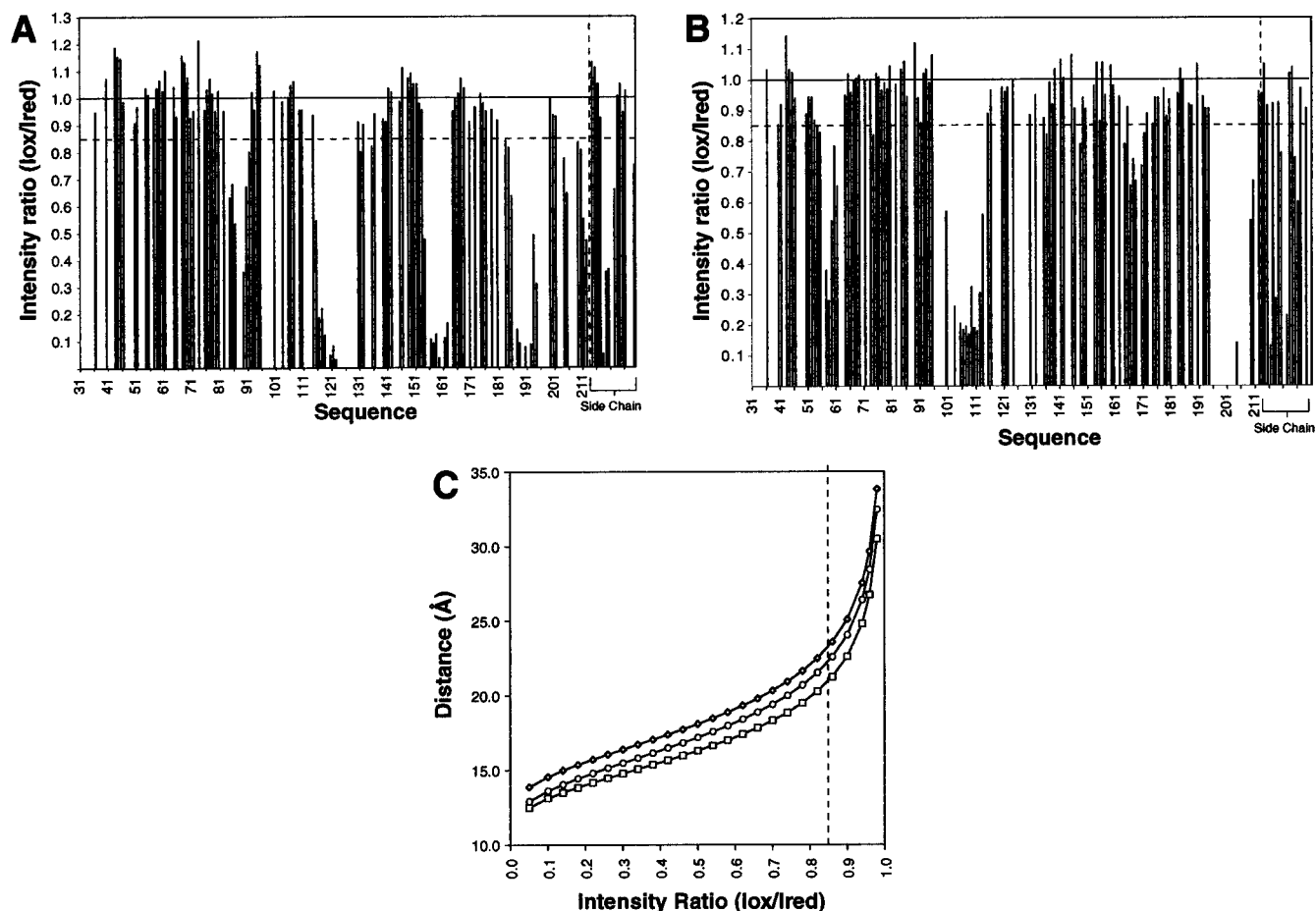
FIGURE 3: Intensity ratios and conversion into distances. Shown are bar graphs of intensity ratios of ^{15}N$-$HSQC cross-peaks vs primary sequence for two spin label positions, (A) R120C and (B) S200C. Ratios are determined as described under Materials and Methods. The first 30 residues, which are disordered in eIF4E, are not shown. The *x*-axis is the backbone amides from 31 to 213, followed on the right by the values for the side chain HNs of N40, W43, W46, W58, Q69, W75, Q79, N80, N98, W104, W115, W130, W166, N150, R157, N161, and Q184, sequentially (separated from backbone amides by vertical dashed line). An intensity ratio of 0.85 is shown by the horizontal dashed line, which is the cutoff for use of an intensity reduction due to spin-label broadening as a modeling restraint. The absence of a bar at any position indicates overlap or lack of assignment that prevents quantitation of an intensity ratio. (C) Simulated curves of theoretical equations used for conversion of intensity ratios into distances for modeling calculations. Curves are shown for different intrinsic half-height line widths [Δv (hertz) $= \pi/R2$] and correlation times (τ_c) to illustrate the dependence of the calculated distance on these input variables. The three curves are for line widths and correlation times of 40 Hz, 16 ns ($\square$); 20 Hz, 16 ns ($\bigcirc$), and 40 Hz, 30 ns ($\lozenge$). A global correlation time of 16 ns for eIF4E in 25 mM CHAPS is used for this study. The intensity ratio cutoff of 0.85 is indicated by a dashed line.

insensitive to large errors in relaxation rates or τ_c (Figure 3C). Nevertheless, the qualitative use of the extracted distances is the most conservative and prudent approach.

Three classes of distance restraints were used for molecular modeling calculations. For cross-peaks undetectable in the oxidized HSQC spectrum, an upper limit for the target distance was estimated from the noise. The target distance was restrained with no lower bound and a fixed upper bound of a few angstroms. For detectable resonances with I_{ox}/I_{red} less than 0.85, a target distance was calculated and restrained with upper and lower bounds of a few angstroms. If I_{ox}/I_{red} was >0.90, it was assumed that the spin label was sufficiently far away to not affect the amide relaxation and could be used as a "negative" restraint. The intensity ratio cutoffs (e.g., 0.85) were empirically estimated from plots of intensities vs amino acid sequence (Figure 3) and the known three-dimensional structure of eIF4E. The apparent noise or variation of intensity ratios for cross-peaks that should be far away from the spin-label nitroxide (no broadening effect) is approximately 10$-$15%. In addition, an intensity ratio cutoff of 0.85 produced distance restraints that were all

qualitatively consistent with the three-dimensional structure of eIF4E. The only exception was the ϵ-HN of the W75 side chain, which was substantially broadened by all spin-label substitutions, not in agreement with the three-dimensional structure. W75 is solvent-exposed on a hydrophobic patch that is the binding surface for the proteins 4E-BP and eIF4G and also appears to be the binding surface for the CHAPS micelle (*18*). Thus, it appears likely that this broadening is from intermolecular association of the hydrophobic nitroxide and W75 ring. The broadening effect appears to be correlated to the protein concentration, consistent with an intermolecular association (data not shown). Also, the backbone amide of W75 only 5$-$7 Å away was not affected, indicating r^{-6} averaging effects that would be expected for a weak nonspecific association. It is not clear how common non-specific binding of the nitroxide to hydrophobic patches on the protein surface will be. However, the problem should be localized to side-chain HN groups and fairly easy to identify. In addition, it probably warrants using as low a protein concentration as possible to minimize intermolecular effects.
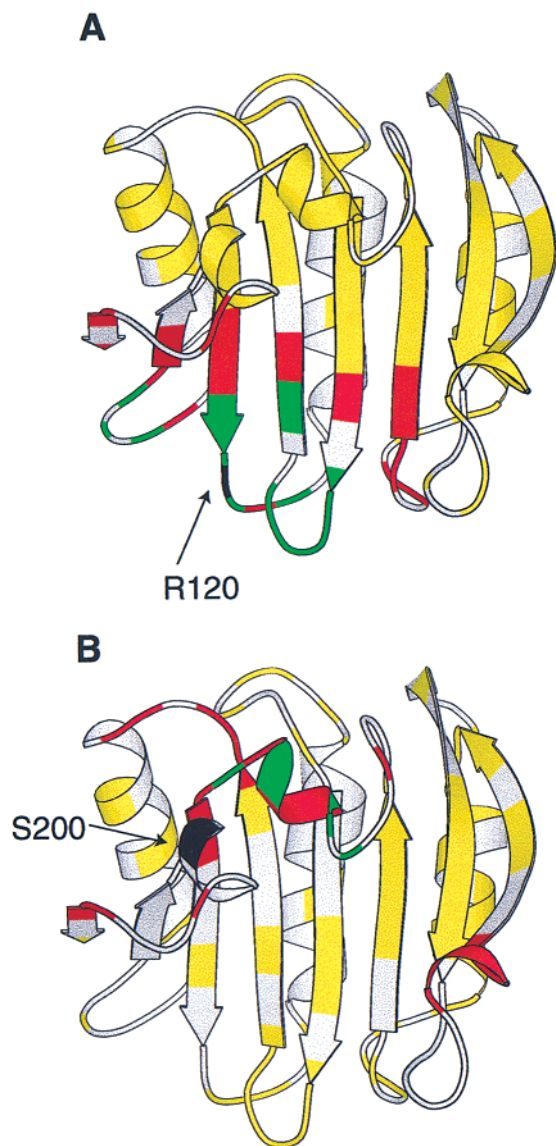
FIGURE 4: Summary of backbone amides that are broadened by a spin label at positions (A) R120C and (B) S200C. Results are shown on the ribbon diagram of the previously determined NMR structure of eIF4E. The position of the spin-label residue is shown in black and marked by an arrow. Amides that are broadened to undetectable levels are shown in green (distance $< \sim$14 Å). Amides that are broadened with measurable intensity ratios < 0.85 are shown in red (distance $\sim$14$-$23 Å). Amides with intensity ratios > 0.90 are shown in yellow (distance $> \sim$23 Å). Uncolored residues are either overlapped in ^{15}N$-$HSQC spectra, not assigned, or have intensity ratios between 0.85 and 0.90.

A graphical depiction of the three PBE distance restraint classes observed for two spin-label positions is shown on the NOE-derived three-dimensional structure of eIF4E (Figure 4). The boundaries of the restraint classes agree very well with the eIF4E structure and appear to be qualitatively sensitive to the orientation of the backbone to the spin-label side chain. For instance, only the side of the short α3 helix facing the S200C spin label is in the shortest distance restraint class (green, no observable cross-peak, <14 Å). Note that the free radical of the nitroxide is typically 8$-$10 Å from the nearest backbone amide (including its own residue), since it is extended into solvent. Therefore, the fact that cross-peaks completely broadened cannot be quantified other than to place an upper bound of approximately 14 Å is not as

Table 1: Precision and Accuracy of Structures Calculated with NOE and Spin-Label Restraints[a]

data set	precision (Å) res 37–213	precision (Å) sec str	accuracy (Å) res 37–213	accuracy (Å) sec str
cap eIF4E				
NOE (reference)[b]	1.30 ± 0.23	0.60 ± 0.16	*NA*	*NA*
NOE/spin (4 Å bounds)[c]	1.26 ± 0.23	0.58 ± 0.14	0.85	0.39
NOE/spin (3 Å bounds)[c]	1.36 ± 0.18	0.56 ± 0.10	1.36	0.70
Apo eIF4E				
HN–HN NOE[d]	8.31 ± 1.35	7.19 ± 1.55	7.94	5.99
HN–HN/spin (4 Å bounds)[e]	2.34 ± 0.20	1.70 ± 0.39	3.20	2.56

[a] Results are rms deviations for backbone atoms (N, Cα, C) of a family of 20 low-energy structures from a target structure. For precision values, the target is the average structure of the respective family. For accuracy values, the target is the average structure of the reference (full NOE data set). Superpositions were performed with either residues 37$-$213 or the amino acids in regular secondary structure elements (39$-$48, 62$-$69, 71$-$80, 91$-$96, 113$-$119, 126$-$138, 151$-$157, 161$-$167, 173$-$185, 194$-$197, and 211$-$213). For the calculations with the limited NOE data sets (HN$-$HN), the m^{7}GDP "cap" was not included in the modeling, although it was present in the NMR experiments. [b] Full NOE data set (*18*) containing 2014 NOEs (553 intraresidue, 628 sequential, 401 medium-range, 428 long-range, and 24 intermolecular), 58 hydrogen bonds, and 279 dihedral angles (82 φ, 89 ψ, 60 χ^1, and 48 χ^2). [c] Full NOE data set plus 515 spin label distances [53 upper bound only, 121 upper/lower bound, 341 lower bound only ("negative" restraint)]. [d] Limited NOE data set with only 403 HN$-$HN NOEs and 171 angle restraints (82 φ, 89 ψ). [e] Limited NOE data set with 515 spin label distances.

limiting as it might appear. In practical use, the amide resonances broadened beyond detection actually represent a 5 Å range between $\sim$9 and 14 Å. A distinct advantage of the spin-label method is that it is a long range interaction, compared to the NOE. This is exemplified by the distances observed from position S200C to the α1 helix across the nucleotide binding site (Figure 4). The orientation of the loop/helices across the nucleotide binding site could not be directly defined from NOE analysis of protein resonances alone, although it is defined via intermolecular NOEs to the m^{7}GDP nucleotide in the high-resolution NMR structure.

Molecular Modeling Calculations with PBE Restraints. To quantitatively verify that the distances derived from PBEs were consistent with the NOE distances, simulated annealing calculations were performed combining the full NOE and PBE restraint sets (Table 1). This verification is important since there are no covalently "fixed" distances to calibrate the spin label broadening, like there are for NOEs. In addition, it was important to empirically determine the proper bounds to use for the spin-label distances in a highly restrained system. For all modeling calculations in this paper, the covalent structure of all spin-label side chains were included, regardless of whether the PBE distances were utilized in the calculations. Inclusion of the spin-label restraints with 4 Å bounds and the full NOE set resulted in structures nearly identical to the NOE-only structure. Importantly, the structures converged with few NOE violations, indicating that the NOE and PBE distances were consistent with one another. The same calculations performed with tighter bounds on the spin label distances (3 Å) did not converge as readily (data not shown) and often contained one or two NOE violations > 0.5 Å. In addition, the "accuracy" rms deviations from the reference structure got worse (Table 1). Together, this suggested that the 3 Å bounds were too tight for modeling calculations. Bounds of 4 Å
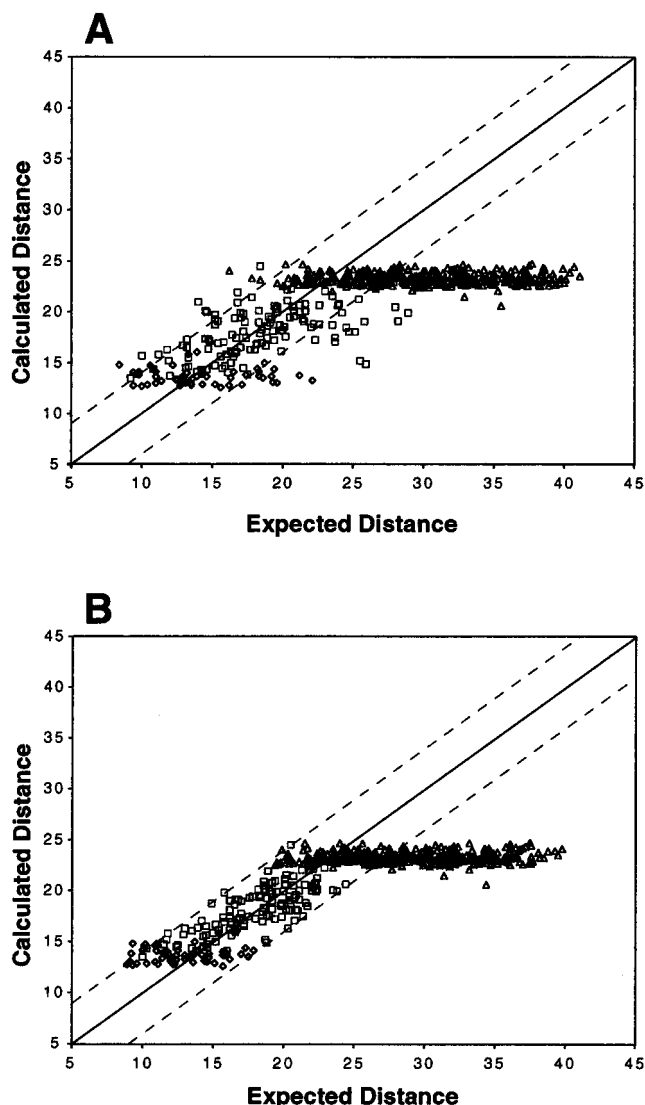
FIGURE 5: Calculated versus expected distances from five spin-labeled samples of eIF4E. "Calculated" distances are determined from NMR spectra of ^{15}N-labeled protein. "Expected" distances are determined from three-dimensional structures calculated with (A) full NOE data set or (B) full NOE data set plus all spin-label distances (4 Å bounds). Each spin-label distance is marked according to three distance restraint classes by (○) upper bound only, (□) upper and lower bounds, and (△) lower bound only. Where "calculated" = "expected" is marked by a solid line, with 4 Å bounds marked by dashed lines.

accommodate up to ~20−30% error in the PBE distances ranging from 14 to 23 Å, which is reasonable when compared to bounds typically employed for NOEs. Unfortunately, the PBE distances did not appear to refine the high-resolution NMR structure (i.e., lower precision).

The consistency of the PBE (4 Å bounds) and NOE data is shown in graphical form in Figure 5. In Figure 5A the expected vs calculated distance for each spin-label restraint is shown for the average structure calculated with the NOE data only. The "calculated" distance is that obtained from the PBE, while the "expected" distance is determined from the reference structure (distance from the nitrogen atom of the nitroxide to the hydrogen atom of the respective amide). Since the position of the nitroxide ring is underdetermined by the NOE data alone (essentially determined by the force field), the "expected" distance from the NOE-only structure is not a rigorous check or calibration of the PBE distances.

Nevertheless, the general correlation between the expected vs calculated distance can be observed. In Figure 5B the expected vs calculated distance is shown for the average structure calculated with all NOEs and PBEs combined. This graphically depicts the result that the PBE distances are consistent with the NOEs and can refine with all PBE distances satisfied within the 4 Å bounds (dotted lines in Figure 5). The distances within the 4 Å boundaries appear to be evenly distributed about the line "calculated" = "expected", indicating that there is not a strong bias or systematic error affecting the distances. It might have been expected that r^{-6} averaging, caused by flexibility of the spin-label side chain, would lead to calculated distances that were shorter than the true or "expected" distance. However, at the large distances probed by the spin label (>14 Å), it would take very large motions to produce significant r^{-6} averaging effects, and the results of Figure 5B suggest that it is not a major problem.

Structure calculations were then performed with the limited NOE and PBE data sets. The limited NOE data set (HN−HN) is from ^{15}N amide protons that could be obtained from a large perdeuterated protein (Table 1). The HN−HN NOE data set is taken from actual 3D and 4D ^{15}N-edited NOESY experiments of partially deuterated eIF4E (18). In addition, ϕ/ψ angles were loosely restrained for secondary structure elements that could be identified from the chemical shift index (CSI) of backbone ^{1}H/^{15}N/^{13}C assignments. Calculations performed with just HN−HN and CSI data resulted in very poor quality structures with the global fold incorrectly determined (defined as proper arrangement/topology of secondary structure elements) for the majority of the family (Figure 6B and Table 1). Seven out of the 20 low-energy structures had the correct global fold (data not shown), yet still have high precision and accuracy rms deviations (2.81 ± 0.55 and 5.51, respectively, for residues 37−213).

Inclusion of the PBE distances to the HN−HN NOEs dramatically improved the quality of structures and correctly determined the overall global fold of the protein (Figure 6C). The precision and accuracy of the PBE-derived structures are similar to what has been achieved for other methods for global fold determination utilizing selective protonation (7, 9). This demonstrates that it should be possible to correctly determine the global fold of a large protein via SDSL given that backbone ^{15}N-proton assignments have been obtained.

Simulated Calculations of eIF4E Using Alternative Method for Global Fold Determination. Simulated restraints for a methyl-protonated sample were generated for eIF4E as described (9) to compare with the spin-label results. The simulated methyl NOEs also correctly determined the global fold of eIF4E with slightly better statistics than the spin-label method (Figure 7A and Table 2). In particular, the packing of the three α-helices against the β-sheet was better defined by the methyl-protonation method. However, the binding site for the m^{7}GDP nucleotide was not as well-defined due to the lack of short-range NOEs between the loops in the cleft (in the absence of the nucleotide in the modeling calculations). This type of long-range restraint is available from the PBEs as was shown in Figure 4. Combining the two methods improved the precision of the structures obtained (Figure 7B). In particular, the binding-site cleft is well-defined even in the absence of the nucleotide in the modeling calculations. The precision of the combined
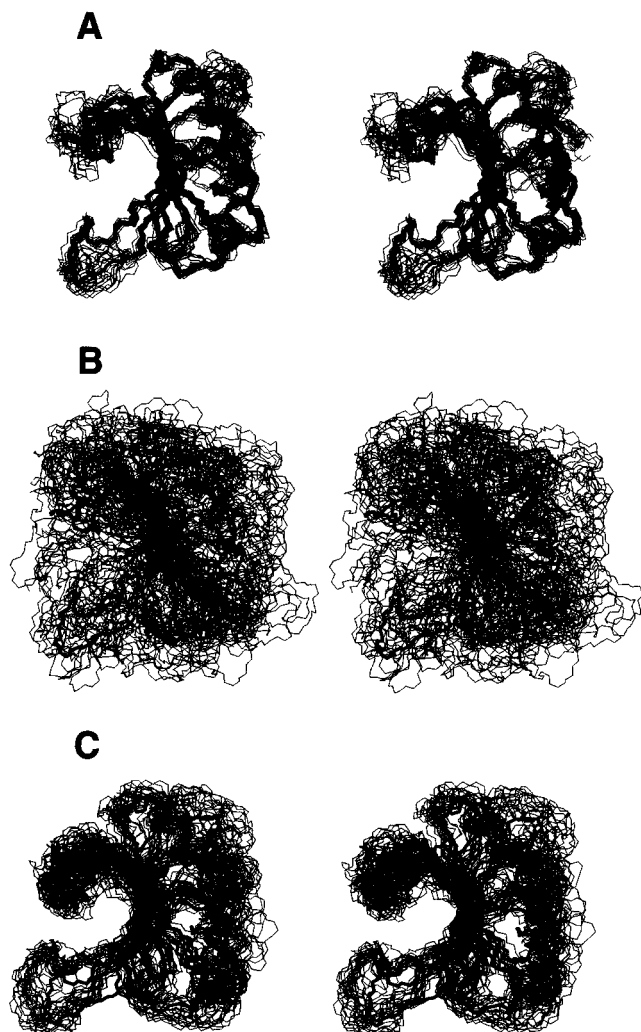
FIGURE 6: Superpositions of backbone atoms for families of structures calculated with (A) full NOE data set, (B) reduced HNHN NOE data set, and (C) reduced HNHN NOE data set plus spin-label distances. The full NOE data set is calculated with m^{7}GDP and intermolecular NOEs, while the reduced NOE data sets are calculated without m^{7}GDP present in the calculations. The family of 20 low-energy structures for each is shown in thin black lines, while the average structure of the "reference" calculation with the full NOE data set is shown by a thick black line in all parts.
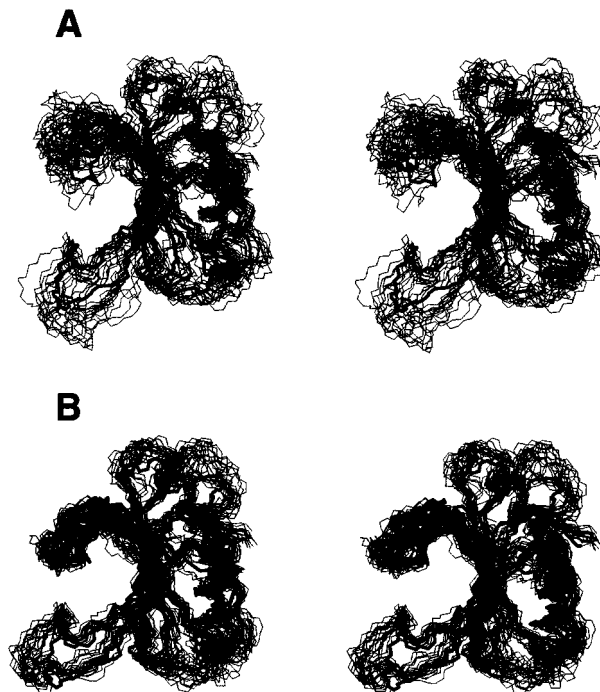

FIGURE 7: Superpositions of backbone atoms for families of structures calculated with simulated restraints for a methyl-protonated, highly deuterated protein sample. (A) Structures calculated with reduced HN−HN NOE data set plus simulated methyl−methyl and methyl−HN NOEs. (B) Structures calculated with methyl NOEs (panel A) plus PBE distances (see Table 2).

Table 2: Precision and Accuracy of Structures Calculated with Simulated Restraints^a

eIF4E (apo) data set	precision (Å)		accuracy (Å)	
	res 37−213	sec str	res 37−213	sec str
met^b	1.96 ± 0.38	0.99 ± 0.15	2.33	1.27
met-spin^c	1.47 ± 0.13	0.87 ± 0.10	2.24	1.19

FPS data set	precision (Å)		precision (Å)	
	res 21−366	sec str	res 21−366	sec str
met^d	8.35 ± 2.67	7.04 ± 2.31	7.90	5.82
spin^e	12.86 ± 1.55	12.33 ± 1.84	8.15	7.58
met-spin^f	2.00 ± 0.22	1.64 ± 0.28	2.37	2.11

^{a} Results are rms deviations of backbone atoms (N, Cα, C) of a family of 20 low-energy structures fro, a target structure. For precision values, the target is the average structure of respective family. For accuracy values, the target is the average structure of the reference for eIF4E (Table 1) or the crystal structure for FPS. For eIF4E, the superpositions are for the same regions as Table 1. For FPS, the superpositions are for residues 21−366 or the secondary structure elements (residues 23−46, 53−66, 73−85, 94−119, 142−158, 167−191, 205−229, 237−262, 284−289, 296−303, 310−320, 325−346, and 352−362). ^{b} Limited HN−HN NOE data plus 241 simulated methyl protonation NOEs (205 HN−methyl and 36 methyl−methyl). ^{c} Limited HN−HN NOE data plus 241 simulated methyl NOEs plus 515 spin-label PBEs (Table 1) plus 196 simulated spin-label−methyl PBEs (21 upper only, 38 upper/lower, and 137 lower only bounds). ^{d} Simulated NMR data containing a total of 868 NOEs (463 HN−HN, 298 HN−methyl, and 102 methyl−methyl) and 456 angle restraints (228 φ and 228 ψ). ^{e} Simulated NMR data containing a total of 463 HN−HN NOEs, 456 angle restraints (228 φ and 228 ψ), and 2311 PBEs (97 upper only, 266 upper/lower, and 1948 lower only bounds). ^{f} Simulated NMR combining met and spin NOEs and PBEs plus 1045 PBEs to methyl groups (52 upper only, 142 upper/lower, and 851 lower only bounds).

restraints is almost comparable to the results obtained with the full NOE data set.

For structural analysis at the level of the backbone fold, these structures all have similar interpretative quality (Figure 8). Noticeable insufficiencies in the accuracy are the loose packing of the helices for the spin-label method (Figure 8B) and a "relaxed" binding-site cleft for the methyl-protonation method (Figure 8C). Both of these problems are mitigated when the methods are combined (Figure 8D), indicating that they contain complementary data. The long-range nature of the PBE is very significant and should be even more important for determining the structures of larger proteins that will often contain multiple domains requiring long-range information for proper orientation.

Simulated Calculations on the Helical Protein FPS. Previous results have demonstrated that the methyl-protonation method worked better for proteins containing high β-sheet content. Helical proteins are problematic since there are virtually no long-range HN−HN or HN−methyl NOEs for amides in helices. This contrasts with β-sheets where

the secondary structure information is inherently long-range, and there can be many observable HN−HN or HN−methyl
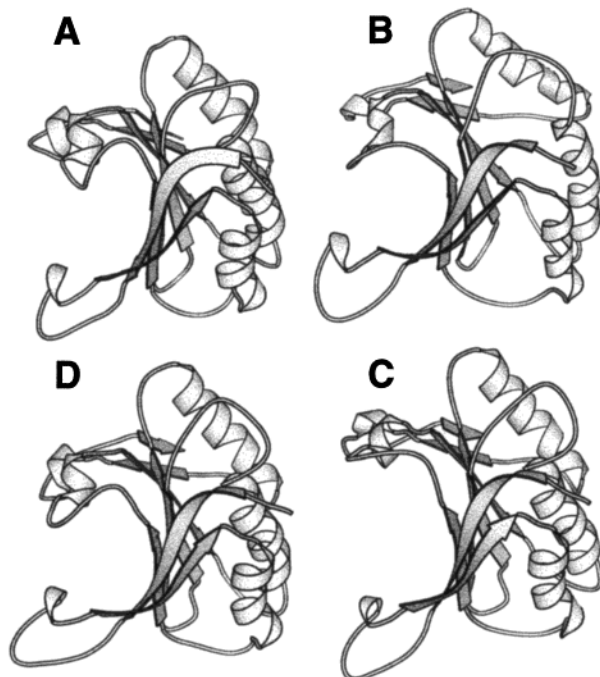
FIGURE 8: Average structures of eIF4E calculated with the following restraint sets: (A) full NOEs, including m⁷GDP and intermolecular NOEs; (B) PBEs and HN−HN NOEs; (C) methyl-protonation NOEs; and (D) PBEs and methyl-protonation NOEs. The backbone is shown from residues 37 to 213.

NOEs between β strands. Therefore, we performed modeling calculations on the 40 kDa protein FPS, which contains exclusively α-helices, to test whether the PBE restraints could help determine the global fold of a helical-rich protein. The crystal structure of FPS is known (*24*) and was used to generate simulated methyl−methyl, methyl−HN, and HN−HN NOEs, as well as simulated PBE restraints from 12 randomly spaced spin labels on the surface of FPS (Table 2). It was estimated that approximately 50% of the amides in an ¹⁵N−HSQC of a 40 kDa protein would be sufficiently resolved for PBE analysis. This number is lower than the approximately 70% of residues available as probes for spin broadening in the structured core of eIF4E (37−213).

As was previously published (*9*), the simulated restraints for the methyl-protonation method did not correctly determine the global fold of FPS, although most of the errors were minor rearrangements of helices. The precision and accuracy for the family of structures was ∼8 Å (Table 2). Use of the spin-label method alone also did not correctly determine the global fold of FPS. The higher precision rms deviations for the spin-label method reflect the presence of two subsets of structures in the low-energy family, one similar to the crystal structure (and methyl-protonation calculations) and the other a mirror image (data not shown). The precisions of these individual subsets are similar to the methyl-protonation method. When the two methods are combined, however, the global fold of FPS is correctly determined with a precision and accuracy of ∼2 Å (Figure 9A). A view of the ribbon diagrams of the crystal structure and average structure from the simulated calculations is shown in Figure 9B,C. Most pertinent structural features of the backbone are correctly determined by the simulated NMR structure. Again, the complementary nature of the two methods for global fold determination of proteins is apparent
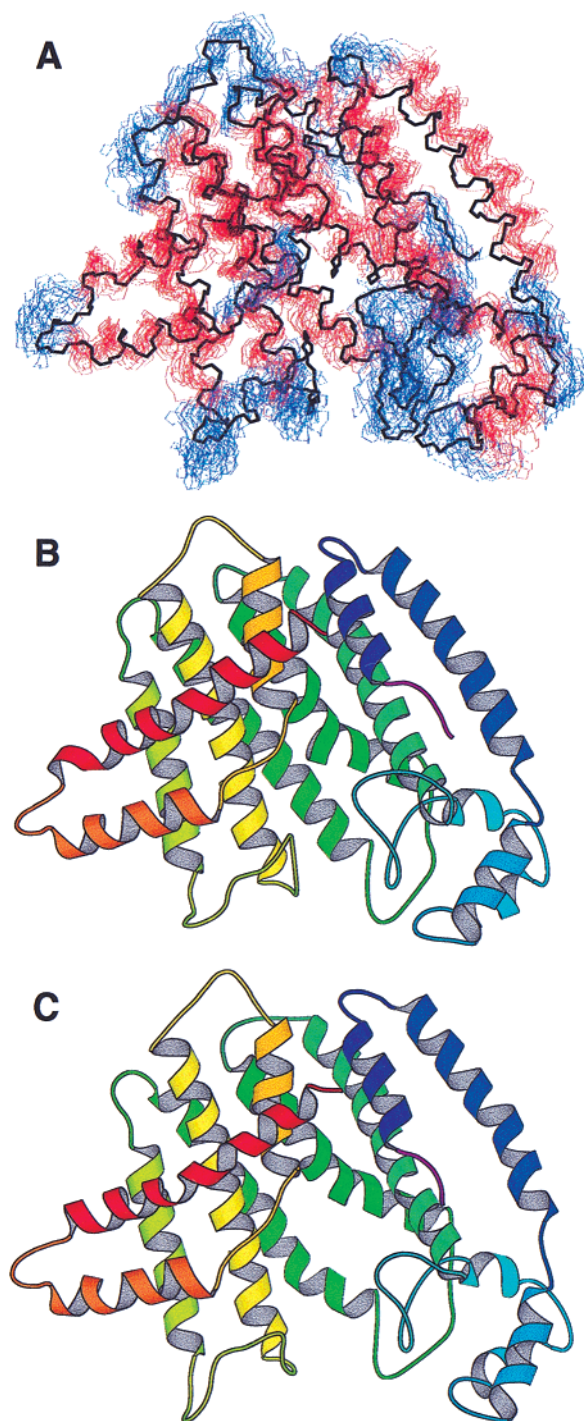

FIGURE 9: Simulated structure calculations of the protein FPS. (A) Family of 20 low-energy structures calculated with simulated PBEs and methyl-protonation NOEs. Helical residues are shown in red and remaining residues in blue. Backbone atoms of the crystal structure from residues 21 to 366 are shown with a thick black line. (B) Ribbon diagram of the crystal structure. (C) Ribbon diagram of the average structure from the simulated calculations. Both ribbon diagrams are rainbow colored from red to blue, progressively, from the N- to C-terminus.

and holds great promise for determining medium-resolution structures of large proteins.

Future Prospects. Current NOE-based methods can determine monomeric protein structures up to ∼30 kDa. From a relaxation standpoint, backbone assignments are obtainable up to at least 70 kDa with perdeuteration; however, spectral overlap practically limits assignments of monomeric proteins

in the 30−40 kDa range in the absence of sophisticated segmental labeling schemes (*25, 26*). In macromolecular complexes, however, overlap can be alleviated by isotopic labeling of only one component. Thus, at this time most NMR methodologies for study of larger macromolecules are more applicable to the study of protein−protein or protein−nucleic acid complexes. There is a disadvantage to NOE-based methodologies for complexes in that isotopic filtering experiments, which are often performed to determine inter-molecular NOEs, are unlikely to be successful at higher molecular weights. Definition of the intermolecular interface, however, should be attainable via spin labeling of the unlabeled component of the complex, while the broadening of the isotopically labeled component is monitored, since the PBE is relatively insensitive to the molecular weight of the system. In addition, there is often a paucity of NOEs at protein−nucleic acid interfaces, and spin labeling may be a useful technique for study of these complexes even in situations where NOE analysis is possible.

As was demonstrated in this paper, PBEs are probably best utilized in combination with other methods, such as methyl protonation, to improve structure quality. In particular, the combination with residual dipolar couplings holds promise to achieve significantly higher quality structures than presented here. Dipolar couplings provide purely angle orientations of H−N or H−C bond vectors to a magnetic susceptibility axis; therefore, some distance information is required for translational orientation of secondary structure elements. Combining the spin labels and dipolar couplings might provide a method of structure determination that was solely dependent on the use of ^{1}H−^{15}N correlation spectra, which are well-resolved and easy to analyze. ^{1}H−^{13}C correlation spectra could also be used in situations where they could be easily obtained. The major drawback to the SDSL methodology is that many different mutations and samples are required. The number of spin-label positions will also need to increase with larger molecular weight systems. However, the advantage is that after initial structure determination, the spin labels will provide detailed structural probes for biochemical studies involving binding of ligands/macromolecules and/or the conformational changes that occur upon binding, which is the traditional use of spin labels with nuclear magnetic resonance.

REFERENCES

1. Kay, L. E., and Gardner, K. H. (1997) *Curr. Opin. Struct. Biol. 7*, 722−731.

2. Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H., and Kay, L. E. (1996) *J. Am. Chem. Soc. 118*, 6570−6579.

3. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 12366−71.

4. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K. (1999) *J. Biomol. NMR 14*, 85−88.

5. Venters, R. A., Metzler, W. J., SPicer, L. D., Mueller, L., and Farmer, B. T. (1995) *J. Am. Chem. Soc. 117*, 9592−9593.

6. Metzler, W. J., Witteking, M., Goldfarb, V., Mueller, L., and Farmer, B. T. (1996) *J. Am. Chem. Soc. 118*, 6800−6801.

7. Smith, B. O., Ito, Y., Raine, A., Teichmann, S., Ben-Tovim, L., Nietlispach, D., Broadhurst, R. W., Terada, T., Kelly, M., Oschkinat, H., Shibata, T., Yokoyama, S., and Laue, E. D. (1996) *J. Biomol. NMR 8*, 360−368.

8. Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T., and Kay, L. E. (1996) *J. Mol. Biol. 263*, 627−636.

9. Gardner, K. H., Rosen, M. K., and Kay, L. E. (1997) *Biochemistry 36*, 1389−1401.

10. Tjandra, N., and Bax, A. (1997) *Science 278*, 1111−1114.

11. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M., and Bax, A. (1997) *Nat. Struct. Biol. 4*, 732−728.

12. Kosen, P. A. (1989) *Methods Enzymol. 177*, 86−121.

13. Krugh, T. R. (1976) in *Spin Labeling: Theory and Applications* (Berliner, L. J., Ed.) pp 339−372, Academic Press, New York.

14. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 170−184.

15. Girvin, M. E., and Fillingame, R. H. (1994) *Biochemistry 33*, 665−674.

16. Gillespie, J. R., and Shortle, D. (1997) *J. Mol. Biol. 268*, 158−169.

17. Hubbell, W. L., Mchaourab, H. S., Altenbach, C., and Lietzow, M. A. (1996) *Structure 4*, 779−783.

18. Matsuo, H., Li, H., McGuire, A. M., Fletcher, C. M., Gingras, A.-C., Sonenberg, N., and Wagner, G. (1997) *Nat. Struct. Biol. 4*, 717−724.

19. Berliner, L. J., Grunwald, J., Hankovszky, H. O., and Hideg, K. (1982) *Anal. Biochem. 119*, 450−455.

20. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.

21. Solomon, I., and Bloembergen, N. (1956) *J. Chem. Phys. 25*, 261−266.

22. Mchaourab, H. S., Lietzow, M. A., Hideg, K., and Hubbell, W. L. (1996) *Biochemistry 35*, 7692−7704.

23. Schmidt, P. G., and Kuntz, I. D. (1984) *Biochemistry 23*, 4261−4266.

24. Tarshis, L. C., Yan, M., Poulter, C. D., and Sacchettini, J. C. (1994) *Biochemistry 33*, 10871−10877.

25. Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., and Uegaki, K. (1998) *J. Am. Chem. Soc. 120*, 5591−5592.

26. Xu, R., Ayers, B., Cowburn, D., and Muir, T. W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 388−393.