NMR structure determination and investigation using a reduced proton (REDPRO) labeling strategy for proteins

Alexander Shekhtman[a], Ranajeet Ghose[a], Michael Goger[a,b], David Cowburn[a,b,*]

[a] *The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399, USA*
[b] *New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027-7556, USA*

Received 24 April 2002; revised 7 June 2002; accepted 14 June 2002

First published online 8 July 2002

Edited by Thomas L. James

Abstract We present here a stable isotope labeling technique for proteins, which seeks the appropriate compromise between the advantages of (a) random isotope labeling, with its large number of protons available for structure determination, and (b) selective labeling to generate isolated proton spins decreasing spectral complexity and improving relaxation properties of NMR experiments. The described reduced proton (REDPRO) procedure results in side-chain specific protonation of overexpressed proteins, which is highly selective. The REDPRO labeling scheme provides a sufficient number of NOE constraints for structure calculation. Dramatically improved relaxation properties of the heteronuclear magnetization transfer coupled with TROSY advantages make the proposed labeling scheme an attractive approach for study of high molecular weight protein targets, their ligand sites, and interdomain interactions.

Key words: Stable isotope labeling;
Nuclear magnetic resonance relaxation;
Solution structure determination

1. Introduction

The study of biological molecules by solution NMR spectroscopy has greatly benefited from the use of isotope labeling. The utility of uniform ^{15}N,^{13}C labeling and triple resonance spectroscopy permits structure determination and functional characterization of proteins on the order of 30 kDa (e.g. [1]). Application of NMR-based techniques of structure determination to proteins and protein complexes of higher molecular weight is hampered mainly by the highly efficient spin relaxation from the ^{1}H–^{1}H dipolar interactions [2,3]. These relaxation pathways lead to decreased magnetization transfer efficiency and broadened spectral lines in heteronuclear multidimensional experiments resulting in a loss of both sensitivity and resolution. An obvious way to alleviate this problem is to substitute non-exchangeable protons by deuterium to take advantage of the 6.5-fold lower gyromagnetic ratio of ^{2}H relative to ^{1}H, suppressing spin relaxation pathways that would otherwise be present in fully protonated samples [4]. Random uniform incorporation of ^{2}H to various degrees has been used in many structural studies (e.g. [4,5]). Several samples with different levels of deuteration are usually required for structure determination. Selective protonation of methyl groups [6] or backbone C^{α} atoms [7] in a highly deuterated environment solves the problem of spectral complexity but leaves very few useful protons for structural studies relying on traditional NOE-based approaches, and can be prohibitively expensive.

Ideally, one would like to have a single NMR sample that contains a sufficiently low proton density to improve relaxation properties of triple resonance NMR experiments and at the same time contains enough protons for structural studies using NOE constraints. We present here a simple labeling method which seeks the appropriate compromise of the advantages of random uniform labeling with its large number of protons available for structure determination, and selective labeling methodology to generate isolated proton spins with decreased spectral complexity and improved relaxation properties of the NMR experiments.

2. Materials and methods

2.1. Protein overexpression

The technique is based on the observation that a culture of *Escherichia coli* strain BL21 containing an overexpression vector for the protein of interest can be grown in minimal medium containing 100% D_2O supplemented with [U-^{13}C,^{1}H]glucose and [U-^{15}N,^{1}H]-ammonium chloride without prior adaptation [8]. We developed a protocol that was tested on five proteins, human ubiquitin, human Abl SH(32), σ^{A} factor from *Thermotoga maritima*, *E. coli* GroEL, and *E. coli* GroES, all of which produced a consistent labeling pattern. The host strain of *E. coli* BL21 (Novagen) was freshly transformed with each expression construct. Cells from overnight culture grown on unlabeled M9 minimal medium in $^{1}H_2O$ were collected by centrifugation, washed in phosphate-buffered saline, resuspended in labeling minimal medium, and grown at 37°C from OD_{600} 0.5 to 0.8. In about 2–3 h the cells adapt to growth in D_2O and reach the indicated cell density. Protein overexpression was induced by addition of 0.5 mM IPTG and the cells were aerated for 20 h at 37°C. Finally, the cells were collected and stored at −80°C for further purification. The protein yield using the reduced proton (REDPRO) labeling scheme in all five cases of protein overexpression was very similar to that of the standard [U-^{13}C,^{15}N] labeling scheme [4]. The specific details of protein purification have been previously published for human ubiquitin [9], human Abl SH(32) [10]. σ^{A} factor from *T. maritima* [11], *E. coli* GroEL and GroES [12].

2.2. NMR spectroscopy

NMR spectra were recorded on Bruker DMX spectrometers operating at ^{1}H frequencies of 500 and 600 MHz and equipped with triple-resonance probes with *z*-axis magnetic field gradient capability. CBCA(CO)NH, CC(CO)NH, HNCACB, ^{1}H-{^{13}C} HSQC and homonuclear NOESY spectra [1] were collected at 27°C, processed using the NMRPipe suite [13] and analyzed using XEASY software [14]. NMR samples of [U-^{15}N,^{13}C] and REDPRO labeled human ubiquitin with concentrations of 1 and 2 mM, respectively, were exchanged into

*Corresponding author. Fax: (1)-646-349 2840.
E-mail address: cowburn@nysbc.org (D. Cowburn).

0014-5793/02/$22.00
PII: S0014-5793(02)03051-X

30 mM sodium acetate pH 4.6, 10% D_2O, 0.01% NaN_3. Deuterium decoupling using the WALTZ16 scheme was employed in all NMR experiments during carbon evolution. The CBCA(CO)NH spectrum was collected with 40, 70 and 512 complex points in the ^{15}N, ^{13}C and direct ^{1}H dimensions, respectively with four scans per transient. The corresponding sweep widths were 32 ppm, 80 ppm and 15 ppm. The data in the ^{15}N dimension were doubled by data mirroring. The data in all dimensions were multiplied by a squared cosine-bell function and zero-filled to 128, 128 and 1024 points prior to Fourier transformation. The HNCACB dataset was collected with 32, 100 and 512 complex points in ^{15}N, ^{13}C and direct dimensions respectively. The sweep width in the ^{15}N dimension was 26 ppm. All other parameters and processing modes were the same as in the case of the CBCA-(CO)NH experiment. The CC(CO)NH experiment was collected and processed the same way as the HNCACB experiment using a 7.8 ms DIPSI2 mixing scheme but no linear prediction was used in the ^{15}N dimension. The homonuclear NOESY spectra were collected with 512 complex points in each dimension with a sweep width of 14 ppm and a mixing time of 500 ms with four scans per transient. The data in both dimensions were multiplied by a squared cosine-bell window function and zero-filled to double their size prior to Fourier transformation.

2.3. Structure calculation

Structure calculations of ubiquitin were performed using DYANA v1.5 [15] using homonuclear NOESY connectivities filtered for the observed labeling pattern and reported proton–proton distances [16] and loose angular (ϕ, ψ, and χ_1) constraints generated from chemical shifts using the TALOS [17] procedure and measured scalar couplings. In all, 1000 structures were generated and subjected to 4000 steps of torsional angle dynamics at an elevated temperature, followed by 16 000 steps of simulated annealing. Finally, the resulting structures were subjected to 8000 steps of conjugate gradient minimization. A typical structure calculation takes 35 s on a Dell Precision 330 equipped with Pentium IV processor operating at a clock frequency of 1.7 GHz. The 10 structures that had the lowest value for the target function were considered to represent the structure of ubiquitin.

3. Results and discussion

The REDPRO procedure results in products with the following composition: the total incorporation of deuterium depends on the amino acid composition of the protein and ranges from 70% to 80% of the chemically non-exchangeable protons. Protonation was sequence independent, side-chain specific and, surprisingly, highly selective. A detailed analysis of the spectrum of ubiquitin (Fig. 1) shows that methyl groups of Ala, Val, Leu, Ile, Met, Thr and methine group of Thr, a total of 44, had intensities equivalent to protonation from 55% to 64%. Methylene groups of Arg (δ), Ile (γ), Lys (γ, δ, ε), Pro (δ), Ser (β), Asp (β), Glu (β, γ), Asn (β), and Gln (β, γ), a total of 69, had equivalent proton intensities to a lesser degree (27–36%). Representative spectra for these isotopic ratios are shown in Fig. 2. For methyl groups (a total of 37), CHD_2 (73%) and CH_2D (22%) were the major isotopomeric fragments and for methylene groups CHD (87%) constituted the major isotopomer. The isotopomer subspectra are discriminated by small ^{13}C and ^{1}H shifts associated with the additional ^{2}H [4]. The fraction of completely protonated methyl and methylene was less that 10%. The existence of methylene isotopomeric fragments CH^RD^S and CH^SD^R results in multidimensional spectra in which two corresponding positions at

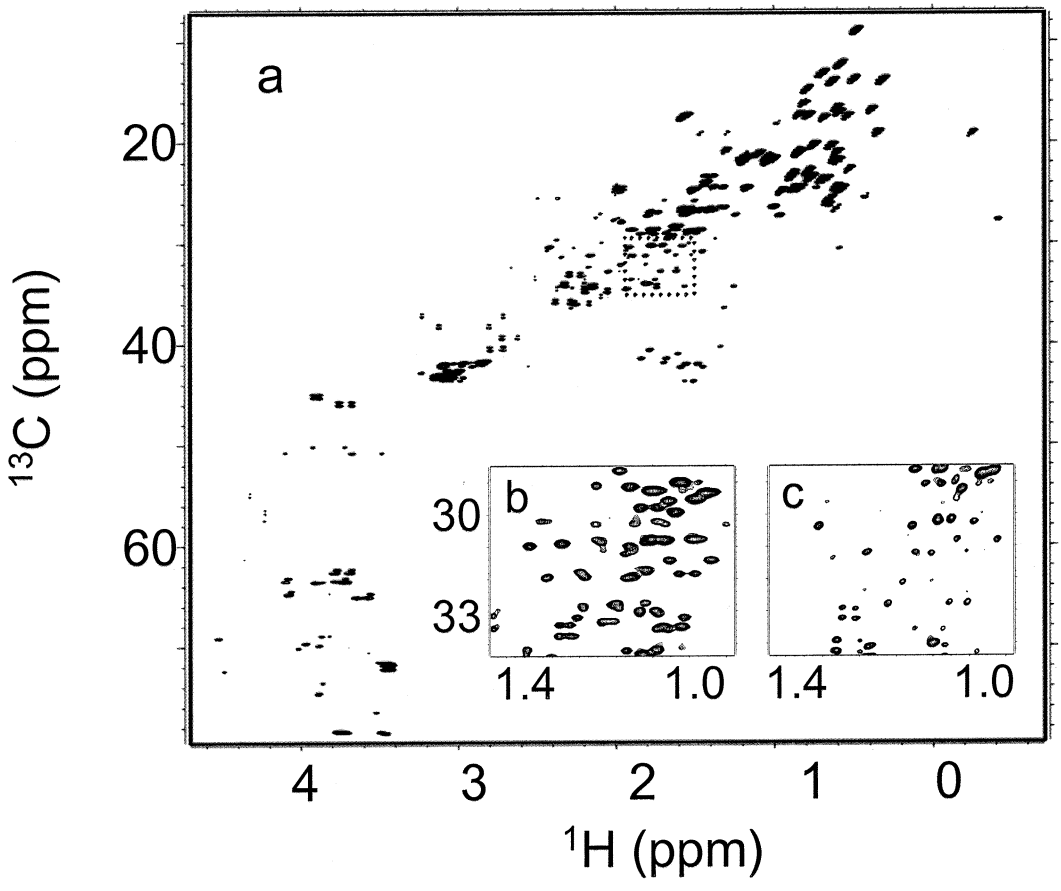

Fig. 1. ^{1}H-{^{13}C} HSQC spectrum of ubiquitin labeled using REDPRO scheme (a and c) and [U-^{15}N,^{13}C]ubiquitin (b). The samples were in 30 mM sodium acetate pH 4.6, 10% D_2O, 0.01% NaN_3. In comparing panels b and c, the ^{13}C and ^{1}H shifts in carbon and proton dimensions are generally ^{2}H isotope shifted upfield.

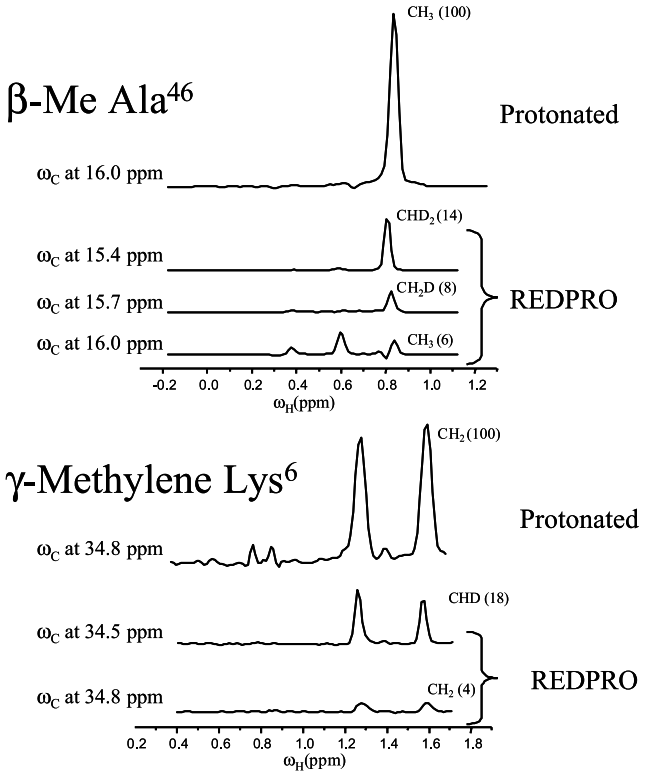

Fig. 2. Representative slices of Ala[46] methyl and Lys[6] γ-methylene peaks of ^{1}H-{^{13}C} HSQC spectra of ubiquitin labeled using the REDPRO scheme and [U-^{15}N,^{13}C]ubiquitin. Numbers in parentheses reflect the peak volume of the isotopomers of methyl and methelene groups of ubiquitin labeled using REDPRO relative to [U-^{15}N, ^{13}C]ubiquitin. In this and several other cases, the ratio of $CH^{R}D^{S}$/$CD^{R}H^{S}$ differs slightly from 1.0.

the same ^{13}C chemical shift represent the isotopic mixture, and the hydrogen dimension has no multiplicity from homonuclear spin coupling [18]. Such spectra are considerably simpler to analyze than equivalent spectra from [U-^{13}C,^{15}N] labeled proteins (Fig. 3). Aromatic side-chains of Tyr (δ, ε), Phe

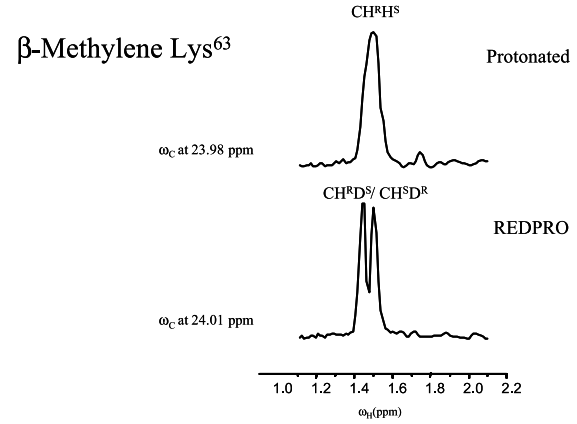

Fig. 3. Representative slice of Lys[63] γ-methylene peak of ^{1}H-{^{13}C} HSQC spectra of ubiquitin labeled using the REDPRO scheme and [U-^{15}N,^{13}C]ubiquitin. The peaks representing CH_2 moieties in the protonated NMR samples are multiplets arising from homonuclear spin coupling [18]. On the other hand, methylene isotopomeric fragments $CH^{R}D^{S}$ and $CH^{S}D^{R}$ in the NMR samples labeled using the REDPRO scheme are singlets and can be resolved in the multidimensional NMR spectra. ^{1}H-{^{13}C} HSQC spectra were apodized using a squared cosine-bell function to enhance resolution in the proton dimension in both panels.

(δ), and His (δ, ε), a total of six, were protonated from 67% to 73% whereas those of Phe (ε, ζ), a total of four, were protonated from 19% to 22%. Overall, we see protonation of about 23% of non-exchangeable protons that is a compromise between low proton density and high information context of the spectrum. Deuterium is present at 95% of all C^{α}s. This labeling pattern is consistent with well-known glucose metabolism and amino acid synthesis pathways. While there is a general expectation that each site per amino acid type would have constant isotopic enrichment, small variations would have little effect on use for structure calculation because the dependence of derived distance on cross peak intensity in the NOE spectrum is typically at the one-sixth power.

One of the apparent challenges of structural studies of proteins with reduced density of protons is resonance assignment.

Table 1
Simulated structure calculations for selected NOE sets in REDPRO labeled ubiquitin[a]

	Set A	Set B
Constraints		
Distance constraints	236	339
ϕ, ψ constraints	92	92
χ_1 constraints	35	35
Hydrogen-bonding constraints	36	92
Structural characteristics		
RMSD (Å)[b,c]		
Backbone	1.7 ± 0.5	0.6 ± 0.1
All heavy atoms	2.3 ± 0.5	1.1 ± 0.2
Dihedral angles		
Most favored	76.2%	74.7%
Additional favored	19.8%	19.2%
Generously favored	2.0%	3.2%
Disallowed	2.0%	2.9%

[a]Set A includes NOEs to the backbone ^{1}H′. Set B includes all NOEs from set A and methyl–methyl and aliphatic–aromatic NOEs.
[b]All RMSDs calculated including residues 2–71.
[c]The backbone RMSD for the final structures for set B with the mean structure from the reference (1D3Z) is 1.1 ± 0.1 Å in regions with definite secondary structure (1.2 ± 0.2 in sheets and 0.7 ± 0.1 in helices).

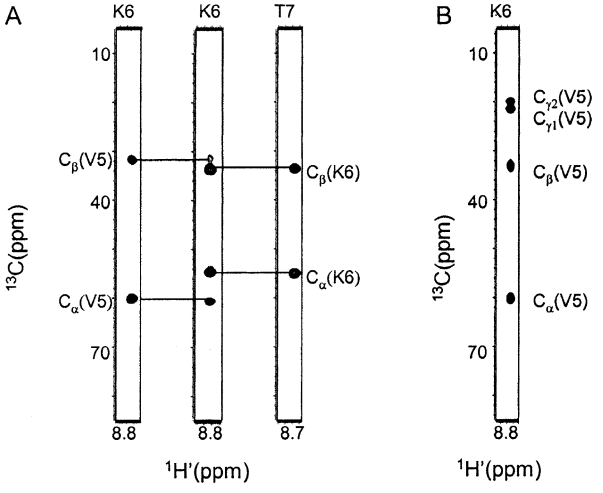

Fig. 4. Representative sequential assignment of 2 mM ubiquitin labeled using REDPRO: (A) strips from the CBCACONH and HNCACB datasets, (B) strip from the CCC(CO)NH dataset. The sample was in 30 mM sodium acetate pH 4.6, 10% D_2O, 0.01% NaN_3.

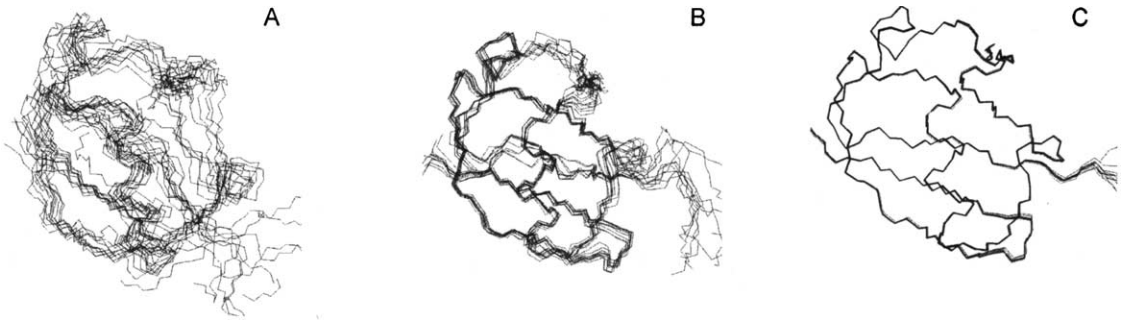

Fig. 5. Cluster of 10 ubiquitin structures with the lowest value of the target function calculated using DYANA based on (A) H′/aliphatic proton distances (excluding H′/Hα distances) (backbone RMSD = 1.7 Å), (B) the same as a plus methyl–methyl and aromatic–aliphatic proton distances (backbone RMSD = 0.6 Å), and (C) solution structures of ubiquitin (PDB code ld3z). The structures in C were calculated using a full set of NOEs as well as ^{1}H–^{15}N, ^{1}H–$^{13}C^{\alpha}$, $^{13}C^{\alpha}$–$^{13}C^{\beta}$ and ^{15}N–$^{13}C'$ residual dipolar couplings in several alignment media [16].

The standard approach of using aliphatic protons to start or end triple resonance NMR experiments fail. We demonstrate (Fig. 4) that triple resonance experiments can be run initiating magnetization transfer from ^{13}C nuclei. The nominal four-fold loss in sensitivity is compensated by significantly improved relaxation properties of the sample in standard CBCACONH and CC(CO)NH experiments. This advantage is expected to be of greatest significance at high molecular weights as observed previously [19].

3.1. *Does the labeling scheme permit sufficient NOEs for the generation of structure?*

Generating several sets of structures using subsets of NOE information tested the utility of the REDPRO labeling scheme to provide sufficient number of NOE constraints for structure calculation. Fig. 5A shows the structures of ubiquitin calculated with H′-aliphatic constraints, excluding H′-H^{α} constraints, and loose (ϕ, ψ, and χ_1) dihedral angle restraints. (ϕ, ψ) restraints were obtained from chemical shifts using

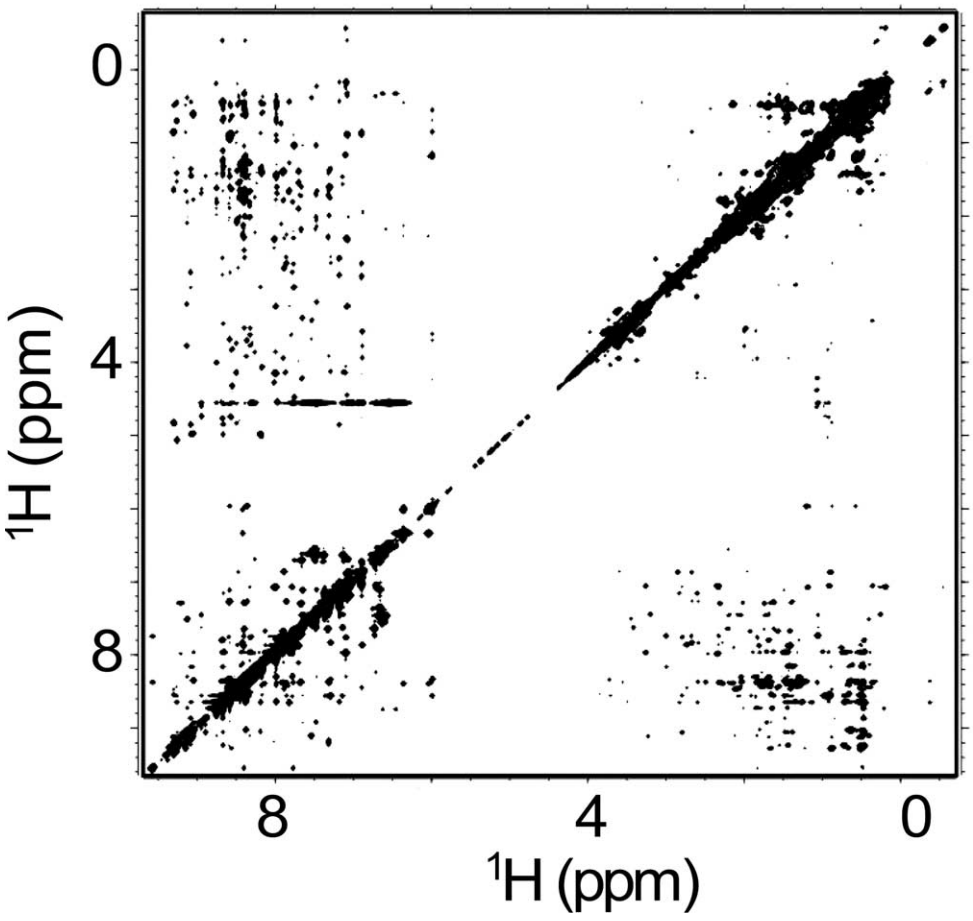

Fig. 6. Homonuclear ^{1}H–^{1}H NOESY spectrum of 2 mM ubiquitin labeled using the REDPRO scheme with a mixing time of 500 ms. The sample was in 30 mM sodium acetate pH 4.6, 10% D_2O, 0.01%. The spectrum is asymmetric because of the variable intensity of labeling, and hence observable projected density in the direct dimension. So, for example, $^{1}H'\{^{1}H^{\alpha}\}$ NOEs are observed although the $^{1}H^{\alpha}$ density is about 5%, while $^{1}H^{\alpha}\{^{1}H'\}$ are not observed.

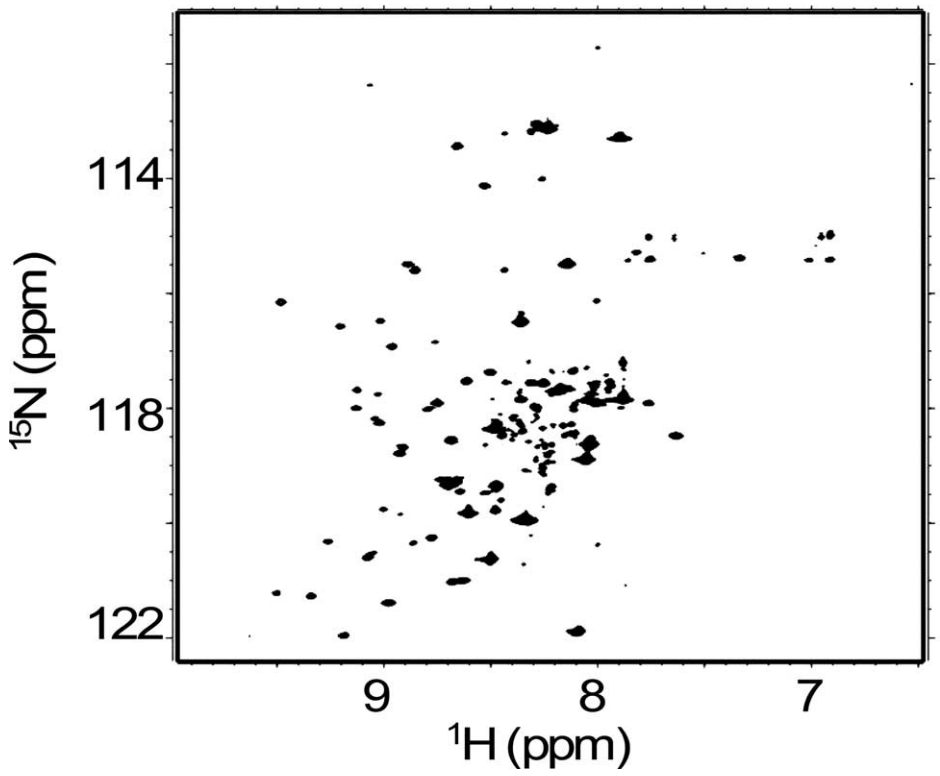

Fig. 7. $^{1}H\{^{15}N\}$ TROSY-HSQC spectrum of 1 mM GroES (MW of the heptamer is 70 kDa) labeled using the REDPRO scheme. The sample of GroES was in 50 mM phosphate buffer, pH 6.5, 100 mM NaCl, 10% D_2O, 0.01% NaN_3, 2 mM EDTA. The data were acquired on a Varian spectrometer operating at 900 MHz at 30°C. 256 t_1 increments were collected with 16 acquisitions per increment.

the TALOS database [17]. These structures revealed an RMS deviation of 1.6 Å for backbone atoms in the regions with definite secondary structure (Table 1). The addition of methyl–methyl and aromatic–aliphatic NOE constraints decreased the RMSD value to 0.6 Å. Addition of dipolar coupling constrains can further improve the quality of the structure and the labeling pattern favors high sensitivity measurement of one bond residual dipolar coupling of backbone and side-chains nuclei. A homonuclear ^{1}H–^{1}H NOESY spectrum of ubiquitin with a 500 ms mixing time reflecting the NOE pattern expected is shown in Fig. 6.

One of the most powerful features of NMR is the ability to generate interaction maps of molecules based on the chemical shift perturbations. Previously used observations of chemical shift changes were focused on observation of the backbone ^{15}N, $^{1}H'$, and $^{13}C'$ nuclei. Though powerful, this approach suffers from the relatively distant position of backbone nuclei from possible interaction sites. This may especially be the case of hydrophobic or predominantly charged interactions. The REDPRO scheme allows us to use the exquisite sensitivity of chemical shift perturbation mapping to study interaction surfaces formed by side chains with the potential of obtaining additional information on the inter and or intra-molecular interactions. An application of this methodology is reported elsewhere [11].

The availability of the isolated proton spins in the highly deuterated environment opens new possibilities for the dynamic characterization of the proteins. Studies to date have mostly focused on backbone dynamics, largely through measurements of backbone ^{15}N relaxation properties in uniformly ^{15}N labeled proteins due to of the predominant 'simple' dipolar relaxation of $^{15}N,H'$ pairs. All relaxation experiments developed for backbone $^{15}N,H'$ pair can be potentially utilized to study side chain dynamics using isolated $^{13}C,^{1}H$ pairs prepared by the scheme illustrated here. The possible deleterious effects of the carbon–carbon coupling could be circumvented using partially enriched ^{13}C glucose.

TROSY based triple resonance techniques which permit the study of proteins with molecular weights exceeding 20 kDa also benefit from high-level deuteration obtained using REDPRO due to greatly reduced ^{1}H homonuclear dipolar interaction [20]. REDPRO appears to have a sufficient spin isolation of $^{1}H's$, presumably because of the low ^{1}H density at H^{α} sites. The TROSY-HSQC of GroES (MW of the heptamer is 70 kDa) at 900 MHz (Fig. 7) demonstrates significant signal, while an HSQC under the same conditions gave no observable signals (data not shown). Dramatically improved relaxation properties of the heteronuclear ^{13}C magnetization transfer coupled with TROSY advantages presents proposed labeling scheme as an attractive approach for study of high molecular weight protein targets, their ligand sites, and interdomain interactions. The REDPRO scheme has also possible applications in recent solid-state NMR techniques which utilize inverse detection [21,22].

In conclusion, we demonstrated that the described REDPRO protein labeling technique allows us to create in a cost and time efficient way ^{2}H, ^{13}C, and ^{15}N NMR samples which contain enough protons for structural characterization and investigation of proteins properties and protein interactions in solution.

Acknowledgements: We acknowledge valuable discussions with David

LeMaster. We would like to thank Varian, Inc. and Oxford Instruments for allocating NMR spectrometer time and Eriks Kupce for collecting the 900 MHz data. This work was supported by NIH Grant GM-47021 and 59908 and NCI Fellowship F037244.

References

[1] Cavanaugh, J., Fairbrother, W.J., Palmer III, A.J. and Skelton, N.J. (1996) Protein NMR Spectroscopy: Principles and Practice, Academic Press, San Diego, CA.
[2] Garrett, D.S., Seok, Y.J., Liao, D.I., Peterkofsky, A., Gronenborn, A.M. and Clore, G.M. (1997) Biochemistry 36, 2517–2530.
[3] Clore, G.M. and Gronenborn, A.M. (1998) Trends Biotechnol. 16, 22–34.
[4] Gardner, K.H. and Kay, L.E. (1998) Annu. Rev. Biophys. Biomol. Struct. 27, 357–406.
[5] LeMaster, D.M. (1990) Q. Rev. Biophys. 23, 133–174.
[6] 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.
[7] Yamazaki, T., Tochio, H., Furui, J., Aimoto, S. and Kyogoku, Y. (1997) J. Am. Chem. Soc. 119, 872–880.
[8] Leiting, B., Marsilio, F. and O'Connell, J.F. (1998) Anal. Biochem. 265, 351–355.
[9] You, J., Cohen, R.E. and Pickart, C.M. (1999) BioTechniques 27, 950–954.
[10] Gosser, Y.Q., Zheng, J., Overduin, M., Mayer, B.J. and Cowburn, D. (1995) Structure 3, 1075–1086.
[11] Camarero, J.A., Shekhtman, A., Campbell, E., Chlenov, M., Gruber, T.M., Bryant, D.A., Darst, S.A., Cowburn, D. and Muir, T.W. (2002) Proc. Natl. Acad. Sci. USA 99, 8536–8541.
[12] Clark, A.C., Ramanathan, R. and Frieden, C. (1998) Methods Enzymol. 290, 100–118.
[13] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) J. Biomol. NMR 6, 277–293.
[14] Bartels, C., Xia, T., Guntert, P., Billeter, M. and Wuthrich, K. (1995) J. Biomol. NMR 5, 1–10.
[15] Guentert, P., Mumenthaler, C. and Wuthrich, K. (1997) J. Mol. Biol. 273, 283–298.
[16] Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) J. Am. Chem. Soc. 120, 6836–6837.
[17] Cornilescu, G., Delaglio, F. and Bax, A. (1999) J. Biomol. NMR 13, 289–302.
[18] Corio, P.L. (1967) Structure of High Resolution NMR Spectra, Academic Press, New York.
[19] Venters, R.A., Farmer, B.T., Fierke 2nd, C.A. and Spicer, L.D. (1996) J. Mol. Biol. 264, 1101–1116.
[20] Pervushin, K. (2000) Q. Rev. Biophys. 33, 161–197.
[21] Ishii, Y., Yesinowski, J.P. and Tycko, R.J. (2001) J. Am. Chem. Soc. 123, 2921–2922.
[22] Schmidt-Rohr, K., Saalwachter, K., Shu-Feng, L. and Hong, M. (2001) J. Am. Chem. Soc. 123, 7168–7169.