JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Subscriber access provided by American Chemical Society

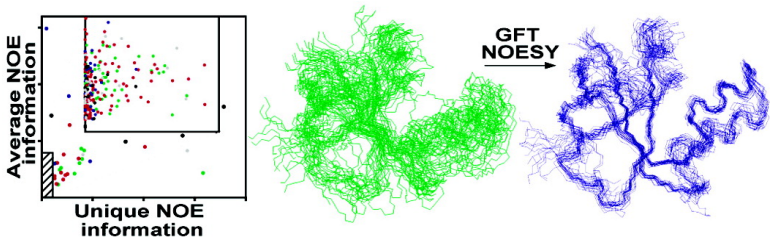
Article

G-Matrix Fourier Transform NOESY-Based Protocol for High-Quality Protein Structure Determination

Yang Shen, Hanudatta S. Atreya, Gaohua Liu, and Thomas Szyperski

J. Am. Chem. Soc., **2005**, 127 (25), 9085-9099• DOI: 10.1021/ja0501870 • Publication Date (Web): 07 June 2005

Downloaded from http://pubs.acs.org on March 25, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 6 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

ACS Publications
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

J|A|C|S
ARTICLES
Published on Web 06/07/2005

G-Matrix Fourier Transform NOESY-Based Protocol for High-Quality Protein Structure Determination

Yang Shen, Hanudatta S. Atreya, Gaohua Liu, and Thomas Szyperski*

Contribution from the Departments of Chemistry and Structural Biology, The State University of New York at Buffalo, Buffalo, New York 14260

Received January 11, 2005; E-mail: szypersk@chem.buffalo.edu

Abstract: A protocol for high-quality structure determination based on G-matrix Fourier transform (GFT) NMR is presented. Five through-bond chemical shift correlation experiments providing 4D and 5D spectral information at high digital resolution are performed for resonance assignment. These are combined with a newly implemented (4,3)D GFT NOESY experiment which encodes information of 4D ^{15}N/^{15}N-, ^{13}C^{alipahtic}/^{15}N-, and ^{13}C^{aliphatic}/^{13}C^{aliphatic}-resolved [^{1}H,^{1}H]-NOESY in two subspectra, each containing one component of chemical shift doublets arising from 4D → 3D projection at ω_1:Ω(^{1}H) ± Ω(X) [X = ^{15}N,^{13}C^{aliphatic}]. The peaks located at the centers of the doublets are obtained from simultaneous 3D ^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}-resolved [^{1}H,^{1}H]-NOESY, wherein NOEs detected on aromatic protons are also obtained. The protocol was applied for determining a high-quality structure of the 14 kDa Northeast Structural Genomics consortium target protein, YqfB (PDB ID 1TE7). Through-bond correlation and NOESY spectra were acquired, respectively, in 16.9 and 39 h (30 h for shift doublets, 9 h for central peaks) on a 600 MHz spectrometer equipped with a cryogenic probe. The rapidly collected highly resolved 4D NOESY information allows one to assign the majority of NOEs directly from chemical shifts, which yields accurate initial structures "within" ~2 Å of the final structure. Information theoretical "QUEEN" analysis of initial distance limit constraint networks revealed that, in contrast to structure-based protocols, such NOE assignment is not biased toward identifying additional constraints that tend to be redundant with respect to the available constraint network. The protocol enables rapid NMR data collection for robust high-quality structure determination of proteins up to ~20−25 kDa in high-throughput.

Introduction

Efficient NMR-based protein structure determination[1] relies on measurement of nuclear Overhauser effects (NOEs), which yield ^{1}H−^{1}H upper distance limit constraints. The assignment of NOEs quite generally depends on having (nearly) complete resonance assignments.[1,2] However, due to degeneracy of chemical shifts, the NOE assignment remains a nontrivial task even when complete resonance assignments are available. Nowadays, two approaches are routinely used to solve this "NOE assignment problem". First, proteins are ^{15}N/^{13}C double labeled[3] so that NOEs can be measured in 3D ^{15}N- or ^{13}C-resolved [^{1}H,^{1}H]-NOE spectroscopy (NOESY).[2] Dispersing NOE signals in a third dimension, which encodes a ^{13}C or a ^{15}N shift, typically allows one to assign for medium-sized proteins ~15−25% of the NOEs directly based on chemical shift data (compared to only few percent in 2D [^{1}H,^{1}H]-NOESY[1,2]). Second, an *initial* structure is calculated which is used in conjunction with the chemical shifts to assign additional NOEs. Several such cycles of structure calculation and NOE assignment are usually performed iteratively until a refined structure is obtained.

Importantly, inaccuracies in the initial fold arising from incorrectly assigned NOEs may result in the mis-assignment of additional NOEs. Hence, proper convergence of the NMR structure determination depends on obtaining an appropriately accurate initial structure; that is, it is advantageous if the bundle of conformers representing the initial solution structure covers a conformational subspace which overlaps with that of the refined ensemble of conformers. This requirement constitutes a key challenge for reliable automated NOE assignment[4] and thus also for the development of a robust and scalable platform for high-throughput structure determination in structural genomics.[5] Several programs have been established to automatically obtain accurate initial folds.[4] Among those are AutoStructure[6] and CYANA,[7] both of which are widely used. Conceptually,

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(2) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Skelton, N. J. *Protein NMR Spectroscopy*; Academic Press: San Diego, CA, 1996.
(3) (a) Kainosho, M. *Nat. Struct. Biol.* **1997**, *4*, 858−861. (b) Acton, T. B. et al. *Methods Enzymol*. **2005**, *394*, 210−243.
(4) (a) Güntert, P. *Prog. NMR Spectrosc.* **2003**, *43*, 105−125. (b) Baran, M. C.; Huang, Y. J.; Moseley, H. N. B.; Montelione, G. T. *Chem. Rev.* **2004**, *104*, 3451−3555. (c) Huang, Y. J.; Moseley, H.; Baran, M. C.; Arrowsmith, C. H.; Powers, R.; Tejero, R.; Szyperski, T.; Montelione, G. T. *Methods Enzymol.* **2005**, *394*, 111−141.
(5) (a) Montelione, G. T.; Zheng, D.; Huang, Y.; Gunsalus, C.; Szyperski, T. *Nat. Struct. Biol.* **2000**, *7*, 982−984. (b) Yee A. et al. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1825−1830.
(6) (a) Moseley, H. N. B.; Monleon, D.; Montelione, G. T. *Methods Enzymol.* **2001**, *339*, 91−108. (b) Huang, Y. J.; Swapna, G. V.; Rajan, P. K.; Ke, H.; Xia, B.; Shukla, K.; Inouye, M.; Montelione, G. T. *J. Mol. Biol.* **2003**, *327*, 521−536. (c) Huang, Y. J.; Powers, R.; Montelione, G. T. *J. Am. Chem. Soc.* **2005**, *127*, 1665−1674.
(7) (a) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298. (b) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209−227. (c) Güntert, P. *Methods Mol. Biol.* **2004**, *278*, 347−372.

AutoStructure mimics the approach an expert usually takes when solving a structure manually. The initial fold is generated based on (i) intraresidue, sequential, and medium-range NOEs considering NOE patterns of secondary structures, and (ii) unique long-range packing constraints. In contrast, CYANA relies on NOE network anchoring and a combination of (ambiguous) upper distance limit constraints. This led Montelione et al.[4b] to classify the AutoStructure and CYANA approaches as being "bottom-up" and "top-down", respectively. Since the two programs use distinctly different algorithms, their coupled operation aiming at a consensus NOE assignment promises to further increase the reliability of initial structure calculations.[8]

In the early 1990s, before the more sophisticated computational techniques[4] alluded to above were established, researchers devised 4D heteronuclear resolved [^{1}H,^{1}H]-NOESY and explored its impact for NMR structure determination of proteins.[9] Such 4D NOESY represents a straightforward and robust approach to tackle the "initial fold problem": dispersing signals in a fourth dimension enables one to assign the majority of NOEs directly based on chemical shift data. This may yield a highly accurate initial structure so that fast and reliable convergence of the structure determination can be accomplished. However, 4D NOESY suffers from two major drawbacks, which gave a competitive edge to computational methods in recent years. First, an additional heteronuclear polarization transfer needs to be inserted in the radio frequency (rf) pulse scheme. This leads to additional losses arising from transverse relaxation and tends to limit the use of 4D NOESY to small and medium-sized proteins. Second, conventional sampling of three indirect dimensions leads to long (minimal) measurement times. Typically, several days to a week are required to collect a single data set, even when accepting comparably short maximal evolution times (which limits the spectral resolution). This drawback is further exacerbated if a (minimal) radio frequency (rf) phase cycle is employed for artifact suppression.[2]

The first drawback of 4D NOESY, that is, its low sensitivity, has been significantly alleviated by the commercial introduction of cryogenic NMR probes,[10] which routinely deliver about 3-fold higher sensitivity compared to that of conventional probes.[11] Among the various options[12] to reduce the long minimal measurement times of heteronuclear NOESY, simultaneous ("time-shared") acquisition of ^{15}N- and ^{13}C-resolved NOESY,[13] extensive signal aliasing,[14] and the employment of the reduced-dimensionality (RD) approach[15] have been proposed.[16,17] Recently, G-matrix Fourier Transform (GFT) NMR,[19,20] a generalization of RD NMR,[15] has been devised to avoid data collection in the "sampling limited data collection regime",[15g] in which a large fraction of the instrument time is invested for sampling indirect dimensions and not for achieving a workable signal-to-noise (S/N) ratio. In GFT NMR, data acquisition is accelerated several fold by phase-sensitive joint sampling of two or more chemical shift evolution periods. This results in "chemical shift multiplets" in which additionally sampled chemical shifts are encoded in signal splittings. The G-matrix transformation enables one to edit the chemical shift multiplets so that each shift multiplet component, and thus each type of linear combination, is registered in a separate subspectrum. Detection of peaks defining the centers of the splittings[15e,20a] represents a straightforward concept[20,21] to retain the full information of the parent high-dimensional experiment. Provided that $m = K + 1$ chemical shift evolution periods of an ND experiment are jointly sampled in a single GFT dimension, central peak spectra up to Kth order[20] (recorded by omission of K shift evolution periods) are required. Then, a total of $p = 2^m - 1$ (N-K)D subspectra define a (N,N-K)D GFT NMR experiment.[20a] Alternatively, one can record the basic spectra[20a] twice (in which all $K + 1$ chemical shift evolution periods are sampled), each time with a (slightly) varied scaling factor for the K projected evolution periods.[12,15a] In this case, 2^m subspectra represent the (N,N-K)D GFT NMR experiment.[21] Furthermore, since chemical shifts are encoded in a redundant manner in the multiplets, multidimensional NMR spectral information can be obtained with increased precision.[19,20]

Here, we describe a through-bond GFT NMR-based resonance assignment protocol[19,20] used in conjunction with a newly implemented NOESY experiment combining simultaneous[13] and GFT[19,20] NMR data acquisition. The experiment is named "GFT (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH]", where the underlined letters denote nuclei for which the shifts are jointly sampled. The NOESY experiment encodes in two subspectra,

(8) For the present study, an evaluation using the program CYANA was best suited. This is because we assigned for YqfB all intraresidue, sequential, and medium-range NOEs by predicting NOESY peak lists from chemical shift data and considering information on secondary structure elements (see Materials and Methods). Hence, a top-down algorithm appeared to be the natural choice to complement our chemical shift-based assignment protocol.

(9) (a) Kay, L. E.; Clore, G. M.; Bax. A.; Gronenborn, A. M. *Science* **1990**, *249*, 411−414. (b) Clore, G. M.; Kay, L. E.; Bax, A.; Gronenborn, A. M. *Biochemistry* **1991**, *30*, 12−18. (c) Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Reizer, J.; Saier, M. H.; Wright, P. E. *Biochemistry* **1992**, *31*, 4413−4425. (d) Grzesiek, S.; Dobeli, H.; Gentz, R.; Garotta, G.; Labhardt, A. M.; Bax, A. *Biochemistry* **1992**, *31*, 8180−8190. (e) Archer, S. J.; Vionson, V. K.; Pollard, T. D.; Torchia, D. A. *Biochemistry* **1993**, *32*, 6680−6687. (f) Vuister, G. W.; Clore, G. M.; Gronenborn, A. M.; Powers, R.; Garrett, D. S.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1993**, *B101*, 210−213.

(10) Styles, P.; Soffe, N. F.; Scott, C. A.; Cragg, D. A.; White, D. J.; White, P. C. *J. Magn. Reson.* **1984**, *60*, 397−404.

(11) Monleon, D.; Colson, K.; Moseley, H. N. B.; Anklin, C.; Oswald, R.; Szyperski, T.; Montelione, G. T. *J. Struct. Funct. Genomics* **2002**, *2*, 93−101.

(12) Atreya, H. S.; Szyperski, T. *Methods Enzymol.* **2005**, *394*, 78−108.

(13) (a) Farmer, B. T.; Mueller, L. *J. Biomol. NMR* **1994**, *4*, 673−687. (b) Pascal, S. M.; Muhandiram, D. R.; Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Magn. Reson.* **1994**, *103*, 197−201. (c) Jerala, R.; Rule, G. S. *J. Magn. Reson.* **1995**, *B108*, 294−298. (d) Uhrin, D.; Bramham, J.; Winder, S. J.; Barlow, P. N. *J. Biomol. NMR* **2000**, *18*, 253−259. (e) Xia, Y.; Yee, A.; Arrowsmith, C. H.; Gao, X. *J. Biomol. NMR* **2003**, *27*, 193−203.

(14) Morshauser, R. C.; Zuiderweg, E. R. P. *J. Magn. Reson.* **1999**, *139*, 232−239.

(15) (a) Szyperski, T.; Wider, G.; Bushweller, J. H.; Wüthrich, K. *J. Am. Chem. Soc.* **1993**, *115*, 9307−9308. (b) Szyperski, T.; Pellecchia, M.; Wüthrich, K. *J. Magn. Reson.* **1994**, *B105*, 188−191. (c) Brutscher, B.; Simorre, J. P.; Caffrey, M. S.; Marion, D. *J. Magn. Reson.* **1994**, *B105*, 77−82. (d) Szyperski, T.; Braun, D.; Fernández, C.; Bartels, C.; Wüthrich, K. *J. Magn. Reson.* **1995**, *B108*, 197−203. (e) Szyperski, T.; Braun, D.; Fernández, C.; Wüthrich, K. *J. Am. Chem. Soc.* **1996**, *118*, 8146−8147. (f) Szyperski, T.; Banecki, B.; Braun, D.; Glaser, R. W. *J. Biomol. NMR* **1998**, *11*, 387−405. (g) Szyperski, T.; Yeh, D. C.; Sukumaran, D. K.; Moseley, H. N. B.; Montelione, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8009−8014.

(16) (a) Brutscher, B.; Morelle, N.; Cordier, F.; Marion, D. *J. Magn. Reson.* **1995**, *B109*, 397−404. (b) Kupce, E.; Freeman, R. *J. Magn. Reson.* **2004**, *172*, 330−333.

(17) Notably, rapid sampling techniques based on shortening of the relaxation delay between scans, such as longitudinal relaxation optimization,[18,19] are not well-suited for NOESY; it is desirable to keep ^{1}H steady-state magnetization close to its thermal equilibrium value in order to avoid an extensive modulation of NOE by T_1(^{1}H) relaxation.

(18) Pervushin, K.; Vogeli, B.; Eletsky, A. *J. Am. Chem. Soc.* **2002**, *124*, 12898−12902.

(19) Atreya, H. S.; Szyperski, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9642−9647.

(20) (a) Kim, S.; Szyperski, T. *J. Am. Chem. Soc.* **2003**, *125*, 1385−1393. (b) Kim, S.; Szyperski, T. *J. Biomol. NMR* **2004**, *28*, 117−130.

(21) For example, the projected chemical shift evolution periods can be scaled once with $\kappa = 1.0$ and once with $\kappa = 0.9$. The latter results in a compression of the chemical shift multiplets by 10% in the second data set and enables one to unambiguously identify all components belonging to the same chemical shift multiplet by comparison of the two data sets (see Figure 5 in ref 12).

the information of 4D $^{15}N/^{15}N$-, $^{13}C^{alipahtic}/^{15}N$-, and $^{13}C^{aliphatic}/^{13}C^{aliphatic}$-resolved [$^1H$,$^1H$]-NOESY. Each of the subspectra contains one component of a chemical shift doublet manifested along the GFT dimension at ω_1:$\Omega(^1H) \pm \Omega(X)$ [X = ^{15}N,$^{13}C^{aliphatic}$]. A third subspectrum, containing peaks located at the centers of the shift doublets at ω_1:$\Omega(^1H)$, encodes the information of 3D ^{15}N- and$^{13}C^{aliphatic}$-resolved [^1H,^1H]-NOESY. Preferably, central peaks are recorded in simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-resolved [$^1H$,$^1H$]-NOESY (named here for brevity 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]); this allows one to detect NOEs on aromatic protons along with the desired central peaks.

The impact of the simultaneous GFT NOESY data collection strategy has been explored here in detail since it is a priori not straightforward to identify sampling-limited[15g] NOESY data acquisition. Longer measurement times and the resulting increased S/N ratios lead to detection of additional NOEs corresponding to longer $^1H-^1H$ distances. This may have significant impact on the precision of the final NMR structure, and the sole analysis of S/N ratio distributions (for example, of intraresidue and sequential NOEs) is not sufficient. Hence, one has to assess for (4,3)D NOESY (i) the relative sensitivity of chemical shift doublet versus central peak detection, (ii) the increase in the fraction of central peak NOEs that can be assigned directly based on chemical shift data when having the additional information encoded in shift doublets, (iii) its impact on precision and accuracy of *initial* NMR structures obtained based on chemical shift data only, (iv) its value relative to computational techniques for generating *initial* structures from central peaks detected in 3D NOESY without reference to an initial structure, and (v) the number of additionally resolved and assigned NOEs allowing one to obtain a well-refined NMR structure. These criteria are central for establishing the role of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] for efficient NOE-based structure determination.

NMR experiments were performed on a 600 MHz spectrometer equipped with a cryogenic probe[10] for the 103-residue target protein YqfB of the Northeast Structural Genomics consortium (NESG; http://www.nesg.org; NESG ID: ET99). Protein YqgB expressed[5b] for the NMR study contained a 22-residue N-terminal tag to facilitate purification resulting in a molecular weight of 14.5 kDa (15.3 kDa with $^{15}N/^{13}C$ double labeling). To ensure that results are representative for medium-sized proteins, NMR experiments were conducted with a 1 mM protein solution at an ambient temperature of 25 °C.

Materials and Methods

NMR Sample Characterization. Uniformly (U)-^{13}C,^{15}N-labeled YqfB was produced as described previously[5b] and provided by Drs. Arrowsmith and Yee, University of Toronto. The protein contained a 22-residue N-terminal tag with sequence MGTSH6SSGRENLYFQGH in order to facilitate purification. Thus, the polypeptide expressed for NMR structural studies had a molecular weight of 15.3 and 14.5 kDa, respectively, with and without $^{13}C/^{15}N$ double labeling. U-^{13}C,^{15}N YqfB was concentrated to a ~1 mM solution in 90% H_2O/10% ^2H_2O (25 mM Na phosphate, pH = 6.5, 400 mM NaCl, 1 mM DTT, 20 mM $ZnCl_2$, 0.01% NaN_3). The approximate isotropic overall rotational correlation time, τ_r, of the protein was determined (at a ^1H resonance frequency of 600 MHz) from ^{15}N $T_1/T_{1\rho}$ nuclear spin relaxation time ratios as described previously.[15g] In agreement with a molecular weight of 15.5 kDa for U-^{13}C,^{15}N YqfB, $\tau_r \sim 7.7$ ns was obtained. This demonstrates that the protein is monomeric in solution.

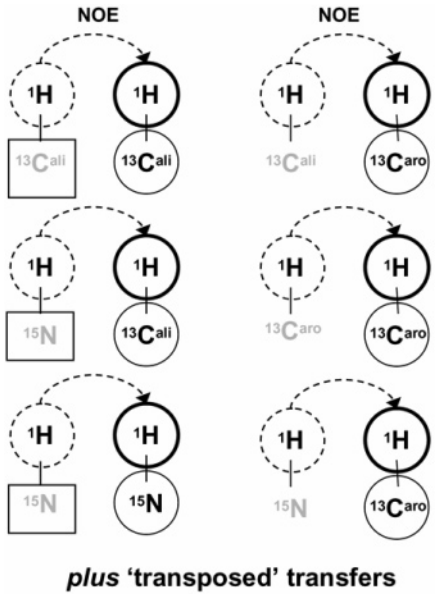


Figure 1. NOEs and chemical shifts are indicated which are measured in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH], provided that central peaks are detected in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]. The proton from which the NOE originates and the proton on which the signal is detected are shown, respectively, in dashed and bold circles. For the heteronuclei (depicted in black), which are attached to the detected proton, the chemical shift is measured in both (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] and 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]. The heteronuclei attached to the originating proton are shown in gray. The boxes around ^{15}N and $^{13}C^{aliphatic}$ indicate that these shifts are measured in shift doublet subspectra of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] (see text), thereby providing 4D NOESY information. For clarity, transposed NOEs are not indicated.

Implementation of GFT (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH]. The chemical shift measurements and NOE transfers taking place in simultaneous GFT (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] are shown in Figure 1. Chemical shift doublets are registered in the GFT dimension at ω_1:$\Omega(^1H) \pm \kappa\cdot\Omega(X)$ [X = ^{15}N,$^{13}C^{aliphatic}$], where κ represents a factor for scaling the projected chemical shift evolution.[15a] Proton polarization is transferred to the attached heteronucleus, frequency labeled with $\Omega(X)$, transferred back to the proton and frequency labeled with $\Omega(^1H)$. The latter shift is detected in quadrature.[2] Hence, with $\kappa = 0.5$, one has that G-matrix transformation yields two subspectra with peaks located either at $\Omega(^1H) + 0.5\cdot\Omega(X)$ or at $\Omega(^1H)$-$0.5\cdot\Omega(X)$ [i.e., the peak separation matches $\Omega(X)$]. After NOE mixing, the polarization is transferred to the second heteronucleus for frequency labeling and is then transferred back to the attached proton for signal detection. As a result, polarization transfers and chemical shift measurements of 4D $^{15}N/^{15}N$-, $^{13}C^{alipahtic}/^{15}N$-, and $^{13}C^{aliphatic}/^{13}C^{aliphatic}$-resolved [$^1H$,$^1H$]-NOESY occur simultaneously. For detection of central peaks located at ω_1:$\Omega(^1H)$, frequency labeling on the first heteronucleus is not required, and hence, the additional simultaneous heteronuclear polarization transfer is omitted (see below). It is advantageous to detect central peaks in conjunction with NOEs on aromatic protons. In this experiment, the polarization transfers of 3D ^{15}N-, $^{13}C^{aliphatic}$-, and $^{13}C^{aromatic}$-resolved [^1H,^1H]-NOESY occur simultaneously. For uniformity of nomenclature, we name this experiment as "3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]", wherein chemical shifts are measured for nuclei indicated in brackets.

The radio frequency (rf) pulse scheme for the detection of shift doublets (Figure 2a) was derived from simultaneous 3D ^{15}N,$^{13}C^{aliphatic}$-resolved [^1H,^1H]-NOESY[13e] by inserting an additional simultaneous [^1H,^{13}C]/[^1H,^{15}N]-HSQC module[2] before NOE mixing. To minimize signal losses arising from $T_2(^1H)$ relaxation, ^1H chemical shift evolution is implemented in a semiconstant time manner[2] during the reverse INEPT.[2] To maximize intensity of $^1H^{aliphatic}-^1H^{aliphatic}$ NOEs, INEPT

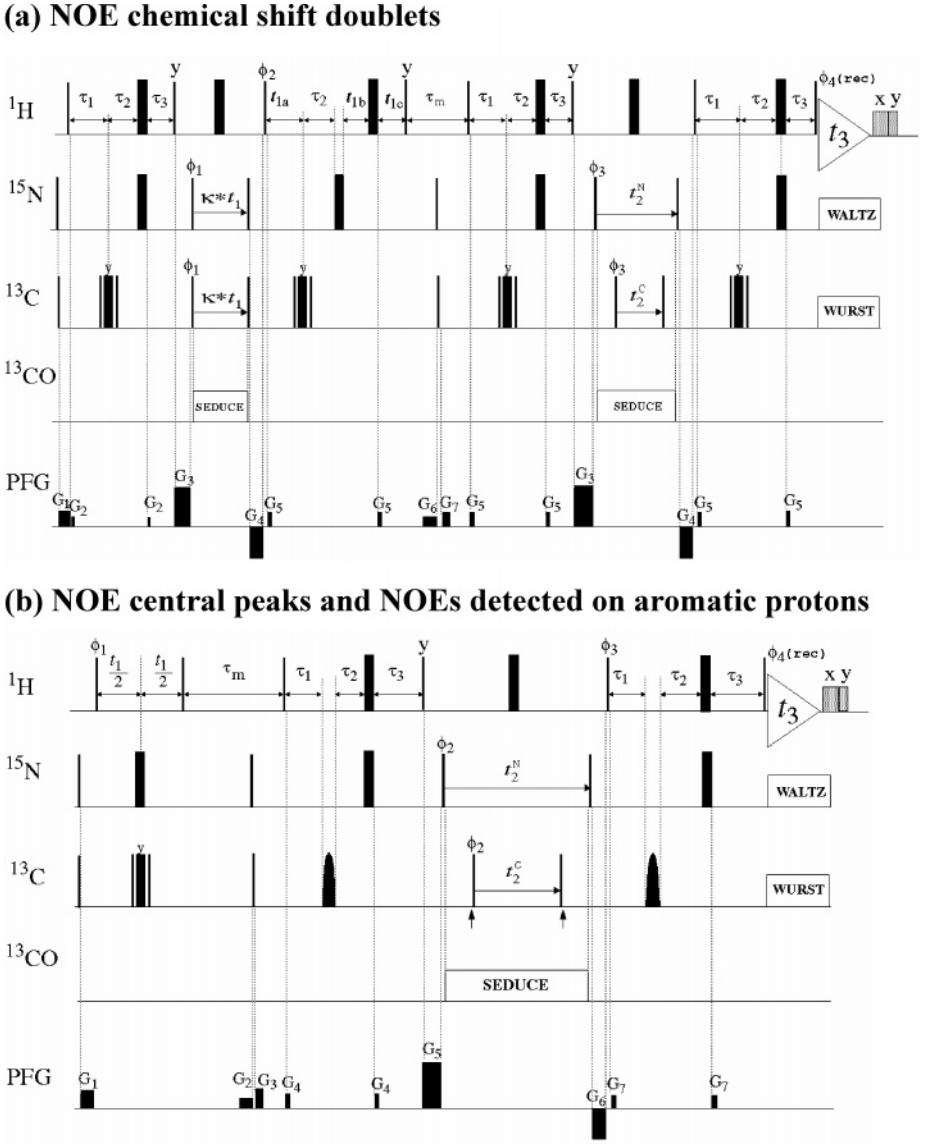


Figure 2. Radio frequency pulse schemes employed for acquisition of NOESY data. (a) Detection of shift doublets in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] and (b) 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]. Both schemes are derived from simultaneous 3D ^{15}N/^{13}C^{aliphatic}-resolved [^{1}H,^{1}H]-NOESY.[13e] Rectangular 90 and 180° pulses are indicated by thin and thick vertical bars, respectively, and phases are indicated above the pulses. Where no rf phase is marked, the pulse is applied along *x*. High-power 90° pulse lengths are: 9.0 μs for ^{1}H, 17.0 μs for ^{13}C, and 40 μs for ^{15}N. WURST[2] is used for decoupling of ^{13}C during acquisition. WALTZ16[2] is employed to decouple ^{15}N (rf = 1.70 kHz) during acquisition. ^{13}C decoupling during indirect ^{1}H chemical shift evolution is achieved using a ($90°_x-180°_y-90°_x$) composite pulse.[2] SEDUCE[2] is used for decoupling of ^{13}C, during t_1(^{1}H/^{13}C/^{15}N) and t_2(^{13}C/^{15}N) (rf field strength = 1 kHz). The duration of ^{1}H spin-lock purge pulses applied immediately after acquisition to improve suppression of the water line[13e] are: SL$_x$, 4.9 ms; SL$_y$: 2.6 ms. The ^{1}H rf and ^{15}N carrier positions are set to 4.78 and 118 ppm, respectively. For t_2(^{13}C/^{15}N), sampling starts at 1/(2·SW(^{13}C/^{15}N)) to ensure 180° first-order phase correction.[2] Parameters specific for the individual rf pulse schemes: For (a) the scaling factor was set to $\kappa = 0.5$ (see text), and the ^{13}C carrier position is set to 36 ppm. ^{1}H−^{13}C INEPT[2] rely on ($90°_x-180°_y-90°_x$) composite pulses,[2] and the ^{1}H−^{13}C INEPT delays are set to 3.9 ms (corresponding to a $^1J_{CHaliphatic} \sim 130$ Hz). ^{1}H frequency labeling (at a ^{1}H resonance frequency of 600 MHz) is achieved in a semiconstant time fashion[2] with t_1^a (0) = 1.9 ms, t_1^b (0) = 1.0 μs, t_1^c (0) = 2.5 ms, $\Delta t_1^a = 40$ μs, $\Delta t_1^b = 35$ μs, $\Delta t_1^c = -15$ μs. Hence, the fractional increase of the semiconstant time period with t_1 equals $\lambda = 1 + \Delta t_1^c/\Delta t_1^a = 0.63$. The duration and strengths of the rectangular pulsed *z*-field gradients (PFGs) are: G_1 (3 ms, 8 G/cm); G_2 (0.5 ms, 6 G/cm); G_3 (4 ms, 30 G/cm); G_4 (3 ms, −20 G/cm); G_5 (0.5 ms, 8 G/cm); G_6 (3.0 ms, 4 G/cm); G_7 (1.0 ms, 8 G/cm). The delays are: $\tau_1 = 2.0$ ms, $\tau_2 = 0.4$ ms, $\tau_3 = 2.4$ ms. The mixing time was set to $\tau_m = 70$ ms. Phase cycling: $\phi_1 = x, -x$; $\phi_2 = x$; $\phi_3 = 2(x), 2(-x)$; ϕ_4(receiver) = $x, -x, -x, x$. Quadrature detection in t_1(^{1}H/^{13}C/^{15}N) and t_2(^{13}C/^{15}N) is accomplished by altering the phases ϕ_2 and ϕ_3, respectively, according to States-TPPI.[2] GFT NMR phase cycle for recording the two basic spectra: $\phi_1 = x, y$. For (b): simultaneous inversion of ^{13}C^{aliphatic} and ^{13}C^{aromatic} spins during ^{1}H−^{13}C INEPT[2] is achieved using adiabatic 180° ^{13}C rf pulses of 1.0 ms duration,[2] and the INEPT delay is tuned to a compromise value of $1/2^1J_{CH} = 3.4$ ms (corresponding to $^1J_{CHaliphatic} < {}^1J_{CH} = 150$ Hz $< {}^1J_{CHaromatic}$) to allow simultaneous detection of NOEs on aliphatic and aromatic protons. The ^{13}C carrier position is initially set to 70 ppm and shifted to 36 ppm at the time point indicated by the first arrow and shifted back to 70 ppm at the time point indicated by the second arrow. Bloch simulations show that the loss of sensitivity for the aromatic signals that arise from off-resonance effects of the two rectangular 90° pulses applied at a ^{13}C carrier position of 36 ppm is less than ∼30% at 600 MHz ^{1}H resonance frequency. [At higher field strengths, it is advantageous to minimize off-resonance effects by placing the ^{13}C carrier between aliphatic and aromatic resonances (at ∼70 ppm) and employing time proportional phase incrementation[2,15d] to shift the ^{13}C carrier position to 36 ppm. At 900 MHz, for example, sensitivity losses at the edges of the ^{13}C spectral range, that is, for methyl groups and downfield aromatic resonances, are then limited to ∼25%]. The spectral widths along ω_2 are: ω_2(^{15}N) = 1680 Hz, ω_2(^{13}C^{aliphatic},^{13}C^{aromatic}) = 4300 Hz. This ensures that NOEs detected on aromatic protons are aliased along ω_2(^{13}C^{aliphatic},^{13}C^{aromatic}) and, provided that a 180° first-order phase correction is applied, have opposite sign than NOEs detected on amide protons to facilitate assignment. Moreover, mutual cancellation of these two types of NOEs is avoided. The duration and strengths of the rectangular pulsed *z*-field gradients (PFGs) are: G_1 (3 ms, 6 G/cm); G_2 (3 ms, 4 G/cm); G_3 (1 ms, 8 G/cm); G_4 (0.5 ms, 8 G/cm); G_5 (4 ms, 30 G/cm); G_6 (3 ms, −20 G/cm); G_7 (0.5 ms, 6 G/cm). The delays are: $\tau_1 = 1.2$ ms, $\tau_2 = 0.2$ ms, $\tau_3 = 2.4$ ms. The mixing time was set to $\tau_m = 70$ ms. Phase cycling: $\phi_1 = 2(x), 2(-x)$; $\phi_2 = x, -x$; $\phi_3 = x$; ϕ_4(receiver) = $x, -x, -x, x$. Quadrature detection in t_1(^{1}H) and t_2(^{13}C/^{15}N) is accomplished by altering the phases ϕ_1 and ϕ_2, respectively, according to States-TPPI.[2]

Table 1. Through-Bond GFT NMR Experiments for Resonance Assignment of Protein YqfB

Experiment	*Polarization transfer pathway*[a]	*Linear combinations of chemical shifts observed along* ω_1	*Minimal measurement time (hrs)*[b] GFT	non-GFT
(4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$	$^1HN_i \rightarrow {}^{15}N_i \rightarrow {}^{13}C^{\alpha}_{i/i-1} \rightarrow {}^{13}C^{\alpha\beta}_{i/i-1} \rightarrow {}^{13}C^{\alpha}_{i/i-1} \rightarrow {}^{15}N_i \rightarrow {}^1HN_i$ (t_1) (t_1) (t_2) (t_3)	$\Omega(^{13}C^{\alpha}_{i/i-1}) \pm \Omega(^{13}C^{\alpha}_{i/i-1})$ $\Omega(^{13}C^{\alpha}_{i/i-1}) \pm \Omega(^{13}C^{\beta}_{i/i-1})$	2.5	38
(4,3)D $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN	$^1H^{\alpha\beta}_{i-1} \rightarrow {}^{13}C^{\alpha\beta}_{i-1} \rightarrow {}^{13}C^{\alpha}_{i-1} \rightarrow {}^{13}C'_{i-1} \rightarrow {}^{15}N_i \rightarrow {}^1HN_i$ (t_1) (t_1) (t_2) (t_3)	$\Omega(^{13}C^{\alpha}_{i-1}) \pm \Omega(^{13}C^{\alpha}_{i-1})$ $\Omega(^{13}C^{\alpha}_{i-1}) \pm \Omega(^{13}C^{\beta}_{i-1})$	2.5	38
(5,2)D HACACONHN	$^1H^{\alpha}_{i-1} \rightarrow {}^{13}C^{\alpha}_{i-1} \rightarrow {}^{13}C'_{i-1} \rightarrow {}^{15}N_i \rightarrow {}^1HN_i$ (t_1) (t_1) (t_1) (t_2)	$\Omega(^{15}N_i) \pm \Omega(^{13}C'_{i-1})$ $\pm \Omega(^{13}C^{\alpha}_{i-1}) \pm \Omega(^1H^{\alpha}_{i-1})$	0.45[c]	102
(4,3)D HCCH aliphatic/aromatic	$^1H^{(1)} \rightarrow {}^{13}C^{(1)} \rightarrow {}^{13}C^{(2)} \rightarrow {}^1H^{(2)}$ (t_1) (t_1) (t_2) (t_3)	$\Omega(^{13}C^{(1)}) \pm \Omega(^1H^{(1)})$ $\Omega(^{13}C^{(2)}) \pm \Omega(^1H^{(2)})$	1.9/0.63[c]	31/2.2

[a] The $i-1$ and i indicate two neighboring amino acid residues along the polypeptide chain; t_1, t_2, and t_3 denote the evolution periods used for frequency labeling of the respective spins. [b] The minimal measurement times for GFT and parent (non-GFT) experiments were calculated by choosing the acquisition parameters of Table 2 and assuming that a single scan per free induction decay (FID) is recorded. Spectral widths: $^{13}C^{\alpha}$ 3000 Hz; $^{13}C^{\alpha\beta}$ 9000 Hz; $^{13}C^{aliphatic}$ 9000 Hz; $^{13}C^{aromatic}$ 4500 Hz; $^1H^{aromatic}$ 1500 Hz; $^1H^{aliphatic}$ 6000 Hz; $^1H^{\alpha}$ 1800 Hz; ^{15}N 1600 Hz; $^{13}C'$ 1400 Hz. Delay between start of FID acquisitions (recycle delays) 0.55 s for (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$and (4,3)D $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN; 0.95 s for (5,2)D HACACONHN, 0.39 s for (4,3)D HCCH (aliphatic), and 0.47 s for (4,3)D HCCH (aromatic). For experiments encoding $K \mp 1$ shifts in the GFT dimension, the spectral width along this dimension (ω_1) is^{20a} $\Sigma^K_{j=0}$ SW$_j$. The measurement times given include the time required for recording both basic and central peak spectra. Note that the third-order central peak spectrum of (5,2)D HACACONHN is equivalent to 2D [^{15}N,^1H] HSQC. [c] Assuming that central peak spectra are acquired by successive omission of indirect evolution periods.

delays are tuned to $1/2^1J_{CHaliphatic}$. Moreover, ^{13}C and ^{15}N chemical shift evolution in the GFT dimension is scaled[15a] down with $\kappa = 0.5$, limiting $t_{1,max}(^{13}C/^{15}N)$ to ~8 ms. This reduces the sensitivity loss arising from both transverse $T_2(^{13}C/^{15}N)$ relaxation and passive aliphatic one-bond $^{13}C-^{13}C$ scalar couplings (~35 Hz). Likewise, starting from the scheme of 3D $^{15}N,^{13}C^{aliphatic}$-resolved [$^1H,^1H$]-NOESY,[13e] 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] was implemented for acquisition of central peaks and NOEs detected on aromatic protons. For completeness, the rf pulse scheme is shown in Figure 2b.

NMR Data Acquisition and Processing. NMR spectra for resonance assignment (Table 1) and structure determination (Tables 1, 2) of YqfB were recorded at 25 °C on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic $^1H\{^{13}C,^{15}N\}$ triple resonance probe and processed using the program PROSA.[22] The cryogenic probe used for the present study delivers a signal-to-noise ratio of 4500:1 for the standard ethylbenzene sample (which is about 4-fold larger than what is measured for our conventional $^1H\{^{13}C,^{15}N\}$ probe). A detailed comparison pursued for 17 kDa NESG target protein YgdK dissolved in a 90% H_2O/10% ^2H_2O buffer (pH 6.5) containing 100 mM NaCl revealed that our cryogenic probe increases (on average) sensitivity by a factor of ~3 for double/triple resonance experiments and NOESY. Due to higher salt content (400 mM NaCl), the gain in sensitivity for the YqfB measurements was decreased to a factor of ~2. Hence, the NMR measurement times reported in this study (Table 2) would have been about 4 times longer if conducted with a conventional probe.

GFT NMR data for resonance assignment were acquired in 16.9 h (Table 2), while NOESY data were recorded with a total measurement time of 39 h; (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] shift doublets were acquired in 30 h (hereafter referred to as "data set I") using the pulse scheme of Figure 2a, and central peaks were obtained from 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] (Figure 2b) acquired in 9.1 h (Table 2). For comparison, a second data set ("data set II') comprising the shift doublets was acquired with twice the measurement time (60 h; Table 2).

Analysis of Through-Bond GFT NMR Correlation Spectra for Resonance Assignment. As a salient feature, GFT NMR affords linear combinations of shifts encoded in chemical shift multiplets (Table 1), and G-matrix transformation warrants editing of the multiplet components into different subspectra.[19,20] Hence, the number of peaks per subspectrum does not increase when increasing the number of jointly sampled indirect chemical shift evolution periods. Concomitantly, the peak dispersion (and thus spectral resolution) increases when compared with that of conventional congeners[2] of the same dimensionality.[19] Hence, the analysis of GFT NMR subspectra is quite generally less challenging than the analysis of the conventional congeners. In the present study, the program XEASY[23] was used for data analysis. Importantly, XEASY can handle linear combinations of shifts by expanding chemical shift lists accordingly. This option was originally introduced for analysis of RD NMR spectra[15,23] and allows one to assign linear combinations of shifts to peak positions. The chemical shifts are readily obtained from peak positions encoding the linear combina-

Table 2. Acquisition Parameters of NMR Experiments

Experiment	*Indirect dimension: t_{max} (ms); Complex points; Digital Resolution (Hz/Pt)*[a]	*Measurement time (hrs)*
(4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$	$\omega_1(^{13}C^{\alpha}, {}^{13}C^{\alpha\beta})$: 6.4; 78; 23 $\omega_2(^{15}N)$: 15.6; 26; 25	5.0
(4,3)D $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN	$\omega_1(^{13}C^{\alpha}, {}^{13}C^{\alpha\beta})$: 6.4; 78; 23 $\omega_2(^{15}N)$: 15.6; 26; 25	5.0
(5,2)D HACACONHN	$\omega_1(^{15}N;{}^{13}C', {}^{13}C^{\alpha}, {}^1H^{\alpha})$: 6.3; 48; 14	1.5[b]
(4,3)D HCCH (aliphatic)	$\omega_1(^{13}C;{}^1H)$: 6.3; 95; 29 $\omega_2(^{13}C)$: 3.6; 16; 70	4.0[c]
(4,3)D HCCH (aromatic)	$\omega_1(^{13}C;{}^1H)$: 4.0; 25; 23 $\omega_2(^{13}C)$: 3.8; 16; 62	1.4[d]
3D [H]-NOESY-[CH^{ali}/ CH^{aro}/NH]	$\omega_1(^1H)$: 16; 128; 15 $\omega_2(^{13}C)$: 6.4; 28; 67 $\omega_2(^{15}N)$: 17; 28; 25	9.1
Shift doublets of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/ NH] ('data sets I/II')	$\omega_1(^1H)$: 16; 200; 24.4 $\omega_1(^{13}C/^{15}N)$: 8; 200; 24.4 $\omega_2(^{13}C)$: 6.4; 28; 67 $\omega_2(^{15}N)$: 16.5; 28; 25	30/60[e]

[a] Direct dimension: $\omega_3(^1H)$ 64; 512; 8. All spectra were recorded with two scans per FID unless indicated differently in a separate footnote. For recycle delays, see footnote b of Table 1. [b] Includes 5 min to record a 2D [^{15}N,^1H]-HSQC, that is, third-order central peak, spectrum with $t_{max}(^{15}N) = 24$ ms and a single scan per FID. Other spectra were acquired with four scans per FID. [c] Includes 12 min to record a 2D constant time [^{13}C,^1H] HSQC[7] with $t_{max}(^{13}C) = 28$ ms. [d] Includes 10 min to record a 2D constant time [^{13}C,^1H] HSQC[7] with $t_{max}(^{13}C) = 18$ ms with four scans per FID. [e] Data sets I and II were recorded with two and four scans per FID, respectively. Hence, the minimal measurement time is 15 h.

(22) Güntert, P.; Dötsch, V.; Wider, G.; Wüthrich, K. *J. Biomol. NMR* **1992**, *2*, 619−629.

(23) Bartels, C.; Xia, T. H.; Billeter, M.; Güntert, P.; Wüthrich, K. *J. Biomol. NMR* **1995**, *6*, 1−10.

tions by use of a simple linear least-squares fit.[20a] Visual inspection of matching linear combinations of shifts, either within one spectrum or between several spectra, was accomplished for conventional spectra by use of XEASY. Peak picking was achieved by (i) generating peak lists from averaged shifts taken from the BioMagResBank or from shifts that were obtained at an earlier stage of the resonance assignment protocol, and (ii) manual adjustment of these lists.

Analysis of NOESY Data. NOESY spectra were analyzed using the program XEASY.[23] 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] comprises the central peaks of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] (Figure 1) and was analyzed as described in the following. First, on the basis of backbone and $^{13}C^{\beta}$ chemical shifts, the location of the regular secondary structure elements of YqfB was identified.[24,25] Subsequently, an initial 3D NOESY peak list was generated that contained the expected intraresidue, sequential, and medium-range NOE peak positions. After manual adjustment of peak positions and identification of other, primarily long-range NOE peaks, all peaks were integrated. To confirm assignments of overlapped peaks, line shapes were compared visually. The final 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] peak list, which yielded a high-quality structure of protein YqfB[24] (PDB ID 1TE7; see Results section), was used as a reference peak list to evaluate the impact of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH]. First, it was determined how many peaks in 3D NOESY can be assigned based on chemical shift data only [matching tolerances: 0.02 ppm for $\Omega(^1H)$; 0.2 ppm for $\Omega(^{15}N)$ and $\Omega(^{13}C)$], that is, without reference to an (initial) structure. Second, the spectra comprising the doublets were analyzed as described in the following. (i) The two subspectra containing the shift doublets were assigned in a bottom-up manner[20a] starting from the 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] peak list; for each peak in the 3D NOESY representing a central peak of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH], the corresponding shift doublet was identified, and the additionally encoded heteronuclear chemical shift was measured. It was then determined which peaks in 3D NOESY can be assigned unambiguously when having the additional, fourth chemical shift (with the matching tolerances indicated above). (ii) The subspectra containing the shift doublets were examined in order to identify new NOEs which could not be resolved and/or assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH], even with reference to a high-quality NMR structure. (iii) The S/N ratios of a large number of peaks were measured in order to assess the relative sensitivity of shift doublet compared to central peak detection. (iv) It was determined which fraction of the NOEs detected and assigned in 3D NOESY was likewise observed in the subspectra containing the shift doublets. This yielded a shift doublet detection yield relative to the (more sensitive) central peak detection.

NMR Structure Calculations. Cross-peak volumes measured in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] were converted into $^1H-^1H$ upper distance limit constraints by using the program DYANA.[7a] Cross-peaks overlapping in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH], but resolved in the shift doublet spectra of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH], were interpreted conservatively and used to derive 5 Å distance limit constraints. Residues of regular secondary structure elements were initially identified by use of the chemical shift index method,[25] and for polypeptide segments with three or more such identified consecutive residues, ϕ and ψ backbone dihedral angle constraints were derived from the chemical shifts by use of the program TALOS.[26] No hydrogen bond constraints were used. The final round of DYANA structure calculations employing torsion angle dynamics was started with 100 random conformers and 30 000 annealing steps, and the 20 structures with the lowest DYANA target functions were selected to represent the solution structure. Stereospecific assignments for calculating the refined reference structure were obtained using the FOUND and GLOMSA modules of DYANA.[7a] The Ramachandran map statistics of NMR structures were evaluated using the program PROCHECK.[27]

Automated NOE Assignment and Structure Calculation with CYANA. Calculations with the program CYANA[7b] were performed for automated NOE assignment and structure calculation using the standard protocol with seven cycles. Matching tolerances for chemical shifts were set to 0.02 ppm for $\Omega(^1H)$ and to 0.2 ppm for $\Omega(^{13}C)$ and $\Omega(^{15}N)$. CYANA structure calculations were started with 100 random conformers and 30 000 annealing steps. The 20 conformers with lowest target function value were selected for the next cycle of NOE assignment. In routine applications, automatically obtained NOE peak assignments are confirmed by visual inspection of spectra. To assess the impact of network anchoring[7b] in a first cycle of a CYANA calculation, the entire 3D NOESY reference peak lists were provided as input and combined with $^1H-^1H$ upper distance limit constraints derived from intraresidue, sequential, and medium-range NOEs (which were assigned as described in the paragraph above) and chemical shift derived[26] dihedral angle constraints. In addition, two CYANA calculations (referred to as 1 and 2 below) were performed to assess the impact of NOEs which could be assigned in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] using chemical shift data only. In these calculations, a peak list containing only the remaining unassigned long-range NOEs was provided as input. Specifically, calculations 1 and 2 were performed with the following input: (i) list of unassigned peaks of 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] reference peak list, (ii) TALOS dihedral angle constraint list which yielded the reference structure of YqfB (1TE7; see Results section), (iii) distance limit constraints representing intraresidue, sequential, and medium-range NOEs (see above), and (iv) distance constraints representing long-range NOEs that were assigned in the 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] reference peak list based on chemical shifts and (1) the shift doublets registered in the (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] subspectra of data set I (recorded in 30 h), or (2) the shift doublet subspectra of data set II (recorded in 60 h).

Evaluation of NOE Information Content with the Program QUEEN. The program QUEEN[28] affords quantitative evaluation of sets of NOEs using criteria originally developed for information theory. We have used this program to analyze the information content of *long-range* NOEs assigned in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] acquired with 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for central peak detection. For each constraint network, the total information content, I_{total} (see eq 8 in ref 28), and for each individual constraint within a given network, the unique information, I_{uni} (eq 10 in ref 28), and the average information, I_{ave} (eq 11 in ref 28), were calculated. The long-range upper distance constraint networks derived from the following NOE peak lists were subject to such an analysis: (1) 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] reference peak list, (2) list comprising peaks assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on shift data only, (3) list comprising peaks assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on shift data only but with reference to 4D information encoded in shift doublet data set I, and (4) same as (3) but with shift doublet data set II.

Results and Discussion

Efficient analysis of NOESY spectra requires resonance assignments.[1] Inspection of the 2D [^{15}N,^1H] HSQC spectrum (Figure 3) reveals that protein YqfB exhibits an overall good $^{15}N/^1H^N$ chemical shift dispersion. However, the central region of the spectrum is crowded, partly due to peaks arising from a 22-residue tag added to facilitate purification. Such $^{15}N/^1H^N$ chemical shift degeneracy poses a challenge for efficient

(24) The structural biology will be published elsewhere (Atreya, H. S.; Shen, Y.; Yee, A.; Arrowsmith, C. H.; Szyperski, T.).
(25) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1994**, *31*, 1647−1650.
(26) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289−302.
(27) Laskowski. R. A.; Rullmann, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477−486.
(28) Nabuurs, S. B.; Spronk, C. A. E. M.; Krieger, E.; Maassen, H.; Vriend, G.; Vuister, G. W. *J. Am. Chem. Soc.* **2003**, *125*, 12026−12034.

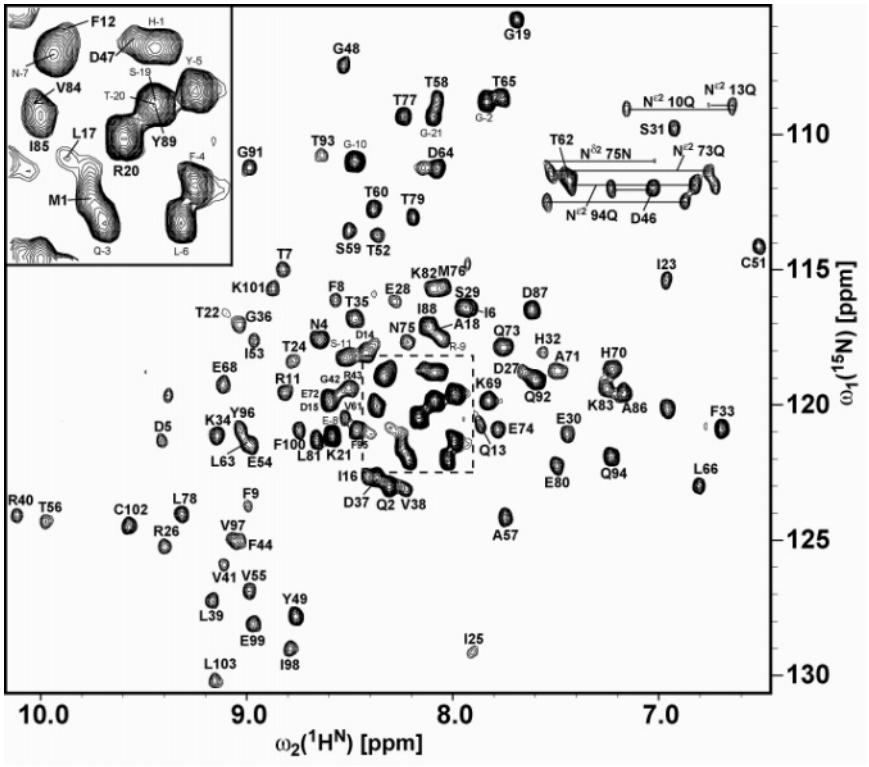


Figure 3. The 2D [^{1}H,^{15}N]-HSQC spectrum recorded for protein YqfB from *Escherichia coli*, a target of the Northeast Structural Genomics consortium. Peaks are labeled with their respective sequential resonance assignments using the one-letter code of amino acids and the amino acid sequence number. Negative numbers identify residues of the N-terminal tag (first residue of the tag is at position −22), and corresponding labels are printed with a smaller font size. For clarity, the assignments of the crowded central region of the spectrum are shown in an insert at the upper left corner. Resonance assignments (see text) were obtained with five through-bond GFT NMR experiments (Figure 4).

backbone resonance assignment and makes YqfB an attractive target for GFT NMR spectroscopy, which provides high dimensional spectral information[19,20] to break shift degeneracy.

Resonance Assignment. For obtaining (nearly) complete resonance assignments, a set of five through-bond GFT NMR experiments (Table 1) were performed in 16.9 h (Table 2): (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$ and $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN for backbone and $^{13}C^{\beta}$ assignment,[19] (5,2)D <u>HACACONHN</u>[20a] for $^1H^{\alpha}$ assignment, and aliphatic/aromatic (4,3)D <u>HC</u>CH, derived from RD 3D <u>HC</u>CH,[15g] for side chain spin system identification. Figure 4 illustrates the resulting GFT NMR-based resonance assignment strategy for protein YqfB. For backbone and $^{13}C^{\beta}$ resonance assignment, joint analysis of two subspectra of (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$ and $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN[19] proceeds conceptually as for 3D HNNCACB/CBCA(CO)NHN.[2] The analysis of (5,2)D <u>HACACONHN</u> described previously[20a] involves generating 15 2D peak lists from which >quintuples of chemical shifts are calculated. For side chain assignment, analysis of each of the three subspectra of (4,3)D <u>HC</u>CH proceeds conceptually as for conventional 3D H(C)CH.[2] First, (4,3)D <u>HC</u>CH and 3D H(C)CH are indistinguishable in both the indirect carbon and direct proton dimensions. Second, ω_1:$\Omega(^{13}C)+\Omega(^1H)$, ω_1:$\Omega(^{13}C)$-$\Omega(^1H)$ (chemical shift doublets) and ω_1:$\Omega(^{13}C)$ (central peaks) are detected along the GFT dimension in the three subspectra of (4,3)D <u>HC</u>CH (Table 1), while $\Omega(^1H)$ is detected along ω_1 in 3D H(C)CH. Hence, spin system identification in (4,3)D <u>HC</u>CH can be visualized as "walking" into the side chains by use of (i) $\Omega(^{13}C)$, (ii) $\Omega(^{13}C+^1H)$, and (iii) $\Omega(^{13}C-^1H)$.

Through-bond GFT NMR[19,20] provided the spectral information of one 5D and four 4D conventional FT NMR spectra with high digital resolution (Table 2; Figure 4), yielding nearly complete resonance assignments (98% of backbone and 95% of the side chain chemical shifts; BMRB ID 6207; Figure 3) with investment of 16.9 h of instrument time. Importantly, the availability of the highest dimensional spectral information ensured that data analysis is (i) robust with respect to occurrence of chemical shift degeneracies, (ii) highly reliable with respect to assignment yield and accuracy, and (iii) amenable to automated protocols due to increased peak dispersion and precision of shift measurements.[12,19,20] Analysis of the two backbone experiments, (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$/$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN, greatly profited from (i) the intraresidue $^{13}C^{\alpha}$—$^{13}C^{\beta}$ shift correlations and (ii) the doubled dispersion manifested, for example, by peaks encoding $2\cdot\Omega(^{13}C^{\alpha})$ instead of $\Omega(^{13}C^{\alpha})$ (see Figure S12 in ref 12). Moreover, it has been shown that 4D information encoded in RD 3D <u>HC</u>CH enables efficient assignment of the aliphatic side chains of proteins with molecular weights up to at least ~18 kDa.[29] This finding is confirmed here with the resonance assignment of YqfB using (4,3)D <u>HC</u>CH, the GFT NMR congener of RD 3D <u>HC</u>CH. Taken together, the through-bond GFT NMR[19,20]-based resonance assignment protocol provides a suitable basis for subsequent NOE assignment.

NOE Peak Assignment and Distance Constraints. A single 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] data set recorded in 9.1 h (Table 2; corresponding to ~1.5−2 days with a conventional probe) provided the information of all three 3D NOESY experiments routinely acquired for structure determination of

(29) Shen, Y.; Atreya, H. S.; Xiao, R.; Acton, T. B.; Shastry, R.; Ma, L.; Montelione, G. T.; Szyperski, T. *J. Biomol. NMR* **2004**, *29*, 549−550.

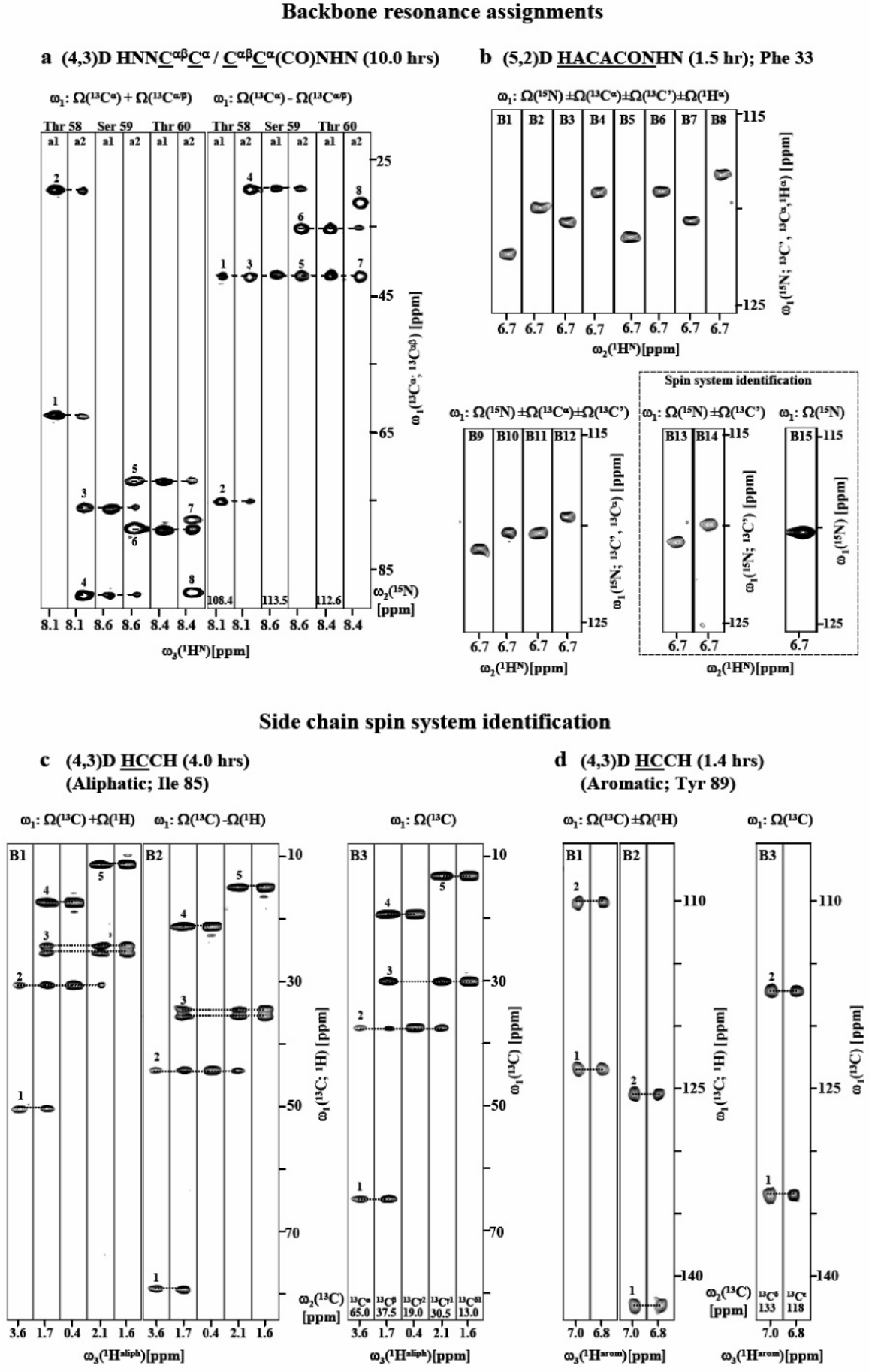

Backbone resonance assignments
a (4,3)D HNNCαβCα / CαβCα(CO)NHN (10.0 hrs)
b (5,2)D HACACONHN (1.5 hr); Phe 33
Spin system identification
Side chain spin system identification
c (4,3)D HCCH (4.0 hrs)
(Aliphatic; Ile 85)
d (4,3)D HCCH (1.4 hrs)
(Aromatic; Tyr 89)

Figure 4. Resonance assignment based on GFT NMR experiments (Tables 1 and 2) exemplified for protein YqfB. (a) [$\omega_1(^{13}C^{\alpha};^{13}C^{\alpha/\beta})$,$\omega_3(^1H^N)$]-strips taken from (4,3)D $C^{\alpha/\beta}C^{\alpha}$(CO)NHN (labeled with "a1") and (4,3)D HNN$C^{\alpha/\beta}C^{\alpha}$ (labeled with "a2"). The strips were taken at $\omega_2(^{15}N)$ (the ^{15}N chemical shifts are indicated at the bottom of the strips) of residues 57−60 (referred to as residue *i*), and are centered along $\omega_3(^1H^N)$ about their backbone $^1H^N$ shifts. Along $\omega_1(^{13}C^{\alpha};^{13}C^{\alpha/\beta})$, peaks are observed at $\Omega(^{13}C^{\alpha}) \pm \Omega(^{13}C^{\alpha/\beta})$ of residue $i-1$ in "a1" and of residue *i* in "a2" (in addition, peaks originating from residue $i-1$ are observed in "a2" if transfer via $^2J_{NC\alpha}$ is sufficiently effective). $\Omega(X)$ ($X = {}^{13}C^{\alpha}, {}^{13}C^{\beta}$) denotes the offset relative to the carrier position [during $t_1(^{13}C^{\alpha/\beta})$, the ^{13}C carrier frequency is placed at 43 ppm; during $t_1(^{13}C^{\alpha})$, $^{13}C^{\alpha}$ is detected in quadrature and the carrier frequency is placed at 56 ppm; see Figures S7 and S9 of ref 19]. The composite plot of strips *on the left* was taken from the GFT subspectrum comprising peaks at $\Omega(^{13}C^{\alpha}) + \Omega(^{13}C^{\alpha})$ (labeled as 1, 3, 5, 7) and $\Omega(^{13}C^{\alpha}) + \Omega(^{13}C^{\beta})$ (labeled as 2, 4, 6, 8), and the composite plot *on the right* was taken from the subspectrum comprising peaks at $\Omega(^{13}C^{\alpha}) - \Omega(^{13}C^{\alpha})$ (1, 3, 5, 7) and $\Omega(^{13}C^{\alpha}) - \Omega(^{13}C^{\beta})$ (2, 4, 6, 8) (the type of linear combination of chemical shifts is indicated above the composite plots). The combined use of (4,3)D $C^{\alpha/\beta}C^{\alpha}$(CO)NHN/HNN$C^{\alpha/\beta}C^{\alpha}$ yields three sequential walks along the polypeptide backbone which are indicated by dashed lines [note: peaks at $\Omega(^{13}C^{\alpha}) - \Omega(^{13}C^{\alpha})$ on the right are all located at the carrier position (43 ppm) and do not provide connectivities]. Peaks were sequentially assigned to the $^{13}C^{\alpha/\beta}$ shifts of Ala 57 (1, 2), Thr 58 (3, 4), Ser 59 (5, 6), and Thr 60 (7, 8) after $^{13}C'_{i-1}$,$^{15}N_i$,$^1H^N$ spin system identification was accomplished in the second- and third-order central peak spectra B13−B15 of (5,2)D HACACONHN,[20a] which represents (3,2)D HNNCO and thus encodes 3D HNNCO[2] spectral information. This is indicated in panel (b) which shows composite plots of strips taken from (5,2)D HACACONHN.[20a] This (5,2)D experiment is used to sequence specifically assign $^1H^{\alpha}$ and ^{13}C shifts after having obtained the sequence-specific assignments of $^{13}C^{\alpha/\beta}$, ^{15}N, and $^1H^N$ from (4,3)D $C^{\alpha/\beta}C^{\alpha}$(CO)NHN/HNN$C^{\alpha/\beta}C^{\alpha}$. The signals in panel (b) arise from magnetization transfer from $^1H^{\alpha}$ of His 32 to $^1H^N$ of Phe 33. The eight basic spectra (labeled B1−B8) encode the chemical shifts of $^1H^{\alpha}_{i-1}$, $^{13}C^{\alpha}_{i-1}$, $^{13}C_{i-1}$, and $^{15}N_i$ in a single GFT dimension (ω_1). Due to the particular choice for quadrature detection of ^{15}N and central peak detection,[20a] combination of chemical shifts is registered as $\Omega_0 \pm \Omega_1 \pm \Omega_2 \pm \Omega_3$ with $\Omega_o = \Omega(^{15}N)$, $\Omega_1 = \Omega(^{13}C)$, $\Omega_2 = \Omega(^{13}C^{\alpha})$ and $\Omega_3 = \Omega(^1H^{\alpha})$. Note that the part per million scale along ω_1 is defined by the type of nucleus detected in quadrature (^{15}N is this case). As a result, the jointly sampled chemical shifts of different types of nuclei (^1H and ^{13}C in this case) are scaled up by the ratio of the gyromagnetic ratios (for example, the ^1H chemical shifts in ppm are scaled up 10-times on the ^{15}N ppm scale). Specifically, the linear combinations observed in subspectra B1−B8 are: B1 $[\Omega_0+\Omega_1+\Omega_2+\Omega_3]$; B2 $[\Omega_0-\Omega_1+\Omega_2+\Omega_3]$; B3 $[\Omega_0+\Omega_1-\Omega_2+\Omega_3]$; B4 $[\Omega_0-\Omega_1-\Omega_2+\Omega_3]$; B5 $[\Omega_0+\Omega_1+\Omega_2-\Omega_3]$; B6 $[\Omega_0-\Omega_1+\Omega_2-\Omega_3]$; B7 $[\Omega_0+\Omega_1-\Omega_2-\Omega_3]$; B8 $[\Omega_0-\Omega_1-\Omega_2-\Omega_3]$. To resolve potential shift degeneracies, first-order central peak spectra are acquired (B9−B12) comprising peaks at: B9 $[\Omega_0+\Omega_1+\Omega_2]$; B10 $[\Omega_0-\Omega_1+\Omega_2]$; B11 $[\Omega_0+\Omega_1-\Omega_2]$; B12 $[\Omega_0-\Omega_1-\Omega_2]$, second-order central peaks (B13−B14) comprise peaks at B13 $[\Omega_0+\Omega_1]$ and B14 $[\Omega_0-\Omega_1]$, and third-order central peaks (B15) are signals of 2D [^{15}N,^1H]-HSQC at Ω_0. (c) Assignment of aliphatic side chains exemplified for Ile 85. On the left, two composite plots show strips taken from the basic subspectra providing $\Omega(^{13}C)+\Omega(^1H)$ (labeled B1) and $\Omega(^{13}C)-\Omega(^1H)$ (labeled B2) along the GFT dimension ω_1 (Table 1). On the right, a composite plot (labeled B3) shows strips taken from the central peak spectrum of (4,3)D HCCH, providing $\Omega(^{13}C)$ along ω_1 [that is, 3D (H)CCH information[2]]. For aliphatic spin system identification, sums and differences of shifts of covalently attached ^{13}C and ^1H nuclei (Table 1) are delineated in B1 and B2, while ^{13}C shifts are matched in B3 (indicated by dashed lines). Sequence-specific assignments are inferred from $^1H^{\alpha}$ and $^{13}C^{\alpha/\beta}$ shifts assigned in (4,3)D HNN$C^{\alpha/\beta}C^{\alpha}/C^{\alpha/\beta}C^{\alpha}$(CO)NHN[19] and (5,2)D HACACONHN[20a] as described in the legend for panels (a) and (b). The five $[\omega_1(^{13}C;^1H),\omega_3(^1H)]$ strips were taken along $\omega_2(^{13}C)$ at the shifts of $^{13}C^{\alpha}$,$^{13}C^{\beta}$,$^{13}C^{\gamma 1}$, $^{13}C^{\gamma 2}$, and $^{13}C^{\delta 1}$ of Ile 85 (indicated at the bottom of the strips of B3). The peaks encode along ω_1 linear combinations of the following shifts: $^{13}C^{\alpha}/^1H^{\alpha}$ (labeled as 1), $^{13}C^{\beta}/^1H^{\beta}$ (2), $^{13}C^{\gamma 12,\ \gamma 13}/^1H^{\gamma 12,\ \gamma 13}$ (3), $^{13}C^{\gamma 2}/^1H^{\gamma 2}$ (4), and $^{13}C^{\delta 1}/^1H^{\delta 1}$ (5). (d) Identification of aromatic spin systems in (4,3)D HCCH demonstrated for Tyr 89. On the left, two composite plots show strips taken from basic subspectra, providing $\Omega(^{13}C)+\Omega(^1H)$ (labeled B1) and $\Omega(^{13}C)-\Omega(^1H)$ (labeled B2). The composite plot shown on the right (labeled B3) shows strips taken from the central peak spectrum of (4,3)D HCCH, providing $\Omega(^{13}C)$ along ω_1. Aromatic spin system identification was accomplished in a manner similar to that described in the legend of panel (c) for aliphatic spin systems, and sequence-specific assignments were inferred from observation[1] of intraresidue NOEs. The two $[\omega_1(^{13}C;^1H),\omega_3(^1H)]$ strips from each of the three subspectra were taken along $\omega_2(^{13}C)$ at the shifts of $^{13}C^{\delta}$ and$^{13}C^{\epsilon}$ of Tyr 89 (indicated at the bottom of the strips of B3). The peaks encode along ω_1 linear combinations of the following shifts: $^{13}C^{\delta}/^1H^{\delta}$ (labeled as 1) and $^{13}C^{\epsilon}/^1H^{\epsilon}$ (2) (since $\Omega(^1H^{\delta}) = 7.2$ ppm and $\Omega(^1H^{\epsilon}) = 6.7$ ppm, the separation of peaks labeled with 1 in B1 and B2 appears to be similar to the separation of peaks labeled with 2).

$^{13}C/^{15}N$-labeled proteins. A total of 4708 NOEs were assigned (Figure 5) and yielded 1453 $^1H-^1H$ upper distance limit constraints (Table 3). Out of the 280 long-range constraints, only 53 (19%) involve aromatic protons. However, calculations with the program QUEEN[28] confirm[1,30] that these aromatic constraints are highly valuable for the three-dimensional structure determination: 8 (27%) out of the 30 most informative constraints with the largest I_{ave}/I_{total} [and 6 (40%) out of 15 most informative] involve aromatic rings (see also Figure 7). This finding emphasizes the importance of including aromatic protons into the simultaneous NOESY data acquisition (Figure 2b) to enable high-quality NMR structure determination.

YqfB Reference Structure Determination. A high-quality NMR structure of YqFB (PDB ID: 1TE7) was obtained with the distance constraints derived from 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] (Figure 5), as is evidenced (see middle column labeled PDB in Table 3) by (i) the small size and number of residual constraint violations, (ii) average rmsd values relative to the mean coordinates of 20 conformers of 0.65 ± 0.18 Å for the backbone and 1.19 ± 0.25 Å for all heavy atoms, (iii) a large fraction of stereospecific assignments for β-methylene and the Val and Leu isopropyl moieties, and (iv) the fact that all ϕ and ψ dihedral angles are located in the allowed (most favored, additionally, or generously allowed) regions of the Ramachandran map.

(30) Skalicky, J. J.; Mills, J. L.; Sharma, S.; Szyperski, T. *J. Am. Chem. Soc.* **2001**, *123*, 388−397.

Sensitivity of NOE Detection in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH]. Sensitivity of NOE detection is critical for identifying the optimal use of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH]. First, the relative intrinsic sensitivity of 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] and *separately* acquired 3D ^{15}N-, $^{13}C^{aliphatic}$-, and $^{13}C^{aromatic}$-resolved [^1H,^1H]-NOESY is evaluated. Complementing previous assessments,[13] differences in sensitivity arise because (i) a ^{15}N,^1H sensitivity enhancement scheme[2] has not been integrated in simultaneous [^{15}N,^1H]/[^{13}C,^1H]-HSQC detection modules, which reduces the sensitivity of the ^{15}N-resolved part to ~70% ($\sim 1/\sqrt{2}$) compared to a sensitivity enhanced ^{15}N-resolved [^1H,^1H]-NOESY, (ii) a compromise value is chosen for the $^{13}C-^1H$ INEPT delay (Figure 2) in order to enable simultaneous detection of NOEs on aliphatic and aromatic protons, which attenuates $^{13}C-^1H$ detected signals by ~5%, and (iii) off-resonance effects of 90° rf pulses on $^{13}C^{aromatic}$ leading to a sensitivity reduction to ~75%. Additional smaller losses arise because (i) the $^1H-^{15}N$ INEPT[2] delay (5.4 ms) is longer than the $^{13}C-^1H$ INEPT delay (3.4 ms), which increases the signal loss for the ^{13}C-resolved part due to $T_2(^1H)$ relaxation during an additional period of 4 ms, and (ii) of longer maximal evolution times in t_2 for ^{15}N frequency (~16 ms) than for $^{13}C^{aliphatic}/^{13}C^{aromatic}$ frequency labeling (~6 ms), which requires that longitudinal two-spin order, H_zC_z, is present prior to and after frequency ^{13}C labeling leading to some signal loss for $^{13}C-^1H$-detected signals due to $T_1(H_zC_z)$ relaxation. Taken together, sensitivity is reduced relative to separate data acquisi-

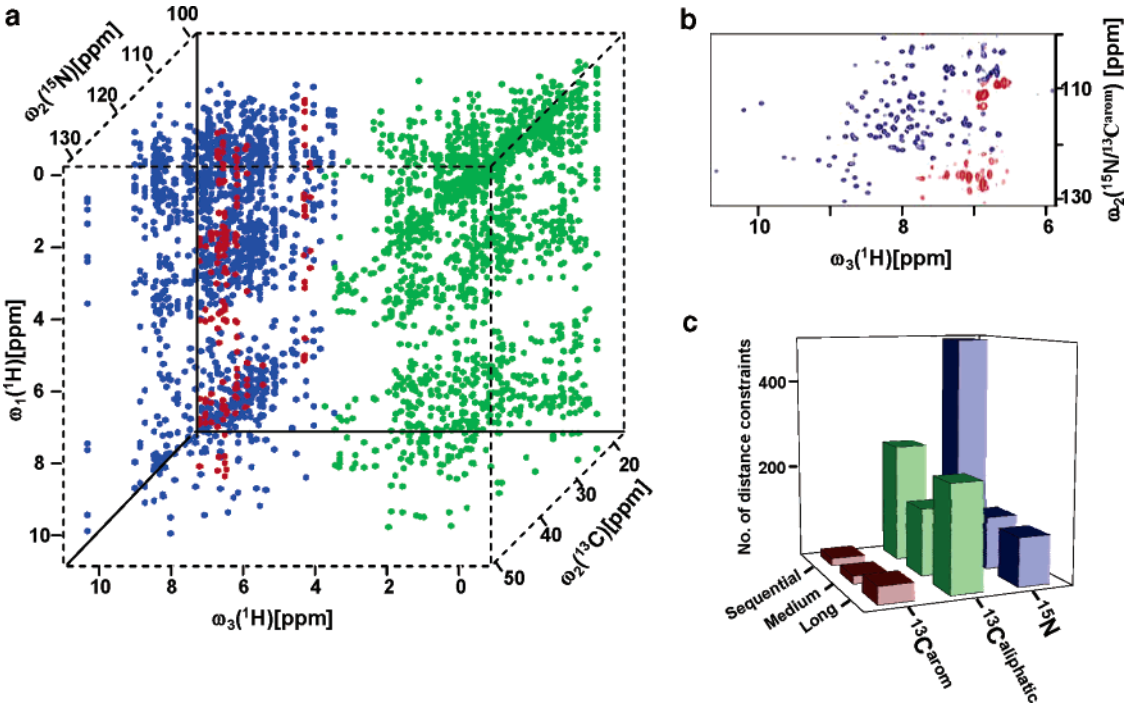


Figure 5. The 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for detection of (i) central peaks of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH], and (ii) NOEs on aromatic protons. (a) Cube with dots representing NOEs that were used for calculating the reference structure of YqfB[24] (PDB ID 1TE7; blue: detected on $^1H^N$; green: detected on $^1H^{aliphatic}$; red: detected on $^1H^{aromatic}$); 1373, 3094 and 241 NOEs (4708 in total, including transposed peaks) were detected on amide, aliphatic, and aromatic protons, respectively, and assigned as described in the Materials and Methods section. (b) Skyline projection along $\omega_1(^1H)$ showing that proper adjustment of spectral width and carrier position along ω_2 (Figure 2b) allows one to avoid mutual cancellation of NOEs detected on $^1H^N$ (blue) and $^1H^{aromatic}$ (red). (c) Bar plot providing a breakdown of sequential, medium-, and long-range NOEs (note that these numbers include transposed peaks): detected on amide protons: 494/114/104; detected on aliphatic protons: 257/145/228; detected on aromatic protons: 16/17/37.

tion to ~70% for the ^{15}N-resolved, to ~80% for the $^{13}C^{aliphatic}$-resolved, and to ~65% for the $^{13}C^{aromatic}$-resolved part. Requiring that the same S/N ratios are obtained in simultaneous 3D NOESY as in the separately acquired spectra, one obtains an effective acceleration of data acquisition speed by ~1.5.[32]

Next, the sensitivity of central peak detection in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] is compared with shift doublet detection in (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH]. Since the shift doublets arise from an in-phase splitting of the central peaks, the intrinsic sensitivity of detecting *each peak* of a shift doublet is a priori reduced to 50% when compared with central peak detection. However, the fact that only shift doublets represent viable signals allows one to identify peaks significantly closer to the noise level.[15a] This compensates, at least partly, for the 2-fold loss in sensitivity. The sensitivity of shift doublet detection is further reduced due to the transverse relaxation occurring during the additional simultaneous [^{15}N,^1H]/[^{13}C,^1H]-HSQC module (Figure 2). This loss depends on protein size and shape, which determine the correlation time(s) for the overall rotational tumbling, as well as internal mobility. For a rigid spherical protein, calculation of $^1H^N$, ^{15}N, $^1H^{aliphatic}$, and $^{13}C^{aliphatic}$ (non-methyl) transverse relaxation rates[2] is feasible. This allows one to estimate, with the transfer delays defined in Figure 2a, the reduction of sensitivity as a function of the correlation time for isotropic reorientation (Figure S1). For YqfB (τ_r ~ 7.7 ns), theory[2] predicts (Figure S1) that relaxation further reduces S/N ratios of shift doublet components to ~38 and ~28%, respectively, of the sensitivity of central peak detection after a splitting with $\Omega(^{15}N)$ and $\Omega(^{13}C)$ is encoded. Additional smaller losses can be expected to arise from (i) signal dephasing due to the presence of passive one-bond $^{13}C-^{13}C$, and one- and two-bond $^{15}N-^{13}C^{\alpha}$ scalar couplings, (ii) rf pulse inhomogenities, and (iii) rf pulse off-resonance effects. (Note that for $^{13}C^{aliphatic}$-resolved shift doublets, a slight gain in sensitivity relative to central peak detection in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] is due to the fact the $^{13}C-^1H$ INEPT delay is not tuned to a compromise value; see Figure 2.) Experimentally, we observed that the S/N of the peaks constituting the shift doublets is reduced to ~30 ± 5% for ^{15}N-resolved, and to ~29 ± 6% (non-methyl) and ~37 ± 4% (methyl) for ^{13}C-resolved central peaks. These values are in good agreement with the theoretical estimates (Figure S1), suggesting that these allow one to assess the role of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH] for larger systems.

Completeness of Central Peak versus Shift Doublet Detection. The 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] was acquired in 9.1 h (Table 2). Considering that the sensitivity for detecting each peak of a shift doublet is only ~30% of the sensitivity for central peak detection (see paragraph above; Figure S1), 10-fold longer measurement time (i.e., about 90 h) would be

(31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(32) (a) Following these sensitivity considerations, the impact of simultaneous NOESY acquisition on NMR data collection speed shall be illustrated. For example, one can assume that ^{15}N-resolved, $^{13}C^{aliphatic}$-resolved, and $^{13}C^{aromatic}$-resolved NOESY are acquired in 12 h each (yielding 36 h total measurement time). About 30−35% sensitivity is lost due to simultaneous acquisition for ^{15}N-resolved as well as for the $^{13}C^{aromatic}$-resolved part. To compensate for these losses, the measurement time needs to be doubled. As a result, data collection speed is effectively increased by a factor of 1.5 (24 instead of 36 h). Notably, use of a simultaneous [^{15}N,^1H]/[^{13}C,^1H]-HSQC detection module[32b] could increase the sensitivity of the ^{15}N-resolved part. (b) Sattler, M.; Maurer, M.; Schleucher, J.; Griesinger, C. *J. Biomol. NMR* **1995**, *5*, 97−102.

Table 3. Statistics of YqfB(1−103) Structure Calculations[a]

	(1) without SSA	(2) PDB	(3) refined
Stereospecific assignments (SSA)[b] [%]			
$^{\beta}CH_2$		34	34
Val and Leu isopropyl groups		58	58
Conformationally restricting distance constraints[c] between residues *i* and *j*			
intraresidue [$i = j$]	463	454	466
sequential [$i - j = 1$]	496	511	527
medium range [$1 < \|i - j\| \leq 5$]	177	208	211
long range [$\|i - j\| > 5$]	304	280	289
total	1440	1453	1493
Number of constraints per residue	15.3	15.4	15.8
Number of long-range constraints per residue	2.9	2.7	2.8
DYANA target function [Å^2]	1.73 ± 0.17	1.89 ± 0.16	1.72 ± 0.20
Average pairwise rmsd [Å] to the mean coordinates for residues 4−101			
backbone N, C^{α}, C′	0.97 ± 0.22	0.65 ± 0.18	0.69 ± 0.16
all heavy atoms	1.58 ± 0.24	1.19 ± 0.25	1.25 ± 0.21
backbone + best defined side chains[d]	1.05 ± 0.15	1.00 ± 0.11	0.79 ± 0.09
Average number of distance constraint violations per DYANA conformer			
0.2−0.5 Å	1.0	1.1	1.0
>0.5 Å	0	0	0
Average number of dihedral angle constraint violations per DYANA conformer >5°	0	0	0
Ramachandran map statistics			
residues in most favored regions (%)	71	73	73
residues in additional allowed regions (%)	23	25	25
residues in generously allowed regions (%)	5	2	2
residues in disallowed regions (%)	1	0	0

[a] Twenty conformers with lowest DYANA target function values out of 100 calculated for YqfB (excluding the N-terminal 22-residue tag). (1) PDB is the reference structure deposited (1TE7) in the PDB;[31] (2) without SSA is the reference structure calculated with the constraint input of the PDB structure omitting the stereospecific assignments (SSA); (3) refined is PDB structure refined with NOEs resolved in the shift doublet spectra of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/ NH]. [b] Relative to pairs with nondegenerate chemical shifts. [c] Note that the numbers vary for without SSA and PDB due to handling of stereospecific assignments.[7a] [d] Best-defined side chains include residues 7, 16, 22−25, 35, 37−39, 52, 55−67, 71, 77, 79, 81, 84−86, 93, 97, 98.

required to achieve comparable completeness of NOE detection in central peak and shift doublet subspectra. This appears to be inappropriately long considering that through-bond GFT NMR and 3D NOESY spectra required for structure determination of YqfB were recorded in 26 h (Table 2). Importantly, it may well suffice to detect shift doublets only for stronger NOEs in order to increase the fraction of assigned long-range NOEs to a level that an accurate initial structure can be calculated. Hence, for shift doublet data sets recorded with 30 h (data set I) and 60 h (data set II) of measurement time, we determined the fraction of NOEs in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for which a corresponding shift doublet peak was detected. In data set I (data set II), 75% (93%) of the ^{15}N-resolved long-range NOEs (total: 104) and 64% (95%) of the $^{13}C^{aliphatic}$-resolved long-range NOEs (total: 225) were detected.[33] Hence, the majority of long-range NOEs were detected after 30 h measurement time, while NOE detection approaches completeness (relative to 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]) only when investing 60+ h of spectrometer time. We thus propose that for proteins up to ∼15−20 kDa ($\tau_r \sim 8-10$ ns), the measurement time for shift doublet detection should be ∼2−3 times longer than that for central peak detection. Considering increased losses due to spin relaxation at longer τ_r (Figure S1), we anticipate that 3−4 times longer measurement times are required for larger proteins in the 20−25 kDa range ($\tau_r \sim 10-13$ ns).

Chemical Shift-Based Unambiguous NOE Assignment. The fraction of NOEs which can be unambiguously assigned directly from chemical shift data represents a key "figure of merit" determining the robustness of an NMR structure determination protocol.[4] Hence, we examined to what extent the measurement of the fourth chemical shift encoded in the splitting of the shift doublets resolve assignment ambiguities encountered in 3D NOESY (Figure 6).

Among the long-range NOEs in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH], only 33% of the ^{15}N-resolved and 18% of the $^{13}C^{aliphatic}$-resolved NOEs can be assigned without reference to an initial structure. When considering also the 4D information encoded in the shift doublets of data set I (data set II), these fractions increase to 74% (88%) and 71% (83%). Hence, most shift doublets can be unambiguously assigned based on chemical shifts; that is, the detection yield of shift doublets largely determines the fraction of unambiguously assigned NOEs (Figure 6).

Additional NOEs Resolved in Shift Doublet Subspectra. Due to the lower sensitivity of shift doublet detection, one expects to register only those NOEs as shift doublets which are also present in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] as central peaks. However, peak overlap and cases of severe chemical shift degeneracy quite generally prevent one from completely assigning peaks in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH], even with reference to a high-quality structure (for YqfB, 86% of all NOE peaks could be assigned). Analysis of the shift doublet subspectra of data set II yielded 22 hitherto unidentified (new) ^{15}N-resolved and 19 $^{13}C^{aliphatic}$-resolved NOEs. Hence, as long as the measurement times invested for shift doublet and central peak acquisition are similar, shift doublets are primarily expected to play a key role in assigning central peak NOEs in 3D

(33) Similar fractions are obtained for medium-range NOEs assigned based on chemical shifts in 3D NOESY/3D NOESY *plus* shift doublet data set I/3D NOESY *plus* shift doublet data set II: ^{15}N-resolved NOEs (total: 114) 21%/53%/72%; $^{13}C^{aliphatic}$-resolved NOEs (total: 90) 14%/64%/76%. Moreover, we obtained 60%/100%/100% for long-range H^N-H^N (total: 25) and 29%/69%/83% for long-range CH_3-CH3 NOEs (total: 29).

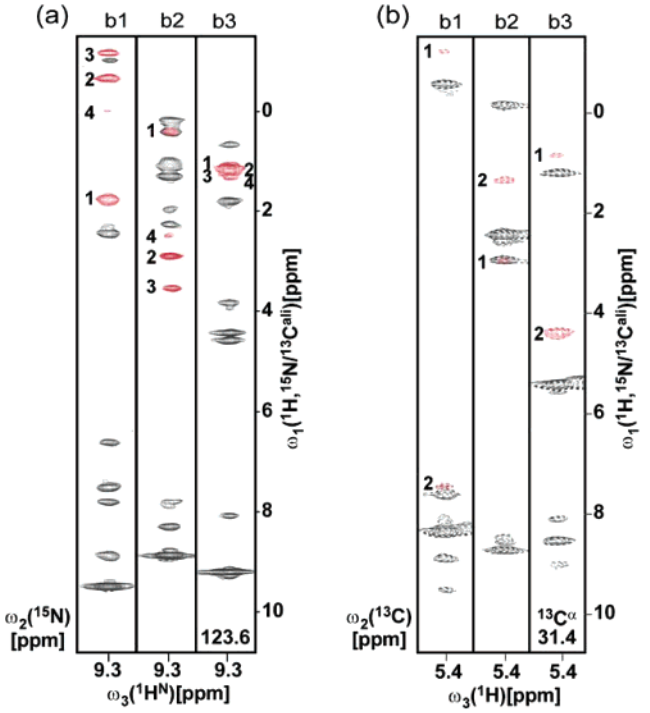


Figure 6. Chemical shift-based NOE assignment in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] acquired with 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for central peak detection. (a) ^{15}N-resolved part: [ω_1,ω_3($^1H^N$)] strips taken from the shift doublet subspectrum exhibiting peaks at $\Omega(^1H)+\Omega(X)$ along ω_1 (X = ^{15}N,^{13}C; shown on the left and indicated as b1), from the shift doublet subspectrum exhibiting peaks at $\Omega(^1H)-\Omega(X)$ along ω_1 (X = ^{15}N,^{13}C; indicated in the middle as b2) and the central peak subspectrum (3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH]; indicated on the right as b3). Signals were detected on the backbone amide proton of Leu 78. Four peaks are shown in red which are overlapped in the central peak spectrum. These peaks are well resolved and can be unambiguously assigned directly from chemical shifts in the shift doublet subspectra (data set I recorded in 30 h). Assignments: NOEs between H^N of Leu 78 and (1) $H^{\beta 2}$ of Leu 78, (2) $CH_3{}^{\gamma 2}$ of Thr 77, (3) $CH_3{}^{\beta}$ of Ala 71, and (4) H^{γ} of Leu 78. (b) ^{13}C-resolved part: [ω_1,ω_3($^1H^{aliphatic}$)] strips taken from the same subspectra. Signals were detected on α-proton of residue Thr 58. Two long-range NOEs are shown in red which can be unambiguously assigned directly from chemical shifts in the shift doublet subspectra. Assignments: NOEs between H^{α} of Thr 58 and (1) $CH_2{}^{\gamma}$ of Ile 98, (2) H^{α} of Ile 98.

NOESY. Otherwise, except for small proteins, comparably few new distance constraints are anticipated.

QUEEN Analysis of NOE Constraint Networks. Depending on their uniqueness, upper distance limit constraints vary in their impact on structure determination and refinement. Hence, reporting the sheer number of NOEs (Table 3) provides only a semiquantitative assessment of the constraint network obtained with a given protocol. Recently, the algorithm and program QUEEN[28] were devised to quantify information contained in experimental NMR data by information theoretical analysis in distance space. We have used QUEEN to evaluate the long-range constraint networks obtained for protein YqfB when recording (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] acquired with 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for central peak detection (Figure 7). As expected, the total information, I_{total}, of the reference constraint list (1.577 bits/atom^2) is significantly higher than I_{total} of the constraint list derived from 3D NOESY with sole reference to chemical shifts (1.327 bits/atom^2). In contrast, we have that the I_{total} values of the constraint lists derived with reference to (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] shift doublet data sets I and II (1.539 and 1.547 bits/atom^2) are nearly the same and also quite similar to I_{total} of the reference list. This shows that shift doublet data set I (30 h measurement time) enables one to derive the major fraction of the final I_{total}, predicting that an accurate initial structure can be obtained from the thus obtained initial constraint network.

Furthermore, a calculation of the distribution of unique information, I_{uni}, and average information, I_{ave}, of the constraints yielding the reference structure (1TE7) confirms that chemical shift-based NOE assignment identifies constraints irrespective of their location in the [I_{uni},I_{ave}] plot; that is, the highly informative constraints are identified with the same probability as those with a low information content (Figure 7a). This feature of shift-based NOE assignment is in contrast to what can be expected for the algorithm implemented in the program CYANA. Network anchoring[7b] favors identification of distance constraints with comparably lower information content since the constraints with high information content are not embedded in a (dense) network which might serve for assignment. As a result, constraints identified for calculating an initial CYANA structure represent less information in the context of the final constraint network (Figure 7b). This might imply that, in general, initial structures derived from (4,3)D NOESY may be more accurate than those obtained with CYANA from 3D NOESY.

Comparative NMR Structure Calculations. Comparative structure calculations were performed to explore the impact of NOEs which could be assigned directly based on chemical shift data when having shift doublet data sets I or II along with 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] (Table 4; Figure S2 and Figure 8). As usual, mean pairwise rmsd values were calculated relative to mean coordinates in order to assess the *precision* of the resulting bundles of conformers. Stereospecific assignments, which are mostly obtained during the last stages of structural refinement by using the GLOMSA module of DYANA,[7a] were not considered. Hence, precision was assessed relative to a YqfB structure obtained with the NOEs used for the PDB structure (1TE7) but after omission of all stereospecific assignments (column 1 in Table 3, denoted "without SSA"). Accuracy was assessed by calculating the rmsd values between the mean coordinates of the resulting bundle of conformers and the mean coordinates of the reference structure of YqfB deposited in the PDB ID 1TE7 (column 2 in Table 3 denoted "PDB"). Assuming a small number of (unavoidable) human errors for NOE assignment (see Materials and Methods), it is certainly so that the "true" structure is somewhere "in between" the manual structure and the structures obtained either with reduced NOE constraint input or with automated methods. However, NOE assignments were carefully double-checked by visual line-shape comparison, which ensured that the remaining number of assignment errors is very small. Clearly, automated methodology relying on chemical shift data only cannot accomplish the same high reliability. Hence, we assumed for the present study that the manual structure represents, in a good approximation, a "gold standard" for benchmarking the alternative structure determination protocols. A summary of rmsd values obtained from the comparative structure calculations is given in Table 4 and Figure S2.

A. Impact on *Precision* of Initial Structure. For comparison of precision, rmsd values were calculated for a structure obtained with the NOE input (Figure 5) used for determining the reference YqfB structure[24] (1TE7; Figure 8a), except that stereospecific

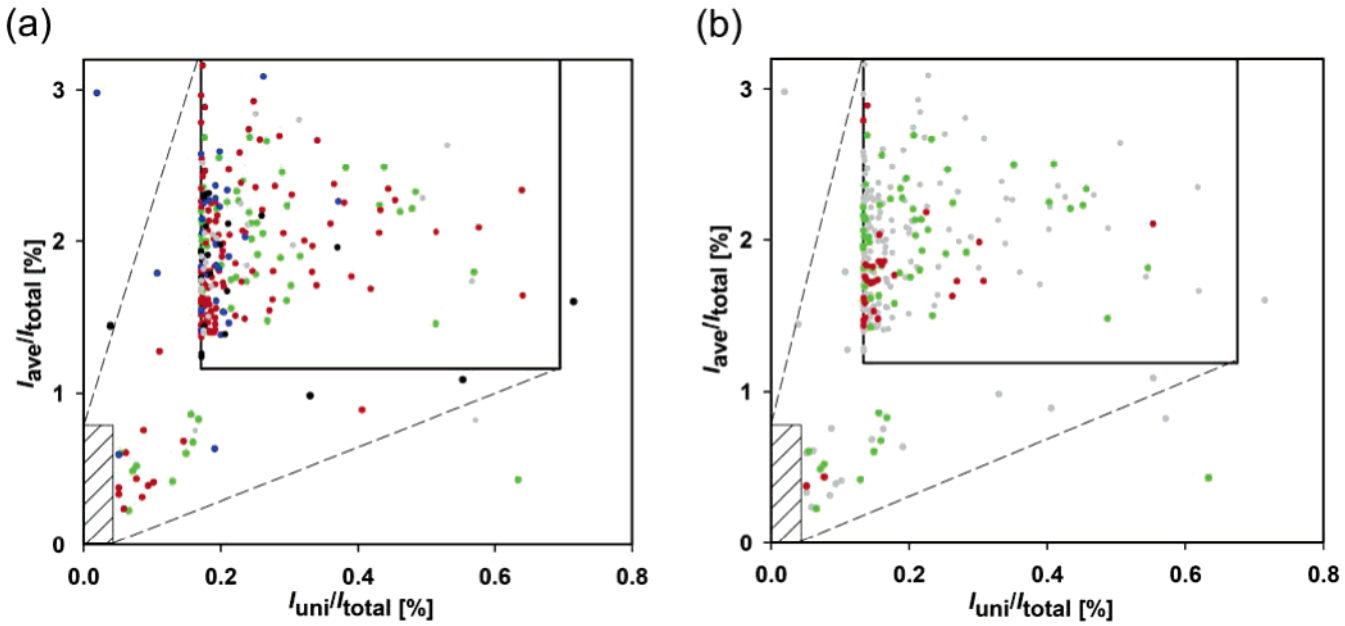


Figure 7. [I_{uni},I_{ave}] plot calculated with the program QUEEN[28] of upper distance limit constraints derived from the 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] peak list (see text) used to calculate the reference YqfB structure (1TE7). In (a), constraints obtained in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on chemical shift data only are depicted as green circles, additional constraints assigned with reference to shift doublet data sets I and II of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH] are shown as red and blue circles, respectively. Aromatic and other aliphatic constraints which could be assigned only with reference to an (initial) structure are displayed as black and gray circles, respectively. The distribution of the circles shows that chemical shift based NOE assignment is, as expected, not dependent on the information content of an NOE. In (b), additional constraints, which were identified by the program CYANA[7b] in the first cycle with a probability of >85%, are depicted as red circles, while those assigned based on shifts are in green [as in (a)], and others are displayed in gray. Comparison with (a) reveals significant bias toward identification of constraints with lower information content. The average I_{uni}/I_{total} and I_{ave}/I_{total} are 0.071 and 0.679% for the red constraints (total: 72) in (a), which were identified with chemical shift doublet data set I. The corresponding values for the red constraints (total: 40) in (b), which were assigned with CYANA, are only 0.011 and 0.254%, respectively.

Table 4. Rmsd Values of Comparative NMR Structure Calculations

		rmsd[a] [Å]			
		Precision[c]		Accuracy[d]	
	structure[b]	backbone	heavy atoms	backbone	heavy atoms
1	reference (PDB ID 1TE7) (X)[e]	0.65 ± 0.18	1.19 ± 0.25		
2	reference (no SSA) (+)[f]	0.97 ± 0.22	1.58 ± 0.24	1.28	1.62
3	3D NOESY and shifts only (■)[g]	4.02 ± 1.03	5.00 ± 0.96	3.22	3.70
4	3D NOESY and shifts + (4,3)D dataset I (●)[h]	1.97 ± 0.44	2.86 ± 0.46	1.58	2.27
5	3D NOESY and shifts + (4,3)D dataset II (○)[h]	1.53 ± 0.28	2.42 ± 0.33	1.35	1.80
6	CYANA (1st cycle) (▼)[i]	2.18 ± 0.49	2.96 ± 0.55	2.81	3.69
7	CYANA starting from (4) (▲)[j]	0.82 ± 0.19	1.35 ± 0.23	1.57	2.17
8	CYANA staring from (5) (△)[j]	0.90 ± 0.18	1.46 ± 0.22	1.58	2.17

[a] Rmsd values were calculated for residues 4−101 of YqfB. For a graphical representation of the values, see Figure S2. [b] Symbols provided in parentheses are those identifying the corresponding bundle in Figure 8. [c] Precision was assessed by calculating the mean pairwise rmsd for the bundle of NMR conformers relative to their mean coordinates. [d] Accuracy was assessed by calculating the rmsd between the mean coordinates of the bundle of NMR conformers and the reference structure. [e] Reference structure. [f] Reference structure calculated after omission of stereospecific assignments (see text and Table 3). [g] Initial structure calculated from NOEs assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on chemical shift data only. [h] Initial structure calculated from NOEs assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on chemical shift data only, but with reference to shift doublet datasets I and II of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH]. [i] Initial structure calculated with the program CYANA in the 1st cycle using the final 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] peak list used for determining the reference structure[24] (PDB ID 1TE7). [j] Structures calculated with the program CYANA when starting with the constraint input yielding the structures indicated with footnote h.

assignments were omitted. This yielded rmsd values of 0.97 ± 0.22 Å (Figure 8b) and 1.58 ± 0.24 Å for backbone and heavy atoms of residues 4−101, respectively (Tables 3 and 4). As expected, a structure of rather low precision is obtained when considering solely long-range NOEs which can be assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on shift data; the corresponding rmsd values are 4.02 ± 1.03 Å (Figure 8c) and 5.00 ± 0.96 Å, respectively. Evidently, if the structure refinement is to be completed using manual methods with the 3D NOESY data set alone, great care would be necessary to ensure proper convergence. When including also NOEs that could be assigned with reference to the shift doublets detected in data set I (or data set II), these values drop to 1.97 ± 0.44 Å (Figure 8d) (1.53 ± 0.28 Å; Figure 8e) and 2.86 ± 0.46 Å (2.42 ± 0.33 Å). This shows that (i) the 4D information encoded in (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH] allows one to generate quite precise initial structures, and (ii) the structure obtained with data set I exhibits rmsd values which are only ∼1 Å above those of the reference structure (calculated after omission of stereospecific assignments; Table 3). The rmsd values are consistent with the finding that ∼70−75% (∼80−90%) of all backbone and aliphatic long-range NOEs can be assigned as central peaks in 3D NOESY (see above) when having the information of shift doublet data set I (data set II). Furthermore,

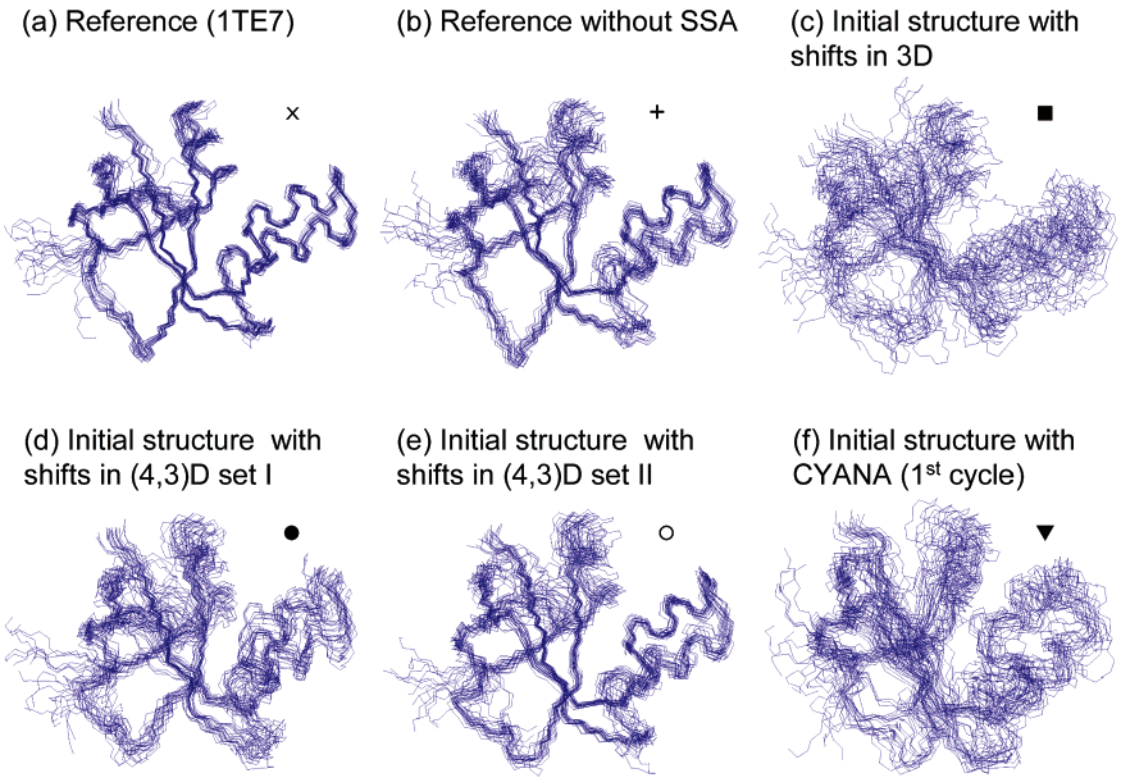


Figure 8. Survey of YqfB structures. (a) Reference structure[24] (PDB ID 1TE7; code in Figure 8: ×). (b) Reference structure calculated after omission of stereospecific assignments (+). (c) Initial structure calculated with NOEs assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] based on chemical shift data only (■). (d) Initial structure calculated with additional NOEs assigned in 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] with reference to shift doublet data sets I of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH] (●). (e) Same as in (d), but with shift doublet data set II (○). (f) Initial structure calculated with the program CYANA[7b] (1st cycle out of 7) with TALOS dihedral angle constraints and intraresidue, sequential, and medium-range distance constraints as input (▼). The 20 DYANA conformers with the lowest residual target function value were chosen to represent the NMR solution structures. These conformers were superimposed for minimal rmsd of backbone heavy atoms N, C^{α}, and C′ of residues 4−101 and the heavy atoms of the best defined side chains (Table 3). The figure was generated using the program Molmol.[34]

the major part of the precision gap to the reference structure is due to the fact that the shift doublet spectra do not exhibit aromatic NOEs; when including all aromatic constraints, the backbone rmsd value drops to 1.41 ± 0.23 Å (1.12 ± 0.17 Å for data set II). This is close to the reference value of 0.97 ± 0.22 Å. We thus conclude that, in agreement with information theoretical QUEEN analysis of the constraint networks (Figure 7), the measurement time invested for recording shift doublet data set I (30 h) represents the most effective approach for obtaining a precise initial fold which ensures smooth convergence of the structure refinement.

B. Impact on *Accuracy* of Initial Structure. The 20 best conformers representing the PDB reference structure were used to assess accuracy and exhibit mean rmsd values of 0.65 ± 0.18 Å (backbone) and 1.19 ± 0.25 Å (all heavy atoms) relative to the mean coordinates of residues 4−101 (Table 3). These rmsd values define, arguably somewhat arbitrarily, the "allowed conformational space" associated with the reference structure. We consider an initial structure as "accurate" if the conformational space associated with the initial structure overlaps with the allowed conformational space. This criterion is fulfilled if the rmsd values calculated between the mean coordinates of initial and reference structure are smaller than the sum of the mean rmsd values to the mean coordinates of the two structure calculations. Root mean square deviation calculations show that initial structures derived from NOEs assigned based on chemical shift data are accurate (Table 4; Figures S2 and 8). In fact, the initial structures are accurate enough to even have the mean coordinates of the reference structure located within their allowed conformational space.

C. Comparison with CYANA's Initial Structure. To compare the impact of (4,3)D [$\underline{HC}^{ali}$/$\underline{HN}$]-NOESY-[CH^{ali}/NH] data acquisition versus an approach to automatically assign NOEs in 3D NOESY, we have performed a CYANA[7b] calculation. The first cycle of a CYANA calculation generates an initial structure by using network anchoring and constraint combination, that is, likewise without reference to three-dimensional structural knowledge.[7b] When providing only TALOS dihedral angle constraints, and intraresidue, sequential, and medium-range $^1H-^1H$ upper distance limit constraints as input, along with the unassigned reference 3D NOESY peak list, the resulting initial CYANA structure exhibits rmsd values of 2.18 ± 0.49 Å (backbone) and 2.96 ± 0.55 Å (all heavy atoms) relative to the mean coordinates (Table 4, Figures S2 and 8). Hence, with the final peak list as input, the algorithm implemented in CYANA for solving the NOE assignment problem results in a precision which is comparable to what is obtained with 4D NOESY information (Table 4). However, the corresponding accuracy after the 1st cycle turns out to be lower than the initial structure obtained with 4D NOESY information based on chemical shifts only (Table 4). This may well be due to the fact that employment of network anchoring may not support the assignment of NOEs with the highest information content (Figure 7). This is because the high information content arises

from *not* being embedded in a (dense) NOE network (which may serve to "anchor" them). Moreover, it might be that use of a larger number of ambiguous long-range constraints simply leads to an overestimation of the precision of the initial structure.

D. CYANA Structures Derived from (4,3)D NOESY-Derived Initial Structures. The high accuracy of the initial folds obtained from (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] (Table 4, Figure S2) allows one to reliably obtain most of the remaining NOE assignments using the program CYANA. Both precision and accuracy of the obtained structures are close to the values of the manually obtained reference structure (a slightly lower accuracy compared to the manually refined structure indicates, however, that manual intervention, such as visual line shape comparison, would be required for finishing the refinement). We thus conclude that the combination of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] with automated structure determination protocols,[6,7] possibly a parallel consensus operation of bottom-up and top-down protocols, represents a powerful approach to ensure fast and robust high-throughput determination of high-quality NMR structures.

E. Refinement of YqfB Reference Structure. The reference structure (1TE7) determined with distance constraints derived from 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] (Figure 5) was further refined by incorporating NOEs that could be resolved in the shift doublet subspectra (data set II; Figure 6). As expected, the comparably few new NOEs result in only a moderate increase in precision; only the rmsd value calculated for backbone and best-defined side chains (that is, the molecular core) is somewhat decreased (column 3 in Table 3 denoted "refined"). This finding supports the view that shift doublet detection in (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] is primarily a valuable tool to assign the majority of the (stronger) NOEs directly based on chemical shift data only.

Conclusions

GFT NMR enables one to "cut the Gordian knot" and combine rapid NMR data collection with robust, high-quality NMR structure determination. This is pivotal for NMR-based structural biology and genomics. The resonance assignment is facilitated by the fact that only five GFT NMR experiments are required, each providing 4D and 5D NMR spectral information at high digital resolution. (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] acquired with 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] for central peak detection affords the information of several 3D and 4D heteronuclear resolved NOESY spectra and enables detection of a dense networks of ^{1}H$-^{1}$H upper distance constraints, as required for high-quality structures. The majority of NOEs detected as shift doublets can be assigned based on chemical shift data only, and assignment of the weaker NOEs, which are often not detectable as shift doublets, is greatly facilitated by having in a single 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] spectrum with each X1$-$H$^{1}\cdots$H$^{1}-$X2 NOE resolved at the chemical shift of X1 and the corresponding to the transposed peak resolved at the chemical shift of X2. Moreover, the impact of distance constraints referred to aromatic rings for structural refinement has long been documented,[1,30] which emphasizes the importance of including ^{13}C^{aromatic}-resolved [^{1}H,^{1}H]-NOESY in the scheme for simultaneous acquisition (see also legend of Figure 5). For proteins comprising a large number of aromatic rings, recording of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{aro}][35] might be advisable. The present study shows that for proteins up to $\sim$15$-$20 ($\sim$20$-$25) kDa, it is recommended to acquire the shift doublet subspectra with $\sim$2$-$3 ($\sim$3$-$4) times of the measurement required for 3D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] (the central peak spectra). This suffices to assign most of the stronger NOEs by detection of shift doublets, and the thus obtainable precise *and* accurate initial structures ensure rapid convergence of the structure refinement. (For smaller proteins, it can be envisaged that (4,2)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] or (3,2)D [H]-NOESY-[CH^{ali}/CH^{aro}/NH] are valuable tools to study protein structure and folding.) In the future, new software for efficiently symmetrizing[15d] (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] subspectra about the position of the central peaks along the GFT dimension will allow one to increase the effective S/N ratio of the shift doublets.[36] This will further reduce the NMR measurement time required for (4,3)D NOESY and/or increase its feasibility for large proteins, and combination of (4,3)D [HC^{ali}/HN]-NOESY-[CH^{ali}/NH] with automated structure determination protocols promises to be an ideal choice for high-throughput determination of proteins.

Acknowledgment. This work was supported by the National Science Foundation (MCB 0075773 and 0416899 to T.S.) and the National Institutes of Health (P50 GM62413). We thank Drs. A. Yee and C. Arrowsmith for providing the sample of protein YqfB.

Supporting Information Available: Figure S1: relative sensitivity of shift doublet versus central peak detection. Figure S2: graphical presentation of the rmsd values of Table 4. Complete refs 3b and 5b. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0501870

(34) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51$-$55.

(35) (4,3)D [HC^{ali},HN]-NOESY-[CH^{aromatic}] can be readily implemented by eliminating the second [^{15}N,^{1}H^{N}]-HSQC module and tuning the second [^{13}C,^{1}H]-HSQC module in Figure 2a for aromatic CH moieties. Except for ^{1}H$^{aromatic}-^{1}$H^{aromatic} NOEs, all NOEs involving aromatic protons are detected in such an experiment and allow their assignment based on detection of shift doublets. Considering that the spectral width of ^{13}C^{aromatic} is about one-half of the spectral width of ^{13}C^{aliphatic}, the minimal measurement time is $\sim$6.5 h at 600 MHz (i.e., about half of the minimal measurement time of the aliphatic congener; Table 2).

(36) With such software, the precision and accuracy of the initial structure obtained with data set II (60 h measurement time; Table 4; Figures S2 and 8) can be expected be obtained with data set I (30 h measurement time).