A standardized protocol enabling rapid NMR data collection for high-quality protein structure determination is presented that allows one to capitalize on high spectrometer sensitivity: a set of five G-matrix Fourier transform NMR experiments for resonance assignment based on highly resolved 4D and 5D spectral information is acquired in conjunction with a single simultaneous 3D 1H,15N,13C aromatic resolved [1H,1H]-NOESY (3D NOESY) (12, 13) for high-quality NMR structure determination. The protocol was used for eight targets of the Northeast Structural Genomics (NESG) consortium (www.nesg.org). Molecular masses of uniformly [13C,15N]-double-labeled polypeptides expressed with tags for structural studies ranged from 10 to 22 kDa (average: 16.2 kDa), and NMR experiments were recorded with ~1 mM protein solutions at ambient temperature. The study demonstrates feasibility and robustness of high-throughput solution NMR structure determination of domain-sized proteins.

Materials and Methods

NMR Sample Preparation. Seven uniformly (U) [13C,15N]-labeled samples were produced at the NESG production site at Rutgers University as described in ref. 14 for targets encoded by genes Pyrococcus furiosus PF0470 (SwissProt accession no. Q8U3J6; NESG ID P1R14), Bacillus cereus BC4709 (Q816V6; BcR68), Bacillus subtilis yqBG (P45923; SR215), Escherichia coli yhgG (P64639; ET95), Methanosarcina mazei rps24q (Q8FPZ95; MaR11), Bacillus halodurans BH1534 (Q9KC55; BhR29), and Homo sapiens UFC1 (Q9Y3C8; HR41). The expressed proteins contained a C-terminal tag with sequence LEH6 to facilitate purification, and ~1 mM solutions were prepared (Table 1) in 95% H2O/5% D2O (20 mM Mes, pH 6.5/100 mM NaCl/10 mM DTT/5 mM CaCl2/0.02% NaN3). The eighth U-[13C,15N]-labeled sample was produced for a target encoded by E. coli gene yqfB (P67603; ET99). The sample was produced at the Toronto site as described in ref. 3, contained a 22-residue N-terminal tag with sequence MGTSHESGRENYLFQGH, and was concentrated to ~1 mM in 90% H2O/10% D2O (25 mM Na phosphate, pH 6.5/400 mM NaCl/1 mM DTT/20 mM ZnCl2/0.01% NaN3). The predicted in vivo molecular masses of the target proteins range from 9 to 20 kDa (average: 14.0 kDa). However, when

Using conventional multidimensional NMR, however, fast data collection for structure determination is impeded by the need to record several spectra, each of which requires sampling of two or more indirect dimensions (7). With highly sensitive instrumentation, this protocol can lead to data acquisition in the “sampling limited” regime (4), in which a large fraction (or even most) of the spectrometer time is invested to sample indirect dimensions and not for achieving sufficient signal-to-noise ratios. G-matrix Fourier transform (GFT) NMR spectroscopy (8–10) offers a solution to this “NMR sampling problem” (11) by joint sampling of several indirect dimensions. This approach leads to detection of “chemical shift multiplets” in which each component encodes a defined linear combination of jointly sampled shifts. To avoid spectral crowding, G-matrix transformation enables one to edit the multiplets; that is, each type of linear combination of shifts is registered in a separate subspectrum.

Here, we present a protocol for rapid NMR data collection based on GFT NMR and simultaneous 3D 15N,13C aromatic resolved [1H,1H]-NOESY (3D NOESY) for high-quality NMR structure determination. This protocol was used for eight targets of the Northeast Structural Genomics (NESG) consortium (www.nesg.org). Molecular masses of uniformly [13C,15N]-double-labeled polypeptides expressed with tags for structural studies ranged from 10 to 22 kDa (average: 16.2 kDa), and NMR experiments were recorded with ~1 mM protein solutions at ambient temperature. The study demonstrates feasibility and robustness of high-throughput solution NMR structure determination of domain-sized proteins.