Supplementary Methods

Sample Preparation. Full-length MM1357 gene from Methanosarcina mazei was cloned into a pET21d (Novagen) derivative, yielding the plasmid MaR30-21.1. The resulting open reading frame contains an additional eight nonnative residues at the C-terminus (LEHHHHHH) of the protein. *E. coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with MaR30-21.1, and cultured in MJ minimal medium containing (15NH₄)₂SO₄ and U-¹³C-glucose as the sole nitrogen and carbon sources. Initial growth was carried out at 37 °C until the OD₆₀₀ of the culture reached ≈0.7 units, followed by overnight induction of protein expression at 17 °C by the addition of isopropyl-β-D-thiogalactopyranoside at a final concentration of ≈1 mM. The cells were harvested by centrifugation and lysed by sonication. U-¹³C,¹⁵N Q8PX65 was purified in a two step protocol consisting of Ni affinity (HisTrap HP, GE Healthcare) and gel filtration (HiLoad 26/60 Sephadex 75, GE Healthcare) chromatography. The final yield of U-¹³C,¹⁵N Q8PX65 (>97% homogeneity by SDS-PAGE; 9.17 kDa by MALDI-TOF mass spectrometry) was ≈42 mg/L.

NMR Spectroscopy. For the conventional experiments, samples of U-¹³C,¹⁵N Q8PX65 for NMR spectroscopy were prepared at a concentration of 0.7 mM in 95% H₂O/5% D₂O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl₂ at pH 6.5. NMR data were collected at 20 °C, processed with NMRPipe² and analyzed by SPARKY (Goddard and Kneller, University of California, San Francisco). NMR data were acquired on Varian INOVA 600 MHz and Bruker AVANCE 500 MHz spectrometers, equipped with 5 mm triple resonance conventional and cryoprobes, respectively. Proton chemical shifts were referenced to DSS, while ¹³C and ¹⁵N chemical shifts were referenced indirectly using the gyromagnetic ratios of ¹³C:¹H (0.251449530) and ¹⁵N:¹H (0.101329118), respectively. Backbone (HN, Hα, N, C', Cα) and Cβ resonance assignments were made with AutoAssign,³ using peak lists from 2D ¹H-¹⁵N HSQC and 3D HNCO, HN(CA)CO, HN(CO)CA, HNCA, HN(CO)CACB, and HNCACB along with spin-system typing assignment constraints obtained from (H)CC(CO)NH TOCSY spectra. Side chain aliphatic assignments were completed manually using 3D (H)CC(CO)NH-TOCSY, H(CCCO)NH-TOCSY, and HCCH-COSY spectra. Side chain aromatic resonances were assigned using 3D HCCH-TOCSY and ¹³C-edited NOESY spectra. Individual assignments for most (16/22) Asn/Gln side chain amide protons (H(E) and H(Z)) were made on the basis of H°/°° NOE intensities.⁴ Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in a high resolution 2D ¹H-¹³C HSQC spectrum of 5%-¹³C,100%-¹⁵N Q8PX65.⁵ All NMR resonance assignments were validated using the AVS software package.⁶ NOE distance constraints were derived from 3D ¹⁵N-edited NOESY (τ_m = 80 ms) and 3D ¹³C-edited aliphatic and aromatic NOESYs (τ_m = 80 ms). Three-bond ³J(H'-H°) scalar couplings were
determined from a 3D HNHA spectrum of $^{13}$C,$^{15}$N Q8PX65. Slowly exchanging backbone amide protons were identified by dissolving lyophilized $^{13}$C,$^{15}$N Q8PX65 in 100% $^2$H$_2$O and monitoring the decay of the $^{15}$N-$^1$H HSQC signal over time.

For the microprobe experiments, samples of $^{13}$C,$^{15}$N Q8PX65 for NMR spectroscopy were prepared at concentrations of 0.7 to 1.4 mM in 95% H$_2$O/5% D$_2$O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl$_2$ at pH 6.5. NMR data were collected at 20 °C, on a Bruker AVANCE 700 MHz spectrometer equipped with a 1 mm TXI HCN z-gradient MicroProbe, with the exception of the 3D HNHA and H/D exchange experiments which were obtained on a Bruker AVANCE 600 MHz spectrometer equipped with a 1 mm TXI HCN z-gradient MicroProbe. Automated backbone and manual side chain assignments as well as NOESY data collection were performed as described above. Side chain aromatic resonances were assigned using the 2D $^1$H-$^{13}$C HSQC, 2D HH plane and 3D aromatic HCCH-TOCSY and 3D aromatic $^{13}$C-edited NOESY spectra.

**Structure Determination.** For the conventional structure determination, structure calculations were performed using the program AutoStructure,$^8$,$^9$ interfaced with XPLOR-NIH 2.0.6.$^{10}$ Briefly, AutoStructure uses an automated “bottom-up” iterative approach of NOESY peak analysis, followed by structure calculation using DYANA$^{11}$ or XPLOR for a user-defined number of cycles. The program first performs a secondary structure analysis and generates a reliable initial protein fold based on the first cycle of NOESY spectral analysis, and chemical shift, scalar coupling constant and slow amide exchange data. AutoStructure automatically generates dihedral angle (using the algorithm HYPER$^{12}$) and iteratively refines the distance (NOE-based) restraints in subsequent cycles. Here, the input for the AutoStructure program consisted of a resonance assignment list, manually edited peak lists with intensities for the 3D $^{15}$N-edited and 3D $^{13}$C-edited NOESY spectra,$^3$ $^3$J(H$_N$-H$_{\alpha}$) values, $\phi,\psi$ angle constraints ($\pm$ 40° and $\pm$ 50°, respectively) derived from chemical shift data using the program TALOS,$^{13}$ and slow amide hydrogen exchange data. TALOS dihedral constraints were used only for residues with confidence scores of 10. Analysis of hydrogen bond constraints is done automatically in the AutoStructure software$^8$,$^9$, based on the combination of short distances in intermediate structures and the experimental amide H/D exchange data. The hydrogen-bond donor is determined by H/D exchange data, and the acceptor is determined by the automated NOE analysis. In each cycle, structure calculations were performed using an XPLOR-NIH 2.0.6 simulated annealing refinement protocol.$^{10}$ The best 10 of 56 structures (lowest energy) from the final cycle of AutoStructure were refined by restrained molecular dynamics in explicit water using CNS,$^{14}$ based on published procedures.$^{15}$

The solution NMR structure of Q8PX65 determined using microprobe NMR data were calculated using AutoStructure interfaced with DYANA 1.5,$^1$ using the assignment list obtained from the microprobe experiments, spectral peak lists from the 3D $^{15}$N-edited and 3D $^{13}$C-edited NOESY
spectra obtained on the microprobe, as well as the $^3\text{J}(\text{H}^\alpha\text{-H}^\text{N})$ and slow N-H exchange information also determined on the microprobe sample. The microprobe data set was acquired with comparable digital resolutions in all dimensions compared to the conventional data set. The final ensemble of microprobe structures was obtained from refinement of the best 10 of 56 structures (lowest target function) from the final cycle of AutoStructure by restrained molecular dynamics in explicit water using CNS.\textsuperscript{15}

**Structure Quality Assessment.** We report the global goodness-of-fit of the final ensemble of structures with the NOESY data using a set of RPF-scores, a formalism recently developed in our laboratory based on information retrieval statistics.\textsuperscript{8} Briefly, Recall measures the percentage of NOESY peaks that are consistent with the interproton distances in the 3D structure, Precision measures the percentage of close distance proton pairs (< 5 Å) in the 3D structure whose back-calculated NOESY cross peaks are observed in the NOESY peak lists, F-measure is the overall performance score calculated from the Recall and Precision, and Discriminating Power (DP)-score is a normalized F-measure which reflects how the query structure is distinguished from the freely-rotating chain model. In practice, DP-scores and F-measures greater than 0.7 and 0.9, respectively, correlate strongly with good structure quality and accuracy. The module for computing protein RPF-scores is included in the AutoStructure software package. RPF-scores for the final ensemble of CNS refined Q8PX65 structures were calculated against the raw 3D NOESY peak list data only. Global structure quality factors for the final ensemble of Q8PX65 structures were determined using the PSVS software package,\textsuperscript{16} which outputs Verify3D,\textsuperscript{17} Prosa II,\textsuperscript{18} PROCHECK,\textsuperscript{19} and MolProbity\textsuperscript{20} raw and statistical Z-scores. In our experience the metrics that are the most sensitive to structure quality and accuracy are the PROCHECK dihedral distribution and MolProbity clash scores.\textsuperscript{21} The final ensemble of conventional structures (minus the C-terminal tag) and complete resonance assignments have been deposited in the Protein Data Bank (PDB ID 1yez; BMRB ID, 6506).

**Mass Sensitivity Experiments.** Relative probehead mass sensitivity experiments (Supplementary Table 1 online) were obtained on a 1.04 mM sample of $^{13}\text{C},^{15}\text{N}$ Q5LST8 from *Silicibacter pomeroyi* (NESG ID: SiR5; 93 a.a.) in 20 mM MES, 100 mM NaCl, pH 6.5 buffer. The protein was expressed and purified following standard protocols of the NESG consortium.\textsuperscript{22} Using identical acquisition parameters (1k x 200; 16 scans per increment), 2D $^1\text{H}-^{15}\text{N}$ HSQC spectra were acquired on a Bruker AVANCE 600 MHz spectrometer equipped with a 1 mm TXI HCN z-gradient MicroProbe, a 5 mm TXI z-gradient probe, and a 5 mm TXI z-gradient CryoProbe. For comparison, we also performed the identical experiments on the system used for the bulk of the microprobe structure work (Bruker AVANCE 700 MHz spectrometer with 1 mm TXI HCN z-gradient MicroProbe). All spectra were
processed identically using NMRPipe\textsuperscript{2} and analyzed using SPARKY. Mass sensitivity values were obtained from the average ratios of signal intensities for well resolved peaks in the microprobe vs. conventional probe spectra, normalized for either the active coil volume or minimum practical sample volume required for the probes.

**Microprobe Performance Experiments.** In order to assess the general applicability of this microprobe to protein structure determination, we obtained a series of identical 2D $^1$H-$^{15}$N HSQC spectra (1k x 200; 16 scans per increment) on NESG targets\textsuperscript{23} of varying molecular weights, whose solution structures have been solved in our consortium, on a Bruker AVANCE 600 MHz spectrometer with 1 mm TXI HCN z-gradient MicroProbe. The following protein samples were used: 1.4 mM $U^{-13}$C,$^{15}$N Q8PX65 from *M. mazei* (NESG ID: MaR30; 68 a.a.) in 20 mM MES, 100 mM NaCl, pH 6.5 buffer; 1.04 mM of $U^{-13}$C,$^{15}$N Q5LST8 from *S. pomeroyi* (TrEMBL ID: Q5LST8_SILPO; NESG ID: SiR5; 93 a.a.) in 20 mM MES, 100 mM NaCl, pH 6.5 buffer; 1.25 mM $U^{-15}$N, 5%-13C ydfO from *Escherichia coli* (TrEMBL ID: Q8FI2_ECOL6; NESG ID: ER251; 150 a.a.) in 20mM NH$_4$OAc, 100 mM NaCl, pH 4.5 buffer; 1.00 mM $U^{-15}$N, 5% $^{12}$C human HSPC155 (TrEMBL ID: Q5VTX1_HUMAN; NESG ID: HR41; 167 a.a.) in 20 mM MES, 100 mM NaCl, pH 6.5 buffer. The proteins were expressed and purified following standard protocols of the NESG consortium.\textsuperscript{22} Average signal to noise ratios were obtained from the first processed 1D increment of each experiment.
References


