SUPPLEMENTARY MATERIAL FOR:

NMR Structure of F-Actin Binding Domain of Arg/Abl2 from *Homo sapiens*

Gaohua Liu,* Rong Xiao, Dongyan Wang, Yuanpeng J. Huang, Thomas B. Acton, Gaetano T. Montelione*

Center for Advanced Biotechnology and Medicine, Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, and Northeast Structural Genomics Consortium

MATERIALS AND METHODS

The Arg (Abl2) F-actin binding domain (FABD) from *Homo sapiens* (UniProtKB/Swiss-Prot ID P42684/ABL2_HUMAN, residues 1058-1182) was cloned, expressed and purified following standard largely automated NESG protocols to produce a uniformly ^{13}C, ^{15}N-labeled protein sample.[1] Briefly, the truncated ABL2_HUMAN (1058-1182) gene was cloned into a pET14-15C (Novagen) derivative, yielding the plasmid pHR5537A-14.12. The resulting construct contains 10 nonnative residues at the N-terminus (MGHHHHHHSH) to facilitate protein purification and one single mutation T1062A was introduced. *Escherichia coli* BL21 (DE3) pMGK cells were transformed with pHR5537A-14.12, and cultured in MJ9 minimal medium[2] containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. *U*-^{13}C, ^{15}N Arg FABD was purified using an ÄKTAxpress™ (GE Healthcare) based two step protocols consisting of IMAC (HisTrap HP) and gel filtration (HiLoad 26/60 Superdex 75) chromatography. The final yield of purified *U*-^{13}C, ^{15}N Arg FABD (> 98% homogeneous by SDS-PAGE; 15.2 kDa by MALDI-TOF mass spectrometry) was ~100 mg/L. In addition, a *U*-^{15}N and 5% biosynthetically directed fractionally ^{13}C-labeled sample[3] was generated for stereo-specific assignment of isopropyl methyl groups. Both *U*-^{13}C,^{15}N and 5%^{13}C,*U*-^{15}N Arg FABD were dissolved, respectively, at concentrations of ~1.2 mM and 1.4 mM in 95% H_2O/5% D_2O (20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3) at pH 4.5. Analytical gel filtration with static light scattering [1] and ^{15}N T_1 and T_2 relaxation data indicate that the protein is monomeric in solution under the conditions used for these NMR studies.

All NMR spectra were recorded at 25 ºC. Triple resonance NMR data were collected on Varian INOVA 600 MHz, a simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-edited NOESY[4] spectrum (mixing time 100 ms) in H_2O and a 3D ^{13}C-edited NOESY in D2O were acquired on a Bruker AVANCE 800 MHz spectrometer. 2D constant-time [^{13}C, ^{1}H]-HSQC spectra with 28 ms and 42

ms constant-time delays were recorded for the 5% biosynthetically directed fractionally ^{13}C-labeled sample on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe in order to obtain stereo-specific assignments for isopropyl groups of valines and leucines.[3] ^{1}H-^{15}N heteronuclear NOE and ^{15}N T_1 and T_2 relaxation measurements were made using gradient sensitivity-enhanced 2D heteronuclear NOE and ^{15}N T_1 and T_2 (CPMG) relaxation experiments, respectively.[5] All NMR data were processed using the program NMRPipe[6] and analyzed using the program XEASY.[7] Spectra were referenced to external DSS. Resonance assignments were achieved as described previously.[8] Sequence specific assignments (H^N, H^α, N, C^α) and H^β/C^β assignments were obtained largely automated with the program AUTOASSIGN,[9] which used to simulate a NOESY peak list that was used to facilitate interactive analysis of side-chain assignments. Simultaneous $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-edited NOESY[4] and CCH-TOCSY spectra were analyzed manually to obtain nearly complete side-chain assignment. Assignments were obtained for 90% of backbone and side-chain chemical shifts assignable with the NMR experiments listed above (excluding N-terminal NH_3^+, Lys NH_3^+, Arg NH_2, OH of Ser, Thr and Tyr, $^{13}C^\gamma$ of Asp and Asn, $^{13}C^\delta$ of Glu and Gln, and aromatic $^{13}C^\gamma$ shifts). Chemical shifts were deposited in the BioMagResBank on 06/14/2009 with accession code 16349. Based on chemical shifts, the locations of regular secondary structure elements were identified.[10] A NOESY peak list containing expected intra-residue, sequential and α-helical medium range NOE peaks was initially generated and was manually edited by visual inspection of the simultaneous NOESY spectra, and subsequent manual peak picking was pursued to identify remaining, primarily long-range NOEs.[8] The programs CYANA[11,12] and AUTOSTRUCTURE[13] were used in parallel to automatically assign long-range NOEs. Assignments identically obtained by both programs ('consensus assignments') were retained and established the starting point for iterative cycles of noise/artefact peak removal, peak picking, NOE assignment and structure calculation.[8] ^{1}H - ^{1}H

upper distance limit constraints for structure calculations obtained from both NOESY were summarized in Table I. In addition, backbone dihedral angle constraints were derived from chemical shifts using the program TALOS+[14] for residues located in well-defined secondary structure elements (Table I). The final structure calculation was performed with CYANA 3.0, and the 20 conformers with the lowest target function value were refined in an 'explicit water bath' [15] using the program CNS.[16] The coordinates were deposited in the Protein Data Bank on 06/14/2009 (accession code 2KK1). Structural statistics and global structure quality factors, including Verify3D,[17] ProsaII,[18] PROCHECK,[19] and MolProbity[20] raw and statistical *Z*-scores, were computed using the PSVS 1.3 software package.[21] The global goodness-of-fit of the final structure ensembles with the NOESY peak list data were determined using the RPF analysis program.[22] All structure figures were made using MOLMOL[23] or PyMOL 1.1.[24]

Table S1. Summary of NMR and structural statistics for *Homo sapiens* Arg FABD[a]

Completeness of resonance assignments[b]	
Backbone (%)	99.0
Side chain (%)	95.8
Aromatic (%)	94.0
Stereospecific methyl (%)	89
Conformationally-restricting constraints[c]	
Distance constraints	
Total	3882
intra-residue ($i = j$)	634
sequential ($\|i- j\| = 1$)	1017
medium range ($1 < \|i – j\| < 5$)	1270
long range ($\|i – j\| \geq 5$)	961
distance constraints per residue	29.4
Dihedral angle constraints	172
Hydrogen bond constraints	
total	84
long range ($\|i – j\| \geq 5$)	0
Number of constraints per residue	31.3
Number of long range constraints per residue	7.3
Residual constraint violations[c]	
Average number of distance violations per structure	
0.1 – 0.2 Å	1.60
0.2 – 0.5 Å	0.5
> 0.5 Å	0
average RMS distance violation / constraint (Å)	0.005
maximum distance violation (Å)	0.32
Average number of dihedral angle violations per structure	
1 – 10°	3.4
> 10°	0
average RMS dihedral angle violation / constraint (degree)	0.43
maximum dihedral angle violation (degree)	6.13
RMSD from average coordinates (Å)[c,e]	
backbone atoms	0.36
heavy atoms	0.68
Ramachandran plot statistics[c,d]	
most favored regions (%)	95.0
additional allowed regions (%)	4.9
generously allowed (%)	0.0
disallowed regions (%)	0.0

Table S1 (Cont'd). Summary of NMR and structural statistics for *E. coli* Spr[37-162][a]

Global quality scores [c]	Raw	*Z*-score
Verify3D	0.30	-2.57
ProsaII	0.55	-0.41
Procheck(phi-psi) [d]	0.30	1.49
Procheck(all) [d]	0.14	0.83
Molprobity clash	18.20	-1.60

RPF Scores[f]	
Recall	0.97
Precision	0.94
F-measure	0.96
DP-score	0.82

[a] Structural statistics were computed for the ensemble of 20 deposited structures (PDB 2KK1).

[b] Computed using AVS software[25] from the expected number of peaks, excluding: highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, non-protonated aromatic carbons, and the N-terminal tag.

[c] Calculated using PSVS 1.3 program.[21] Average distance violations were calculated using the sum over r^{-6}.

[d] Ordered residue ranges [S(phi) + S(psi) > 1.8] : 1065-1066, 1071-1154, 1159-1181.

[e] Regular secondary element: alpha residues 1086-1099, 1106-1123, 1130-1152, 1165-1181

[f] RPF scores[22] reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NMR data.

Figure S1

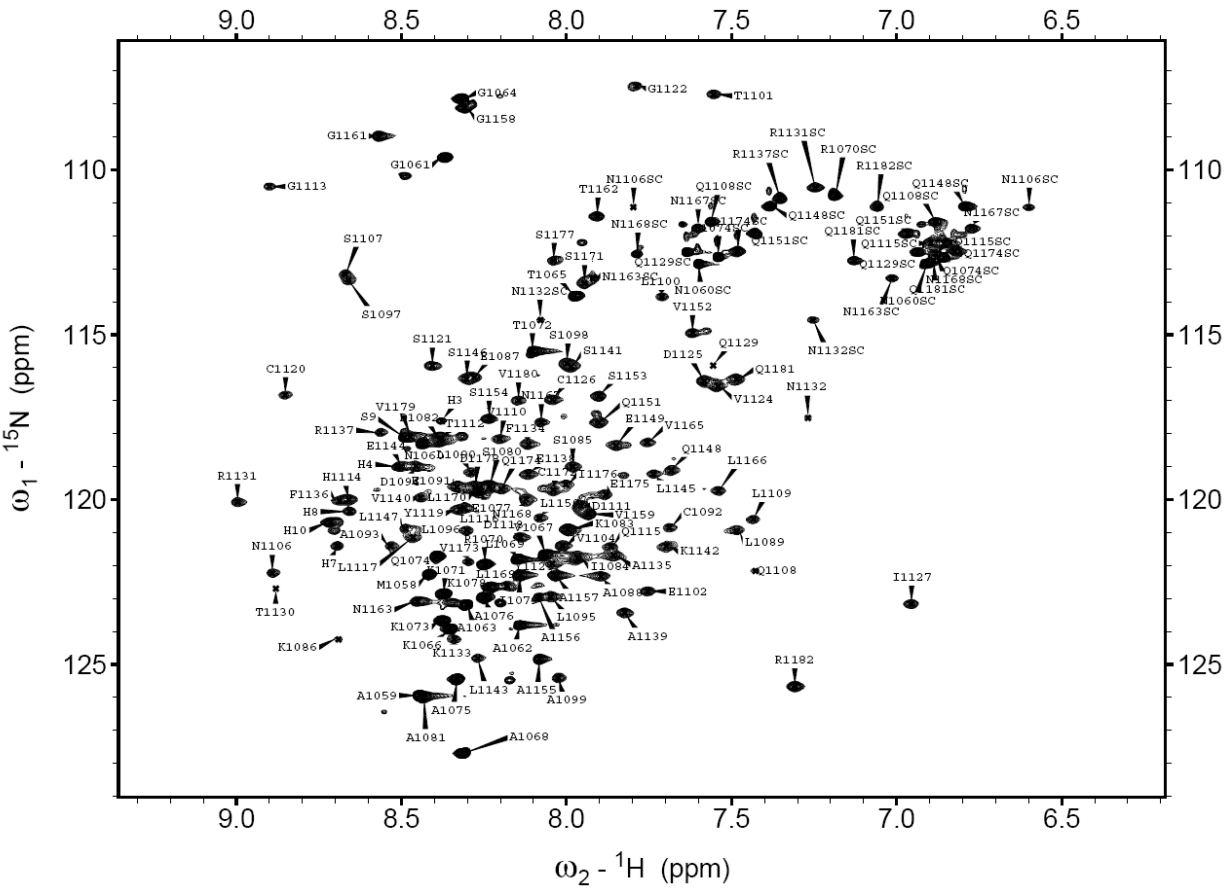


Figure S1: 2D [^{1}H,^{15}N]-HSQC spectrum of ABL2_HUMAN FABD from *H. sapiens*. Peaks are labeled with their respective sequential resonance assignments using the one-letter code of amino acids and the amino acid sequence number. The NMR spectrum was recorded at 25 ºC and pH4.5.

Figure S2.

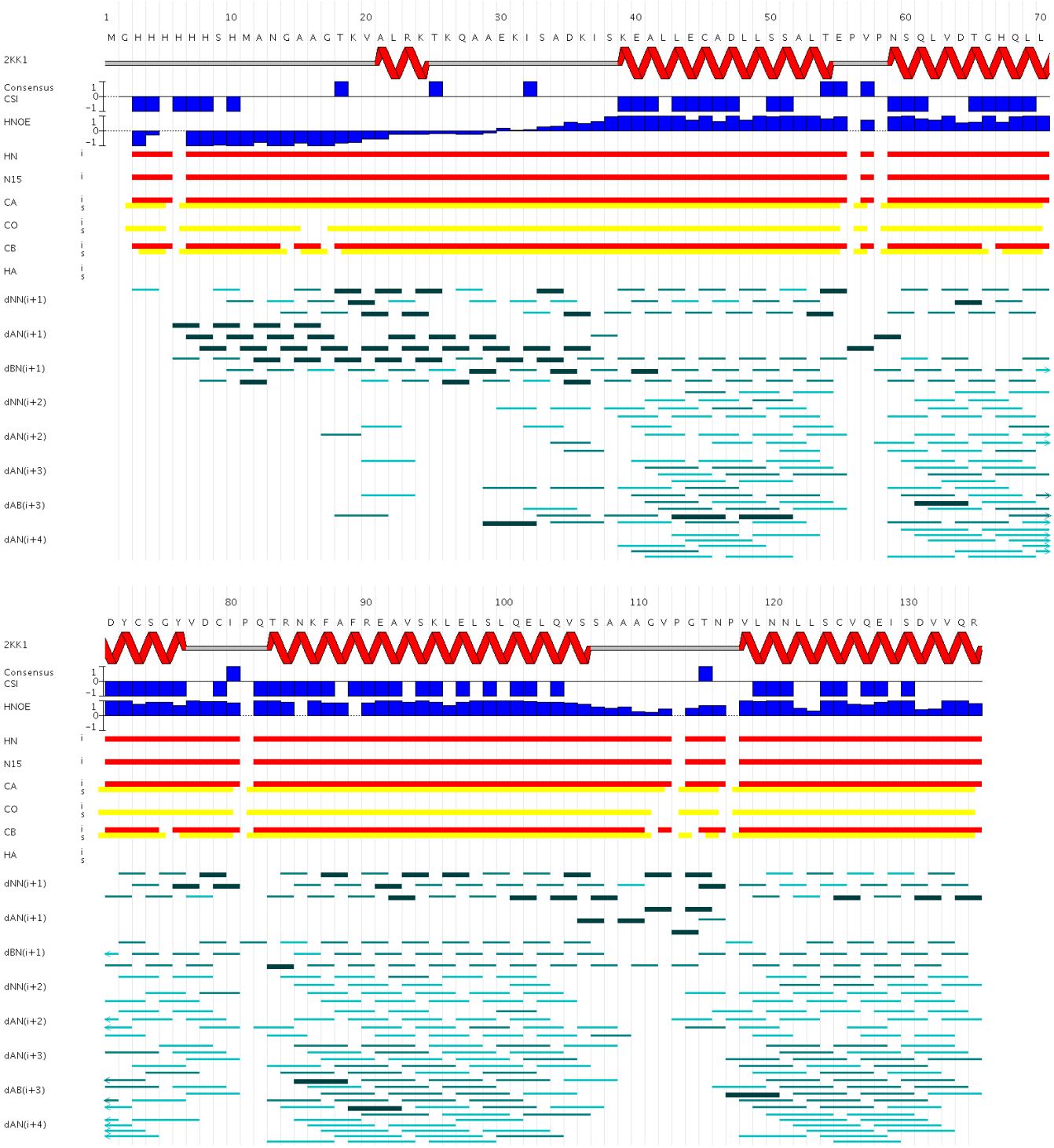

1 10 20 30 40 50 60 70
M G H H H H H H S H M A N G A A G T K V A L R K T K Q A A E K I S A D K I S K E A L L E C A D L L S S A L T E P V P N S Q L V D T G H Q L L
2KK1
Consensus CSI
HNOE
HN
N15
CA
CO
CB
HA
dNN(i+1)
dAN(i+1)
dBN(i+1)
dNN(i+2)
dAN(i+2)
dAN(i+3)
dAB(i+3)
dAN(i+4)
80 90 100 110 120 130
D Y C S G Y V D C I P Q T R N K F A F R E A V S K L E L S L Q E L Q V S S A A A G V P G T N P V L N N L L S C V Q E I S D V V Q R
2KK1
Consensus CSI
HNOE
HN
N15
CA
CO
CB
HA
dNN(i+1)
dAN(i+1)
dBN(i+1)
dNN(i+2)
dAN(i+2)
dAN(i+3)
dAB(i+3)
dAN(i+4)

Figure S2. NMR connectivity map summarizing data used to determine resonance assignments and secondary structure for *H. sapiens* Arg FABD [1058-1182] (labeled as residues 11-135 in the construct used in this work). The Intra-residue (i) and sequential (s) connectivities for the three-rung assignment strategy[9] matching intra-residue and sequential C′, C^{α}, and C^{β} resonances are shown as horizontal red and yellow lines, respectively. Inter-residue NOE connectivities are shown as thin, medium, and thick black lines, corresponding to weak, medium, and strong NOE interactions. Bar graphs of the consensus CSI[26] and ^{1}H-^{15}N heteronuclear NOE data are shown in blue. The secondary structural elements in the final Arg FABD NMR structure (2KK1) are also shown.

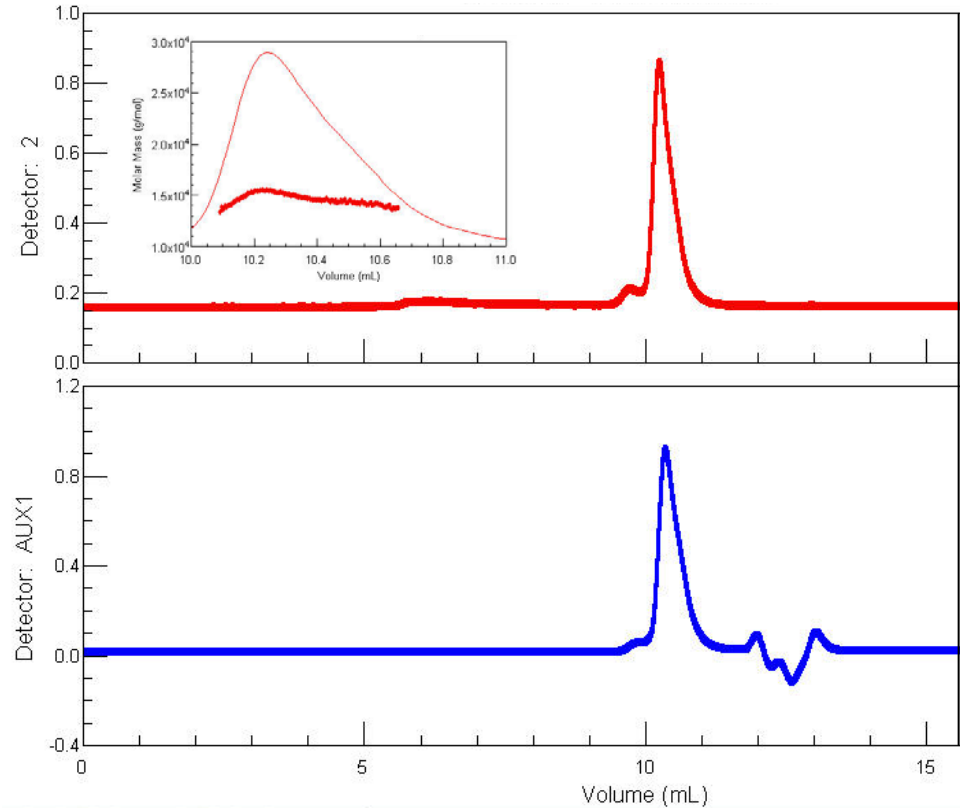
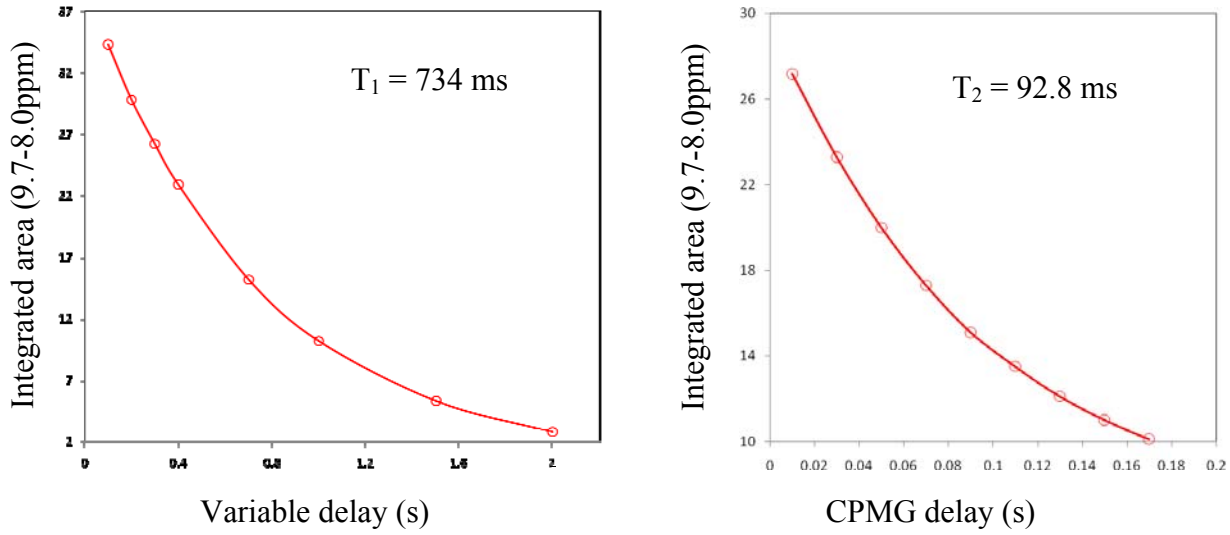


Figure S3. Analytical gel filtraion with static light scattering detection, for *H. sapiens* Arg FABD. Data were collected on a miniDAWN Light Scattering instrument (Wyatt Technology) at λ = 690 nm and at room temperature on an NMR sample of 5%-^{13}C, 100%-^{15}N *H. Sapien* Arg FABD at pH 6.5. Inset: Plot of molar mass versus elution volume. The resulting experimental molecular weight of Arg FABD is 14.7 kDa; the expected MW including affinity tag is 14.6 kDa.

Figure S4.


Integrated area (9.7-8.0ppm)
T1 = 734 ms
Variable delay (s)
Integrated area (9.7-8.0ppm)
T2 = 92.8 ms
CPMG delay (s)

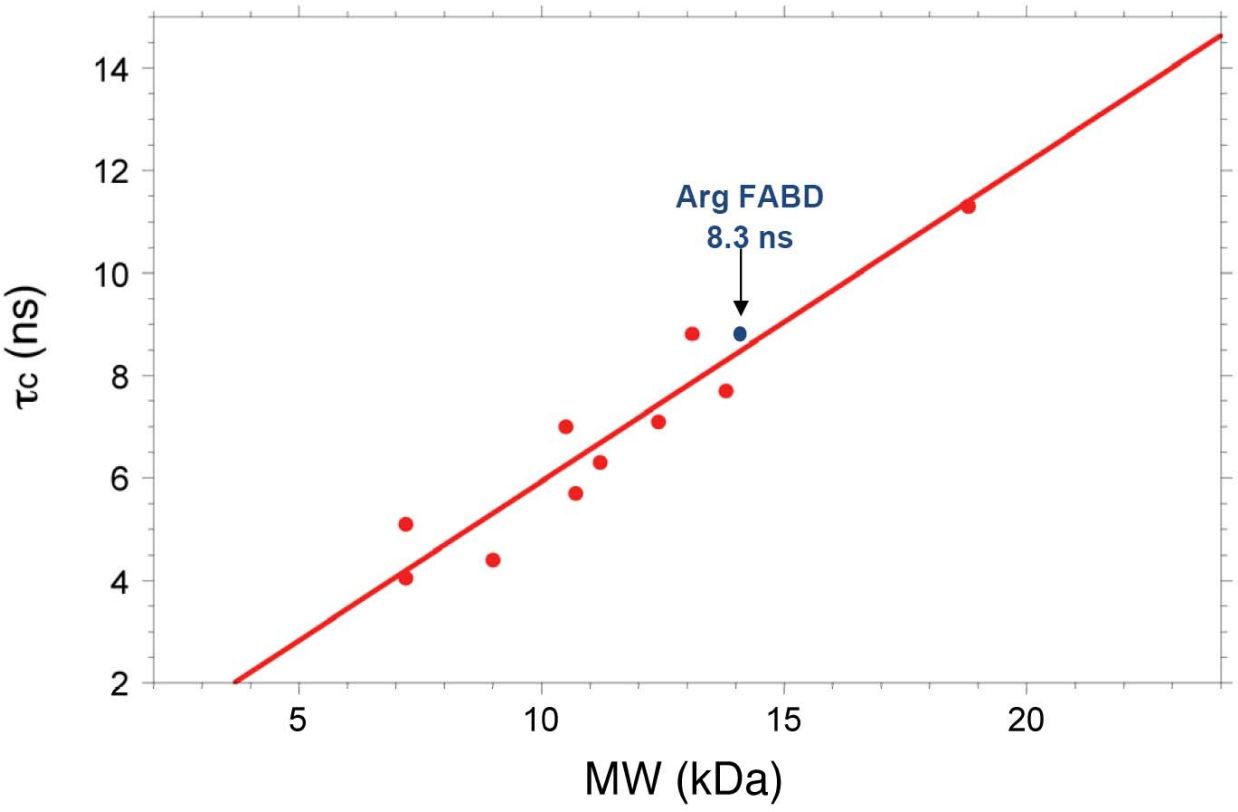

Arg FABD
8.3 ns
τc (ns)
MW (kDa)

Figure S4. ^{15}N T_1 and T_2 relaxation data for 5%-^{13}C, 100%-^{15}N *H. sapiens* Arg FABD. The data were acquired on a Varian Inova 600 MHz spectrometer at 298 K using pseudo-2D ^{15}N T_1 and T_2 gradient experiments.[5] T_1 spectra were acquired with delays, T = 100, 200, 300, 400, 700, 1000, 1500 and 2000 ms, and a relaxation delay of 3s. T_2 spectra were acquired with CPMG delays, T = 10, 30, 50, 70, 110, 130, 150 and 170 ms, and with a relaxation delay of 2s. (Top): ^{15}N T_1 and T_2 values were extracted by plotting the decay of integrated $^1H^N$ intensity between $\delta \approx 8.0 - 9.7$ ppm and fitting the curves with standard exponential equations using the program 'tc' within VNMRJ (Varian). (Bottom): Plot of rotation correlation time, τ_c (ns), versus protein molecular weight (kDa) for known monomeric NESG targets of ranging size (taking into account isotope enrichment as well as affinity tags in the sequence). ^{15}N T_1/T_2 data for all monomeric proteins used for the τ_c vs. MW plot were obtained on 600 MHz spectrometer at 298 K, and analyzed as described above. For each protein, the τ_c was calculated from the ^{15}N T_1/T_2 ratio using the following approximation of literature relaxation equations:[26]

$$\tau_c \approx \left(\sqrt{\frac{6T_1}{T_2} - 7} \right) / 4\pi\nu_N \qquad (2)$$

where ν_N is the resonance frequency of ^{15}N in Hz.

Using this approach, we obtain a τ_c of 8.3 ns for 5%-^{13}C, 100%-^{15}N *H. Sapiens* Arg FABD, shown in blue, which is consistent with a monomer (expected MW = 14.6 kDa, including N-terminal MGHHHHHHSH affinity tag)

Supplementary References

1. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang YW, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GVT, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and Protein Production Platform of the Northeast Structural Genomics Consortium. Method Enzymol 2005;394:210-243.
2. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly N-15- and C-13-enriched fusion proteins in Escherichia coli. Journal of Biomolecular Nmr 1996;7(2):131-141.
3. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling. Biochemistry 1989;28(19):7510-7516.
4. Shen Y, Atreya HS, Liu GH, Szyperski T. G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. Journal of the American Chemical Society 2005;127(25):9085-9099.
5. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 1994;33(19):5984-6003.
6. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. Journal of Biomolecular Nmr 1995;6(3):277-293.
7. Bartels C, Xia TH, Billeter M, Guntert P, Wuthrich K. The Program Xeasy for Computer-Supported Nmr Spectral-Analysis of Biological Macromolecules. Journal of Biomolecular Nmr 1995;6(1):1-10.
8. Liu GH, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. Proceedings of the National Academy of Sciences of the United States of America 2005;102(30):10487-10492.
9. Moseley HNB, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 2001;339:91-108.
10. Wishart DS, Sykes BD. The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. Journal of Biomolecular Nmr 1994;4(2):171-180.
11. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. Journal of Molecular Biology 1997;273(1):283-298.
12. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. Journal of Molecular Biology 2002;319(1):209-227.
13. Huang, Y.J. *et al.*, An integrated platform for automated analysis of protein NMR structures. *Method Enzymol* 394, 111-141 (2005).

14. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44 (4), 213-223 (2009).
15. Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M. Refinement of protein structures in explicit solvent. Proteins 2003;50(3):496-506.
16. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica Section D-Biological Crystallography 1998;54:905-921.
17. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356(6364):83-85.
18. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17(4):355-362.
19. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. Journal of Applied Crystallography 1993;26:283-291.
20. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC. Exploring steric constraints on protein mutations using MAGE/PROBE. Protein Science 2000;9(11):2251-2259.
21. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins 2007;66(4):778-795.
22. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. Journal of the American Chemical Society 2005;127(6):1665-1674.
23. Koradi R, Billeter M, Wuthrich K. MOLMOL: A program for display and analysis of macromolecular structures. Journal of Molecular Graphics 1996;14(1):51-55.
24. DeLano WL. The pymol manual. San Carlos, CA: DeLano Scientific;. 2002.
25. Moseley HN, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 2004;28(4):341-355.
26. Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 1989;28(23):8972-8979.