SUPPLEMENTARY MATERIALS FOR

Human Cyclin Dependent Kinase 2 Associated Protein 1 is Dimeric in its Disulfide-Reduced State, with Natively Disordered N-terminal Region*

Asli Ertekin[1,2], James M. Aramini[1], Paolo Rossi[1], Paul G Leonard[1,3,4], Haleema Janjua[1], Rong Xiao[1], Melissa Maglaqui[1], Hsiau-Wei Lee[5], James H. Prestegard[5], and Gaetano T. Montelione[1,3]

[1]Center for Advanced Biotechnology and Medicine, Northeast Structural Genomics Consortium, Rutgers, The State University of New Jersey, Piscataway, NJ 08854

[2]BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854

[3]Department of Biochemistry, Robert Wood Johnson Medical School-UMDNJ, Piscataway, NJ 08854

[4]Howard Hughes Medical Institute

[5]Complex Carbohydrate Research Center, and Northeast Structural Genomics Consortium, The University of Georgia, Athens, GA 30602

**Running Title: Solution NMR structure of CDK2AP1 from H. sapiens*

To whom correspondence should be addressed: Gaetano T. Montelione CABM, Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854, USA, Tel.: (732) 235-5375; Fax: (732) 235-5779; E-mail: guy@cabm.rutgers.edu

Materials and Methods

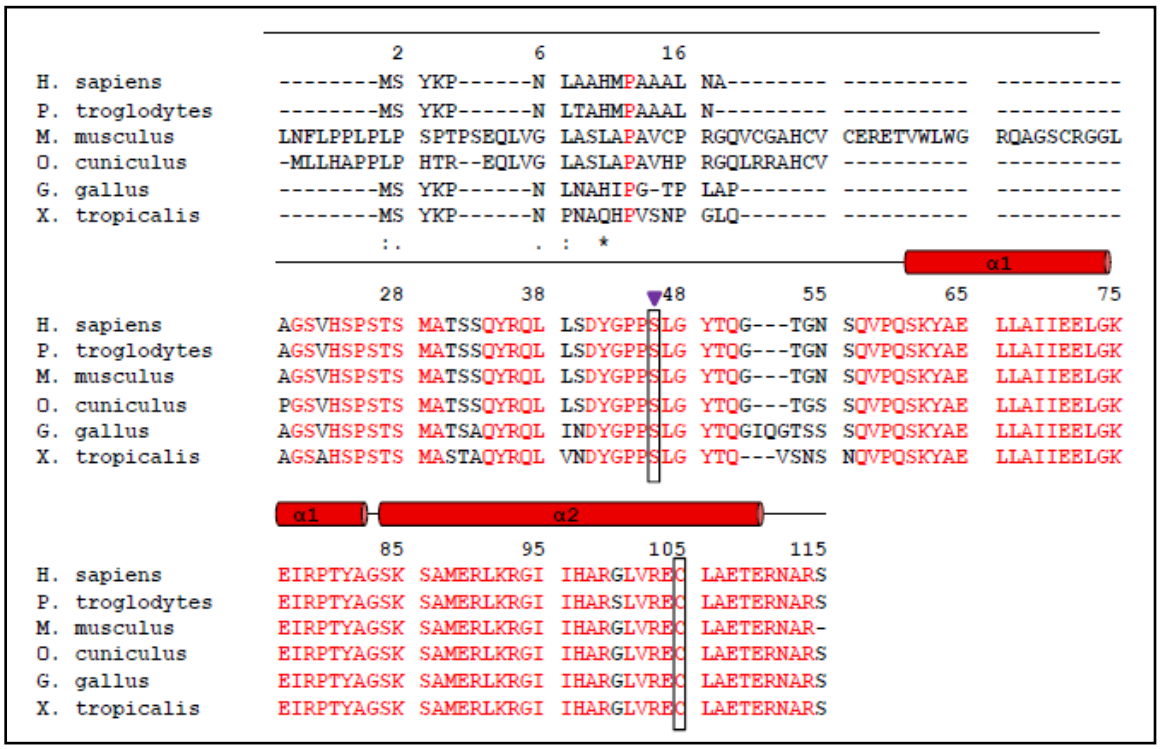
Full-length, truncated and mutant constructs of the *doc-1* gene from *Homo Sapiens* were cloned into pET21 expression vectors (Novagen) containing a N-terminal affinity tag (MGHHHHHHSH), yielding the plasmids HR3057-14, HR3057H-14 and HR3057H-C105A-14. Plasmids were transformed into codon enhanced BL21 (DE3) pMGK *E. coli* cells, which were cultured at 37°C in MJ minimal medium (1) containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as the sole nitrogen and carbon sources. Initial cell growth was carried out at 37°C and protein expression was induced at 17°C by isopropyl-β-D-thiogalactopyranoside (IPTG). Expressed proteins were purified using an AKTAexpress (GE Healthcare) two-step protocol consisting of HisTrap HP affinity chromatography followed directly by HiLoad 26/60 Superdex 75 gel filtration chromatography. Samples of [*U*-^{13}C,^{15}N]-CDK2AP1 and [*U*-^{13}C,^{15}N]-CDK2AP1(61-115) for NMR spectroscopy were concentrated to 0.7 to 0.9 mM in 95% H_2O/5% ^2H_2O solution containing 20 mM MES, 200 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$ at pH 6.5. Sample purity and identity were confirmed by SDS-PAGE, MALDI-TOF mass spectrometry, and NMR spectroscopy. Both proteins are dimers according to static light scattering data.

All NMR data were collected at 25°C on Varian INOVA 600 and Bruker AVANCE 800 NMR spectrometers, processed with NMRPipe (2), and visualized using SPARKY (3). Complete ^{1}H, ^{13}C, and ^{15}N resonance assignments for CDK2AP1(61-115) were determined using conventional (4) triple resonance NMR methods, respectively, and deposited in the BioMagResDB (BMRB accession number 16808). Stereospecific isopropyl methyl assignments for all Val and Leu residues were deduced from characteristic cross-peak fine structures in high resolution 2D ^{1}H-^{13}C HSQC spectra of 5%-^{13}C,100%-^{15}N CDK2AP1(61-115) (5). Resonance assignments were validated using the Assignment Validation Suite (AVS) software package (6). ^{1}H-^{15}N heteronuclear NOEs were measured with gradient sensitivity-enhanced 2D heteronuclear NOE approaches (7,8).

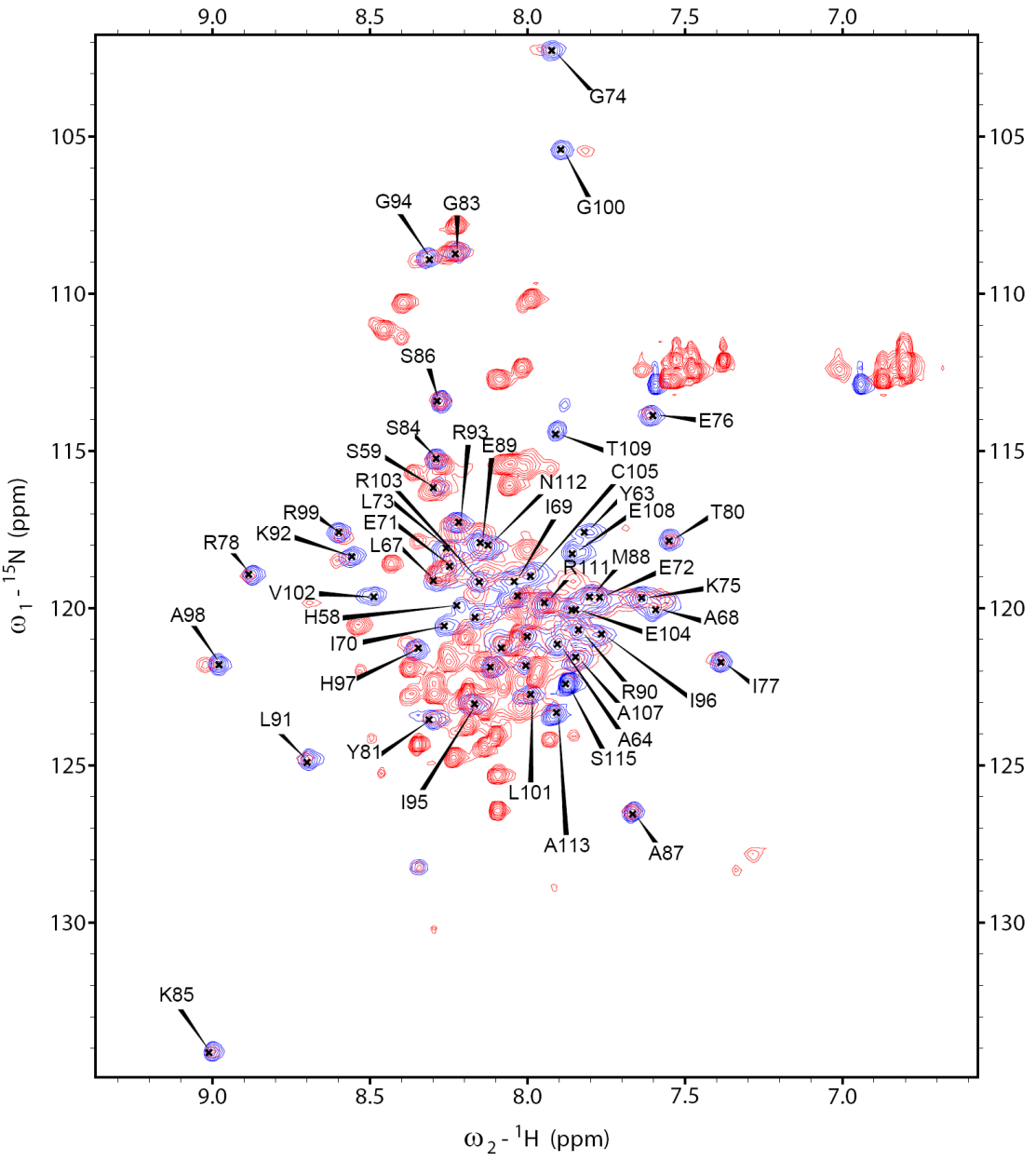
The one bond ^{1}H-^{15}N couplings for isotropic and aligned samples of CDK2AP1(61-115) were measured using an interleaved TROSY HSQC set of experiments (9). The one bond ^{15}N-^{13}C' couplings for isotropic and aligned samples of CDK2AP1(61-115) were measured using NC' J-modulation experiments (10).

CDK2AP1(61-115) was first aligned in a positively charged (50% 3-acrylamidopropyl-trimethylammonium chloride + 50% acrylamide) compressed gel medium.(11) The positively charged gel was initially cast in a 3.2 mm diameter plastic tube. The polymerized gel was first washed extensively in deionized water followed by washing with protein buffer to equilibrate the pH. Finally, the gel was washed with deionized water to remove the buffer. The swelled gel (~7 mm diameter) was trimmed to a length of 35 mm and dried in open air for two days. The gel pellet was swollen in a 5 mm Shigemi NMR tube using the protein solution. The plunger of the Shigemi tube was fixed at a height of 12 mm from the bottom of the tube.

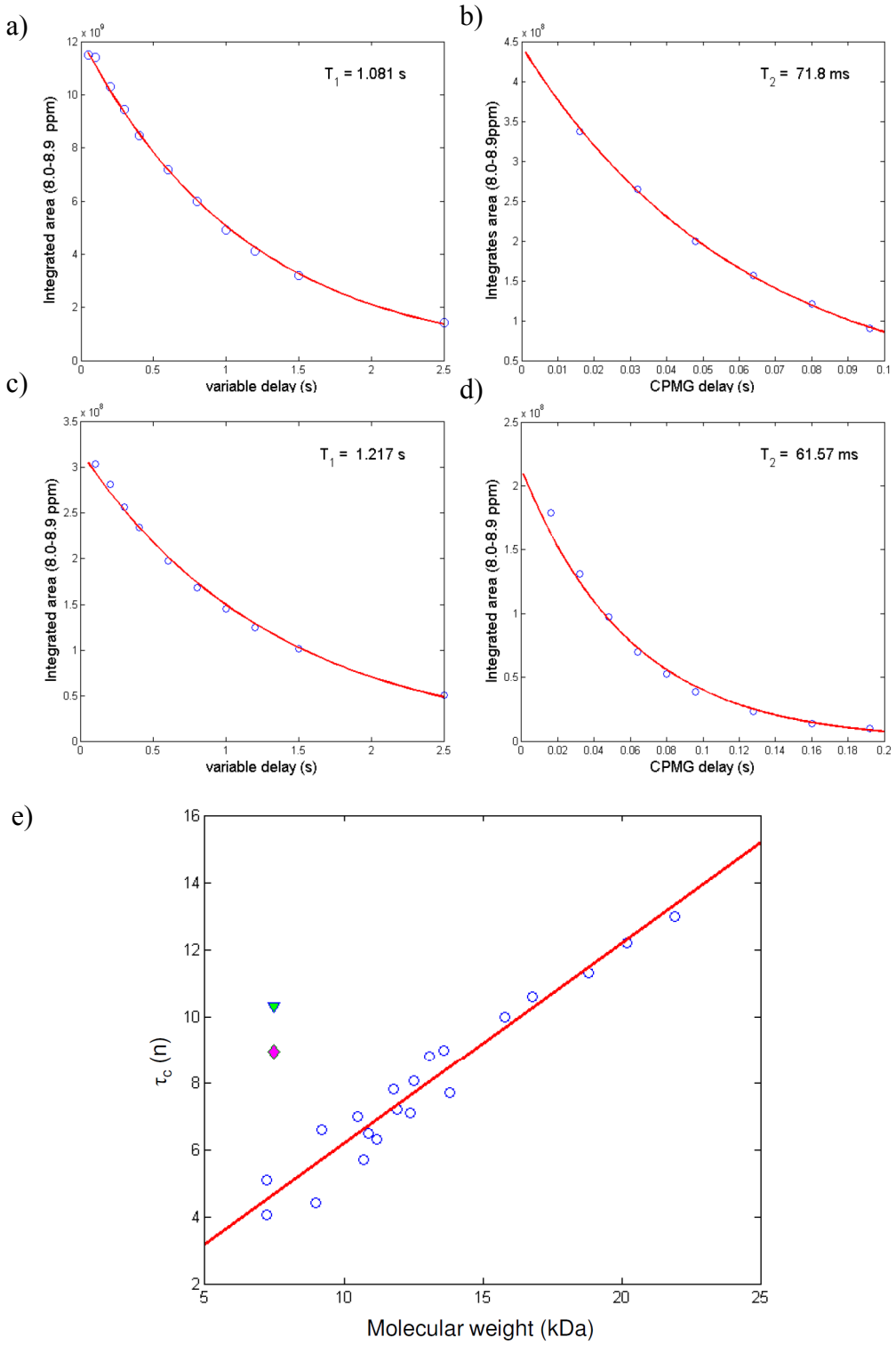
For the truncated CDK2AP1(61-115) structure determination, initial structure calculations were performed by CYANA 3.0 (12,13), using peak intensities from 3D simultaneous CN-NOESY (14) (τ_m = 120 ms), dihedral angle constraints computed by TALOS (ϕ ± 20°; ψ ± 20°) (15), and N-H^N and N-C' RDC constraints. The 20 structures with lowest target function out of 100 in the final cycle were further were refined by restrained molecular dynamics in explicit water using CNS 1.1 (16,17), using the final NOE derived distance, TALOS dihedral angle and RDC constraints. The final refined ensemble of 20 structures was deposited into the Protein Data Bank (PDB ID, 2KW6). Structural statistics and global structure quality factors, including Verify3D (18), ProsaII (19), PROCHECK (20), and MolProbity (21) raw and statistical *Z*-scores, were computed using the PSVS 1.3 software package (22). The global goodness-of-fit of the final structure ensembles with the NOESY peak list data were determined using the RPF analysis program (23). PALES was used to back calculate the alignment tensor and the RDC values from the final structures (24), which allows additional validation statistics for the final ensemble.



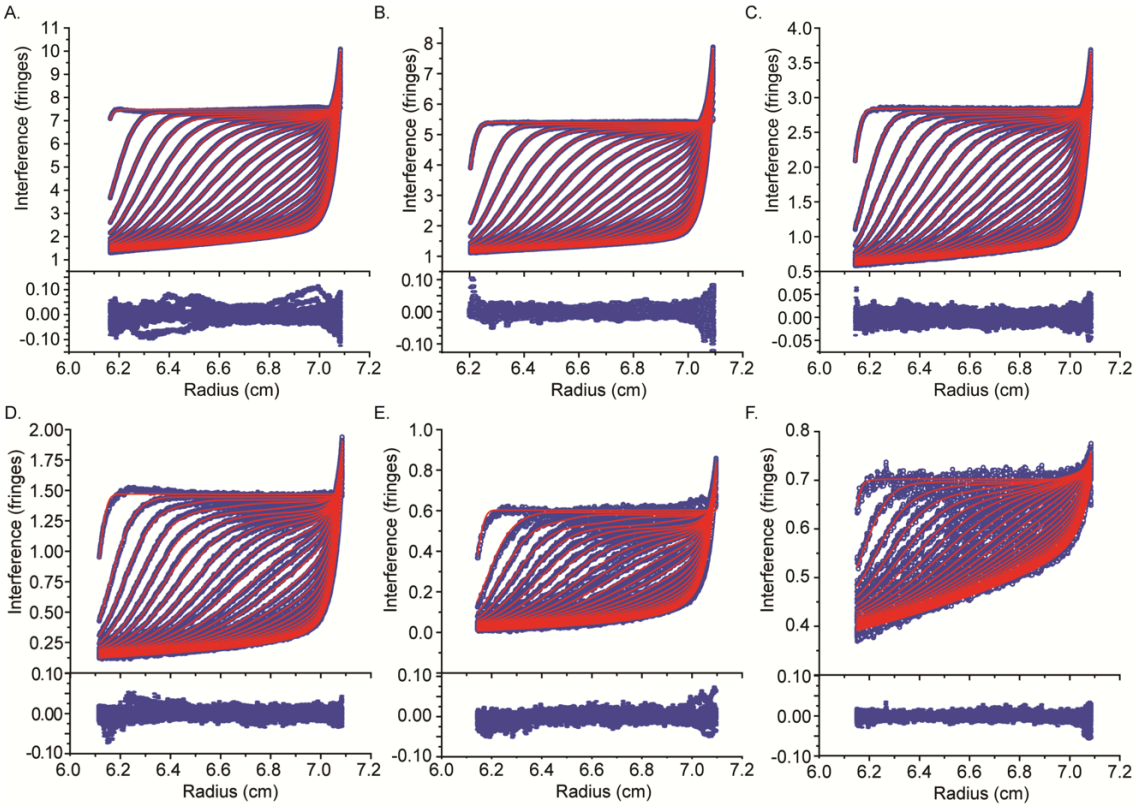
Supplementary Figure S1. Multiple sequence alignment for CDK2AP1 computed by CLUSTALW. The sequence numbering for CDK2AP1 and secondary structure elements are based on the NMR structure for human CDK2AP1(61-115) are shown in rows above the alignment. The phosphorylation site, Ser46 is indicated by a triangle.



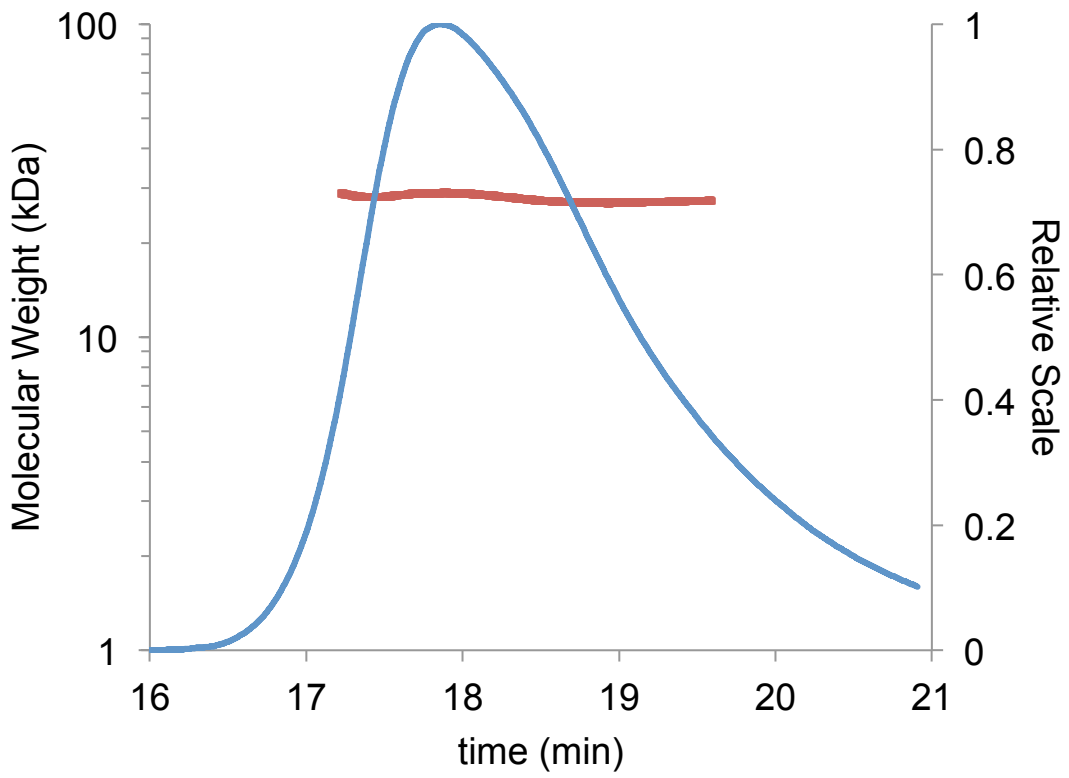
Supplementary Figure S2. Overlay of 600 MHz ^{1}H-^{15}N HSQC spectra of full-length (red) and 61-115 construct (blue) of CDK2AP1. Both spectra were recorded at 298 K in at ~ 0.5 mM protein concentration in 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% $^{2}H_2O$, and 50 µM DSS.



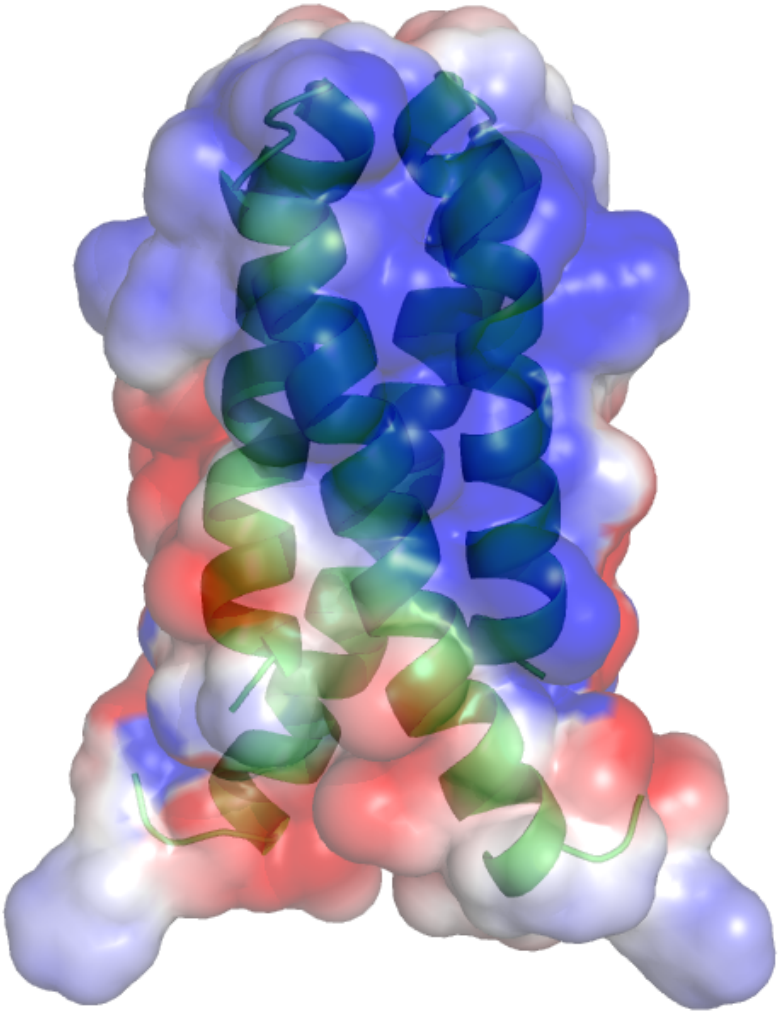
Supplementary Figure S3. ^{15}N NMR relaxation values, T_1 and T_2, at 25 °C for wt CDK2AP1(61-115) (a and b) and CDK2AP1(61-115)-C105A (c and d) are shown. The rotational correlation time, τ_c, calculated from *T1* and *T2* of known monomeric proteins (blue circles), wt CDK2AP1(61-115) (magenta diamond) and CDK2AP1(61-115)-C105A (green triangle) are plotted with respect to their molecular weights (e). CDK2AP1(61-115) samples were prepared in 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% ^2H_2O, and 50 µM DSS.



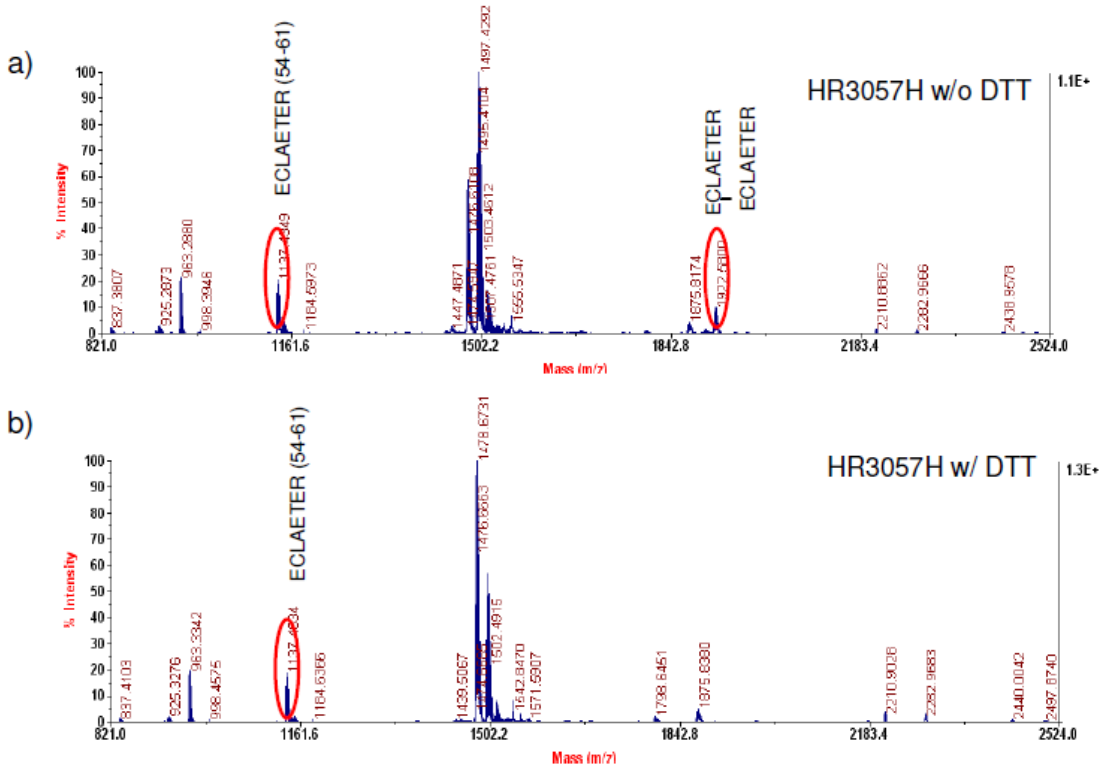
Supplementary Figure S4. Sedimentation velocity analytical ultracentrifugation analysis of CDK2AP1(61-115) using the Sedphat hybrid local continuous distribution model with a single discrete species. Samples of CDK2AP1(61-115) were dissolved in 20 mM HEPES, 200 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT at pH 7.0 were analyzed at six loading concentrations: A. 300 µM, B. 200 µM, C. 100 µM, D. 50 µM, E. 20 µM and F. 10 µM. Boundary fits of the sedimentation scans recorded using interference optics at 50,000 r.p.m. are shown for each protein concentration. For clarity, only every tenth scan of the 300 scans recorded is depicted, with the raw data points shown as blue open circles and the line of best fit shown in red. The residuals for each data set are shown in the lower panel for each dataset.



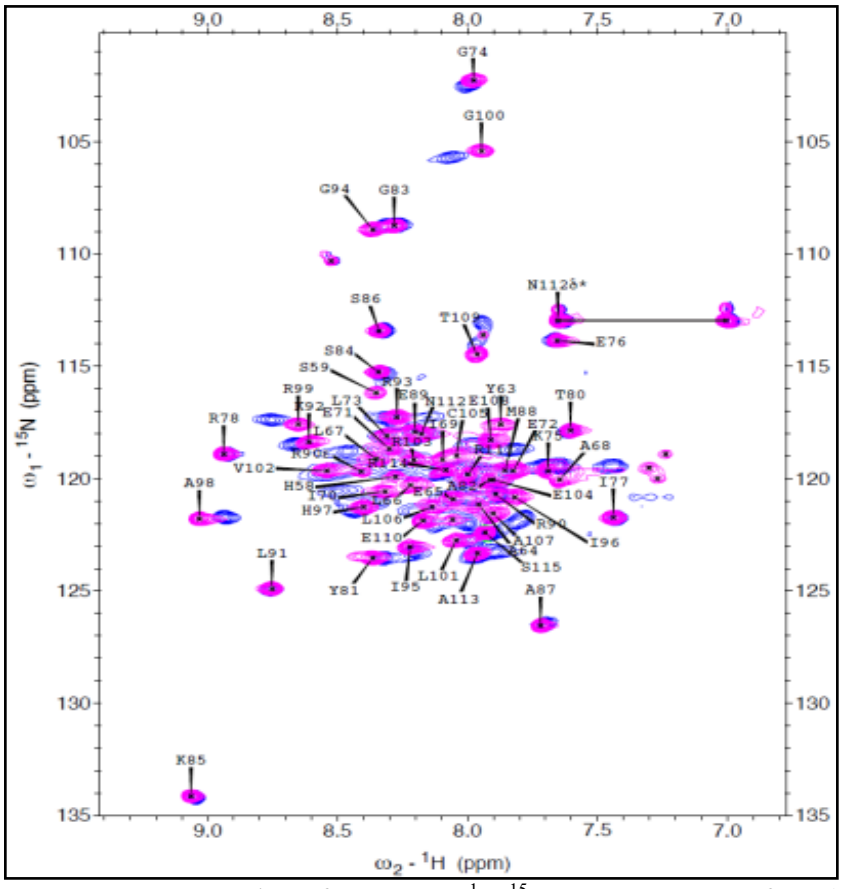
Supplementary Figure S5. Static light scattering results for full-length CDK2AP1. The NMR sample (30 µl) of CDK2AP1 in a buffer containing 20 mM MES (pH 6.5), 200 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, 1X Proteinase Inhibitors, 0.02% NaN_3 was injected onto an analytical gel-filtration column (Protein KW-802.5, Shodex, Japan) with the effluent monitored by refractive index (black trace, Optilab rEX) and 90° static light-scattering detectors. The resulting experimental molecular weight of CDK2AP1 is 27.83 kDa (red), the expected MW of the single chain, including affinity tag, is 13.78 kDa.



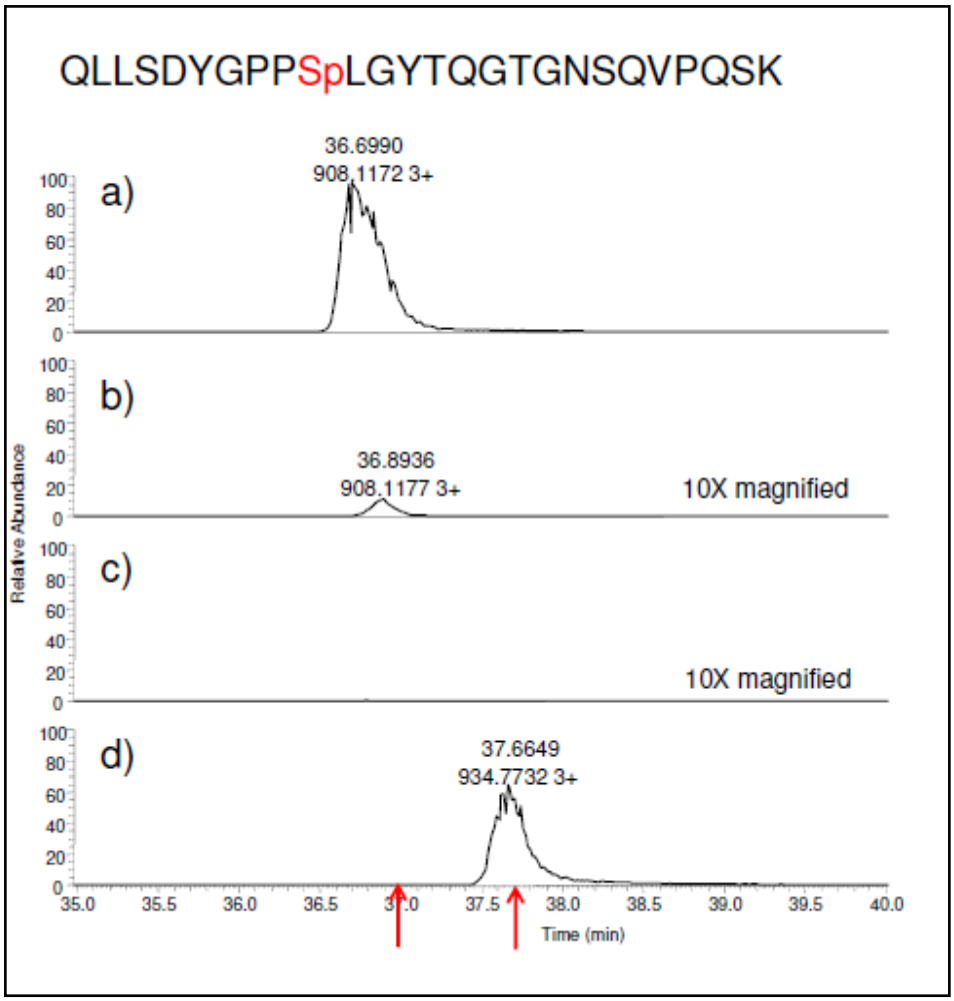
Supplementary Figure S6. Electrostatic potential map for the lowest-energy solution NMR structure of CDK2AP1(61-115), calculated using the APBS (25) add-on in Pymol.



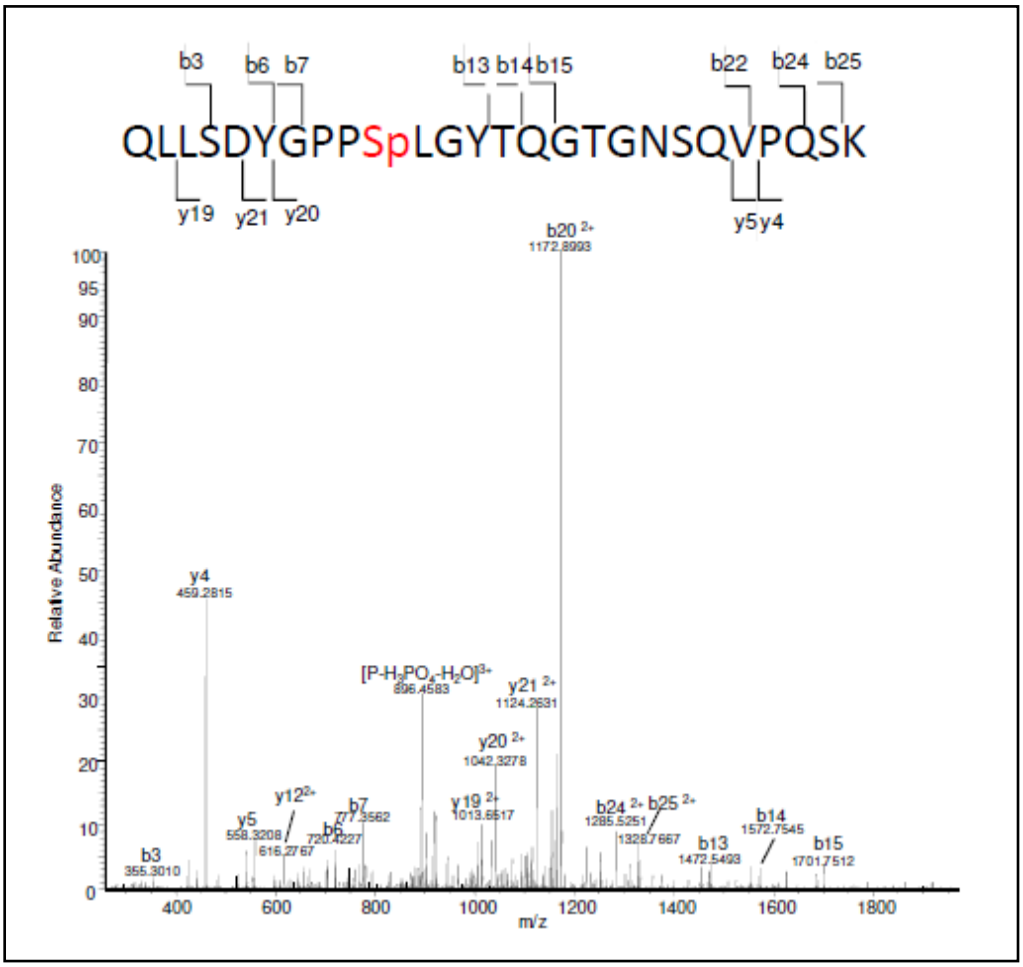
Supplementary Figure S7. MALDI-TOF mass spectrometry disulfide bond mapping results for CDK2AP1(61-115) (a) without DTT and b) with 10 mM DTT. In the absence of DTT (a), peptides containing both reduced and oxidized (circled) cysteine are observed. Upon addition of 10 mM DTT (b), only fragments containing reduced cysteine are observed.



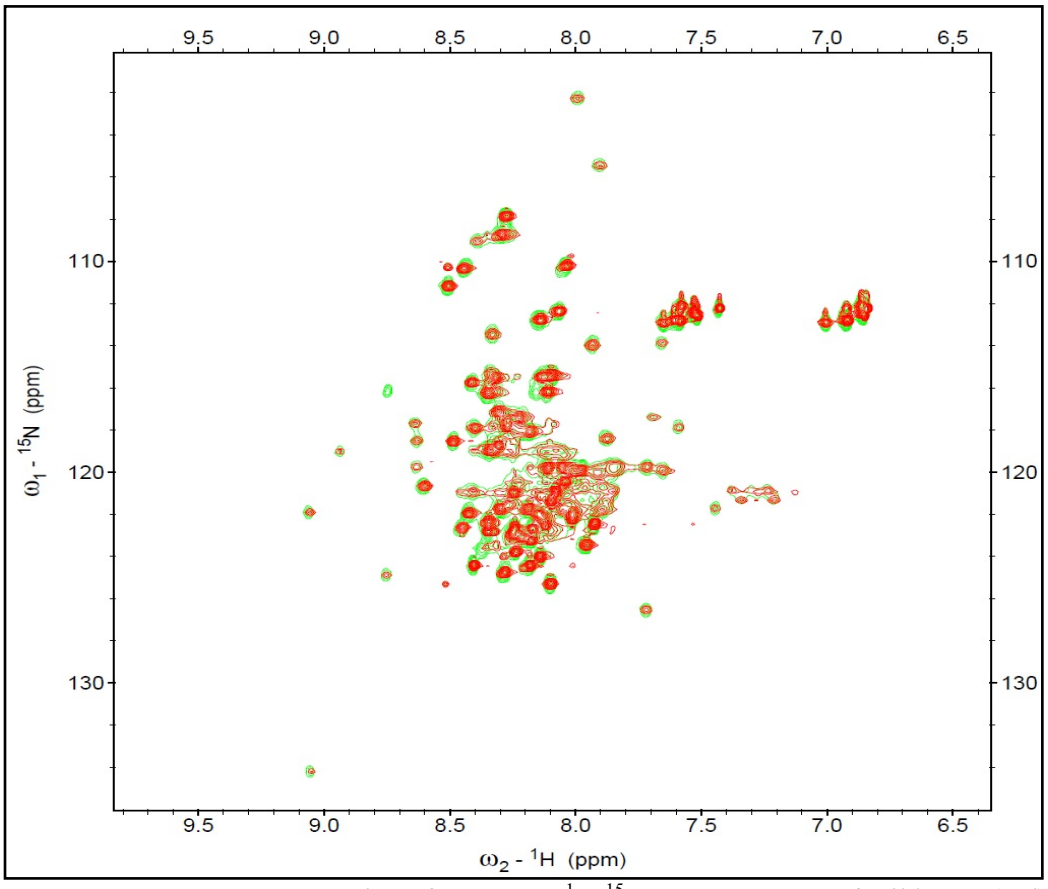
Supplementary Figure S8. Overlay of 600 MHz ^{1}H-^{15}N HSQC spectra of wt (magenta) and C105A mutant (blue) CDK2AP1(61-115) at 25 °C. The sequence-specific ^{1}H-^{15}N resonance for the wt protein are labeled. Samples were prepared at ~ 0.7 mM protein concentration in 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% ^2H_2O, and 50 µM DSS.



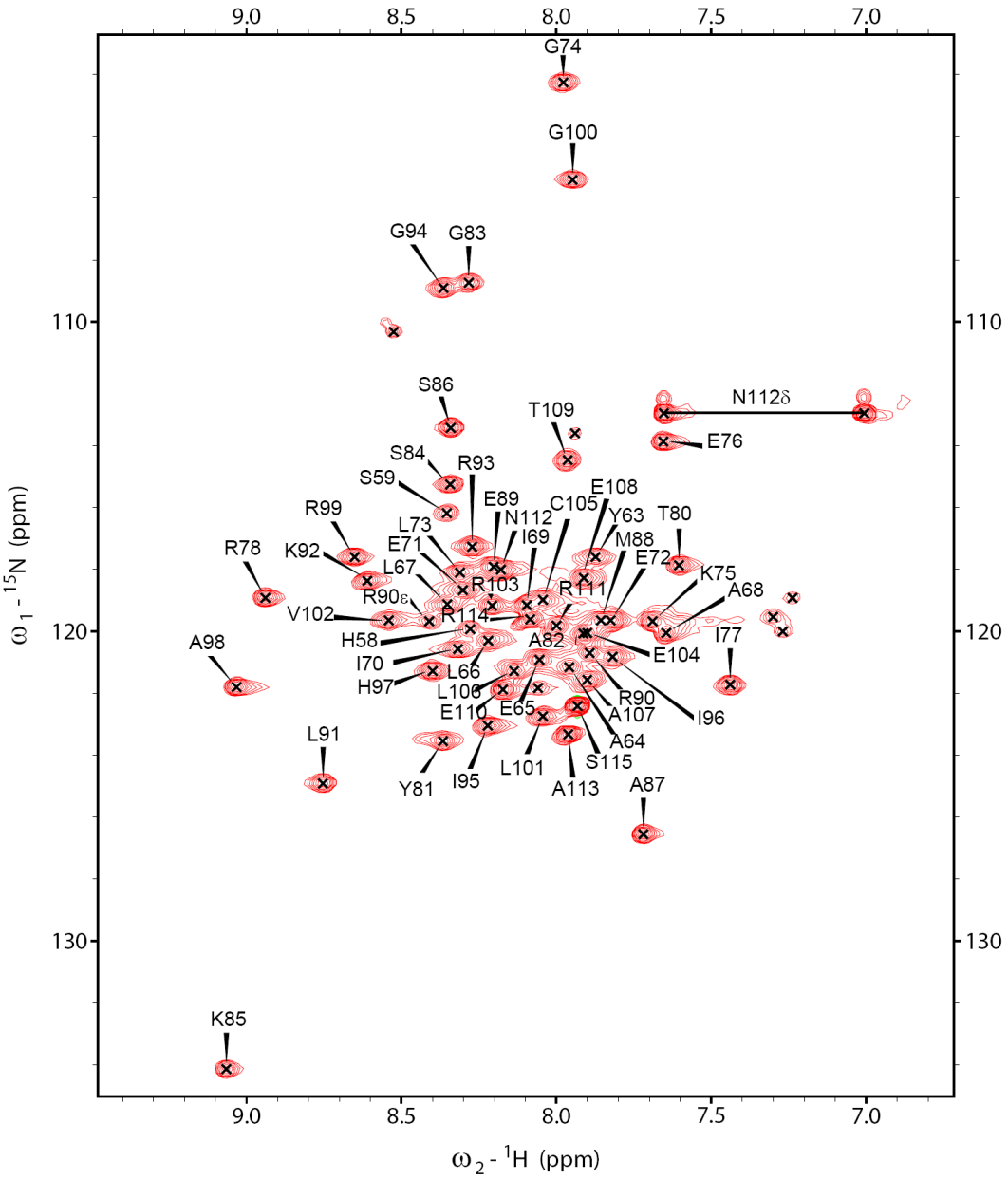
Supplementary Figure S9. The LC/MS/MS chromatograms of peptides after trypsin digestion of phosphorylated (reacted with IKKε) and control (unreacted) samples. Chromatograms of non-phosphorylated peptide from control (a) and reaction samples (b) and phosphorylated peptide from control (c) and reaction sample (d) proteins are shown. The red arrows indicate the retention times for the non-phosphorylated and phosphorylated peptide.



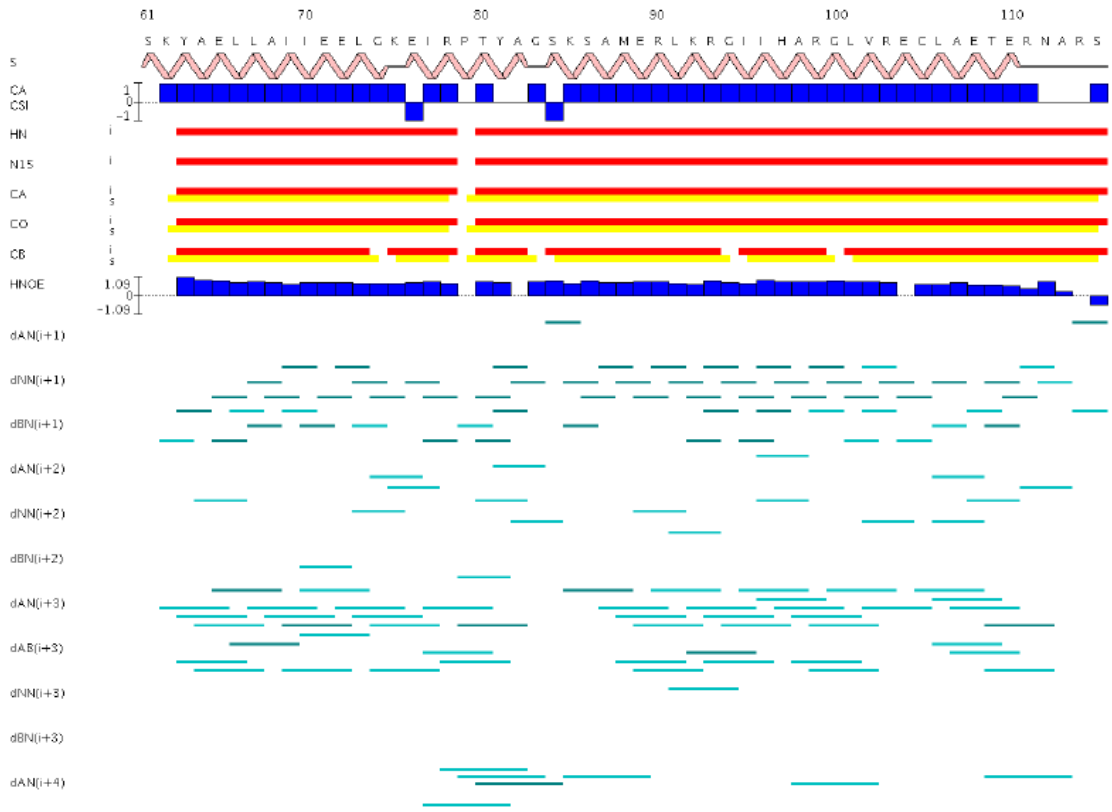
Supplementary Figure S10. MS/MS spectrum of QLLSDYGPPSpLGYTQGTGNSQVPQSK peptide. S46 is identified as the only phosphorylated site after the IKKε phosphorylation reaction.



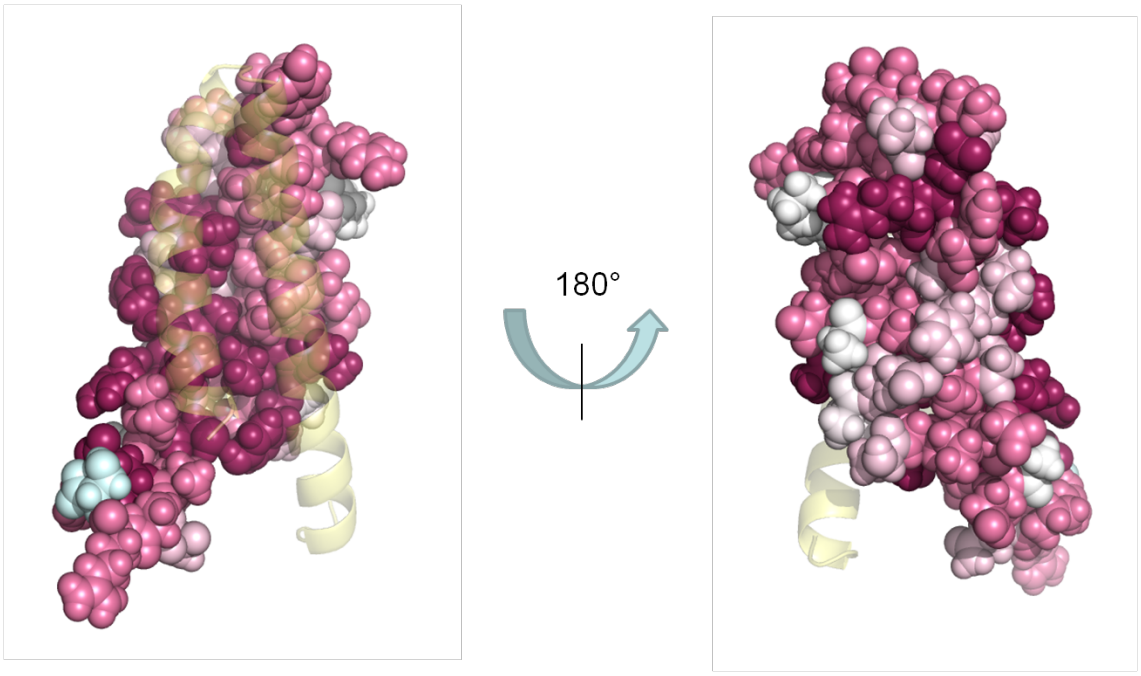
Supplementary Figure S11. Overlay of 600 MHz ^{1}H-^{15}N HSQC spectra of wild-type (red) and phosphorylated (green) full-length CDK2AP1 at 25 °C. Samples were prepared at ~ 0.1 mM protein concentration in 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% $^{2}H_2O$, and 50 µM DSS,.



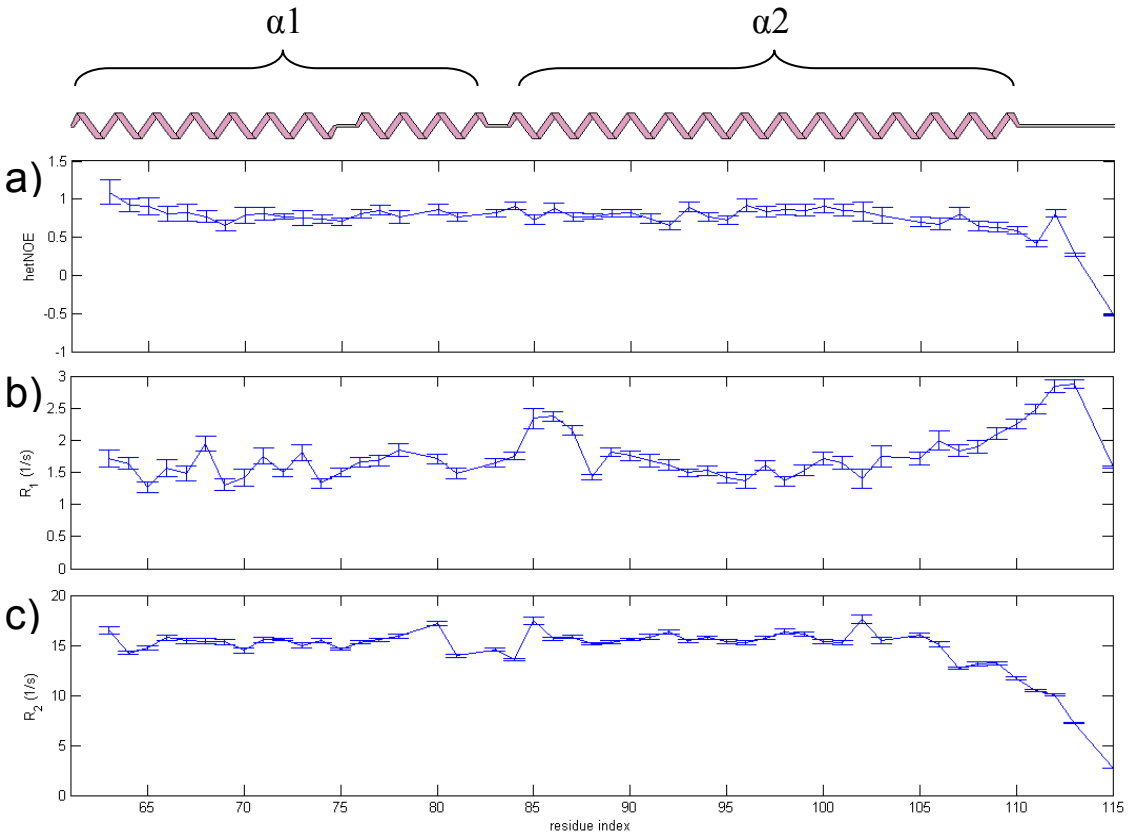
Supplementary Figure S12. Two dimensional 800 MHz ^{1}H-^{15}N HSQC spectrum of CDK1-AP1(61-115) at 25 °C. The sample was prepared at ~ 0.9 mM protein concentration in 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% $^{2}H_2O$, and 50 µM DSS,.



Supplementary Figure S13. Triple resonance NMR connectivity map (6) and summary of other NMR data used to determine resonance assignments and secondary structure for CDK2AP1(61-115). Triple resonance intraresidue (i) and sequential (s) connectivities for matching intraresidue and sequential C', C^{α} and C^{β} resonances are shown as horizontal red and yellow lines, respectively. Bar graphs of Chemical Shift Index (CSI) and ^{1}H-^{15}N heteronuclear NOE data are shown in blue.



Supplementary Figure S14. Consurf analysis based on sequence conservation within the same Pfam domain family (Pfam Id: PF09806) are shown on a single chain of CDK2AP1(61-115) in space filling representation. The darker colors indicate higher level of conservation. The second chain of the dimer is shown in transparent gold color.



Supplementary Figure S15. ^{15}N relaxation rates R_1 (a), R_2 (b) and ^{1}H-^{15}N heteronuclear NOEs (c) of backbone amides for CDK2AP1(61-115) at 600 MHz and 25 °C . The sample was prepared at ~ 0.9 mM protein concentration is 20 mM MES buffer at pH 6.5, containing 0.02% NaN_3, 10 mM DTT, 5 mM $CaCl_2$, 200 mM NaCl, 1x Protease Inhibitors, 10% $^{2}H_2O$, and 50 µM DSS,. The secondary structure elements are shown above.

Supplementary References

1. Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, B. T., and Nilsson, B. (1996) *J Biomol NMR* **7**, 131-141
2. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J BioMol NMR* **6**, 277-293
3. T. D. Goddard, D. G. Kneller. *University of California, San Francisco*
4. Aramini, J. M., Huang, Y. J., Swapna, G. V., Cort, J. R., Rajan, P. K., Xiao, R., Shastry, R., Acton, T. B., Liu, J., Rost, B., Kennedy, M. A., and Montelione, G. T. (2007) *Proteins* **68**, 789-795
5. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989) *Biochemistry* **28**, 7510-7516
6. Moseley, H. N., Sahota, G., and Montelione, G. T. (2004) *J Biomol NMR* **28**, 341-355
7. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* **33**, 5984-6003
8. Li, Y. C., and Montelione, G. T. (1993) *J Magn Reson, Ser B* **101**, 315-319
9. Kontaxis, G., Clore, G. M., and Bax, A. (2000) *J Magn Reson* **143**, 184-196
10. Liu, Y., and Prestegard, J. H. (2009) *J Magn Reson* **200**, 109-118
11. Cierpicki, T., and Bushweller, J. H. (2004) *J Am Chem Soc* **126**, 16259-16266
12. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J Mol Biol* **273**, 283-298
13. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* **319**, 209-227
14. Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Formankay, J. D., and Kay, L. E. (1994) *J Magn Reson, Ser B* **103**, 197-201
15. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J Biomol NMR* **13**, 289-302
16. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D* **54**, 905-921
17. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M., and Nilges, M. (2003) *Proteins* **50**, 496-506
18. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) *Nature* **356**, 83-85
19. Sippl, M. J. (1993) *Proteins* **17**, 355-362
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J Appl Crystallogr* **26**, 283-291
21. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437-450
22. Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007) *Proteins* **66**, 778-795
23. Huang, Y. J., Powers, R., and Montelione, G. T. (2005) *J Am Chem Soc* **127**, 1665-1674
24. Zweckstetter, M., and Bax, A. (2000) *J Am Chem Soc* **122**, 3791-3792
25. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc Natl Acad Sci U S A* **98**, 10037-10041