Supporting Information

Structural basis for PHD_vC5HCH_{NSD1}-$C2HR_{Nizp1}$ interaction: implications for Sotos Syndrome

Andrea Berardi [1,2$,], Giacomo Quilici[1$], Dimitrios Spiliotopoulos[1§], Maria Angeles Corral-Rodriguez[1], Fernando Martin-Garcia[1], Massimo Degano[2], Giovanni Tonon[3], Michela Ghitti[1], and Giovanna Musco[1]

[1] Biomolecular NMR Unit, Division of Genetics and Cell Biology, IRCCS S. Raffaele Scientific Institute, Milan, 20132, Italy

[2] Università degli Studi di Milano. Italy

[3] Biocrystallography, IRCCS S. Raffaele Scientific Institute, Milan, 20132, Italy

[4] Department, Institution, Functional genomics of cancer, IRCCS S. Raffaele Scientific Institute, Milan, 20132, Italy

* To whom correspondence should be addressed. Tel: 0039+0226434824; Fax: 0039+0226434153; Email: musco.giovanna@hsr.it

[‡]The authors wish it to be known that, in their opinion, the first 3 authors should be regarded as joint First Authors

[§]Present Address: Dimitrios Spiliotopoulos, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057

Materials and Methods

Sequence alignment

Sequence search using mouse PHD_vC5HCH_{NSD1} (residues E2117 to P2211) was done using Blast+ (v.2.2.9+) and the non-redundant database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/), In the first round search, performed with blastp (e-value=0.0001), we retrieved 1000 sequences. Alignment of the sequences was done with ClustalO (default options). Sequences belonging to the same organism with a 95% of identity were removed. Visual inspection and sequence removing was done with Jalview. Resulting multiple sequence alignment was used as profile for psi-blast (e-value=0.0001). Several rounds of psi-blast were done, removing redundant sequences (95%) until no new sequence was found. The same protocol was applied to perform $C2HR_{Nizp1}$ alignment. Sequence search was done using mouse $C2HR_{Nizp1}$ (residues V398 to K434).

Histone overlay assays

MODified™ Histone Peptide Arrays were purchased from Active Motif. They enable screening in a single experiment of 59 acetylation, methylation, phosphorylation and citrullination modifications on the entire N-terminal tails of histones H2A, H2B, H3 and H4. A series of synthetic 19mer histone H2A, H2B, H3 and H4 peptides, each of which may contain as many as four modifications, are spotted in duplicate onto a glass slide, generating a total of 384 unique histone modification combinations. Following overnight blocking at 4 °C with 5% milk in TTBS buffer (10 mM Tris/HCl pH 7.4, 150 mM NaCl, 0.05% Tween 20), the array was washed twice with TTBS and once with binding buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 0.1% NP-40, protease inhibitors). The array was then incubated, for 2–4 h at room temperature, with 1 μM solution of GST-tagged PHDv-$C5HCH_{NSD1}$ in binding buffer. After three washes with binding buffer, the array was incubated with primary antibody anti-GST (1 : 1000) in 5% milk/TTBS for 1 h at room temperature. The array was then washed three times with TTBS and incubated for 1 h at room temperature with a secondary antibody HRP-conjugated (1 : 10000) in 5% milk/TTBS. Three washes with TTBS followed, and ECLTM Western Blotting detection solution (GE Healthcare) was added and incubated on the array surface for 5 min

at room temperature. The image was captured by the ImageQuantTM ECL image analysis system (GE Healthcare).

Table S1: Summary of Sotos Syndrome missense-mutations targeting PHD_vC5HCH_{NSD1}

Domain	Mutations
PHD_V	C2124R (1); Y2142N (1); H2143E (3); H2143Y (1); H2143Q (1); C2146R(1); R2152Q (2); C2159Y (1); H2162R (1)
C5HCH	C2164R (1); C2164Y (1); C2167R (3); C2178R (1); F2182I (1); C2183S (1);H2205R (1); C2178Y (1)

1. Tatton-Brown,K., Douglas,J., Coleman,K., Baujat,G., Cole,T.R., Das,S., Horn,D., Hughes,H.E., Temple,I.K., Faravelli,F., et al. (2005) Genotype-phenotype associations in sotos syndrome: An analysis of 266 individuals with NSD1 aberrations. *Am. J. Hum. Genet.,* **77,** 193-204.
2. Melchior,L., Schwartz,M. and Duno,M. (2005) dHPLC screening of the NSD1 gene identifies nine novel mutations--summary of the first 100 sotos syndrome mutations. *Ann. Hum. Genet.,* **69,** 222-226.
3. Kurotaki,N., Harada,N., Shimokawa,O., Miyake,N., Kawame,H., Uetake,K., Makita,Y., Kondoh,T., Ogata,T., Hasegawa,T., et al. (2003) Fifty microdeletions among 112 cases of sotos syndrome: Low copy repeats possibly mediate the common deletion. *Hum. Mutat.,* **22,** 378-387.

Table S2: Summary of Ambiguous Interactions Restraints. Active and passive residues for PHD_vC5HCH_{NSD1} and $C2HR_{Nzip1}$ are indicated

PHD_vC5HCH_{NSD1}	Active residues	D2119, F2122-C2124, H2143, L2147, W2157, E2158, M2177, E2204
	Passive residues	E2120, G2125, A2127, Q2129, V2131, K2134, K2135, D2145, N2148, P2153, G2155, K2156, E2176, R2199, T2203, D2206
$C2HR_{Nzip1}$	Active residues	R415, W416, R417, V418, F420, R422
	Passive residues	K403, K412, N419, I421, L424, R425

Table S3: **HADDOCK cluster statistics.** The table reports the HADDOCK score, the electrostatic and van der Waals energy terms, the ambiguous interaction restraint energy term, the empirical desolvation term and the Buried Surface Area associated to each cluster. Clusters were ranked according to the HADDOCK score. Statistics have been calculated on the five lowest HADDOCK score models in each cluster.

	HADDOCK [a.u.]	**Eelec** [Kcal/mol]	**EvdW** [Kcal/mol]	**EAIR** [Kcal/mol]	**Edesolv** [Kcal/mol]	**BSA** [Å^2]
Cluster1	-133.7 ± 4.5	-529.5 ± 33.5	-52.3 ± 4.1	110.6 ± 34.4	11.3 ± 7.4	1539.0 ± 64.0
Cluster2	-114.7 ± 12.7	-490.3 ± 45.6	-47.6 ± 5.6	137.8 ± 13.3	1.8 ± 6.4	1622.5 ± 58.7
Cluster3	-112.2 ± 4.4	-516.6 ± 32.4	-46.1 ± 2.8	174.4 ± 24.6	9.6 ± 9.3	1413.3 ± 51.7
Cluster4	-105.9 ± 14.7	-350.7 ± 52.0	-60.1 ± 5.7	124.5 ± 15.2	16.3 ± 5.2	1652.7 ± 162.4
Cluster5	-105.1 ± 5.4	-418.6 ± 35.9	-45.6 ± 5.1	114.5 ± 26.2	14.2 ± 9.4	1505.1 ± 87.0
Cluster6	-99.1 ± 4.9	-404.6 ± 45.1	-45.1 ± 4.4	171.2 ± 7.17	9.3 ± 5.6	1354.4 ± 49.8
Cluster7	-94.4 ± 5.4	-390.7 ± 77.6	-46.1 ± 4.6	120.6 ± 16.5	9.4 ± 6.6	1505.5 ± 51.3
Cluster8	-94.7 ± 5.6	-375.6 ± 46.9	-44.4 ± 7.0	100.1 ± 19.2	20.1 ± 10.4	1369.5 ± 80.5
Cluster9	-92.7 ± 6.4	-363.8 ± 48.9	-47.4 ± 9.4	152.0 ± 25.4	21.8 ± 9.8	1413.8 ± 107.4
Cluster10	-81.7 ± 2.1	-415.6 ± 54.5	-33.8 ± 6.7	217.1 ± 17.7	18.9 ± 8.2	1157.8 ± 63.3
Cluster11	-75.6 ± 3.6	-322.2 ± 20.3	-42.2 ± 4.9	194.0 ± 22.7	12.1 ± 4.0	1175.7 ± 86.1

Figure legends

Figure S1 A) Top, alignment of PHD_VC5HCH_{NSD1} with the indicated species. Zn^{2+} binding and conserved hydrophobic residues at the domains interface are shown in red and green, respectively. Sotos mutations described in the literature (*), experimentally tested Sotos mutations (■) and rationally designed mutations (:) are indicated in the alignment. Bottom, PHD_vC5HCH alignment of NSD2 and NSD3. B) Sequence alignment of $C2HR_{Nizp1}$ with the corresponding sequence in other species. Zn^{2+} binding residues and hydrophobic conserved residues are highlighted in red. The RWR signature is highlighted in blue and the Arginine of the C2HR signature is highlighted in bold. Mutations generated to validate the complex model are indicated with ":" An alignment search performed on the domain search (blast+) reveals that this domain is only present in mammals.

Figure S2: Backbone dynamics (R2/R1, R1, R2) of A) PHD_v-$C5HCH_{NSD1}$, free (black line) and in complex with $C2HR_{Nizp1}$ (red line) and of B) $C2HR_{Nizp1}$, free (black line) and in complex with PHD_vC5HCH_{NSD1} (red line). Notably, residue $R425_{Nizp1}$ has a very high R2 rate, it is located at the end of the α-helix and is most likely affected by *fraying* effects of the α-helix that reflect into conformational exchange and line broadening. Conceivably, the conformational exchange observed for R425 is due to the absence of the fourth Zn^{2+} binding residue in position 427, thus reducing by one helical turn the α-helix.

Figure S3: Structural comparison with other PHD tandem domains. A) Superposition of PHD_vC5HCH_{NSD1} structure (blue, lowest energy structure from the NMR bundle) and PHD_vC5HCH_{NSD3} (red, PDB code:4GND). The side-chains of $P1355\text{-}P1356_{NSD3}$ and of $R2152\text{-}P2153_{NSD1}$ are indicated in sticks. B) Superposition of PHD12 tandem domain (orange) of human DPF3b (2KWN) with PHD_vC5HCH_{NSD1} (blue), C) Superposition of PHD12 tandem domain (grey) of human MOZ (2LN0), with PHD_vC5HCH_{NSD1} (blue). Zn^{2+}ions are represented with spheres. D) Comparison between the electrostatic surfaces of PHD_vC5HCH_{NSD1} and PHD_vC5HCH_{NSD3} in complex with H3K9me3 (blue cartoon). E) $^1H\text{-}^{15}N$ HSQC spectra (top) and 1H-1D NMR spectra of WT and mutants PHD_vC5HCH_{NSD1}.

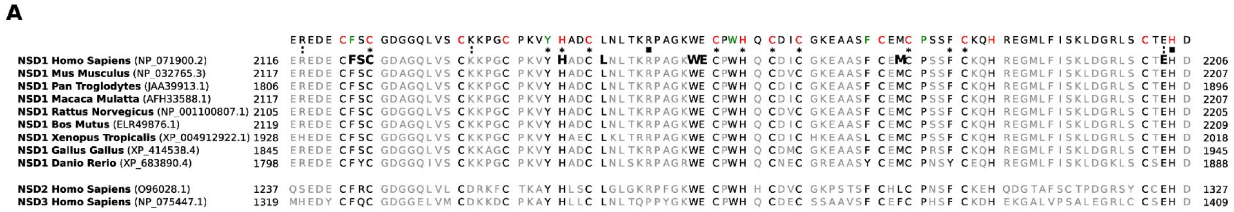
Figure S4

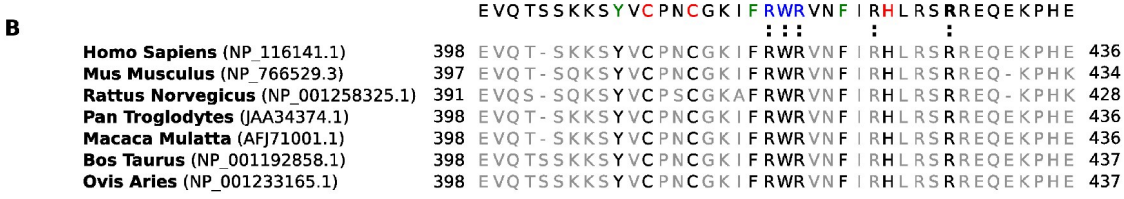
A) ^1H-1D NMR spectra of $C2HR_{Nizp1}$ wild-type and mutants showing that all mutants are well folded. B) Integral of the radial distribution function g(r) of the Sulfur (Sγ of C407 and C410), of the Nitrogen (Nε of H423) and of the water Oxygen (O) atoms around the Zn^{2+} ion during the QM/MM analysis (100 ps). A maximum radius of 5 Å is showed in order to simplify the visualization of the result. The analysis reveals that within a radius of 2.2 Å Zn^{2+} is stably tetra-coordinated by two Sγ atoms, by one Nε and one O atom.

Figure S5 A) Superposition of $^1H\text{-}^{15}N$ spectra of PHD_vC5HCH_{NSD1} in complex with a two-fold excess of $C2HR_{Nizp1}$ (black) and upon addition of a twelve fold excess of $H3_{1\text{-}21}$ to the complex (red). B) Superposition of $^1H\text{-}^{15}N$ spectra of PHD_v-$C5HCH_{NSD1}$ without (black) and with a two-fold excess of W416A-$C2HR_{Nizp1}$ mutant.

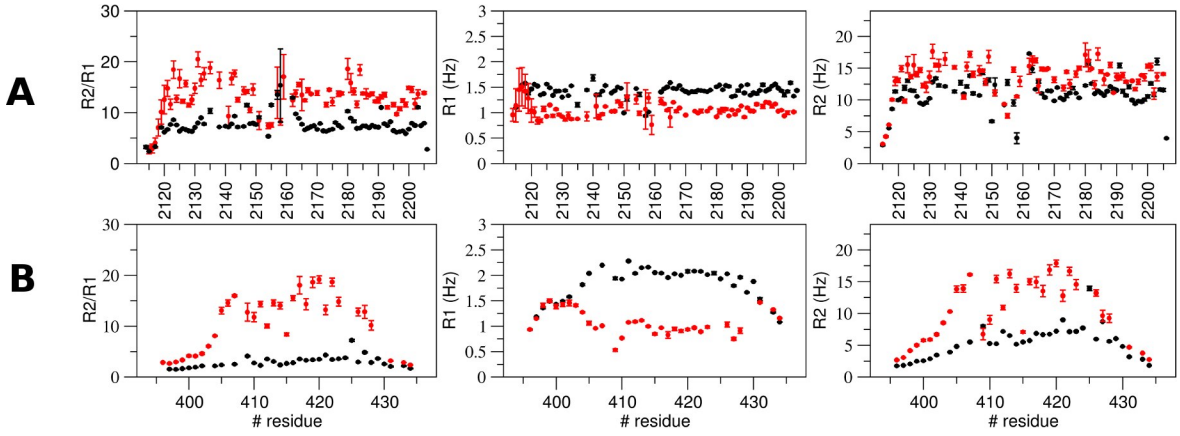
Figures S6 A) Interaction of GST-PHD_VC5HCH_{NSD1} (top) and GST control (bottom) with histone peptide arrays. Experiments have been performed in duplicate. The Histone modification states giving the strongest positive signals involved histone H3 (1-19) with methylation of K4, K9, R8. The identity of all the spots is reported in the Active Motif company website (http://www.activemotif.com). B) Modification preference

(defined as percentage^{PTM}=(hist^{PTM} /his^{tot})*100) as calculated by Array Analyses Software based on two duplicated arrays. C) Superposition of ^{1}H-^{15}N HSQC spectra of PHD$_{v}$-C5HCH$_{NSD1}$ alone (black) and upon addition of twelve-fold excess (red) of H3K4$_{1-21}$, H3K4me3$_{1-21}$, H3K9me3$_{1-21}$-D). Histograms showing the average backbone chemical shift perturbations (CSP) observed in ^{15}N-labelled PHD$_{v}$C5HCH$_{NSD1}$ (0.2mM) upon addition of a twelve-fold excess of H3K4$_{1-10}$, H3K4$_{1-37,}$ H3R8me2$_{1-21}$ and upon addition of 300mM Lysine (K). Titrations with H3R8me2$_{1-21}$ gave a similar profile as the other titrations with 21 amino-acids long histone H3 peptides (see Figure 3A in main text). E) Histograms showing the average backbone chemical shift perturbations (CSP) observed in ^{15}N-labelled R2117A_K2134D PHD$_{v}$C5HCH$_{NSD1}$ mutant (0.2mM) upon addition of a twelve-fold excess of H3K4$_{1-21}$.

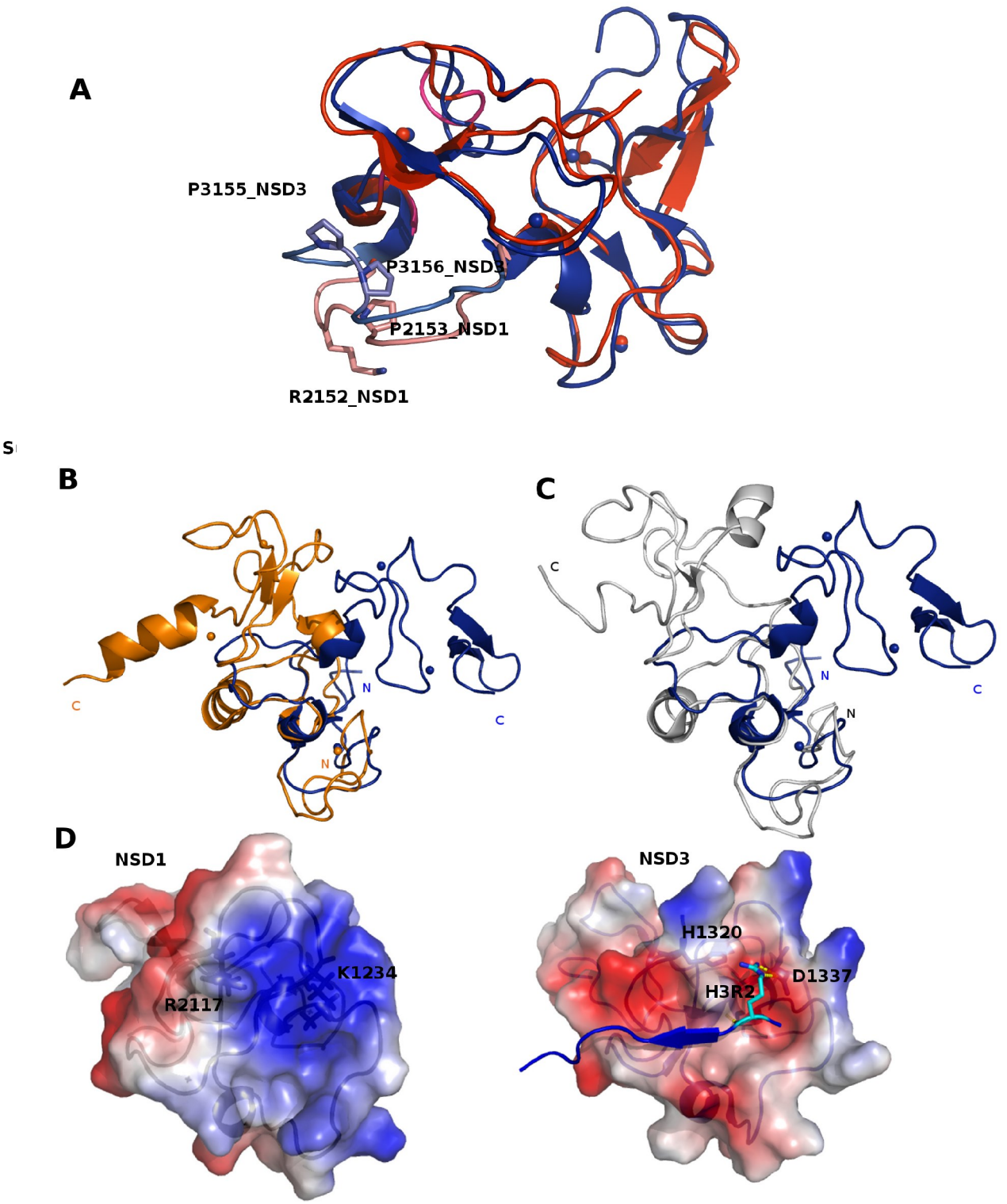




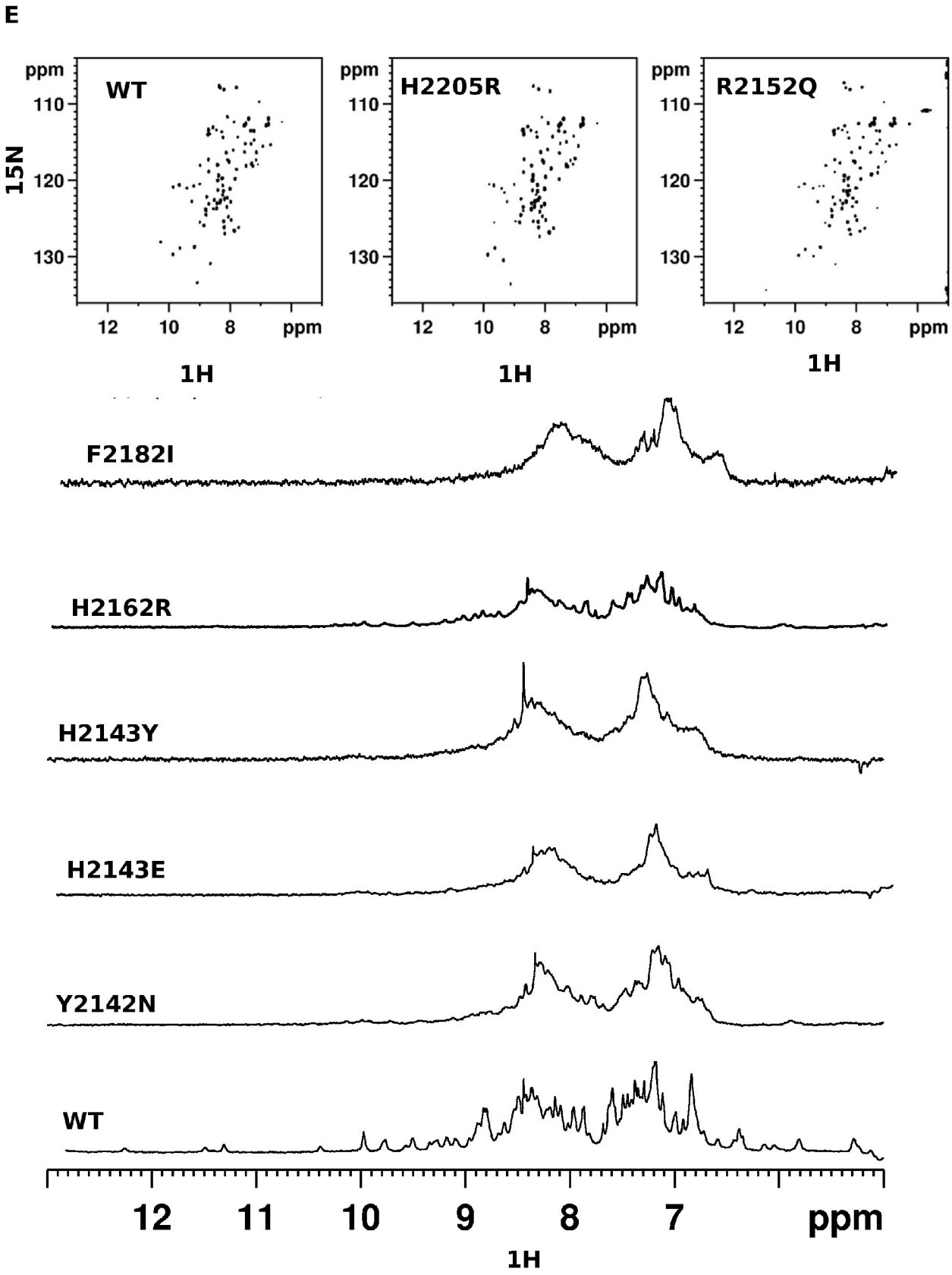
Supplementary Figure S1



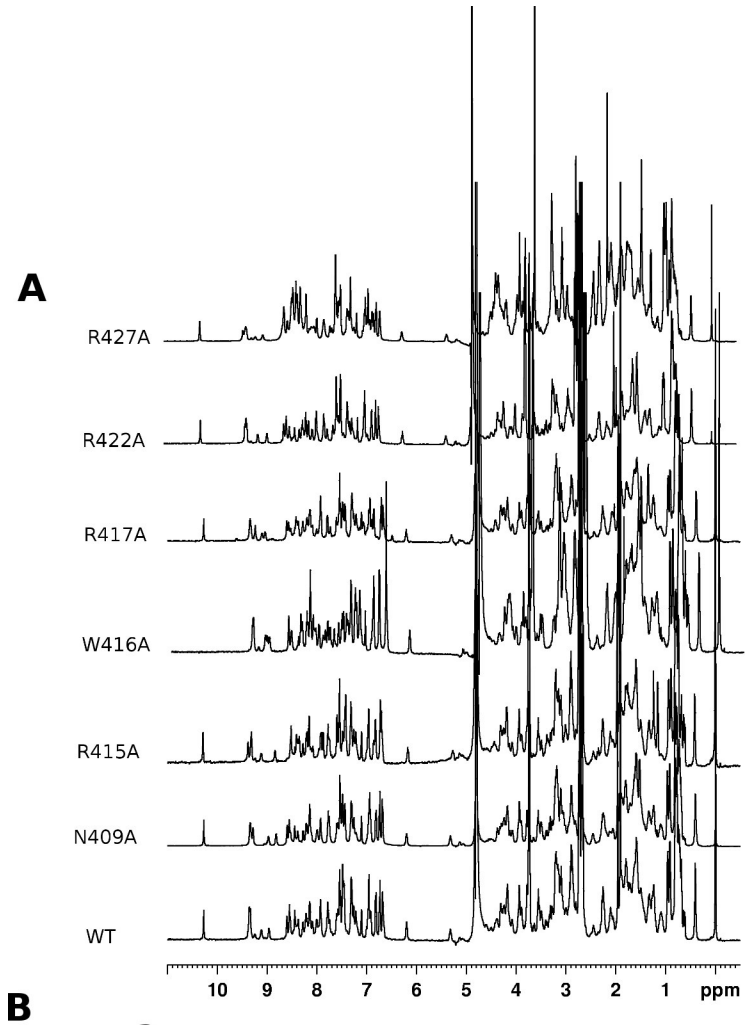
Supplementary Figure S2

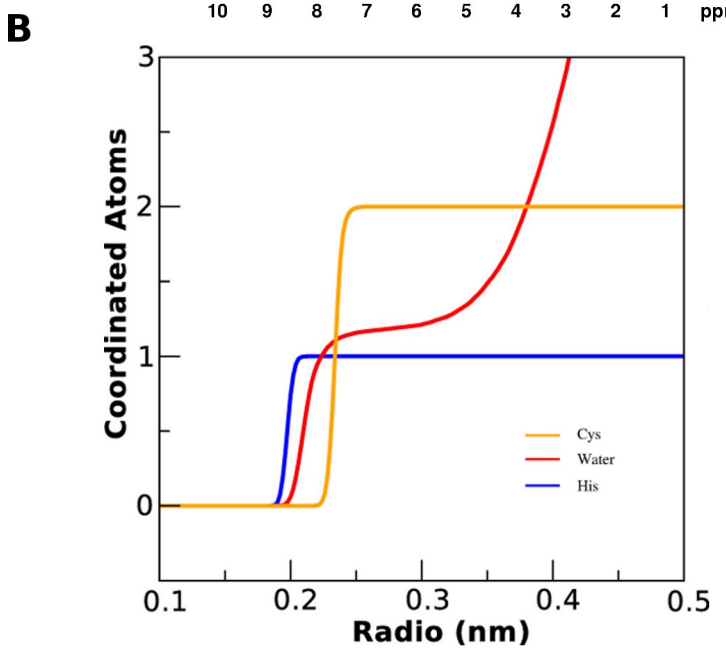


Supplementary Figure S3

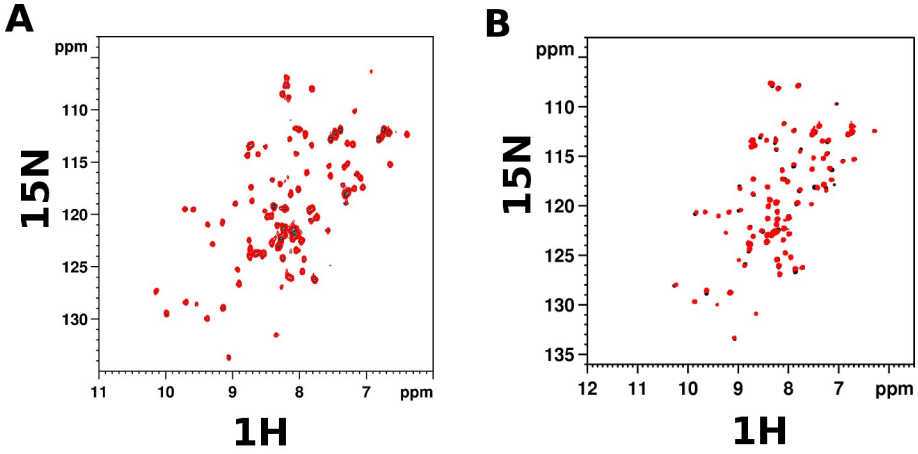


Supplementary Figure S3

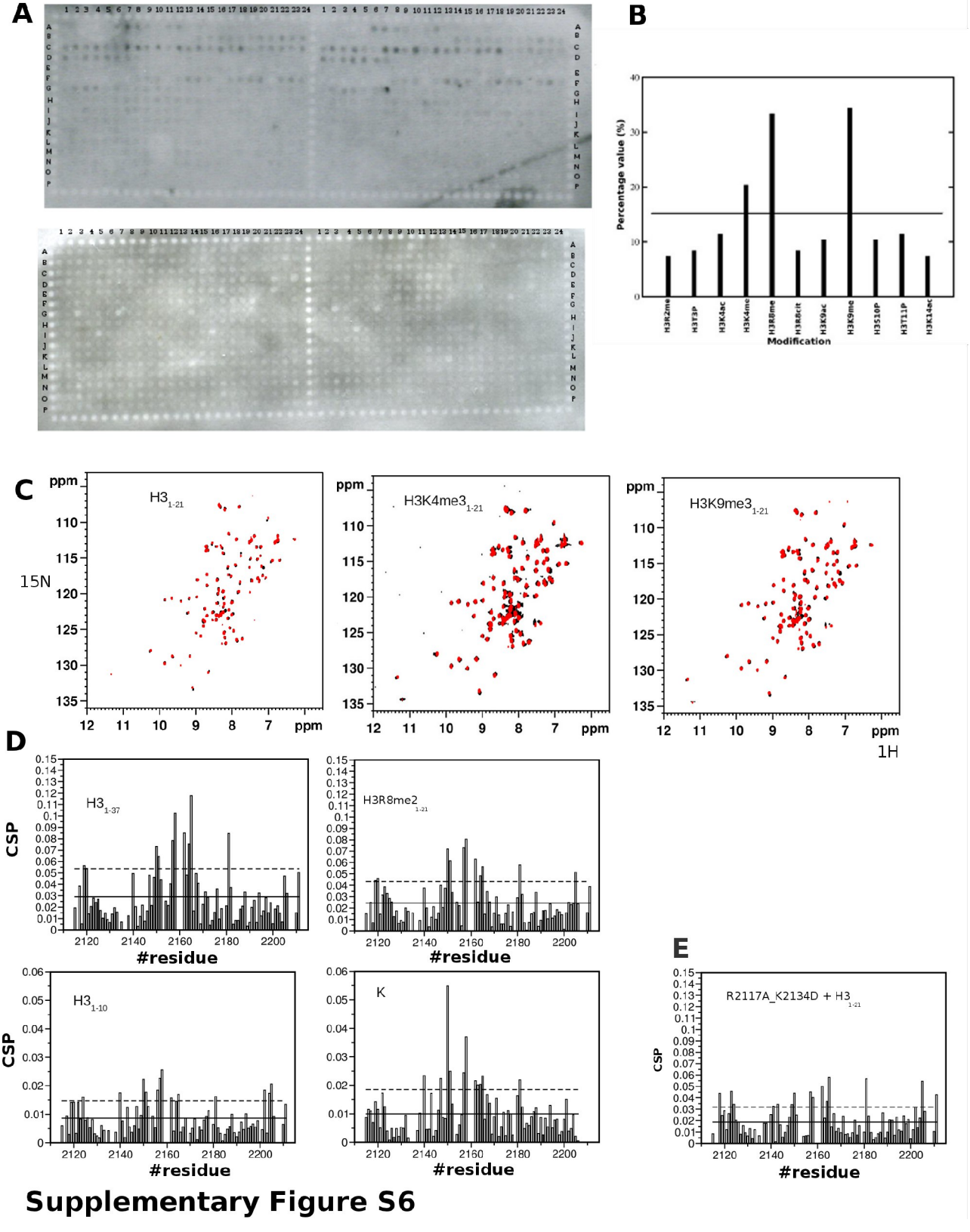




Supplementary Figure S4



Supplementary Figure S5



Supplementary Figure S6