SUPPLEMENTAL INFORMATION

Competitive binding of a benzimidazole to the histone-binding pocket of the Pygo PHD finger

Thomas C. R. Miller[1,3], Trevor J. Rutherford[1,4], Kristian Birchall[2], Jasveen Chugh[2], Marc Fiedler[1,4,5] and Mariann Bienz[1,5]

[1] MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

[2] MRC Technology, 1-3 Burtonhole Lane, Mill Hill, London, NW7 1AD, UK

SUPPLEMENTARY METHODS

In silico screening

For the initial round of virtual screening, the crystal structure of the ternary complex between Pygo1 PHD, BCL9 HD1 and H3K4me2 (2VPE; (*1*) was prepared for docking studies using the Schrodinger molecular modelling suite of software (http://www.schrodinger.com). This involved solvent removal, valence and charge assignment, addition of hydrogen atoms, orientation of ambiguous groups (e.g. amide groups of asparagine and glutamine) and restrained minimization to relieve residual crystallographic strain. To investigate targeting of individual histone-binding pockets, two docking grids were generated, one against the K4me pocket and one against the A1 pocket. The K4me2 pocket grid has a 12 Å inner box centred on T3 and K4me2, with a hydrogen bond constraint set against D352 carboxyl, and a hydrophobic constraint set for the aromatic cage between W366 and Y341. The A1 pocket grid has a 12 Å inner box centred on A1 and with hydrogen bonding constraints to L382, E385 and A388 peptide carbonyl groups, T383 hydroxyl group and E360 carboxyl group. A library containing 225K compounds derived from property and substructure filtering of commercial supplier catalogues had been compiled previously and prepared for docking at MRCT. These were docked flexibly using Glide (*2*) in a hierarchical virtual screening workflow, involving three successive rounds of docking with increasing levels of sophistication (HTVS, followed by SP, followed by XP modes) such that only the best compounds from each round were fed into the next. When docking to the K4me pocket, both constraints were required for compounds to be retained. When docking to the A1 pocket, >1/5 constraints was required. This resulted in 130 compounds shortlisted against the K4me pocket and 280 against the A1 pocket, from which 313 were eventually purchased. A single series of three A1 pocket compounds was subsequently confirmed as binders by NMR (see main **Fig. 2A, B**), corresponding to a hit rate of 0.96%.

For the second round of virtual screening, aimed at identifying larger compounds occupying the full site (i.e. bridging A1 and K4me pockets), the docking grid was centred on the first 7 residues of the H3K4me2 15-mer peptide in the ternary crystal structure (*1*), and the inner box was expanded to 14 Å. Hydrogen bonding constraints were set for L382, E385, A388 and Q354 peptide carbonyl groups, D352 and E360 carboxyl groups, T383 and Y379 hydroxyl groups, and L358 peptide backbone nitrogen. A hydrophobic constraint was set in the vicinity of I357, as occupied by the methyl group of histone H3T3. From a larger 850K updated version of the commercial library at MRCT, 154K compounds with molecular weight 450–550 were docked into this grid using Glide in a hierarchical virtual screening workflow as outlined above, with requirement to fulfil >3/10 constraints. 47 of the top scoring 100 were purchased, resulting in 6 new compounds with detectable though not improved binding (IS16–18 and IS28; **Fig. S1**), corresponding to a hit rate of 12.8%.

For the third round of screening, the structure of the paralog PHD-HD1 complex (2XB1; (*3*) was used, and the strategies for docking and shortlisting were varied. The structure was prepared for docking as outlined above, and alignment to the structure of the ternary complex (*1*) allowed superimposition of the histone peptide onto the 2XB1 structure, to achieve comparable definition of the docking grid, with a 12 Å inner box and centroid based on the ARTKme2 peptide. To limit solubility issues encountered during the previous round of screening, the commercial library of 850K compounds was filtered to 600K based on a minimum solubility of 10 μM, as calculated using Pipeline Pilot 8.0 (Accelrys, San Diego, USA; http://accelrys.com). These were docked into the Pygo2 grid without constraints, using Glide in HTVS mode to select the top 10K scoring species, which were then re-docked and re-scored using the more refined Glide SP method. The results were post-processed to identify those molecules capable of forming similar interactions to those observed in the ternary complex (*1*). 238 molecules were found to occupy both A1 and K4me pockets, forming at least one hydrogen bond in each and also forming a hydrogen bond to the peptide carbonyl of A343. Further manual shortlisting was based on

various criteria, including Glide ligand efficiency, structural diversity, similarity to compounds already screened and commercial availability. 99 compounds were purchased, 13 of which were confirmed to bind (IS5, IS8, IS12, IS14–15, IS19, IS21–23, IS25 and IS27), corresponding to a hit rate of 13.1%. IS19 appeared to be the strongest binder of the whole IS set, based on the magnitude of its CSPs (**Fig. S2**).

Finally, a fourth round of compound purchase was undertaken, to build up the SAR of existing hits. This involved using a combination of substructure and similarity searching (Pipeline Pilot ECFP_4 fingerprints) in the 850K commercial library, using each of the 13 binders from the third round of virtual screening as a query. Following manual inspection of the resultant hit lists, 30 were purchased, yielding 10 confirmed binders (IS4, IS6–7, IS9–11, IS13, IS20, IS24 and IS26).

Processing of X-ray crystallographic data

X-ray diffraction data were processed with Mosflm (*4*) and scaled with Aimless (*5*) (see **Tables S2 & S3**), and the structures were solved by molecular replacement with Phaser (*6*) based on 2XB1 (*3*), and refined with Refmac (*7*). The models were updated with Coot (*8*), and their stereochemistry was verified with MolProbity (*9*), and analyzed with the CCP4i programs (*10*).

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1

***In silico* hits interacting with PHD-HD1.** (**A**) Structures of 25 IS compounds from the second to fourth rounds of refined *in silico* screens (yielding 29 hits), defining three chemical scaffolds (red,

green, and orange) in addition to that defined by IS1–3 identified in the first virtual screen (see main **Fig. 2B**).

Fig. S2

Heat-maps of IS hits. CSPs elicited by the IS hits shown in **Fig. S1** were projected onto the PHD-HD1 structure (2XB1), revealing different modes of binding to the histone-binding surface (see also main **Fig. 2C**–**F**).

Fig. S3

Heat-maps of selected benzimidazole hits. CSPs elicited by selected CF hits from group III (see main **Fig. 4A**) were projected onto the PHD-HD1 structure (2XB1), revealing two distinct binding sites within PHD-HD1 (the distal K4me pocket in its front face, and the benzothiazole cleft in its rear). Note that CF15 and CF18 (with extended N-alkyl side-groups) retain their interaction with the benzothiazole cleft, whereas those with short N-alkyl groups (CF14, CF16 and CF17) only weakly interact with this cleft (below the threshold used to generate these heat-maps), emphasizing that the K4me pocket is the primary binding site for the latter three benzimidazoles. V327 was note included in these heat-maps (due to its overlap with L383 in the assigned HSQC spectrum of this structure; **Fig. 7A**).

Fig. S4

NOESY spectra. (**A**) Double half-filtered 2D ^{1}H-^{1}H NOESY spectrum (250 ms NOE evolution time) from ^{13}C-^{15}N-labelled hPHD-GSG-HD1 (500 μM) probed with 5 mM CF16, annotated with

assignments to specific H-H contacts. Note that the multiplicity of the observed NOE correlation peaks was independent of whether ^{15}N-decoupling pulses were applied in t_1, indicating that all NOE transfer originates from ^{1}H(^{12}C), and not ^{1}H(^{15}N). (**B**) Double half-filtered 2D ^{1}H-^{1}H NOESY spectra for CF16 (same spectrum as in **A**), compared to control spectra derived from protein-only (500 μM PHD-HD1 in deuterated aqueous phosphate buffer; *middle panel*) or CF16 only (5 mM, in deuterated aqueous phosphate buffer; *right panel*); boxed regions reveal intermolecular NOEs from the CF16-PHD-HD1 complex that are absent in the two control spectra. Cross peak intensities were dependent on NOE mixing time (not shown), as expected for peaks deriving from NOE transfer.

Fig. S5

Compilation of restraints from NOESY spectra and CSPs. *Top*, unambiguous restraints with regard to CF16, derived from H-H intermolecular NOEs (including H5/H6 of CF16 to L381 Hδ and A332 Hβ; an NOE from H5/H6 was unassigned, due to overlap of N338 Hα in the K4me pocket and A332 Hα in the CF cleft); *a, b*, *, pairs of proton resonances that are unresolved in NOE spectra; #, pairs of prochiral or symmetry-related resonances that are not stereospecifically assigned; for each designated pair, no restraint violation energy was applied when either proton was within distance bounds. *Bottom*, list of ambiguous restraints, derived from CSPs more than one standard deviation (0.04 ppm) from the mean CSP for all residues (0.02 ppm) and >50% solvent exposed (see also main **Fig. 7B**). The orientation of CF16 within the K4me pocket, as shown in the NOE-based model (**Fig. 6A**), is favored by the docking simulations, given that it is consistently seen in 199/200 models. Note that this model implies proximity of the ethyl group of CF16 to W353 and the aromatic hydrogens of CF16 to D339 Hα, however, the NOEs corresponding to an alternative orientation (with aromatic hydrogens of CF16 in proximity to aromatic hydrogens on the protein, and the CF16 ethyl group in proximity of aliphatic

hydrogens on the protein) would not be resolvable in our NOESY spectra and, consequently, there are no modelling restraints to these centres. To test whether this lack of resolution of potentially decisive NOEs imposed a bias on the docking, these restraints were entered manually into the HADDOCK calculations, which however did not result in consistent clusters of docking poses and orientations. This argues against the presence of these H-H contacts.

Fig. S6

Interactions of histone H3 tail peptides with PHD-HD1. (**A**) Overlay of HSQC spectra of 40 μM ^{15}N-PHD-HD1 alone (black), and with 0.5 mM ARTKme2Q (red) and (**B**) corresponding heat-map, after projection of CSPs onto the PHD-HD1 structure, with front and rear views as shown in main **Fig. 4B**. Note that the incubation of 40 μM ^{15}N-PHD-HD1 with 0.5 mM histone H3 trimers (ART, TKme2Q) produced no detectable CSPs whatsoever, demonstrating complete lack of binding of these trimers, and defining the ARTKme2Q pentamer as the minimal PHD-interacting histone H3 tail peptide. (**C**) Overlay of HSQC spectra of 43.3 μM ^{15}N-PHD-HD1 alone (black), and 21.6 μM 15-mer H3K4me2 peptide (*1*) (red), i.e. at 2:1 stoichiometry, and (**D**) corresponding heat-map, as in (**B**). Note that although (**A**) and (**C**) are both recorded at conditions of 50% bound protein, the extended 15-mer peptide produces mostly peak broadening (but very little CSPs) while the pentamer produces predominantly CSPs. Nevertheless, the comparison between the two heat-maps indicates that the two histone peptides, despite highly dissimilar binding affinities for PHD-HD1 (see main text), interact essentially with the same set of PHD residues in their cognate surface. Given the potential for an intramolecular interaction between T3 and T6 of the histone H3 tail, as seen in the crystal structure (*1*) (see main text), we tested whether this interaction also forms in solution. However, the Cα and Cβ chemical shifts from the 15-mer peptide are all within 0.2 ppm of random coil values (*11*), with the

exception of the single N- and C-terminal amino acids, which argues against residual structure in the free solvated peptide that might potentially affect binding entropy, relative to that of shorter constructs.

Fig. S7

Titrations of histone peptide and CF16 binding. (**A**) Assigned HSQC spectrum for the complex between PHD from hPygo2 (residues 327–387) and HD1 from hB9L (residues 235–263), as previously shown (*3*). (**B, C**) Titrations of 50 μM ^{15}N-labelled PHD-HD1 alone (blue), and with increasing concentrations of (**B**) ARTKme2Q (red, no peptide; orange, 30 μM; green, 60 μM; blue, 120 μM; purple, 250 μM) or (**C**) CF16 (red, no peptide; orange, 600 μM; green, 900 μM; blue, 2200 μM; purple, 3200 μM); plotting of CSPs against ligand concentration (see main **Fig. 3C**) led to an estimated K_d of 528 ± 32 μM for the pentamer peptide, and of 7.3 mM ± 2.0 for CF16 (**see Table S1**).

Table S1

Affinities of CF compounds for PHD-HD1. K_d values for selected CF hits were derived from HSQC CSPs (either $\Delta\delta^1$H or $\Delta\delta^{15}$N/5, according to which perturbed the greatest) and are recorded as the median value obtained for 5–7 different resonances (see **Fig. 3A**); SE, spread of values measured.

Table S2

Data collection and refinement statistics for 4UP5 (PHD-HD1-CF4)

Table S3

Data collection and refinement statistics for 4UP0 (PHD-HD1-H3K4me2K9ac)

SUPPLEMENTARY REFERENCES

1. Fiedler, M., Sanchez-Barrena, M., Nekrasov, M., Mieszczanek, J., Rybin, V., Muller, J., Evans, P., and Bienz, M. (2008) Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex, *Mol Cell 30*, 507–518.
2. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J Med Chem 47*, 1739–1749.
3. Miller, T. C., Rutherford, T. J., Johnson, C. M., Fiedler, M., and Bienz, M. (2010) Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor, *J Mol Biol 401*, 969–984.
4. Leslie, A. G. (2006) The integration of macromolecular diffraction data, *Acta Crystallogr D Biol Crystallogr 62*, 48–57.
5. Evans, P., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr 69*, 1204–1214.
6. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C., and Read, R. J. (2005) Likelihood-enhanced fast translation functions, *Acta Crystallogr D Biol Crystallogr 61*, 458–464.
7. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr D Biol Crystallogr 53*, 240–255.
8. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126–2132.
9. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12–21.
10. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr 67*, 235–242.
11. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J Biomol NMR 5*, 67–81.

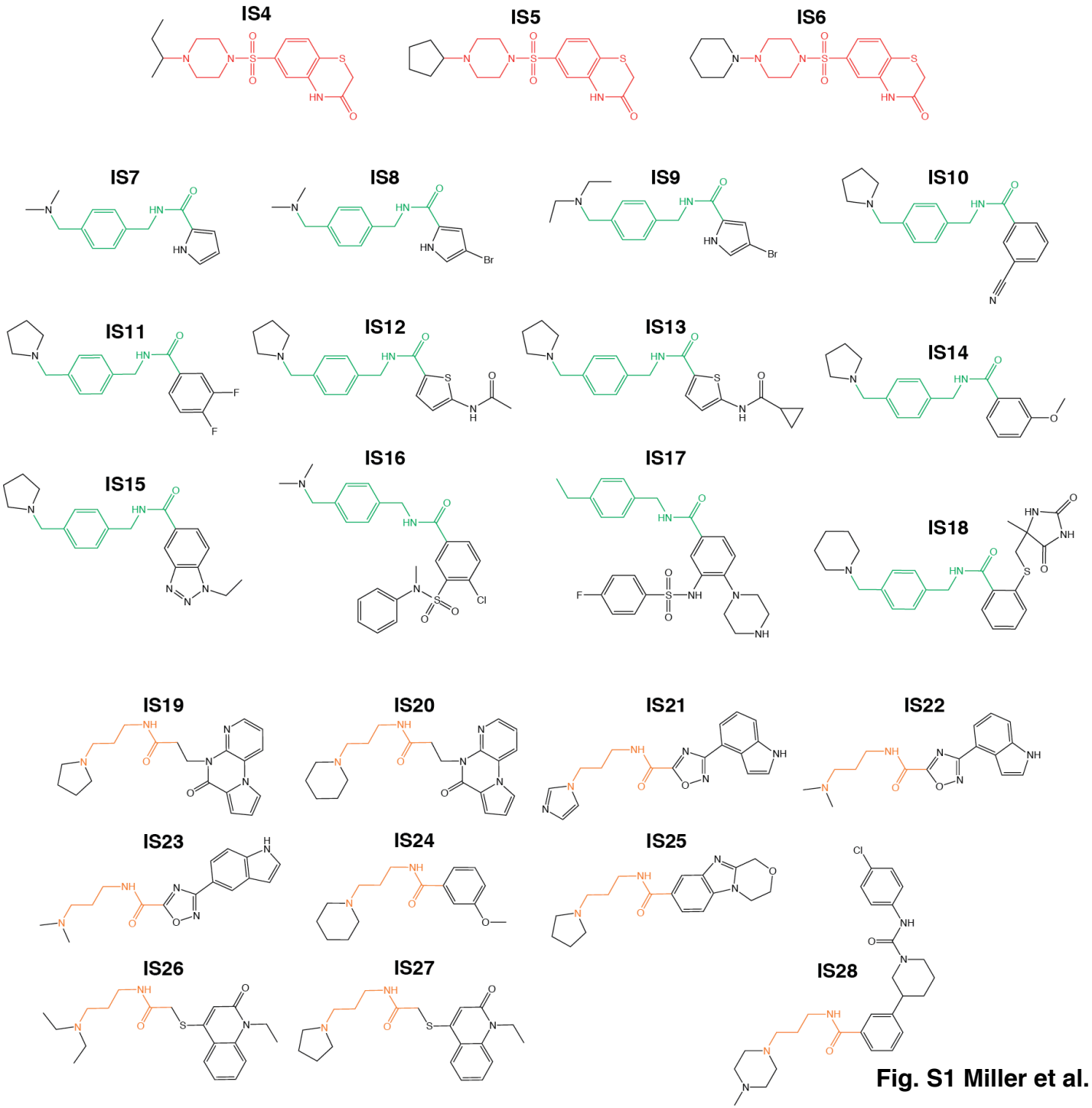


Fig. S1 Miller et al.

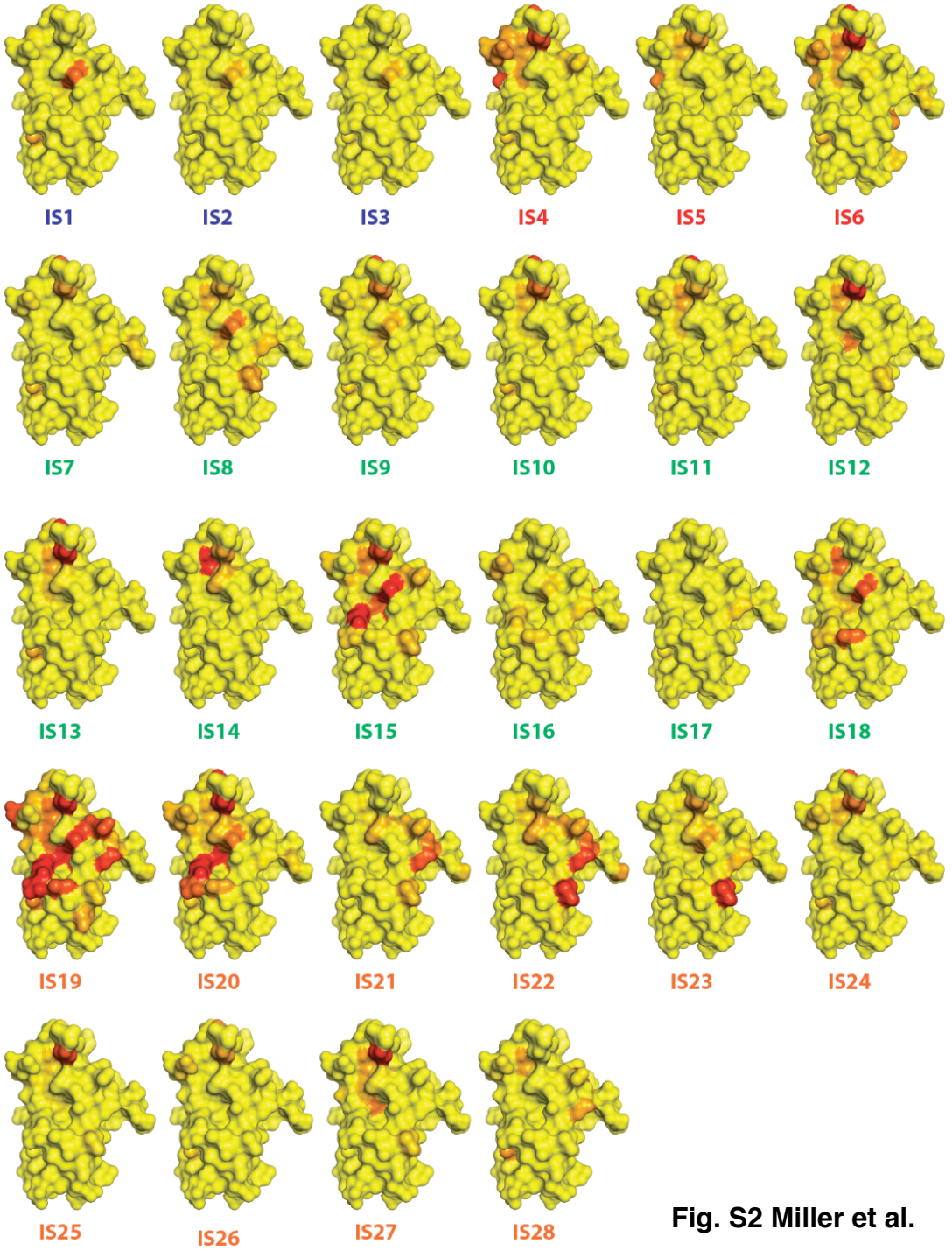


Fig. S2 Miller et al.

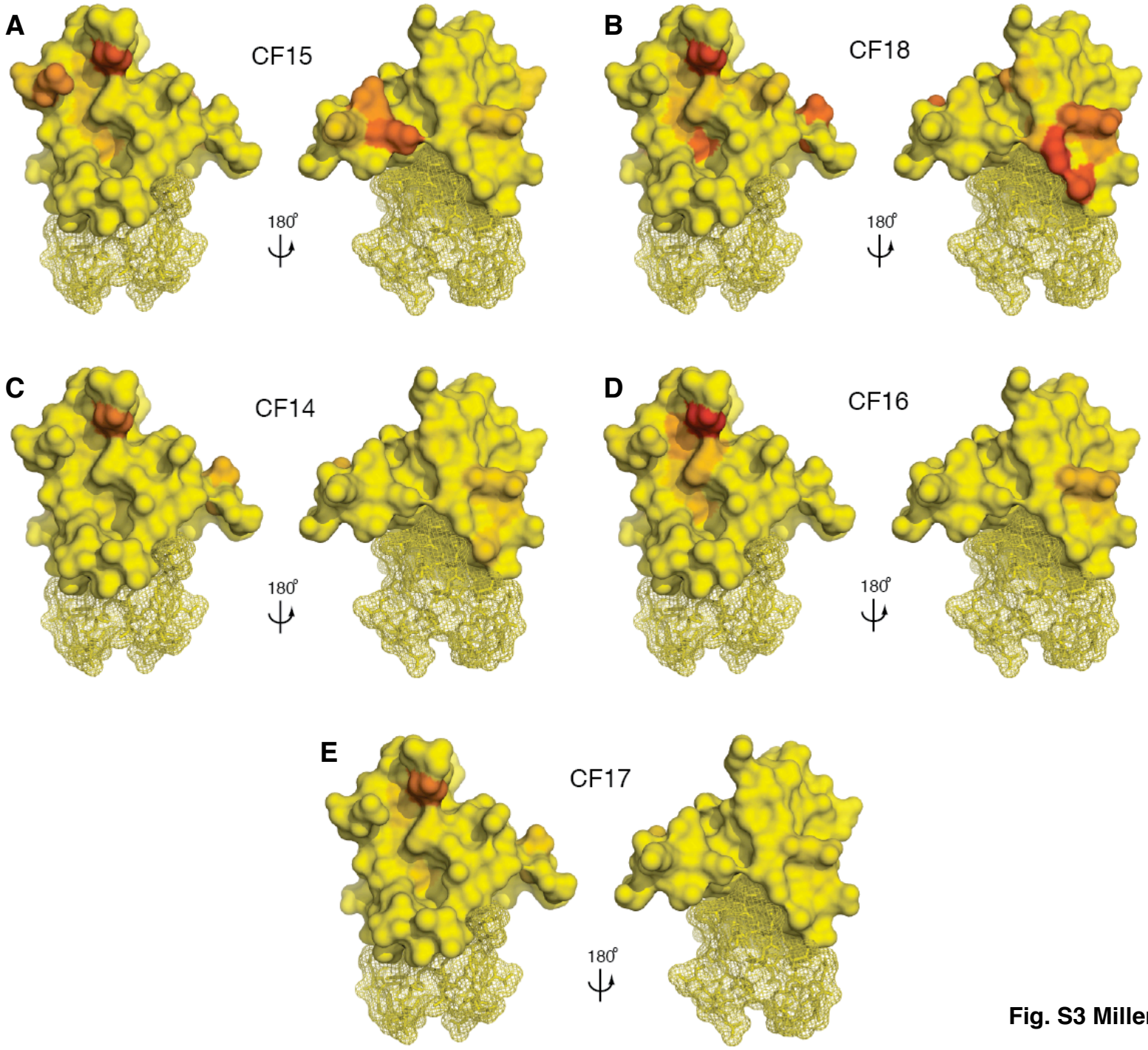


Fig. S3 Miller et al.

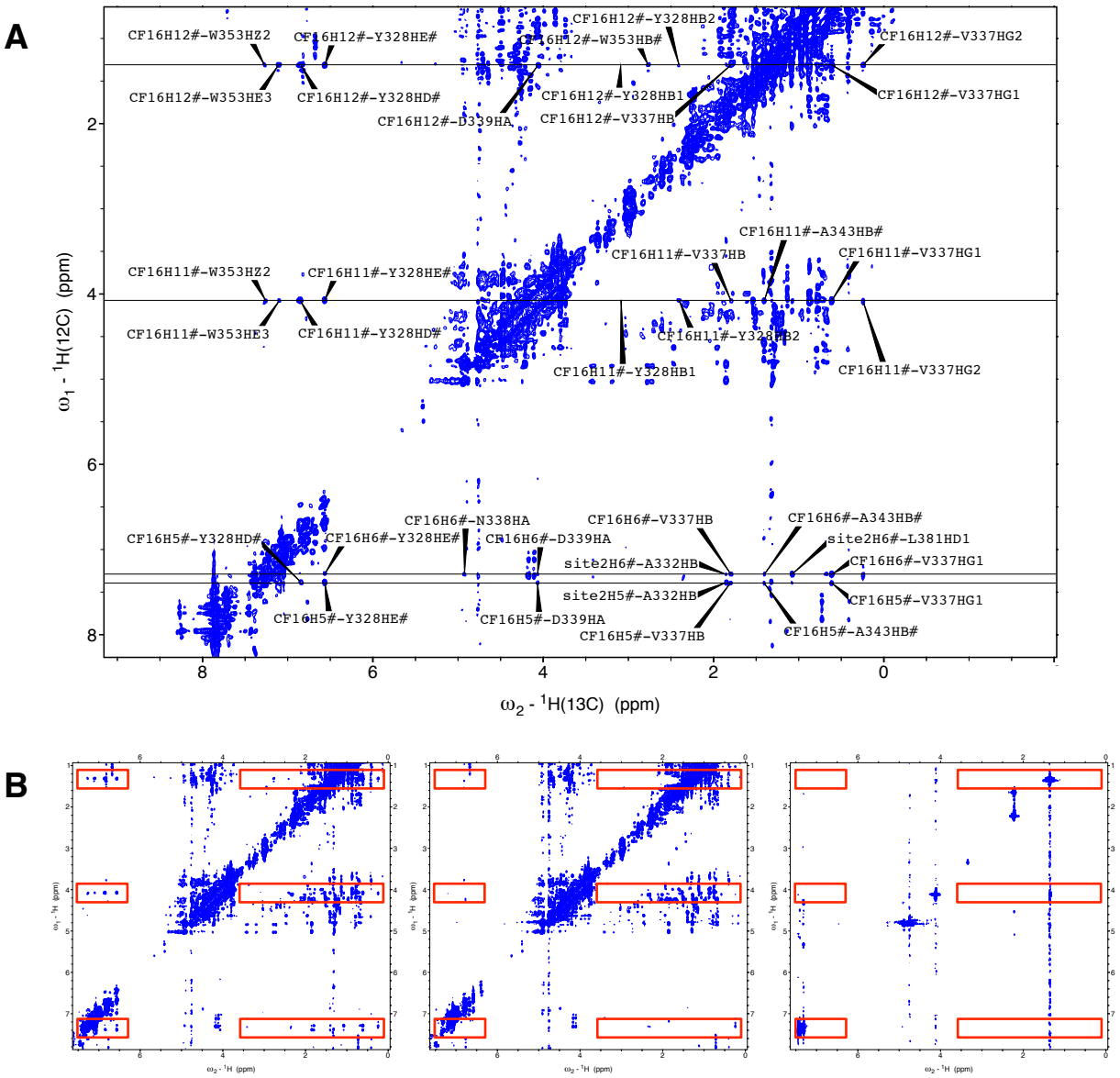
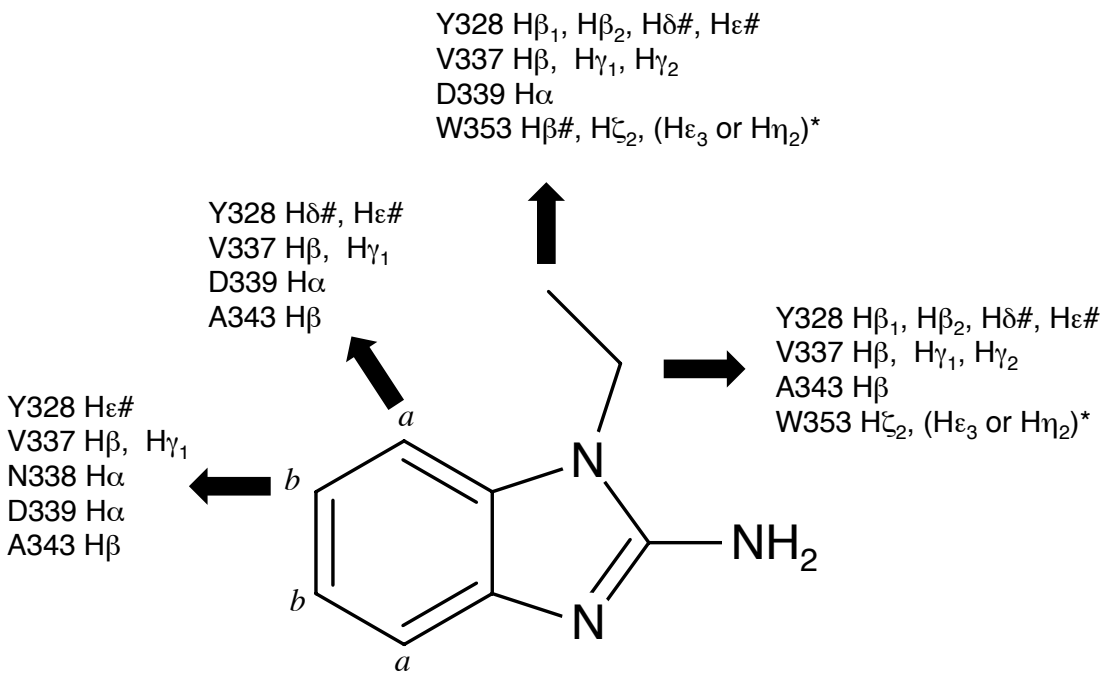


Fig. S4 Miller et al.



HADDOCK ambiguous restraints from CSP in {^{1}H,^{15}N}-HSQC:
active (5): V327, Y328, D340, Q341, W353
passive (6): L326, E336, N338, D339, D342, K352

Fig. S5 Miller et al.

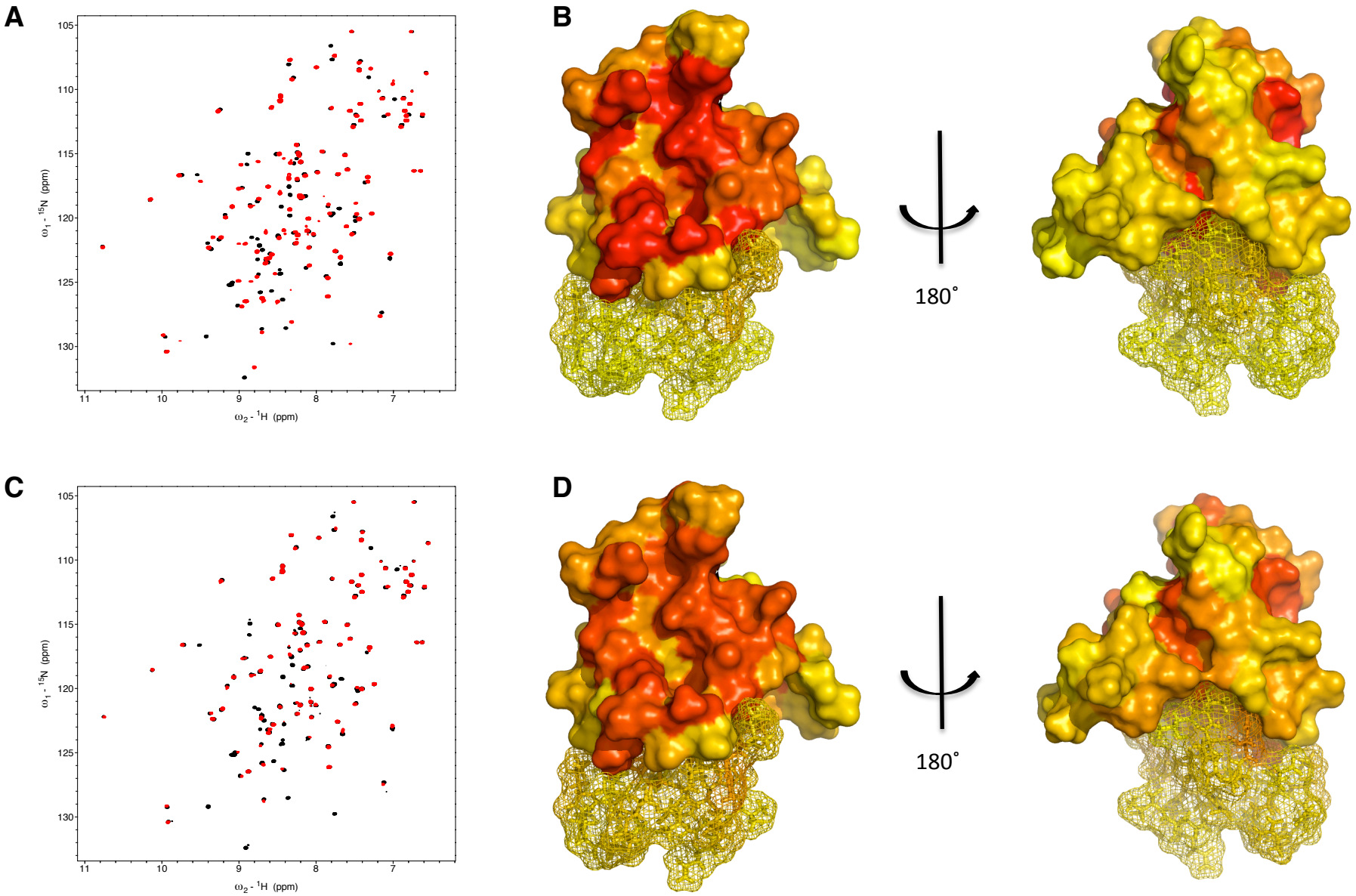


Fig. S6 Miller et al.

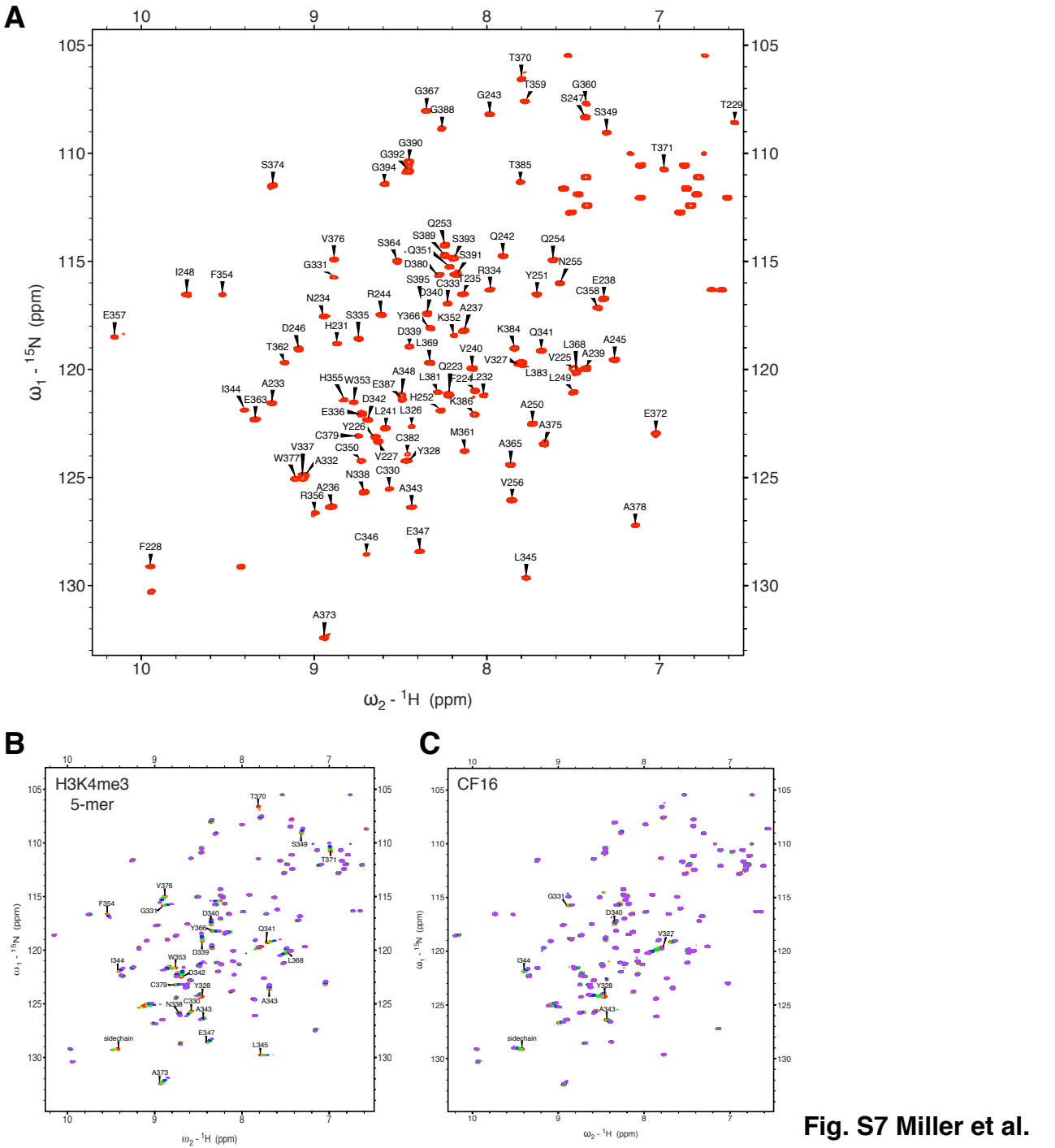


Fig. S7 Miller et al.

	CF1	CF3	CF4	CF6	CF7	CF16	CF18
K_d (mM)	3.1	4.6	2.5	4.3	2.9	7.3	14.7
SE (mM)	1.3	1.6	0.5	2.1	0.6	2.0	5.3

Table S1 Miller et al.

PHD-HD1-CF4 (4UP5)	
Beamline	Diamond, I04
Strategy (Low Resolution)	90°, Δφ 0.5°
Strategy (High Resolution)	90°, Δφ 0.5°
Space Group	$P4_12_12$
a, b, c (Å)	54.27, 54.27, 57.65
α, β, γ (°)	90, 90, 90
Resolution (Å)	31.95 - 1.65 A (1.68-1.65)[a]
R_{merge} (%)[b]	9.5 (141.1)
I/σ(I)	16.7 (2.1)
Mn(I) half-set correlation CC(1/2)	0.999 (0.722)
Completeness (%)	100 (100)
Multiplicity	13 (13.7)
Complexes in A.U.	1
Refinement	
Resolution (Å)	39.52 - 1.65 (1.693-1.650)
Number of reflections	10,327
Test set size (%)	4.8
R_{work}(%)	18.190 (26.8)
R_{free}(%)	21.117 (32.1)
Number of atoms (non-H)	692
Residues (PHD/HD1)	327-387/235-263
<B> ($Å^2$)	28.805
Rmsd	
Bond Length (Å)	0.022
Bond Angle (°)	2.110
Ramachandran plot	
In favoured regions (%)	98.8
In allowed regions (%)	0
Outliers (%)	1.2 (S349[c])

[a]Highest resolution shell (in Å) shown in parentheses

[b]Rmerge = $\Sigma_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \Sigma_{hkl} I_{hkl}$

[c]S349 is present in an unusual conformation (the electron density is clear) due to the adjacent cysteines (333/350) that form part of the second Zn^{2+}-binding site

Table S2 Miller et al.

PHD-HD1-H3K4me2K9ac (4UP0)	
Beamline	ESRF, ID29 – remote data collection
Strategy	180°, Δφ 0.25°
Wavelength (Å)	0.97621
Space Group	$P4_12_12$
a, b, c (Å)	53.91, 53.91, 57.94
α, β, γ (°)	90, 90, 90
Resolution (Å)	39.47 - 1.28 A (1.3-1.28)[a]
R_{merge} (%)[b]	8.9 (154.5)
I/σ(I)	13 (2.4)
Mn(I) half-set correlation CC(1/2)	0.998 (0.753)
Completeness (%)	99.9 (99.6)
Multiplicity	11.6 (12.3)
Complexes in A.U.	1
Refinement	
Resolution (Å)	39.47 - 1.28 A (1.312-1.279)
Number of reflections	21,476
Test set size (%)	5.1
R_{work}(%)	13.797 (25.5)
R_{free}(%)	16.611 (29.9)
Number of atoms (non-H)	789
Residues (PHD/HD1/Histone H3)	327-387/235-263/1-15
<B> (Å^2)	21.159
Rmsd	
Bond Length (Å)	0.021
Bond Angle (°)	1.974
Ramachandran plot	
In favoured regions (%)	98.8
In allowed regions (%)	1.2 (S349[c])
Outliers (%)	0

[a]Highest resolution shell (in Å) shown in parentheses
[b]Rmerge = $\Sigma_{hkl} |I_{hkl} - \langle I_{hkl} \rangle| / \Sigma_{hkl} I_{hkl}$

[c]S349 and R344 are present in an unusual conformation (the electron density is clear) due to the adjacent cysteines (333/350) that form part of Zn^{2+}-binding sites

Table S3 Miller et al.