Supporting Information

Couture *et al.* 10.1073/pnas.0806712105

SI Text

TLC Assay. Purified native SET8 and the Y334F mutant (5 μM) were used in methyltransferase assays with the biotinylated H4-K20 peptide (500 μM). Assays contained 1% (wt/vol) BSA, 20 mM $MgCl_2$, 100 mM Bicine (pH 8.0), 1.0 mM [^{3}H-methyl] AdoMet [diluted to a specific activity of 0.2 μCi/nmol with purified AdoMet (1)], 3 μM of *Sulfolobus solfataricus* AdoHcy hydrolase (2), and 2 units of adenosine deaminase (Sigma) in a final volume of 20 μL and incubated at 37 °C. Assays were terminated by the addition of an equal volume of 200 mM MES, pH 5.1 before the addition of a 2-fold molar excess of immobilized avidin resin (Pierce Biotechnology). The resin was then collected by centrifugation, washed three times with 300 mM NaCl, and boiled in 2% SDS for 5 min to liberate the bound histone H4 peptide. The recovered samples were then hydrolyzed under acidic conditions (6 M HCl) at 110 °C for 24 h. After treatment with KOH to remove the SDS, the samples were taken to dryness in vacuo. Subsequently, the samples were reconstituted with 50 μL of 100 mM Bicine, pH 8.0 and standards of lysine and monomethylated, dimethylated, and trimethylated lysine (commercial sources; K, Sigma; Kme1 and Kme3, Bachem; and Kme2, Chem-Impex International) added to each for determining Rf positions on the silica gel TLC plates (Sigma; 20 $\times$ 20 cm, 250-μm layer of silica gel matrix on glass support, 5- to 17-μm particle size) after chromatography by using methanol/ammonium hydroxide (75:25, vol/vol) and ninhydrin staining [0.25% (wt/vol) in acetone] (3). After applying the entire reaction 2 cm from the bottom and developing with solvent for $\approx$3 h at 22 °C to within 1 cm of the top, the TLC plate was thoroughly dried before placing it in contact with the tritium Phosphor screen (GE Healthcare) for 15.5 days in the dark. Rf regions corresponding with the methylated lysine derivatives were excised from the plates, and the quantity of tritium incorporated during the assay was measured by liquid scintillation spectroscopy and corrected for recovery.

1. Chirpich TP, Zappia V, Costilow RN, Barker HA (1970) Lysine 2,3-aminomutase. Purification and properties of a pyridoxal phosphate and *S*-adenosylmethionine-activated enzyme. *J Biol Chem* 245:1778–1789.
2. Collazo E, Couture JF, Bulfer S, Trievel RC (2005) A coupled fluorescent assay for histone methyltransferases. *Anal Biochem* 342:86–92.
3. Rebouche CJ, Broquist HP (1976) Carnitine biosynthesis in *Neurospora crassa*: Enzymatic conversion of lysine to epsilon-*N*-trimethyllysine. *J Bacteriol* 126:1207–1214.

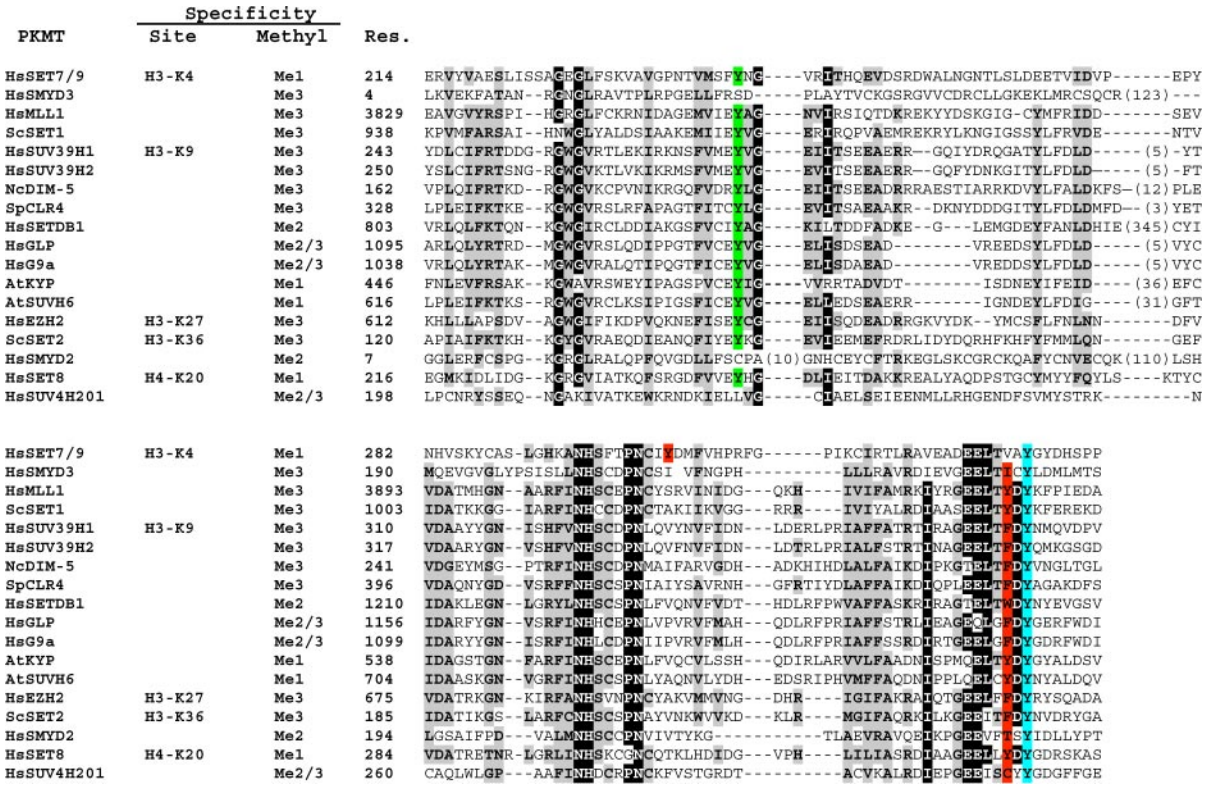

Fig. S1. Sequence alignment of the SET domains of representative histone lysine methyltransferases. The enzymes are categorized according to their site specificity, and their reported product specificities are also delineated. Sequence identity and conservation are delineated by black and gray text backgrounds, respectively. The conserved active-site tyrosine that hydrogen-bonds to the lysine or methyllysine ε-amine group is denoted in green, and the invariant tyrosine residue within the lysine-binding channel is indicated in cyan. Phe/Tyr switch residues are highlighted in red. In SET7/9, Tyr-305 has been shown to function as the Phe/Tyr switch residue. Although not conserved at the primary sequence level, the position of the phenol side chain of Tyr-305 is structurally conserved with the Phe/Tyr switch residues in other histone methyltransferases (Fig. S2 *D* and *E*). The product specificity ''ME2/3'' signifies PKMTs that exhibit weak trimethyltransferase activity. Species abbreviations: *At*, *Arabidopsis thaliana*; *Hs*, *Homo sapiens*; *Nc*, *Neurospora crassa*; *Sc*, *Saccharomyces cerevisiae*; and *Sp*, *Schizosaccharomyces pombe*.

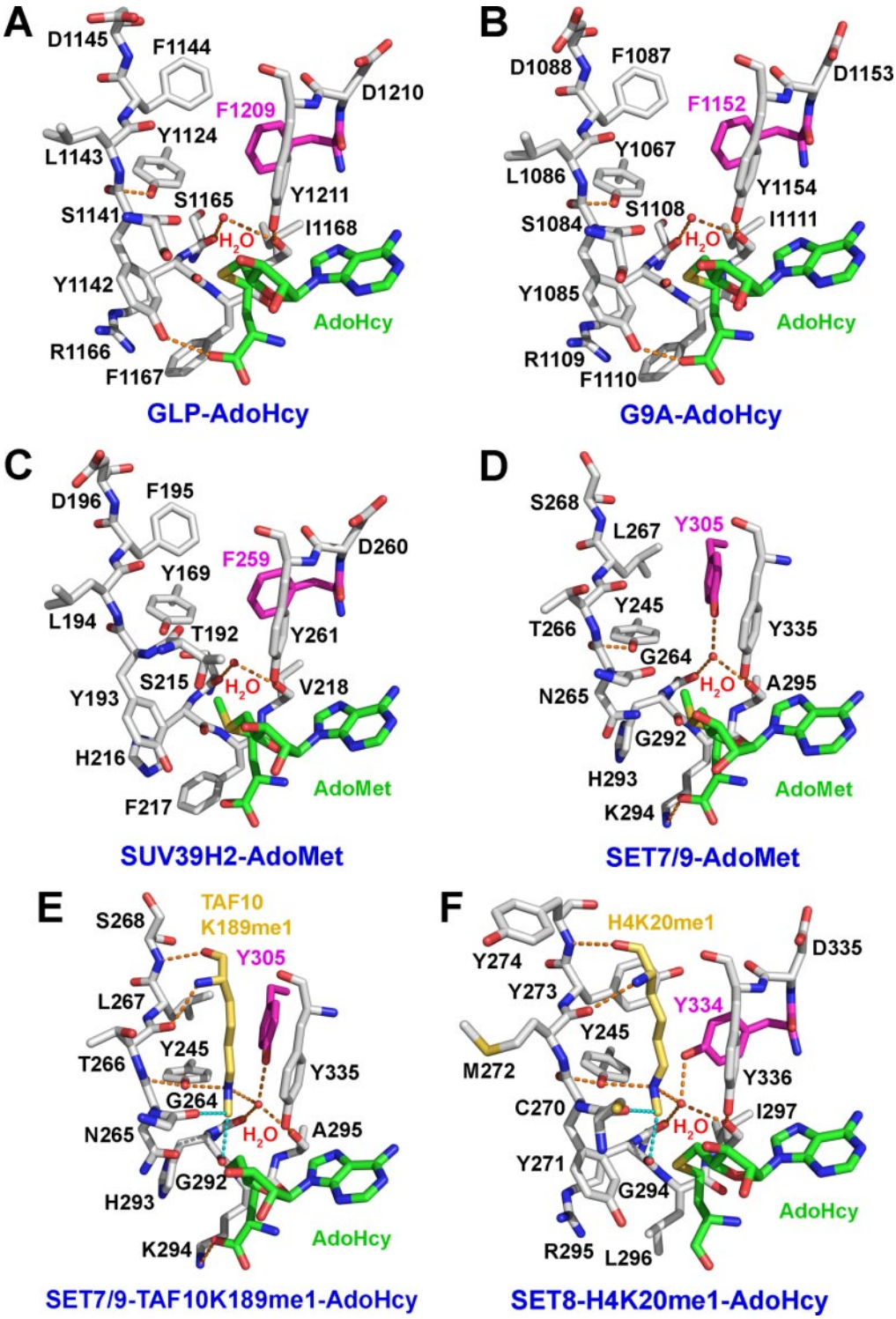

Fig. S2. Binary and ternary complexes of SET domain PKMTs highlighting the structural conservation of the active-site water molecule. The residues comprising the active site of each enzyme (gray carbons atoms), cofactor (green carbons), water (red), and peptidyl lysine or monomethyllysine (yellow carbons) are illustrated. Conventional hydrogen bonds are depicted as dashed orange lines, and CH···O hydrogen bonds are rendered as cyan dashes. Cut-off distances for conventional and CH···O hydrogen bonds are 3.3 and 3.7 Å, respectively. (*A*) GLP–AdoHcy complex (Protein Data Bank ID code 2IGQ). (*B*) G9A–AdoHcy (Protein Data Bank ID code 2O8J). (*C*) SUV39H2–AdoMet (Protein Data Bank ID code 2R3A). (*D*) SET7/9–AdoMet (Protein Data Bank ID code 1N6A). (*E*) SET7/9–TAF10K189me1–AdoHcy (Protein Data Bank ID code 2F69). (*F*) SET8–H4K20me1–AdoHcy (Protein Data Bank ID code 2BQZ).

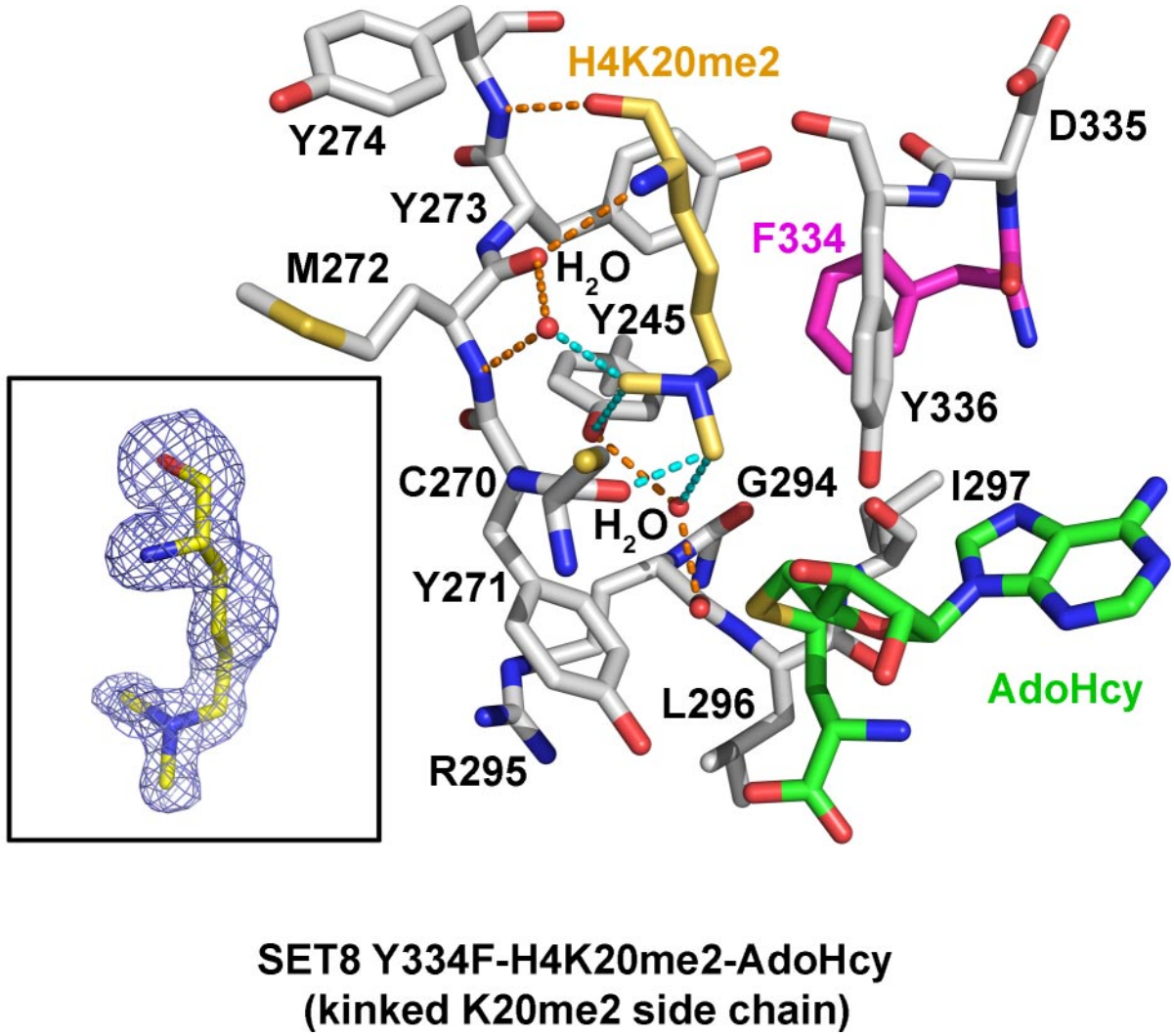

Fig. S3. The SET8 Y334F mutant bound to AdoHcy and the H4K20me2 peptide with the dimethyllysine side chain in a kinked conformation. Residues and hydrogen bonds are illustrated as in Fig. S2. *Inset* for the Y334F mutant represent the $F_O - F_C$ omit map electron density for the H4K20me2 side chain contoured at 2.0 σ.

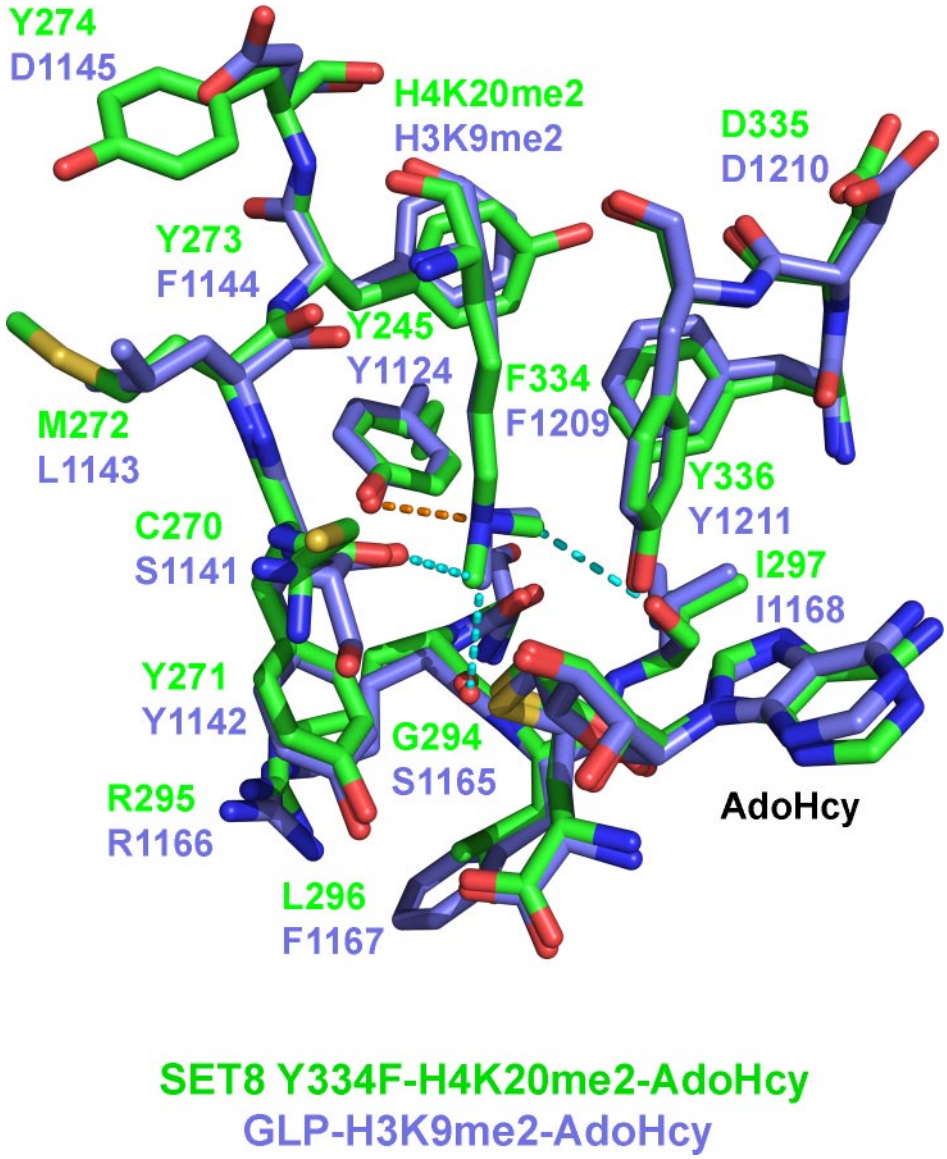

Fig. S4. Superimposition of the active sites of SET8 Y334F–H4K20me2–AdoHcy and GLP–H3K9me2–AdoHcy (Protein Data Bank ID code 2RFI) product complexes. The SET8 Y334F and GLP complexes are distinguished by green and violet carbon atoms, respectively. Hydrogen bonds are depicted as in Fig. S2.

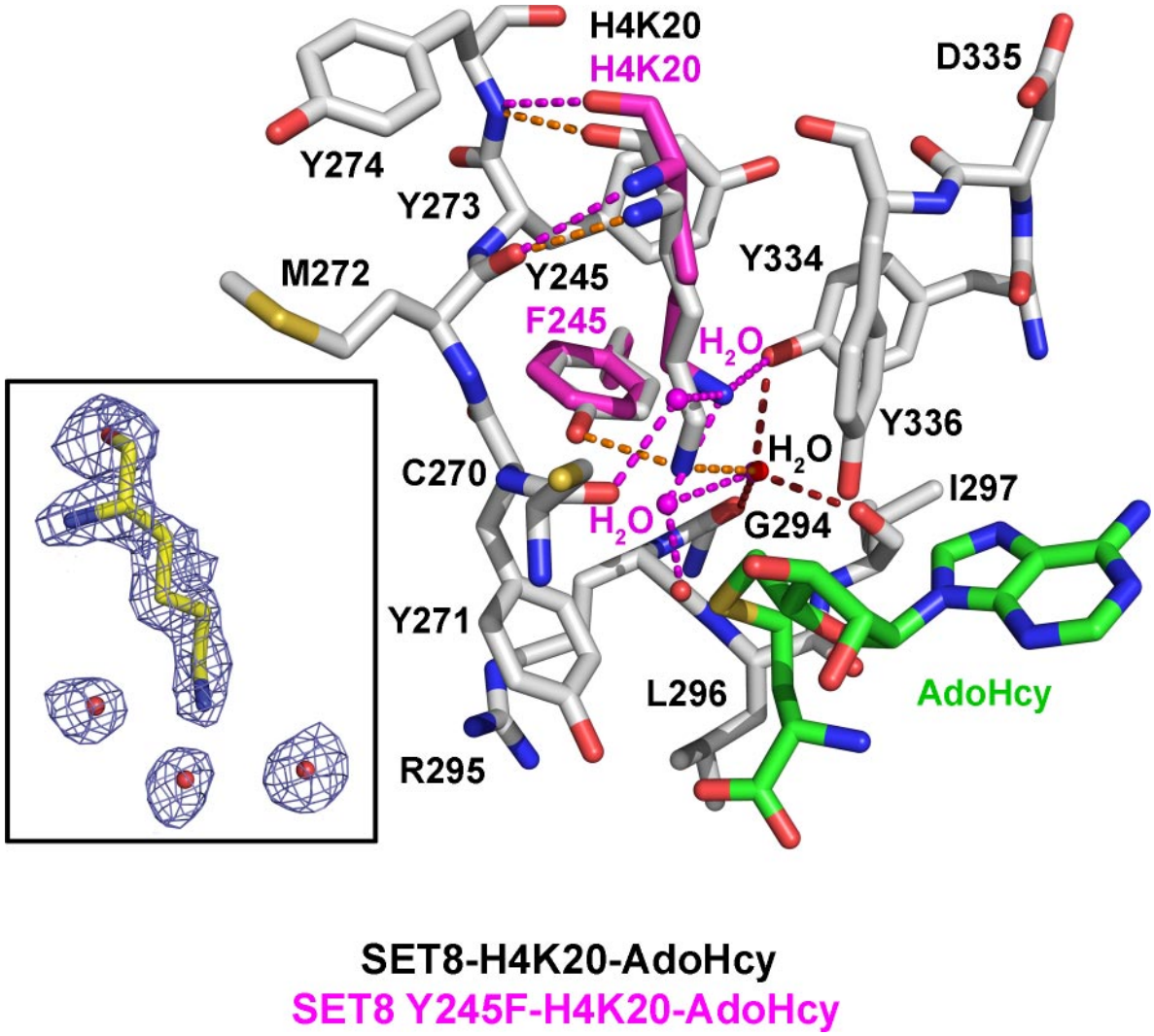

Fig. S5. Superimposition of the active sites of native SET8 and the SET8 Y245F mutant bound to the H4K20 peptide and AdoHcy. The lysine-binding channel of the enzyme is depicted as in Fig. S2. The Y245F mutation, H4K20, and the water molecules corresponding to this complex are depicted in magenta. Orange and magenta dashed lines illustrate hydrogen bonds that are observed in native SET8 and the Y245F mutant, respectively, and hydrogen bonds common to both complexes are shown as red dashes. (*Inset*) The $F_O - F_C$ omit map density (contoured at 2.0 σ) of the H4K20 side chain in the SET8 Y245F ternary complex.

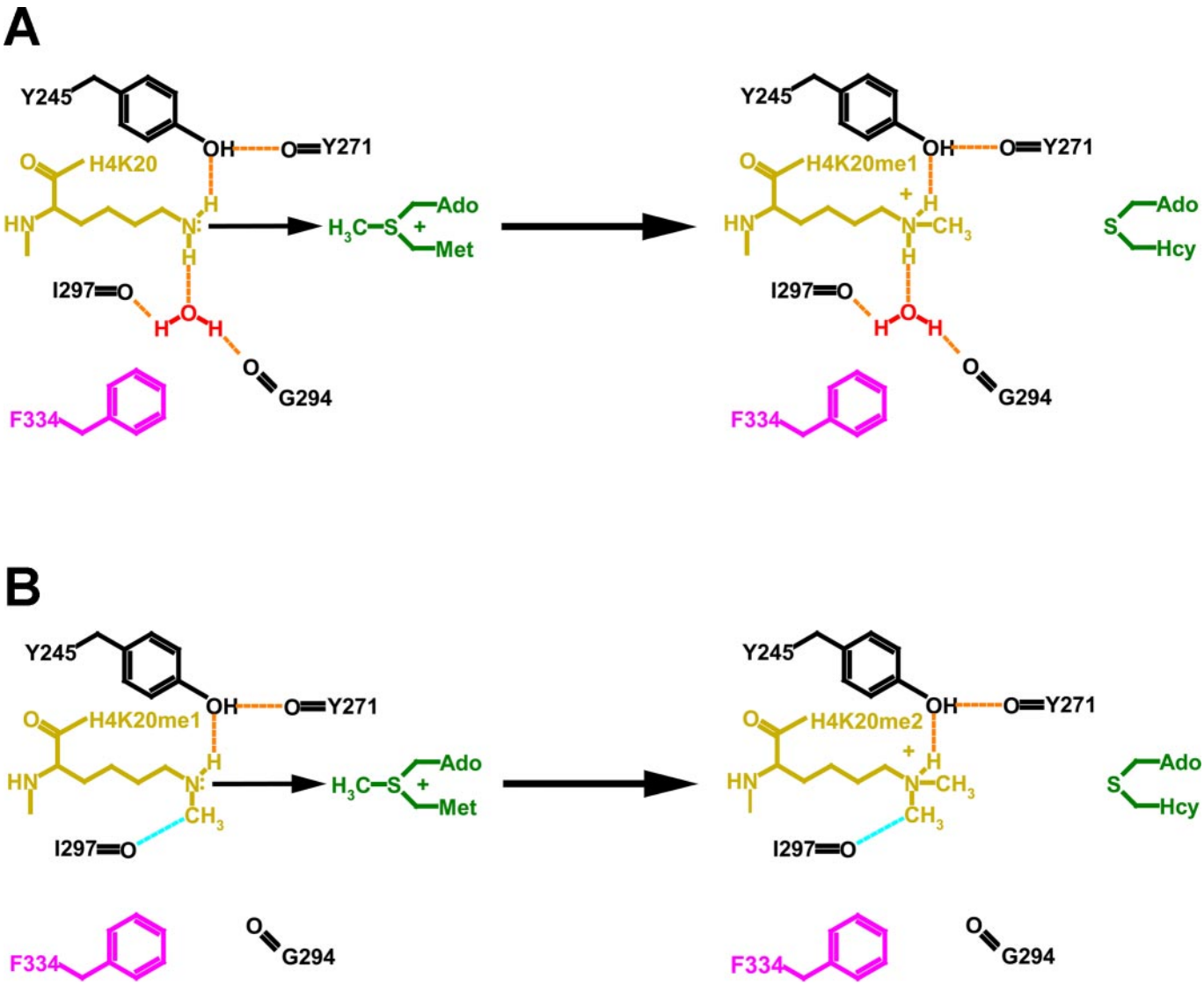

Fig. S6. Schematic representation of hydrogen bonding in the active site of SET8 Y334F during monomethyltransfer and dimethyltransfer. (*A*) In monomethylation, the Tyr-245 hydroxyl group and the active-site water molecule in the enzyme function as hydrogen-bond acceptors to align the deprotonated K20 ε-amine for an S_N2 methyltransfer reaction with the methyl group of AdoMet (same color scheme as in Fig. S2). After methyltransfer, the K20me1 side chain is bound in a protonated state through these two hydrogen bonds. (*B*) During dimethylation, the K20me1 side chain is bound in an alternative conformation with its methyl group positioned in the solvent binding pocket where it forms a CH···O hydrogen bond with Ile-297 in the enzyme. The deprotonated ε-amine is oriented for methyltransfer with AdoMet through a hydrogen bond with the Tyr-245 hydroxyl group. The dimethyl product is bound in a protonated state owing to the hydrogen bond to this tyrosine.

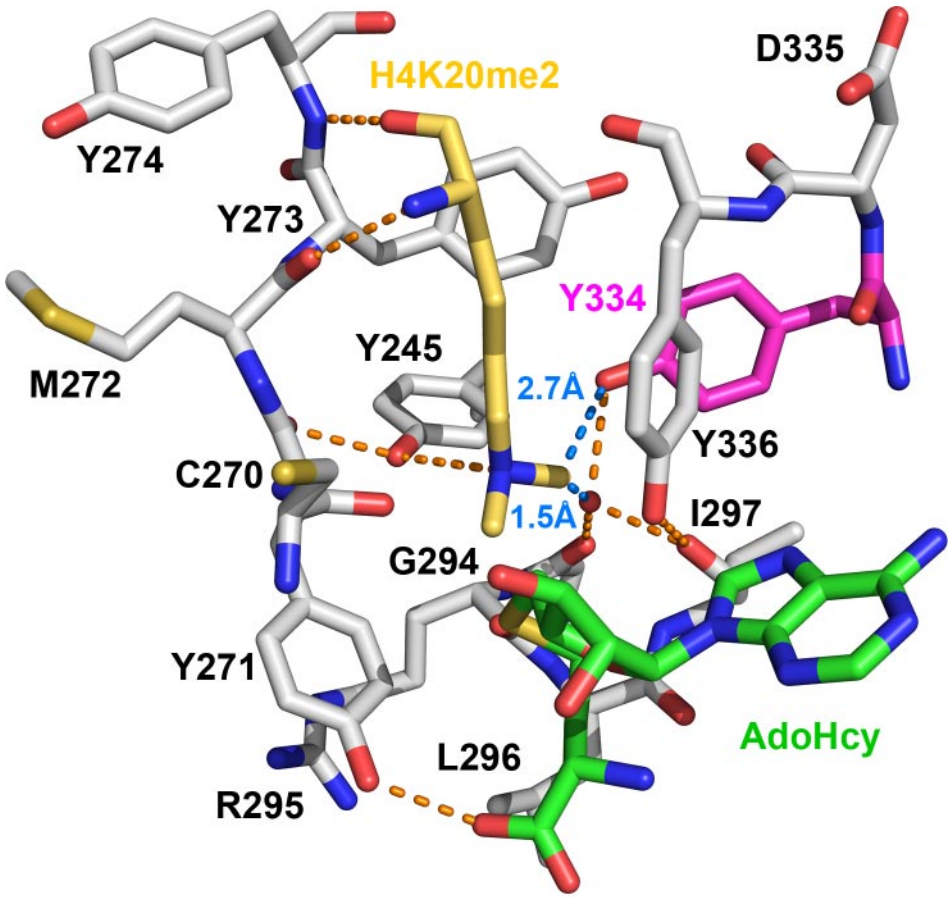

Fig. S7. Model of native SET8 bound to the H4K20me2 peptide. The coordinates of the SET8–H4K20–AdoHcy (Protein Data Bank ID code 1ZKK) and SET Y334F–H4K20me2–AdoHcy complexes were superimposed to generate the docking of the K20me2 in the active site of the native enzyme. Residues and hydrogen bonds are illustrated as in Fig. S2. The interaction distances between the K20me2 methyl group bound in the solvent pocket and the oxygen atoms of the active-site water molecule and Tyr-334 OH group in SET8 are denoted by dashed blue lines with distances shown.

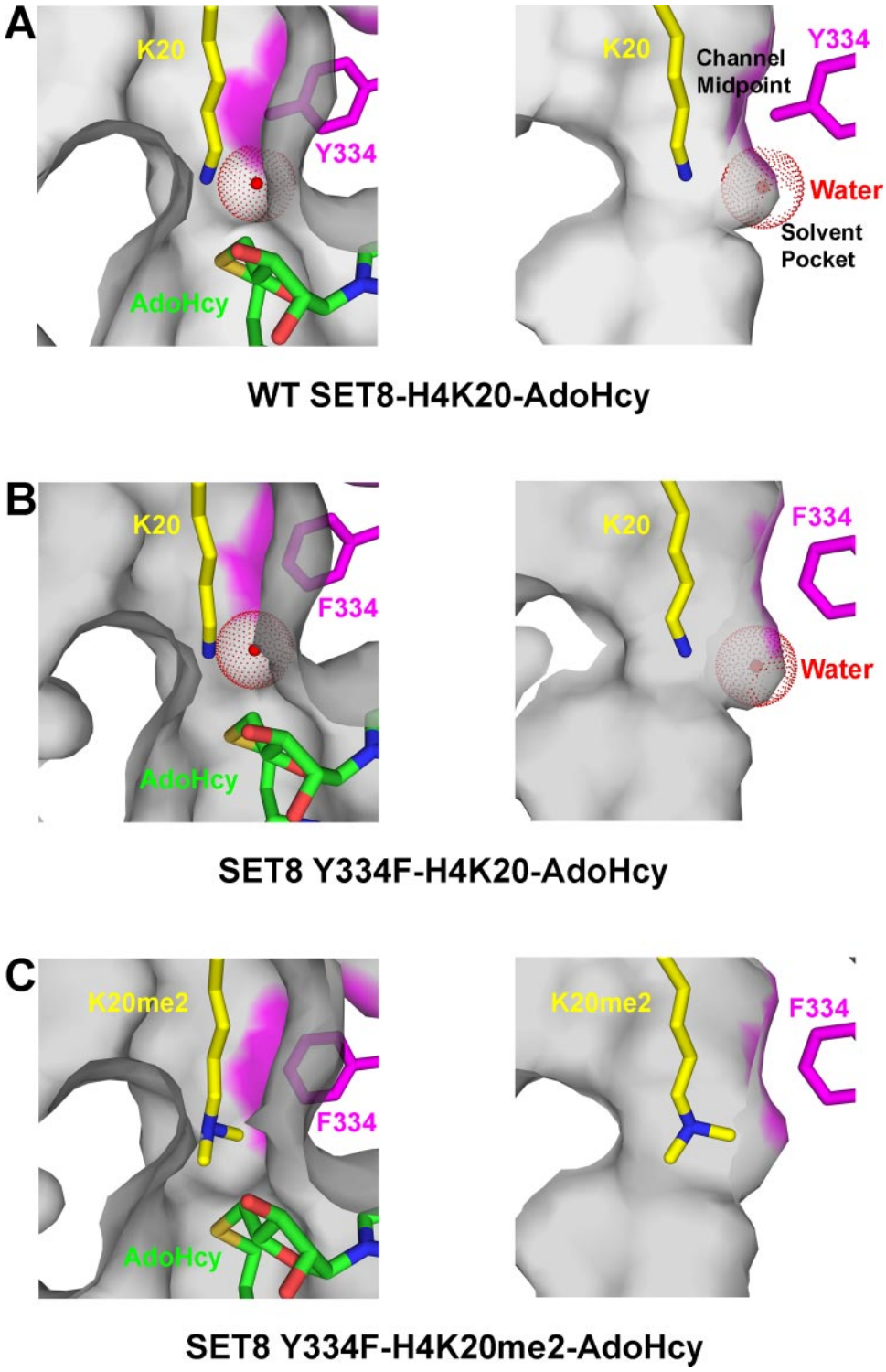

Fig. S8. Dimensions of the lysine-binding channels of native SET8 and the Y334F mutant. (*A*) Two views of the molecular surface of the lysine-binding channel of wild type (WT) SET8 bound to AdoHcy (green carbon atoms) and the H4K20 peptide (yellow carbons). For clarity, only the K20 side chain of the substrate is shown. The Phe/Tyr switch residue Tyr-334 and its contributions to the channel are illustrated in magenta. The active-site water molecule bound in the solvent pocket is rendered as a dotted van der Waals sphere in red. The right and left panels are separated by an approximate 45° rotation along the vertical axis. In the right panel, the cofactor is omitted, and the midpoint of the lysine binding channel and the solvent pocket are labeled for clarity. (*B*) Active site of SET8 Y334F–H4K20–AdoHcy complex illustrated as in *A*. (*C*) Active site of the SET8 Y334F–H4K20me2–AdoHcy complex depicted as in *A*.

Table S1. Crystallographic data and refinement statistics for SET8 mutant complexes

Statistic	SET8-Y334F H4K20/AdoHcy	SET8-Y334F H4K20me1/AdoHcy	SET8-Y334F H4K20me2/AdoHcy	SET8-Y245F, H4K20/AdoHcy
Synchrotron	APS	APS	APS	ESRF
Beamline	17-ID	22-ID	17-ID	BM30A
Crystal parameters				
Space group		P1		
Unit cell				
a, b, and c (Å)	44.0, 45.6, 94.2	44.4, 45.0, 52.4	44.1, 45.7, 94.7	43.9, 45.5, 94.4
α, β and γ (°)	89.2, 87.1, 90.8	91.4, 115.3, 89.8	89.4, 87.6, 89.8	89.4, 87.1, 90.5
No. of molecules in the AU	4	2	4	4
Data collections statistics				
Resolution range (Å)	12.9 – 1.6	25.0 - 1.5	50.0 – 1.25	40.0 – 1.6
Total reflections	363,128	439,614	604,630	334,404
Unique reflections	91,908	55,907	204,489	90,904
R_{sym} (%)	2.7 (14.6)	6.7 (52.8)	2.8 (13.1)	4.1 (17.3)
I /sigma (I)	38.2 (10.6)	22.5 (8.4)	30.4 (9.9)	31.1 (8.4)
Completeness (%)	95.2 (92.2)	95.0 (91.8)	94.3 (86.6)	94.1 (90.6)
Refinement statistics				
Resolution range (Å)	12.9 – 1.6	23.7 – 1.5	47.4 – 1.25	38.9 – 1.6
Reflections ($F_o > 2\sigma$)	87,774	53,407	183,160	87,679
Final model				
rmsd with native SET8	0.18	1.13*	0.55	0.30
Protein/peptide atoms	5,122/308	2,566/135	5,117/324	5,138/286
AdoHcy	104	52	104	104
Water	1,022	227	864	589
R factors†				
$R_{working}$	16.2	19.3	18.1	21.5
R_{free}	20.6	22.0	20.7	25.6
rms				
Bond length (Å)	0.010	0.019	0.013	0.016
Bond angles (°)	1.254	1.895	1.573	1.600
Average B factors (Å^2)				
Protein/peptide atoms	16.2/17.3	32.5/34.2	34.5/35.6	6.7/8.4‡
AdoHcy	15.1	23.3	38.8	6.6‡
Water	30.7	33.6	40.1	17.4‡

APS, Advanced Photon Source; ESRF, European Synchrotron Radiation Facility. Values in parentheses refer to data in the highest-resolution shell.

*The higher rms value for the H4K20me1 complex is caused in part by differences in the unit cell parameters compared with the other complexes.

†R factor: $R_{working} = \Sigma \mid\mid F_o \mid - \mid F_c \mid\mid / \Sigma \mid F_o \mid$; $R_{free} = \Sigma_T \mid\mid F_o \mid - \mid F_c \mid\mid / \Sigma_T \mid F_o \mid$, where T is a test data set of 5% of the total reflections randomly chosen and set aside before refinement.

‡The B factors of the SET8 Y245F complex differ from the other structures due to TLS refinement using Refmac. TLS groups and tensors are listed in the coordinate file.