Supporting Information

Investigation of Trimethyllysine Binding by the HP1 Chromodomain via Unnatural Amino Acid Mutagenesis

Stefanie A. Baril[a,‡], Amber L. Koenig[a,‡], Mackenzie W. Krone[a], Katherine I. Albanese[a], Cyndi Qixin He[b], Ga Young Lee[b], Kendall N. Houk[b], Marcey L. Waters*[,a], and Eric M. Brustad*[,a]

[a]Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599

[b]Department of Chemistry and Biochemistry, Box 951569, University of California, Los Angeles, Los Angeles, CA 90095

Table of Contents

Materials and General Methods.....S3

Cloning and DNA and Protein Sequences.....S3

Protein Expression.....S4

Protein Purification.....S5

Protein Characterization.....S6

ESI-LCMS confirmation of UAA incorporation.....S6

Circular dichroism (CD) of HP1 mutants.....S6

Peptide Synthesis.....S6

Isothermal titration calorimetry (ITC) binding measurements.....S7

Protein Crystallography.....S7

X-ray Data Collection and Protein Structure Determination.....S8

Verification of the Y24*p*NO_2Phe Mutation in Protein Structure.....S8

Analysis of Y24F and Y24*p*NO_2F Structures vs. Wild Type.....S8

Computational Methods.....S8

Table S1. Extinction Coefficients for UAAs and UAA HP1 variants.....S9

Table S2. ESI-LCMS instrument information.....S10

Table S3. ESI-LCMS method information for method A.....S11

Table S4. ESI-LCMS method information for method B.....S11

Table S5. ESI-LCMS data verifies UAA-incorporation.....S12

Table S6. Binding Constants for HP1 Mutants as Measured by ITC.....S13

Table S7. Data collection and refinement statistics for HP1 mutant crystals……S14
Figure S1. SDS-PAGE analysis of purified HP1 Mutants.....S15
Figure S2. LCMS of HP1 Wild Type using method A.....S15
Figure S3. LCMS of HP1 Y24F using method A.....S16
Figure S4. LCMS of HP1 Y24*p*CH_3Phe using method B.....S16
Figure S5. LCMS of HP1 Y24*p*CF_3Phe using method B.....S16
Figure S6. LCMS of HP1 Y24*p*CNPhe using method A.....S17
Figure S7. LCMS of HP1 Y24*p*NO_2Phe using method A.....S17
Figure S8. LCMS of HP1 Y24TAG with no UAA added using method B.....S17
Figure S9. LCMS of HP1 Y48F using method A.....S18
Figure S10. LCMS of HP1 Y48*p*CH_3Phe using method B.....S18
Figure S11. LCMS of HP1 Y48*p*CF_3Phe using method B.....S18
Figure S12. LCMS of HP1 Y48*p*CNPhe using method A.....S19
Figure S13. LCMS of HP1 Y48*p*NO_2Phe using method A.....S19
Figure S14. LCMS of HP1 Y48TAG with no UAA added using method B.....S19
Figure S15. Contaminant LCMS Peaks.....S20
Figure S16. Circular Dichroism of HP1 mutants.....S21
Figure S17. ITC curves of H3K9me3 binding to HP1 mutants.....S22
Figure S18. LFER plots of ΔG_b vs. σ_m and electrostatic potential (ESP).....S26
Figure S19. Density maps of HP1 Y24 mutants.....S27
Figure S20. Whole protein overlays of HP1.....S28
Figure S21. Overlays of the aromatic cage of various HP1 mutants.....S29
Figure S22. Cation-π distances between Kme3 and 24-and 48-position substituents.....S30
References.....S31

Materials and General Methods

Oligonucleotides were obtained from Integrated DNA Technologies. Enzymes and reagents used for cloning were obtained from New England BioLabs Inc. All other chemical reagents and solvents were obtained from chemical suppliers (Acros, Fisher Scientific, or Sigma-Aldrich) and used without further purification. DNA sequencing was performed by Genewiz. Protein LC-MS analysis was performed on an Agilent 6520 Accurate Mass QToF LC-MS ESI positive in high resolution mode. THE LC-MS was equipped with a Restek Viva C4 column (5 µm, 2.1 x 150 mm) see Table S1 - 3. MALDI mass spectrometry was performed on a SCIEX TOF/TOF 5800. ITC experiments were performed using a Microcal AutoITC200 and CD spectra were obtained using an Applied Photophysics Chiroscan Circular Dicroism Spectrophotometer. See detailed procedures below.

Cloning and DNA and Protein Sequences

pULTRA-*p*CNPheRS[1] was obtained from the lab of Dr. Peter Schultz and is also available from addgene (Plasmid # 48215). HP1 was cloned into a pET11a vector using NdeI and BamHI restriction sites. Mutations to the HP1 gene were generated using standard overlap PCR (primers available upon request). DNA sequences of cloned HP1 mutants from NdeI and BamHI restriction sites are shown below. The underlined portion of the sequence is the HP1 coding sequence. The 6XHis-tag is italicized. 24 and 48 positions have been bolded for clarity. Mutations to the Y24 position are shown in red and mutations to the Y48 position are shown in blue.

HP1 wild type DNA sequence:
CAT ATG AAA AAA *CAC CAC CAC CAC CAC CAC* GCC GAA GAG GAG GAG GAG GAG **TAC** GCC GTG GAA AAG ATC ATC GAC AGG CGG GTG CGC AAG GGA ATG GTG GAG TAC TAT CTG AAA TGG AAG GGC **TAT** CCC GAA ACT GAG AAC ACG TGG GAG CCG GAG AAC AAT CTC GAC TGC CAG GAT CTT ATC CAG CAG TAC GAG GCG AGC CGC AAG GAT TAA GGA TCC

HP1 wild type Protein sequence:
MKKHHHHHHAEEEEEE**Y**AVEKIIDRRVRKGMVEYYLKWKG**Y**PETENTWEPENNLDCQ DLIQQYEASRKD

HP1 Y24F DNA sequence
CAT ATG AAA AAA *CAC CAC CAC CAC CAC CAC* GCC GAA GAG GAG GAG GAG GAG **TTC** GCC GTG GAA AAG ATC ATC GAC AGG CGG GTG CGC AAG GGA ATG GTG GAG TAC TAT CTG AAA TGG AAG GGC **TAT** CCC GAA ACT GAG AAC ACG TGG GAG CCG GAG AAC AAT CTC GAC TGC CAG GAT CTT ATC CAG CAG TAC GAG GCG AGC CGC AAG GAT TAA GGA TCC

HP1 Y24F Protein sequence:
MKKHHHHHHAEEEEEE**F**AVEKIIDRRVRKGMVEYYLKWKG**Y**PETENTWEPENNLDCQ DLIQQYEASRKD

HP1 Y24TAG DNA sequence
CAT ATG AAA AAA *CAC CAC CAC CAC CAC CAC* GCC GAA GAG GAG GAG GAG GAG **TAG** GCC GTG GAA AAG ATC ATC GAC AGG CGG GTG CGC AAG GGA ATG GTG GAG TAC TAT CTG AAA TGG AAG GGC **TAT** CCC GAA ACT GAG AAC ACG TGG GAG CCG GAG AAC AAT CTC GAC TGC CAG GAT CTT ATC CAG CAG TAC GAG GCG AGC CGC AAG GAT TAA GGA TCC

HP1 Y24TAG Protein Sequence: (* represents UAA)
MKKHHHHHHAEEEEEE*****AVEKIIDRRVRKGMVEYYLKWKGYPETENTWEPENNLDCQ DLIQQYEASRKD

HP1 Y48F DNA sequence
CAT ATG AAA AAA *CAC CAC CAC CAC CAC CAC* GCC GAA GAG GAG GAG GAG **TAC** GCC GTG GAA AAG ATC ATC GAC AGG CGG GTG CGC AAG GGA ATG GTG GAG TAC TAT CTG AAA TGG AAG GGC **TTT** CCC GAA ACT GAG AAC ACG TGG GAG CCG GAG AAC AAT CTC GAC TGC CAG GAT CTT ATC CAG CAG TAC GAG GCG AGC CGC AAG GAT TAA GGA TCC

HP1 Y48F Protein sequence:
MKKHHHHHHAEEEEEEYAVEKIIDRRVRKGMVEYYLKWKG**F**PETENTWEPENNLDCQ DLIQQYEASRKD

HP1 Y48TAG DNA sequence:
CAT ATG AAA AAA *CAC CAC CAC CAC CAC CAC* GCC GAA GAG GAG GAG GAG **TAC** GCC GTG GAA AAG ATC ATC GAC AGG CGG GTG CGC AAG GGA ATG GTG GAG TAC TAT CTG AAA TGG AAG GGC **TAG** CCC GAA ACT GAG AAC ACG TGG GAG CCG GAG AAC AAT CTC GAC TGC CAG GAT CTT ATC CAG CAG TAC GAG GCG AGC CGC AAG GAT TAA GGA TCC

HP1 Y48TAG Protein Sequence: (* represents UAA)
MKKHHHHHHAEEEEEEYAVEKIIDRRVRKGMVEYYLKWKG*****PETENTWEPENNLDCQ DLIQQYEASRKD

Protein Expression

For UAA-HP1 variants, pET11a-HP1-Y24TAG or -48TAG was co-transformed with pULTRA-pCNPheRS into BL21-Gold(DE3) competent cells (Agilent Technologies). For HP1 wild type, Y24F and Y48F, pET11a-HP1, -Y24F or -Y48F were transformed into BL21-Gold(DE3) competent cells. Cells were rescued with 1 mL SOC broth and then incubated for 45 min at 37°C with shaking. 50 uL of each rescue was plated as follows: wild type/Y24F/Y48F on LB ampicillin (100 mg/L) agar plates; Y24TAG/Y48TAG contransformed with pULTRA-pCNPheRS on LB ampicillin (100 mg/L) and streptomycin (50 mg/L) agar plates. Plates were incubated overnight at 37°C. Single colonies from the transformation plates were used to inoculate LB with appropriate antibiotic in baffled

flasks (flask volume <4X larger than LB volume). Cultures were grown to saturation overnight at 37°C with shaking at 225 RPM.

All proteins were expressed in 2.5 L Ultra Yield Flask™ (Thompson Instrument Company) containing 500 mL of ZYP-5052 autoinduction media[2,3] supplemented with 5 mM $MgCl_2$, 5 mM $MgSO_4$, and 1:5000 dilution of Antifoam 204 to increase oxygen uptake and prevent foaming over. Each flask also contained appropriate antibiotics (100 mg/L ampicillin (pET-HP1s), 50 mg/L streptomycin (pUltra-pCNPheRS)). For wild type, Y24F, and Y48F expressions, media was inoculated with 2.5 mL of saturated overnight culture. For Y24TAG and Y48TAG expressions, autoinduction media was inoculated with 5 mL of saturated overnight culture to account for the slower initial growth in the presence of two antibiotics. After inoculation, cultures were incubated at 37°C with 310-350 RPM shaking until reaching an OD_{600} between 1 – 2. Dry UAA (Chem Impex International) was added to the appropriate TAG cultures (2.5 mmol UAA for 5 mM final concentrations. *p*NO_2Phe was increased to 10 mmol for 20 mM final concentration to compensate for lower affinity of the *p*CNPheRS for *p*NO_2Phe). Incubator temperature was then dropped to 18°C and the cultures were left to express for 24 hours. For expressions containing Y24*p*NO_2Phe, the incubator was covered with aluminum foil to prevent light degradation of *p*NO_2Phe.

After expression, cultures were pelleted at 4500 RPM for 10 min and the supernatant was decanted. Cell pellets were frozen overnight at -20°C and resuspended in 20 mL lysis buffer (50 mM Tris, pH 8, 150 mM NaCl, 30 mM imidazole, 0.25 mg/mL lysozyme, 1mM phenylmethanesulfonyl fluoride, with cOmplete EDTA-Free Protease Inhibitor Cocktail Tablets (Roche)). The resuspended pellet was incubated at 37°C with 225 RPM shaking for 30 min and cooled on ice for 10 min. Pellets were sonicated on ice for 7.5 min (20% amplitude, 0.5 s on, 0.5 s off) until the lysate appeared homogenous. Lysate was clarified by centrifugation (19,000 RPM, Sorvall SS-34 rotor) for 45 min. Supernatant was decanted and filtered through a 0.45 um syringe filter.

Protein Purification

Filtered lysate was purified on an ÄKTAPurifier UPC 10 (GE) equipped with a HisTrap-5mL HP column (GE). HP1 was 6XHis-tag purified using the buffers previously described[4] and eluted using a step gradient from 0 – 55 % buffer B. Eluted fractions were pooled and concentrated on a 3 kDA Amicon Ultra-15 Centrifugal filter. The concentrated sample was purified by size exclusion chromatography using a Superdex 200 10/300 GL size exclusion column equilibrated in SEC buffer (50 mM sodium phosphate, pH 8, 25 mM NaCl, 2 mM DTT). Eluted fractions (eluted at 15.5 - 18 mL) were pooled, concentrated, and quantified using a Cary 100 UV/Vis Spectrophotometer (Agilent Technologies). Extinction coefficients for UAA proteins were calculated by measuring the extinction coefficient of each free amino acid in solution and adding the free UAA extinction coefficient to the extinction coefficient of wild type HP1 with one tyrosine removed. The extinction coefficient of wild type HP1 with a tyrosine removed

was calculated using the Scripps Protein Calculator (http://protcalc.sourceforge.net/). Extinction coefficients are provided in **Table S1**.

Protein Characterization

Protein purity was confirmed by SDS-PAGE (**Figure S1**) and ESI-LCMS (**Table S5**, **Figure S2 – S14**). Protein folding was confirmed by circular dichroism (**Figure S15**).

ESI-LCMS confirmation of UAA incorporation

1 mL of a 10 μM solution of each protein was exchanged into HPLC-grade water using an Amicon Ultra-15 centrifugal filter and then filtered through glass wool. The samples were run on an Agilent 6520 Accurate-Mass Q-TOF ESI positive LCMS (**Table S2**) using one of two methods: A) or B). Details of each method can be found in **Table S3 and S4**.

All LCMS chromatograms show evidence of the appropriate UAA-incorporation with no detectible canonical amino acid contamination (**Table S5**). Chromatograms from each LCMS can be found in the **Figure S2** – **Figure S14.** Although incorporation of tyrosine or phenylalanine can be detected in TAG mutants expressed in the absence of unnatural amino acid (**Figure S8** and **Figure S14**), no evidence of tyrosine or phenylalanine incorporation is detected in the presence of UAA.

Circular dichroism (CD) of HP1 mutants

CD experiments were performed using an Applied Photophysics Chiroscan Circular Dicroism Spectrophotometer. Spectra were obtained with 30 μM chromodomain in 10 mM sodium phosphate buffer, pH 7.4 with 2 mM dithiothreitol (DTT) at 20°C. All scans were corrected with buffer subtraction. The mean residue ellipticity was calculated using the equation $\theta = \frac{signal}{10lc}\frac{1}{r}$ where θ is MRE, signal is CD signal, l is path length, c is protein concentration, and r is the number of amino acid residues. All spectra were measured using a quartz cuvette with a path length of 0.1 cm.

Peptide Synthesis

H3K9me3 (ARTKQTARK(Me)$_3$STGGKAY) was synthesized using Fmoc protected amino acids and Rink Amide AM resin on a 0.5 mmol scale. The amino acid residues were activated with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) and HOBt (N-hydroxybenzotriazole) in the presence DIPEA (diisopropylethylamine) in DMF (N,N-dimethylformamide). 4 equivalents of the amino acid, HBTU, and HOBt were used for each coupling step, along with 8 equivalents of DIPEA. Double couplings of 30 minutes were used for each residue. Deprotections of Fmoc were carried out in 20% piperidine in DMF, twice for 15 minutes each.

Trimethyllysine was generated during the synthesis of the H3 peptide by first coupling Fmoc-Lys(Me)$_2$-OH·HCl for 5 hours with HBTU/HOBt activation. 2 equivalents of

dimethyllysine, HBTU, and HOBt were used, along with 4 equivalents of DIPEA. Immediately after coupling, the resin was washed with DMF and the residue was methylated to form trimethyllysine with 7-methyl-1,5,7-triaza-bicyclo[4.4.0]dec-5-ene (MTDB, 1.2 eq) and methyl iodide (10 eq) in DMF for 6 hours. The resin was washed with DMF and peptide synthesis was continued with aforementioned conditions.

Peptides were cleaved with 95:2.5:2.5 trifluoroacetic acid (TFA):water:triisopropylsilane (TIPS) for 4 hours. The TFA was evaporated and products were precipitated with cold diethyl ether. The resulting peptides were extracted with water and lyophilized. Crude peptide material were purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. The purified peptides were lyophilized. The peptide was desalted for ITC using a Sephadex G-24 column from GE in water and lyophilized to a powder. Identity was confirmed by MALDI mass spectrometry. Calculated $M+H^+$: 1765.02 Da Observed: 1765.95 Da.

Isothermal titration calorimetry (ITC) binding measurements

ITC experiments were performed by titrating H3K9me3 peptide (2.5-7.47 mM) into HP1 mutants (160-290 μM) in 50 mM sodium phosphate, pH 7.4, 150 mM NaCl, 2 mM TCEP at 25°C using a Microcal AutoITC200. Peptide and protein concentrations were determined by measuring absorbance at 280 nm on a Cary 100 UV/Vis Spectrophotometer (Agilent Technologies). Heat of dilution was accounted for by subtracting the endpoint ΔH value from each prior injection. Data was analyzed using the One-Site binding model supplied in Origin software. While the binding stoichiometry is know to be 1:1, at the high concentrations used here active protein concentration may differ from measured concentration. When ITC experiments were run under low c-value conditions ($c \leq 4$, $c = \frac{[protein]}{K_D}$), the stoichiometry parameter (N) of the non-linear fitting function was fixed to 1.[5,6] Data shown in Table S6 is the average of 3 runs unless otherwise noted.

Polarizability and Log P Values

Polarizability and Log P were calculated using Spartan 16 at the DFT B3LYP 6-31G* level of theory.[7]

Protein Crystallography

HP1 Y24F and Y24*p*NO_2Phe protein was diluted to a concentration of 10 mg/mL in 10 mM potassium phosphate, pH 7, 2mM TCEP. The diluted protein was then spiked with 8.6 mg/mL H3K9me3 peptide (~70% pure) in a 4:1 peptide:HP1 ratio. Crystals were grown by sitting drop vapor diffusion at 4°C. Cryschem Plates (Hampton Research) were set up on ice by mixing 1uL of the protein-peptide dilution and 1 uL of reservoir solution. Crystal growth was typically observed within 12 - 72 hours. Crystals were harvested and flash-frozen in liquid nitrogen with no supplementary cryoprotectant necessary.

Reservoir solution for Y24Phe: 0.1 M MES, pH 6.3; 3.4 M $(NH_4)_2SO_4$
Reservoir solution for Y24*p*NO_2Phe: 0.1 M MES, pH 5.8; 3.0 M $(NH_4)_2SO_4$

X-ray Data Collection and Protein Structure Determination

X-ray diffraction data were collected at Southeast Regional Collaborative Access Team (SER-CAT) at the Advanced Photon Source (Argonne National Laboratory) using beamline 22-ID and a MAR300HS CCD detector. Data were collected at 100 K. Statistics for data collection and refinement are listed in Table S6. Diffraction data sets were integrated and scaled with the automated data processing software KYLIN provided by SER-CAT.[8] Initial phases were determined by molecular replacement against the wild type HP1 structure (PDB accession code 1KNE)[9] using Phenix Phaser.[10] Refinement was accomplished by iterative cycles of manual model building with *Coot*[11] and automated refinement using Phenix Refine.[10] Model quality was assessed with the Phenix Validation tool. All of the protein structure figures and alignments were generated using PyMOL software (The PyMOL Molecular Graphics System, Version 1.8, Schrödinger LLC.).

Verification of the Y24*p*NO_2Phe Mutation in Protein Structure

For the Y24*p*NO_2Phe structure, a phenylalanine was first modeled in at the Y24 residue. After refinement, the mFo-DFc map showed extra electron density near the *para*-position of the phenylalanine ring (**Figure S18A**). When the phenylalanine is mutated to *p*NO_2Phe, the mFo-DFc density fits the UAA's R-group well (**Figure S18B**). Once the Y24*p*NO_2Phe mutation model is refined, the 2mFo-DFc density fits the UAA well (**Figure S18C**). The 2mFo-DFc density from the Y24F structure also matches the Y24F mutation, but lacks the *para*-electron density of the Y24*p*NO_2Phe structure (**Figure S18D**).

Analysis of Y24F and Y24*p*NO_2F Structures vs. Wild Type

The HP1 wild type structure (1KNE) was overlaid with the Y24F (6ASZ) and Y24*p*NO_2Phe (6AT0) structures (**Figure S19 and 20**). RMS values were calculated using the align feature of PyMOL. Alignment of wild type and Y24F gave an RMSD of 0.216, wild type and Y24*p*NO_2Phe gave a RMSD of 0.196, and Y24F and Y24*p*NO_2Phe gave an RMSD of 0.098. Based on the RMSD values, differences in binding are not likely due to changes in structure.

Computational Methods

E_{int} calculations between the wild type protein and trimethyllysine (Kme_3)

The structure of the Y24–Y48–Kme_3 complex was extracted and truncated from the crystal structure of the wild type protein (PDB: 1KNE). Each terminus of the fragments

was capped with a hydrogen atom at 1.09 Å. The cation/π interaction for each of the two tyrosine residues with the lysine ammonium ion was computed by single-point energy calculations at the M06/6-31G(d,p) level of theory.[12] M06/6-31G(d,p) was recently shown to model cation/π interactions well by Dougherty, et al.[13] The interaction energy is defined as the energy difference between the dimer and each amino acid monomers: $E_{int} = E_{dimer} - (E_{Kme3} + E_{Y})$. All quantum chemical calculations were performed using *Gaussian 09*.[14] All graphics on optimized structures were generated with *CYLview*.[15]

Table S1. Extinction Coefficients for UAAs and UAA HP1 variants

Mutant Name	Extinction Coefficients at 280 nm ($cm^{-1}M^{-1}$)	MW (Da)
wild type	17780.0	8569.4
Y24F or Y48F	16500.0	8553.4
Y24*p*CNPhe or Y48*p*CNPhe	17169.4	8578.4
Y24*p*NO_2Phe or Y48*p*NO_2Phe	24817.3	8598.4
Y24*p*CH_3Phe or Y48*p*CH_3Phe	16632.8	8567.5
Y24*p*CF_3Phe or Y48*p*CF_3Phe	16504.4	8621.4
UAA	Free UAA Extinction Coefficients	MW
*p*CNPhe	669.4	190.2
pNO_2Phe	8317.3	210.2
pCH_3Phe	132.8	179.2
pCH_3Phe	4.4	233.2

Table S2. ESI-LCMS instrument information

Column	Restek Viva C4 5 µm 150 x 2.1 mm
Solvent A	0.1 % formic acid in water
Solvent B	0.1 % formic acid in acetonitrile
Temperature	35°C
Ion Source	Dual ESI
Ion Polarity	Positive
Abs. Threshold	200
Rel threshold (%)	0.01
Cycle Time	1 s
Gas Temp	350 °C
Drying gas	12 l/min
Nebulizer	50 psig
Fragmentor	200 V
Skimmer	65 V
OCT 1 RF VPP	750
Min Mass Range	100 m/z
Max Mass Range	3200 m/z
Acquisition Rate	1 spectra/s
Acquisition time	1000.2 ms/spectrum
Transients/spectrum	9898

Table S3. ESI-LCMS method information for method A

Method A	
Solvent A	Water
Solvent B	Acetonitrile
Flowrate	0.4 mL/min
Gradient	
Time (min)	%B
0	5
2	5
8	30
22	60
23	60
35	70
40	95
42	95
44	5

Table S4. ESI-LCMS method information for method B

Method B	
Solvent A	Water
Solvent B	Acetonitrile
Flowrate	0.3 mL/min
Gradient	
Time (min)	%B
0	5
15	95
20	95
20.01	5
25	5

Table S5. ESI-LCMS data verifies UAA-incorporation

Sample	Expected Mass (Da)	Observed Masses (Da)	Difference (Da)	% Difference
Wild Type	8569.30	8569.67	0.29	3.4×10^{-5}
24F	8553.31	8553.78	0.29	3.4×10^{-5}
24*p*CH_3Phe	8567.46	8567.11	0.35	4.1×10^{-5}
24*p*CNPhe	8578.43	8578.96	0.55	6.4×10^{-5}
24*p*CF_3Phe	8621.42	8622.13	0.71	8.2×10^{-5}
24*p*NO_2Phe	8598.39	8598.80	0.41	4.8×10^{-5}
48F	8553.31	8553.85	0.54	6.3×10^{-5}
48*p*CH_3Phe	8567.46	8568.10	0.64	7.5×10^{-5}
48*p*CNPhe	8578.43	8579.01	0.58	6.8×10^{-5}
48*p*CF_3Phe	8621.42	8622.09	0.67	7.8×10^{-5}
48*p*NO_2Phe	8598.39	8598.71	0.32	3.7×10^{-5}

Table S6. Binding Constants for HP1 Mutants as Measured by ITC

Protein	Cation-π Energy[a] (kcal/mol)	K_d (μM)[b]	ΔG_b (kcal/mol)
Wild Type	26.6	14.4 ± 1.9	-6.6 ± 0.1
Y24*p*CH_3Phe	28.3	19.5 ± 1.3[c]	-6.4 ± 0.1
Y24F	26.9	19.0 ± 0.6	-6.4 ±0.1[d]
Y24*p*CF_3Phe	19.4	51.8 ± 5.2	-5.9 ± 0.1[d]
Y24*p*NO_2Phe	14.0	91.7 ± 0.1[c]	-5.5 ± 0.1[d]
Y48*p*CH_3Phe	28.3	16.7 ± 3.0	-6.5 ± 0.1
Y48F	26.9	15.8 ± 2.2	-6.5 ± 0.1
Y48*p*CF_3Phe	19.4	24.0 ± 0.8	-6.3 ± 0.1[d]
Y48*p*CNPhe	16.0	44.2 ± 1.7	-5.9 ± 0.1[d]
Y48*p*NO_2Phe	14.0	44.9 ± 14.3[c]	-5.9 ± 0.2

[a]Values taken from Wheeler et al.[16] [b]Values are an average of 3 runs unless otherwise noted. Errors are calculated from standard deviation. [c]Average of 2 runs. [d]Errors are calculated from error in fit given by Origin software.

Table S7. Data collection and refinement statistics for HP1 mutant crystals

	HP1 Y24F	HP1 Y24*p*NO_2Phe
PDB accession #	6ASZ	6AT0
Data collection		
Space group	C 2 2 21	C 2 2 21
Wavelength	1.000	1.000
Cell dimensions		
a, *b*, *c* (Å)	34.52 76.78 75.51	34.42 76.86 76.48
a, b, g (°)	90	90
Resolution (Å)	10.89 − 1.52 (1.57 − 1.52)*	11.22 − 1.285 (1.33 − 1.285)*
R_{merge}	8.0(47.4)	4.5 (46.86)
I / σI	4.4(1.5)	12.8 (2.56)
Completeness (%)	98.2(99.8)	96.9 (93.9)
Redundancy	5.7(5.2)	5.36(4.14)
Refinement		
Resolution (Å)	1.57 − 1.52	1.33 − 1.285
No. reflections	15582	25417
R_{work} / R_{free}	0.25 / 0.27	0.24 / 0.26
No. atoms		
Protein	448	483
Ligand/ion	49	56
Water	19	40
B-factors		
Protein	26.6	27.9
Ligand/ion	29.5	30.8
Water	29.1	37.5
R.m.s. deviations		
Bond lengths (Å)	0.006	0.004
Bond angles (°)	0.76	0.75
Ramachandran outliers	0%	0 %

*All data sets were collected from single crystals. Highest-resolution shell is shown in parentheses.

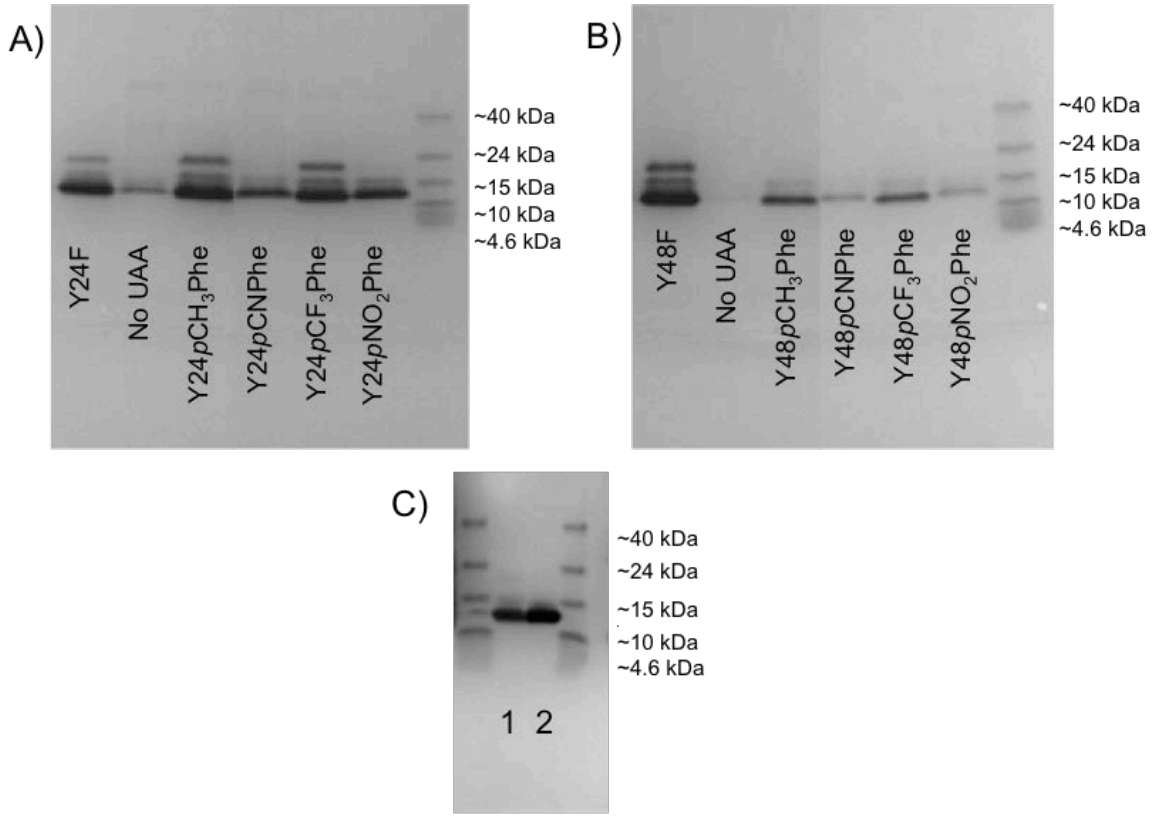


Figure S1. SDS-PAGE analysis of purified HP1 Mutants. 6XHis-Tagged purification of Y24 (A) and Y48 (B) mutants. C) Impurities present after his-tag purification (1) are removed after size-exclusion chromatography (2).

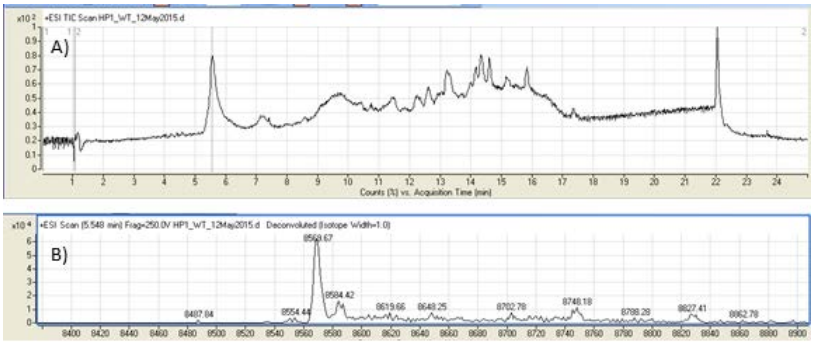


Figure S2. LCMS of HP1 Wild Type using method A. ESI scan (A) and corresponding m/z deconvolution (B).

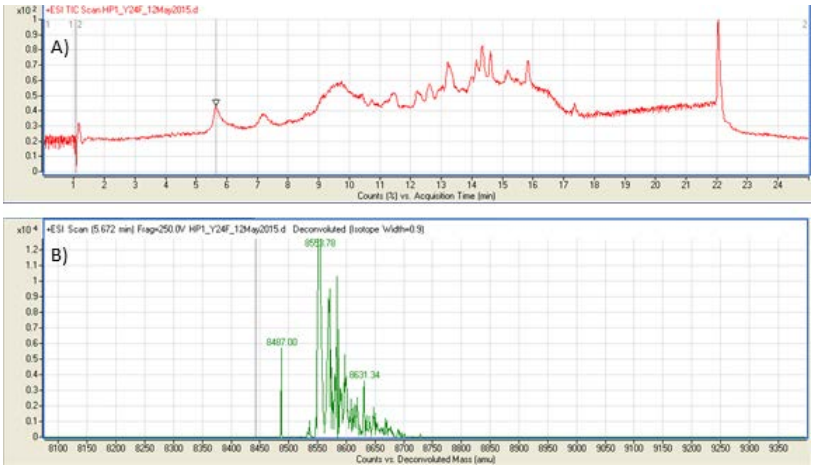


Figure S3. LCMS of HP1 Y24F using method A. ESI scan (A) and corresponding m/z deconvolution (B).

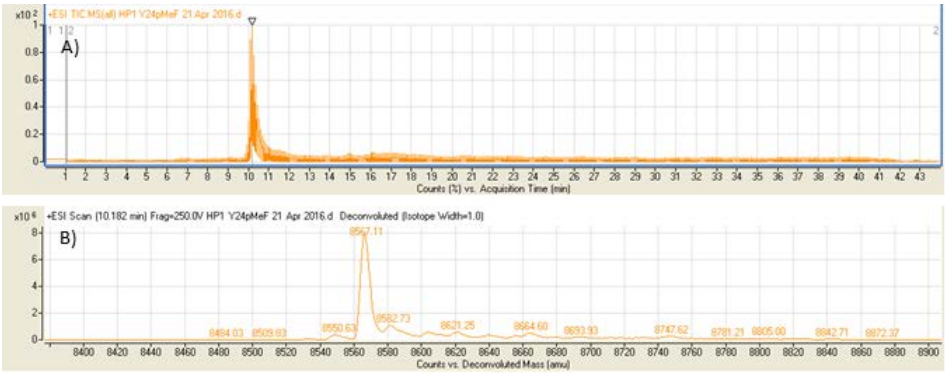


Figure S4. LCMS of HP1 Y24pCH3Phe using method B. ESI scan (A) and corresponding m/z deconvolution (B).

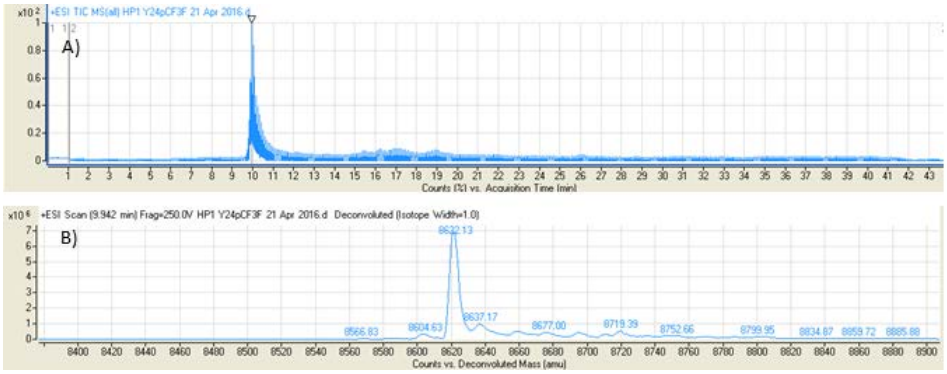


Figure S5. LCMS of HP1 Y24pCF3Phe using method B. ESI scan (A) and corresponding m/z deconvolution (B).

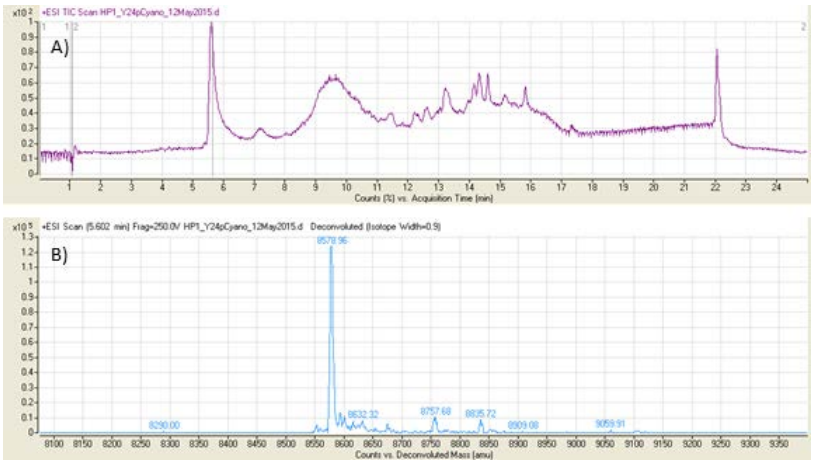


Figure S6. LCMS of HP1 Y24pCNPhe using method A. ESI scan (A) and corresponding m/z deconvolution (B).

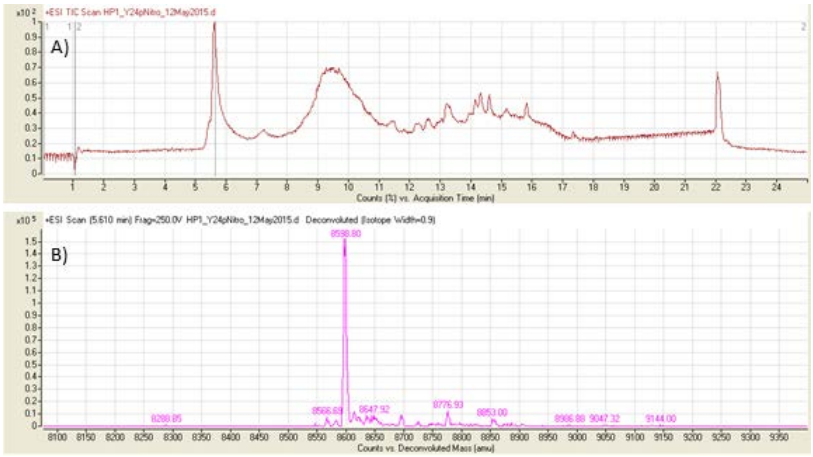


Figure S7. LCMS of HP1 Y24pNO2Phe using method A. ESI scan (A) and corresponding m/z deconvolution (B).

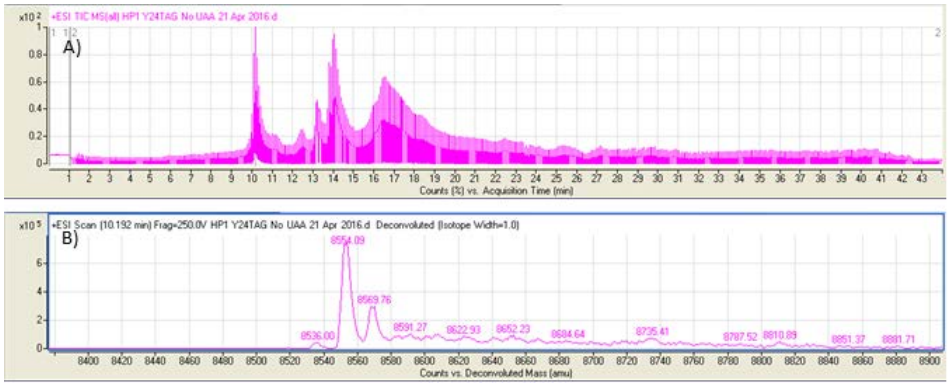


Figure S8. LCMS of HP1 Y24TAG with no UAA added using method B. ESI scan (A) and corresponding m/z deconvolution (B). Wild type HP1 and Y24F are produced in the absence of UAA, but when UAA is added wild type and Y24F are not detected.

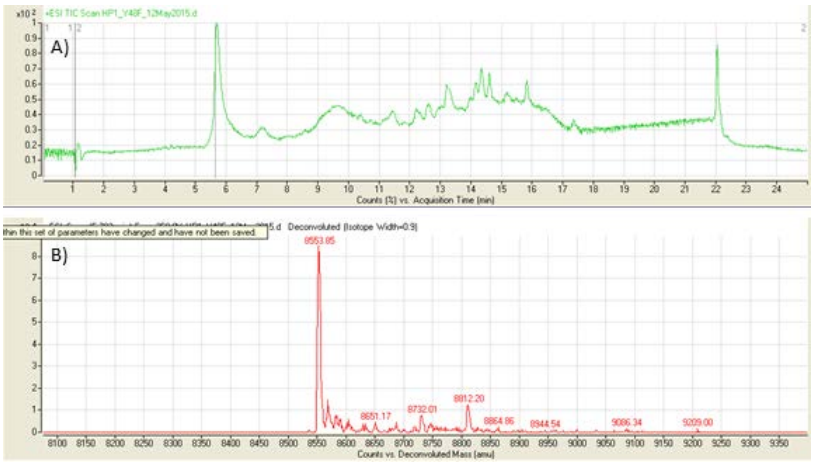


Figure S9. LCMS of HP1 Y48F using method A. ESI scan (A) and corresponding m/z deconvolution (B).

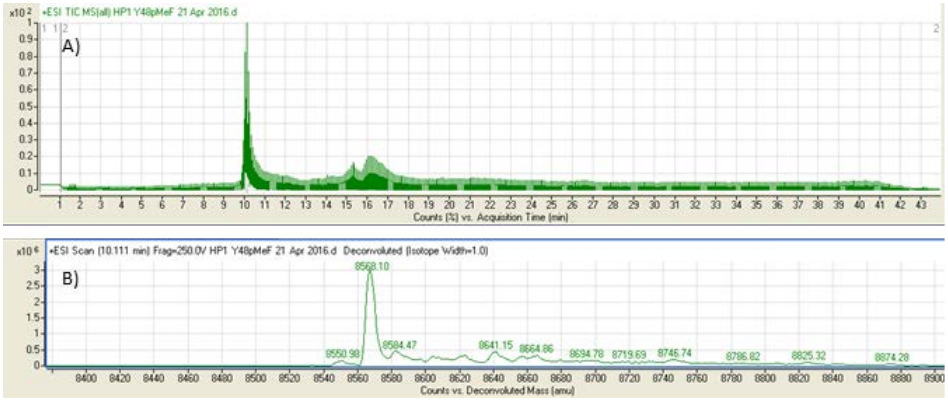


Figure S10. LCMS of HP1 Y48pCH$_3$Phe using method B. ESI scan (A) and corresponding m/z deconvolution (B).

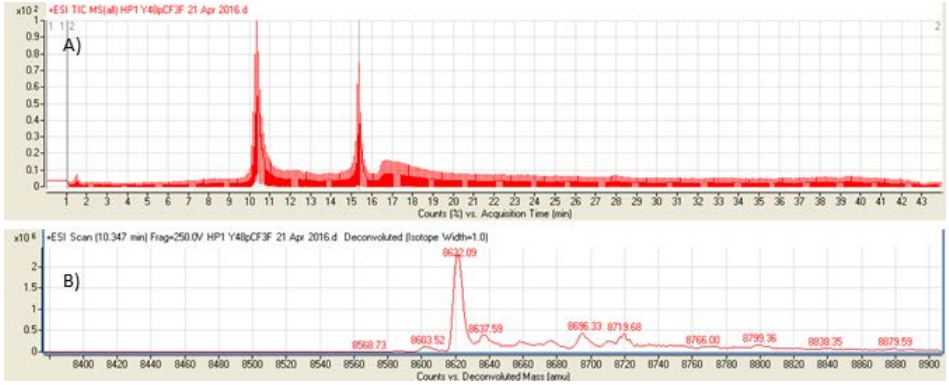


Figure S11. LCMS of HP1 Y48pCF$_3$Phe using method B. ESI scan (A) and corresponding m/z deconvolution (B).

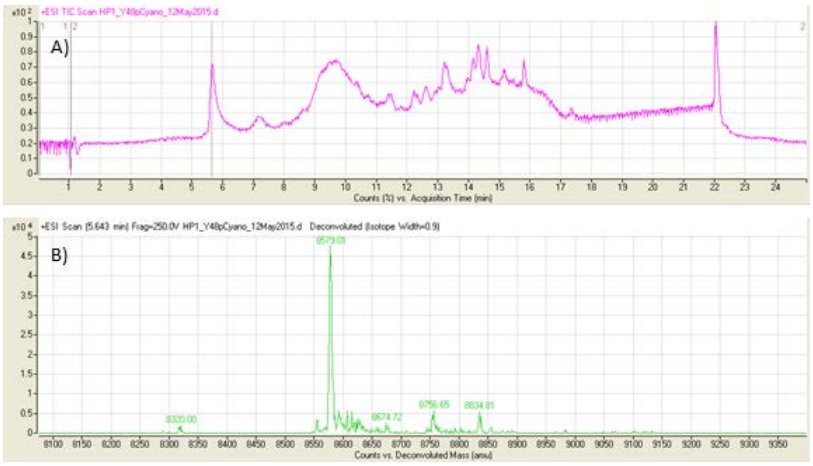


Figure S12. LCMS of HP1 Y48pCNPhe using method A. ESI scan (A) and corresponding m/z deconvolution (B).

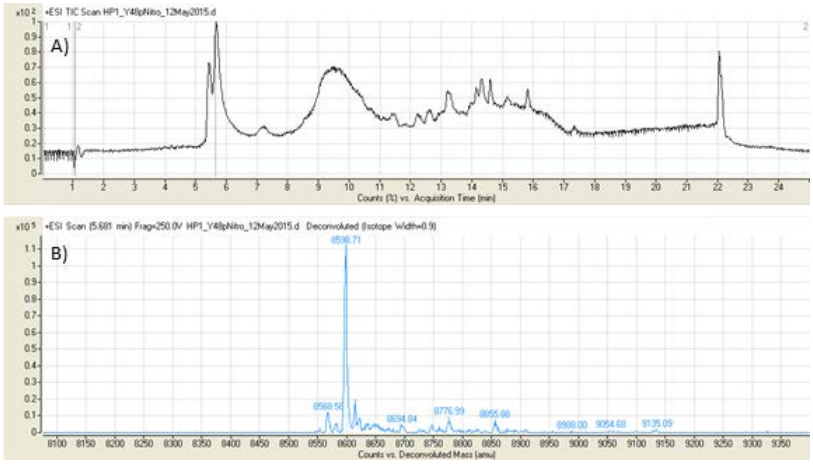


Figure S13. LCMS of HP1 Y48pNO2Phe using method A. ESI scan (A) and corresponding m/z deconvolution (B).

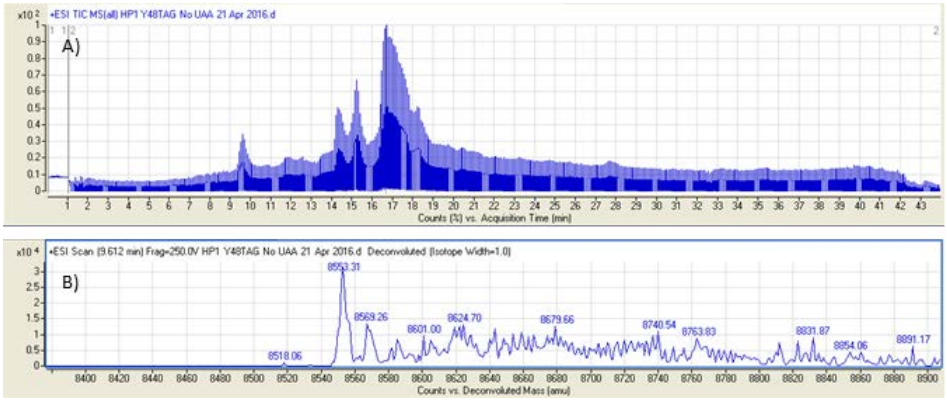


Figure S14. LCMS of HP1 Y48TAG with no UAA added using method B. ESI scan (A) and corresponding m/z deconvolution (B). Wild type HP1 and Y48F are produced in the absence of UAA, but when UAA is added wild type and Y48F are not detected.

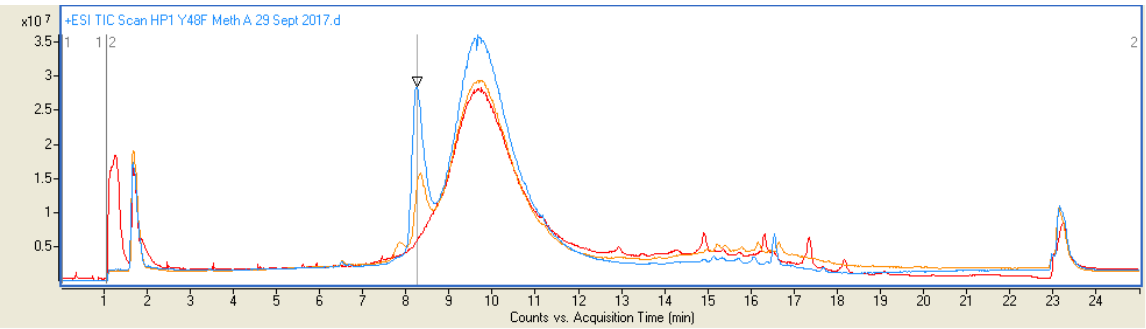
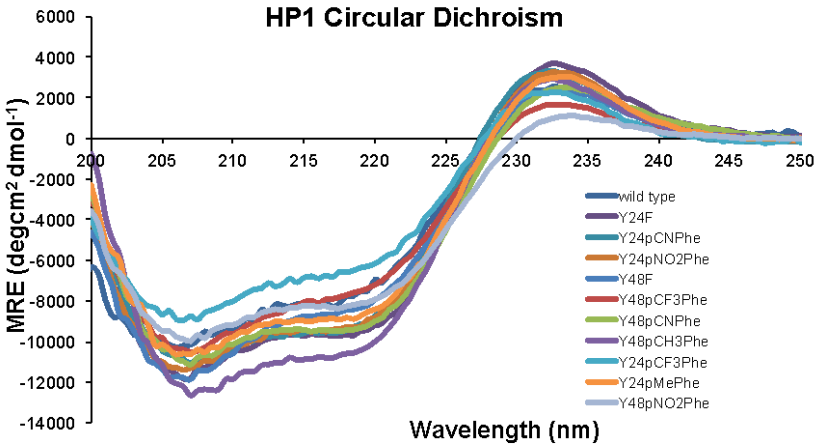


Figure S15. Contaminant LCMS Peaks.

While LCMS traces shown in Figures S2 – S14 are intended to show the fidelity of unnatural amino acid incorporation, other higher retention peaks are also observed. Comparison of LCMS traces from representative protein samples (shown by the arrow; Y48F (blue) and Y48*p*NO2Phe (orange)) and free buffer (red) show that these impurities are independent of the protein sample and are likely highly ionizable contaminants in the LCMS, including small molecule plasticizers.

A)

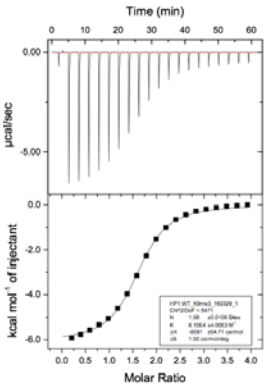


B)

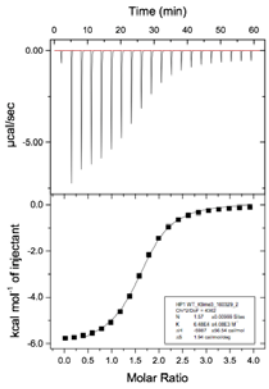
protein	MRE (deg*cm^{2}*dmol^{-1}) 207 nm	218 nm	ratio 207/218	% difference from mean
wild type	-10141.26	-7755.30	1.31	5.11
Y24pClPhe	-11428.75	-8990.84	1.27	1.46
Y48pClPhe	-12226.36	-9754.94	1.25	0.32
Y24F	-11787.34	-9655.03	1.22	3.57
Y24pCNPhe	-11065.97	-9442.52	1.17	8.46
Y24pNO2Phe	-11285.79	-9278.91	1.22	4.03
Y48F	-11869.15	-8428.22	1.41	15.17
Y48pCF3Phe	-10513.59	-7704.74	1.36	10.80
Y48pCNPhe	-11129.10	-9427.92	1.18	7.61
Y48pCH3Phe	-12658.25	-10586.50	1.20	6.08
Y24pCF3Phe	-8739.70	-6579.11	1.33	7.19
Y24pMePhe	-10624.58	-8802.82	1.21	4.96
Y48pNO2Phe	-9975.28	-8247.21	1.21	4.70

Figure S16. Circular Dichroism of HP1 Variants: (A) CD spectra of HP1 variants in 10 mM sodium phosphate, pH 7.4, 2 mM DTT. (B) While some variation in the intensity of CD spectra is observed, the ratio between the characteristic double minima at 207 nm and 218 nm remains consistent between variants (< 15 % error from the mean across all mutants). These data suggest that differences in CD spectra result from small errors in protein concentration determination, and are not representative of altered protein structure.

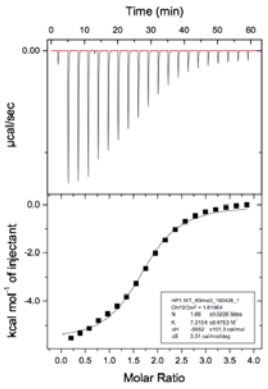
Figure S17. ITC curves of H3K9me3 binding to HP1 mutants



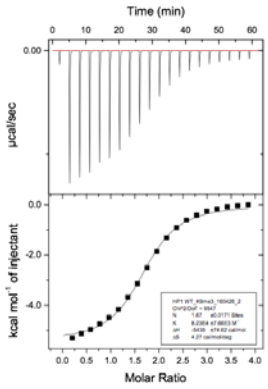
HP1 Wild Type - 1
[Peptide]: 4.95 mM
[Protein]: 255 µM
K_D: 16.39 µM
C: 15.6
N: 1.58



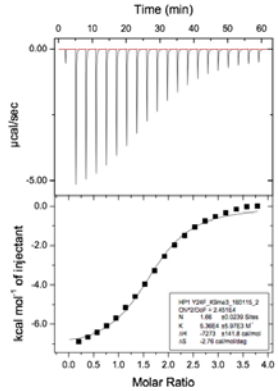
HP1 Wild Type - 2
[Peptide]: 4.95 mM
[Protein]: 255 µM
K_D: 15.43 µM
C: 16.5
N: 1.57



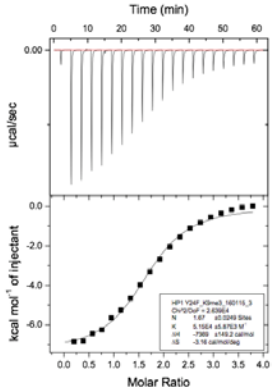
HP1 Wild Type - 3
[Peptide]: 2.86 mM
[Protein]: 150 µM
K_D: 13.68 µM
C: 11.0
N: 1.69



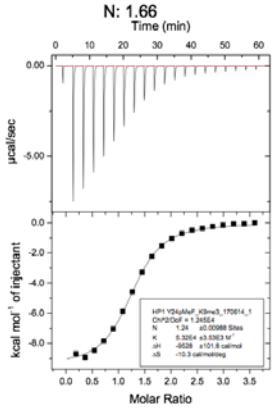
HP1 Wild Type - 4
[Peptide]: 2.86 mM
[Protein]: 150 µM
K_D: 12.15 µM
C: 12.3
N: 1.67



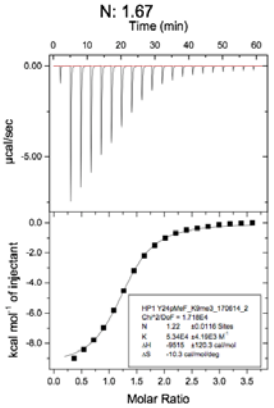
HP1 Y24F - 1
[Peptide]: 3.00 mM
[Protein]: 160 µM
K_D: 18.66 µM
C: 8.6
N: 1.66



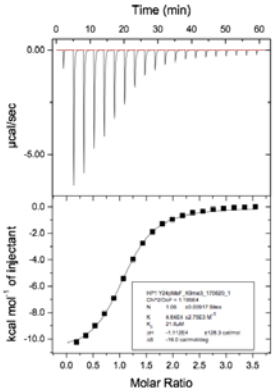
HP1 Y24F - 2
[Peptide]: 3.00 mM
[Protein]: 160 µM
K_D: 19.42 µM
C: 8.2
N: 1.67



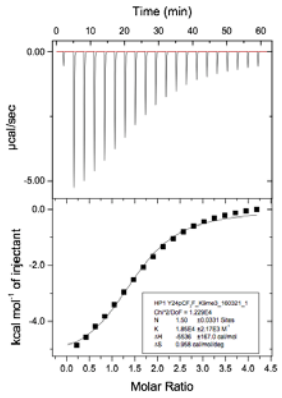
HP1 Y24pMeF - 1
[Peptide]: 4.75 mM
[Protein]: 267 µM
K_D: 18.80 µM
C: 14.2
N: 1.24



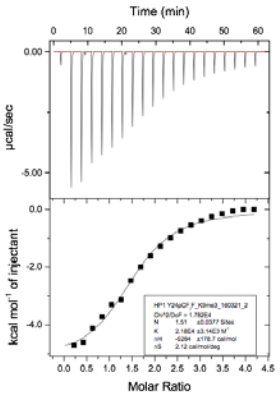
HP1 Y24pMeF - 2
[Peptide]: 4.75 mM
[Protein]: 267 µM
K_D: 18.73 µM
C: 14.3
N: 1.22



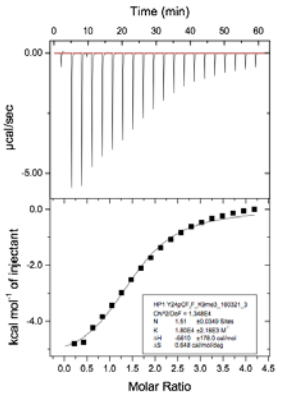
HP1 Y24pMeF - 3
[Peptide]: 4.25 mM
[Protein]: 242 µM
K_D: 21.55 µM
C: 11.2
N: 1.06



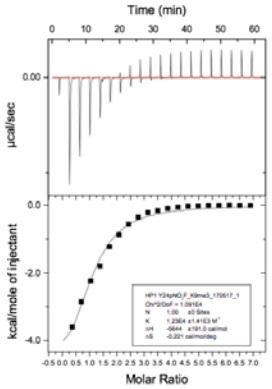
HP1 Y24pCF$_3$F - 1
[Peptide]: 5.18 mM
[Protein]: 250 µM
K_D: 54.05 µM
C: 4.6
N: 1.50



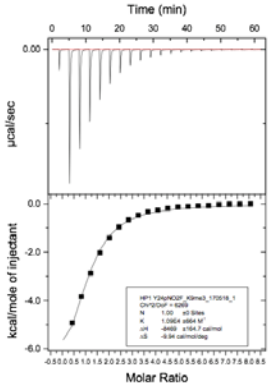
HP1 Y24pCF$_3$F - 2
[Peptide]: 5.18 mM
[Protein]: 250 µM
K_D: 45.87 µM
C: 5.5
N: 1.51



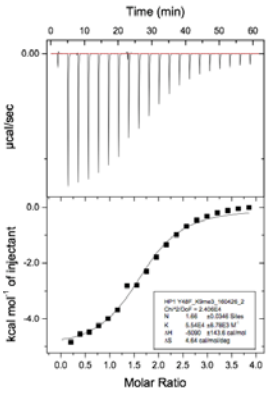
HP1 Y24pCF$_3$F - 3
[Peptide]: 5.18 mM
[Protein]: 250 µM
K_D: 55.56 µM
C: 4.5
N: 1.51



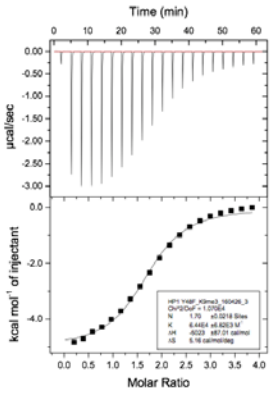
HP1 Y24pNO$_2$F - 1
[Peptide]: 7.00 mM
[Protein]: 205 µM
K_D: 91.74 µM
C: 2.5
N: 1.00



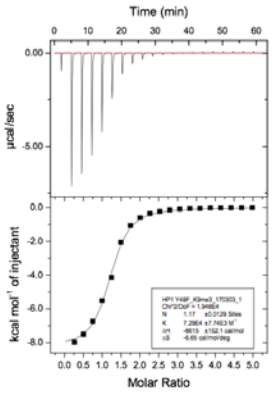
HP1 Y24pNO$_2$F - 2
[Peptide]: 7.47 mM
[Protein]: 187 µM
K_D: 91.74 µM
C: 2.0
N: 1.00



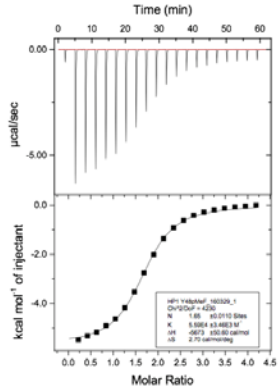
HP1 Y48F - 1
[Peptide]: 2.86 mM
[Protein]: 150 µM
K_D: 18.05 µM
C: 8.3
N: 1.66



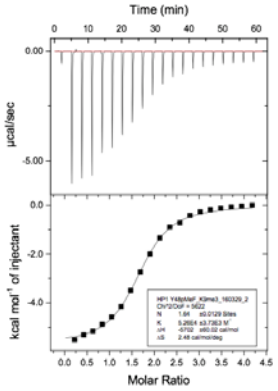
HP1 Y48F - 2
[Peptide]: 2.86 mM
[Protein]: 150 µM
K_D: 15.53 µM
C: 9.7
N: 1.70



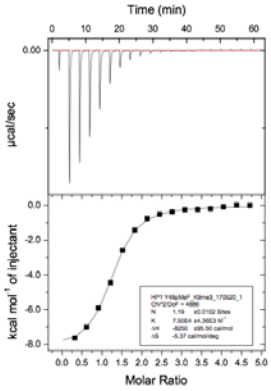
HP1 Y48F - 3
[Peptide]: 5.20 mM
[Protein]: 210 µM
K_D: 13.72 µM
C: 15.3
N: 1.17



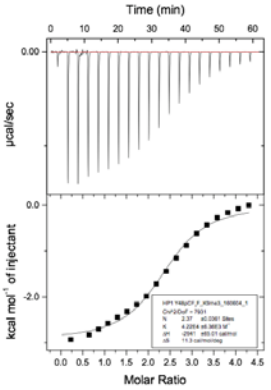
HP1 Y48pMeF - 1
[Peptide]: 4.95 mM
[Protein]: 240 µM
K_D: 17.89 µM
C: 13.4
N: 1.65



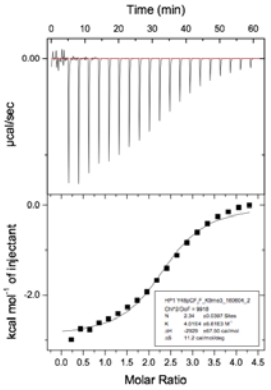
HP1 Y48pMeF - 2
[Peptide]: 4.95 mM
[Protein]: 240 µM
K_D: 19.01 µM
C: 12.6
N: 1.64



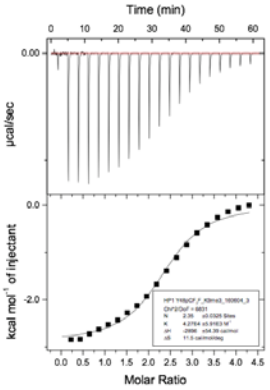
HP1 Y48pMeF - 3
[Peptide]: 4.60 mM
[Protein]: 153 µM
K_D: 13.33 µM
C: 11.5
N: 1.19



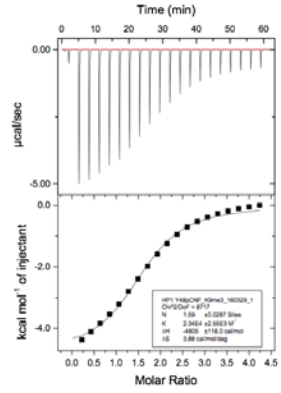
HP1 Y48pCF$_3$F - 1
[Peptide]: 5.00 mM
[Protein]: 235 µM
K_D: 23.70 µM
C: 9.9
N: 2.37



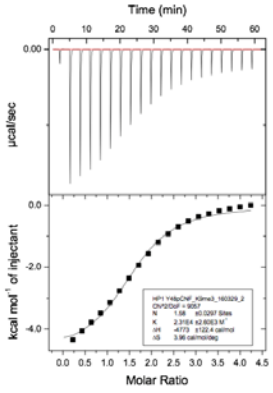
HP1 Y48pCF$_3$F - 2
[Peptide]: 5.00 mM
[Protein]: 235 µM
K_D: 24.94 µM
C: 9.4
N: 2.34



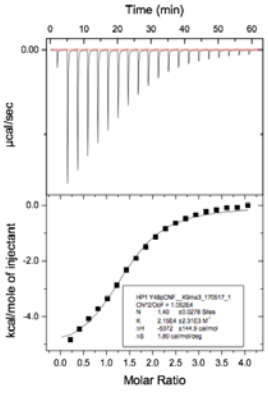
HP1 Y48pCF$_3$F - 3
[Peptide]: 5.00 mM
[Protein]: 235 µM
K_D: 23.42 µM
C: 10.0
N: 2.35



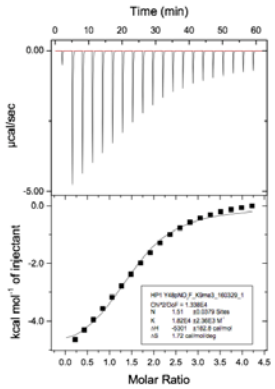
HP1 Y48pCNF - 1
[Peptide]: 4.95 mM
[Protein]: 240 µM
K_D: 42.74 µM
C: 5.6
N: 1.59



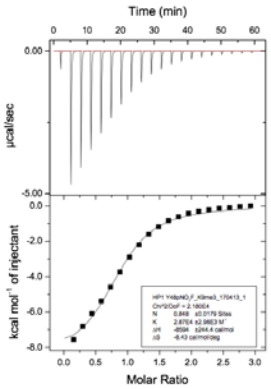
HP1 Y48pCNF - 2
[Peptide]: 4.95 mM
[Protein]: 240 µM
K_D: 43.29 µM
C: 5.5
N: 1.58



HP1 Y48pCNF - 3
[Peptide]: 5.21mM
[Protein]: 259 µM
K_D: 46.51 µM
C: 5.6
N: 1.40



HP1 Y48pNO$_2$F - 1
[Peptide]: 4.95 mM
[Protein]: 240 µM
K_D: 54.95 µM
C: 4.4
N: 1.51



HP1 Y48pNO$_2$F - 2
[Peptide]: 4.10 mM
[Protein]: 280 µM
K_D: 34.8 µM
C: 6.8
N: 0.856

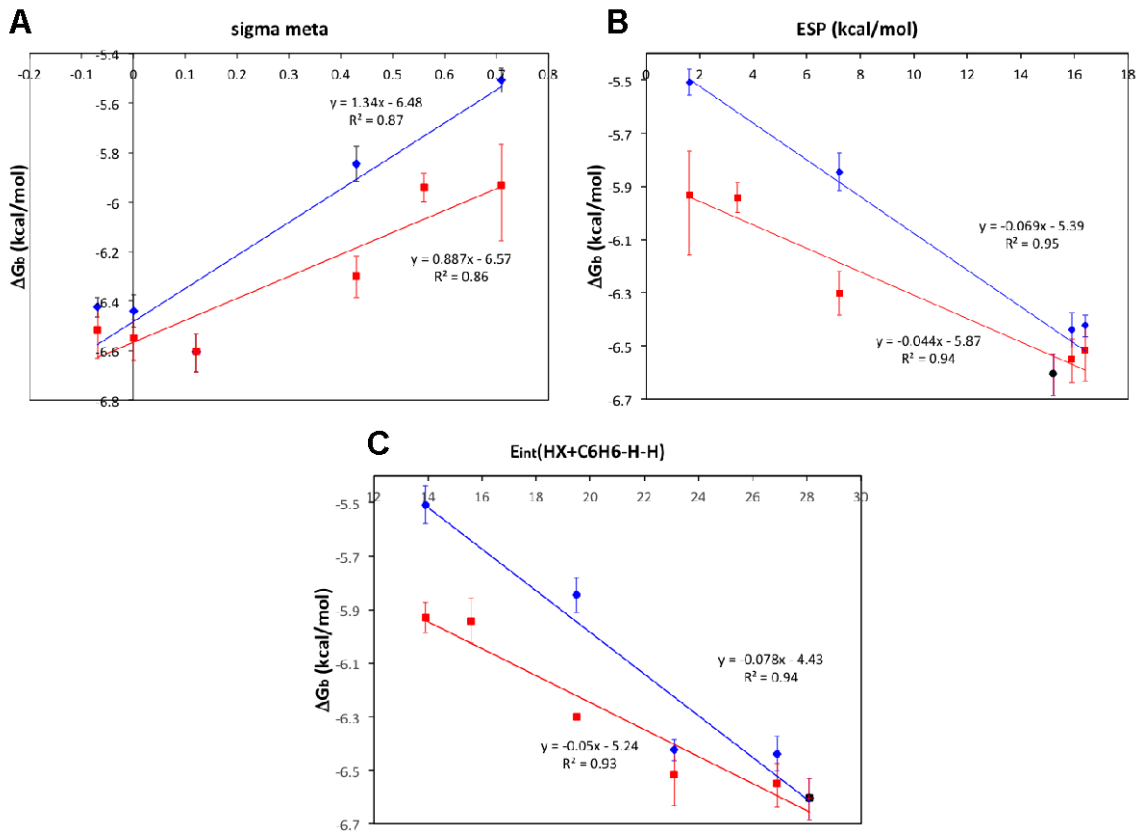


Figure S18. LFER plots of ΔG_b vs. σ_m (A), electrostatic potential (ESP, B), and sum of through space interaction of substituent (HX) plus benzene (C).[16]

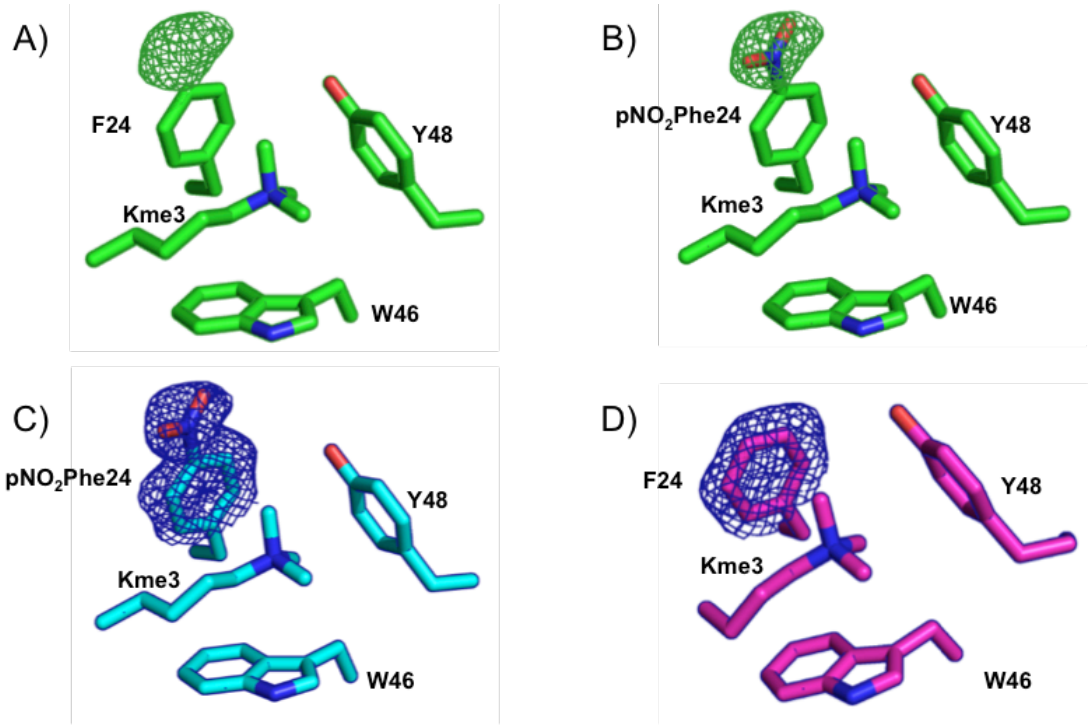


Figure S19. Density maps of HP1 Y24 mutants. A) mFo-DFc map of Y24*p*NO_2Phe density with Y24F mutation shows additional density at *para*-position. B) When Y24*p*NO_2Phe mutation is modeled into the Y24*p*NO_2Phe mFo-DFc density, the nitro group fits the density well. C) 2mFo-DFc density map of the Y24*p*NO_2Phe density with Y24*p*NO_2Phe mutation shows *p*NO_2Phe mutation is present. D) 2mFo-DFc density of Y24F shows the differences in density for the F and Y24*p*NO_2Phe amino acids.

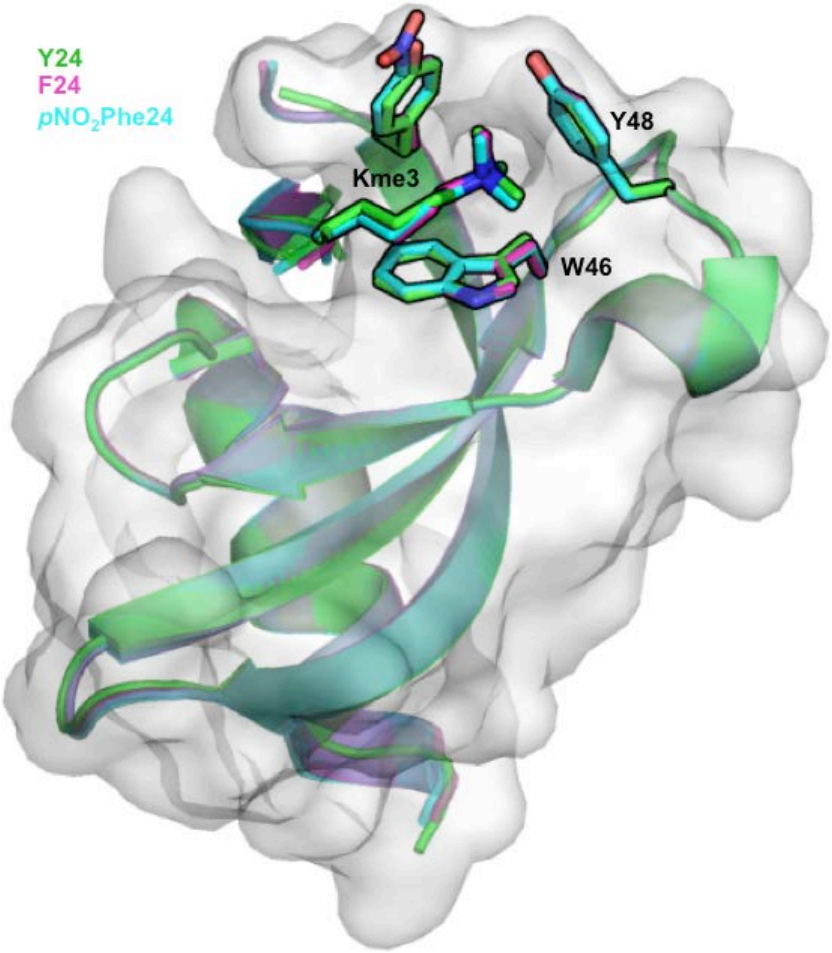


Figure S20. Whole protein overlays of HP1 wild type (green), Y24F (magenta), and Y24*p*NO_2Phe (cyan). Residues of the aromatic cage have been shown as sticks.

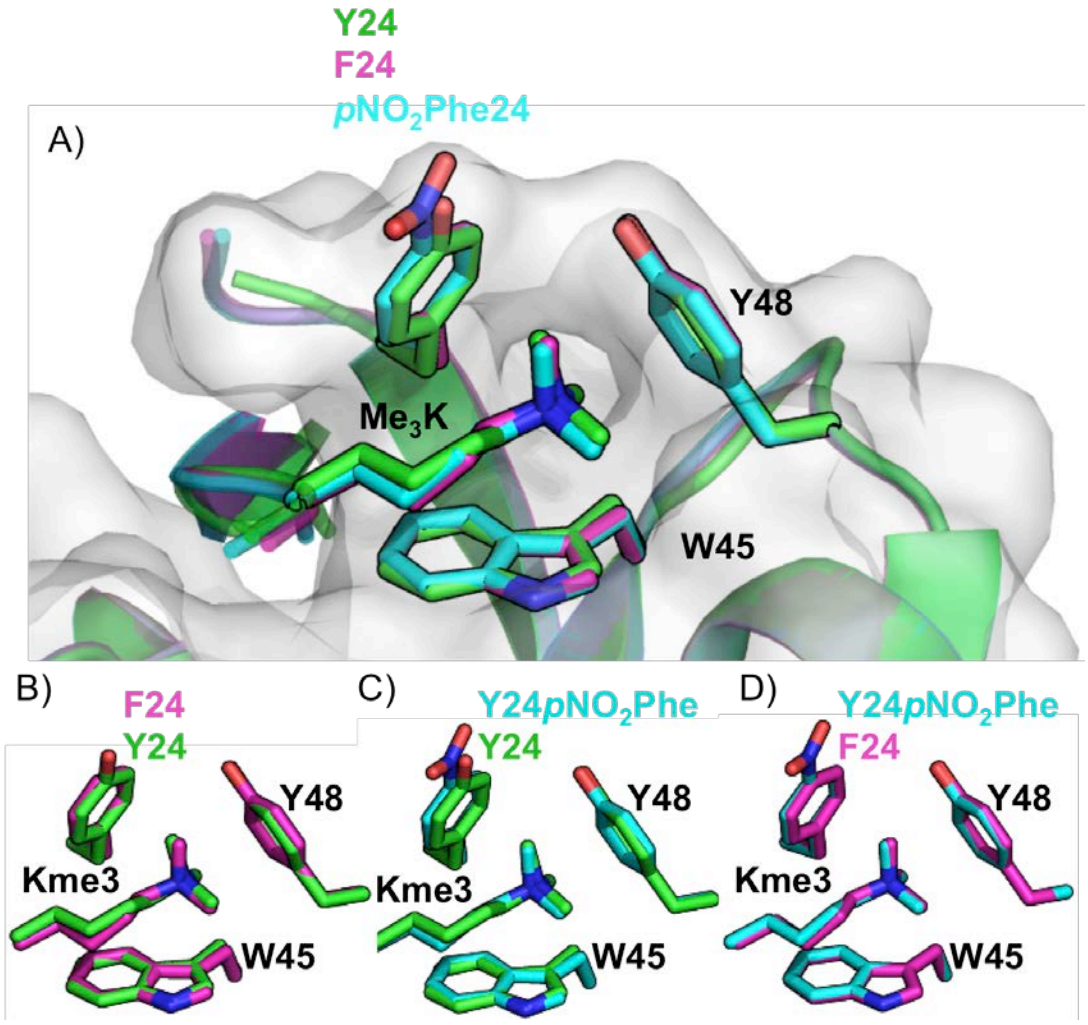
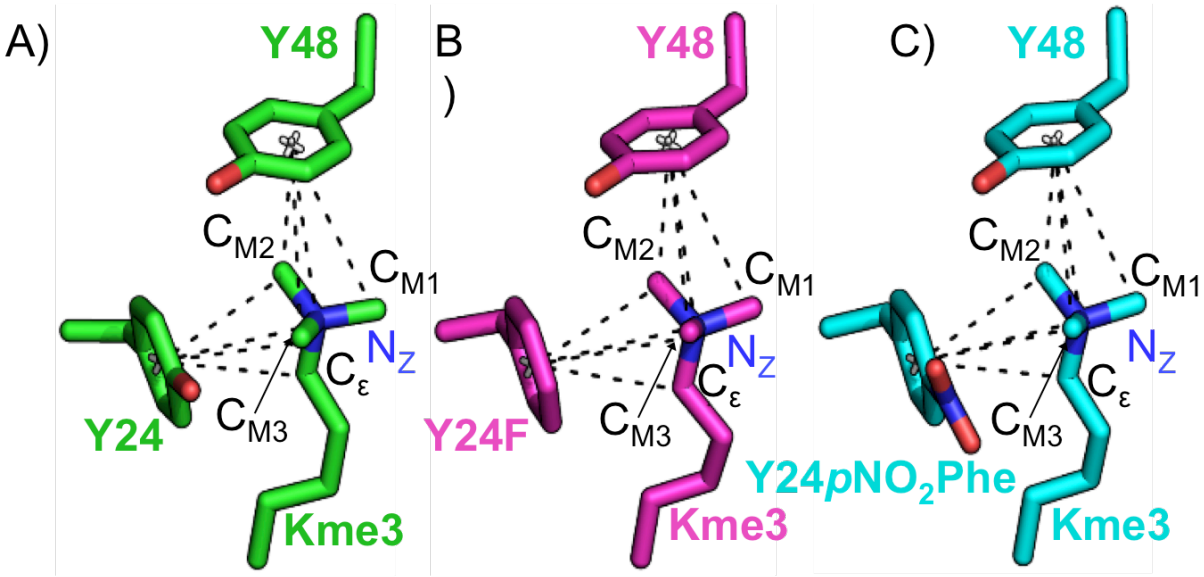


Figure S21. Overlays of the aromatic cage of various HP1 mutants. A) Overlay of Wild type (green), Y24F (magenta), and Y24*p*NO_2Phe (cyan) shows minimal perturbation of the aromatic cage. Wild type surface shown for orientation. B) Wild type and Y24F cage overlay, RMS = 0.222 Å, C) Wild type and Y24*p*NO_2Phe cage overlay, RMS = 0.211 Å, D) Y24F and Y24*p*NO_2Phe cage overlay, RMS = 0.058 Å.



WT Y24F Y24*p*NO_2Phe

D)

Mutant	Distance to 24 Position (Å)				Distance to 48 Position (Å)				
	C_{M2}	C_{M3}	C_ε	N_Z	C_{M1}	C_{M2}	C_{M3}	C_ε	N_Z
WT	4.0	3.6	4.1	4.2	4.5	3.2	4.3	5.5	4.2
Y24F	3.9	4.0	4.1	4.3	4.5	3.5	4.5	5.8	4.4
Y24NO_2Phe	4.1	4.0	4.2	4.4	4.4	3.3	4.3	5.6	4.2

Figure S22. Cation-π distances between Kme3 and 24-and 48-position substituents. A) HP1 wild type (PDB 1KNE), B) HP1 Y24F (PDB 6ASZ), and C) HP1 Y24*p*NO2Phe (PDB 6AT0). D) Measured cation-π distances between Kme3 and center of the aromatic ring at positions 24 or 48.

References

(1) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. *Biochemistry* **2011**, *50*, 1894.
(2) Studier, F. W. *Protein Expr Purif* **2005**, *41*, 207.
(3) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nat Protoc* **2007**, *2*, 2601.
(4) Eisert, R. J.; Waters, M. L. *Chembiochem* **2011**, *12*, 2786.
(5) Turnbull, W. B.; Daranas, A. H. *J Am Chem Soc* **2003**, *125*, 14859.
(6) Sokkalingam, P.; Shraberg, J.; Rick, S. W.; Gibb, B. C. *J Am Chem Soc* **2016**, *138*, 48.
(7) Hwang, J.; Dial, B. E.; Li, P.; Kozik, M. E.; Smith, M. D.; Shimizu, K. D. *Chemical Science* **2015**, *6*, 4358.
(8) Fu, Z. Q. *Acta Crystallogr D Biol Crystallogr* **2005**, *61*, 1643.
(9) Jacobs, S. A.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080.
(10) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213.
(11) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486.
(12) Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2008**, *120*, 215.
(13) Davis, M. R.; Dougherty, D. A. *Phys Chem Chem Phys* **2015**, *17*, 29262.
(14) Gaussian 09, R. A., M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
(15) Legault, C. Y. C., version 1.0b, Université de Sherbrooke, 2009.
(16) Wheeler, S. E.; Houk, K. N. *J Am Chem Soc* **2009**, *131*, 3126.