Supplementary Information

More than π-π-π stacking: Contribution of Amide-π and CH-π Interactions to Crotonyllysine Binding by the AF9 YEATS Domain

Mackenzie W. Krone[1], Christopher R. Travis[1], Ga Young Lee[2], Hannah J. Eckvahl[2], K. N. Houk[2], Marcey L. Waters[*,1]

[1]Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States

[2]Department of Chemistry and Biochemistry, University of California at Los Angeles, 607 Charles E. Young Drive East, Box 951569, Los Angeles, CA 90095-1569, USA

ORCID

Marcey L. Waters: 0000-0002-4917-5755

Kendall N. Houk: 0000-0002-8387-5261

Mackenzie W. Krone: 0000-0002-8602-3742

Christopher R. Travis: 0000-0003-0673-6978

Ga Young Lee: 0000-0002-7453-7148

Materials and Methods

General:

Oligonucleotides were obtained from Integrated DNA Technologies. Enzymes and reagents used for cloning were obtained from New England BioLabs Inc. Unnatural amino acids were purchased from Chem-Impex. 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. The plasmid pULTRA-*p*CNFRS was obtained from the laboratory of Prof. Peter Schultz (Scripps Research Institute). Mass spectrometry was performed using an Agilent Technologies Rapid Resolution LC-MSD system (1200 Series).

Computational:

The geometries of the AF9–H3K9cr and AF9–H3K9ac complexes were extracted from previously reported crystal structures (PDB: 5HJB and 4TMP, respectively) and truncated to include only the amino acid side chains.[1, 2] Each C and N terminus of the amino acid residues was replaced with a hydrogen atom with C-H bond distance of 1.09 Å. Then, electronic interaction energies, E_{int}, were calculated for each and all three aromatic amino acids with the substrates (dimer and tetramer, respectively) using single-point energy calculations at the M06-2X/6-311+G(d,p) level of theory.[3] Implicit solvent model (SMD)[4] was employed to account for solvation energies, where pentyl ethanoate was applied for the complex and each amino acid to mimic the hydrophobic cavity while water was used for the substrate. The interaction energies are defined as the energy difference between the complex and each amino acid(s) and the substrate (**Equation 1**).

$$E_{int} = E_{complex}(pentyl\ ethanoate) - [E_{a.a.\ sidechain}(pentyl\ ethanoate) + E_{substrate}(water)]$$

(Equation 1)

All quantum chemical calculations were performed using Gaussian 16.[5] All graphics on optimized structures were generated with CYLview.[6]

Cloning:

The human AF9 YEATS domain gene (residues 1-149) was obtained from the lab of Dr. Brian Strahl (UNC-Chapel Hill) and inserted into a pET28a(+) vector (Addgene) with the BamHI and XhoI restriction sites and an N-terminal His tag. For NMR experiments, the AF9 gene was truncated to residues 1-138 and cloned into the NdeI and XhoI restriction sites of the pET28a vector. Canonical mutations of the three aromatic cage residues (F28Y, F59Y, and Y78F) were generated using overlap extension PCR methods. These methods were also used to install TAG stop codons at each of these positions (F28TAG, F59TAG, and Y78TAG) in order to incorporate unnatural amino acids.

Protein Sequences:

AF9 wild type (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**F**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S**F**PRPKRVCKDPPYKVEESG**Y**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLR

CEKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 F28Y (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**Y**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S**F**PRPKRVCKDPPYKVEESG**Y**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLR

CEKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 F28• (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG•THDWMVFVRGPEHSNIQHFVEKVVFHLHES

FPRPKRVCKDPPYKVEESG**Y**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLRC

EKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 F59Y (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**F**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S**Y**PRPKRVCKDPPYKVEESG**Y**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLR

CEKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 F59• (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**F**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S•PRPKRVCKDPPYKVEESG**Y**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLRC

EKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 Y78F (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**F**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S**F**PRPKRVCKDPPYKVEESG**F**AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLR

CEKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

AF9 Y78• (1-149) protein sequence:

MASSCAVQVKLELGHRAQVRKKPTVEG**F**THDWMVFVRGPEHSNIQHFVEKVVFHLHE

S**F**PRPKRVCKDPPYKVEESG•AGFILPIEVYFKNKEEPRKVRFYDLFLHLEGHPPVNHLRC

EKLTFNNPTEDFRRKLLAKAGGDPNRSIHTS

Protein Expression:

Plasmids pET28a_AF9_WT, pET28a_AF9_F28Y, pET28a_AF9_F59Y, and pET28a_AF9_Y78F were each transformed into BL21(DE3)-RIL cells (Agilent). Single colonies from each LB-agar plate were grown and used to inoculate 1 L expression cultures of LB media supplemented with kanamycin (50 μg/mL). Cultures were allowed to shake at 215 rpm at 37 °C until cells grew to an OD_{600} = 0.7, at which point 0.5 mM IPTG was added to induce protein expression. The temperature was decreased to 18 °C, and cells were allowed to continue shaking for 16 hours post-induction. Cells were harvested via centrifugation.

Plasmids pET28a_AF9_F28TAG, pET28a_AF9_F59TAG, and pET28a_AF9Y78TAG were each co-transformed with plasmid pULTRA-*p*CNFRS into BL21(DE3)-RIL cells. Single colonies from each LB-agar plate were grown to density and used to inoculate expression cultures. Cells were grown in 1 L of LB media, supplemented with kanamycin (50 μg/mL) and streptomycin (50 μg/mL). Cultures were allowed to shake at 215 rpm at 37 °C until cells grew to an OD_{600} = 0.5, at which point 2 mM of the desired unnatural amino acid (*p*CH_3F, *p*ClF, *p*CNF, or *p*NO_2F) was added as a dry solid. After cells had grown to an OD_{600} = 0.7, 0.5 mM IPTG was added to induce protein expression. The temperature was decreased to 18 °C, and cells were allowed to continue shaking for 16 hours post-induction. Cells were harvested via centrifugation.

Protein Purification:

The following procedure was used to purify all AF9 protein constructs. Cell pellets were resuspended in lysis buffer (50 mM Tris, 500 mM NaCl, 30 mM imidazole, 2 mM EDTA, 0.1% Triton, 0.5 mg/mL lysozyme, pH 7.4) and allowed to shake at 37 °C for 30 minutes. Samples were then sonicated on ice and centrifuged at 15000 rpm for 1 hour. Following centrifugation, supernatant was filtered to 0.22 μm. Filtered lysate was then purified on an AKTA (GE) FPLC using a HisTrap 5 mL HP column (GE). Using buffer A (50 mM Tris, 500 mM NaCl, 30 mM imidazole, 2 mM DTT, pH 7.4) and buffer B (50 mM Tris, 500 mM NaCl, 300 mM imidazole, 2 mM DTT, pH 7.4), samples were 6xHis-tag purified using a gradient of 0-55% buffer B. Fractions of interest were pooled and concentrated on a 3 kDa Amicon Ultra-15 Centrifugal filter. For NMR samples, the pooled His fractions were exchanged into PBS, and the 6xHis tag was removed by overnight thrombin cleavage at room temperature (Sigma). Concentrated samples were then purified by size exclusion chromatography (SEC) using a Superdex 75 Increase size exclusion column equilibrated in SEC buffer (50 mM sodium phosphate, 500 mM NaCl, 2 mM EDTA, 2 mM β-mercaptoethanol, pH 7.4). For wild type AF9 as well as all mutant proteins, separation of desired monomeric AF9 protein from a larger mass peak presumed to be aggregated protein occurred during SEC. Protein identity and purity was confirmed through SDS-PAGE and ESI-LCMS. Fractions of interest were pooled, concentrated, and stored at -20 °C until further use.

Peptide Synthesis and Purification:

Residues 4-15 of H3 with crotonylated lysine at position 9 (Sequence: KQTARKcrSTGGKAY, hereafter denoted as H3K9cr) was synthesized on a CEM Liberty Blue peptide synthesizer using Fmoc protected amino acids and Rink Amide AM resin (CEM). A

tyrosine residue was placed at the C-terminus for concentration determination. Amino acid residues were activated with Oxyma (ethyl cyanohydroxyiminoacetate) and DIC (N,N'-diisopropycarbodiimide). This synthesis was done on a 0.25 mmol scale. For each coupling step, 5 equivalents of each the desired amino acid, Oxyma, and DIC were used. Two coupling cycles of 4 minutes each were performed at 90 °C in DMF (dimethylformamide) for each residue. Deprotections of Fmoc were carried out in a solution of 20% piperidine in DMF for 1 minute and were repeated twice. The resin was washed with DMF before each coupling cycle and deprotection cycle.

Crotonyllysine was incorporated into the H3 peptide at position 9 by first double coupling 2 equivalents of Fmoc-Lys(Mtt)-OH for 4 minutes at 90 °C in DMF. Immediately after this coupling, the resin was removed from the CEM Liberty Blue peptide synthesizer, transferred to a peptide flask, and washed with DMF followed by DCM (dichloromethane). Then, an Mtt deprotection solution comprised of 1% TFA (trifluoroacetic acid) and 2% TIPS (triisopropylsilane) in DCM was added and the solution was bubbled with N_2 for 30 minutes. The resin was then washed with DCM three times and then returned to the CEM Liberty Blue peptide synthesizer. Crotonic acid (2 equivalents) was double coupled for 4 minutes at 90 °C in DMF. Following this, synthesis of the peptide continued with the previously described conditions.

The peptide was then cleaved from the resin by adding a solution of 2.5% water and 2.5% TIPS in TFA and allowing it to bubble with N_2 for 3 hours. The TFA was then evaporated and cold diethyl ether was added. The desired peptide was then extracted into water and lyophilized. Crude peptide was purified by reverse phase HPLC using a C18 XBridge 5 uM column (Waters) with a gradient of 0-20% solvent B in 60 minutes. Solvent A was comprised of 95% water, 5%

acetonitrile, and 0.1% TFA. Solvent B was comprised of 95% acetonitrile, 5% water, and 0.1% TFA. The purified peptides were lyophilized, and product was confirmed by ESI-LCMS.

Residues 4-15 of H3 with acetylated lysine at position 9 (Sequence: KQTARKacSTGGKAY, hereafter denoted as H3K9ac) was synthesized in the same fashion, with the only exception at position 9. For the acetylated peptide, Fmoc-Lys(ac)-OH was double coupled for 4 minutes at 90 °C in DMF at position 9. Following this, peptide synthesis, deprotection, and purification were executed in the same manner. Product purity was confirmed by ESI-LCMS.

Circular Dichroism (CD):

Thermal denaturation experiments were performed using an Applied Photophysics Chirascan with 10 μM AF9 protein in CD buffer (25 mM sodium phosphate, 500 mM NaF, pH 7.4). Measurements were taken at 208 nm between 20 °C to 85 °C using a quartz cuvette with a path length of 0.1 cm, with a 1 degree temperature ramp and 30 second settling time. The mean residue ellipticity (MRE) was calculated using **Equation 2**, where θ is MRE, *signal* is millidegree measurement, *l* is path length in cm, *c* is protein concentration in mM, and *r* is the number of amino acid residues.

$$\theta = \frac{signal}{10 \cdot l \cdot c \cdot r} \qquad \textbf{(Equation 2)}$$

Melting curves were then generated using **Equation 3** to show the fraction folded (α). θ represents the calculated mean residue ellipticity (MRE), θ_D represents the MRE for the fully denatured protein, and θ_F represents the MRE for the fully folded protein.

$$\boldsymbol{\alpha = \frac{(\theta - \theta_D)}{(\theta_F - \theta_D)}} \qquad \textbf{(Equation 3)}$$

Protein NMR:

^{15}N-labeled wild-type and mutant AF9 proteins were expressed and purified as described earlier with the exception of the bacterial growth media. Cultures were grown in ^{15}N-enriched M9 media (1g of ^{15}N labelled NH_4Cl per 1L of M9) in addition to the corresponding antibiotics. All other parameters of expression and purification are equivalent to those reported for the non-isotopically labelled proteins.

^{15}N-labeled purified proteins (400 μM) were exchanged into NMR buffer (10 mM Na_2HPO_4 pH 7.4; 500 mM NaCl; 2 mM EDTA; 2mM DTT; 90% H_2O; 10% D_2O) using Amicon filters (3K MWCO). The HSQC spectra were collected at 298K on a Bruker Avance III 850 MHz spectrometer equipped with a cryogenic QCI probe (Figure S6). For ligand-bound experiments, H3K9cr or H3K9ac was added (1 mM) to achieve a 1:2.5 protein:peptide molar ratio (Figure S7). Data was processed using NMRPipe and analyzed using the CcpNMR Analysis software.[7] The resonance assignments for AF9-YEATS bound to H3K18ac and H3K18cr were used as a reference for identifying the resonance peaks of interest in the collected spectra (BMRB ID: 26059 and 26060).

Chemical shift differences (CSD) between AF9 WT and F28Y spectra were calculated using **Equation 4** with a scaling factor (α) of 0.14.[8] *d* is CSD (in ppm), δ_H is the chemical shift change for hydrogen nuclei, and δ_H is the chemical shift change for nitrogen nuclei. CSD values greater than 0.1 ppm were considered significant.

$$d = \sqrt{\frac{1}{2}\left[\delta_H^2 + (\alpha \cdot \delta_N^2)\right]} \quad \textbf{(Equation 4)}$$

Isothermal Titration calorimetry (ITC):

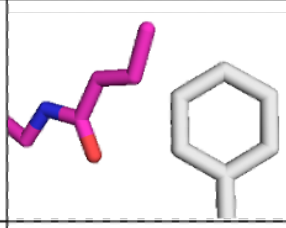
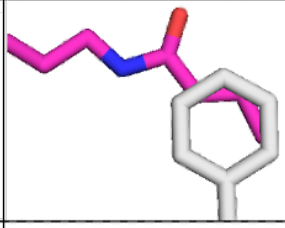
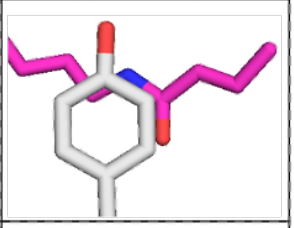
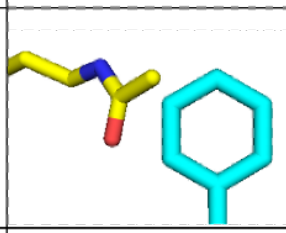
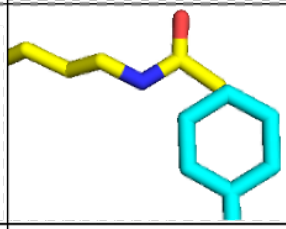
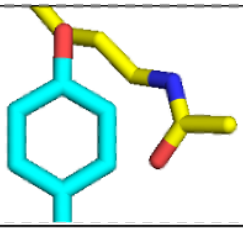
Both peptide and protein samples were prepared in AF9 SEC buffer (50 mM sodium phosphate, 500 mM NaCl, 2 mM EDTA, 2 mM β-mercaptoethanol, pH 7.4) and concentrations

were determined via measuring absorbance at 280 nm with a Nanodrop 2000 (Thermo). ITC experiments were performed on a MicroCal PEAQ-ITC Automated (Malvern Instruments) by titrating peptide (either H3K9cr or H3K9ac) into AF9 wild type or mutant protein at 25 °C. Data was modeled using the One-Site binding model supplied in MicroCal PEAQ-ITC Analysis software (version 1.1.0.1262) with the fitted offset control. Replicate ITC experiments were performed for each protein/peptide pair. Errors given for K_d, $\Delta G_{binding}$, ΔH, and TΔS are the standard deviation from three runs.

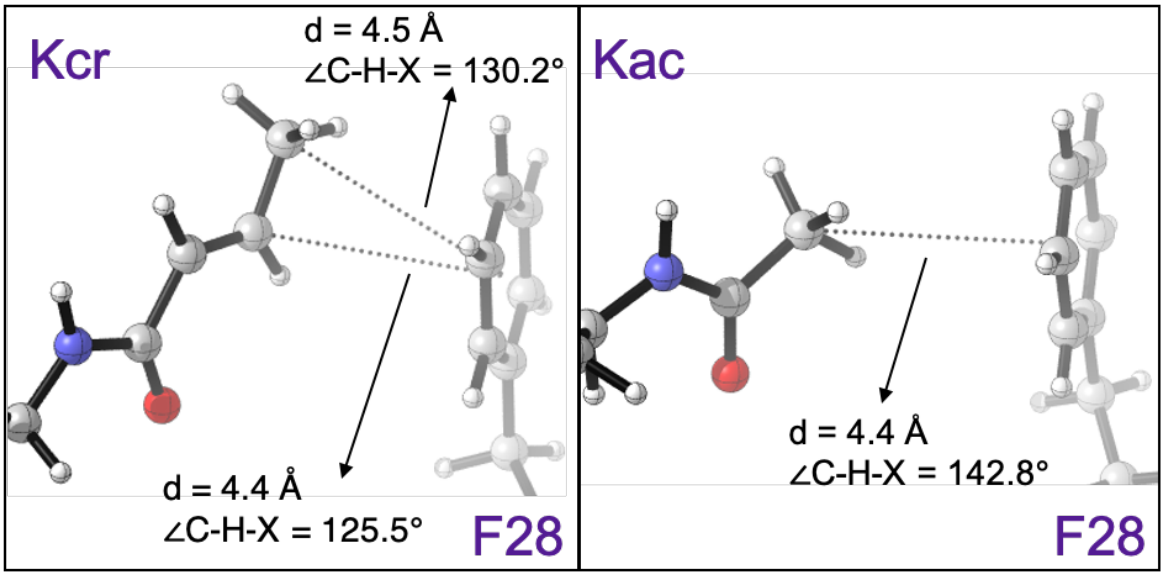
Supplementary Figures

Supplementary Figure S1. (a) Images of Kcr and Kac interacting with residues 28, 59, and 78 in AF9 viewed normal to the plane of the aromatic ring. Reported crystal structures were used (PDB: 5HJB, 4TMP).[1, 2] (b) and (c) CH-π orientations for Kcr and Kac interacting with Phe28 and a table of statistically preferred versus observed geometries in AF9.[9] Distance is shown between carbon atom and center of Phe28 ring (*d*) and angle of H atom between carbon atom and centroid (∠C-H-X)

a.

AF9 YEATS Domain	Residue Number: Phe28	Phe59	Tyr78
w/ H3K9cr			
	E_{int} = -2.2 kcal/mol	E_{int} = -3.8 kcal/mol	E_{int} = -3.3 kcal/mol
w/ H3K9ac			
	E_{int} = -0.5 kcal/mol	E_{int} = -2.0 kcal/mol	E_{int} = -1.1 kcal/mol

b.



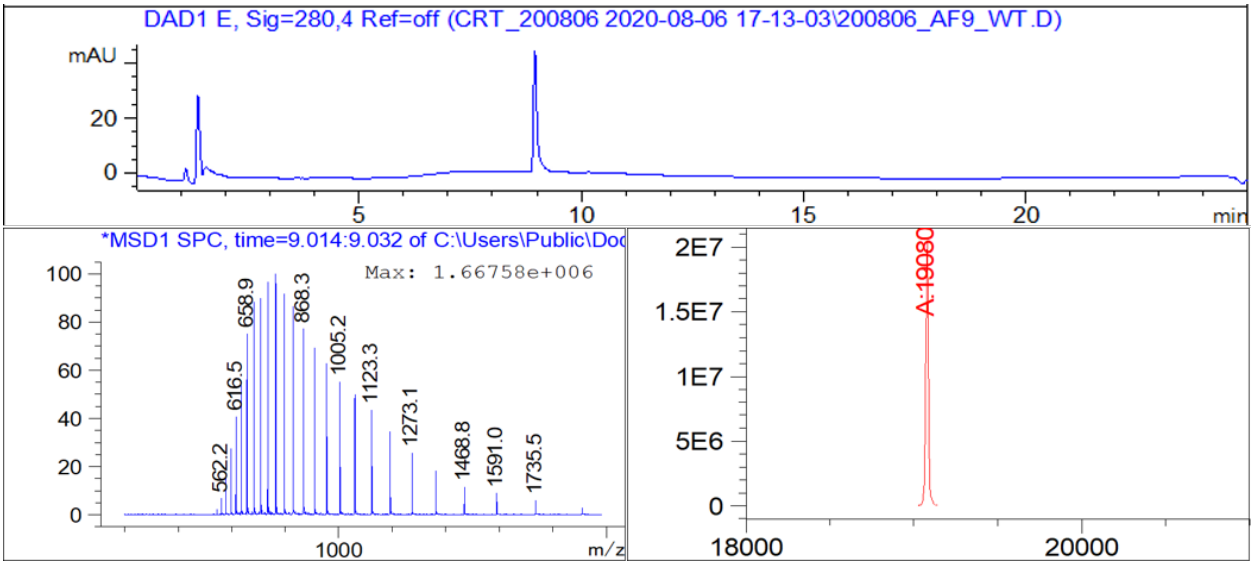
(c) Statistically Preferred and AF9 CH-π Interaction Geometries

Interaction	∠C-H-X (Å)	d_{CH-X}	d_{Hp-X} (Å)
Statistical	≥ 120°	≤ 4.5	1.0-1.2
Kcr CH_3	130.2°	4.5	2.2
Kcr CH	125.5°	4.4	3.0
Kac CH_3	142.8°	4.4	3.0

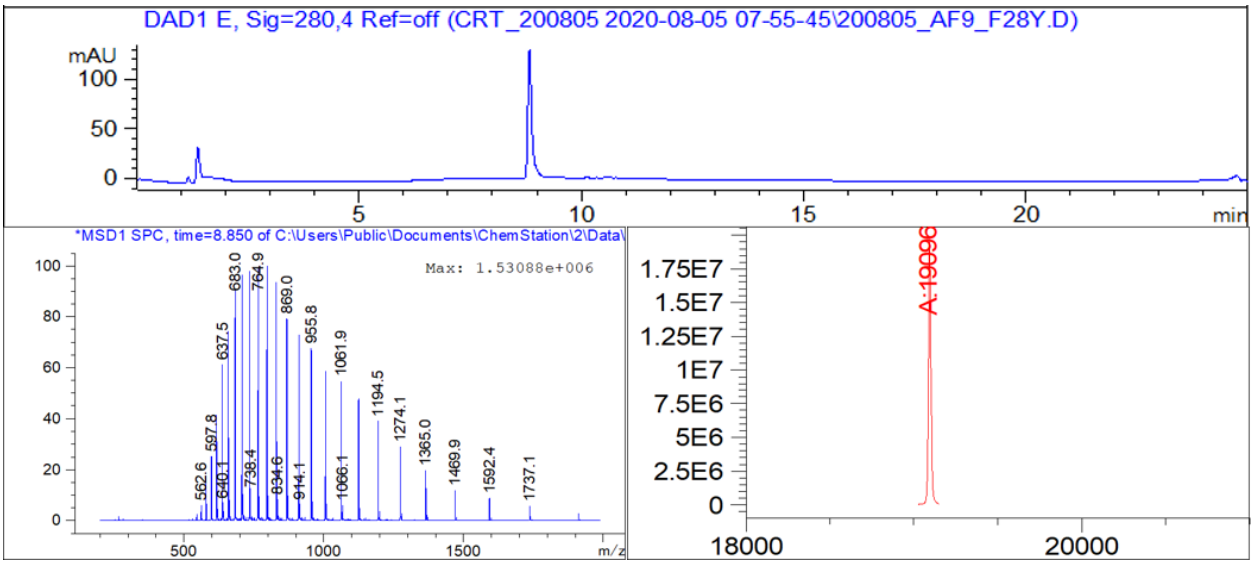
Supplementary Figure S2. ESI-LCMS data for AF9 proteins. Observed masses correlate well with expected masses.

Protein	**Expected Mass (amu)**	**Observed Mass (amu)**
AF9 WT	19080.8	19079.8
AF9 F28Y	19096.8	19095.8
AF9 F28*p*CH_3F	19094.8	19093.8
AF9 F28*p*ClF	19114.3	19114.2
AF9 F28*p*CNF	19105.8	19104.2
AF9 F28*p*NO_2F	19125.8	19124.6
AF9 Y78*p*CH_3F	19078.8	19077.7
AF9 Y78*p*ClF	19098.3	19098.2
AF9 Y78*p*CNF	19089.8	19088.4
AF9 Y78*p*NO_2F	19109.8	19108.2

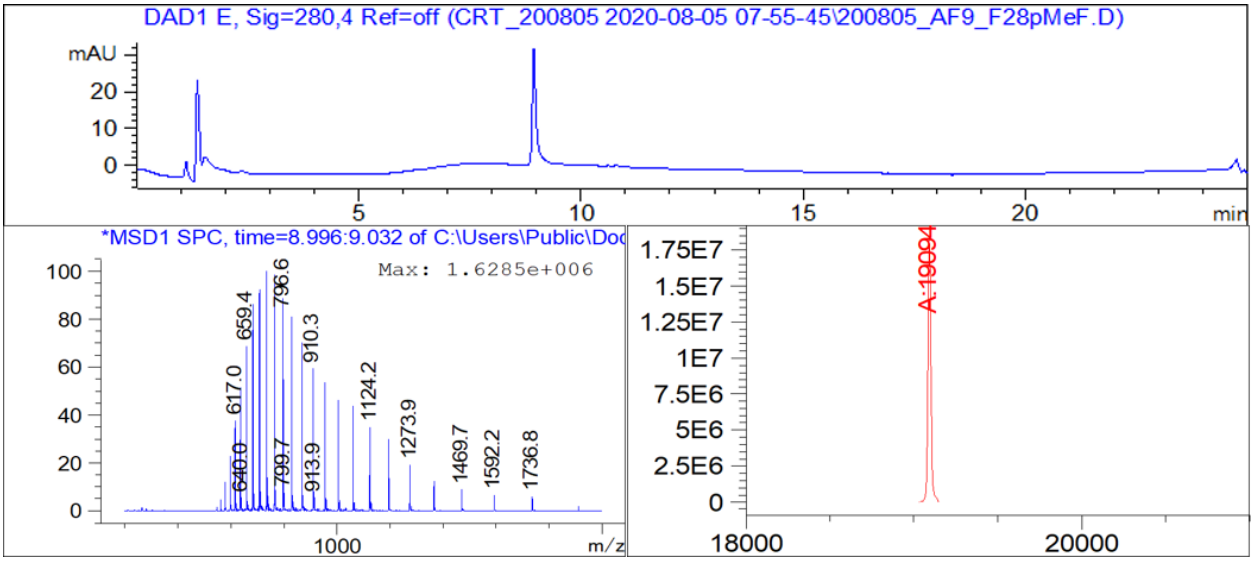
AF9 WT



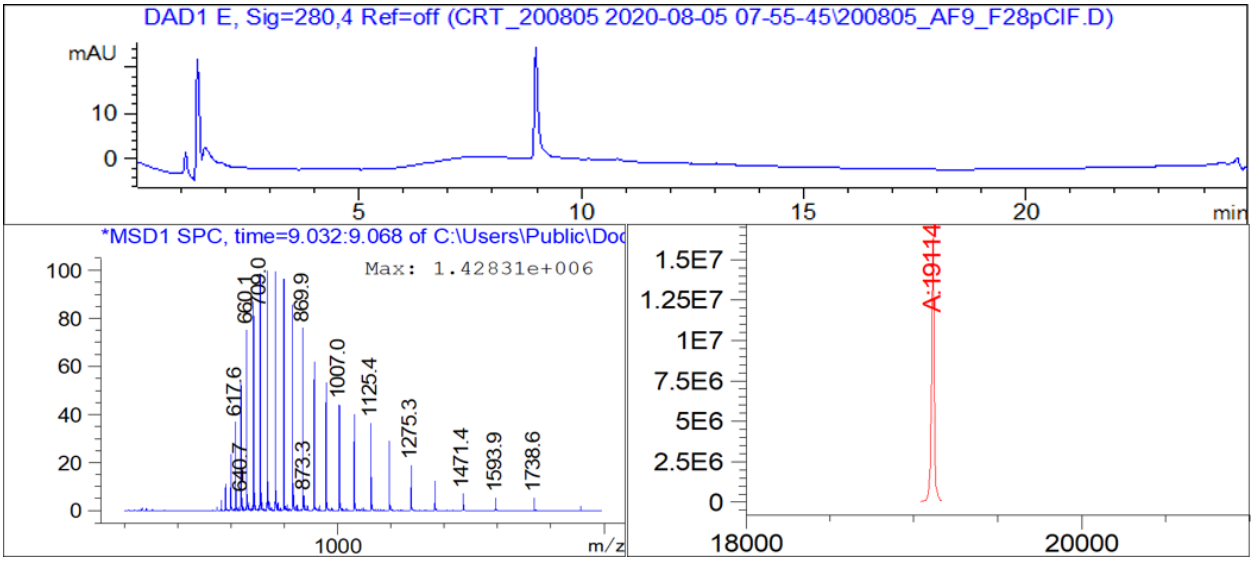
AF9 F28Y



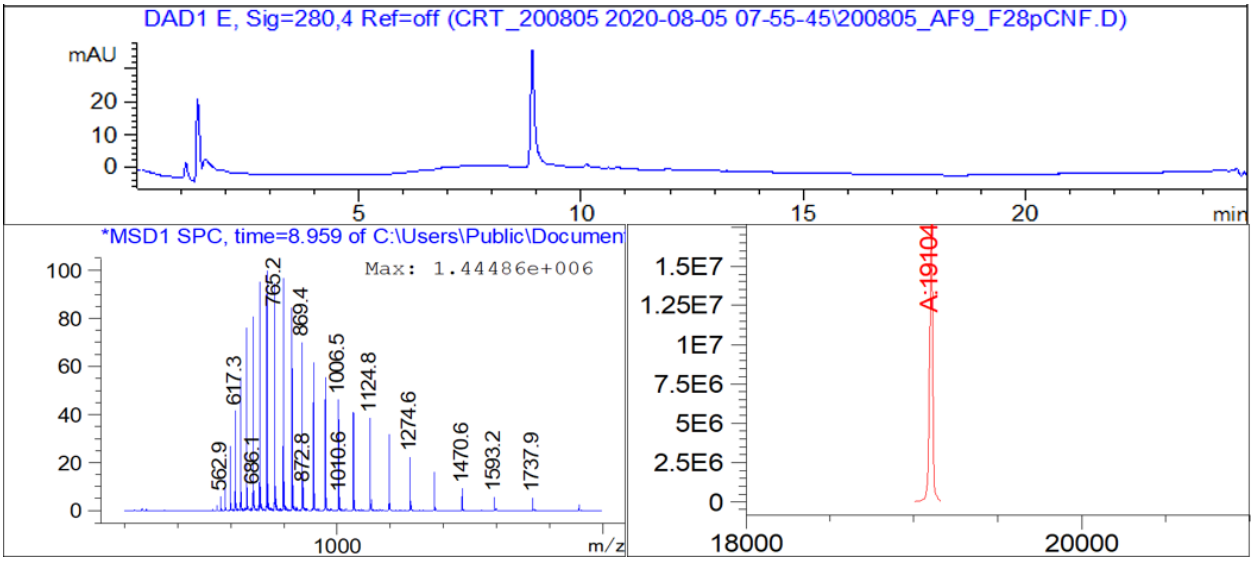
AF9 F28*p*CH$_3$F



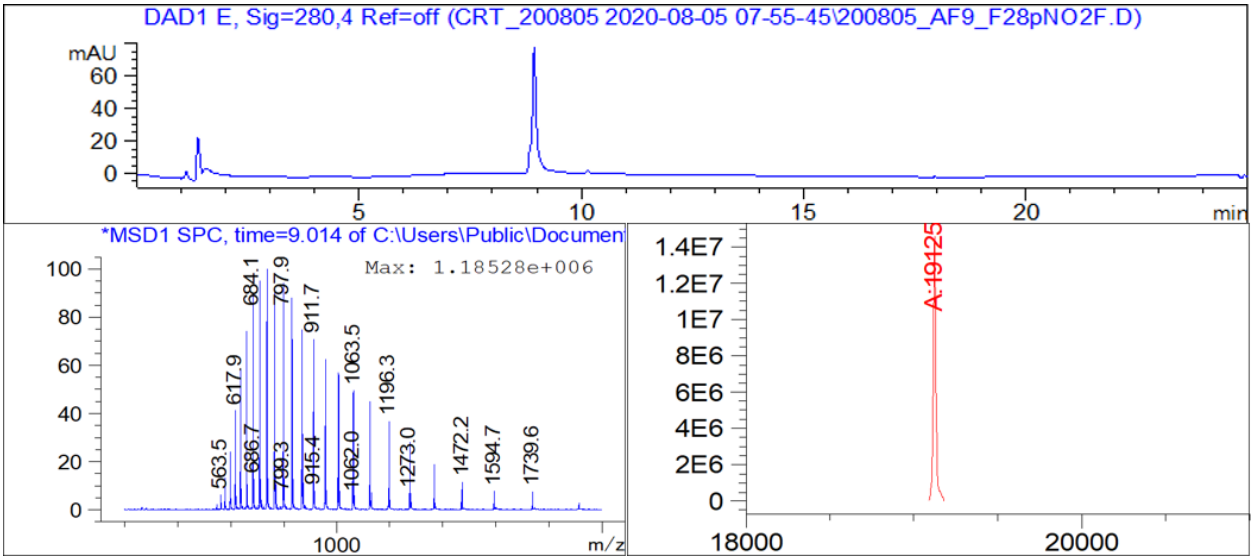
AF9 F28*p*ClF



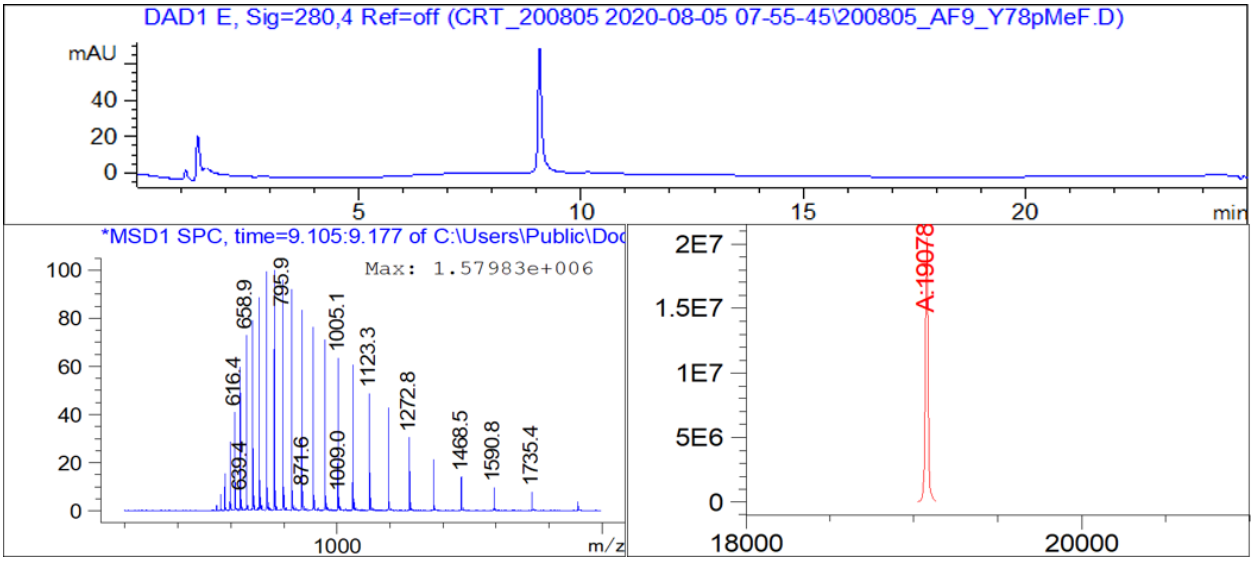
AF9 F28*p*CNF



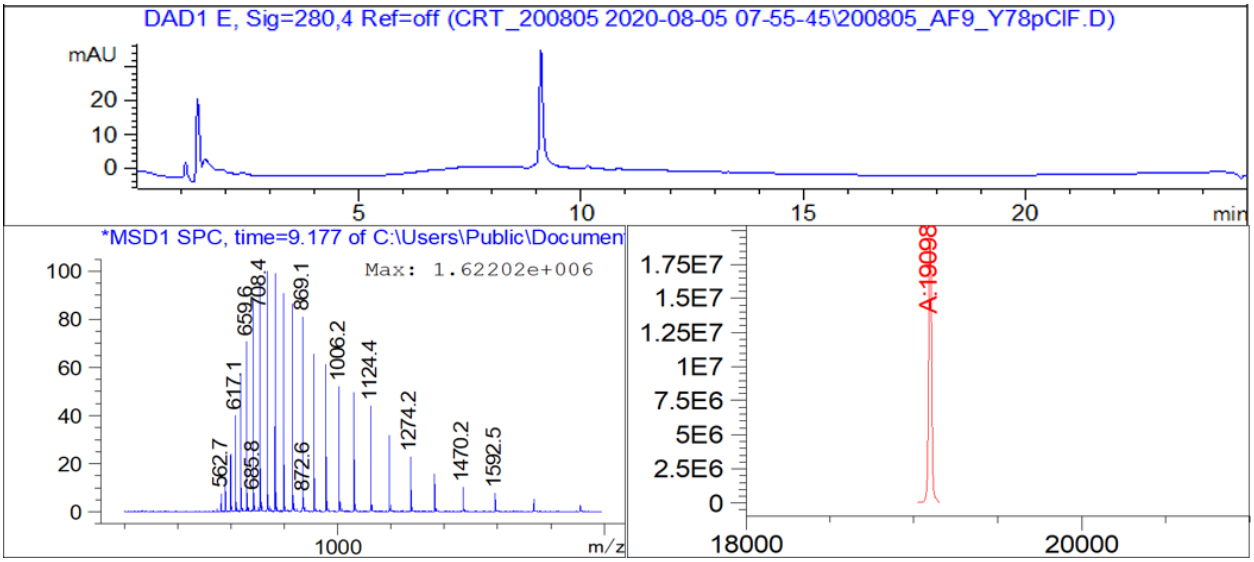
AF9 F28*p*NO$_2$F



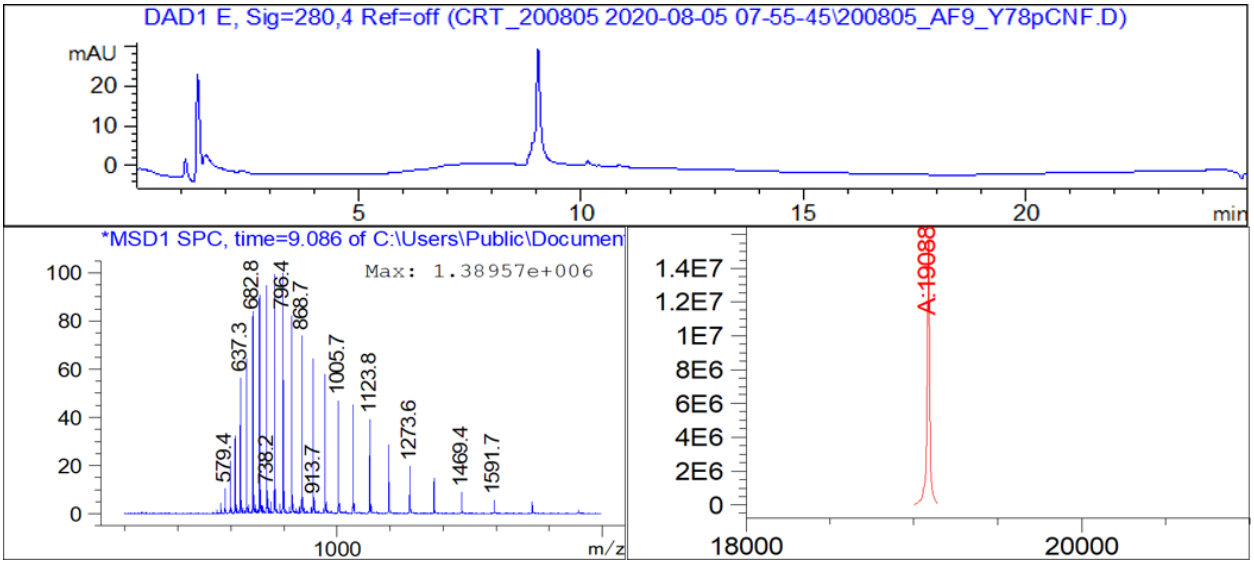
AF9 Y78*p*CH$_3$F



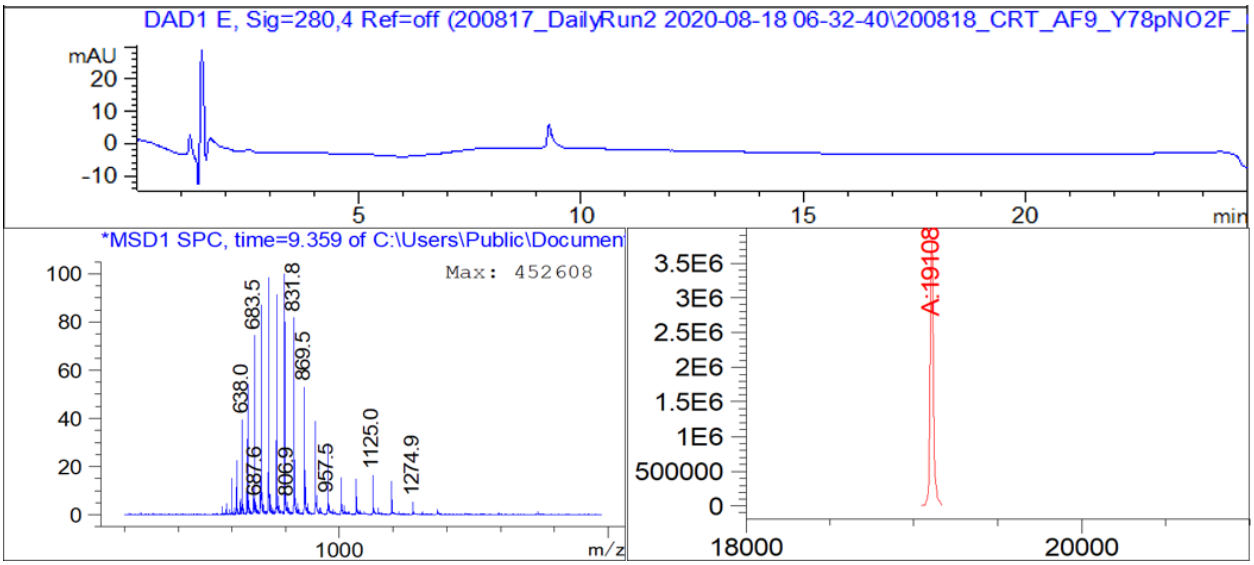
AF9 Y78*p*ClF



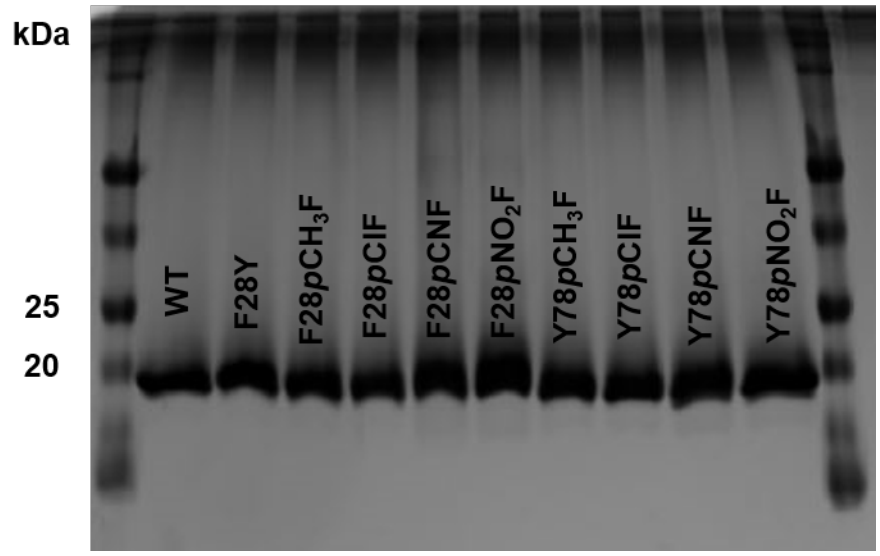
AF9 Y78*p*CNF



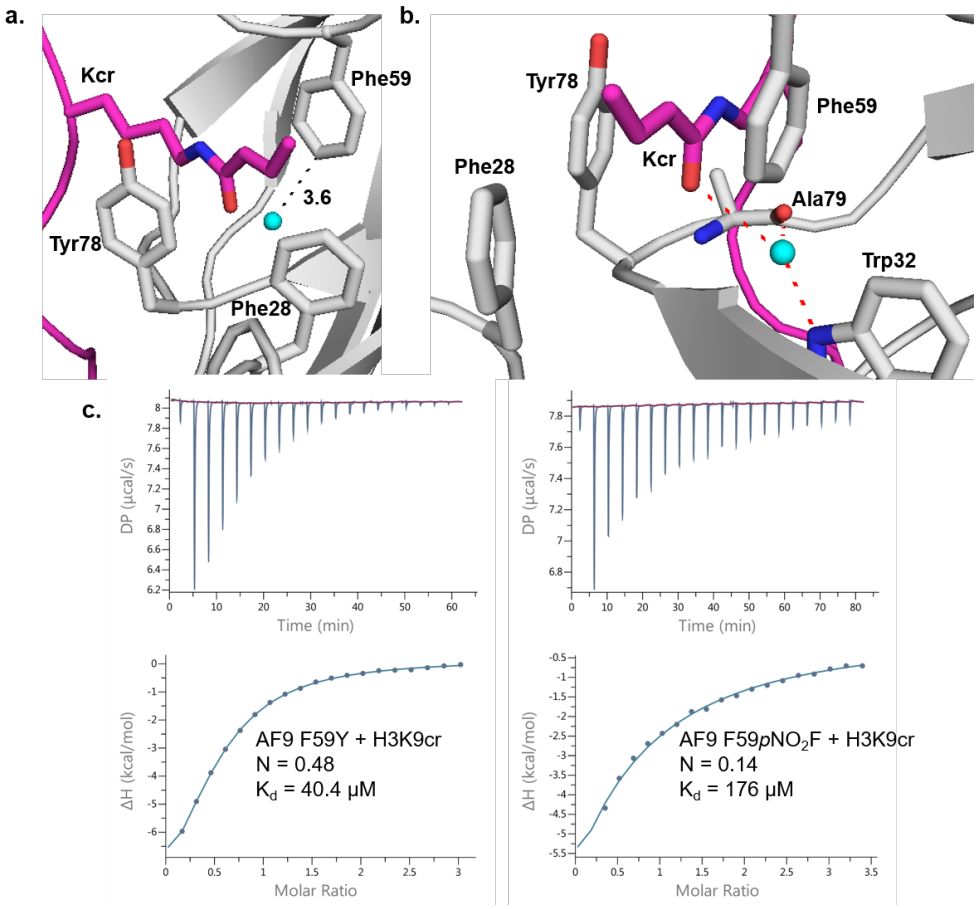
AF9 Y78*p*NO$_2$F



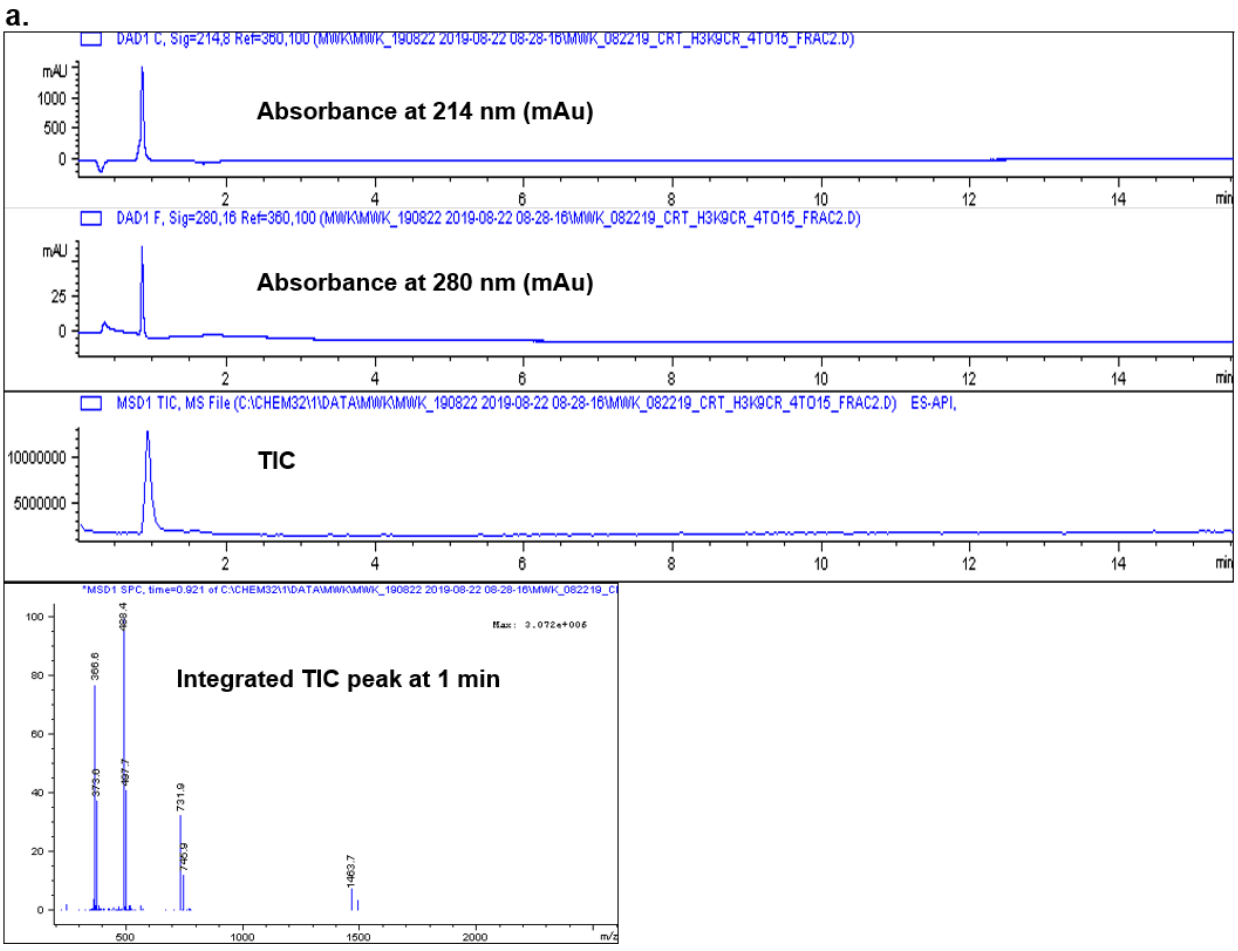
Supplementary Figure S3. SDS-PAGE gel of AF9 proteins. Precision Plus Dual Xtra protein ladder was used.



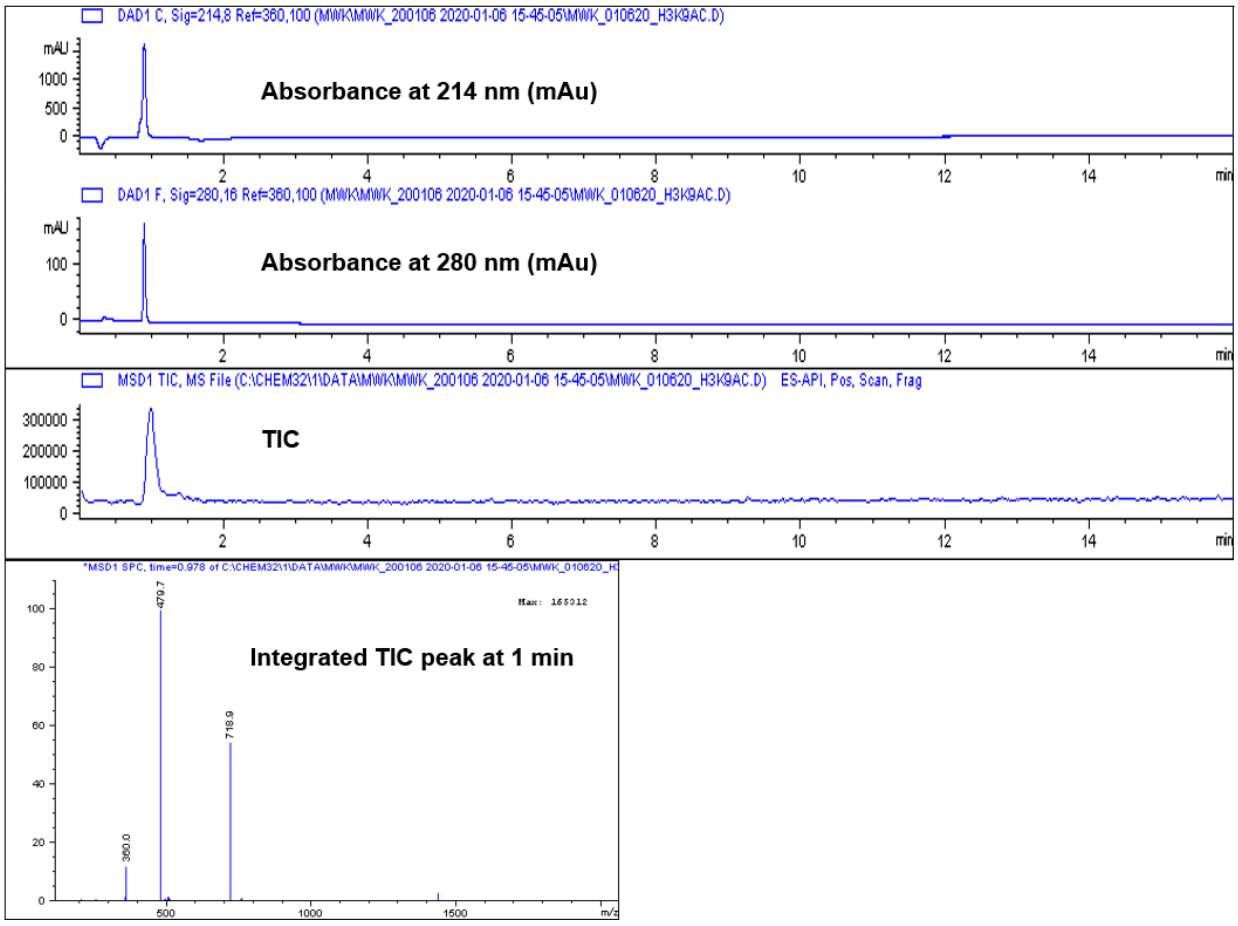
Supplementary Figure S4. (a) AF9 Phe59 mutants (gray) did not express well, which we hypothesize to be due to steric clash between a *para* substituent and a highly conserved water molecule (cyan) located 3.6 Å away from the residue. (b) This conserved water molecule is within hydrogen bonding distance (red lines) to the Trp32 indole sidechain, the backbone of Ala79 in AF9, and to the carbonyl oxygen in the acyl group of the peptide ligand (PDB ID: 5HJB)[1]. Thus, displacement of this water molecule via incorporation of a *para*-substituted Phe at residue 59 would likely result in a perturbed binding pocket and reduced affinity for acyllysines due to a loss of key hydrogen bonds. (c) Despite extremely low expression yields, binding to H3K9cr was assayed via ITC, resulting in poorly fit curves and low binding stoichiometry (N), indicating reduced quantities of active protein and perturbation of the binding pocket.



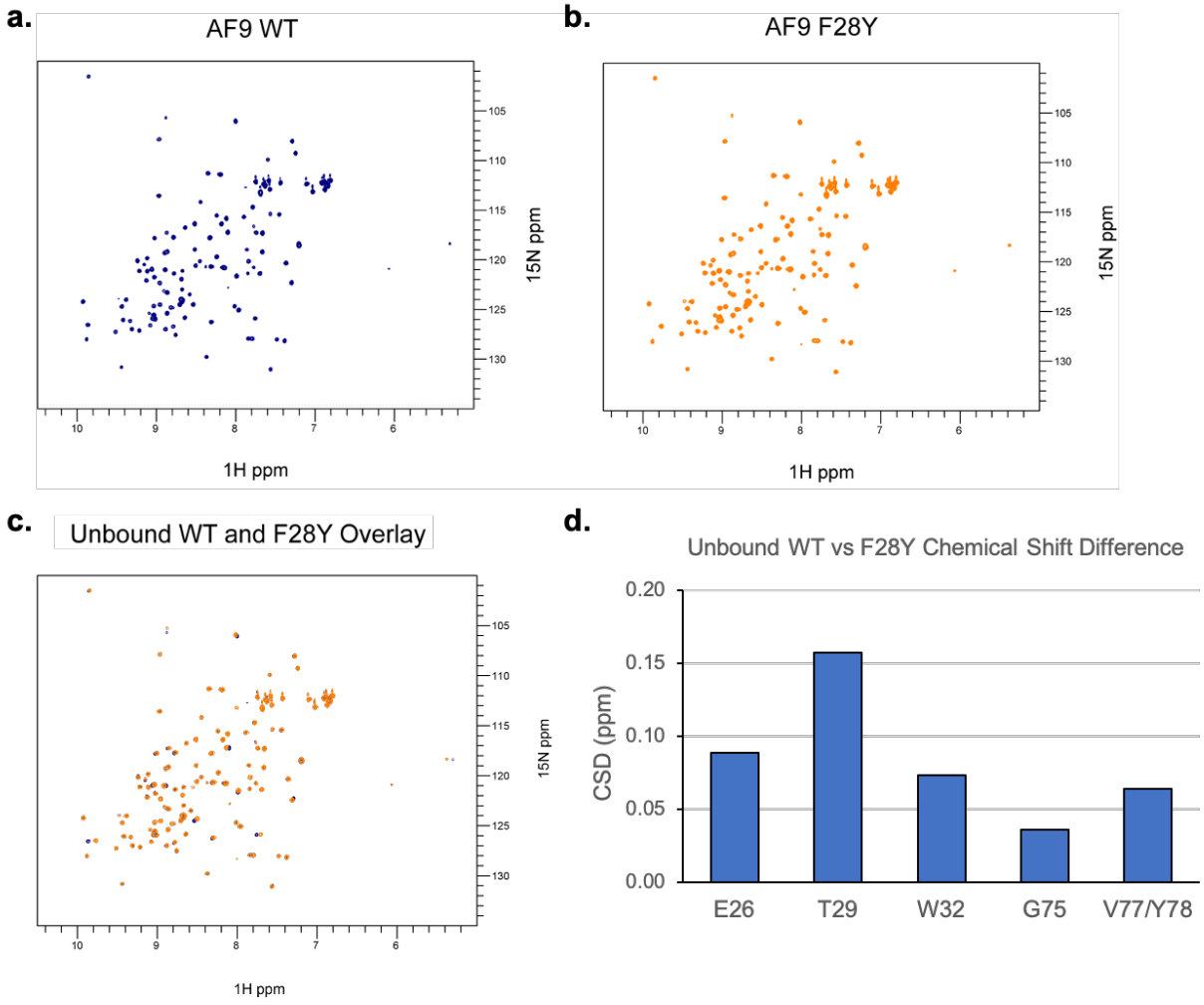
Supplementary Figure S5. ESI-LCMS data for H3K9cr (a) and H3K9ac (b) peptides. **H3K9cr** Expected m/z: 1463.7 (M+H), 732.3 (M+2H), 488.9 (M+3H), 366.9 (M+4H). Observed m/z: 1485.7 (M+H+Na), 1463.7 (M+H), 745.9 (M+2H+Na), 731.9 (M+2H), 497.7 (M+3H+Na), 488.4 (M+3H), 373.0 (M+4H+Na), 366.6 (M+4H). **H3K9ac** Expected m/z: 1437.6 (M+H), 719.3 (M+2H), 479.8 (M+3H), 360.0 (M+4H). Observed m/z: 1437.7 (M+H), 718.9 (M+2H), 479.7 (M+3H), 360.0 (M+4H).



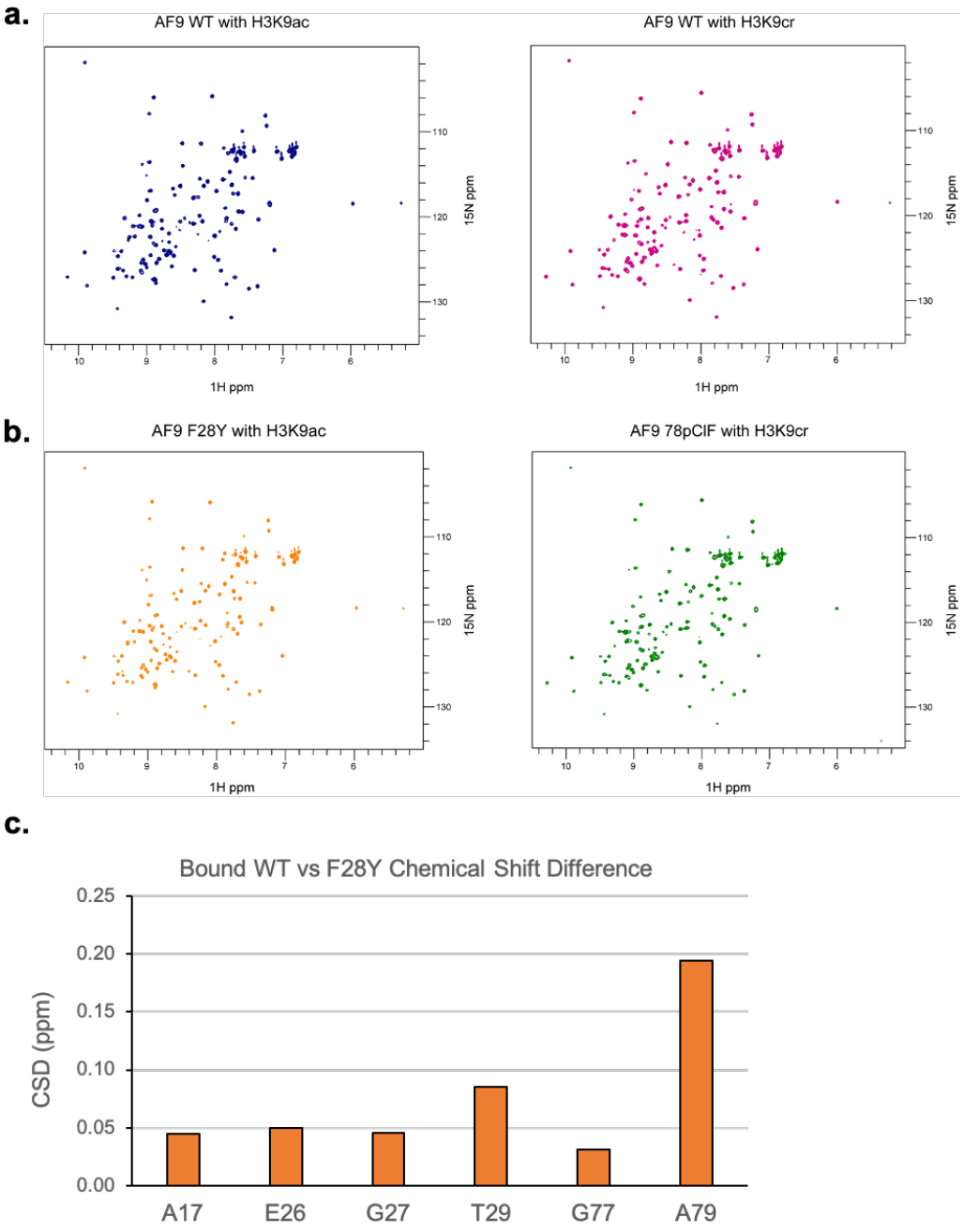
b.


DAD1 C, Sig=214,8 Ref=360,100 (MWK\MWK_200106 2020-01-06 15-45-05\MWK_010620_H3K9AC.D)
mAU
Absorbance at 214 nm (mAu)
DAD1 F, Sig=280,16 Ref=360,100 (MWK\MWK_200106 2020-01-06 15-45-05\MWK_010620_H3K9AC.D)
Absorbance at 280 nm (mAu)
MSD1 TIC, MS File (C:\CHEM32\1\DATA\MWK\MWK_200106 2020-01-06 15-45-05\MWK_010620_H3K9AC.D) ES-API, Pos, Scan, Frag
TIC
Integrated TIC peak at 1 min
479.7
718.9
360.0
min
m/z

Supplementary Figure S6. HSQC NMR spectra of unbound AF9 WT (a) and AF9 F28Y (b). Spectra overlay (c) very well, indicating mutation at the 28 position does not perturb protein folding in its unbound state. (d) Chemical shift differences (CSD) between unbound WT and F28Y spectra were calculated for residues with largest shifts.[8] Only T29 shows a significant perturbation (CSD > 0.1 ppm) which is likely due to a combination of through-bond and through-space effects. Residues were assigned based on examination of AF9 bound to K18ac (BMRB ID: 26059).

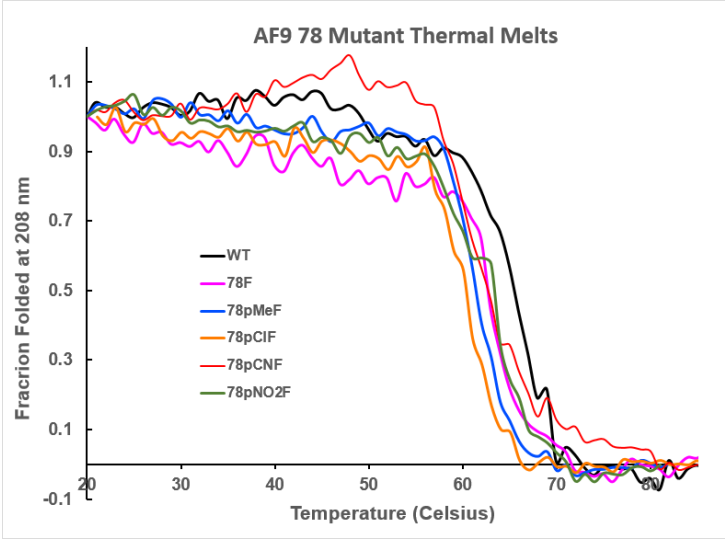


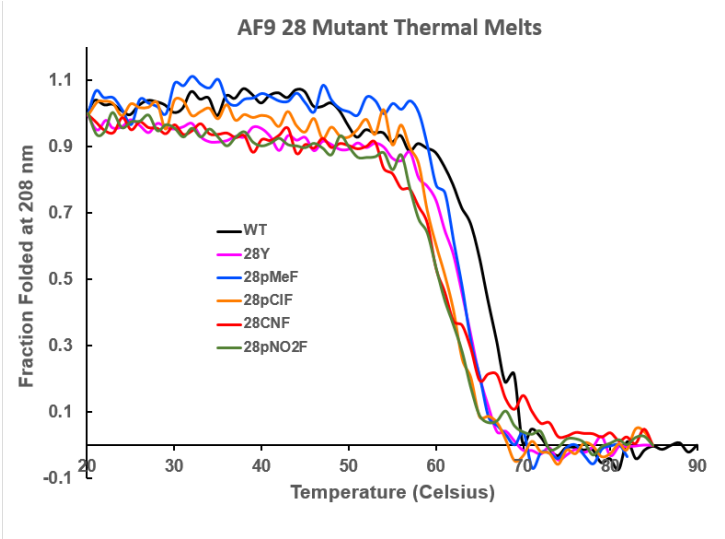
Supplementary Figure S7. (a) HSQC NMR spectra of AF9 WT bound to H3K9ac (blue) and H3K9ac (red-violet) (b) HSQC NMR spectra of AF9 F28Y bound to H3K9ac (orange). and AF9 Y78*p*ClF bound to H3K9cr (green). (c) Chemical shift differences (CSD) between bound WT and F28Y spectra were calculated for residues with largest shifts.[8] Only A79 shows a significant perturbation (CSD > 0.1 ppm), which is likely due to through-space effects.



Supplementary Figure S8. CD thermal melt data for wild type and mutant AF9 proteins.

Protein	**T_m (°C)**
WT	65
Y78F	63
Y78*p*CH_3F	61
Y78*p*ClF	60
Y78*p*CNF	63
Y78*p*NO_2F	63
F28Y	62
F28*p*CH_3F	63
F28*p*ClF	61
F28*p*CNF	61
F28*p*NO_2F	61

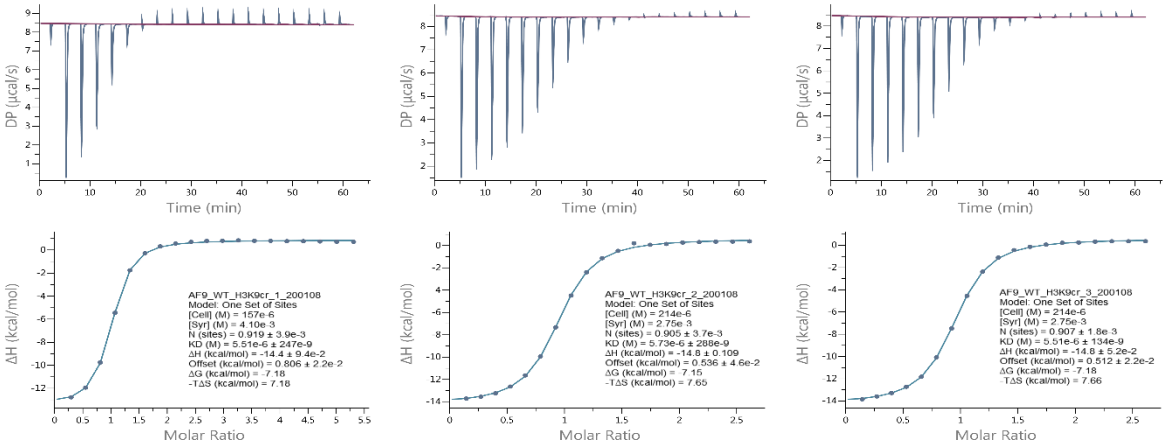




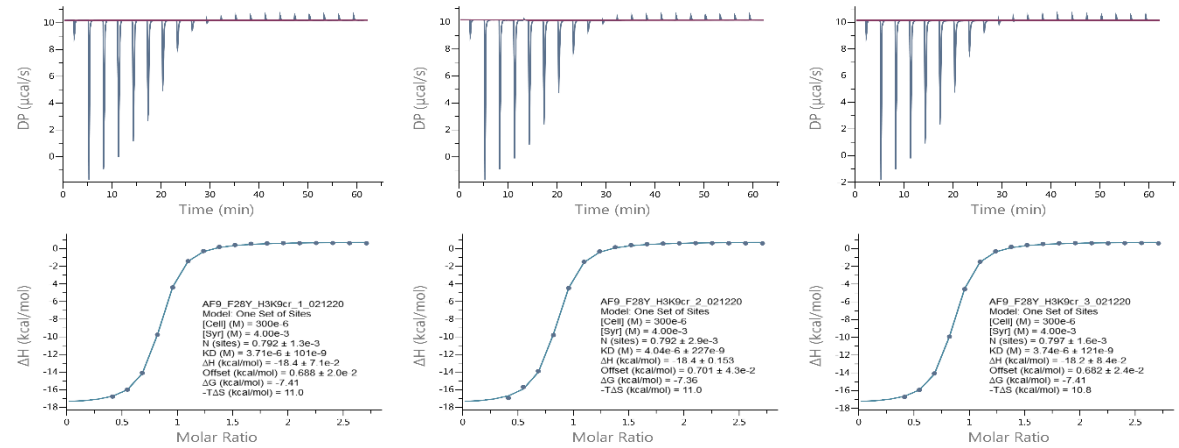
Supplementary Figure S9. (a) ITC curves for each run completed as part of this study. (b) Table displaying averaged ΔH and -TΔS data from each set of ITC binding experiments.

a.

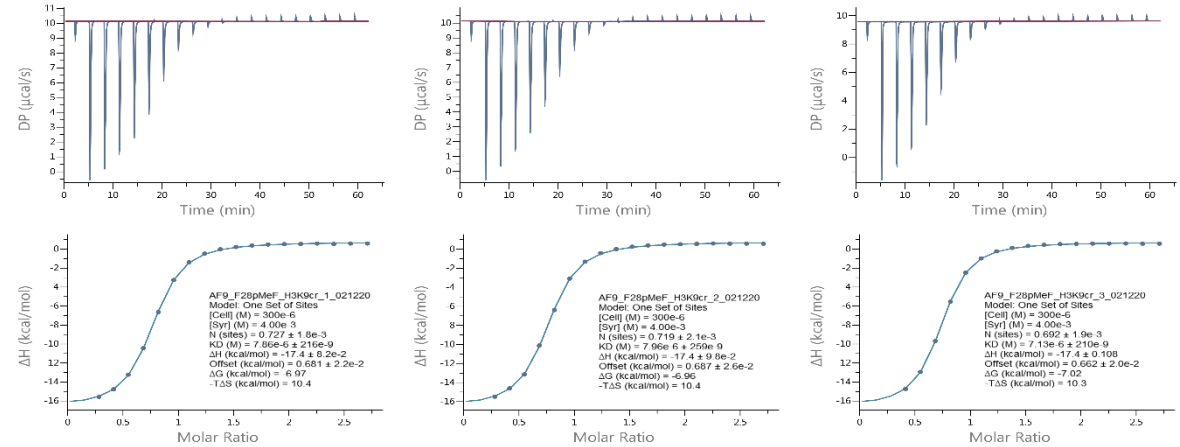
AF9 WT + H3K9cr



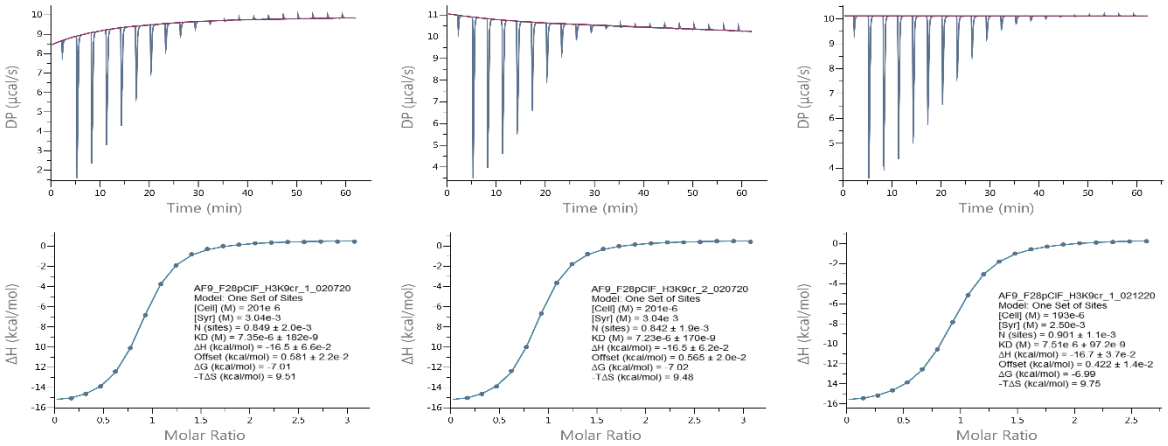
AF9 F28Y + H3K9cr



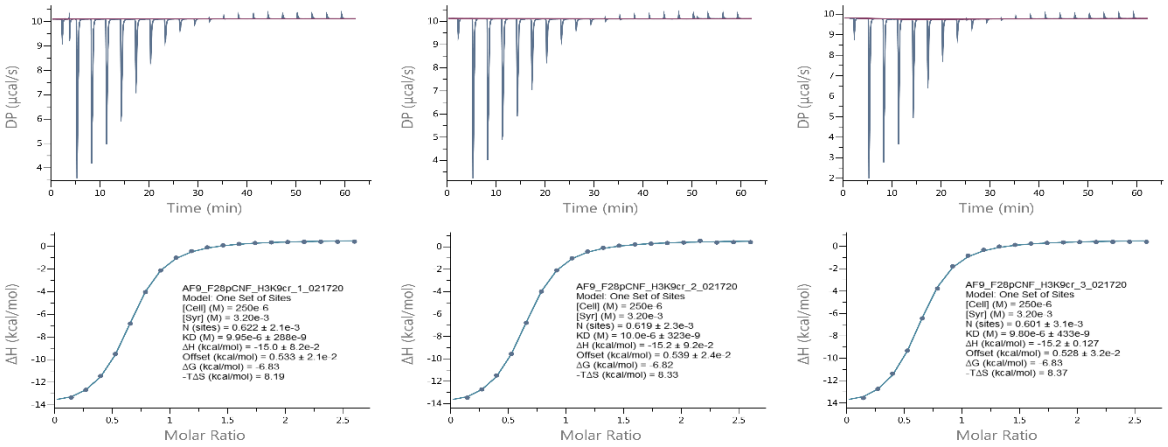
AF9 F28*p*CH_3F + H3K9cr



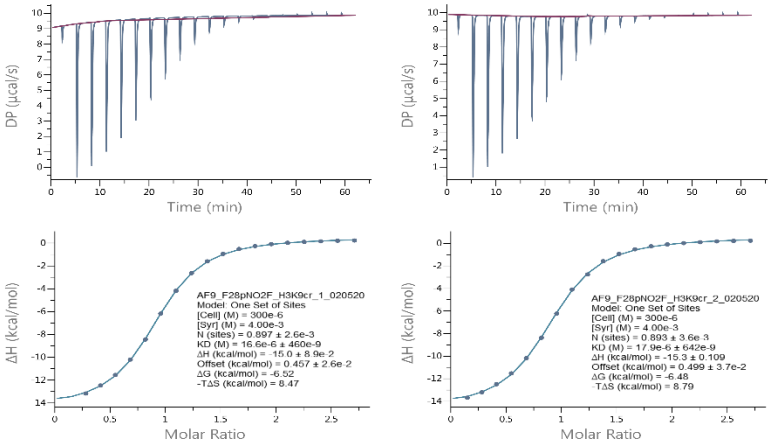
AF9 F28*p*ClF + H3K9cr



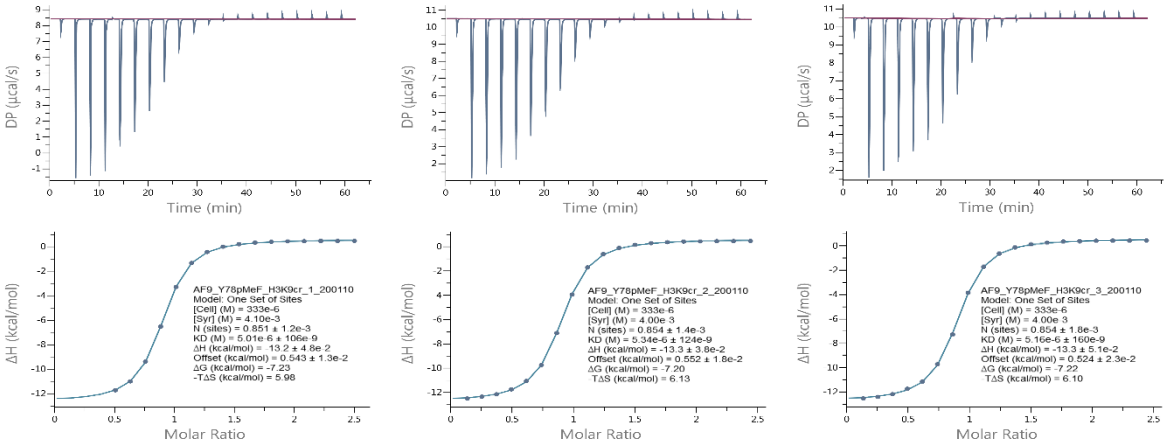
AF9 F28*p*CNF + H3K9cr



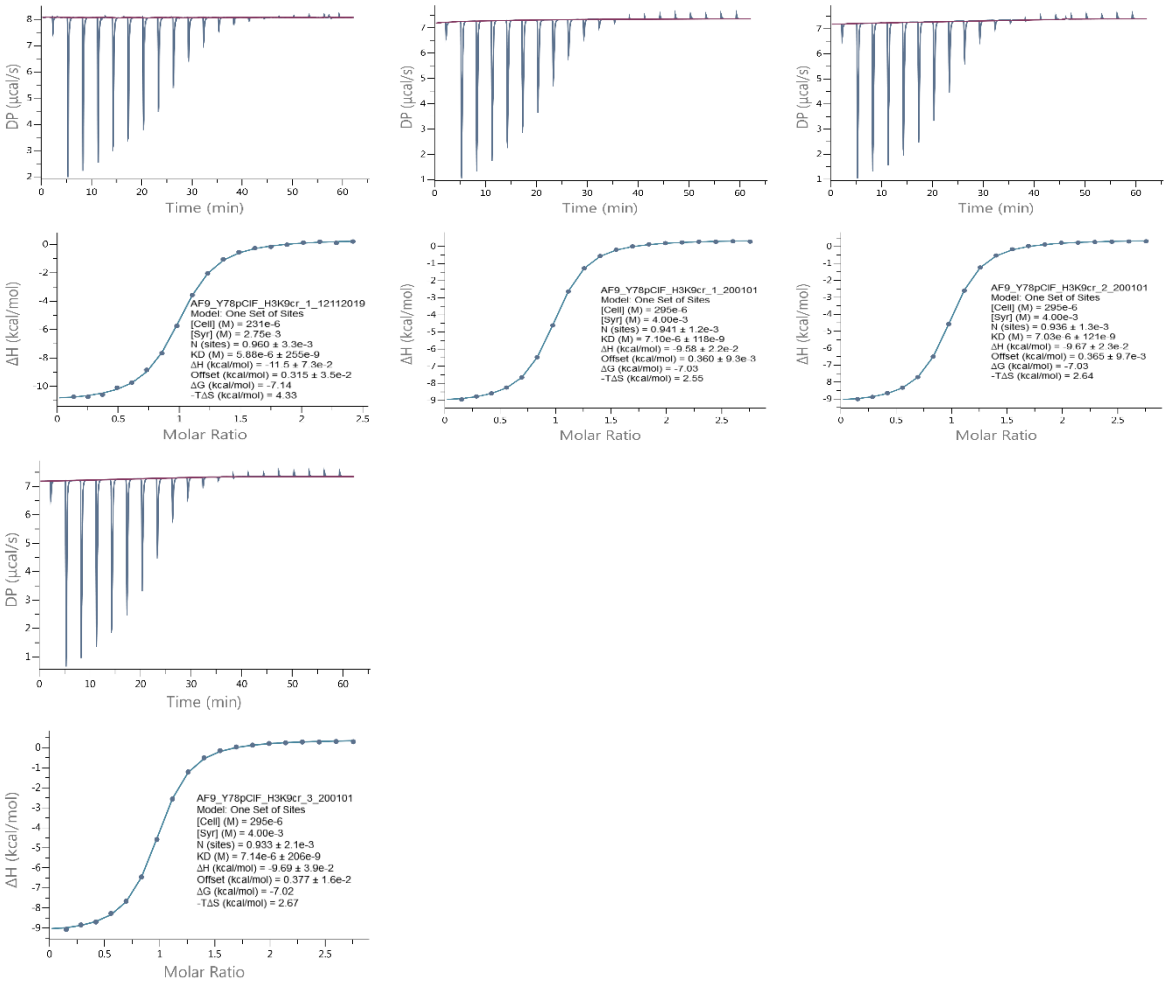
AF9 F28*p*NO_2F + H3K9cr



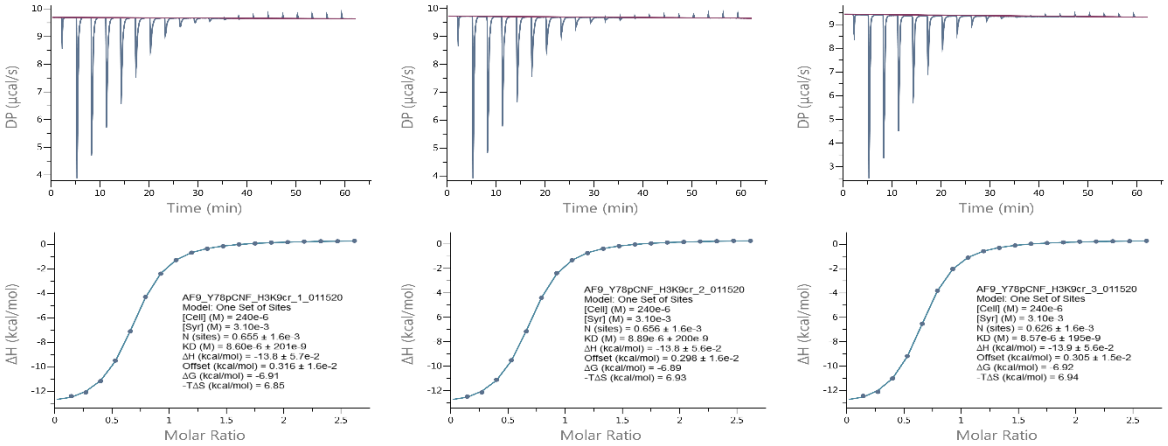
AF9 Y78*p*CH_3F + H3K9cr



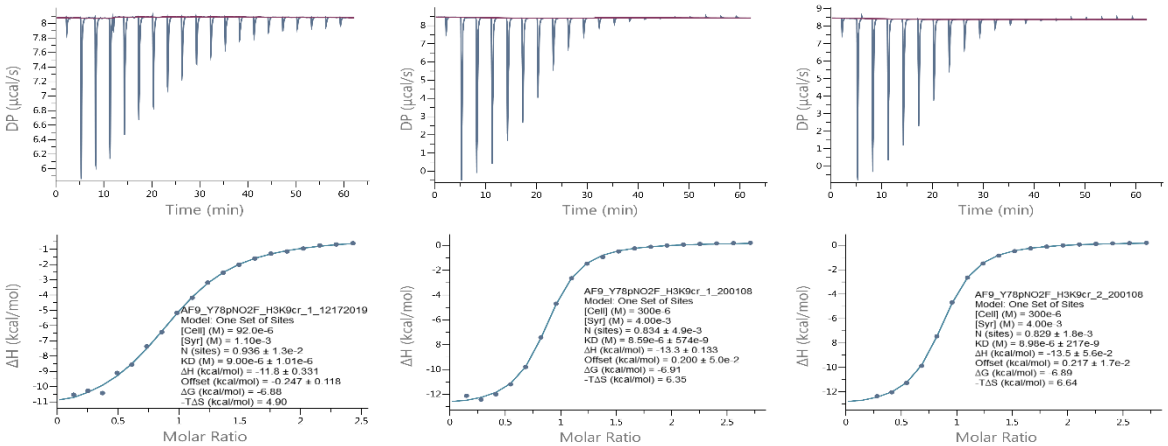
AF9 Y78*p*ClF + H3K9cr



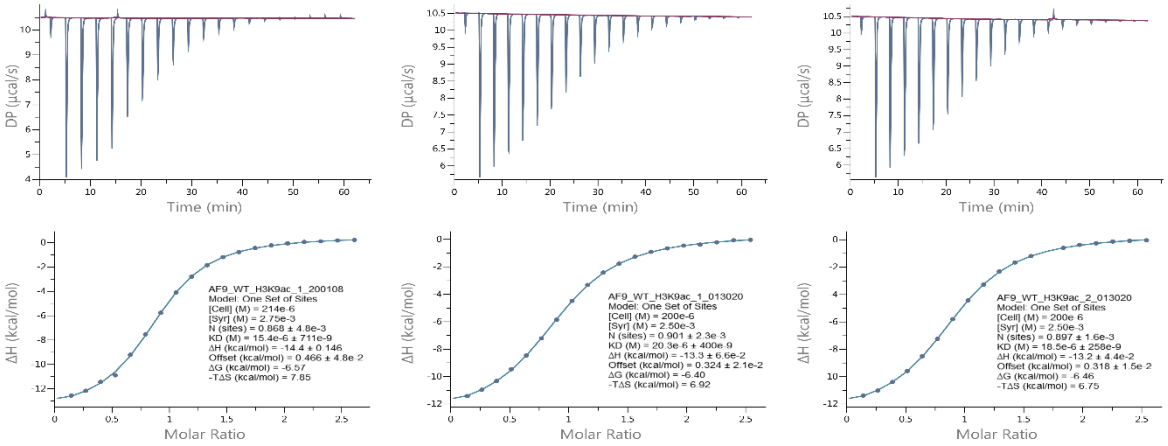
AF9 Y78*p*CNF + H3K9cr



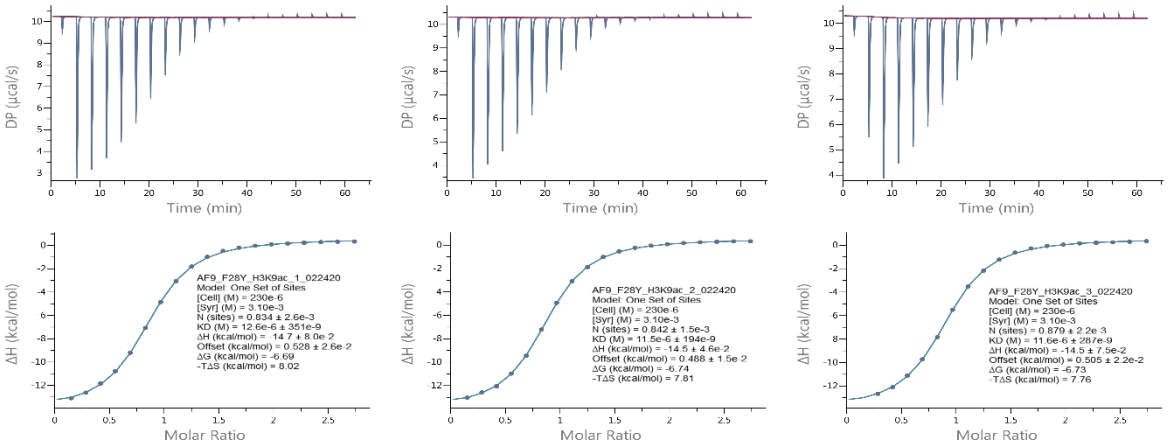
AF9 Y78*p*NO_2F + H3K9cr



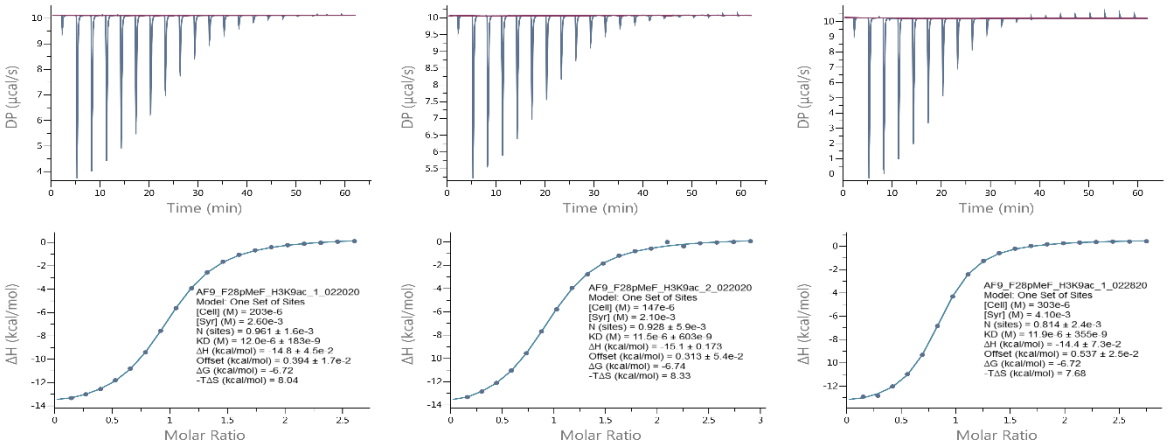
AF9 WT + H3K9ac



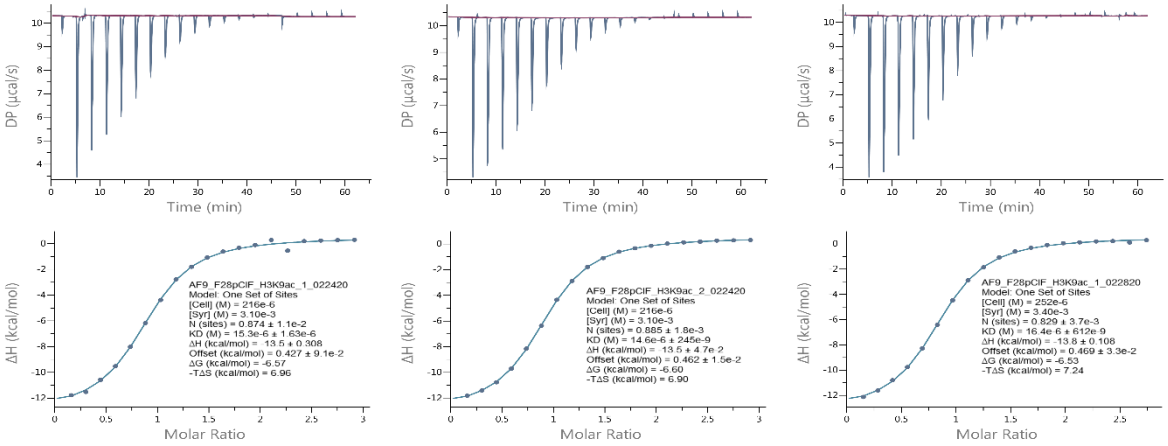
AF9 F28Y + H3K9ac



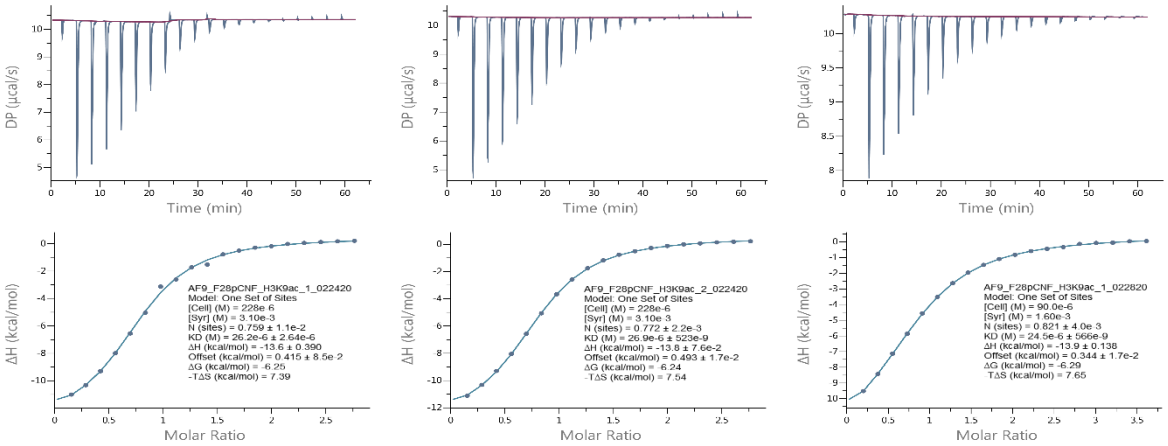
AF9 F28*p*CH$_3$F + H3K9ac



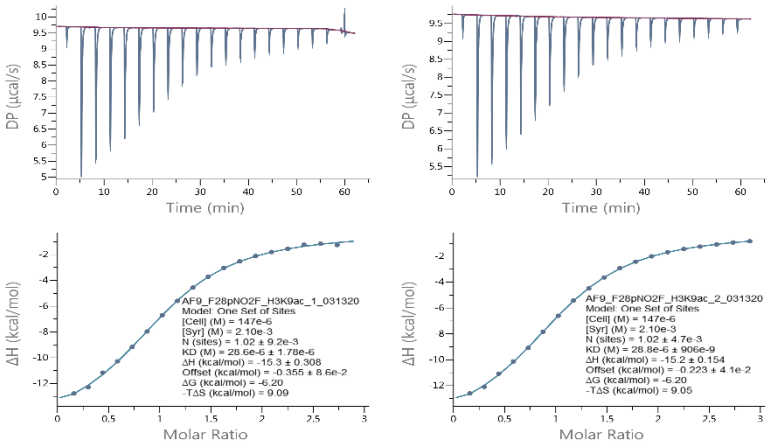
AF9 F28*p*ClF + H3K9ac



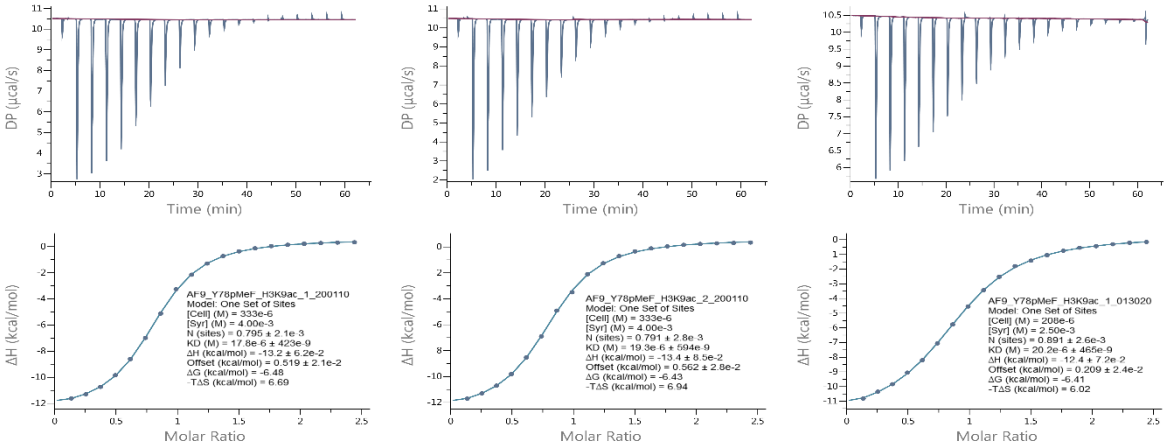
AF9 F28*p*CNF + H3K9ac



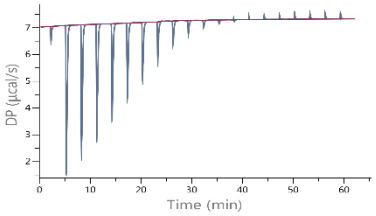
AF9 F28*p*NO$_2$F + H3K9ac

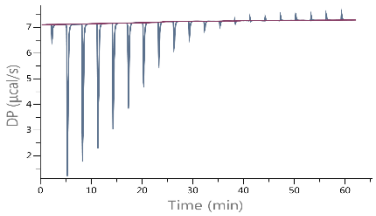


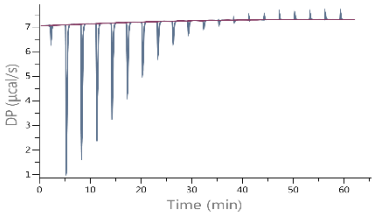
AF9 Y78*p*CH$_3$F + H3K9ac

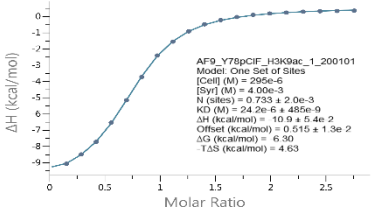


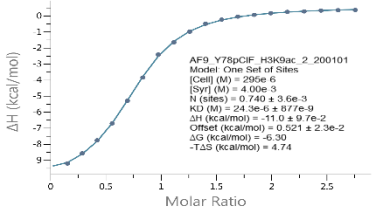
AF9 Y78*p*ClF + H3K9ac

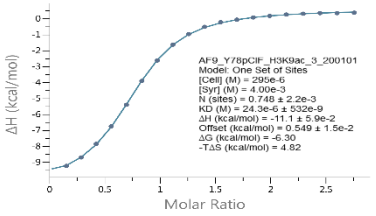


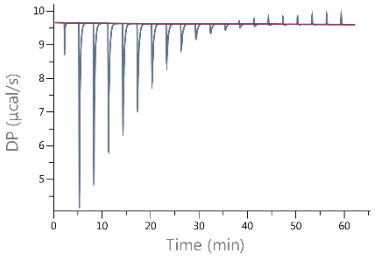


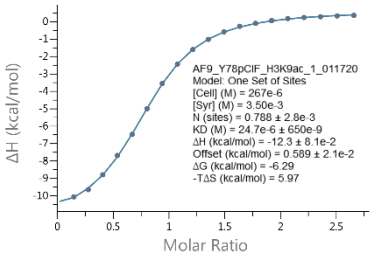




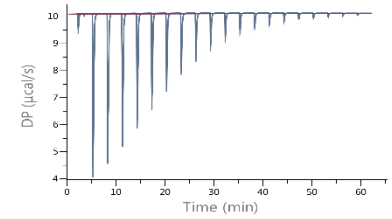


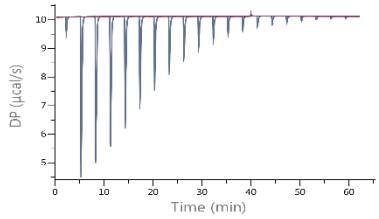


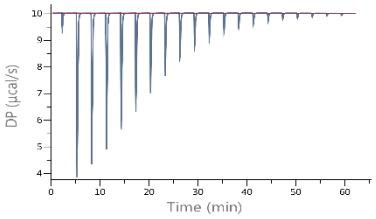


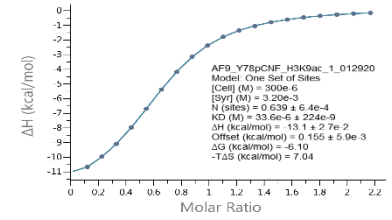


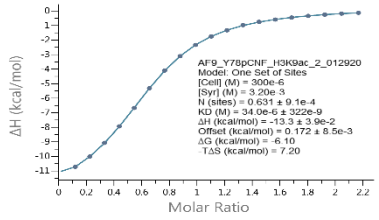
AF9 Y78*p*CNF + H3K9ac

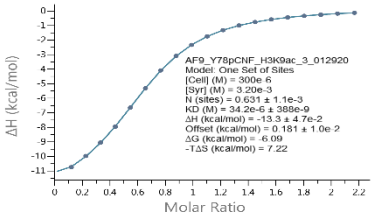




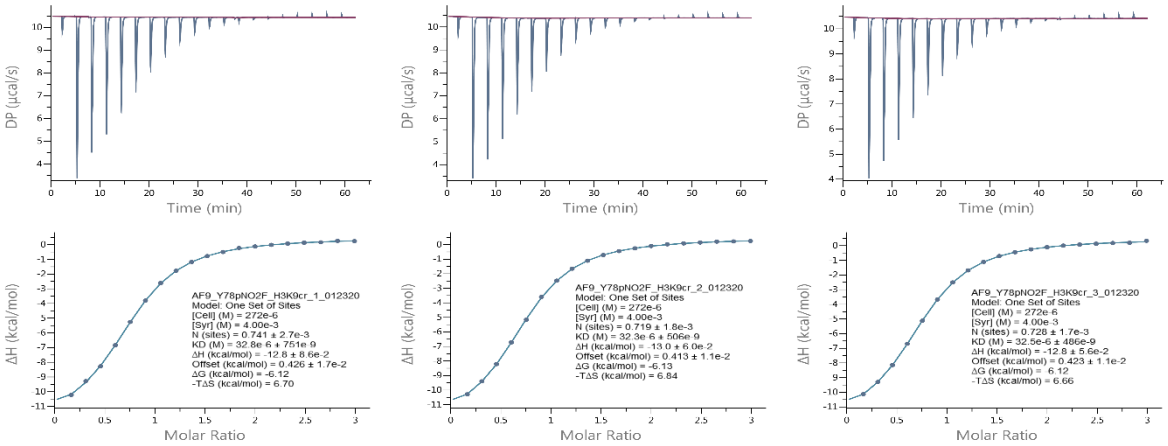








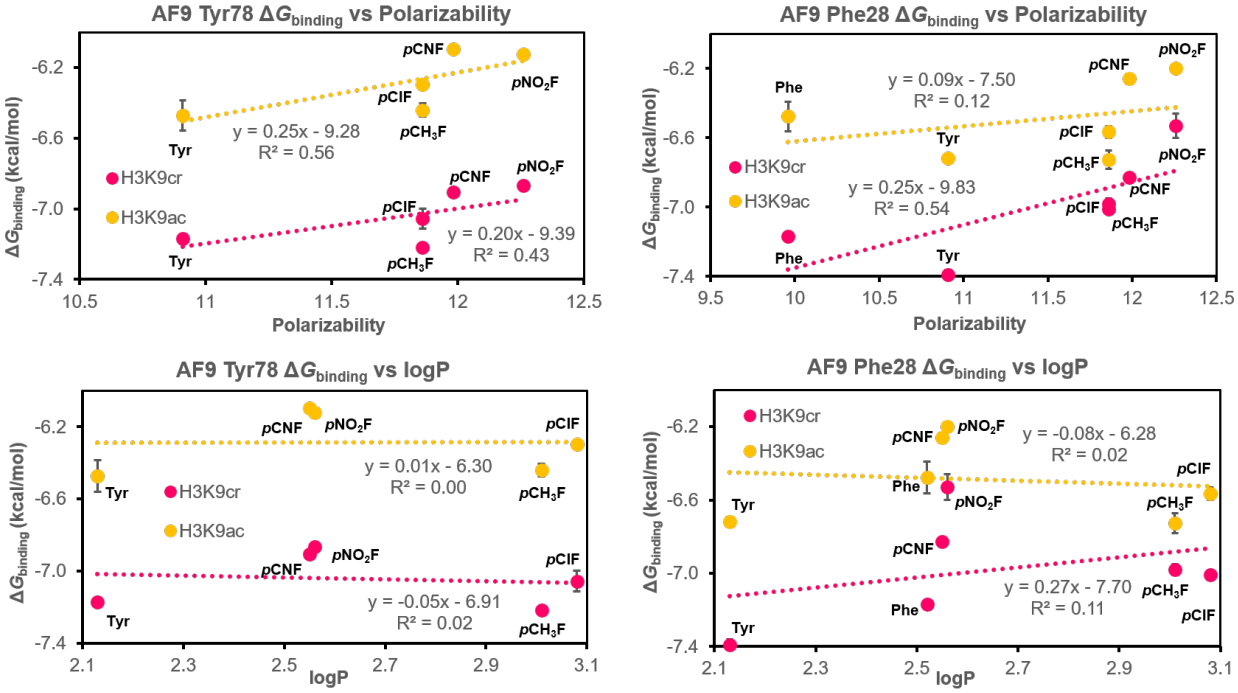
AF9 Y78*p*NO$_2$F + H3K9ac



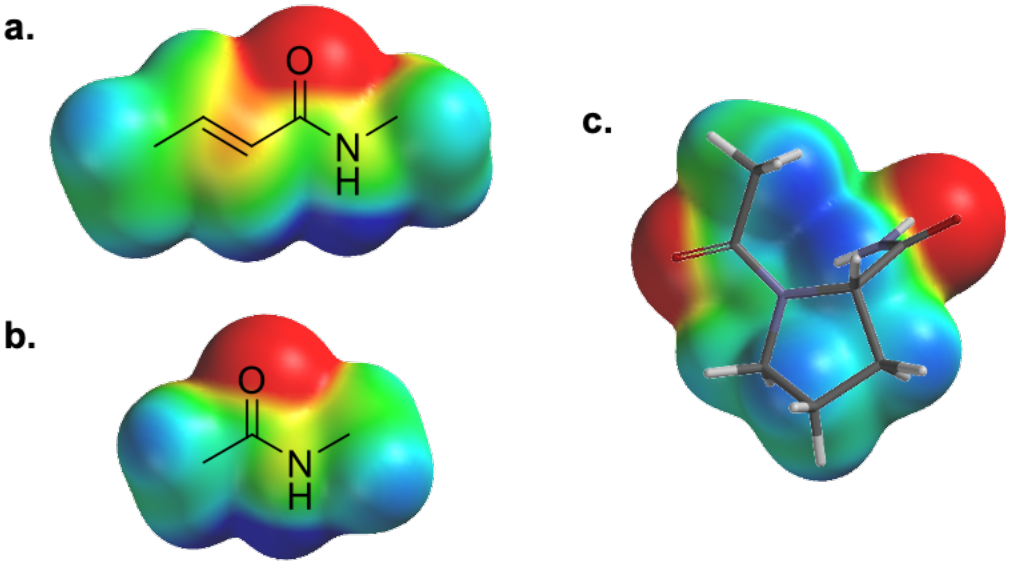
b.

H3K9cr			H3K9ac	
ΔH (kcal/mol)	**-TΔS (kcal/mol)**	**Protein**	**ΔH (kcal/mol)**	**-TΔS (kcal/mol)**
-15 ± 1	8 ± 1	**WT**	-14 ± 1	7 ± 1
-18 ± 1	11 ± 1	**F28Y**	-15 ± 1	8 ± 1
-17 ± 1	10 ± 1	**F28*p*CH$_3$F**	-15 ± 1	8 ± 1
-17 ± 1	10 ± 1	**F28*p*ClF**	-14 ± 1	7 ± 1
-15 ± 1	8 ± 1	**F28*p*CNF**	-14 ± 1	8 ± 1
-15 ± 1	9 ± 1	**F28*p*NO$_2$F**	-15 ± 1	9 ± 1
-13 ± 1	6 ± 1	**Y78*p*CH$_3$F**	-13 ± 1	7 ± 1
-10 ± 1	3 ± 1	**Y78*p*ClF**	-11 ± 1	5 ± 1
-14 ± 1	7 ± 1	**Y78*p*CNF**	-13 ± 1	7 ± 1
-13 ± 1	6 ± 1	**Y78*p*NO$_2$F**	-13 ± 1	7 ± 1

Supplementary Figure S10. LFER plots showing little to no correlation of $\Delta G_{binding}$ and calculated polarizability or hydrophobicity values for a range of AF9 position 78 and 28 mutants binding H3K9cr and H3K9ac. Experimental polarizability values reported by NIST (Computational Chemistry Comparison and Benchmark Database) for *para*-substituted toluenes were used to model *para*-substituted amino acids.



Supplementary Figure S11. ESP maps for the crotonyl γ-methyl (a), the acetyl α-methyl (b), and the axial CH-groups in proline (c) indicate that the methyls in the crotonyl and acetyl are very similarly polarized to the prolyl CHs. ESP maps were all calculated in Spartan using DFT wB97X-D 6-31G* level of theory with energetic range of -100 to +100 kcal/mol, where red indicates negative electrostatic potential, blue indicates positive electrostatic potential, and green is neutral.



Supplementary Information References

1. Li, Y.; Sabari, B. R.; Panchenko, T.; Wen, H.; Zhao, D.; Guan, H.; Wan, L.; Huang, H.; Tang, Z.; Zhao, Y., Molecular coupling of histone crotonylation and active transcription by AF9 YEATS domain. *Molecular cell* **2016,** *62* (2), 181-193.
2. Li, Y.; Wen, H.; Xi, Y.; Tanaka, K.; Wang, H.; Peng, D.; Ren, Y.; Jin, Q.; Dent, S. Y.; Li, W., AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **2014,** *159* (3), 558-571.
3. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1-3), 215-241.
4. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B* **2009,** *113* (18), 6378-6396.
5. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16, revision C. 01. Wallingford: Gaussian. Inc: 2016.
6. Legault, C., CYLview, 1.0 b. *Université de Sherbrooke* **2009**.
7. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: structure, function, and bioinformatics* **2005,** *59* (4), 687-696.
8. Williamson, M. P., Using chemical shift perturbation to characterise ligand binding. *Progress in nuclear magnetic resonance spectroscopy* **2013,** *73*, 1-16.
9. Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R., CH··· π-interactions in proteins. *Journal of molecular biology* **2001,** *307* (1), 357-377.