Supporting Information

Molecular Recognition of Methylated Amino Acids and Peptides by Pillar[6]MaxQ

David King,[a] Chelsea R. Wilson,[b] Lukas Herron,[c,d] Chun-Lin Deng,[a] Shams Mehdi,[c,d] Pratyush Tiwary,[a,d,*] Fraser Hof,[b,*] Lyle Isaacs[a,*]

[a]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, United States; [b]Department of Chemistry, University of Victoria, Victoria, BC, V8W 3V6, Canada; [c]Biophysics Program, University of Maryland, College Park, MD 20742, United States; [d]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, United States

Table of Contents **Pages**

General experimental details. Amino acids and amino acid amides were purchased from commercial suppliers and were used without further purification. 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI) was purchased from Sigma-Aldrich (D9542, ≥98%). Trimethyllysine was purchased from Sigma-Aldrich (55528-53-5, ≥97%). Peptide sequences were purchased from GenScript with the exception of H3K4Me$_2$ which was obtained as described in literature (*Org. Biomol. Chem.* **2021**, *19*, 4691-4696). P5MQ and P6MQ were synthesized according to the literature procedures (*Angew. Chem. Int. Ed.* **2020**, *59*, 13313). ^{1}H NMR spectra were recorded on a commercial spectrometer operating at 600 MHz. ITC data was collected on a Malvern Microcal PEAQ-ITC instrument and analyzed using the software provided by the manufacturer. All fluorescence assays were performed in a Thermo Scientific™ Nunc MicroWell 96-Well Optical-Bottom, Black-walled Plate (265031). All plates were read on a BioTek Cytation-5 cell imaging and multi-mode reader plate reader.

Peptide sequences: H3K4 = (ARTKQTAY), H-K-OH = lysine, H3K4Me = (ART(KMe)QTAY), H-K(Me)-OH = monomethyl lysine, H3K4Me$_2$ = (ART(KMe$_2$)QTAY), H-K(Me$_2$)-OH = dimethyl lysine, H3K4Me$_3$ = (ART(KMe$_3$)QTAY), H-K(Me$_3$)-OH = trimethyl lysine, H-K-NH$_2$ = lysine amide, Ac-K-NH$_2$ = acetyl lysine amide, H-R-OH = arginine, H-R(Me)-OH = monomethyl arginine, H3R2Me$_2$a = (A(RMe$_2$a)TKQTAY), Asym-H-R(Me$_2$) = asymmetric dimethylarginine, H3R2Me$_2$s = (A(RMe$_2$s)TKQTAY), Sym-H-R(Me$_2$) = symmetric dimethylarginine, H-R-NH$_2$ = arginine amide, Ac-R-NH$_2$ = acetyl arginine amide.

Determination of the thermodynamic parameters of binding between P5MQand the guests by isothermal titration calorimetry

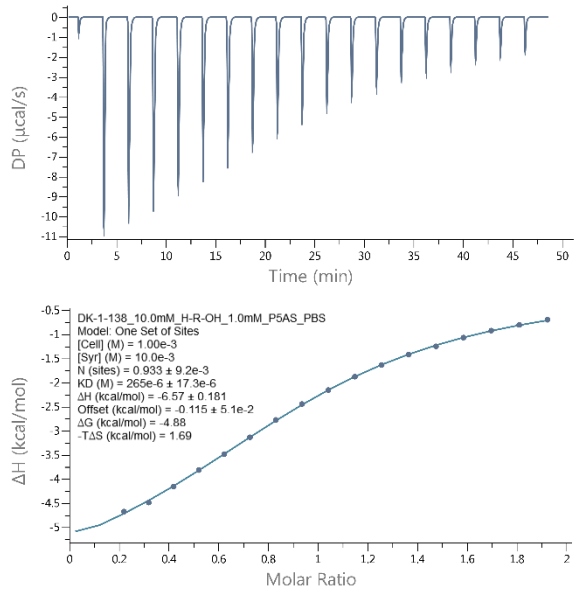

Figure S1. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (1.00 mM) in the cell was titrated with H-R-OH (10.0 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (3.77 ± 0.24) x 10^3 M^{-1}.

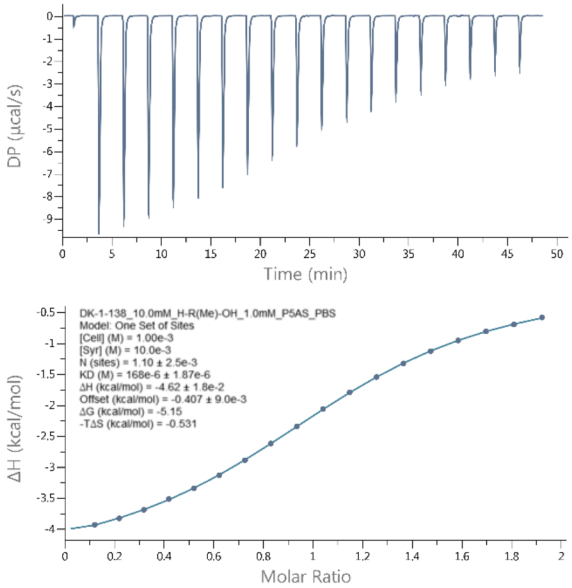

Figure S2. ITC curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with H-R(Me)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (5.95 ± 0.07) x 10^3 M^{-1}.

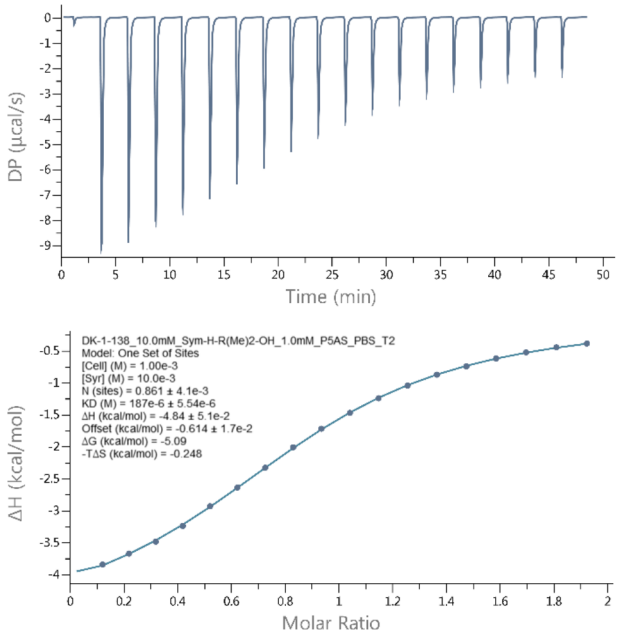

Figure S3. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with Sym-H-R(Me_2)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (5.35 ± 0.16) x 10^3 M^{-1}.

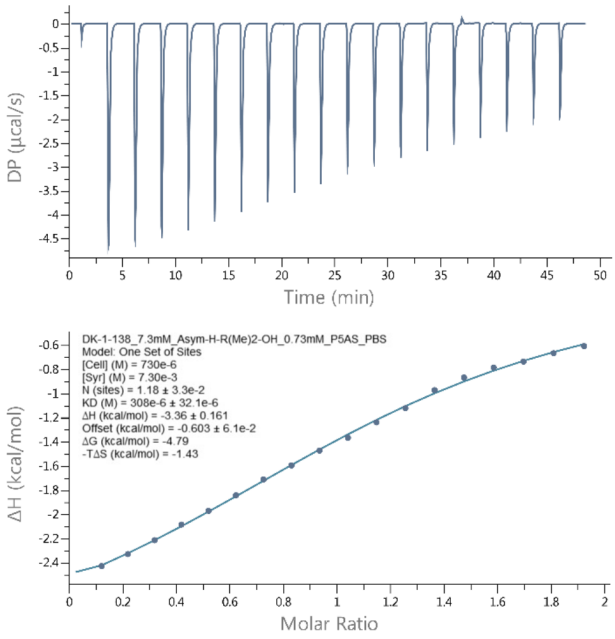

Figure S4. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.73 mM) in the cell was titrated with Asym-H-R(Me_2)-OH (7.30 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (3.25 ± 0.34) x 10^3 M^{-1}.

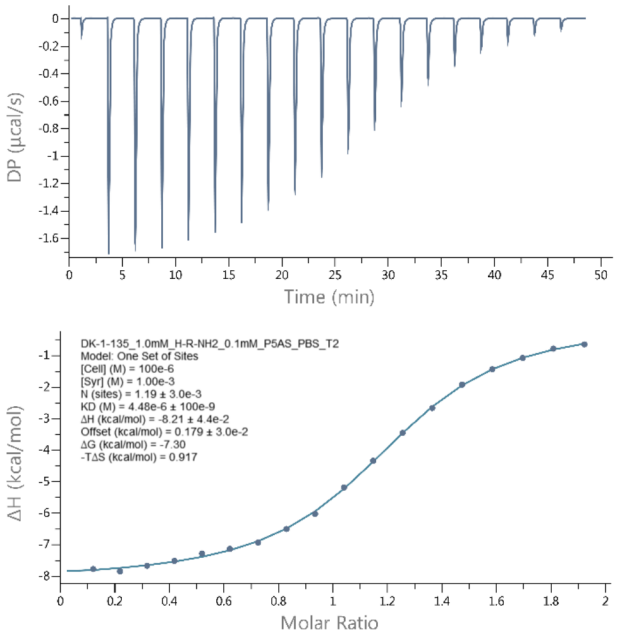

Figure S5. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with H-R-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (2.23 ± 0.05) x 10^5 M^{-1}.

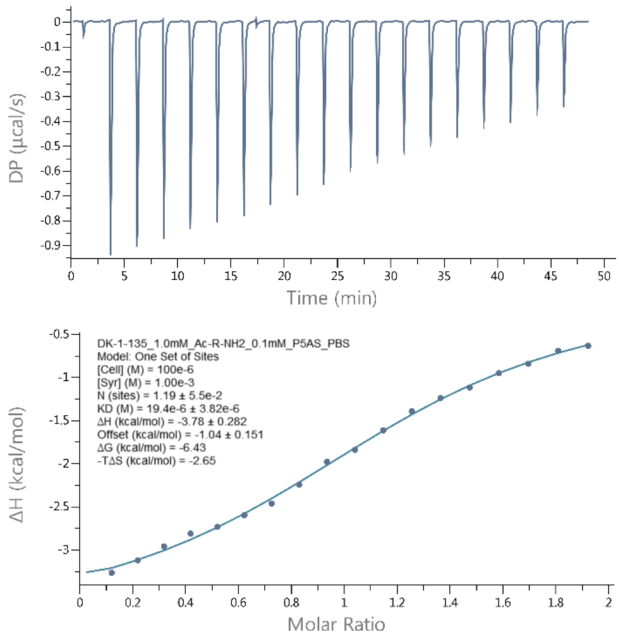

Figure S6. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with Ac-R-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (5.15 ± 1.01) x 10^4 M^{-1}.

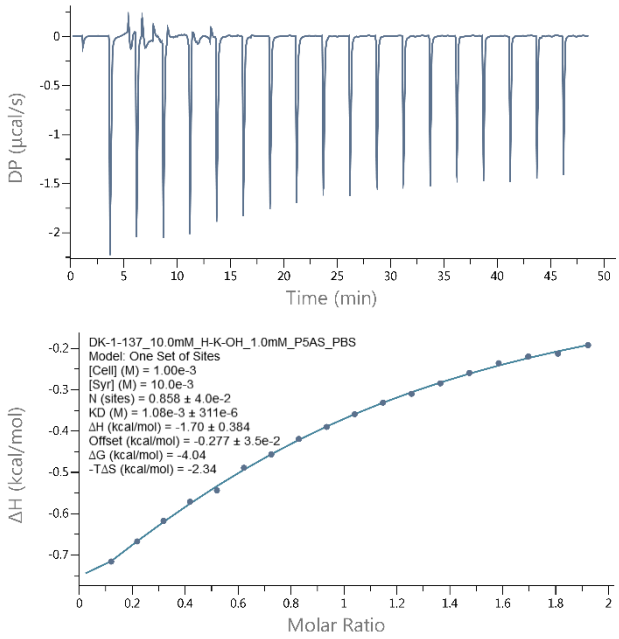

Figure S7. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (1.0 mM) in the cell was titrated with H-K-OH (10.0 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (9.26 ± 2.67) x 10^2 M^{-1}.

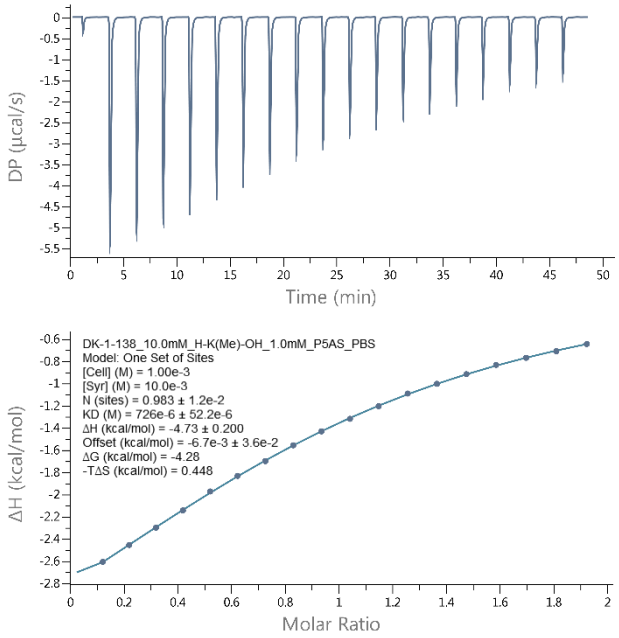

Figure S8. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (1.0 mM) in the cell was titrated with H-K(Me)-OH (10.0 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.38 ± 0.10) x 10^3 M^{-1}.

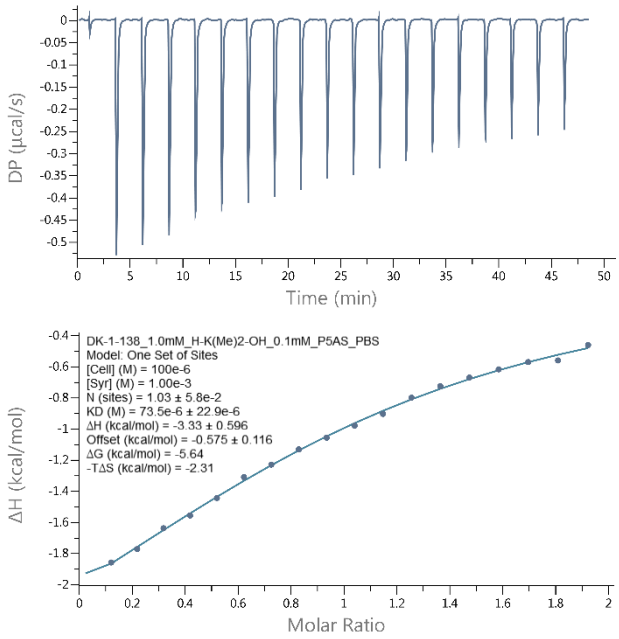

Figure S9. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with H-K(Me_2)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.36 ± 0.42) x 10^4 M^{-1}.

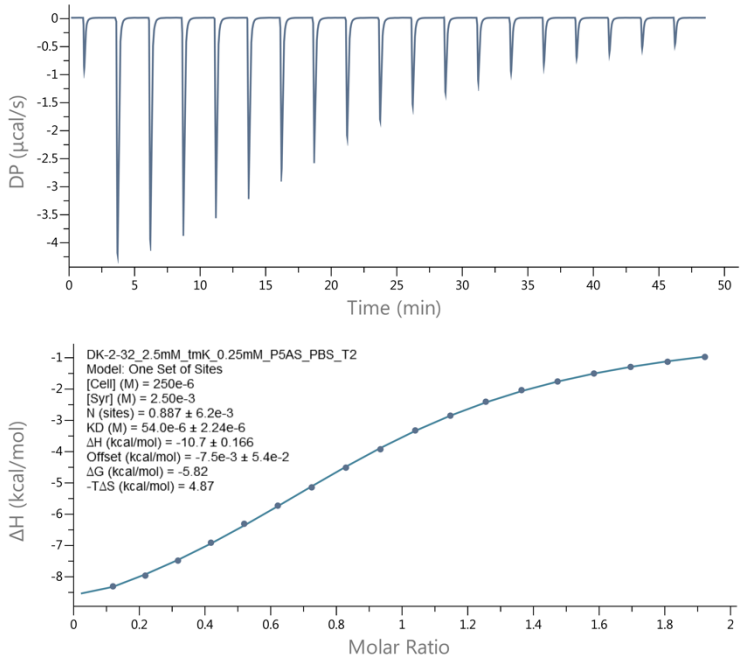

Figure S10. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.25 mM) in the cell was titrated with H-K(Me_3)-OH (2.50 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.81 ± 0.26) x 10^4 M^{-1}.

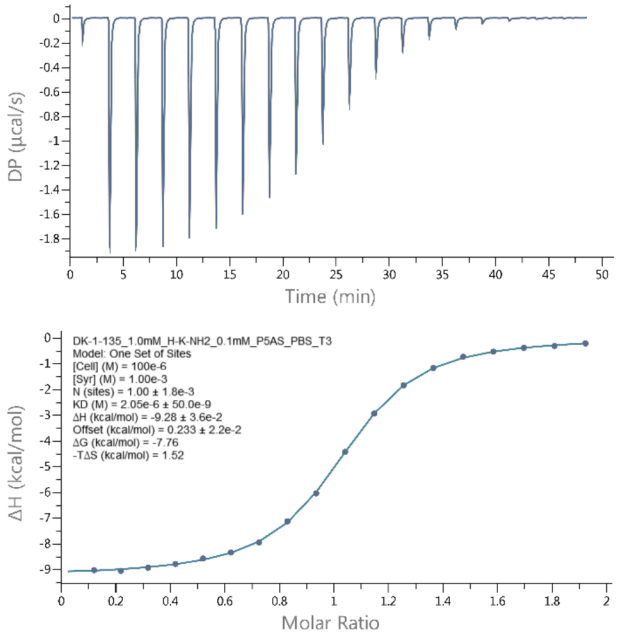

Figure S11. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with H-K-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (4.88 ± 0.12) x 10^5 M^{-1}.

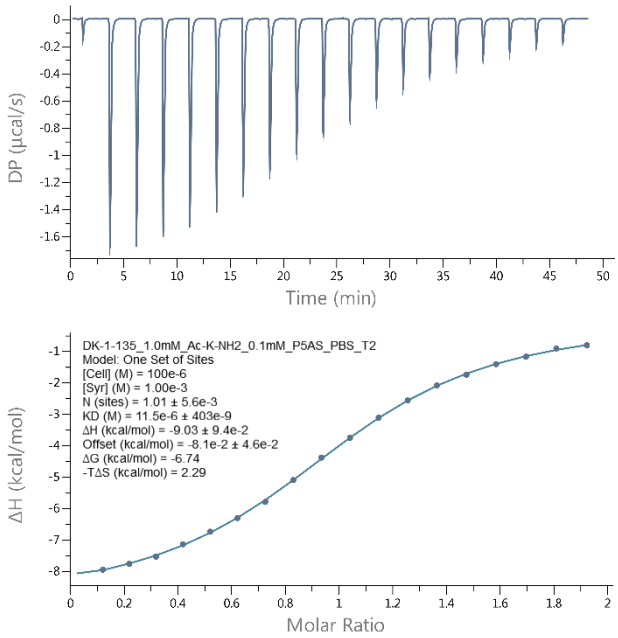

Figure S12. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P5MQ (0.1 mM) in the cell was titrated with Ac-K-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (8.70 ± 0.30) x 10^4 M^{-1}.

Determination of the thermodynamic parameters of binding between P6MQ and the guests by isothermal titration calorimetry

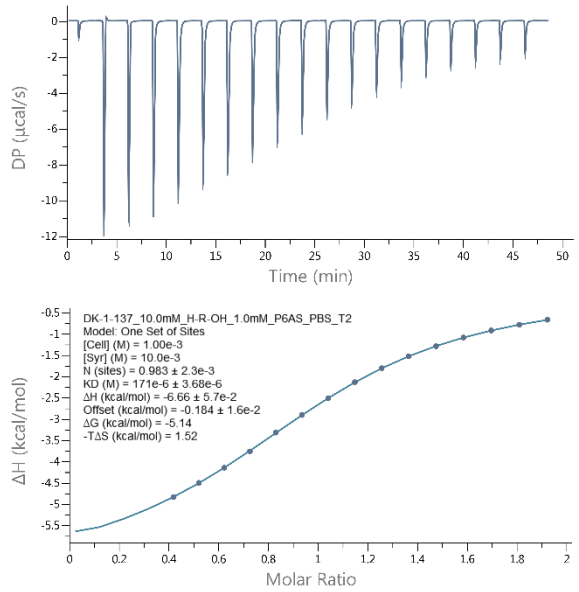

Figure S13. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (1.0 mM) in the cell was titrated with H-R-OH (10.0 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (5.85 ± 0.13) x 10^3 M^{-1}.

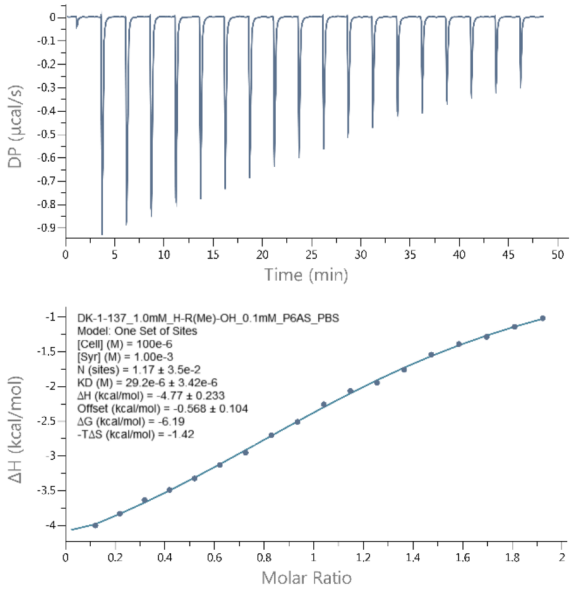

Figure S14. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with H-R(Me)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (3.42 ± 0.40) x 10^4 M^{-1}.

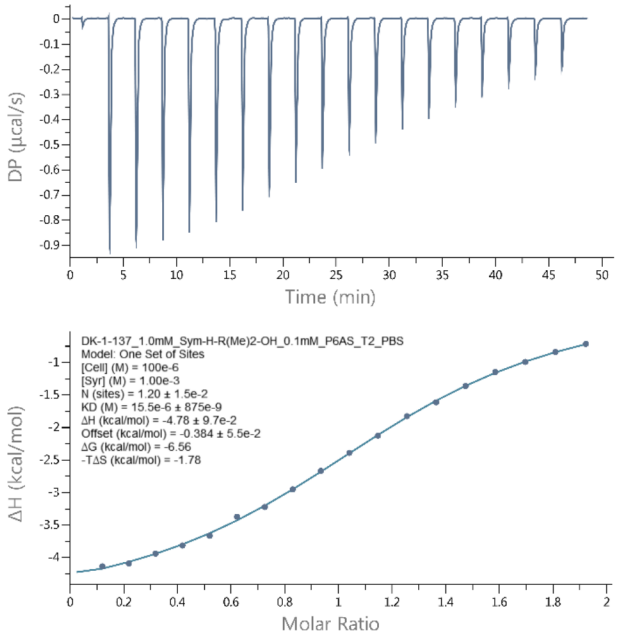

Figure S15. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with Sym-H-R(Me_2)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (6.45 ± 0.36) x 10^4 M^{-1}.

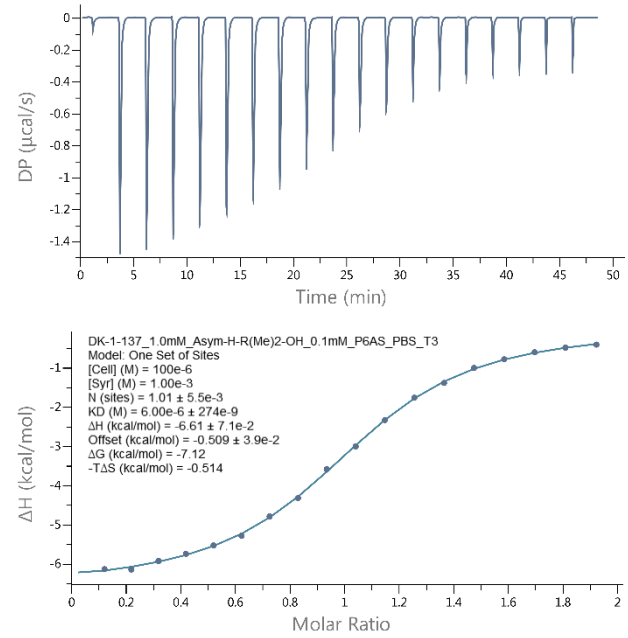

Figure S16. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with Asym-H-R(Me_2)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.67 ± 0.08) x 10^5 M^{-1}.

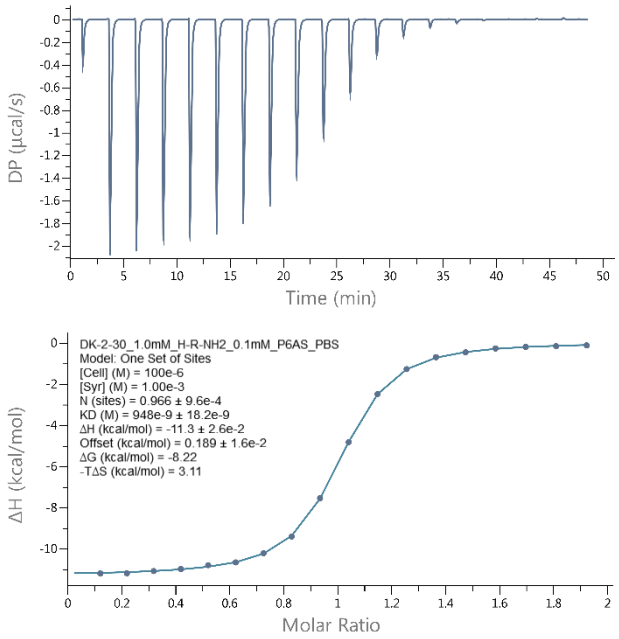

Figure S17. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with H-R-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.05 ± 0.02) x 10^6 M^{-1}.

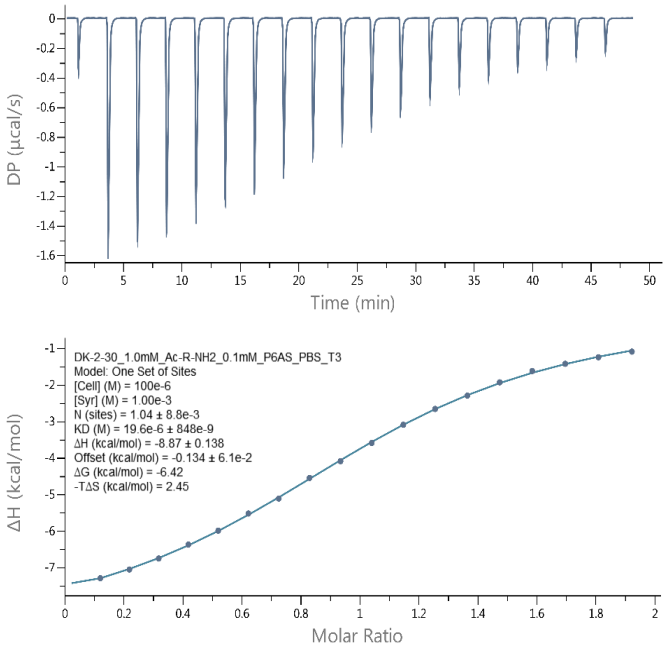

Figure S18. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with Ac-R-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (5.10 ± 0.02) x 10^4 M^{-1}.

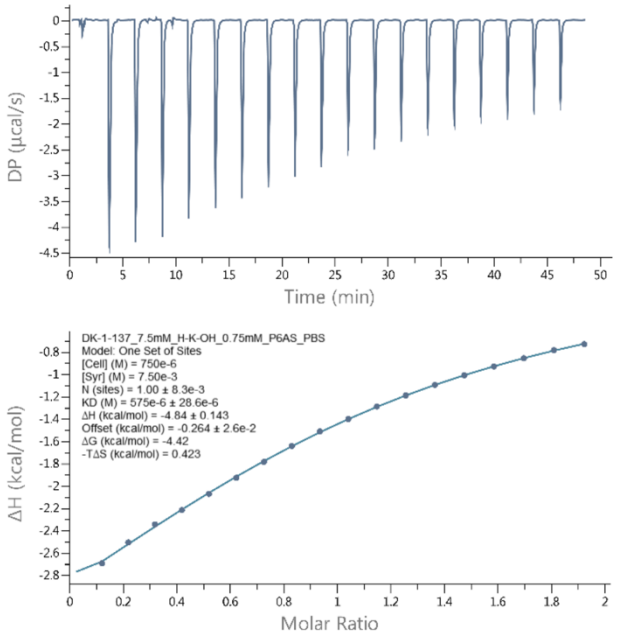

Figure S19. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.75 mM) in the cell was titrated with H-K-OH (7.5 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.74 ± 0.09) x 10^3 M^{-1}.

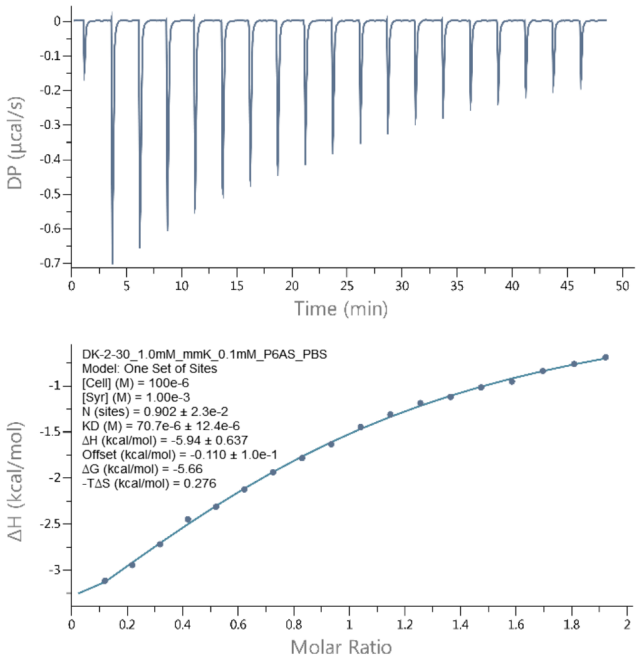

Figure S20. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with H-K(Me)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.40 ± 0.03) x 10^4 M^{-1}.

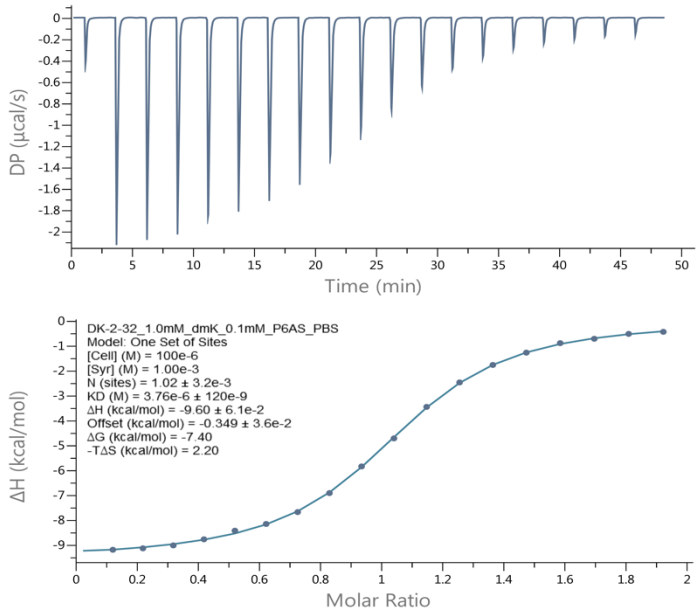

Figure S21. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with H-K(Me_2)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (2.39 ± 0.07) x 10^5 M^{-1}.

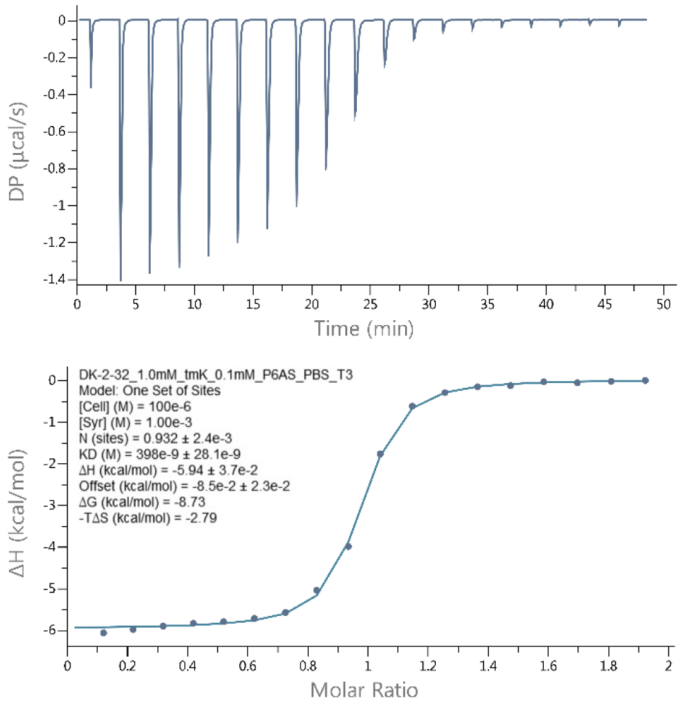

Figure S22. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with H-K(Me_3)-OH (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (3.39 ± 0.15) x 10^6 M^{-1}.

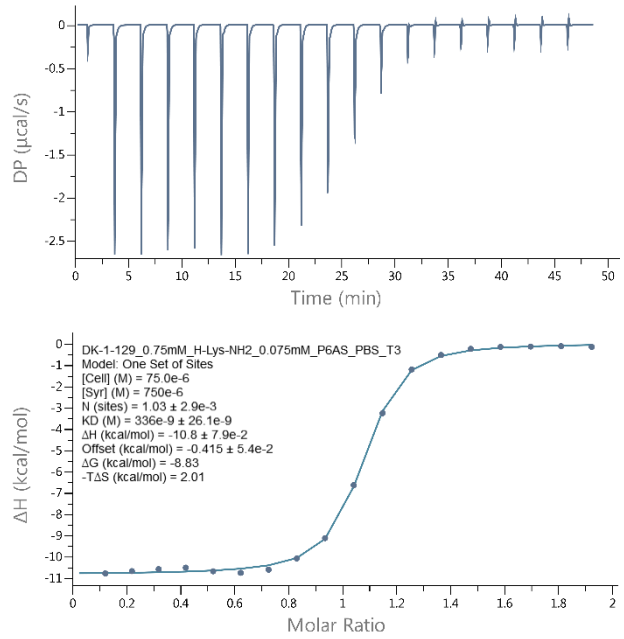

Figure S23. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.075 mM) in the cell was titrated with H-K-NH_2 (0.75 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (4.78 ± 0.80) x 10^6 M^{-1}.

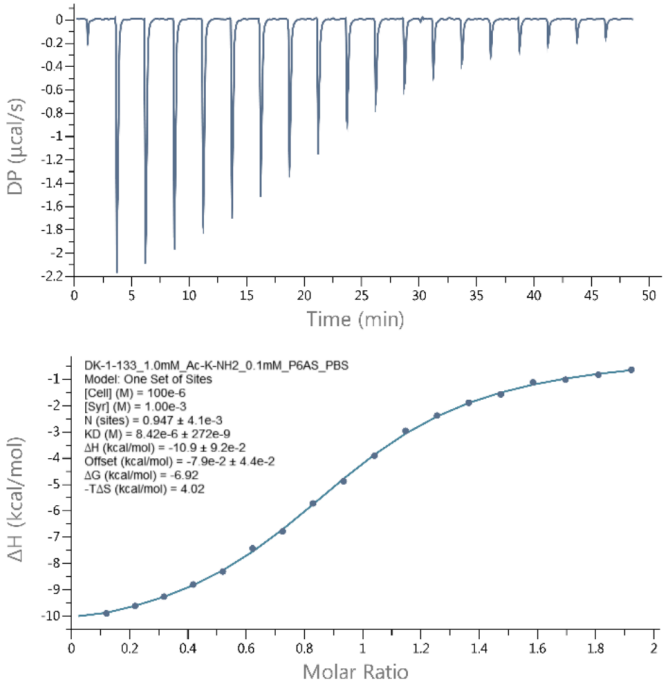

Figure S24. Isothermal Titration Calorimetry (ITC) curve obtained through direct binding titration studies. A solution of P6MQ (0.1 mM) in the cell was titrated with Ac-K-NH_2 (1.00 mM) in the syringe at 298.0 K in 10 mM sodium phosphate buffered saline at pH 7.4. K_a = (1.19 ± 0.04) x 10^5 M^{-1}.

Direct and competitive fluorescence binding assays of P6MQ with H3K4 peptides.

Direct fluorescence titration assays. A direct titration of P6MQ into indicator was performed via serial dilutions as follows. All solutions were made using Milli-Q® ultrapure water and 10 mM PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) at pH 7.4. In the 96-well plate, row 12 contained 200 µL of a solution of 100 nM DAPI and 4 µM P6MQ in 10 mM PBS buffer. Row 1 – 11 contained 100 µL of a solution of 100 nM DAPI in 10 mM PBS buffer. 100 µL of the solution in well 12 was added to well 11 and mixed, then 100 µL was taken from row 11 and added to row 10. These dilutions continued to well 1, whereupon 100 µL was discarded after mixing, leaving the total volume of all wells (1 – 12) at 100 µL. In a separate column of the 96-well plate, three wells contained the buffer blank (10 mM PBS), three wells contained the host blank (4 µM P6MQ in 10 mM PBS) and three wells contained the dye blank (100 nM DAPI in 10 mM PBS) all at a volume of 100 µL. Following the serial dilution of P6MQ across the plate, the plates were centrifuged (1300 rpm, pulse approximately 10 seconds) to remove any air bubbles. The plate was read on a BioTek Cytation-5 cell imaging and multi-mode plate reader, λ_{ex} 360/20 nm and λ_{em} 450/20 nm.

Two sets of triplicates were performed on two different days for a total of six replicates. The data for each triplicate was plotted and a K_d with standard error was determined using GraphPad Prism 9. The fluorescence (RFU) as a function of host concentration was plotted and a direct one site binding equation was used (equation S1). The fitting was constrained to the concentration of DAPI (100 nM). The graphs were plotted with standard error and a second plot of x residuals was determined. The IC_{90} point from the direct titrations was selected as the host concentration used in subsequent competitive titrations, this was determined to be 125 nM. The average K_d values of the two sets of triplicates was reported along with the propagated standard error. For the direct assays, some of the data points (0 to 2 points) representing the highest P6MQ concentrations had to be excluded due to aggregation.

Equation S1: Curve fit for direct titration

$$Y = F_{min} + (F_{max} - F_{min})\left(\frac{(D+x+K_{d\,DAPI}) - \sqrt{(D+x+K_{d\,DAPI})^2 - (4xD)}}{2D}\right) \quad \text{(S1)}$$

Where: Y = Fitted data point; F_{max} = Maximum signal; F_{min} = Minimum signal; D = Concentration of DAPI in micromolar (0.100 μM); x = Concentration of titrant in micromolar; $K_{d\ DAPI}$ = Dissociation constant of DAPI

Competitive fluorescence titration assays. Competitive titration assays were performed in a 96-well plate with 4′,6-diamidino-2-phenylindole (DAPI) as the indicator. Solutions were made using Milli-Q® ultrapure water and 10 mM PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) at pH 7.4. Two solutions were used, solution A and solution B. Solution B contained P6MQ (125 nM), DAPI (100 nM), and the guest of interest (between 1 – 100 μM) in 10 mM PBS buffer. Solution A contained P6MQ (125 nM) and DAPI (100 nM) in 10 mM PBS buffer at pH 7.4. In the 96 well plate, row 12 contained 200 μL solution B, row 1 – 11 contained 100 μL solution A. 100 μL from well 12 was added to row 11 and mixed, then 100 μL was taken from row 11 and added to row 10. These dilutions continued to row 1, whereupon 100 μL was discarded after mixing, leaving the total volume of all wells (1 – 12) at 100 μL. In a separate column of the 96 well plate three wells contained a buffer blank (10 mM PBS), three wells contained a host blank (125 nM P6MQ in 10 mM PBS), three wells contained a dye blank (100 nM DAPI in 10 mM PBS), three wells contained a guest blank (guest at highest concentration used in assay in 10 mM PBS buffer) and three wells contained a solution A blank (solution A not diluted by solution B) all at a volume of 100 μL. Following serial dilution of P6MQ across the plate, the plates were centrifuged (1300 rpm, pulse approximately 10 seconds) to remove any air bubbles. The plate was read on a BioTek Cytation-5 cell imaging and multi-mode reader plate reader, λ_{ex} 360/20 nm and λ_{em} 450/20 nm.

Two sets of triplicates for each guest were performed on two different days for a total of six replicates. The data for each triplicate was plotted and a K_d with standard error was determined. The logarithm of the guest concentration against the fluorescence response from DAPI displacement (Fl-Fl_o, RFU) was plotted. A blank containing only the host (Fl_o) was subtracted from each triplicate experiment. The data was analyzed in GraphPad Prism 9 using the competitive one site binding equation (Equation S2). The fitting was constrained by the direct dissociation constant of DAPI ($K_{d\ DAPI}$, 15 nM) and the concentration of DAPI (100 nM). This model fits the K_d directly and does not report the EC_{50}, EC_{50} in equation S2 is solved by equation S3. This analysis also

assumes a 1:1 stoichiometry. Graphs were plotted with standard error and a second plot of x residuals were determined. The average K_d values of the two sets of triplicates was reported along with the propagated standard error. For the competitive assays with KMe_3, some of the data points (0 to 2 points) representing the highest KMe_3 concentrations had to be excluded due to aggregation.

Equation S2: Curve fit for competitive titration

$$Y = \frac{(F_{max} - F_{min})}{1+10^{(x - logEC_{50})}} + F_{min} \quad \text{(S2)}$$

Where: Y = Fitted data point; F_{max} = Maximum signal; F_{min} = Minimum signal; $logEC_{50}$ is solved by the equation below;

Equation S3: Curve fit for competitive titration

$$logEC_{50} = log\left(10^{logK_{d\ guest}\left(1+\frac{D}{K_{d\ DAPI}}\right)}\right) \quad \text{(S3)}$$

Where: $K_{d\ guest}$ = Dissociation constant for guest; D = Concentration of DAPI in nM; $K_{d\ DAPI}$ = Dissociation constant of the DAPI in nM.

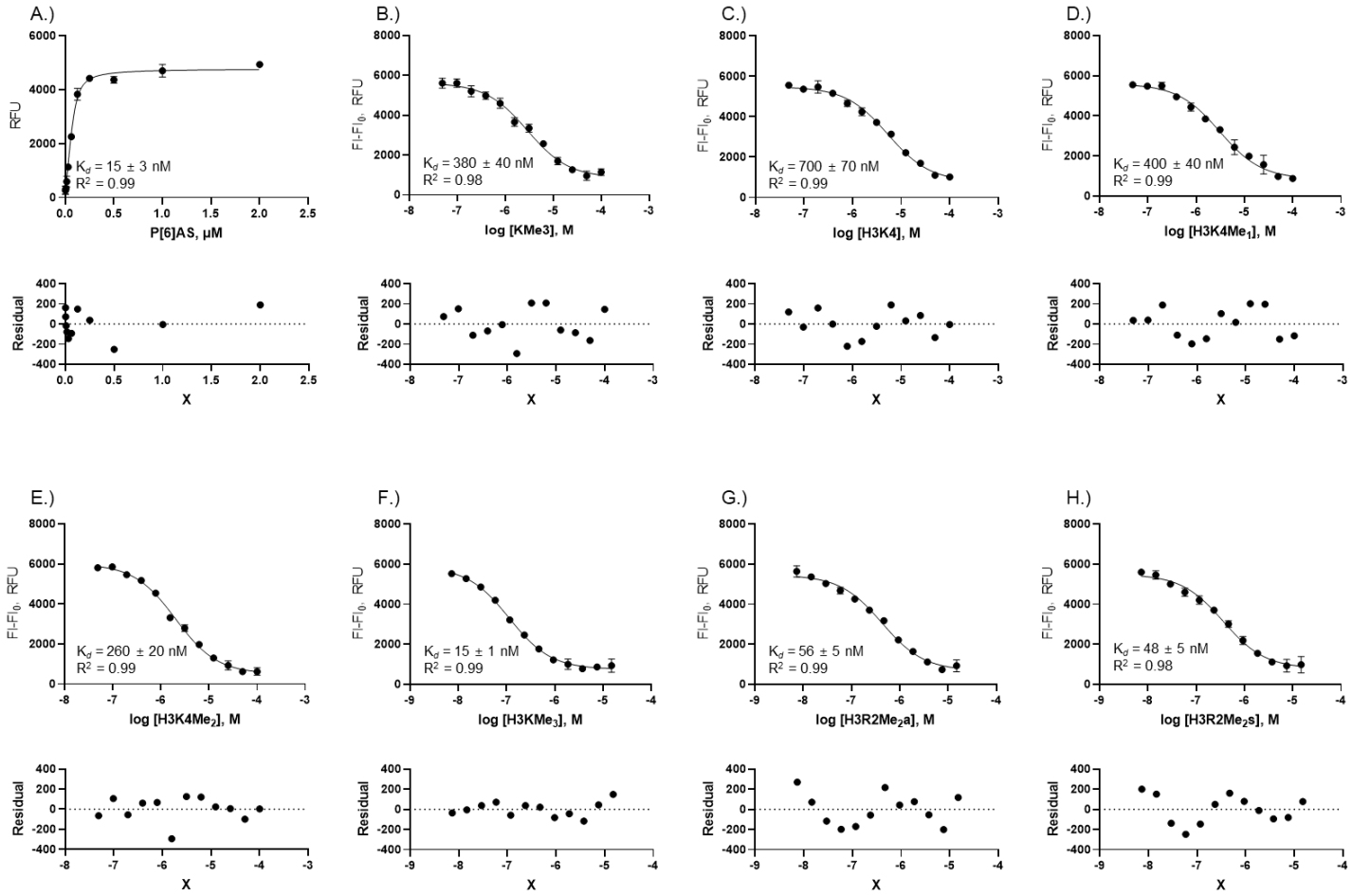

Figure S25. Representative titration curves of direct and competitive titrations with residuals. Reported dissociation constants and R^2 are for an exemplary set of triplicates. A.) Direct titration of P6MQ into DAPI (100 nM). B.) Competitive titration of H-K(Me_3)-OH (100 µM) into P6MQ (125 nM) and DAPI (100 nM). C.) Competitive titration of H3K4 (100 µM) into P6MQ (125 nM) and DAPI (100 nM). D.) Competitive titration of H3K4Me_1 (100 µM) into P6MQ (125 nM) and DAPI (100 nM). E.) Competitive titration of H3K4Me_2 (100 µM) into P6MQ (125 nM) and DAPI (100 nM). F.) Competitive titration of H3K4Me_3 (15 µM) into P6MQ (125 nM) and DAPI (100 nM). G.) Competitive titration of Asym-H3R2Me_2 (15 µM) into P6MQ (125 nM) and DAPI (100 nM) H.) Competitive titration of Sym-H3R2Me_2 (15 µM) into P6MQ (125 nM) and DAPI (100 nM).

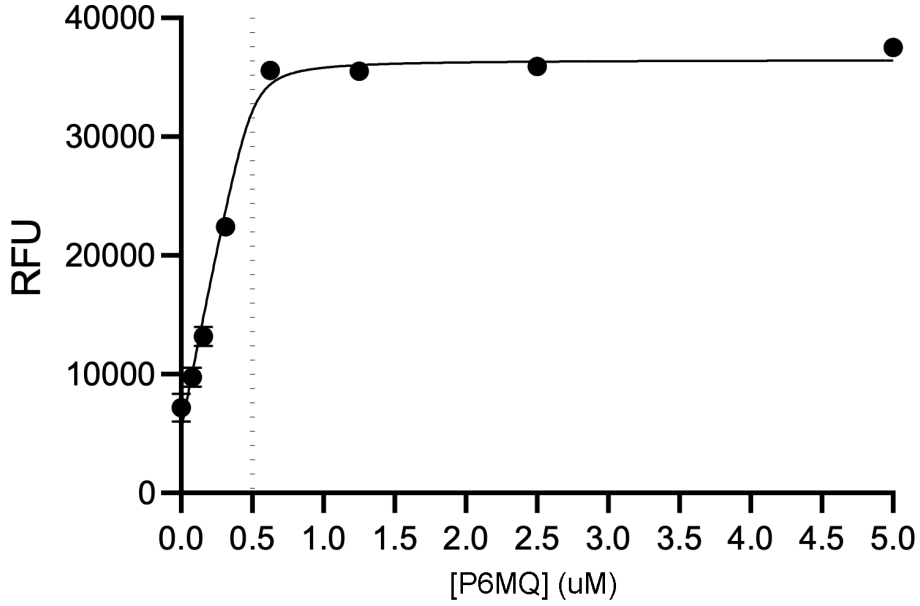

Figure S26. Determination of P[6]AS•DAPI stoichiometry. Direct titration of P[6]AS into DAPI (0.5 μM). Preformed in duplicate in 10 mM phosphate buffer, pH 7.4. Saturation reached at 0.5 μM P[6]AS indicating 1:1 complexation.

^{1}H NMR spectra of P5MQ with selected guests.

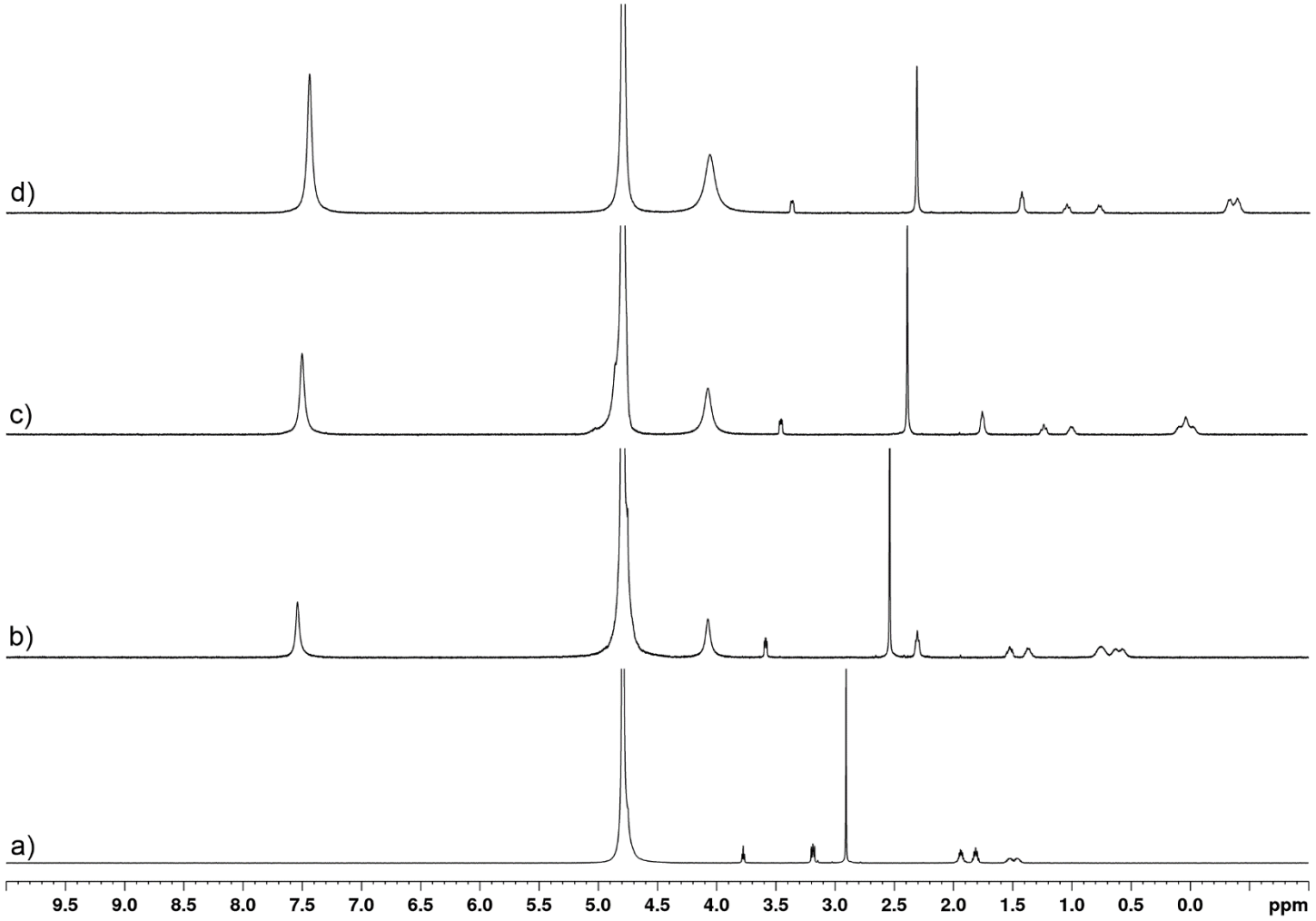

Figure S27. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_2)-OH (1.0 mM), b) a mixture of H-K(Me_2)-OH (1.0 mM) and P5MQ (0.5 mM), c) an equimolar mixture of H-K(Me_2)-OH and P5MQ (1.0 mM), and d) a mixture of H-K(Me_2)-OH (1.0 mM) and P5MQ (2.0 mM).

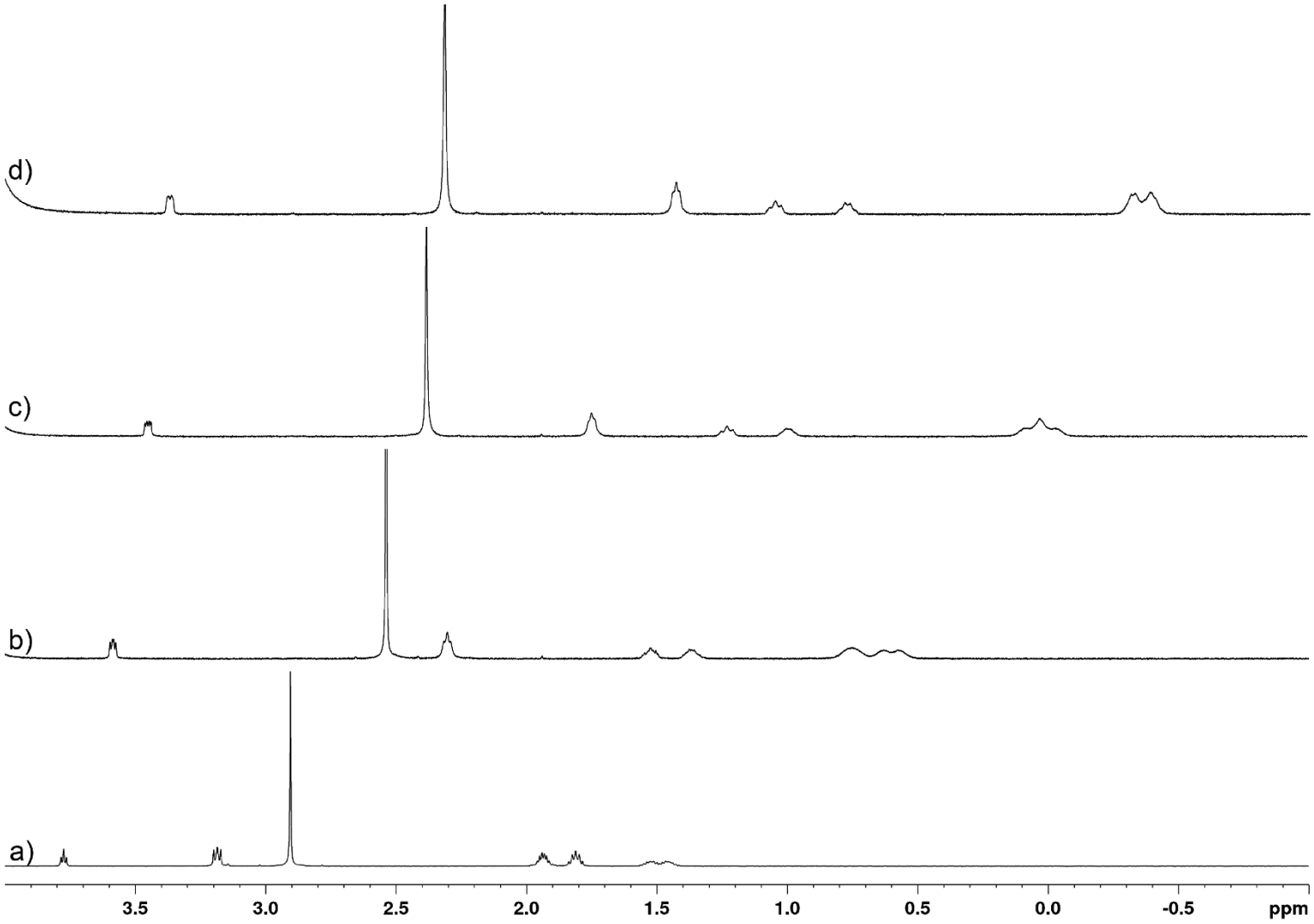

Figure S28. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_2)-OH (1.0 mM), b) a mixture of H-K(Me_2)-OH (1.0 mM) and P5MQ (0.5 mM), c) an equimolar mixture of H-K(Me_2)-OH and P5MQ (1.0 mM), and d) a mixture of H-K(Me_2)-OH (1.0 mM) and P5MQ (2.0 mM).

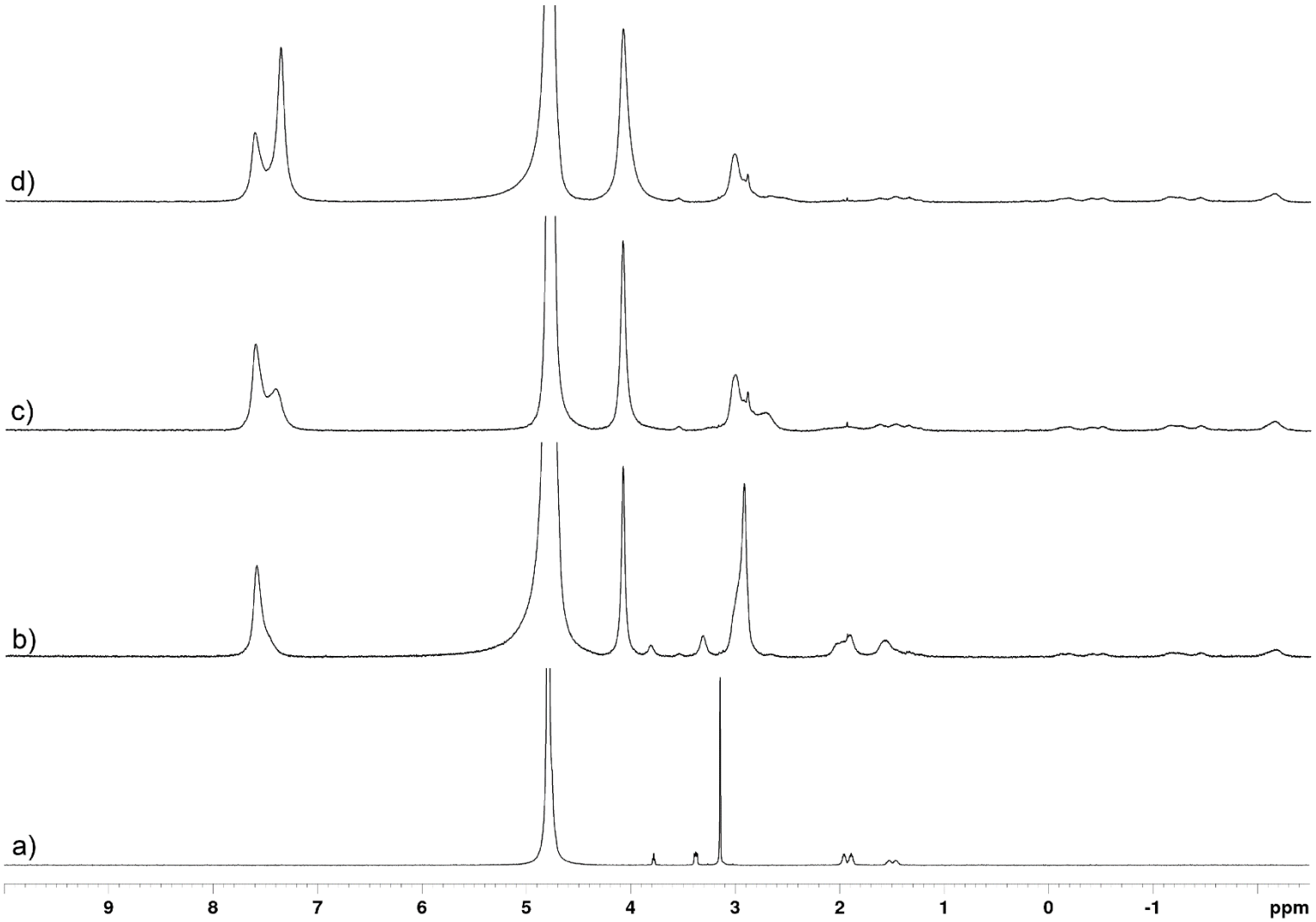

Figure S29. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_3)-OH (1.0 mM), b) a mixture of H-K(Me_3)-OH (1.0 mM) and P5MQ (0.5 mM), c) an equimolar mixture of H-K(Me_3)-OH and P5MQ (1.0 mM), and d) a mixture of H-K(Me_3)-OH (1.0 mM) and P5MQ (2.0 mM).

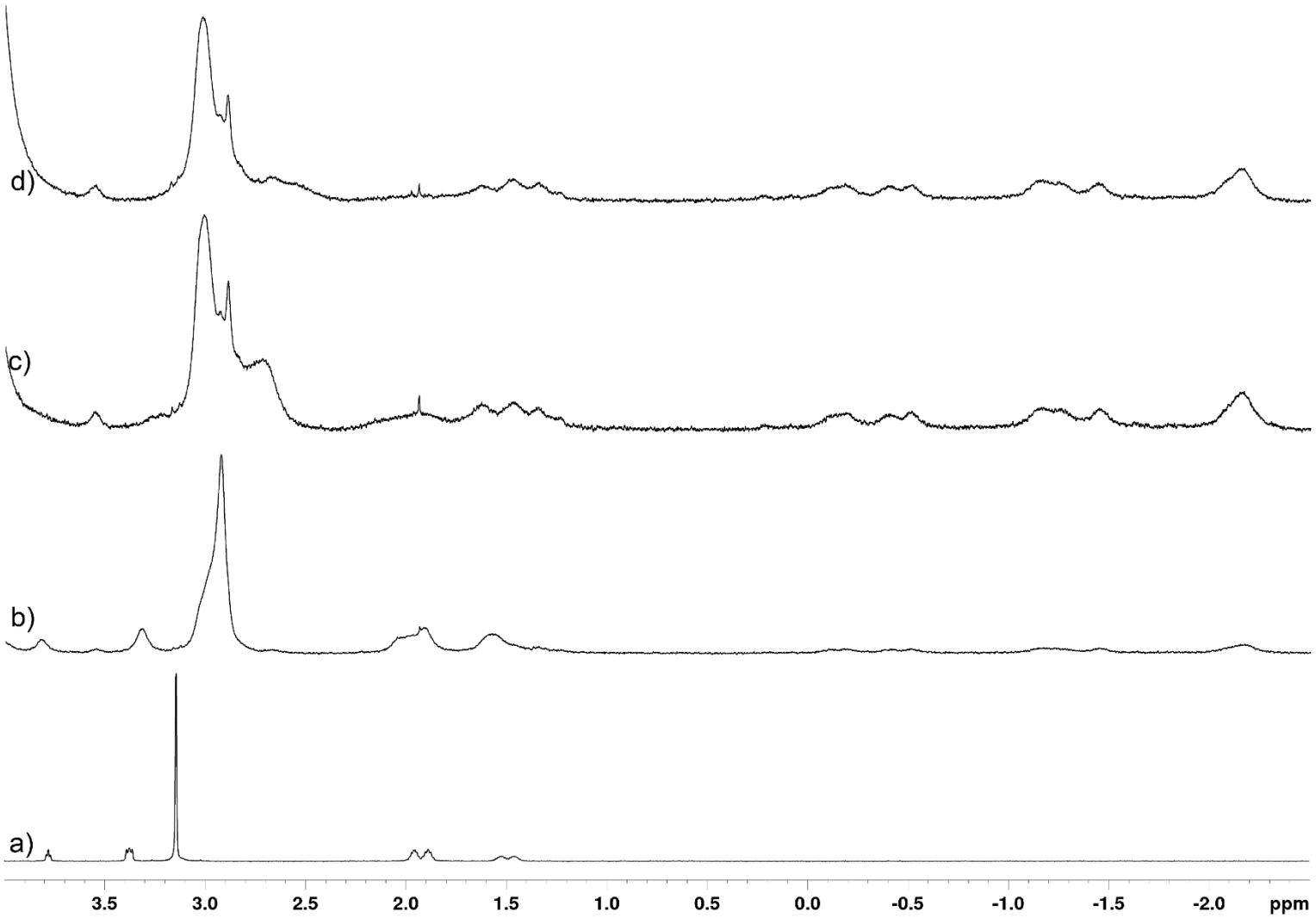

Figure S30. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_3)-OH (1.0 mM), b) a mixture of H-K(Me_3)-OH (1.0 mM) and P5MQ (0.5 mM), c) an equimolar mixture of H-K(Me_3)-OH and P5MQ (1.0 mM), and d) a mixture of H-K(Me_3)-OH (1.0 mM) and P5MQ (2.0 mM).

^{1}H NMR spectra of P6MQ with selected guests.

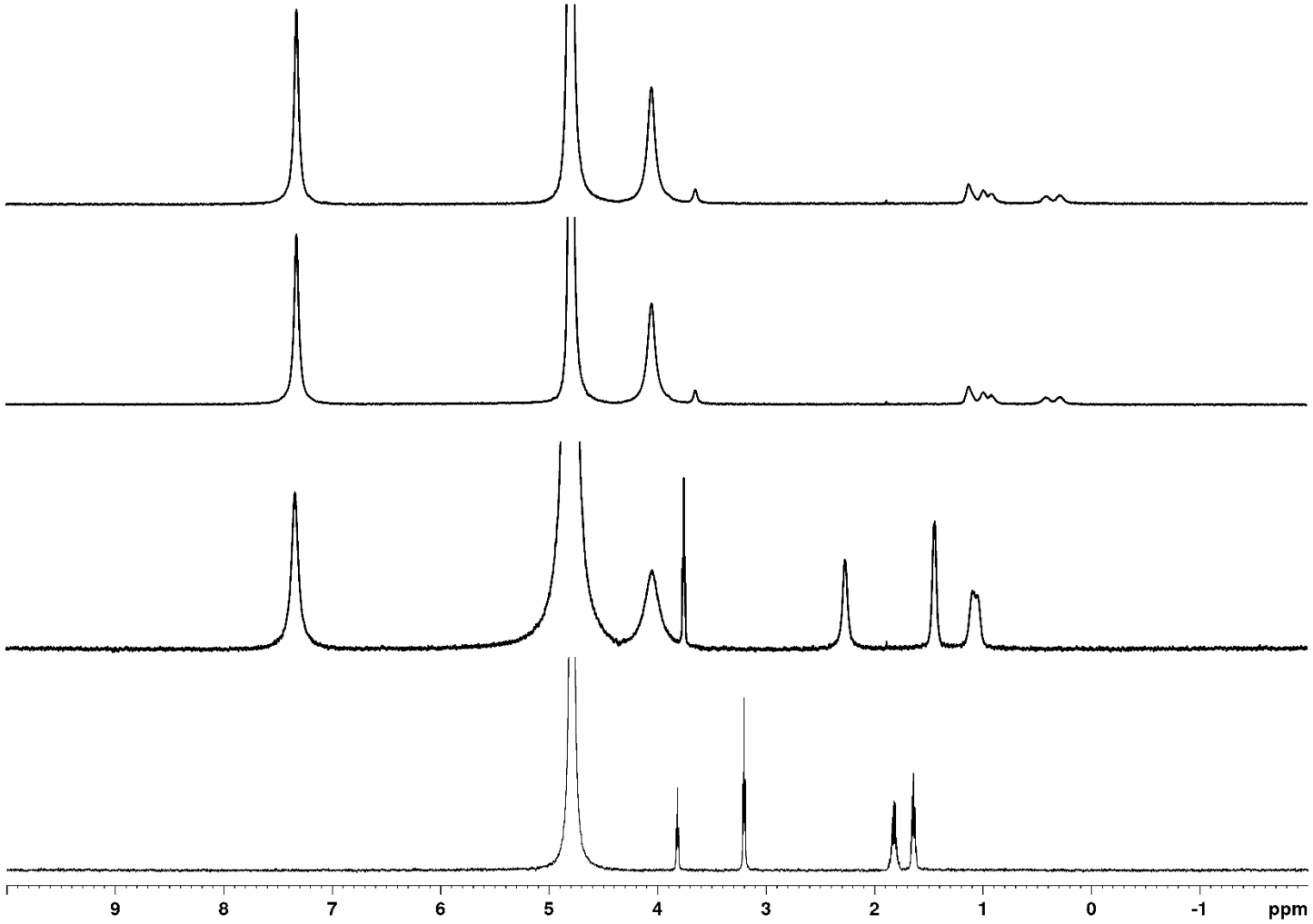

Figure S31. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-R-NH_2 (1.0 mM), b) a mixture of H-R-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-R-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-R-NH_2 (1.0 mM) and P6MQ (2.0 mM).

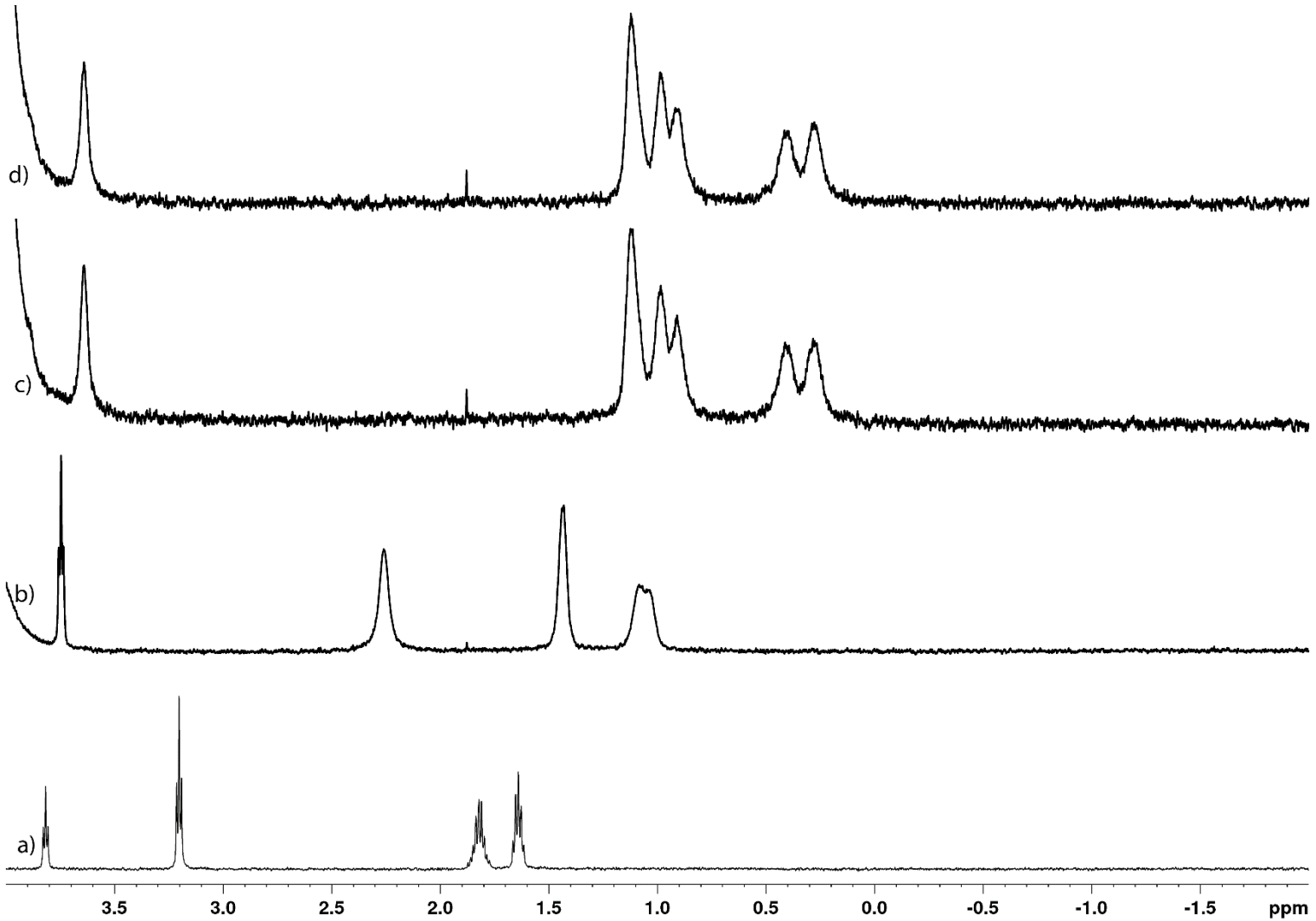

Figure S32. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-R-NH_2 (1.0 mM), b) a mixture of H-R-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-R-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-R-NH_2 (1.0 mM) and P6MQ (2.0 mM).

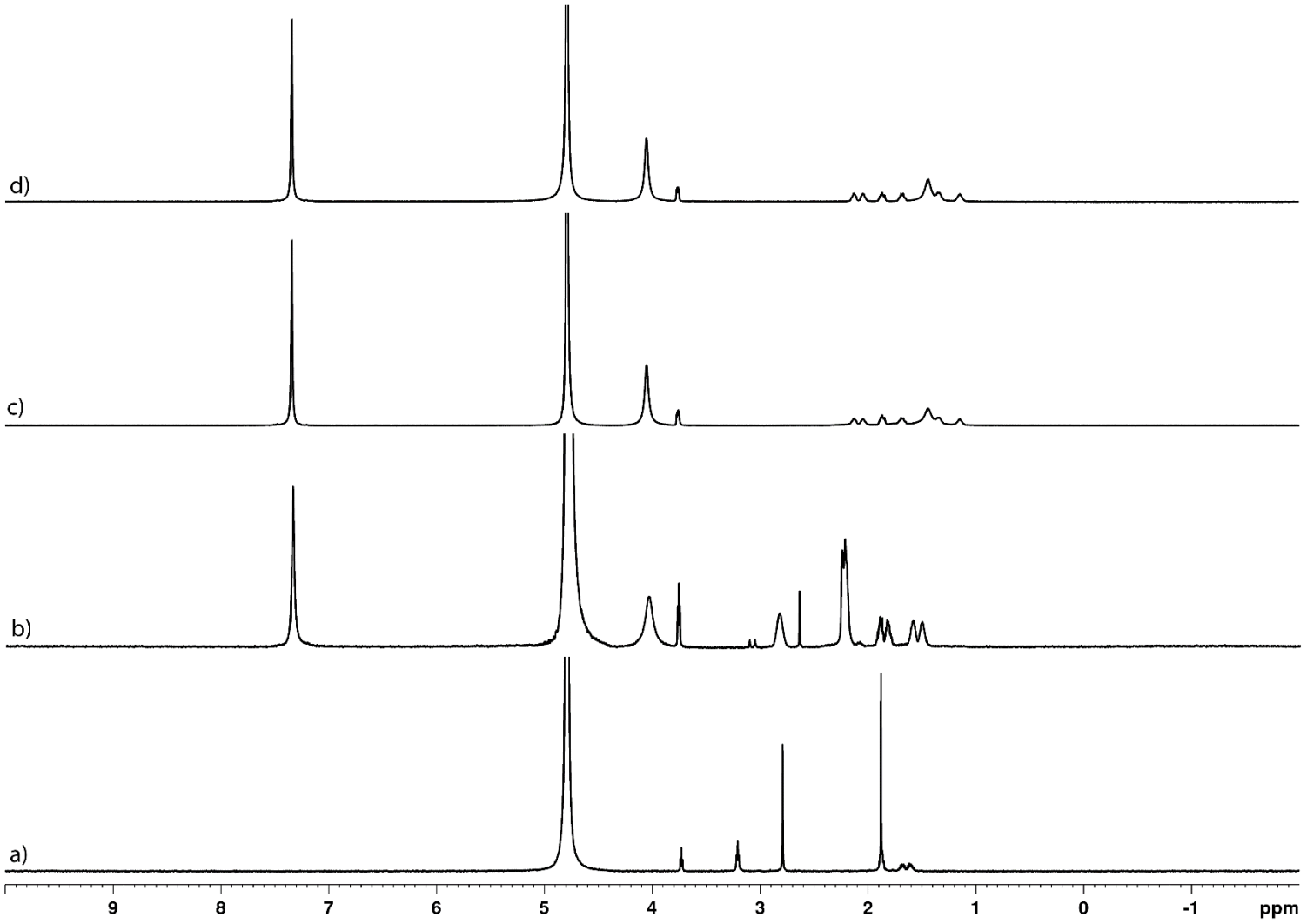

Figure S33. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) Sym-H-R(Me_2)-OH (1.0 mM), b) a mixture of Sym-H-R(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of Sym-H-R(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of Sym-H-R(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

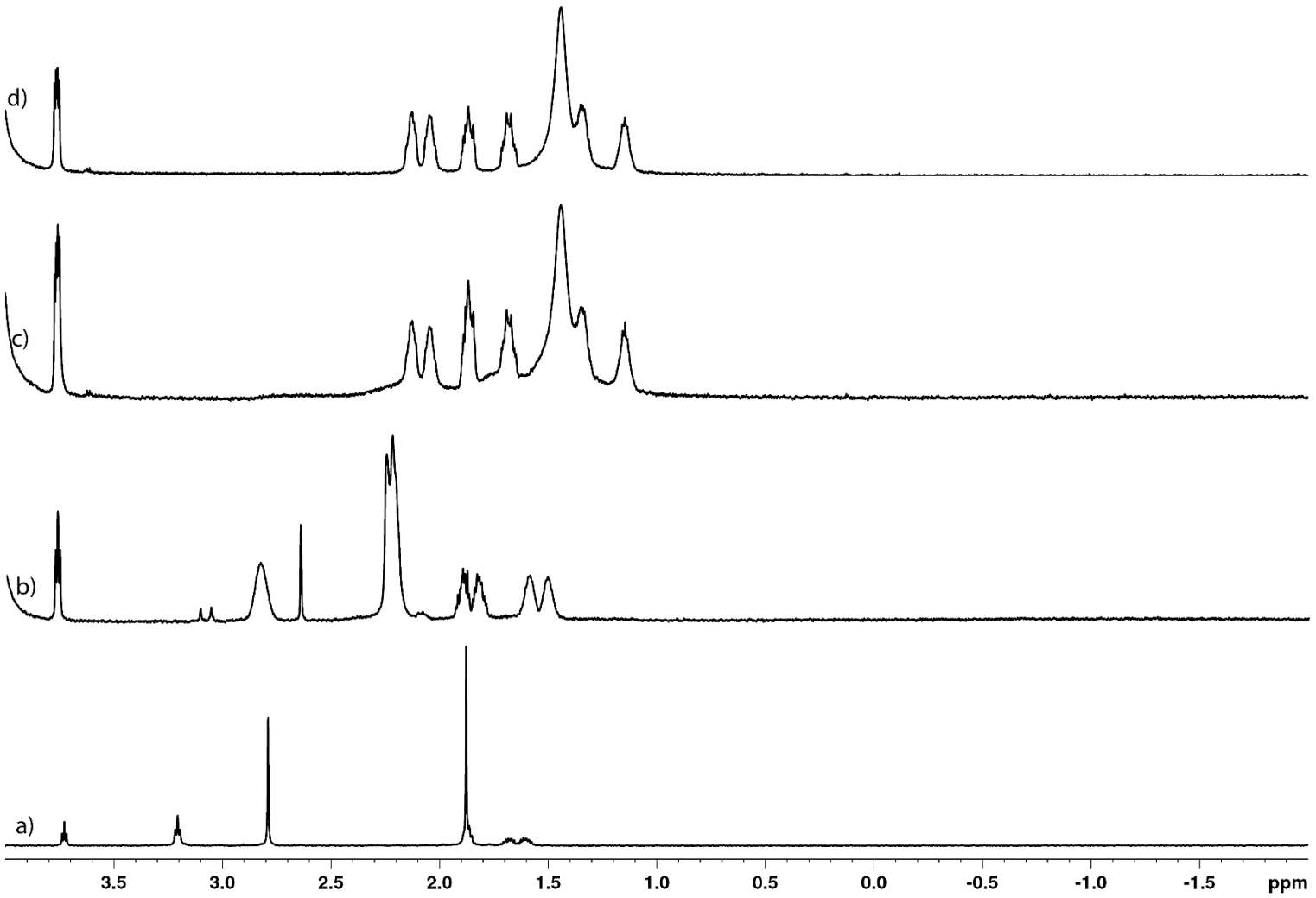

Figure S34. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) Sym-H-R(Me_2)-OH (1.0 mM), b) a mixture of Sym-H-R(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of Sym-H-R(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of Sym-H-R(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

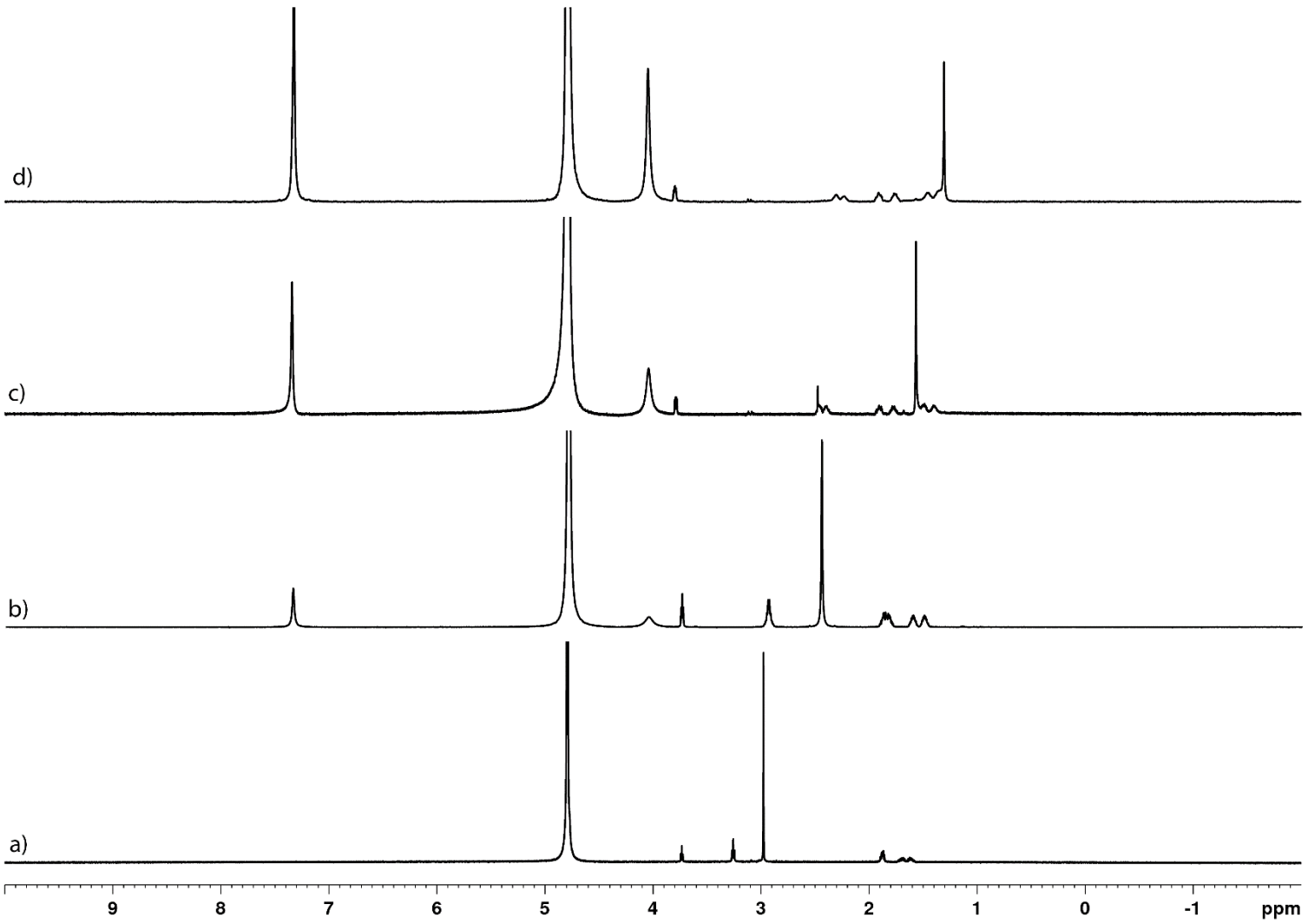

Figure S35. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) Asym-H-R(Me_2)-OH (1.0 mM), b) a mixture of Asym-H-R(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of Asym-H-R(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of Asym-H-R(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

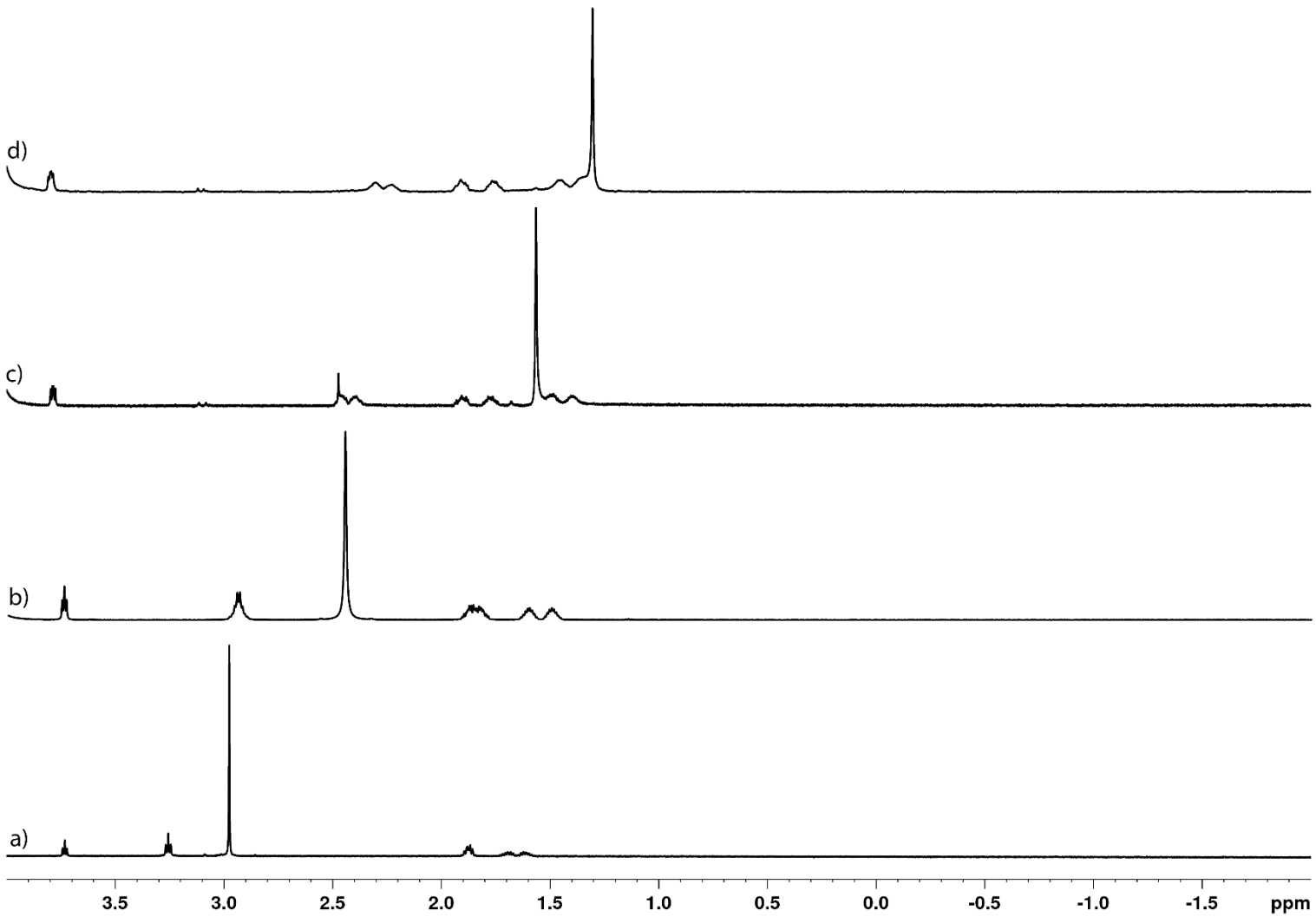

Figure S36. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) Asym-H-R(Me_2)-OH (1.0 mM), b) a mixture of Asym-H-R(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of Asym-H-R(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of Asym-H-R(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

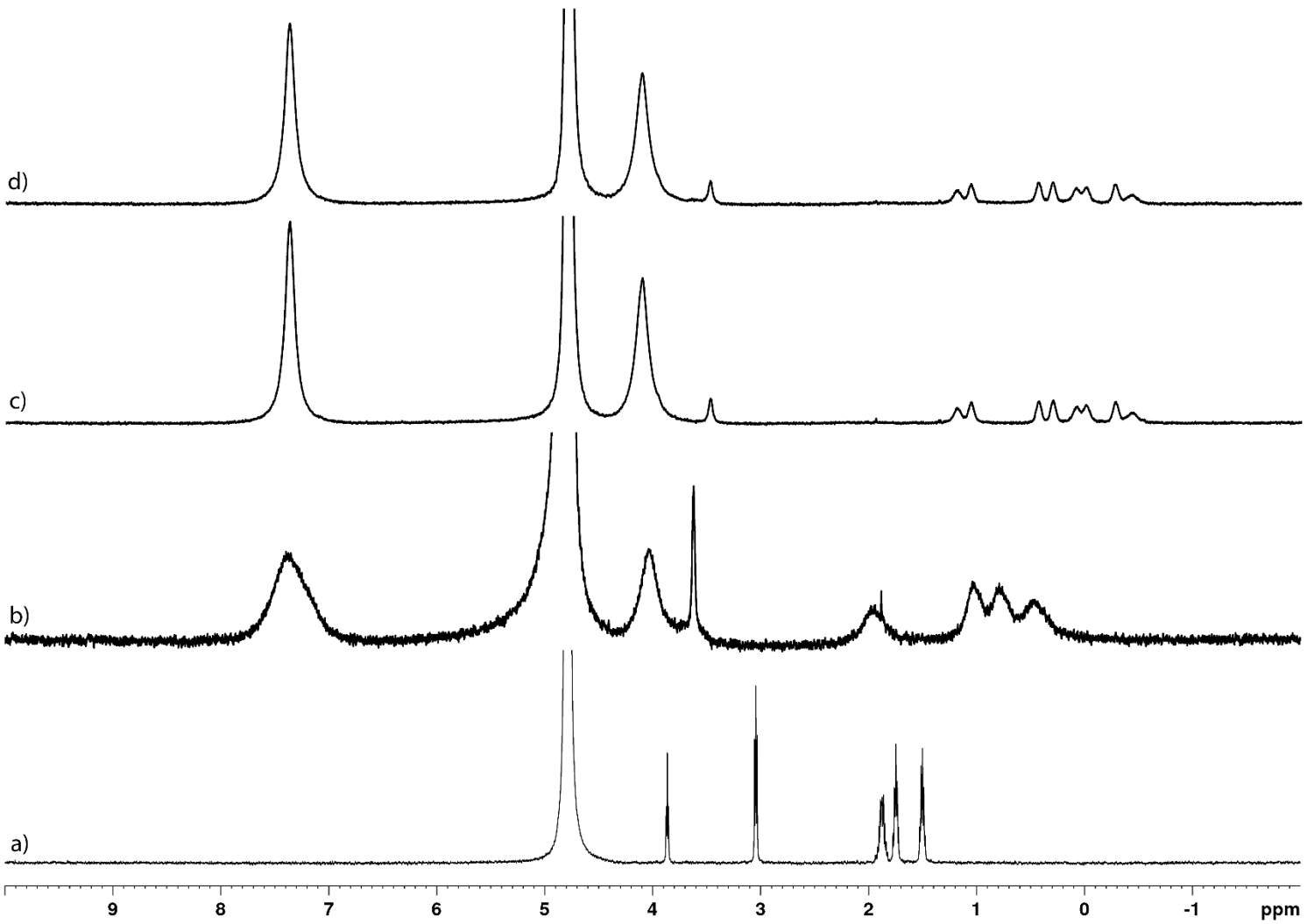

Figure S37. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K-NH_2 (1.0 mM), b) a mixture of H-K-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-K-NH_2 (1.0 mM) and P6MQ (2.0 mM).

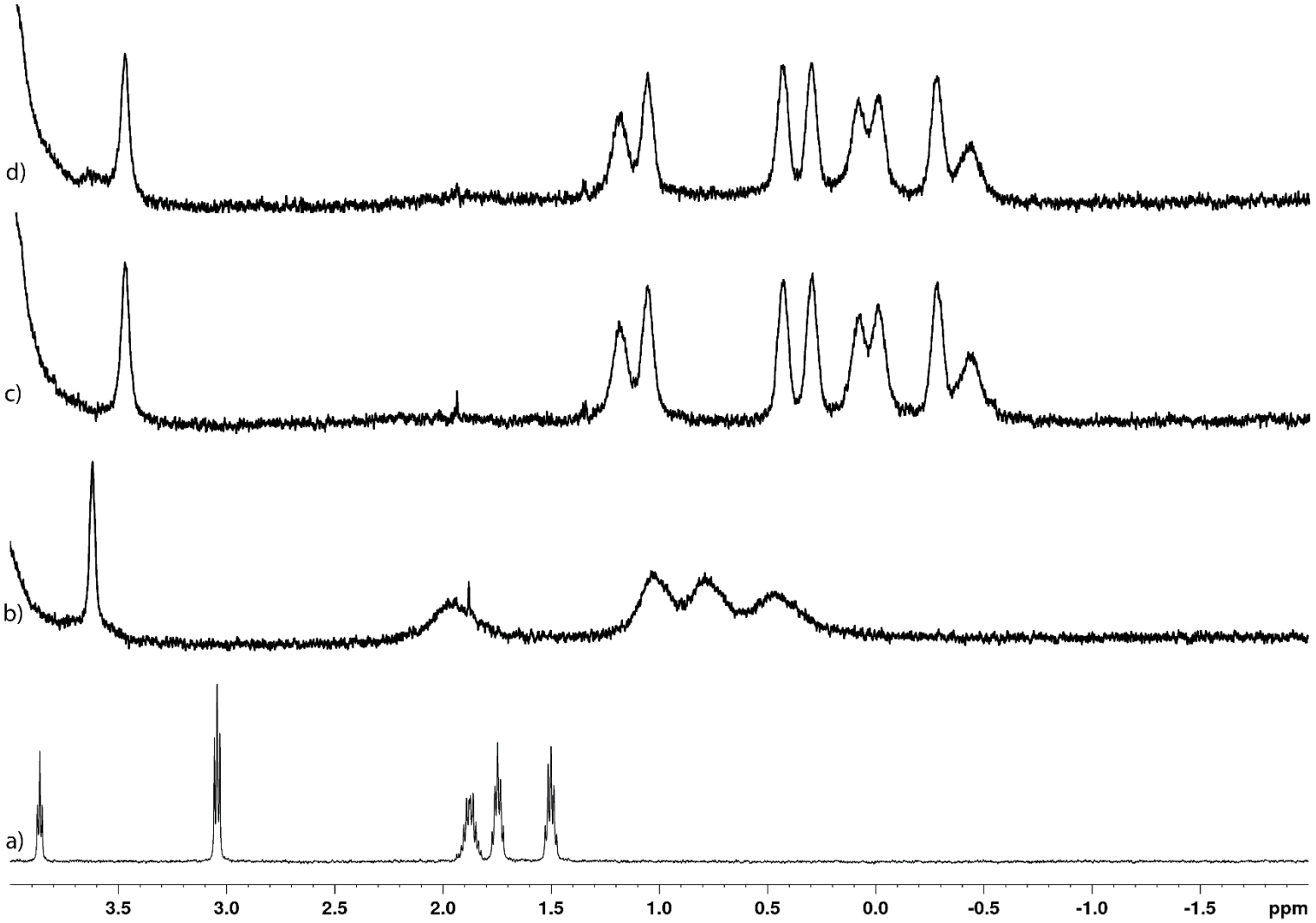

Figure S38. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K-NH_2 (1.0 mM), b) a mixture of H-K-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-K-NH_2 (1.0 mM) and P6MQ (2.0 mM).

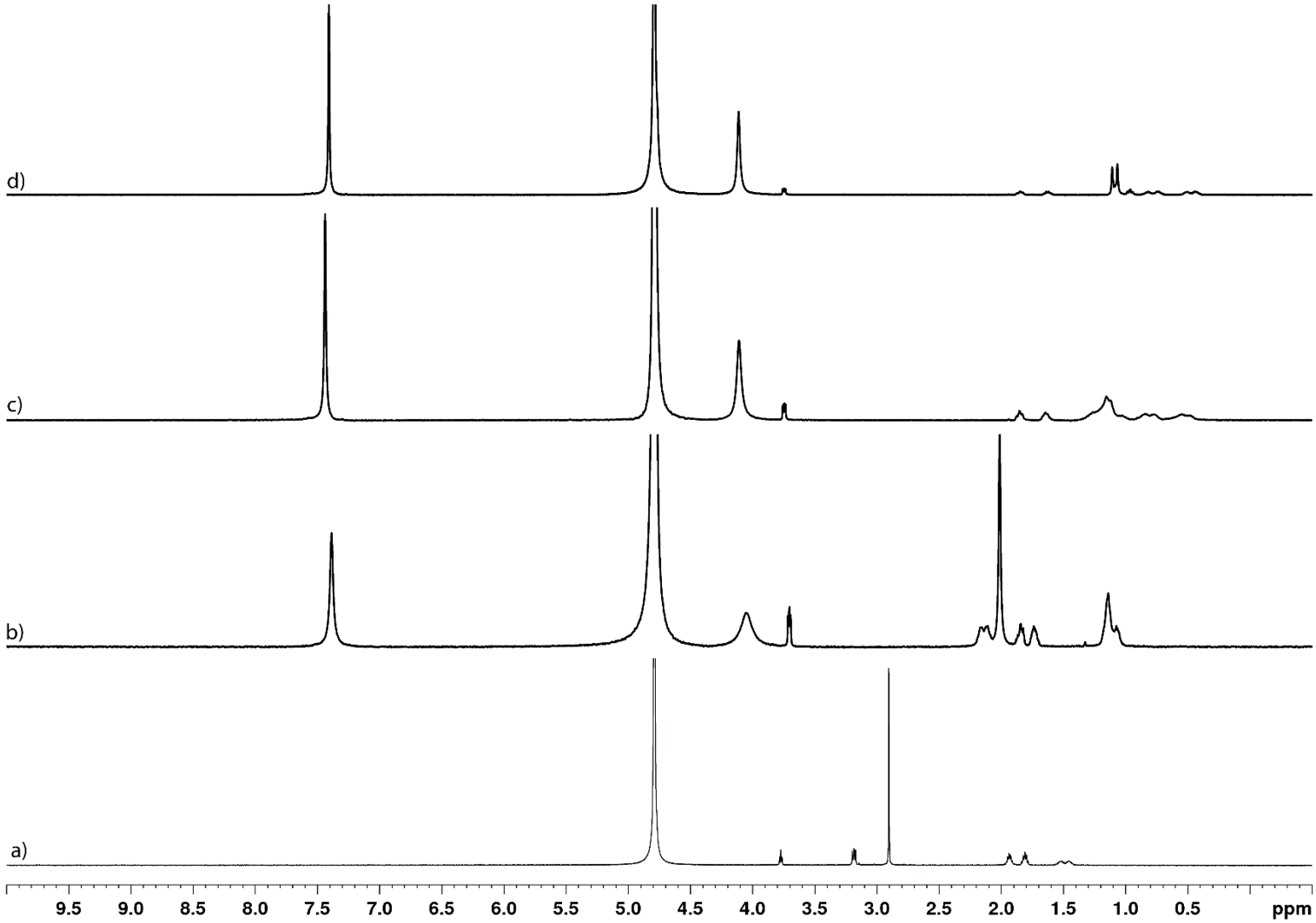

Figure S39. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_2)-OH (1.0 mM), b) a mixture of H-K(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of H-K(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

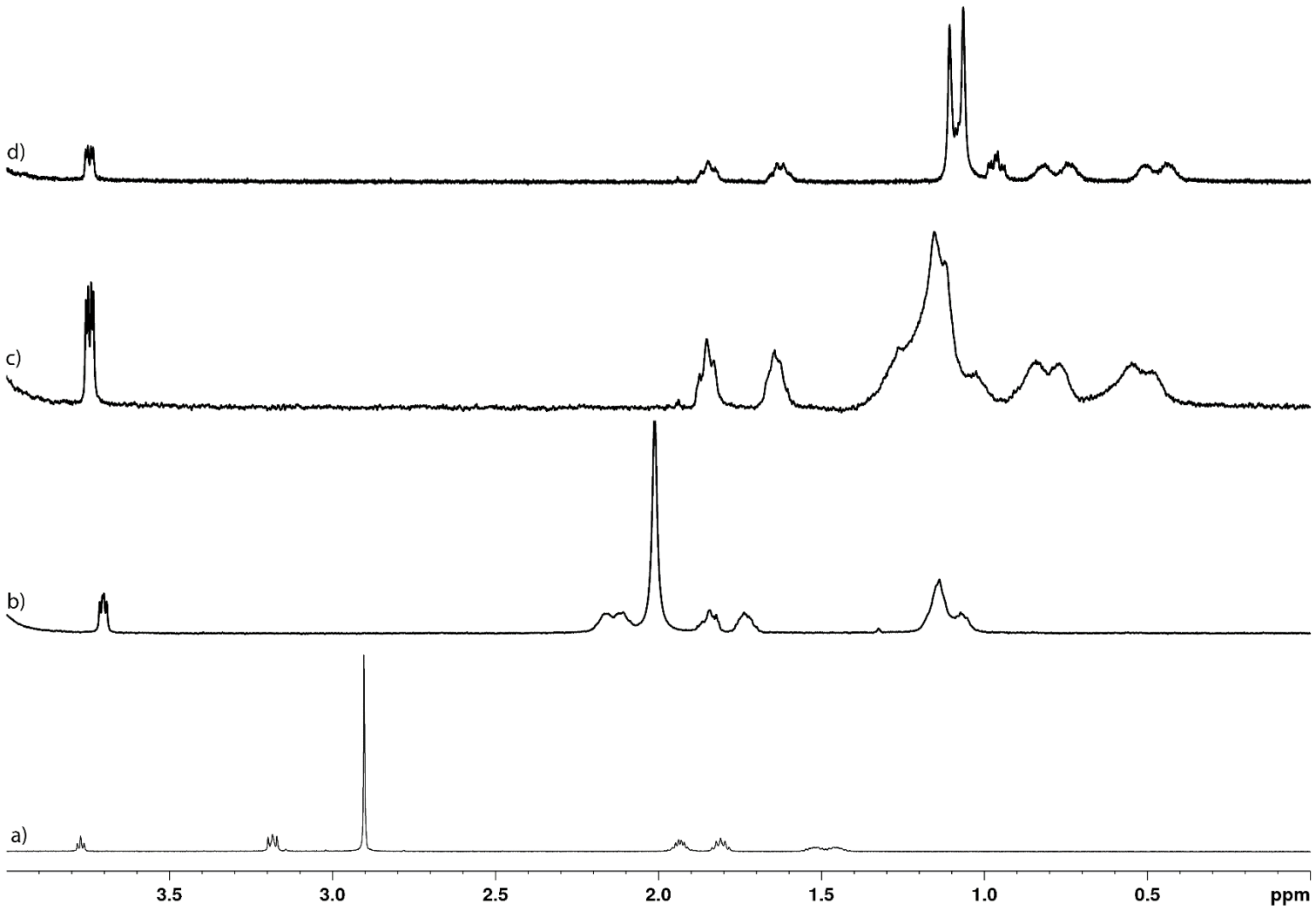

Figure S40. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_2)-OH (1.0 mM), b) a mixture of H-K(Me_2)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K(Me_2)-OH and P6MQ (1.0 mM), and d) a mixture of H-K(Me_2)-OH (1.0 mM) and P6MQ (2.0 mM).

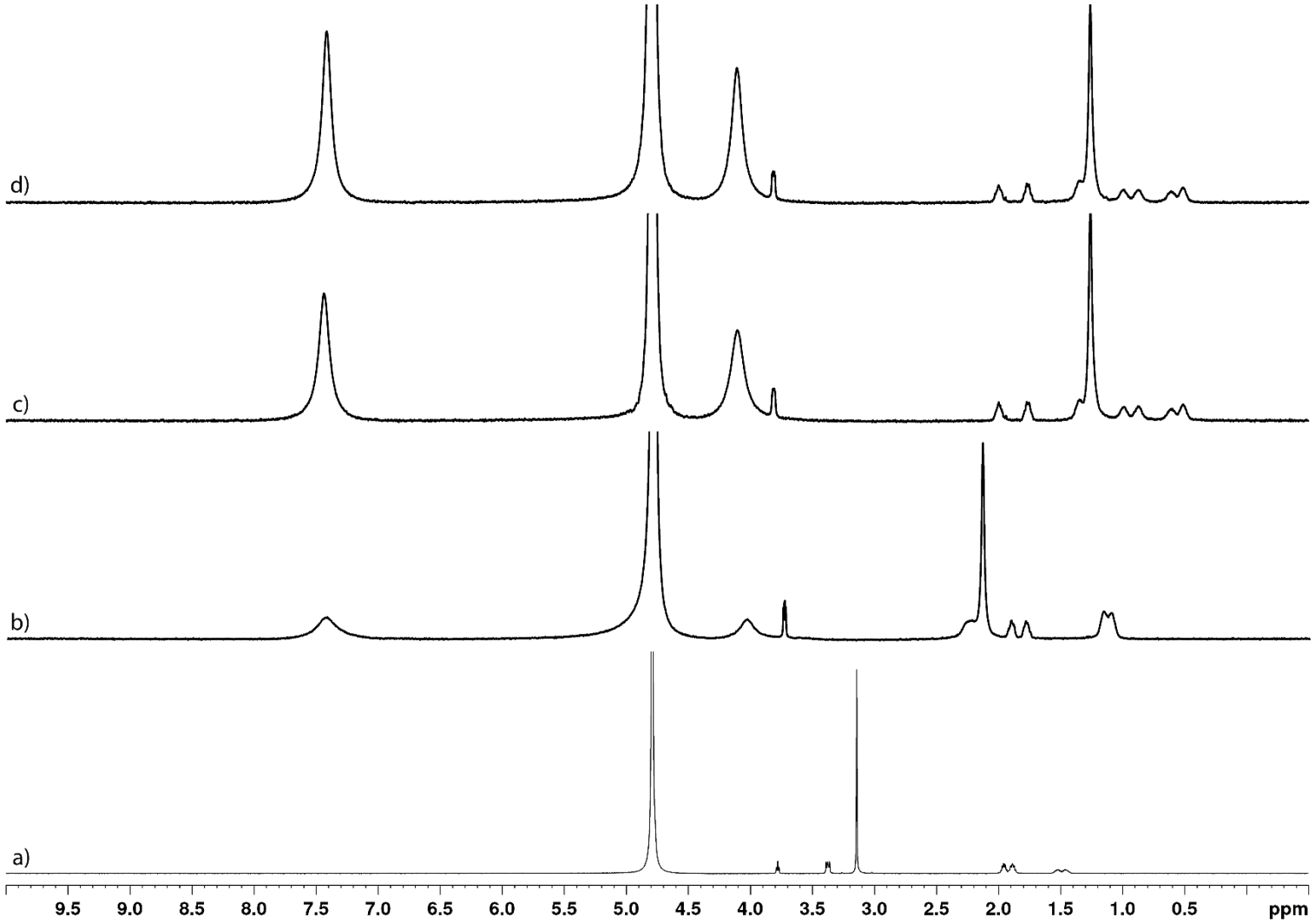

Figure S41. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_3)-OH (1.0 mM), b) a mixture of H-K(Me_3)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K(Me_3)-OH and P6MQ (1.0 mM), and d) a mixture of H-K(Me_3)-OH (1.0 mM) and P6MQ (2.0 mM).

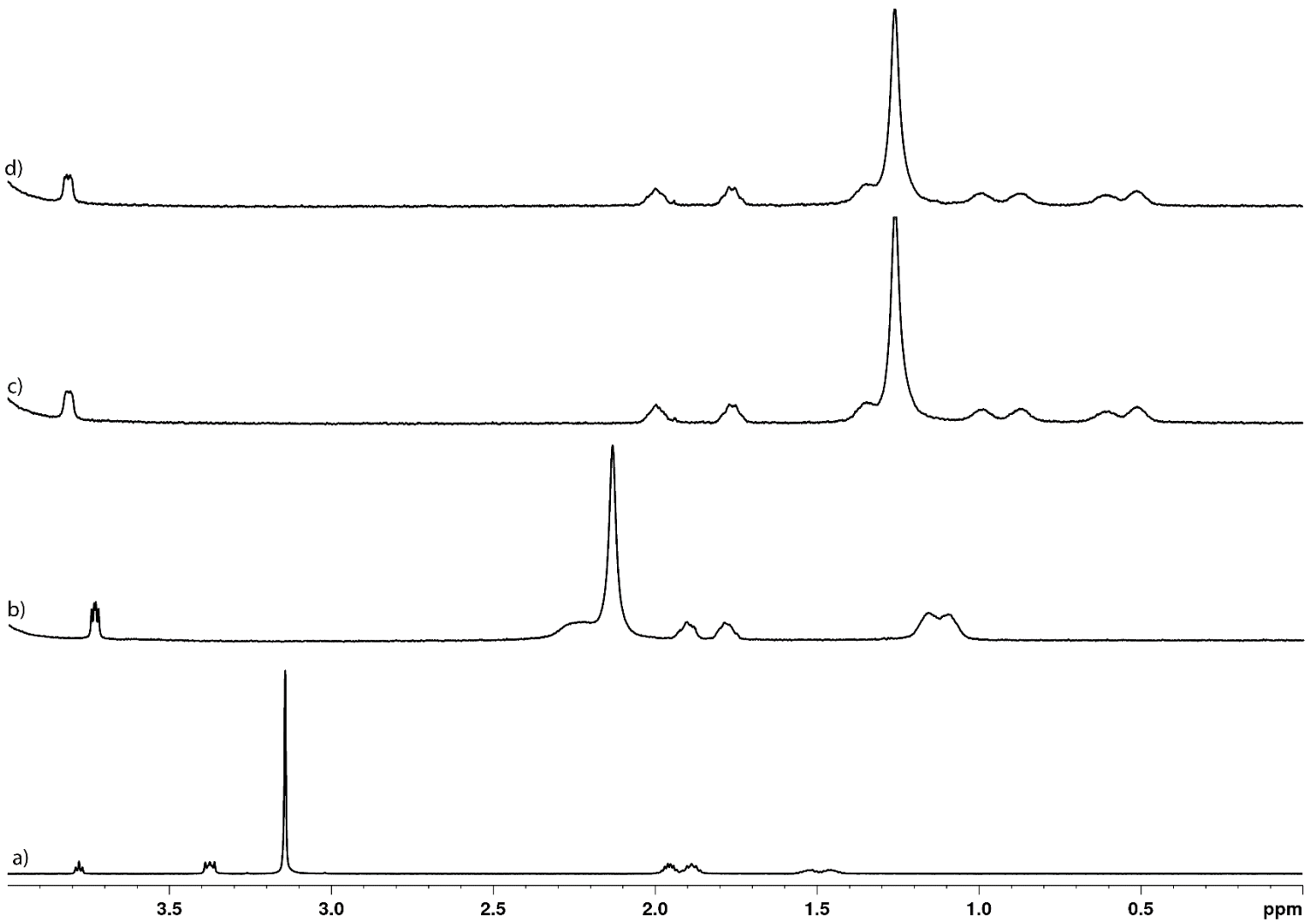

Figure S42. Guest region of ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-K(Me_3)-OH (1.0 mM), b) a mixture of H-K(Me_3)-OH (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-K(Me_3)-OH and P6MQ (1.0 mM), and d) a mixture of H-K(Me_3)-OH (1.0 mM) and P6MQ (2.0 mM).

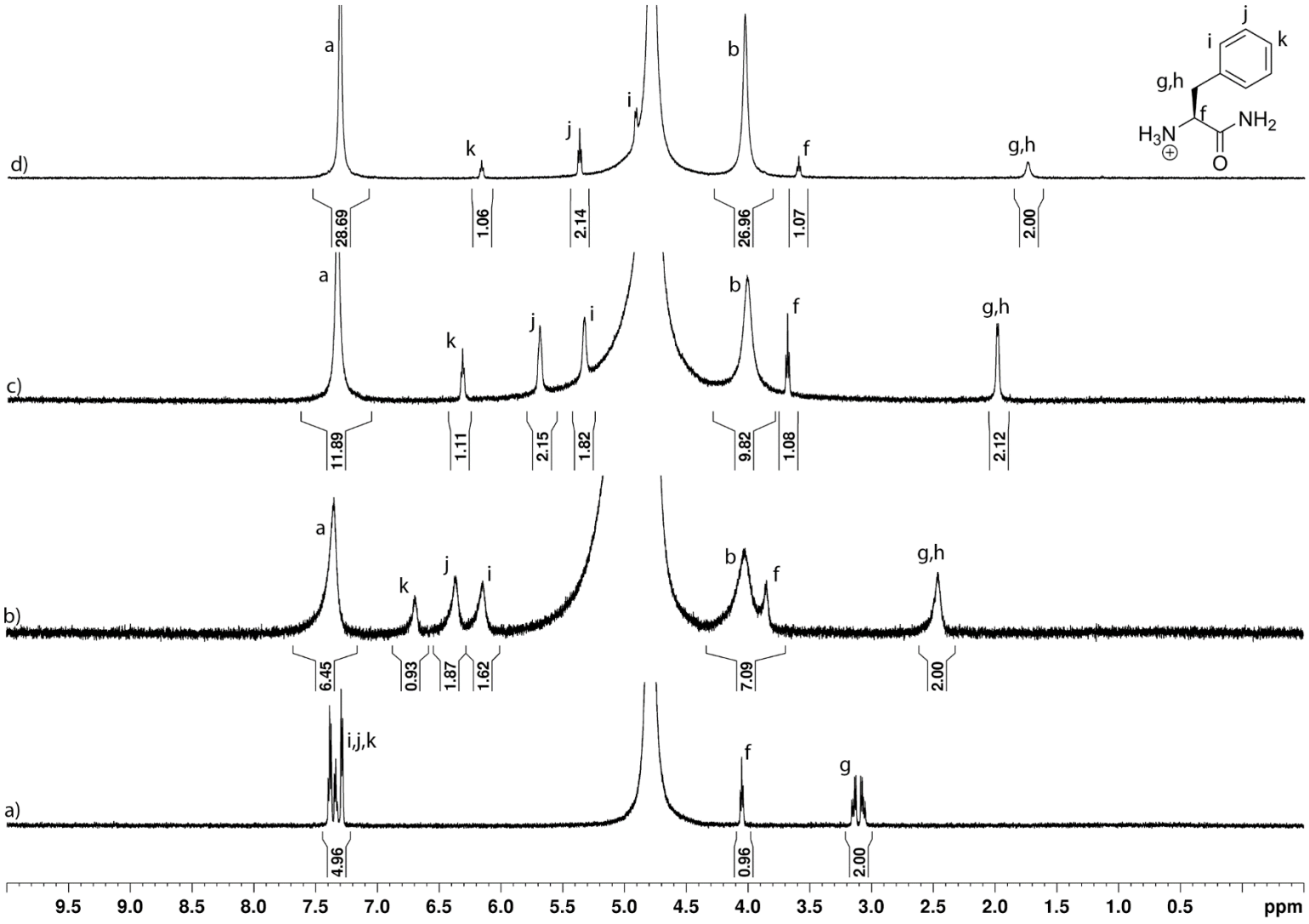

Figure S43. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-F-NH_2 (1.0 mM), b) a mixture of H-F-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-F-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-F-NH_2 (1.0 mM) and P6MQ (2.0 mM).

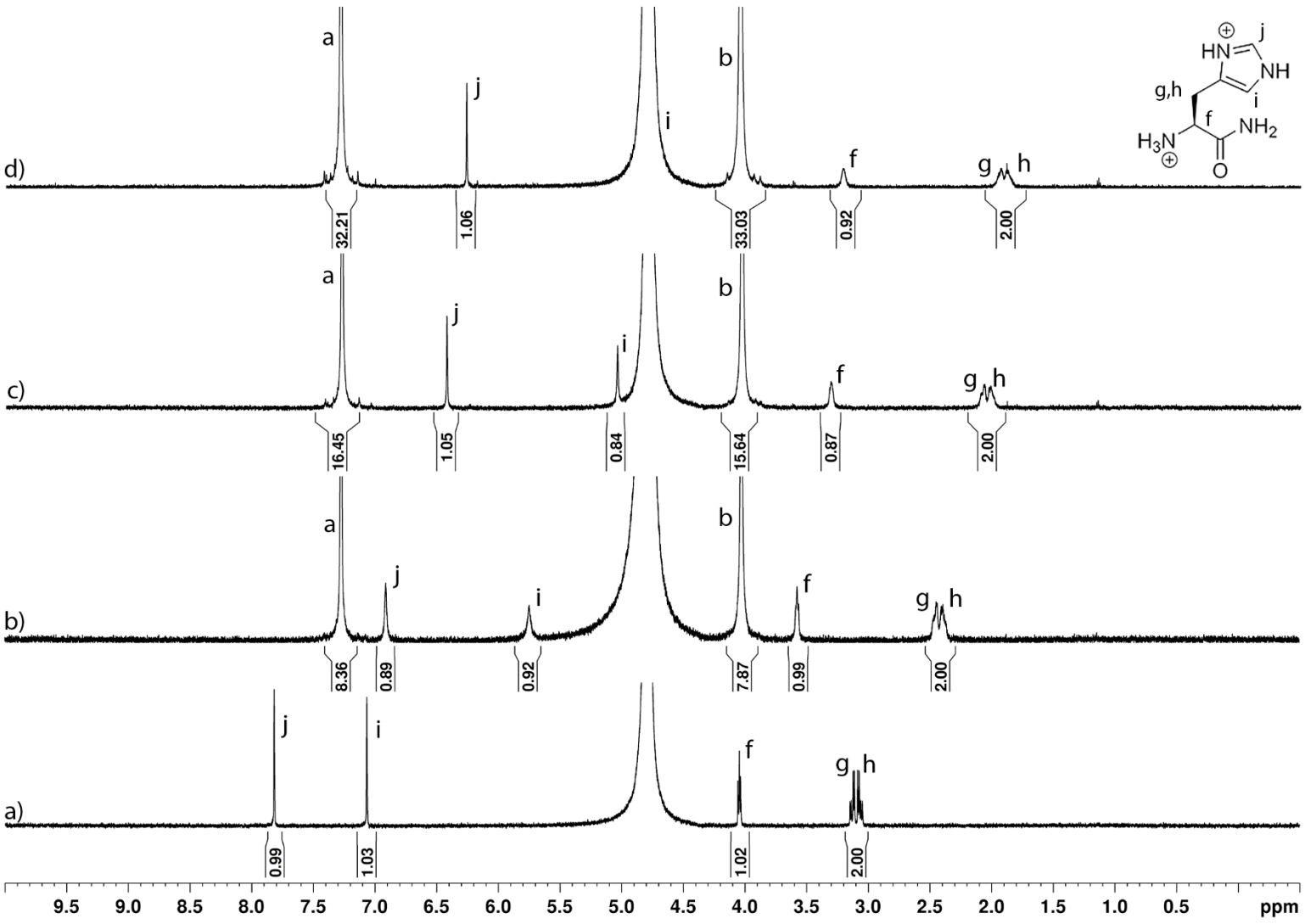

Figure S44. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-H-NH_2 (1.0 mM), b) a mixture of H-H-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-H-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-H-NH_2 (1.0 mM) and P6MQ (2.0 mM).

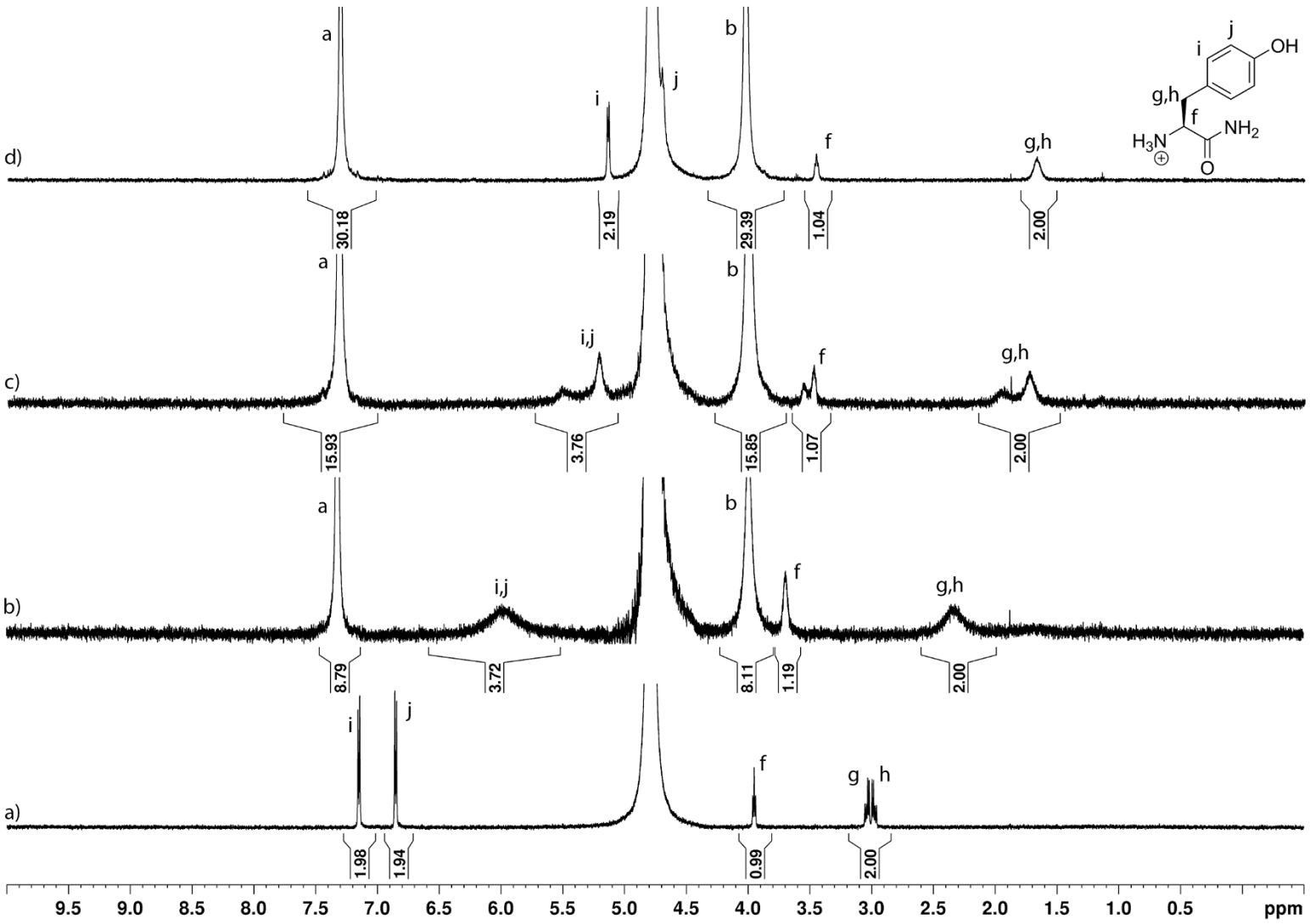

Figure S45. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-Y-NH_2 (1.0 mM), b) a mixture of H-Y-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-Y-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-Y-NH_2 (1.0 mM) and P6MQ (2.0 mM).

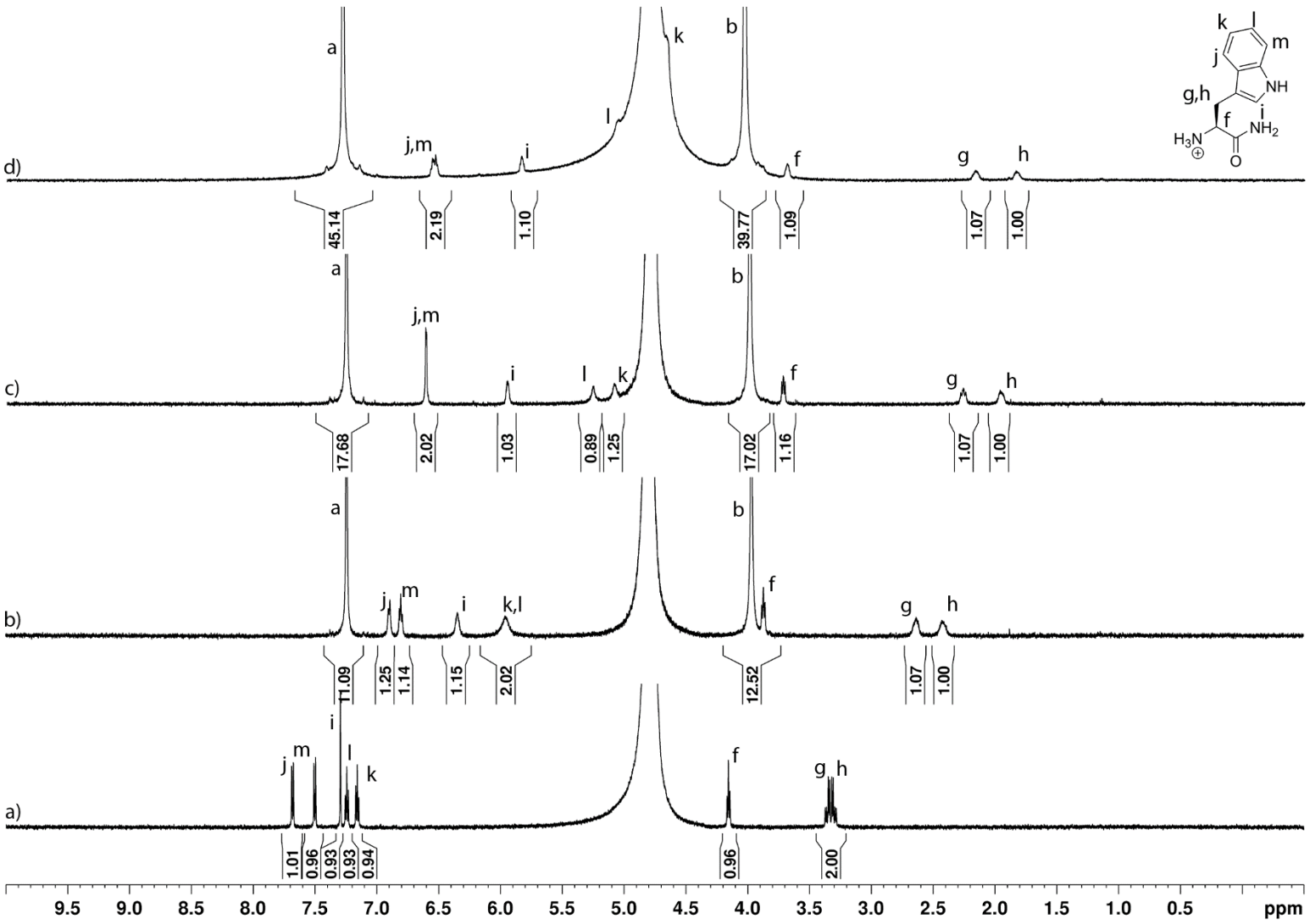

Figure S46. ^{1}H NMR spectra recorded for (600 MHz, RT, phosphate buffered saline D_2O, pD 7.4) for: a) H-W-NH_2 (1.0 mM), b) a mixture of H-W-NH_2 (1.0 mM) and P6MQ (0.5 mM), c) an equimolar mixture of H-W-NH_2 and P6MQ (1.0 mM), and d) a mixture of H-W-NH_2 (1.0 mM) and P6MQ (2.0 mM).

Details of the molecular dynamics simulations

System Setup for Computations. We analyzed a total of eight systems to characterize P6MQ binding. We studied an empty P6MQ host, three free amino acids, H3K4 and H3K4Me$_3$ system. The H3K4 and H3K4Me$_3$ systems are bound in two different configurations, and the different bound configurations are treated as separate systems.

Using the NMR studies as a guide, all systems were initialized in a physically reasonable state. The H3 systems were further minimized from the initial structure using SPARTAN software. The systems were then prepared for further minimization, equilibration, and production simulation via the following procedure.

The systems were parameterized using the AMBER GAFF2 force field, the state-of-the-art force field for parameterizing novel systems. Subsequently, each system was solvated in TIP3P water and neutralized with NaCl counter-ions at a concentration of 0.15 mM to emulate the experimental conditions. Due to the method used to calculate electrostatic interactions during simulation, the system is required to be electrically neutral. The system was prepared for simulation with periodic boundary conditions, and the size of the box is chosen such that the Van der Waals interaction strength is zero between periodic images of the system. Specific values for the simulation box size, and water, and ion counts can be found in Table S1.

The variation in the number of water molecules in Table S1 is due to the Monte-Carlo algorithm used to fill the simulation box. The difference between the number of Na^+ and Cl^- ions is such that the system is electrically neutral and the desired concentration of ions is retained.

Table S1. Values for box size and number of water and ions in each simulation

System	Box Size (Å)	Water Count	Na^+ Count	Cl^- Count
P6MQ	37	4308	16	4
H-K-OH	37	4302	16	4
H-K-NH2	37	4314	14	4
H-K(Me_3)-OH	37	4740	15	4
H3K4 (Lys bound)	42	6501	15	6
H3K4 (Arg bound)	42	6492	15	6
H3K4Me_3 (TMK bound)	42	5985	14	5
H3K4Me_3 (Arg bound)	42	5967	14	5

Simulation Details. Each of the systems were minimized using the steepest descent algorithm, and the equilibrated to 298K over 125ps. Then a 1μs production simulation was performed at 298K. In both the equilibration and production simulations a Nose-Hoover thermostat was used along with a Parrinello-Rahman barostat held at 1bar, placing the simulation under the NPT ensemble. In all cases the GROMACS simulation engine was used to perform simulations. Calculations were performed over recorded trajectories consisting of 100,000 frames each using PLUMED except where noted.

Order Parameters. We examined order parameters which aimed to characterize geometry, electrostatics, and the role of water. All calculations except for the H-bond calculations were performed using PLUMED. The order parameters included

- Bound depth: depth of the ligand in the binding cavity relative to the P6MQ center of mass
- Electrostatic distances: minimum distance between functional groups on the ligand and sulfates on P6MQ
- Na+ Distance: minimum distance between a Na^+ ion and the P6MQ center of mass
- H-bonds: average number of H-bonds formed in the P6MQ binding cavity
- Local Density: radial distribution of water molecules within the cavity

Bound depth.

The bound depth of the ligand is characterized by

$$d_{bound} = D(COM_{LIG}, COM_{OSO3^-,BOT}) - D(COM_{LIG}, COM_{OSO3^-,TOP}), \quad \text{(Eq. S1)}$$

where $COM_{OSO3^-,TOP}$ and $COM_{OSO3^-,BOT}$ refer to the center of mass of sulfates along the top portal and bottom portals respectively, and COM_{LIG} refers to the center of mass of the bound ligand. The function $D(X, Y)$ computes the 2 distance between points X and Y. The orientation of the molecule is such that the top portal is taken to be the one containing the C_α of the bound sidechain. Thus, the quantity defined in Eq. S1 is positive when the center of mass of the bound sidechain is closer to the top portal and negative when the center of mass is closer to the bottom portal. The numerical value of the quantity will be accurate so long as the center of mass is located between and co-linear with the top and bottom portal. In simulations these conditions were observed to hold

approximately for all systems except for $H3K4Me_3$ with arginine bound. When the center of mass of the ligand is located outside the binding cavity the displacements are accurate in a relative sense.

Electrostatic distances.

Functional groups were identified on each ligand and considered to study the electrostatic interaction between the host-guest complexes. These include amine groups, amide groups, and the backbone nitrogen where applicable. The minimum distance between the functional groups and P6MQ sulfates is calculated as

$$d_{FG} = \min D(COM_{FG},\ O_{P6MQ}), \quad \{FG\} \qquad \text{(Eq. S2)}$$

where {FG} represents the set of functional groups for a ligand. The purpose of this order parameter is to quantify electrostatic interactions between the host and bound ligand.

Na^+ Distance. Similarly, the minimum distance between the center of P6MQ and a sodium ion is calculated as:

$$dNa = \min D(Na^+,\ COM_{P6MQ}), \quad \{Na^+\} \qquad \text{(Eq. S3)}$$

where COM_{P6MQ} represents the center of mass of the host P6MQ, and $\{Na^+\}$ represents the set of sodium ions present in the system. The presence of sodium in the binding cavity was observed to play a role in deciding the hydrogen bonding structure of water in the cavity.

H-bonds. We analyzed the number of hydrogen bonds, n_{HB}, each water molecule participates in within a distance r of the P6MQ center of mass as:

$$n_{HB}(r) = 2 \times \text{Hbonds}(H_2O_{int}, H_2O_{int}) + \text{Hbonds}(H_2O_{int}, H_2O_{ext}) +$$

$$\text{Hbonds}(H_2O_{int}, \text{P6MQ}) + \text{Hbonds}(H_2O_{int, \text{LIG}}), \quad \text{(Eq. S4)}$$

where the function Hbonds(X, Y) counts the number of hydrogen bonds formed between the groups X and Y. In Eq. S4 P6MQ and LIG represent the atoms of the host and bound ligand which are able to accept and donate hydrogen bonds (N, O, H), and H_2O_{int} and H_2O_{ext} respectively refer to water molecules within and outside of the region of interest defined by r. The number of hydrogen bonds between groups was calculated using GROMACS' built-in hbond tool. Importantly, the number of hydrogen bonds formed between water molecules within the region of interest must be counted twice since two water molecules participate in a hydrogen bond.

Local Density. We also computed local density plots derived from radial distribution functions which aim to quantify the structure of water within the binding cavity. Since the radial distribution function $g(r)$ is related to the local density $\rho(r)$ through $\rho(r) = \rho_{bulk} g(r)$, we may directly interpret $g(r)$ as the local density $\rho(r)$ because $\rho_{bulk} = 1$ for water.

Free Energy Calculation: The free energy for a collective variable $\boldsymbol{z}$ is defined as

$$F(\boldsymbol{z}) = -k_B T \log p(\boldsymbol{z}), \quad \text{(Eq. S5)}$$

where $p(\boldsymbol{z})$ is the Boltzmann probability of observing $\boldsymbol{z}$.

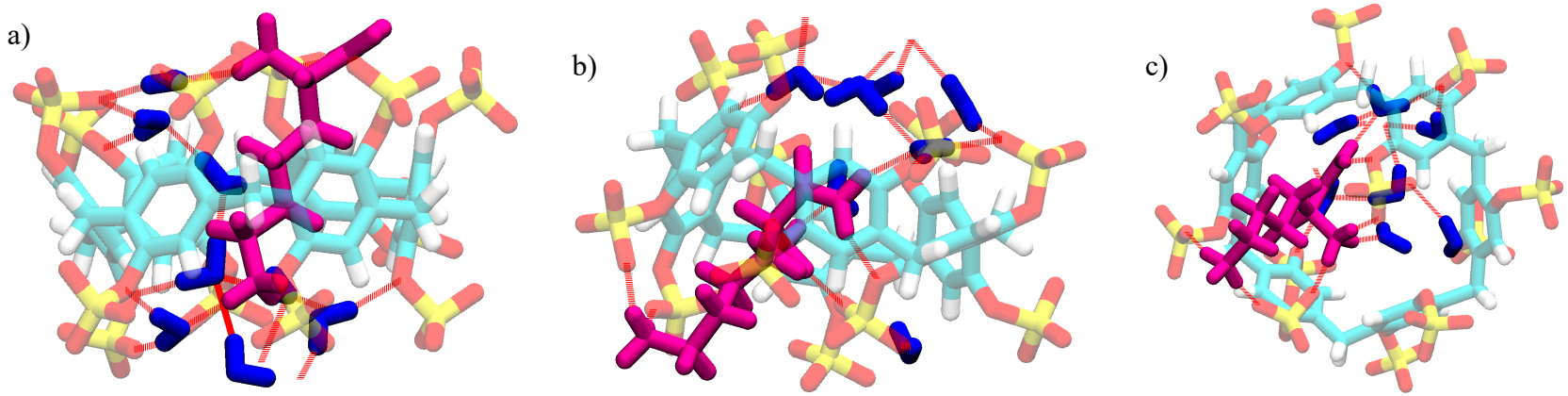

Figure S47. Representative structures of P6MQ•H-K-NH_2. In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) Side view of H-K-NH_2 interacting with the upper portal of P6MQ. b) Side view of H-K-NH_2 interacting with the lower portal of P6MQ. c) View of H-K-NH_2 interacting with the lower portal of P6MQ. H-K-NH_2 is capable of forming multiple H-bonds with P6MQ sulfate groups and is permissive of water in the cavity. In simulations the conformations observed in a) and b) were observed with approximately equal frequency.

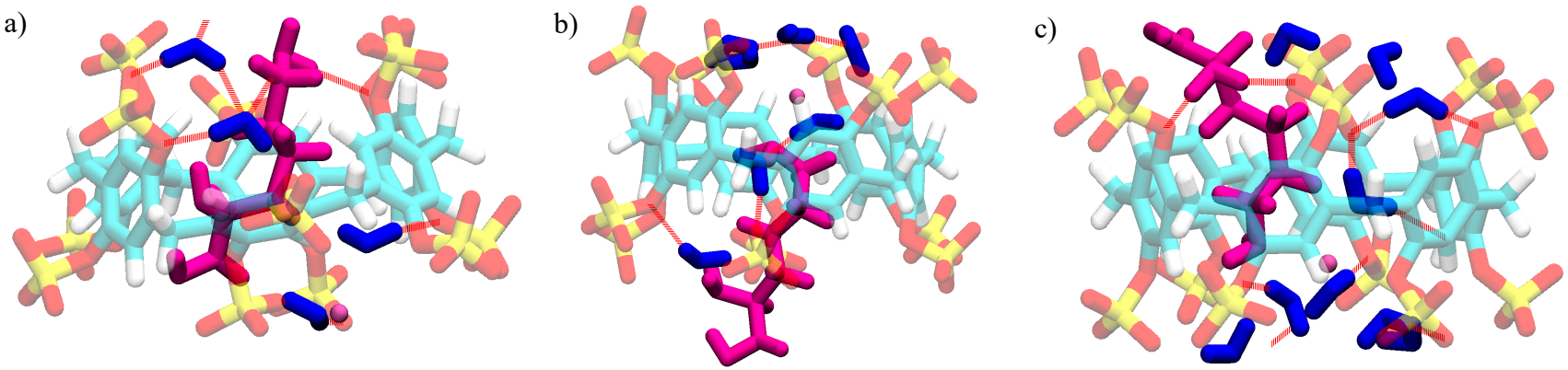

Figure S48. Representative structures of P6MQ•H-K-OH. In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) Side view of H-K-OH bound to the upper portal of P6MQ. c) Side view of H-K-OH bound to the lower portal of P6MQ. c) Side view illustrating that H-K-OH is permissive of water and sodium in the binding cavity. H-K-OH often binds to one side of the binding cavity allowing water and ions to enter the cavity. The states depicted in a) and b) were observed with approximately equal frequency during simulation.

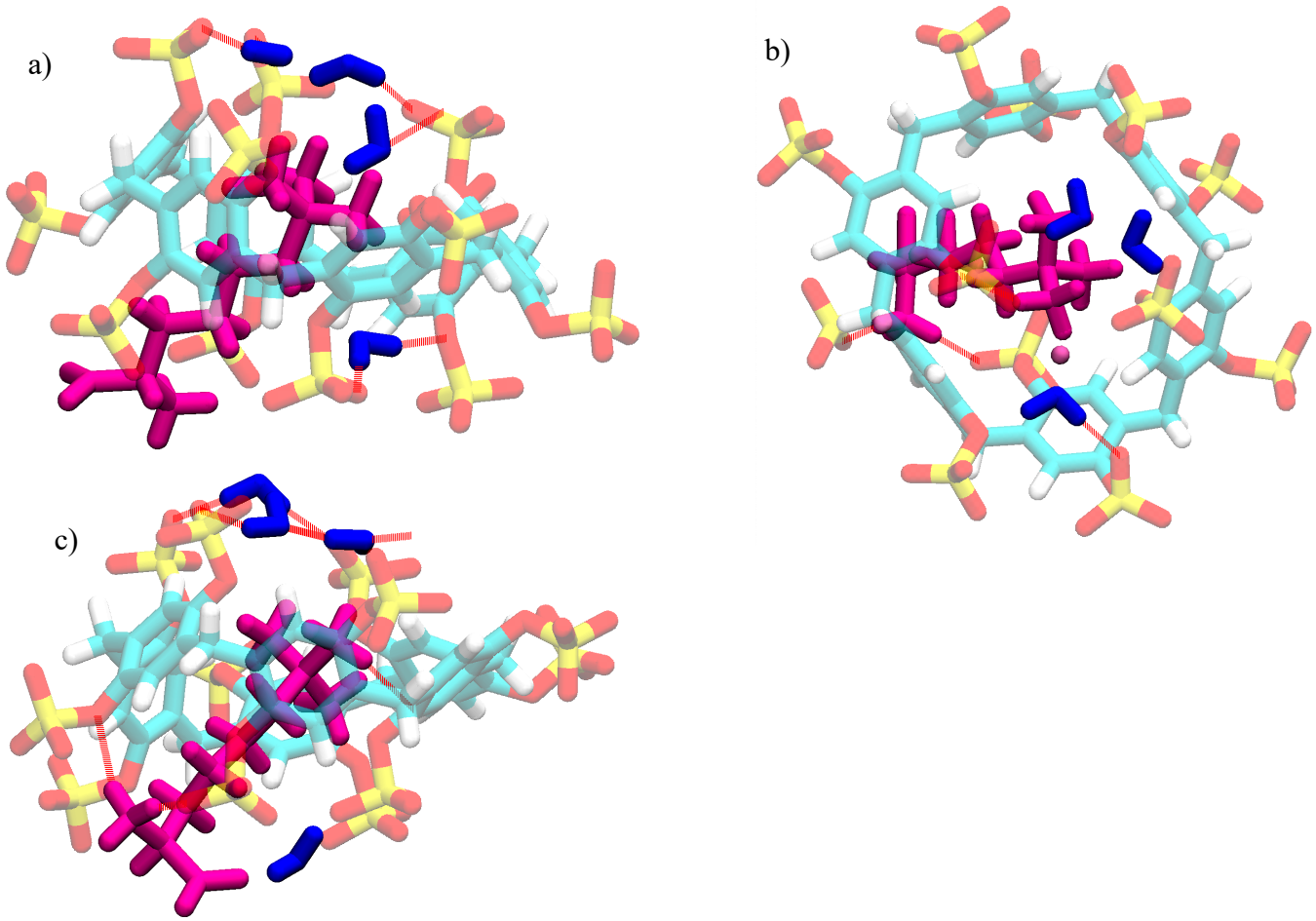

Figure S49. Representative structures of P6MQ•H-K(Me_3)-OH. In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) Side view of H-K(Me_3)-OH bound to the lower portal of P6MQ. b) Top view of H-K(Me_3)-OH bound to the lower portal of H-K(Me_3)-OH. The ligand is bound at an angle in the cavity. c) Side view of H-K(Me_3)-OH bound to the lower portal of P6MQ highlighting the angle at which the ligand is bound, as well as the resulting deformation of the P6MQ host. The NMe_3 group largely excludes water from the cavity (c.f. H-K-NH_2 and H-K-OH). The conformation depicted in (a-c) was the only conformation observed in simulation.

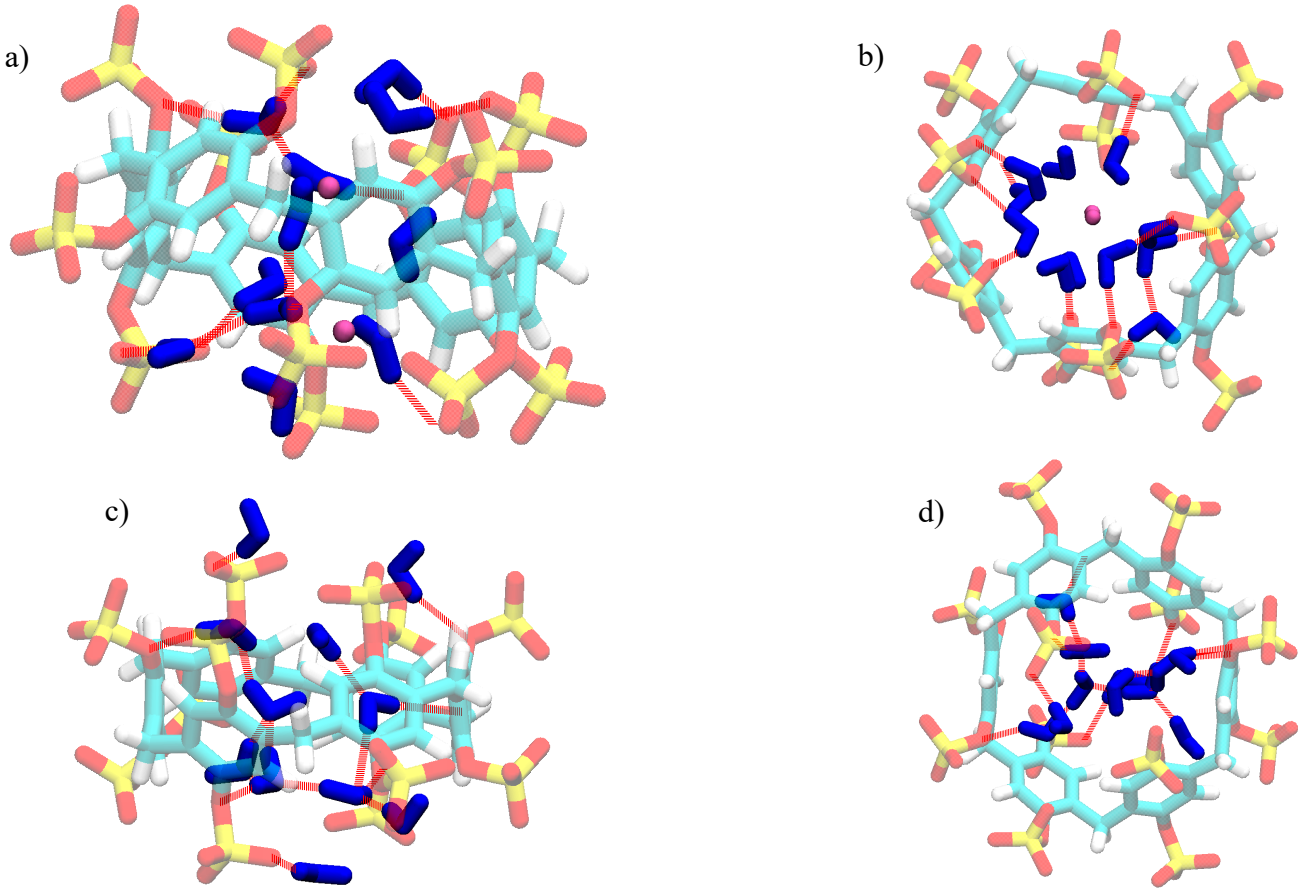

Figure S50. Representative structures of unbound P6MQ. In all panes water molecules are depicted in blue, sodium ions in pink, and H-bonds in red. a) Side view of unbound P6MQ with two sodium ions in the binding cavity. b) Top view of sodium ions in the unbound P6MQ cavity. Observe how the sodium ions coordinate the water oxygens so that H-bonds are mainly formed between water hydrogens and P6MQ sulfate oxygens. c) Side view of the unbound P6MQ cavity in the absence of sodium ions. d) Top view of the unbound P6MQ cavity in the absence of sodium. Note that more hydrogen bonds are formed between waters when sodium ions are absent (c-d) compared to when sodium is present (a-b). The resulting network of hydrogen bonds is more connected in panels (c-d). The unbound state where sodium ions are absent is observed exceedingly rarely in simulation.

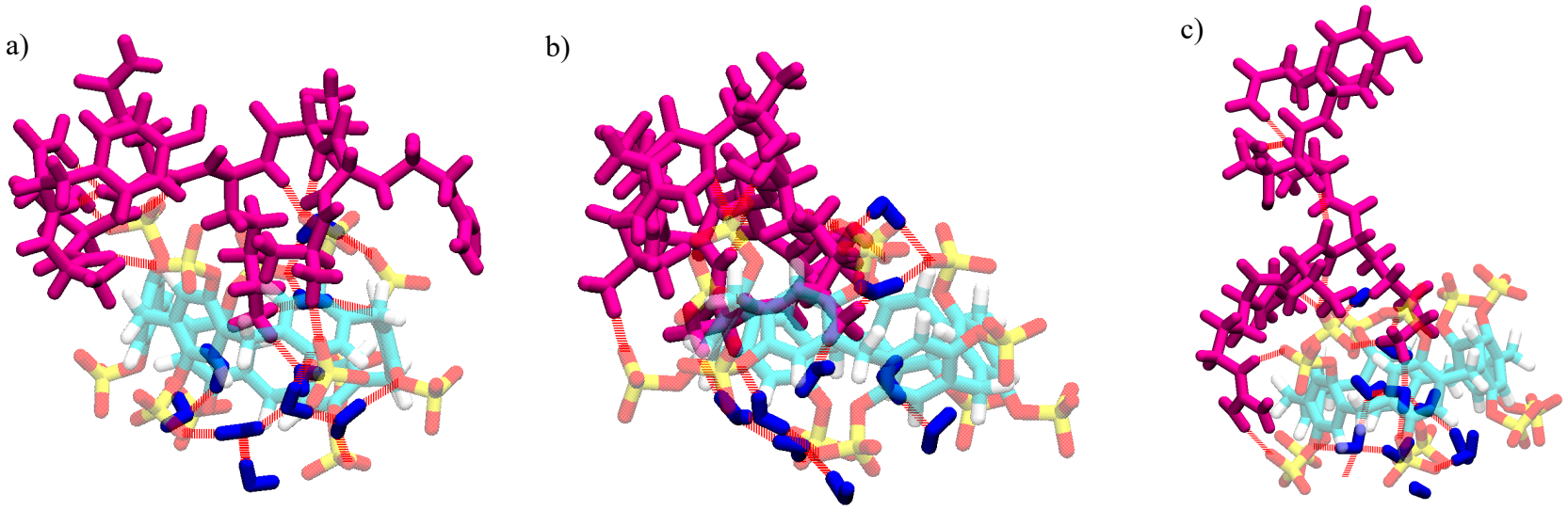

Figure S51. Representative structures of P6MQ•H3K4 (Lys bound). In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) The first of two metastable states observed in simulation. The backbone and unbound sidechains participate in H-bonds with P6MQ. b) Alternative view of metastable state depicted in (a). Note the deformation of P6MQ due to the H-bond interactions between H3K3 and P6MQ. c) The second metastable state is depicted. Here the half of the ligand containing the tyrosine ring is not interacting with P6MQ and is stabilized by H-bonds between the unbound sidechains and the backbone. Here the arginine sidechain is within π-stacking distance of the P6MQ aromatic walls. H-bonding between the arginine sidechain and P6MQ further stabilizes the π-stacking interaction.

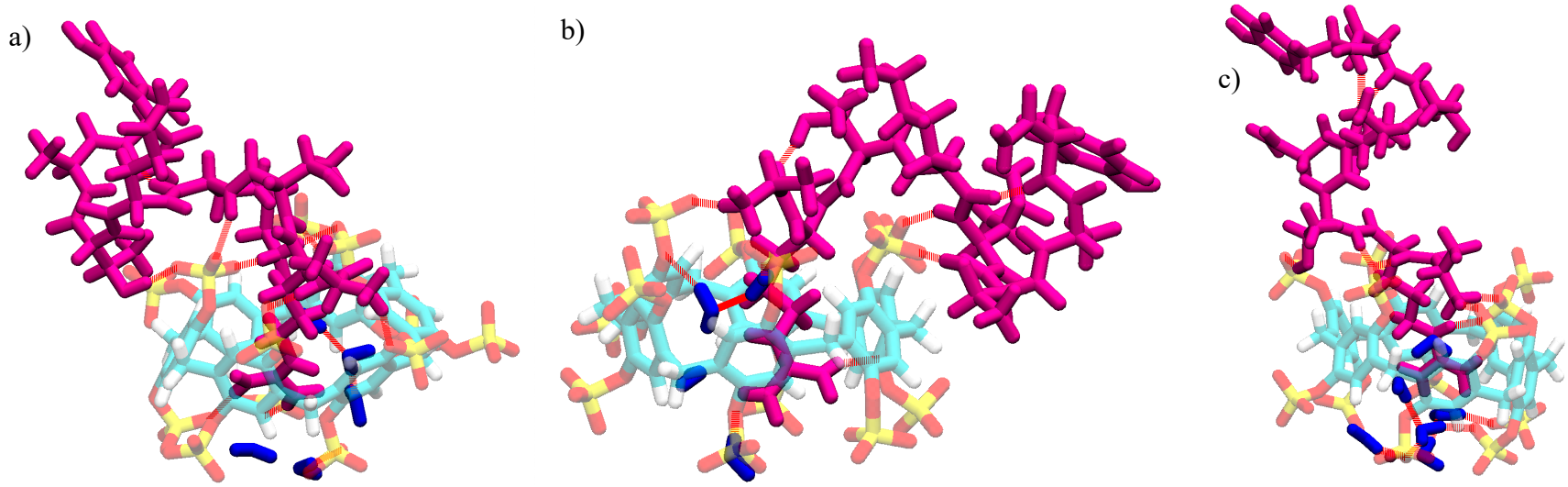

Figure S52. Representative structures of P6MQ•H3K4 (Arg bound). In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) One of two metastable states observed in simulation. The bound arginine sidechain is H-bonded to the lower portal of P6MQ and the unbound sidechains and backbone for H-bonds with the upper and lower portals. b) Alternative view of the first metastable state. c) Second metastable state observed in simulation. Here the unbound sidechains and residues are stabilized away from P6MQ from H-bonds between the sidechains and backbone. H-bonding may still occur between the unbound sidechains and backbone and P6MQ.

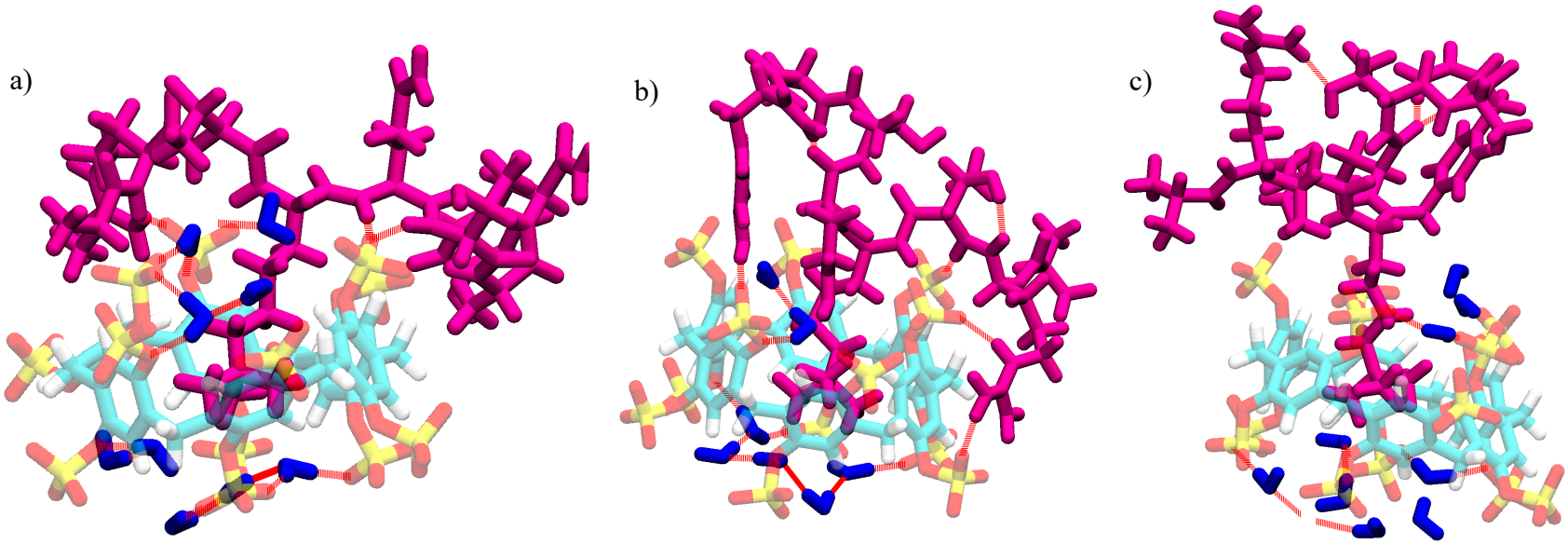

Figure S53. Representative structures of P6MQ•H3K4Me$_3$ (KMe$_3$ bound). In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) The first of two metastable states observed in simulation. The backbone and unbound sidechains engage in H-bonding with P6MQ. b) Alternative view of the first metastable state where arginine is within π-stacking distance of the P6MQ aromatic walls. Arginine is observed within π-stacking distance more rarely when KMe$_3$ is bound compared to lysine. c) The second metastable state observed. The unbound portion of the molecule is stabilized by H-bonds between the sidechains and backbone. H-bond interactions are still present between the unbound sidechains and the top portal of P6MQ.

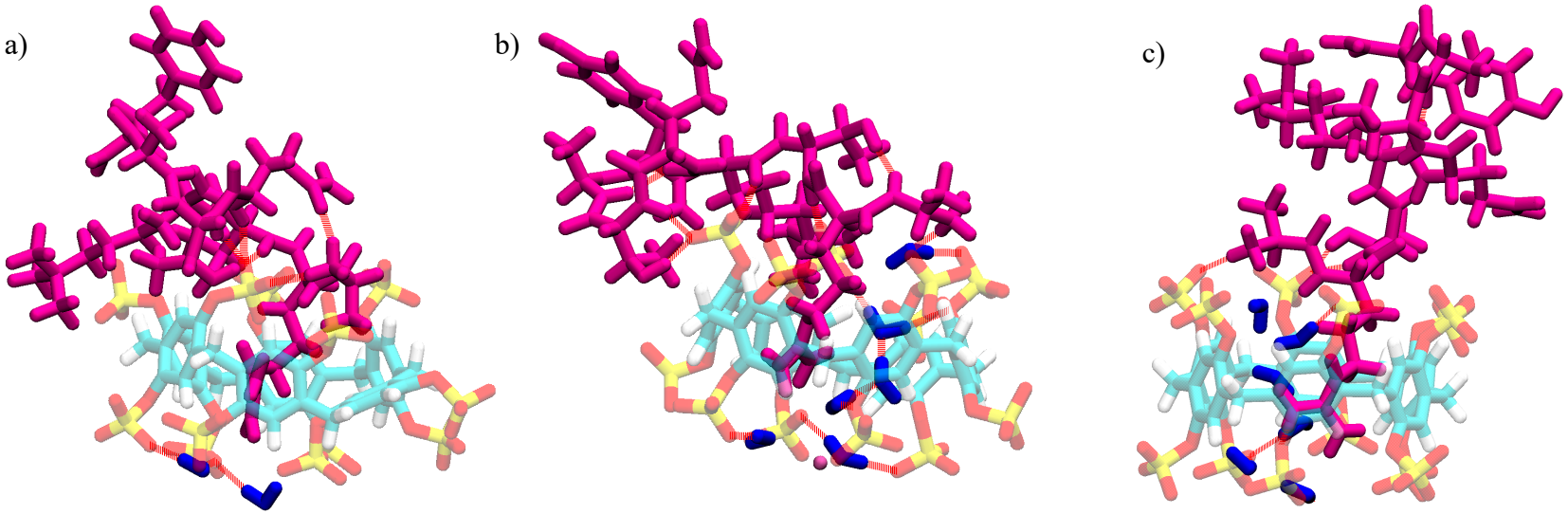

Figure S54. Representative structures of P6MQ•H3K4Me$_3$ (Arg bound) In all panes water molecules are depicted in blue, the ligand in magenta, and sodium ions in pink, and H-bonds in red. a) The first of two metastable states observed in simulation. The backbone and unbound sidechains engage in H-bonding with P6MQ. Here the deformation of P6MQ due to the ligand is evident. b) Alternative view of the first metastable state. c) The second metastable state observed. Here the unbound portion of the molecule is stabilized by H-bonds between the sidechains and backbone. H-bond interactions are still present between the unbound sidechains and the top portal of P6MQ. It is clear from (a-c) that arginine is bound deeper within the cavity compared to Lys/KMe$_3$ for H3K4 and H3K4Me$_3$.

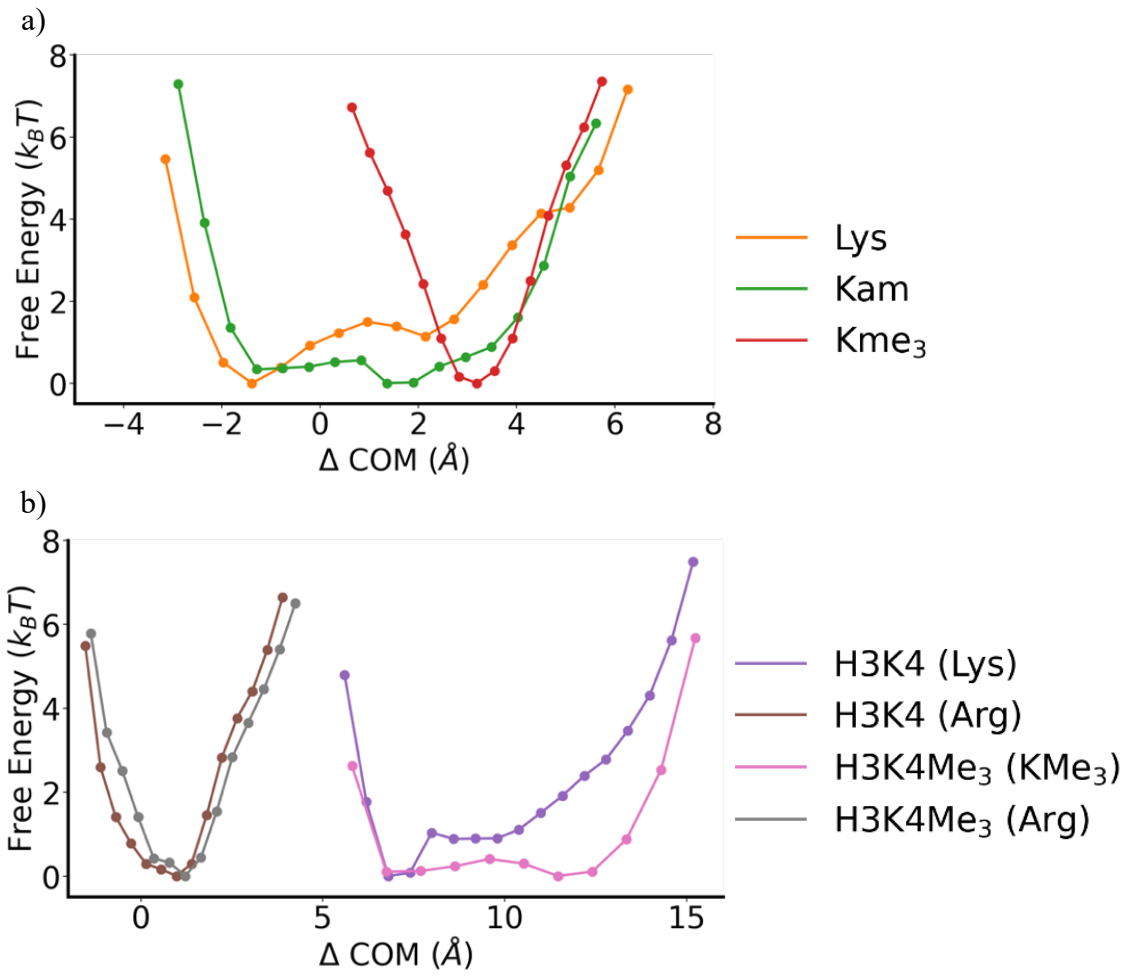

Figure S55. Bound depth of P6MQ complexes. a) Bound depth for amino acids (Lys = H-K-OH; Kam = H-K-NH_2; KMe_3 = H-K(Me_3)-OH). b) Bound depth for H3K4 and H3K4Me_3. Note that the bound depth calculation is only accurate when the center of mass of the bound ligand is between the P6MQ portals. For H3K4 (Lys) and H3K4Me_3 (KMe_3) bound the center of mass is outside of the portals, so the distances are accurate only in a relative sense. That is, arginine is bound deeper in the P6MQ cavity compared to lysine and KMe_3.

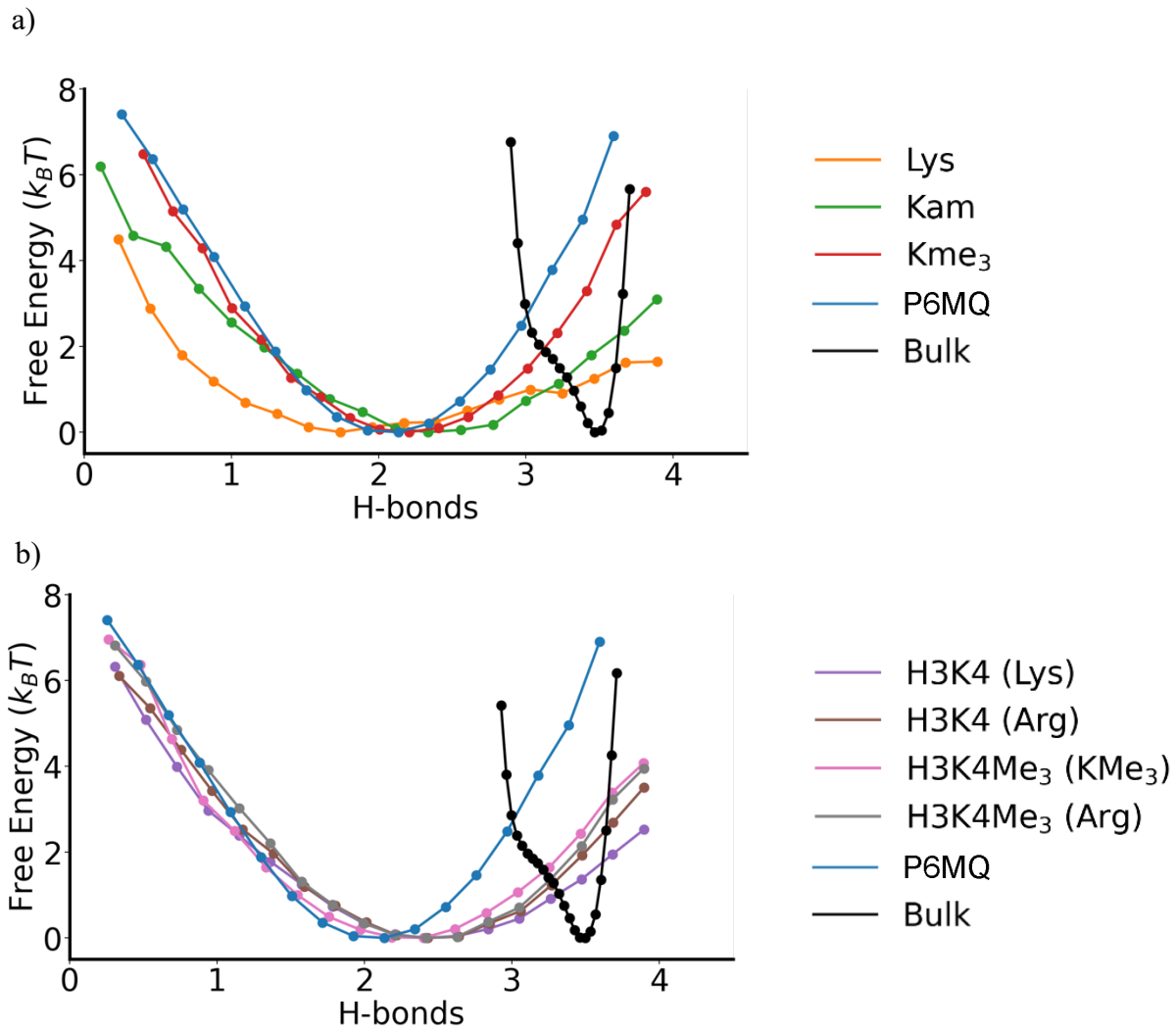

Figure S56. Number of H-bonds within P6MQ cavity. a) The average number of H-bonds formed per water molecule within 6 Angstroms of the P6MQ center of mass is depicted for amino acids bound to P6MQ, unbound P6MQ, and a bulk-water reference. (Lys = H-K-OH; Kam = H-K-NH_2; KMe_3 = H-K(Me_3)-OH). b) The number of H-bonds is depicted for P6MQ•H3K4 complexes, unbound P6MQ, and a bulk-water reference.

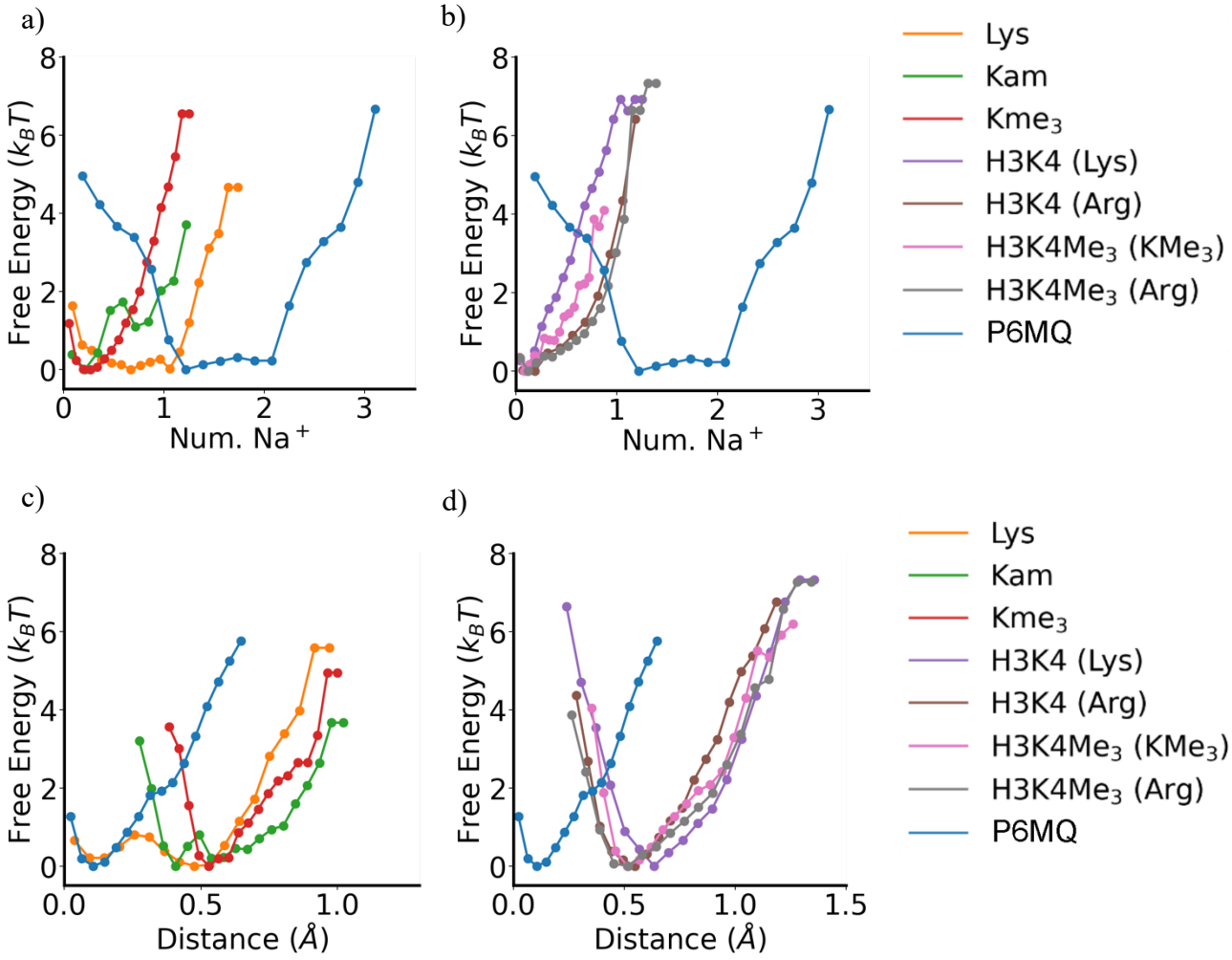

Figure S57. Quantifying Na^+ presence in P6MQ. a) Number of Na^+ ions present within 4 Angstroms of the P6MQ center of mass for amino acid complexes and unbound P6MQ. (Lys = H-K-OH; Kam = H-K-NH_2; KMe_3 = H-K(Me_3)-OH). b) Number of Na^+ ions present within 4 Angstroms of the P6MQ center of mass for H3 complexes and unbound P6MQ. c) Minimum distance between any Na^+ ion and the P6MQ center of mass for amino acid complexes and unbound P6MQ. d) Minimum distance between any Na^+ ion and the P6MQ center of mass for H3 complexes and unbound P6MQ.

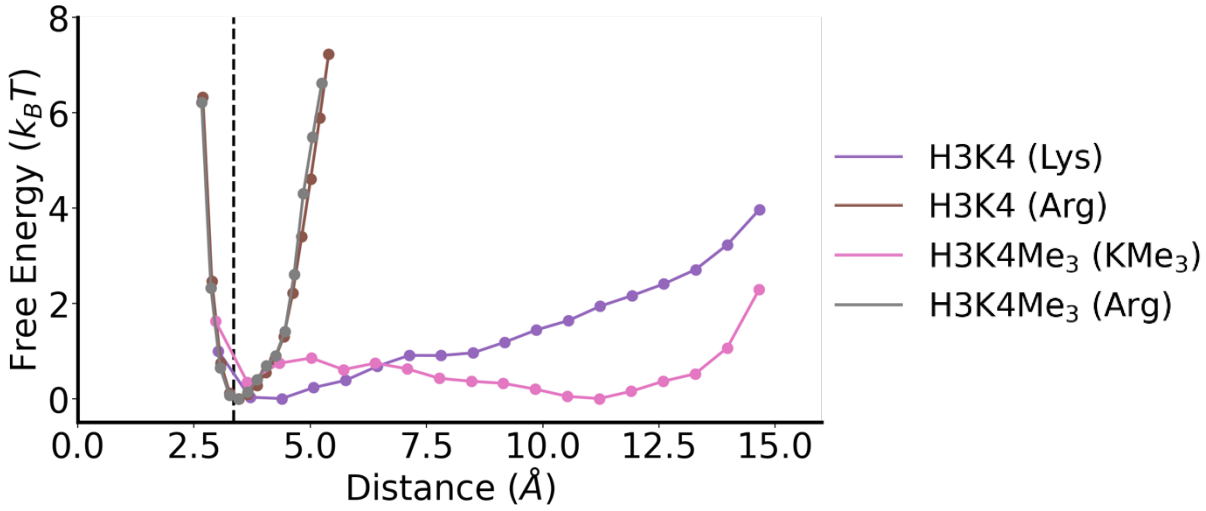

Figure S58. Arginine π-stacking in H3K4 complexes. The minimum distance between the arginine π-bond and the center of mass of the P6MQ aromatic walls is shown. The characteristic distance of π-stacking interactions (3.35 Angstroms) is depicted by the dashed line.

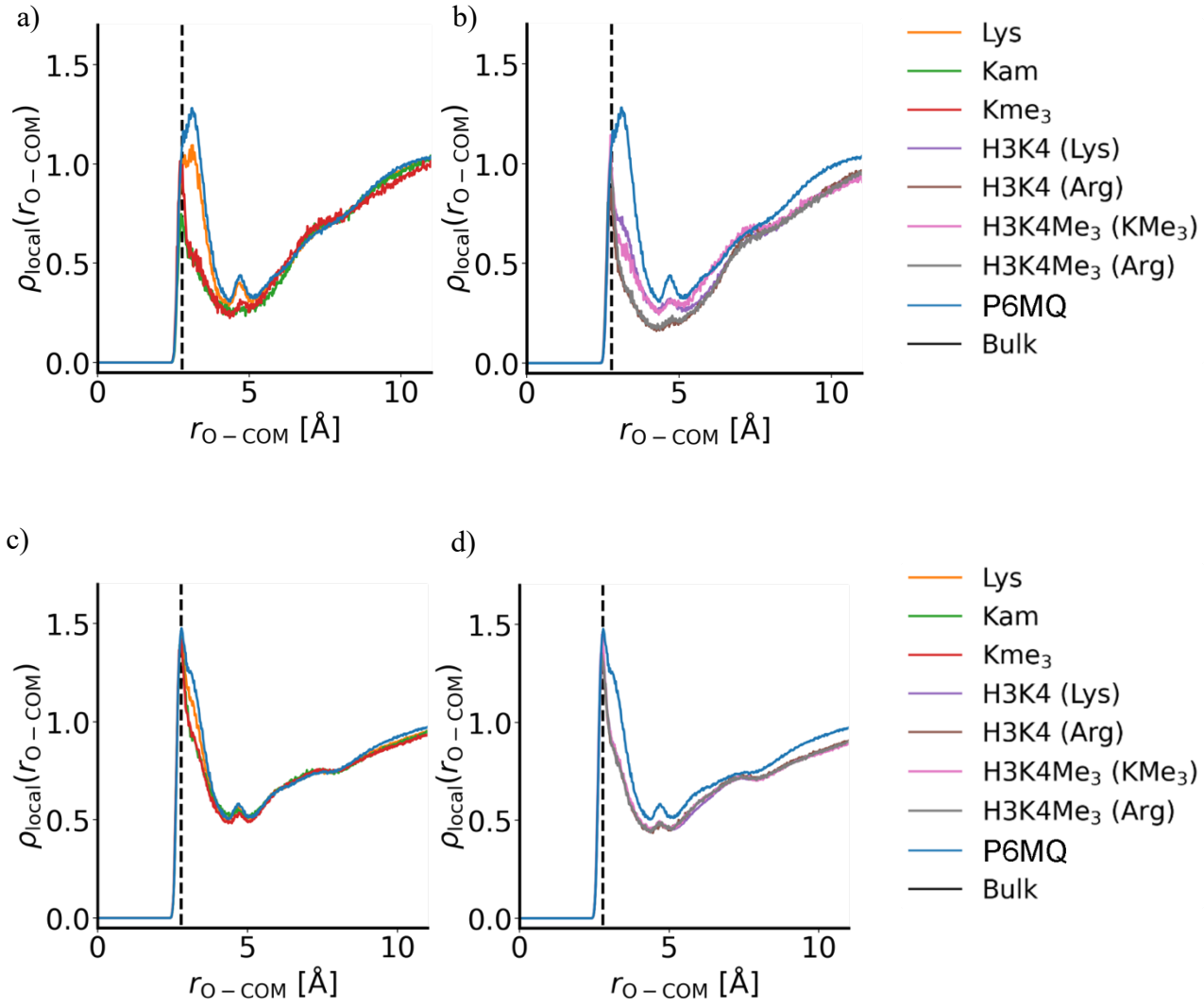

Figure S59. Local density of water measured from P6MQ center of mass**.** a) Local density of water oxygens within 4 Angstroms of the P6MQ center of mass for amino acid complexes. (Lys = H-K-OH; Kam = H-K-NH_2; KMe_3 = H-K(Me_3)-OH). b) Local density of water oxygens within 4 Angstroms of the P6MQ center of mass for H3K4 complexes. c) Local density of water oxygens between 4 and 7 Angstroms of the P6MQ center of mass for amino acid complexes. d) Local density of water oxygens between 4 and 7 Angstroms of the P6MQ center of mass for H3K4 complexes. The black dashed line represents the dominant peak for bulk water at 2.78 Angstroms, and unbound P6MQ is included as a reference for all panes.

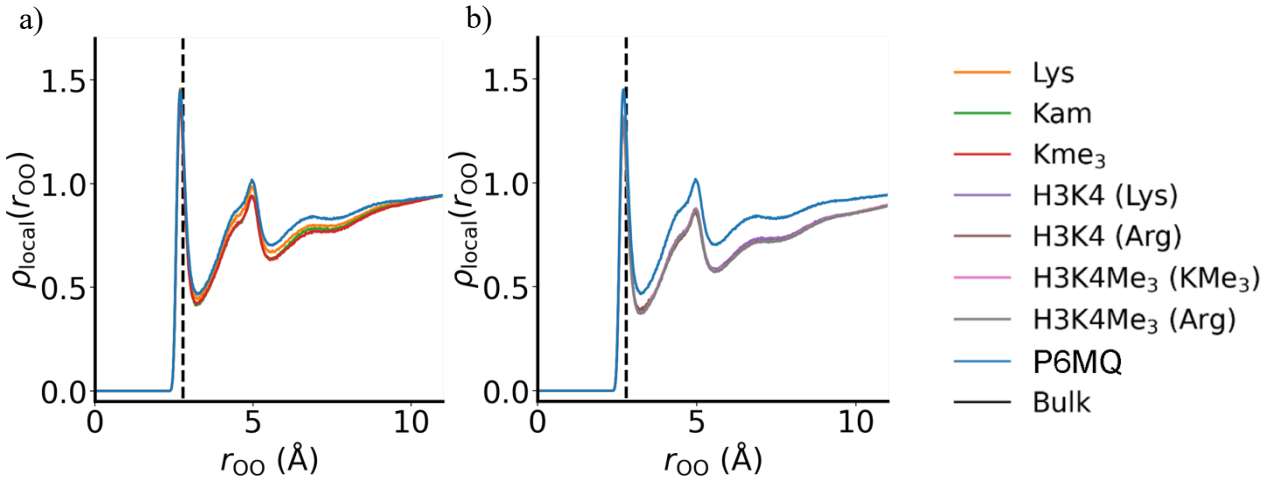

Figure S60. Local density of water measured from P6MQ sulfate oxygens. a) The local density of water is measured for oxygens within 7 Angstroms of a P6MQ sulfate oxygen for amino acid complexes. (Lys = H-K-OH; Kam = H-K-NH_2; KMe_3 = H-K(Me_3)-OH). b) The local density of water is measured for oxygens within 7 Angstroms of a P6MQ sulfate oxygen for H3K4 complexes. The black dashed line represents the dominant peak for bulk water at 2.78 Angstroms, and unbound P6MQ is included as a reference for both panes.