Electronic Supplementary Material (ESI) for ChemComm.

Supporting information for

A Squaramide Cage Capable of Binding and Extracting $H_2PO_4^-$ and $HP_2O_7^{3-}$ in Highly Polar Protic Media

Ju Ho Yang and Sung Kuk Kim*

Department of Chemistry and Research Institute of Natural Science, Gyeongsang National University, Jinju, 52828, Korea,

1. General experimental and synthetic details S1 – S3

2. ^{1}H NMR spectroscopic data for anion binding studies S4 – S13

3. NMR and HRMS Spectra for characterization S14 – S19

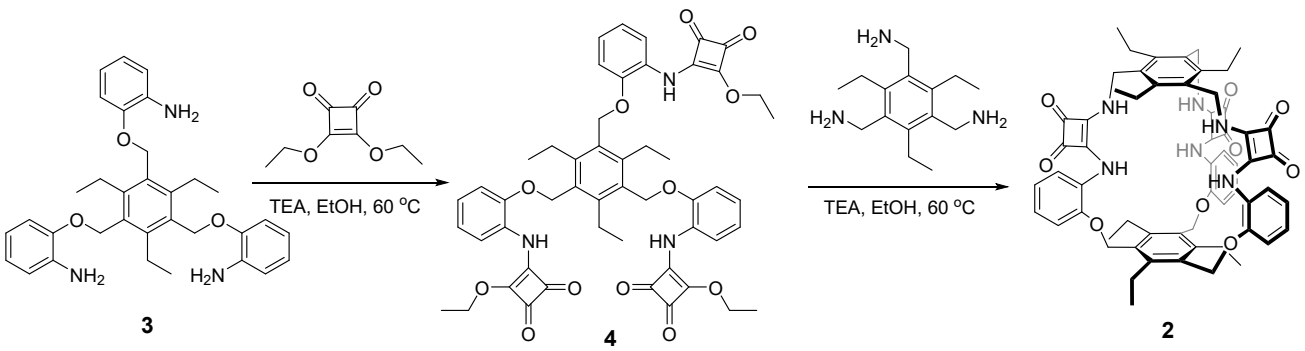
4. Single crystal X-ray diffraction analysis for **2**·$H_2PO_4^-$ S20 – S22

5. References S23

1. General experimental and synthetic details

Solvents and reagents used for the synthetic work were purchased from Aldrich, TCI, or Alfa Aesar and used without further purification. Compound **2** was prepared as reported previously.[1] NMR spectra were recorded on a Bruker Advance-300 MHz instrument. The NMR spectra were referenced to residual solvent peaks and the spectroscopic solvents were purchased from either Cambridge Isotope Laboratories or Aldrich. Fast atom bombardment (FAB) mass spectra (MS) were recorded on a JMS-700 (JEOL) spectrometer. TLC analyses were carried out using Sorbent Technologies silica gel (200 mm) sheets. Column chromatography was performed on Sorbent silica gel 60 (40–63 mm).

Scheme S1. Synthetic scheme for cage **2**



Compound 4

A reaction mixture of compound **3** (0.5g, 0.95 mmol) and diethyl squarate (0.81g, 4.76 mmol) in EtOH (50 mL) was stirred for 3 days at 60 ºC in the presence of 6 equiv of triethylamine (0.8ml). After the reaction removed solvent using reduced pressure evaporation. After the reaction was completed, the volatile solvent and reagent were removed by evaporation *in vacuo*. To the resulting crude reaction mixture, ethyl acetate was added, and the organic layer was washed twice with water. The organic layer was separated off and then dried over anhydrous $MgSO_4$. The organic solvent was evaporated *in vacuo* to afford a brownish solid. The resulting crude mixture was purified by column chromatography over silica gel using an eluent of CH_2Cl_2/EA (9:1) to give 0.41g (48% yield) of **4** as an ivory yellow solid. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.85 (s, 3H), 7.61 (d, J = 8.0 Hz 3H), 7.25 – 7.15 (m, 6H), 7.13 – 7.00 (m, 3H), 5.20 (s, 6H), 4.86 (q, J = 7.1 Hz 6H), 2.83 (q, J = 7.4 Hz 6H), 1.48 (t, J = 7.1 Hz 9H), 1.25 (t, J = 7.4 Hz 9H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 183.4, 147.3, 145.9, 129.4, 125.7, 124.4, 120.7, 118.6, 111.0, 76.2, 69.2, 64.2, 22.0, 15.7, 14.8 ppm. HR FAB-MS m/z 898.3473 $[M]^+$ calc. for $C_{51}H_{51}N_3O_{12}$, found 898.3561.

Compound 1

A reaction mixture of compound **4** (0.21g, 0.23 mmol) and 1,3,5-tris(aminomethyl)-2,4,6-triethylbenzene (0.06g, 0.23 mmol) in EtOH (40 mL) was stirred for 3 days at 60 ºC in the presence of 6 equiv of triethylamine (0.2ml). After the reaction was completed, the reaction solution was concentrated under reduced pressure. To the resulting crude product, dichloromethane and water were added. The organic layer was separated off, washed twice with water. Then, the organic layer

was dried over anhydrous $MgSO_4$ and evaporated in vacuo to give a brownish solid. The crude product was purified by column chromatography over silica gel (eluent: CH_2Cl_2/MeOH, 9:1) to give 0.06g (25% yield) of cage molecule **1** as an ivory yellow solid. ^{1}H NMR (300 MHz, DMSO-d_6) δ 9.15 (s, 3H), 8.01 (s, 3H), 7.47 (d, J = 8.6 Hz 3H), 7.41 (d, J = 7.9 Hz 3H), 7.16 (t, J = 7.8 Hz 3H), 7.00 (t, J = 7.6 Hz 3H), 5.06 (s, 6H), 4.89 (s, 6H), 2.66 (s, 12H), 7.16 (q, J = 7.5 Hz 18H). ^{13}C NMR (75 MHz, DMSO-d_6) δ 185.0, 180.4, 168.5, 164.1, 150.1, 147.5, 143.3, 132.85, 130.7, 126.7, 125.3, 123.0, 121.1, 112.3, 65.2, 41.4, 23.6, 23.4, 16.8, 16.4 ppm. HR FAB-MS m/z 1009.4422 $[M]^+$ calc. for $C_{60}H_{60}N_6O_9$, found 1009.4515.

2. ^{1}H NMR spectral data

a) Treatment used where binding between receptors and anions is governed by a binding/release equilibrium that is slow on the NMR time scale:

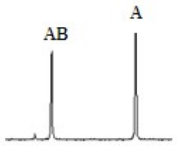
Equilibrium: $A + B \underset{}{\overset{K_a}{\rightleftharpoons}} AB$

Equilibrium constant: $K_a = \dfrac{[AB]}{[A][B]} = \dfrac{[AB]}{(c(A) - [AB])(c(B) - [AB])}$ (1)

c(A) and c(B) are the initial concentrations of A and B, and [A], [B] and [AB] are the equilibrium concentrations of the three species.

A and B is in slow exchange with the complex AB on the ^{1}H NMR time scale.

Two signals for one specific proton on A can be seen in the spectrum, corresponding to complexed and uncomplexed forms of A:



Single-point Methods

K_a is determined from the integrals of complexed and uncomplexed A. If I(A) denotes the integral of a signal for one specific proton of A and I(AB) the integral for the same proton in the complex, the concentration of AB at equilibrium is shown by eq 2. The equilibrium expression is obtrained after substituting into eq. (1):

$$[AB] = \frac{I(AB)}{I(A) + I(AB)} c(A) \quad (2)$$

$$K_a = \frac{\dfrac{I(AB)}{I(A) + I(AB)} c(A)}{I(A)\left(c(B) - \dfrac{I(AB)}{I(A) + I(AB)} c(A)\right)} \quad (3)$$

b) Treatment used where binding between receptors and anions is governed by a binding/release equilibrium that is fast on the NMR time scale:

Binding constants of receptors with anions were calculated using the equation $y = (b \times x)/(1 + x \times Ka)$, where x = [G], y = $\left| \delta_0 - \delta \right|$ (δ is the chemical shift of an indicative

receptor proton signal at a certain anion concentration and δ_0 is the chemical shift of the receptor signal for the anion-free form).

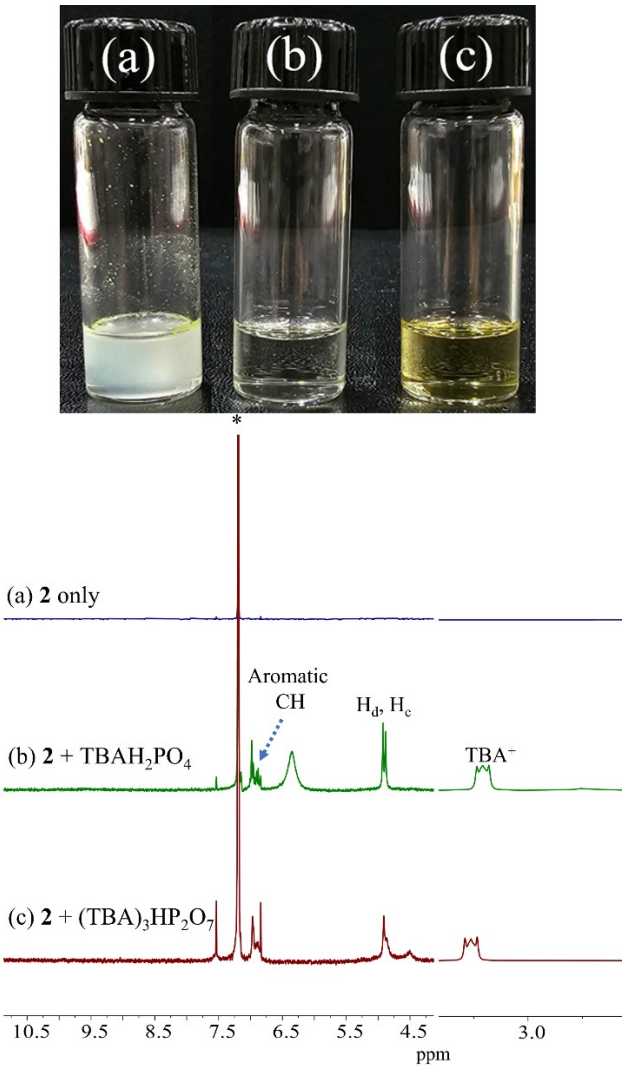


Figure S1. Top: A photo of chloroform solutions containing (a) **2** only, (b) **2** + $TBAH_2PO_4$ (10.0 equiv.), and (c) **2** + $(TBA)_3HP_2O_7$ (10.0 equiv.), respectively. Bottom: Partial ^1H NMR spectra of receptor **2** before and after the addition of 10 equiv of $TBAH_2PO_4$ (tetrabutylammonium dihydrogen phosphate) and 10 equiv of $(TBA)_3HP_2O_7$ (tris(tetrabutylammonium)hydrogen pyrophosphate) in $CDCl_3$. The asterisk denotes a residual NMR solvent peak.

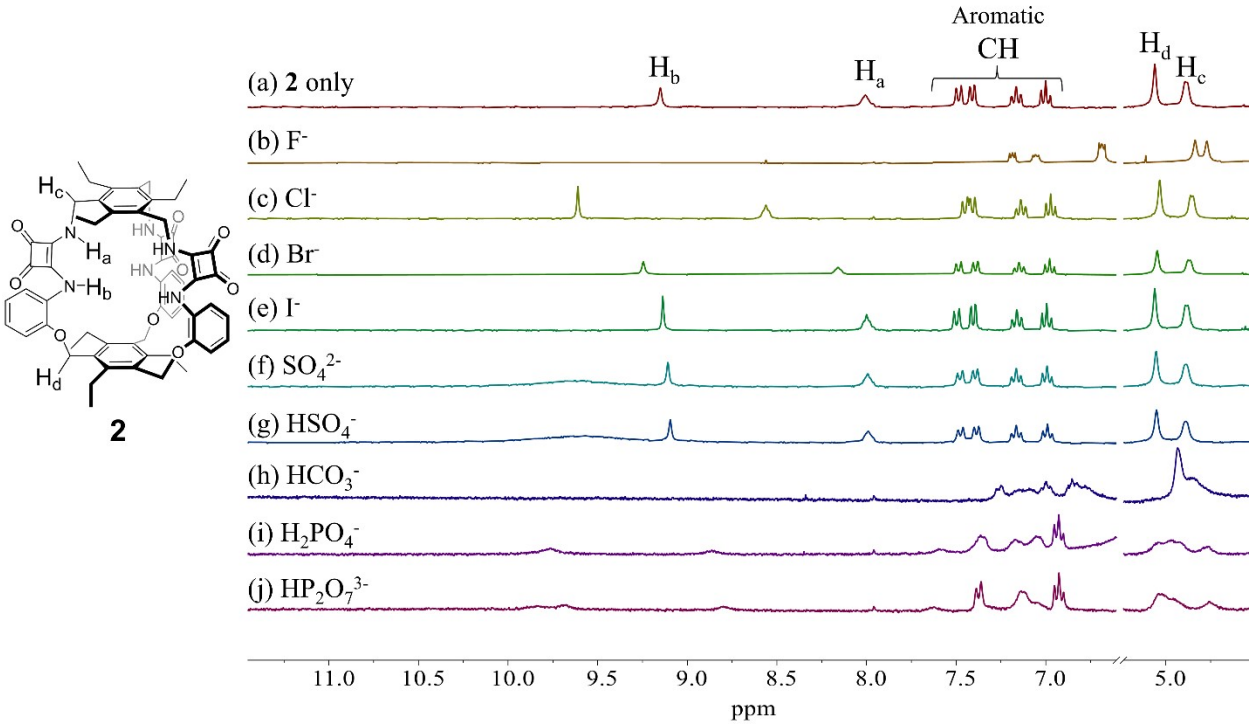


Figure S2. Partial views of ^{1}H NMR spectra of cage **2** (3 mM) recorded in DMSO-d_6 in the presence of the indicated anions as their respective TBA^+ salts except for HCO_3^- in the TEA^+ salt form.

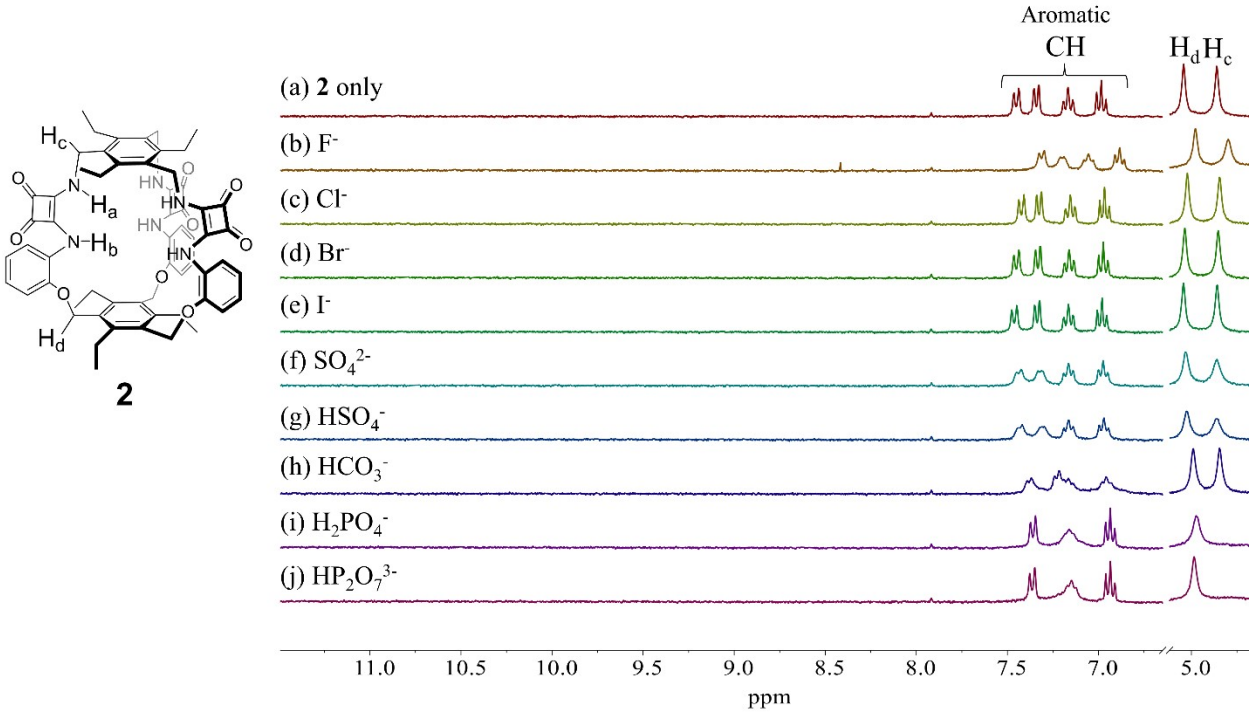
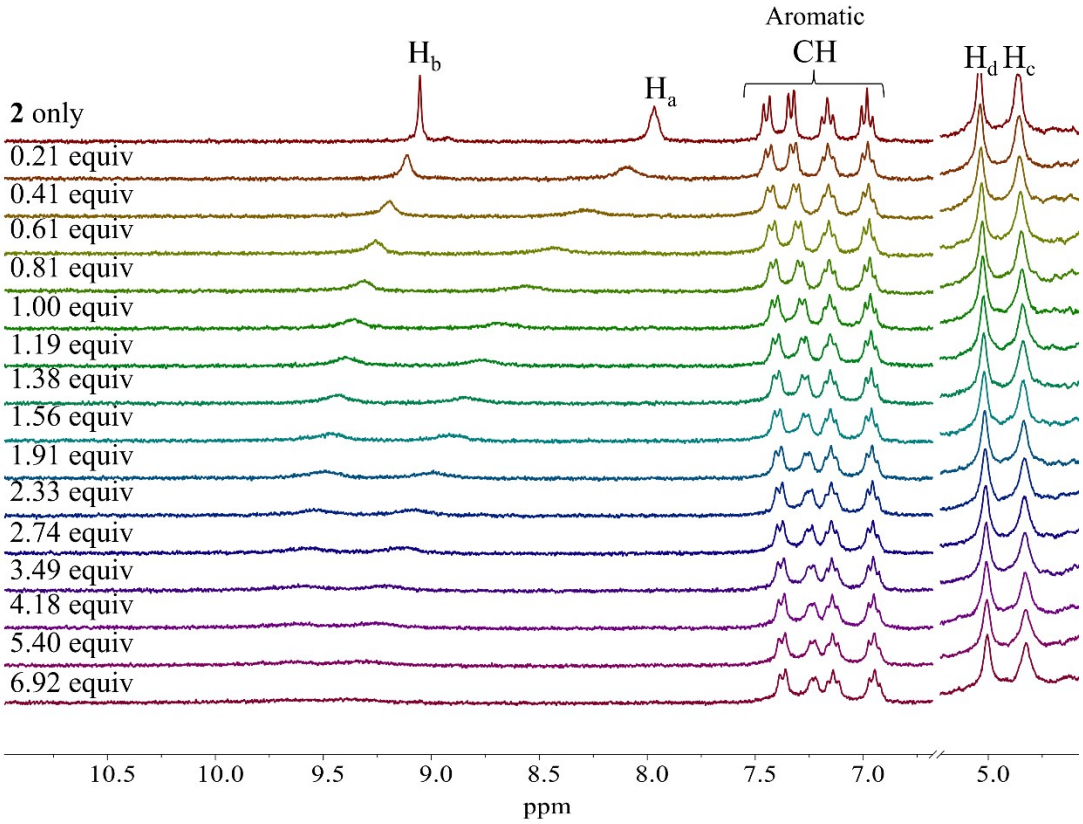


Figure S3. Partial views of ^{1}H NMR spectra of cage **2** (3 mM) recorded in 10% D_2O in DMSO-d_6 in the presence of the indicated anions as their respective TBA^+ salts except for HCO_3^- in the TEA^+ (tetraethylammonium) salt form.



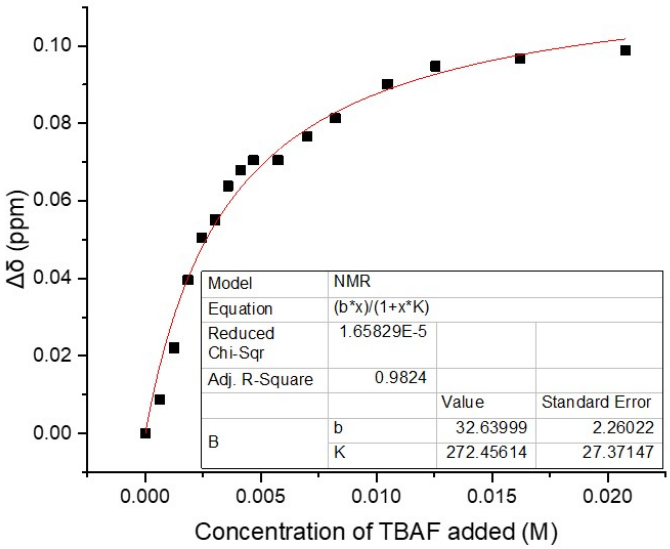


Figure S4. Partial ^{1}H NMR spectra recorded during the titration of **2** (3 mM) with TBAF (tetrabutylammonium fluoride) in 10% H_2O in DMSO-d_6 and the corresponding binding isotherm.

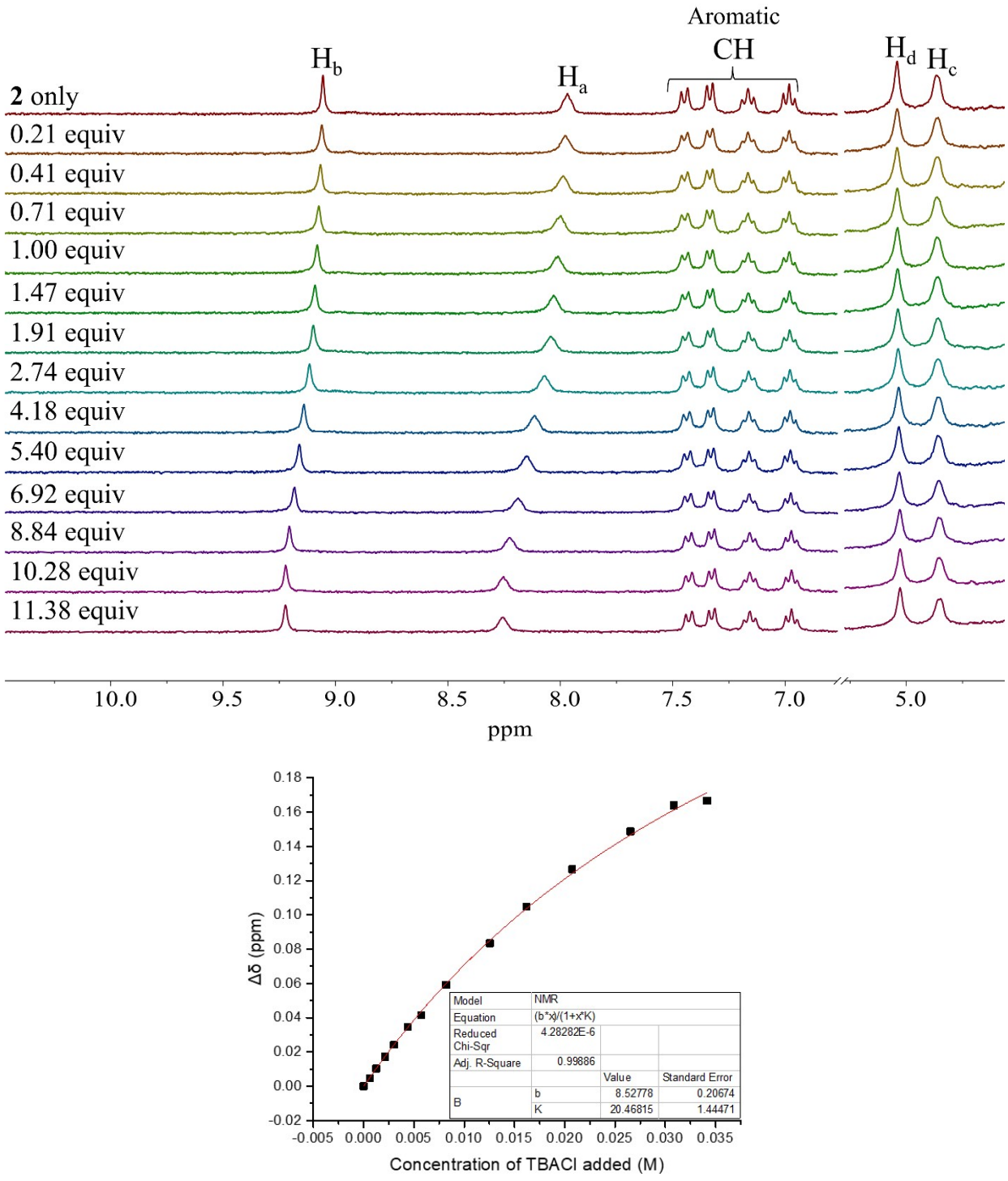


Figure S5. Partial ^{1}H NMR spectra recorded during the titration of **2** (3 mM) with TBACl (tetrabutylammonium chloride) in 10% H_2O in DMSO-d_6 and the corresponding binding isotherm.

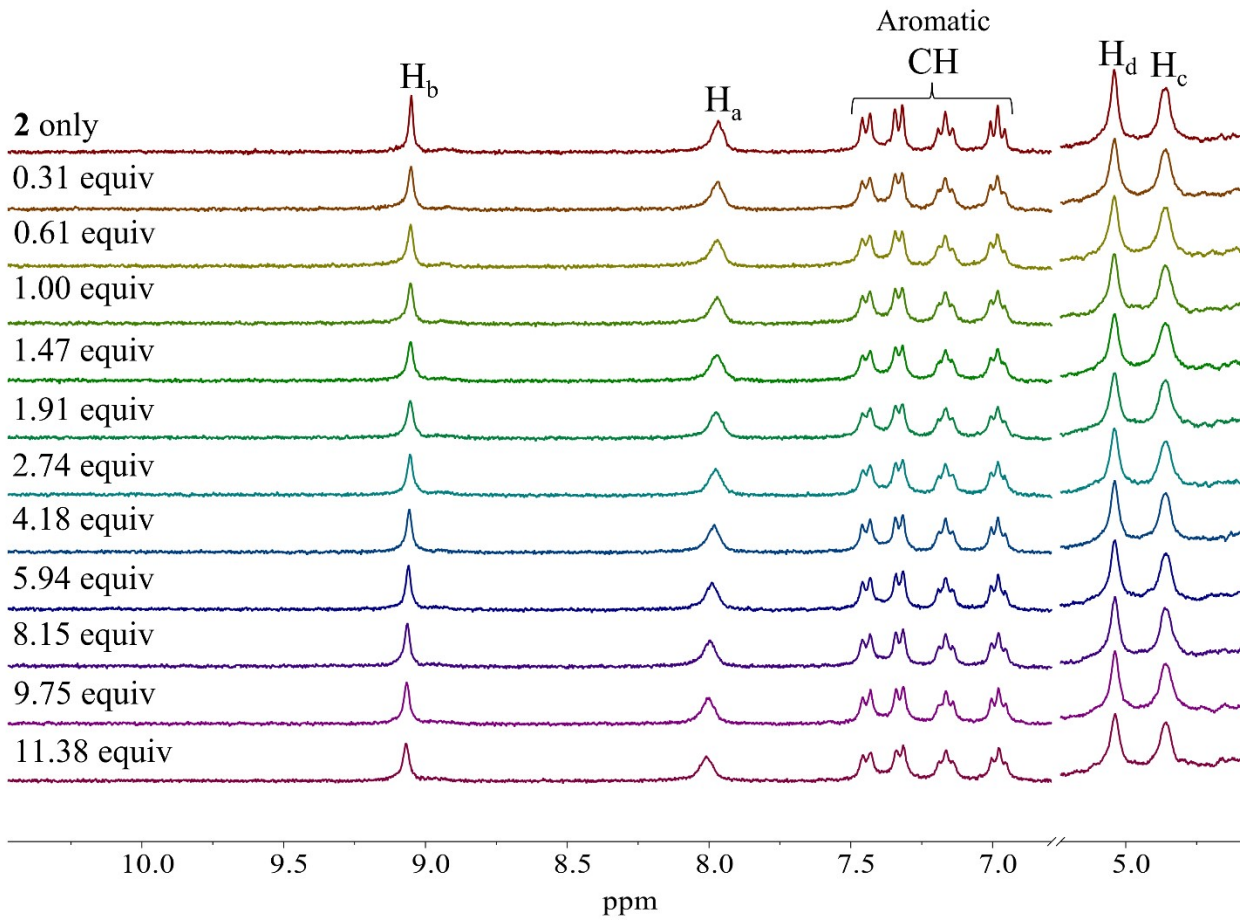


Figure S6. Partial ^{1}H NMR spectra recorded during the titration of **1** (3 mM) with TBABr (tetrabutylammonium bromide) in 10% H_2O in DMSO-d_6.

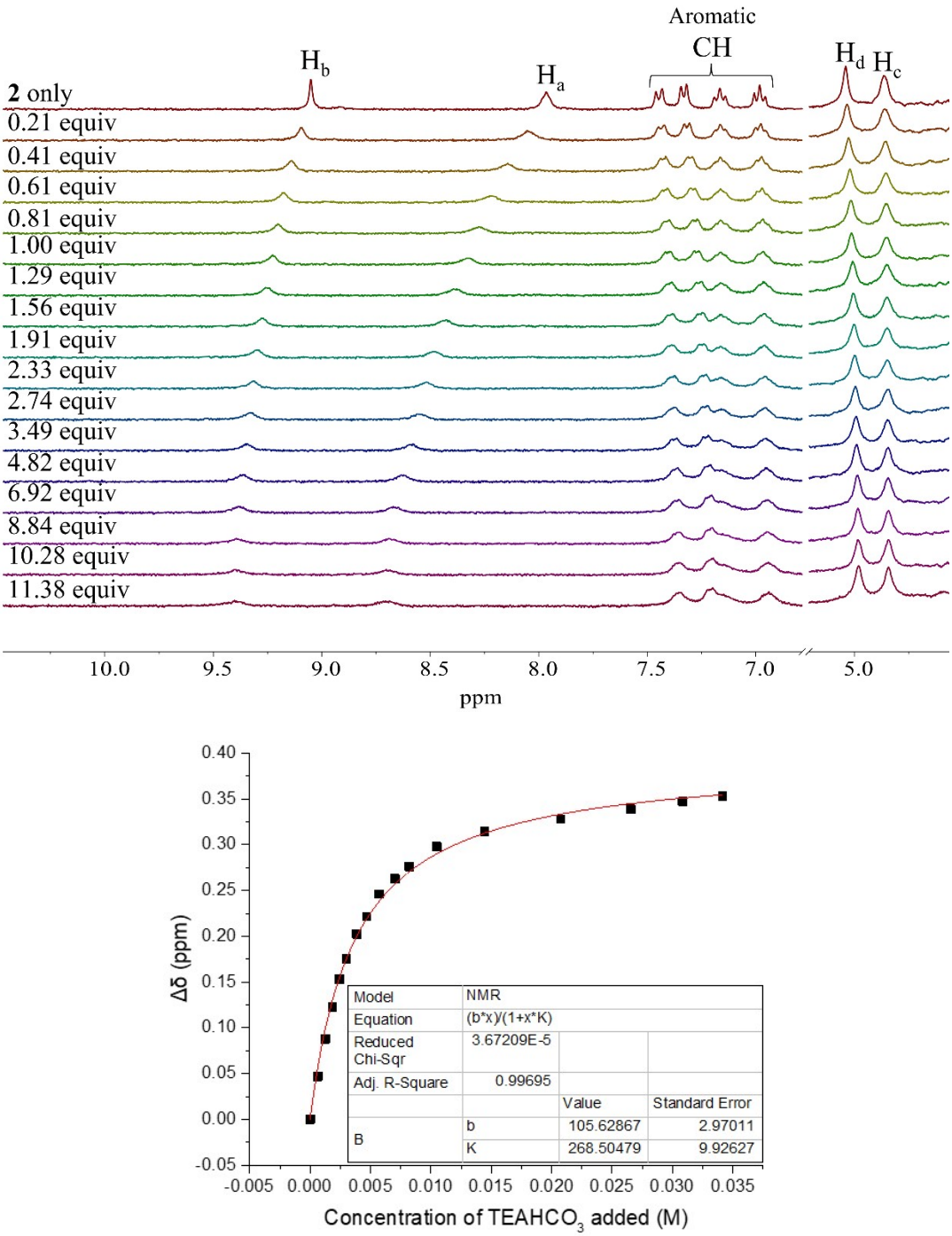


Figure S7. Partial ^{1}H NMR spectra recorded during the titration of **1** (3 mM) with $TEAHCO_3$ (tetraethylammonium bicarbonate) in 10% H_2O in DMSO-d_6 and the corresponding binding isotherm.

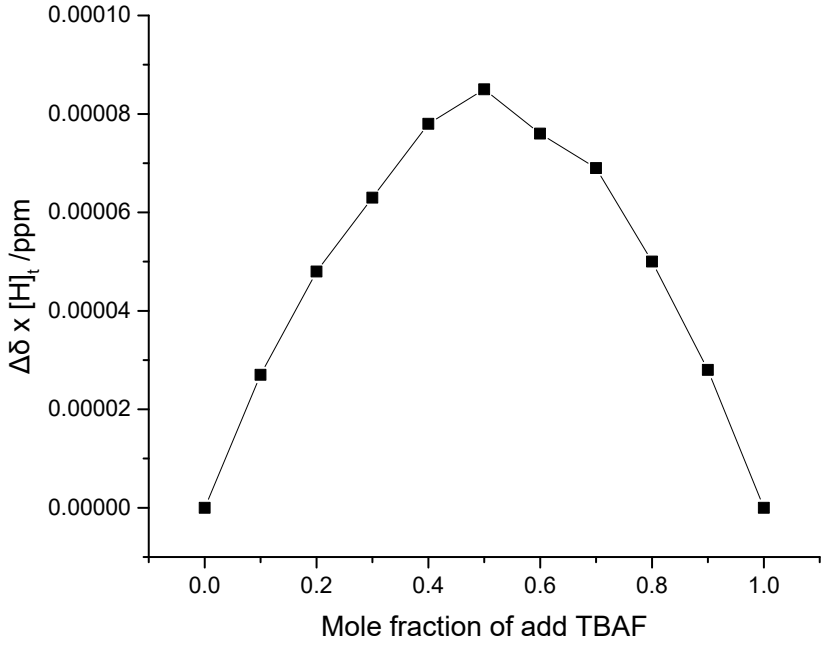


Figure S8. Job's plots for the interaction of receptor **2** with TBAF.

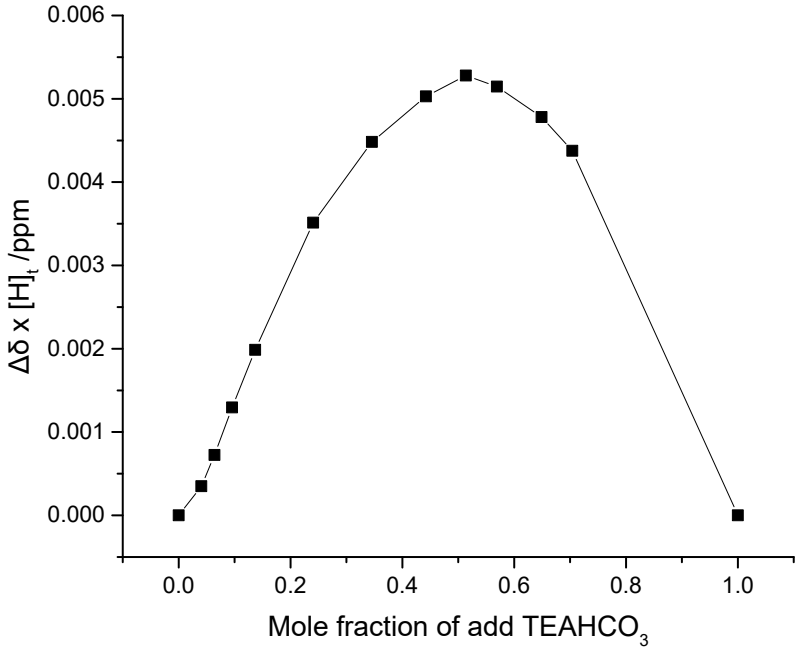


Figure S9. Job's plots for the interaction of receptor **2** with $TEAHCO_3$.

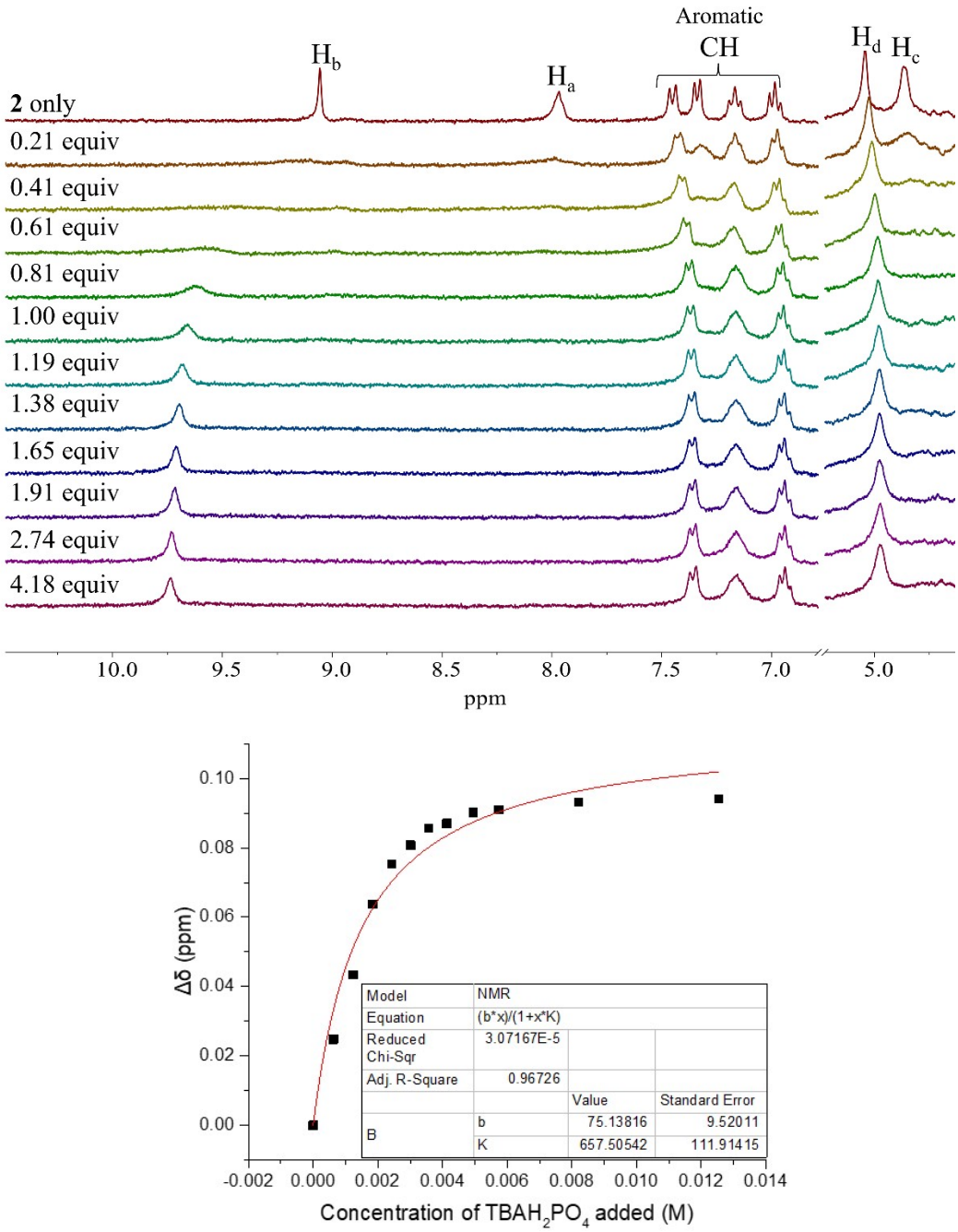


Figure S10. Partial ^{1}H NMR spectra recorded during the titration of **1** (3 mM) with $TBAH_2PO_4$ (tetrabutylammonium dihydrogen phosphate) in 10% H_2O in DMSO-d_6 and the corresponding binding isotherm.

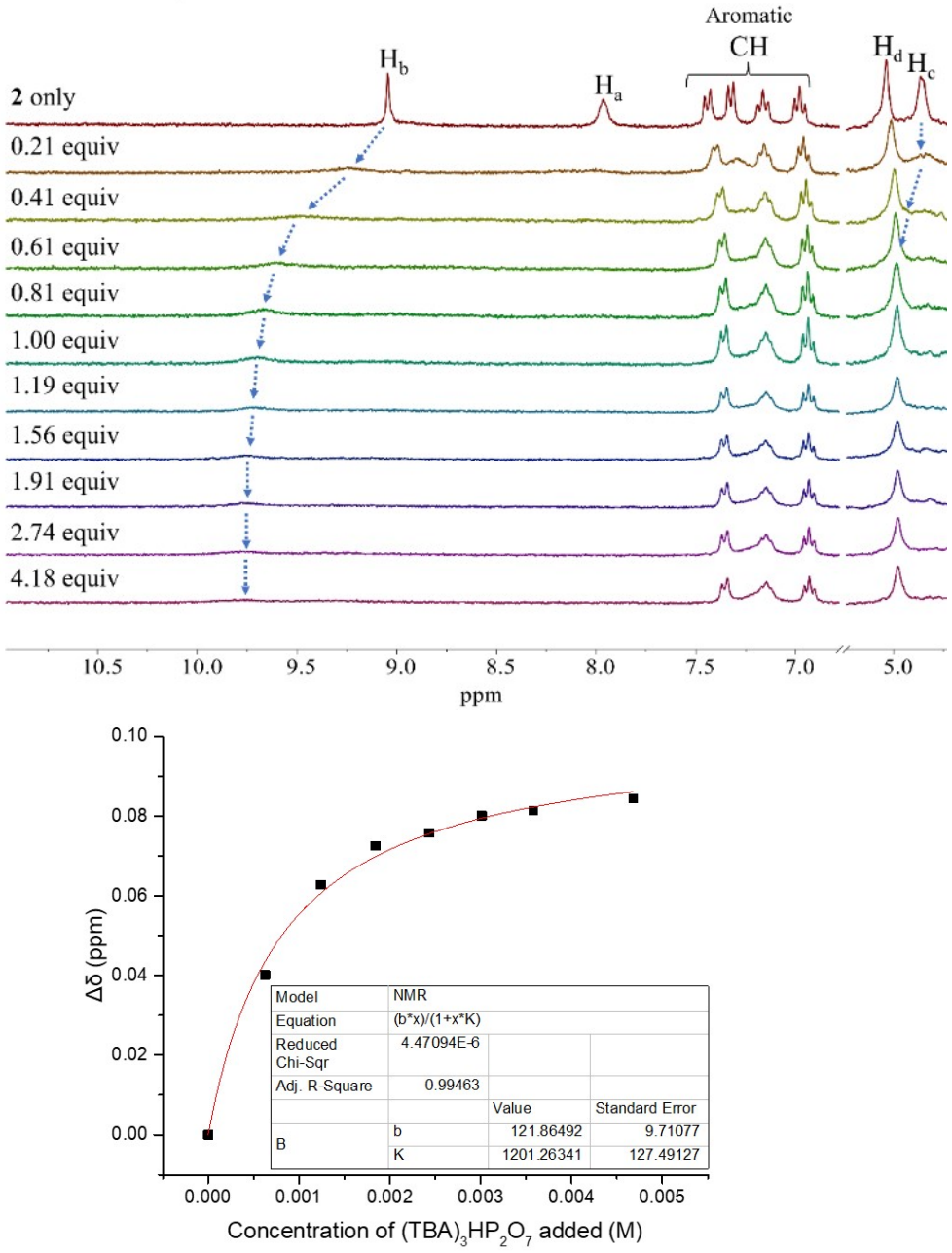


Figure S11. Partial ^{1}H NMR spectra recorded during the titration of **1** (3 mM) with $(TBA)_3HP_2O_7$ (tris(tetrabutylammonium)hydrogen pyrophosphate) in 10% H_2O in DMSO-d_6 and the corresponding binding isotherm.

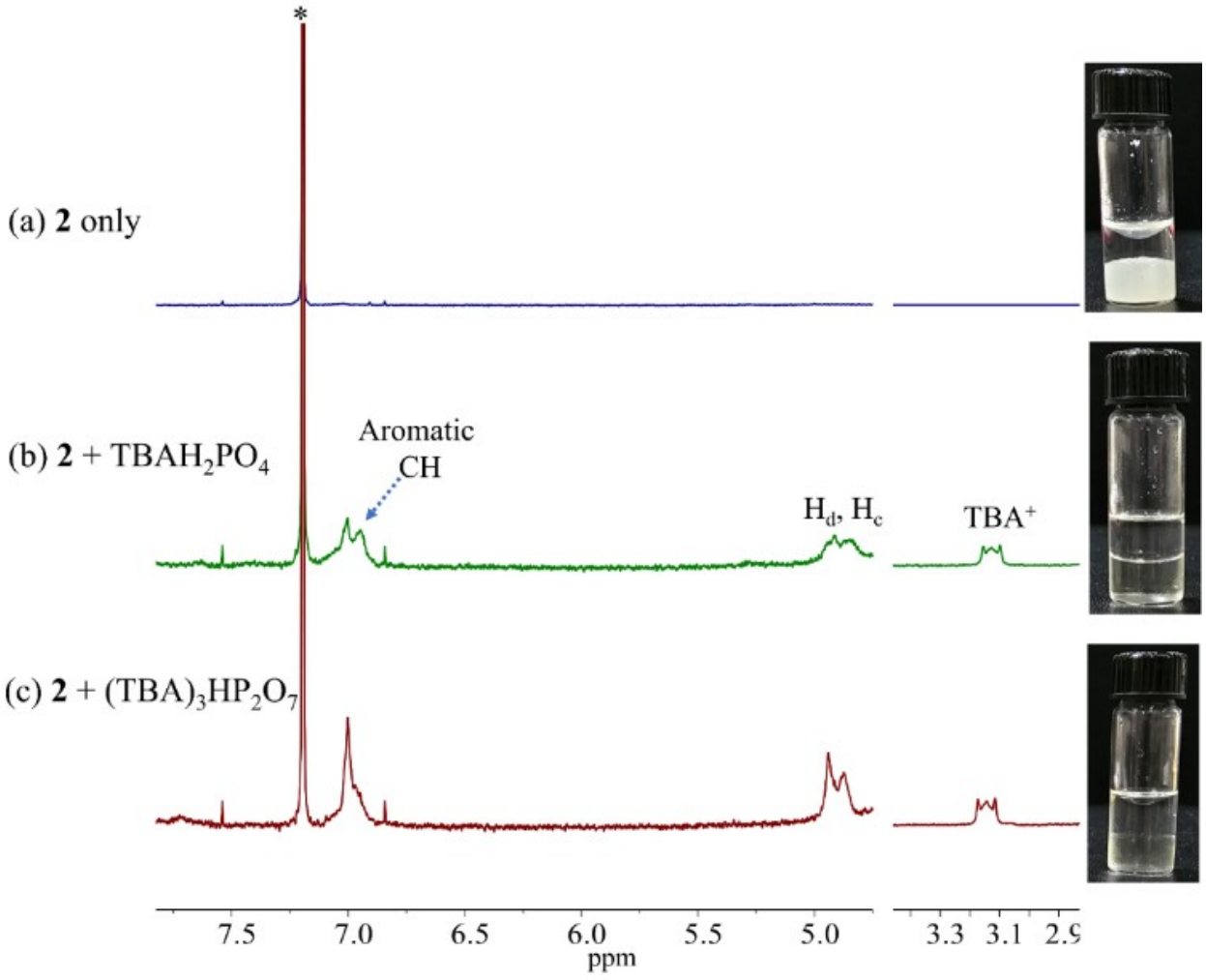


Figure S12. Partial ^{1}H NMR spectra of $CDCl_3$ solutions of receptor **2** after contacting with (a) an ion-free D_2O solution, (b) a D_2O solution containing 30 equiv of $TBAH_2PO_4$, and (c) a D_2O solution containing 30 equiv of $(TBA)_3HP_2O_7$ and the photos of the resulting two phase solutions. The asterisk denotes a residual NMR solvent peak.

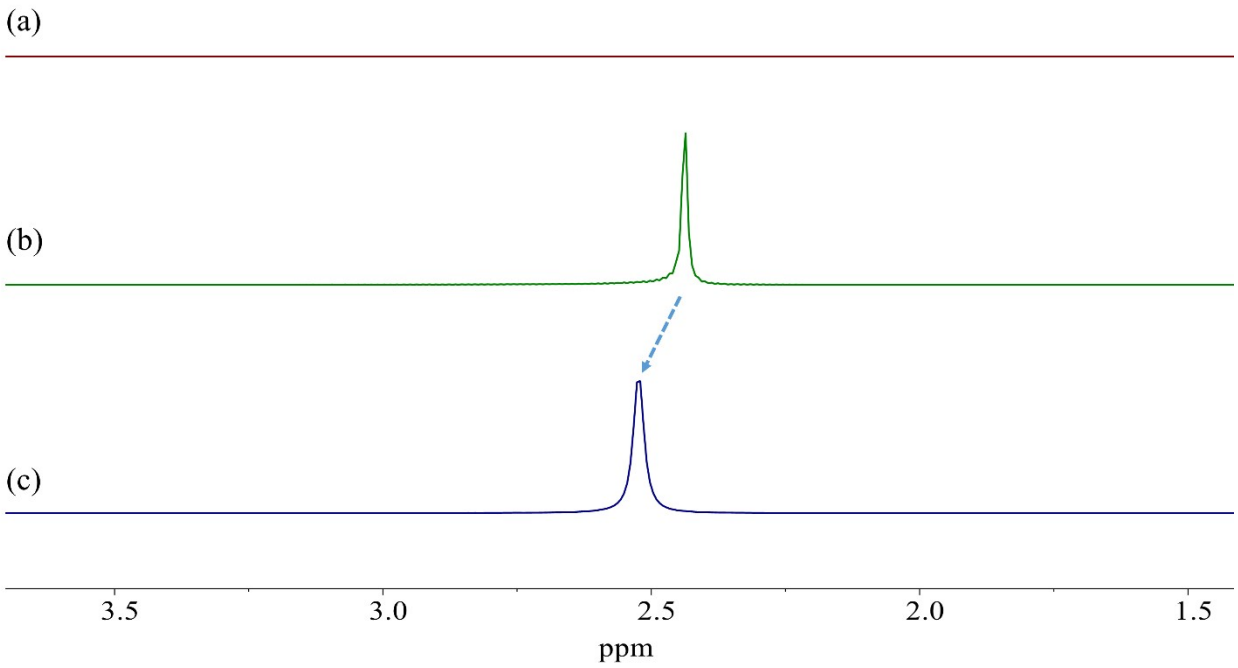


Figure S13. ^{31}P NMR spectra of (a) a receptor-free $CDCl_3$ solution after contacting with a D_2O solution containing $TBAH_2PO_4$ (90 mM), (b) a $CDCl_3$ solution of $TBAH_2PO_4$ (90 mM) without contacting with a D_2O solution, and (c) a $CDCl_3$ solution of receptor **2** (3 mM) after contacting with a D_2O solution containing $TBAH_2PO_4$ (30 mM).

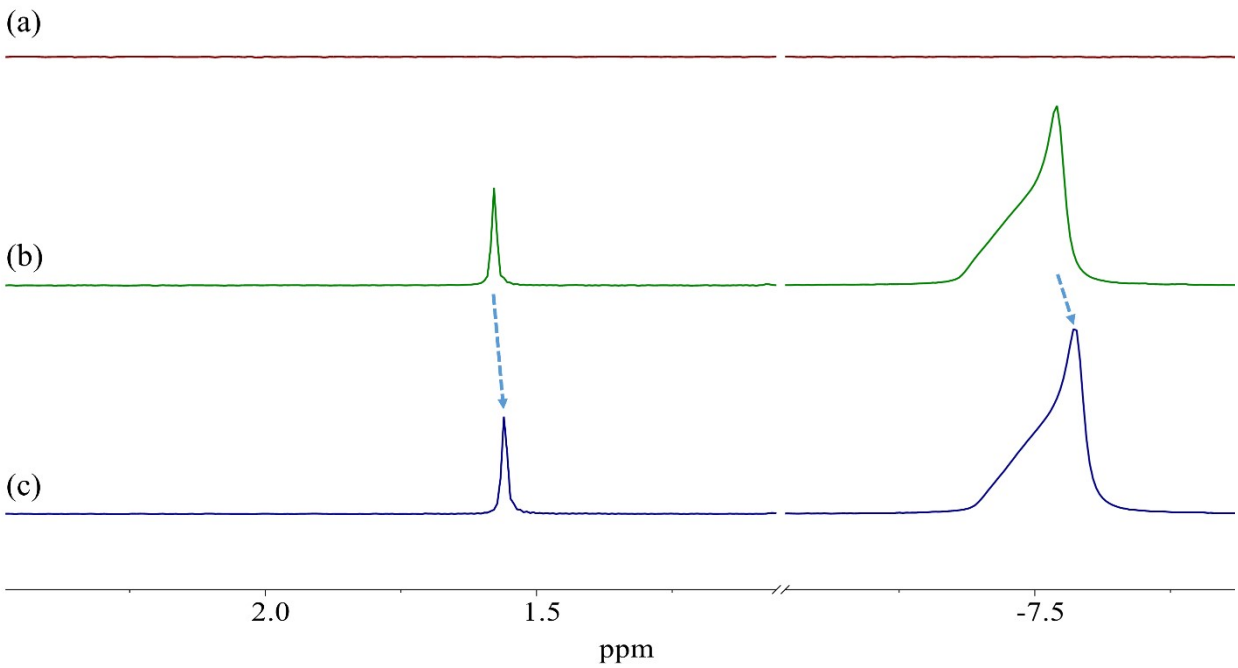


Figure S14. ^{31}P NMR spectra of (a) a receptor-free $CDCl_3$ solution after contacting with a D_2O solution containing $(TBA)_3HP_2O_7$ (90 mM), (b) a $CDCl_3$ solution of $(TBA)_3HP_2O_7$ (90 mM) without contacting with a D_2O solution, and (c) a $CDCl_3$ solution of receptor **2** (3 mM) after contacting with a D_2O solution containing $(TBA)_3HP_2O_7$ (90 mM).

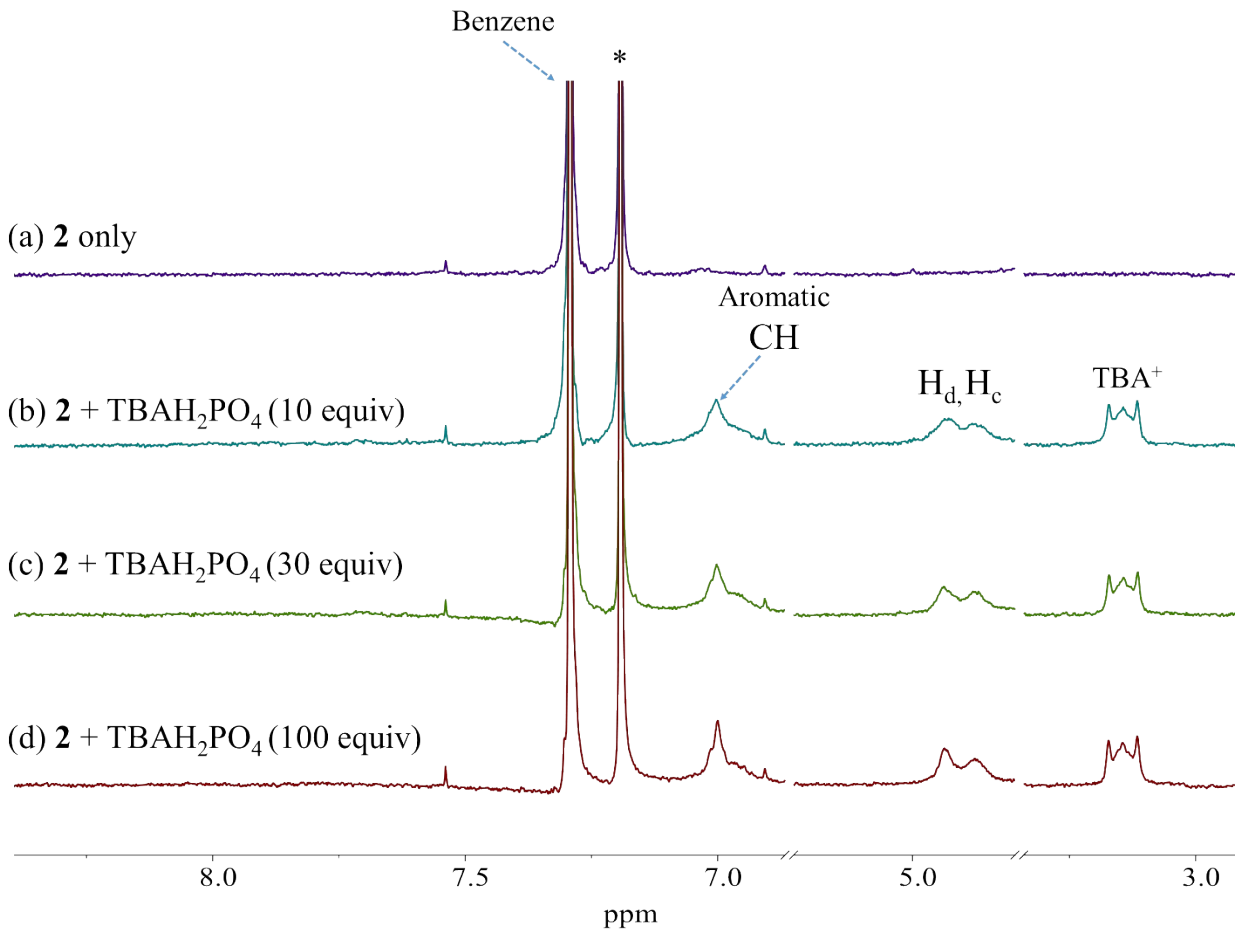


Figure S15. Partial ^{1}H NMR spectra of $CDCl_3$ solutions of receptor **2** (1 mM) recorded after contacting with (a) an ion-free D_2O solution, (b) a D_2O solution containing 10 equiv of $TBAH_2PO_4$ (*ca.* 56% extraction efficiency), (c) a D_2O solution containing 30 equiv of $TBAH_2PO_4$ (*ca.* 71% extraction efficiency), and (d) a D_2O solution containing 100 equiv of $TBAH_2PO_4$ (*ca.* 84% extraction efficiency). The asterisk denotes a residual NMR solvent peak. Benzene (10 mM) was used as an internal reference.

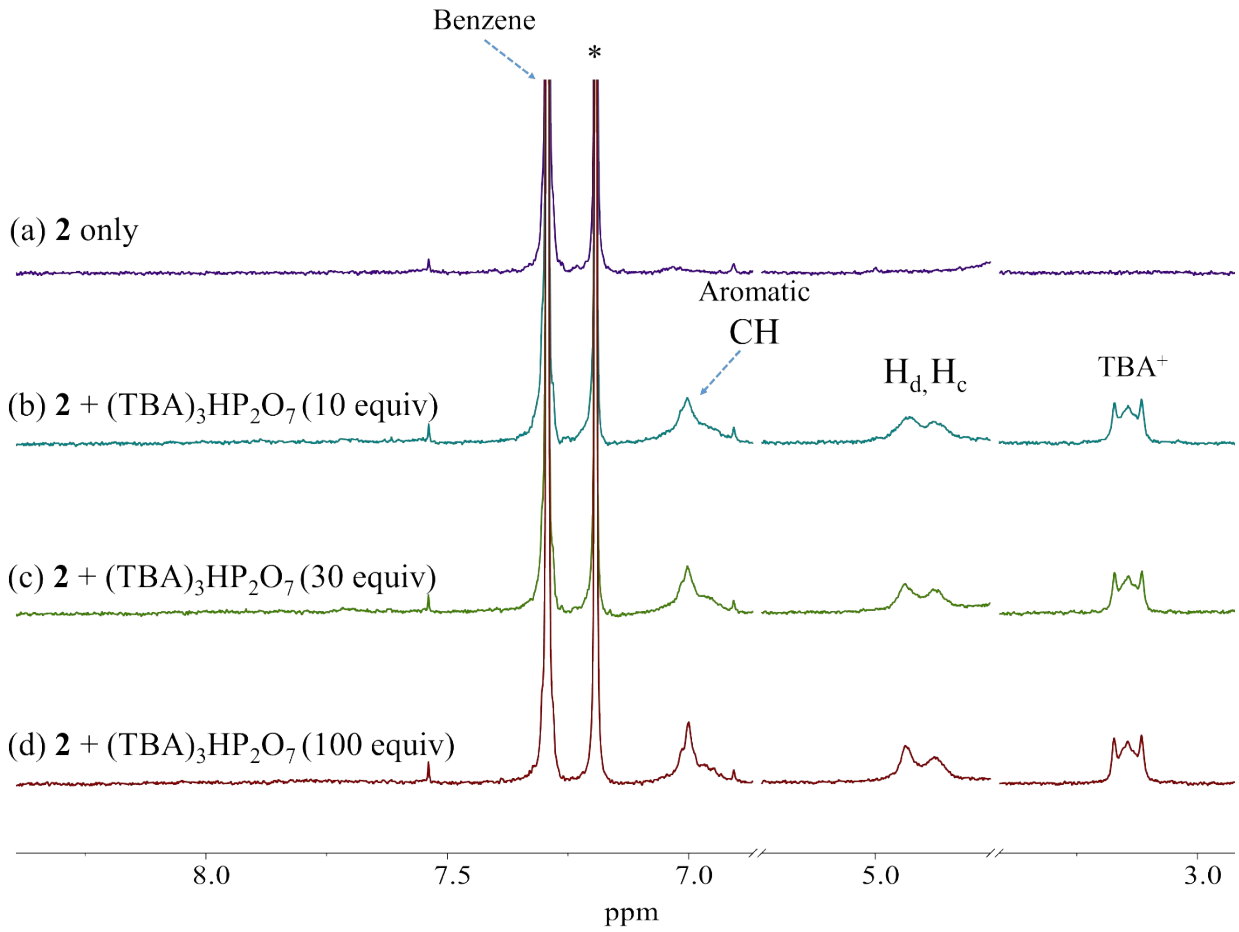


Figure S16. Partial ^{1}H NMR spectra of $CDCl_3$ solutions of receptor **2** (1 mM) recorded after contacting with (a) an ion-free D_2O solution, (b) a D_2O solution containing 10 equiv of $(TBA)_3HP_2O_7$ (*ca.* 23% extraction efficiency), (c) a D_2O solution containing 30 equiv of $(TBA)_3HP_2O_7$ (*ca.* 27% extraction efficiency), and (d) a D_2O solution containing 100 equiv of $(TBA)_3HP_2O_7$ (*ca.* 30% extraction efficiency). The asterisk denotes a residual NMR solvent peak. Benzene (10 mM) was used as an internal reference.

3. NMR Spectra and HRMS spectra

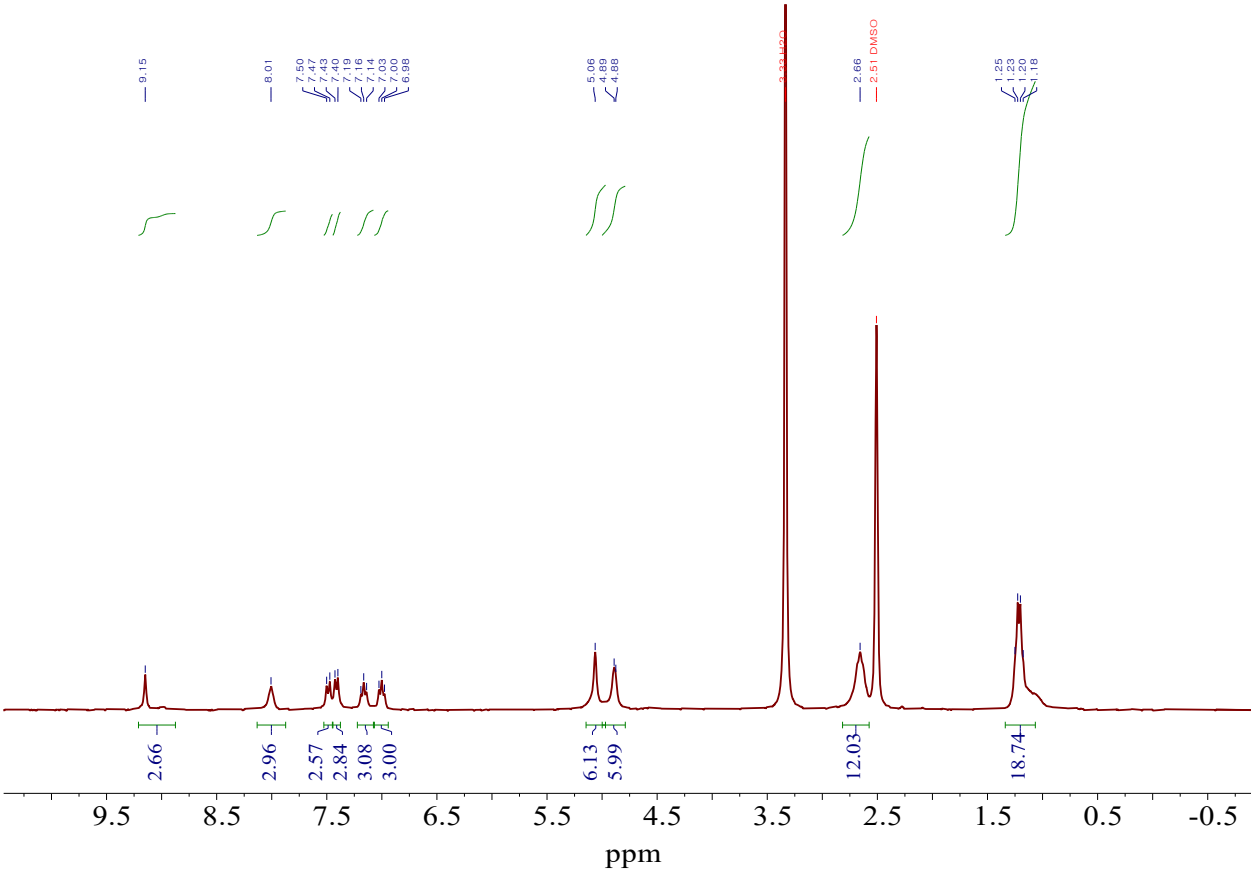


Figure S17. ^{1}H NMR spectrum of **2** recorded in DMSO-d_6.

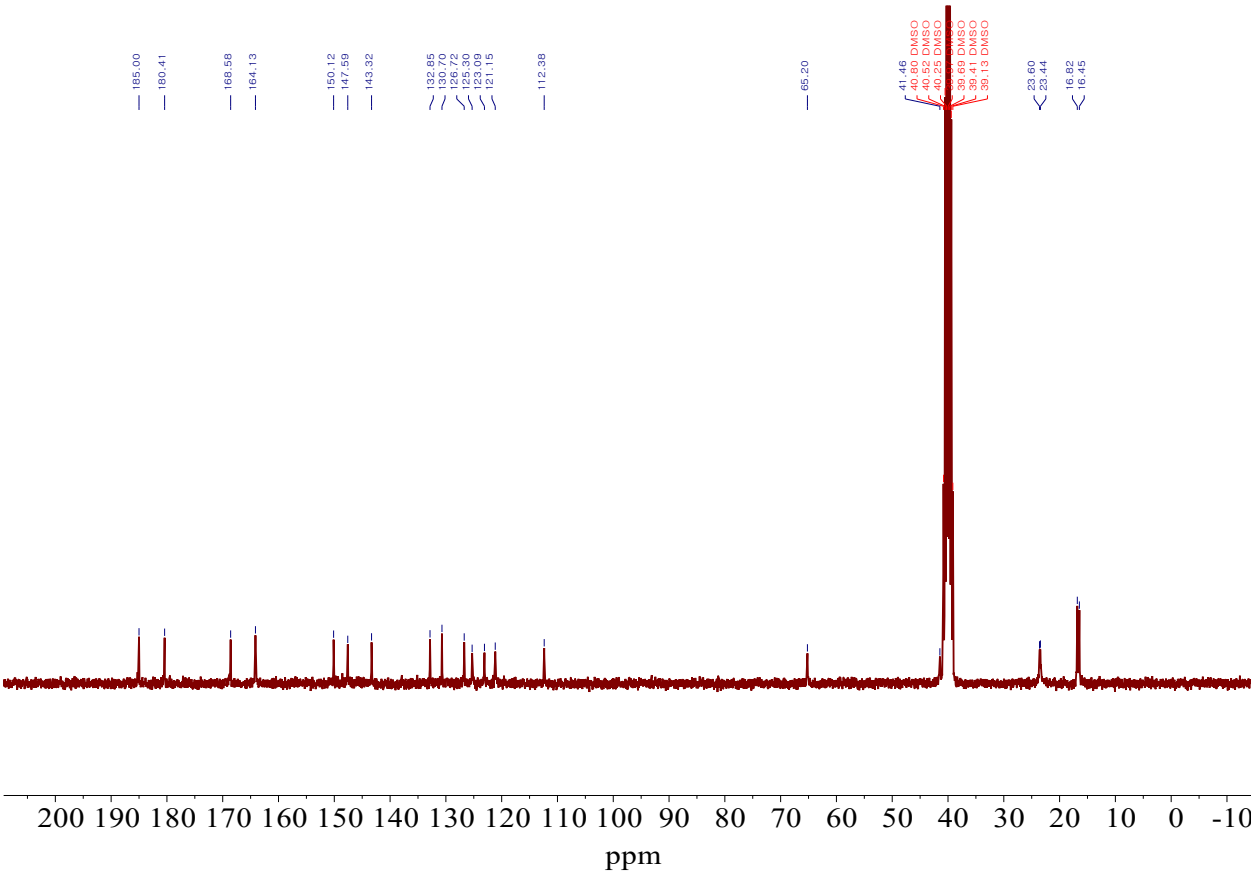


Figure S18. ^{13}C NMR spectrum of **2** recorded in DMSO-d_6.

Data : 230109_YJH cage_HR Date : 09-Jan-2023 14:40
Instrument : MStation
Sample : -
Note : -
Inlet : Direct Ion Mode : FAB+
RT : 2.40 min Scan# : 73
Elements : C 100/1, H 100/1, N 10/1, O 10/1
Mass Tolerance : 1000ppm, 3mmu if m/z > 3
Unsaturation (U.S.) : -0.5 - 34.0

	Observed m/z	Int%	Err[ppm / mmu]	U.S.	Composition
1	1009.4515	100.00	+2.8 / +2.8	34.0	C58 H59 N9 O8
2			+1.5 / +1.5	33.5	C60 H61 N6 O9
3			+0.2 / +0.2	33.0	C62 H63 N3 O10

Figure S19. HR FAB mass spectral data of **2**.

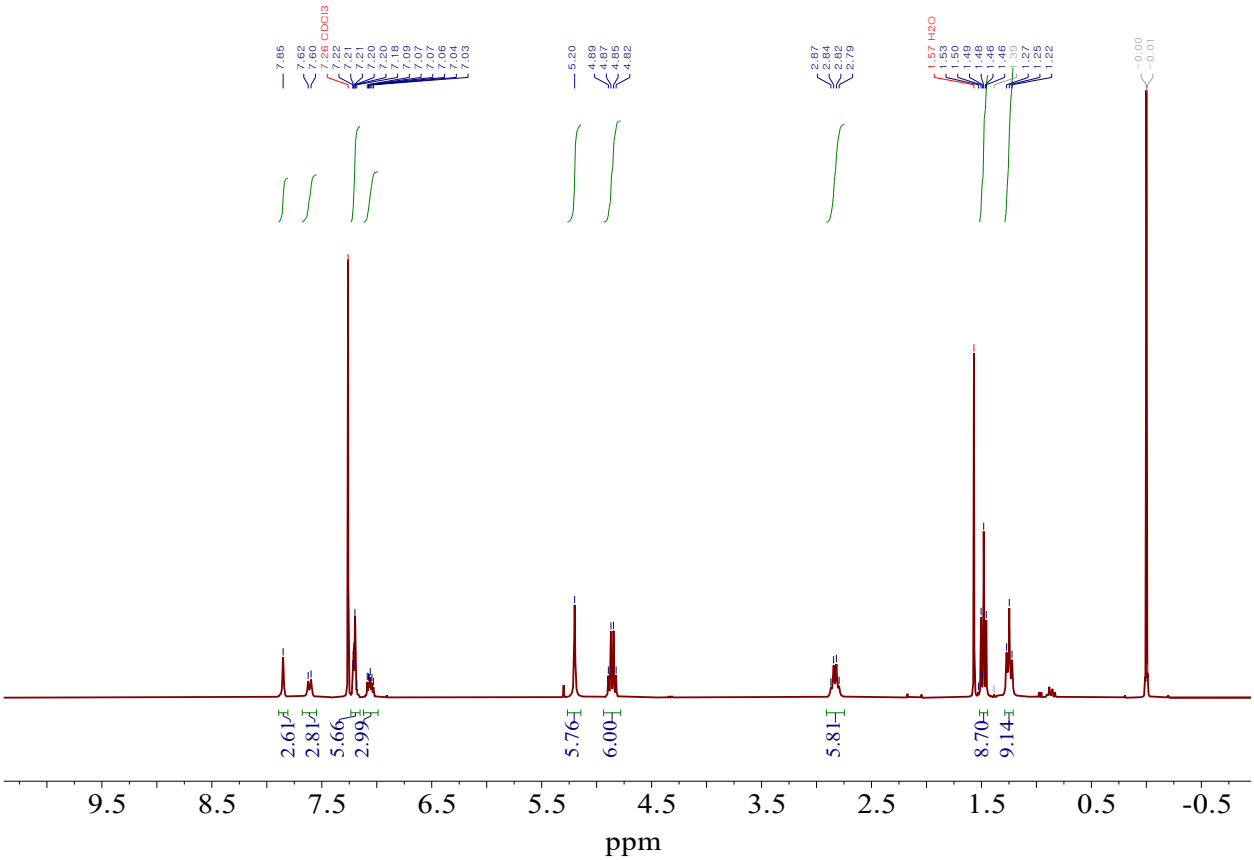


Figure S20. ^{1}H NMR spectrum of **4** recorded in $CDCl_3$.

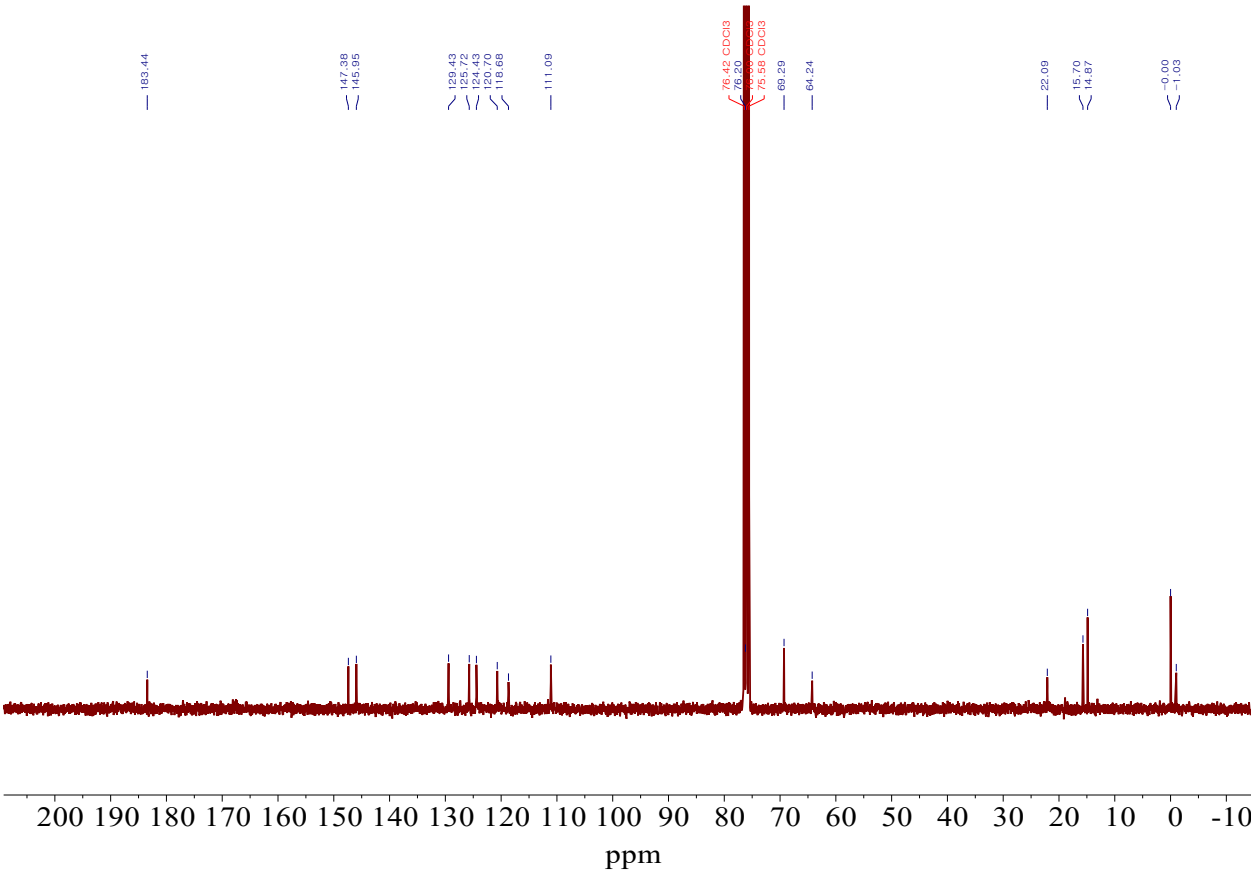


Figure S21. ^{13}C NMR spectrum of **4** recorded in $CDCl_3$.

Data : 230109_YJH Tri_HR Date : 09-Jan-2023 16:46
Instrument : MStation
Sample : -
Note : -
Inlet : Direct Ion Mode : FAB+
RT : 3.00 min Scan# : 91
Elements : C 100/1, H 100/1, N 10/1, O 15/1
Mass Tolerance : 1000ppm, 3mmu if m/z > 3
Unsaturation (U.S.) : -0.5 - 28.0

	Observed m/z	Int%	Err[ppm / mmu]	U.S.	Composition
1	898.3561	100.00	+2.6 / +2.3	28.0	C49 H50 N6 O11
2			+1.1 / +1.0	27.5	C51 H52 N3 O12
3			-2.4 / -2.2	19.5	C40 H52 N9 O15

Figure S22. HR FAB mass spectral data of **4**.

4. X-ray crystal data

X-ray Experimental for 2·TBAH_2PO_4-2$CHCl_3$-3CH_3OH: Crystals grew as colorless plates by vapor diffusion of methanol into a chloroform solution. The data crystal had approximate dimensions; 0.18 x 0.15 x 0.058 mm. The data were collected on an Rigaku Oxford Diffraction SuperNova Dual Source diffractometer using a μ-focus Cu Kα radiation source (λ = 1.5418Å) with collimating mirror monochromators. A total of 2322 frames of data were collected using ω-scans with a scan range of 0.5° and a counting time of 27 seconds per frame for frames collected with a detector offset of +/- 48.7° and 103 seconds per frame with frames collected with a detector offset of 107.8°. The data were collected at 100 K using an Oxford Cryostream low temperature device. Details of crystal data, data collection and structure refinement are listed in Table 1. Data collection, unit cell refinement and data reduction were performed using Rigaku Oxford Diffraction's CrysAlisPro V 1.171.41.123a.[2] The structure was solved by direct methods using SHELXT[3] and refined by full-matrix least-squares on F^2 with anisotropic displacement parameters for the non-H atoms using SHELXL-2018/3.[4] Structure analysis was aided by use of the programs PLATON[5] and OLEX2.[6]

A molecule of chloroform and a molecule of methanol were disordered. The disorder was modeled by setting the sum of the site occupancy factors for the appropriate atoms to 1 while refining the displacement parameters. For both molecules, geometric restraints were applied during refinement. There was a fairly large peak about 1.7Å from C8 that interpreted to be due to a small amount of Cl atom impurity. The site occupancy was refined to approximately 4%.

The function, $\Sigma w(|F_o|^2 - |F_c|^2)^2$, was minimized, where $w = 1/[(\sigma(F_o))^2 + (0.0642*P)^2 + (6.9455*P)]$ and $P = (|F_o|^2 + 2|F_c|^2)/3$. $R_w(F^2)$ refined to 0.123, with R(F) equal to 0.0444 and a goodness of fit, S, = 1.03. Definitions used for calculating R(F), $R_w(F^2)$ and the goodness of fit, S, are given below.[7] The data were checked for secondary extinction effects but no correction was necessary. Neutral atom scattering factors and values used to calculate the linear absorption coefficient are from the International Tables for X-ray Crystallography (1992).[8] All figures were generated using SHELXTL/PC.[9] Tables of positional and thermal parameters, bond lengths and angles, torsion angles and figures can be obtained from the Cambridge Crystallographic Centre by referencing CCDC number 2266290.

Table S1. Crystal data and structure refinement for **2**·$TBAH_2PO_4$-2$CHCl_3$-3CH_3OH

Empirical formula	C81 H112 Cl6 N7 O16 P	
Formula weight	1685.32	
Temperature	100(1) K	
Wavelength	1.54184 Å	
Crystal system	monoclinic	
Space group	P 1 21/n 1	
Unit cell dimensions	a = 13.13170(10) Å	α= 90°.
	b = 19.00400(10) Å	β= 96.6710(10)°.
	c = 34.9608(2) Å	γ = 90°.
Volume	8665.57(10) $Å^3$	
Z	4	
Density (calculated)	1.292 Mg/m^3	
Absorption coefficient	2.538 mm^{-1}	
F(000)	3572	
Crystal size	0.182 x 0.15 x 0.058 mm^3	
Theta range for data collection	2.545 to 77.117°.	
Index ranges	-14<=h<=16, -23<=k<=18, -41<=l<=44	
Reflections collected	65351	
Independent reflections	17522 [R(int) = 0.0320]	
Completeness to theta = 67.684°	99.2 %	
Absorption correction	Gaussian	
Max. and min. transmission	1.000 and 0.653	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	17522 / 79 / 1097	
Goodness-of-fit on F^2	1.032	
Final R indices [I>2sigma(I)]	R1 = 0.0444, wR2 = 0.1205	
R indices (all data)	R1 = 0.0482, wR2 = 0.1229	
Extinction coefficient	n/a	
Largest diff. peak and hole	0.963 and -0.654 $e.Å^{-3}$	

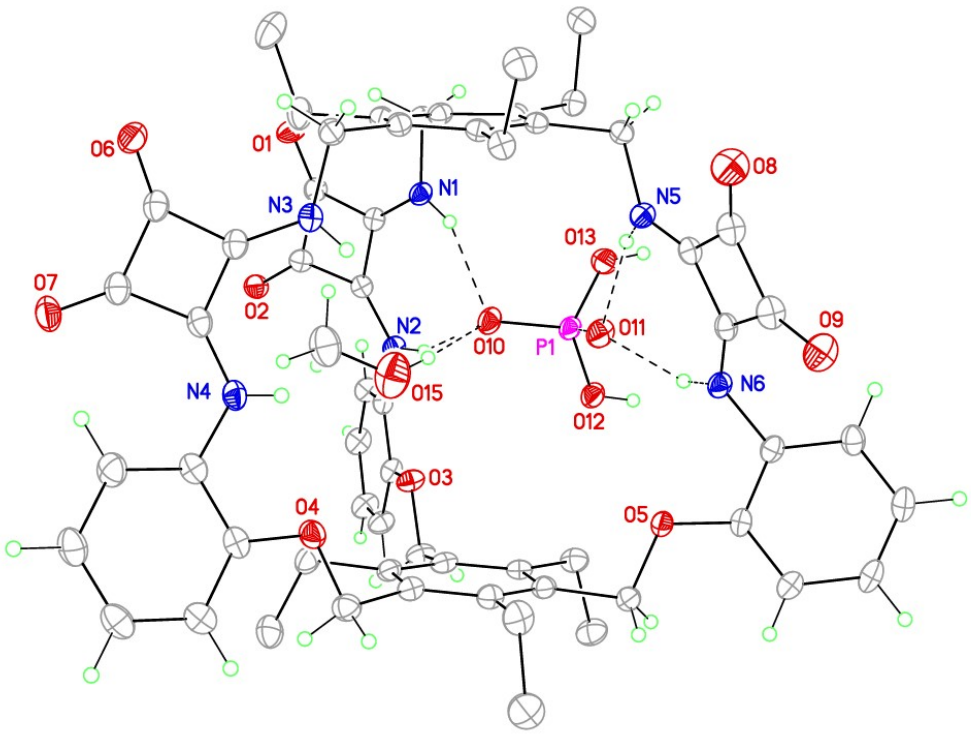


Figure S23. View of the $H_2PO_4^-$ complex in **1** showing the heteroatom labeling scheme. Displacement ellipsoids are scaled to the 50% probability level. The methyl group H atoms were omitted for clarity.

5. References

1. Yamanaka, M.; Nakagawa, T.; Aoyama, R.; Nakamura, T. *Tetrahedron*, **2008**, *64*, 11558-11567.
2. CrysAlisPro. Rigaku Oxford Diffraction (2021). CrysAlisPro Software System, 1.171.41.123a.
3. SHELXT. (2015). G. M. Sheldrick. A program for crystal structure solution. Acta Cryst. A71, 3-8.
4. Sheldrick, G. M. (2015). SHELXL-2018/3. Program for the Refinement of Crystal Structures. Acta Cryst., C71, 3-8.
5. Spek, A. L. (2009). PLATON, A Multipurpose Crystallographic Tool. Utrecht University, The Netherlands. Acta Cryst. D65, 148-155.
6. OLEX2. Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. and Puschmann, H. A Complete Structure Solution, Refinement and Analysis Program. J. Appl. Cryst. 42, 339-341.
7. $R_W(F^2) = \{\Sigma w(|F_o|^2 - |F_c|^2)^2/\Sigma w(|F_o|)^4\}^{1/2}$ where w is the weight given each reflection.

   $R(F) = \Sigma (|F_o| - |F_c|)/\Sigma|F_o|\}$ for reflections with $F_o > 4(\sigma (F_o))$.

   $S = [\Sigma w(|F_o|^2 - |F_c|^2)^2/(n - p)]^{1/2}$, where n is the number of reflections and p is the number of refined parameters.
8. International Tables for X-ray Crystallography (1992). Vol. C, Tables 4.2.6.8 and 6.1.1.4, A. J. C. Wilson, editor, Boston: Kluwer Academic Press.
9. Sheldrick, G. M. (1994). SHELXTL/PC (Version 5.03). Siemens Analytical X-ray Instruments, Inc., Madison, Wisconsin, USA.