Supporting information

Measurement of Nucleobase pK_a Values in Model Mononucleotides shows RNA-RNA Duplexes to be More Stable than DNA-DNA Duplexes

P. Acharya, P. Cheruku, S. Chatterjee, S. Acharya and J. Chattopadhyaya*

Department of Bioorganic Chemistry, Box 581, Biomedical Center,
Uppsala University, S-751 23 Uppsala, Sweden

E-mail: jyoti@boc.uu.se

Table of contents

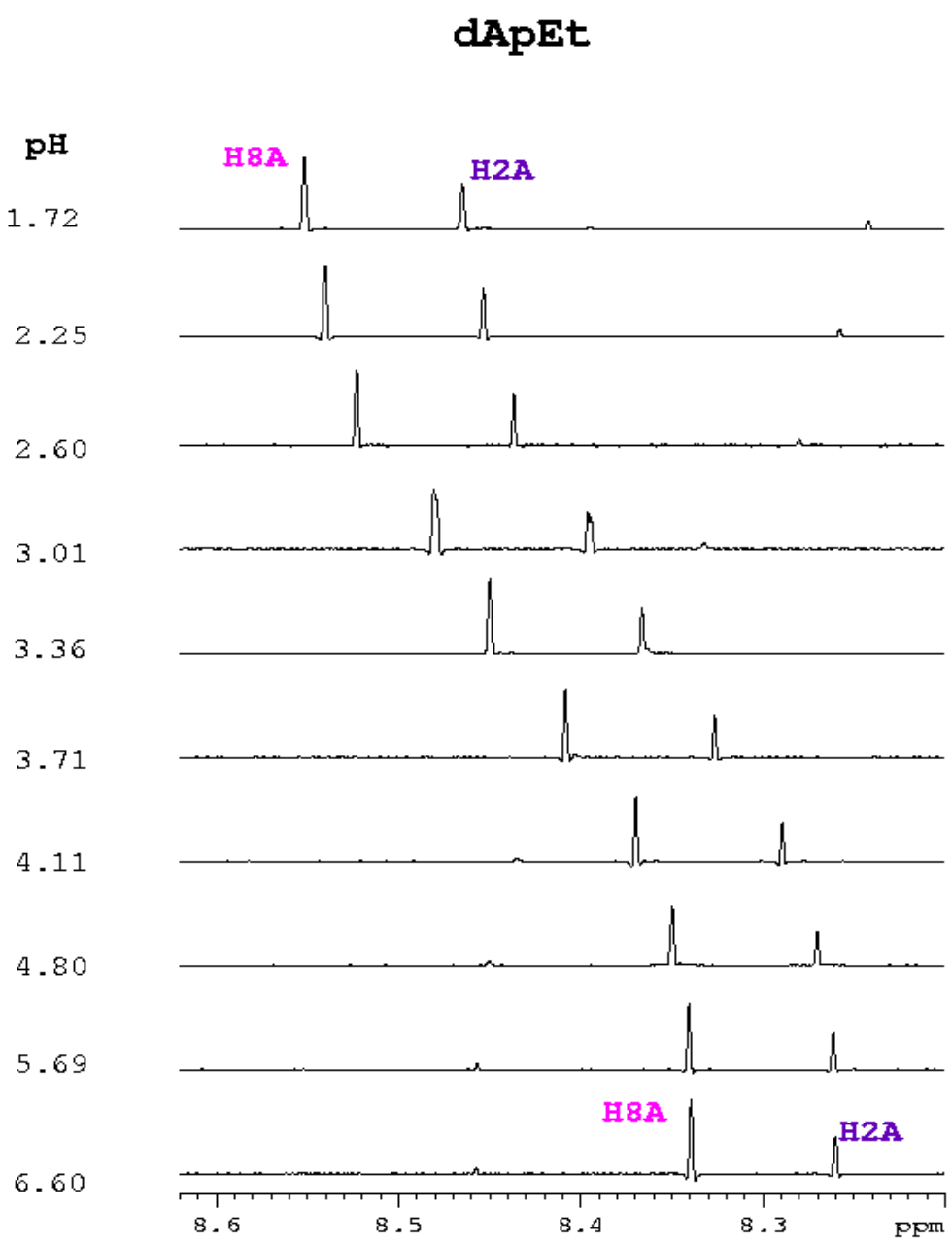
4. **Table S1**: ^{1}H NMR chemical shift for the aromatic protons at 298 K for **1a – 10a** and **1b** – **10b** at the protonated (P), the neutral (N) and the deprotonated (D) states (*i.e.* two plateaus of the titration plots at two extreme pHs).

5. **Table S2**: ^{1}H NMR chemical shift for the sugar protons at 298 K for **1a – 10a** and **1b** – **10b** at the protonated (P), the neutral (N) and the deprotonated (D) states (*i.e.* pH corresponding to the two plateaus of the titration plots).

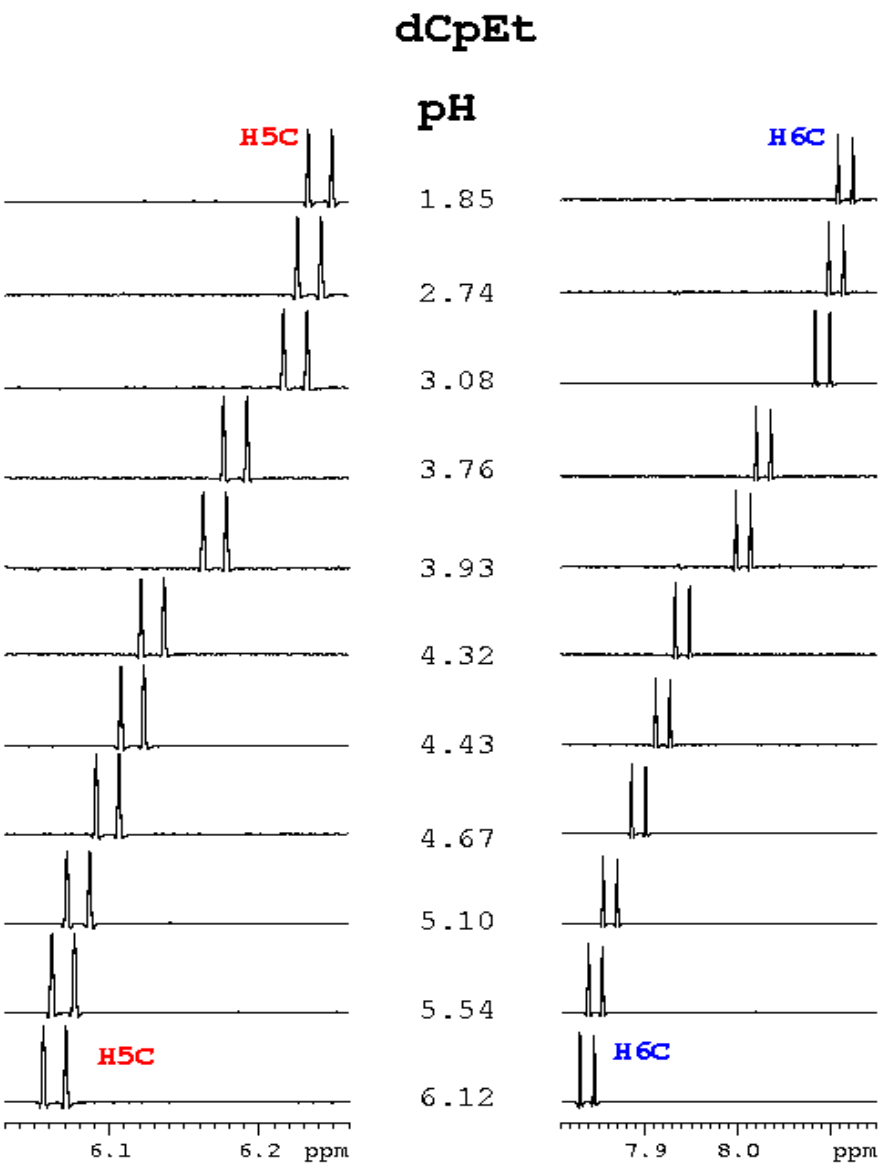
6. **Table S3**: **Table S3**. The total number of the middle (Mid) and the terminal (Term, both $^{3'}_{5'}$ and $^{5'}_{3'}$) basepairing [both r(A-U) and d(A-T) as well as r(G-C) and d(G-C) bp] in duplexes **1** – **14** .

7. **Table S4**: Dimerisation shift ($\delta_{\mathbf{Np}Et} - \delta_{\mathbf{Np}N'}$, in ppm) estimated from ^{1}H chemical shift at 298 K for aromatic protons of monomers, **7** – **10**, in comparison with dimers, **1** – **6**, at acidic (pH = 1.9), neutral (pH = 6.6) and alkaline (pH = 10.3) state.

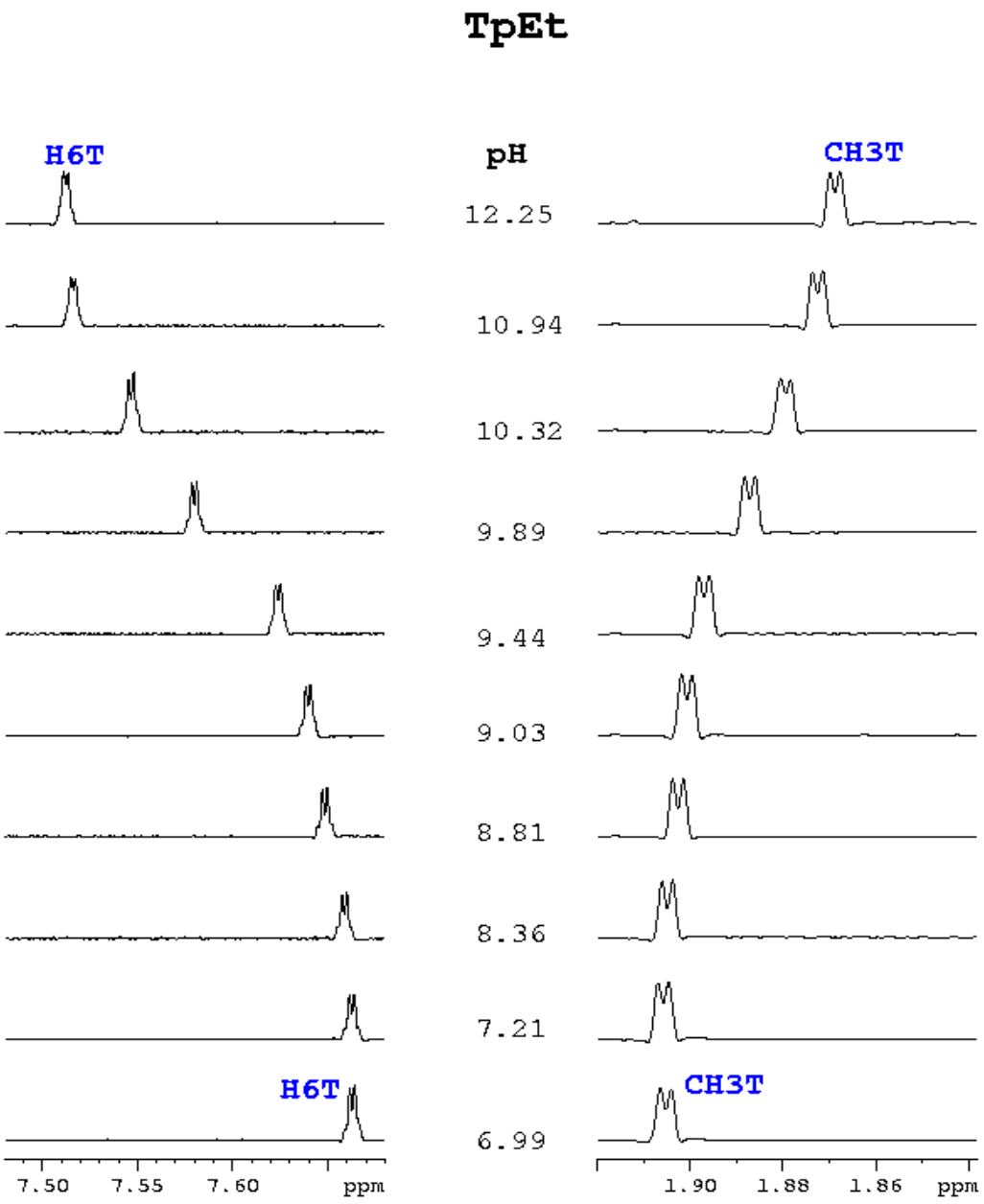
(A) pH-dependent ^{1}H chemical shift (in D_2O) of dApEt (**1a**) at 298 K



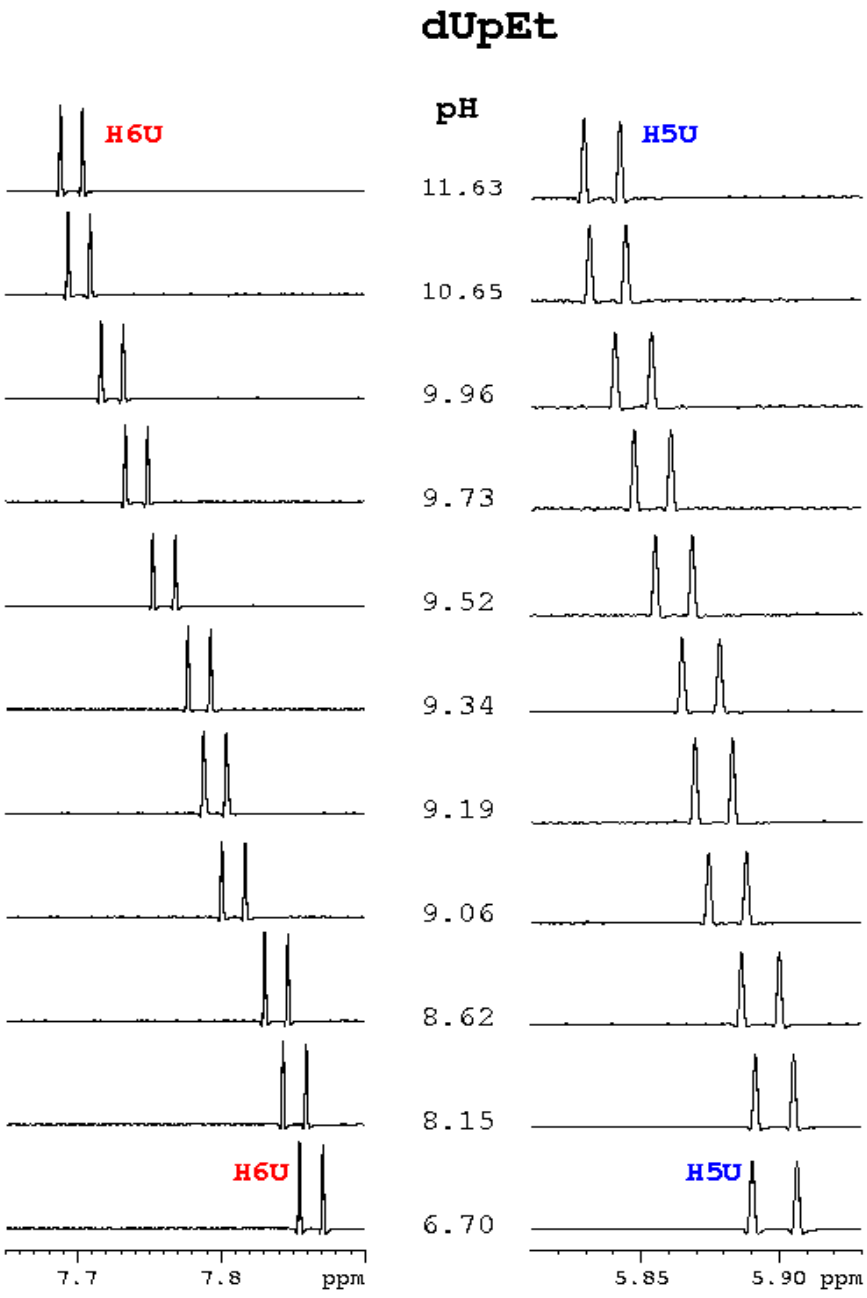
(B) pH-dependent ^{1}H chemical shift (in D_2O) of dCpEt (**2a**) at 298 K



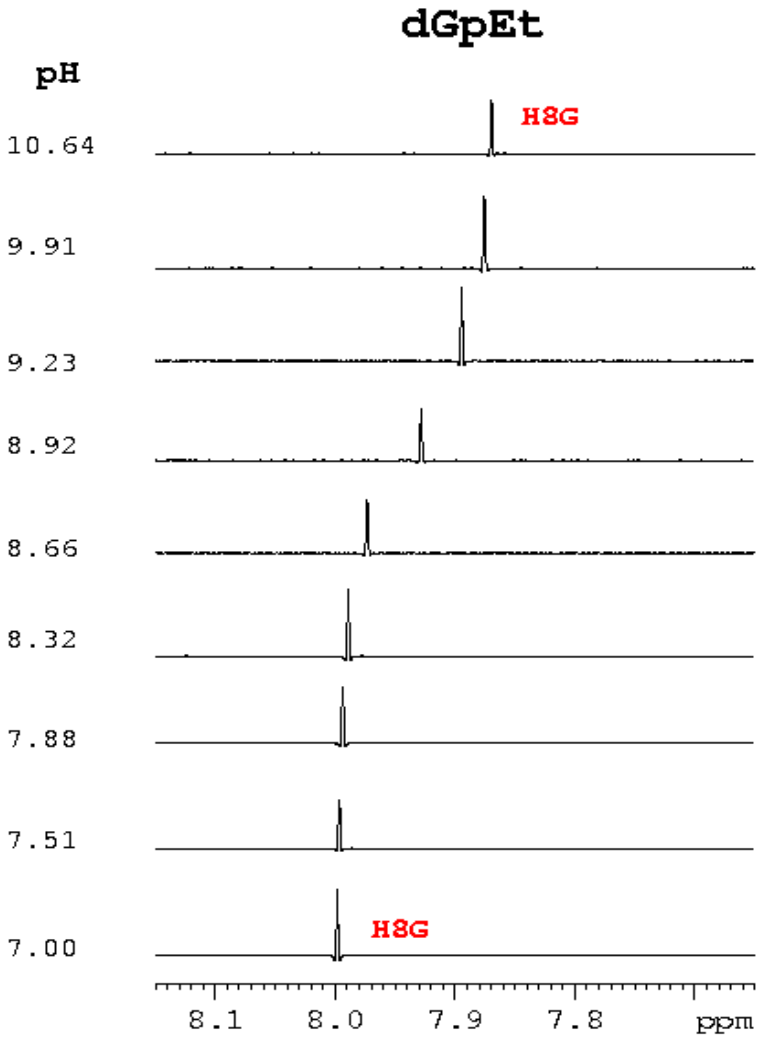
(C) pH-dependent ^{1}H chemical shift (in D_2O) of TpEt (**3a**) at 298 K



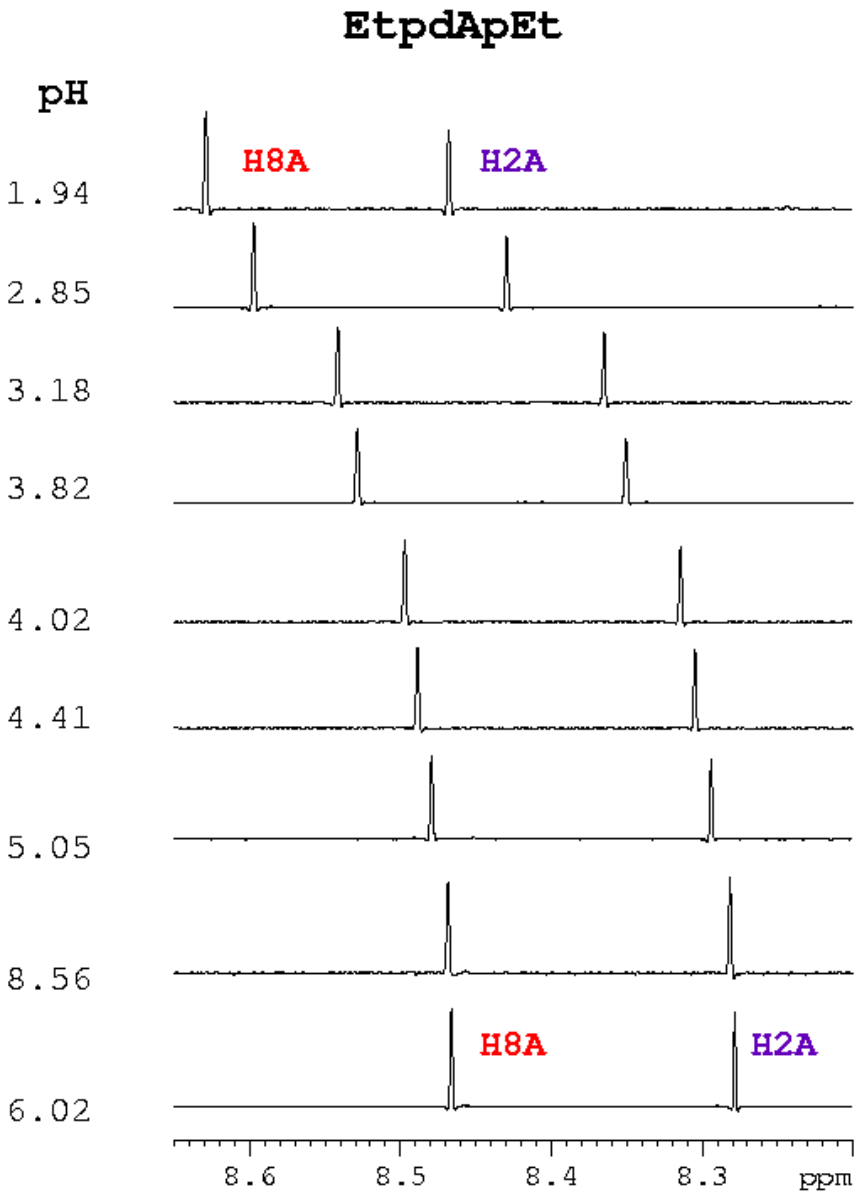
(D) pH-dependent ^{1}H chemical shift (in D_2O) of dUpEt (**4a**) at 298 K



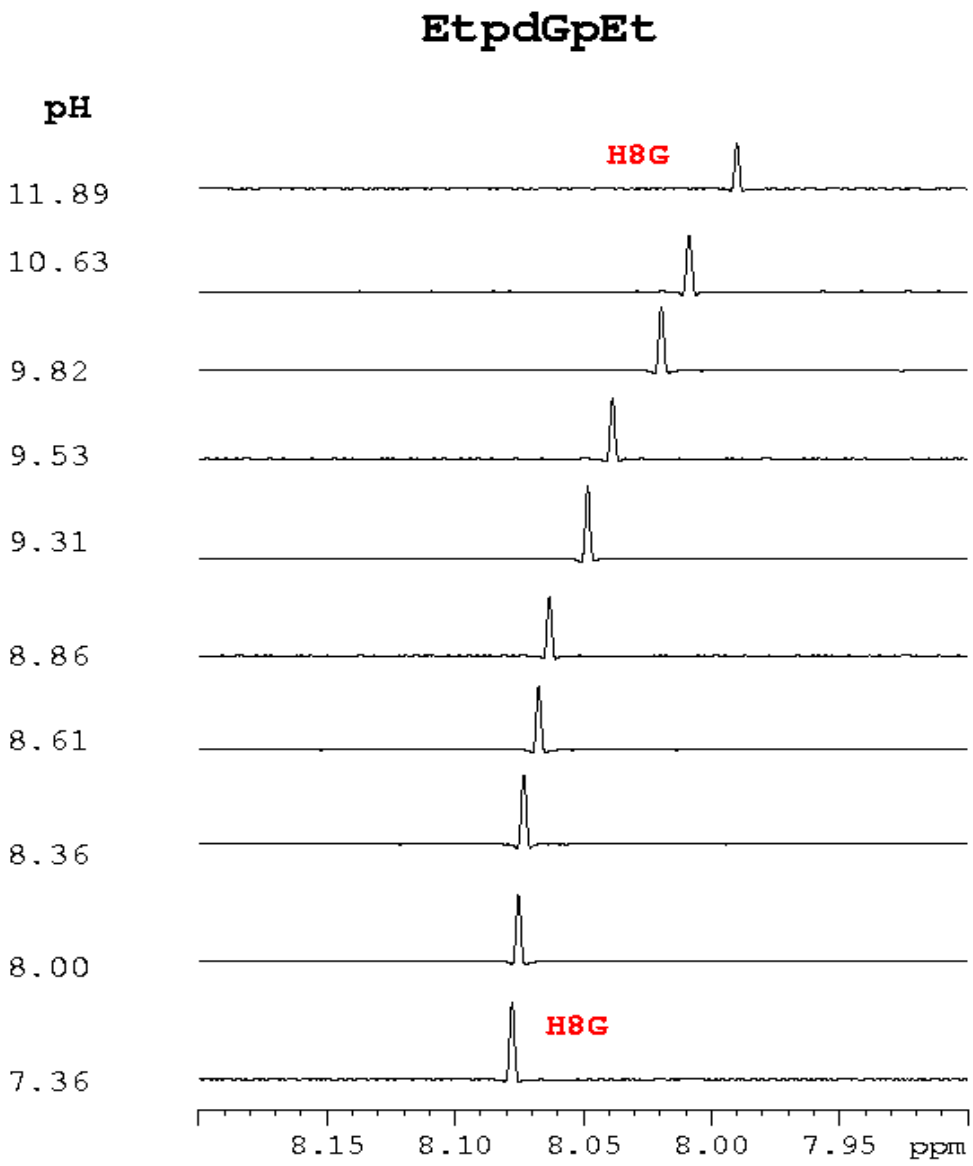
(E) pH-dependent ^{1}H chemical shift (in D_2O) of dGpEt (**5a**) at 298 K



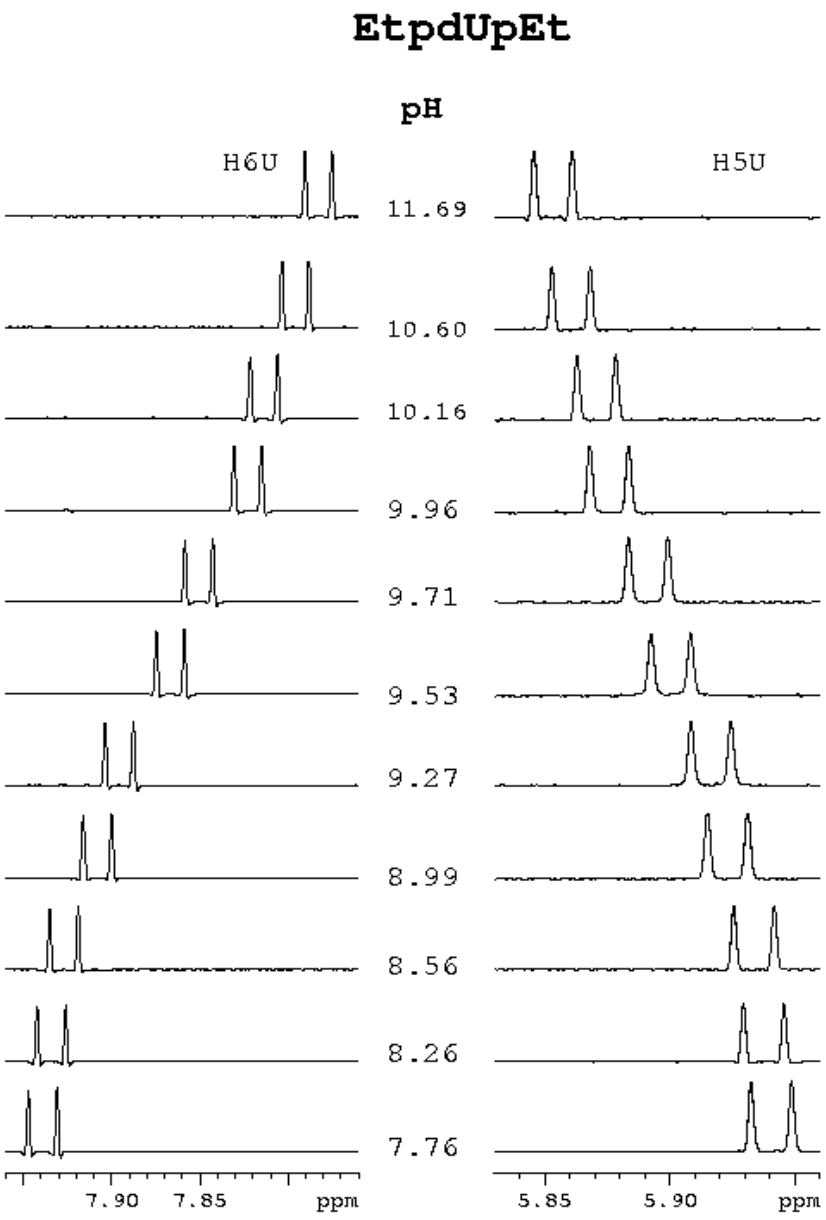
(F) pH-dependent ^{1}H chemical shift (in D_2O) of EtpdApEt (**6a**) at 298 K



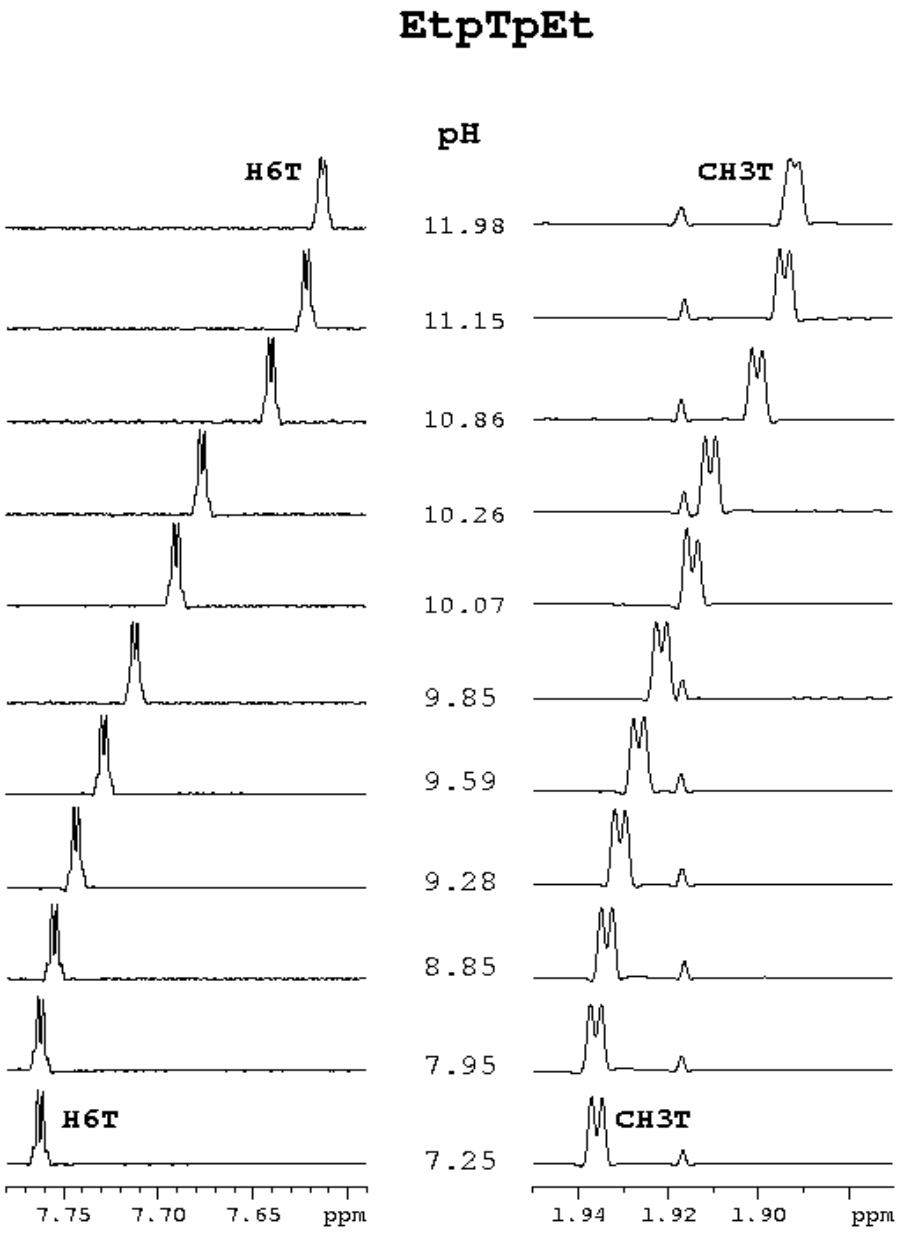
(G) pH-dependent ^{1}H chemical shift (in D_2O) of EtpdGpEt (**7a**) at 298 K



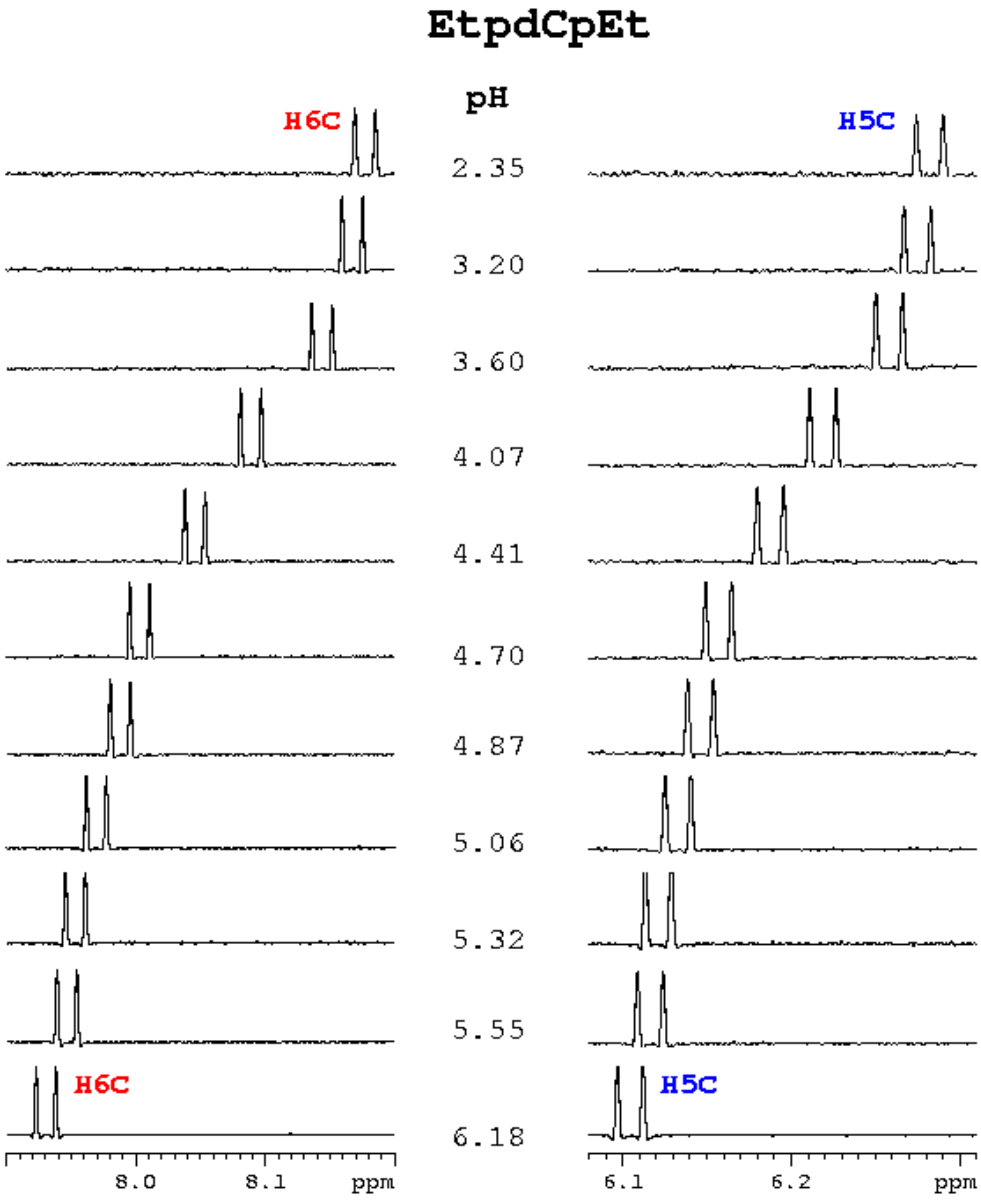
(H) pH-dependent ^{1}H chemical shift (in D_2O) of EtpdUpEt (**8a**) at 298 K



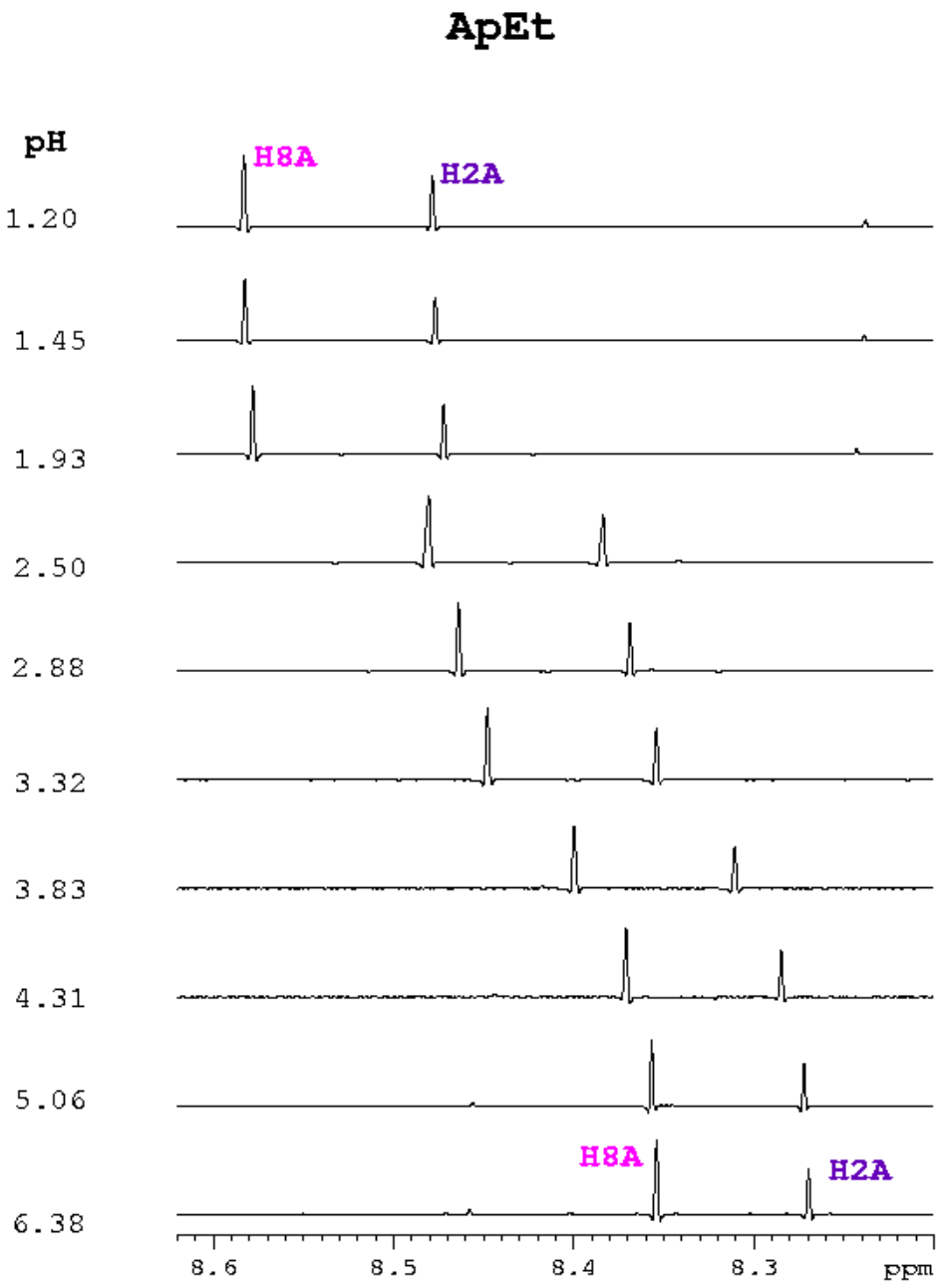
(I) pH-dependent ^{1}H chemical shift (in D_2O) of EtpTpEt (**9a**) at 298 K



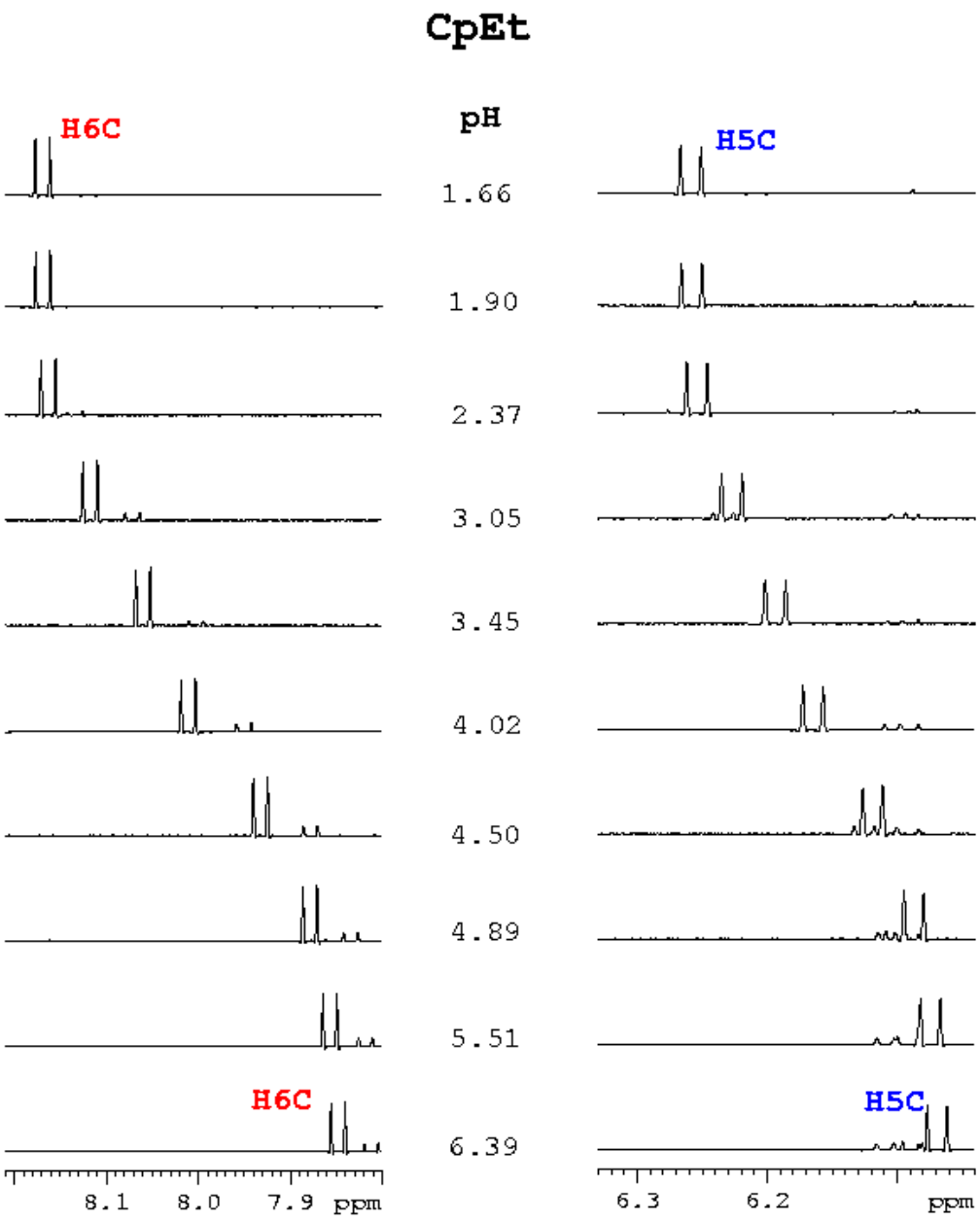
(J) pH-dependent ^{1}H chemical shift (in D_2O) of EtpdCpEt (**10a**) at 298 K



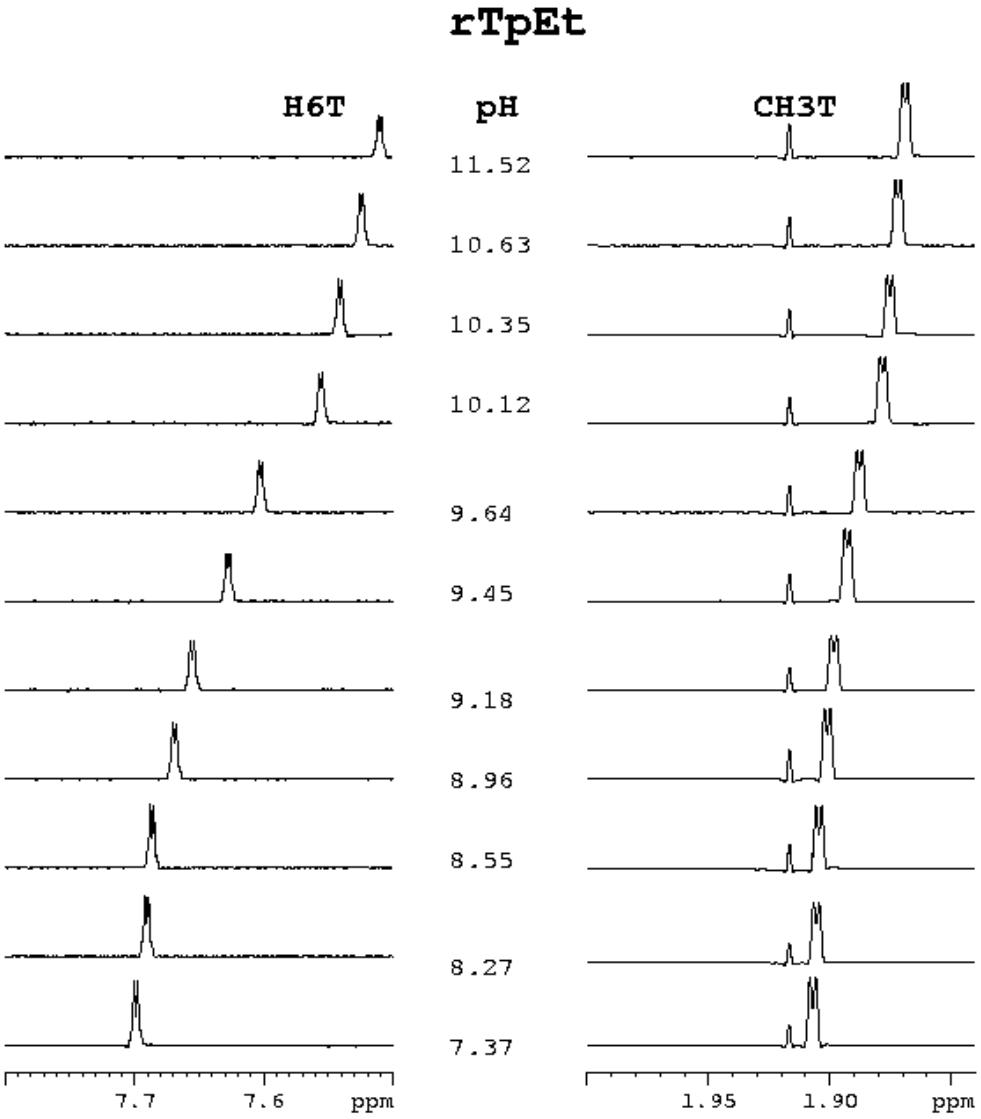
(K) pH-dependent ^{1}H chemical shift (in D_2O) of ApEt (**1b**) at 298 K



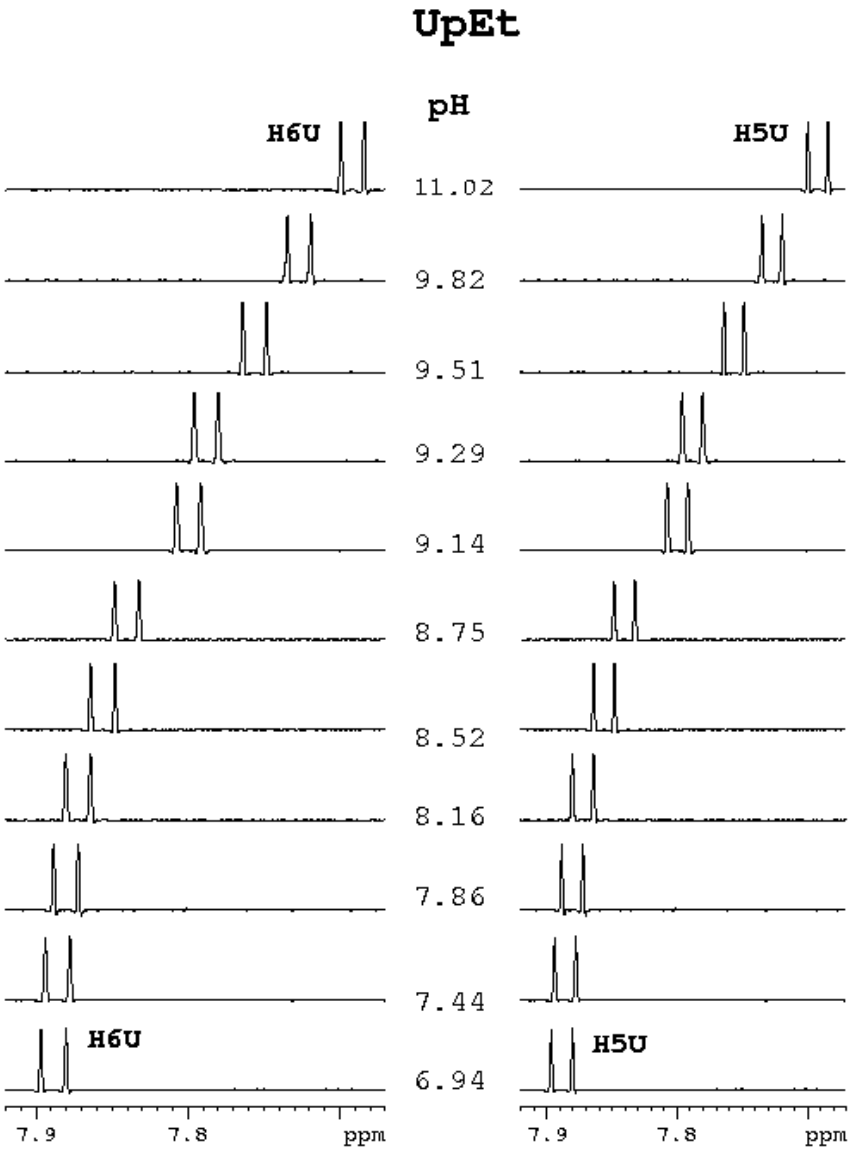
(L) pH-dependent ^{1}H chemical shift (in D_2O) of CpEt (**2b**) at 298 K



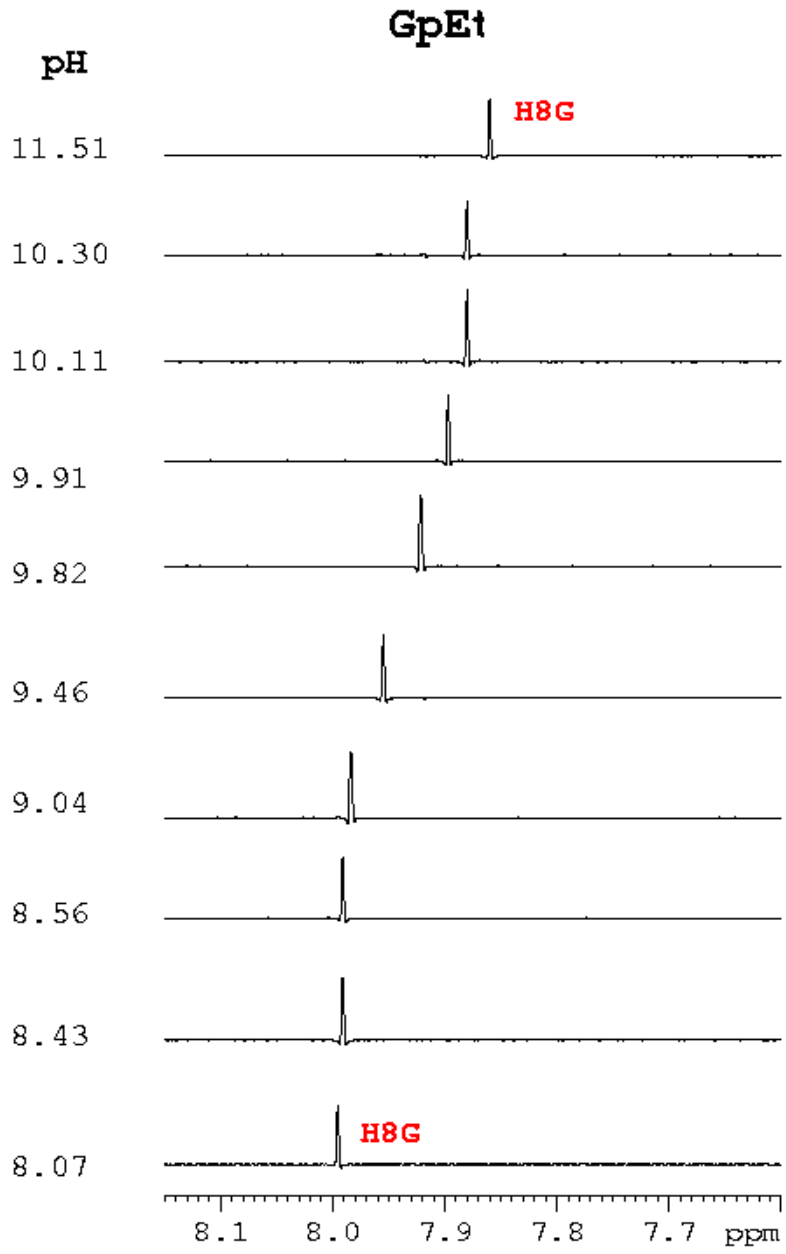
(M) pH-dependent ^{1}H chemical shift (in D_2O) of rTpEt (**3b**) at 298 K



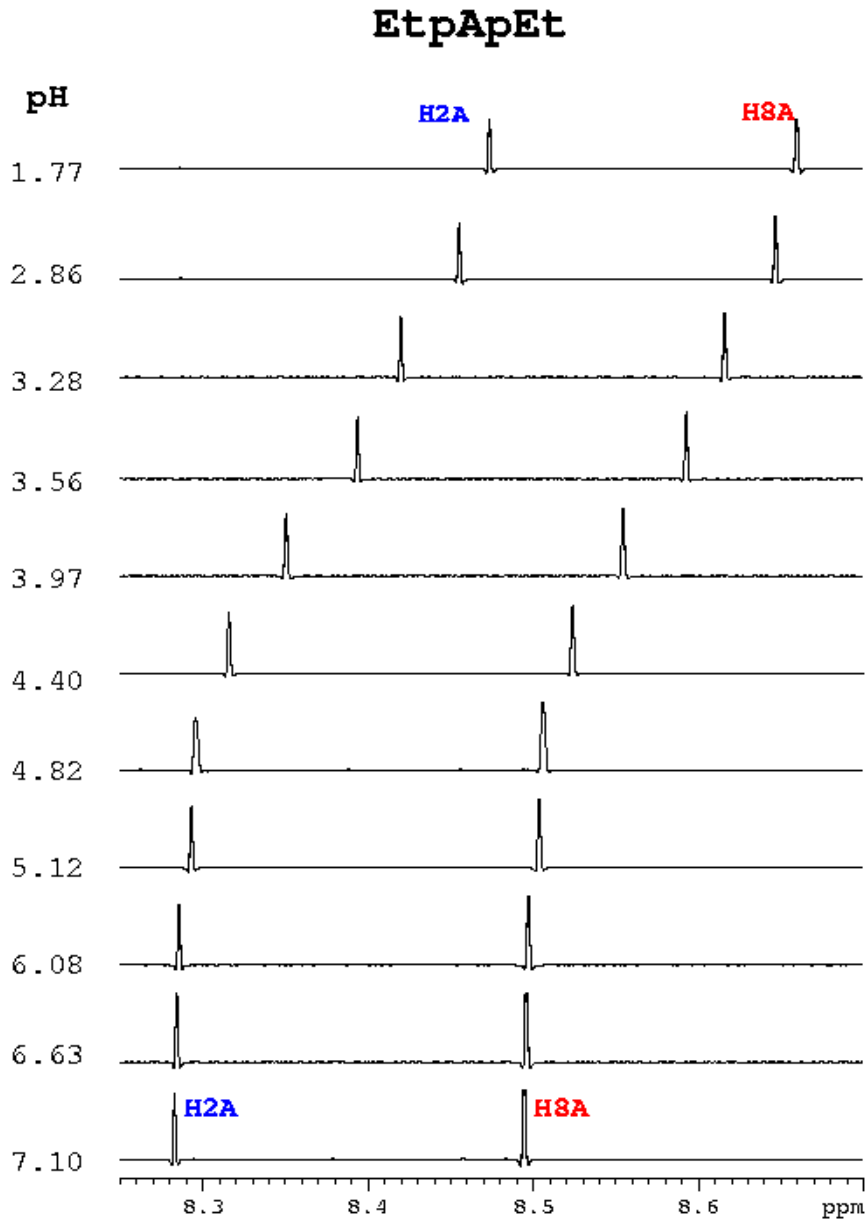
(N) pH-dependent ^{1}H chemical shift (in D_2O) of UpEt (**4b**) at 298 K



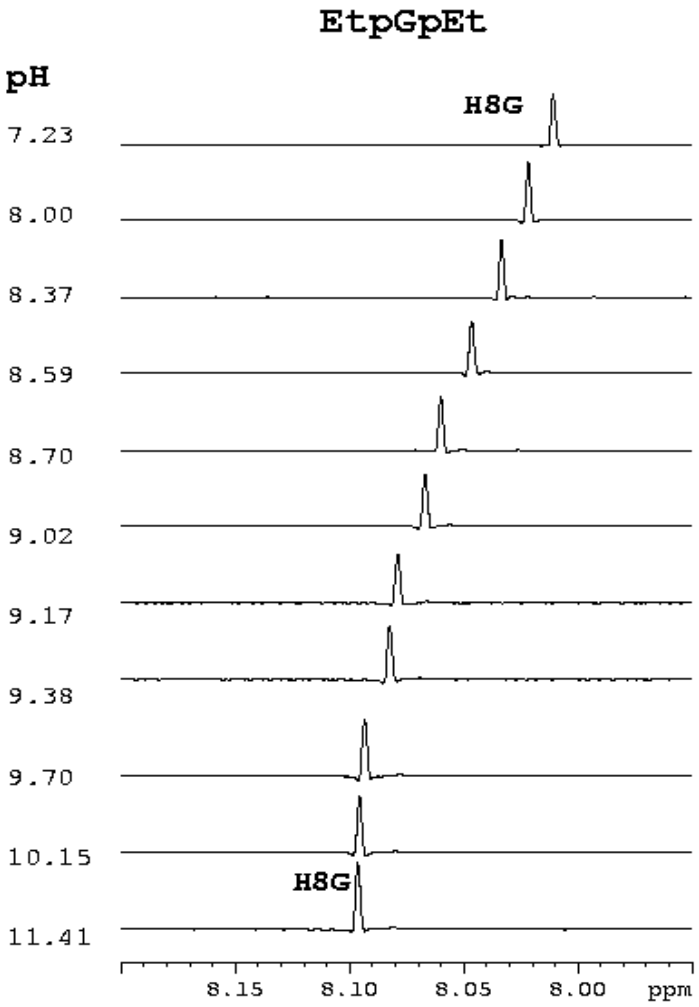
(O) pH-dependent ^{1}H chemical shift (in D_2O) of GpEt (**5b**) at 298 K



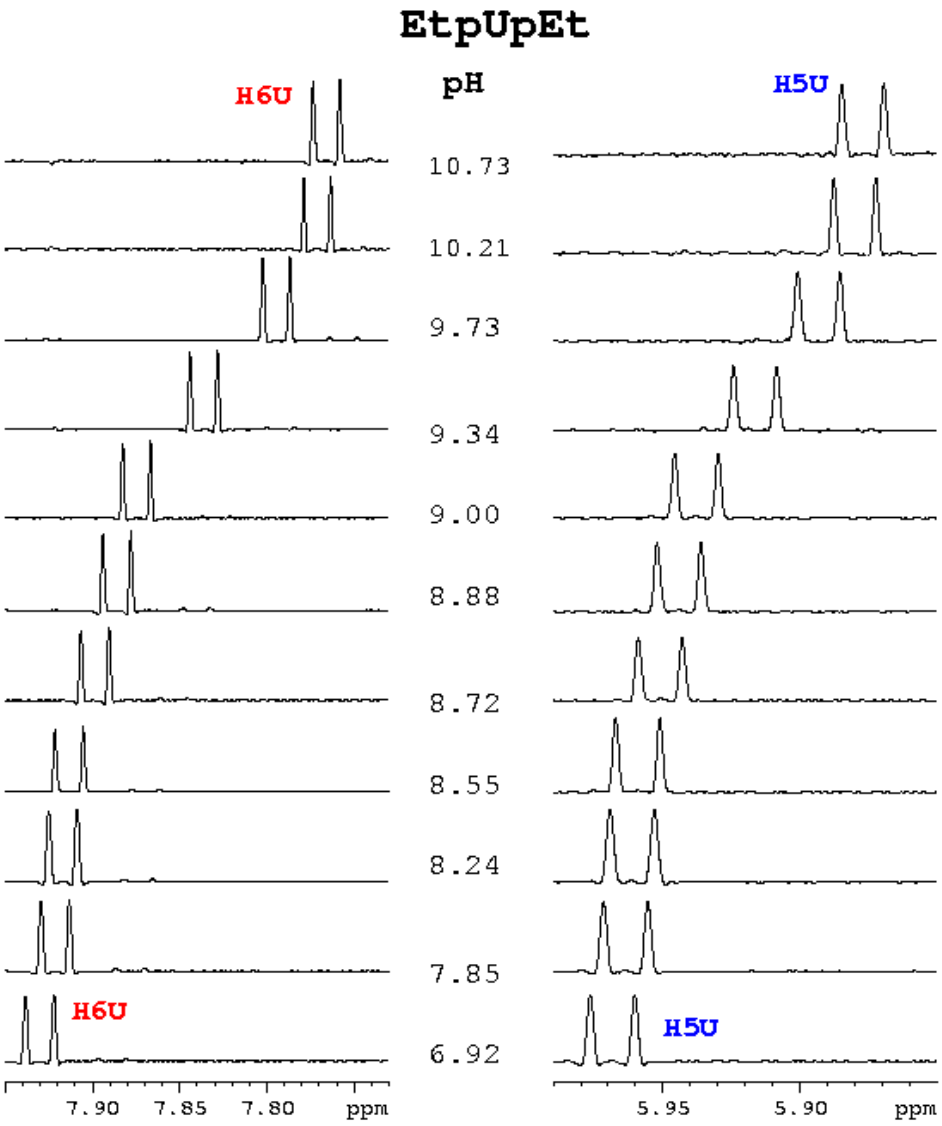
(P) pH-dependent ^{1}H chemical shift (in D_2O) of EtpApEt (**6b**) at 298 K



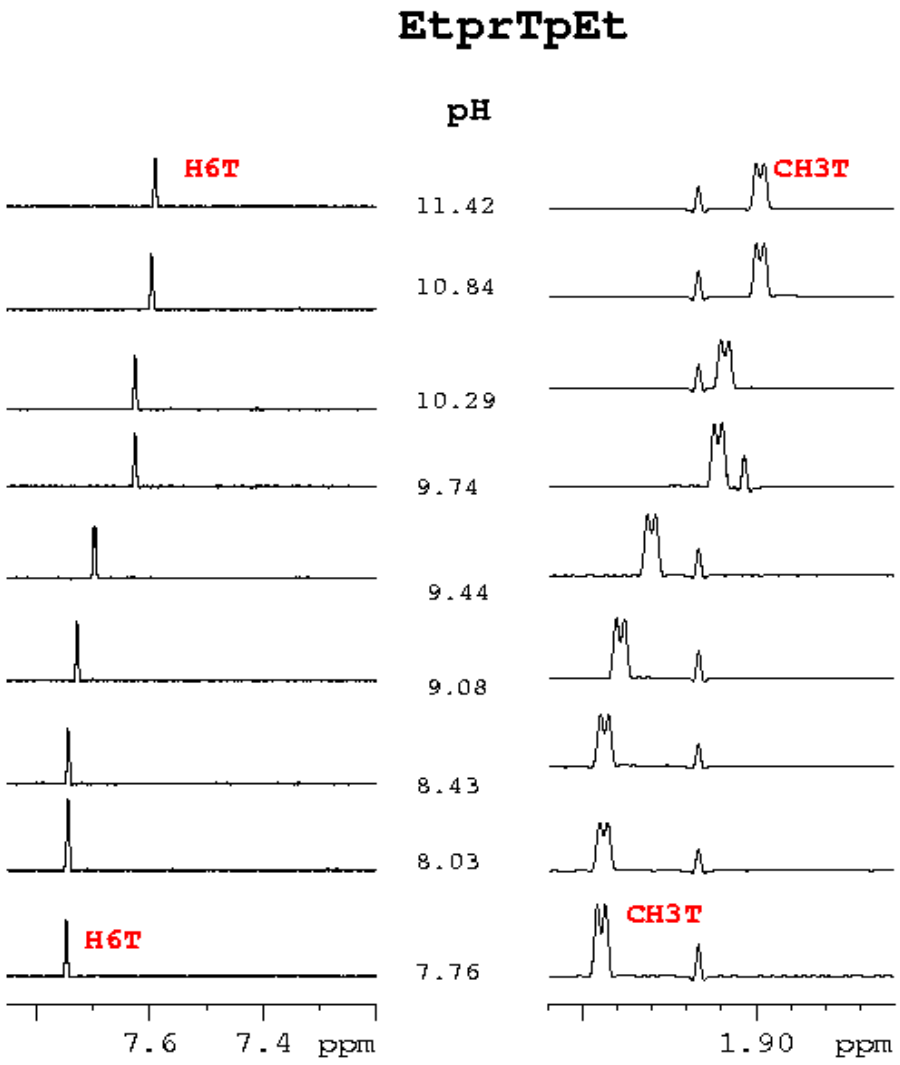
(Q) pH-dependent ^{1}H chemical shift (in D_2O) of EtpGpEt (**7b**) at 298 K



(R) pH-dependent ^{1}H chemical shift (in D_2O) of EtpUpEt (**8b**) at 298 K



(S) pH-dependent ^{1}H chemical shift (in D_2O) of EtprTpEt (**9b**) at 298 K



(T) pH-dependent ^{1}H chemical shift (in D_2O) of EtpCpEt (**10b**) at 298 K

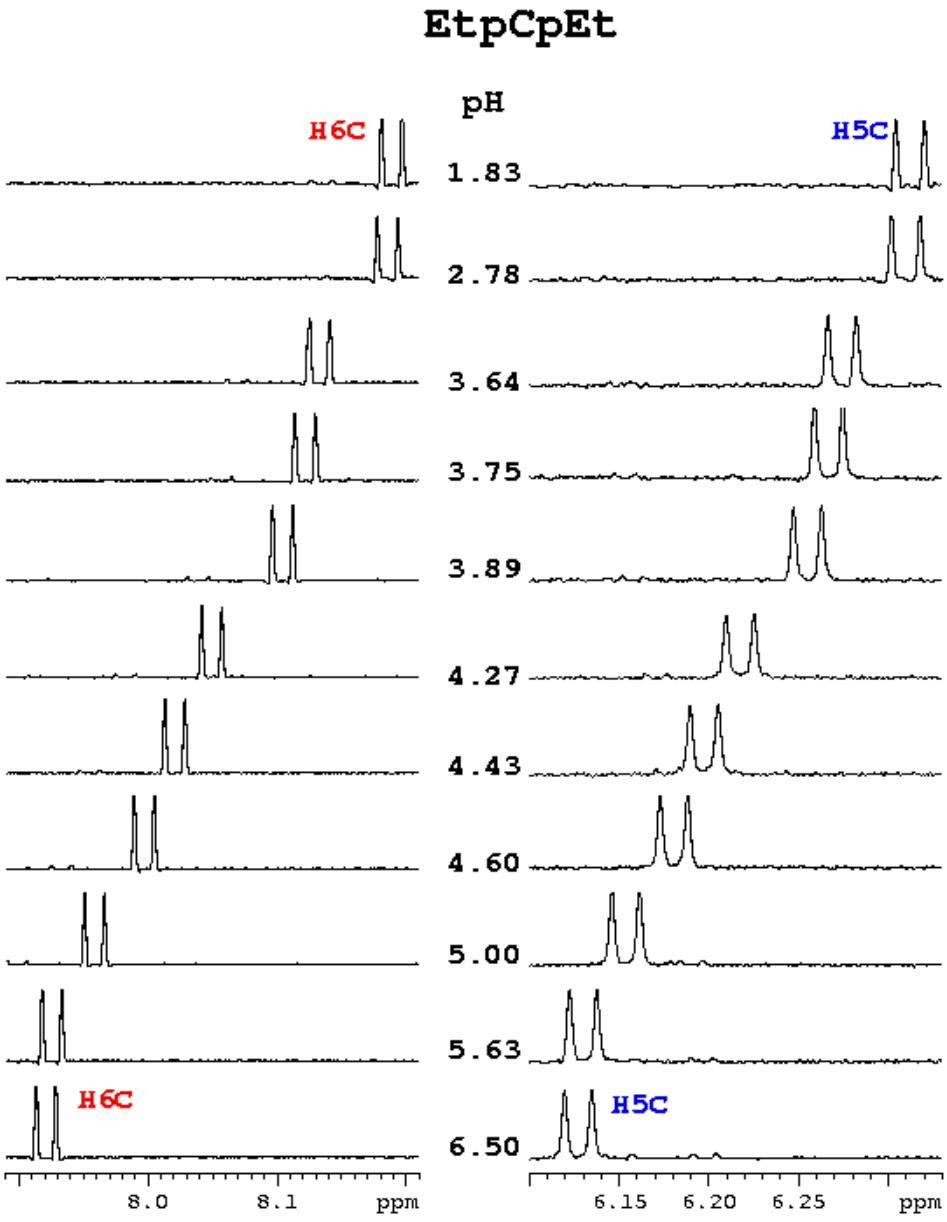
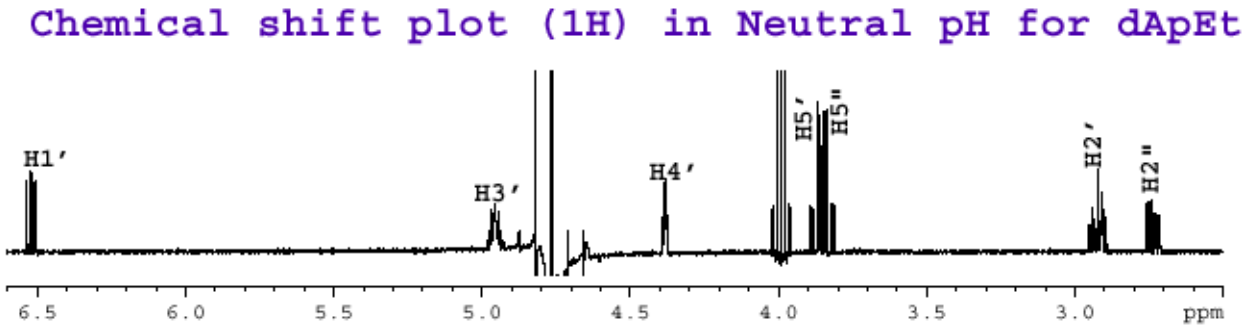
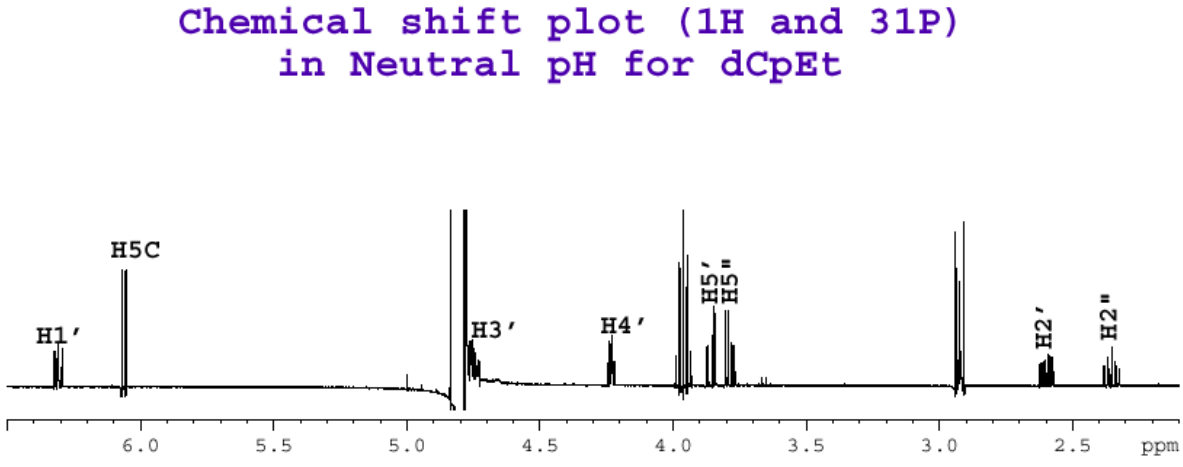


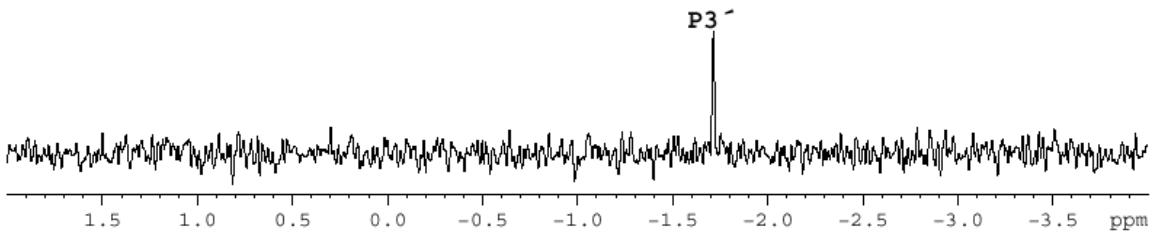
Figure S1. Panels (**A**) – (**T**) show stack plots of the pH-dependent ^{1}H NMR chemical shifts (in D_2O) of the aromatic protons for compounds **1a – 10a** and **1b** – **10b** (Figure 1) at 298 K. It is only 10 – 11 pHs (including two plateaus at two extreme pHs) those are shown out of total ~20 – 33 pHs (data points used for the titration plots) in panel (**A**) for dApEt (**1a**), panel (**B**) for dCpEt (**2a**), panel (**C**) for TpEt (**3a**), panel (**D**) for dUpEt (**4a**), panel (**E**) for dGpEt (**5a**), panel (**F**) for EtpdApEt (**6a**), panel (**G**) for EtpdGpEt (**7a**), panel (**H**) for EtpdUpEt (**8a**), panel (**I**) for EtpTpEt (**9a**), panel (**J**) for EtpdCpEt (**10a**), panel (**K**) for rApEt (**1b**), panel (**L**) for rCpEt (**2b**), panel (**M**) for rTpEt (**3b**), panel (**N**) for rUpEt (**4b**), panel (**O**) for rGpEt (**5b**), panel (**P**) for EtprApEt (**6b**), panel (**Q**) for EtprGpEt (**7b**), panel (**R**) for EtprUpEt (**9b**), panel (**S**) for EtprTpEt (**9b**) and panel (**T**) for EtprCpEt (**10b**). All NMR measurements have been performed in 500 MHz and/or 600 MHz spectrometers with δ_{DSS} = 0.015 ppm as internal standard. The pH values are corrected with the deuterium effect (see experimental section for details).



1a.1 1H spectrum

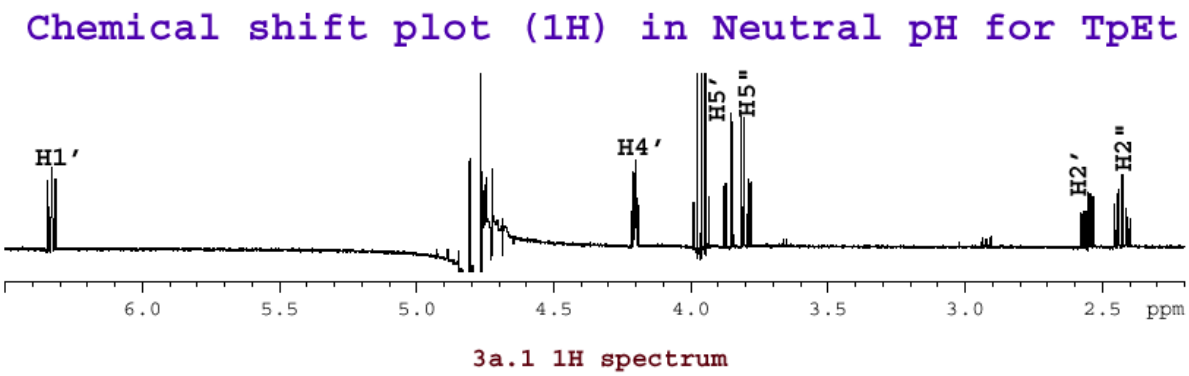


2a.1 1H spectrum

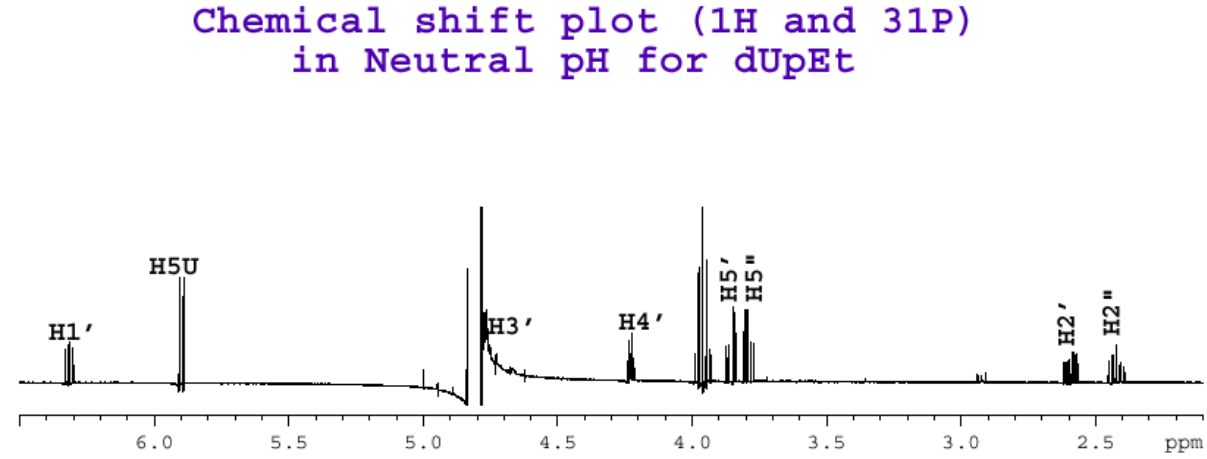


2a.2 31P spectrum

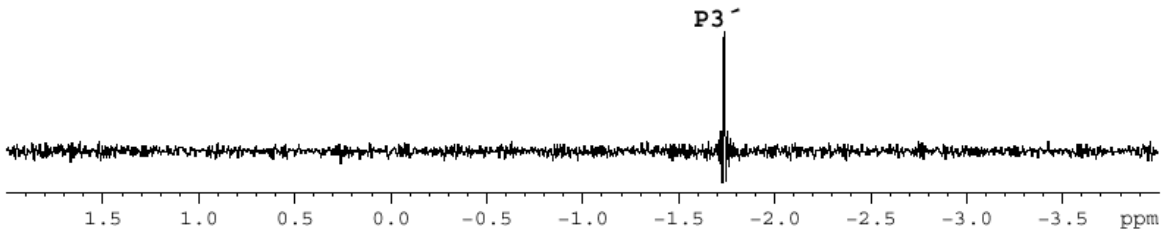
Figure S2 contd.



3a.1 1H spectrum



4a.1 1H spectrum



4a.2 31P spectrum

Figure S2 contd.

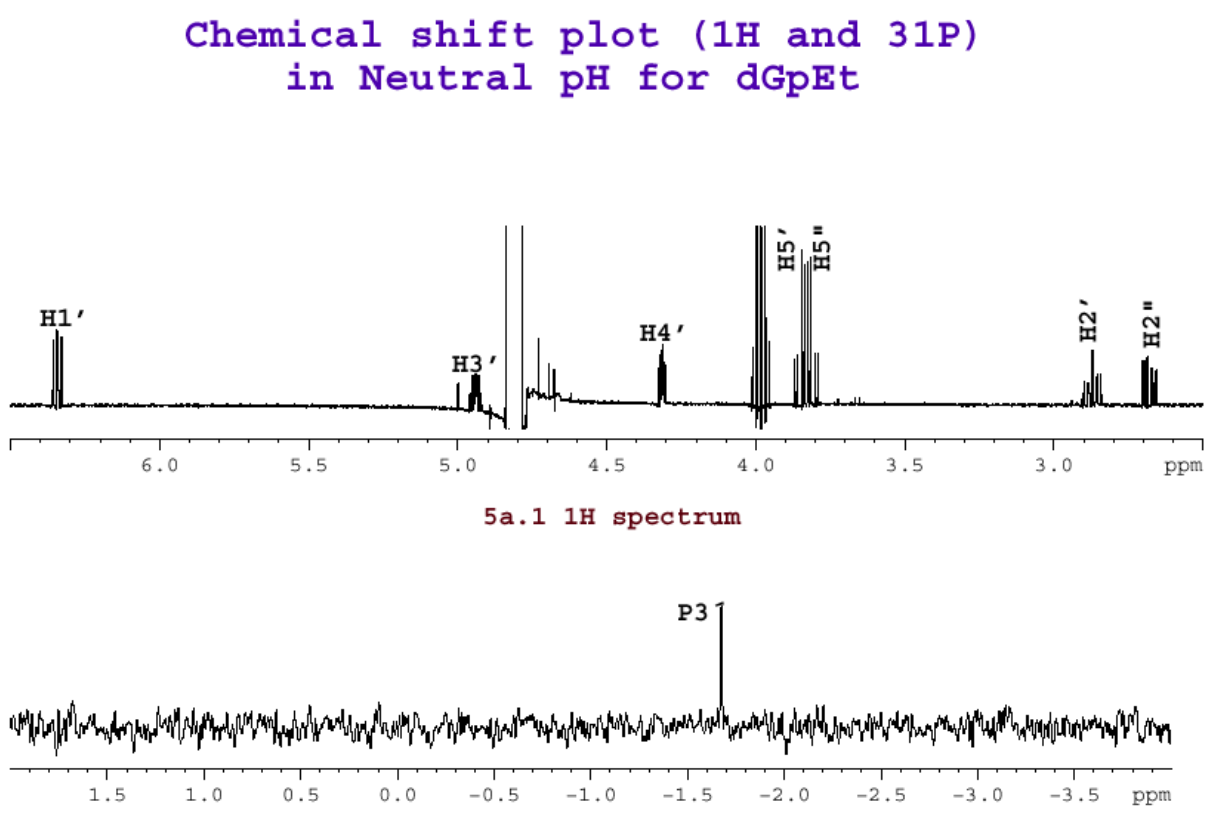


Figure S2 contd.

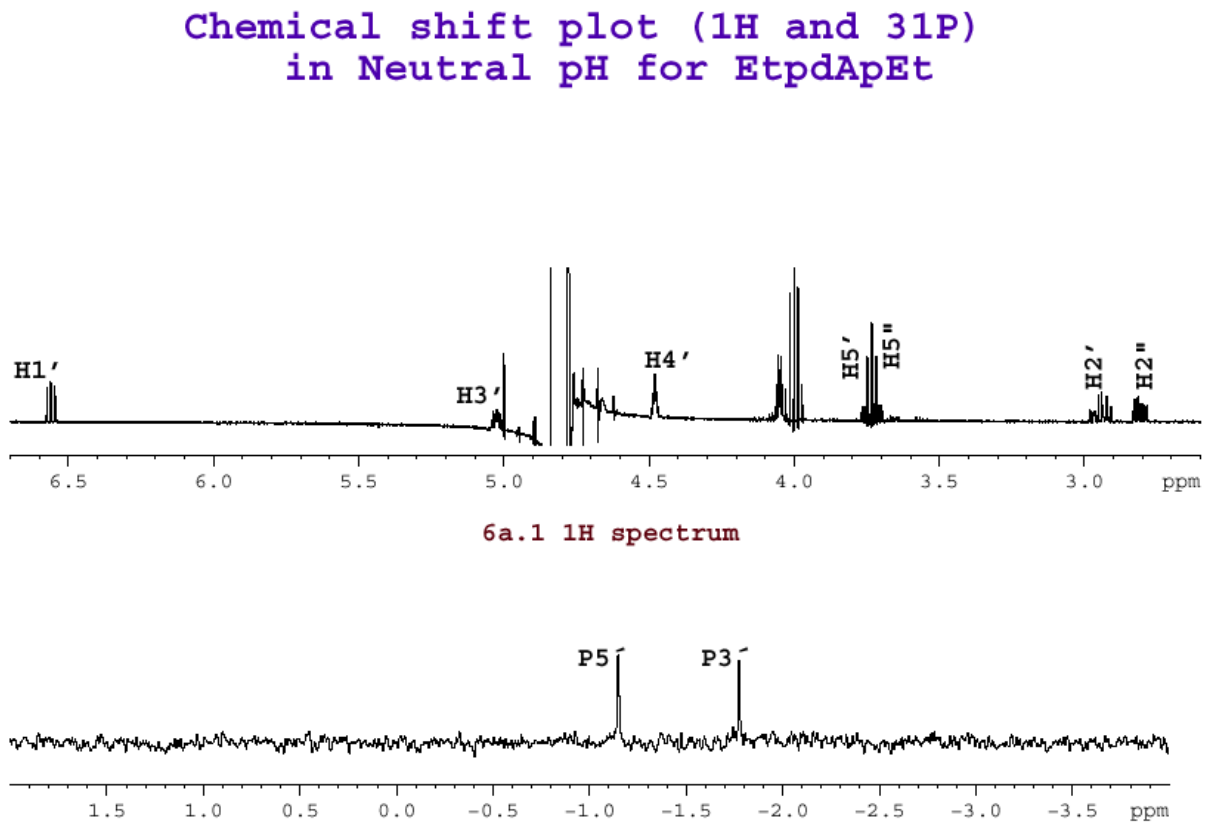


Figure S2 contd.

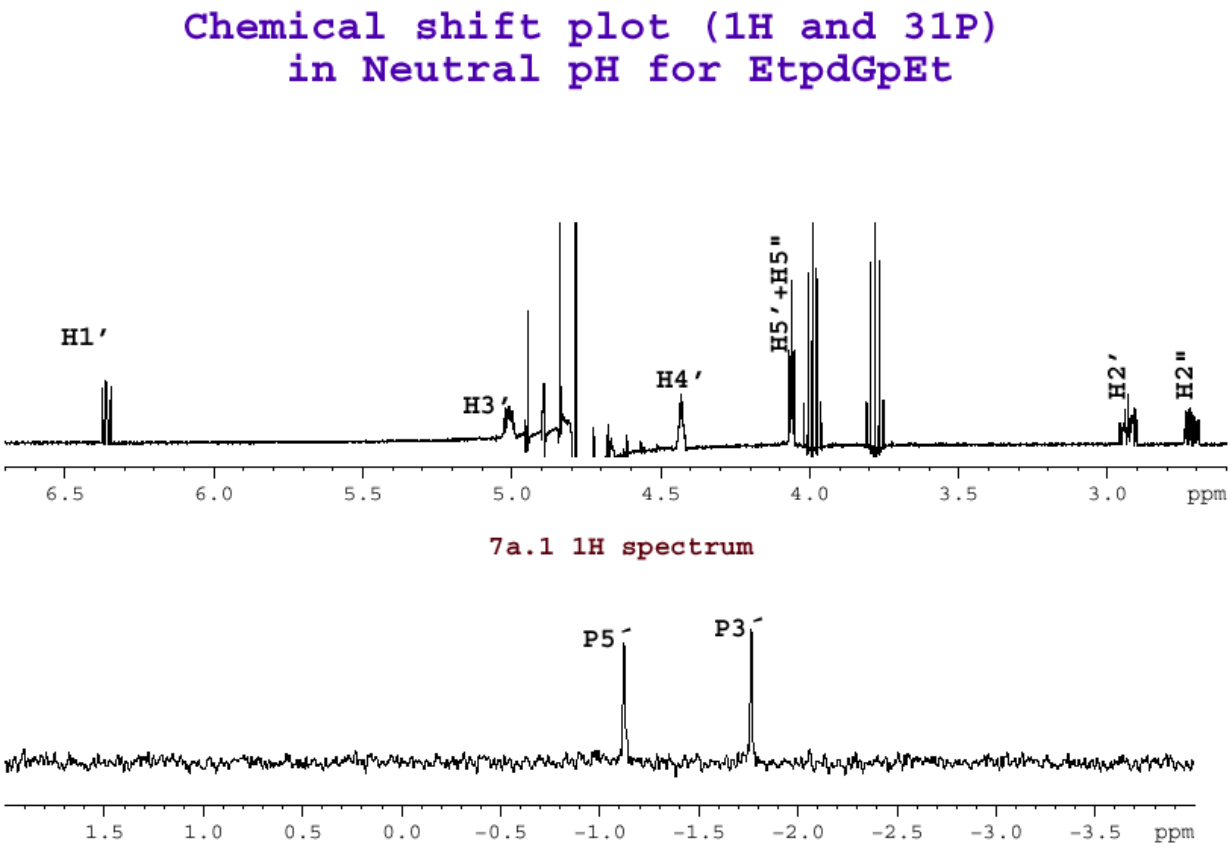


Figure S2 contd.

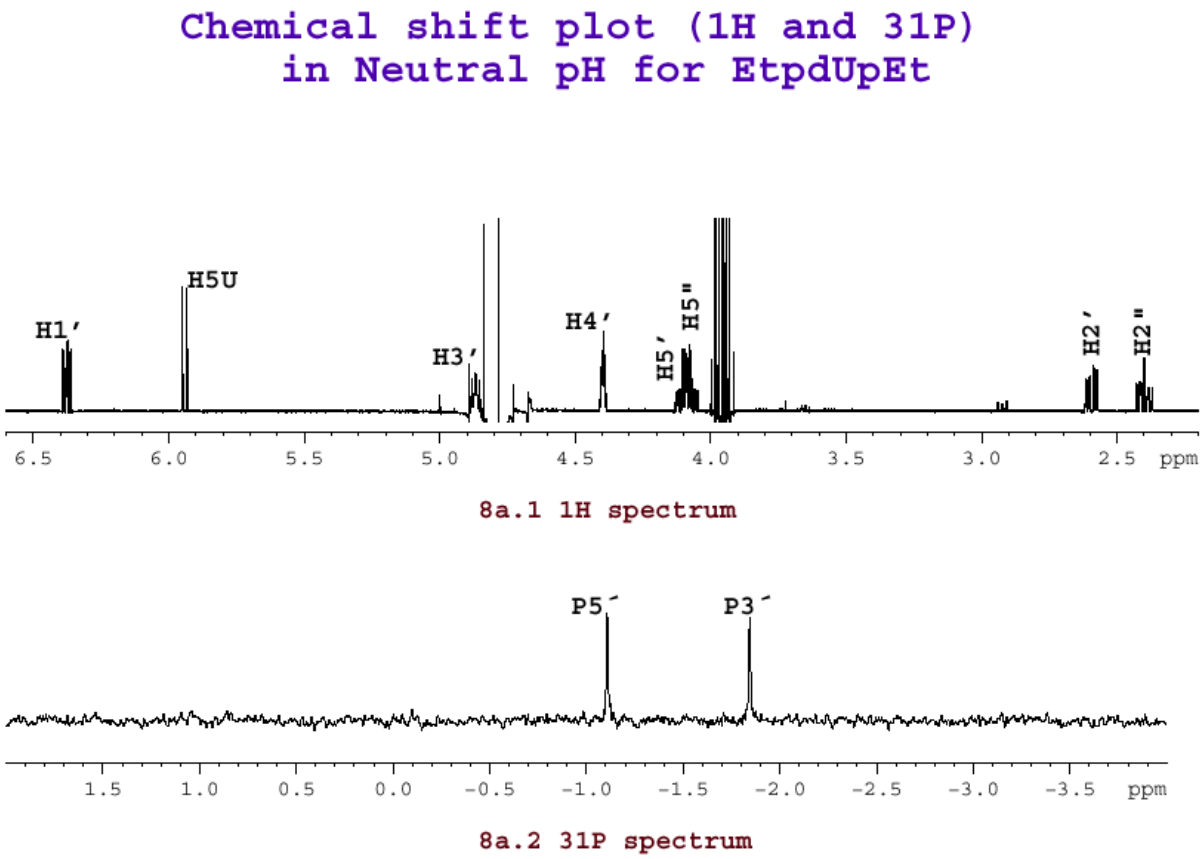


Figure S2 contd.

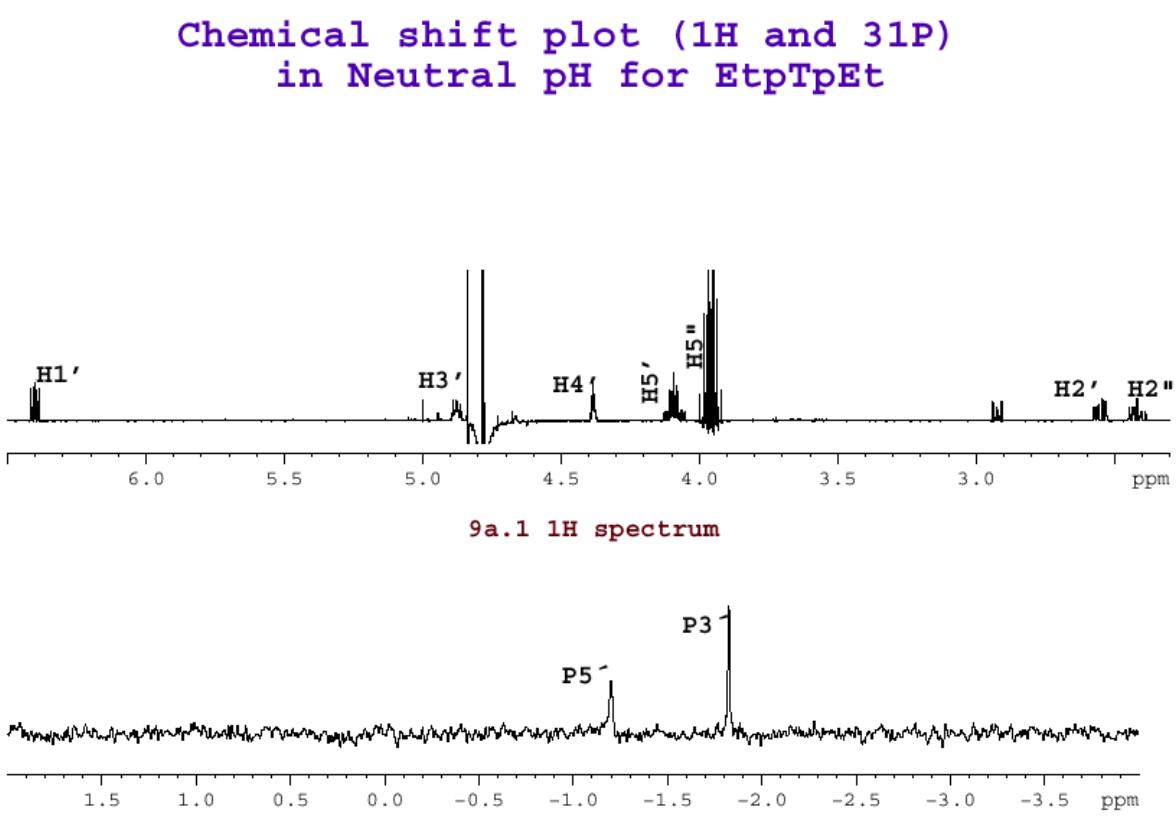
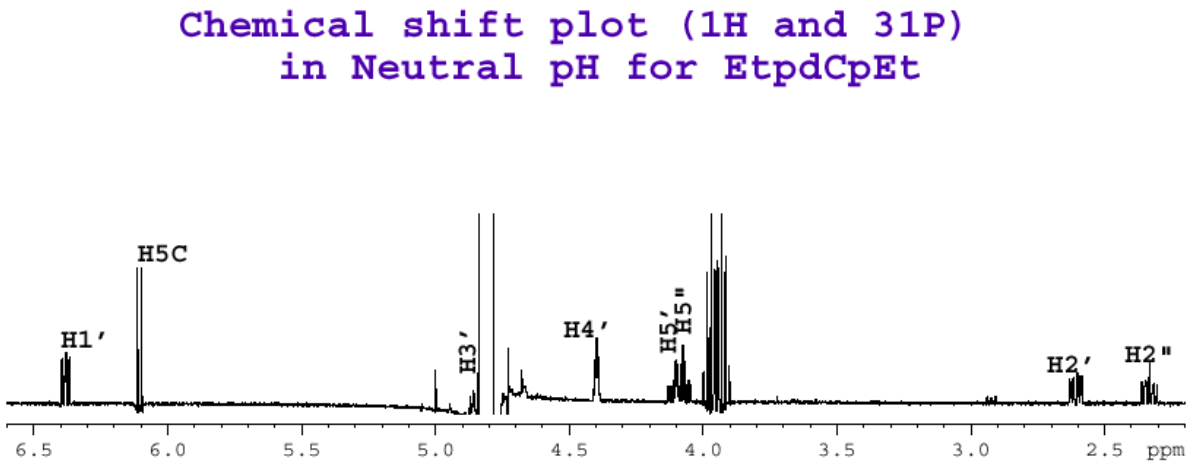
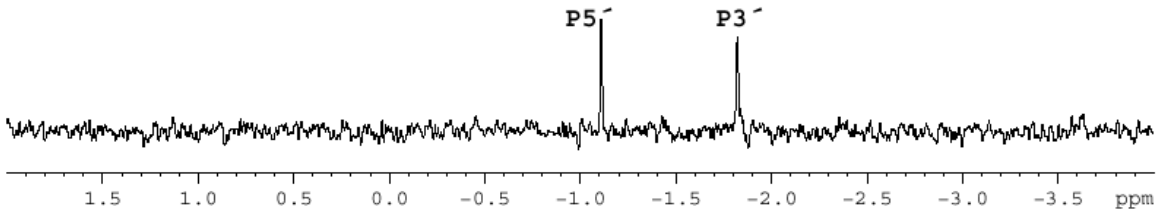


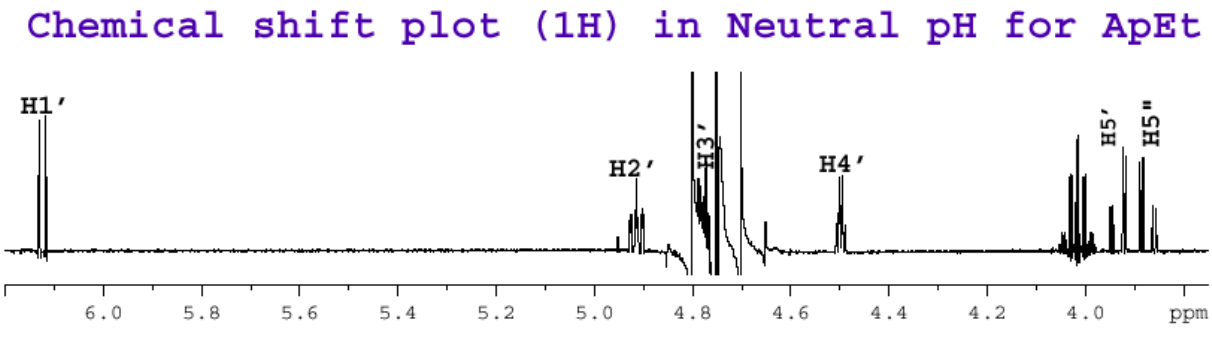
Figure S2 contd.



10a.1 1H spectrum

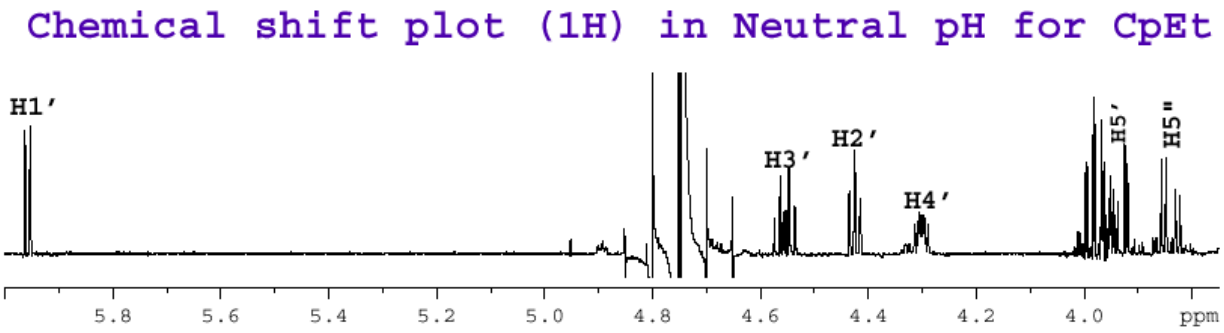


10a.2 31P spectrum

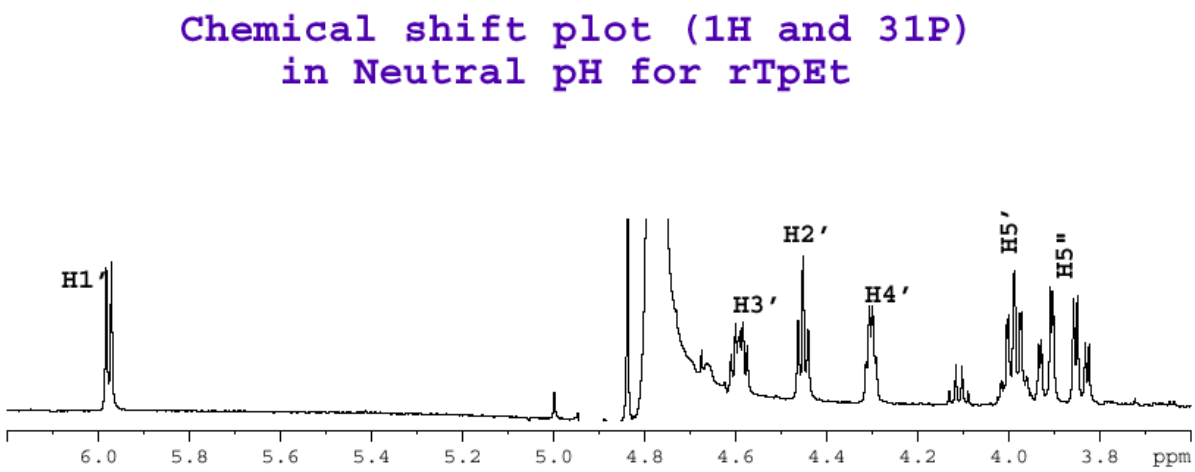


1b.1 1H spectrum

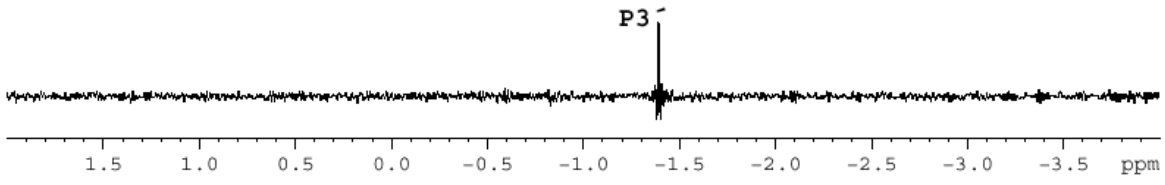
Figure S2 contd.



2b.1 1H spectrum

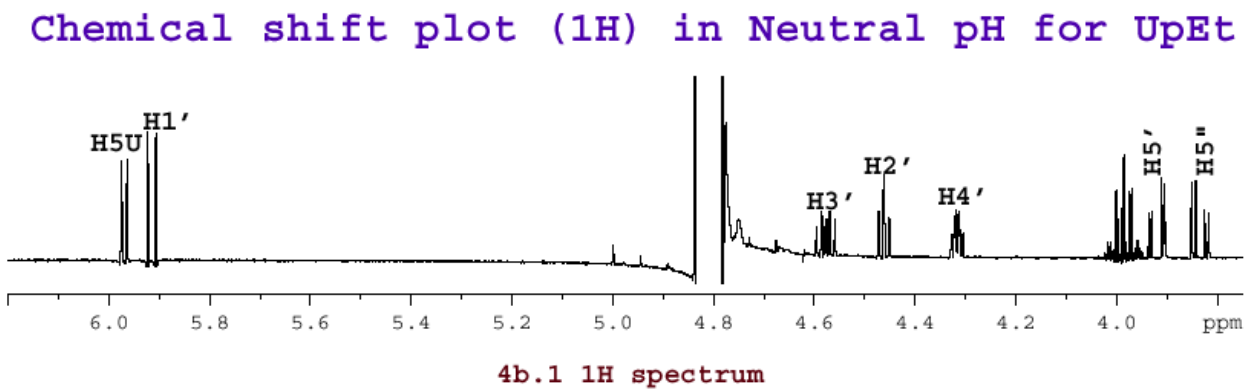


3b.1 1H spectrum

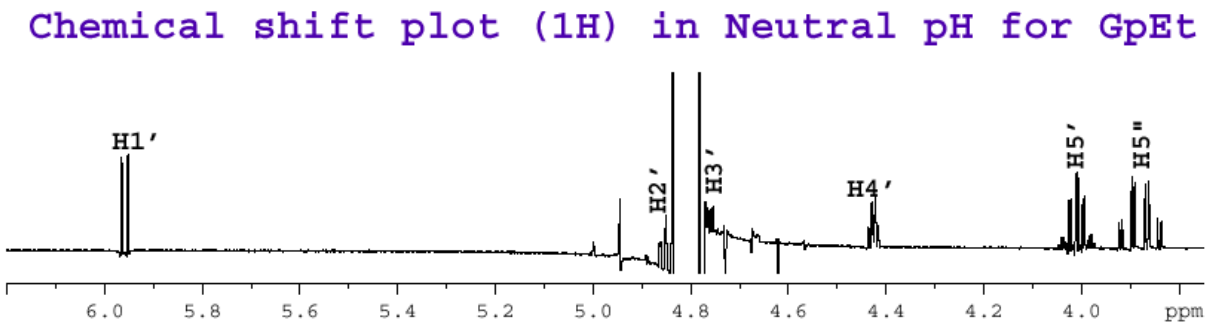


3b.2 31P spectrum

Figure S2 contd.

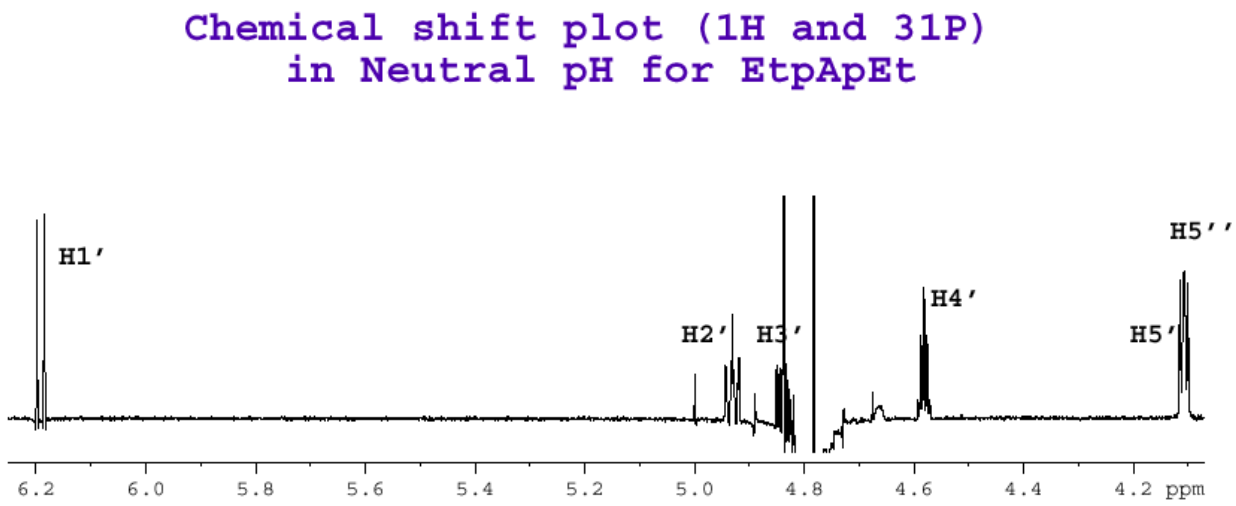


4b.1 1H spectrum

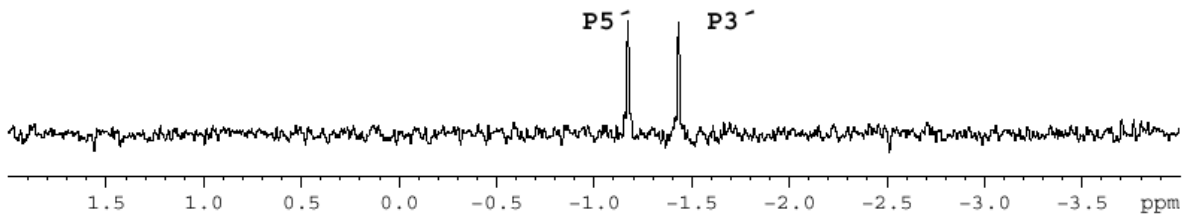


5b.1 1H spectrum

Figure S2 contd.



6b.1 1H spectrum



6b.2 31P spectrum

Figure S2 contd.

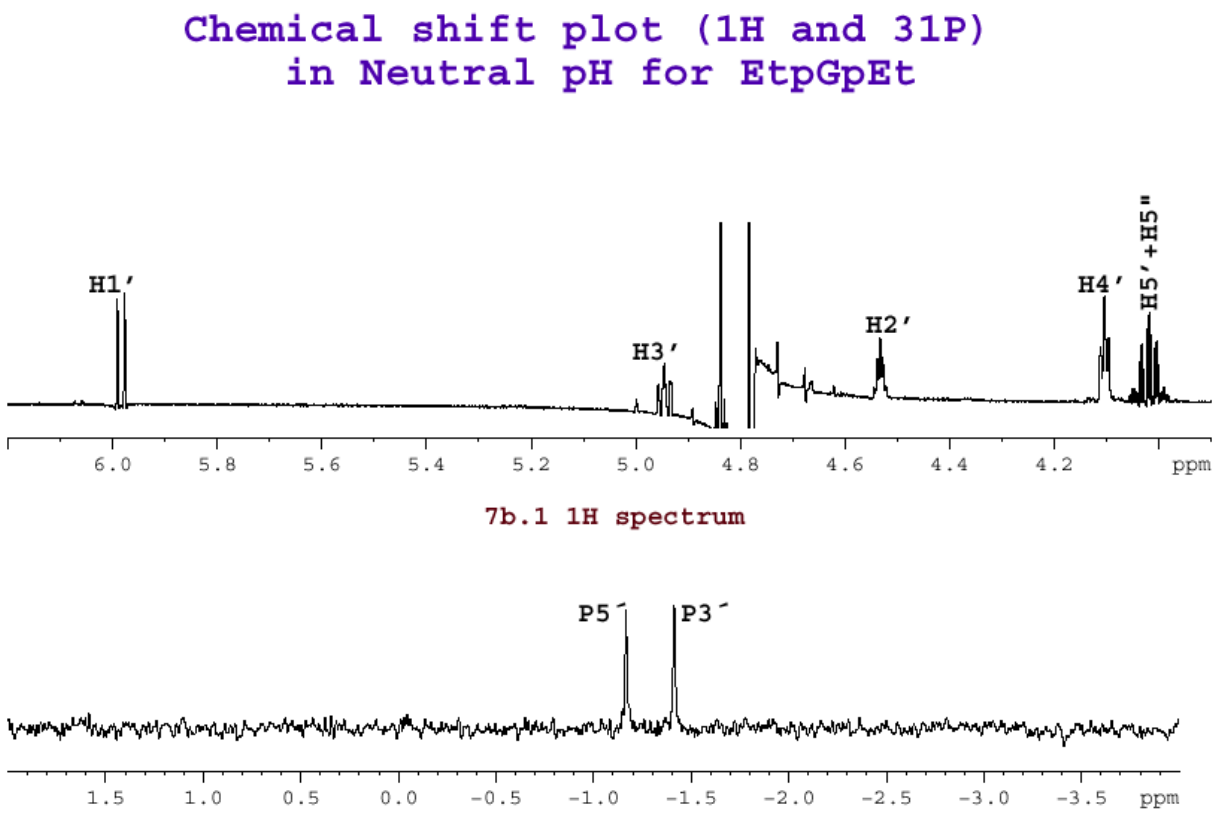


Figure S2 contd.

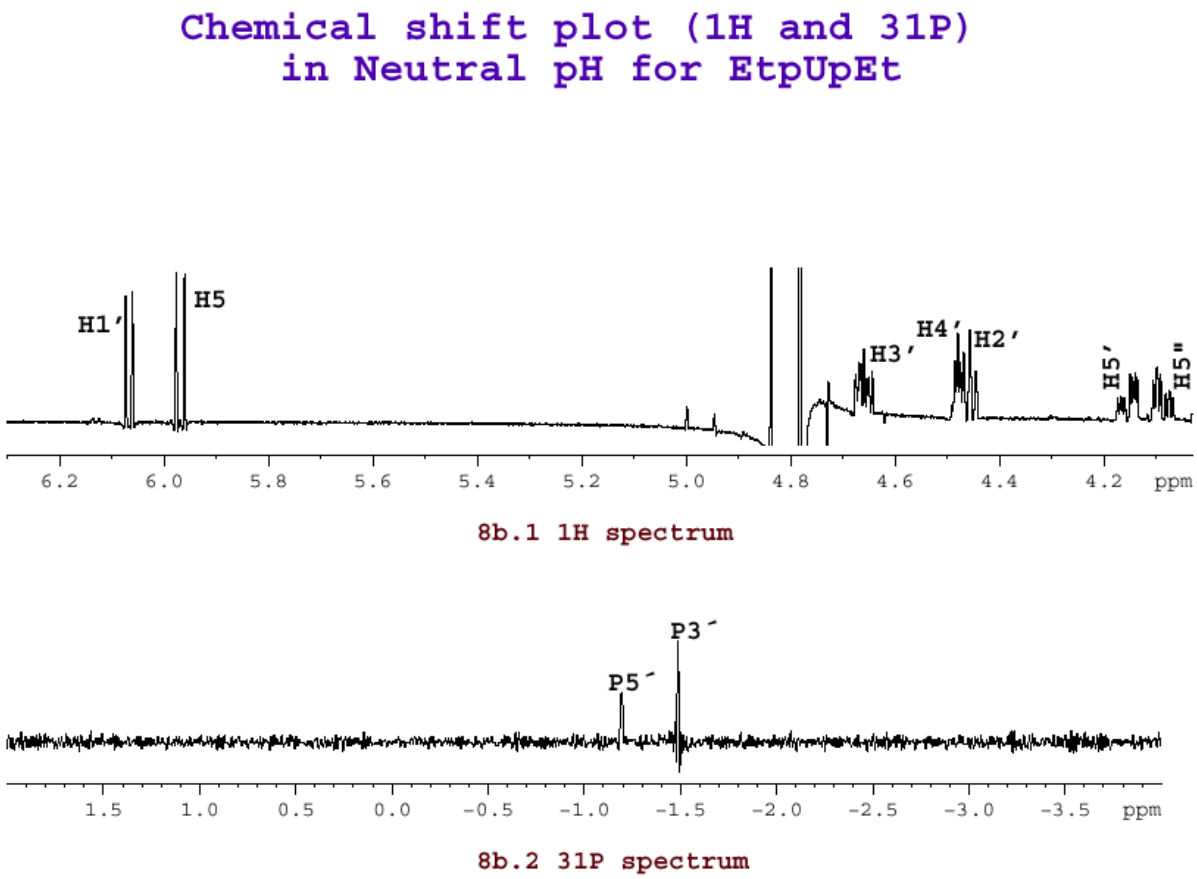


Figure S2 contd.

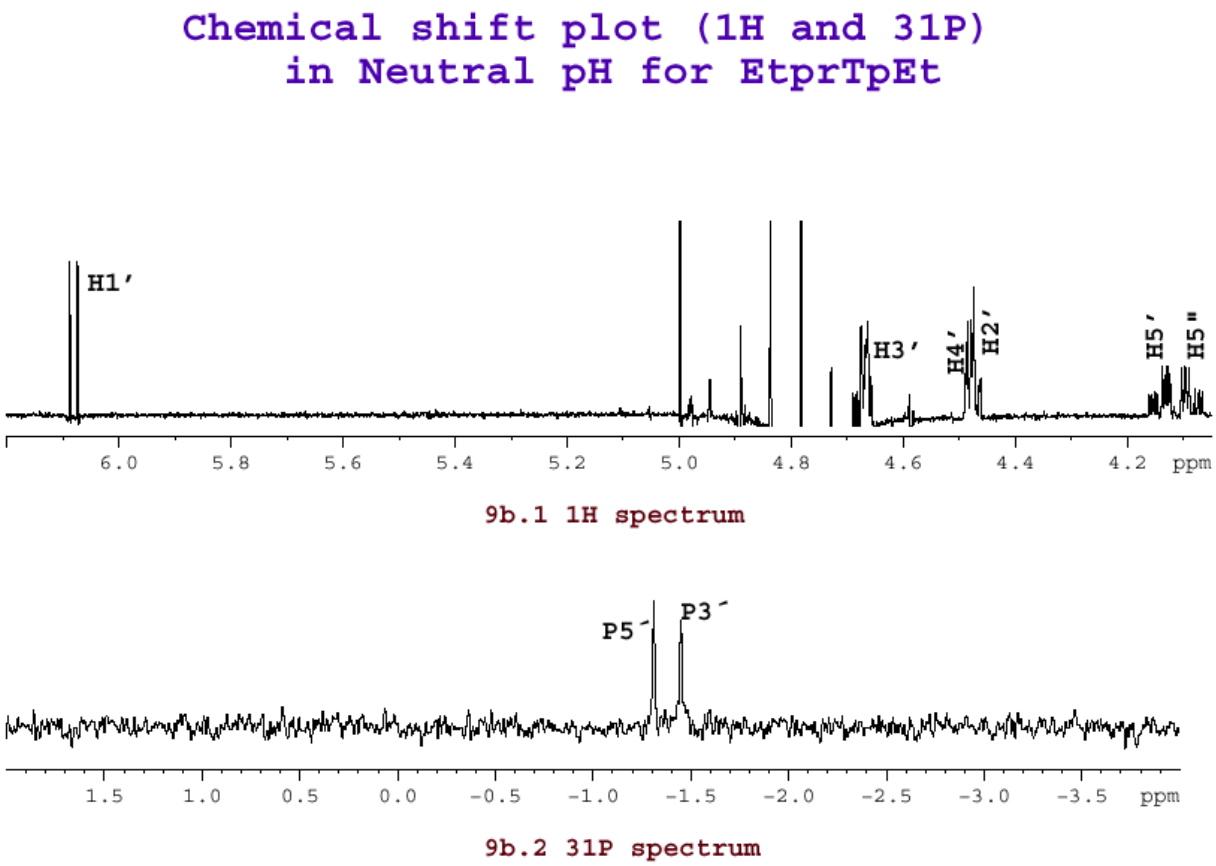


Figure S2 contd.

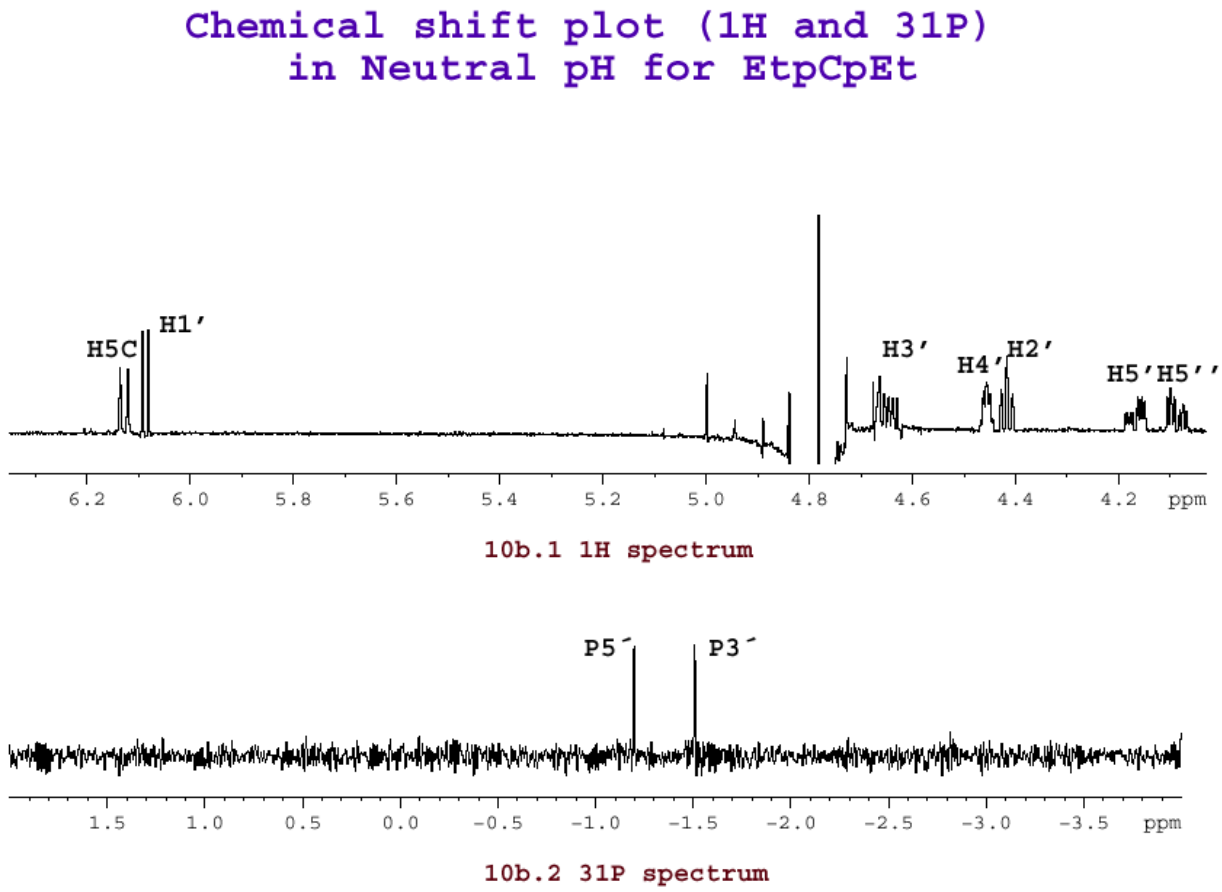


Figure S2. Panels **1a.1** – **10b.2** show the ^{1}H NMR chemical shifts at the neutral pH of the sugar protons [~ 6.5 ppm δ_H ~ 2.0 ppm] and ^{31}P NMR chemical shifts [~ -4.0 ppm δ_{31P} ~ 1.5 ppm] for compounds **1a – 10a** and **1b** – **10b** at 298 K. The type of the spectrum for each compound with the corresponding panel number is given at the bottom of each spectrum.

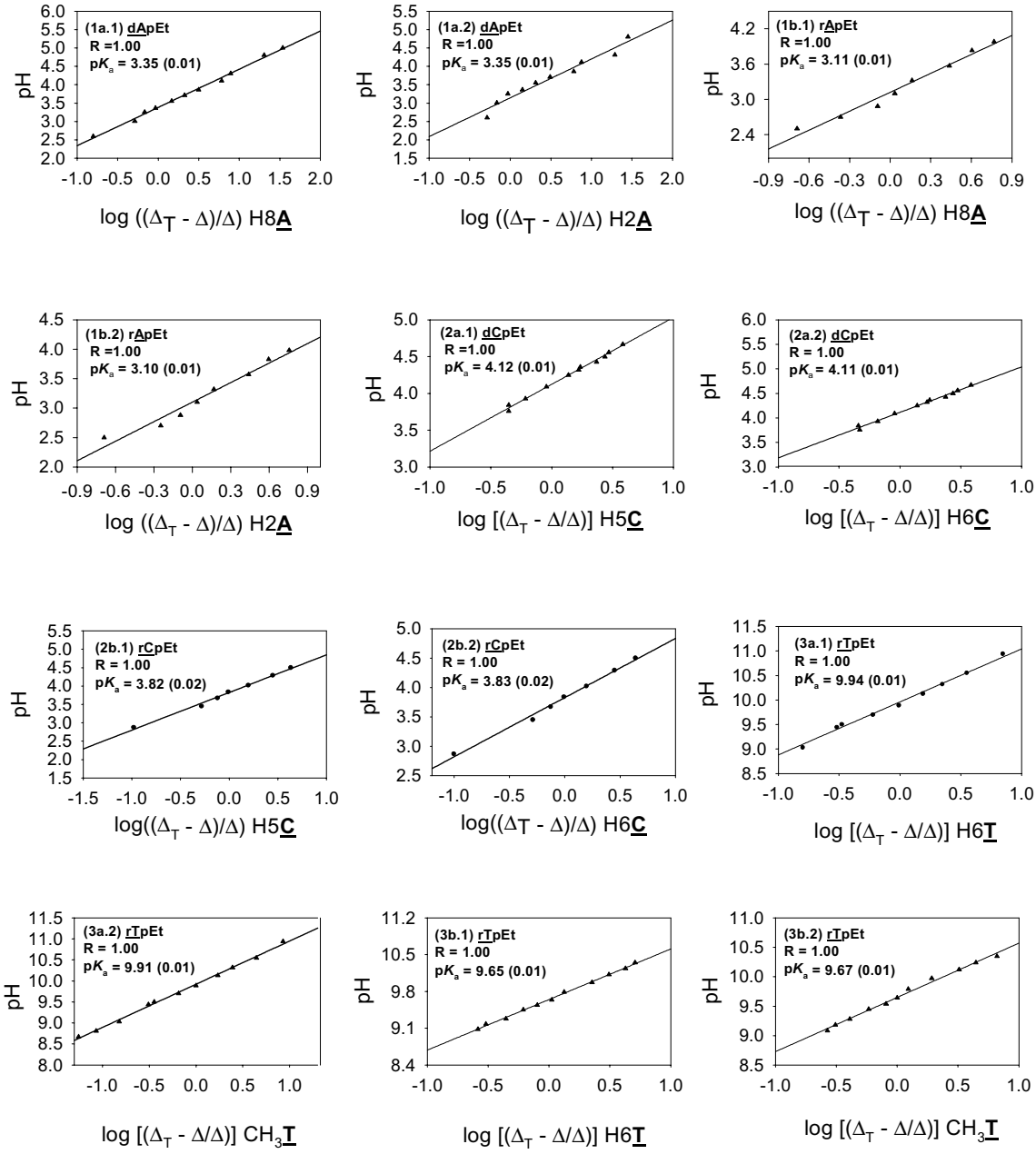


Figure S4 contd.

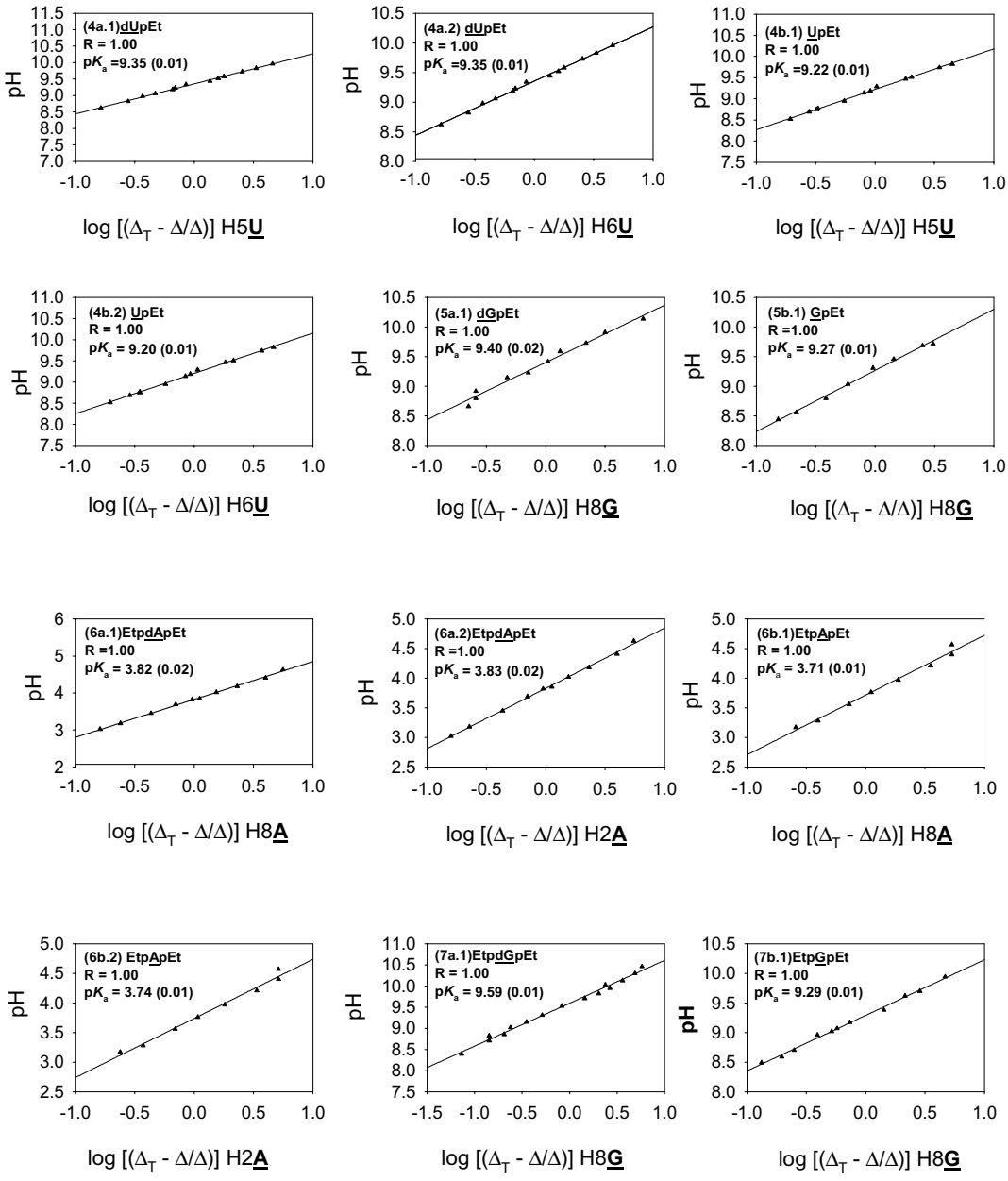
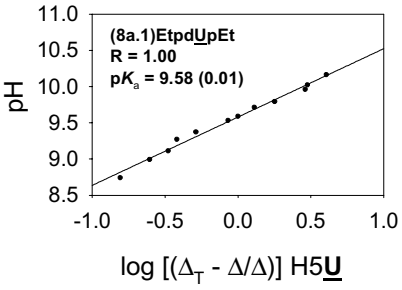
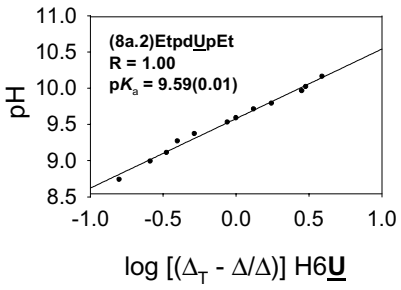
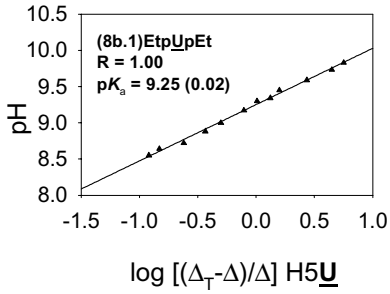
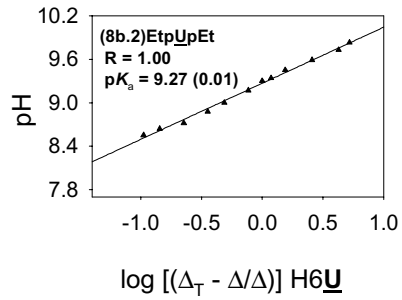


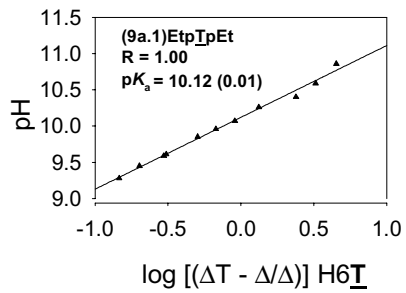
Figure S4 contd.

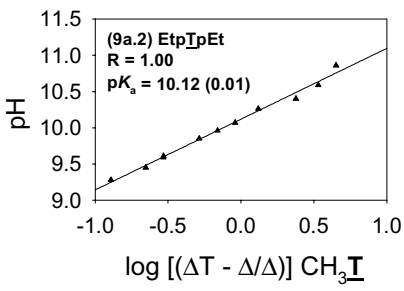

(8a.1)EtpdUpEt
R = 1.00
pKa = 9.58 (0.01)
pH
log [(ΔT - Δ/Δ)] H5U

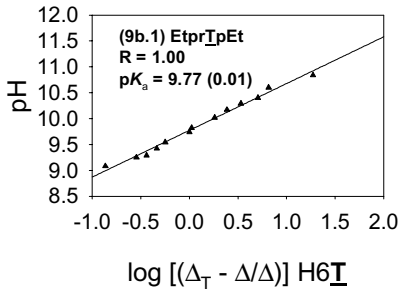

(8a.2)EtpdUpEt
R = 1.00
pKa = 9.59(0.01)
pH
log [(ΔT - Δ/Δ)] H6U

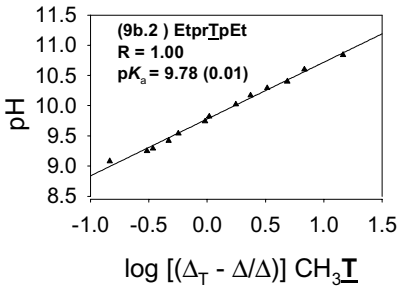

(8b.1)EtpUpEt
R = 1.00
pKa = 9.25 (0.02)
pH
log [(ΔT-Δ)/Δ] H5U

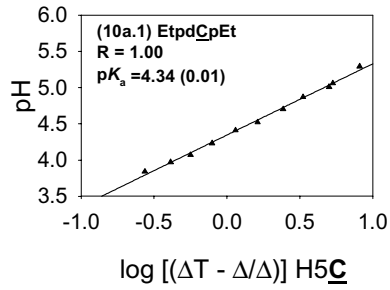

(8b.2)EtpUpEt
R = 1.00
pKa = 9.27 (0.01)
pH
log [(ΔT - Δ/Δ)] H6U

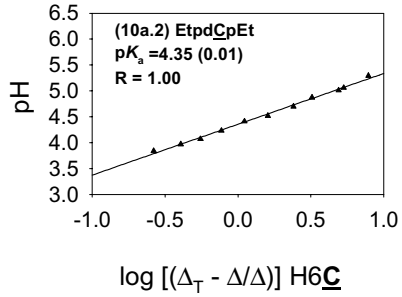

(9a.1)EtpTpEt
R = 1.00
pKa = 10.12 (0.01)
pH
log [(ΔT - Δ/Δ)] H6T

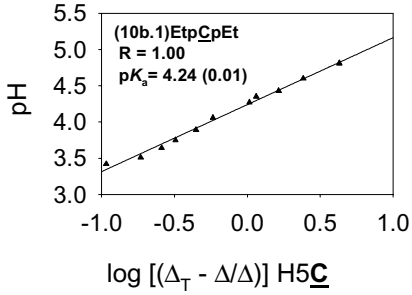

(9a.2) EtpTpEt
R = 1.00
pKa = 10.12 (0.01)
pH
log [(ΔT - Δ/Δ)] CH3T


(9b.1) EtprTpEt
R = 1.00
pKa = 9.77 (0.01)
pH
log [(ΔT - Δ/Δ)] H6T


(9b.2) EtprTpEt
R = 1.00
pKa = 9.78 (0.01)
pH
log [(ΔT - Δ/Δ)] CH3T


(10a.1) EtpdCpEt
R = 1.00
pKa =4.34 (0.01)
pH
log [(ΔT - Δ/Δ)] H5C


(10a.2) EtpdCpEt
pKa =4.35 (0.01)
R = 1.00
pH
log [(ΔT - Δ/Δ)] H6C


(10b.1)EtpCpEt
R = 1.00
pKa= 4.24 (0.01)
pH
log [(ΔT - Δ/Δ)] H5C

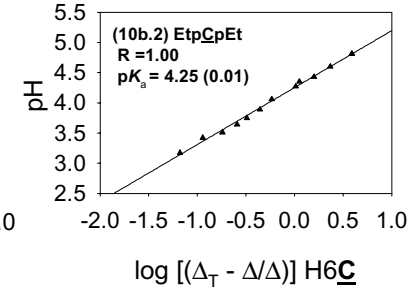

(10b.2) EtpCpEt
R =1.00
pKa = 4.25 (0.01)
pH
log [(ΔT - Δ/Δ)] H6C

Figure S3. Panels **1a.1** – **10b.2** show the Hill plots for the aromatic protons of the compounds **1a – 10a** and **1b** – **10b** (Scheme 1) at 298 K for calculating the pK_a. The plot of log $((\Delta_T - \Delta)/\Delta)$ of the aromatic protons as a function of pH gave a straight line with pK_a as intercept of the straight line. The name of the compounds with corresponding panel number, values for the correlation coefficient (R) and the pK_a obtained from the regression analyses are shown in the respective graphs [see experimental section for details].

Table S1. ^{1}H NMR chemical shift for the aromatic protons (δH8, δH2, δH5, δH6 and δC$\underline{H}_3$, in ppm) at 298 K for **1a – 10a** and **1b** – **10b** at the protonated (P), the neutral (N) and the deprotonated (D) states (*i.e.* two plateaus of the titration plots at two extreme pHs).

Compounds	**δH8**		**δH2**		**δH5**		**δH6**		**δCH$_3$**	
	N	**P/D**	**N**	**P/D**	**N**	**P/D**	**N**	**P/D**	**N**	**P/D**
dApEt (**1a**)	8.340	8.552	8.260	8.465	-	-	-	-	-	-
dCpEt (**2a**)	-	-	-	-	6.062	6.241	7.836	8.116	-	-
TpEt (**3a**)	-	-	-	-	-	-	7.663	7.495	1.905	1.867
dUpEt (**4a**)	-	-	-	-	5.898	5.820	7.863	7.696	-	-
dGpEt (**5a**)	7.998	7.861	-	-	-	-	-	-	-	-
EtpdApEt (**6a**)	8.464	8.629	8.276	8.465	-	-	-	-	-	-
EtpdGpEt **(7a)**	8.078	7.990	-	-	-	-	-	-	-	-
EtpdUpEt (**8a**)	-	-	-	-	5.941	5.853	7.939	7.783	-	-
EtpTpEt (**9a**)	-	-	-	-	-	-	7.762	7.613	1.936	1.892
EtpdCpEt (**10a**)		-	-	-	6.105	6.282	7.931	8.178	-	-
rApEt (**1b**)	8.350	8.583	8.270	8.479	-	-	-	-	-	-
rCpEt (**2b**)	-	-	-	-	6.069	6.259	7.848	8.169	-	-
rTpEt (**3b**)	-	-	-	-	-	-	7.699	7.510	-	-
rUpEt (**4b**)	-	-	-	-	5.914	5.829	7.888	7.691	-	-
rGpEt (**5b**)	8.011	7.860	-	-	-	-	-	-	-	-
EtprApEt (**6b**)	8.494	8.660	8.283	8.473	-	-	-	-	-	-
EtprGpEt (**7b**)	8.096	8.011	-	-	-	-	-	-	-	-
EtprUpEt (**8b**)	-	-	-	-	5.969	5.876	7.930	7.762	-	-
EtprTpEt (**9b**)	-	-	-	-	-	-	7.746	7.588	1.945	1.898
EtprCpEt (**10b**)	-	-	-	-	6.127	6.312	7.915	8.189	-	-

Table S2. ^{1}H NMR chemical shift (in ppm) for the sugar protons (δH1', δH2', δH3', δH4' and δH5' and δH5") at 298 K for **1a – 10a** and **1b** – **10b** at the protonated (P), the neutral (N) and the deprotonated (D) states (*i.e.* pH corresponding to the two plateaus of the titration plots).

Compounds	δH1'		δH2'		δH2"		δH3'		δH4'		δH5'		δH5"	
	N	P/D	N	P/D	N	P/D	N	P/D	N	P/D	N	P/D	N	P/D
dApEt (**1a**)	6.521	6.586	2.921	2.948	2.734	2.810	4.956	4.986	4.382	4.636	3.874	3.874	3.829	3.828
dCpEt (**2a**)	6.308	6.285	2.594	2.675	2.351	2.428	4.751	-[a]	4.232	4.276	3.358	3.877	3.785	3.797
TpEt (**3a**)	6.328	6.338	2.554	2.479	2.426	2.353	-[a]	-[a]	4.203	4.155	3.863	3.830	3.798	3.767
dUpEt (**4a**)	6.316	6.345	2.590	2.535	2.420	2.357	4.764	-[a]	4.227	4.201	3.855	3.843	3.787	3.778
dGpEt (**5a**)	6.341	6.337	2.868	2.864	2.676	2.629	4.939	-[a]	4.314	4.349	3.852	3.868	3.807	3.812
EtpdApEt (**6a**)	6.558	6.564	2.937	2.937	2.804	2.808	5.023	5.025	4.478	4.480	3.754	3.761	3.707	3.716
EtpdGpEt (**7a**)	6.360	6.346	2.929	2.892	2.714	2.709	5.009	-[a]	4.430	4.415	4.059*	4.023*	4.059*	4.023*
EtpdUpEt (**8a**)	6.375	6.424	2.592	2.536	2.397	2.341	4.865	-[a]	4.396	4.356	4.109	4.068*	4.066	4.068*
EtpTpEt (**9a**)	6.400	6.450	2.549	2.494	2.414	2.371	4.877	-[a]	4.382	4.339	4.110	4.073*	4.071	4.073*
EtpdCpEt (**10a**)	6.380	6.333	2.606	2.692	2.331	2.394	4.856	-[a]	4.397	4.453	4.111	-[‡]	4.063	-[‡]
ApEt (**1b**)	6.123	6.217	4.914	4.909	-[a]	-[a]	4.777	-[a]	4.497	4.470	3.933	3.945	3.871	3.885
CpEt (**2b**)	5.957	5.948	4.424	4.463	-[a]	-[a]	4.554	4.562	4.301	4.342	3.931	-[‡]	3.838	3.856
rTpEt (**3b**)	5.976	5.996	4.451	4.441	-[a]	-[a]	4.594	4.584	4.302	4.271	3.917	3.896	3.840	3.824
UpEt (**4b**)	5.968	5.984	4.461	4.445	-[a]	-[a]	4.477	4.567	4.316	4.284	3.919	3.898	3.834	3.820
GpEt (**5b**)	5.957	5.925	4.850	-[a]	-[a]	-[a]	4.762	-[a]	4.425	4.471	3.906	3.911	3.851	3.842
EtpApEt (**6b**)	6.192	6.263	4.931	4.931	-[a]	-[a]	4.835	4.869	4.582	4.605	4.108*	4.160*	4.108*	4.160*
EtpGpEt (**7b**)	5.983	4.976	4.945	-	-[a]	-[a]	4.825	-[a]	4.532	4.520	4.103*	4.066*	4.103*	4.066*
EtpUpEt (**8b**)	6.166	6.117	4.458	4.425	-[a]	-[a]	4.659	4.651	4.479	4.440	4.154	4.102*	4.085	4.102*
EtprTpEt (**9b**)	6.079	6.132	4.464	4.429	-[a]	-[a]	4.674	-[a]	4.479	4.464	4.143	4.092*	4.084	4.092*
EtpCpEt (**10b**)	6.088	6.036	4.416	4.452	-[a]	-[a]	4.650	4.645	4.456	4.499	4.168	4.227	4.089	4.101

[a] The chemical shift of the proton could not be assigned due to overlap with the water resonance. * The H5'/H5" chemical shift could not be distinguished as they become almost isochronous. [‡] The H5'/H5" chamical shift could not be distinguished as they are overlapped with the O-C$\underline{H}_2$-CH_3 of the ethylphospahte moiety.

Table S3. The total number of the middle (Mid) and the terminal (Term, both $^{3'}_{5'}$ and $^{5'}_{3'}$) basepairing [both r(A-U) and d(A-T) as well as r(G-C) and d(G-C) bp] in duplexes **1** – **14**.[a]

No.	Duplex Sequence [a]	Total A-T/U bp	Total G-C bp	Mid A-T/U bp	Mid G-C bp	Term $^{5'}_{3'}$ bp	Term $^{3'}_{5'}$ bp
(1)	**5'-TCCCTCCTCTCC-3' 3'-AGGGAGGAGAGG-5'**	4	8	3	7	**T A**	**C G**
(2)	**5'-CCTTCCCTT-3' 3'-GGAAGGGAA-5'**	4	5	3	4	**C G**	**T A**
(3)	**5'-TTCCCTTCC-3' 3'-AAGGGAAGG-5**	4	5	3	4	**T A**	**C G**
(4)	**5'-GCTCTCTGGC-3' 3'-CGAGAGACCG-5'**	3	7	3	5	**G C**	**C G**
(5)	**5'-CTCGTAC CTTCCGGTCC-3' 3'-GAGCATGGAAGGCCAGG-5'**	6	11	6	9	**C G**	**C G**
(6)	**5'-CTCGTACCTTTCCGGTCC-3' 3'-GAGCATGG AAAGGCCAGG-5'**	7	11	7	9	**C G**	**C G**
(7)	**5'-TAGTTATCTCTATCT- 3' 3'-ATCAATAGAGATAGA-5'**	11	4	9	4	**T A**	**T A**
(8)	**5'-GCACAGCC-3' 3'-CGTGTCGG-5'**	2	6	2	4	**G C**	**C G**
(9)	**5'-GAGCTCCCAGGC-3' 3'-CTCGAGGGTCCG-5'**	3	9	3	7	**G C**	**C G**
(10)	**5'-CCGAGGTCCATGTCGTACGC3' 3'-CGGCTCCAGGTACAGCATGCG-5**	7	14	7	12	**G C**	**C G**
(11)	**5'-TGTACGTCACAACTA-3' 3'-ACATGCAGTGTTGAT-5'**	9	6	7	6	**T A**	**A T**
(12)	**5'-TATACAAGTTATCTA-3' 3'-ATATGTTCAATAGAT-5'**	12	3	10	3	**T A**	**A T**
(13)	**5'-CGCCTATGCAAAAAC-3' 3'-GCTGATACGTTT TTG-5**	9	6	9	4	**C G**	**C G**
(14)	**5'-CGCAAAAAAAAAACGC-3' 3'-GCGTTTTTTTTTTGCG-5'**	10	6	10	4	**C G**	**C G**

[a] The duplex sequences (1 – 14) are taken from ref. 4. See Table 3 for total free energy calculation for basepairing ($[\Delta G^{o}_{bp}]_{RR-DD}$, in kcal mol^{-1}). See ref. 8 for equation used for $[\Delta G^{o}_{bp}]_{RR-DD}$ calculations.

Table S4. **Table S3**. The total number of the middle (Mid) basepairing and the calculations for $\sum\Delta pK_a$ in both DNA-DNA and RNA-RNA duplexes **1** – **14**.[a]

No.	Duplex Sequence [a]	Mid [b] A-T/U bp	Mid [c] G-C bp	$\sum\Delta pK_a$ [d]	$\sum\Delta pK_a$ [e]
(1)	**5'-TCCCTCCTCTCC-3'** **3'-AGGGAGGAGAGG-5'**	3	7	55.55	51.87
(2)	**5'-CCTTCCCTT-3'** **3'-GGAAGGGAA-5'**	3	4	39.83	36.75
(3)	**5'-TTCCCTTCC-3'** **3'-AAGGGAAGG-5**	3	4	39.83	36.75
(4)	**5'-GCTCTCTGGC-3'** **3'-CGAGAGACCG-5'**	3	5	45.07	41.79
(5)	**5'-CTCGTAC CTTCCGGTCC-3'** **3'-GAGCATGGAAGGCCAGG-5'**	6	9	84.90	78.54
(6)	**5'-CTCGTACCTTTCCGGTCC-3'** **3'-GAGCATGG AAAGGCCAGG-5'**	7	9	91.19	84.07
(7)	**5'-TAGTTATCTCTATCT- 3'** **3'-ATCAATAGAGATAGA-5'**	9	4	- [¥]	- [¥]
(8)	**5'-GCACAGCC-3'** **3'-CGTGTCGG-5'**	2	4	33.54	31.22
(9)	**5'-GAGCTCCCAGGC-3'** **3'-CTCGAGGGTCCG-5'**	3	7	55.55	51.87
(10)	**5'CCGAGGTCCATGTCGTACGC-3'** **3'-CGGCTCCAGGTACAGCATGCG-5**	7	12	106.91	99.19
(11)	**5'-TGTACGTCACAACTA-3'** **3'-ACATGCAGTGTTGAT-5'**	7	6	75.47	68.95
(12)	**5'-TATACAAGTTATCTA-3'** **3'-ATATGTTCAATAGAT-5'**	10	3	- [¥]	- [¥]
(13)	**5'-CGCCTATGCAAAAAC-3'** **3'-GCTGATACGTTT TTG-5**	9	4	77.57	69.93
(14)	**5'-CGCAAAAAAAAAACGC-3'** **3'-GCGTTTTTTTTTTGCG-5'**	10	4	83.86	75.46

[a] The RNA-RNA (RR) and DNA-DNA (DD) duplexes, **1** – **14**,are taken from ref. 4 and $\Sigma\Delta pK_a$ for DD and RR are used in the plot for Figure 3. [b] The total number of middle d(A-T) basepairs (bp) for DD and r(A-U) basepairs (bp) for RR duplexes in each sequence. ΔpK_a for r(U-A): $[pK_a]_{\mathbf{8b}} - [pK_a]_{\mathbf{6b}} = 5.53$ and ΔpK_a for d(T-A): $[pK_a]_{\mathbf{9a}} - [pK_a]_{\mathbf{6a}} = 6.29$. [c] The total number of middle d(G-C) basepairs for DD and r(G-C) basepairs for RR duplexes in each sequence. ΔpK_a for r(G-C): $[pK_a]_{\mathbf{7b}} - [pK_a]_{\mathbf{10b}} = 5.04$, ΔpK_a for d(G-C): $[pK_a]_{\mathbf{7a}} - [pK_a]_{\mathbf{10a}} = 5.24$. [d] $\Sigma\Delta pK_a$ for DD = [No. of d(A-T) bp * ΔpK_a for d(A-T)] + [No. of d(G-C) bp * ΔpK_a for d(G-C)]. [e] $\Sigma\Delta pK_a$ for RR = [No. of r(A-U) bp * ΔpK_a for r(A-U)] + [No. of r(G-C) bp * ΔpK_a for r(G-C)].