SUPPLEMENTARY DATA

Tethered imidazole mediated duplex stabilization and its potential for aptamer stabilization

Lars Verdonck[1,2], Dieter Buyst[2], Anne-Mare de Vries[1,2], Vicky Gheerardijn[1], Annemieke Madder[1,*] and José C. Martins[2,*]

[1] Department of Organic and Macromolecular Chemistry, Organic and Biomimetic Chemistry Research Group, Ghent University, Gent, Oost-Vlaanderen, 9000, Belgium
[2] Department of Organic and Macromolecular Chemistry, NMR and Structure Analysis Research Group, Ghent University, Gent, Oost-Vlaanderen, 9000, Belgium

*To whom correspondence should be addressed. Tel: +32 9 264 44 69; Fax: +32 9 264 49 98; Email: Jose.Martins@UGent.be, Correspondence may also be addressed to Annemieke.Madder@UGent.be

Table of contents

Building block synthesis

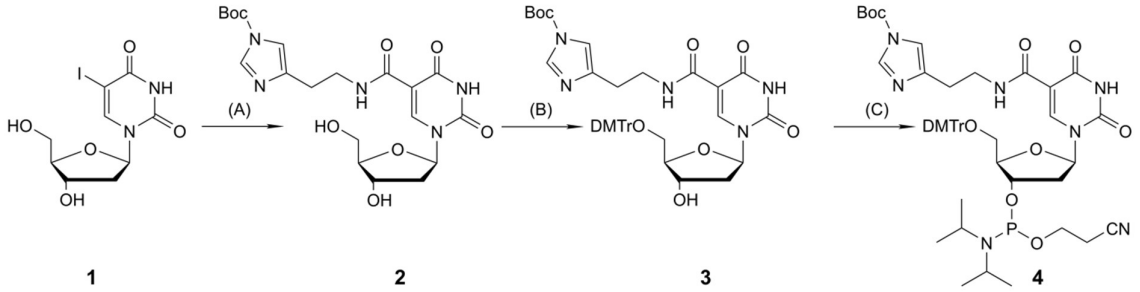

Figure S1. Synthesis of 5-(2-(1-(tert-Butoxycarbonyl)imidazole-4-yl)ethylaminocarbonyl)-5'-O-(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl-N,N'-diisopropylphosphoramidite)-2'-deoxyuridine; (A) (i) Histamine, Et_3N, tetrakis(triphenylphosphine)palladium(0), CO, DMF; (ii) di-tert-butyl dicarbonate, Et_3N, DMF; (B) DMTrCl, pyridine; (C) DIPEA, 2-cyanoethoxy-diisopropylaminochlorophosphine, DCM

All used reagents and solvents were purchased from Sigma-Aldrich and Fluka except the anhydrous solvents (Acros Organics), deuterated solvents (Eurisotop) and 5-iodo-2'-deoxyuridine (Chem-Impex International). Dichloromethane was distilled from CaH_2, all other solvents were used without further purification. All reactions were carried out under argon or nitrogen atmosphere with magnetic stirring. TLC was performed on pre-coated silica gel plates (SIL G-25 UV254, 0.25 mm thickness) and compounds were stained using phosphomolybdic acid (5 % in ethanol) or anisaldehyde (5 % in ethanol, 1 % sulfuric acid). Silica gel 60 (0.063–0.200 mm particle size, 230-400 mesh) was used for flash column chromatography. ^1H, ^{13}C and ^{31}P NMR spectra were recorded at 300 MHz, 75 MHz and 121 MHz, respectively. Chemical shifts (δ) are reported in units of parts per million (ppm), with the residual ^1H or ^{13}C peaks of the solvent used as internal standards ($(CD_3)_2SO$: δ_H = 2.50 ppm and δ_C = 39.52 ppm; $CDCl_3$: δ_H = 7.26 ppm and δ_C = 77.16 ppm). The following abbreviations are used to explain the observed multiplicities: s, singlet; d, doublet; t, triplet; q, quadruplet; m, multiplet; br, broad; band, several overlapping signals; AB, AB system with strongly skewed signals; app, indicates an "apparent" multiplicity, for which only the observed average coupling constant can be quoted, in absence of information on the real J-values. Where given, assignments of resonances were confirmed by standard COSY and HSQC 2D NMR experiments. High resolution mass spectra (HRMS) were recorded with an Agilent Accurate-Mass Quadrupole Time-of-Flight mass spectrometer.

5-{2-[1-(tert-Butoxycarbonyl)imidazole-4-yl]ethylaminocarbonyl}-2'-deoxyuridine, **2**. 5-Iodo-2'-deoxyuridine **1** (1 g, 2.8 mmol) was placed in a Parr high pressure vessel and suspended in anhydrous DMF (20 mL). After addition of histamine (0.944 g, 8.5 mmol, 3 eq.), Et_3N (2.4 mL, 17 mmol, 6 eq.) and tetrakis(triphenylphosphine)palladium (0.162 g, 0.14 mmol, 0.05 eq.), the temperature was set to 70 °C and a constant carbon monoxide pressure was applied (90 psi, 6.2 bar). After 48 h, the reaction mixture was cooled to room temperature before being transferred to a flask and concentrated *in vacuo* to an oil. The oil was redissolved in anhydrous DMF (10 mL) and Et_3N (1.2 mL, 8.5 mmol, 3 eq.) and di-tert-butyl dicarbonate (0.930 mg, 4.2 mmol, 1.5 eq.) were added. The reaction was stirred at room temperature for 15-30 min before being quenched with methanol (15 mL). The solvent was evaporated and the residue was purified with silica column chromatography (gradient 5-10 % MeOH in DCM) to give a white foam (800 mg, 65 %). R_f = 0.23 (CH_2Cl_2: CH_3OH 9:1). ^1H NMR ($(CD_3)_2SO$, 300 MHz): δ 11.84 (s (br),1H), 8.78 (t, *J*= 6.5 Hz,1H), 8.68 (s, 1H), 8.11 (d app, *J*= 1.2 Hz, 1H), 7.31 (d app, *J*= 1.2 Hz, 1H), 6.11 (t app, *J*= 6.6 Hz, 1H), 5.28 (d, *J*= 4.1 Hz, 1H), 5.00 (t, *J*= 4.14 Hz, 1H), 4.22 (m, 1H), 3.86 (q, *J*= 3.7 Hz, 1H), 3.56 (t, *J*= 3.7 Hz, 2H), 3.52 (q, *J*= 6.5 Hz, 2H), 2.69 (t, *J*= 6.5 Hz, 2H), 2.21 (ddd AB, *J*= 3.6, 6.6 and 13.7 Hz, 1H), 2.12 (dt AB, *J*= 6.6 and 13.7 Hz, 1H), 1.56 (s, 9H). ^{13}C APT NMR ($(CD_3)_2SO$, 75 MHz): δ 163.1, 161.4, 149.5, 145.8, 140.8, 123.9, 113.6, 105.3, 87.9, 85.5, 85.1, 70.4, 61.2, 45.7, 37.9, 27.8, 27.4; HRMS (ESI): *m/z*: calculated for $C_{20}H_{26}O_8N_5^-$: 464.1781 $[M-H]^-$; found: 464.1791

5-{2-[1-(tert-Butoxycarbonyl)imidazole-4-yl]ethylaminocarbonyl}-5'-dimethoxytrityl-2'-deoxyuridine, **3**. The starting nucleoside **2** (1.0 g), dried by repeated coevaporation with anhydrous pyridine (3 x 10 mL) was dissolved in a cooled (0 °C) mixture of anhydrous pyridine (12 mL) and anhydrous DCM (8 mL) and placed under an Argon atmosphere. Dimethoxytrityl chloride (0.9 g) was dissolved in a mixture of anhydrous pyridine (4 mL) and DCM (4 mL) and added slowly to the cooled nucleoside solution. After stirring for 18h at room temperature, the reaction was quenched by the addition of anhydrous methanol (5 mL, 0 °C). The reaction mixture was concentrated to an oil and re-dissolved in DCM (40 mL). The solution was washed with a saturated $NaHCO_3$ solution (1 x 40 mL) and brine (2 x 40 mL) and dried with Na_2SO_4. Finally, it was reconcentrated to an oil before purification with silica gel column chromatography (gradient 0-5 % MeOH in DCM, 1 % Et_3N). The collected fractions were concentrated under vacuum and gave a white foam (445 mg, 70 %). R_f = 0.53 (CH_2Cl_2: CH_3OH 9:1). ^1H NMR ($CDCl_3$, 300 MHz): δ 9.17 (t, *J*= 6.9 Hz, 1H), 8.54 (s, 1H), 8.11 (d app, *J*= 1.3 Hz, 1H), 7.17-7.42 (band, 10H), 6.83 (m, 4H), 6.20 (t app, *J*= 6.5 Hz, 1H), 4.35 (dt, *J*= 5.0 and 6.5 Hz, 1H), 3.95 (q app, *J*= 5.0 Hz, 1H), 3.78 (s, 6H), 3.66 (q, *J*= 6.9 Hz, 2H), 3.51 (dd AB, *J*= 5.0 and 10.2 Hz, 1H), 3.36 (dd AB, *J*= 5.0 and 10.2 Hz, 1H), 2.87 (t, *J*= 6.9 Hz, 2H), 2.46 (ddd AB, *J*= 5.0, 6.5 and 13.7 Hz, 1H), 2.23 (dt AB, *J*= 6.5 and 13.7 Hz, 1H), 1.60 (s, 9H). ^{13}C APT NMR ($CDCl_3$, 75

MHz): δ 163.3, 161.8, 158.7, 150.0, 147.0, 145.7, 144.7, 140.8, 135.8, 130.2, 128.2, 128.1, 127.1, 114.0, 113.4, 106.8, 87.0, 85.8, 85.7, 85.4, 72.6, 63.9, 55.4, 46.1, 40.3, 39.5, 28.4, 28.1; HRMS (ESI): *m/z*: calculated for $C_{41}H_{44}O_{10}N_5^-$: 766.3088 $[M-H]^-$; found: 766.3081

5-(2-(1-(tert-Butoxycarbonyl)imidazole-4-yl)ethylaminocarbonyl)-5'-O-(4,4'-dimethoxytrityl)-3'-(2-cyanoethyl-N,N'-diisopropylphosphoramidite)-2'-deoxyuridine, **4**. The dimethoxytrityl-protected nucleoside **3** (395 mg, 0.51 mmol), dried by repeated co-evaporation with anhydrous DCM (3 x 10 mL) was further dried *in vacuo* (oil pump, overnight). The resulting off-white foam was dissolved in anhydrous DCM (15mL) and cooled to 0 °C before adding DIPEA (4 eq., 0.36 mL, 2 mmol) and 2-cyanoethyl *N,N'*-diisopropylchlorophosphoramidite dropwise. After completion of the reaction (1 h) excess phosphoramidite reagent was quenched with 5 % $NaHCO_3$ at 0 °C. After addition of 30 mL DCM the mixture was washed with 5 % $NaHCO_3$ (1 x 40 mL) and brine (3 x 40 mL). The organic phase was dried on Na_2SO_4 and concentration under vacuum to yield a white foam. The product was used in solid phase oligonucleotide synthesis without further purification (428 mg, quantitative conversion). R_f = 0.62, 0.66 (CH_2Cl2: CH_3OH 9:1). ^{31}P NMR ($CDCl_3$, 121 MHz): δ 149.14, 148.74 ppm. HRMS (ESI): m/z: calculated for $C_{50}H_{61}O_{11}N_7P^-$: 966.4166 $[M-H]^-$; found: 966.4153.

Synthesis of the T_x^{ImH+} modified oligodeoxyribonucleotides

All phosphoramidites were purchased from Glen Research, other reagents for ODN synthesis were obtained from Proligo Reagents (Sigma Aldrich). Modified oligodeoxynucleotides (ODNs) were synthesized on an ABI 394 DNA synthesizer and couplings proceeded using the standard automated phosphoramidite-based solid phase synthesis protocol for 1 µmol synthesis scale except for the coupling of the modified phosphoramidite (**3**). After mixing a 0.2 M solution of modified phosphoramidite in dry acetonitrile (0.4 mL) with the 4,5-dicyanoimidazole activator solution (0.1 M in dry acetonitrile, 0.5 mL) this coupling mixture was pushed manually through the synthesis column on a gradual basis (0.05 mL/min). The column was then reinstalled on the synthesizer and automated synthesis was resumed. After completion, the 5'-O-DMTr protected ODNs were cleaved from solid support and deprotected by incubation in a concentrated aqueous ammonia solution at 55 °C (overnight). The oligonucleotides were purified by solid phase extraction (SPE) using Waters C18 Sep-Pak cartridges followed by precipitation of the ODNs in a 95/5 mixture of isopropanol and a 3 M NaOAc solution to exchange the triethylammonium counter ions. Mass analysis of the modified oligonucleotides was carried out on an Agilent G1946C LC/MSD-VL 1100 Series HPLC equipped with API-ESI source. All spectra were acquired in negative ion mode from m/z = 50 to m/z = 1500. The experimental masses were obtained from the mass spectra using the built-in deconvolution algorithm of the Agilent LC/MSD Chemstation software (version A.08014). The calculated molecular weights are the exact molecular masses of the most abundant isotopic species calculated via the Perkin Elmer ChemBioDraw software (version 13.0.2.3021).

Selection of the doubly imidazole modified constructs

Based on the analysis made in the main text using an otherwise unspecified dsDNA sequence (Figure 2), several possibilities exist with regards to the motif configurations that can be considered when two T^{ImH+} are introduced in a dsDNA. First it can be noted that two *non-motif* T^{ImH+} can be combined at will, i.e. no limitations are imposed as to their relative position in the sequence (positions labelled 'b' or 'n' in Figure 2A of the manuscript). Second, introduction of an *in-motif* T^{ImH+} in a strand already featuring one T^{ImH+} is generally possible irrespective of the motif adopted by the latter, provided this *in-motif* T^{ImH+} does not precede a *non-motif* T^{ImH+} (corresponding to positions *n-1* and *n-2* Figure 2A) or the entire three base pair pKa-motif sequence box of an *in-motif* T^{ImH+} (positions *n+1* and *n+2* Figure 2B) by one or two places in the 5'-3' direction. When the *in-motif* T^{ImH+} is introduced on the other, complementary strand, the same holds except that the position prior to the pKa-motif (marked 'x' in Figure 2B) should be excluded to avoid an ambiguous state. Thus, introducing a mixed motif combination (non/in) imposes some limitations compared to the non/non combination. Finally, the strongest limitations occur when a double *in-motif* combination is to be introduced. Now the overlap of the respective pKa-motif sequence boxes must be avoided. Practically speaking (Figure S2), the latter implies that both T^{ImH+} should be separated by at least 2 nucleotides when placed on the same strand. When placed on opposite strands, there should be a separation of at least four nucleotides when the second T^{ImH+} precedes the complementary base of the first T^{ImH+} in the 5' to 3' direction. When the second T^{ImH+} immediately follows this complementary base, all positions are possible, imposing no restrictions (Figure S2).

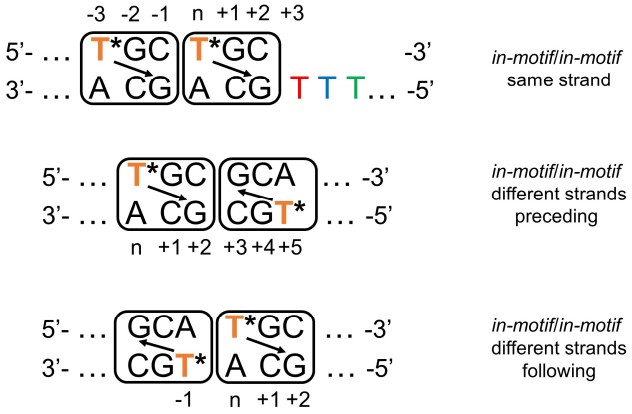

Figure S2. Overview of closest positions available for two T^{ImH+} nucleotides, forming a pKa motif in an otherwise unspecified duplex composition. The three base pair box contains the pKa motif sequence, with the arrow indicating the hydrogen bonding interaction of the T^{ImH+} at position n with the Hoogsteen side of the required guanine on the opposite strand at position *n+2*.

Hydrogen bond analysis

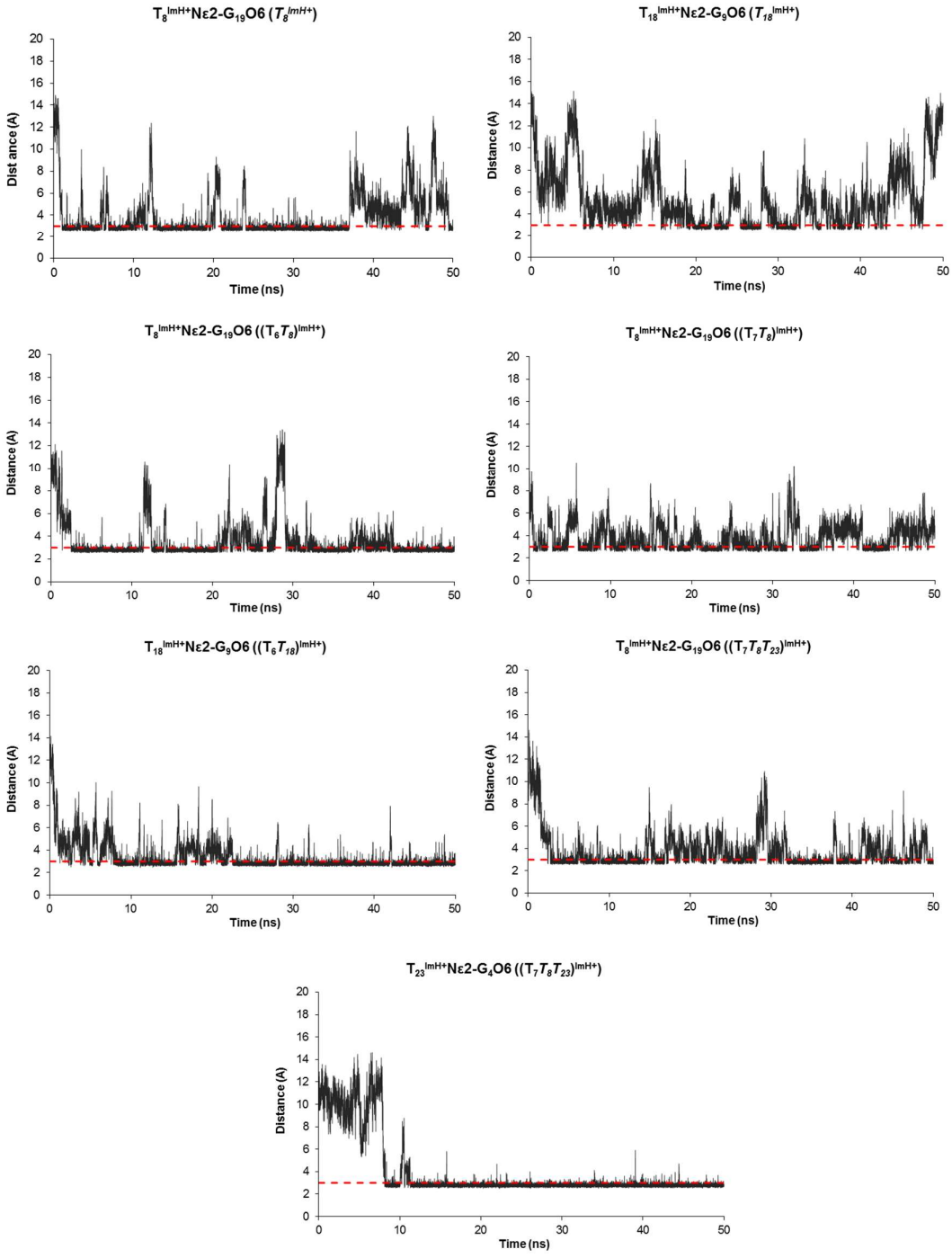

Figure S3. Overview of the distances between the donor heavy atom, T_l^{ImH+} Nε2, and the acceptor heavy atom, G_j O6, extracted from the 50 ns trajectory runs of the systems featuring at least one *in-motif* T^{ImH+}. The red dotted line at 3 Å indicates the distance criterion for successful hydrogen bond formation.

Chemical shift perturbation analysis of the T^{ImH+}-modified duplexes

In the following figures (S4 – S10) the chemical shift perturbation analysis of each modified duplex with respect to the corresponding non-modified reference sequences is shown for the non-exchangeable sugar and aromatic base protons ($\Delta\delta = \delta_{reference\ duplex} - \delta_{modified\ duplex}$) . In each case, the perturbation mapping of the non-exchangeable protons is performed using data recorded at 25°C and pD 5.0 in 100 % D_2O.

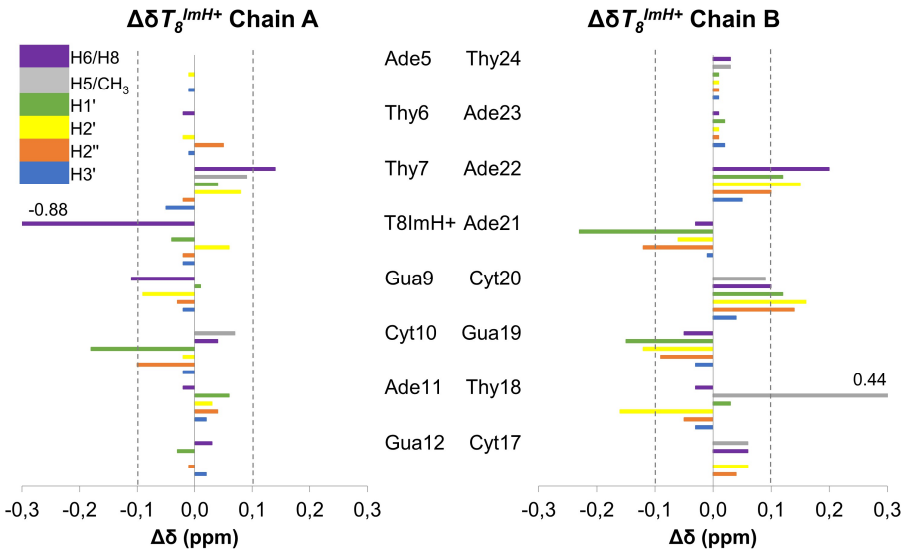

Figure S4. Chemical shift perturbation analysis of the non-exchangeable protons of the T_8^{ImH+} duplex with respect to the unmodified sequence.

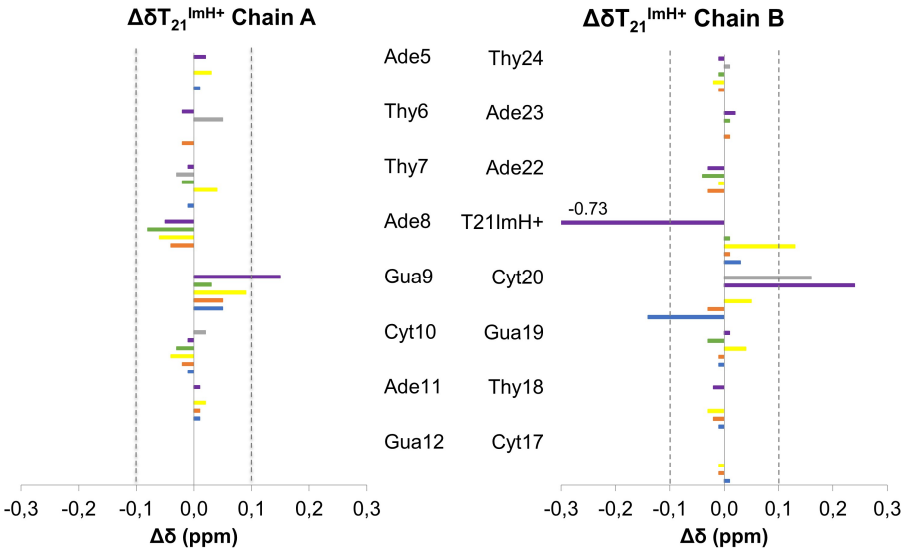

Figure S5. Chemical shift perturbation analysis of the non-exchangeable protons of the T_{21}^{ImH+} duplex with respect to the unmodified sequence.

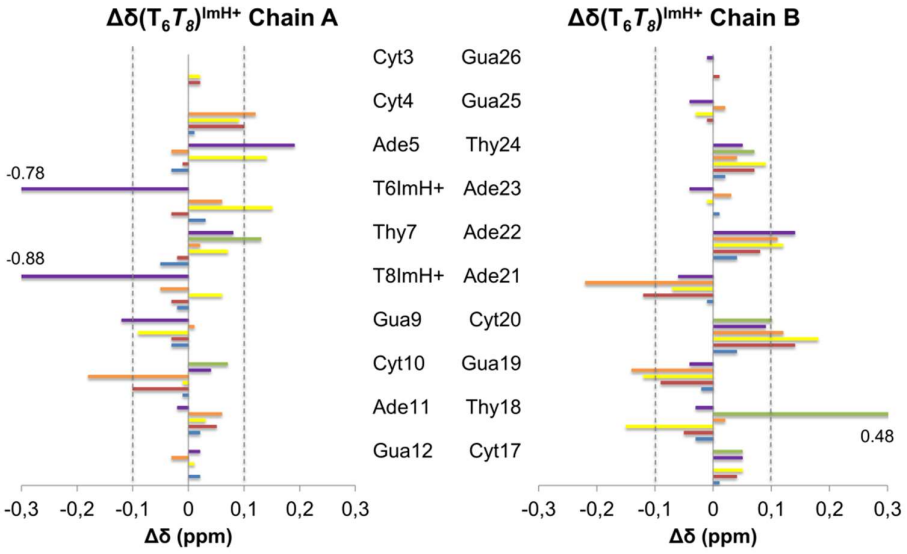

Figure S6. Chemical shift perturbation analysis of the non-exchangeable protons of the $(T_6T_8)^{ImH+}$ duplex with respect to the unmodified sequence.

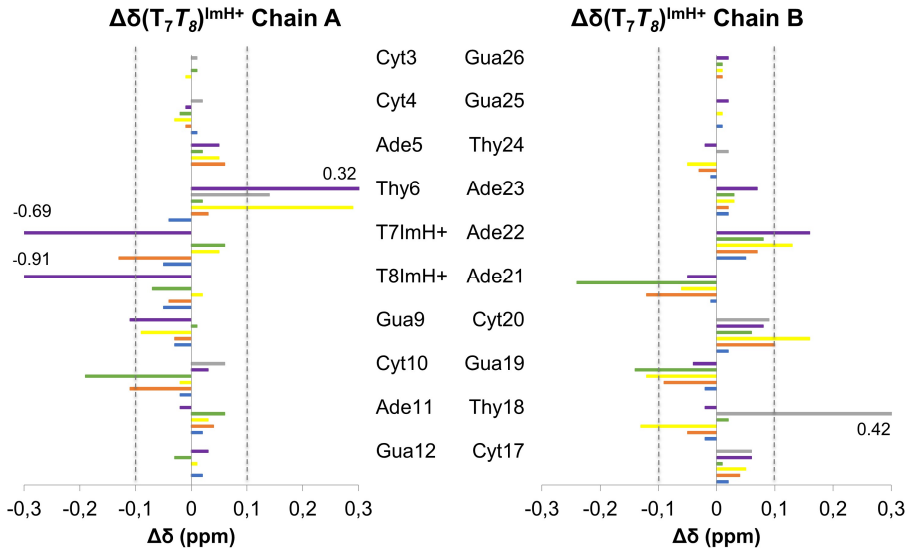

Figure S7. Chemical shift perturbation analysis of the non-exchangeable protons of the $(T_7T_8)^{ImH+}$ duplex with respect to the unmodified sequence.

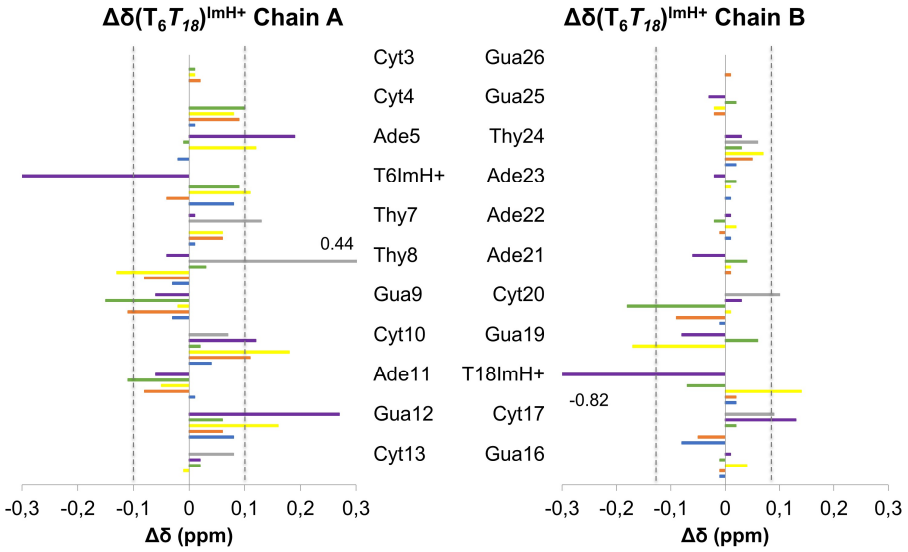

Figure S8. Chemical shift perturbation analysis of the non-exchangeable protons of the $(T_6T_{18})^{ImH+}$ duplex with respect to the unmodified sequence.

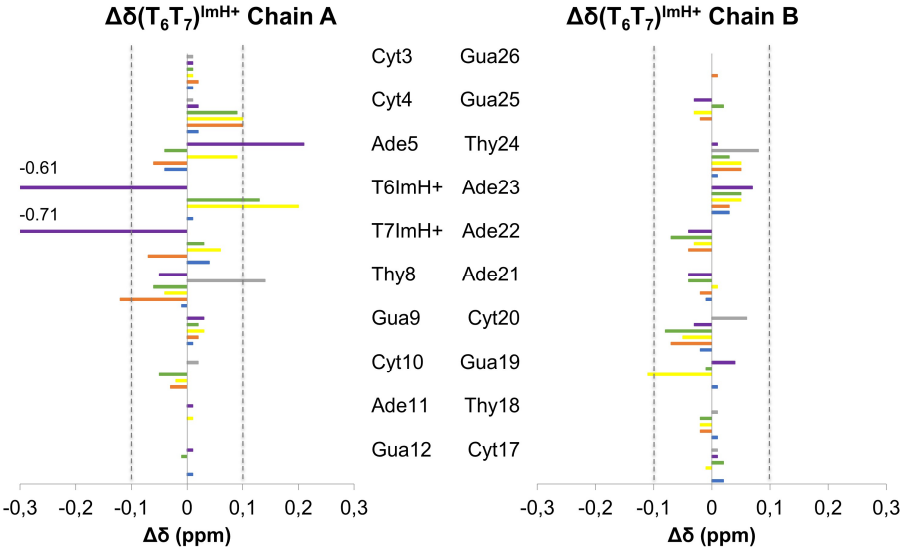

Figure S9. Chemical shift perturbation analysis of the non-exchangeable protons of the $(T_6T_7)^{ImH+}$ duplex with respect to the unmodified sequence.

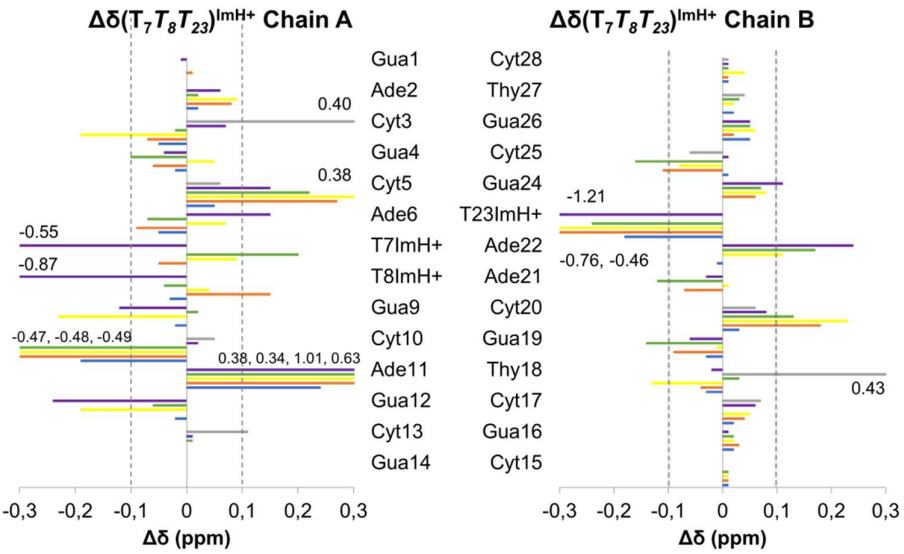

Figure S10. Chemical shift perturbation analysis of the non-exchangeable protons of the $(T_7T_8T_{23})^{ImH+}$ duplex with respect to the unmodified sequence.

Chemical shift perturbation analysis of the T^{ImH+}-modified aptamers

In the following figures (S11. – S12.) the chemical shift perturbation analysis of the imidazole-modified aptamers with respect to the corresponding non-modified reference sequence are shown for the non-exchangeable sugar and aromatic base protons ($\Delta\delta = \delta_{reference\ duplex} - \delta_{modified\ duplex}$). In each case, the perturbation mapping of the non-exchangeable protons is performed using data recorded at 25°C and pD 5.0 in 100 % D_2O.

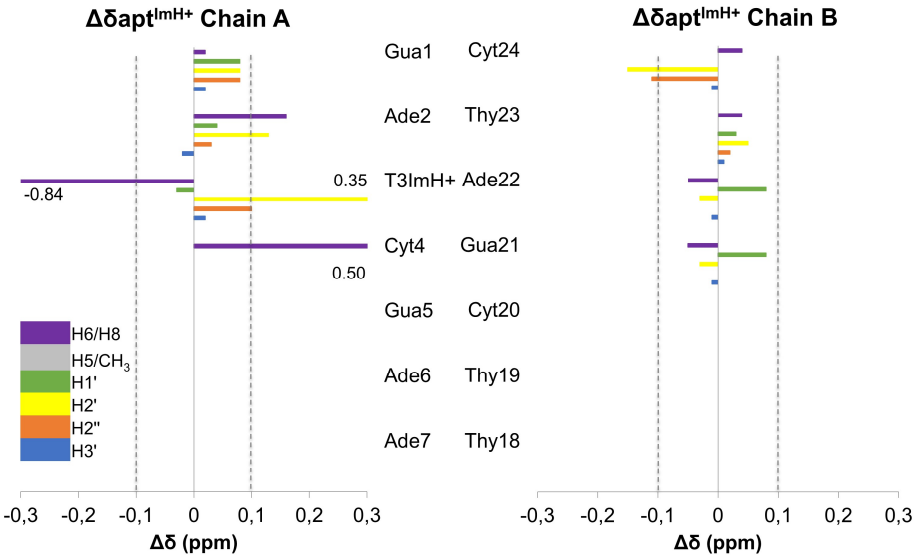

Figure S11. Chemical shift perturbation analysis of the non-exchangeable protons of the apt^{ImH+} duplex with respect to the unmodified apt sequence.

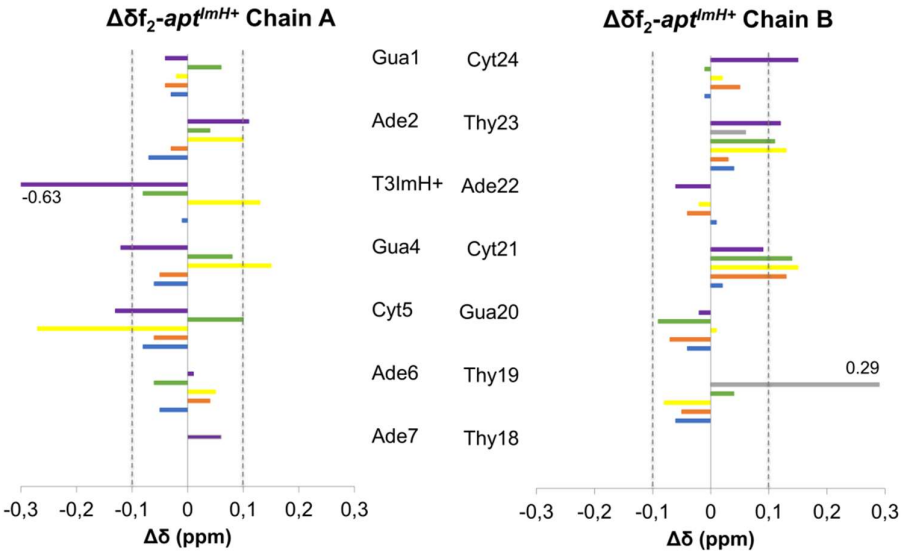

Figure S12. Chemical shift perturbation analysis of the non-exchangeable protons of the f_2-*Apt*$^{ImH+}$ duplex with respect to the unmodified f_2-Apt sequence.

Experimental pKa-titration curves

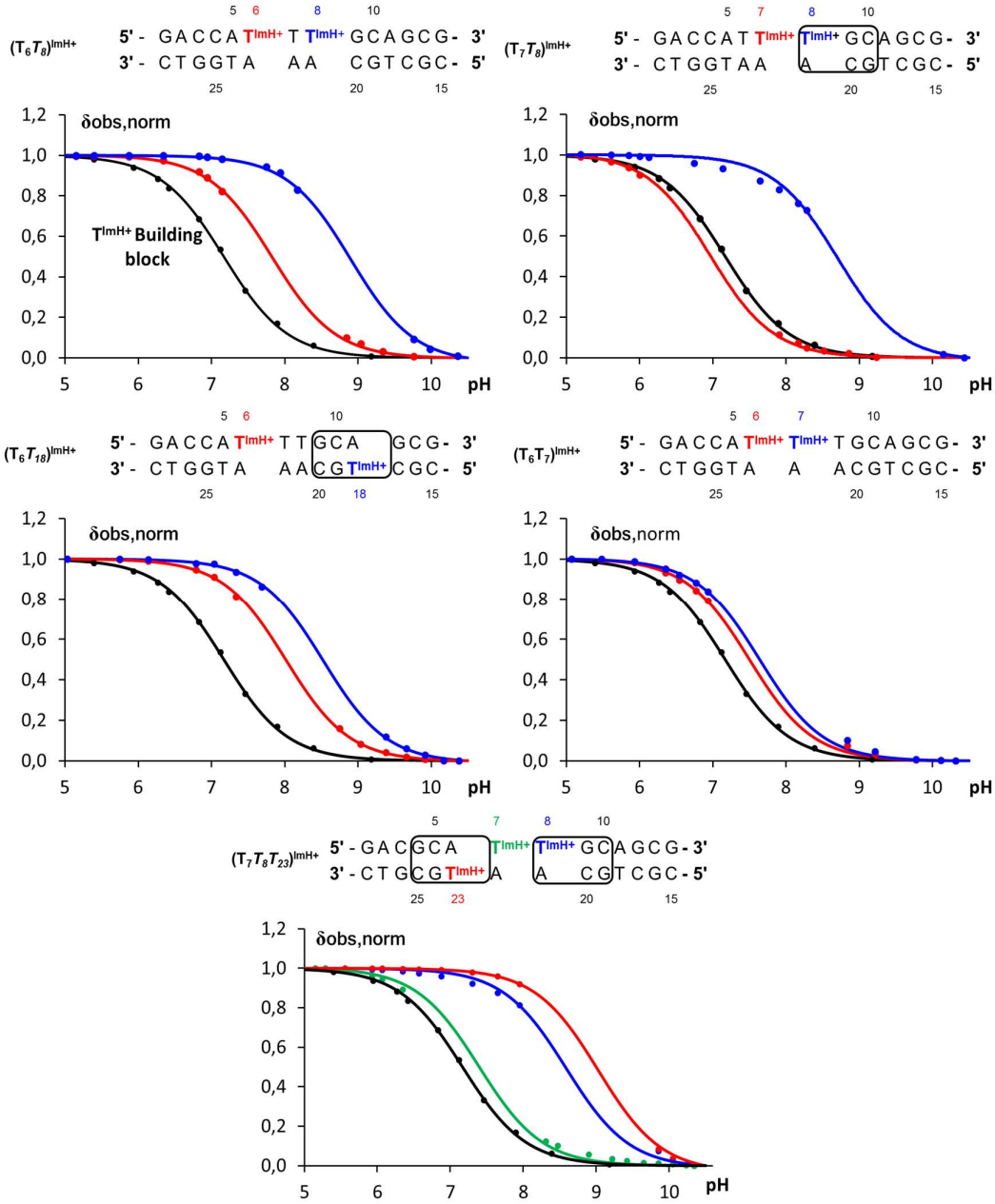

Figure S13. Normalized observed chemical shifts as a function of pH for the isolated building block together with all studied doubly and triply modified T_x^{ImH+} duplexes. Both the experimental data and fitted curves are shown.

Remark: In case of the modified duplexes, (de)protonation of the imidazole functionality is only moderately fast on the NMR time-scale. Therefore, the line-width is dependent on the size of the acidic/basic populations. This causes line broadening in the range where the pH of the sample approximates the pKa of the imidazolium hence the lack of data points in this region (51). The T^{ImH+} nucleoside does not suffer from this problem thereby allowing full sampling of the titration curve (black curve). When up to 5 points are removed from the transition part of this curve, the pKa obtained through the fit remains identical except for a somewhat larger standard deviation. Thus asng as the initial and final deviations from linearity are sufficiently sampled the pKa value reflects the same value as when all points would be available.