

# Identification of a pK<sub>a</sub>-regulating motif stabilizing imidazole-modified double-stranded DNA

Dieter Buyst<sup>1,†</sup>, Vicky Gheerardijn<sup>2,†</sup>, Krisztina Fehér<sup>1</sup>, Bjorn Van Gasse<sup>1</sup>, Jos Van Den Begin<sup>2</sup>, José C. Martins<sup>1,\*</sup> and Annemieke Madder<sup>2,\*</sup>

<sup>1</sup>Department of Organic and Macromolecular Chemistry, NMR and Structure Analysis Unit, Ghent University, Ghent, Oost-Vlaanderen 9000, Belgium and <sup>2</sup>Department of Organic and Macromolecular Chemistry, Organic and Biomimetic Chemistry Research Group, Ghent University, Ghent, Oost-Vlaanderen 9000, Belgium

Received July 18, 2014; Revised November 24, 2014; Accepted December 02, 2014

## ABSTRACT

The predictable 3D structure of double-stranded DNA renders it ideally suited as a template for the bottom-up design of functionalized nucleic acid-based active sites. We here explore the use of a 14mer DNA duplex as a scaffold for the precise and predictable positioning of catalytic functionalities. Given the ubiquitous participation of the histidine-based imidazole group in protein recognition and catalysis events, single histidine-like modified duplexes were investigated. Tethering histamine to the C5 of the thymine base via an amide bond, allows the flexible positioning of the imidazole function in the major groove. The mutual interactions between the imidazole and the duplex and its influence on the imidazolium pK<sub>aH</sub> are investigated by placing a single modified thymine at four different positions in the center of the 14mer double helix. Using NMR and unrestrained molecular dynamics, a structural motif involving the formation of a hydrogen bond between the imidazole and the Hoogsteen side of the guanine bases of two neighboring GC base pairs is established. The motif contributes to a stabilization against thermal melting of 6°C and is key in modulating the pK<sub>aH</sub> of the imidazolium group. The general features, prerequisites and generic character of the new pK<sub>aH</sub>-regulating motif are described.

## INTRODUCTION

The imidazole moiety contributed by histidine residues is a ubiquitous functionality in proteins and peptides and is known to be involved in many binding and catalysis events. Prominent examples of catalytic sites where imidazole(s) occupies a central role in the mechanism of action

include serine proteases, such as  $\alpha$ -chymotrypsin (1–3), and ribonucleases, such as RNase A (4–6). With a pK<sub>aH</sub> value around physiological pH that can be additionally fine-tuned through specific interactions, both the charged (protonated ImH<sup>+</sup>) and the neutral (Im) state can occur in natural systems conferring a versatile general acid/general base character whenever required for binding or catalysis. Not surprisingly, therefore, considerable interest exists in the design of artificial enzymes featuring one or multiple imidazole moieties (7–10). Since this requires a precise organization of the catalytically active moieties, considerable effort has been invested in the *de novo* design of artificial, protein-based enzymes, wherein the engineering of a precise structural organization remains a formidable challenge.

While double-stranded DNA lacks the portfolio of protein side chain functionalities, it features a well-defined and mostly predictable 3D structure. While this should simplify the design efforts, the potential of nucleic acids as potential binders or catalysts has only recently been recognized. A series of successful examples can be found in the work of Roelfes *et al.* (11) where ligands intercalating in a double helix are used to carry out asymmetric transformations using mostly unmodified oligonucleotides. Furthermore, the SELEX (Systematic Evolution of Ligands by EXponential enrichment) technology (12–15) has revolutionized the field of nucleic acid-based binders or so-called aptamers. However, the innate absence of functional side chains in natural nucleic acids and the corresponding nucleotides poses clear limitations in the diversity of binders that can be obtained compared to protein-based ones. Examples exist that circumvent this issue through the introduction of modified residues thereby yielding new non-natural binders (16–19) and catalysts (20–24). However, the SELEX protocol becomes more labor intensive and reproducibility of results is often a problem in these cases. Moreover, no control is possible in terms of positioning of the desired extra functionalities.

\*To whom correspondence should be addressed. Tel: +32 9264 44 72; Fax: +32 9 264 49 98; Email: Annemieke.Madder@UGent.be  
Correspondence may also be addressed to José C. Martins. Email: Jose.Martins@UGent.be  
Present address: Bjorn Van Gasse, Chevron Technology Ghent, Ghent, Oost-Vlaanderen 9000, Belgium.

†These authors contributed equally to the paper as first authors.

As opposed to the SELEX-based *top-down* approach where a rather random and combinatorial method is applied in the search for enhanced nucleic acid-based binders and catalysts, we became interested in a bottom-up approach for the rational introduction of relevant protein-like functionalities in nucleic acid-based systems. This can be enabled by techniques, such as solid phase synthesis, post-synthetic modifications or enzymatic incorporation of modified analogs (25,26). Here, the systematic introduction via solid phase synthesis of a single imidazole moiety in a duplex DNA structure is presented using T<sup>Im</sup>, a thymine nucleoside bearing an imidazole at C5 (Figure 1).

Because of the stable and predictable Watson–Crick base pairing and stacking interactions, oligonucleotide duplexes allow the design of functional sites in one of the grooves through the carefully planned introduction of functional side chains onto the nucleoside building block. However, one must also consider the impact on the duplex integrity, stability and local structure induced by the modification. A complete picture therefore requires a detailed structural analysis and verification of the modified duplexes in order to validate the design approach. Here, we present a short sequence motif capable of regulating the pK<sub>aH</sub> of an imidazole functionality tethered to thymine using a short but flexible linker. Using nuclear magnetic resonance (NMR) spectroscopy and unrestrained molecular dynamics (MD), the influence of an incorporated imidazole moiety on duplex stability and structure are delineated, while reciprocally, the impact of the DNA context on the imidazole pK<sub>aH</sub> through specific interactions involving base pairs in the major groove are presented. The here described pK<sub>aH</sub>-regulating motif, its portability across other sequences and the procedures used to extract these findings can be of interest for their application in nucleic acid-based binders with enhanced affinities, artificial ribonucleases and potentially for oligonucleotide-based artificial catalysts.

## MATERIALS AND METHODS

### Synthesis of T<sup>Im</sup>-modified DNA

The synthesis of the imidazole-modified nucleoside from thymidine is achieved in three steps as described by Holmes and Gait (27). In a first step, a palladium-catalyzed one-pot carboxamidation reaction is used to attach histamine (Sigma-Aldrich) to 5-iodo-2'-deoxyuridine (Sigma-Aldrich) via an amide bond. Subsequent *t*Boc-protection of the imidazole N-H and DMTr-protection of the hydroxyl group on the 5'-position completes the synthesis of a T<sup>Im</sup> nucleotide building block suitable for incorporation into a DNA sequence via phosphoramidite chemistry using standard protocols (28). Resonance assignment of the imino protons in the T<sub>8</sub>-T<sub>21</sub> mismatch was achieved by incorporating 5–10% uniformly <sup>15</sup>N-labeled thymidine (CIL/Buchem) at T<sub>8</sub> in the unmodified duplex, and opposite T<sup>Im</sup> in the modified duplexes (29,30). All non-modified, non-labeled strands were purchased from Integrated DNA technologies (IDT, Belgium). Both purchased and synthesized oligonucleotide strands were purified using an ion exchange method (31) in which the oligonucleotides were dissolved in a 95/5 mixture of isopropanol and a 3M NaOAc

solution. The mixture was centrifuged (10 min, 7000 revolutions per minute (rpm)) and the oligonucleotide precipitate was resuspended in pure isopropanol and subsequently freeze-dried. The molar quantities of single strands were determined spectrophotometrically by measuring the absorbance at room temperature at 260 nm using a  $\epsilon = 99600 \text{ M}^{-1}\text{cm}^{-1}$  (Trinean DropSense 96). The desired double-stranded DNA was obtained from annealing equimolar amounts of the complimentary single strands by heating to 95°C for 30 min, followed by gradual cool-down to room temperature. The annealed duplex sample was dialyzed extensively against H<sub>2</sub>O, using a dialysis membrane with a 3.5 kDa molecular weight cut-off.

### Ultraviolet (UV) thermal denaturation experiments

UV measurements for the determination of the T<sub>m</sub> values were performed on a Varian Cary 300 Bio UV/VIS spectrophotometer equipped with a six-cell thermostatted cell holder and N<sub>2</sub>-supply. Melting curves were monitored at 260 nm from 5°C to 90°C with a heating/cooling rate of 0.3°C/min. Three consecutive runs were performed for each duplex, and values represent averages with standard deviations of these runs. The buffer solution contained 100 mM NaCl and 10 mM phosphate (pH 7.0). The oligonucleotide concentration was 1  $\mu\text{M}$  for each strand. The melting temperatures were calculated from the first derivative of the heating curves using the Cary 300 Bio software.

### NMR studies of the modified oligonucleotides

DNA duplexes were dissolved in 550  $\mu\text{l}$  90/10 H<sub>2</sub>O/D<sub>2</sub>O containing 100 mM NaCl, 0.1 mM EDTA, 0.05 mM NaN<sub>3</sub> and 0.05 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) for internal chemical shift referencing. For the assignment of the non-labile protons, the samples were freeze-dried and resuspended in 550  $\mu\text{l}$  of D<sub>2</sub>O sample solution. The pH or pD (corrected for the isotope effect) of the samples was adjusted by adding small aliquots of diluted NaOH and HCl or deuterated equivalents, respectively. Assignment was performed at pH 5.0. All spectra were recorded on a Bruker Avance II spectrometer operating at a <sup>1</sup>H frequency of 700.13 MHz operating under Topspin 3.1pl6 and using a standard 5 mm inverse TXI-Z ATMA probe head. All measurements for the assignment of the non-labile protons were performed at 25°C (298.15 K), while the 2D NOESY spectra for the assignment of the exchangeable protons were recorded at 5°C (278.15 K). Standard pulse sequences from the Bruker library were used throughout. In D<sub>2</sub>O solution, the residual HDO signal was suppressed using excitation sculpting when necessary (32). All spectra extended over 25.0 ppm (90/10 H<sub>2</sub>O/D<sub>2</sub>O) or 12.0 ppm (D<sub>2</sub>O) along the <sup>1</sup>H dimension. For the <sup>1</sup>H 1D spectra (zgesgp), 128 scans of 16K data points were accumulated. Prior to Fourier transformation, the FID's (Free Induction Decay) were zero filled to 32K and were apodized with a squared cosine bell function. Zero-order polynomial baseline corrections were applied. For the COSY (cosydfesgpppp), TOCSY (clmlevesgpph) and NOESY (noesyfesgpph) spectra, 512 t1 increments of 4K data points, 64 scans each, were recorded. Mixing times for the TOCSY and NOESY were



**Figure 1.** Sequence of the wt (unmodified) DNA template. ‘Up’ follows increasing sequence numbers in the 5′ to 3′ direction and ‘down’ is used in the 3′ to 5′ direction, with respect to a specific position in the strand. The sites where  $T^{Im}$  is introduced and the T-T mismatch are highlighted.

75 and 200 ms, respectively. Apodization with a squared cosine bell function, followed by zero filling and Fourier transformation to a  $2K \times 2K$  data matrix, followed by baseline correction yielded the final spectra.

For the spectra recorded in 90:10  $H_2O/D_2O$ , the  $H_2O$  signal was suppressed using WATERGATE (33) to account for the larger spectral width (25.00 ppm) 32K data points were collected. For the assignment of imino protons contributed by  $^{15}N$ -labeled thymidines,  $^{15}N$ -decoupled  $^1H$  1D (zggpwgcf3) spectra were recorded and compared to regular spectra to detect the  $^{15}N$ -satellites in a 0.5 mM duplex solution. For the assignment of the imino protons, a NOESY spectrum (noesyfpgpphwg) with 80 ms mixing time was recorded, with 512 t1 increments of 4K data points, 64 scans each and processed as described above. The CCPN data model was used for the complete assignment of all duplexes and provides chemical shift data for both exchangeable and non-exchangeable protons (34). This afforded the collection of  $^1H$  chemical shift perturbations caused by introduction of the imidazolium group and the identification of nOe contacts involving the  $T^{Im}$  residue in the duplexes.

### $T^{Im}$ $pK_{aH}$ determination

The  $pK_{aH}$  of the imidazole in the  $T_x^{Im}$  was determined by monitoring the pH dependence of one or both of the aromatic proton resonances of the imidazole ring, i.e.  $He1$  and  $H\delta2$ , with 1D  $^1H$  NMR (35,36). In spite of the considerable  $^1H$  background signal from the duplex, the lack of other similar functionalities within the DNA structure is exploited to monitor these characteristic resonances. The  $He1$  resonance occurs mostly downfield from the DNA spectral background, above 8 ppm, while the  $H\delta2$  resonance resides mostly upfield from the aromatic  $H_6$  and  $H_8$  protons of the nucleobases. When the imidazole resonances are partially obscured by the duplex background, the mutual long-range  $^4J(H\delta2, He1)$  coupling ( $\sim 1$  Hz) is exploited to generate a characteristic cross-peak in a 2D TOCSY spectrum with 60 ms spinlock time. Since no other protons in this spectral region show any  $^1H$ - $^1H$  couplings long range or otherwise, the unambiguous assignment of both the  $He1$  and  $H\delta2$  protons from this cross-peak is straightforward. As a result, complete *a priori* spectral assignment of the duplex is not necessary for  $pK_{aH}$  determination. In all cases,  $pK_{aH}$  values were obtained by fitting the chemical shift data obtained between pH values 5 and 10 to a Henderson–Hasselbalch equation. Error values were calculated by using an in-house

written Monte Carlo-based algorithm (37). All reported  $pK_{aH}$  values were derived from the  $He1$  chemical shift data since this proton can mostly be observed without any signal overlap. Values obtained, whenever possible, from  $H\delta2$ -derived data agreed within error limits with that from  $He1$ .

### Generation of DNA models

Initial non-modified DNA structures for all systems were built using the standard B-DNA parameters in Discovery Studio (Accelrys Software Inc., San Diego, USA, release 4.0). The  $T^{Im}$  nucleoside was created starting from thymidine as a template using the builder functionality. Topologies were constructed using the xLEAP interface and AmberTools 12 (38). The parmbc0 refinement was applied for all nucleic acids, whereas the protonated imidazole ring and subsequent linker parameters were taken from the relevant histidine force field parameters for proteins. The amide bond, attaching the protein-like modification to the thymine base, is described by parameters derived from the AMBER ff12SB force field. Each oligomer was solvated in a truncated octahedral TIP3P water box and placed at least 13 Å (1 Å = 0.1 nm) removed from the edges of the water box. Sodium counter ions were inserted in the water box both for the modified and non-modified DNA structures (25 respectively 26) in order to ensure charge neutrality of the complete system.

### Molecular simulations

All unrestrained MD simulations started following a short standard equilibration protocol that consisted of submitting the initial DNA structures to a two-step energy minimization, first with harmonic restraints (500  $kcal.mol^{-1}.\text{Å}^{-2}$ ) on all nucleic acid atoms, then without restraints. Each system was subsequently heated to 300K during a 20 ps NVT run, with weak restraints (10  $kcal.mol^{-1}.\text{Å}^{-2}$ ) on the solute. Next, the density of the solvent was allowed to relax during a subsequent 100 ps run, using constant pressure and without imposing any restraints on the DNA structure. Finally, a 50 ns production MD was performed using the NPT ensemble, with snapshots collected every 2 ps.

All MD simulations were run on a single Nvidia GTX680 GPU using the GPU implementation of pmemd provided in the AMBER12 simulation package (39,40). Each simulation was performed using periodic boundary conditions

and a 15 Å cut-off for the particle mesh Ewald summations (41) of non-bonding interactions. The SHAKE algorithm (42) was applied to bonds involving hydrogen atoms, allowing a 2 fs time step. Temperature was regulated using Langevin dynamics (43) with a collision frequency of  $1.0 \text{ ps}^{-1}$ . Finally, pressure regulation was obtained using isotropic position scaling with a pressure relaxation time of 2.0 ps. Using this setup, the complete calculation took about 96 h for each DNA duplex.

The calculated trajectories were visualized using VMD 1.9.1 (44) while further in-depth analysis was performed using 3DNA (45) for the structural and conformational DNA parameters and the HBonanza (46) python script for hydrogen bond analysis. Further data post-processing and analysis utilized various VMD plugins complemented with custom-made MATLAB scripts.

Persistence and quality of base pair hydrogen bond formation, the extent of the  $\alpha/\gamma$  conformational space sampled and the torsional parameters ( $\epsilon/\zeta$ ) indicative of  $B_I/B_{II}$  conformations were all monitored (see Supplementary Figures S14–S17) for the non-modified wild-type (wt)-duplex and all  $T_x^{\text{Im}}$ -modified systems (47–49). In addition, the integrity of the B-DNA duplex form during simulation was checked and found to be respected (see Supplementary Figures S18 and S19).

## RESULTS AND DISCUSSION

### Introducing an imidazole function into the major groove

Equipping DNA duplexes with additional functionalities can be most simply achieved by the introduction of custom-made nucleosides through modification at the level of sugar, phosphate backbone or base. Introduction of functionalities at C5 of pyrimidine bases positions the newly introduced moieties in the chiral and hydrophobic microenvironment of the major groove where sufficient space is available for accommodation of multiple functionalities and/or binding of potential substrate molecules. Here, histamine is tethered to the C5 of thymidine via an amide bond, placing the imidazole function in the major groove. Only minimal disruption of the helical structure is expected, in contrast with, for instance, modifications at the C2' of the ribose sugar (50). This is validated from the MD results described further.

### Selection of the DNA template sequence

The sequence of the DNA template used throughout this work was selected based on three criteria. First, insertion of a modified nucleotide can reduce the overall duplex stability. Therefore, the sequence should be of adequate length and composition to offset such effects. Second, the ideal template should have sufficient AT base pairs distributed throughout its sequence in order to assess the impact of the modification's location on the properties of the duplex. In this respect, varying the  $T^{\text{Im}}$  location on the same template allows for comparative analysis. Furthermore, a judicious distribution should allow to envisage a variety of catalytic configurations when exploring multiple  $T^{\text{Im}}$ -modified duplexes. Third, choices regarding the sequence length and composition should aim to facilitate spectral interpretation

by NMR, e.g. by avoiding local and global symmetries. Also, sequences of considerable length will have a negative impact on synthesis efficiency, further compromising NMR analysis due to limited sample availability.

Taking these considerations into account, the sequence shown in Figure 1 was chosen. As fraying destabilizes base pairs at the duplex ends and a full turn of B-DNA requires about 11 base pairs, a 14 base pair duplex was used as template. To limit the extent of fraying, the duplex was prolonged at each end with one and two GC base pairs, respectively.

To meet the design criteria, a variety of DNA sequences were tested for their overall stability using UV measured melting temperature values ( $T_m$ ). In addition, inspection of the number and line-width of resonances in the imino region of  $^1\text{H}$  NMR spectra (5–65°C, see Supplementary Figure S4) was used to establish individual base-pair stability in a qualitative manner, together with overall spectral dispersion and appearance of the  $^1\text{H}$  NMR spectrum. The final duplex sequence, hereafter referred to as the wt-duplex shown in Figure 1, represents a trade-off with respect to duplex stability and spectral simplicity, while maintaining sufficient modification points to study the impact of a single modification on the properties of the imidazole function and the duplex template. Since ultimately two (or more)  $T^{\text{Im}}$  modified building blocks are to be introduced, the location of AT base pairs was also optimized to allow variation in mutual spatial location. With the introduction of multiple  $T^{\text{Im}}$  building blocks at the same level in the DNA sequence in mind, a  $T_8-T_{21}$  mismatch was introduced at the center of the sequence. At 48.6°C, the  $T_m$  value drops by  $\sim 10^\circ\text{C}$  compared to the full Watson–Crick paired duplexes  $A_8T_{21}$  and  $T_8A_{21}$  (Table 1 (51,52)). However, it enables future investigations of double modification at the same base pair level. As an added bonus, the central mismatch leads to an increase in spectral resolution beneficial for the NMR assignments.

Initially, four modified duplexes featuring one  $T^{\text{Im}}$  were investigated and their properties compared to the wt-duplex at pH 7.0. They are referred to as the  $T_6^{\text{Im}}$ ,  $T_7^{\text{Im}}$ ,  $T_8^{\text{Im}}$  and  $T_{21}^{\text{Im}}$  duplexes, indicating the incorporation of the imidazole functionalized thymidine at positions 6, 7, 8 and 21 (Figure 1).  $T_x^{\text{Im}}$  will be used when simultaneously referring to several of these modified duplexes.

### The introduction of $T^{\text{Im}}$ does not destabilize the DNA duplex

The  $T_m$  values measured by UV for all  $T_x^{\text{Im}}$  at pH 7.0 (Table 1) demonstrate that none of the modified systems are destabilized by the introduction of the imidazole-tethered thymidine. A modest increase is found for the  $T_6^{\text{Im}}$  (+1.9°C) and  $T_7^{\text{Im}}$  (+1.5°C) duplexes compared to the wt-duplex. In contrast, introducing  $T^{\text{Im}}$  at the level of the T-T mismatch results in a significant increase in  $T_m$  of 6°C for the  $T_8^{\text{Im}}$  and 4.7°C for  $T_{21}^{\text{Im}}$  duplexes. This implies that the overall stability of the modified DNA duplex is enhanced to a varying degree by a contribution of the modification and that part of the loss in duplex stability resulting from the presence of the mismatch is undone when introducing one  $T^{\text{Im}}$  at this position. More insight into the latter follows from comparing fully Watson–Crick paired unmodified and modified  $AT^{\text{Im}}$  duplexes, all lacking the T-T mismatch. In the case

**Table 1.** Overview of melting temperatures and pK<sub>aH</sub> values for all systems studied

System	T <sup>Im</sup>	T-T mismatch				
		WT	T <sub>6</sub> <sup>ImH+</sup>	T <sub>7</sub> <sup>ImH+</sup>	T <sub>21</sub> <sup>ImH+</sup>	T <sub>8</sub> <sup>ImH+</sup>
T <sub>m</sub> (°C) <sup>a</sup>	n.r. <sup>c</sup>	48.6 ± 0.5	50.5 ± 0.2	50.1 ± 0.3	53.3 ± 0.3	54.6 ± 0.1
pK <sub>aH</sub> <sup>b</sup>	7.17 ± 0.02	/	8.13 ± 0.04	7.92 ± 0.04	7.70 ± 0.02	8.88 ± 0.05
		TA		AT		AT scrambled
System	T <sub>8</sub> A <sub>21</sub>	T <sub>8</sub> <sup>ImH+</sup> A <sub>21</sub>	A <sub>8</sub> T <sub>21</sub>	A <sub>8</sub> T <sub>21</sub> <sup>ImH+</sup>	sT <sub>21</sub>	sT <sub>21</sub> <sup>ImH+</sup>
T <sub>m</sub> (°C)	58.9 ± 0.2	64.1 ± 0.6	57.2 ± 0.2	59.3 ± 0.5	53.2 ± 0.3	59.4 ± 0.2
pK <sub>aH</sub>	/	8.72 ± 0.02	/	7.62 ± 0.03	/	9.07 ± 0.02

Notes:<sup>a</sup>T<sub>m</sub> values were determined in 100 mM NaCl, 10 mM phosphate buffer, 1 μM duplex at pH = 7.0.

<sup>b</sup>pK<sub>aH</sub>-values for the imidazolium group were established using NMR monitoring of selected resonances, as explained in the text.

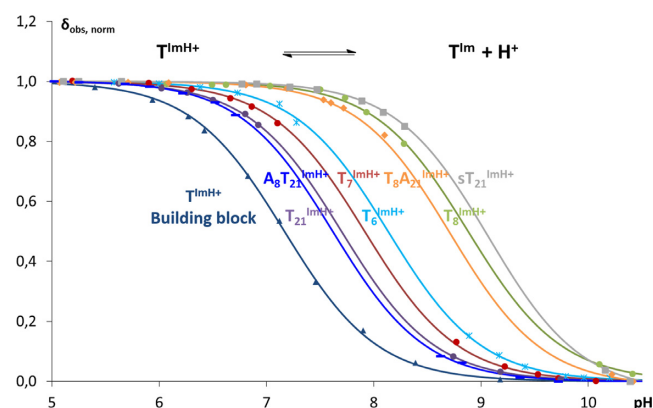
<sup>c</sup>n.r. not relevant, since T<sup>Im</sup> represents the building block itself rather than a DNA duplex and therefore there is no melting process. In addition, see also Supplementary Table S1.

of A<sub>8</sub>T<sub>21</sub><sup>Im</sup>, stabilization is reduced from 4.7°C to 1.9°C, a value similar to T<sub>6</sub><sup>Im</sup> and T<sub>7</sub><sup>Im</sup> duplexes. This supports a view where introduction of T<sup>Im</sup> at position 21 in the mismatched sequence partially compensates the destabilizing effect caused by the mismatch at this position. For T<sub>8</sub><sup>Im</sup>A<sub>21</sub>, however, stabilization remains high at 5.2°C compared to 6°C in the original mismatched T<sub>8</sub><sup>Im</sup> duplex (Table 1). The fact that stabilization is maintained irrespective of the presence of a mismatch in the duplex sequence points to a fundamentally different origin in T<sub>8</sub><sup>Im</sup> compared to all other T<sub>x</sub><sup>Im</sup> sequences.

### pK<sub>aH</sub> study of T<sup>Im</sup>-modified duplexes

The stabilization apparent in T<sub>x</sub><sup>Im</sup> duplexes suggests the possibility of an interaction involving the imidazole function, the tether or both. Given the negative charge of the phosphate backbone, an electrostatic interaction involving ImH<sup>+</sup> appears reasonable as a stabilizing interaction that should contribute to all systems. If present, a concurrent shift in the deprotonation equilibrium of ImH<sup>+</sup> should be apparent, leading to a rise in pK<sub>aH</sub> that may correlate with the trend in melting temperature. Since the large UV background of the DNA structure precludes UV-based pK<sub>aH</sub> determination, 1D <sup>1</sup>H NMR was used instead. pH-induced chemical shift changes of the non-exchangeable Hε1 and Hδ2 imidazole resonances are followed, as detailed in Materials and Methods. The experimental data and fitted titration curves are presented in Figure 2 while pK<sub>aH</sub> values are collected in Table 1. At 7.17, the pK<sub>aH</sub> of the T<sup>Im</sup> nucleoside falls within the pK<sub>aH</sub> range reported for imidazole (7.0) and 2- or 4-methylimidazole (7.75 and 7.9) respectively (53,54).

Once incorporated in the duplex sequence, the pK<sub>aH</sub> value in the T<sub>x</sub><sup>Im</sup> systems increases between 0.53 and 1.75 units compared to the T<sup>Im</sup> nucleoside. All pK<sub>aH</sub> values show that, at pH 7.0, the imidazole group is protonated for ~90% or more in all T<sub>x</sub><sup>Im</sup> duplexes. Thus, a positively charged imidazolium is present in all cases, capable of contributing stabilizing electrostatic interactions with the negatively charged phosphate backbone. At 8.88 ± 0.18, the pK<sub>aH</sub> of T<sub>8</sub><sup>Im</sup> is 0.9 units higher than the average pK<sub>aH</sub> value of the other three T<sub>x</sub><sup>Im</sup> duplexes (7.92 ± 0.03). This significantly higher pK<sub>aH</sub> is well maintained (8.72 ± 0.02)



**Figure 2.** Normalized observed chemical shifts as a function of pH for the isolated building block as well as for all studied T<sub>x</sub><sup>Im</sup> duplexes. Both the experimental data and fitted curves are shown. T<sup>Im</sup> and T<sup>Im</sup>H<sup>+</sup> refers to the neutral imidazole protonated and charged imidazolium state of the function, respectively.

in the T<sub>8</sub><sup>Im</sup>A<sub>21</sub> duplex, underlining that its higher value is independent from the T-T mismatch at this position. T<sub>21</sub><sup>Im</sup>, on the other hand, displays the lowest pK<sub>aH</sub> value, a feature maintained in the A<sub>8</sub>T<sub>21</sub><sup>Im</sup> duplex. Thus, introducing T<sup>Im</sup> on opposing sides of the T-T mismatch leads to a quite different effect on the imidazolium group's acidity. For T<sub>6</sub><sup>Im</sup> and T<sub>7</sub><sup>Im</sup>, the pK<sub>aH</sub> remains close to that observed for T<sub>21</sub><sup>Im</sup>. Most notable, however, is the existence of a correlation between the significantly higher pK<sub>aH</sub> and T<sub>m</sub> values irrespective of the mismatch when introducing T<sup>Im</sup> at position 8. This marked position-dependence suggests a specific interaction somehow tied into the local base pair context. This proposal is further supported by noting that when a T<sup>Im</sup> is introduced at the same base-pair level but on opposing strands, the space available for the imidazolium group within the major groove must necessarily be different when the integrity of the duplex is maintained.

### nOe-contact mapping and chemical shift perturbation of the $T_x^{\text{Im}}$ duplexes

The assignment of imino- and non-exchangeable resonances for all  $T_x^{\text{Im}}$  was achieved using established methods for unlabeled oligonucleotides (55–57), and was performed at pD/pH 5 to ensure full protonation of the imidazolium group. The NMR spectra, in terms of sequential nOe contacts and intraresidue COSY intensities, are consistent with a right-handed double helix in the B-family of conformations. From these, chemical shift perturbations and nOe contacts involving the modified  $T^{\text{Im}}$  could be analyzed. In the case of the  $T_8^{\text{Im}}$  duplex, clear chemical shift perturbations larger than 0.1 ppm extend up-sequence from the modified base pair to the  $A_{11}$ - $T_{18}$  base pair, i.e. 3 base pairs from the  $T_8$ - $T_{21}$  mismatch (see Supplementary Figure S8). A marked up-field chemical shift change of 0.46 ppm is observed for the  $T_{18}$  methyl group, indicating a major change in the local environment of the major groove. This is in line with a set of distinct nOe contacts (Figure 3B) that connects the He1 and Hδ2 protons of the imidazole ring to H1 and H8 of  $G_{19}$  and H1 of  $G_9$  (position +1), H5 of  $C_{10}$  (position +2) and H3, H1', H2' and H2'' of  $T_{18}$  (position +3). These indicate that the imidazolium group extending from the  $T_8^{\text{Im}}$  base folds back into the major groove toward the base pairs up-sequence from  $T_8^{\text{Im}}$ . A long-range nOe to the  $T_{18}$  methyl group accompanies the aforementioned up-field shift for the latter. The extent of chemical shift perturbation (Supplementary Figure S8) and cluster of nOe's observed for the  $T_8^{\text{Im}}$  duplex contrasts sharply with the  $T_6^{\text{Im}}$ ,  $T_7^{\text{Im}}$  and  $T_{21}^{\text{Im}}$  ones, where chemical shift perturbations quickly wane in number and magnitude beyond the immediate neighboring base-pairs. Even more indicative are the nOe's involving the imidazolium that are scarce and very weak. For instance,  $T_7^{\text{Im}}$  only displays a nOe contact between He1 and the methyl of the neighboring  $T_{21}$  base (see Figure 3A). Another element differentiating  $T_8^{\text{Im}}$  is that it is the only modified duplex where the exchangeable Ne2 proton is in slow exchange with the solvent, allowing its tentative assignment at pH 5 and lower temperature (5°C) and the extraction of a number of nearest neighbor nOe contacts (see Supplementary Figure S23). The interactions in the  $T_6^{\text{Im}}$ ,  $T_7^{\text{Im}}$  and  $T_{21}^{\text{Im}}$  systems are too transient to allow the visualization and identification of the exchangeable imidazole protons under these conditions. Combined with the melting temperature and pK<sub>aH</sub> studies, all data available for the  $T_8^{\text{Im}}$  system are compatible with a preferred and stable conformation of the imidazolium functionality inside the major groove, placing it in the vicinity of the  $C_{10}G_{19}$  and  $A_{11}T_{18}$  base pair steps. The formation of a specific interaction linked to this conformation could not only contribute to an increased thermal stability but also enhance the proton acceptor capacity of the imidazole and hence account for the observed significant increase in pK<sub>aH</sub> for the  $T_8^{\text{ImH}^+}$ . This hypothesis also implies that systems that lack the increase in stability and pK<sub>aH</sub> lack this specific interaction, as shown by the dearth of nOe contacts for the other  $T_x^{\text{Im}}$  systems.

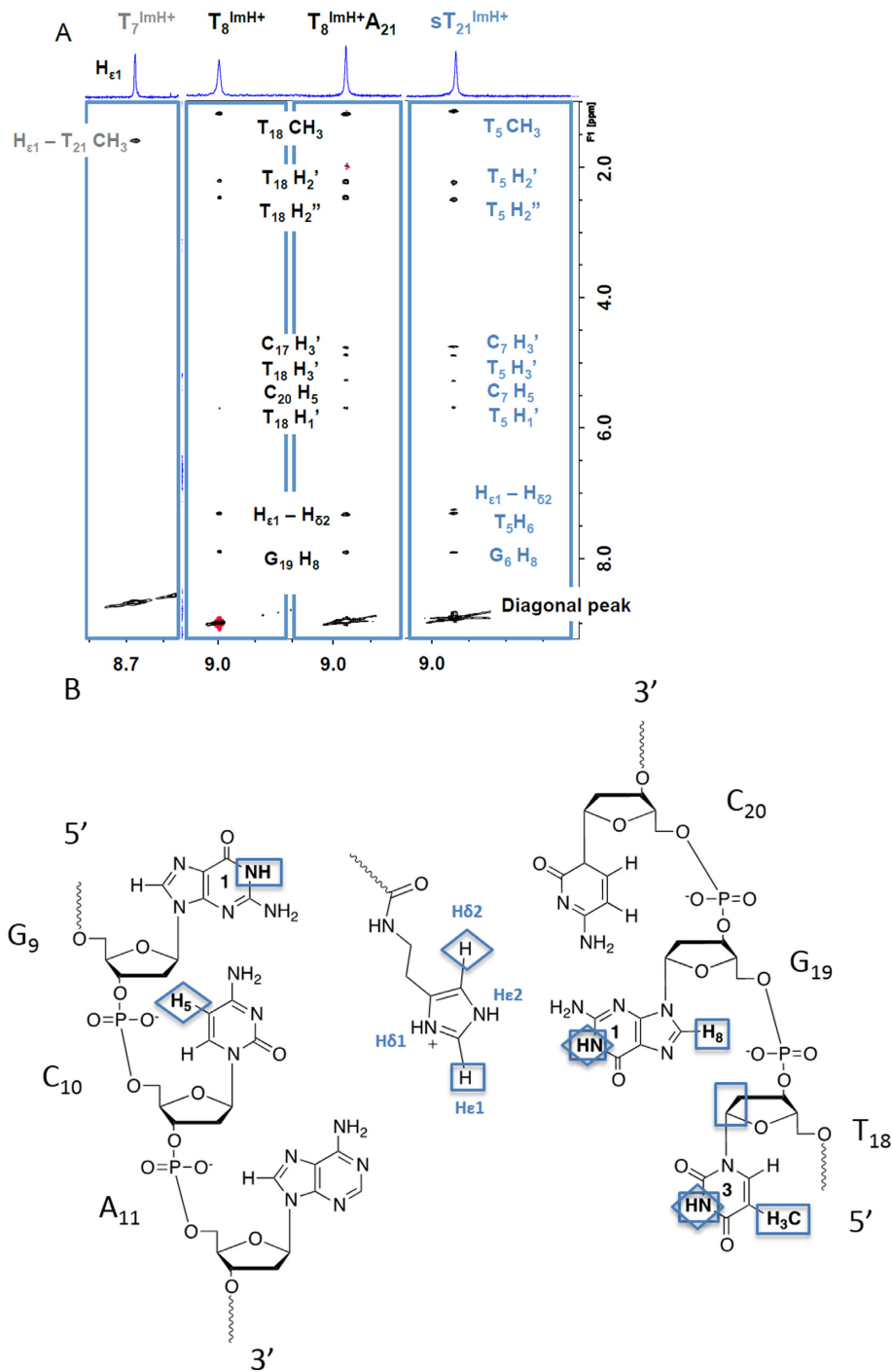
While distance restraints extracted from nOe data can be used for structure calculations, the sparse nature of the long-range nOe contacts along the linear DNA sequence,

combined with the inability to use  $^{13}\text{C}$  and  $^{15}\text{N}$  editing to harvest most of the nOe data would undoubtedly prevent to produce an ensemble of sufficient precision and accuracy to investigate the structural details of the modified duplexes, and the nature of the specific interaction in sufficient detail. As an alternative, therefore, we turned to unrestrained MD simulations, which were confronted with the nOe data, to further explore the specific nature of the DNA-imidazolium interactions in the  $T_x^{\text{Im}}$  duplexes.

### MD of the wt sequence and general impact of the $T^{\text{Im}}$ modification

MD is established as a reliable simulation method to study a broad range of biomolecules including DNA, with simulation times recently extending up to the millisecond timescales (58). Using a protocol similar to that described by Pauwels *et al.* (59), we used MD simulations of all  $T_x^{\text{Im}}$  duplexes in explicit solvent to investigate the conformational behavior of the imidazole moiety. Since all NMR data were recorded at pD 5.0, simulations used a charged imidazolium group, which is explicitly referred to as  $T^{\text{ImH}^+}$ . To minimize bias, all simulations started with the tether and imidazolium extended outwards from the major groove into the solvent and ran over 50 ns to allow sufficient sampling of conformational space to reveal specific interactions. Subsequently, the trajectories were also confronted with the experimental NMR data to obtain a complete picture.

In order to estimate the impact of the modification on the DNA structure as a whole, a reference simulation of the same time span was performed on the original non-modified wt sequence. This first simulation was also used to evaluate the overall quality of the MD trajectory and stability of the B-DNA type helix during the simulation by extracting a number of parameters known to be good indicators of the simulation quality (*vide supra*, (60)). When all these parameters are considered (Supplementary Figures S14, S16 and S18), the wt trajectory shows no unexpected behavior. The end base pairs show indications of fraying as can be expected, imino hydrogen bonds within the duplex sequence are persistent and the conformational space explored remains within the regions expected for the B-DNA conformational space. A notable exception is the  $G_9$  base neighboring the  $T_8$ - $T_{21}$  mismatch, which shows two excursions from the global minimum in the  $\alpha/\gamma$  conformational space. These occur toward the end of the simulation trajectory and involve two known and allowed local minima (47). A few other bases show similar tendencies of much shorter duration, but overall the structure and individual base pairs remain well within the regions expected of the B-DNA conformational space. Identical analysis of the relevant parameters for the  $T_8^{\text{Im}}$  duplex (Supplementary Figures S15, S17 and S19) demonstrates that the simulation trajectory of the modified systems features an overall conformational sampling that is highly similar to that of the wt trajectory. Changes are only apparent for the modified base itself ( $T_8^{\text{Im}}$ ), the opposite base ( $T_{21}$ ) and the base pairs in the immediate vicinity ( $G_9$ ,  $C_{10}$ ,  $T_6$  and  $A_{23}$ ). For instance, when analyzing the  $B_1/B_{II}$  torsional parameter for the  $T^{\text{Im}}$ -modified base, the majority of structures still feature the DNA backbone either in the  $B_1$  or  $B_{II}$  conformation, how-



**Figure 3.** (A) Overview of the specific nOe contacts of the He1 proton of  $T_8^{Im}$ ,  $T_8^{Im}A_{21}$ ,  $sT_{21}^{Im}$  and  $T_7^{ImH}$  to the DNA backbone (2D NOESY, 200 ms, 700 MHz, 25°C,  $D_2O$ ). (B) Schematic overview of the specific nOe-contacts from the  $T_8^{ImH^+}$  He1 (□) and Hδ2 (○) protons to the nucleotides up-stream from the modification site.

ever, some excursions outside these areas can be observed as well. Once a base pair is more than two positions up- or down-sequence from the modification site, no unexpected features can be discerned (see also Supplementary Figure S24 for the corresponding  $^{31}P$  spectra). Thus, the MD simulation trajectories are proposed to provide a good basis to explore the conformation and interaction of the ImH<sup>+</sup> moi-

ety in the  $T_8^{Im}$  duplex and by extension in all modified duplexes.

### MD reveals a specific interaction in the $T_8^{Im}$ duplex

Analysis of the hydrogen bond network involving the modified nucleoside over the MD simulation trajectory of  $T_8^{Im}$  shows that a first hydrogen bond is initially formed after 2 ns

between NH $\epsilon$ 2 of the imidazolium and N7 of G<sub>19</sub>, rapidly followed by hydrogen bond formation with the O6 along the Hoogsteen edge of G<sub>19</sub> (Figure 4). Both the donor-acceptor distance (<3 Å) and donor-hydrogen-acceptor angle (>150°) in the latter are satisfied for 30% of the simulation time. Even when both hydrogen bond criteria are not met, the donor-acceptor distance indicates that the imidazolium remains within the major groove, in close vicinity to the Hoogsteen edge of the G<sub>9</sub>C<sub>20</sub> and C<sub>10</sub>G<sub>19</sub> base pairs for over 90% of the total simulation time (Figure 5). This would also account for the 0.46 ppm upfield chemical shift change of the T<sub>18</sub> methyl group, which is situated close to the face of the imidazolium group inside the major groove. Alternative but very low persistence hydrogen bond interactions involving N7 of G<sub>19</sub> (1.3%) and O6 (1.8%) of G<sub>9</sub> complete the hydrogen bond interaction pattern involving T<sup>Im</sup> (Figures 4 and 5). In contrast to the T<sub>8</sub><sup>Im</sup> duplex, the imidazolium group in the other T<sub>x</sub><sup>Im</sup> duplexes remains overall solvent accessible only showing transient or low persistence interactions with the major groove of the duplex. These include low persistence hydrogen bonds formed between the H $\delta$ 1 of T<sup>Im</sup> and the phosphate backbone oxygen of the preceding base: A<sub>5</sub> (1.6%) in the T<sub>6</sub><sup>Im</sup> and T<sub>6</sub> (8%) in the T<sub>7</sub><sup>Im</sup> duplexes. In addition, the imidazolium H $\epsilon$ 2 proton in T<sub>6</sub><sup>Im</sup> appears involved in a hydrogen bond with the O4 carbonyl oxygen of the mismatched T<sub>21</sub> base on the opposite strand for 11.2% of the total simulation time, while for T<sub>7</sub><sup>Im</sup> the O4 of the T<sub>8</sub> base is involved, albeit only 2% of the simulation time (Figure 5).

When no productive hydrogen bond is formed, the imidazolium moiety is seen to make frequent (T<sub>6</sub><sup>Im</sup>) or prolonged (T<sub>7</sub><sup>Im</sup>) excursions away from the interaction sites toward the solvent, underlining the transient nature of the interactions. This compares well with the chemical shift perturbation data (see Supplementary Figures S5 and S6), where perturbations are limited to the base pair and the immediately neighboring base pairs. As for the nOe data, only few nOe-contacts are present in these T<sub>x</sub><sup>Im</sup> duplexes, which supports the transient nature of the hydrogen-bonding interactions inferred from the MD simulation. Thus, the behavior of both the T<sub>6</sub><sup>Im</sup> and T<sub>7</sub><sup>Im</sup> duplexes appears quite similar. In the MD simulation trajectory of T<sub>21</sub><sup>Im</sup>, transient hydrogen bonds involving NH $\delta$ 1 from the imidazole and a phosphate oxygen of either T<sub>21</sub> or C<sub>20</sub> as acceptor can be found for 7.5% and 12.8% of the simulation time. No other acceptors for hydrogen bonds involving the surrounding bases could be identified, in contrast with T<sub>6</sub><sup>Im</sup> and T<sub>7</sub><sup>Im</sup>. Again, frequent excursions of the imidazole moiety toward the linker are present (see Supplementary Figure S20). Since the chemical shift perturbation and nOe data is likewise limited to the effects mentioned for the T<sub>6</sub><sup>Im</sup> and T<sub>7</sub><sup>Im</sup> duplexes, good agreement is maintained between the experimental data and the simulation.

### Confrontation of nOe and simulation data for the T<sub>8</sub><sup>Im</sup> duplex

A total of 13 nOe's unambiguously assigned from the NOESY spectrum of T<sub>8</sub><sup>Im</sup> as shown in Figure 3B were matched to the simulation by extracting the associated interproton distance as a function of simulation time (Figure 6). These distances should consistently dip below the

5 Å mark in order to generate the observed nOe's. Since the simulation starts with an extended conformation for the T<sup>ImH+</sup> modification, each distance starts around 10–12 Å, and rapidly decreases as the imidazole moiety folds back on the duplex. Over the entire simulation trajectory, matching of the interproton distances corresponding to the eight nOe contacts involving the resolved H $\epsilon$ 1 resonance establishes that the one to the T<sub>18</sub> methyl averages below 5 Å (Figure 6A). The other distances associated with clearly discernable nOe contacts make regular excursions below 5 Å but average above the threshold value. The distance to G<sub>19</sub> H8 satisfies the threshold for half the simulation time in good agreement with the nOe data (Figure 6B). The H $\epsilon$ 1–G<sub>19</sub> H1 (imino) distance, generally close and on average above the threshold is indeed one of the weaker nOe contacts (Figure 6C). The same holds true for the T<sub>8</sub><sup>Im</sup> H $\epsilon$ 1–T<sub>18</sub> H3 (imino) (Figure 6D) and T<sub>8</sub><sup>Im</sup> H $\epsilon$ 1–G<sub>9</sub> H1 (imino) distance (Figure 6E). Only the T<sub>8</sub><sup>Im</sup> H $\epsilon$ 1–T<sub>18</sub> H1' distance (Figure 6F, green), also showing the weakest nOe contact, does not drop below 5 Å. Overall, the average distance values are in good agreement with the relative intensities of the nOe contacts. In view of the weakness of some nOe contacts, especially those involving the T<sub>18</sub> deoxyribose protons (Figure 6F), the more important deviations from the threshold value could actually result from spin-diffusion via common neighboring spins thereby underestimating the actual distance.

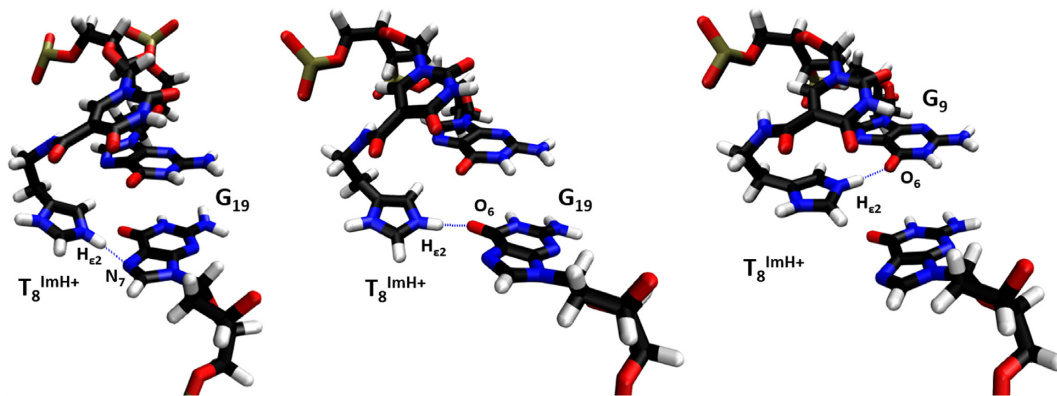
H $\delta$ 2, the other non-exchangeable proton present on the imidazole functionality, also shows a number of characteristic nOe contacts. Because this resonance is located in the H8-H6 region of the <sup>1</sup>H spectrum, assignment of relevant nOe contacts for analysis is difficult. Nevertheless, three characteristic through space contacts can be accessed: T<sub>8</sub><sup>Im</sup> H $\delta$ 2–C<sub>10</sub> H5, T<sub>8</sub><sup>Im</sup> H $\delta$ 2–G<sub>19</sub> H1 and T<sub>8</sub><sup>Im</sup> H $\delta$ 2–T<sub>18</sub> H3 (imino). As is the case with H $\epsilon$ 1, the contacts of H $\delta$ 2 concerning the imino protons (G<sub>19</sub> H1 and T<sub>18</sub> H3 (imino)) are weak, hereby suggesting the extracted distances with the respective protons should be close to 5 Å (Figure 6H and I). These average to 6.17 and 7.19 Å, respectively. On the other hand, the clear nOe to C<sub>10</sub> H5 is associated with an average distance of 5.75 Å (Figure 6G), and regularly dips below 5 Å in the simulation.

The final relevant nOe contacts concern the HA1 and HA2 protons of the ethylene tether closest to the amide bond, connecting the imidazole ring to the T<sub>8</sub><sup>Im</sup> base. As could be expected, these protons are in spatial proximity to neighboring T<sub>7</sub> methyl (Figure 6J, HA1 red and HA2 black), but yield distances averaging between 6.38 and 7.44 Å.

Thus, of the set of 13 nOe's, the most intense ones are seen to correspond to distances close to or below the 5 Å threshold, while weaker nOe's generate distances above, which correlate well with the ranking of the nOe intensities. Taking into account the unrestrained nature of the MD simulation, the overall agreement with the nOe data appears very satisfactory.

For all the T<sub>x</sub><sup>ImH+</sup> systems studied so far, the unrestrained MD confirms that only the T<sub>8</sub><sup>ImH+</sup> system shows a sustainable and specific hydrogen bond pattern of its imidazole functionality with the DNA scaffold. T<sub>6</sub><sup>ImH+</sup>, T<sub>7</sub><sup>ImH+</sup> and even T<sub>21</sub><sup>ImH+</sup> on the other hand do not. These observations substantiate the proposal that the markedly higher pK<sub>aH</sub>





**Figure 4.** Three specific hydrogen bond patterns of the  $T_8^{\text{ImH}^+}$  functionality with the  $G_9$  and  $G_{19}$  bases from the Hoogsteen side. The respective hydrogen bond donor and acceptor are indicated in each snapshot. Visualized using VMD 1.9.1.

for  $T_8^{\text{ImH}^+}$  is linked to the existence of a specific interaction motif. This motif apparently arises in a three base pair cassette composed of the modified  $T^{\text{Im}}$  base pair followed by a GC/CG sequence allowing for interaction between the imidazole and the hoogsteen side of a guanine at position +2 (opposing strand, here  $G_{19}$ ) or the guanine at position +1 (same strand, here  $G_9$ ) from the modification.

#### Importance of the $T_8$ - $T_{21}$ mismatch: should it stay or can it go?

An obvious question is whether a T-T mismatch is a prerequisite for the occurrence of this motif. Considering  $T_8^{\text{Im}}A_{21}$  in Table 1, the obtained  $\text{pK}_{\text{aH}}$  of  $8.72 \pm 0.02$  is very close to the original value of  $8.88 \pm 0.05$  for  $T_8^{\text{ImH}^+}$ . Similarly, with respect to the unmodified  $T_8A_{21}$  wt sequence, the melting temperature at pH 7.0 has increased with  $5.2^\circ\text{C}$  thereby indicating a significant rise in stability. The chemical shift of the  $T_{18}$  methyl resonance, a fingerprint for the interacting imidazolium group, also moves 0.44 ppm up-field as observed before in the  $T_8^{\text{Im}}$  duplex. Analyzing the nOe data involving the imidazolium modification for  $T_8^{\text{Im}}A_{21}$ , a set of near-identical nOe contacts can be found (Figure 3A), indicating the integrity and similarity of the interaction in the absence of the T-T mismatch. When mapping the specific nOe contacts involving  $T^{\text{ImH}^+}$  to the duplex, all the contacts present in the  $T_8^{\text{Im}}$  system also can be seen with  $T_8^{\text{ImH}^+}A_{21}$ . Three additional nOe contacts are visible as well (Figure 3A). The nOe connecting  $T_8^{\text{Im}}$  He1 to  $C_{20}$  H5 averages to  $6.67 \pm 2.30 \text{ \AA}$ , and makes regular excursions below the 5 Å threshold. The other two nOe's also involving He1 are either satisfied to a lesser extent or deemed to far as indicated from their average distances found to be  $8.84 \pm 2.47 \text{ \AA}$  for  $T_{18}$  H3' and  $9.70 \pm 1.72 \text{ \AA}$  for  $C_{17}$  H3', respectively. It is probable that spin diffusion, possibly enhanced due to rigidification as a result of removing the T-T mismatch, is the underlying reason for the appearance of these last two nOe contacts. Nevertheless, the new  $C_{20}$  H5 nOe contact appears to indicate the possibility of a somewhat different local conformation that still includes the interaction to the  $G_9$ - $C_{20}$  base pair.

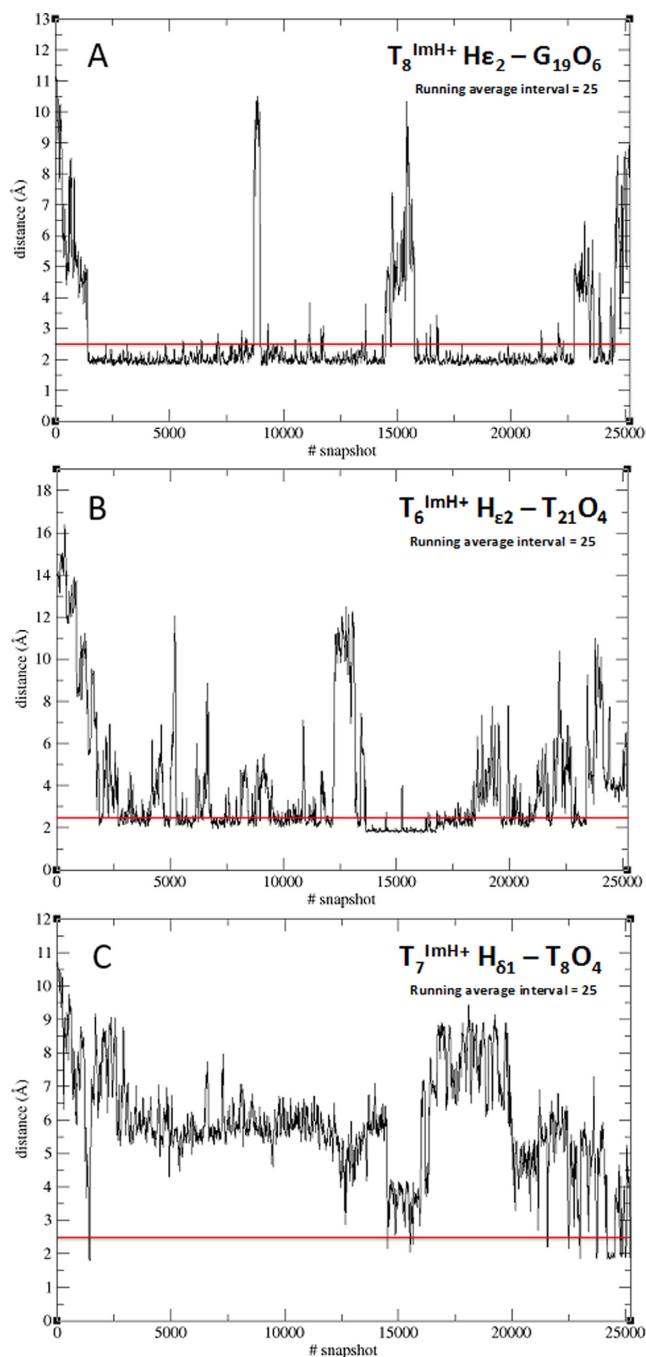
The comparison of the hydrogen bond pattern obtained from the complete MD trajectories for the  $T_8^{\text{ImH}^+}$  and

$T_8^{\text{ImH}^+}A_{21}$  duplexes does not reveal a significant difference (see Supplementary Figure S21). During the trajectory, the  $T^{\text{ImH}^+}$  functionality in the  $T_8^{\text{Im}}$  duplex engages in three distinct hydrogen bond patterns of which the  $T_8^{\text{Im}}$   $\text{NH}\epsilon 2$ - $\text{O}6$   $G_{19}$  is most persistent (29.7%), with transiently populated alternatives (<2%) involving the  $\text{O}6$  of  $G_9$  or  $\text{N}7$  of  $G_{19}$  (*vide supra*). In case of the  $T_8^{\text{Im}}A_{21}$  duplex, only two hydrogen bonds are present: the  $T_8^{\text{ImH}^+}$   $\text{NH}\epsilon 2$ - $\text{O}6$   $G_{19}$  one, which is almost equally persistent (27.1%), and  $T_8^{\text{ImH}^+}$   $\text{NH}\epsilon 2$ - $\text{O}6$   $G_9$  that is again only transiently populated (1.5%). Thus, no notable differences are observed with regards to the MD trajectories of both duplexes. In conclusion, all experimental and simulation data indicate the preservation of the specific imidazolium-DNA hydrogen bond interaction in  $T_8^{\text{ImH}^+}A_{21}$ , proving that the  $T_8$ - $T_{21}$  mismatch, though present in the current system for a well-defined reason (*vide supra*), is not a necessity for an imidazolium-mediated duplex interaction and concomitant stabilization. This further confirms the exact composition of the motif to consist of a 3 base pair 'cassette' wherein the  $T^{\text{Im}}$  nucleotide is followed by a GC/CG sequence and where the exact nature of the  $T^{\text{Im}}$  pairing partner base can be either T or A.

#### General applicability of the new motif: built to last?

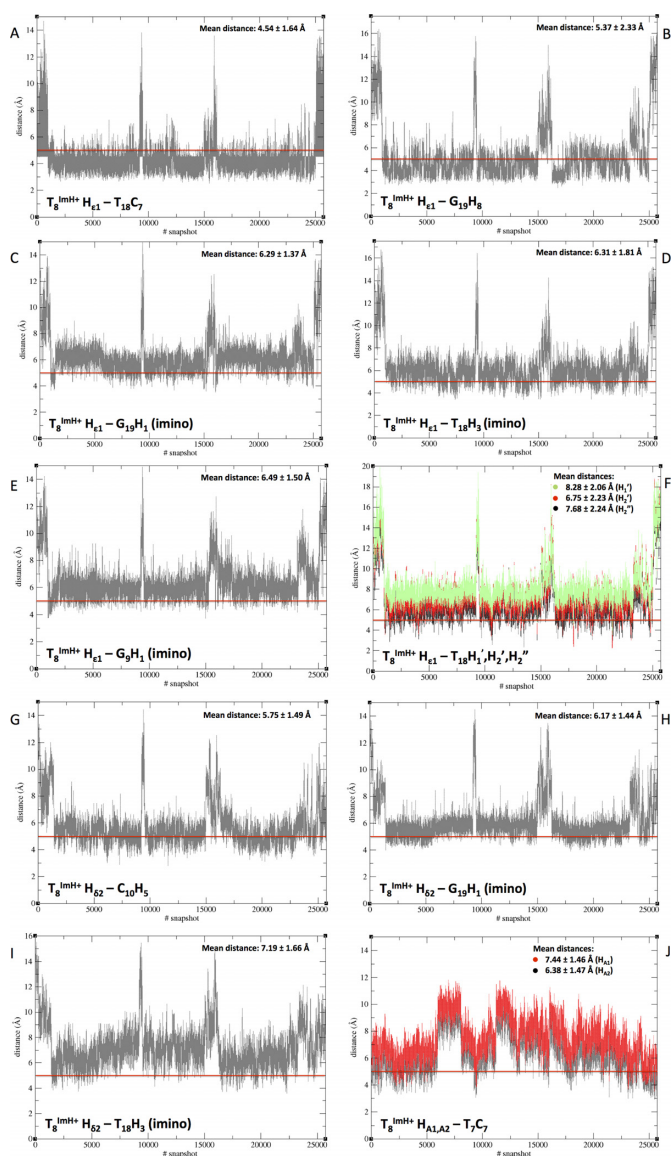
How robust is this potentially  $\text{pK}_{\text{aH}}$ -regulating motif? In other words, if this cassette would be introduced in a different sequence, will the specific interaction still be intact? This was addressed via a 'scrambled' modified and non-mismatched duplex,  $sT_{21}^{\text{Im}}$  (Figure 7).

Compared to the original  $T_8^{\text{Im}}A_{21}$  duplex, the cassette was translated three positions and rotated  $180^\circ$  to yield a  $G_6C_7A_8 \bullet T_{21}G_{22}C_{23}$  sequence, expected to generate the  $\text{pK}_{\text{aH}}$ -regulating motif. The  $A_5T_{24}$  base pair, preceding the motif in the wt-duplex, was flipped to  $T_5A_{24}$ , as this would position the  $T_{24}$  methyl in a similar position for an up-field chemical shift as observed in the  $T_8^{\text{Im}}$  and  $T_8^{\text{Im}}A_{21}$  duplexes. The melting temperature of  $sT_{21}^{\text{Im}}$  shows an increase of  $6.2^\circ\text{C}$  compared to the unmodified  $sT_{21}$  duplex, and the  $\text{pK}_{\text{aH}}$  is 9.07 (Table 1). Both values are clearly in line with those for the  $T_8^{\text{Im}}A/T$  duplexes indicating the presence of the interaction with the imidazolium group. In addition, a similar chemical shift perturbation profile (see supplement-

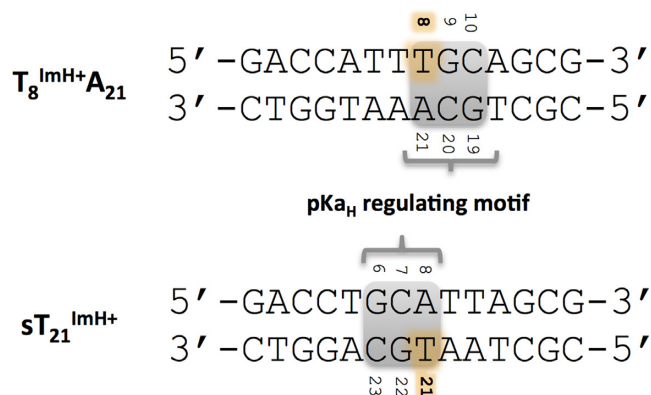


**Figure 5.** Overview of the persistence of the observed specific hydrogen bonds involving the  $T_6^{\text{ImH}^+}$ ,  $T_7^{\text{ImH}^+}$  and  $T_8^{\text{ImH}^+}$  bases in the corresponding duplexes. A running average over 25 structures is presented over the duration of the entire 50 ns trajectories.

tary Figure S10), including a 0.44-ppm up-field shift of the  $T_5$  methyl, and a set of specific nOe contacts can be observed (Figure 3A), concomitant with the expected  $T_{21}^{\text{Im}}$   $\text{NH}\epsilon_2\text{-O}\delta_6$   $G_6$  hydrogen bond in the MD trajectory, with a 35.3% persistence. These data provide final confirmation of the viability of the cassette as a  $\text{pK}_{\text{aH}}$ -regulating motif in DNA duplex sequences.



**Figure 6.** Overview of the extracted distances from the 50 ns trajectory run of  $T_8^{\text{Im}}$  which are confronted with the nOe contact limit at 5 Å, indicated by the red line on each graph.  $T^{\text{ImH}^+}$  explicitly refers to the protonated imidazolium state of the  $T^{\text{Im}}$  nucleoside at position 8. The abscissa refers to snapshots and can be converted to time (ps) by multiplying by two (one shot per two seconds).



**Figure 7.** Comparison of the original motif in the  $T_8^{\text{Im}}A_{21}$  system and the same motif in a new scrambled sequence  $ST_{21}^{\text{Im}}$ .

## CONCLUSIONS AND OUTLOOK

Imidazole residues have been engineered into the major groove of a DNA duplex. As demonstrated by detailed molecular modeling studies and experimentally verified by NMR data, a new functional imidazolium-based motif has been uncovered that increases the imidazolium  $pK_{aH}$ , with respect to the free modified nucleoside, by more than one unit. Having established the general applicability of this  $pK_{aH}$ -regulating motif, a summary of the different properties can be extracted. These in turn can be used as a checklist to confirm whether the specific interaction is present or absent. The modified system suspected of having a specific hydrogen bond interaction between the imidazole functionality and its DNA backbone must feature (i) a general increase in melting temperature of  $6 \pm 1^\circ\text{C}$  with respect to the non-modified wt sequence, (ii) show a clear downfield shift in the 1D  $^1\text{H}$  spectrum of the He1 resonance to  $\approx 9.0$  ppm in the presence of the interaction as opposed to  $\approx 8.7$  ppm without the hydrogen bridge, (iii) demonstrate a clear and specific set of 2D nOe contacts starting from He1 of the modification at position  $n$  to the base at position  $n + 3$  of the opposite chain, (iv) have a significant increase in  $pK_{aH}$  of the  $T^{\text{ImH}^+}$ -modification to  $\approx 8.7$ – $9.0$  or higher, where in the absence of the motif the  $pK_{aH}$  is located around a lower value of  $\approx 7.9$ . Fulfilling these requirements guarantees that the system at hand does exhibit the specific imidazole  $pK_{aH}$ -increasing interaction. The exploration of duplexes carrying multiple  $T^{\text{Im}}$  modifications wherein one or more  $pK_{aH}$  motifs contribute will be the object of future investigations using our combined NMR and MD approach.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

The 700MHz is part of the INMRF jointly operated by UGent, UA and VUB and funded by the Flemish Government via the FFEU–ZWAP initiative.

## FUNDING

University of Ghent [1.5.186.03 to V.G., B.V.G. and J.V.D.B.]; Research Foundation – Flanders (FWO) [G.0422.13 to D.B.]. Funding for open access charge: Flanders (FWO) [G.0422.13].

Conflict of interest statement. None declared.

## REFERENCES

1. Matthews, B.W., Sigler, P.B., Henderso, R. and Blow, D.M. (1967) 3-dimensional structure of Tosyl-Alpha-Chymotrypsin. *Nature*, **214**, 652–656.
2. Blow, D.M. (1976) Structure and mechanism of Chymotrypsin. *Accounts Chem. Res.*, **9**, 145–152.
3. Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.*, **102**, 4501–4523.
4. Raines, R.T. (1998) Ribonuclease A. *Chem. Rev.*, **98**, 1045–1065.
5. Barnard, E.A. (1969) Ribonucleases. *Annu. Rev. Biochem.*, **38**, 677–732.
6. Wyckoff, H.W., Hardman, K.D., Allewell, N.M., Inagami, T., Johnson, L.N. and Richards, F.M. (1967) Structure of Ribonuclease-S at 3.5 Å resolution. *J. Biol. Chem.*, **242**, 3984–3984.
7. Perrin, D.M., Garestier, T. and Helene, C. (2001) Bridging the gap between proteins and nucleic acids: A metal-independent RNaseA mimic with two protein-like functionalities. *J. Am. Chem. Soc.*, **123**, 1556–1563.
8. Razkin, J., Nilsson, H. and Baltzer, L. (2007) Catalysis of the cleavage of uridine 3'-2,2,2-trichloroethylphosphate by a designed Helix-Loop-Helix motif peptide. *J. Am. Chem. Soc.*, **129**, 14752–14758.
9. Razkin, J., Lindgren, J., Nilsson, H. and Baltzer, L. (2008) Enhanced complexity and catalytic efficiency in the hydrolysis of phosphate diesters by rationally designed helix-loop-helix motifs. *Chembiochem*, **9**, 1975–1984.
10. Thomas, J.M., Yoon, J.K. and Perrin, D.M. (2009) Investigation of the catalytic mechanism of a synthetic DNAzyme with protein-like functionality: an RNaseA mimic? *J. Am. Chem. Soc.*, **131**, 5648–5658.
11. Roelfes, G., Boersma, A.J. and Feringa, B.L. (2006) Highly enantioselective DNA-based catalysis. *Chem. Commun.*, 635–637.
12. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
13. Klug, S.J. and Famulok, M. (1994) All you wanted to know about SELEX. *Mol. Biol. Rep.*, **20**, 97–107.
14. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment - RNA ligand to bacteriophage-T4 DNA-polymerase. *Science*, **249**, 505–510.
15. Joyce, G.F. (2004) Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, **73**, 791–836.
16. Famulok, M. (1994) Molecular recognition of amino-acids by RNA-aptamers - an L-Citrulline binding RNA motif and its evolution into an L-Arginine binder. *J. Am. Chem. Soc.*, **116**, 1698–1706.
17. Nieuwlandt, D., Wecker, M. and Gold, L. (1995) In-vitro selection of RNA ligands to Substance-P. *Biochemistry*, **34**, 5651–5659.
18. Liew, F.F., Hayashi, H., Nakano, S., Nakata, E. and Morii, T. (2011) A ribonucleopeptide module for effective conversion of an RNA aptamer to a fluorescent sensor. *Bioorg. Med. Chem.*, **19**, 5771–5775.
19. Kasahara, Y. and Kuwahara, M. (2012) Artificial specific binders directly recovered from chemically modified nucleic acid libraries. *J. Nucleic Acids*, **2012**, 156482.
20. Sidorov, A.V., Grasby, J.A. and Williams, D.M. (2004) Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. *Nucleic Acids Res.*, **32**, 1591–1601.
21. Santoro, S.W., Joyce, G.F., Sakthivel, K., Gramatikova, S. and Barbas, C.F. (2000) RNA cleavage by a DNA enzyme with extended chemical functionality. *J. Am. Chem. Soc.*, **122**, 2433–2439.
22. Kuwahara, M., Nagashima, J., Hasegawa, M., Tamura, T., Kitagata, R., Hanawa, K., Hososhima, S., Kasamatsu, T., Ozaki, H. and Sawai, H. (2006) Systematic characterization of 2'-deoxynucleoside-5'-triphosphate analogs as substrates for DNA polymerases by polymerase chain reaction and kinetic studies on enzymatic production of modified DNA. *Nucleic Acids Res.*, **34**, 5383–5394.
23. Wiegand, T.W., Janssen, R.C. and Eaton, B.E. (1997) Selection of RNA amide syntheses. *Chem. Biol.*, **4**, 675–683.
24. Zinnen, S.P., Domenico, K., Wilson, M., Dickinson, B.A., Beaudry, A., Mokler, V., Daniher, A.T., Burgin, A. and Beigelman, L. (2002) Selection, design, and characterization of a new potentially therapeutic ribozyme. *RNA-Publ. RNA Soc.*, **8**, 214–228.
25. Verma, S. and Eckstein, F. (1998) Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.*, **67**, 99–134.
26. Lercher, L., McGouran, J.F., Kessler, B.M., Schofield, C.J. and Davis, B.G. (2013) DNA modification under mild conditions by Suzuki-Miyaura cross-coupling for the generation of functional probes. *Angewandte Chemie-Int. Edition*, **52**, 10553–10558.
27. Holmes, S.C. and Gait, M.J. (2005) Syntheses and oligonucleotide incorporation of nucleoside analogues containing pendant imidazolyl or amino functionalities - the search for sequence-specific artificial ribonucleases. *Eur. J. Org. Chem.*, **2005**, 5171–5183.
28. Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859–1862.

29. Phan, A.T. and Patel, D.J. (2002) Differentiation between unlabeled and very-low-level fully N-15, C-13-labeled nucleotides for resonance assignments in nucleic acids. *J. Biomol. NMR*, **23**, 257–262.
30. Phan, A.T. and Patel, D.J. (2002) A site-specific low-enrichment N-15, C-13 isotope-labeling approach to unambiguous NMR spectral assignments in nucleic acids. *J. Am. Chem. Soc.*, **124**, 1160–1161.
31. Andrus, A. and Kuimelis, R.G. (2001). *Current Protocols in Nucleic Acid Chemistry*. John Wiley & Sons, Inc., Hoboken, New Jersey.
32. Stott, K., Stonehouse, J., Keeler, J., Hwang, T.L. and Shaka, A.J. (1995) Excitation sculpting in high-resolution nuclear-magnetic-resonance spectroscopy - application to selective NOE experiments. *J. Am. Chem. Soc.*, **117**, 4199–4200.
33. Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
34. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, P., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
35. Markley, J.L. and Porubcan, M.A. (1976) Charge-relay system of serine proteinases - proton magnetic-resonance titration studies of 4 Histidines of Porcine trypsin. *J. Mol. Biol.*, **102**, 487–509.
36. Roberts, G.C.K., ed. (1993) *NMR of Macromolecules: A Practical Approach*. Oxford University Press, Oxford.
37. Alper, J.S. and Gelb, R.I. (1990) Standard errors and confidence intervals in nonlinear-regression - comparison of Monte-Carlo and parametric statistics. *J. Phys. Chem.*, **94**, 4747–4751.
38. Case, D.A., Cheatham, T.A.D. T.E. III, Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B. et al. (2012). University of California, San Francisco.
39. Gotz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S. and Walker, R.C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.*, **8**, 1542–1555.
40. Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S. and Walker, R.C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle Mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.
41. Darden, T., York, D. and Pedersen, L. (1993) Particle Mesh Ewald - an n.log(n) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
42. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical-integrations of Cartesian equations of motion of a system with constraints - molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.
43. Uberuaga, B.P., Anghel, M. and Voter, A.F. (2004) Synchronization of trajectories in canonical molecular-dynamics simulations: observation, explanation, and exploitation. *J. Chem. Phys.*, **120**, 6363–6374.
44. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
45. Lu, X.J. and Olson, W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
46. Durrant, J.D. and McCammon, J.A. (2011) HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J. Mol. Graph.*, **31**, 5–9.
47. Varnai, P., Djuranovic, D., Lavery, R. and Hartmann, B. (2002) Alpha/gamma transitions in the B-DNA backbone. *Nucleic Acids Res.*, **30**, 5398–5406.
48. Hartmann, B., Piazzola, D. and Lavery, R. (1993) B-I-B-II transitions in B-DNA. *Nucleic Acids Res.*, **21**, 561–568.
49. Djuranovic, D. and Hartmann, B. (2004) DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, **73**, 356–368.
50. Catry, M.A., Gheerardijn, V. and Madder, A. (2010) Surprising duplex stabilisation upon mismatch introduction within triply modified duplexes. *Bioorgan. Chem.*, **38**, 92–97.
51. Aboulela, F., Koh, D., Tinoco, I. and Martin, F.H. (1985) Base-base mismatches - thermodynamics of double helix formation for DCA3XA3G + DCT3YT3G (X, Y = A, C, G, T). *Nucleic Acids Res.*, **13**, 4811–4824.
52. Bhattacharya, P.K., Cha, J. and Barton, J.K. (2002) H-1 NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches. *Nucleic Acids Res.*, **30**, 4740–4750.
53. Broo, K., Brive, L., Lundh, A.C., Ahlberg, P. and Baltzer, L. (1996) The mechanism of self-catalyzed site-selective functionalization of a designed helix-loop-helix motif. *J. Am. Chem. Soc.*, **118**, 8172–8173.
54. Bruice, T.C. and Schmir, G.L. (1958) Imidazole catalysis. 2. The reaction of substituted imidazoles with phenyl acetates in aqueous solution. *J. Am. Chem. Soc.*, **80**, 148–156.
55. Wütrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley Interscience, New York.
56. Boelens, R., Scheek, R.M., Dijkstra, K. and Kaptein, R. (1985) Sequential assignment of imino-proton and amino-proton resonances in H-1-NMR spectra of oligonucleotides by two-dimensional NMR-spectroscopy - application to a Lac operator fragment. *J. Magn. Reson.*, **62**, 378–386.
57. Wijmenga, S.S. and van Buuren, B.N.M. (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.*, **32**, 287–387.
58. Klepeis, J.L., Lindorff-Larsen, K., Dror, R.O. and Shaw, D.E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.*, **19**, 120–127.
59. Pauwels, E., Claeys, D., Martins, J.C., Waroquier, M., Bifulco, G., Van Speybroeck, V. and Madder, A. (2013) Accurate prediction of H-1 chemical shifts in interstrand cross-linked DNA. *RSC Adv.*, **3**, 3925–3938.
60. Mura, C. and McCammon, J.A. (2008) Molecular dynamics of a kappa B DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res.*, **36**, 4941–4955.