Synthesis, Physicochemical Properties, and Hydrogen Bonding of 4(5)-Substituted 1-*H*-Imidazole-2-carboxamide, a Potential Universal Reader for DNA Sequencing by Recognition Tunneling

Feng Liang, Shengqing Li, Stuart Lindsay,* and Peiming Zhang*[a]

Abstract: We have developed a chemical reagent that recognizes all naturally occurring DNA bases, a so called universal reader, for DNA sequencing by recognition tunneling in nanopores.[1] The primary requirements for this type of molecules are the ability to form non-covalent complexes with individual DNA bases and to generate recognizable electronic signatures under an electrical bias. 1-*H*-imidazole-2-carboxamide was designed as such a recognition moiety to interact with the DNA bases through hydrogen bonding. In the present study, we first furnished a synthetic route to 1-*H*-imidazole-2-carboxamide containing a short ω-functionalized alkyl chain at its 4(5) position for its attachment to metal and carbon electrodes. The acid dissociation constants of the imidazole-2-carboxamide were then determined by UV spectroscopy. The data show that the 1-*H*-imidazole-2-carboxamide exists in a neutral form between pH 6–10. Density functional theory (DFT) and NMR studies indicate that the imidazole ring exists in prototropic tautomers. We propose an intramolecular mechanism for tautomerization of 1-*H*-imidazole-2-carboxamide. In addition, the imidazole-2-carboxamide can self-associate to form hydrogen bonded dimers. NMR titration found that naturally occurring nucleosides interacted with 1-*H*-imidazole-2-carboxamide through hydrogen bonding in a tendency of $dG > dC \gg dT > dA$. These studies are indispensable to assisting us in understanding the molecular recognition that takes place in the nanopore where routinely used analytical tools such as NMR and FTIR cannot be conveniently applied.

Keywords: DNA · hydrogen bonds · recognition tunneling · self-association · universal reader

Introduction

DNA sequencing is a pivotal tool in rapidly expanding personal genomics where individual genomes are genotyped and bioinformatically analyzed to find disease association of genes and loci. The sequencing of DNA and RNA will eventually form the basis of personalized medicine.[2] In spite of its monumental accomplishments culminating in the completion of Human Genome Project, the classical Sanger method[3] has limited use where a large number of genome sequences are involved due to its high cost and low throughput. Current advances in next generation sequencing (NGS) have brought the whole genome cost down to thousands of dollars and the sequencing timescale down to a number of days. There are several different types of NGS DNA sequencers available on the market,[4] all of which are based on enzymatic reactions and are dominated by optical detection. DNA segments (or "reads") sequenced by these NGS technologies vary in length from 35 to < 1000 base pairs. Short reads hamper de novo assembly of complex genomes from scratch, suffering from duplications and repeats, as well as the fragmentation of genomic sequences.[5] As a result, data from NGS are often assembled using published human genomes as references, an approach that may create distorted personal genomes.[6] To overcome these caveats of NGS, nanopores with diameters comparable to that of a DNA double helix (ca. 2 nanometers) have been investigated as a physical means for the sequencing.[7] A nanopore should be able to read very long segments, meanwhile simplifying sample preparation and eliminating the use of costly biochemical reagents, such as polymerases or ligases. In principle, a single stranded DNA, while passing through a nanopore, blockades ionic current and generates electronic signals corresponding to its base compositions and the order of bases in the sequence. It has been demonstrated that the ion current blockades can used to distinguish all four nucleotides and 5-methylcytidine monophosphate (5-methyl C) if they are individually passed into the pore,[8] and single bases in a DNA oligomer can be read if trapped in an MspA pore at a double strand-single strand junction.[9] Nevertheless, nanopores to date have not shown the capability of reading a DNA sequence. One of the reasons is that the current blockade is the sum of effects from all the DNA bases inside the nanopore,[10] complicating a single base resolution.

[a] Dr. F. Liang, Dr. S. Li, Prof. S. Lindsay, Prof. P. Zhang
Center for Single Molecule Biophysics
Biodesign Institute, Arizona State University
Tempe, Arizona 85287 (USA)
Fax: (+1)480-727-2378
E-mail: Stuart.Lindsay@asu.edu
 peiming.zhang@asu.edu

Measuring the electron tunneling current that is created transversely across a translocating DNA molecule has been proposed as an alternative to the measurement of ion current blockades.[7a, 11] We have recently demonstrated by means of a Scanning Tunneling Microscope (STM) setup that a DNA base flanked by other bases in a short DNA oligomer can be identified in a nanogap consisting of two gold electrodes functionalized with mercaptobenzamide in aqueous solution.[12] The benzamide specifically functions as a recognition moiety, which presumably interacts with DNA bases through hydrogen bonding. We found that the benzamide can read all DNA bases, A, C, ^{m}C, G, except T. In addition, there are significant overlaps of electric signal peaks among A, ^{m}C, G, which would compromise the base-calling accuracy. Although the selectivity of benzamide can be improved by optimizing the experimental conditions and including other characteristics of the tunneling noise, a highly discriminating universal reader molecule will ultimately fulfill a better resolution and simplified data analysis.

In one configuration of the nanopore sequencing devices under development, atomically thin electrodes are embedded in the dielectric film with their ends exposed to the nanopore.[13] We envision that as a single stranded DNA molecule translocates through a nanopore, the inside water molecules are squeezed out of the nanopore, rendering the interior of the nanopore less hydrophilic in comparison to its outside. Thus, unlike in the aqueous solution, the hydrogen bonding interactions should be conceivable inside the nanopore. Based on this hypothesis, we have drafted a set of general guidelines for design of a univesal reader. The molecule should contain multiple hydrogen bonding sites with well-balanced numbers of donors and acceptors, have high conductance, and be small enough in size to sit in a nanopore without impeding the DNA translocation. In addition, it should have variable conformations in order for one to fit all. The universal reader must also be electrochemically stable under physiological conditions. Our first choice was five-membered aromatic nitrogen heterocycles because of their π-excessive (6 π electrons shared by 5 atoms) nature[14] and built-in hydrogen bonding sites. Bergstrom and co-workers have synthesized 4(5)-(2′-deoxy-β-D-ribofuranosyl)-1-*H*-imidazole-2-carboxamide as a degenerate DNA base

analog.[15] It was unclear how this molecule interacts with DNA bases since its base pairing properties have not been examined. At first glance, 1-*H*-imidazole-2-carboxamide is an aromatic molecule that contains several hydrogen bonding sites. However, when the imidazole-2-carboxamide moiety is attached to an electrode through a flexible linker, it can freely rotate and its amide group can rotate as well, as illustrated in Figure 1. In a nanopore with a diameter of
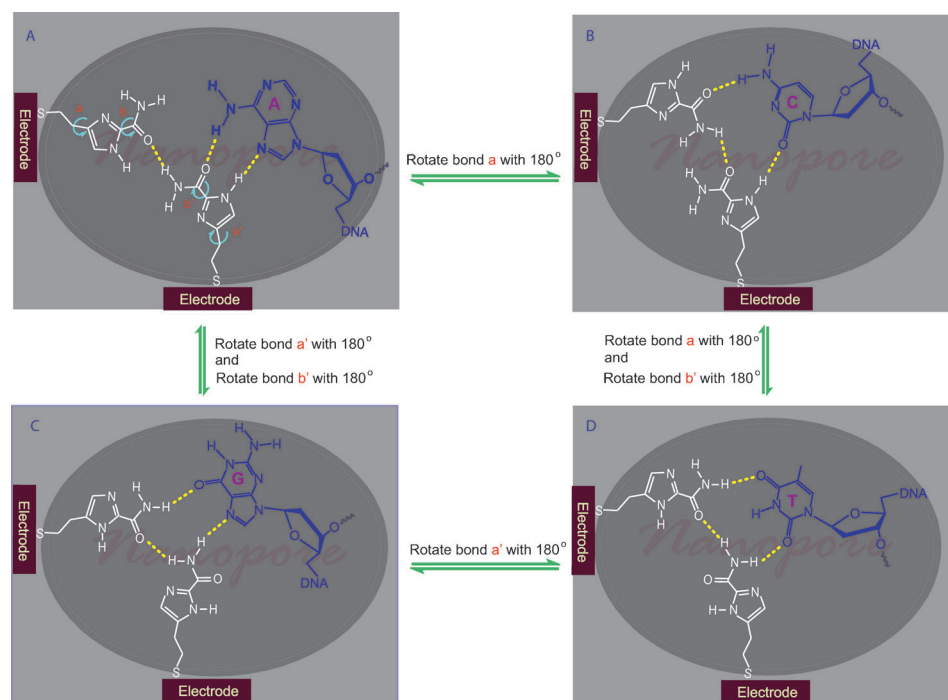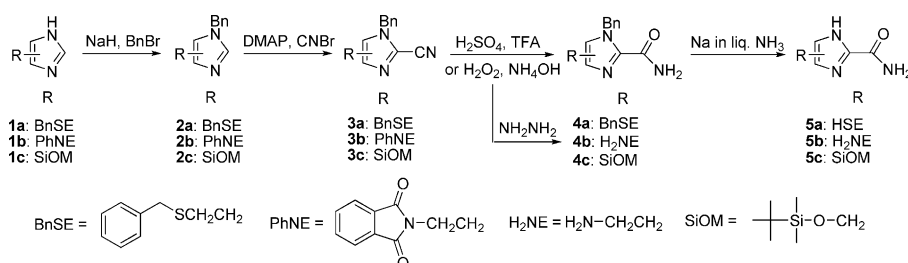

Figure 1. An illustration of 1-*H*-imidazole-2-carboxamide as a universal reader, which reads each of DNA bases by changing its conformations to form a different hydrogen bonding pattern (in an actual device fabricated by E-beam lithography and atomic layer deposition (ALD) processes, the atomically thin electrodes are embedded in a dielectric film with their ends exposed to the nanopore).

2 nm, an imidazole-2-carboxamide molecule in one electrode can reach the same molecule in another electrode to form a molecular bridge through hydrogen bonding for the electron tunneling. Due to the changeable conformations of the imidazole-2-carboxamide molecule, the bridge can exist in multifaceted configurations, each of which should selectively recognize one of DNA bases and form an electrically distinguishable triplet structure (Figure 1 A, B, C, D). With two free rotating bonds, two imidazole-2-carboxamides can statistically form as many as 16 different facets. Thus, it should be possible for such a single molecule to read all the DNA bases and their modifications, such as 5-methylcytosine and N^6-methyladenine which have different π-electron distributions from their parent bases. In the present work, we have developed an effective way to synthesize 4(5)-substituted-1-*H*-imidazole-2-carboamides and studied physicochemical and hydrogen bonding properties of the imidazole-2-carboxamide moiety in solution. The goal was to chemically understand the imidazole molecule, laying down a basis

for studying its interactions with DNA bases and electron tunneling in the nanopore.

Results and Discussion

Synthesis of 1-*H*-imidazole-2-carboxamide containing a short linker at its 4(5) position: We have synthesized a number of 1-*H*-imidazole-2-carboxamides composed of a short ω-functionalized alkyl chain at their 4(5) positions for attachment to metal and carbon electrodes. In general, the 1-*H*-imidazole derivatives exist in tautomeric forms that are chromatographically inseparable. No attempt was made to separate the tautomers during the synthesis. Since a variety of 4(5)-alkylated imidazoles were commercially available or reported in literature, the main efforts of the synthesis was finding a facile method to introduce an amide group to the 2 position of the imidazole ring. As illustrated in Scheme 1, 4(5)-(2-thioethyl)-1-*H*-imidazole-2-carboxamide

reaction. The cyano group was introduced into the 2 position of the imidazole ring of **2a**, **2b**, and **2c** by treating them with 1-cyano-4-(dimethylamino)pyridinium bromide (CAP). CAP was generated in situ by mixing equivalent amounts of cyanogen bromide and 4-(dimethylamino)pyridine in dimethyl formamide (DMF) at 0 °C. A 2.5 fold of CAP gave the best yield. The cyano group of **3a** was converted into an amide (**4a**) by hydrolysis in a mixture of sulfuric acid (20 % by volume) and trifluoroacetic acid (18 % by volume) in a fair yield (46 %).[18] Under the basic condition and in the presence of hydrogen peroxide,[19] the hydrolysis could give us **4b** from **3b** (the phthaloyl group was completely removed by hydrazine) and **4c** from **3c**, but failed to furnish the desired product **4a** from **3a** because the benzylthio group was oxidized and eliminated from **3a**. Final products **5a**, **5b**, and **5c** were obtained by removing the benzyl groups with sodium in liquid ammonia. The *tert*-butyldimethylsilyloxymethyl group survived from the debenzylation conditions. **5c** was separated in a fair yield (35 %).
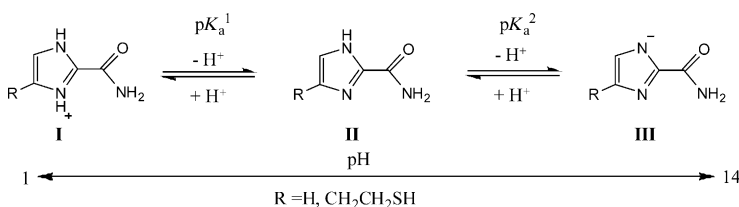
Scheme 1. Synthetic route to 4(5)-substituted-1-*H*-imidazole-2-carboxamides. BnBr = benzyl bromide; DMAP = 4-(dimethylamino)pyridine; TFA = trifluoroacetic acid.

Acid dissociation constants (pK_a) of 1-*H*-imidazole-2-carboxamides: Imidazole is amphoteric, which can either gain a proton via its imine nitrogen N3 (as a base) to form an imidazolium, or lose its NH proton (as an acid) to form an imidazolide in aqueous solution, as shown in Scheme 2.[20] The introduction of an amide group

(**5a**) and 4(5)-(2-aminoethyl)-1-*H*-imidazole-2-carboxamide (**5b**) were synthesized from 4(5)-(2-(benzylthio)ethyl)-1-*H*-imidazole (**1a**) and *N*-phthaloyl histamine (**1b**), respectively. The thiol and amine function as anchor groups for attaching the molecule to either metal or carbon electrodes. In the same way, 4(5)-(*tert*-butyldimethylsilyloxymethyl)-1-*H*-imidazole-2-carboxamide (**5c**) was synthesized from 4(5)-(*tert*-butyldimethylsilyloxymethyl)-1-*H*-imidazole (**1c**), and used as a model compound for NMR studies on hydrogen bonding properties of 1-*H*-imidazole-2-carboxamide in organic solvents (vide infra). We have explored two routes for synthesizing these 1-*H*-imidazole-2-carboxamide compounds mentioned above. The 2 position of imidazole can be substituted with a formate ester[16] or a cyano group,[17] both of which can readily be converted into an amide group. We found that the cyano route gave us a better yield overall. First, compound **1a**, **1b**, and **1c** were converted into the 1-*H* nitrogen protected products (**2a**, **2b**, **2c**) by reacting their sodium salts (generated in situ by treating them with NaH) with benzyl bromide. NMR confirms that each of them is a mixture of two isomers with the benzyl group at N^1 and N^3. We have tried different protecting groups such as trityl and *tert*-butyloxycarbonyl (Boc), finding that benzyl was the most effective protecting group for the following cyanating

Scheme 2. Protonation and deprotonation equilibria of 1-*H*-imidazole-2-carboxamide in aqueous solution.

to the imidazole should alter both its acidity and basicity. We determined acid dissociation constants of 1-*H*-imidazole-2-carboxamide in buffered solution by means of UV spectroscopy (Figure S1 in the Supporting Information), giving pK_a^1 = 3.29 and pK_a^2 between 12 and 13. As expected, they are smaller than those of the parent imidazole molecule (pK_a^1 = 7.0, pK_a^2 = 14). To examine effects of 4(5) substitution on acidity and basicity of the imidazole, we first determined pK_as of **5a** in aqueous solutions (not buffered). The UV spectra of **5a** was recorded at different pH values ranging from 1 to 13 (Figure 2a). The measured pK_as in this pH range should directly be related to protons on the nitrogen atoms of the imidazole ring. In general, an amide proton has its pK_a value below 0 as a base and above 15 as
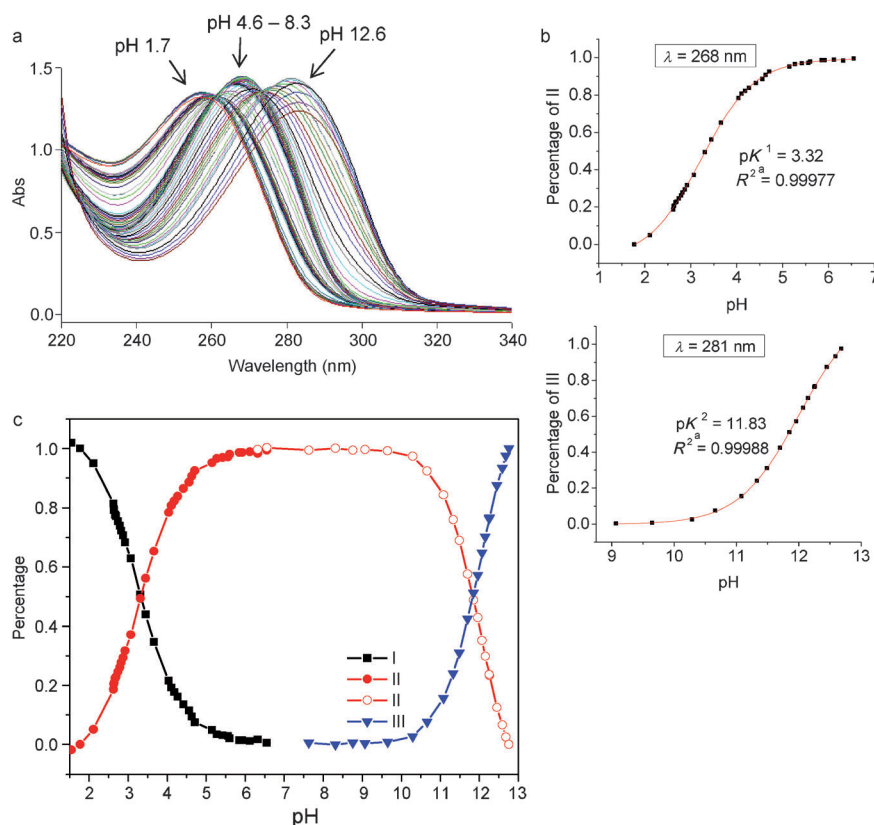
Figure 2. a) UV spectra of **5a** at different pH values ranging from 1.7 to 12.6. b) Degree of dissociation/pH profile of **5a** at 268 and 281 nm. c) Distribution of the three forms of the imidazole with pH, calculated based on UV absorbance at 281 nm, 268 nm, and 258 nm from Figure 2a, respectively (see the Experimental Section for details).

tions (pH ca. 7). It should be noted that the thiol group in **5a** does not interfere the pK_a measurement significantly. Thiol (RSH) is a weak acid with pK_a about 10 and forms thiolate (RS⁻) under basic conditions. Although the thiolate has strong UV absorbance at 238 nm, the UV absorbance of thiol above 250 nm and thiolate above 275 nm is negligible.[22]

Structure, tautomerism, and self-association of 1-*H*-imidazole-2-carboxamide: Owing to its suitable solubility in organic solvents, **5c** was employed as a model compound to study the hydrogen bonding interactions of imidazole-2-carboxamide. We first studied its molecular structure by DFT. Since the amide group is connected to the imidazole through a σ bond, it can rotate freely and exist in a conformation with its NH₂ either *cis* or *trans* to the NH of the imidazole ring (Figure 3). The DFT calculation indicates
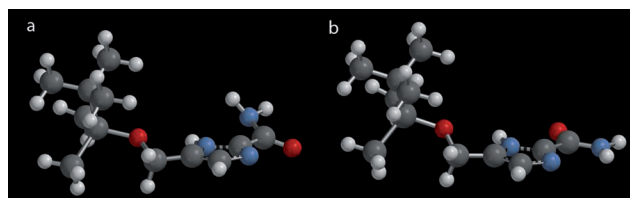
an acid.[21] Our NMR data show that the NH proton ($\delta = 10.1$ ppm) of the imidazole-2-carboxamide appears at the downfield of its amide protons ($\delta = 6.2$ ppm) in chloroform (see below). That indicates that the NH proton is more acidic and should be deprotonated ahead of the amide in the process of the titration with base. From Figure 2a, we observe bathochromic shifts of the λ_{max} from 258 to 268 to 281 nm with increasing pH and two isosbestic points respectively at 261 and 274 nm, indicating that there are two UV absorbents existing in each of the equilibriums. The degree of dissociation versus pH curves were well fit to a sigmoidal function ($R^2 > 0.99$), which yielded $pK_a^1 = 3.32$ and $pK_a^2 = 11.83$ for **5a** (Figure 2b). We also determined the pK_as of **5a** in buffered aqueous solutions with the result that $pK_a^1 = 3.33$ and pK_a^2 between 12 and 13 (Figure S2 in the Supporting Information). These studies show that the 4(5) substitution has no noticeable effects on pK_a values of 1-*H*-imidazole-2-carboxamide. However, the pK_a^2 value of **5a** in the non-buffered solution is smaller than in the buffered solution. This indicates that ions may impede release of the NH proton from the imidazole ring. Figure 2c shows the distribution of the three forms of the 1-*H*-imidazole-2-carboxamide versus pH. Between pH 6 to 10, the imidazole-2-carboxamide exists solely in a neutral form, which is important for tunneling experiments carried out in physiological condi-

Figure 3. DFT model of **5c** with a) N−H of the imidazole *cis* with respect to the NH₂ group of the carboxamide. b) N−H of the imidazole *trans* with respect to the NH₂ group of the carboxamide.

that the amide would primarily take the *trans* conformation (Figure 3b). As shown in Figure 3a, in the *cis* conformation, the NH₂ group is pushed out of the imidazole plane by the NH group, losing conjugation with the imidazole ring. There is a difference of approximately 10.5 kcal mol⁻¹e in total energy between the *cis* and *trans* conformations (Table S1 in the Supporting Information).

Another important feature of the 1-*H*-imidazole-2-carboxamide molecule is its tautomeric proton (NH) that sways between two nitrogen atoms of the imidazole ring. We studied the molecular structures of two prototropic tautomers of **5c** with the amide situated at its *trans* conformation in detail by DFT calculations using B3LYP in combination with 6-311 + +G** basis sets in vacuum, followed by B3LYP/6-31G* to
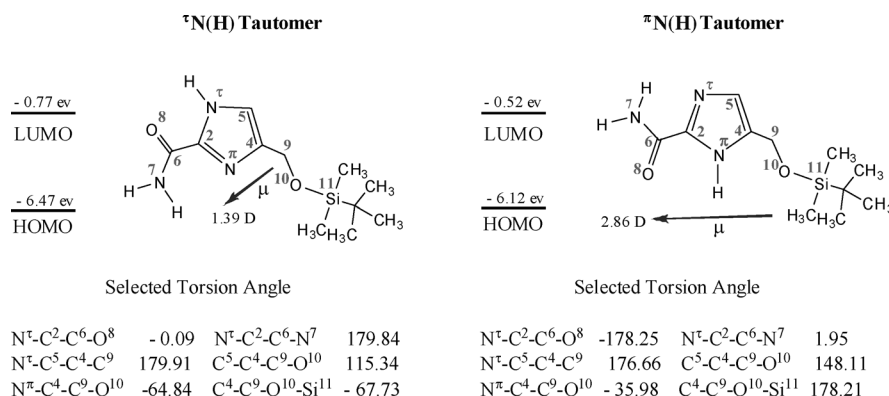
Figure 4. Molecular orbital energies (eV), dipole moments (μ, D), and torsion angles (deg) for the two tautomers of **5c** in chloroform from the DFT calculation. The two tautomers of **5c** are designated according to nomenclature of amino acids and peptides recommended by IUPAC and IUBMB (http://www.chem.qmul.ac.uk/iupac/index.html) for histidine.

calculate their NMR chemical shifts in chloroform. Some of characteristic structure parameters are summarized in Figure 4. In the DFT model, the SiOM group stays on the top of the imidazole amide plane, projected in the direction opposite to the amide and away from the imidazole ring (see Figure 3) so that the SiOM group would not interfere in hydrogen bonding interactions related to the amide group. Furthermore, the NH proton does not form intramolecular hydrogen bonds with either the SiOM or the amide group, referring to the geometric definition for hydrogen bonds: $d \leq 2.35$ Å and $100° \leq \alpha \leq 180°$ ($d =$ length of the hydrogen bond, $\alpha =$ angle of the hydrogen bond).[23] However, the SiOM group has some weak interactions most likely through its oxygen with the NH in the $^{\pi}$N(H) tautomer. In the DFT structure of $^{\pi}$N(H) tautomer, the torsion angle N^{π}-C^4-C^9-O^{10} is $-35.98°$, much smaller than the same one ($-64.84°$) in the $^{\tau}$N(H) tautomer. This places the oxygen near the NH with the O to H distance about 2.6 Å in the $^{\pi}$N(H) tautomer, which deshields the proton and moving its chemical shift downfield a bit in comparison to the shift in the NH in the $^{\tau}$N(H) tautomer in NMR spectroscopy (see Scheme 3). The DFT calculation shows that the $^{\pi}$N(H) tautomer is slightly more stable than the $^{\tau}$N(H) tautomer by -1.08 kcal mol^{-1} in the total energy (Table S1 in the Supporting Information), which is so small that we conclude that both tautomers should concomitantly exist in solution without a marked preference. In a 2D NOSEY proton NMR spectrum of **5c** (15 mM in

chloroform), we observe the NOE cross peaks of the NH correlating with C^5H, C^9H, and CH$_3$ protons, and no cross peaks from correlation of the NH with the amide NH$_2$ (Figure S3 in the Supporting Information). These results can be interpreted by the presence of the tautomers and the amide in its *trans* conformation. **5c** was also studied by variable temperature (VT) NMR. In a low concentration (2 mM), both NH and amide protons appear as single peaks from 273 to 309 K (Figure 5a), indicating that the tautomeric proton are in the fast exchange and the amide bond in the fast rotation on the NMR time scale. Meanwhile, there are no significant upfield chemical shifts caused by elevated temperatures. In a high concentration (15 mM), both NH and amide protons appears as broad peaks below $-30°$C (243 K). With increasing temperatures, the NH proton peak disappears while the two amide proton peaks become much sharper and shift systematically upfield (Figure 5b). The VT NMR studies reveal a dynamic slowdown of the tautomerization and rotation with the increased concentration.

To understand these NMR data, we propose a scheme to depict the chemical equilibriums of 1-*H*-imidazole-2-carboxamide taking place in chloroform (Scheme 3). The 1-*H*-imidazole-2-carboxamide molecule can spontaneously undergo both tautomerization within the imidazole ring and dimerizations (or self-associations) between the amides through hydrogen bonding interactions. It has been reported that the
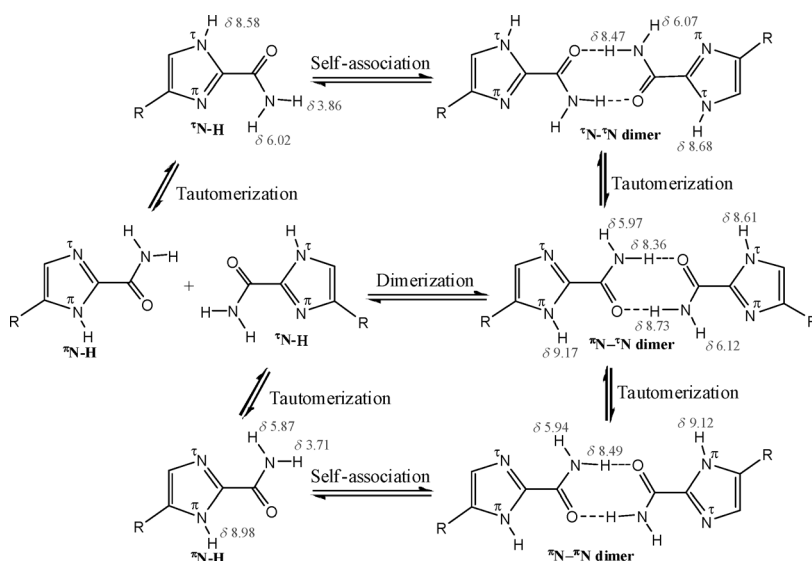
Scheme 3. Chemical equilibria of 4(5)-SiOM-1-*H*-imidazole-2-carboxamide in chloroform (numbers: calculated chemical shifts (ppm) of protons in **5c** dimers).
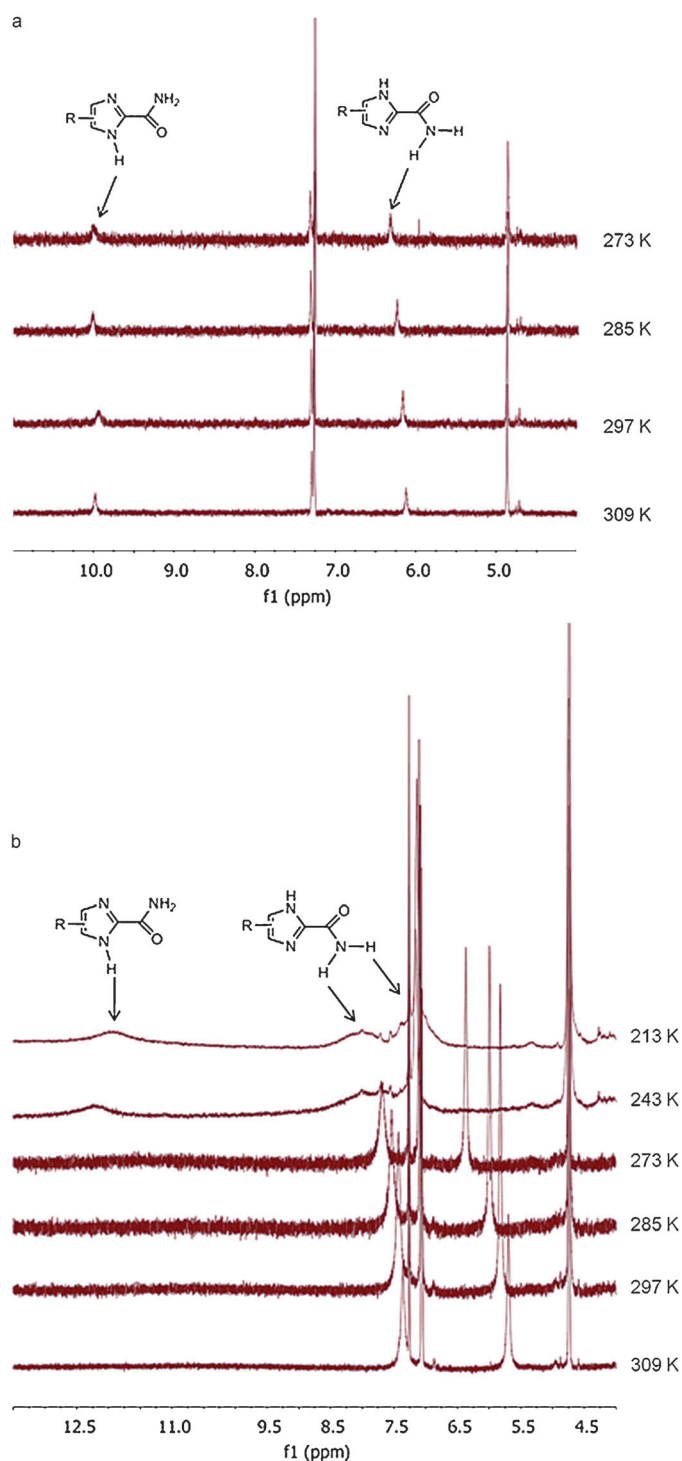
Figure 5. VT NMR spectra of **5c** in CDCl₃ at a concentration of a) 2 mm and b) 15 mm.
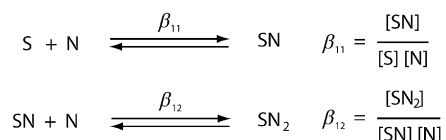
tautomerization rate decreased with an increase in the size of an N-substituent on carboxamide at the 2 position of imidazole, presuming that the formation of tautomers was through an intermolecular proton exchange mechanism.[24] Although the intermolecular proton transfer in nonprotic solvents is an energetically favorable process in the proto-

tropic tautomerization of imidazole, an intramolecular proton transfer is also feasible in imidazole,[25] supported by low-temperature IR experiments in inert gas matrices that indicate the existence of two imidazole tautomers in isolated adenine and guanine.[26] Since our NMR data indicate that the NH proton was not involved in hydrogen bonding, we believe that an intramolecular mechanism may play a major role in tautomerization of 1-*H*-imidazole-2-carboxamide. The amide group functions as a carrier for transportation of the proton from one nitrogen atom to another by rotating around the imidazole plane. To do so, the NH proton first migrates to the nearby oxygen of amide, and then the protonated amide turns 180° to other side of the imidazole ring and releases the proton to the nearby nitrogen. Thus, formation of a hydrogen bonded dimer slows down the rotation of the amide and reduces the exchange rate of the proton, resulting in the broadening (or disappearing into the baseline) of the NH proton signal in NMR spectra. The signal reappears when the temperature continues lowering to further slow down the proton exchange. That is a phenomenon commonly observed in dynamic NMR. As shown in Scheme 3, there are three different dimers that can be formed from two tautomers. Our DFT calculation shows that there are no significant differences in total energy and dimerization energy among these dimers (Table S1 in the Supporting Information), which implies that there is no preference to form these dimers. The calculated chemical shifts of the hydrogen bonded amide proton and the NH proton are close to each other (see Scheme 3) so that these dimers may not readily be distinguished by routine NMR experiments and the NH proton may coalesce with the hydrogen bonding amide proton. That requires more sophisticated NMR techniques, such as dynamic NMR at very low temperatures, to identify these individual dimers.[27] Through NMR dilution experiments, the self-association constant of **5c** was determined (Figure S4 in the Supporting Information). The curve fitting was carried out in the computer program HypNMR 2008. The first attempt to fit data from the hydrogen bonded amide proton into a dimerization (R_2) model was failed. However, the data from the non-hydrogen-bonded amide proton fit into the R_2 model convergently, giving the constant 2.0 (1) (in log β). To explain the non-convergence of the hydrogen bonded amide proton, we deduce that its chemical shifts in the lower concentrations were coalesced with those of the NH proton, based on observation that the amide peak were broadened out and shifted downfield with decreasing concentrations (Figure S4 in the Supporting Information).

As far as physical properties are concerned, two tautomers of **5c** have relatively small dipole moments (Figure 4), compared with DNA bases (G: 6.55 D > C: 6.39 D > T: 4.31 D > A: 2.56 D).[28] This may make it less effective in the base stacking. The HOMO energy of the 1-*H*-imidazole-2-carboxamide molecule (6.47 eV for ᵗN(H) tautomer and −6.11 eV for ᵖN(H) tautomer) is higher than those of DNA bases (G: −7.75 eV; A: −8.24 eV; C: −8.87 eV; T: −9.14 eV), closer to the Fermi level of gold (−5 to

−5.31 eV);[29] however, its HOMO–LUMO gap (Figure 4) is much larger than those of DNA bases (3–4 eV). Thus, 1-*H*-imidazole-2-carboxamide may be more feasible for charge transfer than the DNA bases, while less effective for electron transfer due to its higher LUMO energy (−0.77 for $^\text{T}$N(H) tautomer and −0.50 eV for $^\pi$N(H) tautomer).

Interactions of 1-*H*-imidazole-2-carboxamide with DNA bases: We have studied the hydrogen bonding interactions of **5c** with DNA bases by NMR titration in chloroform. Four naturally occurring nucleosides (referred as to dA, dC, dG, and dT) were used as the model compounds of the bases imbedded in a single stranded DNA, which were made soluble in chloroform by silylating their hydroxyl groups with *tert*-butyldimethylsilyl chloride (TBDMSCl),[30] and **5c** was only used as a substrate of the titration due to its limited solubility in chloroform, the concentration of which was held constant at 5 mM so that it existed mainly in a monomer form. It was found that the chemical shift of the NH proton changed irregularly with concentrations of the titrants and the analyzable data were obtained from monitoring changes in chemical shifts of the amide protons. The control experiments were titrating dT with dA and dG. The program HypNMR 2008 was employed to determine equilibrium constants from the titration data. First, stoichiometries of the interactions between **5c** and nucleosides were estimated by plotting the NMR titrating data as nucleoside to **5c** mole ratios versus chemical shifts of the amide proton (Figure S5 in the Supporting Information). From the mole ratio plots, we deduce that **5c** may form 1:1 complexes with the nucleoside dA and dT, and 1:2 complexes with nucleoside dC and dG. Thus, the associations between **5c** (S) and the nucleosides (N) can be expressed in multistep equilibriums as shown below:

$$S + N \underset{}{\overset{\beta_{11}}{\rightleftharpoons}} SN \qquad \beta_{11} = \frac{[SN]}{[S][N]}$$

$$SN + N \underset{}{\overset{\beta_{12}}{\rightleftharpoons}} SN_2 \qquad \beta_{12} = \frac{[SN_2]}{[SN][N]}$$

Based on this thermodynamic process, the titration data were then fit into the 1:1, 1:2, or 1:1/1:2 binding isotherm by nonlinear regression in the HypNMR program. It was found that the data from titrating with dA and dT were fit into the 1:1 model and those from titrating with dC and dG were fit into the 1:1/1:2 model convergently (Figure S6 in the Supporting Information). The association constants derived from the curving fitting are listed in Table 1. Following the same procedure just mentioned above, we determined the association constant of the Watson–Crick base pair dA–dT, which is close to the value reported in literature,[31] and the association constant of a mismatched base pair dG–dT, which is consistent with one derived from curving fitting in the program Origin 8 ($\text{Log} \beta_{11} = 2.52$). These indicate that our data analysis process is reasonably reliable. From the association constants, we see that the nucleosides form the 1:1 complexes with **5c** in a tendency of dG > dC ≫ dT > dA in

Table 1. Association constants of **5c** with nucleosides determined from curve fitting of NMR titration data.

	Log β_{11}	Log β_{12}
5c-dA	1.08(9)	–
5c-dC	3.1(1)	5.04(9)
5c-dG	3.31(2)	5.49(1)
5c-dT	1.46(5)	–
dA-dT	1.62(1)	–
dG-dT	2.56(6)	–

chloroform. Furthermore, dG and dC can also form 2:1 complexes with **5c**, which are more stable than the 1:1 complexes. The data reported here demonstrate that **5c** can interact with the nucleosides through their nucleobases in the organic phase although the interactions do not, in themselves, represent strong recognition. We predict that the 1-*H*-imidazole-2-carboxamide molecules tethered to two electrodes should form a more stable 2:1 complex with each of the DNA bases in a nanogap than its 1:1 counterpart in the solution. In our previous study, we found that bond lifetimes with DNA bases were greatly enhanced when the benzamide molecules were tethered in a nanogap for the recognition.[12] We recently observed the same phenomenon in a simpler case of a nucleobase recognizing its complement in a nanogap.[32] Since 1-*H*-imidazole-2-carboxamide interacts with each of DNA bases differently, the DNA bases should be electronically discriminable in the nanogap functionalized with the reading molecules.

Conclusion

To complement our recognition tunneling technique, we have designed 1-*H*-imidazole-2-carboxamide as a potential universal reader to read all of the possible individual DNA bases in a single stranded DNA. A synthetic route has been developed for introducing a linkage to the 4(5)-position of 1-*H*-imidazole-2-carboxamide. Our data show that 1-*H*-imidazole-2-carboxamide is suitable to work under the physiological conditions. In the present study, we demonstrate by NMR titration that the 1-*H*-imidazole-2-carboxamide molecule can form hydrogen bonded complexes with all four naturally occurring DNA bases in the dielectric and aprotic environments. The recognition of DNA bases by the imidazole molecules immobilized in nanogaps is the subject of a manuscript in preparation.

Experimental Section

General information: All reagents were obtained from commercial suppliers unless otherwise stated. Where necessary, organic solvents were routinely dried and/or distilled prior to use and stored over molecular sieves under nitrogen. All reactions requiring anhydrous conditions were performed under a nitrogen atmosphere. Reactions were monitored by TLC on silica gel plates. Flash chromatography was performed in an automated flash chromatography system (CombiFlash *R*$_f$, Teledyne Isco, Inc.) with silica gel columns (60–120 mesh). Proton NMR (^{1}H) spectra

were recorded on a Varian 400 MHz spectrometer or a Varian 500 MHz spectrometer. ^{1}H chemical shifts in chloroform were referenced relative to deuterated solvents (CDCl$_3$: $\delta_H = 7.26$ ppm, CD$_3$OD: $\delta_H = 3.31$ ppm, and DMSO: $\delta_H = 2.50$ ppm). 2D NOE were recorded at 400 MHz on a Bruker 400 MHz spectrometer. In VT NMR, temperature was calibrated with a standard solution of 100% CH$_3$OH and regulated to an accuracy of ± 0.1 °C by a Eurotherm Variable Temperature Unit on the Bruker NMR or a Highland Technologies Temperature unit on the Varian NMR System. Below zero Celsius temperatures were achieved with a Liquid Nitrogen Heat Exchanger on the Bruker and FTS Cooling System (Stone Ridge, New York) on the Varian. HRMS spectra were recorded using the atmospheric pressure chemical ionization (APCI) technique. Mettler Toledo Portable Meter was used for pH determination. UV spectra were recorded on a Varian Cary 300 UV spectrophotometer. Mathematical analysis of NMR data and graphic presentation of results were performed using the program HypNMR2008, purchased from Protonic Software, an upgraded release of HYPNMR handling general host-guest association equilibria.[33]

4(5)-(2-(Benzylthio)ethyl)-1-*H*-imidazole (1a): Sodium hydride (60% in mineral oil, 810 mg, 20.2 mmol) was added to a solution of 4(5)-(2-bromoethyl)-1-*H*-imidazole hydrochloride and 4(5)-(2-chloroethyl)-1-*H*-imidazole hydrochloride[34] (3.32 g) in DMF (30 mL) at 0 °C under nitrogen. The resulting mixture was stirred for 30 min, followed by adding a solution of benzyl mercaptan (5.05 g, 40.6 mmol) in DMF (30 mL) dropwise. The mixture was stirred at 0 °C for 2 h, and then allowed to warm to room temperature and stirred overnight. The solvent was removed by rotary evaporation under reduced pressure, and then the residue was suspended in water (100 mL), extracted with chloroform (4×50 mL). The combined organic layer was washed with water (3×50 mL), brine (100 mL), and then concentrated. The crude product was purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:10 to 10:100. Compound **1a** was obtained as a white powder (605 mg). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.73$ (t, 2H, ^3J (H,H) = 7.2 Hz), 2.86 (t, 2H, ^3J (H,H) = 7.2 Hz), 3.72 (s, 2H), 6.79 (s, 1H), 7.23–7.31 (m, 5H), 7.54 ppm (s, 1H); HRMS (APCI+): *m/z* calculated for C$_{12}$H$_{14}$N$_2$S+H: 219.0956; found: 219.0962.

1-Benzyl-4(5)-(2-(benzylthio)ethyl)-1*H*-imidazole (2a): Sodium hydride (60% in mineral oil, 20 mg, 0.5 mmol) was added to a solution of 4(5)-(2-(benzylthio)ethyl)-1-*H*-imidazole (**1a**) (109 mg, 0.5 mmol) in dioxane (2 mL) at 0 °C under nitrogen. The resulting mixture was stirred for 30 min, and then the benzyl bromide (77 mg, 0.45 mmol) was added dropwise. The mixture was heated to 90 °C and stirred overnight. After cooling to room temperature, the reaction was quenched with water (5 mL) and extracted with chloroform (5×5 mL). The combined organic layer was washed with brine (10 mL) and concentrated. The crude product was purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 8:100. Compound **2a** was obtained as clear oil (129 mg, 84%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.72$ (m, 2H), 2.80 (m, 2H), 3.61 and 3.69 (s, s, 2H), 4.98 and 5.02 (s, s, 2H), 6.61 and 6.82 (s, s, 1H), 7.43 and 7.44 (s, s, 2H), 7.12–7.36 ppm (m, 10H); HRMS (APCI+): *m/z* calculated for C$_{19}$H$_{20}$N$_2$S+H: 309.1425.1; found: 309.1420.

1-Benzyl-4(5)-(2-(benzylthio)ethyl)-1*H*-imidazole-2-carbonitrile (3a): Cyanogen bromide (111 mg, 1.05 mmol) was added to a solution of 4-dimethylaminopyridine (128 mg, 1.05 mmol) in DMF (1.5 mL) at 0 °C under nitrogen. After 30 min, a solution of **2a** (129 mg, 0.42 mmol) in DMF (0.6 mL) was added dropwise to the CAP solution at 0 °C. The resulting mixture was stirred at 40 °C for 16 h and then quenched with sat. sodium bicarbonate solution (10 mL), and extracted with ethyl acetate (4×5 mL). The combined organic layer was concentrated, and the residue was purified by flash chromatography in a silica gel column with a gradient of ethyl acetate in hexanes from 0:100 to 20:100. Compound **3a** was obtained as yellowish oil (117 mg, 84%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.49$–2.78 (m, 4H), 3.61 and 3.67 (s, s, 2H), 5.16 and 5.18 (s, s, 2H), 6.77 and 6.96 (s, s, 1H), 6.99–7.38 ppm (m, 10H); IR (Contact Smart Orbit): $\tilde{v} = 2201$ cm^{-1} (CN stretching); HRMS (APCI+): *m/z* calculated for C$_{20}$H$_{19}$N$_3$S+H: 334.1378; found: 334.1366.

1-Benzyl-4(5)-(2-(benzylthio)ethyl)-1-*H*-imidazole-2-carboxamide (4a): A mixture of **3a** (48 mg, 0.14 mmol) in sulfuric acid (20% by volume, 0.3 mL) and TFA (167 uL) was stirred overnight at 75 °C. After cooling, the solvent was removed by rotary-evaporation, and the residue was purified by flash chromatography in a silica gel column with a gradient of ethyl acetate in hexanes from 0:100 to 20:100. Compound **4a** was obtained as a white powder (23 mg, 46%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.44$ (t, 2H, ^3J (H,H) = 7.6 Hz), 2.59 (t, 2H, ^3J (H,H) = 7.6 Hz), 3.55 (s, 2H), 5.63 (s, 2H), 6.78 (s, 1H), 6.92–7.24 ppm (m, 12H); HRMS (APCI+): *m/z* calculated for C$_{20}$H$_{21}$N$_3$OS+H: 352.1484; found: 352.1493.

4(5)-(2-Mercaptoethyl)-1-*H*-imidazole-2-carboxamide (5a): A solution of **4a** (34 mg, 0.097 mmol) in diethyl ether (0.2 mL) was added to a solution of sodium (40 mg) in ammonia (anhydrous) at - 76 °C under nitrogen. During the reaction time, more sodium was added to keep the solution blue. After 5 h, the reaction was quenched by addition of ammonium chloride (40 mg), allowing ammonia to evaporate at room temperature. The residue was washed with hexanes and diethyl ether, and separated by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 50:50. Compound **5a** was obtained as a white powder (15 mg, 77%). ^{1}H NMR (400 MHz, CD$_3$OD): $\delta = 3.04$–3.15 (m, 4H), 4.14 (s, 1H), 7.32 ppm (s, 1H); HRMS (APCI+): *m/z* calculated for C$_6$H$_9$N$_3$OS+H: 172.0545; found: 172.0522.

N-Phthaloyl histamine (1b): The compound was synthesized following a procedure reported in literature.[35] Finely powdered N-ethoxycarbonylphthalimide (12.5 g, 55 mmol) was added to a vigorously stirred solution of histamine dihydrochloride (9.2 g, 50 mmol) and sodium carbonate (10.6 g, 100 mmol) in water (250 mL) at room temperature in a period of 30 min. After the addition, the mixture was stirred for an additional hour, filtered, and the product recrystallized from aqueous ethanol to give compound **1b** as a white power (9.2 g, 76%). ^{1}H NMR (400 MHz, DMSO): $\delta = 2.93$ (t, 2H, ^3J (H,H) = 6.8 Hz), 3.90 (t, 2H, ^3J (H,H) = 6.8 Hz), 6.76 (s, 1H), 7.46 (s, 1H), 7.74–7.82 ppm (m, 4H).

1-Benzyl-4(5)-(2-(phthaloyl)ethyl)-1-*H*-imidazole (2b): Sodium hydride (60% in mineral oil, 200 mg, 5 mmol) was added to a solution of N-phthaloylhistamine **1b** (1.21 g, 5 mmol) in dioxane (10 mL) at 0 °C under nitrogen. The resulting mixture was stirred for 30 min, followed by adding a solution of benzylbromide (0.81 g, 4.75 mmol) in dioxane (5 mL) dropwise. The mixture was stirred overnight at room temperature, quenched with water (20 mL), and extracted with chloroform (3× 20 mL). The combined organic layer was washed with water (3×50 mL), brine (50 mL), and concentrated. The crude product was purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 10:100. The compound **2b** was obtained as yellowish oil (1.56 g, 94%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.96$ (t, 2H, ^3J (H,H) = 7.4 Hz), 3.98 (t, 2H, ^3J (H,H) = 7.4 Hz), 5.02 (s, 1H), 6.68 (s, 1H), 7.10–7.30 (m, 5H), 7.44 (s, 1H), 7.68–7.81 ppm (m, 4H).

1-Benzyl-4(5)-(2-(phthaloyl)ethyl)-1-*H*-imidazole-2-carbonitrile (3b): Cyanogen bromide (2.48 g, 23.5 mmol) was added to a solution of 4-dimethylaminopyridine (2.86 g, 23.5 mmol) in DMF (23 mL) at 0 °C under nitrogen. After 30 min, a solution of **2b** (1.56 g, 4.7 mmol) in DMF (10 mL) was added dropwise to the CAP solution. The resulting mixture was stirred at 40 °C for 16 h, quenched with sat. sodium bicarbonate solution (50 mL), and then extracted with ethyl acetate (4×50 mL). The combined organic layer was concentrated, and purified by flash chromatography in a silica gel column with a gradient of ethyl acetate in hexanes from 0:100 to 50:50. Compound **3b** was obtained as yellowish oil (1.17 g, 70%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 2.91$ (t, 2H, ^3J (H,H) = 7.0 Hz), 3.95 (t, 2H, ^3J (H,H) = 7.0 Hz), 5.20 (s, 2H), 6.89 (s, 1H), 7.09–7.36 (m, 5H), 7.69–7.82 ppm (m, 4H); IR (Contact Smart Orbit): $\tilde{v} = 2230$ cm^{-1} (CN stretching); HRMS (APCI+): *m/z* calculated for C$_{21}$H$_{18}$N$_4$O$_3$+H: 357.1351; found: 357.1359.

1-Benzyl-4(5)-(2-(amine)ethyl)-1-*H*-imidazole-2-carboxamide (4b): A solution of **3b** (166 mg, 0.47 mmol) in methanol (13 mL) was basified to pH ~9 by addition of NH$_3$·H$_2$O (28–30%), to which H$_2$O$_2$ (30%, 1.25 mL) was added dropwise. The resulting mixture was stirred overnight at room temperature, and then concentrated. The residue was dissolved in hydrazine (1.0 M in THF, 2.2 mL) and refluxed overnight. After removing the solvent, the residue was purified by flash chromatography

in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 50:100. Compound **4b** was obtained as a white powder (77 mg, 32%). ^{1}H NMR (400 MHz, CD$_3$OD): $\delta = 2.89$ (t, 2H, ^3J (H,H) = 6.6 Hz), 3.24 (t, 2H, ^3J (H,H) = 6.6 Hz), 5.69 (s, 2H), 7.18 (s, 1H), 7.26–7.46 ppm (m, 5H); HRMS (APCI+): m/z calculated for C$_{13}$H$_{16}$N$_4$O+H: 245.1402; found: 245.1402.

4(5)-(2-(Amine)ethyl)-1-*H*-imidazole-2-carboxamide (5b): A solution of **4b** (40 mg, 0.16 mmol) in diethyl ether (0.2 mL) was added to a solution of sodium (40 mg) in ammonia (anhydrous) at - 76°C under nitrogen. During the reaction time, more sodium was added to keep the solution blue. After 5 h, the reaction was quenched with ammonium chloride (40 mg), and then the ammonia was allowed to evaporate at room temperature. The residue was wished with hexanes and diethyl ether, and purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 50:50. Compound **5b** was obtained as a white powder (21 mg, 85%). ^{1}H NMR (400 MHz, CD$_3$OD): $\delta = 2.98$ (t, 2H, ^3J (H,H) = 6.6 Hz), 3.26 (t, 2H, ^3J (H,H) = 6.6 Hz), 7.13 (s, 1H), 7.37 ppm (br s, 2H); HRMS (APCI+): m/z calculated for C$_6$H$_{10}$N$_4$O+H: 155.0933; found: 155.0929.

4(5)-((*tert*-butyldimethylsilyloxy)methyl)-1-*H*-imidazole (1c): The compound was synthesized following a procedure reported in literature.[36] To a solution of 4(5)-hydroxymethyl-1-*H*-imidazole hydrochloride (200 mg, 1.48 mmol) in DMF (3 mL) were successively added imidazole (334 mg, 4.90 mmol), TBDMSCl (246 mg, 1.64 mmol) and 4-dimethylaminopyridine (36.4 mg, 0.30 mmol). The resulting mixture was stirred at room temperature overnight and then at 60°C for 3 h, diluted with water (2 mL), and then extracted with a co-solvent of hexanes and ether (v/v = 1:1, 6 mL). The organic layer was washed with water, brine, and then concentrated. Compound **1c** was obtained as a colorless solid (250 mg, 79%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 0.09$ (s, 6H), 0.92 (s, 9H), 4.74 (s, 2H), 6.92 (s, 1H), 7.60 ppm (s, 1H).

1-Benzyl-4(5)-((*tert*-butyldimethylsilyloxy)methyl)-1-*H*-imidazole (2c): Sodium hydride (60% in mineral oil, 88 mg, 2.2 mol) was added to a solution of **1c** (0.47 g, 2.2 mmol) in dioxane (4.2 mL) at room temperature. After 30 min, benzyl bromide (248 μL, 2.09 mmol) was added into the mixture, which was stirred at 90°C for 18 h and then the solvent was removed by rotary evaporation under reduced pressure. The residue was diluted with water (5 mL) and extracted with chloroform (3×5 mL). The combined organic layer was washed with water (3×5 mL), brine (10 mL), and then concentrated. The crude product was purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 10:100. Compound **2c** was obtained as yellowish oil (0.44 g, 66%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 0.09$ (s, 6H), 0.86 and 0.91 (s, s, 9H), 4.52 and 4.69 (s, s, 2H), 5.06 and 5.20 (s, s, 2H), 6.82–7.46 ppm (m, 6H).

1-Benzyl-4(5)-((*tert*-butyldimethylsilyloxy)methyl)-1-*H*-imidazole-2-carbonitrile (3c): Cyanogen bromide (132 mg, 1.25 mmol) was added to a solution of 4-dimethylaminopyridine (152 mg, 1.25 mmol) in DMF (2.5 mL) at 0°C under nitrogen. After 30 min, a solution of **2c** (151 mg, 0.5 mmol) in DMF (0.6 mL) was added dropwise to the CAP solution. The resulting mixture was stirred at 40°C for 16 h, quenched with sat. sodium bicarbonate solution (14 mL), and extracted with ethyl acetate (3×10 mL). The combined organic layer was concentrated, and the residue was purified by flash chromatography on a silica gel column with a gradient of ethyl acetate in hexanes from 0:100 to 20:100. Compound **3c** was obtained as yellowish oil (133 mg, 85%). IR: $v = 2240$ cm^{-1} (CN stretching); ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 0.08$ (s, 6H), 0.88 (s, 9H), 4.66 (s, 2H), 5.25 (s, 2H), 6.99 (s, 1H), 7.24–7.40 ppm (m, 5H); HRMS (APCI+): m/z calculated for C$_{18}$H$_{25}$N$_3$OSi+H: 328.1845; found: 328.1847.

1-benzyl-4(5)-((*tert*-butyldimethylsilyloxy)methyl)-1-*H*-imidazole-2-carboxamide (4c): A solution of **3c** (133 mg, 0.41 mmol) in methanol (10 mL) was basified to pH 9–10 by addition of ammonium hydroxide. To this solution, H$_2$O$_2$ (30%, 1.25 mL) was added dropwise at 0°C. The resulting mixture was stirred for 16 h at room temperature. After the solvent was removed, the residue was purified by flash chromatography in a silica gel column with a gradient of ethyl acetate in hexanes from 0:100 to 50:100. Compound **4c** was obtained as a white powder (90 mg, 64%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 0.08$ (s, 6H), 0.90 (s, 9H), 4.53 and 4.64

(s, s, 2H), 5.26 (br s, 1H), 5.68 and 5.87 (s, s, 2H), 6.92 and 6.99 (s, s, 1H), 7.08–7.33 ppm (m, 6H); HRMS (APCI+): m/z calculated for C$_{18}$H$_{27}$N$_3$O$_2$Si+H: 346.1951; found: 346.1949.

4(5)-((*tert*-Butyldimethylsilyloxy)methyl)-1-*H*-imidazole-2-carboxamide (5c): A solution of **4c** (78 mg, 0.23 mmol) in diethyl ether (12 mL) was added to a solution of sodium (80 mg) in ammonia at −76°C under nitrogen. During the reaction time, more sodium was added to keep the solution blue. After 2 h, the reaction was quenched with ammonium chloride (80 mg), and the ammonia allowed evaporating at room temperature. The residue was washed with hexanes and diethyl ether, and purified by flash chromatography in a silica gel column with a gradient of methanol in dichloromethane from 0:100 to 50:100. Compound **5c** was obtained as a white powder (18 mg, 35%). ^{1}H NMR (400 MHz, CDCl$_3$): $\delta = 0.10$ (s, 6H), 0.91 (s, 9H), 4.73 (s, 2H), 5.93 (br s, 1H), 7.08 (s, 1H), 7.64 ppm (br s, 1H); HRMS (APCI+): m/z calculated for C$_{11}$H$_{21}$N$_3$O$_2$Si+H: 256.1481; found: 256.1488.

UV determination of acid dissociation constants (p*K*$_a$): for the p*K*a determination, a set of UV spectra were first collected at different pH values, from which data was then extracted and analyzed according to the method reported in literature.[37] The p*K*$_a$ values of 1-*H*-imidazole-2-carboxamide were determined from Method a, and those of **5a** from both Method a and b.

Method a: 14 buffer solutions with different pH values were prepared using KCl/HCl (pH 1), C$_6$H$_8$O$_7$/HCl/NaCl (pH 2), C$_6$H$_8$O$_7$/NaOH/NaCl (pH 3), C$_6$H$_8$O$_7$/NaOH/NaCl (pH 4), C$_6$H$_8$O$_7$/NaOH (pH 5, 6), KH$_2$PO$_4$/Na$_2$HPO$_4$ (pH 7), Na$_2$B$_4$O$_7$/HCl (pH 8, 9), Na$_2$B$_4$O$_7$/NaOH (pH 10), H$_3$BO$_3$/NaOH/KCl (pH 11), Na$_2$HPO$_4$/NaOH (pH 12), NH$_2$CH$_2$COOH/NaCl/NaOH (pH 13), NaOH (pH 14). Stock solutions of 1-*H*-imidazole-2-carboxamide and 4(5)-(2-mercaptoethyl)-1-*H*-imidazole-2-carboxamide (**5a**) were prepared in DI water with concentration of 1.65×10^{-3} M and 7.53×10^{-4} M, respectively. Working solutions were prepared by 20-fold dilution of the stock solutions with buffers. Measurements were carried out in the same day as fresh working solutions were prepared.

Method b: An aliquot (0.6 mL) of a **5a** stock solution (2×10^{-3} M) in aq. HCl (0.1 M) was added to a solution of **5a** (2×10^{-3} M) in aq. NaOH (0.1 M), followed by recording the UV spectrum and measuring the pH value of the resultant solution. The same procedure was repeated in the same solution until a set of UV spectra were collected with the pH ranging from1.0 to 13.0. The temperature was controlled between 25 and 27°C.

Data analysis: The p*K*$_a$s were determined based on the Henderson-Hasselbalch equation, which can be expressed as below:

$$pK_a = pH - \log (\beta/1-\beta)$$

where β is the degree of dissociation.

Given that UV absorption is additive, the total absorbance of a weak acid solution can be expressed as the sum of those of the acid and its conjugate base. The UV absorbances of the imidazole-2-carboxamide at the wavelength of 258 nm ($A_{\lambda 1}$), 268 nm ($A_{\lambda 2}$), and 281 nm ($A_{\lambda 3}$) are presented in following equations. C$_I$, C$_{II}$, C$_{III}$ are concentrations of the form I, II, III, respectively, and b is the light path length. Accordingly, the extinction coefficients of I, II, III at 258 nm are designated as ε_{1-I}, ε_{1-II}, and ε_{1-III} for example.

$$A_{\lambda 1} = \varepsilon_{1-I} bC_I + \varepsilon_{1-II} bC_{II} + \varepsilon_{1-III} bC_{III}$$

$$A_{\lambda 2} = \varepsilon_{2-I} bC_I + \varepsilon_{2-II} bC_{II} + \varepsilon_{2-III} bC_{III}$$

$$A_{\lambda 3} = \varepsilon_{3-I} bC_I + \varepsilon_{3-II} bC_{II} + \varepsilon_{3-III} bC_{III}$$

Assuming that the imidazole exists in a positively charged form (I) at pH 1.7, neutral form (II) at pH 8.3, and negatively charged form (III) at pH 12.2, the extinction coefficient of each form at its absorption maximum was estimated from the UV absorbance (Table 2). Using these equations, the degrees of dissociation (β) were calculated from the total absorbances at different pH values with the total concentration of the species constant. The β values were plotted against corresponding pH

Table 2. Extinction coefficients (ε, $mM^{-1}cm^{-1}$) of three forms of **5a** at their maximum wavelengths.

	ε_{258} [nm]	ε_{268} [nm]	ε_{281} [nm]
I	6.761	5.481	1.816
II	6.039	7.231	4.908
III	3.796	5.697	7.158

values, which was fit with a sigmoidal function (Boltzmann fitting) in the Origin software. The pK_a value was directly read out from pH at $\beta = 0.5$ on the fitting curve. To simplify the calculation, C_{III} was set to zero at pH lower than 7.0 and C_I was set to zero at pH higher than 7.0. The distribution of three forms of **5a** at different pH values was calculated in the same manner, shown in Figure 2c.

Computational methods: All the DFT calculations were performed in the Spartan'10 for Windows, Wavefunction Inc. Individual 2D chemical structures were drawn in ChemBioDraw Ultra 11 and exported to Spartan'10 to generate corresponding 3D structures from which hydrogen bonding dimers were generated as well. The geometries of these structures were subjected to energy minimization using the built-in MMFF molecular mechanics prior to the DFT calculations. All the structures were first calculated using the B3LYP with the 6–311++G** basis set in vacuum, followed by B3LYP/6–31G* in chloroform, from which NMR chemical shifts were simultaneously calculated. The DFT calculations for hydrogen bonding interactions of the imidazole-2-carboxamide with DNA bases were performed using B3LYP/6–31G* basis set in vacuum.

^{1}H NMR titrations: ^{1}H NMR titration experiments were carried out at 24 ± 1°C using either a Varian 500 spectrometer operating at 500 MHz or a Varian 400 spectrometer operating at 400 MHz. All chemical shifts are reference to CDCl$_3$ and reported in ppm. The **5c** solution with a known concentration, typically 5×10^{-3} M, was made up in an NMR tube using deuterated chloroform (0.5 mL). The titrating solutions of TBDMS silylated nucleosides were made up in volumetric flasks (2 mL) with a concentration much greater than that of **5c**. In a typical titration, one of the titrating solutions was added to the **5c** solution in 7–14 separate portions. Following each of additions, NMR spectrum was recorded. For example, the dT nucleoside solution with the concentration of 800×10^{-3} M was added incrementally in the order of 1.25, 1.25, 1.25, 1.25, 2.5, 2.5, 2.5, 2.5, 5, 5, and 5 µL aliquots, up to a total volume of 30 µL. The titration data were analyzed by nonlinear regression analysis using the HypNMR2008 program. The stoichiometry of **5c**-nucleoside complexes was established by mole ratio plots.[38]

Acknowledgements

[1] S. Chang, J. He, P. Zhang, B. Gyarfas, S. Lindsay, *J. Am. Chem. Soc.* **2011**, *133*, 14267–14269.

[2] B. Yngvadottir, D. G. MacArthur, H. Jin, C. TylerSmith, *GenomeBiology* **2009**, *10*, 237231–237234.

[3] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.

[4] R. Mukhopadhyay, *Anal. Chem.* **2009**, *81*, 1736–1740.

[5] C. Alkan, S. Sajjadian, E. E. Eichler, *Nat. Methods* **2011**, *8*, 61–65.

[6] a) M. Snyder, J. Du, M. Gerstein, *Genes Dev.* **2010**, *24*, 423–431; b) J. C. Venter, *Nature* **2010**, *464*, 676–677.

[7] a) M. Zwolak, M. D. Ventra, *Rev. Mod. Phys.* **2008**, *80*, 141–165; b) D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. D. Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V.

Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1146–1153.

[8] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, H. Bayley, *Nat. Nanotechnol.* **2009**, *4*, 265–270.

[9] I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16060–16065.

[10] J. J. Kasianowicz, E. B. Eric, D. Branton, D. W. Deamer, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13770–13773.

[11] a) M. Zwolak, M. D. Ventra, *Nano lett.* **2005**, *5*, 421–424; b) A. P. Ivanov, E. Instuli, C. McGilvery, G. Baldwin, D. W. McComb, T. Albrecht, J. B. Edel, *Nano Lett.* **2011**, *11*, 279–285.

[12] S. Huang, J. He, S. Chang, P. Zhang, F. Liang, S. Li, M. Tuchband, A. Fuhrman, R. Ros, S. Lindsay, *Nat. Nanotechnol.* **2010**, *5*, 868–873.

[13] S.-W. Nam, M. J. Rooks, K.-B. Kim, S. M. Rossnagel, *Nano Lett.* **2009**, *9*, 2044–2048.

[14] A. Richaud, N. Barba-Behrens, F. Mendez, *Org. Lett.* **2011**, *13*, 972–975.

[15] D. E. Bergstrom, P. Zhang, J. Zhou, *J. Chem. Soc. Perkin Trans. 1* **1994**, 3029–3034.

[16] K. L. Kirk, *J. Org. Chem.* **1978**, *43*, 4381–4383.

[17] J. R. McCarthy, D. P. Matthews, J. P. Whitten, *Tetrahedron Lett.* **1985**, *26*, 6273–6276.

[18] J. N. Moorthy, N. Singhal, *J. Org. Chem.* **2005**, *70*, 1926–1929.

[19] D. E. Bergstrom, P. Zhang, *Tetrahedron Lett.* **1991**, *32*, 6485–6488.

[20] T. L. Gilchrist, *Heterocyclic Chemistry*, 3rd ed., Longman, **1997**.

[21] R. B. Homer, C. D. Johnson, in *The Chemistry of Amides* (Ed.: J. Zabicky), Wiley, **1970**, pp. 187–243.

[22] B. Roig, E. Chalmin, E. Touraud, O. Thomas, *Talanta* **2002**, *56*, 585–590.

[23] B. Kuhn, P. Mohr, M. Stahl, *J. Med. Chem.* **2010**, *53*, 2601–2611.

[24] E. P. Papadopoulos, U. Hollstein, *Organic Magnetic Resonance* **1982**, *19*, 188–191.

[25] a) G. N. Ten, V. I. Baranov, *Journal of Applied Spectroscopy* **2008**, *75*, 168–173; b) G. N. Ten, T. G. Burova, V. I. Baranov, *Journal of Structural Chemistry* **2007**, *48*, 623–633.

[26] a) S. G. Stepanian, G. G. Sheina, E. D. Radchenko, Y. P. Blagoi, *Journal of Molecular Structure* **1985**, *131*, 333–346; b) G. G. Sheina, S. G. Stepanian, E. D. Radchenko, Y. P. Blagoi, *Journal of Molecular Structure* **1987**, *158*, 275–292.

[27] K. Weisz, J. Jähnchen, H.-H. Limbach, *J. Am. Chem. Soc.* **1997**, *119*, 6436–6437.

[28] J. Sponer, J. Leszczynski, P. Hobza, *Biopolymers* **2002**, *61*, 3–31.

[29] V. Apalkov, X.-F. Wang, T. Chakraborty, in *Charge Migration in DNA* (Ed.: T. Chakraborty), Springer, Berlin Heidelberg, **2007**, pp. 77–119.

[30] S. Chang, S. Huang, J. He, F. Liang, P. Zhang, S. Li, X. Chen, O. Sankey, S. Lindsay, *Nano Lett.* **2010**, *10*, 1070–1075.

[31] J. Sartorius, H.-J. Schneider, *Chem. Eur. J.* **1996**, *2*, 1446–1452.

[32] A. Fuhrmann, S. Getfert, Q. Fu, P. Reimann, S. Lindsay, R. Ros, *Biophysical Journal*, **2011**, submitted.

[33] a) C. Frassineti, L. Alderighi, P. Gans, A. Sabatini, A. Vacca, S. Ghelli, *Anal. Bioanal. Chem.* **2003**, *376*, 1041–1052; b) C. Frassineti, S. Ghelli, P. Gans, A. Sabatini, M. S. Moruzzi, A. Vacca, *Anal. Biochem.* **1995**, *231*, 374–382.

[34] R. Ting, L. Lermer, D. M. Perrin, *J. Am. Chem. Soc.* **2004**, *126*, 12720–12721.

[35] J. C. Emmett, F. H. Holloway, J. L. Turner, *J. Chem. Soc., Perkin Trans. 1* **1979**, 1341–1344.

[36] Y. Amino, H. Eto, C. Eguchi, *Chemical & Pharmaceutical Bulletin* **1989**, *37*, 1481–1487; Pharmaceutical Bulletin **1989**, 37, 1481–1487.

[37] C. Rether, C. Schmuck, *Eur. J. Org. Chem.* **2011**, 1459–1466.

[38] P.-N. Cheng, C.-F. Lin, Y.-H. Liu, C.-C. Lai, S.-M. Peng, S.-H. Chiu, *Org. Lett.* **2006**, *8*, 435–438.