Organic & Biomolecular Chemistry

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 8654

www.rsc.org/obc

PAPER

1,8-Naphthyridine-2,7-diamine: a potential universal reader of Watson–Crick base pairs for DNA sequencing by electron tunneling†

Feng Liang, Stuart Lindsay* and Peiming Zhang*

Received 2nd August 2012, Accepted 13th September 2012
DOI: 10.1039/c2ob26529j

With the aid of Density Functional Theory (DFT), we designed 1,8-naphthyridine-2,7-diamine as a recognition molecule to read DNA base pairs for genomic sequencing by electron tunneling. NMR studies show that it can form stable triplets with both A : T and G : C base pairs through hydrogen bonding. Our results suggest that the naphthyridine molecule should be able to function as a universal base pair reader in a tunneling gap, generating distinguishable signatures under electrical bias for each of DNA base pairs.

Introduction

Next generation DNA sequencing (NGS) has revolutionized many aspects of biological science, ranging from human disease analyses[1–5] to drug discovery[6] to environmental monitoring.[7] Today, sequencing a human genome can be completed in a week and with several thousand dollars. Despite these great advances, the NGS technologies are limited by their notorious short read length and low accuracy in comparison to the conventional Sanger sequencing.[8,9] In addition, the NGS sequencers rely on delicate biochemical reagents for sequencing reactions and require sophisticated optical instruments for signal readouts (apart from the Ion PGM™ from Life Technologies which electronically measures protons[10]). All of these can be roadblocks in further reducing the cost for their use in clinics. In parallel with NGS, nanopore-based devices have being developed as a disruptive platform to sequence single DNA molecules electronically, reagent-free, with long read lengths.[11,12] Although steady progress has been made,[13,14] the nanopore sequencing has yet to achieve a single base resolution. To address this issue, we have recently demonstrated that a STM tip functionalized with benzamide can sense individual DNA bases in a short oligonucleotide on a gold substrate functionalized with a benzamide monolayer.[15] This sub-nanometer spatial resolution opens a door to sequencing DNA by electron tunneling. In order to sequence single DNA molecules with high accuracy, we devised a recognition scheme in a tunneling gap that is incorporated into a nanopore, as shown in Fig. 1. When single stranded DNA translocates through the nanopore, each of its bases is sequentially trapped in a tunneling gap by forming a triplet complex with two electrodes that are functionalized with a base reader and a base pair reader respectively. This complicated recognition system in a nanogap is challenging to construct. However, we were able to demonstrate that the naturally occurring nucleosides can be recognized by nucleobases in the tunneling nanogap following the Watson–Crick base pairing rule.[16] In this present study, we have identified a naphthyridine molecule that can interact with Watson–Crick base pairs through hydrogen bonding, laying down the foundation for us to investigate triplexation through electron tunneling.

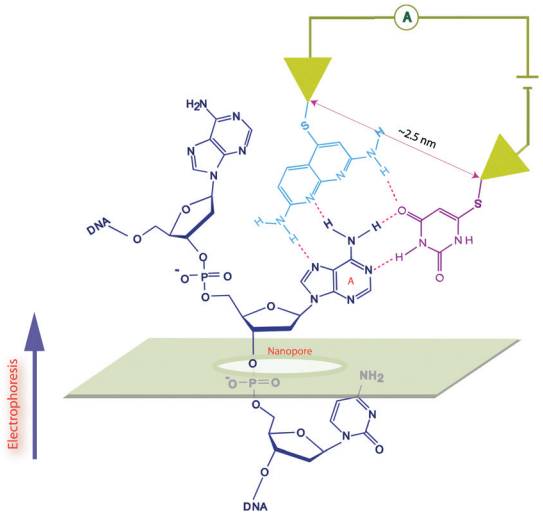


Fig. 1 The concept of DNA sequencing by trans-base electron tunneling. The nanoelectrodes can be embedded in or laid on a nanopore using semiconductor nanotechnologies.

Center for Single Molecule Biophysics, The Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA.
E-mail: stuart.lindsay@asu.edu, peiming.zhang@asu.edu;
Tel: +1 480-965-4691; +1 480-965-5491

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c2ob26529j

Results and discussion

We chose 1,8-naphthyridine-2,7-diamine (designated as **N**) as a candidate to read the base pairs. There are varied structures that have been investigated as motifs of triplex forming oligonucleotides (TFO) to recognize the Watson–Crick base pairs in double stranded DNA.[17] Two typical examples are shown in Fig. 2a, which exhibit fair affinity and selectivity to the DNA base pairs when incorporated into TFO.[18,19] One common feature of these structures is that they are configured either by connecting two aromatic rings together through a free rotating sigma bond or by modifying the DNA base with a functional tail to match the hydrogen bonding patterns of the Watson–Crick base pairs in the major groove side. However, such flexibility results in a loss of entropy when a hydrogen bonded complex is formed because the freely rotating bonds become fixed. This would not be suitable for our recognition scheme in which the hydrogen bonding may be a dominating force for formation of a stable triplet and the important base stacking interactions may be not as effective as that in the double stranded DNA. We postulate that a rigid and planar scaffold with the right geometry will have an advantage in this regard. Furthermore, if a tandem hydrogen bonding site array is constructed along one edge of the scaffold without any C–H interruption, it should increase the hydrogen bonding cooperativity. 1,8-Naphthyridine-2,7-diamine is a molecule that contains an aromatic plane composed of two fused pyridyl rings and two amines aligned with the ring nitrogen atoms to form an array of four hydrogen bonding sites (see Fig. 1). It has been exploited as a moiety to create new base pairs in DNA.[20,21] In addition, the amine derivatives of naphthyridine form stable complexes with guanine[22,23] and deaza guanine,[24] and have been used as a fluorescent dye to stain nucleoli in the nucleus of MDCK-cells.[25] However, the interactions of 1,8-naphthyridine-2,7-diamine with the DNA base pairs have not been explored. With the aid of Density Functional Theory (DFT), we first scrutinized the structural fitness through computer modeling. As illustrated in Fig. 2b, 1,8-naphthyridine-2,7-diamine can form hydrogen bonded complexes of **N**–T : A and **N**–C : G with the Watson–Crick base pairs from the major groove sides. In the computer simulation, a methyl group was placed at the 3-position of **N** as a prospective site for attachment, and all the sugars connected to DNA bases were substituted with methyl groups in order to reduce the computing time. Note that the methyl substitution would not exert a game changing influence on our computing results since the hydrogen bonding interactions take place on the opposite sides of the sugars. The DFT calculations show that there is a gain of ~15 kcal mol^{-1} in energy when **N** hydrogen-bonds to either T : A or C : G base pairs in a vacuum, which is slightly higher than the hydrogen-bonding energy of the T : A Watson–Crick base pair (Table S1 in ESI†). The distance between the two amino nitrogen atoms (~6.86 Å) of **N** is fairly matched to that from N-7 of adenine to O-4 of thymine (~6.26 Å) in the T : A base pair, resulting in formation of a good fit **N**–T : A triplet. In the **N**–C : G triplet, **N** is twisted out of the C : G base pair plane due to a steric hindrance between the 2-position amine of **N** and the 5-position hydrogen of cytosine. The DFT solvation calculation indicates that the triplets are slightly less stable in DMSO, a solvent that has a dielectric constant (ε = 46.7 D) comparable to one in the major groove of DNA (ε = 55 D).[26] We believe that the hydrogen bonding interactions should prevail in the nanogap, which has a local environment less hydrophilic than the bulk aqueous solution especially when it is functionalized with organic molecules.

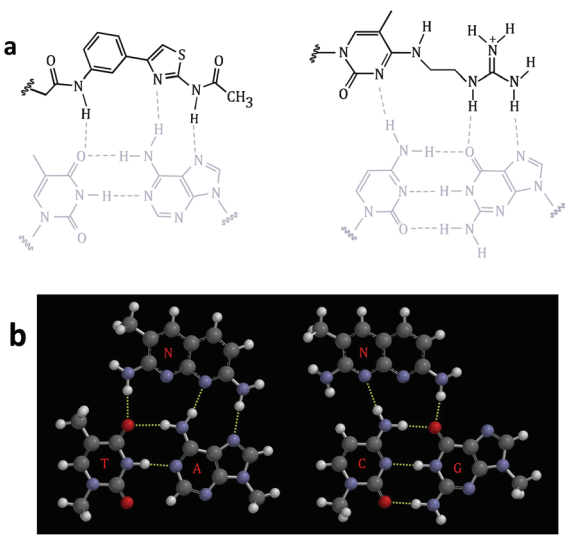


Fig. 2 (a) DNA base pair recognition molecules for TFO and their hydrogen bonding interactions with DNA base pairs; (b) DFT models of the complexes of 1,8-naphthyridine-2,7-diamine with the T : A and C : G base pairs.

Following the computer modeling studies, we investigated the hydrogen bonding interactions of **N** with the DNA base pairs formed by individual nucleosides in solutions using NMR techniques. To adequately dissolve these entities into chloroform (a commonly used solvent for hydrogen bonding studies), **N** was converted to an amine-acetylated derivative (**N^{ac}**), and the hydroxyl groups of the naturally occurring nucleosides were silylated with *tert*-butyldimethylsilyl chloride (designated as dA, dC, dG, dT, and dU). Our primary focus was on the hydrogen bonding of **N^{ac}** with the A : T base pair because the computer modeling showed a good fit between these two entities. First, an NMR spectrum of a mixture of dA and dT in deuterated chloroform (in a 1 : 1 molar ratio) was recorded to confirm the base pairing (upper spectrum in Fig. 3a). It shows that the imino proton peak of dT has not only shifted downfield but also split into two with an integration ratio of 1 : 2, compared to that (δ_H = 9.8 ppm) in a dT only solution. This indicates that there were two different hydrogen-bonding interactions involved between dA and dT. By means of ^{1}H–^{1}H NOESY NMR (Fig. S1 in ESI†), we found that the peak at 11.8 ppm cross-talks to the H_{A2} peak of dA, and the one at 11.5 ppm cross-talks to the H_{A8} peak of dA. The NMR data can be best explained by coexistence of both Hoogsteen (HG) and Watson–Crick (WC) base pairs in equilibrium as delineated in Scheme 1. The HG base pair has been observed in crystals of alkylated nucleobase complexes by X-ray diffraction,[27,28] and in a dA : dU (silylated 2′-deoxyuridine) solution by NMR.[29] It could even exist with a 1% probability in DNA.[30,31] The early calculations predicted that the HG base pair would be slightly more stable than the WC base pair.[32,33] Our data show that the HG base pair is a preferred form in the chloroform solution. Thus, we believe that the HG base

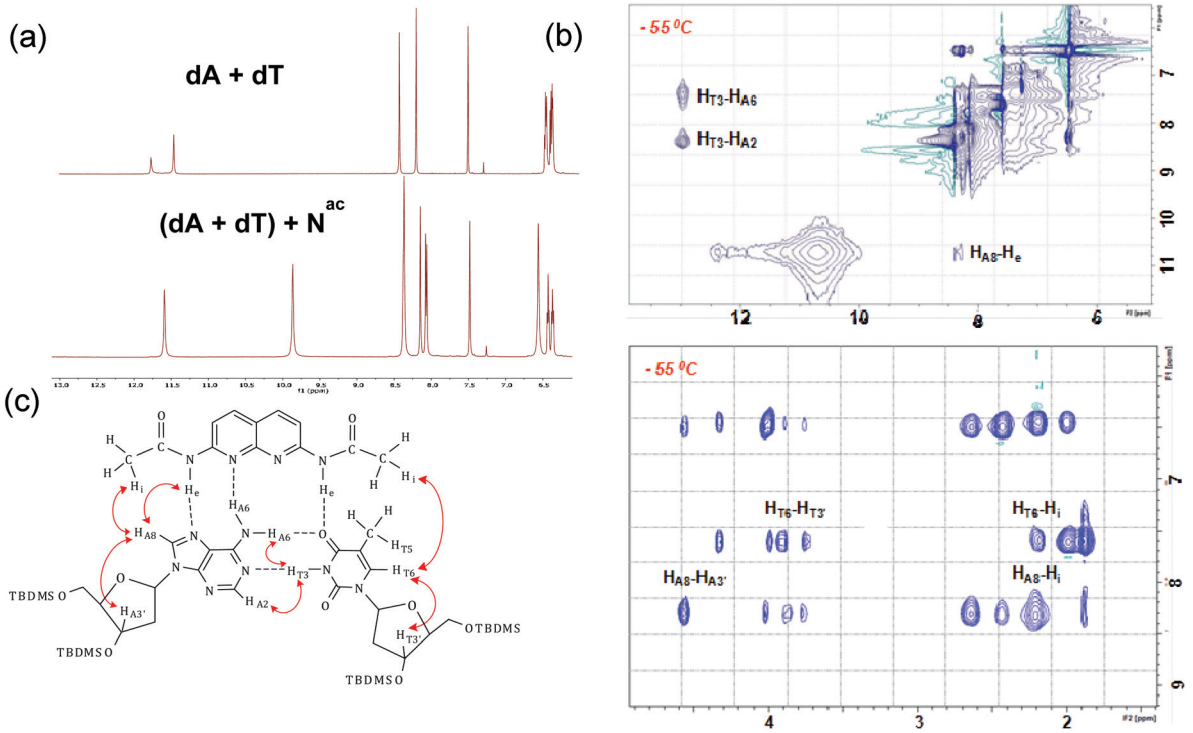
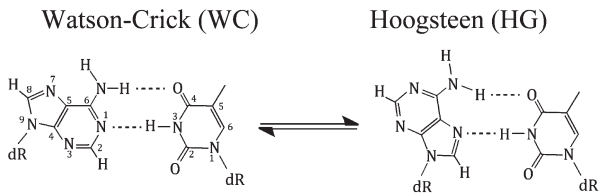


Fig. 3 (a) ^{1}H NMR spectra of a mixture of dA and dT in a ratio of 1 : 1 and a mixture of dA, dT and **N**^{ac} in a ratio of 1 : 1 : 1 at room temperature; (b) 2D ^{1}H–^{1}H NOESY NMR spectra of a mixture of dA, dT and **N**^{ac} in a 1 : 1 : 1 ratio and proton correlation assignments. (c) A schematic of molecular connections in the dA, dT and **N**^{ac} complex.



Scheme 1

pairing is intrinsic at least to A and T bases. This may be one of reasons why it can occur even in DNA double helices where the geometry is constrained to favor the WC base pairing. Compared to the WC base pair, the HG base pair has a shorter distance from N9 of dA to N1 of dT.[34] As a result, the HG base pair may not fit well with **N** to form a stable triplex.

When mixing **N**^{ac} with dA and dT in a 1 : 1 : 1 ratio, a single imino proton peak was only observed in the NMR spectrum (lower spectrum in Fig. 3a). The variable concentration NMR showed that amide protons of **N**^{ac} and amino protons of the adenine were also involved in the hydrogen-bonding interactions (Fig. S2 in ESI†). Formation of a **N**^{ac}–dA : dT triplet was confirmed by ^{1}H–^{1}H NOESY NMR. In the 2D NOESY spectrum (Fig. 3b), we only observed the cross peaks from the dT imino proton (H_{T3}) correlating with H_{A6} and H_{A2} of dA (see Fig. 3c for designation of each proton), implying that only the WC base pair was formed in the complex. Furthermore, both H_i and H_e of **N**^{ac} are correlated to H_{A8} of dA, and the H_i is correlated to H_{T6} of dT as well. These NMR data allow us to sketch a connection among the three molecules shown in Fig. 3c. Due to crowding in the region of the methyl groups at the dT end of the complex, we cannot unambiguously assign the H_{T5}–H_i cross peak. The DFT modelling shows that the amide ends of **N**^{ac} are pushed out of the dA : dT base pair plane in the triplet because of the steric hindrance between the acetyl groups of **N**^{ac} and the methyl group of thymine and the H_{A8} of dA (see Fig. 4 for the triplet conformation from computer modeling). This may explain why H_i is correlated with H_{T6} in the NOESY NMR. We noticed that the two amide protons of **N**^{ac} appeared as a single peak in the NMR spectrum of the complex. They were split into two broad peaks when the temperature was lowered to −55 °C in an 800 MHz NMR spectrometer (Fig. S3 in ESI†), indicating that these two protons were in a fast exchange within the NMR timeframe at room temperature. In contrast, guanine and cytosine only form a stable WC base pair under the same conditions for the A : T base pair. The formation of a triplet between **N**^{ac} and the dG : dC base pair was confirmed by 2D NOESY NMR (Fig. S4 in ESI†).

Based on the principle of complexation-induced chemical shifts (CIS),[35–38] we have determined association constants of **N**^{ac} with DNA bases and base pairs by NMR titration (see Table 1). The silylated nucleosides (dA, dC, dG, dT, and dU) were used as either titrants or substrates, and **N**^{ac} was only used as a substrate due to its limited solubility in chloroform. In a typical NMR titration experiment, a titrant was incrementally added to a substrate solution, and a proton NMR spectrum was recorded following each addition. In general, protons directly involved in hydrogen bonding exhibited downfield chemical shifts, resulting in positive CIS values. The proton we closely monitored in the NMR titration is listed in parentheses under the

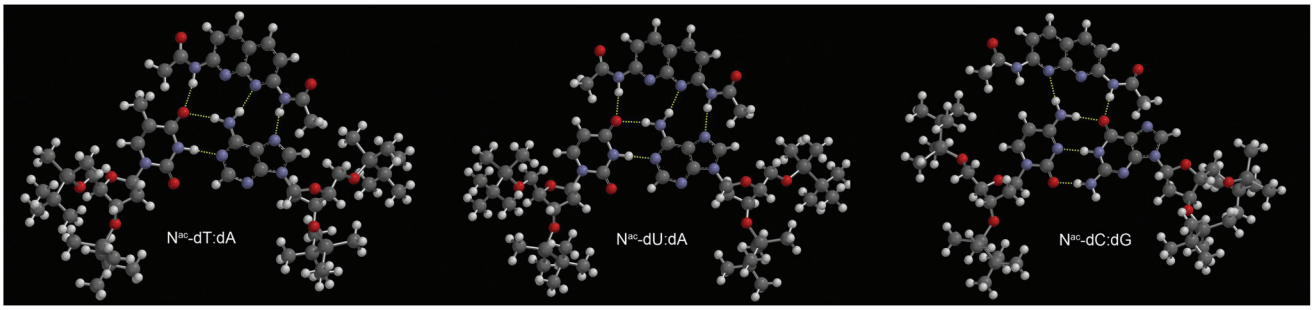


Fig. 4 DFT models of hydrogen bonded triplets of $\mathbf{N^{ac}}$ with dT : dA (A), dU : dA (B), and dC : dG (C) calculated by B3LYP in combination with 6-31G* basis sets in a vacuum.

Table 1 Complexation induced chemical shift (CIS, ppm) and association constants (K_{ass}, M^{-1}) of $\mathbf{N^{ac}}$ with individual nucleosides and nucleoside pairs derived from curve fitting of NMR titration data

		Titrant							
Substrate		dA	dC	dG	dT	dU	dA : dT	dA : dU	dG : dC
$\mathbf{N^{ac}}$ (H_e)	CIS[a]	2.1 ± 0.4	0.9 ± 0.1	1.1 ± 0.0	3.5 ± 0.4	2.0 ± 0.1	1.5 ± 0.0	1.5 ± 0.1	0.7 ± 0.0
	K_{ass}	6.3 ± 1.4	53 ± 9	3889 ± 833	18 ± 3	31 ± 3	59 ± 1	62 ± 11	525 ± 166
dG-$\mathbf{N^{ac}}$ (H_{G1})	CIS		1.2 ± 0.0						
	K_{ass}		710 ± 151						
dA-$\mathbf{N^{ac}}$ (H_{A6})	CIS				0.8 ± 0.0				
	K_{ass}				115 ± 11				
dT-$\mathbf{N^{ac}}$ (H_{T3})	CIS	0.4 ± 0.0							
	K_{ass}	126 ± 4							

[a] CIS = $\delta_{max} - \delta_{initial}$ in ppm, which was determined from the fit titration curve.

substrate in Table 1. First, the complexing stoichiometries between titrants and substrates were determined using the mole ratio plot.[39,40] We found that all of these complexes were fairly close to a 1 : 1 binding mode (Fig. S5 in ESI†). The association constants (K_{ass}) were then derived from curve fitting datasets of chemical shift *vs.* concentration in HypNMR 2008, a program to analyze NMR titration data. All of our NMR data were best fit to a 1 : 1 binding isotherm. We tested the reproducibility of our experimental process by performing the reverse titration. For example, titrating dA with dT yielded a virtually identical result as that from titrating dT with dA. The K_{ass} value (~40 M^{-1}) of the dA : dT base pairing we determined is close to that reported in the literature.[41,42] As shown in Table 1, $\mathbf{N^{ac}}$ was titrated with a series of individual nucleosides and their mixtures. Clearly, it formed a more stable complex with dC : dG than with dA : dT. Nonetheless, K_{ass} of $\mathbf{N^{ac}}$ complexing to dA : dT is comparable to that for the dA : dT base pairing. This K_{ass} value, may underestimate the actual stability of $\mathbf{N^{ac}}$ complexing to dA : dT because the dA and dT mixture mainly exists in a HG base pairing form in chloroform so that there is a free energy penalty to convert the HG base pair to the WC base pair for formation of the $\mathbf{N^{ac}}$–dA : dT triplet. When titrating a mixture of $\mathbf{N^{ac}}$ and dT with dA or a mixture of $\mathbf{N^{ac}}$ and dA with dT, the K_{ass} values (~120 M^{-1}) derived from monitoring H_{A6} and H_{T3} in these two mixed substrates are about three times higher than that of the dA : dT base pairing. This indicates that $\mathbf{N^{ac}}$ could stabilize the dA : dT base pair. It has been known that the G–C base pair is very stable in a nonpolar solvent, such as chloroform (K_{ass} = ~10^{4-5} M^{-1}).[41,43] A 1 : 1 mixture of dG and dC is often treated as a single component in NMR titration experiments.[44–47] Titrating $\mathbf{N^{ac}}$ with dG : dC yielded a relatively stable complex with an association constant of K_{ass} ~ 467 M^{-1}. However, $\mathbf{N^{ac}}$ prevents dG from base pairing with dC because it forms a very stable complex with dG (K_{ass} ~ 3990 M^{-1}). As a result, when titrating a $\mathbf{N^{ac}}$ and dG mixture with dC, a positive CIS on the imino proton of dG was obtained, indicating that there was a hydrogen bonding interaction between dG and dC, but the resultant K_{ass} value (~710 M^{-1}) is significantly smaller than one of the normal dG : dC base pairing. Thus, we have to follow an appropriate route to assemble a $\mathbf{N^{ac}}$–dG : dC triplet. We also notice that the 5-methyl group of dT did not cause any significant steric hindrance to the formation of an $\mathbf{N^{ac}}$–dT : dA triplet because there is a negligible difference in K_{ass} between $\mathbf{N^{ac}}$–dT : dA and $\mathbf{N^{ac}}$–dU : dA.

To interpret the NMR data, we constructed molecular models for the complexes of $\mathbf{N^{ac}}$ with the nucleoside pairs (Fig. 4), which were optimized by DFT calculations using B3LYP in combination with 6-31G* basis sets. As shown in Fig. 4, $\mathbf{N^{ac}}$ forms the hydrogen bonded triplets with the Watson–Crick base pairs from the major groove side. The DFT calculation indicates that $\mathbf{N^{ac}}$–dC : dG is more stable than $\mathbf{N^{ac}}$–dT : dA and $\mathbf{N^{ac}}$–dU : dA in terms of their complexing energies (ΔE^{Tot} in Table 2). This matches the results from the NMR titrations. The higher stability of the $\mathbf{N^{ac}}$–dC : dG triplet may be attributed to the

Table 2 DFT energies (kcal mol^{-1}) of the $\mathbf{N^{ac}}$ triplets

	ΔE^{Tot}	$\Delta E(\mathbf{N^{ac}})$
$\mathbf{N^{ac}}$–dT : dA	−39.3	−22.4
$\mathbf{N^{ac}}$–dU : dA	−40.6	−22.8
$\mathbf{N^{ac}}$–dC : dG	−54.5	−20.7

ΔE^{Tot}: complexing energy calculated from energy of (complex − individual monomers constituting the complex). ΔE ($\mathbf{N^{ac}}$): hydrogen bonding energy of $\mathbf{N^{ac}}$ calculated from energy of (complex − $\mathbf{N^{ac}}$ − nucleoside pair).

strong dG : dC base pair since the calculated hydrogen bonding energy of $\mathbf{N^{ac}}$ with the dG : dC pair is close to those of $\mathbf{N^{ac}}$ with the dU : dA and dT : dA pairs.

Conclusions

Our DFT calculations and NMR studies reveal that 1,8-naphthyridine-2,7-diamine can form hydrogen bonded triplets with both A : T and G : C Watson–Crick base pairs, which are as stable as the A : T base pair or more so. We have found that the naphthyridine molecule has a number of unique features: it tends to stabilize the A : T Watson–Crick base pair, block the A : T Hoogsteen base pairing, and form a stable complex with guanine to prevent the G : C base pairing. Due to differences in their structures, these triplet complexes should create different pathways for electron tunneling, resulting in distinguishable electrical signals for readout of the DNA base pairs in the tunneling gap, which makes it a universal base pair reader. We are developing chemistry to attach the naphthyridine molecules to the metal electrodes for the tunneling measurements.

Experimental section

General information

Proton NMR (^{1}H) spectra were recorded at 400 MHz on a Varian 400 MHz spectrometer, and carbon NMR (^{13}C) spectra were recorded at 100 MHz on a Varian 400 MHz spectrometer. HRMS spectra were recorded using the atmospheric pressure chemical ionization (APCI) technique. Flash chromatography was performed using automated flash chromatography (Teledyne Isco, Inc. CombiFlash Rf). 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.

Synthesis of *N*,*N*′-(1,8-naphthyridine-2,7-diyl)diacetamide ($\mathbf{N^{ac}}$)

A mixture of 1,8-naphthyridine-2,7-diamine[48] (280 mg, 1.75 mmol) in acetic anhydride (3 mL) was heated at reflux for 2.5 h. After cooling, the excess solvent was removed, and the residue was purified by flash chromatography (on silica gel with a gradient of dichloromethane–methanol from 100 : 0 to 100 : 10) to give 180 mg (42%) of the product as a yellow powder. ^{1}H NMR (400 MHz, DMSO-d_6): δ 2.16 (s, 6H), 8.21 (d, J = 9.2 Hz, 2H), 8.26 (d, J = 9.2 Hz, 2H), 10.77 (s, 2H); ^{13}C NMR (100 MHz, DMSO-d_6): δ 24.7, 113.3, 117.7, 139.2, 154.3, 154.8, 170.5. HRMS (APCI+): found, 245.1038 (calcd for $C_{12}H_{13}N_4O_2$, 245.1043).

Silylation of nucleosides

All the nucleosides were silylated with *tert*-butyldimethylsilyl chloride following our reported method.[49]

Computation

DFT calculations were performed using the program Spartan'10 for Windows, Wavefunction Inc. Individual 2D chemical structures were drawn in ChemBioDraw Ultra 11 and exported to Spartan'10 to generate the respective 3D structures, from which hydrogen bonded complexes were constructed. The geometry of each complex was first subjected to energy minimization using the built-in MMFF94s molecular mechanics, and then calculated using B3LYP/6-311++G (2df, 2p) in a vacuum. All of the calculations were successfully converged and no BSSE corrections were carried out for such a large basis set. Following completion of the calculation in a vacuum, the complex was solvated with DMSO using B3LYP/6-31G** based on a SM8 model.[50]

^{1}H NMR binding studies

Proton NMR (^{1}H) spectra were recorded at 400 MHz on a Varian 400 MHz, 500 MHz on a Varian 500 MHz or 800 MHz on a Varian 800 MHz spectrometer. All ^{1}H NMR chemical shifts were referenced to the residual non-deuterated solvent peak as 7.26 ppm in chloroform-d ($CDCl_3$). 2D NOE spectra were recorded at 400 MHz on a Bruker 400 MHz spectrometer. For VT NMR, temperature was calibrated with a standard of 100% CH_3OH 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. Temperatures below zero Celsius were achieved with a Liquid Nitrogen Heat Exchanger on the Bruker and FTS Cooling System (Stone Ridge, New York) on the Varian. $CDCl_3$ was purchased from Sigma-Aldrich, used as received without further purification. Volumetric flasks and syringes for preparing the stock solutions were rinsed with $CDCl_3$ and dried under vacuum prior to use. For NMR titration, samples were prepared from stock solutions, transferred to NMR tubes using a syringe, and diluted following the method in the previous report.[51] Association constants reported were averages of two or more repeats and were derived from fitting NMR titration data to a 1 : 1 binding isotherm using the HypNMR program.

Acknowledgements

This project is funded by the DNA sequencing technology program of the National Institute of Human Genome Research (HG004378). We thank Dr Brian Cherry of the magnetic resonance research center at ASU for his technical support and informative discussion during the NMR experiments.

Notes and references

1 J. Bras, R. Guerreiro and J. Hardy, *Nat. Rev. Neurosci.*, 2012, **13**, 453–464.
2 M. Meyerson, S. Gabriel and G. Getz, *Nat. Rev. Genet.*, 2010, **11**, 685–696.
3 M. R. Nelson, D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli, J. Shen, Z. Tang, S. A. Bacanu, D. Fraser, L. Warren, J. Aponte, M. Zawistowski, X. Liu, H. Zhang, Y. Zhang, J. Li, Y. Li, L. Li, P. Woollard, S. Topp, M. D. Hall, K. Nangle, J. Wang, G. Abecasis, L. R. Cardon, S. Zollner, J. C. Whittaker, S. L. Chissoe, J. Novembre and V. Mooser, *Science*, 2012, **337**, 100–104.
4 G. E. Palomaki, C. Deciu, E. M. Kloza, G. M. Lambert-Messerlian, J. E. Haddow, L. M. Neveux, M. Ehrich, D. van den Boom, A. T. Bombard, W. W. Grody, S. F. Nelson and J. A. Canick, *Genet. Med.*, 2012, **14**, 296–305.
5 L. Ding, T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendl, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson and J. F. DiPersio, *Nature*, 2012, **481**, 506–510.
6 P. M. Woollard, N. A. Mehta, J. J. Vamathevan, S. Van Horn, B. K. Bonde and D. J. Dow, *Drug Discovery Today*, 2011, **16**, 512–519.
7 M. Hajibabaei, S. Shokralla, X. Zhou, G. A. Singer and D. J. Baird, *PLoS One*, 2011, **6**, e17497.
8 J. Kuczynski, C. L. Lauber, W. A. Walters, L. W. Parfrey, J. C. Clemente, D. Gevers and R. Knight, *Nat. Rev. Genet.*, 2012, **13**, 47–58.
9 O. Morozova and M. A. Marra, *Genomics*, 2008, **92**, 255–264.
10 J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, J. Hoon, J. F. Simons, D. Marran, J. W. Myers, J. F. Davidson, A. Branting, J. R. Nobile, B. P. Puc, D. Light, T. A. Clark, M. Huber, J. T. Branciforte, I. B. Stoner, S. E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J. A. Fidanza, E. Namsaraev, K. J. McKernan, A. Williams, G. T. Roth and J. Bustillo, *Nature*, 2011, **475**, 348–352.
11 M. Zwolak and M. D. Ventra, *Rev. Mod. Phys.*, 2008, **80**, 141–165.
12 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 and J. A. Schloss, *Nat. Biotechnol.*, 2008, **26**, 1146–1153.
13 E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis and J. H. Gundlach, *Nat. Biotechnol.*, 2012, **30**, 349–353.
14 G. M. Cherf, K. R. Lieberman, H. Rashid, C. E. Lam, K. Karplus and M. Akeson, *Nat. Biotechnol.*, 2012, **30**, 344–348.
15 S. Huang, J. He, S. Chang, P. Zhang, F. Liang, S. Li, M. Tuchband, A. Fuhrman, R. Ros and S. Lindsay, *Nat. Nanotechnol.*, 2010, **5**, 868–873.
16 S. Chang, J. He, A. Kibel, M. Lee, O. Sankey, P. Zhang and S. Lindsay, *Nat. Nanotechnol.*, 2009, **4**, 297–301.
17 Y. Hari, S. Obika and T. Imanishi, *Eur. J. Org. Chem.*, 2012, 2875–2887.
18 D. A. Rusling, V. E. C. Powers, R. T. Ranasinghe, Y. Wang, S. D. Osborne, T. Brown and K. R. Fox, *Nucleic Acids Res.*, 2005, **33**, 3025–3032.
19 A. Semenyuk, E. Darian, J. Liu, A. Majumdar, B. Cuenoud, P. S. Miller, A. D. MacKerell Jr. and M. M. Seidman, *Biochemistry*, 2010, **49**, 7867–7878.
20 K. Kuramoto, N. Tarashima, Y. Hirama, Y. Kikuchi, N. Minakawa and A. Matsuda, *Chem. Commun.*, 2011, **47**, 10818–10820.
21 S. Ogata, K. Kuramoto, N. Inoue, N. Minakawa and A. Matsuda, *Nucleic Acids Symp. Ser.*, 2006, **50**, 153–154.
22 Q. Gao, H. Satake, Q. Dai, K. Ono, S. Nishizawa and N. Teramae, *Nucleic Acids Symp. Ser.*, 2005, **49**, 219–220.
23 K. Nakatani, S. Sando, H. Kumasawa, J. Kikuchi and I. Saito, *J. Am. Chem. Soc.*, 2001, **123**, 12650–12657.
24 Y. Li, T. Park, J. K. Quansah and S. C. Zimmerman, *J. Am. Chem. Soc.*, 2011, **133**, 17118–17121.
25 C. Hoock, J. Reichert and M. Schmidtke, *Molecules*, 1999, **4**, 264–271.
26 D. A. Barawkar and K. N. Ganesh, *Nucleic Acids Res.*, 1995, **23**, 159–164.
27 F. S. Mathews and A. Rich, *J. Mol. Biol.*, 1964, **8**, 89–95.
28 K. Hoogsteen, *Acta Crystallogr.*, 1959, **12**, 822–823.
29 A. Dunger, H.-H. Limbach and K. Weisz, *J. Am. Chem. Soc.*, 2000, **122**, 10109–10114.
30 E. N. Nikolova, F. L. Gottardo and H. M. Al-Hashimi, *J. Am. Chem. Soc.*, 2012, **134**, 3667–3670.
31 E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei and H. M. Al-Hashimi, *Nature*, 2011, **470**, 498–502.
32 I. R. Could and P. A. Kollman, *J. Am. Chem. Soc.*, 1994, **116**, 2493–2499.
33 K. I. Trollape, I. R. Gould and I. H. Hillier, *Chem. Phys. Lett.*, 1993, **209**, 113–116.
34 B. Honig and R. Rohs, *Nature*, 2010, **470**, 472–473.
35 M. J. Packer, C. Zonta and C. A. Hunter, *J. Magn. Reson.*, 2003, **162**, 102–112.
36 V. Rudiger and H.-J. Schneider, *Chem.–Eur. J.*, 2000, **6**, 3771–3776.
37 C. A. Hunter and M. J. Packer, *Chem.–Eur. J.*, 1999, **5**, 1891–1897.
38 D. H. Brouwer, S. Alavi and J. A. Ripmeester, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3857–3860.
39 C. D. Chriswell and A. A. Schilt, *Anal. Chem.*, 1975, **47**, 1623–1629.
40 K. A. Connors, *Binding Constants*, John Wiley & Sons, New York, Chichester, Brisbabe, Toroto, Singapore, 1987.
41 J. Sartorius and H.-J. Schneider, *Chem.–Eur. J.*, 1996, **2**, 1446–1452.
42 T. Caruso, A. Capobianco and A. Peluso, *J. Am. Chem. Soc.*, 2007, **129**, 15347–11535.
43 Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1969, **179**, 10–17.
44 E. Mertz, S. Mattei and S. C. Zimmerman, *Org. Lett.*, 2000, **2**, 2931–2934.
45 W. Wang, M. G. M. Purwanto and K. Weisz, *Org. Biomol. Chem.*, 2004, **2**, 1194–1198.
46 Z. Xiao and K. Weisz, *J. Phys. Org. Chem.*, 2007, **20**, 771–777.
47 S. C. Zimmerman and P. Schmitt, *J. Am. Chem. Soc.*, 1995, **117**, 10769–10770.
48 P. S. Corbin, S. C. Zimmerman, P. A. Thiessen, N. A. Hawryluk and T. J. Murray, *J. Am. Chem. Soc.*, 2001, **123**, 10475–10488.
49 S. Chang, S. Huang, J. He, F. Liang, P. Zhang, S. Li, X. Chen, O. Sankey and S. Lindsay, *Nano Lett.*, 2010, **10**, 1070–1075.
50 A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 2011–2033.
51 F. Liang, S. Li, S. Lindsay and P. Zhang, *Chem.–Eur. J.*, 2012, **18**, 5998–6007.