ChemComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: P. GUO, A. Paul, A. Kumar, N. K. Harika, S. Wang, A. A. Farahat, D. W. Boykin and W. D. Wilson, *Chem. Commun.*, 2017, DOI: 10.1039/C7CC06246J.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemcomm

View Article Online
DOI: 10.1039/C7CC06246J

Chem Comm

COMMUNICATION

A Modular Design for Minor Groove Binding and Recognition of Mixed Base Pair Sequences of DNA

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Pu Guo,[a] Ananya Paul,[a] Arvind Kumar,[a] Narinder K. Harika,[a] Siming Wang,[a] Abdelbasset A. Farahat,[a] David W. Boykin,[a] and W. David Wilson*[a]

The design and synthesis of compounds that target mixed, AT/GC, DNA sequences is described. The design concept connects two *N*-methyl-benzimidazole-thiophene single GC recognition units with a flexible linker that lets the compound fit the shape and twist of the DNA minor groove while covering a full turn of the double helix.

Small, synthetic molecules that selectively target biological macromolecules in cells and induce specific responses, such as changes in gene expression, are a central goal of biomolecular compound design and synthesis research as well as therapeutic development.[1-8] While there is significant progress in this area with proteins, there is limited progress in the variety of compound designs to target mixed base pair (bp) DNA sequences. Approaches for compounds that can target the DNA component of DNA-protein complexes, for example, would help remove this block to progress and provide an important step forward in the area.[9-12] In order to effectively target DNA in cells, compounds with appropriate physical properties must be designed along with the development of synthetic approaches that are reasonable in cost and effort.

Heterocyclic-cation, minor-groove binders that can selectively bind to DNA have become a valuable resource in therapeutics, and compounds that successfully act on specific cancers, parasitic organisms and bacteria are now in animal and clinical testing as well as therapeutic use.[1,13-15] These successful agents have provided proof of concept that selective and functional recognition of the DNA minor groove with a set of modules, combined in different ways for different sequences, is possible. The relatively simple agents available, however, primarily recognize only A·T bp sequences and have off-target effects. [16]

Our approach to overcome this limitation is design and preparation of modular minor-groove, sequence-specific compounds that can recognize G·C in addition to A·T bp in longer DNA sequences than most current agents. The design strategy starts with heterocyclic cationic molecular modules that have effective DNA binding affinity in AT sequences, cell and nuclear uptake, reasonable syntheses, relatively low human toxicity and a range of physical properties to increase the chances of finding active compounds for the treatment of a variety of diseases.[11] Using this approach a thiophene-*N*-methylbenzimidazole (thiophene-*N*-MeBI) compound (DB2429, Figure1A) was designed according to the "σ-hole" preorganizing principle for macromolecular targeting.[17] This effect is based on the presence of low-lying C–S σ* orbitals on the thiophene S that possess positive electrostatic potential for interaction with electron donating atoms such as the unsubstituted N in *N*-MeBI.[18,19] Such rational control of the conformation of small molecules is a key concept of optimized molecular design for targeting macromolecules and can lead to significantly improved binding affinity and specificity that is independent of compound–DNA contacts. Compounds with furan or pyridine in place of the thiophene in DB2429 bind with less affinity and selectivity in support of the role of σ-hole interactions with the thiophene.[17]

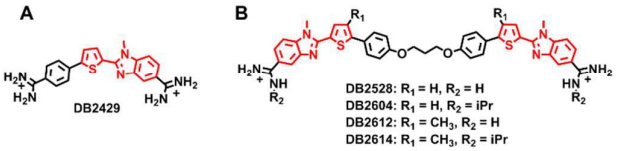
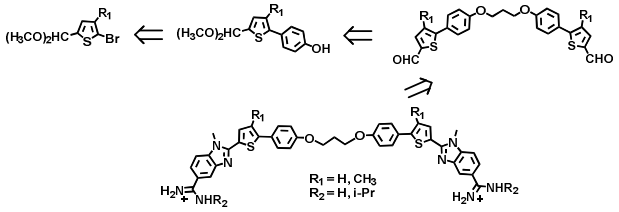


Fig. 1 A) The compound DB2429 with best binding affinity and selectivity for a single G·C bp sequence found in our initial research. B) Chemical structure of new compounds designed to link two GC recognition modules. Red: G·C bp recognition module.

[a] *Department of Chemistry and Center for Diagnostics and Therapeutics Georgia State University, 50 Decatur St Se, Atlanta, GA 30303-3083 (USA) E-mail: wdw@gsu.edu.*

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C7CC06246J



Scheme 1 Retrosynthetic analysis for bis-amidine-*N*-Methylbenzimidazole-thiophene compounds from bis-thiophenecarboxaldehydes that are prepared by linking arylthiophenecarboxaldehydes. Full synthetic details are in Supplementary Materials.

The critical question at this point, which is addressed in this report for the first time, is how to extend relatively simple compounds, such as DB2429, to selectively recognize longer, more complex DNA sequences with additional G·C bp. Our approach combines modular GC recognition units to prepare agents that can selectively recognize two G·C bps in DNA sequences such as $(A/T)_3$-(G/C)-$(A/T)_n$-(G/C)-$(A/T)_3$ where "n" can be 0-5 bps (Figure 1). For the initial test of this approach, a flexible linker, –O-$(CH_2)_3$-O-, for DB2429-type GC recognition molecular modules was used. The design strategy includes enough flexibility to match the shape and curvature of the DNA minor groove and an appropriate length with modules spaced to bind to the GAAAC DNA sequence and a full turn of the DNA helix.

Our first transcription factor (TF)-DNA complex for targeting was the PU.1 TF that is of critical importance in the development of acute myeloid leukemia (AML). Compounds that bind AT or single G·C bp sequences have been successful at targeting the PU.1 promoter and inhibiting AML in cells.[20] To increase the effectiveness of the TF-promoter DNA targeting approach, we wish to target additional TFs, such as the Blimp-1 TF protein that impairs T cell function and is correlated with AML progression.[21] Blimp-1 binds to a consensus GAAAC/G promoter sequence,[22,23] and agents that bind strongly to that sequence have the potential to inhibit the TF and provide new anti-AML activity. Scheme 1 shows a retrosynthetic analysis which allows modular assembly of small molecules designed to bind to DNA sequences with two G·C bps. The target bis-amidino-*N*-methylbenzimidazolethiophenes are prepared from amidino-*N*-methylphenylenediamines by coupling with the appropriate bis-thiophenecarboxaldehydes in the presence of selected oxidizing agents. The required bis-aldehydes are obtained by linking the protected 5-(4-hydroxyphenyl) thiophene-2-carboxaldehdyes with 1, 3-dibromopropane employing standard Williamson Ether synthesis methodology. The protected 5-(4-hydroxyphenyl) thiophene-2-carboxaldehdyes are readily accessible using standard Suzuki coupling reactions between various substituted bromothiophenes and 4-hydroxyphenylboronic acid. A set of related compounds prepared using this approach are shown in Figure 1B. Their synthesis routes are shown in Supporting Information, Scheme S1. The resulting agents are particularly striking because relatively simple changes in chemical groups in AT specific compounds, the addition of specific H-bond accepting groups for G·C bp recognition, converts them into compounds (Figure 1B) that strongly and specifically recognize mixed bps sequences of DNA much more strongly than the original AT sequences. This is the first example where heterocyclic cations linked in this manner have been systematically designed and prepared to recognize G·C bp in complex DNA sequences.

Thermal melting (T_m) provides a rapid, evaluation screen of compounds with selected DNA sequences (Supporting Information, Table S1, S2).[24,25] Binding to DNA gives an increase in the melting temperature that is related to binding affinity. Hairpin DNA oligomers, which have monomolecular melting transitions, were chosen for the screen. DB2429 (Figure 1A) shows our approach of design of single G recognition compounds and for linking these modular units to recognize more complex sequences. DB2528 and analogs are the first test of this concept and T_m results are shown in Table S2. DB2528, which combines two DB2429 units with a flexible –O-$(CH_2)_3$-O- linker (Figure 1B), does not bind well to a single G segment (AAAAGTTTT, ΔT_m = 2℃) but binds with a ΔT_m of 9℃ with the target two G·C bps sequence, GAAAC. It has lower T_m with longer, GAAAAC and GAAAAAC sequences (6℃). It is very encouraging that it shows no significant binding with an all AT sequence (ΔT_m = 1℃), and closer G·C bps (GC, GAC, GAAC), all with ΔT_m <2℃ (Table S1, S2, Figure S1). The ΔT_m for GAAAC and CAAAG are essentially identical but the less symmetrical sequence GAAAG gives a 3℃ lower ΔT_m (Figure S1). Modification of the amidine (DB2604 and 2614) or thiophene (DB2612 and 2614) groups can enhance interactions with the minor groove in cases where self-association limits DNA interactions, but the additions did not enhance binding with DB2528 (Figure 1B, Table S2). The replacement of *N*-MeBI in DB2528 with benzimidazole (BI) gives a more classical minor groove binder that shows strong binding with the AT sequences and weaker binding to the G sequences (not shown). This result clearly confirms the crucial importance of *N*-MeBI to the G·C bp recognition. Circular dichroism (CD) results with DB2528 and the GAAAC sequence show a strong positive induced CD peak at the compound absorption wavelength (Figure S2). This result clearly indicates that the compound is a minor groove binder. Addition of a pure AT sequence to DB2528 results in very little spectral change in agreement with quite weak binding to pure AT sequence.

Biosensor-SPR methods provide an excellent way to quantitatively evaluate the interaction of small molecules with

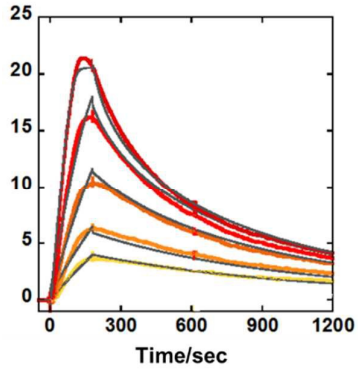


Fig. 2 Representative SPR sensorgrams for DB2528 in the presence of GAAAC DNA, concentrations of DB2528 from bottom to top are 10, 15, 30, 50, and 100 nM. The solid black lines are best-fit values for the global kinetic fitting of the results with a single site function.

View Article Online
DOI: 10.1039/C7CC06246J

immobilized biomolecules.[26] SPR provides sensitive, real-time progress of the binding reaction with binding affinity, kinetics, and stoichiometry of complex formation.[27,28] Based on the T_m results, the interactions of DB2528 were evaluated by SPR with G·C bp sequences (Table S3). As can be seen in Figure 2, DB2528 binds strongly with GAAAC and global kinetics fitting yielded a single binding site with a K_D of 5 nm, a rapid on-rate (k_a = 2.83±0.5×10^6 M^{-1}s^{-1}) and a very slow off-rate (k_d = 1.7±0.4×10^{-2} s^{-1}). The association and dissociation of DB2528 with the shorter GAAG sequence are quite fast with a high K_D = 149 nM. Pure AT and single G sequences were also studied by SPR, and it is encouraging that DB2528 has no detectable binding with either of them. These results indicate excellent selectivity and strong binding for a two G sequence that has a specific distance between the two G·C bps. As a test of the SPR binding results, the K_D was determined by a fluorescence anisotropy titration of the compound with the GAAAC DNA, and the K_D value is in agreement with the results from SPR (Table S3, and Figure S3) and validates the SPR results for binding affinity. Clearly, the combination of thiophene-*N*-MeBI and a flexible –O-(CH$_2$)$_3$-O- linker creates an effective module for strong two G·C bps specific recognition of a full turn of the DNA helix.

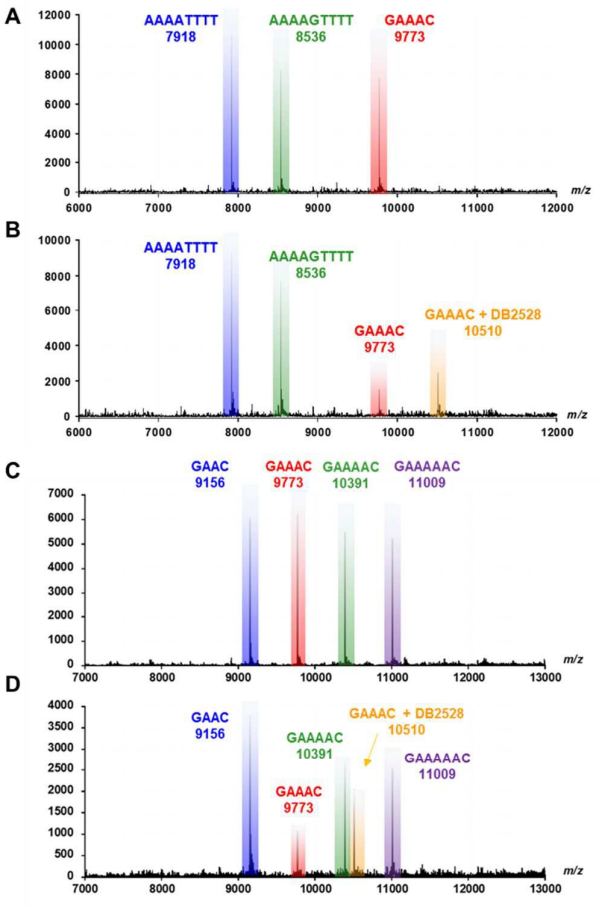


Fig. 3 ESI-MS negative mode spectra of the competition binding of sequences A), B) AAAATTTT, AAAAGTTTT and GAAAC(10 μM each) ; C), D) GAAC, GAAAC, GAAAAC and GAAAAAC (10 μM each); with 40 μM DB2528 in buffer (50 mm ammonium acetate with 10% methanol(v/v), pH 6.8). A), C) The ESI-MS spectra of free DNA mixture. B), D) The ESI-MS spectra of DNA mixture with DB2528. The ESI-MS results shown here are deconvoluted spectra and molecular weights are shown with each peak. Full DNA sequences are in Table S1.

Competition ESI-mass spectrometry (MS) provides a direct analysis of relative binding affinity, stoichiometry and specificity for small molecule DNA complexes.[29,30] The use of several DNA sequences simultaneously mixed with a compound creates a competitive binding environment for direct comparison of DNA interactions. The competition ESI-MS analysis results of DB2528 with DNA sequences AAAATTTT, AAAAGTTTT, and GAAAC, are shown in Figure 3 A, B. The upper plot shows three peaks for the three DNA sequences. On addition of DB2528, the peak for GAAAC (9773) decreases with the simultaneous appearance of a new peak at *m/z*=10510 that is characteristic of a 1:1 GAAAC-DB2528 complex (Figure 3B). There is no appearance of any complex peak with the other DNA sequences. Results for competition ESI-MS analysis of DB2528 with sequences GAAC, GAAAC, GAAAAC, and GAAAAAC are shown in Figure 3 C, D to test the sequence specificity. The upper plot shows four peaks for the four sequences. With the addition of DB2528, only the peak for GAAAC (9773) decreases with the simultaneous appearance of a new peak at *m/z*=10510 for a 1:1 GAAAC-DB2528 complex. There is no significant complex peak with any of the other tested sequences. In agreement with the binding results, the competition ESI-MS results clearly indicate that DB2528 binds with GAAAC very strongly and specifically in a 1:1 complex.

DB2528 was designed to H-bond with two G·C bps separated by three A·T bp and for terminal amidines H-bonds with AT. The compound binds strongly and specifically to this sequence but how does it fit to the minor groove and A·T/G·C bps? To answer this question, the molecular dynamics structures for the DB2528 complex with the minor groove of AGAAACT (Figure 4) were determined with the AMBER 14

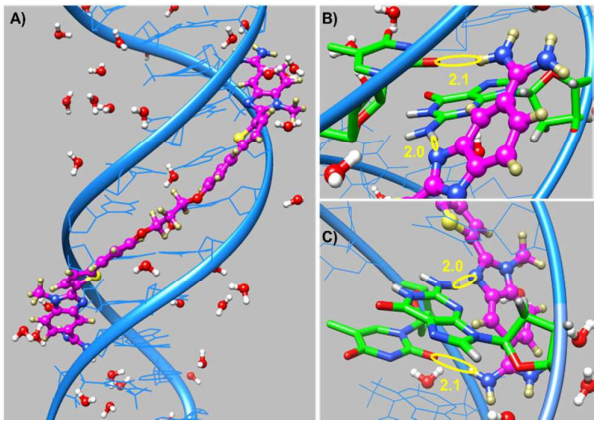


Fig. 4 A) Minor groove view of the DB2528-DNA complex with proximal water molecules. B) Hydrogen bonding the upper G-NH and an opposite strand T-O2 with N of *N*-MeBI and NH of an amidine of DB2528 respectively are shown in yellow circles (in Å). C) View of hydrogen bonding of the lower G-NH and an opposite strand T-O2 with N of *N*-MeBI and NH of an amidine of DB2528, respectively is shown in yellow circles. Upper and lower are with respect to the model in A. DNA is represented in ribbon style (in blue) whereas G and T involved in H-bonding are shown in stick representation (B and C in green). DB2528 is shown in the ball and stick representation (B and C in magenta).

software suite and the ff99 force field. Force constants for

View Article Online
DOI: 10.1039/C7CC06246J

DB2528 were added and the structure determined as previously described.[31-33] The full view in Figure 4A and Figure S4A show DB2528 is able to match the curvature of the DNA minor groove and can cover a full turn of the double helix. The same view is shown in space filling model in Figure S4B and shows the excellent contacts and van der Waals interactions that the compound makes with the minor groove molecular walls. A view of the complex structure at the upper part of the model is shown in Figure 4B. This view shows strong H-bond between the N of *N*-MeBI and the G-NH that points out into the groove. There is also an H-bond between the amidine-NH that points to the floor of the groove and a T=O. The view in Figure 4C shows the same H-bond pattern with the other end of the complex. The amidine-*N*-MeBI-thiophene-phenyl-O- can easily be seen tracking along the minor groove. Again, there are strong H-bonds from the G-NH to the N of *N*-MeBI and from the amidine to a T=O. As can be seen, both amidines are strongly hydrated and this certainly contributes to the energetics of binding and the fit to the groove (Figure 4A and Figure S4). The compound curves around the groove and there is an exact indexing of two *N*-MeBI-G-NH and amidine-A·T bp.

In conclusion, the initial goal of this work was to link previously designed single G·C bp binding modules with a flexible linker for recognition of two G·C bps in the core sequence AGAAACT that has functional significance. With optimized design, synthesis and analysis a linked heterocyclic-diamidine has been obtained that has an excellent match to the DNA minor groove shape and registry of the compound with DNA functional groups. These results show, for the first time, what can be accomplished with modular compounds of this type in selective targeting of complex DNA. The compound has synthesis that is reasonable in cost and time and offers a promising route to develop a broad array of modular agents for control of gene expression.

This work was supported by National Institutes of Health Grant GM111749 to WDW and DWB.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 S. Neidle, D. E. Thurston, *Nat. Rev. Cancer* 2005, **5**(4), 285–296.
2 L. A. Howell, M. Searcey, *ChemBioChem* 2009, **10**(13), 2139–2143.
3 J. Mantaj, P. J. M. Jackson, K. M. Rahman, D. E. Thurston, *Angew. Chem. Int. Ed.* 2017, **56**(2), 462-488.
4 J. B. Chaires, *Biopolymers* 2015, **103**(9), 473–479.
5 G. Savreux-Lenglet, S. Depauw, M. H. David-Cordonnier, *Int. J. Mol. Sci.* 2015, **16**(11):26555-26581.
6 M. P. Barrett, C. G. Gemmell, C. J. Suckling, *Pharmacol. Ther.* 2013, **139**(1):12-23.
7 C. Dardonville , J. J. N. Martínez, *Curr. Med. Chem.* 2017, **24**, DOI: 10.2174/0929867324666170623091522
8 H. Aikawa, A. Yano and K. Nakatani, *Org. Biomol. Chem.* 2017, **15**(6), 1313–1316.
9 J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vázquez, J. L. Mascareñas, *Nat. Commun.* 2013, **4**, 1874.
10 G. He, A. Tolic, J. K. Bashkin, G. M. K. Poon, *Nucleic Acids Res.* 2015, **43**(8), 4322-4331.
11 M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson, G. M. K. Poon, *Nucleic Acids Res.* 2014, **42**(2), 1379–1390.
12 C. R. Millan, F. J. Acosta-Reyes, L. Lagartera, G. U. Ebiloma, L. Lemgruber, J. J. Nué Martínez, N. Saperas, C. Dardonville, H. P. de Koning, J. L. Campos, *Nucleic Acids Res.* 2017, DOI 10.1093/nar/gkx521.
13 J. S. Kang, P. B. Dervan, *Q. Rev. Biophys.* 2015, **48**(4), 453–464.
14 X. Cai, P. J. Gray, D. D. Von Hoff, *Cancer Treat. Rev.* 2009, **35**(5), 437–450.
15 C. Lin, R. I. Mathad, Z. Zhang, N. Sidell, D. Yang, *Nucleic Acids Res.* 2014, **42**(9), 6012–6024.
16 R. Nanjunda, W. D. Wilson, in *Curr. Protoc. Nucleic Acid Chem.*, 2012, Chapter 8, Unit8.8.
17 P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin, W. D. Wilson, *Chem. Eur. J.* 2016, **22**(43), 15404–15412.
18 B. R. Beno, K.-S. Yeung, M. D. Bartberger, L. D. Pennington, N. A. Meanwell, *J. Med. Chem.* 2015, **58**(11), 4383–4438.
19 D. Tilly, F. Chevallier, F. Mongin, *Synthesis* 2016, **48**(2), 184–199.
20 I. Antony-Debré, A. Paul, J. Leite, K. Mitchell, H. M. Kim, L. A. Carvajal, T. I. Todorova, K. Huang, A. Kumar, A. A. Farahat, B. Bartholdy, S.-R. Narayanagari, J. Chen, A. Ambesi-Impiombato, A. A. Ferrando, I. Mantzaris, E. Gavathiotis, A. Verma, B. Will, D. W. Boykin, W. D. Wilson, G. M. K. Poon, U. Steidl, *J. Clin. Invest.* 2017, In Press, Manuscript No.: 92504-JCI-RG-1
21 L. Zhu, Y. Kong, J. Zhang, D. F. Claxton, W. C. Ehmann, W. B. Rybka, N. D. Palmisiano, M. Wang, B. Jia, M. Bayerl, T. D. Schell, R. J. Hohl, H. Zeng, H. Zheng, *J. Hematol. Oncol.* 2017, **10**(124), 1-13.
22 T. C. Kuo, K. L. Calame, *J. Immunol.* 2004, **173**(9), 5556–5563.
23 S. Deng, T. Yuan, X. Cheng, R. Jian, J. Jiang, *Mol. Biol. Rep.* 2010, **37**(8), 3747–3755.
24 M. Munde, A. Kumar, P. Peixoto, S. Depauw, M. A. Ismail, A. A. Farahat, A. Paul, M. V. Say, M.-H. David-Cordonnier, D. W. Boykin, W. D. Wilson, *Biochemistry* 2014, **53**(7), 1218–1227.
25 X. Shi, J. B. Chaires, *Nucleic Acids Res.* 2006, **34**(2), e14.
26 R. Nanjunda, M. Munde, Y. Liu, W. Wilson, in (Ed.: Y. Wanunu), Methods for Studying Nucleic Acid/Drug Interactions, 2011, pp. 92–122.
27 Y. Liu, A. Kumar, S. Depauw, R. Nhili, M.-H. David-Cordonnier, M. P. Lee, M. A. Ismail, A. A. Farahat, M. Say, S. Chackal-Catoen, A. Batista-Parra, S. Neidle, D. W. Boykin, W. D. Wilson, *J. Am. Chem. Soc.* 2011, **133**(26), 10171–10183.
28 R. D. Taylor, S. Asamitsu, T. Takenaka, M. Yamamoto, K. Hashiya, Y. Kawamoto, T. Bando, H. Nagase, H. Sugiyama, *Chem. Eur. J.* 2014, **20**(5), 1310–1317.
29 S. Laughlin, S. Wang, A. Kumar, A. A. Farahat, D. W. Boykin, W. D. Wilson, *Chem. Eur. J.* 2015, **21**(14), 5528–5539.
30 A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M.-H. David-Cordonnier, D. W. Boykin, W. D. Wilson, *Bioorg. Med. Chem. Lett.* 2015, **25**(21), 4927–4932.
31 P. Athri, W. D. Wilson, *J. Am. Chem. Soc.* 2009, **131**(22), 7618–7625.
32 S. Laughlin-Toth, E. K. Carter, I. Ivanov, W. D. Wilson, *Nucleic Acids Res.* 2017, **45**(3), 1297–1306.
33 N. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza, J. Šponer, *J. Am. Chem. Soc.* 2003, **125**(7), 1759–1769.