Supporting Information

Engineered Modular Heterocyclic-Diamidines for Sequence-Specific Recognition of Mixed AT/GC Base Pairs at the DNA Minor Groove

Pu Guo,[a] Abdelbasset A. Farahat,[a,b] Ananya Paul,[a] David W. Boykin,[a] and W. David Wilson*[a]

[a]*Department of Chemistry and Center for Diagnostics and Therapeutics Georgia State University, 50 Decatur St SE, Atlanta, GA 30303, USA.*

[b]*Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt*

Address correspondence to this author

***W. David Wilson Tel: 404-413-5503; Fax: 404-413-5505; Email: wdw@gsu.edu

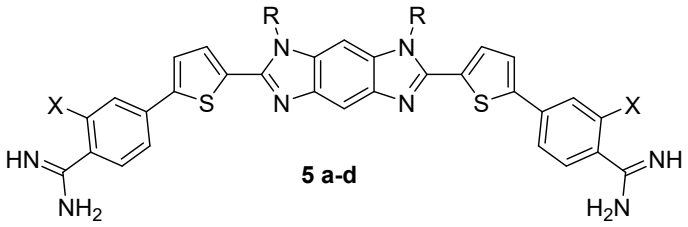
EXPERIMENTAL METHODS

Synthesis

Synthesis:

The melting points of newly synthesized compounds were determined by a Mel-Temp 3.0 melting point apparatus and are uncorrected. NMR spectra and mass spectra for compounds were recorded as previously reported.[1]. If the compounds reported as salts contain a fraction of water and/or solvents, these fractions are seen in HNMR spectra. Elemental analysis of the compounds were performed by Atlantic Microlab Inc., Norcross, GA. The synthesis and of compounds 2 through 4 will be published somewhere else.

.Synthesis of the diamidines (5a-f)

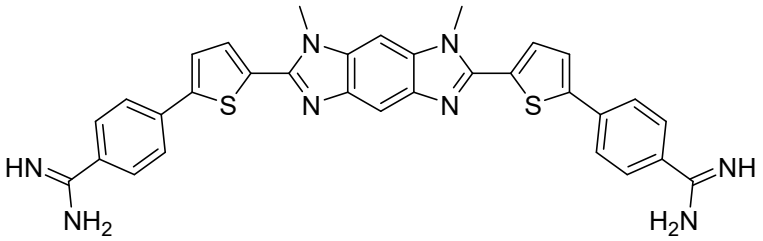


5 a-d

5 e,f

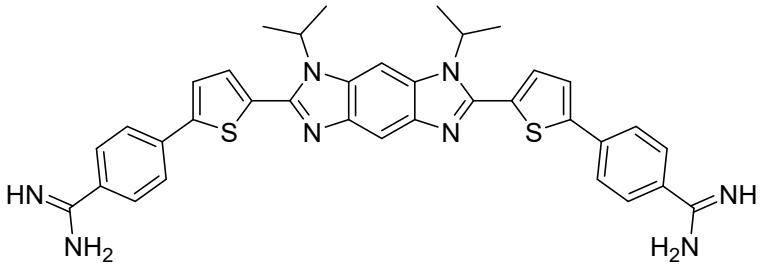
The above bis-nitriles **4a-d and 4e-f (**1 mmol) were suspended in freshly distilled THF (10 mL), and treated with a 1M $LiN(TMS)_2$ in THF solution (6 mL, 6.0 mmol) and the mixture was stirred for 24 h at room temperature. The reaction mixture was cooled to 0 ^oC and HCl saturated ethanol (3 mL) was carefully added. The mixture was stirred for 2 h, ether was added and the resultant solid was collected by filtration. The diamidine was purified by neutralization with 1M sodium hydroxide solution followed by filtration of the resultant solid, washed with water and dried. The free base was stirred with ethanolic HCl for 24 h, acetone was added, and the solid that formed was filtered and dried under vacuum at 100 ^oC for 24 h.

4,4'-((1,7-Dimethyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-5,2-diyl))dibenzimidamide (5a, DB2838).



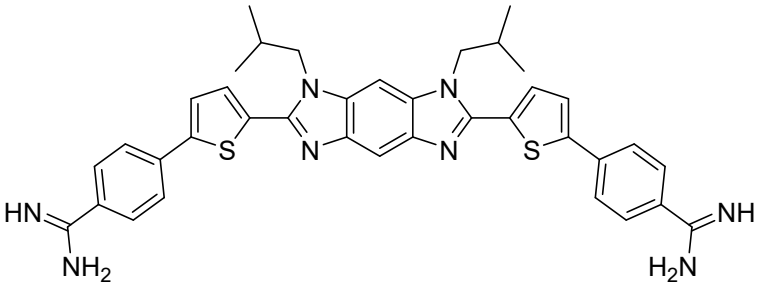
Yellow solid (0.36 gm, 49 %), mp > 300 °C. ^{1}HNMR (DMSO-d_6): δ 9.44 (s, 4 H), 9.10 (s, 4 H), 8.19 (br s, 1 H), 8.08 (m, 6 H), 8.01 (d, *J* = 3.6 Hz, 2H), 7.95 (m, 5H), 4.22 (s, 6H); ESI-HRMS: m/z calculated for $C_{32}H_{28}N_8S_2$: 294.0934, found: 294.0944 (Double charged amidine base M^+ + 2). Anal. Calcd. For $C_{32}H_{26}N_8S_2$. 3HCl. $3H_2O$: C, 51.32; H, 4.71; N, 14.97. Found: C, 51.57; H, 4.83; N, 14.66; Extinction Coefficients: 5.0 x 10^4 L/(mole•cm) at 407 nm

4,4'-((1,7-Diisopropyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-5,2-diyl))dibenzimidamide (5b, DB2831).



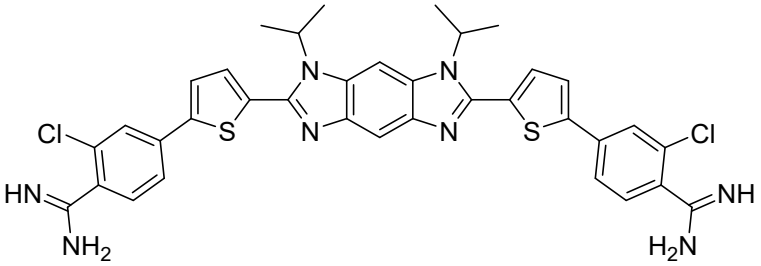
Yellow solid (0.45 gm, 59 %), mp > 300 °C. ^{1}HNMR (DMSO-d_6): δ 9.48 (s, 4 H), 9.22 (s, 4 H), 8.04 (d, *J* = 8 Hz, 4H), 7.96 (d, *J* = 8 Hz, 4H), 7.92 (m, 4H), 7.69 (d, *J* = 4 Hz, 2H), 5.29 (t, *J* = 6.4 Hz, 2H), 1.80 (d, *J* = 6.8 Hz, 12H); ESI-HRMS: m/z calculated for $C_{36}H_{35}N_8S_2$: 643.2426, found: 643.2402 (amidine base M^+ + 1). Anal. Calcd. For $C_{36}H_{34}N_8S_2$. 3HCl. $1.5H_2O$: C, 55.58; H, 5.18; N, 14.41. Found: C, 55.83; H, 5.07; N, 14.31; Extinction Coefficients: 6.5 x 10^4 L/(mole•cm) at 390 nm.

4,4'-((1,7-Diisobutyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-5,2-diyl))dibenzimidamide (5c, DB2833).



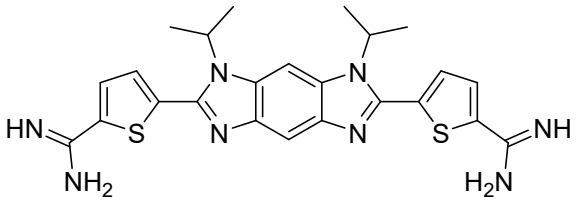
Yellow solid (0.56 gm, 67 %), mp > 300 °C. ^{1}HNMR (DMSO-d_6): δ 9.43 (s, 4 H), 9.25 (s, 4 H), 8.03 (d, *J* = 8.4 Hz, 4H), 7.98 (d, *J* = 8.4 Hz, 4H), 7.91 (m, 6H), 4.53 (d, *J* = 7.4 Hz, 4H), 2.27 (t, *J* = 6.8 Hz, 2H), 0.95 (d, *J* = 6.8 Hz, 12H); ESI-HRMS: m/z calculated for $C_{38}H_{39}N_8S_2$: 671.2739, found: 671.2729 (amidine base M^+ + 1). Anal. Calcd. For $C_{38}H_{38}N_8S_2$. 3HCl. $3H_2O$: C, 54.79; H, 5.69; N, 13.46. Found: C, 54.47; H, 5.58; N, 13.53; Extinction Coefficients: 6.1 x 10^4 L/(mole•cm) at 410 nm.

4,4'-((1,7-Diisopropyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-5,2-diyl))bis(2-chlorobenzimidamide) (5d, DB2836).



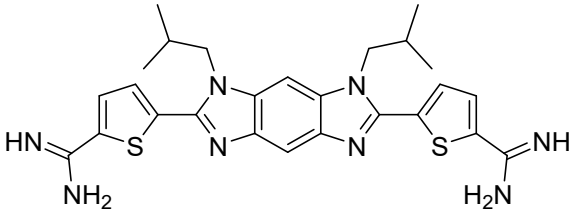
Yellow solid (0.58 gm, 69 %), mp > 300 °C. ^{1}HNMR (DMSO-d_6): δ 9.61 (s, 4 H), 9.54 (s, 4 H), 8.17 (m, 2H), 8.05 (br s, 1H), 8.01 (m, 3H), 7.95 (d, *J* = 8 Hz, 2H), 7.78 (m, 4H), 5.29 (t, *J* = 6.8 Hz, 2H), 1.80 (d, *J* = 7.2 Hz, 12H); ESI-HRMS: m/z calculated for $C_{36}H_{33}Cl_2N_8S_2$: 711.1641, found: 711.1665 (amidine base M^+ + 1). Anal. Calcd. For $C_{36}H_{32}Cl_2N_8S_2$. 3HCl. $2H_2O$: C, 50.57; H, 4.60; N, 13.11. Found: C, 50.44; H, 4.38; N, 13.39; Extinction Coefficients: 3.3 x 10^4 L/(mole•cm) at 380 nm.

5,5'-(1,7-Diisopropyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-2-carboximidamide) (5e, 2834).

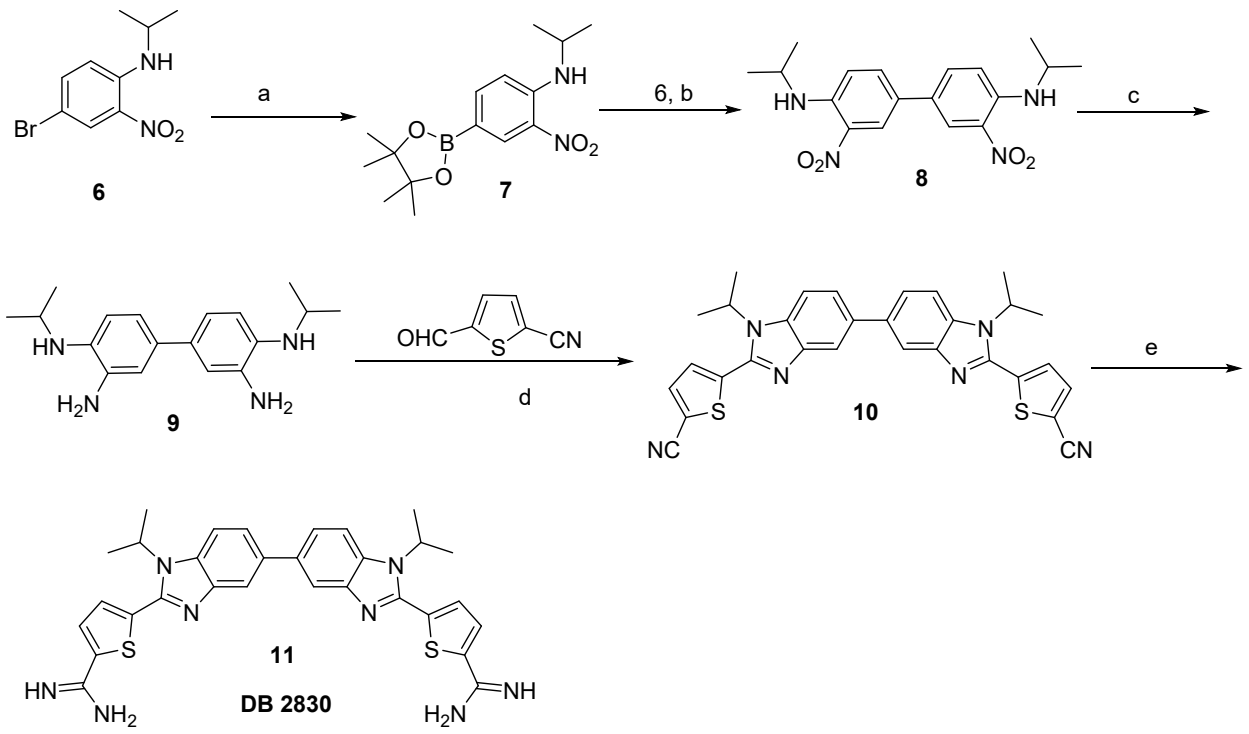


Yellow solid (0.46 gm, 76 %), mp > 300 ^{o}C. ^{1}HNMR (DMSO-d_6): δ 9.78 (s, 4 H), 9.46 (s, 4 H), 8.32 (d, *J* = 4 Hz, 2 H), 8.16 (br s, 1 H), 8.10 (br s, 1 H), 7.91 (d, *J* = 4 Hz, 2 H), 5.22 (m, 2 H), 1.78 (d, *J* = 6.8 Hz, 12 H); ESI-HRMS: m/z calculated for $C_{24}H_{27}N_8S_2$: 491.1800, found: 491.1789 (amidine base M^+ + 1). Anal. Calcd. For $C_{24}H_{26}N_8S_2$. 2HCl. $3H_2O$: C, 46.74; H, 5.56; N, 18.18. Found: C, 46.97; H, 5.26; N, 17.89; Extinction Coefficients: 6.0 x 10^4 L/(mole•cm) at 360 nm.

5,5'-(1,7-Diisobutyl-1,7-dihydrobenzo[1,2-d:4,5-d']diimidazole-2,6-diyl)bis(thiophene-2-carboximidamide) (5f, 2835).



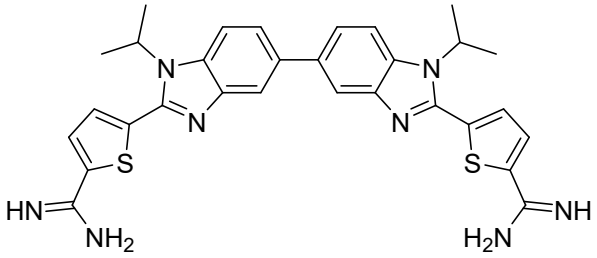
Yellow solid (0.38 gm, 61 %), mp > 300 ^{o}C. ^{1}HNMR (DMSO-d_6): δ 9.69 (s, 4 H), 9.38 (s, 4 H), 8.26(d, *J* = 4 Hz, 2 H), 8.23 (br s, 1 H), 8.03 (d, *J* = 4 Hz, 2 H), 7.98 (br s, 1 H), 4.54 (d, *J* = 7.2 Hz, 4 H), 2.17 (m, 2 H), 0.90 (d, *J* = 6.8 Hz, 12 H); ESI-HRMS: m/z calculated for $C_{26}H_{31}N_8S_2$: 519.2113, found: 519.2117 (amidine base M^+ + 1). Anal. Calcd. For $C_{26}H_{30}N_8S_2$. 2HCl. $2.25H_2O$: C, 49.47; H, 5.83; N, 17.76. Found: C, 49.19; H, 5.74; N, 17.17.41; ; Extinction Coefficients: 4.1 x 10^4 L/(mole•cm) at 380 nm.



Scheme S1: Reagents and conditions: a) Bis(pinacolato)diboron, $PdCl_2(PPh_3)_2$, K acetate, Dioxane; b) $Pd(PPh_3)_4$, K_2CO_3/H_2O, dioxane, reflux; c) $NaBH_4$, Pd(C), CH_2Cl_2,MeOH; d) $Na_2S_2O_5$/ DMSO, 130 ^oC; e) i- $LiN(TMS)_2$/THF, ii- HCl/EtOH

The synthesis of compounds 6 through 10 will be published somewhere else.

5,5'-(1,1'-Diisopropyl-1H,1'H-[5,5'-bibenzo[d]imidazole]-2,2'-diyl)bis(thiophene-2-carboximidamide) (11).



This compound was prepared using $LiN(TMS)_2$ method mentioned before starting from the bis-cyano compound **10.**

Yellow solid (68 %), mp > 300 ^oC. ^{1}HNMR (DMSO-d_6): δ 9.78 (s, 4 H), 9.46 (s, 4 H), 8.29 (d, *J* = 4 Hz, 2 H), 8.06 (m, 4 H), 7.88 (d, *J* = 4 Hz, 2 H), 7.75 (m, 2 H), 5.13 (t, *J* = 6.8 Hz, 2 H), 1.72 (d, *J* = 6.8 Hz, 12 H); ESI-HRMS: m/z calculated for $C_{30}H_{31}N_8S_2$: 567.2113,

found: 567.2128 (amidine base M^+ + 1). Anal. Calcd. For $C_{30}H_{30}N_8S_2$. 3HCl. $1H_2O$: C, 52.01; H, 5.09; N, 16.18. Found: C, 52.24; H, 5.29; N, 15.86.

Biophysical Experimental

Materials and sample preparation

In the DNA thermal melting (T_m), biacore-SPR and circular dichroism (CD) experiments, hairpin oligomer sequences were used (Figure 1D). All DNA oligomers were obtained from IDT, Coralville, IA, with reverse-phase HPLC purification and mass spectrometry characterization. The UV-vis thermal melting (T_m), circular dichroism (CD) and fluorescence spectroscopy experiments were carried out in 50 mM Tris-HCl, 100 mM to 400 mM NaCl, 1 mM EDTA, pH 7.4 (TNE 100). The buffer for biosensor-surface plasmon resonance (SPR) experiments were made in double distalled, filtered, degassed TNE 100 with 0.05% (v/v) surfactant P20.

UV-vis thermal melting for native hairpin DNA and complexes:

The thermal melting experiments were performed on a Cary 300 Bio UV-vis spectrophotometer (Cary). The hairpin DNA concentration of was 3 μM in TNE 100 using 1 cm quartz cuvettes. The solutions of DNA and ligands were tested with a ratio of 2:1 [ligand] / [DNA]. All samples were heated to 95 °C and cooled down to 25 °C slowly before each experiment. The spectrophotometer method and T_m determination were done set up as previously described procedure.[2]

Biosensor-Surface Plasmon Resonance (SPR) Experiments to Determine Ligand-DNA Binding Constant:

Biacore SPR measurements were performed with streptavidin-derivatized (SA) CM5 sensor chips by using four-channel Biacore T200 optical biosensor system (Cytiva, Global life science solutions USA LLC). The procedure of the preparation of SA chips and immobilization of biotinylated-DNAs (**AAAGTTT**: 5'-biotin-CC**AAAGTTT**GCTCTCAAACTTTGG-3'; **AAAGGTTT**: 5'-biotin-CC**AAAGGTTT**GCTCTCAAACCTTTGG-3'; **AAAGTGTTT**: 5'-biotin-CC**AAAGTGTTT**GCCTCTGCAAACACTTTGG-3') on-chip surface (cell 2-4 respectively) were described before. Ligands solutions were prepared with degassed and filtered 50 mM Tris-HCl-buffer with varied NaCl concentrations (100 mM NaCl to 400 mM NaCl) pH 7.4 with 0.05% (v/v) surfactant P20. A series of ligand concentrations (2 nM to 500 nM) were injected over the DNA-immobilized sensor chip with the flow rate of 100 µL/min for 180 s, followed by buffer flow for ligand dissociation (600−1800 s). After each sample run, the sensor chip surface was regenerated by injecting acidic 10 mM glycine solution (pH 2.5) for 30 s followed by several buffer injections to establish a stable baseline for the subsequent cycles. The data analysis was followed by a previously described method where reference response from the blank cell (cell 1) was subtracted from the response in each flow cell containing DNA to give a signal (RU_{obs}, response units) directly related to the amount of bound ligand. The expected maximum response (RU_{max}) per bound ligand in the steady-state region was determined from the molecular weight of the DNA, the ligand molecular weight, and the refractive index gradient ratio of the ligand and DNA. KaleidaGraph 4.0 software was used to plot RU_{obs} versus free ligand concentration (C_{free}). The equilibrium binding constants (K_1) were determined with a one-site binding model. In this model, r =

(RU_{obs}/RU_{max}) represents the moles of bound compound/mol of DNA hairpin duplex, and K_1 is macroscopic binding constant.

$$\boldsymbol{r = K_1 * C_{free} / 1 + K_1 * C_{free}} \qquad (1)$$

To evaluate the stoichiometry of ligand-DNA complex, RU_{max} in the equation was used as a fitting parameter and compared that value with predicted maximal response per bound ligand. Kinetic analysis was achieved by globally fitting the ligand-binding sensorgrams by using a standard 1:1 kinetic model with incorporated mass transport-limited binding parameters as described previously.[3]

Circular Dichroism (CD) Experiments to Determine the Ligand-DNA binding Mode:

Circular dichroism experiments were performed on a Jasco J-1500 spectrometer as previously described.[2] The hairpin DNA sequence **AAAGGTTT**: 5'-CC**AAAGGTTT**GCTCTCAAACCTTTGG-3' (5 μM), Figure 1D, in TNE 100 was used in this experiment

Fluorescence Spectroscopy Binding Determinations

Fluorescence spectra were recorded on a Cary Eclipse Spectrophotometer, with excitation and emission slit width as determined depending on the concentrations of ligands. The free compound solutions at different concentrations were prepared in an appropriate buffer in TNE 100, and DNA (**AAAGTTT**: 5'-CC**AAAGTTT**GCTCTCAAACTTTGG-3'; **AAAGGTTT**: 5'-CC**AAAGGTTT**GCTCTCAAACCTTTGG-3') aliquots were added from a concentrated

stock. All titration spectra were collected after allowing an incubation time of 10 min. DB2831 was excited at 390 nm based on molecular absorbance from UV-vis spectroscopy. Emission spectra of these compounds were monitored from 200 nm to 650 nm wavelength range. All the fluorescence titrations were fitted to equation 2 as previously described.[2]

$$I = I_f + (I_b - I_f) \times \{L_T + K_D + R_T - [(L_T + K_D + R_T)^2 - 4L_T R_T]^{1/2}\}/2L_T \quad (2)$$

where I represents the measured intensity, I_f and I_b are the intensity of free ligand and fully bound ligand. L_T and R_T are the concentrations of total ligand and receptor.

Competition Electrospray Ionization Mass Spectrometry (ESI-MS) to Determine the Sequence selectivity of the Ligand:

Competition Electrospray Ionization Mass Spectrometry (ESI-MS) analyses were performed as previouly descrived method on a Waters Q-TOF micro–Mass [4] Spectrometer (Waters Corporate, Milford, MA) equipped with an electrospray ionization source (ESI) in a negative ion mode. DNA sequences **AAATTT** (5'-CC**AAATTT**GCCTCTCGAAATTTGG-3'); **AAAGTTT** (5'-GCC**AAAGTTT**GCCTCTGCAAACTTTGGC-3'); **AAAGCTTT** (5'-CC**AAAGCTTT**GCTCTCAAAGCTTTGG-3'); **AAAGGTTT** (5'-GCC**AAAGGTTT**GCCTCTGCAAACCTTTGGC-3'); and **AAAGTGTTT** (5'-GCC**AAAGTGTTT**GCCTCTGCAAACACTTTGGC-3'), for ESI-MS experiments were purified by dialyzing in 50 mM ammonium acetate buffer (pH 6.7) at 4 ºC with 3x buffer exchange. The competitive experiments were done by mixing a ligand and DNAs with different sequences at different ratios.

Molecular Curvature Determination: Method Description:

The method developed here for relative curvature determination is illustrated and calibrated with two AT specific binding compounds, DB75, which has a moderate binding constant (K_A = 1 × 10^7 M^{-1}) in our standard conditions: TNE 100 and DB818, a similar compound with a significantly stronger binding constant (K_A = 11 × 10^7 M^{-1}) (Figure S2). Both compounds have been crystallized with the AATT binding site *ds*[CGCGAATTCGCG]$_2$ and the compound structures are is shown in Figure S2.[5] A third compound also used as a calibration control is DB2457, one of the first GC recognition compounds that we designed. It has a binding constant of (K_A = 25 × 10^7 M^{-1}) with the AAGTT sequence in a hairpin duplex (Figure 1A). [2]

To determine comparative molecular curvature values for the compounds, A reference circle is first defined that passes through both amidine carbons (Figure 5). These two points are then connected with the reference circle that has a radius that allows them to pass as closely as possible through the center of each molecular unit of the entire molecule and the two amidine carbons. This is illustrated with the three control compounds in Figure S2. Their angles of curvature are DB75, 134°; DB818, 142°; and DB2457, 135° with an error of +/- 2° for all. These values are in agreement with the relative binding constants of DB75 and DB818 for the same sequence. As the angle of curvature approaches low values, the compounds are quite curved, while higher numbers approaching 180 are more linear. A value of around 145 is ideal for recognizing the minor groove curvature based on binding constants versus relative curvature values. For the compounds in this paper and using the same procedure, we find angles of curvature of 136° for DB2831 and 113° for DB2830.

Molecular Dynamics (MD) Simulations

Structure optimization of DB2831 was performed by using DFT/B3LYP theory with the 6-31+G* basis set in Gaussian 09 (Gaussian, Inc., 2009, Wallingford, CT) with Gauss-view 5.09.[6] Partial charges were derived using the RESP fitting method (restrained electrostatic potential).[7] AMBER 16 (Assisted Model Building with Energy Refinement) software suite was used to perform molecular dynamics (MD) simulations. [8] Canonical *B*-form ds[5'-CCAAAGGTTTGG-3'][5'-CCAAACCTTTGG-3'] DNA was built in the Nucleic Acid Builder (NAB) tool in AMBER. The ANTECHAMBER Tools were used to create *LEaP* Input topology files for the ligand DB2831 which was used in the AMBER simulation programs.[8] Specific atom types assigned for the DB2831 molecule were adapted from the *ff99* force field. Most of the force field parameters for DB2831 molecule were derived from the existing set of bonds, angles, and dihedrals for similar atom types in *parm99* and *GAFF* force fields.[9] Some dihedral angle parameters were obtained from previously reported parametrized data.[9] The molecular structure with specific atom types used for the DB2831 molecule is shown in Figure S4. Parameters of DB2831 in .frcmod file are listed in Table S1.

The AutoDock Vina program was used to dock DB2831 in the minor groove of DNA to obtain the initial structure for the DB2831-DNA complex.[10] The AMBER16 package was used to equilibrate the DB2831−DNA complex system using OL15 force field modifications for DNA. MD simulations were performed in explicit solvation conditions where DNA and DNA-ligand complexes were solvated in a 72 Å x 72 Å x 72 Å truncated octahedron box filled with approximately 5000 TIP3P water [11] molecules by using *TLeap* [12] program in AMBER16. To reach physiological salt concentration, 150

mM Na+ and Cl-, were added to the systems.Na+ and Cl-, an appropriate number of ions were added to the systems. This is a higher salt concentration than required to achieve electrical neutrality but is more biologically relevant. The particle mesh Ewald (PME) [13] method was used to handle Coulombic interactions, and a 10 Å cutoff was applied on all van der Waals interactions. The MD simulations were performed by using Sander module with the SHAKE [14] algorithm applied to constrain all bonds involving hydrogen atoms with an integration time step of 2 fs. In the multistage equilibration protocol, the system was relaxed with 500 steps of steepest-descent energy minimization. The temperature of the system was increased from 0 K to 310 K for over 10 ps under constant-volume conditions. In the final step, the production runs on the system was subsequently performed for 600 ns under NPT (constant-pressure) conditions on the PMEMD CUDA module of AMBER16.[8,12] Trajectories were post-processed using the CPPTRAJ module of AMBERTOOLS16 [8,12] to produce 25000 snapshots for analysis and visualization in UCSF Chimera visualization software.[15] The steepest descent algorithm is useful for quickly removing the largest strains in the system, but it also converges slowly when close to a minimum.

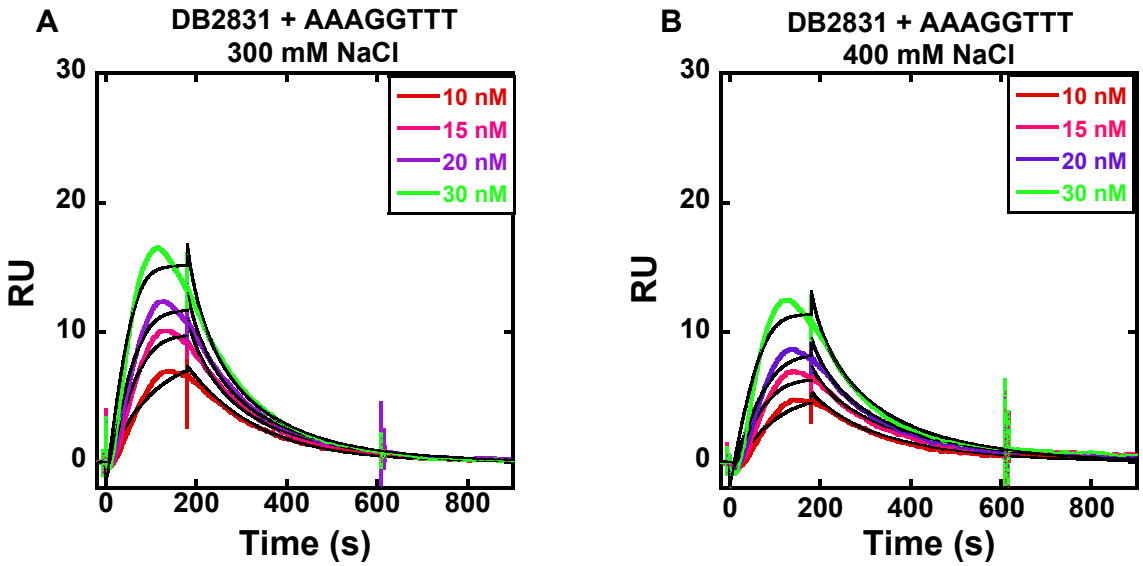


Figure S1: (A) and (B) SPR sensorgrams (color) and kinetic fits (black overlays) for DB2831 with the AAAGGTTT DNA sequence at 300 mM and 400 mM NaCl concentrations.

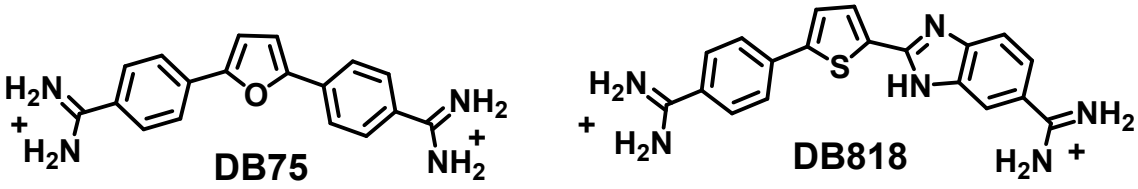


Figure S2: Chemical structure of DB75 and DB818

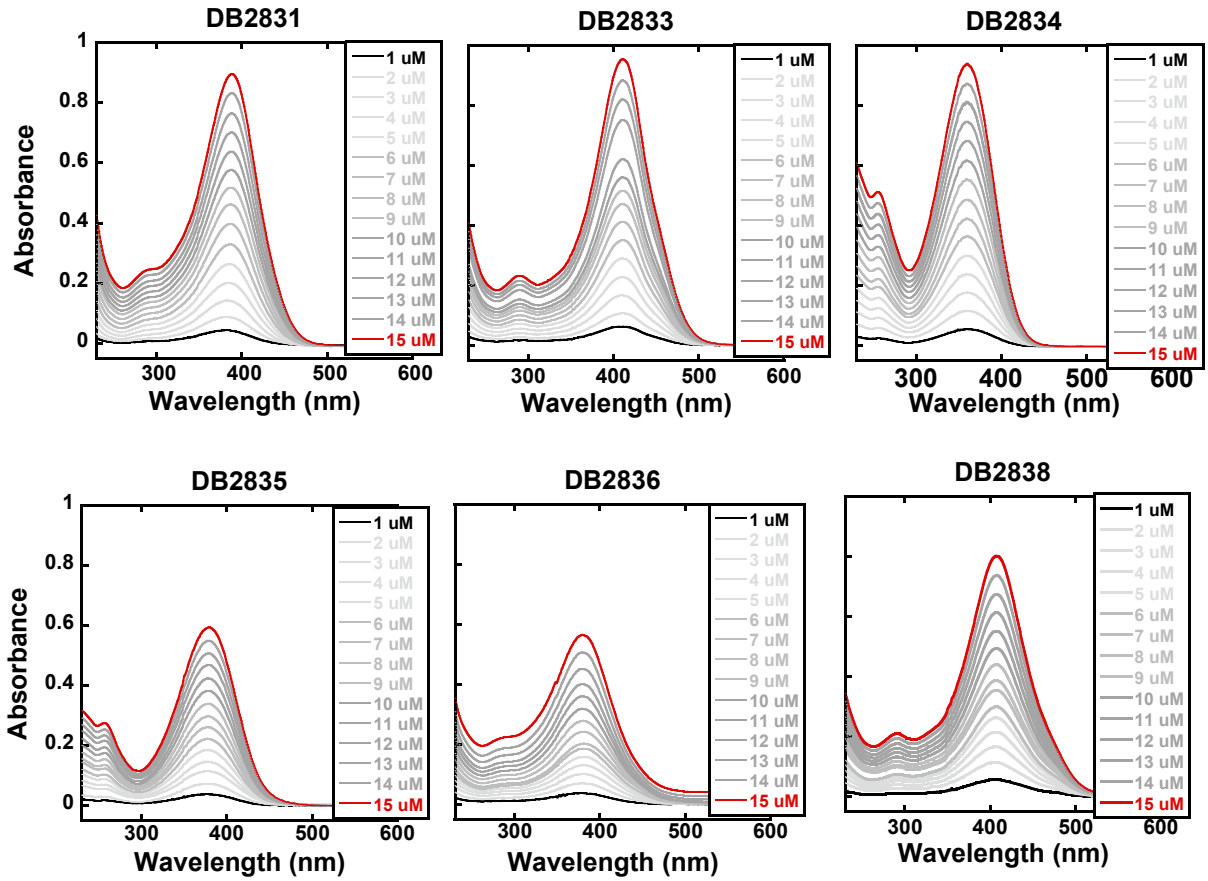


Figure S3. Absorption spectra of two adjacent GC bps binding diamidine compounds in water.

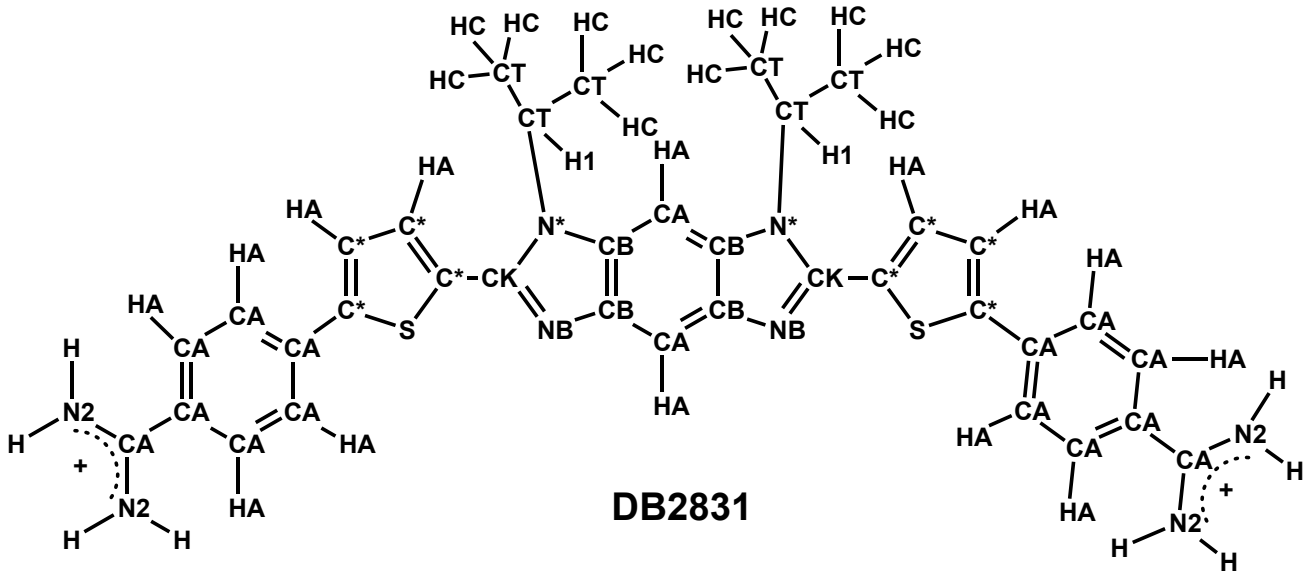


Figure S4. Molecular structure with specific atom types used for the DB2831 molecule

Table S1. .Frcmod file of the DB2831 molecule

MASS

N2 14.01	0.530	parm99
CA 12.01	0.360	parm99
CB 12.01	0.360	parm99
C* 12.01	0.360	gaff SP2 carbon at non-pure aromatic system
CK 12.01	0.360	parm99
CT 12.01	0.878	parm99
HA 1.008	0.167	parm99
H 1.008	0.161	parm99
HC 1.008	0.135	parm99
H1 1.008	0.135	parm99
N* 14.01	0.530	parm99
NB 14.01	0.530	parm99
S 32.06	2.900	gaff

BOND

CA-CA	469.0	1.400	parm99
CA-CB	469.0	1.404	parm99
CB-CB	520.0	1.370	parm99
CB-N*	436.0	1.374	parm99
CB-NB	414.0	1.391	parm99
CT-N*	337.0	1.475	parm99
CT-CT	310.0	1.526	parm99
CA-HA	367.0	1.080	parm99

```
CT-H1  340.0    1.090       parm99
CT-HC  340.0    1.090       parm99
CK-N*  440.0    1.371       parm99
CK-NB  529.0    1.304       parm99
CK-C*  418.3    1.4290      SOURCE1   740 0.0069 cc-cc gaff similar to gaussian bond
C*-C*  418.3    1.4290      SOURCE1    740          0.0069  gaff
C*-HA  347.2    1.0850      SOURCE3    740          0.0039  gaff
CA-C*  411.7    1.4340      SOURCE1     80 0.0000  gaff
CA-N2  481.0    1.340       parm99
S -C*  279.3     1.7370     SOURCE3      52          0.0194  gaff
H -N2  434.0    1.010       parm99   for plain unmethylated bases ADE,CYT,GUA,ARG

ANGLE
CA-CA-CA    63.0      120.00    parm99
CA-CA-HA    50.0      120.00    parm99
CA-CB-CB    63.0      117.30    parm99
CB-CA-HA    50.0      120.00    parm99
CA-CB-NB    70.0      132.40    parm99
CA-CB-N*    70.0      132.40    parm99 CA-CB-NB parm99
CB-CB-N*    70.0      106.20    parm99
CB-CB-NB    70.0      110.40    parm99
CT-CT-CT    40.0      109.50    parm99
CT-CT-H1    50.0      109.50    parm99
CB-N*-CT    70.0      125.80    parm99
CT-CT-N*    50.0      109.50    parm99
N*-CK-NB    70.0      113.90    parm99
CB-NB-CK    70.0      103.80    parm99
CB-N*-CK    70.0      105.40    parm99
C*-CK-NB    67.53     121.69    CORR       105 cc-cc-nc GAFF
C*-CK-N*    67.53     121.69    CORR       105 cc-cc-nc GAFF
C*-C*-CK    66.24     121.77    CORR  c2-cc-cc GAFF
C*-C*-C*    67.880    110.700   SOURCE3          54 3.4091 gaff
CK-N*-CT    70.0      128.80    parm99
CB-N*-CT    70.0      125.80    parm99
H1-CT-N*    50.0      109.50    parm99
HC-CT-HC     35.0       109.50    parm99
HC-CT-CT     50.0       109.50    parm99
C*-C*-HA     47.14      120.86     CORR       1751 cc-cc-ha GAFF
C*-S -C*     41.930     89.910          SOURCE3          11   2.2164 cc-ss-cc gaff
CA-C*-S      78.690     120.980         SOURCE4          28   1.8865 ca-cc-ss gaff
S -C*-C*     80.780     115.020         SOURCE3           2   0.0000 cc-cc-ss gaff
CA-C*-C*     67.660     111.040         SOURCE3           9   7.9455 ca-cc-cc gaff
CA-CA-N2     70.0       119.99    parm99, CM-CA-N2, Gaussian-angle
N2-CA-N2     70.0       120.00    parm99
H -N2-H      35.0       120.00    parm99
CA-N2-H      50.0       120.00    parm99
```

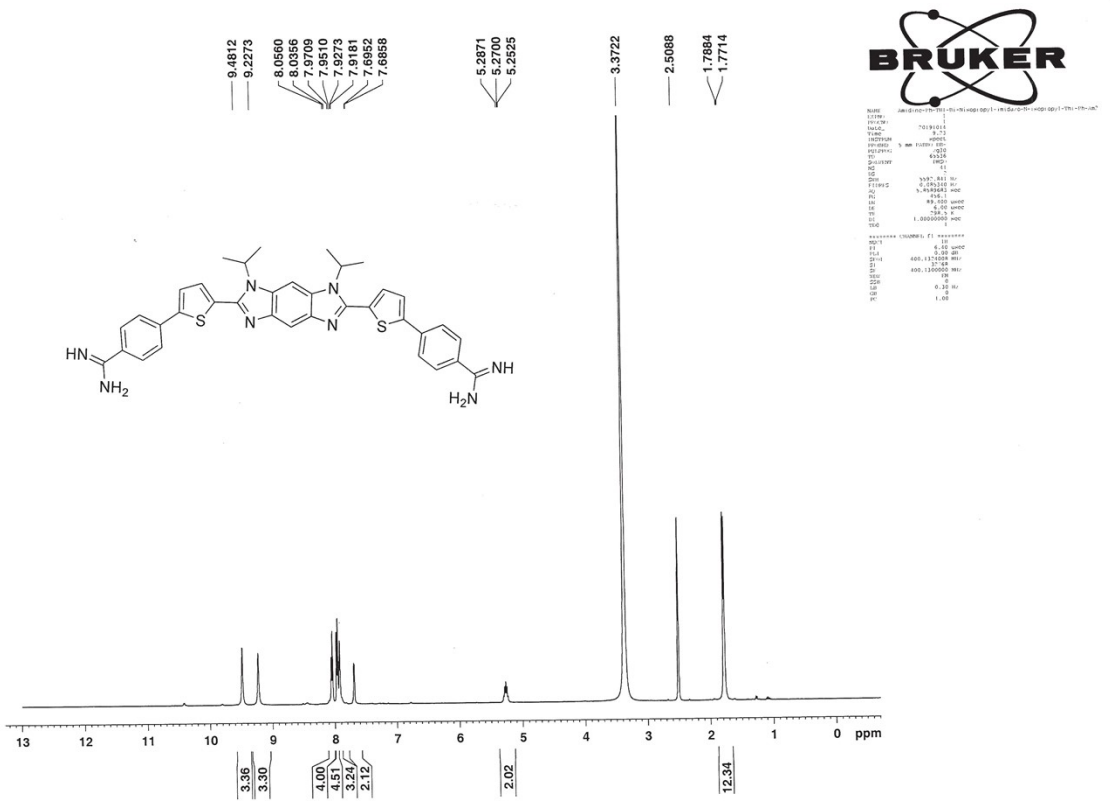
CA-CA-C* 5.99 120.10 SOURCE3 103 0.3451 ca-ca-cc
CK-C*-S 78.460 120.940 SOURCE4 31 1.2422 ce-cc-ss--gaff
CB-CA-CB 63.0 120.00 parm99

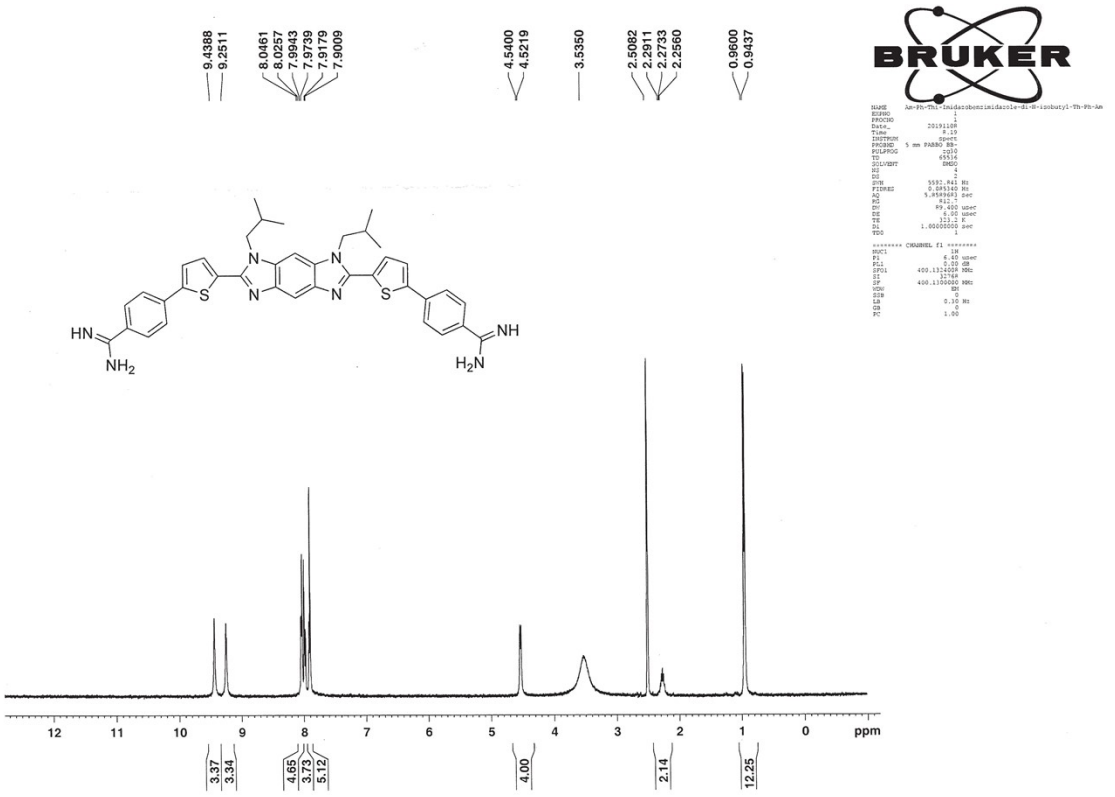
DIHE
N2-CA-N2-H 4 9.60 180.0 2.0 parm 99, X -CA-N2-X
H -N2-CA-CA 4 9.60 180.0 2.0 parm 99, X -CA-N2-X
N2-CA-CA-CA 4 -3.118 0.000 -2.0 DB921
N2-CA-CA-CA 4 0.789 327.0 -4.0 DB921
N2-CA-CA-CA 4 0.609 90.000 1.0 DB921
CA-CA-CA-CA 4 14.50 180.0 2.0 parm99, X -CA-CA-X
CA-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X
HA-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X
HA-CA-CB-N* 4 14.00 180.0 2.0 parm99, X -CA-CB-X
HA-CA-CB-NB 4 14.00 180.0 2.0 parm99, X -CA-CB-X
HA-CA-CB-CB 4 14.00 180.0 2.0 parm99, X -CA-CB-X
CB-CB-N*-CK 4 6.60 180.0 2.0 parm99, X -CB-N*-X
CB-CB-N*-CT 4 6.60 180.0 2.0 parm99, X -CB-N*-X
H1-CT-N*-CB 1 0.00 000.0 -2. parm98, TC,PC,PAK FOR OS-CT-N*CK
H1-CT-N*-CB 1 2.50 0.0 1. parm98, TC,PC,PAK FOR OS-CT-N*CK
CA-CB-N*-CT 4 6.60 180.0 2.0 parm99, X -CB-N*-X
CA-CB-CB-NB 4 21.80 180.0 2.0 parm99, X -CB-CB-X
CA-CB-CB-N* 4 21.80 180.0 2.0 parm99, X -CB-CB-X
CB-N*-CT-CT 1 0.00 000.0 -2. parm98, TC,PC,PAK FOR OS-CT-N*CK
CB-N*-CT-CT 1 2.50 0.0 1. parm98, TC,PC,PAK FOR OS-CT-N*CK
N*-CT-CT-HC 9 1.40 0.0 3. JCC,7,(1986),230, X -CT-CT-X
CA-CB-NB-CK 2 5.10 180.0 2.0 parm99, X -CB-NB-X
CA-CB-N*-CK 4 6.60 180.0 2. JCC,7,(1986),230, X -CB-N*-X
NB-CK-N*-CB 4 6.80 180.0 2.0 parm99, X -CK-N*-X
N*-CK-NB-CB 2 20.00 180.0 2.0 parm99, X -CK-NB-X
C*-CK-NB-CB 2 20.00 180.0 2.0 parm99, X -CK-NB-X
C*-CK-NB-CB 2 20.00 180.0 2. JCC,7,(1986),230, X -CK-NB-X
NB-CK-C*-S 4 -0.6 180.0 -4.0 DB921 for NB-CK-CA-CA
NB-CK-C*-S 4 3.1 180.0 -2.0 DB921 for NB-CK-CA-CA

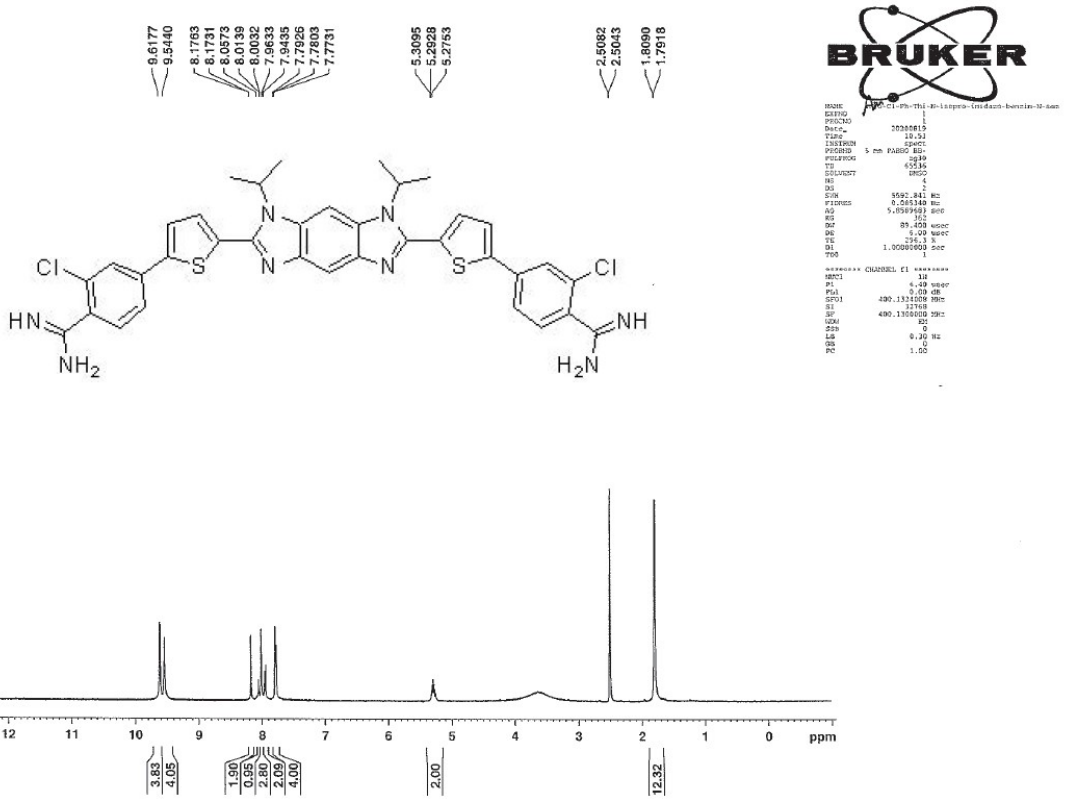
References

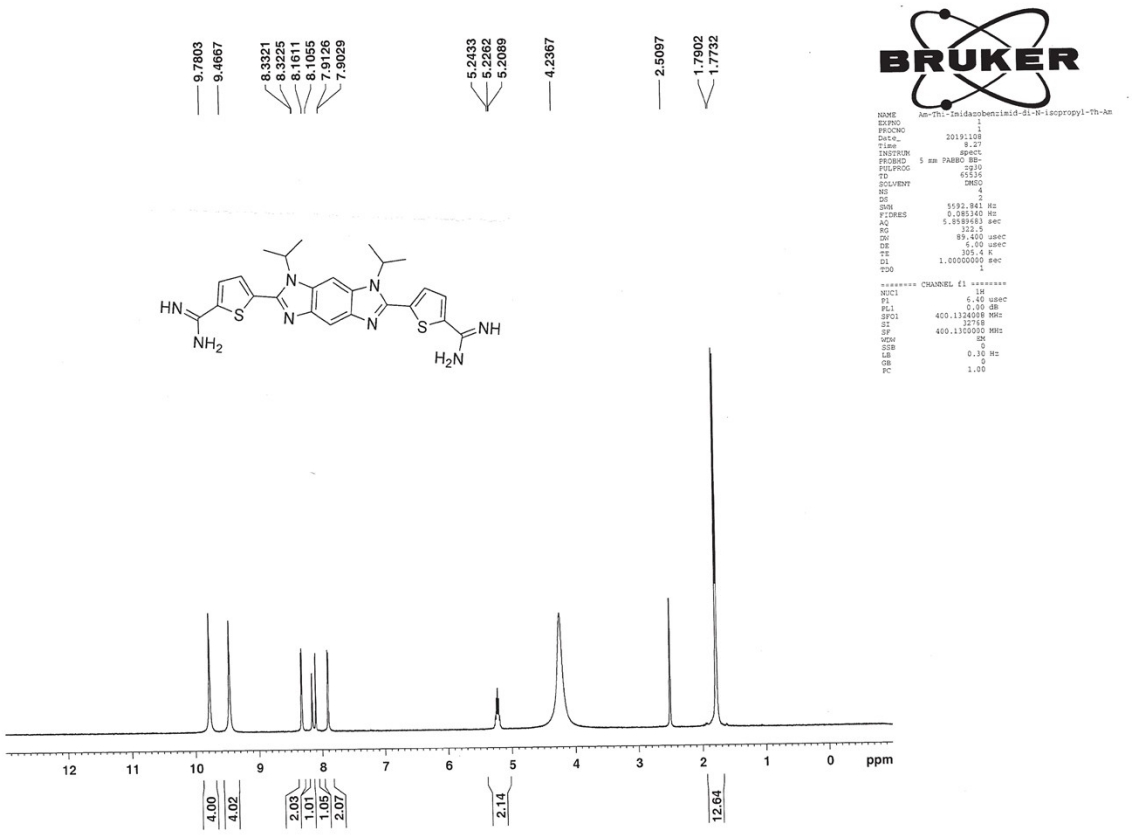
1. A. A. Farahat, C. Bennett-Vaughn, E. M. Mineva, A. Kumar, T. Wenzler, R. Brun, Y. Liu, W. D. Wilson and D. W. Boykin. *Bioorg. Med. Chem. Lett.*, 2016. **26**, 5907; A. A. Farahat, A. Kumar, M. Say, A. E.-D. M. Barghash, F. E. Goda, H. M. Eisa, T. Wenzler, R. Brun, Y. Liu, L. Mickelson, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem*. 2010, **18**, 557; D. Branowska, A. A. Farahat, T. Wenzler, R. Brun, Y. Liu, W. D. Wilson and D. W. Boykin, *Bioorg. Med. Chem*., 2010, **18**, 3551.
2. Y. Chai, A. Paul, M. Rettig, W. D. Wilson and D. W. Boykin, *J. Org. Chem*, 2014, **79**, 852; P. Guo, A. A. Farahat, A. Paul, N. K. Harika, D. W. Boykin and W. D. Wilson, *J. Am. Chem. Soc*., 2018, **140**, 14761; A. A. Farahat, P. Guo, H. Shoeib, A. Paul, D. W. Boykin and W. D. Wilson, *Chem. Eur.*

J., 2020, **26**, 4539; P. Guo, A. Paul, A. Kumar, A. A. Farahat, D. Kumar, S. Wang, D. W. Boykin and W. D. Wilson, *Chem. Eur. J.*, 2016, **22**, 15404
3. B. Nguyen, F. A. Tanious and W. D. Wilson, *Methods,* 2007, **42**, 150.
4. S. Laughlin and W. D. Wilson, *Int. J. Mol. Sci*., 2015, **524**, 506
5. B. Nguyen, S. Neidle and W. D. Wilson, *Acc. Chem. Res*., 2009, **42**, 11.
6. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.
7. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem*., 1993, **97**, 10269; U. C. Singh, P. A. Kollman, *J. Comput. Chem*., 1984, **5**, 129.
8. D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, I. A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and Kollman, P. A. AMBER 14, 2014.
9. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model*, 2006, **25**, 247; N. K. Harika, M. W. Germann and W. D. Wilson, *Chem. Eur. J*., 2017, **23**, 17612; P. Athri and W. D. Wilson, *J. Am. Chem. Soc*., 2009, **131**, 7618; N. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza and J. Šponer, *J. Am. Chem. Soc*., 2003, **125**, 1759.
10. O. Trott and A. J. Olson, *J. Comput. Chem*., 2009, **65**, 455.
11. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
12. T. J. Macke, D. A. Case, *Molecular Modeling of Nucleic Acids*, *American Chemical Society, Washington, D.C., USA,* 1998, 379.
13. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
14. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327.
15. G. S. Couch, D. K. Hendrix and T. E. Ferrin, *Nucleic Acids Res*. **2006**, *34*, e29.


9.4812
9.2273
8.0560
8.0356
7.9709
7.9510
7.9273
7.9161
7.6952
7.6858
5.2871
5.2700
5.2525
3.3722
2.5088
1.7884
1.7714
HN
NH2
NH
H2N
13
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm
3.36
3.30
4.00
4.51
3.24
2.12
2.02
12.34


9.4388
9.2511
8.0461
8.0257
7.9943
7.9739
7.9179
7.9009
4.5400
4.5219
3.5350
2.5082
2.2911
2.2733
2.2560
0.9600
0.9437
BRUKER
HN
NH2
NH
H2N
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm
3.37
3.34
4.65
3.73
5.12
4.00
2.14
12.25

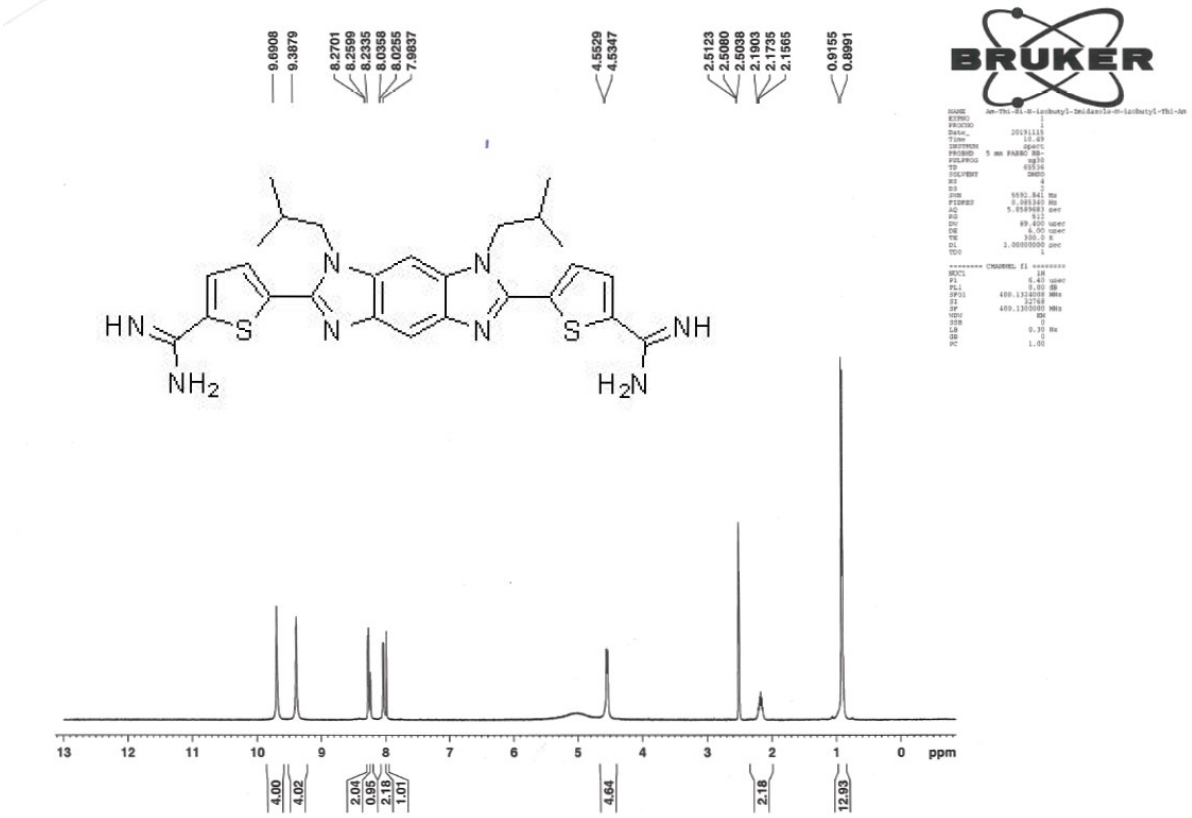

9.6177
9.5440
8.1763
8.1731
8.0573
8.0139
8.0032
7.9633
7.9435
7.7926
7.7803
7.7731
5.3095
5.2928
5.2753
2.5082
2.5043
1.8090
1.7918
Cl
HN
NH2
S
N
N
N
S
Cl
NH
H2N
BRUKER
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm
3.83
4.05
1.90
0.95
2.80
2.09
4.00
2.00
12.32

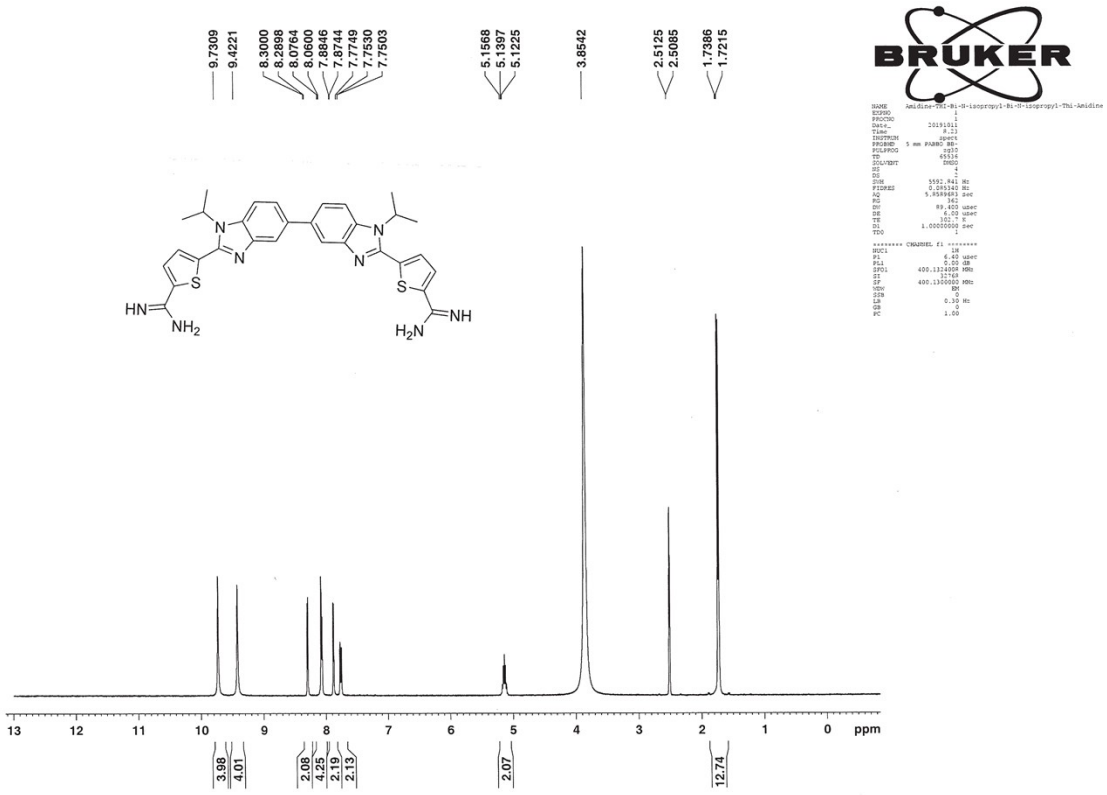


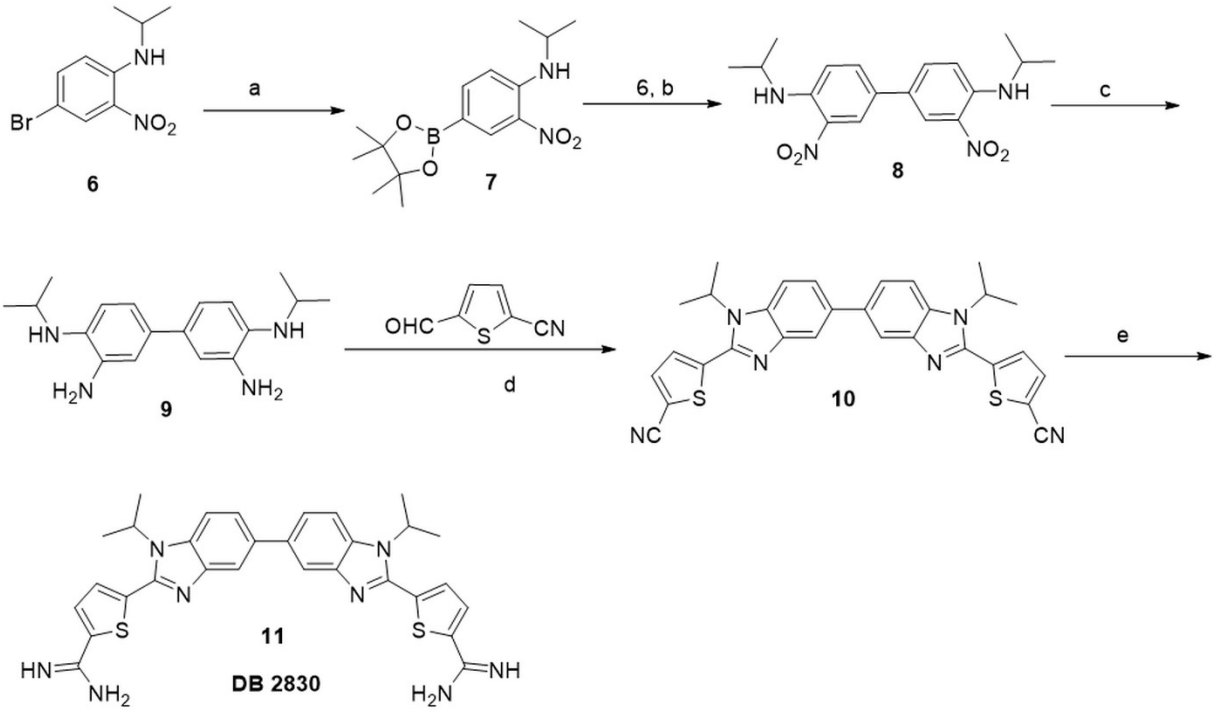
BRUKER

```
NAME            Am-Thi-Imidazobenzimid-di-N-isopropyl-Th-Am
EXPNO                    1
PROCNO                   1
Date_         20191108
Time                 8.27
INSTRUM              spect
PROBHD    5 mm PABBO BB-
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                       4
DS                       2
SWH              5592.841 Hz
FIDRES           0.085340 Hz
AQ               5.8589683 sec
RG                   322.5
DW                 89.400 usec
DE                  6.00 usec
TE                 305.4 K
D1               1.00000000 sec
TD0                      1

======== CHANNEL f1 ========
NUC1                    1H
P1                   6.40 usec
PL1                  0.00 dB
SFO1          400.1324008 MHz
SI                   32768
SF            400.1300000 MHz
WDW                     EM
SSB                      0
LB                   0.30 Hz
GB                       0
PC                    1.00
```


9.6908
9.3879
8.2701
8.2599
8.2335
8.0358
8.0255
7.9837
4.5529
4.5347
2.5123
2.5080
2.5038
2.1903
2.1735
2.1565
0.9155
0.8991
BRUKER
HN
NH2
S
N
N
N
N
S
NH
H2N
13
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm
4.00
4.02
2.04
0.95
2.18
1.01
4.64
2.18
12.93


9.7309
9.4221
8.3000
8.2898
8.0764
8.0600
7.8846
7.8744
7.7749
7.7530
7.7503
5.1568
5.1397
5.1225
3.8542
2.5125
2.5085
1.7386
1.7215
HN
NH2
S
N
N
N
S
NH
H2N
13
12
11
10
9
8
7
6
5
4
3
2
1
0
ppm
3.98
4.01
2.08
4.25
2.19
2.13
2.07
12.74



Scheme S2: Reagents and conditions: a) Bis(pinacolato)diboron, $PdCl_2(PPh_3)_2$, K acetate, Dioxane; b) $Pd(PPh_3)_4$, K_2CO_3/H2O, dioxane, reflux; c) $NaBH_4$, Pd(C), CH_2Cl_2,MeOH; d) $Na_2S_2O_5$/ DMSO, 130˚C; e) i- $LiN(TMS)_2$/THF, ii-HCl/EtOH