SUPPLEMENTARY DATA FOR

Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines

Manoj Munde,1 Shuo Wang,1 Arvind Kumar,1 Chad E. Stephens,1,a Abdelbasset A. Farahat,1,b David W. Boykin,1 W. David Wilson,1 Gregory M. K. Poon2,*

September 25, 2013

1 Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA
2 Department of Pharmaceutical Sciences, Washington State University, Pullman, WA 99164-6534, USA

a Current address: Department of Chemistry and Physics, Georgia Regents University, 2500 Walton Way, Augusta, GA 30904, USA

b Current address: Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt

* Address correspondence to: Gregory Poon (gpoon@wsu.edu).

SUPPLEMENT INDEX

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary methods and references</td>
<td>1 to 7</td>
</tr>
<tr>
<td>Summary of Supplementary Figures</td>
<td>8</td>
</tr>
<tr>
<td>Supplementary Figures SD1 to SD6</td>
<td>9 to 14</td>
</tr>
<tr>
<td>References for Supplementary Figures</td>
<td>15</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY METHODS

Synthesis of heterocyclic selenophene dications.


Regents and conditions: a) 4-bromobenzonitrile, Pd(PPh₃)₄, toluene, 110°C. b) Br₂, NaOAc, HOAc, c) 4-cyanophenylboronic acid, Pd(OAc)₂, (n-Bu)₄NF, DME, K₂CO₃, 90°C, d) LiN(TMS)₂, THF.

2-(4-Cyanophenyl)selenophene.

A mixture of 2-(tri-n-butylstannyl)selenophene (8.82 g, 21 mmole), 4-bromobenzonitrile (3.64 g, 20 mmole) and Pd(Ph₃)₄ (0.30 g) in toluene (50 mL) was heated at 100-110°C (oil bath) under nitrogen for 24 hr. The reaction mixture was then directly purified by chromatography by adding silica gel, concentrating to dryness, and eluting on a column with hexanes, followed by a gradient of diethyl ether in hexanes. Combination of pure fractions gave a yellow/tan solid (4.5 g). Recrystallization from diethyl ether/hexanes gave an off white solid (4.00 g, 82%), mp 84.5-85 °C. ¹H-NMR (CDCl₃): 7.36 (dd, J = 5.5 and 3.9 Hz, 1H), 7.58 (dd, J = 3.9 and 1.0 Hz, 1H), 7.64 (s, 4H), 8.07 (dd, J = 5.5 and 1.0 Hz, 1H). IR (cm⁻¹): 3095, 3056, 2223, 1599, 1496, 1433, 1409, 1258, 1209, 1175, 689, 820, 704, 561. HRMS (ESI): Theory calcd for C₁₁H₈NSe: 233.9822. Found: 233.9820. Analysis calcd for C₁₁H₇NSe (232.14): C, 56.91; H, 3.04; N, 6.03. Found: C, 56.90; H, 2.93; N, 5.96.

2-Bromo-5-(4-cyanophenyl)selenophene.

To a solution of 2-(tri-n-butylstannyl)selenophene (8.82 g, 21 mmole) and anhydrous sodium acetate (0.86 g, 10.5 mmole) in AcOH (60 mL) was added dropwise over 10 minutes a solution of bromine (1.7 g, 10.6 mmole) in AcOH (5 mL). After stirring for 20 min at rt, the mixture was chilled and then diluted with water to give a thick ppt, which was collected and washed with water. Recrystallization from MeOH (100 mL) with addition of water (20 mL) gave an off-white
solid (2.77 g, 86%), mp 115-116.5 °C. $^1$H-NMR (DMSO-$d_6$): 7.49 (d, $J = 4.0$ Hz, 1H), 7.67 (d, $J = 4.0$ Hz, 1H), 7.76 (d, $J = 8.4$ Hz, 2H), 7.84 (d, $J = 8.4$ Hz, 2H). $^{13}$C-NMR (DMSO-$d_6$): 149.0, 138.6, 134.6, 132.6, 128.2, 125.7, 118.2, 116.3, 109.5. IR (cm$^{-1}$): 3066, 3026, 2228, 1603, 1495, 1435, 1203, 1178, 945, 905, 832, 799, 720, 556. HRMS (ESI): Theory calcd for C$_{11}$H$_6$BrNSe: 311.8927. Found: 311.8929. Analysis calcd for C$_{11}$H$_6$BrNSe (311.03): C, 42.48; H, 1.94; N, 4.50. Found: C, 42.44; H, 1.72; N, 4.46.

**2,5-Bis(4-cyanophenyl)selenophene.**

To a mixture of 2-bromo-5-(4-cyanophenyl)selenophene (0.622 g, 2.0 mmole), 4-cyanophenylboronic acid (0.32 g, 2.18 mmole), palladium acetate (0.012 g), and tetra-n-butylammonium fluoride (0.62 g) in dimethoxyethane (20 mL) was added a solution of K$_2$CO$_3$ (0.56 g) in H$_2$O (4 mL) and the mixture was heated under nitrogen at 90 °C (oil bath) for 2 hr. The resulting suspension was then cooled and filtered by suction to give green solid which was rinsed with diethyl ether and water. The product was recrystallized from EtOH/acetone to give a yellow fluffy solid (0.42 g). A second recrystallization from n-BuOH gave yellow micro-needles (0.36 g, 54%), mp 248-250 °C. $^1$H-NMR (DMSO-$d_6$): 7.87 (s, 8H), 7.97 (s, 2H). IR (cm$^{-1}$): 3050, 2219, 1599, 1557, 1492, 1454, 1409, 1286, 1276, 1173, 1124, 1108, 834, 799, 553. Theory calcd for C$_{18}$H$_{10}$N$_2$Se: 335.0087. Found: 335.0095 Theory calcd for C$_{11}$H$_6$BrNSe: 311.8927. Found: 311.8929. Theory calcd for C$_{11}$H$_6$BrNSe: 311.8927. Found: 311.8929. Analysis calcd for C$_{18}$H$_{10}$N$_2$Se (333.24): C, 64.87; H, 3.02; N, 8.41. Found: C, 64.42; H, 2.95; N, 8.23.

**2,5-Bis(4-amidinophenyl)selenophene Dihydrochloride (DB1213).**

2,5-Bis(4-cyanophenyl)selenophene (0.145 g, 0.435 mmole) was suspended in dry THF (4 mL) and treated with lithium hexamethyldisilazane (LiHMDS, 1.06M in THF) (2.0 mL) and the resulting yellow/orange solution was stirred at room-temperature overnight (18 hr). The solution was then chilled and treated with HCl-saturated EtOH (2 mL). The resulting suspension was again stirred overnight at room-temperature, then concentrated in vacuo. The resulting solid was stirred with a small volume of water, filter by suction, and dried in vacuo to give the title compound as an orange solid (0.13 g, 67%), mp > 370 °C dec. $^1$H-NMR (DMSO-$d_6$): 7.91 (s, 8H), 8.01 (s, 2H), 9.32 (br s, 8H). Analysis calcd for C$_{18}$H$_{16}$N$_4$Se$\cdot$2HCl$\cdot$0.25H$_2$O (444.73): Theory: C, 48.61; H, 4.19; N, 12.60. Found: C, 48.73; H, 4.11; N, 12.38.
Scheme 2. Synthesis of DB1281.

Regents and conditions: a) POCl₃, DMF. b) 4-cyano-1,2-phenylenediamine, DMF, nitrobenzene, 140°C. c) i) HCl, EtOH ii) NH₃, EtOH.

2-(4-Cyanophenyl)-5-formylselenophene.
To DMF (15 mL), chilled on an ice-water bath, was added dropwise POCl₃ (12 mL). A solution of 2-(4-cyanophenyl)selenophene (2.30 g, 10 mmole) in CH₂Cl₂ (3 mL) was then added dropwise to the chilled mixture, and the mixture was heated at ~40 °C overnight (18 hr). The mixture was then poured into ice water and extracted with EtOAc containing a trace amount of acetone to help dissolve the solid. After concentrating the extract to near dryness in vacuo, it was diluted with hexanes and allowed to stand for 30 min to give the product as a yellow solid. This solid was purified by boiling it in hexanes and filtering it while warm. The solid was then recrystallized from 2-PrOH/Acetone to give yellow microneedles (1.55 g, 60%), mp 195-195.5 °C. ¹H-NMR (CDCl₃): 7.69 (d, J = 4.0 Hz, 1H), 7.71 (s, 4H), 8.02 (d, J = 4.0 Hz, 1H), 9.80 (s, 1H). IR (cm⁻¹): 3092, 3046, 2853, 2820, 2219, 1657, 1603, 1451, 1409, 1212, 1179, 1059, 797, 725, 602, 561. HRMS (ESI): Theory calcd for C₁₂H₈NOSe: 261.9771. Found: 261.9774. Analysis calcd for C₁₂H₇NOSe (260.15): Theory: C, 55.40; H, 2.71; N, 5.38. Found: C, 55.18; H, 2.76; N, 5.41.

2-(5-Amidino-2-benzimidazolyl)-5-(4-amidino)selenophene Diformate (DB1281).
2-(4-Cyanophenyl)-5-formylselenophene (0.65 g, 2.5 mmole), 4-cyano-1,2-phenylenediamine (0.167 g, 2.5 mmole) and nitrobenzene (0.92 g, 3 eq) were combined in DMF (4 mL) and heated at 140°C for 20 hr under N₂. The mixture was then diluted with water (40 mL) to give an orange solid, which was collected and rinsed with hexanes. Recrystallization from Acetone/EtOH, with concentration to remove most of the acetone, gave a deep orange/red solid (0.35 g), which was collected and identified as the intermediate, non-cyclized imine product. The recrystallization
filtrate was concentrated to give 0.5 g of additional solid, which appeared to consist mainly of desired product. The two solids were combined and heated again with nitrobenzene (0.9 g) in DMF (4 mL) at 160-165 °C until TLC (Hexanes:EtOAc, 1:1) showed that the reaction was complete (about 6-7 hr). The mixture was then diluted with water and hexanes and filtered to give a yellow solid (0.92 g). This solid was purified by column chromatography using Hexanes: EtOAc (1:1) as eluent to give a fine orange crystalline solid (0.5 g), mp 315-320°C. Recrystallization from Acetone/EtOH, with concentration to remove the EtOH, gave a small amount of orange solid after standing in the freezer overnight. This solid was removed by decanting, and the remaining filtrate was diluted with water to give a fine yellow solid (0.33 g, 35%), mp 320-321°C. The bisnitrile was then subjected to the standard Pinner conditions (reaction with HCl in anhydrous EtOH, followed by reaction with NH₃/EtOH) to give, after neutralization with NaOH, the crude diamidine as the free base. This product (0.085 g) was purified by HPLC using a gradient of formic acid, ammonium formate and triethylamine in water to give the diformate salt of DB1281 as a yellow solid (0.022 g). 

\[ \text{H-NMR (DMSO-}d_6\text{): 7.57 (d, J= 8.4 Hz, 1H), 7.70 (d, J = 8.4 Hz, 1H), 7.85 (d, J = 8.7 Hz, 2H), 7.92 (d, J = 8.7 Hz, 2H), 8.01 (d, J = 4.0 Hz, 1H), 8.05 (s, 1H), 8.12 (d, J = 4.0 Hz, 1H), 8.47 (s, 2H, diformate Hs) (acidic Hs exchanging with water signal). Analysis calcd for C}_{19}H_{18}N_{6}Se.2HCO_2H.3H_2O (553.43): Theory: C, 45.58; H, 4.74; N, 15.19. Found: C, 45.50; H, 4.53; N, 15.10.


Regents and conditions: a) 4-amidino-1,2-phenylenediamine hydrochloride, 1,4-benzoquinone, EtOH, reflux. b) 4-(2-imidazoliny)-1,2-phenylenediamine hydrochloride, 1,4-benzoquinone, EtOH, reflux.
2, 5-Bis {2[(5(6)amidino-benzimidazolyl)]selenophene tetrahydrochloride (DB1976)

A well stirred solution of selenophene-2,5-dicarboxaldehyde (0.187 g, 0.001 mole), 4-amidino-1,2-phenylenediamine hydrochloride (0.374 g, 0.002 mole) and 1, 4-benzoquinone (0.216 g, 0.002 mole in anhydrous ethanol (40 ml) (under nitrogen) was heated under reflux for 8-10 h (TLC shows disappearance of aldehyde, hexane-EtOAc 7:3). The reaction mixture was cooled, filtered and stirred in 50 ml acetone and 5 ml ethanol. The solid was filtered, washed with dry ether and dried to yield the salt. The solid was dissolved in a 1:1 mixture of hot ethanol-methanol (175 ml), filtered, the volume reduced to 60 ml and acidified with HCl-saturated ethanol (6 ml). After stirring overnight and diluting with ether the dark green solid was filtered, washed with ether, and dried under vacuum at 50°C (12 h) yielding 0.39 g (66 %); mp >300°C dec. 1H NMR (DMSO-d6): 9.30 (s, 4H), 9.07 (s, 4H), 8.39 (s, 2H), 8.17 (d, 2H, J = 1.5 Hz), 7.78 (d, 2H, J = 8.7 Hz), 7.71 (dd, 2H, J = 1.5 Hz, J = 8.7 Hz); 13C NMR (DMSO-d6): 165.9, 150.5, 142.0, 138.9, 131.4, 122.4, 121.7, 115.9, 114.7; MS (FAB): m/z 449 (M++1); Anal. calculated for C20H16N8Se.4HCl.0.2 H2O: C, 40.25; H, 3.46; N, 18.77; Found: C, 40.77; H, 3.85; N, 18.35.

2, 5-Bis {2[(5(6)-2(imidazo-benzimidazolyl)]selenophene tetrahydrochloride (DB1977).

Selenophene-2,5-dicarboxaldehyde (0.187 g, 0.001 mole), 4-(2-imidazolinyl)-1,2-phenylenediamine hydrochloride (0.425 g, 0.002 mole) and 1, 4-benzoquinone (0.216g, 0.002 mole) in ethanol (30 ml) was heated at reflux for 8h (under nitrogen) following the standard workup described above yielded 0.42 g (65 %) dark green tetrahydrochloride salt; mp >320°C dec. 1H NMR (DMSO-d6): 10.61 (br, 4H), 8.40 (s, 2H), 8.37 (brs, 2H), 7.88 (d, 2H, J = 1.8 Hz, J= 8.4 Hz), 7.79 (d, 2H, J = 8.4Hz), 4.02 (s, 8H); 13C NMR (DMSO-d6): 165.2, 150.8, 142.5, 140.0, 138.9, 131.8, 122.9, 116.5, 115.8, 115.0, 44.1; MS (FAB): m/z 501 (M++1); Anal. calculated for C24H20N8Se.4HCl.0.5 H2O: C, 40.25; H, 3.46; N, 18.77; Found: C, 40.77; H, 3.85; N, 18.35.

SPR assay and data analysis. SPR experiments were performed at 25 ºC in filtered and degassed 25 mM Na2HPO4 (pH 7.4) containing 0.05% P20, 400 mM NaCl, and 1 mM EDTA. Steady state binding analysis was performed with multiple injections of different compound concentrations over the immobilized DNA surface at a flow rate of 30ul/min and 25 ºC. Solutions of known ligand concentration were injected through the flow cells until a constant steady-state response was obtained. Compound solution flow was then replaced by buffer flow resulting in dissociation of the complex. The reference response from the blank cell was subtracted from the response in each cell containing the DNA to give a signal (RU, response units) that is directly
proportional to the amount of bound compound. The predicted maximum response per bound compound in the steady-state region (RU\textsubscript{max}) was determined from the DNA molecular weight, the amount of DNA on the flow cell, the compound molecular weight, and the refractive index gradient ratio of the compound and DNA, as previously described(1). The number of binding sites and the equilibrium constant were obtained from fitting plots of RU versus C\textsubscript{free}.

The stoichiometry \( r \) of the reaction is calculated as follows:

\[
\frac{RU}{RU_{\text{max}}} = r
\]

RU is the observed (experimental) response in the plateau region and RU\textsubscript{max} is the predicted maximum response for a monomer protein binding to a nucleic acid site (Review). Dividing the observed steady-state response RU by the calculated RU\textsubscript{max} yields a stoichiometry normalized binding isotherm.

Binding results from the SPR experiments were fit with either a single site model (\( K_2 = 0 \)) or with a two site model:

\[
r = \frac{K_1 C_{\text{free}} + 2K_1K_2C_{\text{free}}^2}{1 + K_1 C_{\text{free}} + 2K_1K_2C_{\text{free}}^2}
\]

where \( r \) represents the moles of bound compound per mole of the DNA hairpin duplex, \( K_1 \) and \( K_2 \) are macroscopic binding constants, and \( C_{\text{free}} \) is the free compound concentration in equilibrium with the complex.

**Electrophoretic mobility shift assay.** A ~130-bp fragment harboring a single copy of the \( \lambda B \) motif from the \( \lambda \)\(\lambda\)2-4 enhancer was cloned between the EcoRI/HindIII sites of pUC19. The insert was amplified by PCR using standard M13 primers to generate a 209-bp amplicon which was purified from an agarose gel and quantified by UV absorption (1 A\textsubscript{260} = 50 ng/µL; calculated MW = 129,021). Inhibition of PU.1 ETS binding to this fragment was effected by titrating 10 nM of purified PU.1 ETS domain and 1 nM of DNA fragment with graded concentrations of compounds. The solution condition was 10 mM TrisHCl, pH 7.4, 150 mM Na\textsuperscript{+} (total), 0.1 mM EDTA, and 100 ng/µL acetylated BSA (Promega). At equilibrium, samples were loaded onto a running 6% polyacrylamide gel (20 V/cm). After electrophoresis, the gel was stained with SYBR
Gold (Invitrogen) and digitized using a Storm 860 instrument (GE Healthcare). Fractional PU.1-bound DNA $\theta$ as a function of total drug concentration was fitted to the Hill Equation:

$$\theta = \frac{[DB]^{n_H}}{IC_{50}^{n_H} + [DB]^{n_H}}$$  \hspace{1cm} (3)

where $IC_{50}$ is the apparent concentration of DB compound at 50\% inhibition, and $n_H$ is the Hill coefficient. Nonlinear regression was performed with Origin (Originlab; Northampton, MA).
SUMMARY OF SUPPLEMENTARY FIGURES

| Figure SD1 | DB293 binds a single site upstream of the 5'-GGAA-3' consensus at the λB motif, a PU.1 binding site. |
| Figure SD2 | How do the compounds bind to the PU.1 DNA recognition sequence? |
| Figure SD3 | Representative sensorgrams of the PU.1-DNA complex inhibited by DB1976 and DB1281. |
| Figure SD4 | EMSA reveals heterogeneity in PU.1 inhibition by the dications. |
| Figure SD5 | Hydroxyl radical footprints of the λB motif saturated with DB293 or DB1281. |
| Figure SD6 | DNase I and hydroxyl radical footprints of the λB motif site saturated with DB1976 and DB1977. |
Figure SD1. **DB293 binds predominantly upstream of the 5'-GGAA-3' consensus at the λB motif, a PU.1 binding site.** A DNA fragment harboring the λB site was titrated with DB293 (0.1 mM) probed by DNase I. The 5'-TTCC-3' strand was radiolabeled at the 5' end. Uncut as well as cut unbound DNA were included as controls. Footprints are marked by square brackets and the core consensus for PU.1 is marked “CC”. Note the footprint at S1 and negligible protection at S2 as seen for DB1281 (c.f. Fig. 3 in the main text). Additional samples containing PU.1 ETS domain were also included to demonstrate inhibition of PU.1 ETS binding (with the characteristic DNase-I hypersensitive site asterisked). Note the 5' → 3' direction from bottom to top of the gels.
Figure SD2. **How do the compounds bind to the PU.1 DNA recognition sequence?** The furan derivatives of Figure 1, as well as DB1213, have been investigated in detail previously and have been found to bind as 1:1 complexes in the minor groove of four or more DNA AT base pairs (2,3). The minor groove complex formation by these and other similar compounds has been correlated with a significant positive-induced CD spectrum in the absorption region of the compounds above 300 nm where there is no competition from DNA absorption (3,4). Although the structures of the benzimidazole-selenophene derivatives in Figure 1 are quite similar to the furans and would indicate a similar binding mode, CD studies have been conducted to experimentally evaluate their mode of binding to the λB sequence. Shown are representative CD titrations of λB site (3 μM) by representative selenophenes DB1281 (A; up to 6.5 μM) and DB1976 (B; up to 4.0 μM). Both compounds have strong, positive-induced CD signals in the compound absorption region between 350-400 nm that indicate a minor groove binding mode for the Se derivatives.

The λB site (5'-AAATAAAAAGGAAATGAAAC-3') has an 8-bp AT sequence at the 5' side of the GGAA consensus that forms an excellent binding site for the heterocyclic diamidines. On compound titration into the sequence, the first molecule to bind in the minor groove in the AT site will have several fairly equivalent AT binding sites. In principle a second compound should be able to bind to the eight base pair sequence but this second equivalent would be less favored due to the close approach of charged amidine groups, as well as entropic restrictions on the ways in which the two compounds can bind to an eight base pair sequence. There are also favorable mismatch sites at the 3' side of the GGAA consensus: GAAA and AAAC. Thus, the sequence would be expected to have one strong binding site for the compounds of Figure 1 and one to two weaker sites that would be populated at higher concentrations.
Figure SD3. Representative sensorgrams of the PU.1-DNA complex inhibited by (A) DB1976 and (B) DB1281. The red sensorgram in each figure is for free PU.1 binding to DNA. Gray sensorgrams from top to bottom represent PU.1 ETS plus increasing concentrations of compound. The blue sensorgram in each figure is for free compound binding to DNA.
Figure SD4. **EMSA reveals heterogeneity in PU.1 inhibition by the dicitations.** Binding of the PU.1 ETS domain to a DNA fragment harboring the high-affinity λB site was competed with graded concentrations of two pairs of homologs: (A) DB293/DB1281, and (B) DB1976/1977. ETS-bound and unbound DNA fragments were separated at equilibrium by electrophoretic mobility shift. Representative gel images for each compound are shown: U, unbound (no ETS present); ø, ETS present but without compound. Curves represent the best-fit of the Hill equation to the pooled data from two experiments. The fitted estimates of IC50 and n_H are given in Table I.
Figure SD5. Hydroxyl radical footprints of the λB motif saturated with DB293 or DB1281. A DNA fragment harboring the λB site was saturated with DB293 (0.1 mM) or DB1281 (1 µM) and probed by hydroxyl radicals. The two complementary strands (5'-GGAA-3' and 5'-TTCC-3') were separately radiolabeled at the 5' end. Uncut (−) as well as cut unbound DNA (ø) were included as controls. Footprints are marked by square brackets ("S1" and "S2") and the core consensus for PU.1 is marked "CC". Note the 5' → 3' direction from bottom to top of the gels. Intensity traces for the unbound and DB-bound lanes are shown in Fig. 3 in the main text.
Figure SD6. DNase I and hydroxyl radical footprints of the λB motif site saturated with DB1976 and DB1977. A DNA fragment harboring the λB site was saturated with DB1976 (1 µM) or DB1977 (0.1 µM) and probed by DNase I or hydroxyl radicals. The two complementary strands (5'-GGAA-3' and 5'-TTCC-3') were separately radiolabeled at the 5' end. Uncut (−) as well as cut unbound DNA (ø) were included as controls. A, Footprints are marked by square brackets ("S1" and "S2") and the core consensus for PU.1 is marked "CC". Note the 5' → 3' direction from bottom to top of the gels. Intensity traces for the unbound and DB-bound lanes are shown in Fig. 4 in the main text.
REFERENCES FOR SUPPLEMENTARY FIGURES


