

Appendix C

CSI Microarray and Quantitative DNase I Footprinting Data for an Internally Cy3-labeled Linear β -linked Polyamide Targeting GAA Repeats

The experiments in this chapter were performed in collaboration with Christopher L. Warren (University of Wisconsin, Madison), Professor Aseem Z. Ansari (University of Wisconsin, Madison), and Professor Peter B. Dervan (California Institute of Technology).

Abstract

A high-throughput Cognate Site Identity (CSI) microarray platform interrogating all 524,800 10-base pair variable sites is correlated to quantitative DNase I footprinting data of DNA binding pyrrole-imidazole polyamides. Two linear β -linked polyamides programmed to target (GAA)₃ repeat sequences and labeled with Cy3 at an internal position within the polyamide and a terminal position are compared with one another. The internally Cy3-labeled polyamide revealed a microarray-derived sequence motif of 5'-AAGAWGWWS-3,' (W = A,T; S = C,G), slightly different from the previously reported terminally labeled sequence motif of 5'-AARAARWWG-3' (R = A,G). Correlation of microarray intensities from the two polyamide experiments revealed a good, positive correlation (R = 0.91) that suggests the polyamide core and not choice of dye placement drives the sequence recognition preferences.

Introduction.

Cell-permeable small molecules which bind specific DNA sequences and are able to interfere with protein-DNA interfaces would be useful in modulating eukaryotic gene expression. For targeting the regulatory elements of eukaryotic genes, knowledge of the preferred binding landscape of the ligand and the energetics of each site would guide gene regulation studies. Pyrrole-imidazole polyamides are a class of cell permeable oligomers which can be programmed, based on simple aromatic amino acid pairing rules, to bind a broad repertoire of DNA sequences.^{1,2} Knowledge of polyamide match sites has allowed us to pursue the characterization of the equilibrium association constants and hence, free energies, of hairpin polyamides for cognate DNA sites by quantitative footprint titration methods. Despite the predictive power of simple pairing rules, the sequence dependent variability of DNA minor groove shape affords significant variability in the range of affinities for match as well as all formal single and double base pair mismatch sites.^{1,2}

CSI Microarray Platform. Several high-throughput platforms have been developed to characterize the binding properties of ligand-DNA interactions.³⁻⁶ Of these, two have explicitly studied the binding preferences of polyamides. The fluorescence intercalator displacement assay has interrogated polyamide binding to 512 unique 5 bp sequences in a microplate format.⁴ The most recently developed cognate site identifier (CSI) microarray platform presents all 524,800 unique ten-mers to fluorescently labeled polyamides, enabling an unbiased interrogation of binding preference.⁶ By coupling DNase I footprinting with the CSI microarray data, the binding affinities (K_a values) of a given DNA-binding molecule (such as a polyamide) for any DNA sequence can be determined. We have previously examined and experimentally defined the relationship between Cy3-labeled polyamides and their corresponding CSI Microarray intensity values.⁶ In this study, we ask whether the point at which the Cy3 dye has been attached significantly alters sequence specificity.

Two Cy3-labeled polyamides are examined on the CSI microarray that displays all unique 10 base pair sequences, one previously published⁶ and one new to this study, both of

core structure Im-Py- β -Im-Py- β -Im- β -Dp. In order to correlate the CSI relative affinities (intensities) to absolute affinities (K_a values), DNase I footprinting was performed on a subset of these sequences for both the Cy3-polyamide conjugate and the related, unlabeled polyamide of known biological activity.

Results and Discussion.

Polyamide Design. The linear β -linked polyamide architecture has been chosen to interrogate the effects of Cy3-dye placement on CSI microarray sequence specificity and quantitative DNase I footprint specificity. Polyamide **1**, Im-Py- β -Im-Py- β -Im- β -Dp (Figure C.1), is believed to target the intronic 5'-(GAA)_n-3' repeat hyperexpansion in cell culture, enabling 2.5-fold upregulation of the frataxin gene, whose deficiency causes the neurodegenerative disorder Friedreich's Ataxia.⁷ Limited knowledge about the linear β -linked class of polyamides^{6,8-11} precludes the existence of binding rules. The linear β -linked architecture has the added complexity of binding in 1:1 and 2:1 ligand/DNA stoichiometries, and we would anticipate that this class will be generally less useful due to sequence promiscuity resulting from multiple binding modes. Its 1:1 binding preferences for purine tracts, such as (GAA)_n, likely reflect shape selectivity for sequences with narrow DNA minor groove conformations.¹⁰ In a 2:1 binding stoichiometry, polyamide **1** would

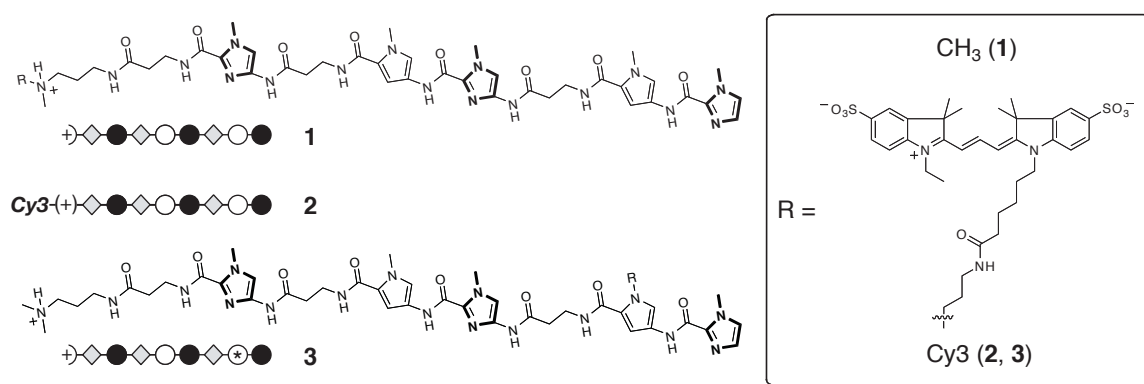


Figure C.1. Polyamides utilized to study the effects of dye placement on CSI microarray determined binding preferences

be predicted to target 5'-WGCWGCWGCW-3'.⁸ Remarkably, relatively few genes are affected from cell culture studies of **3** suggesting that this polyamide may be specific for 5'-AAGAAGAAG-3'.⁷ The Cy3 fluorophore has been conjugated to the C-terminal 3,3'-diamino-*N*-methyldipropylamine tail (polyamide **2**) in previous work.⁶ Current work examines Cy3 fluorophore conjugation to an internal pyrrole, replacing the *N*-methyl moiety with *N*-(propyl-3-amino) as a linker (polyamide **3**).¹²

CSI Microarray design and results. CSI microarrays were synthesized using maskless array synthesis (MAS) technology¹³ to display all 524,800 unique 10-base pair sites in quadruplicate across six microarrays. Replicates of individual hairpins occur on separate microarrays. Each hairpin on the chip consists of a self complementary palindromic sequence interrupted by a central 5'-GGA-3' sequence to facilitate hairpin formation: 5'-GCGC-N¹N²N³N⁴N⁵N⁶N⁷N⁸N⁹N¹⁰-GCGC-GGA-GCGC-N¹⁰'N⁹'N⁸'N⁷'N⁶'N⁵'N⁴'N³'N²'N¹'-GCGC-3' (N=A,T,C,G). Previous experiments have found that 95% of the oligonucleotides on the array form duplexes.⁵

Polyamide **3** was slowly titrated onto the arrays and imaged at each concentration until saturation of the highest intensity binding sites was observed, 250 nM concentration for **3**. After each small addition of polyamide, the arrays were washed prior to imaging. The data for each of the arrays was then normalized as previously described⁵ to give averaged sequence intensities of the 524,800 10-base pair sites for **3**. As found with previously reported CSI arrays,^{5,6} histograms of the probe intensities for **3** display a strong right-handed tail (Figure C.2). The fractional standard deviations among probe replicates (standard deviation of replicates / average normalized intensity) average 0.16 ± 0.10 , for intensities exceeding 1×10^3 (Figure C.3).

Plasmid Design. Two plasmids have been designed based on output from the CSI microarray intensities (Figure C.4). Because of our interest in calibrating the CSI microarray to K_a values, we chose a broad range of intensities to interrogate over two plasmids, pJWP16 and pJWP18. A single binding site (**IIa** and **IIIb**) was held constant

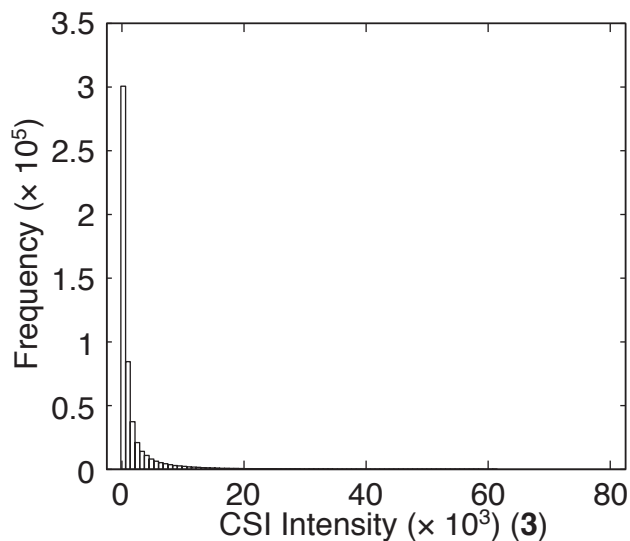


Figure C.2. CSI Intensity Histogram for polyamide 3

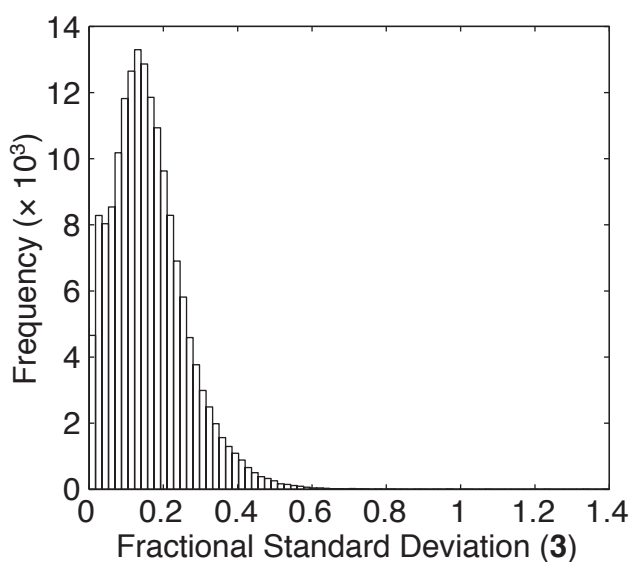


Figure C.3. CSI Fractional Standard Deviation Histogram for polyamide 3

between pJWP16 and pJWP18 to enable interplasmid comparison of binding affinities.

We designed each plasmid binding site to mimic the full 10 base pair binding site from the array in addition to two flanking base pairs on either side of the binding site: 5'-GC-(N)₁₀-GC-3' (N = A,T,C,G). Attempts to fully replicate the 5'-GCCG-(N)₁₀-GCCG-3' binding site from the array exhibited secondary structure formation when the

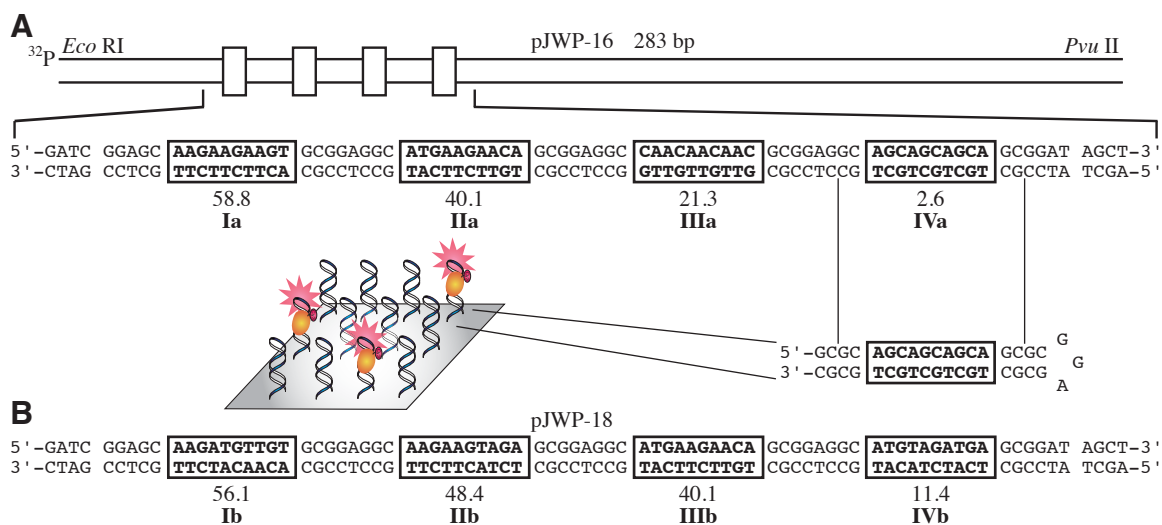


Figure C.4. Plasmids utilized to footprint polyamides **1** and **3**. Designed binding sites are boxed with a rectangle. Beneath each rectangle is the corresponding CSI microarray intensity ($\times 10^3$) and the Roman numeral label for the binding site. a) Structure of pJWP16. b) Structure of pJWP18

respective amplicons were sequenced and separated by denaturing gel electrophoresis.

Quantitative DNase I Footprint Titrations: Affinity and Specificity

Determination. Linear β -linked polyamides **1** and **3** were each incubated for 14 h with pJWP16 and pJWP18 prior to DNase I cleavage. They each bound seven unique 10-base pair sites in the same rank-order, preferentially binding 5'-AAGAAGAAGT-3' (Table C.1 and Figure C.5).

Appending the Cy3 dye to polyamide **1** (giving polyamide **3**) typically reduced binding affinity by one order of magnitude (Table C.1). Polyamide **1** bound all four binding sites over an 830-fold range in affinity, 60% broader than for polyamide **3**.

Calibrating microarrays for K_a prediction. The DNase I footprinting data (Table C.1) was mapped to CSI microarray intensities using Eq 1.⁶

$$\text{Intensity} = c \times \Theta = c \times \frac{K_a[\text{PA}]}{1 + K_a[\text{PA}]} = c \times \frac{[\text{PA}]}{K_d + [\text{PA}]} \quad (1)$$

The line was fit with an R^2 of 0.95, where $c = 59.152$ and $[\text{PA}] = 4.547 \times 10^{-8}$ M. Eq 1 will be rearranged to map CSI microarray intensities to interpolated K_a values.

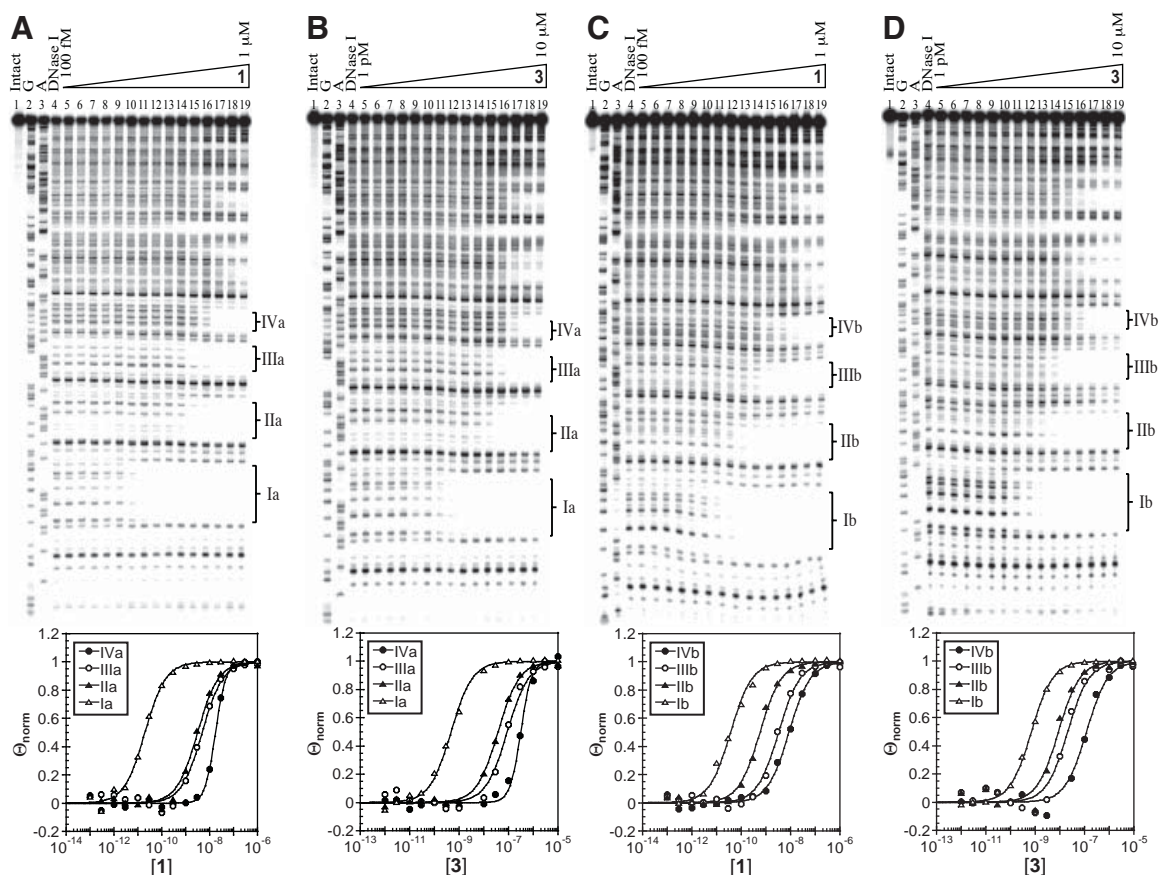


Figure C.5. Quantitative DNase I footprint titrations of polyamides **1** and **3** on plasmids pJWP16 and pJWP18. a) Polyamide **1** on pJWP16. b) Polyamide **3** on pJWP16. c) Polyamide **1** on pJWP18. d) Polyamide **3** on pJWP18

Correlating Binding Between Cy3-labeled Polyamides, Labeled at Different Positions on the Polyamide. Previous work has validated the correlation of Cy3-labeled polyamides with their unlabeled counterparts. A scatter plot of polyamide **1** versus **3** was best fit by a power relationship of $y = ax^n$, where (x,y) denotes the K_a values for (**1**, **3**) (Figure C.7). The R^2 between **1** and **3** is 0.94, and (a,n) were empirically determined as (0.49, 0.90).

To determine the extent which terminally-labeled and internally-labeled Cy3-polyamide conjugates **2** and **3** had correlated CSI Microarray intensity profiles, a scatter plot of all positive microarray intensities was generated (Figure C.8). A line ($y = mx + b$) was fit to the data, and a correlation coefficient (R) of 0.91 was observed. The (m, b) pair

Table C.1. Quantitative DNase I Footprinting Derived K_a values (M^{-1}) for Polyamides **1** and **2**. The 10 base pair binding sites and corresponding CSI Microarray intensities are found in each column. All footprinting incubations were conducted at a minimum in triplicate at 23 °C for 14 h. Standard deviations are shown in parentheses. The bracketed numbers are $K_{a-max}/K_{a-current}$ to compare values within each polyamide series (interplasmid comparisons are made).

pJWP-16		Ia	IIa	IIIa	IVa
Polyamide		AAGAAGAAGT	ATGAAGAACA	CAACAACAAC	AGCAGCAGCA
1	↻◊●◊◊◊●◊◊◊●	$5.4 (\pm 0.5) \times 10^{10}$ [1]	$4.6 (\pm 1.7) \times 10^8$ [120]	$2.5 (\pm 0.7) \times 10^8$ [220]	$6.5 (\pm 1.6) \times 10^7$ [830]
3	↻◊●◊◊◊◊●◊◊◊●	$1.7 (\pm 0.4) \times 10^9$ [1]	$2.5 (\pm 0.2) \times 10^7$ [68]	$1.3 (\pm 0.1) \times 10^7$ [130]	$3.1 (\pm 0.3) \times 10^6$ [550]
CSI Intensity ($\times 10^3$)		58.8 (± 2.7)	40.1 (± 10.3)	21.3 (± 0.3)	2.6 (± 1.6)
pJWP-18		Ib	IIb	IIIb	IVb
Polyamide		AAGATGTTGT	AAGAAGTAGA	ATGAAGAACA	ATGTAGATGA
1	↻◊●◊◊◊◊●◊◊◊●	$2.3 (\pm 0.4) \times 10^{10}$ [2.3]	$1.6 (\pm 0.2) \times 10^9$ [34]	$2.8 (\pm 0.2) \times 10^8$ [190]	$9.8 (\pm 1.4) \times 10^7$ [550]
3	↻◊●◊◊◊◊◊●◊◊◊●	$1.3 (\pm 0.6) \times 10^9$ [1.3]	$1.0 (\pm 0.3) \times 10^8$ [17]	$4.6 (\pm 1.8) \times 10^7$ [37]	$9.8 (\pm 1.8) \times 10^6$ [173]
CSI Intensity ($\times 10^3$)		56.1 (± 9.3)	48.4 (± 19.6)	40.1 (± 10.3)	11.4 (± 1.6)

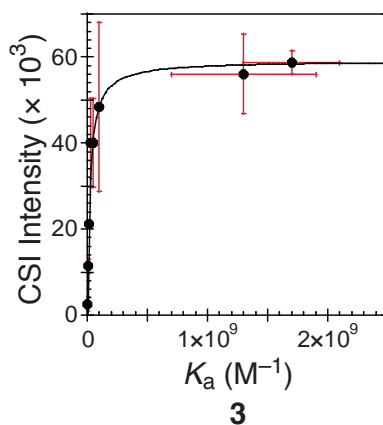


Figure C.6. CSI array intensities correlate well with DNase I footprinting-determined K_a values. Polyamide **3** versus CSI array fit to Eq 1

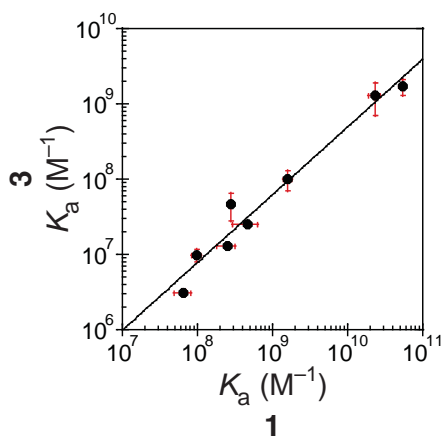


Figure C.7. Cy3-labeled polyamide **3** and unlabeled polyamide **1** correlate well

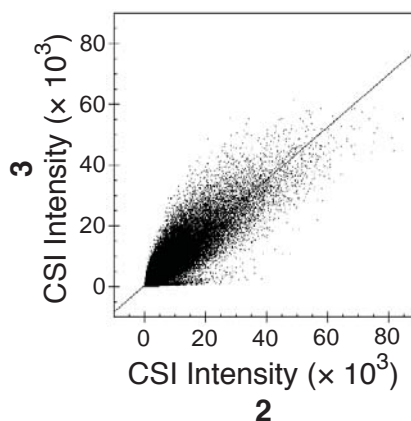


Figure C.8. Scatter Plot Correlation of End-labeled Cy3-polyamide **2** with internally labeled Cy3-polyamide **3**

was empirically determined to be (0.87, 0.25). This high correlation is certainly within the realm of array-to-array technical variability for CSI microarrays, previously reported as an average correlation coefficient of 0.88. This would suggest that Cy3 dye positioning is irrelevant to determining sequence preferences of the polyamide, and that the polyamide amino acid content drives the recognition event's specificity.

Sequence Analysis. To graphically represent the binding preferences of polyamide **3**, a sequence logo has been generated (Figure C.9).

In all cases, the motif finding program MEME¹⁴ was utilized to extract sequence motifs from the CSI binding intensities. The position specific probability matrices output

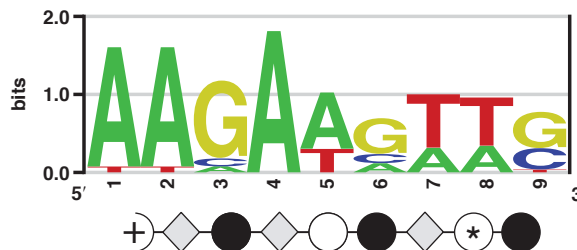


Figure C.9. Sequence logo for polyamide **3**

by MEME were used as inputs to enoLOGOS¹⁵ to generate a sequence logo. The motif search utilized the 10 variable bases of the microarray and two fixed bases, each flanking the variable site 5' and 3'. Because of the change in GC-content, a background GC-content of 58% was utilized in the searches. The logo for polyamide **3** interrogated the 48 highest intensity sequences (a 20-fold range in K_a) of the CSI microarray. We examined the K_a -weighted sequence logo for polyamide **3** and found minimal differences in the resulting logo.

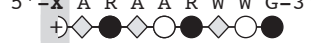
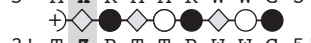
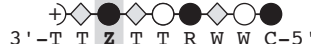



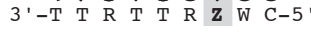
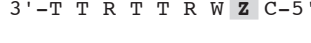
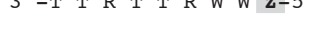
Polyamide **1** specifies 9 base pairs based on MPE footprinting data (Chapter 2B). Polyamide **3** elicits a 9 bp motif that is best represented as 5'-AAGAWGWS-3' (Figure C.9; S = G,C and W = A,T).

Quantitative Profiling of Single Base Pair Mismatches. While sequence logos provide a visual representation of sequence specificity, traditional studies on polyamides quantitate the specificity of a ring pairing at a selected base pair. We have examined a comprehensive single base pair mutational analysis of polyamide **3** using K_a values interpolated from the calibrated CSI microarrays (Tables C.2 and C.3) using the relationship defined in Eq 1. The examination of polyamide **3** complements the comprehensive sequence specificity study of linear β -linked polyamide **2**. In the 5'-A¹A²G³A⁴W⁵G⁶W⁷W⁸S⁹-3' sequence (S = G,C; W = A,T), positions 4, 5, and 7, each containing either a Py or a β , exhibit the greatest specificity for W over S (S = C,G). Intriguingly, the β at position 4 prefers A·T over T·A, an unexpected specificity. The sequence logo for polyamide **3** indicates that Im has modest preference for G·C or A·T over other base pairings—in this

Table C.2. Microarray-Derived Binding Affinities and Specificities of All Single Base Pair Mismatch Sites for Polyamide 3 Using 5'-AAGAWGWS-3' as the Preferred Motif. All K_a values are derived from the geometric average of all CSI binding site intensities on the array containing a specified sequence, converted to a K_a value using eq 3, corrected to include an error term ϵ . The values in parentheses are the geometric standard deviations for each K_a value.

Polyamide 3	X·Z	K_a (M^{-1})
5' - X A G A W G W W S-3' +) 3' - Z T C T W C W W S-5'	A·T T·A C·G G·C	$4.9 (3.2) \times 10^7$ $2.1 (2.1) \times 10^7$ $7.1 (2.2) \times 10^6$ $6.9 (2.3) \times 10^6$
5' -A X G A W G W W S-3' +) 3' -T Z C T W C W W S-5'	A·T T·A C·G G·C	$4.9 (3.2) \times 10^7$ $2.2 (2.2) \times 10^7$ $2.9 (2.3) \times 10^6$ $4.7 (2.4) \times 10^6$
5' -A A X A W G W W S-3' +) 3' -T T Z T W C W W S-5'	A·T T·A C·G G·C	$1.6 (1.9) \times 10^7$ $5.9 (2.0) \times 10^6$ $2.2 (2.1) \times 10^7$ $4.9 (3.2) \times 10^7$
5' -A A G X W G W W S-3' +) 3' -T T C Z W C W W S-5'	A·T T·A C·G G·C	$4.9 (3.2) \times 10^7$ $1.0 (2.1) \times 10^7$ $6.9 (1.9) \times 10^5$ $3.8 (2.2) \times 10^5$
5' -A A G A X G W W S-3' +) 3' -T T C T Z C W W S-5'	A·T T·A C·G G·C	$6.2 (3.7) \times 10^7$ $3.8 (2.5) \times 10^7$ $5.4 (1.7) \times 10^5$ $1.1 (2.1) \times 10^6$
5' -A A G A W X W W S-3' +) 3' -T T C T W Z W W S-5'	A·T T·A C·G G·C	$1.9 (2.1) \times 10^7$ $1.2 (1.9) \times 10^7$ $2.6 (2.5) \times 10^7$ $4.9 (3.2) \times 10^7$
5' -A A G A W G X W S-3' +) 3' -T T C T W C Z W S-5'	A·T T·A C·G G·C	$3.3 (2.8) \times 10^7$ $7.3 (3.1) \times 10^7$ $9.2 (1.6) \times 10^5$ $1.0 (1.5) \times 10^6$
5' -A A G A W G W X S-3' +) 3' -T T C T W C W Z S-5'	A·T T·A C·G G·C	$5.0 (3.2) \times 10^7$ $4.8 (3.2) \times 10^7$ $3.1 (1.9) \times 10^6$ $5.7 (2.1) \times 10^6$
5' -A A G A W G W W X -3' +) 3' -T T C T W C W W Z -5'	A·T T·A C·G G·C	$2.2 (1.7) \times 10^7$ $1.9 (2.2) \times 10^7$ $6.6 (3.1) \times 10^7$ $3.6 (2.8) \times 10^7$

Table C.3. Microarray-Derived Binding Affinities and Specificities of All Single Base Pair Mismatch Sites for Polyamide 3 Using 5'-AARAARWWG-3' as the Preferred Motif. All K_a values are derived from the geometric average of all CSI binding site intensities on the array containing a specified sequence, converted to a K_a value using eq 3, corrected to include an error term ϵ . The values in parentheses are the geometric standard deviations for each K_a value.

Polyamide 3	X·Z	K_a (M ⁻¹)
5' - X A R A A R W W G-3' (+)  3' - Z T R T T R W W C-5'	A·T T·A C·G G·C	2.6 (2.5) × 10 ⁷ 1.3 (2.9) × 10 ⁷ 4.0 (1.8) × 10 ⁶ 4.7 (1.9) × 10 ⁶
5' -A X R A A R W W G-3' (+)  3' -T Z R T T R W W C-5'	A·T T·A C·G G·C	2.6 (2.5) × 10 ⁷ 1.0 (2.1) × 10 ⁷ 2.2 (2.0) × 10 ⁶ 3.3 (2.3) × 10 ⁶
5' -A A X A A R W W G-3' (+)  3' -T T Z T T R W W C-5'	A·T T·A C·G G·C	1.7 (2.0) × 10 ⁷ 4.5 (1.7) × 10 ⁶ 1.2 (1.6) × 10 ⁷ 3.9 (2.6) × 10 ⁷
5' -A A R X A R W W G-3' (+)  3' -T T R Z T R W W C-5'	A·T T·A C·G G·C	2.6 (2.5) × 10 ⁷ 5.0 (1.8) × 10 ⁶ 1.1 (2.5) × 10 ⁶ 9.8 (3.8) × 10 ⁵
5' -A A R A X R W W G-3' (+)  3' -T T R T Z R W W C-5'	A·T T·A C·G G·C	2.6 (2.5) × 10 ⁷ 1.1 (2.2) × 10 ⁷ 1.2 (2.9) × 10 ⁶ 1.2 (2.6) × 10 ⁶
5' -A A R A A X W W G-3' (+)  3' -T T R T T Z W W C-5'	A·T T·A C·G G·C	2.3 (2.1) × 10 ⁷ 8.4 (1.7) × 10 ⁶ 1.5 (2.1) × 10 ⁷ 2.8 (2.9) × 10 ⁷
5' -A A R A A R X W G-3' (+)  3' -T T R T T R Z W C-5'	A·T T·A C·G G·C	2.9 (2.5) × 10 ⁷ 2.3 (2.5) × 10 ⁷ 2.8 (2.4) × 10 ⁶ 2.6 (2.6) × 10 ⁶
5' -A A R A A R W X G-3' (+)  3' -T T R T T R W Z C-5'	A·T T·A C·G G·C	2.6 (2.6) × 10 ⁷ 2.5 (2.5) × 10 ⁷ 5.0 (2.0) × 10 ⁶ 4.7 (1.9) × 10 ⁶
5' -A A R A A R W W X -3' (+)  3' -T T R T T R W W Z -5'	A·T T·A C·G G·C	2.1 (1.9) × 10 ⁷ 1.8 (1.9) × 10 ⁷ 3.3 (2.5) × 10 ⁷ 2.6 (2.5) × 10 ⁷

mutational study, however, imidazole is generally degenerate. The wide range of K_a values comprising each motif (high geometric standard deviation) make the statistical significance of any specificities under 8 relatively small. In general, the geometric standard deviations for polyamide **3** were higher than those for polyamide **2** within this single base pair mismatch study.

To compare sequence preferences directly with the published mismatch study of polyamide **2**,⁶ 5'-A¹A²R³A⁴A⁵R⁶W⁷W⁸G⁹-3' (R = G,A; Table C.3) was used as a seed from which individual mutations would be studied for polyamide **3**. The results were the same as for 5'-AAGAWGWS-3,' with positions 4, 5, and 7, showing the greatest specificity for W over S. The β at position 4 prefers A·T over T·A.

With the sequence logo (approximated as 5'-AAGAWGWS-3') as a snapshot of the highest affinity binding sites for polyamide **3** and the footprint titration binding isotherms for determining DNA binding mode, we confirm a preference for the 1:1 binding stoichiometry. Previous data characterizing the linear β -linked polyamide Im- β -Im-Py- β -Im- β -Im-Py- β -Dp demonstrated a 30-fold energetic preference for the 1:1 versus 2:1 binding stoichiometry, presumably due to the increased entropic cost of the 2:1 binding mode.⁸ Here, we have observed a nearly 830-fold preference for the favorable match binding site **Ia** versus the 2:1 stoichiometry binding at binding site **IVa**.

Conclusion.

Correlating the sequence preference landscape present on the CSI microarray to quantitative footprinting enables energetic studies using global binding information. The relationship between Cy3-labeled and unlabeled polyamides have previously been well-correlated, but the question of effects of dye label position has now been addressed. The strong correlation between polyamides labeled at different positions confirms that polyamide core sequence drives sequence preferences. DNase I footprinting-calibrated CSI microarrays will be an effective technique for determining the binding affinities of

DNA-binding ligands for a vastly expanded repertoire of DNA sequences.

Experimental Methods.

Materials. Boc- β -Ala-Pam resin (0.81 mmol/g), HBTU, and Boc₂O were purchased from Peptides International. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All solvents were purchased from Aldrich or EMD Biosciences. *rac*-Dithiothreitol (DTT) was purchased from ICN. Cy3-NHS ester was purchased from Invitrogen. RNase-free DEPC water was purchased from US Biochemicals. Water (18.2 M Ω) was purified using a Millipore water purification system.

The pH of buffers was adjusted using a Beckman 340 pH/temp meter. All buffers were sterilized by filtration through either a Nalgene 0.2 μ m cellulose nitrate filtration device or a 0.2 μ m Whatman cellulose acetate disposable syringe filter. DNA oligonucleotides were ordered PAGE-purified from Integrated DNA Technologies. [γ -³²P] adenosine 5'-triphosphate (≥ 7000 Ci/mmol) was obtained from MP Biomedicals. Sonicated, phenol-extracted calf thymus DNA was from Amersham, and all enzymes and molecular biology grade glycogen (20 mg/mL) were purchased from Roche.

Methods. UV spectra were recorded in water using an Agilent 8453 UV-Vis spectrophotometer for polyamide **3**. To solvate polyamide **3** in water at sufficiently high concentrations for DNase I footprinting, 5 μ L of DMSO was added to a 20 nmol aliquot of polyamide. The solution was then diluted using RNase-free DEPC water. UV spectra of **3** were blanked against solutions containing appropriate amounts of DMSO. The concentrations of polyamides **1** and **3** were determined using a local λ_{max} of 288 nm, $\epsilon = 43,125 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$. LASER desorption / ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro spectrometer. Analytical and preparative high-pressure liquid chromatography (HPLC) were performed with a Beckman Gold system equipped with a diode array (analytical) or single-wavelength (preparative) detector as previously described.⁶

Synthesis. Polyamides **1**, **2**, and **3** were synthesized on Boc- β -Ala-PAM resin,¹⁶ using previously described methods.⁶

1: (MALDI-TOF) [M+H]⁺ calcd for C₄₁H₅₆N₁₇O₈⁺ 914.4, observed 914.4

2: (MALDI-TOF) [M+H]⁺ calcd for C₄₁H₅₆N₁₇O₈⁺ 1569.7, observed 1569.7

3: (MALDI-TOF) [M+H]⁺ calcd for C₇₄H-N₂₀O₁₅S₂⁺ 1569.7, observed 1569.6

Plasmid Preparation. Plasmids were constructed by ligating the following hybridized inserts into the BamHI / HindIII polycloning site in pUC19:

pJWP16. 5'-GATCGGAGCAAGAAGAAGTGC GGAGGCATGAAGAACAGCG
GAGGCCAACAAACGCGGAGGCAGCAGCAGCAGCAGCGGAT-3'·5'-AGCTA
TCCGCTGCTGCTGCTGCCTCCGCGTTGTTGTTGGCCTCCGCTGTTCTTCA
TGCCTCCGCACTTCTTCTTGCTCC-3'

pJWP18. 5'- GATCGGAGCAAGATGTTGTGCGGAGGCAAGAAGTAGAGCG
GAGGCATGAAGAACAGCGGAGGCATGTAGATGAGCGGAT-3'·5'-AGCTATC
CGCTCATCTACATGCCTCCGCTGTTCTTCATGCCTCCGCTCTACTTCTTGC
CTCCGCACAACATCTTGCTCC-3.'

The ligated plasmid was then transformed into JM109 subcompetent cells. Colonies were selected for α -complementation on agar plates containing 50 mg/L ampicillin, 120 mg/L IPTG, and 40 mg/L X-gal after overnight growth at 37 °C. Cells were harvested after 16 h growth at 37 °C in LB medium containing 50 mg/L ampicillin. Plasmids were then purified by mini-prep kits. The presence of the desired inserts was determined by capillary electrophoresis dideoxy sequencing methods.

Preparation of 5'-Labeled DNA for DNase I Footprinting. Two primer oligonucleotides, 5'-AATTCGAGCTCGGTACCCGGG-3' (forward) and 5'-CTGGCACGACAGGTTTCCCGA-3' (reverse) were constructed for PCR amplification. The forward primer was radiolabeled using [γ -³²P]-dATP and polynucleotide kinase, followed by purification using ProbeQuant G-50 spin columns. The desired DNA region was amplified as previously described.¹⁷ The labeled fragment was loaded onto

a 7% nondenaturing preparatory polyacrylamide gel (5% cross-link), and the desired 283 (pJWP16, pJWP18) base-pair band was visualized by autoradiography and isolated. Chemical sequencing reactions were performed according to published protocols.¹⁷

Quantitative DNase I Footprint Titrations. All reactions were carried out in a volume of 400 μ L according to published protocols. Polyamides were equilibrated with the radiolabeled DNA for 14 h prior to DNase I cleavage at 23 °C. Quantitation by storage phosphor autoradiography and determination of equilibrium association constants were as previously described.¹⁷

Microarray Procedures. Microarrays were synthesized by using a Maskless Array Synthesizer (NimbleGen Systems, Madison, WI). Homopolymer (T_{10}) linkers were covalently attached to monohydroxysilane glass slides. Oligonucleotides were then synthesized on the homopolymers to create a high-density oligonucleotide microarray. The array surface was derivatized such that the density of oligonucleotides was sufficiently low within the same feature so that no one oligonucleotide would hybridize with its neighbors. Four copies of each hairpin containing a unique 10 bp site (5'-GCGC-N¹N²N³N⁴N⁵N⁶N⁷N⁸N⁹N¹⁰-GCGC-GGA-GCGC-N^{10'}N^{9'}N^{8'}N^{7'}N^{6'}N^{5'}N^{4'}N^{3'}N^{2'}N^{1'}-GCGC-3') required a total of 2,099,200 features, divided among six microarrays.

Binding Assay. Microarray slides were immersed in 1x PBS and placed in a 90 °C water bath for 30 min to induce hairpin formation of the oligonucleotides. Slides were then transferred to a tube of nonstringent wash buffer (saline/sodium phosphate/EDTA buffer, pH 7.5/0.01% Tween 20) and scanned to check for low background (<200 intensity). Microarrays were scanned by using an Axon 4000B, and the image files were extracted with GENEPIX PRO Version 3.0 (Axon Instruments, Foster City, CA).

Polyamide Binding. Microarrays prepared as above were placed in the microarray hybridization chamber and washed twice with nonstringent wash buffer. Polyamide was diluted to 250 nM (for **3**) in Hyb buffer (100 mM Mes/1 M NaCl/20 mM EDTA, pH 7.5/0.01% Tween 20). Polyamide was then added to the hybridization chamber and

incubated at room temperature for 1 h. Finally, the microarrays were washed twice with nonstringent wash buffer and scanned.

Data Processing. For each replicate, global mean normalization was used to ensure the mean intensity of each microarray was the same. Local mean normalization¹⁸ was then used to ensure that the intensity was evenly distributed throughout each sector of the microarray surface. Outliers between replicate features were detected by using the Q test at 90% confidence and filtered out. The replicates were then quantile-normalized¹⁹ to account for any possible nonlinearity between arrays. Duplicate features were then averaged together. The median of the averaged features was subtracted to account for background.

Acknowledgment. This work was supported by the National Institutes of Health (GM27681 to P.B.D. and A.Z.A). We thank the Beckman Institute Sequence Analysis Facility for DNA sequencing.

References

- (1) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–99.
- (2) Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C.; Dervan, P. B. *Tetrahedron* **2007**, *63*, 6146–6151.
- (3) Bulyk, M. L. *Methods Enzymol.* **2006**, *410*, 279–99.
- (4) Tse, W. C.; Boger, D. L. *Acc. Chem. Res.* **2004**, *37*, 61–9.
- (5) Warren, C. L.; Kratochvil, N. C.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N., Jr.; Ansari, A. Z. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 867–72.
- (6) Puckett, J. W.; Muzikar, K. A.; Tietjen, J.; Warren, C. L.; Ansari, A. Z.; Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 12310–9.
- (7) Burnett, R.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11497–502.
- (8) Dervan, P. B.; Urbach, A. R. In *Essays in Contemporary Chemistry*; Quinkert, G., Kisakürek, M. V., Eds.; Verlag Helvetica Chimica Acta: Zurich, 2000, p 327–339.
- (9) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4343–8.
- (10) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. *J. Mol. Biol.* **2002**, *320*, 55–71.
- (11) Marques, M. A.; Doss, R. M.; Urbach, A. R.; Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485–4517.
- (12) Arndt, H.-D.; Hauschild, K. E.; Sullivan, D. P.; Lake, K.; Dervan, P. B.; Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322–13323.
- (13) Singh-Gasson, S.; Green, R. D.; Yue, Y.; Nelson, C.; Blattner, F.; Sussman, M. R.; Cerrina, F. *Nat. Biotechnol.* **1999**, *17*, 974–8.
- (14) Bailey, T. L.; Elkan, C. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*; Altman, R., Brutlag, D., Karp, P.,

- Lathrop, R., Searl, D., Eds.; AAAI Press: Menlo Park, 1994, p 28–36.
- (15) Workman, C. T.; Yin, Y.; Corcoran, D. L.; Ideker, T.; Stormo, G. D.; Benos, P. V. *Nucleic Acids Res.* **2005**, *33*, W389–92.
- (16) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146.
- (17) Trauger, J. W.; Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450–66.
- (18) Colantuoni, C.; Henry, G.; Zeger, S.; Pevsner, J. *Bioinformatics* **2002**, *18*, 1540–1.
- (19) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, *19*, 185–93.