Chapter 2

SELECTIVITY OF PY-IM POLYAMIDES IN TISSUE CULTURE

ABSTRACT

Py-Im polyamides have excellent sequence specificity *in vitro*, yet little is known about their selectivity in the nuclei of mammalian cells. In this chapter the extent of the functional selectivity of polyamides is assessed in regulation of gene expression in Glucocorticoid signaling. First, mathematical modeling was used to find the most common GRE sequences that can be bound with 8-ring hairpin polyamides. Then a panel of 12 genes and a focused library of polyamides targeting 7 DNA different sequences was used in evaluation of polyamides as a tool for linking sequence of a response element with the gene it controls. Concurrent nuclear localization studies and in-vitro assessment of DNA binding affinity were performed on the library of polyamides to connect chemical properties of polyamides with their gene regulation patterns. Polyamides show a small degree of selectivity; however, the differences are hard to elucidate because of the low potency of some of the compounds. The potent compounds, on the other hand, show few differences in gene expression patterns. Further steps will need to be taken to increase polyamide specificity, without sacrificing potency; in particular more genes may need to tested, e.g. by using RNA-sequencing. Another possibility is using multiple compounds to target the same regulatory sequence and thus increase the specificity of Py-Im polyamides in tissue culture.

Introduction

Binding of Py-Im polyamides to DNA is sequence-specific (1, 2). While in vitro experiments have shown that a single sequence can be targeted, achieving site-specificity in a mammalian cell nucleus is a significantly more challenging task. The main problem in sequence specificity in mammalian cells is the sheer amount genetic material enclosed in the nucleus. For example, DNA in human cells contains $3x10^9$ base pairs, and a 6-base pair sequence would be expect to occur once in every 4^6 bases, assuming every base pair can be recognized. The typically used Py-Im polyamides, however, only recognize between 3 different base pairs, G, C, and W, which means an average frequency of the DNA sequence bound specifically by an 8-ring Py-Im polyamide is expected to be once in every 3^6 basepairs, an equivalent 4.1 million matched binding sites for an average Py-Im polyamide. Another factor present in mammalian cells, but not *in vitro,* is accessibility of DNA in the nucleus. Not every site in the genomic DNA is equally accessible; some of the DNA is densely packed as heterochromatin. It is currently unclear how the binding properties of polyamides change depending on the density of DNA-packing in the nuclei; however, we do know that Py-Im polyamides are capable of binding to nucleosomes (3). Finally, the time of dissociation of a commonly use hairpin polyamide and DNA match site is long (k_{off} $= 10^{-3} – 10^{-4}$ s^{-1}), half the time of dissociation which ranges from minutes to hours (4), which limits diffusion of polyamides within the nucleus. Thus, sequences most frequently bound by Py-Im polyamides might simply be those that are most accessible thanks to diffusion, or DNA packing. Recent experiments evaluated some aspects of the selectivity of Py-Im polyamides in the genome, showing that sequence-specificity might be just one factor in their genomic-DNA binding profile and the chromatin accessibility may also be important (5). In this chapter we investigated the selectivity of Py-Im polyamides in living cells, by testing the expression of a number of genes related to Glucocorticoid Receptor (GR) in the A549 cell line. We also built theoretical kinetic models of DNA-polyamide binding and calculated possible sequence specificities of Py-Im polyamides within the genome.

Background

Mammalian genes are regulated thanks to a complex network of transcription factors (6) and proteins regulating chromatin accessibility (7, 8). How transcription factors bind and control

gene expression is one of the main questions in molecular biology. Investigating the DNA sequence binding to transcription factors historically has been done through DNA-sequencing of the purified DNA bound to transcription factors (9), and subsequently by Electrophoretic Mobility Shift Assay (EMSA) (10) and DNAseI footprinting (11). These methods allowed for study of a single transcription factor binding site at a time. As a result they yielded information about binding affinity between a transcription factor and a DNA sequence, but failed to inform us about the genomic frequency and positions of these sites. It was not until the advent of high throughput genome sequencing and microarray technology that we were able to do this. Currently the most common method of determining transcription factor binding sites is Chromatin Immunoprecipitation followed by sequencing (12) (Chip-Seq), which can inform us about the position of both genes and transcription factor binding sites in the whole genome. Regrettably this method is incapable of establishing a functional link between the transcription factor sites and the genes they control. While in prokaryotes the transcription factors bind proximally to the genes, in mammalian cells this is not always the case (13, 14).

Large distance between regulatory sequences and their gene targets poses a challenging problem in identifying a functional link between them. Currently there are three Chromosome Conformation capture (3C, 4C, 5C) methods that allow one to connect the particular regulatory element with a particular gene (15-17), however the execution of these assays is often complicated (18, 19). Additionally no other method exists that could confirm the findings, and suggested problems with these methods remain untested (18). Consequently, as of now, reliably matching a transcription binding site to the gene it controls requires knocking out the regulatory sequence in cells. Unfortunately, this method requires prior knowledge of both the gene and its' regulatory sequence and many genes are controlled by multiple regulatory sequences. Because of those issues, targeted knockdown is unsuitable for genome-wide mapping. Pyrrole-Imidazole polyamides could be useful in relating regulatory DNA sequences with the gene expression patterns in a high throughput fashion.

Py-Im polyamides bind the minor groove in double-stranded DNA with affinities and specificities comparable to transcription factors. It has been achieved by combining aromatic amino acids, N-methylpyrroles (Py), N-methylimidazoles (Im), and 3-hydroxy-1-methylpyrroles

(Hp), in a short oligomer. A pair of monomers placed over each other in a minor groove can determine pairing rules between polyamides and DNA. According to those rules a Py/Im pair will recognize a C°G pair, Im/Py a G°C, whereas Py°Py pair will be capable of recognizing W°W pairs. Including a hydroxypyrrole instead of a pyrrole will bias binding of a polyamide towards T°A in the case of a Hp/Py pair and A°T in the case of a Py/Hp (1, 20) (Figs. 1.2 and 1.3). Their capability of sequence specific displacement of transcription factors from their binding sites results in an inhibition of gene expression establishing a functional link between regulatory sequences and the genes they control. However, the question that needs to be answered is whether their sequence specificity is high enough for sequence-specific gene regulation in a large mammalian genome.

Glucocorticoid receptor (GR) response pathway is a common model system for gene expression regulation in mammalian cells (7, 14). GR is a cell permeable steroid receptor binding directly (21) to a well-defined Glucocorticoid Response Element (GRE) (Fig. 2.1). There are several thousand GREs scattered across the mammalian genome implicated in gene regulation (14), each containing three highly degenerate nucleotides at positions 7,8, and 9 (Fig. 2.1B,C). These three nucleotides alone constitute 64 distinct classes of GREs that can be targeted with sequence-matched pyrrole-imidazole polyamides developed in our lab. Additionally, the other bases also show sequence variability that can be utilized for that purpose. In an observed case of one GRE driven gene (GILZ) displacement of transcription factors through polyamides results in an inhibited gene expression (22), and if that will be the case with other GRE controlled genes, we will be able to match classes of GREs to the genes they control.

Glucocorticoid receptor pathway is a useful drug target. Glucocorticoids are widely used in medicine as immunosuppresants and are some of the most potent anti-inflammatory drugs on the market (23). These effects, however, come at a price. Glucocorticoids have significant side effects, such as bone and muscle loss, psychoses, cataract and glaucoma, among many others (23). In children, prolonged use of glucocorticoids may negatively affect bone development (24). Many of those side effects, e.g. glaucoma or diabetes, are mediated through transactivation, or expression of anti-inflammatory proteins. This fact has galvanized the development of more selective glucocorticoid receptor agonists (SEGRas) which aim at decreasing transactivation

without affecting transrepression. One SEGRA (ZK 216348) has shown in animal models that the negative side effects of glucocorticoid treatment can be reduced while maintaining anti-inflammatory effects (23).

Likewise, polyamide are effective in downregulation of gene expression. While the majority of side effects of glucocorticoid treatment are due to transactivation, some of them are not. Thus targeting different sequences within GREs by polyamides administered along with standard glucocorticoids can fine-tune the effects of this anti-inflammatory treatment to minimize the side effects and maximize potency for the specific disease.

Evaluation of the genomic landscape of the GREs and polyamide binding sites

Chromatin Immunoprecipitation followed by sequencing (Chip-Seq) identified 4392 loci in the genome that are occupied by GR in human lung adenocarcinoma cells (A549) (14). The positions of these loci are, however, distant from 234 genes that are highly induced upon treatment with 100nM Dexamethasone (Dex), a synthetic agonist of GR. For genes with Dex induced expression, the median distance between the nearest GRE and a transcription start site (TSS) was 11kb, and those genes that were repressed had a median distance of 146 kb (14). The large TSS-GRE distance and its significant variability suggests that one cannot predict which genes are controlled by which GREs based solely on their relative position. The response of the genes also varies in time; particularly repressed genes are affected later in time than induced ones. This and the large distance between nearest GREs suggest that repressed genes are not controlled by promoter-proximal GR binding.

In order to further analyze the dependence of position of GREs and transcription start, I wrote a simple mathematical model assuming their random distribution. For downregulated genes, I generated random locations for both TSS and GREs and then measured their distance in the aproximately 2.1×10^9 basepairs in non-repetitive parts of the human genome (25). Even this crude estimate of the genome size and a very basic model gives a median nearest neighbour distance between TSS and GREs of 164kb, as compared to 146kb in Chip-Seq study (14). This result suggests that the position of GREs and the genes they repress are independent of each other. In this case a common assumption that the gene is controlled by its nearest neighbour is

most likely unfeasible, further suggesting that repressed genes are controlled independent of proximal GRE-promoter binding. The activated genes, on the other hand, show dependence of the position of GREs and TSSes.

I modeled their relative positions by generating set of gene positions and a random distribution of distances (GRE positions) over the mean length of a chromosome (123kb). It appears that the median distance between a TSS and the nearest neighbor (10.7 kb), assuming their random distribution, once again is very close to 11kb, as found by Chip-seq (14). The distribution of distances as modeled also matches the Chip-Seq data (Fig. 2.2). This result suggests that distance between GREs and genes they control may be distributed randomly within the chromosome. This model further supports current belief that Glucocorticoid Receptor signaling occurs through an exceptionally long range interactions (14). The code and parameters used in writing the models can be found in appendix G. Such quantitative considerations show that one cannot assume that a position of a GRE relative to TSS can predict a functional link between the two and Chromosome Conformation Capture methods maybe be necessary to establish such a link.

In order to make an informed decision on which compounds should be synthesized to exert a specific control of gene expression in A549 cells, I analyzed the GR binding sites for enrichment upon Dex induction in Chip-Seq data set from Myers lab (14). If a rare sequence is targeted with a polyamide, it is unlikely that a large number of genes will respond to it. If, on the other hand, a compound binds a wide array of sequences a larger fraction of genes in a panel is expected to be downregulated. In order to establish which sequences are most common among active GREs, I chose to computationally analyze the genome-wide occurence of sequences compatible with DNA-binding profiles of 8-ring hairpin polyamides.
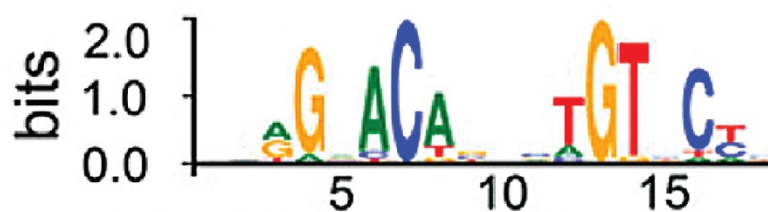
A
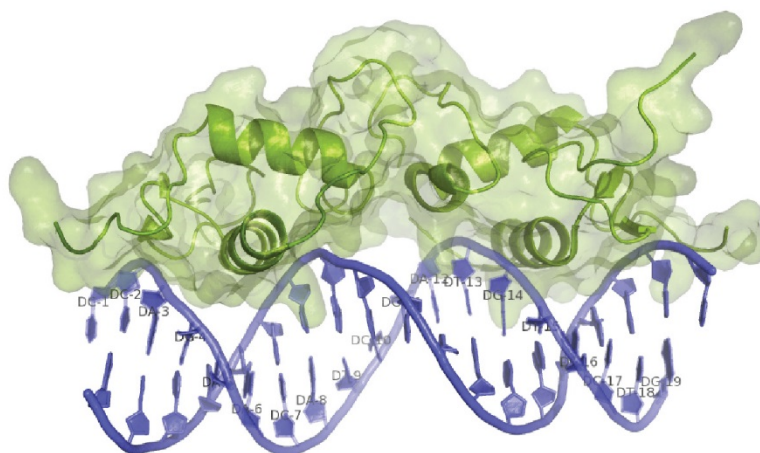
B
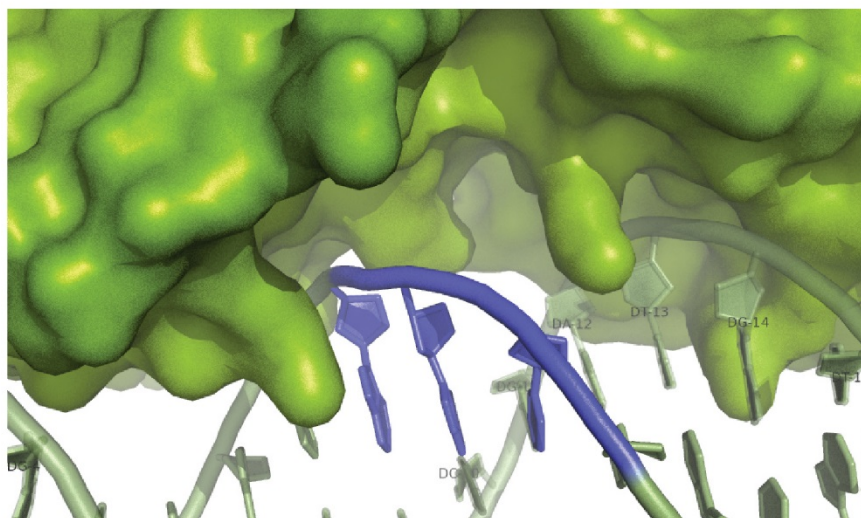
C

Figure 2.1 X-ray crystal structure of a Glucocorticoid Receptor (GR) bound to DNA (PDB 1R4O). (A) GR binds two DNA as a dimer. Its recognition sequences are nearly palindromic and are separated by a 3-base-pair gap, colored blue on the second inset (B). This gap corresponds to an area without physical GR-DNA interaction. (C) GRE binding motif obtained through a custom analysis of GR Chip- Seq data (14). The sequence variability of this motif allows for sequence-specific targeting of subsets of Glucocorticoid Response Elements (GREs).
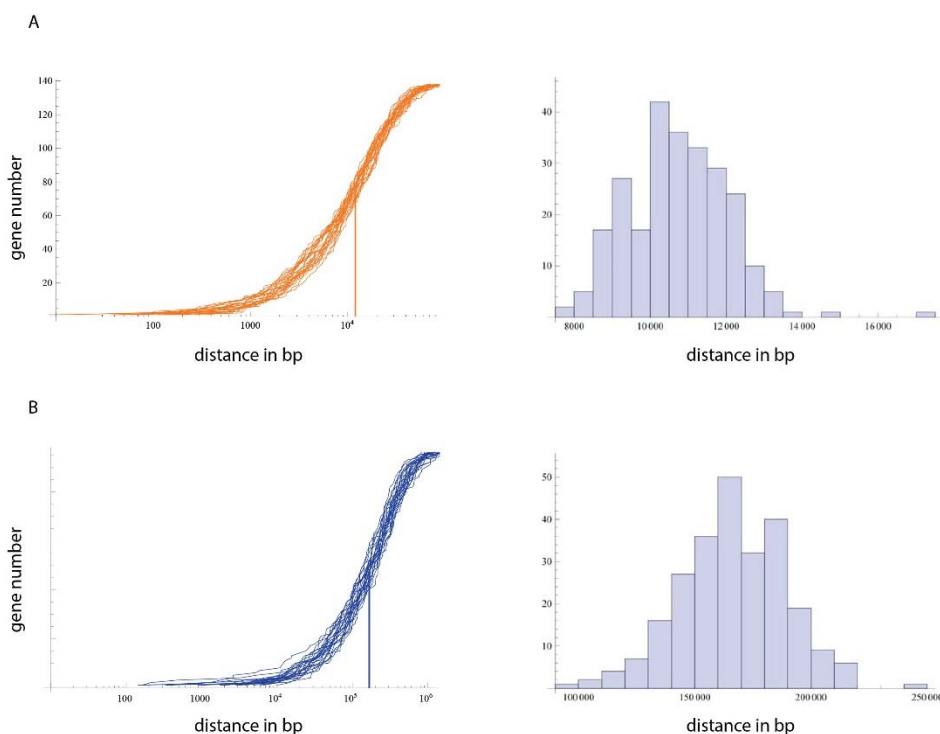
Figure 2.2 Modeling genomic distribution of GREs in relation to transcription starting sites (TSS). (A) modeling distance between GREs and Dexamethasone upregulated genes by placing GREs and TSSes at random within the average size of a human chromosome (123kb$_2$) yields comparable distribution and median distance (10.7kb, n=250) as observed by Chip-Seq (11kb). (B) Modeling distance between GREs and dexamethasone repressed genes by distributing the TSSes and GREs randomly across the whole genome yields similar distribution and median distance. (146kb for ChipSeq, versus 166.6kb for random distribution model, n=250).

The top 100 most enriched regions were scanned for a GRE consensus sequence (Fig. 2.3A), which yielded practically an identical motif as found by Chip-Seq (Fig. 2.3B) (14). I then extracted sequences with 95% homology to the GRE consensus sequence from the most enriched regions (fig 3a) and obtained 405 sequences. Using custom scripts (code in appendix D) I analyzed the frequency of motifs that can be targeted with Py-Im polyamides (Fig 2.3C). The most common sequence can be targeted by a polyamide used previously in our lab (**1**, targeted to 5'-WGWWCW-3') both *in-vitro* (26) and in gene regulation studies (22, 27). The second and third most common are targeted by the same polyamide (**2**, targeted to 5'-WGGWCW-3'), a sequence that also has gene targeted in our group previously (26). The fourth sequence (**3**, targeted to 5'-WWCWGW-3') has not been yet tested. The orthogonality of binding of polyamides **1**-**3** (Fig. 4) was determined by GRE sequence analysis, based on the previous ChIP-seq experiments (14). Comparing these three polyamides in gene regulation studies will narrow our focus to the most commonly found sequences that can be bound by 8-ring hairpin polyamides. The three

compounds can bind different sites within a GREs (Fig. 4a) and some of those sites are more conserved than others – in particular the bases 7-9, show a particularly high variability. This analysis informed the decision on which compounds should be synthesized, to exert a specific control of gene expression in A549 cells. The methods developed allow to perform this analysis for other systems, cells and polyamides.
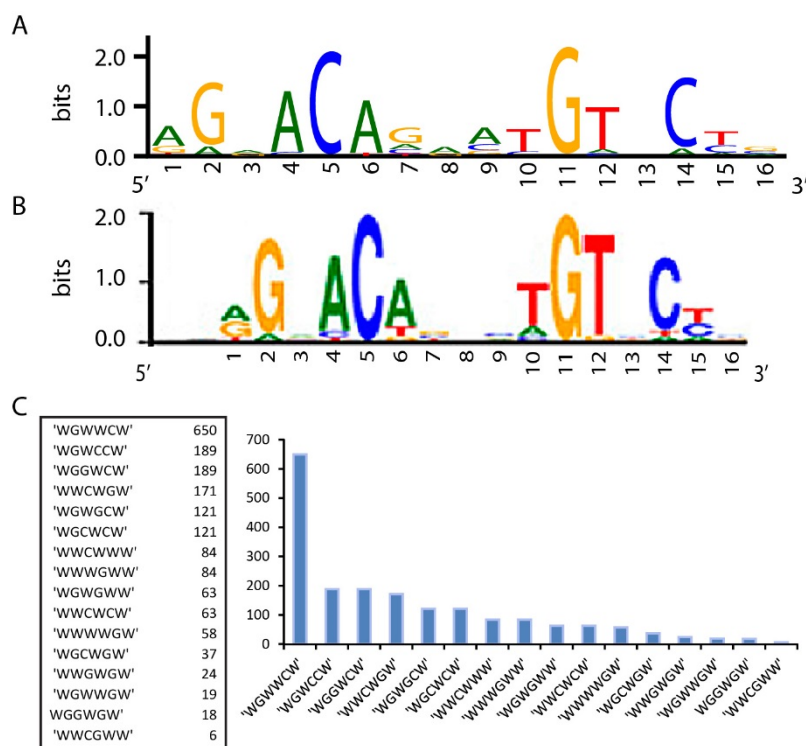

Figure 2.3 Characterizing DNA sequences binding GR. (A) Top 100 most enriched regions in Dex induced samples returned a consensus sequence that is practically identical to one obtained from uninduced cells (B). (C) The frequency of 6-basepair sequences that can be targeted by polyamides reveals WGWWCW is the most common motif among 405 GRE in a 100 regions most enriched upon Dex treatment.

Selectivity of polyamides in A549 lung adenocarcinoma cells: Gene regulation studies

The selectivity of Pyrrole-Imidazole polyamides has been tested rigorously *in vitro* (1, 28, 29); however, many questions need to be answered in the case of polyamide selectivity in cells. In order to address this issue I began gene regulation studies in A549 lung adenocarcinoma cells used previously in gene regulation studies with Py-Im polyamides (22). Since little is known

about a functional link between TSSs and GREs, I decided to investigate effects of polyamides
on

A

5'A**GAACA**GAATGTCCT 3'

1

5'A**GGACA**GCCTGTCCT 3'

2

5'A**GAACA**GAACGTTCT 3'

3

B

C

		1 GWWC	2 GGWC	3 WCWG	4 WCWW	5 WCWC/GWGW
A	GWWC	650	64	85	60	44
B	GGWC	64	189	37	13	24
C	WCWG	85	37	171	43	14
D	WCWW	60	13	43	84	3
E	WCWC/GWGW	44	24	14	3	63

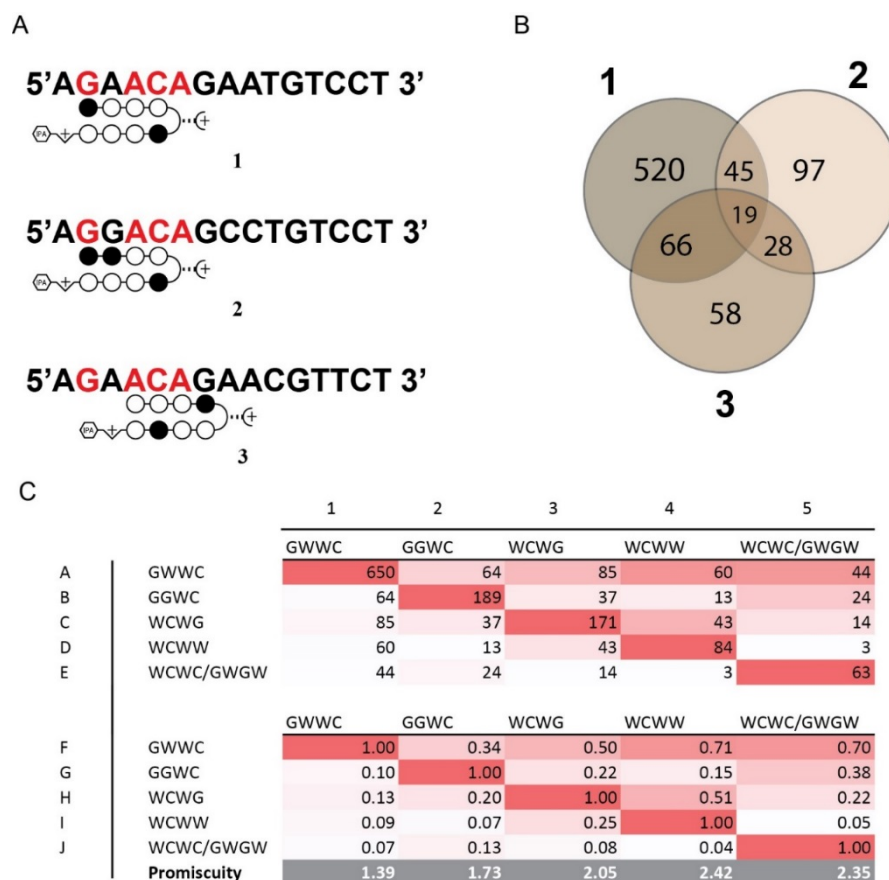|		1 GWWC	2 GGWC	3 WCWG	4 WCWW	5 WCWC/GWGW
F	GWWC	1.00	0.34	0.50	0.71	0.70
G	GGWC	0.10	1.00	0.22	0.15	0.38
H	WCWG	0.13	0.20	1.00	0.51	0.22
I	WCWW	0.09	0.07	0.25	1.00	0.05
J	WCWC/GWGW	0.07	0.13	0.08	0.04	1.00
	Promiscuity	1.39	1.73	2.05	2.42	2.35

Figure 2.4 Orthogonality and targeting sites of three polyamides recognizing the most GREs according to the data in Figure 8. (A) **1** and PA2 bind nucleotides 1-6 differ in the 3rd base of the GRE. Polyamide **3** binds a different site and shows variability in 6th and 7th bases. (B) **1** targets most sequences of the three polyamides. **3** targets mostly a subset of the sequences that are also targeted by **1**, and **2** binds mostly orthogonal sequences. (C) Polyamide specificity table shows orthogonality for other polyamides used in the study. Entries on the diagonal represent absolute number of match sequences for each polyamide. For example, entry A1 shows there are 650 WGWWCW binding sites in the tested Chip-Seq regions. The numbers of the diagonal represent a subset of GREs that can bind two different polyamides. For example, entry B2 shows there are 64 GREs that can bind both WGWWCW and WGGWCW polyamides. The bottom table (entries F1-J5) summarizes the relative promiscuity of each polyamide in the study. Each column in the top table (entries A1-E5) has been normalized to the entry on the diagonal. Promiscuity coefficients have been obtained by summing every entry in the column.

expression of a panel of genes significantly induced by GR agonist Dexamethasone (Dex), thus yielding distinguishable changes in GR-driven gene expression. According to the current models of gene expression in GR system, each of these genes should be regulated by a single or small number of GREs (14). Even though I did not know their sequences, I knew the distribution of GRE sequences genome-wide. Assuming perfect sequence specificity of polyamides, we should be able to elucidate the sequences of those GREs by observing the patterns of gene expression inhibition in a randomly selected subset of genes. I began with testing Dexamethasone induced genes identified by microarray (30) and RNA-sequencing (14). A panel of 17 genes was tested using quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR). However, four of these genes were not upregulated significantly (Fig. 5).
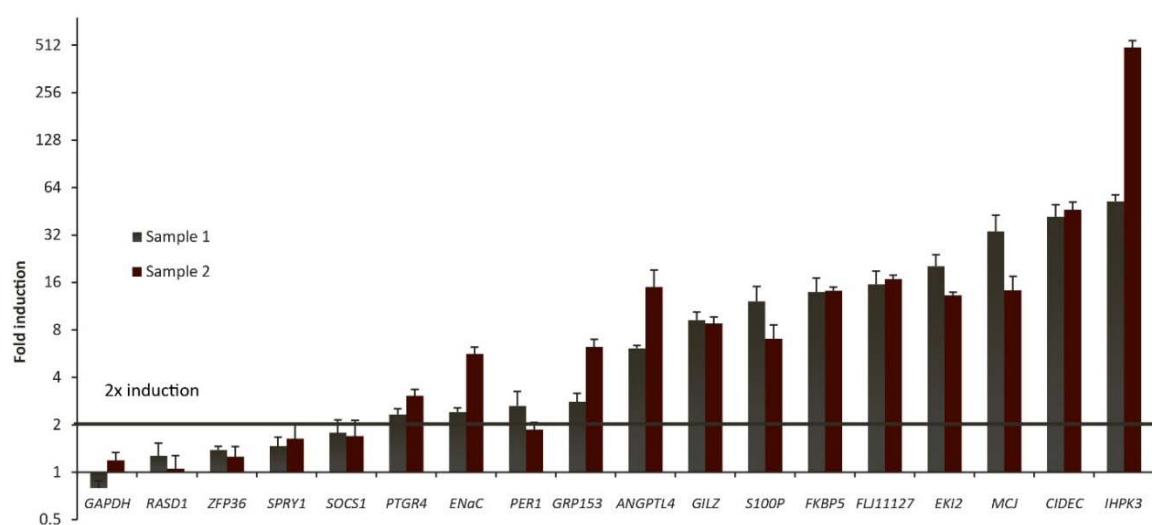

Figure 2.5 Analysis of levels of expression induced by Dexamethasone. The levels of expression were obtained by RT-qPCR. The fold induction values were obtained by dividing levels of mRNA expression obtained for Dex induced samples by uninduced ones. Genes in this panel were identified previously from microarray[7] and RNA-sequencing studies[8]. Twelve Genes that were induced at least two-fold were used for the further studies. GAPDH is a housekeeping gene and acts as a negative control. Each samples has been normalized to expression to a housekeeping gene (GUSB).
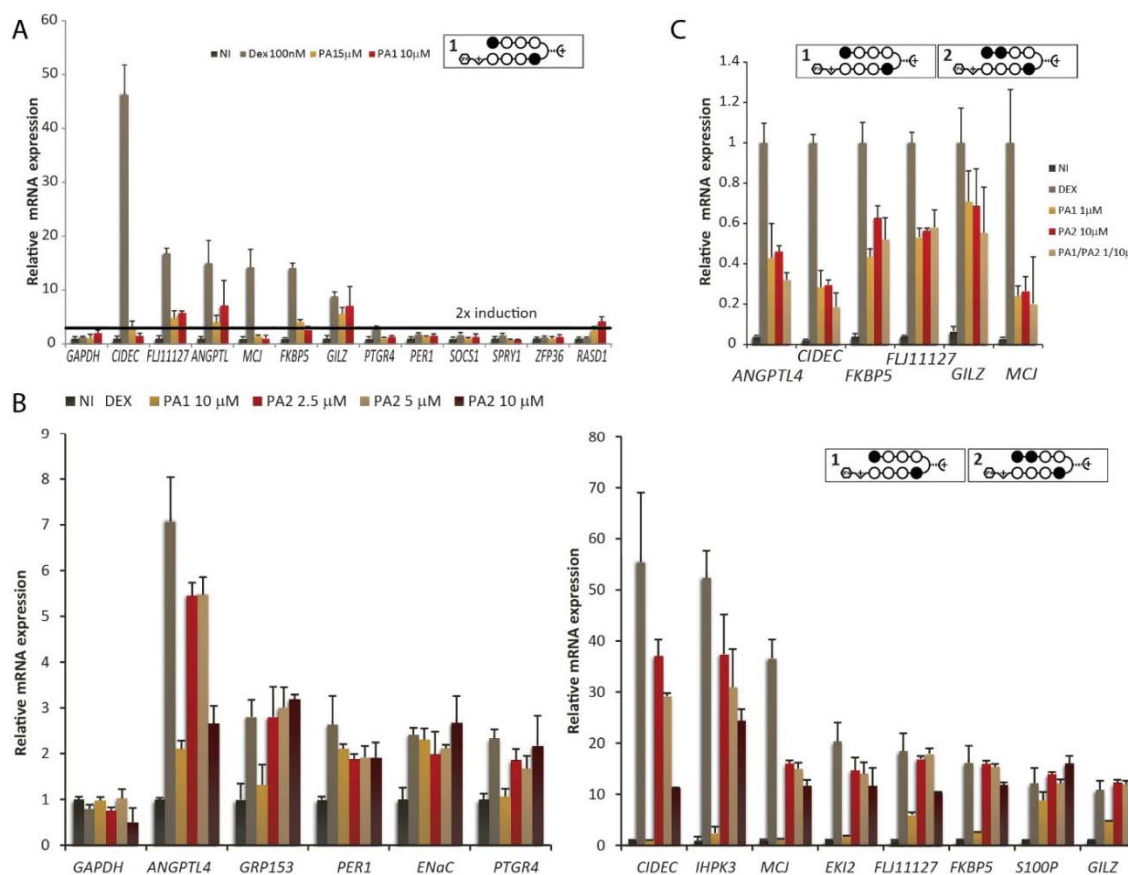
Figure 2.6 Analysis of levels of expression of genes induced by Dex and inhibited by polyamides 1 and 2. (a) Exploratory study showed that well induced genes are strongly inhibited by 1. (b) Polyamide 1 is significantly more potent than 2, but both of them downregulate the same genes. (c) Polyamide 2 is approximately 10 times less potent than 1, but the polyamides downregulated the gene expression in the same way. The correlation between the fold-induction and fold-inhibition was 0.86 for 1 and 0.74 for 2, suggesting relatively non-specific inhibition of gene expression.

I followed with the panel of 13 significantly upregulated (at least 2-fold) genes to measure the effect of polyamides on their expression. The timecourse and initial dosing was consistent with our previous studies with compound **1** in A549 cells (22). In short, the protocol included plating 12000 A549 cells/cm2 in 12 or 24 well plates for 24 hours in F12-K medium supplemented with 10% FBS, then the cells were washed twice with 1x PBS and the medium replaced with F-12K medium supplemented with 10% Charcoal Treated (CT) FBS, including the desired concentration of polyamides. After 48 hours of incubation, 100 nM dexmethasone was added directly to the medium for 6 hours, after which cells were harvested for RNA extraction. In

order to test this dosing on various genes I run a limited test on 11 genes, both upregulated and unaffected, to see if polyamides downregulated either of them (Fig. 2.6A). Confirming that all uninduced genes, with an exception of SPRY1, were unaffected by polyamide **1**, while all well induced genes were inhibited by it, I decided to compare the gene regulation capability of polyamides **1** and **2** (Fig. 2.6B).

Polyamide **1** (targeting 5' WGWWCW 3') downregulated the expression of all Dex induced genes, except two: S100P and ENaC. One of the genes, PER1, was downregulated slightly (approximately 30%) and 6 other genes, FLJ11127, FKBP5, GILZ, ANGPTL4, IHPK3, and PTGR4, were downregulated at least two-fold. Polyamide **1** at 10 μM had also completely abolished the effects of Dex in the remaining four genes, CIDEC, GRP153, MCJ and EKI2. This widespread action of polyamide 1 was not unexpected, given that the motif it targets is very common (Fig 2.3C), for the same reason one would expect polyamide **2** to target only a subset of genes, or set of genes that is different from the one downregulated by **1**. However, the dose response of polyamide **2** (Fig. 2.6B) suggests that polyamide **2** targets the same sequences, although with less potency. The same genes whose expression was most downregulated by polyamide 1, are also downregulated by polyamide **2**, albeit to a lesser extent. By running a series of exploratory experiments I was able to determine that potency of polyamide **2** is comparable to 10-fold lower concentration of polyamide **1** added to the cell media. Dosing cells with 1 μM polyamide **1** and 10 μM polyamide **2** yielded identical reponses (Fig. 2.6C), for the panel of 6 genes that were significantly downregulated in the previous experiment. It is not possible to tell whether this widespread response of both genes is due to non-specificity of polyamides in the cell nucleus, or because each of the genes inhibited happend to be regulated by several GREs containing both 5'WGWWCW3' and 5'WGGWCW3' motifs. However, when both polyamides were dosed at the same time, there was no synergistic effect (Fig. 2.6C) suggesting non-specific polyamide binding as a culprit. The extent of polyamide-mediated gene expression downregulation was correlated with the fold-induction with 100 nM dexamethasone (0.86 for **1** and 0.74 for **2**). This suggests that the most induced genes are the ones most affected by polyamides, possibly regardless of their sequence. No rigorous test exists as of now to determine the reasons for this high correlation.

Next, I tested the specificity of polyamides targeting 7th basepair in GRE motif. At first, I decided to measure the effects of polyamide **3** on the gene expression because the sequence it targets is approximately as common as for **2**. Despite following the same treatment as for compounds **1** and **2**, only two genes, ANGPTL4 and CIDEC, were affected by **3** (Fig. 2.7A) at low concentrations 2.5 μM or 5 μM. Bringing up the concentrations of these two genes showed further inhibition: ANGPTL4 was inhibited by over 50% and CIDEC by over 40%. without affecting four other highly induced genes: FKBP5, FLJ11127, GILZ, and MCJ (Fig 2.7B). Further increase in polyamide 3 concentration did not increase polyamide potency significantly, potentially because of instrument noise or polyamide solubility problems (Fig. 2.7C). Thus 10 μM concentration is either the most effective, or nearly the most effective in gene downregulation. At this concentration, the correlation between fold-induction and fold inhibition was low at 0.25, suggesting that polyamide **3** targets genes more independently of their induced activity than **1** or **2**. In an effort to improve potency of 3, I followed with its acetylation (31). Previous experience in the group suggested that this modification can improve gene downregulation (32). However, in the case of acetylated polyamide 3 (12) the gene downregulation profile was identical in both selectivity and potency.

The specificity of polyamide 3 suggested synthesis of other compounds targeting the 7th base in the GRE with N-methylpyrrole at the last position (cap). I expanded the library of compounds to target these sequences. The next sequences most commonly found in GREs can be targeted by polyamides **4** and **5** (Fig. 2.8), namely, sequences 5'WWCWW3' and 5'WWCWCW3', respectively. These compounds as well proved to be less potent than **1** or **2** but also more selective in gene downregulation. Compound **4** caused downregulation of three genes at two-fold or more: ANGPTL4, CIDEC, and MCJ (Fig. 2.9). Two of these genes, ANGPTL4 and CIDEC, were also downregulated by **3**; however, MCJ was not. Thus there is a distinguishable difference in downregulation profile between 3 and 4, even though both of them target the GRE consensus sequence. It is possible that this effect is due to DNA-binding independent events, and thus it is not yet clear if their differences in gene downregulation are due to differences in sequences of GREs influence on the gene downregulation patterns. Polyamide 6 is a potent compound with little controlling each of these three genes. It is, however, a useful feature of

these compounds and a step towards making polyamides that downregulate small subsets of genes.

Compound **5** targets a less common sequence (5'WWCWCW3') that occurs three times less commonly in the assessed region than 5'WGGWCW3' or 5'WWCWGW3' and 25% less commonly than 5'WWCWWW3'. It also happens to be even less potent than 3 and 4. Similarly to these compounds it downregulates ANGPTL4 and CIDEC, however, at 10 mM only by approximately 1.6- and 2-fold, respectively. Interestingly two other genes were downregulated slightly (albeit within the noise of the method) and showed a dose response: MCJ and FKBP5 (Fig. 2.10A). Higher concentration dosing, or additional modification to the compound 5 will be necessary to improve this potency and establish whether downregulation of these two genes is specific.

In order to distinguish whether the effects of gene expression are solely based on the sequences targeted by polyamides, or if the structural properties of the polyamide itself also matter, I decided to examine the gene regulation capabilities of polyamide **6**, which targets the same sequence as **5**, but has a different structure. It appears that the structure of the polyamide has a significant on gene expression: compound **6** affects a larger number of genes at 10 μM (Fig 2.10B), whereas polyamide **5** is more specific, but not as potent (Fig 2.10A). The former downregulated genes in a similar fashion as other N-methylimidazole capped compounds: most genes are downregulated to nearly their baseline levels, and thus compound 6, at 5 μM and 10 μM, mimicks the downregulation patterns of compound **1** (Fig. 2.10B).
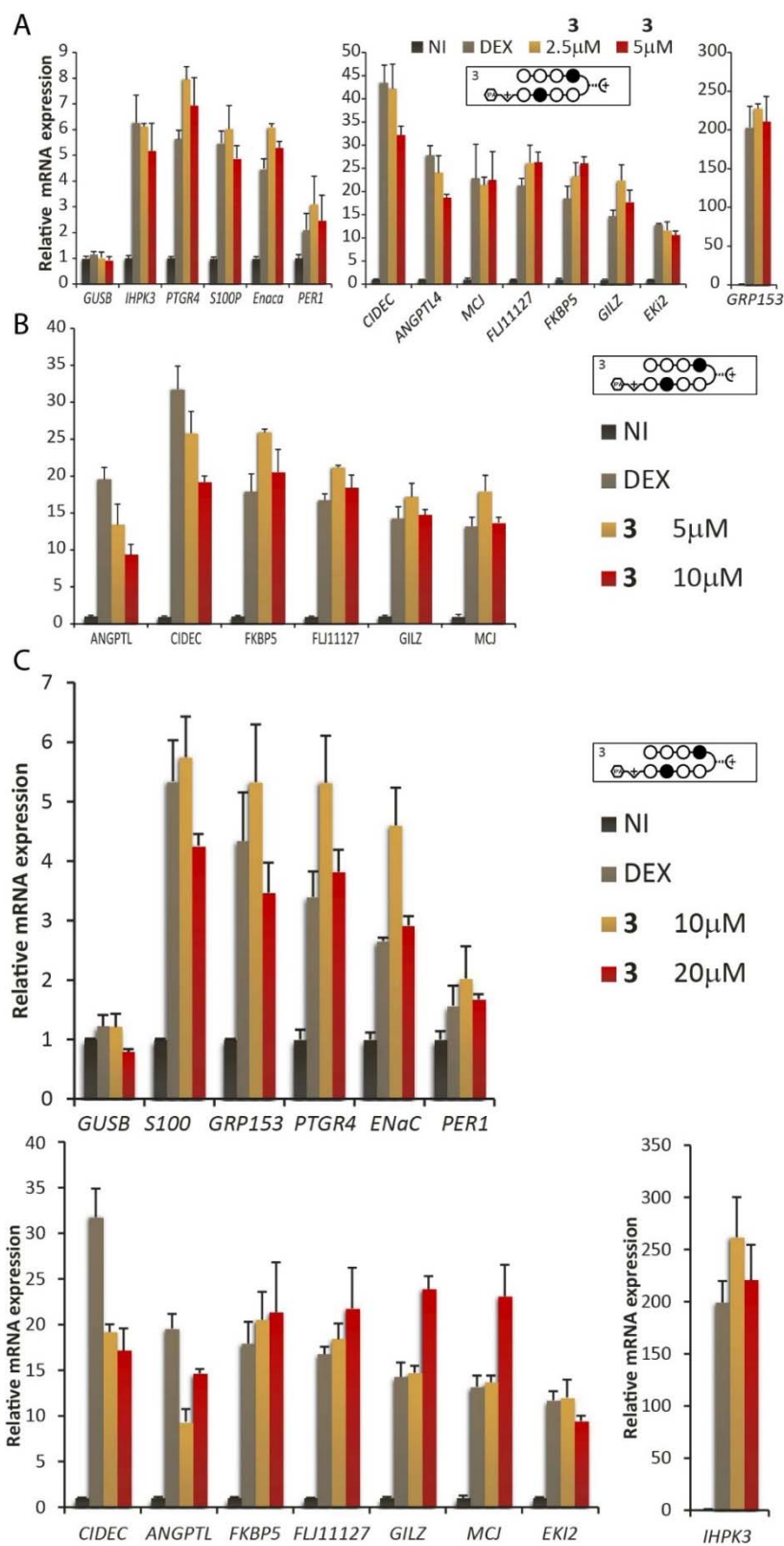
Figure 2.7 Analysis of levels of expression of genes induced by Dex and inhibited by polyamide 3. (a)At 2.5 mM and 5 mM only two genes are affected by polyamide 3. The low concentrations have been used to promote selectivity at a cost of potency. (b) Increasing the concentrations to 5 mM and 10 mM increases potency of polyamide 3 without sacrificing selectivity. Only two genes, that were affected at lower concentrations are inhibited by 3. The remaining four are negative controls and were not affected. (c) Adding even more elevated concentrations of polyamide 3 does not increase potency significantly. This could be due to a cytotoxic effect or polyamide insolubility. However, if the changes in relative mRNA expression are small, they might also not be within the instrument's sensitivity. Unlike in case of polyamides 1 and 2, the correlation between fold-induction and fold-inhibition is low at Rsq=0.25 suggesting that action of 3 is more independent on activation levels, and less on the specifics of the gene.
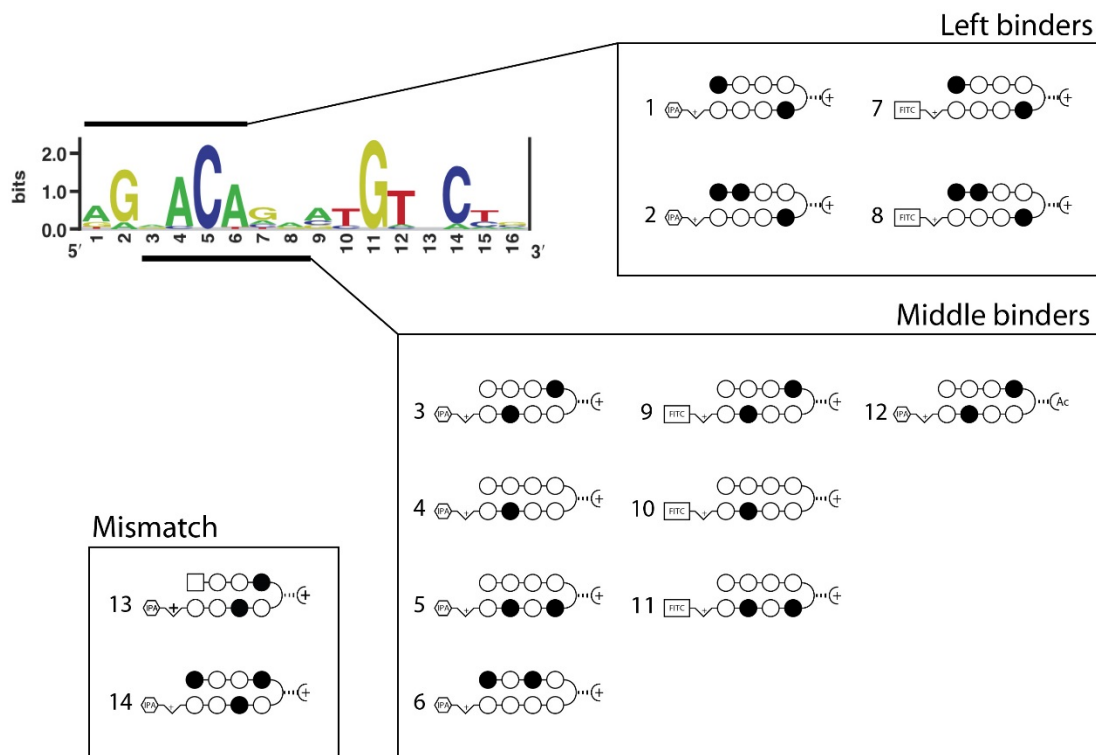
Figure 2.8 Library of synthesized polyamides. The 'left binders' include polyamides spanning bases 1 through 6 of GREs with a 3rd variable base. 'Middle binders' target bases 3-8 and vary at 7th nucleotide in the GREs. Structures of polyamides drawn in Appendix A.
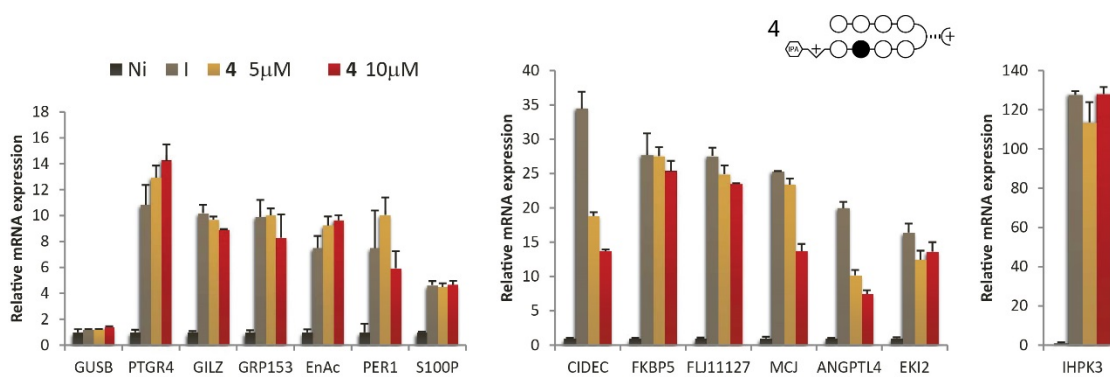

Figure 2.9 Inhibition of a panel of Dex induced genes by polyamide 4 (targeting 5'WWCWWW3' DNA sequence). The polyamide downregulated the three genes two-fold or more at 10 mM. Of those genes CIDEC and ANPGTL4 were downregulated by 3, but MCJ was not. Thus there is a difference in specificity of gene downregulation between 3 and 4.
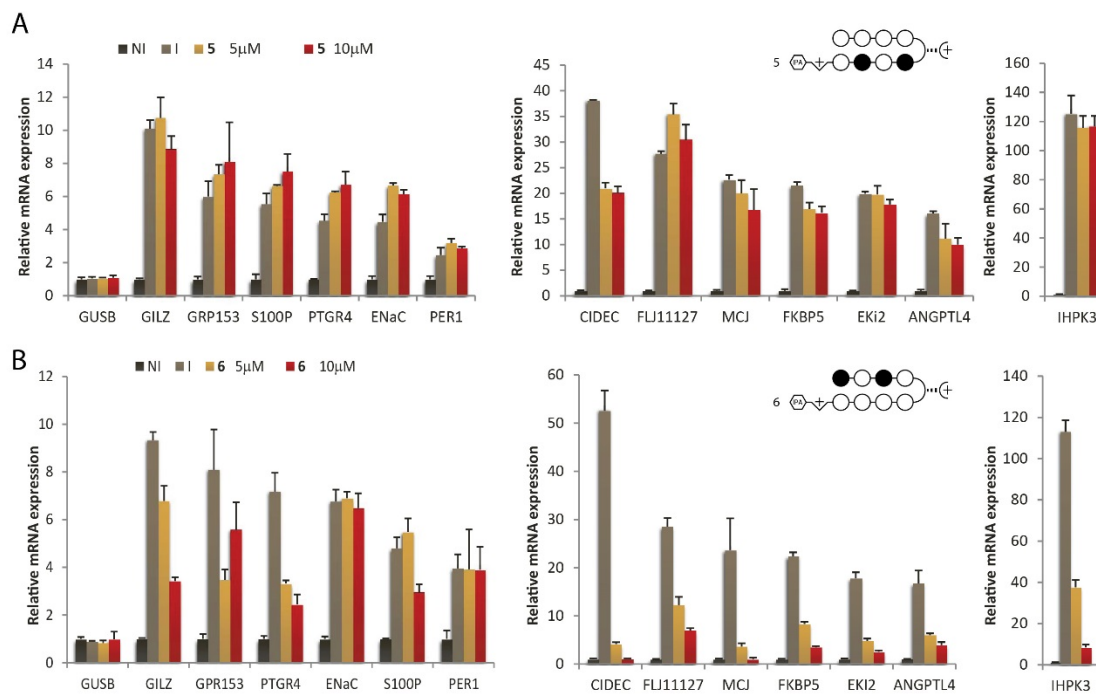
Figure 2.10 Downregulation of Dex induced genes using polyamide targeting the same sequence, but with different structures. (A) n-methylpyrrole capped polyamide, 5, downregulates the genes slightly, but selectively. (B) n-methyl-imidazole capped compound 6 downregulates the genes efficiently, but not selectively.

A negative control of sequence-specific gene regulation could be compounds targeting sequences that are not found within GREs. For example, a 5'WTWCGW3' sequence does not appear in 405 GRE motifs most responsive to Dex (Appendix E, Fig. 3C) and it can be targeted with compound **13**. In tissue culture studies **13** downregulated all genes but two whose expression was unaffected by polyamides **1** and **2** as well: S100P and ENaC. At 10 µM the most genes were downregulated to basal level. The exceptions: IHPK3, GRP153, and GILZ, were also significantly downregulated (Fig. 2.11). At 5 µM the gene downregulation effects were less prominent: most of the genes were downregulated to 4-fold the baseline expression level and two of them, IHPK3 and PTGR4, were downregulated to 10- and 2-fold the baseline. One gene, PER1, was unaffected. As a result downregulation level of most of the genes was highly correlated with the induction level: at 10 µM, the correlation coefficient between fold-induction and fold-inhibition by treatment with **4** was Rsq=0.95 at 10 µM and Rsq=0.96 at 5 µM.
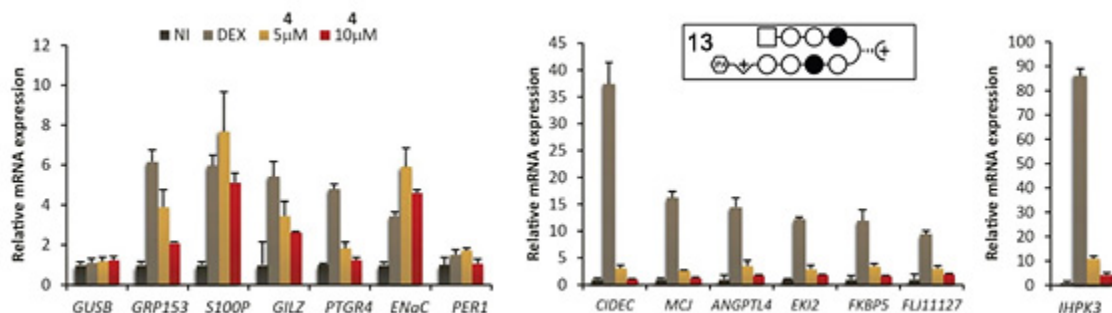
Figure 2.11 Analysis of levels of expression of genes induced by Dex and inhibited by polyamide 13. All genes, except S100P and ENaC were significantly downregulated by both 5 μM and 10 μM polyamide 13. At 10 μM most genes were downregulated to 1-2fold their baseline levels and at 5 μM - to 4-fold baseline expression. The remaining two, IHPK3 and PTGR4, were downregulated to 10-fold and 2-fold their baseline and PER1 was downregulated slightly. Overall polyamide 13 shows excellent correlation between fold-induction and fold-inhibition at both 10 μM (Rsq=0.95) and 5 μM (Rsq=0.96). This suggests that relative induction dictated polyamide inhibtioin levels, regardless of which gene was affected.

In order to compare the trends in gene downregulation studies, I pooled the RT-qPCR results into a matrix (Fig. 2.12). Three most potent polyamides (1,6 and 13, targeting 5'WGWWCW3', 5'WGWGWW3' and 5'WTWCGW3', respectively) downregulated CIDEC significantly better than ANGPTL4. However, only two of them (1,13) downregulated IHPK3 more potently than MCJ (Fig. 20a). Three less potent polyamides downregulated MCJ more potently than IHPK3 (3,4 and 5, targeting 5'WGGWCW3', 5'WWCWWW3' and 5'WWCWCW3', respectively). All but one (compound **2**) of the less potent polyamides downregulated CIDEC less potently than ANGPTL4. Polyamide **2** was overall less selective but also more potent than the other polyamides in the low-potency group (Fig. 2.6B). Only one of these compounds (**3)** downregulated ANGPTL4 and CIDEC but not MCJ.

These results show that in the GR system there is a difference between polyamides in their gene regulation patterns. These differences highlight a balance between specificity and potency. This balance could partially be due to native binding selectivity of the polyamides and partially due to potential cross talk between the genes. Interestingly, Py-Im polyamide sequence specificity alone

does not explain the patterns of gene regulation. A good example of the lack of correlation between targeted sequence and gene expression changes within the library is a mismatch compound (**13**) having a significant effect on gene expression, while a many of the 'match' polyamides such as compounds **3**, **4,** or **5** had less potent gene downregulation activity**.** Additionally, two compounds targeted to the same sequence – **5** and **6 –** show different gene regulation profiles, suggesting that Py-Im polyamide structure may play a significant role in their *in vivo* activity**.** This brought upon a question, what governs the polyamide potency and activity in tissue culture. Two possibilities include the binding affinity of Py-Im polyamides and their cellular uptake.

One explanation for potency of the synthesized compounds is their binding affinity to DNA. A compound with weaker binding could possibly have less binding energy than GR and thus could be incapable of displacing transcription factors. The binding affinity of compounds can be measured using a DNA thermal denaturation assay (32). Two sets of olignucleotids have been used for the assay: published sequences that can be compared to previous results (5' CGAnnnnnnAGC 3', where n are specific to a polyamide) and 5 oligonucleotides with sequences of GREs occuring in A549 cells. The latter have been chosen to assure that each GRE can only contain one polyamide binding site and has a match sequence of only one of the studied polyamides.

The tested polyamides included compounds **1**- **6** (Fig. 2.13). Each of the polyamides stabilized the DNA duplex in thermal denaturation thus proving its DNA-binding capability (Fig. 2.13). In order to test the specificity of polyamides in DNA thermal denaturation assay, the shorter oligonucleotides were used, to lower the melting temperatures. Both polyamides tested show a degree of sequence specificity and stabilize the match-sequence the best (Fig. 20). Importantly, however, the Py-Im polyamides with lower thermal stabilization values, and thus lower affinities (ΔTm < 5.0° C), had lower potency in gene expression inhibition. This suggests a positive correlation between Py-Im polyamide binding affinity and ability to regulate gene expression. Interestingly, however, polyamides 5 and 6 had similar effects on thermal stabilization of DNA duplexes, and thus binding affinity is not the only variable that governs potency of these compounds. Another factor that could contribute to ability of polyamides to regulate gene
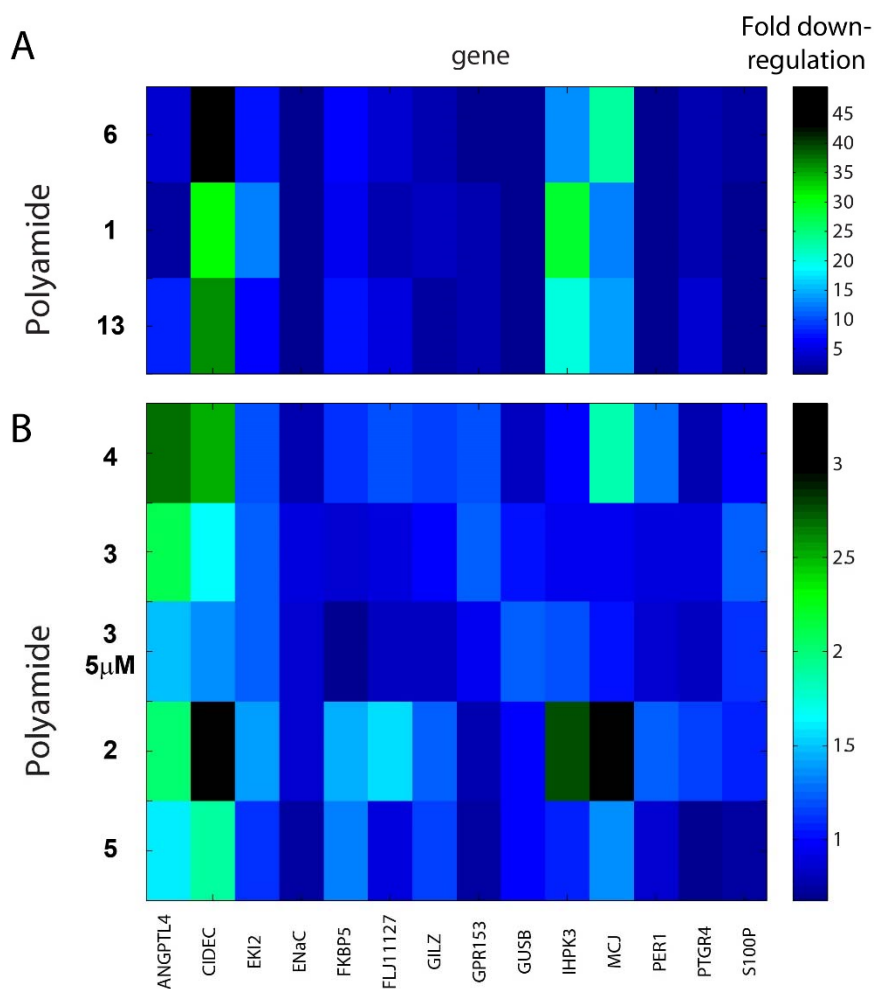
expression is their cellular uptake.

Figure 2.12 Downregulation of Dex-induced genes using Py-Im polyamides. (A) Highly potent polyamides downregulated CIDEC more efficiently than ANGPTL4. Also two out of the total number downregulated IHPK3 more potently than MCJ. (B) All but one of the less potent polyamides downregulated ANGPTL4 more potently (or comparably to) than CIDEC. Polyamide **5** downregulated MCJ as well, but 3 did not. Polyamide **2** downregulated MCJ and CIDEC more potently, but also affected several other genes. Overall, this library of GR-targeting polyamides allows for selective targeting of genes, but suffers from poor potency. Alternatively it can target a large number of genes, while suffering from poor selectivity. This issue will need to be addressed with improved dosing schemes and alternative polyamide structures.

A

Polyamides	T_m (°C)	ΔT_m (°C)
DNA = Respective GREs		
1	90+	?
3	80.6 (±0.3)	2.5
4	78.8 (±0.1)	4.5
5	81.9 (±0.5)	4.7
6	82.6 (±0.3)	5.4

B

Polyamides	T_m (°C)	DT_m (°C)
DNA = 5 - CGA TGTTCA AGC -3		
—	53.4 (±0.2)	—
1	72.7 (±0.4)	19.2
2	55.3 (±0.2)	1.9

C

Polyamides	T_m (°C)	DT_m (°C)
DNA = 5 - CGA TGGTCA AGC-3		
—	57.1 (±0.4)	—
1	69. 4 (±0.3)	16.1
2	73.6 (±0.3)	20.2

Figure 2.13 Thermal denaturation assay on GREs and DNA oligos. (A) Each of the compounds synthesized stabilizes DNA duplex in thermal denaturation. Compound **3** is by far the weakest stabilizer, followed by **4** and **5** and then by **6**. Compound **1** is likely a strong stabilizer: however, it is not possible to estimate the stabilization temperature, because it's out of the range of the assay. Compound three has been tested twice to show that ΔT from polyamide binding is higher for oligos with lower melting temperarture. (B) Assay showing thermal stabilization of compound 1 and its match sequence, and compound **2** with one nucleotide mismatch. (C) Assay showing thermal stabilization of compound **2** to its match sequence, and compound **1** with one nucleotide mismatch. Compound **1** appears to be a more promiscuous binder than **2**.

Cellular uptake of polyamides

The potency of polyamides in gene regulation depend both on their DNA binding characteristics and cellular uptake. In order to decouple those two variables I performed cell-uptake studies by conjugating a Fluorescein Isothiocyanate (FITC) molecule to a polyamides' C-terminus (33). This conjugation has not affected the gene regulation potency in the past (34, 35) and is used as

a golden standard of cellular uptake in our group. We use confocal microscopy and either qualitative or software analysis of images to determine nuclear concentration of polyamides using a protocol described before (36) that has since been slightly simplified. In short, 60000 cells (A549) are plated on an optical glass in 35mm culture dishes, incubated in medium for 24 hours, and consequently dosed with Fluorescein-conjugated polyamides for 24-48 hours. Subsequently cells are washed twice with PBS and imaged on a Zeiss LSM 5 Exciter with a 63x F/1.4 objective. All tested compounds were visualized under microscope showing their nuclear uptake (Fig 2.14). The nuclear uptake of compound **7**, a FITC-conjugated version of **1**, was comparable to **9**, **10** and **11**, fluorescent counterparts of **3,4,** and **5**, respectively. Thus the low potency of the latter three cannot be explained by poor uptake and is more likely due to their binding properties or biological function of the sequences they bind. The poorer uptake compound **8**, on the other hand, could explain less potency in gene downregulation, as compared to **1**.
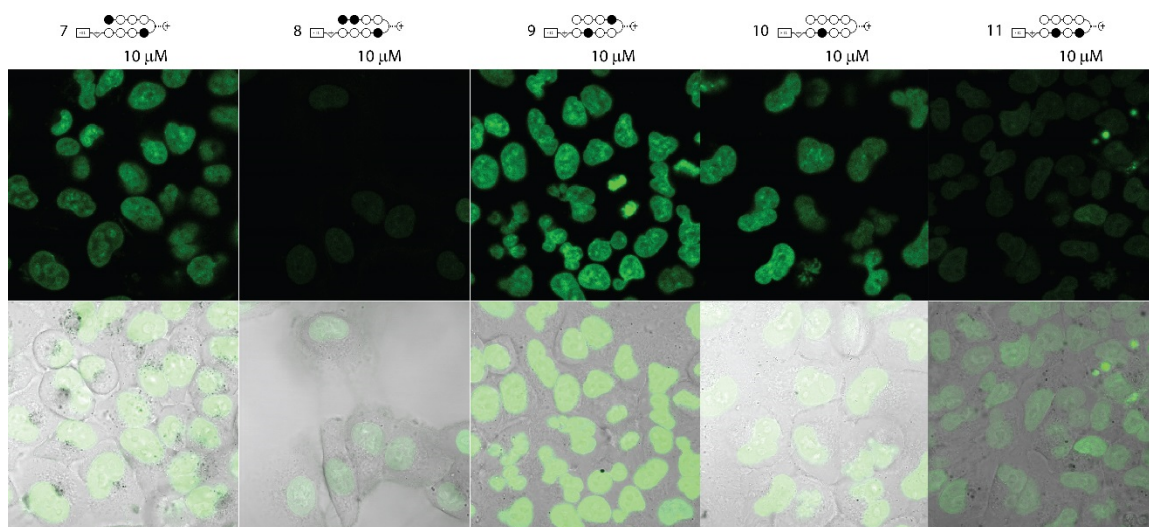

Figure 2.14 Nuclear uptake of polyamides. Compounds **7-11** were dosed at 10 μM and uptaken by A549 cells over the course of 48 hours. They were subsequently imaged on a Zeiss Exciter LSM confocal microscope. The nuclear uptake of **7** was comparable to the uptake of **9** and **10** suggesting that poor uptake cannot explain low potency of their FITC-unconjugated counterparts: **3, 4** and **5**. Poorer nuclear uptake of **8** and **11** could, however, explain low potency of **2** and **5**, respectively.

Discussion and strategies for improving polyamides specificity in-cells.

Dynamics simulations are a useful tool to understand biological processes. Polyamide binding can be modeled to give a quantitative insight into their specificity and kinetics in living cells. The developed model suggests a possible route for improvement of polyamide specificity, through elongating the targeted sequence. Currently the most attractive option is synergistic binding of multiple polyamides. In this approach no new types of molecules need to be developed: we can use molecules with well-known specificity and excellent nuclear uptake. Additionally, using several polyamides in synergy, rather than a single long polyamide, enables targeting disjoint sequences and may enable a of linear to exponential increase in selectivity with each added polyamide.

Modeling kinetics and thermodynamics of polyamide binding in-cells.

The human genome consists of approximately 3 billion bases. Most of the genome is part of intergenic sequences that are thought to have little influence on the gene expression 1 (37). A typical 8-ring hairpin polyamide can recognize 6-base pair regions that appear commonly within the genome. For example, the pattern 5'-WGWWCW-3' occurs 15 million times in the human genome, while 5'-WGGWCW-3' -- 13 million times. As a resul, targeting an expression of a small subsets of genes is difficult and requires a sufficient amount of polyamides inside the nucleus as well as proper kinetics and thermodynamics of their binding. In order to investigate the properties of polyamides that would contribute to the most selective and potent gene regulation I constructed a kinetic model of polyamide uptake and binding (Fig 2.15).
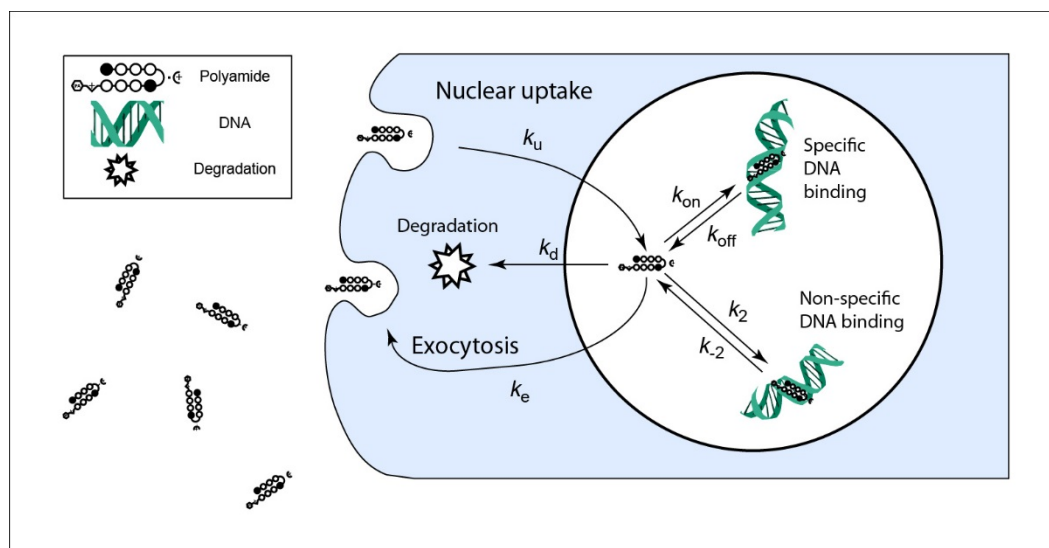
Figure 2.15 Trafficking dynamics of polyamides in living cells. Polyamides are internalized and exocytosed by cells at rates of k_u and k_e, respectively. These rates were estimated based on previous cell-uptake studies in the group. Polyamides uptaken into the nucleus can bind to either the DNA they target (rate constant of association is kon and constant of dissociation is k_{off}) or to mismatch sequences with rate of association of k_2 and rate of dissociation of k_{-2}. The rate of degradation (k_d) was added to account for thermal degradation and irreversible binding of polyamides to proteins and membranes.

Knowing the proper range of binding affinities required for selective gene regulation can guide the choice of polyamide molecules used in the study. Likewise, knowing the qualitative relation between selectivity and potency of polyamides can help in choosing the dosing scheme for a desired effect. Kinetic parameters (listed in Table 1) of several compounds have been published previously (38, 39) and I used them in simulation. In order to match cellular uptake properties I used radiography data performed in our lab in the past (Melander 2001, unpublished data). The concentrations of polyamides binding sites in A549 cell nucleus have been calculated by dividing the whole genome into 6 basepair pieces and measuring their concentration in an average volume of an A549 cell nucleus equal to 466 μm^3. This is a simple assumption that doesn't include influence of polyamides on binding in the proximity of other molecules. Even though polyamides have been shown to bind to nucleosomes (3, 40) little is known about their binding to more organized DNA structures, such as 30 nm fibers. In order to accommodate for genome accessibility, a parameter multiplying DNA concentration has been added for both mismatched and matched DNA. The lower bound of accessible genome was assumed to be the amount of

DNAse I hypersensitive DNA in A549 cells, (7) equal to 2.1% . The upper bound was the whole genome of A549 cell and midrange value chosen for representation was 10%. These three data points allow us to show trends in the effect that the amount of DNA in the nucleus has on the polyamides specificity (Fig. 16). The less DNA there is, the more likely high uptake of polyamides is going to fill up all the match-sequences. Once that happens all additional polyamide uptake decreases the specificity of binding. Depending on the amount of accessible genome, such decrease in specificity can happen at low micromolar concentrations, which is the dose we currently observe in A549 cells (41).

Table 1   Kinetic constants for polyamide binding and mismatch binding used in modeling

	K_D (M)	K_d (1/s)	K_a (1/Ms)	Specificity
15	$(5.12) \times 10^{-9}$	$(5.20) \times 10^{-4}$	$(9.76) \times 10^{4}$	149
16	$(5.57) \times 10^{-7}$	$(1.31) \times 10^{-3}$	$(2.39) \times 10^{3}$	
17	$(2.85) \times 10^{-11}$	$(2.00) \times 10^{-3}$	$(7.00) \times 10^{7}$	88
17 mismatch	$(2.52) \times 10^{-9}$	$(1.51) \times 10^{-1}$	$(6.00) \times 10^{7}$	
18	$(1.33) \times 10^{-11}$	$(2.00) \times 10^{-3}$	$(1.50) \times 10^{8}$	61
18 mismatch	$(8.10) \times 10^{-10}$	$(6.50) \times 10^{-2}$	$(8.00) \times 10^{7}$	

Abbreviations: K_A, association equilibrium constant; K_a, association rate constant; K_D, dissociation equilibrium constant; K_d, dissociation rate constant. Specificity is defined as K_A (polyamide binding)/ K_A (mismatch binding).

Table 2   Initial conditions and constants for modeling

	Value
Nuclear concentration of polyamides (M)	(0.00)
Match DNA concentration (M)	$(7.53) \times 10^{-5}$
Mismatch DNA concentration (M)	$(2.51) \times 10^{-3}$
Rate of nuclear uptake (1/s)	$(2.89) \times 10^{-6}$
Rate of nuclear exocytosis (1/s)	$(2.89) \times 10^{-4}$
Rate of degradation of free polyamide (1/s)	$(5.78) \times 10^{-6}$

Therefore once all match sites are filled, the selectivity of polyamides will get worse, but their potency will stay the same. There could thus be an upper bound at which polyamides should be dosed for maximum selectivity. Similarly, as expected, there is an upper bound to the affinity of

polyamides. Once the polyamides have very slow rate of dissociation, they take a long time to reach equilibrium and during 48 hour induction the majority of polyamides are bound in mismatch DNA regions. If their affinity is above a certain threshold, their binding could be strong enough to displace transcription factors even when bound to mismatch-sequences.

Polyamides typically show between 50-100 fold selectivity in binding different DNA sequences (39). However, given that mismatch DNA is over 30-fold more concentrated than the match sites, it is not possible to avoid nonspecific binding. Since each gene in mammalian cells can affect other ones, even few nonspecific binding events can significantly affect cellular signaling. In the case of polyamides, the extent of non-specific binding makes it very likely that polyamides change signaling in many pathways, not only the one they target. In order to alleviate this problem, we will need to drastically increase polyamide binding specificity *in vivo*.

One possibility is to increase the length of the binding site of polyamides. In the past both ours and other groups targeted DNA regions longer than 6-basepairs (7, 38, 42). However, increasing the length of sequences does not necessarily increase specificity and few studies have been done to investigate it. The polyamides recognizing 10-basepairs are shown to have very different binding affinities when binding to different sequences (38, 43). However, sequence selectivity of these molecules has not been shown.

An alternative possibility is to make use of the synergistic targeting of multiple polyamides at once. Many genes in mammalian cells are controlled by more than one transcription factor. For example, circadian gene regulatory network of tens of functionally interconnected genes (45). One of them, PER2, is controlled by GREs and other two regulatory elements: CEBP and EBOX (44). In this scenario we would use polyamides targeting 5- or 6-basepairs that show good nuclear uptake, strong potency in gene downregulation, and have known sequence specificity. Using the fact that the regulatory elements of most genes are longer than 5 -or 6-basepairs (for example GRE is 15 basepairs long) we can target each of these elements with multiple short polyamides (Fig. 2.17A). In this case the fractional occupancy at the targeted site will be significantly higher than if a single polyamide was targeted at higher concentration. Simply the fact that polyamides can bind concurrently, or one at a time, yields more possible binding
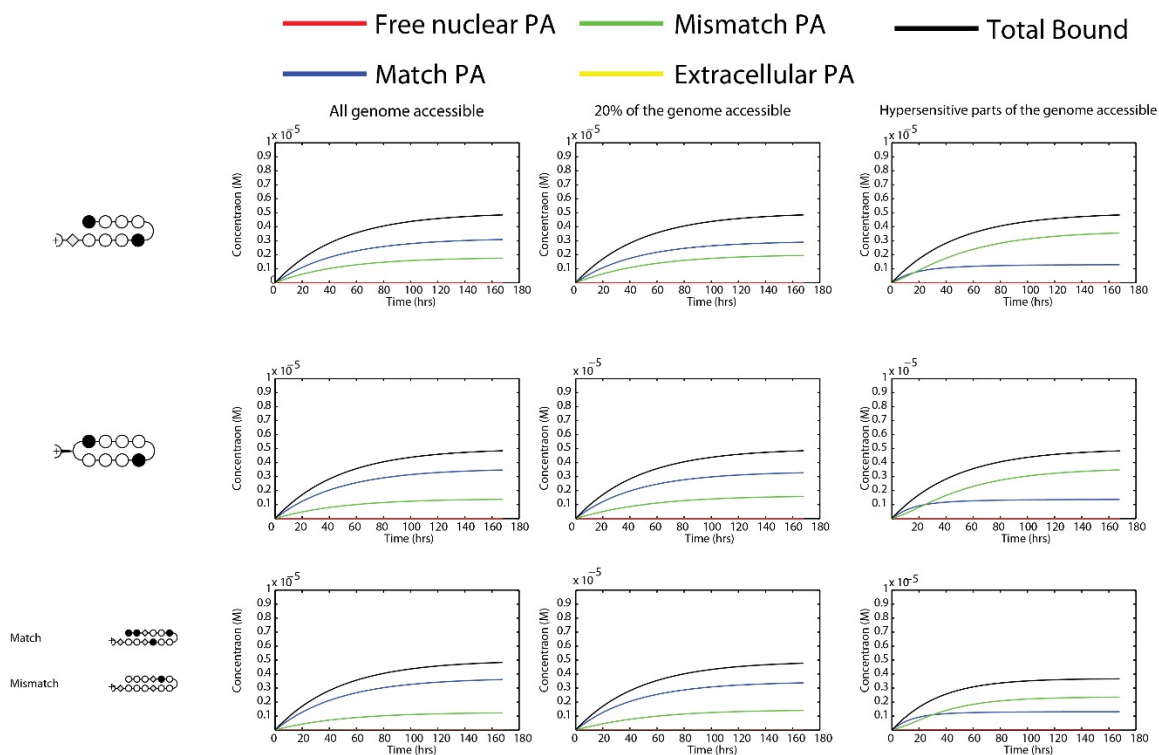
Figure 2.16 Modeling results for the three published compounds ((38, 39)) in A549 cells: a significant amount of polyamides is bound to mismatch DNA and high uptake of polyamides can lead to lower specificity. The first column shows that when all of the genome is accessible there is a constant uptake, and DNA binding, of all three polyamides. Because of the high concentration of both polyamide and nuclear DNA, a majority of polyamides are bound to DNA. This results in an initial linear accumulation of polyamides in the nucleus in time, which is consistent with the radiography data (Melander, 2001; unpublished data). All compunds reach equilibrium of binding relatively quickly and ratio of match-binding (complex Match-DNA-polyamide depicted in blue) to mismatch binding (complex of mismatch-DNA-polyamide depicted in green) is determined by the relative dissociation constants for polyamides. Even though a significant fraction of polyamides bind to (much more abundant) mismatch sites, the main species in the nucleus is DNA bound to the match polyamide. The second column represents situation when only 20% of the genome is accessible. In that case, the match sites quickly become filled with polyamide and its further uptake only increases non-specific binding. This effect is aggravated even further when only the DNAse I hypersensitive DNA is assumed to be accessible (third column). All three compounds yielded similar qualitative characteristics of nonspecific and specific binding. In conclusion, significant fraction of polyamides binds to nonspecific DNA, which will affect cellular signaling in an unpredictable fashion.
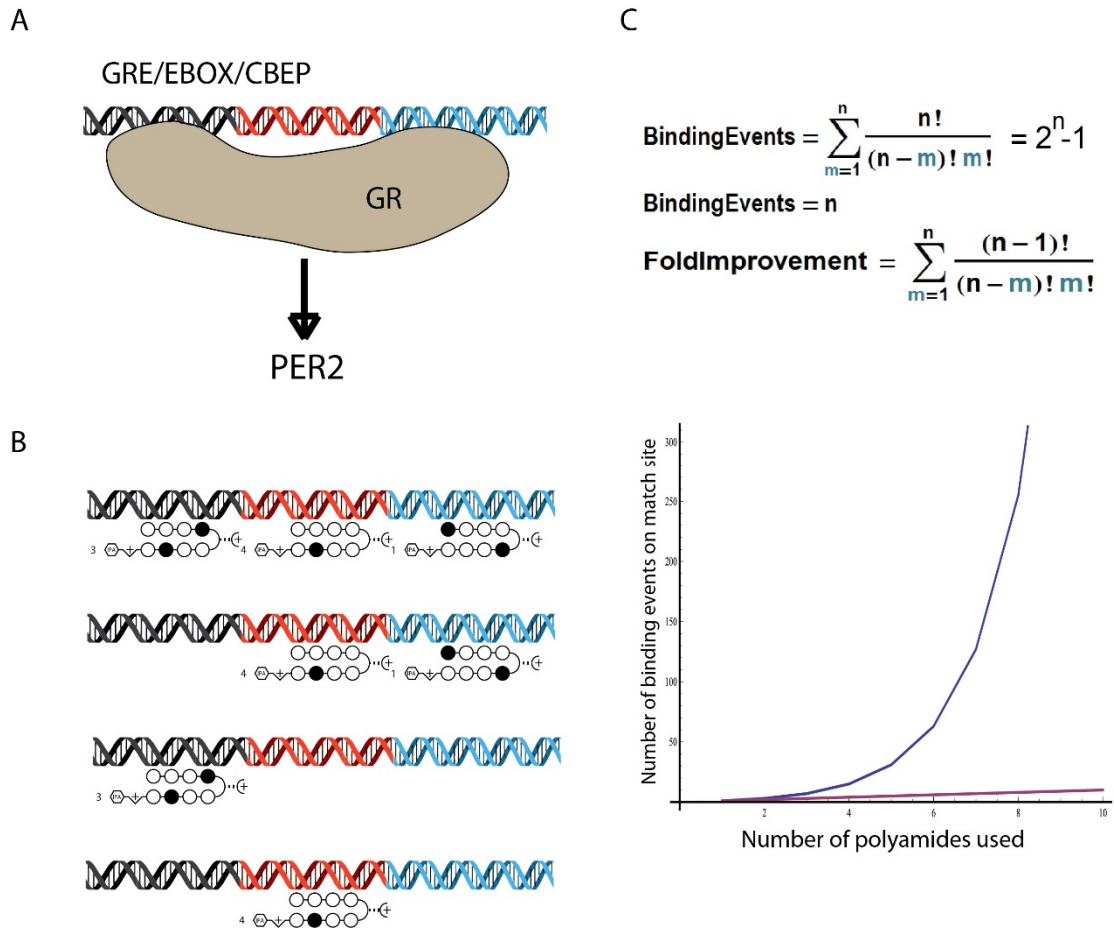
A

GRE/EBOX/CBEP

GR

PER2

C

$$BindingEvents = \sum_{m=1}^{n} \frac{n!}{(n-m)!\,m!} = 2^n - 1$$

$$BindingEvents = n$$

$$FoldImprovement = \sum_{m=1}^{n} \frac{(n-1)!}{(n-m)!\,m!}$$

B

Number of binding events on match site

Number of polyamides used

Figure 2.17 Combinatorial targeting of polyamides. (A) The circadian gene regulatory network in mouse (44) is an example of the many networks. where genes are controlled by multiple transcription factors. (B) One of the genes PER2 is controlled by GRE, EBOX and CBEP regulatory elements. (C) Each of these elements can be targeted by multiple polyamides. Since the presence of even one precludes TF binding, it is likely that more than one polyamide bound will do that as well. When multiple polyamides target long sequence, the time where at least one polyamide resides is increased exponentially for a given concentration, as opposed to linear increase in time of occupancy (D). For the same concentration of polyamides, if 10 polyamides are used instead of 1 at ten times higher concentrations, there is up to 100-fold increase in relative specificity, proviso that each polyamide is capable of regulating gene expression by itself.

Events, leading to displacement of transcription factors (Fig. 2.17B, C). In fact, for large numbers of polyamides targeted the sequence selectivity grows exponentially (with a base of 2).

This is a preferred solution because it relies on the already existing and well developed molecules. It also results in an exponential increase in specificity while allowing targeting very long sequences with 8-ring hairpin polyamides, leveraging their good cell uptake and pharmacokinetic properties.

Directions for a genome-wide evaluation of polyamide DNA-occupancy and action

The panel of 12 genes tested does not capture the complexity of the whole-genome selectivity of polaymides. Previous studies have shown that only a small number of genes are regulated differentially (22, 27, 34). The probability of finding these genes by qPCR is small, and thus we will need to use RNA-sequence to identify them. This method however is expensive, which is the reason for following with a small exploratory panel of genes at first. Since all the compounds tested have good nuclear uptake and bind DNA *in-vitro*, they are candidates for evaluation by RNA-sequencing. In order to asses the reliability of the high-throughput data, my next step will be to evaluate gene downregulation patterns of two previously published1 polyamides (targeting WGWWCW and WGWCGW). These polyamides have shown approximately forty genes that are more downregulated by the latter compound, despite its overall lower potency. In my next experiments I would like to confirm that there exist genes that are downregulated differentially. Since the second compound is overall lest potent than one targeting WGWWCW , the control we used before, showing that a mismatch polyamide (WGWCGW) does not downregulate a subset of genes is not a sufficient proof of sequence selectivity.

Appendix A: materials and methods

Synthesis of Polyamides.
All polyamides were synthesized on solid phase on Kaiser oxime resin (Nova Biochem) using protocols published previously (46, 47). The polyamides were subsequently cleaved subsequently off the resin with 3,3-diamino-N-methyldipropylamine and purified by reverse-phase HPLC. Isophtalic acid was conjugated after prior conjugation by PyBop (Nova Biochem) or PyAop (Oakwood Products, Inc) using an established protocol (34). The mass and purity of polyamides was assesed by MALDI-TOF and reverse phase HPLC.

Measurement of cellular mRNA.

A549 cells (ATCC) were plated in 12-well plates at a density at 60000 cells per well in 1ml of F12-K medium (ATCC) supplemented with 10% (vol/vol) FBS (Irvine Scientific). After 24 h, the medium was replaced with F12-K containing 10% (vol/vol) charcoal treated (CT) FBS and polyamides at desired concentrations. Cells were incubated with polyamides for 48 h and were subsequently treated with 100nM Dexamethasone (Dex). RNA was then extracted using RNEasy kit (Qiagen) and reverse-transcribed using Transcriptor first-strand kit (Roche). Quantitative real-time RT-PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems) with 5 mL of 12-fold diltued RT-reaction, 1.8 µM primer and 1x master mix concentration on an ABI 7300 qPCR instrument (Applied Biosystems). The expression levels of all samples were normalised to a GUSB gene. Primer sequences were obtained from Harvard Primer bank and previously published reports (30).

DNA melting temperature assay

DNA melting assay was performed on a thermally controlled cell in a Cary 100 UV-Vis spectrophotometer in quartz cuvettes. The DNA oligonucleotides and polyamides were added to an aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ at pH 7.0, at 2 mM and 4 mM respectively. The data was recorded in a thermal cycle, where the solution was first heated to 90 C (to anneal oligonucleotides) and cooled to 25 C at 5 C/minute. Subsequently, the temperature was increased from 25C to 90C at rate of 0.5C/min and for each temperature the absorbance at 260 nm was recorded for solutions containing only oligonucleotides or both oligonucleotides and polyamides. The reported melting temperatures were defined as the maximum of the first derivative in temperature of the absorption at 260nm between 25C and 90C. The value of melting temperatures for samples with only DNA were subtracted from the samples containing both polyamides and DNA to extract the ΔT_m, or the stabilization temperature.

Confocal microscopy imaging

For nuclear uptake imaging 12000 A549 cell (ATCC) were plated on a glass optical window in 35mm dishes (MatTek product no. P35G-0-14-C). Cells were grown in 200ul of F12-K medium (ATCC) for 24 h. The cells were then washed with PBS twice and the medium was replaced with 1 ml of a fresh F12-K CT medium supplemented with FITC-labeled polyamides (at 10 mM) in DMSO (final concentration 0.1%). Subsequently the cells were incubated at 37C for 48 h washed twice with PBS and imaged in a Zeiss LSM exciter inverted confocal microscpe with a 63x/F1.4 oil immersion lens.

FITC-conjugated polyamides were imaged in multi-track mode using 488nm laser excitation with a pinhole of 375 mm and standard filter for fluorescein. All images were analyzed using Zeiss LSM software.

Appendix B: Structures of GRE targeting polyamides
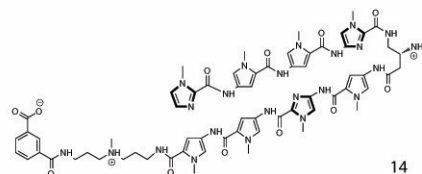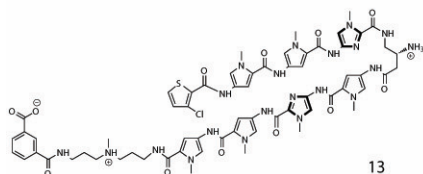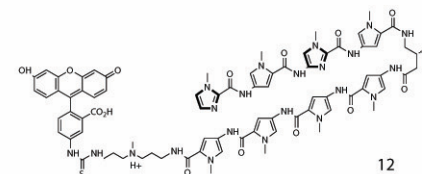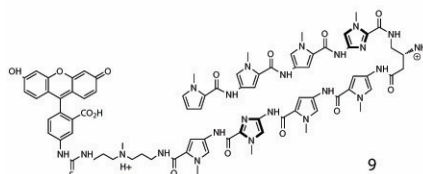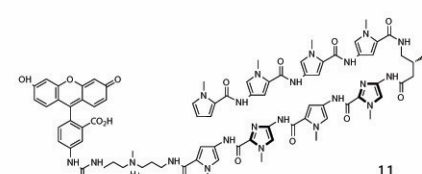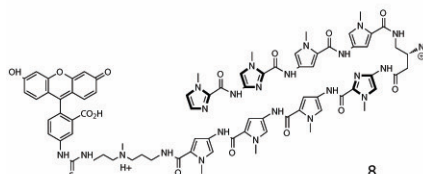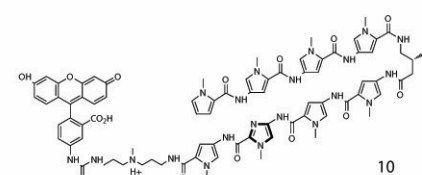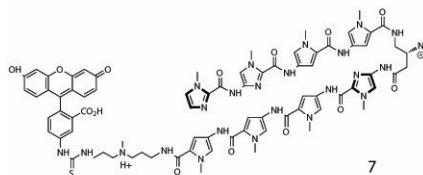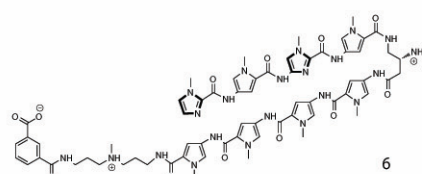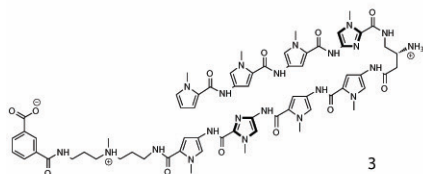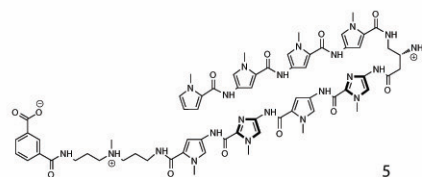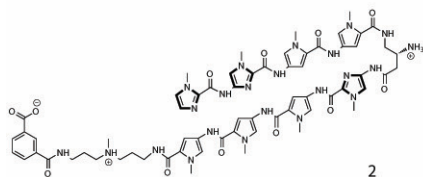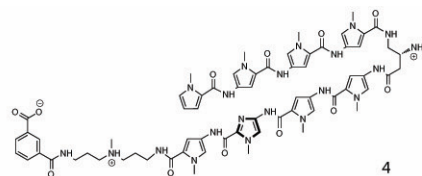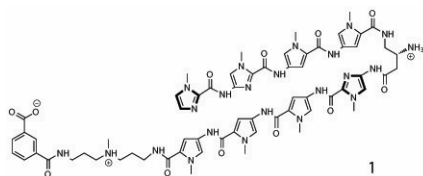
Appendix C: RT-qPCR primers used

PER1_1F CATCATGTTCTCTTGGCTGGTGG
PER1_2F GATCTTTCTTCCCCTACTCCCCG
PER1_3F GATTGGCTGGGGATCTCTTCC
PER1_1R AGGACGGCTGTCGTTTTGTTG
PER1_2R GGCGCTCAGAAAATGCTCAGTAG
PER1_3R CAGCCCTGACCTTAGTGGAGACC
ZFP36_1F CCTCTCCCTCAGTCCTTCCTGAC
ZFP36_1R GACACAGAGGAGTGGCACACAGA
PTGR4_1Fw CATCATCTGCGCCATGAGTGT
PTGR_14Re GCTTGTCCACGTAGTGGCT
PTGR4_2Fw CACTACGTGGACAAGCGATTG
PTGR4_2Re CATAGACTGCAAAGAGCGTGAG
SOCS1_Fw TTTTCGCCCTTAGCGTGAAGA
SOCS1_Re GAGGCAGTCGAAGCTCTCG
SPRY1_1Fw GCAGTGGCAGTTCGTTAGTTG
SPRY1_1Re TCTCTGACGGCTATCCAAAGAA
SPRY1_2Fw GGATAGCCGTCAGAGATTAGACT
SPRY1_2Re GCTGCCTCTTATGGCCTTGA
SPRY1_3Fw TTCGGTGGTGAAAAGACCTGC
SPRY1_3Re CCCTGGCATTACTTGGGAGT
RASD1_1Fw AGCTGAGTATCCCGGCCAA
RASD1_1Re CGAGGATGACCATGCGATAGC
RASD1_2Fw TACACGCCTACCATCGAGGA
RASD1_2Re ACACCAGGATGAAAACGTCTC
p57KIP2_FO AACGCCGAGGACCAGAA
p57KIP2_RE ACCGAGTCGCTGTCCACTT
FKBP5_FO AGGCTGCAAGACTGCAGATC
FKBP5_RE CTTGCCCATTGCTTTATTGG
ANGPTL4_FO ATTCTTTCCAGCGGCTTCTG
ANGPTL4_RE GAGGACTGGAGACGCGGAG
FLJ11127_FO GACAAGGAGCCCCACG
FLJ11127_RE GCTTGTAGCTAGCATCCAGGA
MCJ_FO AGGAGGATTTGAACAGAAAATG
MCJ_RE CTATGAGCTGTTCTAATCTTAG
Cidec1Fw TCCCTTAGCCTTCTCTACCCC
Cidec1Re AGGTACGCACTGACACATGC
Cidec2Fw TGTGTCAGTGCGTACCTCTG

Cidec2Re CCTTCCTCACGCTTCGATCC
Cidec3Fw ATTGATGTGGCCCGTGTAACG
Cidec3Re CAGCAGTGCAGATCATAGGAAA
GilzFw TGG TGG CCA TAG ACA ACA AG
GIlzRe TGC TCC TTC AGG ATC TCC AC
IHPK3_Fw TTCTCGCTGGTGGAAGACAC
IHPK3_Re CAGCAACAAGAACCGATGC
ENaC_Fw AGCACAACCGCATGAAGAC
ENaC_Re TGAGGTTGATGTTGAGGCTG
EKI2_Fw CTACTGCACCTTCCAGAATGG
EKI2_Re CCGTTGGCGTGGATAGTATG
S100P_Fw TGATGGAGAAGGAGCTACCAG
S100P_re ACTTGTGACAGGCAGACGTG
GPR153_Fw CTGGATGGTGTCCTTCATCC
GPR153_Re GATCTCAGCCACGATGAAGC

Appendix D: Code

Dynamics modeling of polyamide uptake and binding.

```
% filename: PolySim.m
%close all; % close all figures
clear; % clear variables
% Initial Conditions:
accessible=0.1;
NucPoly = 0; % Polyamide in nucleus, units M
OutPoly = 10^-5; % Polyamide in the medium, units M
DNA_match=53.8*10^-6*1.4*accessible; % Concentration of match sequences in A549,
multiplied by 1.4 because cells are hypotriploid with 65ish chromosomes
DNA_mis=1.792*10^-3*1.4*accessible; % Concentration of mismatch sequences in A549,
multiplied by 1.4 because cells are hypotriploid with 65ish chromosomes
selectivity=50; % Polyamide selectivity
ku=1/(96*3600); % units 1/s
ke=100*ku; % units 1/s
kd=1/(48*3600); % units 1/s
tstart = 0; % simulation start time (seconds)
tfinish_values = [3600*7*24]; % simulation end time (seconds)
options=odeset('maxstep',10); % (insures the the integration has a maximum stepsize of 1)

for i = 1:1,
```

```
% %Cycle, Baliga et al, 2001
% kon=15*10^7; % units 1/(M*s)
% koff=2*10^-3; % units 1/s
% k2=8*10^7; % units 1/(M*s)
% k_2=0.065; % units 1/s
% %2B3, Chen 2010, Molecular Therapy
% kon=9.76*10^4; % units 1/(M*s)
% koff=5.2*10^-4; % units 1/s
% k2=2.39*10^3; % units 1/(M*s)
% k_2=1.31*10^-3; % units 1/s
%
% %Hairpin, Baliga et al, 2001
% kon=7*10^7; % units 1/(M*s)
% koff=2*10^-3; % units 1/s
% k2=6*10^7; % units 1/(M*s)
% k_2=0.151; % units 1/s
%Weak binder
kon=2*10^3; % units 1/(M*s)
koff=2*10^-3; % units 1/s
k2=10^2;                    %              units              1/(M*s)
k_2=10^-2; % units 1/s

tfinish = tfinish_values(i);
%-----------------------------------------------------------------------
% Simulate and Plot the Full ODE solution to E + S <--> ES --> E + P
initial_conditions = [ NucPoly; % initial amount of Nuclear Polyamide
OutPoly; % initial amount of E
0; % initial amount of DNAPmatch
0; % initial amount of DNAPmmismatch
DNA_match; %initial amount of DNA match
DNA_mis]'; %initial amount of DNA mismatch
tic % keep time
[t y]=ode23t('PolyFunc_nondeplet',[tstart tfinish],initial_conditions,options,ku,ke,kd,kon,koff,
k2,k_2);
t_elapse = toc; % keep time
% display the time it took to integrate full ODEs

sprintf('Time to integrate full ODEs was %3.3f seconds.',t_elapse)
```

```
Pnuc=y(:,1); % Nuclear polyamide
Pout=y(:,2); % External Polyamide
Pmatch=y(:,3); % DNA match bound polyamide
Pmis=y(:,4); % DNA mismatch bound polyamide
% plot the concentrations of S, E, ES, P as a function of time
t=t/3600;
figure(i),
plot(t,Pnuc,'-r',t,Pmatch,'-b',t,Pmis,'-g', t, Pout, '-y',t, Pmis+Pmatch,'k');
legend('Pnuc','Pmatch','Pmis','Pout','Total Bound');
xlabel('Time (hrs)');
ylabel('Concentration (M)');
titletext = ['Full ODE Solution to Polyamide binding and uptake'];
title(titletext)
hold on
figure(i+1),
plot(t, Pmis./Pmatch,'k');
legend('Pnuc','Pmatch','Pmis','Pout','Total Bound');
xlabel('Time (hrs)');
ylabel('Concentration (M)');
titletext = ['Full ODE Solution to Polyamide binding and uptake'];
title(titletext)

hold on


end
function dydt = mmfunc(t,y,flag,ku,ke,kd,kon,koff, k2,k_2)
Pinside=y(1);
Poutside=y(2);
DNAPmatch=y(3);
DNAPmis=y(4);
DNAmatch=y(5);
DNAmis=y(6);
dydt        =        [ku*Poutside-ke*Pinside-kon*DNAmatch*Pinside+koff*DNAPmatch-
k2*DNAmis*Pinside+k_2*DNAPmis-kd*Pinside;

0;
```

kon*DNAmatch*Pinside-koff*DNAPmatch;

k2*DNAmis*Pinside-k_2*DNAPmis;

-kon*DNAmatch*Pinside+koff*DNAPmatch;

-k2*DNAmis*Pinside+k_2*DNAPmis]; % d[P]/dt

Appendix E: Full list of sequences of top 10% most Dexamethasone-induced GREs (5'-3'):

GAGAACAGTATGTCCT	AAGGACACCGTGTGCT	CAGAACTTTCTGTACT	AGAACAACGTGTTCTG
AGAACAGTATGTCCTC	AGGACACCGTGTGCTA	CAGTACACAGTGTGCT	CGGAACACCGTGTTCT
GGGAACACTGTGTCCT	GGGACAGGATGTTCCT	AGTACACAGTGTGCTC	AAGCACATACTGTTCA
GGGAACACTATGTCCT	CAGAACATCCTGTTCT	CAGGACATTTTGTCCT	GGAACAGAATGTCCTG
AGGAACATCATGTCCT	AGAACATCCTGTTCTT	AGGACAGAATGTTCCA	AGCACATGATGTGCCC
AGGAACACTGTGTCCT	TGGAACACTCTGTCCT	GGGAACATTTTGTGCT	CAGCACAGAATGTTCT
AGGACACAGTGTTCCT	CTGGACACTCTGTCCT	TGGAACACTCTGTTCT	AGCACAGAATGTTCTG
AGGACACAGTGTTCCT	CAGGACAGCGTGTGCT	TAGGACATGGTGTTCT	AGCACACACTGTCCCA
GAGGACATACTGTTCT	AGGACAGCGTGTGCTC	AGGACATGGTGTTCTG	CAGTACACAGTGTCCT
AGGACATACTGTTCTC	AGCACAGTGTGTTCCC	AAGGACATGGTGTTCT	AGTACACAGTGTCCTC
GTGGACATGGTGTACT	AGAACAAAGTGTGCCA	AGGACATGGTGTTCTT	GAGGACACATTGTCCT
GGGACAGAATGTGCCT	TGGCACACTTTGTTCT	CTGAACAGCCTGTCCT	AGGACACATTGTCCTC
CAGAACACTATGTCCT	CTGGACATAGTGTGCT	GGGAACAGCCTGTCCT	CAGCACCTTCTGTTCT
AGAACACTATGTCCTG	AGCACAGCGTGTTCCT	AGGCACATGCTGTTCT	CAGAACATGGTGTCCT
CAGAACATTCTGTGCC	AGAACAGCATGTGCAT	TTGTACATGTTGTTCT	AGAACATGGTGTCCTG
GGTACACAATGTCCTG	CAGAACATCCTGTACC	CTGCACACTCTGTTCT	TTGAACAGGCTGTTCT
AGAACAGCCTGTACAG	CAGAACATCCTGTACC	AAGGACAGCCTGTCCT	TAGGACAATTTGTACT
AGAACAGTATGTGCAA	TGGAACATCCTGTACC	AGGACAGCCTGTCCTC	AGCACAGGGTGTTCCC
GGGTACATTCTGTCCT	CAGTACACACTGTACT	CGGTACACTGTGTCCT	GGAACAGGATGTCCTG
CAGGACACTGTGTCCT	AGTACACACTGTACTT	AAGAACAGACTGTCCT	AAGAACATGGTGTCCT
AGGACACTGTGTCCTG	GAGTACAGCTTGTCCT	AGAACAGACTGTCCTT	AGAACATGGTGTCCTG
GAGAACAGCATGTCCT	AGTACAGCTTGTCCTG	ATGTACAGCCTGTGCT	GGAACAGGATGTTCTT
AGAACAGCATGTCCTG	TGGTACACTGTGTTCT	GAGAACAAGCTGTGCT	AGCACAGACTGTTCCC
AGTACACGATGTACAA	CTGTACATACTGTGCT	AGAACAAGCTGTGCTT	AAGTACAGTATGTACT
AGAACAAGCTGTTCCT	CAGAACATCCTGTTCC	GAGGACAGTGTGTTCT	AGTACAGTATGTACTG
GGGGACATGCTGTCCT	AGGGACATTCTGTTCT	AGGACAGTGTGTTCTA	GGAACAGGATGTGCTC
CAGGACACTGTGTCCT	AGAACAGTGTGTTCAT	AGCACAGAAAGTTCTG	GGTACAGAATGTTCCA
AGGACACTGTGTCCTC	TGGGACAGCCTGTGCT	AGGGACACGCTGTTCT	AAGGACATCCTGTGCT
CAGGACACTGTGTCCT	GTGAACAGTCTGTACT	AGAACAGAGTGTTCGA	AGGACATCCTGTGCTG
AGGACACTGTGTCCTC	AGAACAGTACGTTCTG	TTGCACATGCTGTTCT	GAGAACAAGGTGTCCT
CAGGACACTGTGTCCT	AGGACAGGCTGTTCCA	GGGACAGAATGTTCAG	AGAACAAGGTGTCCTG
AGGACACTGTGTCCTC	TAGGACATGCTGTTCC	AGCACAGGACGTGCTC	AAGTACAGGGTGTTCT

```
CAGGACACTGTGTCCT    CAGAACACCCTGTTCT    GGAACAGAATGTTCCC    AGTACAGGGTGTTCTG
AGGACACTGTGTCCTC    AGAACACCCTGTTCTG    GAGGACATGCTGTGCT    GAGGACATACTGTACC
CAGCACAGTCTGTCCC    AAGAACATCCTGTGCC    AGGACATGCTGTGCTC    AGGACATACTGTACCT
AGCACAGTCTGTCCCC    GAGTACAACCTGTTCT    GAGGACAGAATGTTCT    GGGCACAATCTGTACT
TAGAACATTCTGTGCT    AGTACAACCTGTTCTC    AGGACAGAATGTTCTG    AGGGACATAATGTGCT
AGAACATTCTGTGCTC    GGGACACAATGTCCTC    GAGGACAGAATGTGCT    GGGACATAATGTGCTG
AGGACAGAATGTTCCA    AAGTACAATGTGTGCT    AGGACAGAATGTGCTT    CAGAACAAACTGTGCT
GGGACAGAGTGTCCTC    AGTACAATGTGTGCTA    AAGGACATTTTGTGCT    AGAACAAACTGTGCTC
TGGAACACTATGTACT    CAGAACAGCATGTACT    AAGTACATTTTGTTCT    TTGCACATGGTGTTCT
AGTACAGCCTGTTCCC    AGAACAGCATGTACTC    GGGACAGAATGTGCCT    GGGCACATGCTGTCCT
GAGAACATTGTGTTCT    AGGGACAGTTTGTTCT    GGGTACAGGCTGTACT    CTGTACACACTGTCCT
AGAACATTGTGTTCTG    AAGAACAGAATGTTCT    GAGAACAGCGTGTTCT    GAGCACATTTTGTCCT
GAGGACAGTCTGTGCT    AGAACAGAATGTTCTT    AGAACAGCGTGTTCTT    AGCACAGGATGTGCCC
AGGACAGTCTGTGCTC    CAGTACATTATGTTCC    TAGAACAGCCTGTCCT    AGTACATCATGTACAT
AAGTACAACTTGTCCT    AGTACATTATGTTCCC    AGAACAGCCTGTCCTC    GGAACAGCATGTGCTA
GAGGACACTGTGTCCT    CAGAACACTTTGTCCT    GAGTACAGAGTGTTCT    AGGACAGGATGTTCCA
AGGACACTGTGTCCTG    GAGGACAGCCTGTTCT    AGTACAGAGTGTTCTG    AGGGACACCCTGTCCT
AGAACAGTATGTTCAA    AGGACAGCCTGTTCTT    CGGAACATTTTGTCCT    CAGAACATCCTGTGCC
AGAACACACTGTACCC    AGGCACATTCTGTACT    AGGTACAGACTGTTCT    CAGAACATCCTGTGCC
TGGTACACTCTGTACT    AGGACAGAATGTTCCG    GGTACAGACTGTTCTT    CAGAACATCCTGTGCC
GAGAACACAGTGTTCT    CTGCACAATCTGTCCT    AGAACAGCATGTTCCT    CAGAACATCCTGTGCC
AGAACACAGTGTTCTA    AGGCACCATGTTCCT     CAGAACATGCTGTGCT    CAGAACATCCTGTGCC
AGCACAGAGTGTGCCA    CAGAACATCTTGTTCC    AGAACATGCTGTGCTC    CAGAACATCCTGTGCC
AGAACAGGATGTGCAT    AAGCACAGCCTGTCCT    GGAACACAATGTCCTG    CAGAACATCCTGTACC
CAGGACAGGCTGTTCT    AGCACAGCCTGTCCTT    GGGAACATCATGTTCT    CAGAACATCCTGTGCC
AGGACAGGCTGTTCTT    GAGAACAGGCTGTTCT    CAGAACAGGATGTTCT    AGCACAGGATGTCCCT
GGGACAGAATGTCCTG    AGAACAGGCTGTTCTC    AGAACAGGATGTTCTG    CAGGACATTGTGTCCA
AAGGACAGGGTGTTCT    CAGGACATCGTGTACC    AGGACAGGATGTCCCA    CAGTACATACTGTACT
AGGACAGGGTGTTCTA    CTGTACACTCTGTTCT    CAGAACACCCTGTACT    AGTACATACTGTACTG
AGGACAGTACGTTCTG    ATGAACATAATGTTCT    AGAACACCCTGTACTG    TTGAACATGTTGTACT
TGGAACACTCTGTCCT    CAGAACAGAGTGTCCT    CAGCACATTCTGTTCC    GAGCACATTTTGTTCA
CAGAACATTTTGTACC    AGAACAGAGTGTCCTG    TAGAACATTATGTTCT    TTGTACATGTTGTTCT
CAGTACAGTGTGTGCT    CAGAACAATTTGTTCT    AGAACATTATGTTCTA    AAGTACACTCTGTACT
AGTACAGTGTGTGCTT    AGAACAATTTGTTCTC    AGTACAGGGTGTCCCA    AGTACACTCTGTACTC
TTGAACATACTGTGCT    CAGAACACTGTGTACT    AGGACACCATGTCCAT    AAGAACAGAGTGTGCT
AAGTACATTTTGTCCT    AGAACACTGTGTACTC    GAGAACAGCATGTTCT    AGAACAGAGTGTGCTG
TAGGACATTTTGTTCC    GAGTACACAATGTGCT    AGAACAGCATGTTCTT    AGGACAAGTTGTACTT
GGCACAGGATGTCCTC    AGTACACAATGTGCTA    CAGCACATTCTGTCCC    GGGGACATTGTGTCCT
AGAACAGGTTGTGCCG    AGTACAAGATGTGCCC    CAGGACAGCTTGTCCT    AGGCACAGTTTGTTCT
AAGAACAGACTGTCCT    AAGAACACAATGTCCT    AGGACAGCTTGTCCTG    TGGAACATTCTGTTCC
AGAACAGACTGTCCTA    AGAACACAATGTCCTG    TAGAACATTCTGTCCC    GGAACAGAATGTACTT
```

AAGCACAGGATGTTCT	GGAACAGCATGTTCCA	AGGCACAGTGTGTACT	AAGAACATGATGTGCT
AGCACAGGATGTTCTT	AGGACAGGGTGTCCCG	AGGCACAGTGTGTACT	AGAACATGATGTGCTC
CAGCACAGAATGTTCT	AGCACAGTGTGTTCCA	AGGCACAGTGTGTACT	GGGGACACTCTGTTCT
AGCACAGAATGTTCTT	GGGAACAGTCTGTGCT	AGGCACAGTGTGTACT	AGCACACAATGTTCGT
AGAACAGGGTGTTCCC	GAGTACATCTTGTCCT	GGCACAAAATGTGCTT	GGGACAGCATGTTCCA
GGGACAGGATGTCCCC	CAGGACATTCTGTTCT	GGGAACAGGCTGTTCT	GGGGACATTTGTGCT
CAGCACATTCTGTTCC	AGGACATTCTGTTCTT	AGTACACACTGTTCCT	AGAACAGAACGTTCTC
TAGGACATTCTGTTCT	GAGAACAATTTGTGCT	CAGGACATGCTGTGCT	GGAACAGGATGTTCCA
AGGACATTCTGTTCTT	AGAACAATTTGTGCTG	AGGACATGCTGTGCTG	AAGCACATCATGTCCT
GGCACAGAATGTCCCC	AGAACAGCATGTGCAG	CAGAACAGCATGTCCT	AGCACATCATGTCCTT
AAGAACAGGATGTTCT	AGCACAGGCTGTTCCT	AGAACAGCATGTCCTG	TGAACAGGATGTCCTA
AGAACAGGATGTTCTG	AGGGACAGGCTGTGCT	AGGACAGAGTGTACCG	CAGAACATGCTGTGCA
TAGGACATTCTGTGCC	AGGACAGAGTGTGCAT	GAGTACACGGTGTGCT	TGTACAGCATGTACTG
GGGGACAGCCTGTTCT	AAGTACACTCTGTCCT	AGTACACGGTGTGCTG	TTGAACATGGTGTGCT
CAGTACACTCTGTGCT	AGTACACTCTGTCCTG	GAGAACATAATGTTCT	TGGGACACACTGTCCT
AGTACACTCTGTGCTT	AGAACAAAATGTGCAT	AGAACATAATGTTCTA	TGGGACACACTGTCCT
TGGTACAGTTTGTACT	TGGAACAGTTTGTCCT	CAGGACAGGCTGTGCT	TTGTACATGGTGTTCT
TAGGACAGTGTGTGCT	GGAACAGAATGTGCTT	AGGACAGGCTGTGCTT	CAGGACAAACTGTTCT
AGGACAGTGTGTGCTC	AGCACAGGGTGTACAG	AAGTACATTCTGTTCT	AGGACAAACTGTTCTC
AGGACAGAATGTCCAC	TGGACAAAATGTACTA	AGTACATTCTGTTCTT	AGAACAGCATGTGCAA
AGTACAGCCTGTTCAG	AGGAACAGTCTGTTCT	GGGAACAGCCTGTGCT	CAGGACGTTCTGTACT
CAGGACAGACTGTGCT	CAGCACATTCTGTTCT	GAGCACATGCTGTTCA	CTGTACATGGTGTGCT
AGGACAGACTGTGCTT	AGCACATTCTGTTCTG	TGAACAGCATGTGCTC	GGAACAGAGTGTACTC
GGGCACACCCTGTGCT	GGTACAGAATGTACCC	CTGTACATTTTGTGCT	AGCACAAGATGTCCAC
TGGACAGAATGTTCTG	GGGAACATTCTGTACA	AGGACACAGTGTTCCT	TGAACAGAATGTACCG
CTGAACATTCTGTACC	CAGGACATTCTGTTCG	ATGAACATGTTGTTCT	AGGTACATACTGTCCT
AGAACAAGGTGTGCTG	AGTACAAGATGTACCT	AGGACAGGGTGTCCAT	
AGAACAGGCTGTACCC	CAGCACAGTCTGTACC	AGGGACATTCTGTGCT	
AGGACAAACTGTCCCA	AGCACAGTCTGTACCC	CAGAACAACGTGTTCT	

Appendix F: DNA oligomer sequences for thermal denaturation assay

WCWW_GRE_EMSA	GCATTGCTAGAACATTATGTTCTGCTCTCCC
WCWG_GRE_EMSA	GCATTGCAGAACAGTTTGTCCTGGCTCTCCC
WCWC_GRE_EMSA	GCATTGCAGTACACAGTGTTCTGGCTCTCCC
GWWC_Single_EMSA	GCATTGCAGTACAGGGTGTCCCAGCTCTCCC
GWWC_Double_EMSA	GCATTGCCAGAACATCCTGTTCTGCTCTCCC
WCWW_RT_GRE_EMSA	GGGAGAGCAGAACATAATGTTCTAGCAATGC
WCWG_RT_GRE_EMSA	GGGAGAGCCAGGACAAACTGTTCTGCAATGC
WCWC_RT_GRE_EMSA	GGGAGAGCCAGAACACTGTGTACTGCAATGC
GWWC_RT_Single_EMSA	GGGAGAGCTGGGACACCCTGTACTGCAATGC

GWWC_RT_Double_EMSA	GGGAGAGCAGAACAGGATGTTCTGGCAATGC
WCWC_TM_Fw	CGATACTCAAGC
WCWC_Tm_Re	GCTTGAGTATCG
WCWW_Tm_Fw	CGATACTTAAGC
WCWW_Tm_Re	GCTTAAGTATCG
WCWG_TM_Fw	CGATACTGAAGC
WCWG_Tm_Re	GCTTCAGTATCG
GWWC_TM_Fw	CGATGTTCAAGC
GWWC_TM_Re	GCTTGAACATCG
GGWC_TM_Fw	CGATGGTCAAGC
GGWC_TM_Re	GCTTGACCATCG
GCWC_TM_Fw	CGATGCTCAAGC
GCWC_TM_re	GCTTGAGCATCG
GWWC_GRE_Fw	AAGAACATCCTGTGCC
GWWC_GRE_Re	GGCACAGGATGTTCTT
GGWC_GRE_Fw	AAGGACACCGTGTGCT
GGWC_GRE_Re	AGCACACGGTGTCCTT
GW/GWC_GRE_Fw	AGGACATTCTGTTCTT
GW/GWC_GRE_Re	AAGAACAGAATGTCCT

Appendix G: Code for modeling genomic distribution of GREs and Transcription Starting Sites

(Mathematica)
Modeling distribution of genes and GREs in the genome
genomesizeminustelos = 0.7*3*10^9
2.1*10^9
downreg = Round[0.41*234]
96
upreg = 234 - downreg
138
plotsdistances =
Table[genes = Table[Random[], {i, 1, downreg}];
GREs = Table[Random[], {i, 1, 4392}]; genomesize = 0.7*3*10^9;
distances =
Sort[Table[
Min[Table[Min[Abs[genes[[i]] - GREs[[j]]]], {j, 1, 4392}]], {i,
1, downreg}]] ; mediandistance = 0.7*3*10^9*Median[distances];
realdistances = distances*genomesizeminustelos;
realdistancestoplot =
Table[{realdistances[[i]], i}, {i, 1, downreg}];
logtwo =

```
ListLogLinearPlot[{{10^3, .20*downreg}, {2*10^4, .50*
downreg}, {10^5, .85*downreg}, {2*10^6, downreg}},
AxesOrigin -> {10, 0}, PlotStyle -> Hue[0.1], Joined -> True];
logone =
ListLogLinearPlot[Reverse[realdistancestoplot],
AxesOrigin -> {10, 0}, Joined -> True]; logone, {k, 1, 25}];
mediandistancesplot =
Table[genes = Table[Random[], {i, 1, downreg}];
GREs = Table[Random[], {i, 1, 4392}]; genomesize = 3*10^9;
distances =
Sort[Table[
Min[Table[Min[Abs[genes[[i]] - GREs[[j]]]], {j, 1, 4392}]], {i,
1, downreg}]] ; notelomeres = 0.7;
mediandistance = notelomeres*genomesize*Median[distances];
mediandistance, {k, 1, 250}];
Mean[mediandistancesplot]
166589.
Print[Show[plotsdistances]]
-------------------------------------------
plotsdistances =
Table[genes = Table[Random[], {i, 1, downreg}];
GREs = Table[Random[], {i, 1, 4392}]; genomesize = 0.7*3*10^9;
distances =
Sort[Table[
Min[Table[Min[Abs[genes[[i]] - GREs[[j]]]], {j, 1, 4392}]], {i,
1, downreg}]] ; mediandistance = 0.7*3*10^9*Median[distances];
realdistances = distances*genomesizeminustelos;
realdistancestoplot =
Table[{realdistances[[i]], i}, {i, 1, downreg}];
logtwo =
ListLogLinearPlot[{{10^3, .20*downreg}, {2*10^4, .50*
downreg}, {10^5, .85*downreg}, {2*10^6, downreg}},
AxesOrigin -> {10, 0}, PlotStyle -> Hue[0.1], Joined -> True];
logone =
ListLogLinearPlot[Reverse[realdistancestoplot],
AxesOrigin -> {10, 0}, Joined -> True]; logone, {k, 1, 25}];
mediandistancesplot =
Table[genes = Table[Random[], {i, 1, downreg}];
GREs = Table[Random[], {i, 1, 4392}]; genomesize = 3*10^9;
distances =
```

```
Sort[Table[
Min[Table[Min[Abs[genes[[i]] - GREs[[j]]]], {j, 1, 4392}]], {i,
1, downreg}]] ; notelomeres = 0.7;
mediandistance = notelomeres*genomesize*Median[distances];
mediandistance, {k, 1, 250}];
Mean[mediandistancesplotup]
10732.4
Print[Show[plotsdistancesflat]]
```

References

1. White S, Szewczyk JW, Turner JM, Baird EE, & Dervan PB (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* 391(6666):468-471.

2. Dervan PB & Edelson BS (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Current opinion in structural biology* 13(3):284-299.

3. Gottesfeld JM, *et al.* (2001) Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J. Mol. Biol.* 309(3):615-629.

4. Meier JL, Yu AS, Korf I, Segal DJ, & Dervan PB (2012) Guiding the design of synthetic DNA-binding molecules with massively parallel sequencing. *Journal of the American Chemical Society* 134(42):17814-17822.

5. Erwin GS, Bhimsaria D, Eguchi A, & Ansari AZ (2014) Mapping polyamide-DNA interactions in human cells reveals a new design strategy for effective targeting of genomic sites. *Angew Chem Int Ed Engl* 53(38):10124-10128.

6. Bell O, Tiwari VK, Thoma NH, & Schubeler D (2011) Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12(8):554-564.

7. John S, *et al.* (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* 43(3):264-268.

8. Mitchell PJ & Tjian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245(4916):371-378.

9. Gilbert W & Maxam A (1973) The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U S A* 70(12):3581-3584.

10. Garner MM & Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res* 9(13):3047-3060.

11. Galas DJ & Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res* 5(9):3157-3170.

12. Johnson DS, Mortazavi A, Myers RM, & Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830):1497-1502.

13. Carter D, Chakalova L, Osborne CS, Dai YF, & Fraser P (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32(4):623-626.

14. Reddy TE, *et al.* (2009) Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* 19(12):2163-2171.

15. Dekker J, Rippe K, Dekker M, & Kleckner N (2002) Capturing chromosome conformation. *Science* 295(5558):1306-1311.

16. Dostie J & Dekker J (2007) Mapping networks of physical interactions between genomic elements using 5C technology. *Nat Protoc* 2(4):988-1002.

17. Zhao Z*, et al.* (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 38(11):1341-1347.

18. Simonis M, Kooren J, & de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4(11):895-901.

19. Gondor A, Rougier C, & Ohlsson R (2008) High-resolution circular chromosome conformation capture assay. *Nat Protoc* 3(2):303-313.

20. Kielkopf CL*, et al.* (1998) A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* 282(5386):111-115.

21. Luisi BF*, et al.* (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352(6335):497-505.

22. Muzikar KA, Nickols NG, & Dervan PB (2009) Repression of DNA-binding dependent glucocorticoid receptor-mediated gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 106(39):16598-16603.

23. Buttgereit F, Burmester GR, & Lipworth BJ (2005) Optimised glucocorticoid therapy: the sharpening of an old spear. *Lancet* 365(9461):801-803.

24. Leonard MB*, et al.* (2004) Long-term, high-dose glucocorticoids and bone mineral content in childhood glucocorticoid-sensitive nephrotic syndrome. *N Engl J Med* 351(9):868-875.

25. Brown TA (2007) *Genomes 3* (Garland Science Pub., New York) 3rd Ed pp xxii, 713 p.

26. Gearhart MD*, et al.* (2005) Inhibition of DNA binding by human estrogen-related receptor 2 and estrogen receptor alpha with minor groove binding polyamides. *Biochemistry* 44(11):4196-4203.

27. Nickols NG & Dervan PB (2007) Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U. S. A.* 104(25):10418-10423.

28. Dervan PB & Burli RW (1999) Sequence-specific DNA recognition by polyamides. *Curr Opin Chem Biol* 3(6):688-693.

29. Farkas ME, Li BC, Dose C, & Dervan PB (2009) DNA sequence selectivity of hairpin polyamide turn units. *Bioorg Med Chem Lett* 19(14):3919-3923.

30. Wang JC*, et al.* (2004) Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. *Proc. Natl. Acad. Sci. U. S. A.* 101(44):15603-15608.

31. Dose C, Farkas ME, Chenoweth DM, & Dervan PB (2008) Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J Am Chem Soc* 130(21):6859-6866.

32. Chenoweth DM, Harki DA, Phillips JW, Dose C, & Dervan PB (2009) Cyclic pyrrole-imidazole polyamides targeted to the androgen response element. *J Am Chem Soc* 131(20):7182-7188.

33. Best TP, Edelson BS, Nickols NG, & Dervan PB (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A* 100(21):12063-12068.

34. Nickols NG, Jacobs CS, Farkas ME, & Dervan PB (2007) Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem. Biol.* 2(8):561-571.

35. Olenyuk BZ*, et al.* (2004) Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc. Natl. Acad. Sci. U. S. A.* 101(48):16768-16773.

36. Hsu CF & Dervan PB (2008) Quantitating the concentration of Py-Im polyamide-fluorescein conjugates in live cells. *Bioorg Med Chem Lett* 18(22):5851-5855.

37. Wiley HS, Shvartsman SY, & Lauffenburger DA (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol* 13(1):43-50.

38. Chen M, *et al.* (2010) Pretranscriptional regulation of Tgf-beta1 by PI polyamide prevents scarring and accelerates wound healing of the cornea after exposure to alkali. *Mol. Ther.* 18(3):519-527.

39. Baliga R, *et al.* (2001) Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* 40(1):3-8.

40. Suto RK, *et al.* (2003) Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* 326(2):371-380.

41. Meier JL, Montgomery DC, & Dervan PB (2012) Enhancing the cellular uptake of Py-Im polyamides through next-generation aryl turns. *Nucleic Acids Res.* 40(5):2345-2356.

42. Turner JM, Baird EE, & Dervan PB (1997) Recognition of Seven Base Pair Sequences in the Minor Groove of DNA by Ten-Ring Pyrrole−Imidazole Polyamide Hairpins. *Journal of the American Chemical Society* 119(33):7636-7644.

43. Matsuda H, *et al.* (2006) Development of gene silencing pyrrole-imidazole polyamide targeting the TGF-beta1 promoter for treatment of progressive renal diseases. *J Am Soc Nephrol* 17(2):422-432.

44. Yan J, Wang HF, Liu YT, & Shao CX (2008) Analysis of Gene Regulatory Networks in the Mammalian Circadian Rhythm. *PLoS Comput. Biol.* 4(10).

45. Yan J, Wang H, Liu Y, & Shao C (2008) Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLoS Comput Biol* 4(10):e1000193.

46. Belitsky JM, Nguyen DH, Wurtz NR, & Dervan PB (2002) Solid-phase synthesis of DNA binding polyamides on oxime resin. *Bioorg. Med. Chem.* 10(8):2767-2774.

47. Puckett JW, Green JT, & Dervan PB (2012) Microwave assisted synthesis of Py-Im polyamides. *Org. Lett.* 14(11):2774-2777.