Inhibition of major-groove-binding proteins by pyrrole–imidazole polyamides with an Arg–Pro–Arg positive patch

Ryan E Bremer, Eldon E Baird and Peter B Dervan

Background: Gene-specific targeting of any protein–DNA complex by small molecules is a challenging goal at the interface of chemistry and biology. Polyamides containing *N*-methylimidazole and *N*-methylpyrrole amino acids are synthetic ligands that have an affinity and specificity for DNA comparable to many naturally occurring DNA-binding proteins. It has been shown that an eight-ring hairpin polyamide targeted to a specific minor-groove contact within a transcription factor binding site can inhibit protein–DNA binding and gene transcription. Polyamides and certain major-groove-binding proteins have been found to co-occupy the DNA helix, however. To expand the number of genes that can be targeted by pyrrole/imidazole polyamides, we set out to develop a class of polyamides that can selectively inhibit major-groove-binding proteins.

Results: An eight-ring hairpin polyamide conjugated to a carboxy-terminal Arg–Pro–Arg tripeptide was designed to deliver a positive residue to the DNA backbone and interfere with protein–phosphate contacts. Gel mobility shift analysis demonstrated that a polyamide hairpin–Arg–Pro–Arg binding in the minor groove selectively inhibits binding of the transcription factor GCN4 (222–281) in the adjacent major groove. Substitution within the Arg–Pro–Arg revealed that each residue was required for optimal GCN4 inhibition.

Conclusions: A pyrrole–imidazole polyamide that binds to a predetermined site in the DNA minor groove and delivers a positive patch to the DNA backbone can selectively inhibit a DNA-binding protein that recognizes the adjacent major groove. A subtle alteration of the DNA microenvironment targeted to a precise location within a specific DNA sequence could achieve both gene-specific and protein-specific targeting.

Addresses: Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, CA, 91125, USA.

Correspondence: Peter B Dervan
E-mail: dervan@cco.caltech.edu

Key words: DNA, hydrogen bond, minor groove, molecular recognition, transcription factor

Received: **28 November 1997**
Revisions requested: **30 December 1997**
Revisions received: **15 January 1998**
Accepted: **16 January 1998**

Published: **27 February 1998**

Chemistry & Biology March 1998, **5**:119–133
http://biomednet.com/elecref/1074552100500119

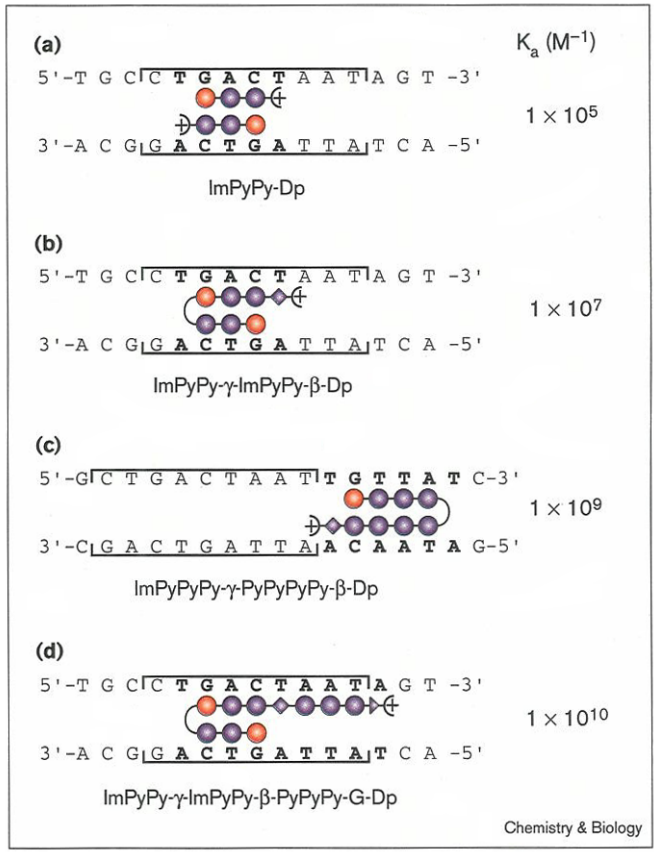
Introduction

Small molecules specifically targeted to any predetermined DNA sequence would be useful tools in molecular biology and, potentially, in human medicine. Polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids bind to predetermined sequences in the minor groove of DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins [1–3]. Sequence specificity is determined by a code of oriented side-by-side pairings of the Py and Im amino acids [4–9]. An Im/Py pairing targets a G–C base pair, while Py/Im pair recognizes C–G [4–6]. The Py/Py pair is degenerate and targets both A–T and T–A base pairs [4–12]. The validity of the pairing rules for ligand design is supported by a variety of polyamide structural motifs which have been characterized by footprinting, affinity cleaving, two-dimensional nuclear magnetic resonance (NMR) and X-ray methods [1–14]. Polyamides have been found to be cell permeable and to inhibit transcription-factor binding and expression of a designated gene [15,16]. Py/Im polyamides offer a potentially general approach for gene regulation, provided that efficient inhibition of DNA binding can be achieved for a variety of transcription factors.

Protein inhibition by synthetic ligands

Several approaches for the development of synthetic ligands that interfere with protein–DNA recognition have been reported. Oligodeoxyribonucleotides that recognize the major groove of double-helical DNA via triple-helix formation bind to a broad range of sequences with high affinity and specificity [17,18]. Although oligonucleotides and their analogs have been shown to disrupt protein–DNA binding [19–21], the triple-helix approach is limited to purine tracts and suffers from poor cellular uptake of oligodeoxyribonucleotides. There are a few examples of carbohydrate-based ligands that interfere with protein–DNA recognition, but oligosaccharides cannot currently recognize a broad range of DNA sequences [22,23]. Perhaps most relevant to this work, analogs of distamycin (PyPyPy) appended with multiple cationic substituents have been found to inhibit protein

Figure 1



Schematic models of polyamides targeted to the binding site of the bZIP transcriptional activator GCN4. **(a)** ImPyPy-Dp [4,35], **(b)** ImPyPy-γ-ImPyPy-β-Dp [56], **(c)** ImPyPyPy-γ-PyPyPyPy-β-Dp [1], **(d)** ImPyPy-γ-ImPyPy-β-PyPyPy-G-Dp [56]. The nine base pair (5′-CTGACTAAT-3′) GCN4-binding site is indicated by brackets above and below the base pairs. Red and blue circles represent imidazole (Im) and pyrrole (Py) polyamide rings, respectively. Blue diamonds and triangles represent β-alanine (β) and glycine (G), respectively. γ-Aminobutyric acid (γ) and dimethylaminopropylamide (Dp) are depicted as a curved line and a plus sign, respectively. Polyamide-binding sites are shown in bold. Equilibrium association constants (K_a) for each polyamide binding to the indicated match site are shown on the right. Association constants were determined by DNase I footprinting; simultaneous binding was determined by gel mobility shift assay.

binding [24–26]. On the basis of these encouraging results, we wished to identify similar charged residues that could be appended to a Py/Im polyamide via linear solid-phase synthesis and that would not compromise polyamide-binding specificity.

Pyrrole–imidazole polyamides inhibit proteins that recognize the minor groove

Proteins use a diverse structural library to recognize their target sequences [27]. Proteins such as TATA-box-binding protein (TBP) bind exclusively in the minor groove [28], others, such as the transcriptional activator GCN4 [29–31], bind exclusively in the major groove, and certain proteins such as Hin recombinase recognize both grooves [32,33]. Polyamides have been found to interfere with protein–DNA recognition in cases where contacts in the minor groove are important for protein–DNA binding affinity. For example, within the nine zinc-finger transcription factor TFIIIA, fingers 4 and 6 bind in or across the minor groove and are required for high affinity binding ($K_a = 5 \times 10^9\ M^{-1}$). An eight-ring hairpin polyamide ($K_a = 3 \times 10^{10}\ M^{-1}$) targeted to the minor-groove contact region of finger 4 has been recently found to efficiently inhibit protein binding [15,16].

Pyrrole–imidazole polyamides bind simultaneously in the minor groove with proteins that recognize the major groove

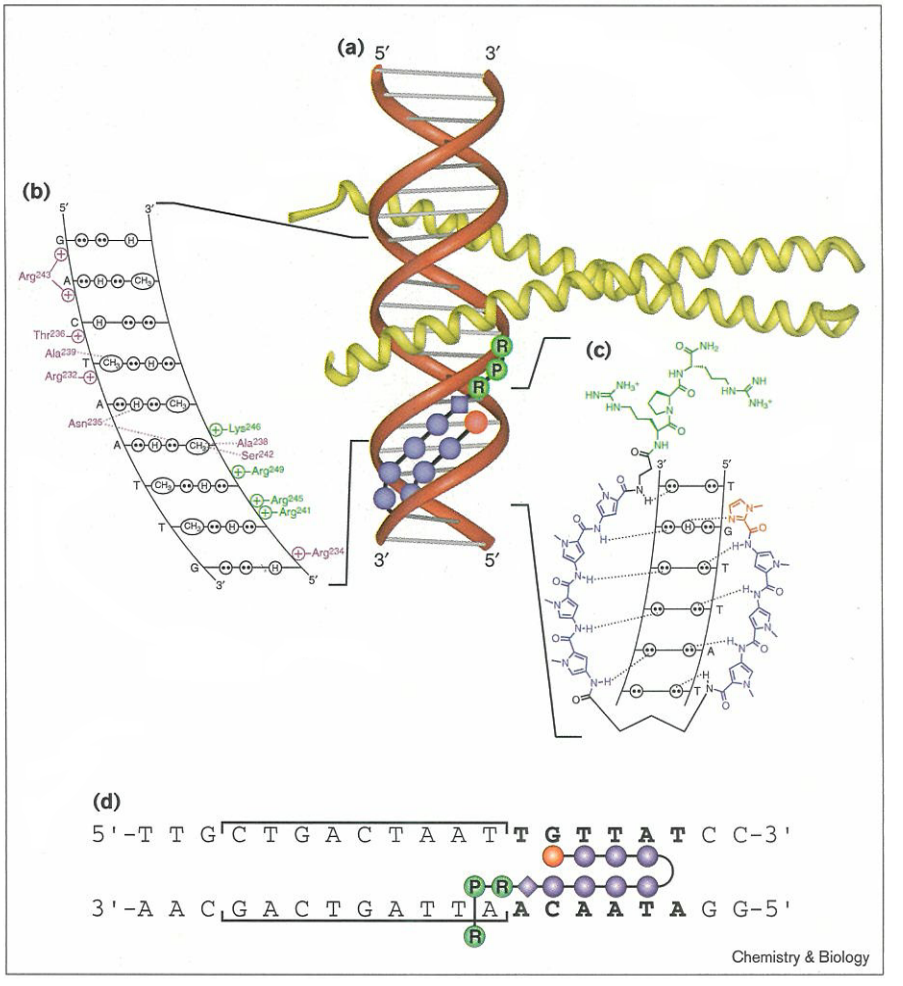
X-Ray crystallography studies reveal that DNA bound by a four-ring homodimeric polyamide is unaltered from its natural B-form structure; all polyamide–DNA contacts are confined to the minor groove [34]. Not surprisingly, polyamides have been found to bind simultaneously with ligands that exclusively occupy the major groove [16,35,36]. For example, an eight-ring hairpin polyamide and a recombinant protein containing only the three amino-terminal zinc fingers of TFIIIA that are in the major groove were found to co-occupy the TFIIIA-binding site [16]. Similarly, the three-ring homodimer ImPyPy bound simultaneously with the basic-region leucine zipper (bZIP) protein GCN4 (226–281) [35]. We found subsequently that a variety of polyamide motifs co-occupy the DNA helix at sites both overlapping and adjacent to GCN4 (Figure 1). The ubiquity of major-groove contacts in protein–DNA recognition provides the impetus to develop approaches for the inhibition of major-groove proteins by polyamides that bind in the minor groove.

Design of the positive patch

The negatively charged DNA phosphate backbone provides a target for ligands designed to disrupt the unique microenvironment that represents a protein's binding site on the DNA double helix. Polyamides that deliver a positive patch to the DNA backbone may destabilize contacts between protein sidechains and phosphate residues and consequently inhibit protein binding [24–26] (Figure 2). We looked to nature for an α-amino acid model of a domain bound to the minor groove that delivers a cationic sidechain to the phosphate backbone.

Homeodomain proteins recognize the minor groove of DNA via a highly conserved Arg–X_{aa}–Arg (where X_{aa} is any amino acid) [37,38]. In Hin recombinase, the corresponding Arg140–Pro141–Arg142 domain serves as a bridge between the amino-terminal arm in the minor groove and the helix–turn–helix motif that recognizes the major groove (Figure 3) [32,33]. Minor groove contacts made by the sidechain of Arg140 direct the peptide chain up from

Figure 2



A schematic model of Arg–Pro–Arg (RPR) polyamides targeted to the major groove transcription factor GCN4. **(a)** The α-helical GCN4 dimer (yellow) is shown binding to adjacent major grooves [30]. The Arg–Pro–Arg–hairpin polyamide is shown as red, blue and green balls which represent imidazole, pyrrole and Arg–Pro–Arg amino acids, respectively. The blue diamond represents β-alanine. γ-Aminobutyric acid is designated as a curved line. **(b)** The contacts between one GCN4 monomer and the major groove of one half-site of 5′-CTGACTAAT-3′ are depicted (adapted from [30]). Circles with two dots represent the lone pairs of the N7 of purines, the O4 of thymine and the O6 of guanine. Circles containing an H represent the N6 and N4 hydrogens of the exocyclic amines of adenine and cytosine, respectively. The C5 methyl group of thymine is depicted as a circle with CH_3 inside. Protein sidechains that make hydrogen bonds or van der Waals contacts to the bases are shown in purple and connected to the DNA via a dotted line. Green and purple plus signs represent protein residues that electrostatically contact the phosphate backbone. The residues that are predicted to be disrupted by an Arg–Pro–Arg polyamide are shown in green. **(c)** The hydrogen-bonding model of the eight-ring hairpin polyamide ImPyPyPy-γ-PyPyPyPy-β-RPR bound to the minor groove of 5′-TGTTAT-3′. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. Py and Im rings are represented as blue and red rings, respectively. The Arg–Pro–Arg moiety is green. **(d)** The model of the polyamide binding its target site (bold) adjacent to the GCN4 binding site (brackets). Polyamide residues are as in (a).

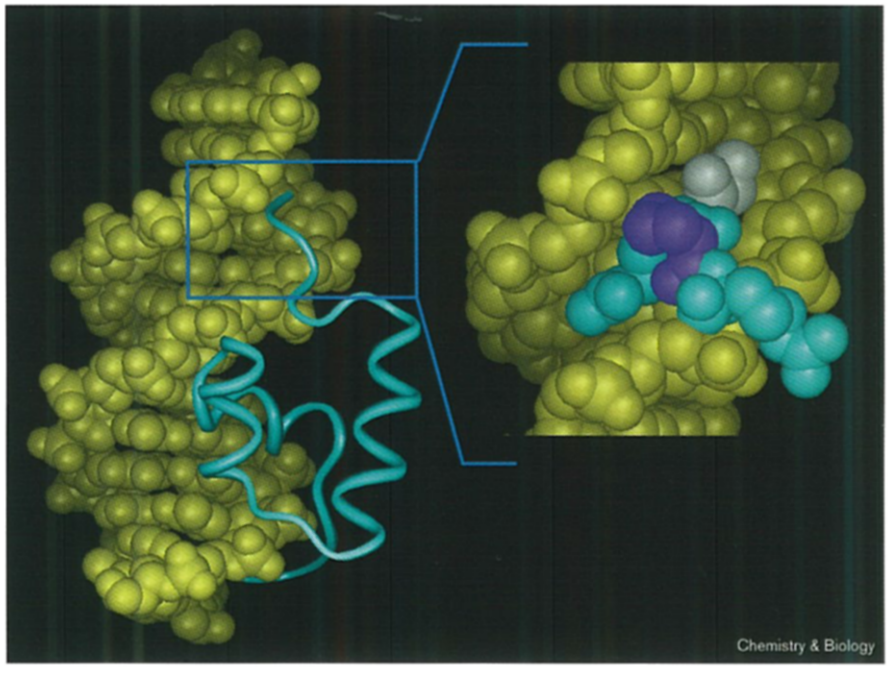
the floor of the minor groove, towards the backbone, where the guanidinium of Arg142 makes electrostatic contact with phosphates. Upon interaction with DNA, the Arg–Pro–Arg (RPR) domain achieves a stable, local tertiary structure which is potentially based solely on the primary sequence. It was postulated that Arg–Pro–Arg attached at the carboxyl terminus of a polyamide would adopt a similar structure to that of Arg140–Pro141–Arg142 in Hin recombinase. The resulting Arg–Pro–Arg polyamide could be used to place a neutralizing positive charge at a predetermined phosphate on the DNA backbone (Figure 2).

Arg–Pro–Arg polyamides

A series of polyamides that have Arg–Pro–Arg tripeptides at the carboxyl terminus have been synthesized by solid-phase methods. The polyamides were evaluated as inhibitors of the major-groove transcription factor GCN4, the prototypical member of the bZIP family of transcriptional regulators [39–41]. The carboxy-terminal 60 amino acids (222–281) of GCN4 contain the 'leucine zipper' dimerization domain and the 'basic region' which is responsible for DNA binding. GCN4 (222–281) has been shown to be sufficient for sequence-specific binding [29,42,43]. The basic region of each α-helical monomer makes specific hydrogen bonds, van der Waals contacts and phosphate interactions with one half-site of the nine base pair pseudosymmetrical GCN4-binding site (Figure 2a,b) [29–31]. The protein–DNA electrostatic interactions that are targeted for disruption by the Arg–Pro–Arg polyamides are highlighted in Figure 2b.

We report here the ability of Arg–Pro–Arg polyamides to selectively inhibit DNA binding by the major-groove transcription factor GCN4 as measured by gel mobility

Figure 3



The structure of the DNA-binding domain of Hin recombinase bound to DNA. On the left is a ribbon model of residues 139–190 of Hin recombinase (blue) bound to the 5′-GTTTTT-GATAAGA-3′ DNA duplex (yellow) [33]. On the right is an enlarged view of the amino-terminal Gly139–Arg140–Pro141–Arg142 domain reaching from the floor of the minor groove to the phosphate backbone. Glycine, arginine, and proline are white, blue and purple, respectively.

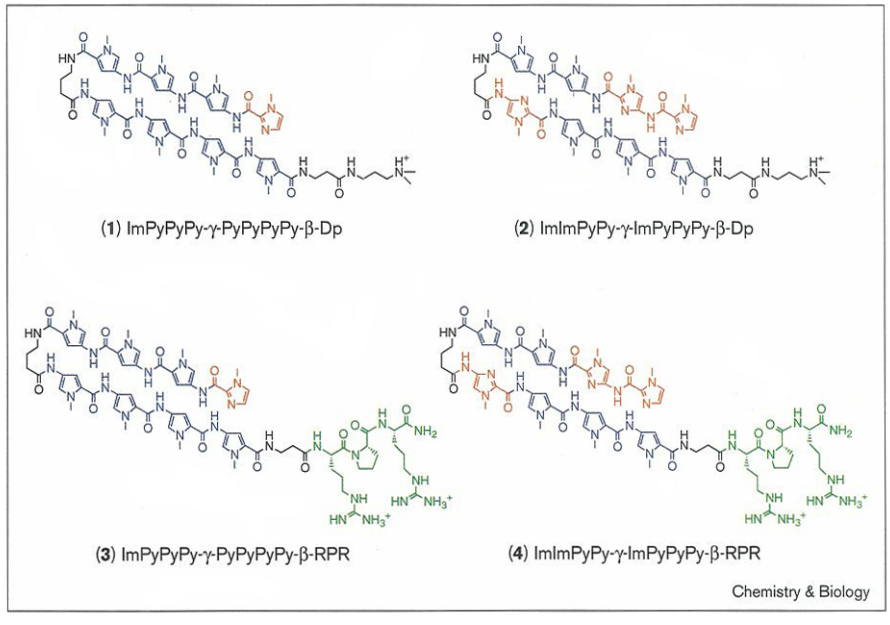
shift analysis. Substitutions in the Arg–Pro–Arg domain revealed the relative importance of each residue for protein inhibition. Quantitative DNase I footprint titration experiments were performed in order to measure the effect of net ligand charge on both the DNA-binding affinity and specificity of the hairpin polyamide.

Results

Synthesis of Arg–Pro–Arg polyamides

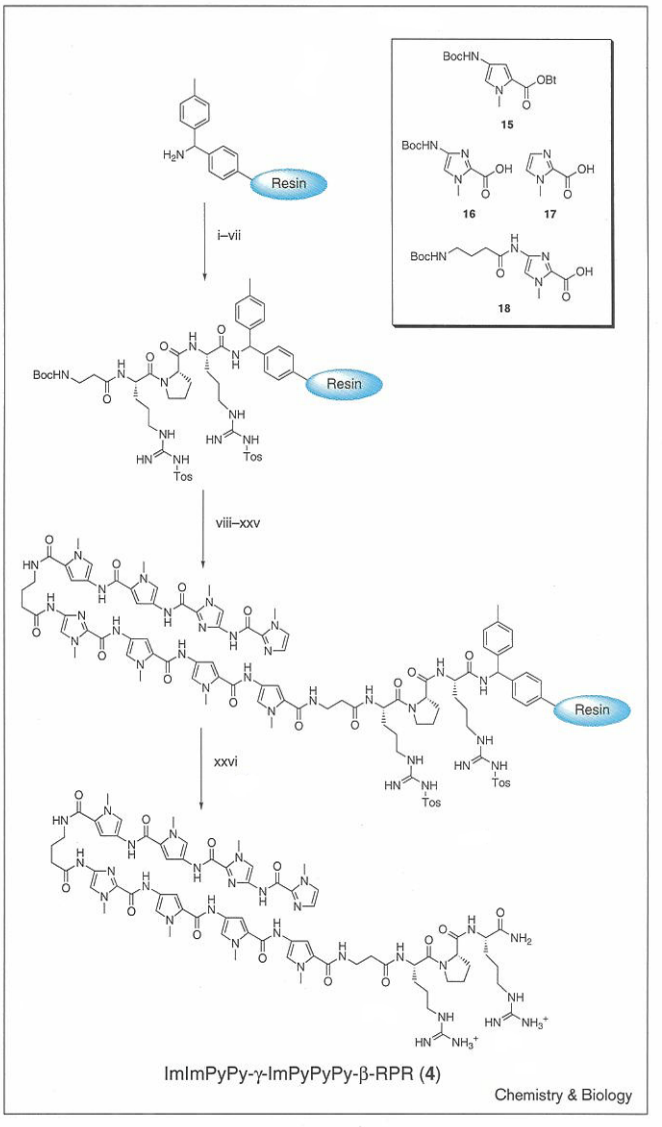
The polyamides ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and ImImPyPy-γ-ImPyPyPy-β-Dp (**2**) were synthesized in a stepwise manner from Boc–alanine–Pam resin using Boc-chemistry machine-assisted protocols as described

Figure 4



Structures of eight-ring hairpin polyamides, ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) and ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), and their Arg–Pro–Arg (RPR) analogs, ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) and ImImPyPy-γ-ImPyPyPy-β-RPR (**4**). Imidazole and pyrrole rings and the Arg–Pro–Arg domain are shown in red, blue and green, respectively. MALDI-TOF mass spectrometry (MS): **2**, 1223.6 (1223.6 calc'd); **3**, 1546.3 (1545.7 calc'd); **4**, 1547.8 (1547.7 calc'd).

Figure 5



Synthetic scheme for solid phase preparation of Arg–Pro–Arg polyamides. Cycling protocols consist of trifluoroacetic acid (TFA) deprotection followed by coupling with HOBt-activated Boc–pyrrole (**15**), Boc–imidazole (**16–18**), or aliphatic amino acid ester (steps i–xxv). The MBHA resin is then cleaved by treatment with HF:cresol (9:1) and purified by reverse phase HPLC (step xxvi). The synthesis of ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) is shown as a representative example.

previously [44]. Polyamides that have carboxy-terminal aliphatic amino acids were synthesized on 4-methylbenzhydrylamine (MBHA) resin from Im and Py monomer units and commercially available aliphatic amino acids in 26 steps (Figures 4 and 5). Treatment with HF:*p*-cresol (9:1) followed by precipitation with ethyl ether and extraction with 0.1% TFA:CH_3CN (50:50) afforded the deprotected polyamide which was purified by reverse phase high performance liquid chromatography (HPLC).

Arg–Pro–Arg polyamides selectively inhibit GCN4 (222–281) binding

Synthetic radiolabeled DNA duplexes, ARE-1 and ARE-2, containing a GCN4-binding site (5′-CTGACTAAT-3′) [29,35], were bound near saturation at 200 nM GCN4 (222–281) as revealed by gel mobility shift analysis (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM dithiothreitol (DTT), 1 mM EDTA, 50 μg/ml poly(dI-dC)–poly(dI-dC), 22°C). ImPyPyPy-γ-PyPyPyPy-β-polyamides (**1** and **3**) target the six-base-pair 5′-TGTTAT-3′ site adjacent to the GCN4-binding site of ARE-1 (Figures 2, 4 and 6a,c). ImImPyPy-γ-ImPyPyPy-β-polyamides (**2** and **4**) were designed to bind 5′-TGGTCT-3′ adjacent to the GCN4 site in ARE-2 (Figures 4 and 6b,f).

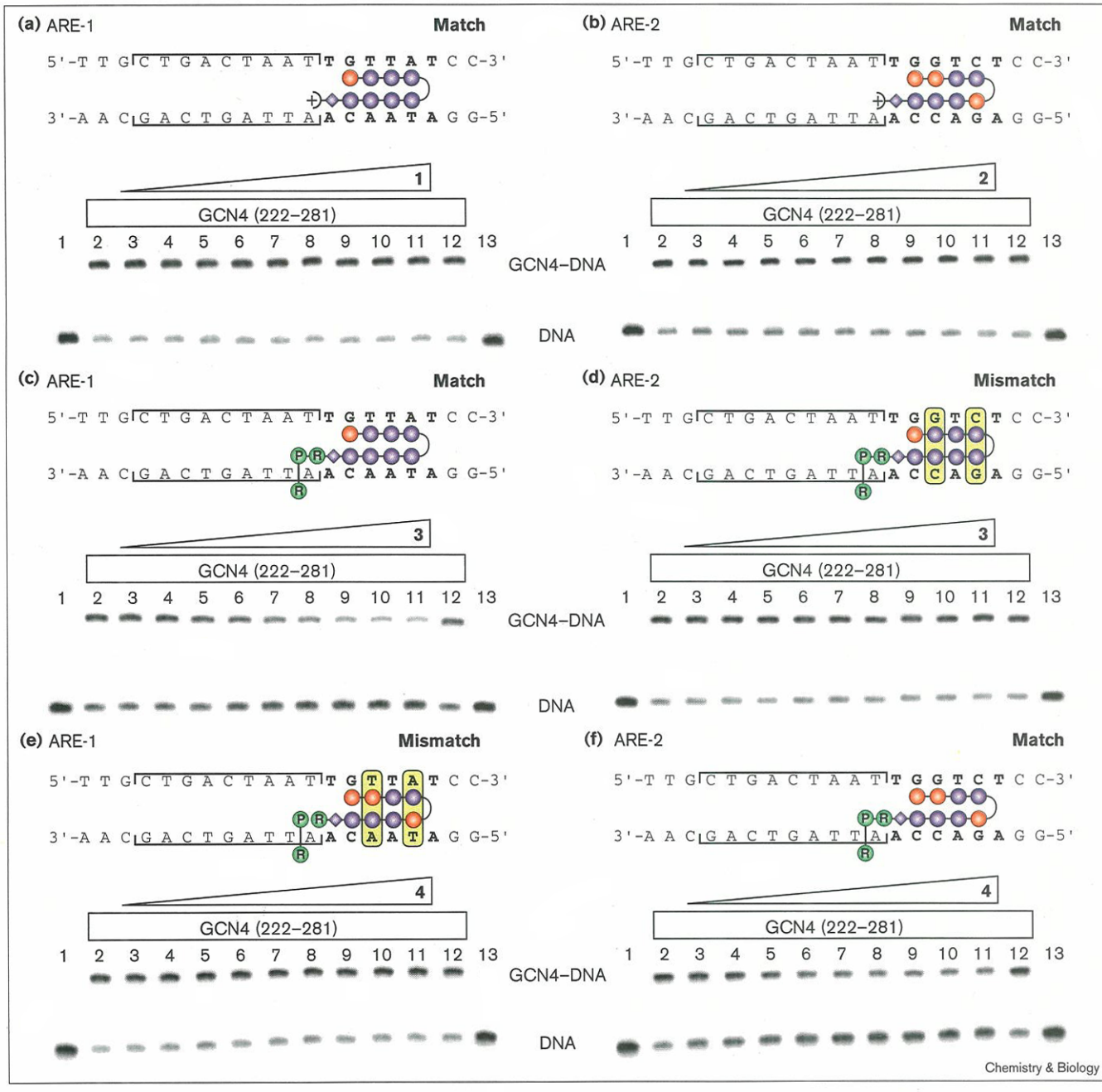
The ability of polyamides to inhibit GCN4 (222–281) binding was evaluated using the gel mobility shift assay. Increasing concentrations of polyamide were incubated with the desired radiolabeled synthetic DNA duplex followed by the addition of a constant concentration of 200 nM GCN4 (222–281). DNA fragments bound and unbound by GCN4 were separated using nondenaturing polyacrylamide gel electrophoresis.

Polyamides that lack the Arg–Pro–Arg moiety were unable to inhibit GCN4 binding (**1** and **2**, Figure 6a,b). The upper band in Figure 6a is the ARE-1 fragment bound by GCN4 (222–281) (lanes 2–12). Lanes 3–11 show that GCN4 binding was unaffected by the addition of ImPyPyPy-γ-PyPyPyPy-β-Dp. ImPyPyPy-γ-PyPyPyPy-β-RPR, however, which differs from **1** by the addition of the carboxy-terminal Arg–Pro–Arg, inhibited GCN4 binding to ARE-1 (Figure 6c). When bound to its match site on ARE-2, ImImPyPy-γ-ImPyPyPy-β-RPR also successfully inhibited GCN4 binding (Figure 6f).

On the basis of the pairing rules for polyamide–DNA complexes, the sites 5′-TGTTAT-3′ (ARE-1 fragment) and 5′-TGGTCT-3′ (ARE-2 fragment) are for ImPyPyPy-γ-PyPyPyPy-β-RPR 'match' and 'double-base-pair mismatch' sites, respectively, and for ImImPyPy-γ-ImPyPyPy-β-RPR 'double-base-pair mismatch' and 'match' sites, respectively. Incubation of GCN4 and up to 2 μM ImPyPyPy-γ-PyPyPyPy-β-RPR with the double mismatch ARE-2 fragment resulted in no inhibition of GCN4 binding (Figure 6d). Likewise, ImImPyPy-γ-ImPyPyPy-β-RPR did not inhibit GCN4 binding to the mismatched ARE-1 fragment (Figure 6e).

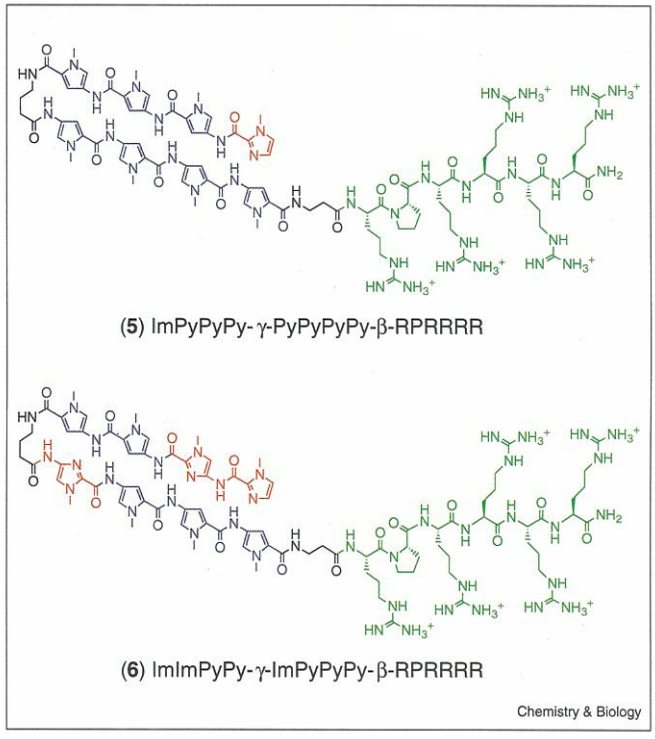
When bound to their respective match sites, ImPyPyPy-γ-PyPyPyPy-β-RPRRRR and ImImPyPy-γ-ImPyPyPy-β-RPRRRR, which contain three more carboxy-terminal arginine residues than **3** and **4**, were found to fully inhibit GCN4 (222–281) binding (Figures 7 and 8). The gel mobility shift experiments shown in Figure 8 demonstrate that **5** and **6** selectively provided complete inhibition of

Figure 6



GCN4 (222–281) gel mobility shift experiments in the presence of Dp and Arg–Pro–Arg polyamides. At the top of each panel is a model of polyamide binding the match or mismatch site of DNA fragment ARE-1 or ARE-2 adjacent to the GCN4-binding site. The polyamide is colored as in Figure 1a. Mismatches are indicated by a yellow box. At the bottom is a storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222–281) binding to the radiolabeled ARE-1 or ARE-2 fragments in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 μg/ml poly(dI-dC)–poly(dI-dC), 22°C). Upper band, GCN4(222–281)–DNA complex; lower band, free DNA; lanes 1 and 13, DNA only. Lanes 2 and 12, DNA incubated with 200 nM GCN4 (222–281). (**a**,**c**,**d**) Lanes 3–11 are 200 nM GCN4 (222–281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM and 2 μM **1** or **3**. (**b**,**e**,**f**) Lanes 3–11, 200 nM GCN4 (222–281) and 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM and 500 nM **2** or **4**. **(a)** ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**) binding the match site 5′-TGTTAT-3′ of ARE-1. **(b)** ImImPyPy-γ-ImPyPyPy-β-Dp (**2**) binding the match site 5′-TGGTCT-3′ of ARE-2. **(c)** ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) binding the match site 5′-TGTTAT-3′ of ARE-1. **(d)** ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) incubated with the double-mismatch site 5′-TGGTCT-3′ of ARE-2 (mismatch sites italicized). **(e)** ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) incubated with the double-mismatch site 5′-TG*T*T*A*T-3′ of ARE-1. **(f)** ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) binding the match site 5′-TGGTCT-3′ of ARE-2. Lower K_a values ($K_a \cong 1 \times 10^7\ M^{-1}$) are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide-binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.

Figure 7



Structures of the Arg–Pro–Arg–Arg–Arg–Arg polyamides ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**). Imidazole and pyrrole rings and the Arg–Pro–Arg–Arg–Arg–Arg domain are shown in red, blue and green, respectively. MALDI-TOF MS: **5**, 2014.0 (2014.1 calc'd); **6**, 2017.1 (2016.0 calc'd).

GCN4 binding with no apparent loss in affinity for double-base-pair mismatches.

Arg–Pro–Arg is the optimum tripeptide for GCN4 inhibition

Polyamides that have deletions and/or substitutions in the Arg–Pro–Arg domain were prepared in order to determine the elements that were essential for GCN4 inhibition (Figure 9). Each of these polyamides was based on the ImPyPyPy-γ-PyPyPyPy-β polyamide targeted to 5′-TGTTAT-3′ of ARE-1 (Figure 9). The ability of polyamides **7–14** to bind to their DNA target sites and inhibit GCN4 binding to ARE-1 was evaluated using DNase I footprinting and gel mobility shift analysis.

DNase I footprinting of polyamides **1–14** was performed on restriction fragments containing the appropriate ARE-1 or ARE-2 sequences under conditions identical to those used for the gel mobility shift experiments. In every case (except **11**, see below), the polyamide was found to specifically bind to the target site with $K_a \cong 1 \times 10^7\ M^{-1}$. Lower K_a values are observed for polyamides under the gel shift conditions because the carrier DNA artificially depresses polyamide-association constants. The polyamide concentrations required for GCN4 inhibition are within the expected range on the basis of the K_a under gel shift conditions.

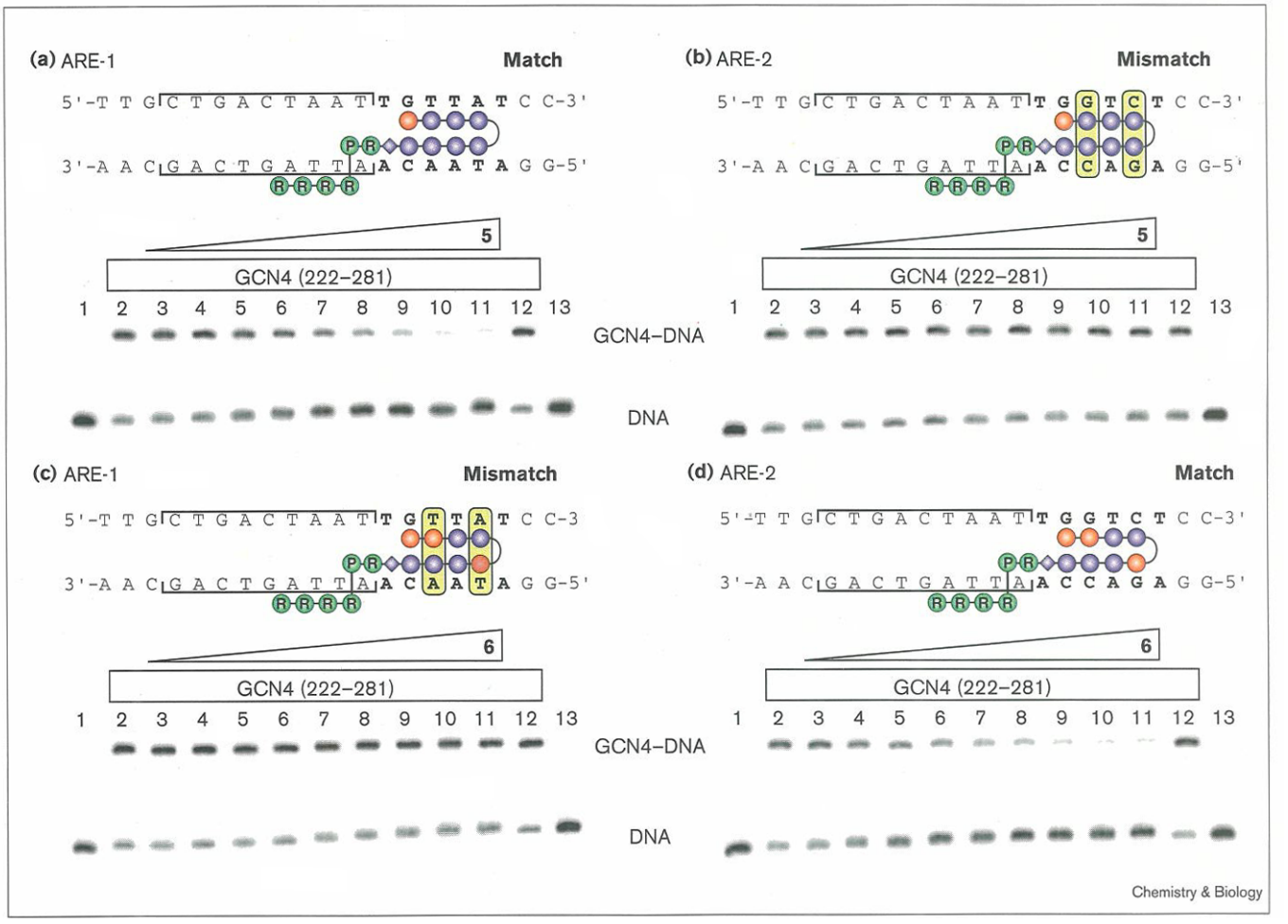
Deletion of the terminal Pro–Arg or Arg, as in ImPyPyPy-γ-PyPyPyPy-β-R (**7**) and ImPyPyPy-γ-PyPyPyPy-β-RP (**8**), results in polyamides that are unable to inhibit GCN4 binding. Substituting the proline with glycine afforded ImPyPyPy-γ-PyPyPyPy-β-RGR (**9**) which did not effectively inhibit GCN4 (222–281). At 1 μM of **9**, < 50% of the GCN4 was inhibited. No inhibition of GCN4 binding was observed for ImPyPyPy-γ-PyPyPyPy-β-R^{D}PR (**10**) which contained a single inversion of stereochemistry relative to **3**.

The internal arginine residue was replaced with an alanine residue to provide ImPyPyPy-γ-PyPyPyPy-β-APR (**11**). **11** was unable to inhibit GCN4 binding under these conditions. This Arg→Ala substitution was the only alteration that was found to affect polyamide-binding affinity. As measured by DNase I footprinting, **11** binds the 5′-TGTTAT-3′ target site with tenfold lower affinity than **3** under conditions identical to those used for gel shift analysis. The conservative substitution of lysine (K) for arginine in ImPyPyPy-γ-PyPyPyPy-β-KPR (**12**) also compromised the polyamide's ability to inhibit GCN4. At 1 μM **12**, < 50% of the bound GCN4 was inhibited, similar to **9**. The identical substitution in the carboxy-terminal position afforded ImPyPyPy-γ-PyPyPyPy-β-RPK (**13**), however, which inhibited GCN4 binding identically to the Arg–Pro–Arg polyamide **3**. The amino acid linkage between the final Py amino acid and the initial arginine residue was also essential for GCN4 inhibition. A polyamide in which the β-alanine linker was replaced with a 7-aminoheptanoic acid linker, ImPyPyPy-γ-PyPyPyPy-C7-RPR (**14**), was unable to inhibit GCN4 binding. Protein inhibition did not require prebinding of the polyamide. Preincubation of ARE-1 with GCN4 followed by addition of **5** afforded inhibition identical to that of prebound polyamide.

Salt dependence of GCN4 inhibition

In order to evaluate the sensitivity of major-groove-protein inhibition mediated by positive patch polyamides to the nature of the compensating electrolyte, as well as the overall ionic strength, gel mobility shift analysis was performed using a buffer that might model the environment of a cellular nucleus (20 mM MOPS, pH 7.2, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine) [45]. Arg–Pro–Arg polyamide **3** was found to inhibit GCN4 (222–281) binding under the ionic conditions in which KCl is the primary compensating electrolyte (Figure 10) and the conditions optimized for protein binding in which NaCl is the predominant compensating electrolyte (Figure 6c). Further biophysical characterization of major-groove-protein inhibition by positive patch polyamides will be reported in due course.

Figure 8



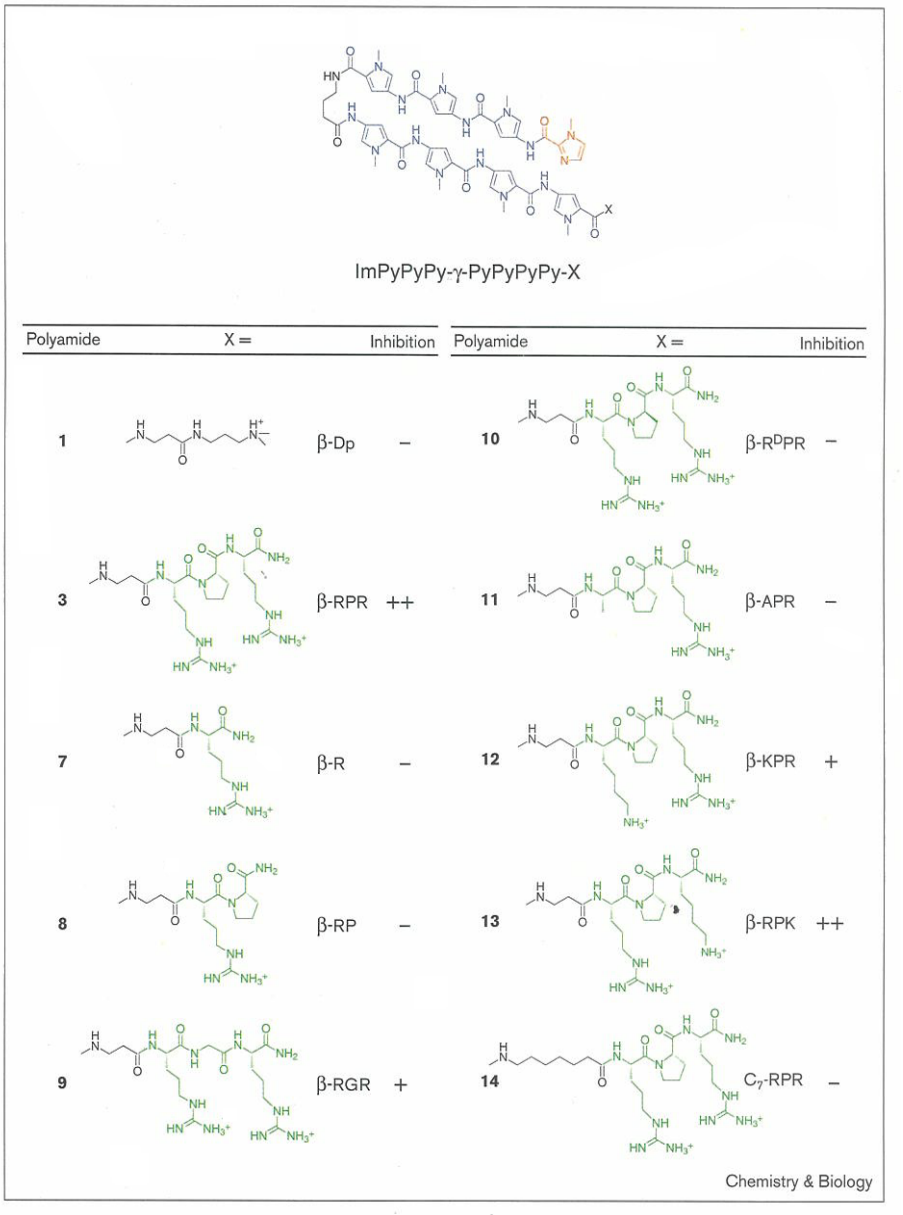
GCN4 (222–281) gel mobility shift experiments in the presence of Arg–Pro–Arg–Arg–Arg–Arg polyamides. **(a–d)** Top, model of polyamide binding the match or mismatch site of DNA fragment ARE-1 or ARE-2 adjacent to the GCN4-binding site. The polyamide is colored as in Figure 1a. Mismatches are indicated by a yellow box. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222–281) binding to the DNA fragment in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 μg/ml poly(dI-dC)•poly(dI-dC), 22°C). The upper band is the GCN4 (222–281)–DNA complex and the lower band is free DNA. Lanes 1 and 13 are DNA only. Lanes 2 and 12 contain DNA incubated with 200 nM GCN4 (222–281). **(a,b)** Lanes 3–11 are 200 nM GCN4 (222–281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM and 2 μM **5**. **(c,d)** Lanes 3–11 are 200 nM GCN4 (222–281) and 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM and 200 nM **6**. (a) ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) binding the match site 5′-TGTTAT-3′ of ARE-1. (b) ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) incubated with the double-mismatch site 5′-TGGTCT-3′ of ARE-2 (mismatch sites italicized). (c) ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) incubated with the double-mismatch site 5′-TG*T*T*A*T-3′ of ARE-1. (d) ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) binding the match site 5′-TGGTCT-3′ of ARE-2. Lower K_a values ($K_a \cong 1 \times 10^7\ M^{-1}$) are observed for polyamides under the gel shift conditions due to the carrier DNA which artificially depresses polyamide binding constants. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.

Polyamide binding affinity and specificity are unaltered by Arg–Pro–Arg

In order to evaluate the effects of the Arg–Pro–Arg moiety on the DNA-binding properties of the polyamides, quantitative DNase I footprint titration experiments were performed to determine the equilibrium association constants of polyamides **1**–**6** for their respective six base pair match and single base pair mismatch sites (10 mM Tris pH 7.0, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 22°C) [46]. DNase I footprinting of ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), ImImPyPy-γ-ImPyPyPy-β-RPR (**4**) and ImImPyPy-γ-ImPyPyPy-β-RPRRRR (**6**) was performed on the 250 base pair *Eco*RI/*Pvu*II restriction fragment of pJK6 (Figure 11) [47]. **2** bound to the match sites, 5′-TGGTCA-3′ and 5′-TGGACA-3′, with identical affinities within experimental error

Figure 9

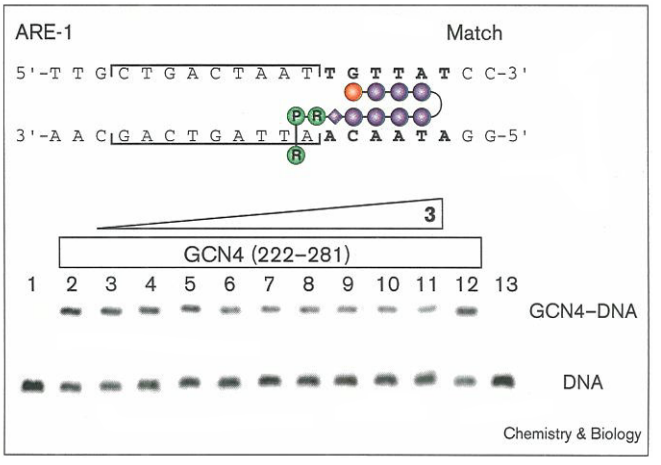
Aliphatic amino acid substitutions in the Arg–Pro–Arg domain. The structure of the ImPyPyPy-γ-PyPyPyPy-X polyamide is shown where X = β-Dp (β-Dp, **1**), β-Arg–Pro–Arg (β-RPR, **3**), β-Arg (β-R, **7**), β-Arg–Pro (β-RP, **8**), β-Arg–Gly–Arg (β-RGR, **9**), β-Arg–D–Pro–Arg (β-R^{D}PR, **10**), β-Ala–Pro–Arg (β-APR, **11**), β-Lys–Pro–Arg (β-KPR, **12**), β-Arg–Pro–Lys (β-RPK, **13**) or C_7-Arg–Pro–Arg (C_7-RPR, **14**). The ability of each polyamide to inhibit GCN4 (222-281) binding in a gel mobility shift assay performed identical to Figure 6a is reported. A lack of protein inhibition is indicated by a minus sign, reduced protein inhibition relative to **3** is shown by a plus sign, and two plus signs indicate protein inhibition identical to **3**. MALDI-TOF MS: **7**, 1292.6 (1292.6 calc'd); **8**, 1389.7 (1389.6 calc'd); **9**, 1505.6 (1505.7 calc'd); **10**, 1545.9 (1545.7 calc'd); **11**, 1461.2 (1460.7 calc'd); **12**, 1517.8 (1517.7 calc'd); **13**, 1517.6 (1517.7 calc'd); **14**, 1601.8 (1601.8 calc'd).



(K_a = 1.3 (± 0.1) × 10^{10} M^{-1} and 6.4 (± 1.2) × 10^9 M^{-1}, respectively). **2** also demonstrated greater than 100-fold specificity relative to binding at a single base pair mismatch site 5′-TG*T*ACA-3′ ($K_a \leq 5 \times 10^7$ M^{-1}, mismatched base pair in italics). Similar affinity and a slight increase in specificity were observed for Arg–Pro–Arg polyamide **4**. 5′-TGGTCA-3′ and 5′-TGGACA-3′ were bound by **4** (K_a = 4.6 (± 0.2) × 10^9 M^{-1} and 6.6 (± 1.0) × 10^9 M^{-1}, respectively) with greater than 450-fold specificity relative to binding at the mismatch site ($K_a \leq 1 \times 10^7$ M^{-1}). The additional three terminal arginine residues of **6** generated a tenfold increase in affinity relative to **2**, coupled with a significant loss in specificity for a single base pair mismatch. **6** bound 5′-TGGTCA-3′, 5′-TGGACA-3′ and 5′-TG*T*ACA-3′ with affinities of 2.6 (± 0.4) × 10^{10} M^{-1}, 2.8 (± 0.5) × 10^{10} M^{-1} and 1.9 (± 0.8) × 10^{10} M^{-1}, respectively.

Corresponding results were observed for DNase I footprinting of ImPyPyPy-γ-PyPyPyPy-β-Dp (**1**), ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**) and ImPyPyPy-γ-PyPyPyPy-β-RPRRRR (**5**) on the 229 base pair *Af*lII/*Fsp*I restriction fragment of pJT8 [1]. **1** has been shown to bind its six base pair match site, 5′-AGT*T*TAT-3′, with an affinity of 3.5 (± 0.8) × 10^9 M^{-1} and a sevenfold specificity relative to binding to a single base pair mismatch site 5′-AGTA*C*T-3′ (K_a = 5.0 (± 0.5) × 10^8 M^{-1}, Table 1) [1]. The Arg–Pro–Arg

Figure 10



GCN4 (222–281) gel mobility shift experiments in the presence of ImPyPyPy-γ-PyPyPyPy-β-RPR under putative physiological ionic conditions. Top, model of polyamide binding the match site of DNA fragment ARE-1 adjacent to the GCN4 binding site. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (222–281) binding to the radiolabeled ARE-1 fragment in the presence of increasing concentrations of **3** under ionic conditions modeling those found *in vivo* (20 mM MOPS pH 7.2, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine, 50 μg/ml poly(dI-dC)•poly(dI-dC), 22°C). The upper band is the GCN4 (222–281)–DNA complex and the lower band is free DNA. Lanes 1 and 13 are DNA only. Lanes 2 and 12 contain DNA incubated with 200 nM GCN4 (222–281). Lanes 3–11 are 200 nM GCN4 (222–281) and 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM and 2 μM **3**. Lower K_a values ($K_a \cong 1 \times 10^7\ M^{-1}$) are observed for polyamides under the gel shift conditions. The polyamide concentrations required for GCN4 inhibition are within the expected range based on the K_a under gel shift conditions.

polyamide **3** demonstrated only a slight loss in affinity and a similar specificity ($K_a = 5.5\ (\pm 1.5) \times 10^8\ M^{-1}$ for 5′-AGTATT-3′ and $9.2\ (\pm 0.4) \times 10^7\ M^{-1}$ for 5′-AGTA*C*T-3′). Analogous to **6**, the additional terminal arginines of **5** provided a tenfold increase in affinity for the match site ($K_a = 1.0\ (\pm 0.2) \times 10^{10}\ M^{-1}$) and a severe loss in specificity for a single-base-pair mismatch ($K_a = 3.4\ (\pm 0.5) \times 10^9\ M^{-1}$).

Discussion

By targeting an eight-ring Arg–Pro–Arg-polyamide adjacent to a GCN4 binding site, selective inhibition of DNA binding by a protein that exclusively contacts the major groove is achieved (Figures 6 and 8). The polyamide domain binds sequence specifically in the minor groove with double base pair mismatches preventing GCN4 inhibition.

The inability of truncated analogs **7** (R) or **8** (RP) to inhibit GCN4 binding indicates that the carboxy-terminal arginine residue in **3** (RPR) is essential for GCN4 inhibition. On the basis of the Hin recombinase model, this arginine is expected to make nonspecific contacts with the DNA phosphate backbone. The ability of **13** (RPK) to inhibit GCN4 in an identical manner to **3** (RPR) supports this model. The neutralization of a portion of the backbone is the most likely mechanism by which Arg–Pro–Arg polyamides achieve GCN4 inhibition. Other models, such as steric blockage of the major groove or DNA distortion, cannot be ruled out [48,49]. Modeling suggests that Arg–Pro–Arg is insufficient to cross the DNA backbone and block the major groove. Determination of the exact mechanism of inhibition awaits high-resolution structure studies which are in progress.

Arg–Pro directs the terminal arginine to the phosphate backbone

The results of GCN4 inhibition experiments with polyamides **7–14** suggest that the Arg–Pro–Arg of **3** and **4** adopts a stable and well-defined structure similar to that of Arg140–Pro141–Arg142 of Hin recombinase (Figure 12). The internal Arg–Pro of **3** and **4** (**RP**R) is required for GCN4 inhibition. Polyamide **8** (RP) does not inhibit GCN4, suggesting that these two residues play a structural role in the placement of the terminal arginine near the phosphate backbone. Replacing the rigid proline of **3** (RPR) with a flexible glycine (**9**; RGR) allows a significant amount of GCN4 to remain bound in the presence of a saturating concentration of **9**. The glycine in **9** may permit the carboxy-terminal arginine to shift to a position that permits simultaneous binding with GCN4.

ImPyPyPy-γ-PyPyPyPy-β-R^{D}PR (**10**) is a diastereomer of ImPyPyPy-γ-PyPyPyPy-β-RPR (**3**), but it is unable to inhibit GCN4. Modeling indicates that substitution of D-proline for L-proline may in fact direct the neutralizing terminal arginine to the backbone of the opposite DNA strand. Confirmation of this prediction awaits studies with other protein systems.

Figure 11

Quantitative DNase I footprint titration experiments with **(a)** ImImPyPy-γ-ImPyPyPy-β-Dp (**2**), **(b)** ImImPyPy-γ-ImPyPyPy-β-RPR (**4**), and **(c)** ImImPyPy-γ-ImPyPyPy-β-RPRRR (**6**), on the 3′-^{32}P-end-labeled 250 base pair *Eco* RI/*Pvu* II restriction fragment from plasmid pJK6 [47]. Left, storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated from DNase I digestion. All reactions contained 20,000 cpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$ and were performed at 22°C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5–23, 1 pM, 2 pM, 5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 150 pM, 250 pM, 400 pM, 650 pM, 1 nM, 1.5 nM, 2.5 nM, 4 nM, 6.5 nM, 10 nM, respectively. Right, models of polyamide bound to the match or mismatch sites on the DNA fragment. I (5′-TGGTCT-3′) and II (5′-TGGACT-3′) are both match sites and III (5′-TG*T*ACA-3′) is a single base pair mismatch site (indicated by a yellow box). The polyamide is colored as in Figure 1a. The restriction fragment was 3′-^{32}P-end labeled on the bottom strand. Equilibrium association constants of each polyamide binding to the match or mismatch site is reported on the right. Each value is the average of at least three experiments.

Figure 11

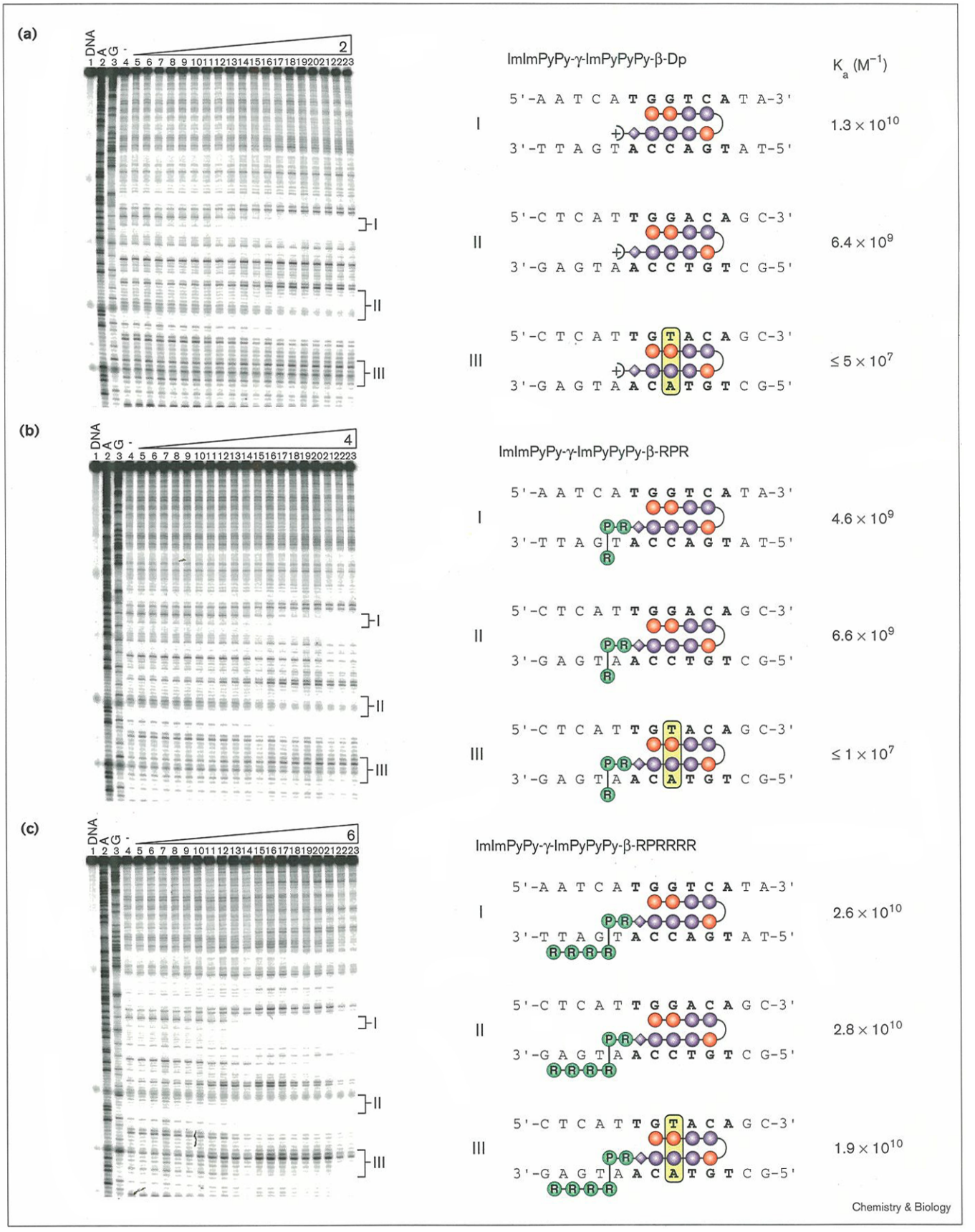

(a)
DNA
A
G
2
I
II
III
ImImPyPy-γ-ImPyPyPy-β-Dp
K_a (M^{-1})
5'-A A T C A T G G T C A T A-3'
3'-T T A G T A C C A G T A T-5'
1.3×10^{10}
5'-C T C A T T G G A C A G C-3'
3'-G A G T A A C C T G T C G-5'
6.4×10^{9}
5'-C T C A T T G T A C A G C-3'
3'-G A G T A A C A T G T C G-5'
$\leq 5 \times 10^{7}$
(b)
4
ImImPyPy-γ-ImPyPyPy-β-RPR
4.6×10^{9}
6.6×10^{9}
$\leq 1 \times 10^{7}$
(c)
6
ImImPyPy-γ-ImPyPyPy-β-RPRRRR
2.6×10^{10}
2.8×10^{10}
1.9×10^{10}
Chemistry & Biology

Table 1

Equilibrium association constants (M^{-1}).

	Binding site	
Polyamide	5′-AGTATT-3′	5′-AGTACT-3′
ImPyPyPy-γ-PyPyPyPy-β-Dp	3.5×10^{9}*	5.0×10^{8}*
ImPyPyPy-γ-PyPyPyPy-β-RPR	5.5×10^{8}	9.2×10^{7}
ImPyPyPy-γ-PyPyPyPy-β-RPRRRR	1.0×10^{10}	3.4×10^{9}

Values reported for the six-base-pair match (5′-AGTATT-3′) and mismatch (5′-AGTACT-3′) sites are the mean values obtained from three DNase I footprint titration experiments on the *AflII*/*FspI* restriction fragment of pJT8. The assays were carried out at 22°C at pH 7.0 in the presence of 10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$. *From [1].

Replacement of the internal arginine with alanine, as in **11**, reduces binding affinity by a factor of ten and prevents GCN4 inhibition. Furthermore, the Lys–Pro–Arg polyamide, **12**, has a comparable binding affinity to **3**, yet it is a less effective inhibitor of GCN4. Together, these results suggest that the guanidinium of the internal arginine residue makes specific contacts with the DNA that are required for the proper positioning of the remaining residues. Replacement of the β-alanine linker (**3**) with 7-aminoheptanoic acid (**14**) eliminates inhibition, further implicating the placement of the Arg–Pro–Arg moiety as a requirement for effective inhibition.

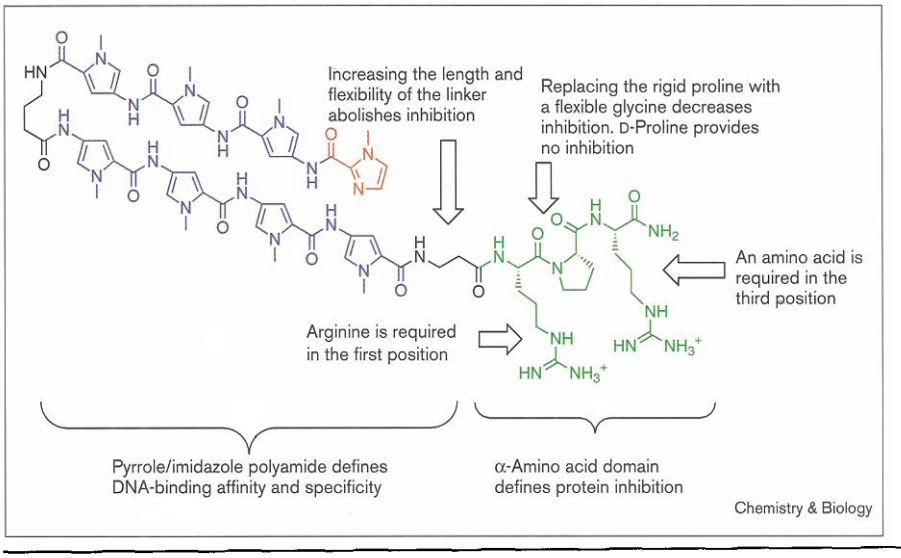
Arg–Pro–Arg polyamides retain high DNA-binding affinity and specificity

Quantitative DNase I footprinting demonstrates that the addition of a carboxy-terminal Arg–Pro–Arg tripeptide, as in **3** and **4**, does not alter the DNA-binding properties of eight-ring hairpin polyamides (**1** and **2**). Arg–Pro–Arg–Arg–Arg–Arg polyamides, **5** and **6**, however, have a increased binding affinity but they have no specificity relative to a single base pair mismatch site. DNase I footprinting and gel mobility shift analysis show that **5** and **6** have similar affinities for match and double base pair mismatch sites. These results indicate that synthetic ligands may balance the benefits of additional charge with the consequence of lowered sequence specificity [50]. For example, a distamycin analog modified with a decaaza decabutylamine moiety on a pyrrole nitrogen interferes with the binding of a major-groove transcription factor [26]. Unfortunately, the sequence specificity of this molecule, which contains potentially 11 positive charges, has not been reported. The results described here suggest that such a molecule may bind DNA with reduced sequence specificity.

Significance

The functional repertoire of polyamides as synthetic ligands for the control of transcription-factor binding has been expanded to include proteins that bind exclusively in the major groove of DNA. The addition of an Arg–Pro–Arg tripeptide results in no significant alteration of polyamide–DNA binding affinity or specificity. By targeting sequences outside the protein-binding site, selective inhibition of the major-groove transcription factor GCN4 at two sites was demonstrated using Arg–Pro–Arg polyamides. The Arg–Pro–Arg domain appears to adopt a stable and defined structure that delivers a neutralizing positive charge to the DNA backbone where it competes with GCN4 for contact with the phosphates. The broad targetable sequence repertoire of polyamides, coupled with the ubiquity of backbone contacts in protein recognition of DNA, makes phosphate neutralization by a positive patch a promising approach for inhibition of major-groove transcription factors.

Figure 12



Summary of the structural requirements for optimal GCN4 inhibition by Arg–Pro–Arg polyamides.

Materials and methods

All buffers for gel mobility shift and footprinting experiments were prepared from J.T. Baker reagents and 0.2 μm filtered. EDTA and dithiothreitol (DTT) were obtained from Gibco BRL. Poly(dI-dC)•poly(dI-dC) was from Pharmacia Biotech. Ficoll (MW 400,000) was purchased from Sigma. T4 polynucleotide kinase, *Eco*RI, *Pvu*II, and DNase I were from Boehringer Mannheim. *Afl*II and *Fsp*I were purchased from New England Biolabs. Sequenase (version 2.0) was obtained from United States Biochemical. [α-^{32}P]-Thymidine-5′-triphosphate (≥ 3000 Ci/mmol) and [α-^{32}P]-deoxyadenosine-5′-triphosphate (≥ 6000 Ci/mmol), and were purchased from Du Pont/NEN. [γ-^{32}P]-Adenosine-5′-triphosphate (≥ 7000 Ci/mmol) was obtained from ICN. Standard methods were used for all DNA manipulations [51]. GCN4 (222–281) prepared by solid phase synthesis was generously provided by Martha G. Oakley [29,35]. MBHA resin (0.57 mmol/g) was from Calbiochem. Boc–Ala, Boc-γ-aminobutyric acid, Boc-(Tos)Arg, Boc-Ala, Boc(CBz)Lys, Boc-Gly, Boc-Pro and Boc-D-Pro were from Peptides International. *p*-Cresol was purchased from Aldrich. All other chemicals, as well as the purification and characterization of polyamides were as previously described [44]. Arg–Pro–Arg polyamide synthesis is exemplified here for polyamide **3**.

ImPyPyPy-γ-PyPyPyPy-β-RPR (3)

ImPyPyPy-γ-PyPyPyPy-β-RPR–MBHA–resin was synthesized in a stepwise fashion by machine-assisted solid phase methods [44] from MBHA resin (600 mg, 0.57 mmol/g, calculated as $L_{new}(mmol/g) = L_{old}/(1 + L_{old}(W_{new}-W_{old}) \times 10^{-3})$, where L is the loading (mmol of amine per gram of resin) and W is the weight (g/mol) of the growing peptide attached to the resin [52]). A sample of polyamide resin (300 mg, 0.30 mmol/g) was placed in a Kel-F reaction vessel, *p*-cresol (1 g) added, and the vessel was cooled to −60°C. HF was then condensed into the vessel. The solution was stirred vigorously for one hour (0°C) and the excess HF was removed *in vacuo*. The reaction mixture was then treated with cold ethyl ether (50 ml) and the resulting resin/polyamide coprecipitate was collected by vacuum filtration. The polyamide was extracted with CH_3CN:H_2O:TFA (10:9:1), 0.1% (w/v) TFA added (6 ml) and the resulting solution purified by reverse phase HPLC using a Waters DeltaPak 25 × 100 mm 100 μm C_{18} column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH_3CN. ImPyPyPy-γ-PyPyPyPy-β-RPR-NH_2 was recovered upon lyophilization of the appropriate fractions as a white powder (84 mg, 60% recovery); UV (H_2O) λ_{max} 244, 312 (66,000); ^{1}H NMR (DMSO-d_6): δ 10.49 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.90 (s, 3H), 9.84 (s, 1H), 8.19 (d, 1H, J = 7.6 Hz), 8.07 (m, 1H), 8.02 (m, 1H), 7.95 (m, 1H), 7.91 (m, 1H), 7.53 (m, 3H), 7.38 (s, 1H), 7.26 (d, 1H, J = 1.6 Hz), 7.25 (s, 1H), 7.21 (d, 1H, J = 1.6 Hz), 7.20 (s, 1H), 7.19 (s, 1H), 7.14 (m, 3H), 7.03 (m, 4H), 6.87 (d, 1H, J = 1.6 Hz), 6.85 (m, 2H), 4.48 (t, 1H, J = 4.8 Hz), 4.27 (q, 1H, J = 4.4 Hz), 4.77 (m, 1H), 3.96 (s, 3H), 3.81 (m, 12H), 3.80 (s, 3H), 3.77 (s, 3H), 3.76 (s, 3H), 3.60 (m, 2H), 3.32 (q, 2H, J = 4.8 Hz), 3.17 (q, 2H, J = 6.1 Hz), 3.0 (m, 4H), 2.36 (t, 2H, J = 6.9 Hz), 2.25 (t, 2H, J = 6.9 Hz), 2.0 (m, 2H), 1.77 (m, 4H), 1.64 (m, 2H), 1.46 (m, 6H).

Preparation of ^{32}P-labeled synthetic duplex DNA for gel mobility shift assay

Synthetic DNA fragments were prepared on an ABI 380B Automated DNA Synthesizer and purified by preparative denaturing polyacrylamide gel electrophoresis. ARE-1 and ARE-2, 5′-CCGGATCCATGGTTGCTGACTAATTGTTATCCTCTAGAGTCGACC-3′ and 5′-CCGGATCCATGGTTGCTGACTAATTGGTCTCCTCTAGAGTCGACC-3′, respectively, were radiolabeled at the 5′-terminus with γ-^{32}P-ATP and T4 polynucleotide kinase, annealed to an equimolar amount of the unlabeled complement, and purified by nondenaturing polyacrylamide gel electrophoresis [51].

Gel mobility shift assay

For gel mobility shifts, polyamide was incubated with radiolabeled synthetic DNA duplex (10,000 cpm) in 40 μl reaction volumes of bisTris (10 mM, pH 7.0), NaCl (100 mM), DTT (1 mM), EDTA (1 mM) and poly(dI-dC)•poly(dI-dC) (50 μg/ml) for 16 h at 22°C (20 mM MOPS, pH 7.0, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 1 mM spermine was used to model ionic conditions *in vivo*). GCN4 (222–281) was added and equilibrated for 30 min. Loading buffer (15% Ficoll, 0.025% bromophenol blue) (10 μl) was added and 10 μl was immediately loaded onto a running 8% (29:1, acrylamide:bis-acrylamide) polyacrylamide gel (0.5 × TBE, 280 V, 0.8 mm, 13 cm). Sufficient separation of the free DNA and the DNA–GCN4 (222–281) complexes was achieved within 45 min. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics) [53].

Preparation of ^{32}P-labeled restriction fragments for DNaseI footprinting

The *Afl* II/*Fsp* I restriction fragment of pJT8 [1] was 3′-^{32}P-end-labeled by digesting the plasmid with *Afl*II and *Fsp*I and simultaneously filling in using Sequenase, [α-^{32}P]-deoxyadenosine-5′-triphosphate, and [α-^{32}P]-thymidine-5′-triphosphate, and isolating the 229 base pair fragment by nondenaturing gel electrophoresis. The 250 base pair *Eco*RI/*Pvu*II restriction fragment of pJK6 [40] was prepared in an analogous manner. A and G sequencing were carried out as described [54,55].

Quantitative DNase I footprint titration experiments

All reactions were executed in a total volume of 400 μl. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (20,000 cpm), affording final solution conditions of 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and either (i) 0.001–100 nM polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22°C for 18 h. Footprinting reactions were initiated by the addition of 4 μl of a DNase I stock solution (at the appropriate concentration to give ~55% intact DNA) containing 1 mM DTT and allowed to proceed for 7 min at 22°C. The reactions were stopped by the addition of 50 μl of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/ml glycogen, and 28 μM base-pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 × TBE/80% formamide loading buffer, denatured by heating at 85°C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1 × TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics) [53].

Quantitation and data analysis

Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the equation:

$$\theta_{app} = 1 - \frac{I_{site} / I_{ref}}{I_{site}^{\circ} / I_{ref}^{\circ}} \quad (1)$$

where I°_{site} and I°_{ref} are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (equation 2, n = 1) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{Ka^n[L]^n_{tot}}{1 + Ka^n[L]^n_{tot}} \quad (2)$$

where $[L]_{tot}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a, θ_{max} and θ_{min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. At least three

sets of data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points.

Acknowledgements

We are grateful to the National Institutes of Health General Medical Sciences for research support, to the National Science Foundation for a predoctoral fellowship to R.E.B., and to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

References

1. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1996). Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **382**, 559-561.
2. Swalley, S.E., Baird, E.E. & Dervan, P.B. (1997). Discrimination of 5′-GGGG-3′, 5′-GCGC-3′, and 5′-GGCC-3′ sequences in the minor groove of DNA by eight-ring hairpin polyamides. *J. Am. Chem. Soc.* **119**, 6953-6961.
3. Turner, J. M., Baird, E.E. & Dervan, P.B. (1997). Recognition of 7 base-pair sequences in the minor groove of DNA by 10-ring pyrrole-imidazole polyamides. *J. Am. Chem. Soc.* **119**, 7636-7644.
4. Wade, W.S., Mrksich, M. & Dervan, P.B. (1992). Design of peptides that bind in the minor groove of DNA at 5′-(A,T)$_5$G(A,T)C(A,T) sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783-8794.
5. Mrksich, M., Wade, W.S., Dwyer, T.J., Geirstanger, B.H., Wemmer, D.E. & Dervan, P.B (1992). Antiparallel side-by-side motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl Acad. Sci. USA* **89**, 7586-7590.
6. Wade, W.S., Mrksich, M. & Dervan, P.B. (1993). Binding affinities of synthetic peptides, pyridine-2-carboxamidenetropsin and 1-methylimidazole-2-carboxamidenetropsin, that form 2/1 complexes in the minor groove of double-helical DNA. *Biochemistry* **32**, 11385-11389.
7. Mrksich, M. & Dervan, P.B. (1993). Antiparallel side-by-side heterodimer for sequence-specific recognition in the minor groove of DNA by a distamycin 1-methylimidazole-2-carboxamide pair. *J. Am. Chem. Soc.* **115**, 2572-2576.
8. White, S., Baird, E.E. & Dervan, P.B. (1997). On the pairing rules for recognition in the minor-groove of DNA by pyrrole–imidazole polyamides. *Chem. Biol.* **4**, 569-578.
9. White, S., Baird, E.E. & Dervan, P.B. (1997). Orientation preferences of pyrrole-imidazole polyamides in the minor-groove of DNA. *J. Am. Chem. Soc.* **119**, 8756-8765.
10. Pelton, J.G. & Wemmer, D.E. (1989). Structural characterization of a 2-1 distamycin A•d(CGCAAATTTGGC)$_2$ complex by two-dimensional NMR. *Proc. Natl Acad. Sci. USA* **86**, 5723-5727.
11. Chen, X., Ramakrishnan, B., Rao, S.T. & Sundaralingham, M. (1994). Binding of 2 distamycin-A molecules in the minor groove of an alternating B-DNA duplex. *Nat. Struct. Biol.* **1**, 169-175.
12. White, S., Baird, E.E. & Dervan, P. (1996). Effects of the A•T/T•A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry* **35**, 12532-12537.
13. Pilch, D.S., *et al.*, & Dervan, P.B. (1996). Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. *Proc. Natl Acad. Sci. USA* **93**, 8306-8311.
14. de Clairac, R.P.L., Geierstanger, B.H., Mrksich, M., Dervan, P.B. & Wemmer, D.E. (1997). NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.* **119**, 7909-7916.
15. Gottesfeld, J.M., Nealy, L., Trauger, J.W., Baird, E.E. & Dervan, P.B. (1997). Regulation of gene expression by small molecules. *Nature* **387**, 202-205.
16. Nealy, L., Trauger, J.W., Baird, E.E., Dervan, P.B. & Gottesfeld, J.M. (1997). Importance of minor groove binding zinc fingers within the transcription factor IIIA•DNA complex. *J. Mol. Biol.* **274**, 439-445.
17. Moser, H.E. & Dervan, P.B. (1987). Sequence-specific cleavage of double-helical DNA by triple helix formation. *Science* **238**, 645-650.
18. Thuong, N.T. & Helene, C. (1993). Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **32**, 666-690.
19. Maher, J.L., Dervan, P.B. & Wold, B. (1992). Analysis of promoter-specific repression by triple-helical DNA complexes in a eukaryotic cell-free transcription system. *Biochemistry* **31**, 70-81.
20. Duval-Valentin, G., Thuong, N.T. & Helene, C. (1992). Specific-inhibition of transcription by triple helix-forming oligonucleotides. *Proc. Natl Acad. Sci. USA* **89**, 504-508.
21. Nielsen, P.E. Design of sequence-specific DNA-binding ligands (1997). *Chem. Eur. J.* **3**, 505-508.
22. Ho, S.N., Boyer, S.H., Schreiber, S.L., Danishefsky, S.J., & Crabtree, G.R. (1994). Specific-inhibition of formation of transcription complexes by a calicheamicin oligosaccharide - a paradigm for the development of transcriptional antagonists. *Proc. Natl Acad. Sci. USA* **91**, 9203-9207.
23. Liu, C. *et al.*, & Vogt, P.K. (1996). Sequence-selective carbohydrate DNA interaction - dimeric and monomeric forms of the calicheamicin oligosaccharide interfere with transcription factor function. *Proc. Natl Acad. Sci. USA* **93**, 940-944.
24. Bruice, T.C., Mei. H.-Y., He, G.-X. & Lopez, V. (1992). Rational design of tripyrrole peptides that complex with DNA by both selective minor-groove binding and electrostatic interaction with the phosphate backbone. *Proc. Natl Acad. Sci. USA* **89**, 1700-1704.
25. Chiang, S.-Y., Bruice, T.C., Azizkhan, J.C., Gawron, L. & Beerman, T.A. (1997). Targeting E2F1-DNA complexes with microgonotropen DNA binding agents. *Proc. Natl Acad. Sci. USA* **94**, 2811-2816.
26. Bruice, T.C., Sengupta, D., Blasko, A., Chiang, S.-Y. & Beerman, T.A. (1997). A microgonotropen branched decaaza decabutylamine and its DNA and DNA/transcription factor interactions. *Bioorg. Med. Chem.* **5**, 685-692.
27. Steitz, T. A. (1990). Structural studies of protein-nucleic acid interaction-the sources of sequence-specific binding. *Quart. Rev. Biophys.* **23**, 205-280.
28. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. (1993). Crystal-structure of a yeast TBP TATA-box complex. *Nature* **365**, 512-520.
29. Oakley, M.G. & Dervan, P.B. (1990). Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. *Science* **248**, 847-850.
30. Ellenberger, T., Brandl, C.J., Struhl, K. & Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted α-helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223-1237.
31. König, P. & Richmond, T. (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**, 139-154.
32. Sluka, J.P., Horvath, S.J., Glasgow, A.C., Simon, M.I., & Dervan, P.B. (1990). Importance of minor-groove contacts for recognition of DNA by the binding domain of Hin recombinase. *Biochemistry* **29**, 6551-6561.
33. Feng, J-A., Johnson, R.C. & Dickerson, R.E. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* **263**, 348-355.
34. Kielkopf, C.L., Baird, E.E., Dervan, P.B. & Rees, D.C. (1998). Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.*, **5**, 104-109.
35. Oakley, M.G., Mrksich, M. & Dervan, P.B. (1992). Evidence that a minor groove-binding peptide and a major-groove binding protein can simultaneously occupy a common site on DNA. *Biochemistry* **31**, 10969-10975.
36. Park, Y.W. & Breslauer, K.J. (1992). Drug-binding to higher ordered DNA structures - netropsin complexation with a nucleic-acid triple helix. *Proc. Natl Acad. Sci. USA* **89**, 6653-6657.
37. Gehring, W.J., *et al.*, & Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell* **78**, 211-223.
38. Gehring, W. J., Affolter, M. & Burglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.
39. Hurst, H.C. (1995). Transcription factors. 1. bZIP proteins. *Protein Profile* **2**, 105-168.
40. Struhl, K. (1992). Yeast GCN4 transcriptional activator protein. In *Transcriptional Regulation.* (McKnight, S. L. & Yamamoto, K. R., eds), pp. 833-859, Cold Spring Harbor Laboratory Press, New York.
41. Curran, T. & Vogt, P. (1992). Dangerous liasions: fos and jun, oncogenic transcription factors. In *Transcriptional Regulation.* (McKnight, S.L. & Yamamoto, K.R., eds), pp. 797-832, Cold Spring Harbor Laboratory Press, New York.
42. Hope, I.A. & Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**, 885-894.
43. Paolella D.N., Palmer, C.R. & Schepartz, A. (1994). DNA targets for certain bZIP proteins distinguished by an intrinsic bend. *Science* **264**, 1130-1133.
44. Baird, E.E. & Dervan, P.B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141-6146.

45. Jones, R.J., Lin, K.-Y., Milligan, J.F., Wadawani, S. & Matteucci, M.D. (1993). Synthesis and binding properties of pyrimidine oligodeoxynucleoside analogs containing neutral phosphodiester replacements: the formacetal and 3'-thioformacetal internucleoside linkages. *J. Org. Chem.* **58**, 2983-2991.
46. Brenowitz, M., Senear, D.F., Shea, M.A. & Ackers, G.K. (1986). Quantitative DNase footprint titration - a method for studying protein-DNA interactions. *Methods Enzymol.* **130**, 132-181.
47. Kelly, J.J., Baird, E.E. & Dervan, P.B. (1993). Binding site size limit of the 2/1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl Acad. Sci., USA* **93**, 6981-6985.
48. Strauss, J.K. & Maher, L.J. (1994). DNA bending by asymmetric phosphate neutralization. *Science* **266**, 1829-1834.
49. Welch, J.J., Rauscher III, F.J., & Beerman, T.A. (1994). Targeting DNA-binding drugs to sequence-specific transcription factor•DNA complexes. *J. Biol. Chem.* **269**, 31051-31058.
50. Breslauer, K.J., Ferrante, R., Marky, L.A., Dervan, P.B. & Youngquist, R. S. (1988). The origins of the DNA binding affinity and specificity of minor groove directed ligands: correlations of thermodynamic and structural data. In *Structure and Expression (Vol. 2), DNA and Its Drug Complexes* (Sarma, R.H. & Sarma, M.H. eds), pp. 273-289, Academic Press.
51. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning.* (2nd edn). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
52. Barlos, K., Chatzi, O., Gatos, D., Stravropoulos, G. (1991). 2-Chlorotrityl chloride resin - studies on anchoring of Fmoc - amino acids and peptide cleavage. *Int. J. Peptide Protein Res.* **37**, 513-520.
53. Johnston, R.F., Picket, S.C. & Barker, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis* **11**, 355.
54. Maxam, A.M. & Gilbert, W.S. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499-560.
55. Iverson, B.L. & Dervan, P.B. (1987). Adenine-specific DNA chemical sequencing reaction. *Methods Enzymol.* **15**, 7823-7830.
56. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1996). Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA. *Chem. Biol.* **3**, 369-377.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.