Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron 63 (2007) 6146–6151
Tetrahedron

Completion of a programmable DNA-binding small molecule library

Carey F. Hsu, John W. Phillips, John W. Trauger, Michelle E. Farkas, Jason M. Belitsky, Alexander Heckel, Bogdan Z. Olenyuk, James W. Puckett, Clay C. C. Wang and Peter B. Dervan*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 27 January 2007; revised 2 March 2007; accepted 6 March 2007
Available online 12 March 2007

Abstract—Hairpin pyrrole-imidazole (Py-Im) polyamides are programmable oligomers that bind the DNA minor groove in a sequence-specific manner with affinities comparable to those of natural DNA-binding proteins. These cell-permeable small molecules have been shown to enter the nuclei of live cells and downregulate endogenous gene expression. We complete here a library of 27 hairpin Py-Im polyamides, which bind seven-base-pair sequences of the general form 5′-WWGNNNW-3′ (where W=A or T, N=W, G, or C). Their equilibrium association constants (K_a) range from $K_a=1\times10^8$ to 4×10^{10} M^{-1} with good sequence specificity. A table of binding affinities and sequence contexts for this completed 27-member library has been assembled for the benefit of the chemical biology community interested in molecular control of transcription.

1. Introduction

The biological applications of sequence-specific DNA-binding small molecules are a subject of intense research but still far from being routine.[1–11] Py-Im polyamides have been shown to influence a number of protein–DNA interactions, demonstrating both repression and activation of gene expression. Confocal microscopic experiments have confirmed that polyamide–fluorophore conjugates traffic unaided to the nuclei of living cells.[12–14] Since many diseases are attributed to aberrant gene expression, the regulation of transcriptional pathways with small molecules could have an important effect on human medicine.[15–17] For researchers' interested in selectively targeting protein–DNA interfaces in promoters of specific genes with small molecules, access to well-characterized libraries of polyamides, which bind a repertoire of different sequences with high affinity and specificity could enable development in this area.

Polyamides constructed from *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methylhydroxypyrrole (Hp) amino acids comprise a class of synthetic ligands that bind within the minor groove of DNA in a sequence-specific manner.[18,19] Inspired by the natural products netropsin and distamycin A, these programmed molecules recognize the Watson–Crick base pairs (bp) according to a series of pairing rules, where aromatic heterocycles paired in an antiparallel fashion are able to discriminate one Watson—Crick base pair from the other three combinations. The Py/Py pair recognizes A,T over C,G.[20] The Im/Py pair distinguishes G·C from C·G.[21,22] The exocyclic amine of guanine presents steric hindrance to the C3–H of Py, while the N3 of Im accommodates the amine and accepts one of its hydrogen bonds. The Hp/Py pair discriminates T·A over A·T due to the steric fit of the hydroxy group protruding into the minor groove, thus completing the pairing rules.[23,24] NMR and X-ray crystallographic studies reveal that the crescent-shaped polyamide side-by-side dimer binds B-form DNA, a remarkable example of shape-selective recognition of the deep minor groove of DNA.[20,22,25,26]

Within the framework of the pairing rules, covalent linkages between two antiparallel polyamide strands result in several possible structures, including the hairpin, cycle, H-pin, and U-pin binding motifs.[27–30] These linked structures show improved affinity and specificity when compared with the unlinked dimers. The eight-ring hairpin polyamide provides a good compromise between synthetic ease (linear vs branched oligomers) and molecular recognition properties. In this binding motif, a γ-aminobutyric acid residue connects the carboxylic terminus of one strand to the amino terminus of the other.[27] The turn residue also serves as a DNA recognition element, as it has been shown to bind A·T and T·A base pairs with greater than 25-fold specificity over C·G and G·C, presumably for steric reasons.[31] Use of

Keywords: Molecular recognition; Gene regulation; Polyamide; Small molecule–nucleic acid interaction.
* Corresponding author. Tel.: +1 626 395 6002; fax: +1 626 683 8753; e-mail: dervan@caltech.edu

0040–4020/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tet.2007.03.041

a chiral diaminobutyric acid turn residue increases the overall binding affinity of the molecule by 10-fold without a loss of sequence specificity.[32]

When hairpin polyamides are synthesized using solid-phase methods on Boc-β-Ala-PAM resin and cleaved with 3-dimethylamino-1-propylamine, the product contains a β-alanine residue and a dimethylaminopropylamide tail at the C-terminus. Both of these elements are specific for W (where W=A or T) over G·C and C·G, again for steric reasons. In the hairpin motif, the β-alanine residue exhibits greater than 210-fold specificity for A·T and T·A base pairs over G·C and C·G.[31] The dimethylaminopropylamide tail shows a 20-fold preference for A·T and T·A over G·C and C·G.[31] Therefore, an eight-ring hairpin polyamide can bind seven base pairs with specificity for W over the turn, β-alanine residue, and tail (Fig. 1).

We have excluded Hp in this study in favor of more stable Py residue. The electron-rich Hp ring degrades in the presence of strong acid and free radical impurities. Based on the paradigm of unsymmetrical ring pairings, more robust rings such as the hydroxybenzimidazole/Py pair can replace Hp for T·A recognition.[33] In our work, hairpin polyamide design typically includes an Im/Py pairing at the hairpin terminus in order to impart G·C specificity at this position. As a result, eight-ring pyrrole-imidazole hairpin polyamides can specifically bind seven-bp sequences of the general form 5′-WWGNNNW-3′ (where N=W, G, or C). There are 27 possible permutations that fall within these guidelines. Our group has published the energetics of 11 hairpin polyamides binding 11 distinct cognate sequences. Over half of the 27 sequences remain unreported. Searching our theses and notebooks confirmed seven additional previously unpublished characterizations. Our studies during the past decade left only 9 out of 27 sequences unexamined; these are: 5′-WWGCGWW-3′, 5′-WWGCCWW-3′, 5′-WWGCWCW-3′, 5′-WWGCCCW-3′, 5′-WWGWCGW-3′, 5′-WWGWGCW-3′, 5′-WWGCGGW-3′, 5′-WWGCCGW-3′, and 5′-WWGWGGW-3′. To complete the library, we have synthesized nine hairpin polyamides designed to target these remaining sequences and assayed their binding affinity and sequence specificity by DNase I footprint titration experiments. In this study, we complete the table of polyamide–DNA-binding affinities, indicating each of the 27 general DNA sequences and an eight-ring hairpin polyamide that sequence specifically binds that seven-bp site. We hope that this centralized source of previously unpublished data proves helpful for other research groups currently modulating protein–DNA interfaces with DNA-binding small molecules.

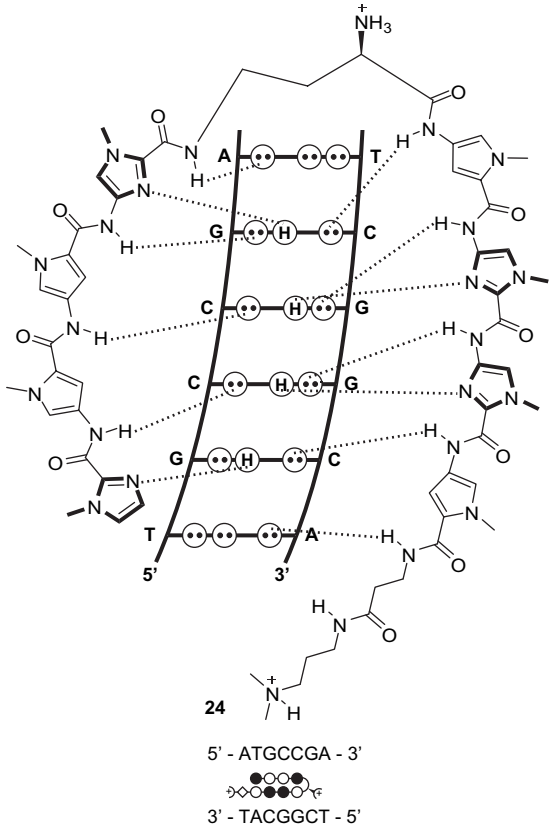

Figure 1. (Top) Model for the complex formed between hairpin polyamide **24** and its match DNA sequence. Circles with two dots represent the lone pairs of N(3) of purines and O(2) of pyrimidines. Circles containing an H represent the N(2) hydrogen of G. Hydrogen bonds are illustrated by dotted lines. (Bottom) Ball-and-stick binding model for the hairpin motif with the polyamide bound to its target DNA sequence. Imidazole and pyrrole are shown as filled and non-filled circles, respectively; β-alanine is shown as a diamond; the dimethylaminopropylamide tail is shown as a half-circle with a plus; and the chiral diaminobutyric acid turn residue is shown as a semicircle linked to a half-circle with a plus connecting the two subunits.

2. Results

2.1. Polyamide synthesis

Nine polyamides **12**–**14**, **16**, **17**, **19**, **21**, **24**, and **27** were synthesized on Boc-β-Ala-PAM resin according to published manual solid-phase synthetic protocols.[34] After cleavage with 3-dimethylamino-1-propylamine and reverse-phase HPLC purification, polyamides were characterized by analytical HPLC, UV–visible spectroscopy, and matrix-assisted laser desorption ionization/time-of-flight mass spectrometry.

2.2. DNA-binding energetics

Quantitative DNase I footprint titration experiments (10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22 °C) were performed on the 295 bp, 5′-^{32}P-end-labeled PCR product of plasmids pCFH2, pCFH3, pCFH4, pCFH5, pPh2, and pMFST.[35] Each of the nine polyamides was assayed on a plasmid containing its seven-bp match site according to the pairing rules, as well as three formal mismatch binding sites (Table 1 and Supplementary data Fig. S1). The energetics of polyamide binding in the minor groove of DNA can be calculated from the Hill equation isotherms following DNase I cleavage and gel separation of the fragments. The equilibrium association constants (K_a) determined in this way provide a quantitative measure of polyamide affinity at a given DNA-binding site. Comparing these constants across the four potential binding sites allows a relative measure of specificity for each base pair at the targeted position.

The nine polyamides **12**, **13**, **14**, **16**, **17**, **19**, **21**, **24**, and **27** bind their respective seven-bp match sites with equilibrium