

Conformational Diversity of Single-Stranded DNA from Bacterial Repetitive Extragenic Palindromes: Implications for the DNA Recognition Elements of Transposases

Tatsiana Charnavets,¹ Jaroslav Nunvar,¹ Iva Nečasová,¹ Jens Völker,² Kenneth J. Breslauer,^{2,3} Bohdan Schneider¹

¹ Institute of Biotechnology, Academy of Sciences of the Czech Republic, Videnska 1083, 142 20 Prague, Czech Republic

² Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Rd., Piscataway, NJ 08854

³ Cancer Institute of New Jersey, Rutgers University, New Brunswick, NJ 08903

Received 20 April 2015; accepted 5 May 2015

Published online 8 May 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22666

ABSTRACT:

Repetitive extragenic palindrome (REP)—associated tyrosine transposase enzymes (RAYTs) bind REP DNA domains and catalyze their cleavage. Genomic sequence analyses identify potential noncoding REP sequences associated with RAYT-encoding genes. To probe the conformational space of potential RAYT DNA binding domains, we report here spectroscopic and calorimetric measurements that detect and partially characterize the solution conformational heterogeneity of REP oligonucleotides from six bacterial species. Our data reveal most of these REP oligonucleotides adopt multiple conformations, suggesting that RAYTs confront a landscape of potential DNA substrates in dynamic equilibrium that could be selected, enriched, and/or induced via differential binding.

Additional Supporting Information may be found in the online version of this article.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Correspondence to: Bohdan Schneider; e-mail: bohdan.schneider@gmail.com

Contract grant sponsor: Czech Science Foundation, project P305/12/1801

Contract grant sponsor: BIOCEV CZ.1.05/1.1.00/02.0109 from the ERDF, RVO 86 652 036, and Biotechnological expert CZ.1.07/2.3.00/30.0020

© 2015 The Authors. Biopolymers Published by Wiley Periodicals, Inc.

Thus, the transposase-bound DNA motif may not be the predominant conformation of the isolated REP domain. Intriguingly, for several REPs, the circular dichroism spectra suggest guanine tetraplexes as potential alternative or additional RAYT recognition elements, an observation consistent with these REP domains being highly nonrandom, with tetraplex-favoring 5'-G and 3'-C-rich segments. In fact, the conformational heterogeneity of REP domains detected and reported here, including the formation of non-canonical DNA secondary structures, may reflect a general feature required for recognition by RAYT transposases. Based on our biophysical data, we propose guanine tetraplexes as an additional DNA recognition element for binding by RAYT transposase enzymes. © 2015 The Authors. Biopolymers Published by Wiley Periodicals, Inc. Biopolymers 103: 585–596, 2015.

Keywords: bacterial repetitive extragenic palindromes (REP); REP associated tyrosine transposases (RAYTs); circular dichroism spectroscopy; landscape of RAYT DNA recognition elements; interstrand guanine tetraplex

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the

preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at biopolymers@wiley.com.

INTRODUCTION

Repetitive extragenic palindrome (REP) elements represent well-characterized noncoding DNA repeats in bacteria. First discovered in *Escherichia coli*,¹ these imperfect palindromic elements are numerous in the genome.² REPs occur predominantly in intergenic regions, where they are mostly found in clusters called REPINs (REP doublet forming hairpins),^{3,4} and BIMEs (bacterial interspersed mosaic elements).⁵ REPs and their clusters serve as binding sites for several proteins in *E. coli*.⁶ REP-dependent regulation of transcription and mRNA turnover has been recorded in *E. coli* and several other gamma proteobacteria.^{7–9} REP elements were hypothesized to be mobile, which was confirmed in *Pseudomonas fluorescens*, where excisions of REPINs were detected.⁴ A class of transposase-related nucleases, termed REP-associated tyrosine transposases (RAYTs), recently were discovered and proposed to function as primary mobilizers of REP elements.¹⁰ RAYTs belong to the HUH superfamily of nucleases whose members promote replication or mobilization of various mobile genetic elements (viruses, plasmids and insertion sequences).¹¹ HUH nucleases are considered promising molecular tools for application in genetic engineering.¹² The ability of RAYTs to catalyze cleavage and recombination of REP elements has been confirmed experimentally for *E. coli*,¹³ and a crystal structure of RAYT in complex with a single-stranded REP DNA substrate has been reported.¹⁴ Due to the mobile nature of highly abundant REP elements and their function in transcription regulation, these repeat elements are believed to represent a dynamic regulatory network.¹¹

Sequence Characteristics of REPs

Table I summarizes the general sequence features of REP elements. Note that at the nucleotide sequence level, REPs are

defined by the presence of an invariant 5'-terminally located tetranucleotide GTAG or GTGG, followed by an imperfect GC-rich palindrome. Whereas the length and actual sequences of REP palindromes are highly variable, only the palindromicity and therefore the potential to fold into a range of single-stranded secondary structures are evolutionarily conserved. Variations of this general REP sequence pattern are found in enterobacteria REPs, which are proposed to form hairpin stems with several noncomplementary bases, and REPs of pseudomonads and xanthomonads, which contain additional conserved nucleotides both 5' and 3' to the pseudo-palindrome domain.

Recognition between REP domains and RAYT is necessary for formation of the nucleoprotein complex and subsequent DNA strand cleavage and transfer. For *E. coli*¹⁴ and for *Haemophilus parasuis* (Nečasová et al. to be published) formation of such nucleoprotein complexes and subsequent DNA strand cleavage has been established. In their single stranded states, all identified palindromic REPs are potentially capable of intramolecularly folding into hairpins that possess a double helical stem linked by a short loop (Scheme 1a). The crystal structure of RAYT bound to a REP from *E. coli* displays such a stem-loop architecture, and identifies specific interactions between the conserved GTAG tetranucleotide of the REP and the binding domain of the RAYT enzyme.¹⁴ Nevertheless, given the potential of such G and C rich palindromic REP domains to form a myriad of other structural elements, it is unlikely that the stem-loop hairpin conformation represents a singular, obligatory RAYT recognition element. In our analysis, we do not assume that all the sequences identified by the selection criteria employed here represent obligatory RAYT binding sites, or that they share a common biological function. Rather, we emphasize that our data reveal an intriguing structural heterogeneity within the class of pseudo-palindromic sequences that fit our selection criteria, and that this heterogeneity provides additional and/or alternative recognition elements for transposase enzymes beyond the simple hairpin conformation.

In this work, we use a combination of spectroscopic and calorimetric techniques to explore the range of conformations

Table I General Sequence Features of REP Elements

Host Bacterial Taxa	REP Sequence Features	Example Sequences ^a	References
Various	GTRG <u>S(7–13)</u> <i>L(2–4)</i> <u>S(7–13)</u>	<i>SM1, SM4, SM8, Chom-22, Hpar1, Hpar2, SNBC-37-1</i>	2
Enterobacteria	GTAG <u>S(7)</u> AA <u>S(3–5)</u> <i>L(2–4)</i> <u>S(3–5)</u> <u>S(7)</u>	<i>Ecol</i>	15, 10
pseudomonads, xanthomonads	GTRG GA <u>S(6)</u> <i>L(2)</i> <u>S(6)</u> GAN	<i>Xcam</i>	8, 10, 2, 16

The 5' recognition tetranucleotides are denoted in bold. Palindromic (complementary) parts are denoted as “S” (stem) and underlined. Potential loop-forming nucleotides are denoted as “L” and italicized. The numbers of nucleotides in each component of the REP sequences are indicated in parentheses.

^a Abbreviations of the REP-related oligonucleotides studied here. Their sequences are listed in Table II.

adopted by single-stranded REP oligonucleotides in solution. To this end, we analyze REP sequences from *Escherichia coli*, *Stenotrophomonas maltophilia*, *Haemophilus parasuis*, *Xanthomonas campestris*, *Sulfurovum sp.*, and *Cardiobacterium hominis*. To assess the conformations of the REP elements and their variability, we employed temperature-dependent circular dichroism (CD) and ultraviolet (UV) spectroscopy. Thermal and thermodynamic properties of select REP elements were further characterized by differential scanning calorimetry (DSC). In the aggregate, our results reveal that, at least *in vitro*, oligonucleotides corresponding to single stranded REP elements from different bacterial species exhibit highly variable conformational behavior, manifesting a host of conformations beyond the expected stem-loop architecture, including tetraplex conformers, thereby expanding the potential recognition elements for transposase enzymes, while also suggesting potential pathways for modulation of transposase activity and biological control.

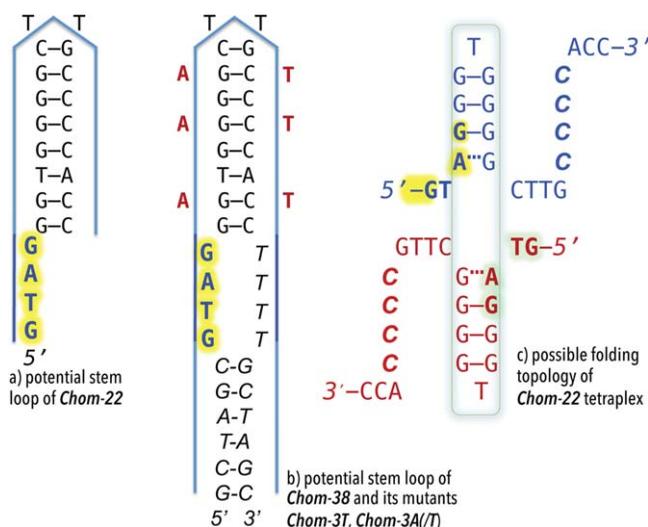
MATERIALS AND METHODS

Identification and Selection of REP Oligonucleotides for Experimental Study

The natural (“wild type”) genome sequences of REP oligonucleotides originating from six bacterial species (*Escherichia coli*, *Stenotrophomonas maltophilia*, *Haemophilus parasuis*, *Xanthomonas campestris*, *Sulfurovum sp.*, and *Cardiobacterium hominis*) were retrieved from bacterial genomes available at the NCBI genomic repository. Association with genes coding for RAYTs guided identification of REP elements, as described previously.^{10,16} The sequences of the REP oligonucleotides analyzed here are listed in Table II. Oligonucleotides were purchased from Integrated DNA Technologies. DNA concentrations were based on extinction coefficients reported by the manufacturer on the basis of nearest neighbor values. In addition to natural REP sequences, some sequences were modified to test properties of the analyzed REPs as discussed in the “Results” section. Oligonucleotides were dissolved in pH 7.4 buffer containing 100 mM Na⁺ cations, prepared by combining appropriate quantities of 59.8 mM NaCl, 20 mM Na₂HPO₄, 0.1 mM Na₂-EDTA mM with 79.8 mM NaCl, 20 mM NaH₂PO₄, 0.1 mM Na₂-EDTA mM. *C. hominis* and *H. parasuis* REP oligonucleotides were dissolved in 100 mM K⁺ buffer, pH 7.4, prepared by combining appropriate quantities of 60 mM KCl, 20 mM K₂HPO₄, 0.1 mM Na₂-EDTA mM with 80 mM KCl, 20 mM KH₂PO₄, 0.1 mM Na₂-EDTA mM. Before starting any experiments, the oligonucleotides were denatured by heating for 5 min at 90°C and allowed to cool to room temperature.

UV Absorption Melting Studies

Temperature-dependent UV absorbance was measured using either an Aviv DS14 UV-VIS spectrophotometer (Aviv Assoc., Lakewood, NJ) equipped with a 5-cell holder or a Chirascan-plusTM spectrometer (Applied Photophysics, Leatherhead, UK). Samples were placed in quartz cuvettes of 0.1–1 cm path length and scanned over the tem-



SCHEME 1 Possible architectures of *C. hominis* REP-related oligonucleotides. The recognition tetranucleotide GTAG is highlighted in yellow. (a) Schematic representation of a hairpin with stem formed by nucleotides forming W-C base pairs, and dinucleotide TT loop as shown for the example of the *Chom-22* sequence. (b) Schematic representation of the hypothetical stem-loop conformation of the 38-mer *Chom-38*. Red letters indicate positions of mutations to form either *Chom-3A*, *Chom-3T*, or *Chom-3A/T*. (c) Proposed pseudoknot-like architecture formed around the central bimolecular G-quadruplex core of *Chom-22*. The tetraplex is formed by the nucleotides highlighted by the gray rectangle; one molecule is drawn in blue, the other in red. The pyrimidine portion of the strands depicted here as unstructured are likely part of an intra- or cross-strand base pairing arrangement that would further stabilize the proposed pseudoknot like structure.

perature range of 5–95°C. The temperature was ramped stepwise in 0.5°C increments. At each temperature, the samples were equilibrated for 1 min and the absorbance at 260 nm (274 nm for *Chom-22* sequences) was recorded with a 10 s integration time. UV melting profiles were measured at different DNA strand concentrations (from 2 to 37 μM). The melting temperatures (T_m) for monomolecular (i.e., strand concentration independent) transitions were obtained from the first derivative of the optical melting curve using the *OriginPro 7.0* software. The temperature-dependent change of the ultraviolet absorbance of oligonucleotides was used to construct melting transition curves of the fraction of the remaining ordered structure.¹⁷ This procedure assumes that the absorbance at any temperature is a combination of contributions from only two spectral components, an ordered state and a disordered state. Using the two-state model, the melting curves were analyzed to determine van’t Hoff transition enthalpy changes of order–disorder transitions of REPs.

Circular Dichroism (CD) Spectra

CD spectra were recorded as a function of temperature using an Aviv DS62 (Aviv Assoc., Lakewood, NJ) and Chirascan-plusTM (Applied Photophysics, Leatherhead, UK) spectropolarimeters in steps of 1 nm over the wavelength range of 205–340 nm with an averaging time of

Table II 23 REP-Related Oligonucleotides Analyzed in this Study

Source Bacterium ^a	Symbol	Sequences of REP Oligonucleotides		
		Sequence	Length	WT ^b
<i>E. coli</i>	<i>Ecoli-cryst</i>	GTAGGACGGATAAGGCCGT TTACGCCGCATCCG	32	Y
<i>E. coli</i> ^c	<i>Ecoli-TT</i>	GTAGGACGGATAAGGCCGT TCACGCCGCATCCG GCA	35	Y
<i>E. coli</i>	<i>Ecoli-TC</i>	GTAGGACGGATAAGGCCGT CCACGCCGCATCCG GCA	35	Y
<i>S. maltophilia</i>	<i>SM1</i>	GGTGGGTGCCGACCG TTGGTCGGCAC	28	Y
<i>S. maltophilia</i> ^c	<i>SM4</i>	GTAGATCCACGCCAT GCGTGGAT	23	Y
<i>S. maltophilia</i>	<i>SM4-A/T</i>	GTAGAT CA ACGCCAT GCGTGGAT	23	N
<i>S. maltophilia</i>	<i>SM8</i>	GGTAGTGCCGGCC GCTGGCCGGCA	24	Y
<i>H. parasuis</i> ^c	<i>Hpar1</i>	GTAGGGTGGGG CTTGACCCACC	22	Y
<i>H. parasuis</i>	<i>Hpar2</i>	GTAGGTCGGGCAT TTATGCCCGAC	24	Y
<i>H. parasuis</i>	<i>Hpar2-A/T</i>	GTAGGTCGAGCAT TTATGCTCGAC	24	N
<i>H. parasuis</i>	<i>Hpar2-T/A</i>	GTAGGTCG TGCATTTATGCACGAC	24	N
<i>Sulfurovum sp.</i> ^c	<i>SNBC-37-1</i>	GTAGGGTGT TGTACCCCTACAACACC	26	Y
<i>Sulfurovum sp.</i>	<i>SNBC-T4</i>	GTAGGGTGT TGTATTTTACAACACC	26	N
<i>X. campestris</i> ^c	<i>Xcam</i>	GTAGGAGCGCG CTTGCCGCGATG	24	Y
<i>X. campestris</i>	<i>Xcam-mod</i>	GTAGGAGCTAG CTTGCTAGCGATG	24	N
<i>X. campestris</i>	<i>Xcam-4T</i>	GTAGGAGCTAG CTTTTGCTAGCGATG	26	N
<i>X. campestris</i>	<i>Xcam-4T-minus</i>	GTAGGAGCTAG CTTTTGCTAGC	22	N
<i>C. hominis</i> ^c	<i>Chom-22</i>	GTAGGGTGGGG CTTGCCCCACC	22	Y
<i>C. hominis</i>	<i>Chom-12</i>	GTAGGGTGGGGC	12	Y
<i>C. hominis</i>	<i>Chom-38</i>	GCTAGCGTAGGGTGGGGCTTGCCCCACCTTTTGCTAGC	38	N
<i>C. hominis</i>	<i>Chom-3A/T</i>	GCTAGCGTAGGATGAGACTTGCTCATCTTTTGCTAGC	38	N
<i>C. hominis</i>	<i>Chom-3A</i>	GCTAGCGTAGGATGAGACTTGCCCCACCTTTTGCTAGC	38	N
<i>C. hominis</i>	<i>Chom-3T</i>	GCTAGCGTAGGGTGGGGCTTGCTCATCTTTTGCTAGC	38	N

The recognition tetranucleotide is highlighted in bold, nucleotides mutated from the natural REP sequences (if any) are labeled in gray, nucleotides forming the stem of a putative hairpin are underlined, and the loop-forming residues are shown in *italics*.

^a *E. coli* = *Escherichia coli*, *S. maltophilia* = *Stenotrophomonas maltophilia*, *H. parasuis* = *Haemophilus parasuis*, *X. campestris* = *Xanthomonas campestris*, *C. hominis* = *Cardiobacterium hominis*.

^b Unmodified (“wild-type”) sequences are labeled “Y”, modified sequences are labeled “N”.

^c CD spectra of these oligonucleotides are shown in Figure 1. CD spectra and UV melting curves of all oligonucleotides are shown in Supporting Information Table S1.

10 s. The optimal signal-to-noise ratio was obtained from samples with concentrations between 2 and 20 μM (in either a 1 mm or a 10 mm path-length quartz cell) producing absorbance between 0.5 and 1 OD₂₆₀. The samples were placed into a thermostated cell holder and spectra were recorded in intervals of 5°C. The CD signal was expressed as the difference between the molar absorption of the right- and left-handed circularly polarized light and the resulting spectra after buffer spectrum subtraction were normalized by oligonucleotide concentration to yield molar ellipticities. To compare oligonucleotides of different lengths, the spectra were normalized on a per nucleotide basis.

To ascertain the minimum number of spectral species (in our case, assumed to be representative of DNA conformers) required to account for the observed global spectral changes, temperature-dependent CD spectra were subjected to single-value decomposition (SVD).^{18,19} Any number greater than two indicates a violation of the two-state assumption; that is, it signals the presence of more than one conformation in the native state or the existence of intermediate species in the order-disorder transition. Demonstration of more than two states obviates the application of the van't Hoff model.²⁰

Differential Scanning Calorimetry (DSC)

Measurements of the excess heat capacity as a function of temperature were performed using a Nano-DSCII differential scanning calorimeter (Calorimetry Science Corporation, UT). Dialyzed samples of nominal concentration 100 μM were placed in the sample cell, with the corresponding buffer solution in the reference cell. Samples were heated and cooled repeatedly at a constant temperature scanning rate of 1°C/min over the temperature range 0–95°C. The resulting power output divided by the scanning rate quantifies excess heat capacity as a function of temperature. Buffer versus buffer scans were subtracted from the sample versus buffer scans. After buffer subtraction, the excess heat capacity data were normalized for oligonucleotide concentration and analyzed as previously described.²¹ The calorimetric enthalpy change (ΔH_{cal}) was determined from the area under the measured excess heat capacity curve. The temperature at the peak of the excess heat capacity curve provides a measure of T_m for monomolecular processes. For two state transitions, the model-independent calorimetric enthalpy change for a given conformational transition must agree, within experimental error, with van't Hoff enthalpy changes determined from the optical melting curves.

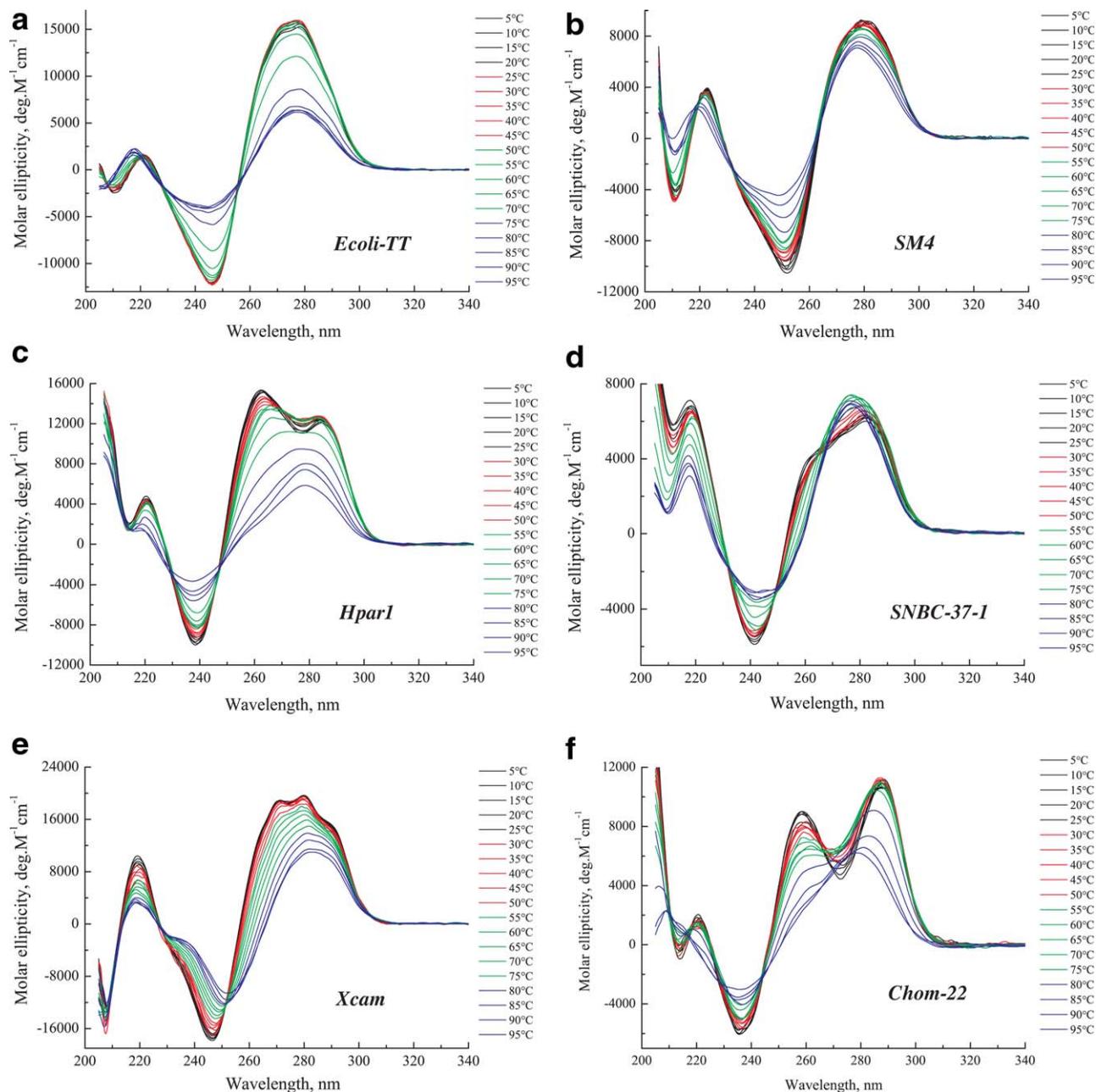


FIGURE 1 Temperature-dependent CD spectra of REP oligonucleotides **Ecoli-TT** (panel a), **SM4** (panel b), **Hpar1** (panel c), **SNBC-37-1** (panel d), **Xcam** (panel e), and **Chom-22** (panel f). Their sequences are listed in Table II. Supporting Information Table SI contains a summary of CD spectra and UV-melting curves of all oligonucleotides listed in Table II.

RESULTS

Overview

Figure 1 displays temperature-dependent CD spectra of six representative REP oligonucleotides, which generally reflect the different spectral types we observe. The corresponding

temperature-dependent CD spectra and UV melting curves for all 23 REP oligonucleotides are provided in Table SI of the Supporting Material. Broad structural characteristics derived from the experimental data are summarized in Supporting Information Table SII. Table III highlights salient general sequence features of a set of 203 REPs we have identified from whole

Table III Nucleotide Composition of 203 REP Sequences Associated with Putative RAYT Genes Observed in 105 Bacterial Species

Segment of the Palindrome	Mononucleotides ^a	G ^b	C ^b	A ^b	T ^b	dinucleotides ^a	GG ^b	CC ^b	tetranucleotide ^a	GGGG ^b	CCCC ^b	CGCG ^b	GCGC ^b	GAGG + GGAG ^b	AAAA + TTTT ^b
5'	1970	909	458	308	295	1767	319	98	1361	11	4	8	25	10	4
3'	1963	543	864	362	194	1649	97	317	1354	5	11	30	30	1	2

Sequences occurring significantly more often than expected in DNA composed of an equimolar mixture of G, A, C, and T nucleotides are shown in **bold**, those occurring less often are in *italics*. Statistical analysis of REP sequences is detailed in Supporting Information Table SIII.

^aThe total number of mono-, di-, and tetranucleotides in the sample of 203 REP sequences.

^bObserved frequencies of individual sequences.

genome searches, a more detailed version of the table can be found in Supporting Information Table SIII.

The sequences of the natural, “wild type,” genomic REP DNAs in part were identified on the basis of being composed of an imperfect palindrome located near a RAYT gene. A simplistic analysis of the sequences listed in Table II suggests that the corresponding oligonucleotides may fold into hairpins containing stems formed exclusively by canonical Watson-Crick (W-C) pairs. However, our spectral data reveal the conformational behavior of these oligonucleotides to be far more complex, with only a minority of these molecules exhibiting properties exclusively consistent with stem-loop structures. Our experimental data suggest that the majority of the oligonucleotides in solution form a mixture of competing secondary conformations. Even oligonucleotides anticipated to fold predominantly into hairpin structures exhibit CD spectra, inconsistent with the presence of a single, homogenous solution structure. Our data reveal the most distinct behavior for REPs of *C. hominis*, notably **Chom-22**. CD spectra of **Chom-22** and related oligonucleotide variants shown in Figure 2, display unique experimental observables that likely reflect solution formation of structures possessing a tetraplex core. Below we discuss the composition and sequence biases in REP elements, and the complex solution behavior we observe for select REP oligonucleotides.

G-Tracts and Alternating GC Dinucleotide Steps are Statistically Overrepresented in the 5' Half of REP Pseudo Palindromes

We identified 203 REP sequences from 105 bacterial species based on the general sequence criteria listed in Table I. This subset extracted from the genetic databases was analyzed in greater detail as to the base composition of each member, with these results being summarized in Table III. For further detail on these distributions, including a statistical *z* score analysis, the reader is referred to Supporting Information Table SIII in the Supporting Material. We separately analyzed the 5'- and 3'-segments of the palindromic domains forming putative stems, and the generally short central part of the palindrome assumed to form a hairpin loop. The universally conserved recognition tetranucleotide GTAG (or GTGG) was excluded from the analysis so as not to bias the results by its constancy.

Our analysis of sequence patterns identified an intriguing commonality amongst the compiled genomic data that goes beyond the imperfect palindrome criteria used initially to identify the relevant sequences. Namely, these sequences tend to have guanine-rich 5'-segments and corresponding complementary cytosine-rich 3'-segments; with adenosines and thymines

being much rarer in both segments. In contrast, the putative loop region, which represents about 20% of the total number of nucleotides in the analyzed REP sequences, has about the same number of all four nucleotides. Certain dinucleotides, trinucleotides, and tetranucleotides occur in the putative stem domains more frequently than expected for a DNA with equimolar G:A:C:T nucleotide composition. The most striking sequence feature is the high frequency of GG, GGG, and GRRG stretches in the 5'-segment of the stem, with the complementary pyrimidine sequences in the palindromic 3'-segment. Another typical feature of the stem domains is repetition of mixed G/C steps. GCGC appears quite frequently in both 5'- and 3'-segment of the stem, while CGCG is overpopulated only in the 3'-segment. We expect that the last two features—G- and C-tracts and/or repeated alternating GC dinucleotides in the predicted stem domain—make REP sequences conformationally complex. More specifically, frequent occurrence of several guanines in a row in the 5'-segment and of several cytosines in the 3'-segment suggests possible formation of folded DNA structures such as G-tetraplexes or i-motives.

Conformational Properties of the REP Oligonucleotides

In the following sections, we discuss conformational preferences of select REP oligonucleotides in terms of their solution behavior as revealed by our spectroscopic and calorimetric measurements. More detailed descriptions of the observed spectral features are given in the supporting information text and in Table SI.

REP Sequences that Predominantly form Stem–Loop Structures

CD spectra of REP oligonucleotides from *E. coli* and *S. maltophilia* reveal features characteristic of B-form DNA.²² Further, their UV denaturation curves exhibit apparent two-state, concentration independent, cooperative melting transitions. These observations, coupled with our SVD analysis revealing two significant spectral components, collectively are consistent with a monomolecular stem–loop architecture as the solution structure for these oligonucleotides. Additionally, for the *Ecoli-TT*, *Ecoli-TC*, and *SM4-A/T* oligonucleotides, we find agreement between the van't Hoff denaturation enthalpy change calculated from the optical melting curves and the enthalpy change directly measured by microcalorimetry, an equivalence reflective of a two-state transition. Taken together, our data are consistent with the presence in solution of a single monomolecular species at low temperatures with a B-form

component dominating the CD spectrum. The apparent hairpin architecture of the three measured *E. coli* sequences studied here (all naturally occurring in *E. coli*) is in accord with the stem–loop structure of *Ecoli-cryst* observed in the crystal structure of PDB code 4er8.¹⁴

The *SM1*, *SM4*, *SM8*, *Hpar2*, *Hpar2-A/T*, *Hpar2-T/A*, *Chom-3AT*, *Chom-3A* oligonucleotides exhibit CD spectral characteristics compatible with the existence of some B-form DNA in solution; namely, two maxima near 220 and 275 nm, and a negative peak near 245 nm (Supporting Information Table SI). Further, for most, but not all, of these REPs, their melting temperatures are concentration independent, consistent with a monomolecular structure. However, SVD analysis of their temperature-dependent CD spectra, and inspection of their UV melting curves reveal the presence of additional conformations besides the anticipated hairpin, a deduction consistent with a more heterogeneous population of conformers for these REP-related oligonucleotides.

The temperature-dependent changes in the CD spectra of the oligonucleotides *SM1*, *SM4*, *SM8*, and *Hpar2* suggest a more heterogeneous population of solution structures. The observed composition bias may contribute to the complex solution behavior of this group of REPs. To test this possibility, we systematically perturbed the properties of *Hpar2* by replacing its GGG and CCC triplets by GAG and CTC triplets in *Hpar2-A/T* and by GTG/CAC triplets in *Hpar2-T/A*. These mutations randomize the stem sequence and disrupt G/C tracts. In principle, these changes should increase the probability of the *Hpar2* variants adopting a hairpin. However, contrary to this expectation, we measure virtually identical spectral properties and melting curves for these three *Hpar2* oligonucleotide variants. We posit that these results suggest that the sequence isomers exist as several conformers in solution.

REP Sequences with Complex Spectral Features

Xanthomonas Campestris. CD spectra and UV melting curves of the REP-related oligonucleotides of *X. campestris* reveal complex conformational behavior. Features of the CD spectra of the natural sequence *Xcam* reflect the presence of some right handed helical structure; however, as revealed by the SVD analysis, this feature does not reflect a unique conformation. The Zuker predicted²³ most stable hairpin form of the *Xcam* oligonucleotide would be stabilized by a W–C stem formed by the hexanucleotide GCGCGC. The apparent contribution of other stem–loop forms to the CD spectra of *Xcam* may be explained by slippage of G–C pairs of the hexanucleotide leading to competition of multiple conformations in solution, thereby rationalizing the observed spectra. To disrupt the

regular GC hexanucleotide repeats and induce unique stem-loop architecture, the native *X. campestris* sequence was mutated to *Xcam-mod* by replacing GC by TA. To further favor hairpin formation, we also elongated the central TT to TTTT in *Xcam-4T* and *Xcam-4T-minus* to stabilize the predicted loop domain. In the latter oligonucleotide, we also attempted to stabilize the stem-loop by removing the 3'-terminal purine-rich tetranucleotide so that it would not interfere with base pairing interactions within the putative hairpin domain. As in the case of the already discussed sequences related to the *Hpar2* oligonucleotide, none of these modifications of the natural sequence led to significant changes in the unusual *Xcam* CD spectra. While an unequivocal interpretation of solution behavior of *Xcam* and its sequence variants is not possible on the basis of our spectral data alone, it is clear that these oligonucleotides form multiple states/conformers in solution beyond the simple stem-loop hairpins.

***Sulfurovum* sp.** The REP *SNBC-37-1* sequence potentially could form a hairpin with an extensive stem composed of nine W-C base pairs and a loop formed by four unpaired cytosines (or thymines in the modified sequence of *SNBC-T4*). However, as in several previous cases discussed, the CD spectra are consistent with more complex equilibria in solution. The *SNBC-37-1* oligonucleotide displays a combination of spectral features suggesting the presence of both B-form conformation and a parallel guanine quadruplex in solution.

REP Sequences Forming Guanine-Rich Tetraplexes

Oligonucleotides related to REP elements of *Cardiobacterium hominis* exhibit the most intriguing behavior. The natural

sequence *Chom-22* consists of a palindrome predicted to form a stem-loop structure with an eight W-C base pair stem linked by two thymines in a putative hairpin loop (Scheme 1a). However, the 5'-segment of the palindrome also contains two consecutive G-tracts separated by a single T. The first of these tracts is an imperfect G-tract containing a 5' A followed by three consecutive G's. The second is a perfect G4-tract. Similar patterns of two consecutive tri- or tetranucleotide G-tracts separated by one or two other nucleotides also are seen in other G rich REP sequences, including *Hpar1*. The low temperature CD spectrum of *Chom-22* is highly unusual, forming two positive peaks at 289 and 259 nm, a positive saddle at 272 nm, and a negative peak at 236 nm (Figure 1). A CD spectrum with two positive peaks around 290 and 260 nm is indicative of an anti-parallel G-tetraplex.^{24,25} However, the spectrum we observe is not fully consistent with CD spectra of the "classic" intramolecular anti-parallel tetraplexes. The most likely reason is the presence of the cytosine-rich 3'-part of the sequence, which will contribute in unknown ways to the CD spectra. The CD spectra of naturally occurring REP *Hpar1* from *H. parasuis* (not further discussed here) also shows CD spectral features suggestive of an antiparallel G-tetraplex conformation (Supporting Information Table SI).

Based on these observations, we propose that the unusual CD spectra of *Chom-22* arise primarily from interactions of its guanine tracks. To test this possibility, we truncated *Chom-22* into *Chom-12*, keeping only the purine-rich 5'-part of the sequence, thereby minimizing possible W-C base pairing. CD spectra of *Chom-12* reveals features different from that of *Chom-22* (Figure 2), with similarities to spectra reported for poly(dG)•poly(dC)²⁶ and for G-rich oligonucleotides forming DNA wire

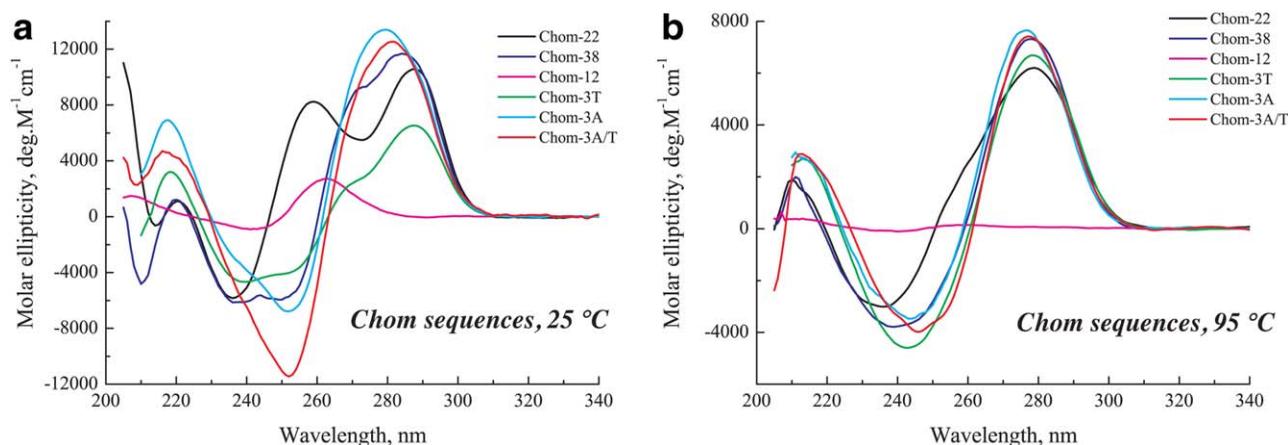


FIGURE 2 Comparison of CD spectra of the REP-related oligonucleotides *Chom-22*, *Chom-38*, *Chom-12*, *Chom-3T*, *Chom-3A*, and *Chom-3A/T* from *Cardiobacterium hominis* at 25°C (left panel) and 95°C (right panel). The spectra measured at 25°C show significant differences, while the spectra of denatured oligonucleotides measured at 95°C are highly similar. Sequences of all oligonucleotides are listed in Table II.

brush structures.²⁷ Because poly(dG)•poly(dC) polymers and DNA wire brushes often contain a high percentage of poly(dG) tetraplex with a parallel arrangement of the strands, we postulate that **Chom-12** also forms an intermolecular parallel stranded tetraplex. Although the structures adopted by **Chom-12** and **Chom-22** are different, they both are consistent with the possibility of structure formation based on G-quartet base pairing.

To test further our prediction that the observed solution behavior of **Chom-22** is dominated by formation of an antiparallel G-tetraplex rather than a stem-loop hairpin, we systematically mutated the **Chom-22** sequence to either enhance or disrupt the probability of tetraplex formation. To mimic longer duplex DNA segments, in which **Chom-22** REP sites are embedded in the natural DNA, we added short self-complementary tail sequences of 6 nucleotides to both the 5'- and the 3'-ends of the original Chom22 sequence (Scheme 1a) to form Chom 38 (Scheme 1b). The nucleotide sequences chosen for the tail segments are arbitrary since no consensus was found during sequence analysis outside the *C. hominis* REP domain. A tetranucleotide T₄ was added as a flexible hinge between the **Chom-22** parent sequence and the tail domain on the 3'-side to account for the recognition sequence GTAG on the 5' side (Scheme 1a,b). To enhance the likelihood of tetraplex formation in **Chom-38**, we strategically inserted 3 T's in place of C's in the 3' half of the palindrome to minimize the Watson-Crick complementarity, thereby forming **Chom-3T**. As a further test of our hypothesis, to disrupt the probability of tetraplex formation we replaced 3 G's in the G-tracts of the purine-rich 5' half of the **Chom-38** palindrome by 3 A's to form **Chom-3A**. We also simultaneously mutated G's to A's as well as the complementary C's to T's to form **Chom-3AT**.

As shown in Figure 2, the CD spectra of **Chom-38** are consistent with the molecule adopting a similar antiparallel tetraplex as the parent **Chom-22** molecule, although its spectral properties are subtly altered, most likely reflecting contributions from the additional bases and the structural constraints they introduce. Importantly, the mutant **Chom-3T**, designed to enhance the probability of tetraplex formation, exhibited CD spectra similar to the spectrum of the **Chom-22** parent, including a positive band near 290 nm. **Chom-3T** does exhibit a cooperative temperature transition at a much lower melting temperature (51°C) than **Chom-22** and **Chom-38**, suggesting that, in addition to the G-tracts, the C₄-tracts in **Chom-22** and **Chom-38** contribute to the stability of the folded conformation in solution. In contrast, and in agreement with our design principles, the CD spectra and sigmoidal melting curves of the **Chom-3A** and **Chom-3AT** mutants lack features typical of antiparallel G-tetraplexes. In fact, their CD spectra more closely resemble conventional B-DNA like spectra, with a prevalence of W-C base pairing, although the spectra are slightly distorted

relative to those of generic B-DNA, perhaps reflecting base composition effects.

Taken together, these observations suggest that the dominant conformation adopted by the **Chom-22** oligonucleotide in solution is not a stem-loop hairpin but rather a pseudoknot-like arrangement formed around a central bimolecular G-tetraplex core, as represented in Scheme 1c. Such a bimolecular pseudoknot-like arrangement leaves the pyrimidine-rich segment nominally unpaired. However, cytosine tracks can also adopt secondary two- and four-stranded structures in solution, especially at low pH.²⁸⁻³⁰ Formation of an ordered cytosine-based structure might further contribute stabilizing interactions to the central tetraplex structure. Importantly, the significant differences observed between the CD spectra of the **Chom** mutants (Figure 2) at low temperatures are largely absent in spectra of the denatured species measured at 95°C. The low temperature spectral differences therefore predominantly reveal differences in conformation of the native state rather than differences in the base composition.

DISCUSSION

REP elements are abundant² and mobile⁴ segments of bacterial genomes, with the latter characteristic possibly due to their association with transposase RAYT,¹⁰ which has been postulated to serve as their mobilizer. The CD, UV melting, and DSC data presented here reveal complex conformational behavior of single stranded repetitive extragenic DNA elements (REPs), which we judge to be important in differential recognition and regulation. We propose that RAYT recognition and processing depends on a complex dynamic equilibrium among hairpin stem-loop, nonhairpin single stranded structure, the duplex, and alternative folding forms, such as a bimolecular guanine tetraplex pseudoknot-like structure. DNA states recognized by RAYTs may be only marginally populated in some cases. RAYT discrimination between REP target sequences based on recognition of alternative structures or on the basis of dynamic equilibria between different DNA conformations potentially provides an additional level of biological control of transposition events. Conventional chemical and/or enzymatic probing cannot be performed to assess the proposed secondary structures of such REP elements, since the cleaving reagents are confronted by complex dynamic ensembles, with differential initial state reactivities, thereby making it impossible to link unambiguously processed product with a specific initial state or reactant conformation.

Dynamic equilibria between different DNA conformations as biological regulators are not without precedent. A related equilibrium between a stem-loop conformation and an

intramolecular G-quadruplex formed from 4 consecutive G tracts very recently has been suggested for oligonucleotides based on the WNT1 gene promoter region.³¹ As elaborated below, there is a long history of diverse, noncanonical DNA states serving important roles in modulating differential recognition and biological regulation.

Soon after the first diffraction studies of DNA fibers in 1950s, diffraction and other biophysical studies have revealed a multitude of well-defined non-B conformations that suggest considerable plasticity of DNA with potentially important genetic implications.^{32,33} Some of these noncanonical forms have been investigated by detailed thermodynamic analysis [e.g., Refs. (34–37)]. Noncanonical structures, such as triplexes,^{38,39} G-tetraplexes,^{40,41} and cruciforms⁴² usually form within specific sequence motifs.³³ These structures may form transiently, but they may also be stabilized under certain cell-mimicking conditions where the DNA duplex is destabilized.⁴³ Noncanonical DNA structures have been shown to play a role as regulators of various molecular processes that involve the unwinding of double-stranded DNA during bacterial conjugation or viral infections,⁴⁴ replication and transcription,⁴⁵ and/or recombination.⁴⁶ It has been demonstrated that gene transcription can be arrested or paused at or near stable noncanonical structures. Z-form duplexes,⁴⁷ triplexes,^{48,49} or G-quadruplexes^{50,51} all have been implicated in transcription arrest, while a cytosine tetraplex structure, the so called *i*-motif,^{52,53} has been shown to regulate gene expression.⁵⁴

Guanine tetraplexes have emerged as some of the more important noncanonical DNA structures in genomes. It has been reported that intramolecular G-tetraplex structures are present inside cells,^{55,56} and become targets of helicases.⁵⁷ Bioinformatic evidence has provided insight into sequence diversity and categories of G-tetraplex-forming sequences in intronic regions of the human genome⁵⁸ and their sequence, topology, and structural properties have been reviewed.⁵⁹ Such structures also are enriched upstream and within gene promoters, which can lead to inhibition of transcription initiation,⁶⁰ and may regulate retro-transposition in plants.⁶¹ G-tetraplex structures have become a promising target of anticancer drugs,⁶² which suggests that tetraplex prone REP elements represent potential drug targets as well.

The conformationally rich behavior of the REP elements described here originates from their highly nonrandom sequences (Table III and Supporting Information Table SIII). Notably, some oligonucleotides, which could adopt extensive stems formed by 8–9 Watson–Crick base pairs, exhibit complex CD spectral and thermal characteristics incompatible with the hairpin architecture, likely reflecting the existence of several competing conformers (*Xcam*, *SNBC-37-1*), and/or the

presence of tetraplexes (*Chom-22*, *Hpar1*). On the other hand, the *E. coli* REP oligonucleotides studied here form hairpins with stems containing non-W–C pairs, which stabilize the single stranded DNA structure similar to what occurs in RNA single strands.

The patterns of guanine tracks encountered in some of the REP sequences analyzed here are reminiscent of those found in telomere sequences that fold into intramolecular tetraplexes, where there are 4 such repeats in a row. The subgroup of REP sequences studied here has only two such repeats, but given their pseudo-palindromic nature, two additional guanine track repeats are located in the opposing DNA strand. We therefore propose that these REP sequences form a pseudoknot-type arrangement tying the two opposing strands together, a configuration which would represent a novel recognition element within genomic DNA (Scheme 1c). That such an arrangement is possible recently has been demonstrated in a crystal structure of a bimolecular G-tetraplex with antiparallel arrangement of strands folding on itself.⁶³ A pseudoknot type G-tetraplex of two REP sequences would create a unique topology, starting at the GTAG recognition sequence, which might “prime” these sites for cleavage by the transposases. Such a construct opens the possibility that in some bacteria the recognized structure of the REP is a guanine tetraplex or pseudoknot type arrangement. We postulate that structures similar to the proposed pseudoknot-like structures form around a G-quartet with guanines from opposing DNA strands, and constitute a new class of recognition and control elements within bacterial DNA. Because pseudo-palindromes also can fold into hairpin stem-loop structures of the kind seen in the *E. coli* RAYT crystal structure,¹⁴ it is conceivable, as noted above, that RAYT recognition and processing depends on a complex dynamic equilibrium between duplex, bimolecular guanine tetraplex pseudoknot-like structure, and hairpin stem-loop structures, with the latter state perhaps only being marginally populated in some cases.

In summary, in contrast to the singular, static “hairpin stem-loop” recognition motif commonly referred to in the literature, the results reported here reveal that REP elements can be conformationally heterogeneous, reflective of a complex conformational landscape with multiple competing secondary structures from which RAYT’s select, bind, and process their cognate recognition state. Our results broaden the conformational landscape implicated in the control, recognition, and function of such elements.

CONCLUSION

We have presented evidence that oligonucleotides composed of REP sequences from six bacterial species (*Escherichia coli*,

Stenotrophomonas maltophilia, *Haemophilus parasuis*, *Xanthomonas campestris*, *Sulfurovum* sp., and *Cardiobacterium hominis*) adopt a range of noncanonical DNA conformations. This propensity to adopt alternative DNA conformations is a consequence of sequence features characteristic of REPs, which we identified in over a hundred bacterial species. In particular, the frequent occurrence of consecutive G tracts in the 5' half of the REP pseudo-palindrome predisposes these DNA domains to adopt a pseudoknot like arrangement around a bimolecular tetraplex core. We propose that such a bimolecular tetraplex pseudoknot represents a novel DNA recognition element that adds to the range of DNA tetraplex structures reported to have important functions in DNA transcription and regulation. We further propose that the detected rich conformational diversity of REP-related pseudo-palindromic oligonucleotides, which exist in multiple conformations in dynamic equilibrium, represents a critical feature of RAYT recognition elements. The complex conformational landscape of REP sequences provides additional context, within which to understand the recognition and regulatory roles of these abundant genetic bacterial elements.

We thank Dr. Eric Plum for his critical reading of the manuscript.

REFERENCES

- Higgins, C. F.; Ames, G. F.; Barnes, W. M.; Clement, J. M.; Hofnung, M. *Nature* 1982, 298, 760–762.
- Di Nocera, P. P.; De Gregorio, E.; Rocco, F. *BMC Genomics* 2013, 14, 522.
- Bertels, E.; Rainey, P. B. *PLoS Genet* 2011, 7, e1002132.
- Bertels, E.; Rainey, P. B. *Mobile Genet Elem* 2011, 1, 262–268.
- Gilson, E.; Saurin, W.; Perrin, D.; Bachellier, S.; Hofnung, M. *Nucleic Acids Res* 1991, 19, 1375–1383.
- Bachellier, S.; Clement, J. M.; Hofnung, M. *Res Microbiol* 1999, 150, 627–639.
- Espeli, O.; Moulin, L.; Boccard, F. *J Mol Biol* 2001, 314, 375–386.
- Aranda-Olmedo, I.; Tobes, R.; Manzanera, M.; Ramos, J. L.; Marques, S. *Nucleic Acids Res* 2002, 30, 1826–1833.
- Rocco, F.; De Gregorio, E.; Di Nocera, P. P. *FEMS Microbiol Lett* 2010, 308, 185–192.
- Nunvar, J.; Huckova, T.; Licha, I. *BMC Genomics* 2010, 11, 44.
- Chandler, M.; de la Cruz, F.; Dyda, F.; Hickman, A. B.; Moncalian, G.; Ton-Hoang, B. *Nat Rev Microbiol* 2013, 11, 525–538.
- Gonzalez-Prieto, C.; Agundez, L.; Linden, R. M.; Llosa, M. *Trends Biotechnol* 2013, 31, 305–312.
- Ton-Hoang, B.; Siguiet, P.; Quentin, Y.; Onillon, S.; Marty, B.; Fichant, G.; Chandler, M. *Nucleic Acids Res* 2012, 40, 3596–3609.
- Messing, S. A.; Ton-Hoang, B.; Hickman, A. B.; McCubbin, A. J.; Peaslee, G. F.; Ghirlando, R.; Chandler, M.; Dyda, F. *Nucleic Acids Res* 2012, 40, 9964–9979.
- Gilson, E.; Bachellier, S.; Perrin, S.; Perrin, D.; Grimont, P. A.; Grimont, F.; Hofnung, M. *Res Microbiol* 1990, 141, 1103–1116.
- Nunvar, J.; Licha, I.; Schneider, B. *BMC Genomics* 2013, 14, 385.
- Breslauer, K. J. *Methods Enzymol* 1995, 259, 221–242.
- Henry, E. R.; Hofrichter, J. *Methods Enzymol* 1992, 210, 129–192.
- Johnson, W. C. *Methods Enzymol* 1992, 210, 426–447.
- Gray, R. D.; Chaires, J. B. *Curr Protoc Nucleic Acid Chem* 2011; Chapter 17, Unit 17.14.
- Marky, L. A.; Breslauer, K. J. *Biopolym* 1987, 26, 1601–1620.
- Bush, C. A. In *Basic Principles in Nucleic Acid Chemistry*; T'so, P. O. P., Ed.; Academic: New York, 1974, pp 91–169.
- Zuker, M. *Nucleic Acids Res* 2003, 31, 3406–3415.
- Kypr, J.; Kejnovska, I.; Rencuk, D.; Vorlickova, M. *Nucleic Acids Res* 2009, 37, 1713–1725.
- Karsiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. *Angew Chem Int Ed Engl* 2011, 50, 10645–10648.
- Gray, D. M. *Biopolymers* 1974, 13, 2087–2102.
- Poon, K.; Macgregor, R. B., Jr. *Biophys Chem* 2000, 84, 205–216.
- Mergny, J. L.; Lacroix, L. *Nucleic Acids Res* 1998, 26, 4797–4803.
- Bucek, P.; Jaumot, J.; Avino, A.; Eritja, R.; Gargallo, R. *Chem Eur J* 2009, 15, 12663–12671.
- Kaushik, M.; Suehl, N.; Marky, L. A. *Biophys Chem* 2007, 126, 154–164.
- Kuo, M. H.; Wang, Z. F.; Tseng, T. Y.; Li, M. H.; Hsu, S. T.; Lin, J. J.; Chang, T. C. *J Am Chem Soc* 2015, 137, 210–218.
- Bacolla, A.; Wells, R. D. *Mol Carcinog* 2009, 48, 273–285.
- Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. *Cell Mol Life Sci* 2010, 67, 43–62.
- Leroy, J. L.; Gehring, K.; Kettani, A.; Gueron, M. *Biochemistry* 1993, 32, 6019–6031.
- Frank-Kamenetskii, M. D.; Mirkin, S. M. *Annu Rev Biochem* 1995, 64, 65–95.
- Keniry, M. A. *Biopolymers* 2000, 56, 123–146.
- Völker, J.; Klump, H. H.; Breslauer, K. J. *Proc Natl Acad Sci USA* 2008, 105, 18326–18330.
- Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab, F. *FASEB J* 1988, 2, 2939–2949.
- Mirkin, S. M.; Frank-Kamenetskii, M. D. *Annu Rev Biophys Biomol Struct* 1994, 23, 541–576.
- Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J. L. *Nucleic Acids Res* 2007, 35, 3064–3075.
- Campbell, N.; Collie, G. W.; Neidle, S. *Curr Protoc Nucleic Acid Chem* 2012, Chapter 17, Unit 17.16.
- Murchie, A. I. H.; Lilley, D. M. J. *Methods Enzymol* 1992, 211, 158–180.
- Miyoshi, D.; Karimata, H.; Sugimoto, N. *J Am Chem Soc* 2006, 128, 7957–7963.
- Trakselis, M. A.; Graham, B. W. *Nature* 2012, 492, 195–197.
- Bansal, M.; Kumar, A.; Yella, V. R. *Curr Opin Struct Biol* 2014, 25C, 77–85.
- Bikard, D.; Loot, C.; Baharoglu, Z.; Mazel, D. *Microbiol Mol Biol Rev* 2010, 74, 570–588.
- Ditlevson, J. V.; Tornaletti, S.; Belotserkovskii, B. P.; Teijeiro, V.; Wang, G.; Vasquez, K. M.; Hanawalt, P. C. *Nucleic Acids Res* 2008, 36, 3163–3170.
- Ebbinghaus, S. W.; Gee, J. E.; Rodu, B.; Mayfield, C. A.; Sanders, G.; Miller, D. M. *J Clin Invest* 1993, 92, 2433–2439.

49. Belotserkovskii, B. P.; De Silva, E.; Tornaletti, S.; Wang, G.; Vasquez, K. M.; Hanawalt, P. C. *J Biol Chem* 2007, 282, 32433–32441.
50. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc Natl Acad Sci USA* 2002, 99, 11593–11598.
51. Broxson, C.; Beckett, J.; Tornaletti, S. *Biochemistry* 2011, 50, 4162–4172.
52. Kang, C.; Zhang, X. H.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* 1992, 356, 126–131.
53. Ahmed, S.; Henderson, E. *Nucleic Acids Res* 1992, 20, 507–511.
54. Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J Am Chem Soc* 2014, 136, 4161–4171.
55. Han, H.; Hurley, L. H. *Trends Pharmacol Sci* 2000, 21, 136–142.
56. Lam, E. Y.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. *Nat Commun* 2013, 4, 1796.
57. Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N. *Nat Chem Biol* 2014, 10, 313–318.
58. Todd, A. K.; Neidle, S. *Nucleic Acids Res* 2011, 39, 4917–4927.
59. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res* 2006, 34, 5402–5415.
60. Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat Rev Genet* 2012, 13, 770–780.
61. Lexa, M.; Kejnovsky, E.; Steflava, P.; Konvalinova, H.; Vorlickova, M.; Vyskot, B. *Nucleic Acids Res* 2014, 42, 968–978.
62. Neidle, S.; Read, M. A. *Biopolymers* 2000, 56, 195–208.
63. Wei, D.; Todd, A. K.; Zloh, M.; Gunaratnam, M.; Parkinson, G. N.; Neidle, S. *J Am Chem Soc* 2013, 135, 19319–19329.

Reviewing Editor: Gary Glick