*Published online 07 March 2014*

*Nucleic Acids Research, 2014, Vol. 42, No. 9* ***5755–5764***
*doi: 10.1093/nar/gku185*

# Molecular population dynamics of DNA structures in a bcl-2 promoter sequence is regulated by small molecules and the transcription factor hnRNP LL

**Yunxi Cui[1], Deepak Koirala[1], HyunJin Kang[2], Soma Dhakal[1], Philip Yangyuoru[1], Laurence H. Hurley[2,3,4] and Hanbin Mao[1,*]**

[1]Department of Chemistry and Biochemistry and School of Biomedical Sciences, Kent State University, Kent, OH 44242, USA, [2]College of Pharmacy, University of Arizona, 1703 East Mabel Street, Tucson, AZ 85721, USA, [3]Arizona Cancer Center, 1515 North Campbell Avenue, Tucson, AZ 85724, USA and [4]BIO5 Institute, 1657 East Helen Street, Tucson, AZ 85721, USA

Received November 8, 2013; Revised February 3, 2014; Accepted February 18, 2014

## ABSTRACT

**Minute difference in free energy change of unfolding among structures in an oligonucleotide sequence can lead to a complex population equilibrium, which is rather challenging for ensemble techniques to decipher. Herein, we introduce a new method, molecular population dynamics (MPD), to describe the intricate equilibrium among non-B deoxyribonucleic acid (DNA) structures. Using mechanical unfolding in laser tweezers, we identified six DNA species in a cytosine (C)-rich bcl-2 promoter sequence. Population patterns of these species with and without a small molecule (IMC-76 or IMC-48) or the transcription factor hnRNP LL are compared to reveal the MPD of different species. With a pattern recognition algorithm, we found that IMC-48 and hnRNP LL share 80% similarity in stabilizing i-motifs with 60 s incubation. In contrast, IMC-76 demonstrates an opposite behavior, preferring flexible DNA hairpins. With 120–180 s incubation, IMC-48 and hnRNP LL destabilize i-motifs, which has been previously proposed to activate bcl-2 transcriptions. These results provide strong support, from the population equilibrium perspective, that small molecules and hnRNP LL can modulate bcl-2 transcription through interaction with i-motifs. The excellent agreement with biochemical results firmly validates the MPD analyses, which, we expect, can be widely applicable to investigate complex equilibrium of biomacromolecules.**

## INTRODUCTION

Unlike proteins in which native structures are often the most stable conformation in an amino acid sequence (1), conformation polymorphism with similar stability is often seen in a deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sequence (2–4). The disparity can be rationalized by different organization strategies employed in these macromolecules. The cooperative arrangement in proteins is a result of intricate cross-talks within or between secondary or tertiary structures. Once the cross-talks are compromised, the overall architecture may collapse as neither secondary nor tertiary structures are stable as stand-alone units (5). Nucleic acid structures, however, are hierarchical (6–8). Even without the higher order interactions, stand-alone secondary conformations are stable. Secondary or higher order structures in nucleic acids are stabilized by Watson–Crick (WC) base pairing in duplex strands (9) or Hoogsteen bonding in tetraplex strands (10–12). As energetic difference is small between different WC and Hoogsteen H-bonds, it becomes possible that multiple structures may coexist in the same nucleic acid sequence (13,14).

Not only different conformations exist in the same sequence, but also their interconversions occur frequently (15–17). In the context of naturally existing duplex DNA, each complementary strand can host a different set of structures. For example, in a sequence of more than four tracts of guanine (G)-rich repeats, multiple possibilities of G-quadruplex (GQ) (18) can form. Each GQ requires four tracts of G-repeats to fold into a stack of planar G-quartets, which consists of four guanines cross-linked by Hoogsteen bonding. In the complementary cytosine (C)-rich strand, i-motif structures can form (19). Similar to GQ, each i-motif requires four tracts of C-rich repeats to fold into a stack of hemiprotonated cytosine:cytosine pairs, which are more stable in DNA than RNA (20,21). In promoter sequences upstream of the transcriptional start site, negative

*To whom correspondence should be addressed. Tel: +1 330 672 9380; Fax: +1 330 672 3816; Email: hmao@kent.edu

superhelicity generated by transcriptional firing (22) results in the potential of forming GQs and i-motifs in complementary stands. In the case of the insulin promoter, these have been shown to be mutually exclusive (23). The situation is more complex during the transcription, in which the nascent RNA strand may participate in the equilibrium by forming structures in the RNA strand or those between RNA and DNA strands (24,25). All these possibilities bring a high level of complexity in the investigation of nucleic acids structures.

It is possible that only one or a few species are active in a biological process. To understand the biological roles of active nucleic acid structures, therefore, it is necessary to clarify the entire population equilibrium. Conventional experimental techniques such as nuclear magnetic resonance, Circular Dichroism (CD) and X-ray provide detailed structural information on pure species (26–28). However, when it comes to a mixture, these methods often fail to resolve individual structures due to their ensemble average nature. Single-molecule approaches can provide a unique advantage to decipher individual species in a complex mixture. They also have an excellent capability of probing dynamic processes (29), which is rather challenging for traditional methods due to their reduced temporal resolutions. These capabilities afford single-molecule techniques unparalleled perspectives to probe the population dynamics of a complex system.

In ecological biology, population dynamics have been used to describe the change in biological populations due to processes such as immigration, emigration, birth or death (30). Factors such as climate or geographical locations can be dissected to reveal the specific effect on the population dynamics (31). The folding and unfolding of structures in a nucleotide sequence and the interconversion among these structures closely resemble the population change in biological species. Therefore, we apply the concept of population dynamics to describe the effect of different factors, such as ligands and proteins, on the population pattern of multiple non-B DNA species that can form in a DNA fragment. To differentiate our approach from that used in biology, we name this method as molecular population dynamics (MPD). Compared to current ensemble approaches in which the influence of external factors on population equilibrium is described rather qualitatively, the MPD method allows a quantitative measurement, such as similarity and additivity (see below), to compare these factors in an intuitive and straightforward fashion.

In previous publications, Hurley and coworkers have described that i-motif structures in the bcl-2 promoter region are transcriptional modulators (32,33). The population equilibrium in this system is highly complex. Not only multiple i-motif structures compete for folding in the fragment, 5′-CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT-3′ (see Figure 1a), which contains six C-rich tracts for a minimal of 15 i-motifs, the binding of a transcription factor hnRNP LL or a small molecule (IMC-76 or IMC-48; see Supplementary Figure S1 for structures) can influence the equilibrium as well. With a highly sensitive *Po*pulation *D*econvolution at *Nano*meter resolution method (or PoDNano) recently developed in our lab (13,34), we wish to decipher this complex equilibrium and identify relevant populations responsible for the bcl-2 transcription from the perspective of MPD.

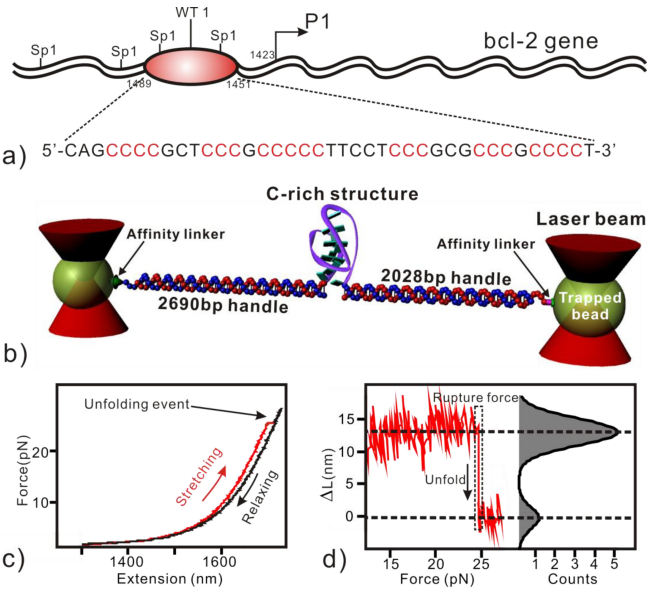


**Figure 1.** Mechanical unfolding of the structures formed in a C-rich fragment of the bcl-2 promoter region by laser tweezers. (a) Location of the C-rich fragment in the upstream of the P1 promoter in the bcl-2 gene. (b) The DNA construct that contains the C-rich fragment is sandwiched between two beads trapped by the laser tweezers. (c) A typical force–extension (*F*–*X*) curve. The red and black curves represent the stretching and relaxing processes, respectively. (d) Left: a plot of change in contour length ($\Delta L$) versus force. The change in extension in (c) has been converted to the $\Delta L$ (see text). Right: histograms of folded (top) and unfolded populations (bottom). The black curve represents two-peak Gaussian fitting.

First, we measured the population pattern of species in the bcl-2 promoter fragment in the presence of modulators IMC-76, IMC-48 or hnRNP LL with 60 s incubation. After comparing the population pattern without a modulator, the effect of each modulator on the MPD is revealed. We confirmed biochemical findings (32,33) that IMC-48 or hnRNP LL stabilizes i-motif species over flexible DNA hairpins while IMC-76 shows an opposite behavior. With a simple pattern recognition algorithm, we estimated 80% similarity between the effects of IMC-48 and hnRNP LL. The similarity drops to 40% between IMC-76 and IMC-48, and 30% between IMC-76 and hnRNP LL. A mixture of IMC-48 and hnRNP LL has an additive effect on the i-motif population dynamics (100% in additive probability), suggesting that they stabilize i-motif structures through different sites. With 120–180 s incubation, hnRNP LL and IMC-48 show a destabilization effect on i-motif populations. These results suggest that hnRNP LL first binds to i-motif species, followed by the unfolding of these structures, which is consistent with those proposed for the activation of bcl-2 transcription (33). Overall, our MPD analyses provide strong support, at the level of population equilibrium, that small molecules (IMC-76 and IMC-48) and the transcription factor hnRNP LL could modulate bcl-2 transcription through interaction with i-motif structures in the bcl-2 promoter region. Such a scenario provides evidence that DNA species may modulate transcription in a fashion similar to that of translational control by riboswitches (35).

## MATERIALS AND METHODS

Unless particularly noted, all DNA oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). All chemicals with >99% purity were purchased from VWR (West Chester, PA, USA). Enzymes were purchased from New England Biolabs (NEB, Ipswich, MA, USA) and surface functionalized beads for the laser tweezers experiments were obtained from Spherotech (Lake Forest, IL, USA). The IMC-48, IMC-76 and hnRNP LL were prepared as described in recently submitted papers (32,33).

### Preparation of DNA constructs

The DNA constructs used for single-molecule mechanical unfolding and refolding experiments in the bcl-2 promoter region were prepared by sandwiching the 5′-CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT-3′ sequence between two double-stranded DNA (dsDNA) handles according to the procedures described previously (36). To reduce the interference from the DNA handles on the bcl-2 fragment, two deoxythymidines were added at each end of the bcl-2 fragment. Briefly, the 2690-bp dsDNA handle was prepared based on the pEGFP vector (Clontech, Mountain View, CA, USA). The vector was first digested by SacI and EagI restriction enzymes and then purified with agarose gel. The SacI end was labeled with digoxigenin by terminal deoxynucleotidyl transferase. The 2028-bp dsDNA handle was amplified from the pBR322 plasmid by polymerase chain reaction (PCR). One end of the handle was labeled with biotin through a modified PCR primer. The other end with an XbaI restriction site was incorporated through another PCR primer. The 2028-bp dsDNA handle was digested with XbaI. A single-stranded DNA (ssDNA) target that contained the bcl-2 fragment (5′-CTA GAC GGT GTG AAA TAC CGC ACA GAT GCG TTC AGC CCC GCT CCC GCC CCC TTC CTC CCG CGC CCG CCC CTT GCC AGC AAG ACG TAG CCC AGC GCG TC-3′) was hybridized with two other DNA oligos (5′-CGC ATC TGT GCG GTA TTT CAC ACC GT-3′ and 5′-GGC CGA CGC GCT GGG CTA CGT CTT GCT GGC-3′), resulting in a dsDNA–ssDNA hybrid assembly with EagI and XbaI overhangs at the two ends and a single-stranded bcl-2 promoter fragment in the middle. Finally, this dsDNA–ssDNA hybrid and the two dsDNA handles were ligated by T4 DNA ligase to obtain the final DNA construct.

### Single-molecule force-ramp assay

To immobilize the DNA construct prepared above on the surface of anti-digoxigenin–antibody-coated polystyrene beads, 0.1 ng ($3.5 \times 10^{-17}$ mol) of DNA was mixed with 1 μl of beads (2.10 μm in diameter, 0.5% w/v) in 5 μl of a 10 mM phosphate buffer supplemented with 100 mM KCl either at pH 5.5 or 6.3. Since no significant difference has been found for ensemble experiments performed at pH 6.6 and 6.3, we carried out single-molecule experiments in pH 6.3 buffers, which allowed more formation of folded species. The mixture was diluted to 750 μl with the same buffer after incubation at room temperature for 30 min. The DNA-immobilized beads were then injected into a custom-made chamber and made ready for laser tweezers experiments.

Home-built dual-trap 1064 nm laser tweezers were used to carry out the force-ramp assay at 23°C (37,38). One laser focus (mobile trap) grabbed the anti-digoxigenin–antibody-coated bead that had already been linked with the DNA construct, while another focus trapped a streptavidin-coated bead (1.87 μm diameter). The mobile trap was controlled by a motorized mirror to move one bead close to or apart from another, which allows the tethering of the DNA construct between the two beads through affinity interactions. After the attachment of the DNA construct between the two beads, similar bead movements were carried out in the force-ramp assay with a loading rate of 5.5 pN/s (see text).

To evaluate transcription modulators on MPD, 10 μM IMC-76 or IMC-48 (32) was incubated with the DNA construct for 60 s. For transcription factor hnRNP LL, 280 nM was used to incubate with the DNA construct for 60 s. To evaluate the temporal effect, 120–180 s incubation was used.

## RESULTS AND DISCUSSION

### DNA population pattern is obtained by a single-molecule mechanical unfolding method

Inspired by the finding that i-motif structures in the promoter region of bcl-2 gene can regulate Bcl-2 expression through the recognition of hnRNP LL, a protein that belongs to a family with transcriptional control activities (33), we set out to evaluate different species involved in this regulation process at the single-molecular level by laser tweezers (37). With procedures described in the Materials and Methods section, we sandwiched the single-stranded C-rich strand with the sequence (Figure 1a), 5′-CAG CCC CGC TCC CGC CCC CTT CCT CCC GCG CCC GCC CCT-3′, between the two dsDNA handles. The free ends of the dsDNA handles were immobilized to two optically trapped beads via digoxigenin–anti-digoxigenin–antibody and biotin–streptavidin interactions, respectively (Figure 1b). By moving one of the trapped beads with a loading rate of 5.5 pN/s in a pH 5.5 phosphate buffer with 100-mM KCl at room temperature, mechanical unfolding experiments were carried out as tethered DNA was stretched below the plateau force (maximum 60 pN; Figure 1c). Unfolding event was indicated by a sudden decrease in tension accompanied by an increase in extension in the force–extension ($F$–$X$) curves. By reversing the direction of the bead movement, tension in the DNA construct can be reduced to 0 pN, allowing structures to refold. Subtraction of the stretching from the relaxing $F$–$X$ curve permits us to obtain the change-in-contour-length ($\Delta L$) through the change in extension ($\Delta x$) between these two curves at a particular force ($F$) using the worm-like-chain model (34,39):

$$\frac{\Delta x}{\Delta L} = 1 - \frac{1}{2}\left(\frac{k_B T}{F P}\right)^{1/2} + \frac{F}{S} \qquad (1)$$

where $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, $P$ is the persistent length (51.95 nm) and $S$ is the stretching modulus (1226 pN) for dsDNA handles (39). A

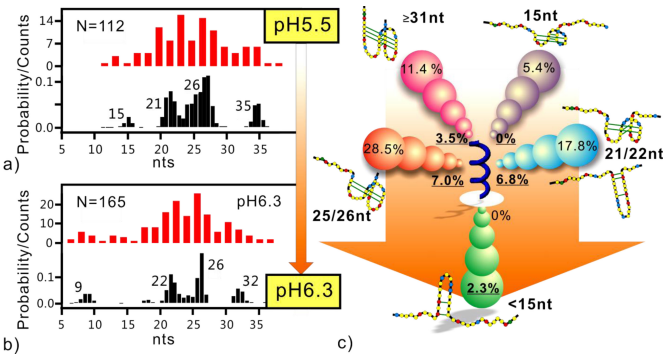


**Figure 2.** Effect of pH on the MPD of the C-rich structures. The $\Delta L$ histograms of C-rich populations at pH 5.5 (a) and pH 6.3 (b). The red and black bars depict $\Delta L$ populations measured at $F_{rupture}$ and deconvoluted from the PoDNano method, respectively. (c) MPD obtained from the differential population pattern between (a) and (b) (see text). The background arrow shows the direction of the population shift. The change of bubble size in each string of bubbles depicts the direction of change in a specific population between two different conditions. Underscored percentage values indicate those at pH 6.3 while the rest depict those at pH 5.5. The random coil structure is shown at the center. See Supplementary Figure S3 for details of other structures.

typical plot of $\Delta L$–$F$ curve is shown in Figure 1d. The $\Delta L$ obtained here reflects the size of a folded structure, and the rupture force ($F_{rupture}$) at which unfolding occurs depicts its mechanical stability. Since only one $\Delta L$ can be obtained at $F_{rupture}$ (Figure 1d), a histogram of $\Delta L_{(F@rupture)}$ only allows a rough estimation of the number of species in the DNA fragment (Figure 2a and b, red bars). To increase the accuracy, we expanded $\Delta L$ measurements to the force regions both below and above the $F_{rupture}$ (Figure 1d, left panel) and then constructed $\Delta L$ histograms for each region (Figure 1d, right panel). From the difference of the Gaussian centers between the two $\Delta L$ populations, the $\Delta L$ for a particular transition can be determined rather accurately from the stretching and relaxing $F$–$X$ curves (Figure 1d, right panel).

The $\Delta L$ thus obtained represents a particular species formed in the DNA sequence. To deconvolute different populations in the bcl-2 promoter DNA, we used kernel density treatment followed by resampling statistical analysis (34). First, we expanded each $\Delta L$ value with a Gaussian kernel, the width of which is the average of the standard errors in the $\Delta L$ measurements immediately before and after the unfolding event (Figure 1d, left panel). With a set of these Gaussian kernels, we then constructed a probability density distribution of populations from randomly selected Gaussian kernels in each of 1000 resamplings. Three predominant populations in each probability distribution were grouped to construct a histogram (Figure 2a and b, black histograms). Due to the significant expansion of $\Delta L$ measurements by experimentally determined Gaussian kernels, this PoDNano approach affords ~0.5 nm spatial resolution in the population identification (34). With the information of the number and the size of species, we were able to estimate the abundance of each population from the original $\Delta L$ histogram (see Supplementary Figure S2 for details) (38). The percentage of each species (Table 1), the number and the size of species constitute three major features of a population pattern (Figure 2a and b, black histograms).

### Flexible hairpin and i-motif species are present in the C-rich strand of the bcl-2 promoter

The population pattern in Figure 2a reveals DNA species that involve 15, 21, 26 and full-length (≥31) nucleotides (nts) in the C-rich fragment (see Supplementary Information for the conversion of $\Delta L$ to the number of nucleotides). These species have respective abundance of 5.4, 17.8, 28.5 and 11.4% in a pH 5.5, 10 mM phosphate buffer with 100 mM KCl (Table 1 and Figure 2c). Due to the hemiprotonated nature in the intercalating cytosine:cytosine pairs, i-motif structures are well known to be pH sensitive. To test whether these species can form i-motif structures, we repeated the experiment at pH 6.3, a condition similar to that employed by Hurley and coworkers [see the Materials and Methods section and (33)]. Application of the PoDNano analysis reveals a quite different population pattern at this pH (Figure 2b and Table 1). While there is a new 9-nt species with 2.3% in population, the percentage population for the 15-, 21-, 26-nt and full-length species is reduced to 0, 6.8, 7.0 and 3.5%, respectively. Such a pH dependency suggests that the species larger than 15 nts contain i-motif structures. The appearance of the 9-nt species at pH 6.3, but not at pH 5.5, suggests that it should not be an i-motif. In fact, no i-motif structures known so far with less than 15 nts can form in this sequence. Instead, the 9-nt species could be a flexible hairpin that is stable at the higher pH (See Supplementary Figure S3 for a possible structure).

To evaluate the effect of pH on the population distribution, we subtracted the percentage of each population at pH 5.5 (Figure 2a, bottom panel and Table 1) from that at pH 6.3 (Figure 2b, bottom panel and Table 1). As most $F$–$X$ curves contain only one rupture event, we argue that structures revealed by mechanical unfolding may not have intermediates, which would lead to more than one unfolding event. From topology perspective, it is not possible to interconvert between different i-motifs and flexible hairpins without unfolding to random coils first, although partially unfolded species could be generated without such interconversion (40,41). For simplicity, we constructed an MPD diagram centered on the unfolded DNA fragment (Figure 2c). This diagram gives a clear visualization on the change of population with pH. For example, it clearly shows that the 15-, 21-, 26-nt and the full-length species (≥31 nts) reduce their populations at higher pH (Figure 2c) and therefore have i-motif elements in their structures. We assigned the full-length population as a fully folded i-motif structure (see Supplementary Figure S3 for a possible structure), since no stable flexible hairpin of similar size can be found in mfold® calculations (Supplementary Figure S4). Assuming i-motifs in this C-rich DNA fragment behave similarly with pH increase, we reasoned that species with population reduction no smaller than that of the full-length i-motif (60% in reduction; Table 1) are likely to contain i-motif elements in their structures. Therefore, we assigned the 15-nt (5.4% → 0% reduction) and 26-nt (75% in reduction) species as i-motifs [see Supplementary Figure S3 for possible structures; notice it is possible for the 26-nt to assume partially folded conformation that contains pH-sensitive hemiprotonated C: CH stackings (40)]. Although it is possible that the 22-nt species could be i-motif only, our experiments in a pH 5.5 buffer

**Table 1.** Percentage population of individual species under different conditions

| pH | 5.5 | | | | 6.3 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Time | 60 s | | | | | | | | | | 120 s | | | 180 s | | |
| IMC-76 | – | – | + | + | – | + | – | – | + | – | – | – | – | – | – | – |
| IMC-48 | – | + | – | + | – | – | + | – | – | + | + | – | + | + | – | + |
| hnRNP | – | – | – | – | – | – | – | + | + | + | – | + | + | – | + | + |
| <15 nt | 0.0 | 0.0 | 0.0 | 0.0 | 2.3 | 8.7 | 0.0 | 2.4 | 6.1 | 1.9 | 0.0 | 10.9 | 7.6 | 0.0 | 17.0 | 5.7 |
| 15 nt | 5.4 | 10.3 | 2.0 | 8.6 | 0.0 | 0.0 | 11.2 | 0.0 | 7.7 | 7.0 | 11.8 | 0.0 | 9.2 | 10.3 | 0.0 | 5.3 |
| 22 nt | 17.8 | 0.0 | 26.9 | 17.9 | 6.8 | 2.4 | 0.0 | 4.7 | 0.0 | 0.0 | 7.6 | 5.9 | 3.1 | 7.5 | 8.5 | 5.3 |
| 26 nt | 28.5 | 29.1 | 18.8 | 24.3 | 7.0 | 5.2 | 11.4 | 15.5 | 10.4 | 40.5 | 4.2 | 4.9 | 4.6 | 4.6 | 3.5 | 6.5 |
| >32 nt | 11.4 | 28.3 | 2.9 | 6.4 | 3.5 | 1.6 | 4.3 | 4.7 | 1.2 | 7.0 | 1.4 | 4.2 | 3.1 | 3.4 | 1.0 | 2.0 |

indicated that IMC-76 increases the 22-nt population from 17.8% to 26.9% (Table 1). Since IMC-76 should decrease i-motif populations while increasing those of flexible hairpins (32), the 22-nt could be a mixture of flexible hairpins and i-motifs (see Supplementary Figure S3 for possible structures). Finally, due to the unique presence of the <15-nt species at pH 6.3, this structure has been assigned as flexible hairpins (see above and Supplementary Figure S3 for possible structures).

### MPD of C-rich DNA structures is modulated by small molecules and the transcription factor hnRNP LL

With the assignment of different populations, we proceeded to quantify the effects of modulators identified by Hurley and coworkers (32,33) during the transcription of the bcl-2 gene from the perspective of MPD. First, we conducted mechanical unfolding experiments in the presence of IMC-76 to obtain a population pattern in the DNA fragment at pH 6.3 by the PoDNano approach (Figure 3a). This population pattern is then compared to that without ligand (Figure 2b) to obtain the MPD controlled by the ligand IMC-76. As shown in Figure 3d, the populations of the 22-, 26-nt and the full-length species decrease whereas that of the flexible hairpin (<15-nt) increases. This result suggested that the IMC-76 ligand helps to stabilize the flexible hairpin and shifts the population equilibrium toward smaller species (see Figure 3 in (32) for binding assays). An opposite trend was observed when the IMC-48 was evaluated (Figure 3e). The populations of the 15-, 26-nt and the full-length species increase at the expense of both 22- and <15-nt species. As the former three species contain i-motif structural elements whereas the latter two flexible hairpins, it appears that IMC-48 stabilizes i-motifs over flexible hairpins. Similar results were observed in the presence of transcription factor hnRNP LL, which decreases the 22-nt species while increasing the populations of the 26-nt and the full-length species (Figure 3f).

To quantitatively compare the effects of different factors on the MPD of DNA species in the bcl-2 promoter fragment, we designed a simple pattern recognition algorithm based on pairwise analysis (42). To facilitate the comparison, first, we digitized the change in population of species $i$ ($C_i$). We assigned $C_i = 1$ for an increase in population; $C_i = -1$ for a decrease; and $C_i = 0$ for no change (Supplementary Tables S1 and S5). Pairwise comparison of factors 1 and 2 is carried out by the sum of the multiplication of $C_i$ values for all $n$ species ($\sum_{i=1}^{n} (C_{1,i} C_{2,i})$). Such a treatment gives a similarity score between $n$ (identical patterns) and $-n$ (totally opposite patterns). To compare factors that affect different number of species, we converted the similarity score to percent similarity ($s$) by the expression $S = \left(\frac{n+\sum_{i=1}^{n} (C_{1,i} C_{2,i})}{2n}\right) \times 100\%$. This calculation gives $s = 100\%$ between identical MPD (positive correlation), $s = 0\%$ for two totally opposite MPD (anti-correlation) and $s = 50\%$ for no correlation.

Using this algorithm, we quantified the similarity between different bcl-2 gene modulators (Table 2 and Supplementary Table S2). We found that the percent similarity between IMC-48 and hnRNP LL is 80%, demonstrating similar effects on the MPD between these two modulators. In comparison, the similarity percentage is 30% between IMC-76 and IMC-48 and 40% between IMC-76 and hnRNP LL, indicating that IMC-76 has an opposite effect on the MPD with respect to IMC-48 or hnRNP LL. These results agree very well with Hurley's finding that IMC-76 destabilizes i-motif structures, whereas IMC-48 and hnRNP LL stabilize these structures in biochemical assays (32,33).

To evaluate whether there is an additive effect between different modulators, we performed the same mechanical unfolding experiments in the IMC-48/hnRNP LL mixture (Figure 4) and in the IMC-76/hnRNP LL mixture (Supplementary Figure S5). Using the algorithm for pattern recognition described above, we compared the similarity percentage of each modulator, or their mixture, with respect to IMC-48. The results are shown in a semi-circle similarity plot in Figure 5 (see Supplementary Information for the construction of the similarity plot). As expected, the effects of IMC-48, hnRNP LL and their mixture are highly correlated, while IMC-76 shows an anti-correlated behavior. Interestingly, the mixture of two anti-correlated modulators (IMC-76 and hnRNP LL) shows a non-correlated behavior ($s = 55\%$) with respect to IMC-48. Likewise, two anti-correlated small molecules IMC-48 and IMC-76 show a non-correlated behavior ($s = 50\%$) when they mix together. These results suggest that there is an additive effect between different modulators.

Further evidence for an additive effect comes from the MPD analyses. Figure 4a showed that the net effect of hnRNP LL in the presence of IMC-48 is to further increase the population of the 26-nt and full-length i-motif species. Similarly, more i-motif structures (15-, 26-nt and full-length species) were formed due to the net effect of IMC-48 in the context of hnRNP LL (Figure 4c).

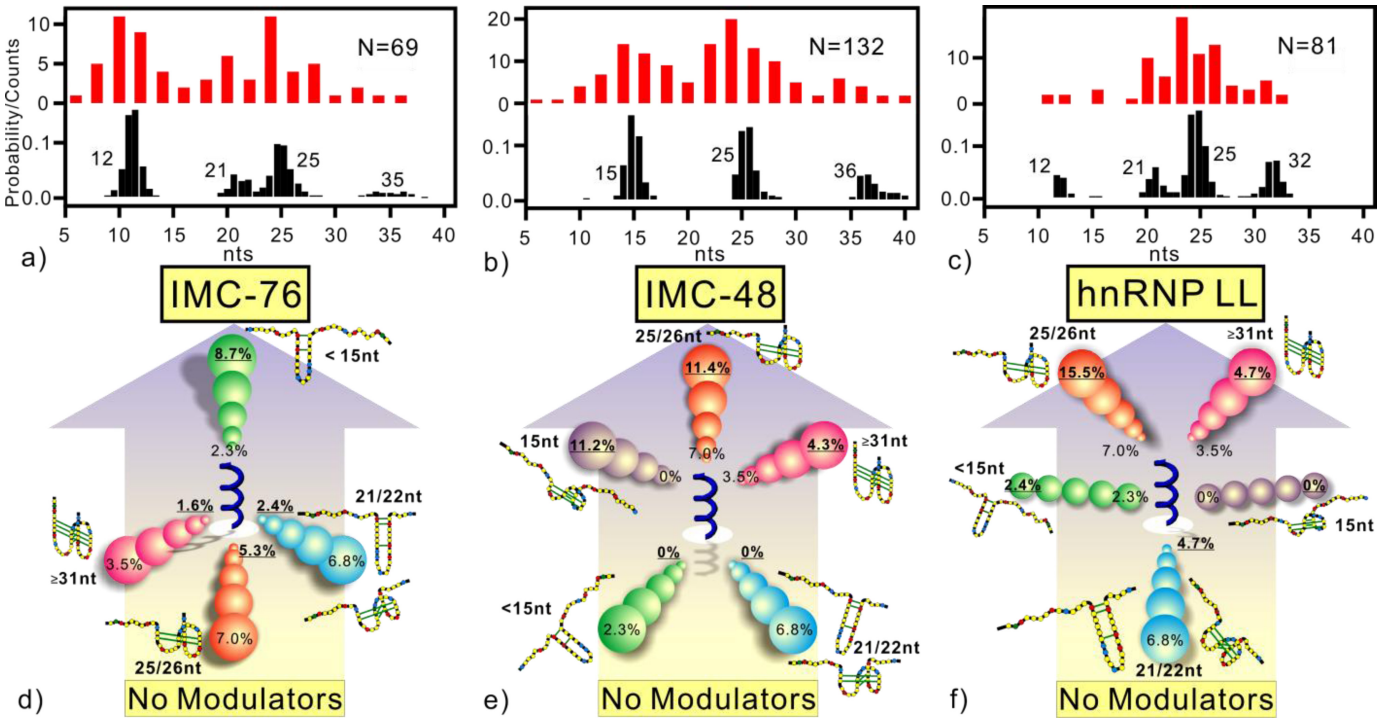


**Figure 3.** Effect of transcription modulators on the MPD of the C-rich structures at pH 6.3. Population patterns for IMC-76 (10 μM) (a), IMC-48 (10 μM) (b) and hnRNP LL (280 nM) (c). The red and black bars depict $\Delta L$ populations measured at $F_{rupture}$ and deconvoluted from the PoDNano method, respectively. MPD for IMC-76 (d), IMC-48 (e) and hnRNP LL (f). Each MPD is obtained after comparison of population patterns with [(a), (b) or (c)] and without (Figure 2b) a particular transcription modulator. The background arrow shows the direction of the population shift. The change of bubble size in each string of bubbles depicts the direction of change in a specific population between two different conditions. Underscored percentage values indicate those with modulators while the rest are those without ligands or proteins. The random coil structure is shown at the center. See Supplementary Figure S3 for details of other structures.

**Table 2.** Comparison of percent similarity in MPD between different factors with 60 s incubation

| | IMC-76 | IMC-48 | hnRNP LL |
|---|---|---|---|
| IMC-76 | 100% | | |
| IMC-48 | 30% | 100% | |
| hnRNP LL | 40% | 80% | 100% |

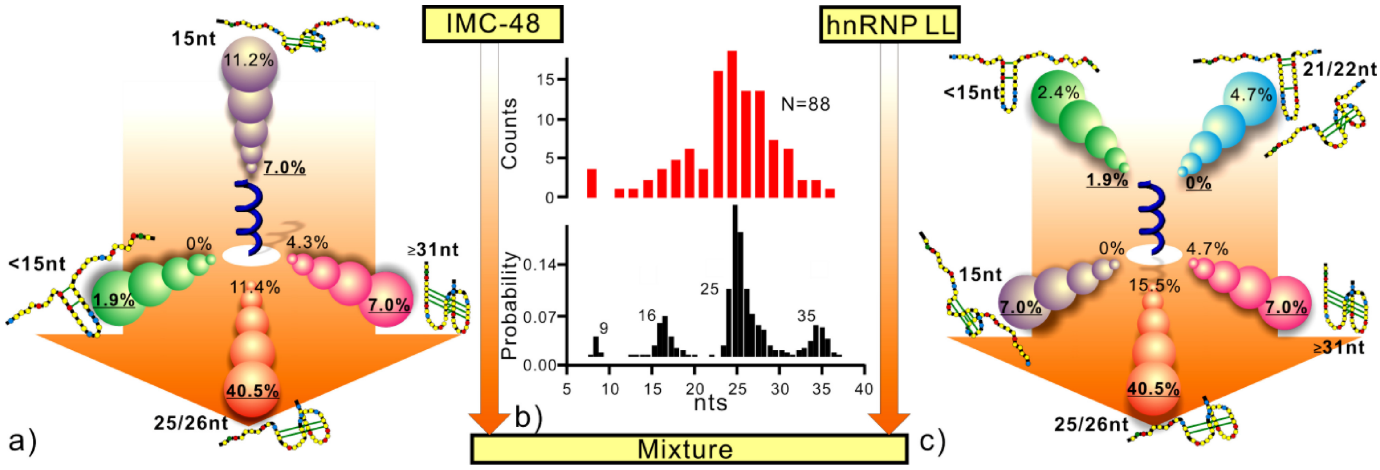


**Figure 4.** Combined effect of IMC-48 and hnRNP LL on the MPD of C-rich structures at pH 6.3. (a) Net effect of hnRNP LL (280 nM) in the presence of IMC-48 (10 μM). (b) Population patterns of the C-rich species in the presence of IMC-48 (10 μM) and hnRNP LL (280 nM) with 60 s incubation. The red and black bars depict $\Delta L$ populations measured at $F_{rupture}$ and those deconvoluted from the PodNano method, respectively. (c) Net effect of IMC-48 (10 μM) in the presence of hnRNP LL (280 nM). Each MPD is obtained after comparison of the population pattern with a particular modulator (Figure 3b or c) and that with the mixture of IMC-48 and hnRNP LL (b). Underscored percentage values indicate those with both modulators while the rest are those with only one modulator. The background arrow shows the direction of the population shift. The change of bubble size in each string of bubbles depicts the direction of a specific population shift. The random coil structure is shown at the center. See Supplementary Figure S3 for details of other structures.

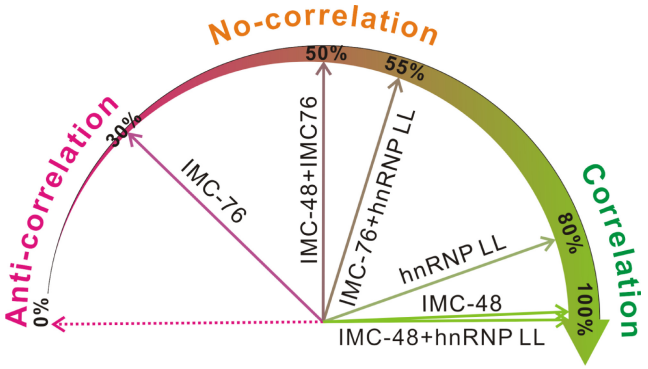


**Figure 5.** Semi-circle similarity plot for different modulators and their mixtures. Percentage similarity for a specific modulator i ($PS_i$, shown inside the gradient arrow) is obtained after comparison with IMC-48. The similarity between two modulators (i and j) can be calculated as $s\% = 100\%-(PS_i-PS_j)$, here $PS_i > PS_j$.

To quantify this additive effect, we evaluated the effect of the IMC-48 and hnRNP LL mixture with respect to IMC-48 or hnRNP LL. For each species in the population pattern, an additivity is confirmed when the effect of the mixture is no smaller than the combined effects of IMC-48 and hnRNP LL (see Supplementary Information and Supplementary Table S3-2). Out of the five species, four have shown this additivity, which is equivalent to 80% in the probability of additivity between IMC-48 and hnRNP LL. The probability of additivity rises to 100% for i-motif species (15-nt to full length). The high additivity suggests separate binding sites for IMC-48 and hnRNP LL. In comparison, the probability of additivity between IMC-76 and hnRNP LL is 40% for all species and 50% for i-motifs (Supplementary Figure S5 and Supplementary Table S3-1); for IMC-48 and IMC-76, it is 60% and 50%, respectively (Supplementary Table S3-3). It is possible that the binding sites may partially overlap for these two pairs of modulators. The different additive effects between the IMC-76/hnRNP LL pair (50%) and the IMC-48/hnRNP LL pair (100%) have been well reproduced qualitatively in the binding assay performed by Hurley and coworkers [see Figure 5 in (33)], which validates the new method described here. Additional validation of the method came from the population analysis of species in the human telomeric RNA (TERRA). It has been well established that TERRA GQ can be bound with ligand carboxypyridostatin (cPDS) or anti-GQ antibody BG4 (43,44). Using the same MPD approach, indeed, we found 100% similarity between the effect of cPDS and BG4 on the structures formed in the TERRA (Supplementary Figure S6).

By using the similarity comparison and additivity calculation in the new MPD method, we have quantified for the first time the effect of transcription modulators on the MPD of non-B DNA structures. With the *in vivo* presence and the biological activity of non-B DNA structures firmly established (43,45), we envision the method is instrumental to understand biological implications of these non-B DNA structures by quantitative evaluation of the population equilibrium affected by specific cellular factors.

### Temporal effect on the MPD of the C-rich DNA structures

It has been reported by Hurley and coworkers that initial binding of hnRNP LL to i-motif structures in the bcl-2 promoter fragment eventually leads to the destabilization of i-motifs with longer incubation time (33). To illustrate this process from a perspective of molecular population dynamics, we first mechanically unfolded DNA structures by laser tweezers. This was followed by incubation with different times (120–180 s versus 60 s) to obtain the temporal effect on the MPD in the presence of IMC-48, hnRNP LL or both.

As shown in Figure 6, with 180 s incubation, populations of species larger than 26 nts decrease whereas that of the 22-nt species increases. As the 26-nt and the full-length species contain i-motif elements whereas the 22-nt species is a mixture of i-motifs and flexible hairpins, this result suggested that IMC-48 and hnRNP LL destabilize i-motifs over flexible hairpins with 180 s incubation. Using the pattern recognition algorithm discussed above (see Supplementary Table S5 for the scores), we found that the similarity between IMC-48 and hnRNP LL is 70% (Table 3 and Supplementary Figure S7). In addition, the temporal effect of the IMC-48 and hnRNP LL mixture has 70% similarity with that of IMC-48 and 90% similarity with that of hnRNP LL (Table 3). As 50% similarity depicts non-correlation between two factors (see Figure 5), the similarity of 70% represents a moderately positive correlation between the IMC-48 and the hnRNP LL. Consistent with this, the additivity calculation (see above) showed that with 180 s incubation, these two factors have a reduced probability of additivity among DNA species (60% for all species and 50% for i-motifs; Supplementary Table S6) compared to short-term incubation. To probe the temporal effect more accurately, we also performed experiments with 120 s incubation. As shown in Table 1 and Supplementary Figure S8, the trend of the population dynamics is well maintained within the experimental error of the measurement.

It has been proposed by Hurley that hnRNP LL binds i-motif structures through the sequences CCCGC and CGCCC in the lateral loops. This is followed by the disassembly of the i-motif into an ssDNA bound with hnRNP LL, which activates the bcl-2 transcription (33). With sufficient template tension, the bound hnRNP LL is expected to be ejected from ssDNA (46). This event gives rise to a rupture transition in the $F$–$X$ curve with small change-in-contour-length ($\Delta L$), which is a result of releasing flexible DNA segments between the two binding sequences for the hnRNP LL [see Figure 8 in (33)]. The observation of small rupture events ($\leq$13-nt transitions; Figure 6b, c, e and f), therefore, supports these sequential events involved in the activation of bcl-2 transcription. It is rather surprising that with long-term incubation, IMC-48 also destabilizes i-motif, although with a weaker effect. Such a result, however, is in agreement with the observation that IMC-48 can activate bcl-2 transcription (32), presumably through destabilization of i-motif structures similar to the hnRNP LL (33). Based on the fact that small transitions ($\leq$13 nt; Figure 6a and d) were not observed in the presence of IMC-48, the detailed mechanisms for the i-motif destabilization are

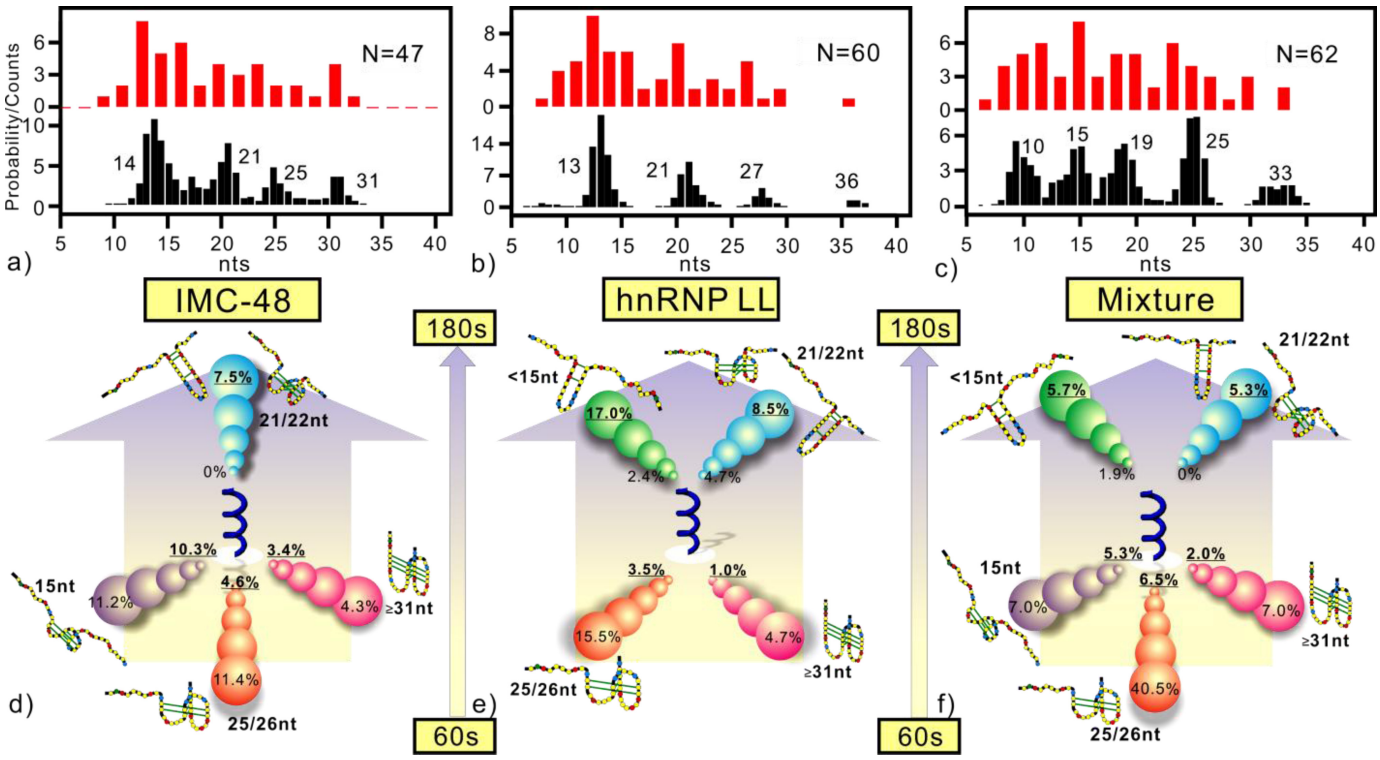


**Figure 6.** Temporal effect of IMC-48 and hnRNP LL on the MPD of the C-rich structures at pH 6.3. Population patterns for IMC-48 (10 μM) (a), hnRNP LL (280 nM) (b) and both (c). The red and black bars depict $\Delta L$ populations measured at $F_{rupture}$ and deconvoluted from the PoDNano method, respectively. MPD for IMC-48 (d), hnRNP LL (e) and both (f). Each MPD is obtained after comparison of the population patterns between 180 s [(a), (b) or (c)] and 60 s incubation (Figure 3b and c and Figure 4b, respectively). Underscored percentage values indicate those with 180 s incubation while the rest are those with 60 s incubation. The background arrow shows the direction of the population shift. The change of bubble size in each string of bubbles depicts the direction of change in a specific population between two different conditions. The random coil structure is shown at the center. See Supplementary Figure S3 for details of other structures.

**Table 3.** Comparison of percent similarity in MPD (180 versus 60 s incubation) between different factors

| | IMC-48 | hnRNP LL | IMC-48+hnRNP LL |
|---|---|---|---|
| IMC-48 | 100% | | |
| hnRNP LL | 70% | 100% | |
| IMC-48+hn RNP LL | 70% | 90% | 100% |

*Note.* The similarity is calculated from digitized change in population (see Supplementary Table S5).

likely to be different between IMC-48 and hnRNP LL during long-term incubation.

The potential effect of the i-motif and flexible DNA structures formed in the bcl-2 promoter on the regulation of bcl-2 transcription and the small-molecule-induced MPD of these non-B DNA species closely resemble the function of riboswitches (35). Riboswitch segments in messenger RNA (mRNA) are known to assume versatile conformations dependent on the presence of small molecules, which are often effector molecules of the protein encoded by the mRNA. Different RNA structures in a riboswitch reach a population equilibrium, which then regulates the production of the encoded protein. Therefore, the MPD under the control of small molecules is highly important for a riboswitch to modulate protein expression. The results presented here provide first evidence that, similar to a riboswitch for translational control, DNA structures, i-motifs in particular, may carry out transcriptional control via MPD modulated by small molecules.

## CONCLUSIONS

By identifying six different populations in a C-rich bcl-2 promoter fragment using mechanical unfolding strategy, we analyzed the population equilibrium of these species with a new method, MPD. With a simple algorithm for pattern recognition, we found IMC-48 and hnRNP LL share a similar effect on the stabilization of i-motifs over flexible DNA hairpins, whereas IMC-76 shows a reversed effect with 60 s incubation. Longer incubation (120–180 s) causes IMC-48 and hnRNP LL to destabilize i-motifs. These results agree strikingly well with those observed in biochemical experiments, validating this new single-molecule method. Our results also provide evidence, at the level of population equilibrium of non-B DNA species, that bcl-2 transcription can be modulated by i-motif populations under the control of small molecules and the transcription factor hnRNP LL. Together with the results from chemosensitization of cancer cells, promoter binding and mRNA production (32,33), the

evidence for the involvement of i-motif structures in the bcl-2 transcription becomes compelling. These findings start to shed light on the potential of non-B DNA structures as new transcriptional regulation elements through MPD controlled by small molecules and transcription factors. As a natural extension, we anticipate the MPD approach developed here can be used as a new tool to delineate complex and dynamic population equilibrium among nucleic acid and protein structures.

## FUNDING

National Science Foundation [CHE-1026532 to H.M.]; National Institutes of Health [GM085585 to L.H.H.]; National Foundation for Cancer Research [VONHOFF0601 to L.H.H.]. Funding for open access charge: University of Arizona Foundation; National Science Foundation [CHE-1026532].

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [47–49].

## ACKNOWLEDGEMENTS

H.M. is thankful for support from the National Science Foundation. L.H.H. acknowledges support from the National Institutes of Health and the National Foundation for Cancer Research.

## REFERENCES

1. Dill,K.A., Ozkan,S.B., Shell,M.S. and Weikl,T.R. (2008) The protein folding problem. *Annu. Rev. Biophys.*, **37**, 289–316.
2. Gilbert,D.E. and Feigon,J. (1999) Multistranded DNA structures. *Curr. Opin. Struct. Biol.*, **9**, 305–314.
3. Wells,R.D. and Harvey,S.C. (1988) *Unusual DNA structures*. Springer, New York, NY.
4. Dethoff,E.A., Chugh,J., Mustoe,A.M. and Al-Hashimi,H.M. (2012) Functional complexity and regulation through RNA dynamics. *Nature*, **482**, 322–330.
5. Shank,E.A., Cecconi,C., Dill,J.W., Marqusee,S. and Bustamante,C. (2010) The folding cooperativity of a protein is controlled by its chain topology. *Nature*, **465**, 637–640.
6. Onoa,G.B., Liphardt,J.T., Smith,S.B., Tinoco,I. and Bustamante,C.J. (2001) RNA tertiary and secondary unfolding monitored by mechanical stretching of single molecules. *Biophys. J.*, **80**, 155a
7. Greenleaf,W.J., Frieda,K.L., Foster,D.A., Woodside,M.T. and Block,S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, **319**, 630–633.
8. Li,P.T., Vieregg,J. and Tinoco,I. Jr (2008) How RNA unfolds and refolds. *Annu. Rev. Biochem.*, **77**, 77–100.
9. Watson,J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
10. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA.*, **48**, 2013–2018.
11. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
12. Davis,J.T. (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.*, **43**, 668–698.
13. Yu,Z., Gaerig,V., Cui,Y., Kang,H., Gokhale,V., Zhao,Y., Hurley,L.H. and Mao,H. (2012) The tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J. Am. Chem. Soc.*, **134**, 5157–5164.
14. Koirala,D., Mashimo,T., Sannohe,Y., Yu,Z., Mao,H. and Sugiyama,H. (2012) Intramolecular folding in three tandem guanine repeats of human telomeric DNA. *Chem. Commun.*, **48**, 2006–2008.
15. Alberti,P. and Mergny,J.L. (2004) DNA structural changes as the basis for a nanomolecular device. *Cell. Mol. Biol.*, **50**, 241–253.
16. Lee,J.Y., Okumus,B., Kim,D.S. and Ha,T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA.*, **102**, 18938–18943.
17. Choi,J., Kim,S., Tachikawa,T., Fujitsuka,M. and Majima,T. (2011) pH-induced intramolecular folding dynamics of i-motif DNA. *J. Am. Chem. Soc.*, **133**, 16146–16153.
18. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
19. Gehring,K., Leroy,J.L. and Guéron,M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561–564.
20. Snoussi,K., Nonin-Lecomte,S. and Leroy,J.-L. (2001) The RNA i-motif. *J. Mol. Biol.*, **309**, 139–153.
21. Lacroix,L., Mergny,J.-L., Leroy,J.-L. and Hélène,C. (1996) Inability of RNA to form the i-motif: implications for triplex formation. *Biochemistry*, **35**, 8715–8722.
22. Kouzine,F., Gupta,A., Baranello,L., Wojtowicz,D., Ben-Aissa,K., Liu,J., Przytycka,T.M. and Levens,D. (2013) Transcription-dependent dynamic supercoiling is a short-range genomic force. Nat Struct Mol Biol., **20**, 396–403.
23. Dhakal,S., Yu,Z., Konik,R., Cui,Y., Koirala,D. and Mao,H. (2012) G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys. J.*, **102**, 2575–2584.
24. Zheng,K., Xiao,S., Liu,J., Zhang,J., Hao,Y. and Tan,Z. (2013) Co-transcriptional formation of DNA: RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.*, **41**, 5533–5541.
25. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
26. Bucek,P., Jaumot,J., Avino,A., Eritja,R. and Gargallo,R. (2009) pH-modulated Watson–Crick duplex-qaudruplex equilibria of guanine-rich and cytosine-rich DNA sequences 140 base pairs upstream of the c-kit transcription intiation site. *Chem. Eur. J.*, **15**, 12663–12671.
27. Phan,A.T., Gueron,M. and Leroy,J.L. (2000) The solution structure and internal motions of a fragment of the cytidine-rich strand of the human telomere. *J. Mol. Biol.*, **299**, 123–144.
28. Cai,L., Chen,L., Raghavan,S., Ratliff,R., Myozis,R. and Rich,A. (1998) Intercalated cytosine motif and novel adenine clusters in the crystal structure of the *Tetrahymena* telomere. *Nucleic Acids Res.*, **26**, 4696–4705.
29. Achillefs,N. and Kapanidis,T.S. (2009) Biology, one molecule at a time. *Trends Biochem. Sci.*, **34**, 234–243.
30. Turchin,P. (2003) *Complex population dynamics: a theoretical/empirical synthesis*. Princeton University Press, Princeton, NJ.
31. William,A. and Nelson,E.M.F.J.W. (2005) Population and genotype dynamics in three externally driven environments. *Nature*, **433**, 413–417.
32. Kendrick,S., Kang,H.-J., Akiyama,Y., Agrawal,P., Gokhale,V., Yang,D., Hecht,S.M. and Hurley,L.H. (2014) The dynamic character of the BCL2 promoter i-Motif provides a mechanism for modulation of gene expression by compounds that bindselectively to the alternative DNA hairpin structure. J. Am. Chem. Soc., doi: 10.1021/ja410934b.
33. Kang,H.-J., Kendrick,S., Hecht,S.M. and Hurley,L.H. (2014) The transcriptional complex between the BCL2 i-motif and hnRNP LL Is a molecular switch to control gene expression. J. Am. Chem. Soc., doi: 10.1021/ja4109352.
34. Yu,Z. and Mao,H. (2013) Non-B DNA structures show diverse conformations and complex transition kinetics comparable to RNA or proteins—a perspective from mechanical unfolding and refolding experiments. *Chem. Rec.*, **13**, 102–116.
35. Winkler,W.C. and Breaker,R.R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.*, **59**, 487–517.
36. Dhakal,S., Schonhoft,J.D., Koirala,D., Yu,Z., Basu,S. and Mao,H. (2010) Coexistence of an ILPR i-motif and a partially folded

structure with comparable mechanical stability revealed at the single-molecule level. *J. Am. Chem. Soc.*, **132**, 8991–8997.
37. Mao,H. and Luchette,P. (2008) An integrated laser tweezers instrument for microanalysis of individual protein aggregates. *Sensors Actuators B*, **129**, 764–771.
38. Yu,Z., Schonhoft,J.D., Dhakal,S., Bajracharya,R., Hegde,R., Basu,S. and Mao,H. (2009) ILPR G-quadruplexes formed in seconds demonstrate high mechanical stabilities. *J. Am. Chem. Soc.*, **131**, 1876–1882.
39. Baumann,C.G., Smith,S.B., Bloomfield,V.A. and Bustamante,C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci. USA.*, **94**, 6185–6190.
40. Dhakal,S., Lafontaine,J.L., Yu,Z., Koirala,D. and Mao,H. (2012) Intramolecular folding in human ILPR fragment with three C-rich repeats. *PLoS ONE*, **7**, e39271.
41. Koirala,D., Ghimire,C., Bohrer,C., Sannohe,Y., Sugiyama,H. and Mao,H. (2013) Long-loop G-quadruplexes are misfolded population minorities with fast transition kinetics in human telomeric sequences. *J. Am. Chem. Soc.*, **135**, 2235–2241.
42. Bradley,R.A. (1952) Rank analysis of incomplete block designs, I. The method of paired comparisons. *Biometrika*, **39**, 324–345.
43. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
44. Yangyuoru,P.M., Zhang,A.Y.Q., Shi,Z., Koirala,D., Balasubramanian,S. and Mao,H. (2013) Mechanochemical properties of individual human telomeric RNA (TERRA) G-quadruplexes. *Chem. Bio. Chem.*, **14**, 1931–1935.
45. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. *Proc. Natl. Acad. Sci. USA.*, **99**, 11593–11598.
46. Koirala,D., Yu,Z., Dhakal,S. and Mao,H. (2011) Detection of single nucleotide polymorphism using tension-dependent stochastic behavior of a single-molecule template. *J. Am. Chem. Soc.*, **133**, 9988–9991.