

Flipons and Condensates Enhances Evolution

Subjects: Cell Biology

Contributor: Alan Herbert

A number of insights derive from viewing flipons as scaffolds for condensates. Flipons provide a controlled way to initiate condensate formation, one subject to natural selection. The alternative conformation localizes required factors needed to regulate transcription, RNA processing, and epigenetic modification, while excluding nucleosomes and other B-DNA- and A-RNA-specific proteins that produce competing outcomes.

Keywords: Z-DNA ; Z-RNA ; flipons ; simple repeats ; condensates ; G4 ; evolution ; MYC ; non-coding RNA ; DNA conformation ; complexity ; phase separation ; enhancersome ; nucleosome

1. Starting Simple

The theme of this article is presented in **Figure 1** . Here, a nucleic acid structural motif is recognized by a structure-specific protein interaction. The nucleic acid acts as a scaffold and the protein as an anchor for cellular machines. Patches on the anchor protein provide a docking site for other proteins (**Figure 1** , left panel). The outcome depends on the functions of the assembled proteins, any one of which may have peptide patches for the attachment of additional proteins. In the simplest case, both the nucleic acid structures and the peptide patches are encoded by simple sequence repeats (SSRs) (**Figure 1** , right panel). The approach to build complex biological machines is adaptative for ever-changing environments.

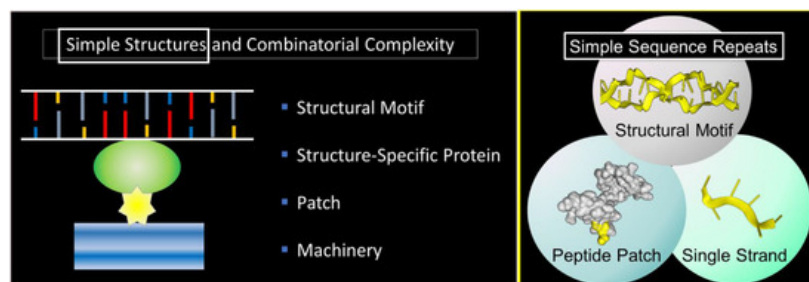


Figure 1. Simple nucleic acid structures and structure-specific proteins with low-complexity patches enable many combinations that enhance rapid evolution based on proven functional domains. The left panel displays flipons that form alternative nucleic acid structures under physiological conditions and are recognized by structure-specific proteins nucleate condensates that perform specific functions. The condensates assembled depend on the flipon conformation and protect the alternative structures by walling them off to maintain genome integrity. The right panel displays how simple sequence repeats impact condensate formation through structural motifs, peptide patches (highlighted in yellow), and sequence-specific interactions with single-stranded binding proteins that can also alter the condensates formed.

The scheme exploits the properties of SSRs that enable them to encode alternative nucleic acid structures, called flipons, to specify simple peptide patches that fold in different ways and to engage sequence-specific, single-stranded binding proteins that play multiple roles in condensate biology. We will focus in this review on the biology of flipons and peptide patches: how they can initiate and how they can promote condensate formation to optimize responses. We will discuss the role of SSRs in disease and their evolutionary impact. To introduce the concepts, we will start with a description of condensates and their role in classical genetics and then move on to SSRs and flipon genetics.

2. Z-Flipons

Z-RNA regulates both type I interferon responses by adenosine deaminase RNA specific (ADAR1, encoded by ADAR) ^[1] and the programmed cell death necroptosis pathway by Z-DNA binding protein 1 (ZBP1) ^[2]. In both cases, the left-handed helix is recognized by a structure specific, winged helix-turn-helix α protein domain without any base-specific contacts involved ^{[3][4]}. With ADAR1, α and another structure-specific, double-stranded RNA (dsRNA)-specific binding motif target

the deaminase domain to modify adenosines in dsRNA, forming inosine that is treated by the downstream processes as the equivalent of guanosine. This process produces non-synonymous codon changes in a limited number of human substrates [5].

It is likely that other Z-binding proteins exist, some of which may also bind B-DNA. Simple peptide repeats, such as peptides with alternating lysine residues, have high specificity for Z-DNA when tested with a methylated polymer [6]. These repeats that are likely IDRs are present in a number of interesting proteins. One example is the DNA methyl transferase I (encoded by DNMT1) that contains a lysine–glycine repeat, 10 amino acids long. The repeat is unstructured in the native DNMT1 crystal (Protein Data Bank (PDB): 5GUV) but interacts with ubiquitin-specific peptidase 7 (USP7) in the co-crystal (PDB: 4YOC). USP7 negatively modulates the activity of DNMT1 [7]. It is also possible that the repeat serves to localize DNMT1 to regions that form Z-DNA to methylate them. Another example is the histone H2A.Z variant 2 (encoded by H2AZ2; NCBI: NP_036544.1) that contains three alternating lysine–alanine repeats and destabilizes nucleosomes.

Less clear cut are the roles of poly-lysine repeats in the large ATP-dependent SWI/SNF chromatin-remodeling proteins BRG1 (encoded by SMARCA4) and BRM (encoded by SMARCA2) (**Figure 2**). The SWI/SNF complex is able to eject nucleosomes from DNA. The stress of DNA negative supercoiling, previously relieved by winding the DNA around the histone octamer, now becomes available to stabilize Z-DNA (**Figure 2**). Here, the left-handed solenoidal winding around the nucleosome is converted into a left-handed DNA twist. Previous reports have revealed that Z-DNA formation induced by BRG1 is associated with activation of the colony-stimulating factor 1 (CSF1) gene [8] and also the transcription of the heme oxygenase 1 gene (encoded by HMOX1) induced by nuclear factor, erythroid 2 like 2 (NRF2, encoded by NFE2L2) in response to oxidative stress [9]. In the CSF1 promoter, Z-DNA formation occurs while DNA is bound by the nucleosome [10], consistent with a model where local induction of Z-formation in a Z-prone sequence by BRG1 initiates nucleosome eviction to create a region of open chromatin. As the flip from B-DNA to Z-DNA is cooperative, the propagation of Z-formation into adjacent segments will dislodge the entire octamer, releasing negative supercoiling that further promotes Z-formation in the region (**Figure 2**) and the binding of structure-specific proteins to that locus. The energy stored as Z-DNA is then available to promote the assembly of new protein complexes on the DNA. The flip to Z-DNA controls both the location and the timing of subsequent events, producing a switch from one genetic program to another. The flip back to B-DNA relieves topological stresses as the new condensate forms.

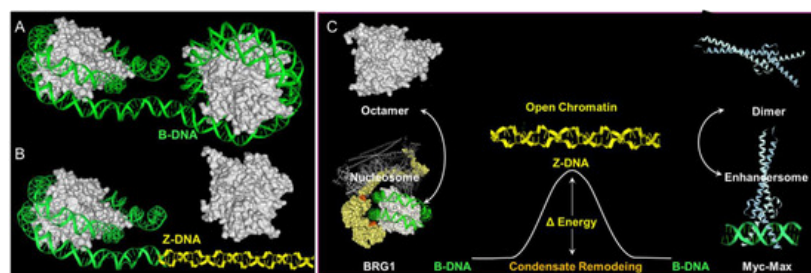


Figure 2 Flipons as genetic switches. The formation of alternative structures by flipons enables condensate remodeling. Here, the formation of Z-DNA is associated with nucleosome ejection and the formation of an enhancersome. The process of ejection is likely initiated by histone-remodeling complexes, such as SWI/SNF. The removal of the histone core then releases negative supercoiling that further promotes Z-DNA formation. (A) DNA bound by two nucleosomes. (B) The ejection of one nucleosome converts the left-handed writhe associated with winding of DNA around an octamer to the negative supercoiling that stabilizes left-handed Z-DNA. (C) In this model, the initial flip from B-DNA to Z-DNA involves lysine patches in the BRG1 protein. The segment 519–755 containing the lysine patch is not defined in the structure, while a KG patch starts at residue 1027. (Purple residues on the nucleosomes show residues 516-518 and 1020-1040 residues from PDB: 6LTJ). The Z-DNA conformation then propagates in a cooperative manner, using the energy released (ΔE) by nucleosome ejection to stabilize longer segments of DNA in the Z-DNA conformation. The Z-DNA formed then nucleates enhancersome formation and powers condensate assembly with the free energy released when DNA flips back to the B-DNA conformation. The SWI/SNF themselves are localized by transcription factors, such as NRF2 that form condensates at other locations in the enhancer region [9]. Here, the binding of the Myc-Max bZIP domain dimer is displayed.

3. G-Flipons

G-flipons can fold in a number of different ways, with strands running either parallel or anti-parallel with loops, providing connections that vary in length and position. The G4 quartet is recognized in a number of different ways [11]. Crystal structures of DDX36 and RAP1 proteins bound to G4 DNA reveal that the G4 caps are contacted by hydrophobic residues present in an α -helix, with engagement of the phosphate backbone by basic residues [12][13]. RAP1 also binds B-DNA through the same helix that it uses to bind G4, but through a different face [13]. A number of other proteins initially

characterized as single-stranded binders have subsequently been shown to bind G4 with high affinity. While their single-strand specificity was evident because they contain the RNA recognition motif (RRM), the role of peptide patches that recognizes G4 was not appreciated initially fell within IDR, lacking structure [14]. In these cases, the entropic cost of binding a disordered, single-stranded RNA is reduced by the RRM structure, while the entropic cost of docking to a peptide patch is lessened by the prepositioned backbone of the G4 motif [15]. Recognition of G-flipons through arginine–glycine motifs within IDR is common [16], with hydrophobic residues such as tyrosine and phenylalanine [14][17] showing different preferences for G4 RNA and G4 DNA [18]. Other modes of binding to G4 structures also exist. Fragile mental retardation protein recognizes the junction between B-DNA and the G4 structure, combining backbone contacts with base-specific ones [19].

The potential biological roles for G-flipons are various. Some are evidenced by both genetic and biochemical studies. There are mendelian diseases caused by defects in DNA repair and replication. The helicase variants involved show altered binding to G4 in vitro [20]. The 425 G4 interacting proteins recently identified using probes containing various constrained G4 structures are enriched for spliceosomes, RNA transport, RNA degradation, mRNA surveillance, DNA replication, and homologous recombination pathways. One G4-binding protein complex, the negative elongation factor (NELF), regulates gene expression in eukaryotic cells by promoting RNA polymerase pausing [21]. Other cell-based studies demonstrate the presence of nuclear G4 structures [22][23]. Collectively, the above studies suggest an important role for G-flipons in localizing cellular machines to regions where outcomes modify both normal function and disease risk [24].

4. When Flipons and Codons Clash

Many simple repeats that undergo expansion produce autosomal dominant disease [25], providing insight into their biology. The adverse outcomes reflect the abilities of repeats to both form alternative structures and encode longer peptide patches that seed aggregates. For example, the hexameric FTD/ALS repeat sequence GGGGCC can be transcribed from both strands to form a G4 quartet from one [26] and an I-motif (**Figure 3**) from the other [27]. The RNA transcribed from both strands is translated without the need for a traditional AUG start codon, a process called repeat-associated non-AUG (RAN) translation [28]. The RNA produced encodes six different dipeptide protein products, depending on the reading frame, with the two containing arginine being the most toxic. Disruption of many fundamental processes by the alternative nucleic conformations, by the peptide repeats, and by loss of function of the protein have been proposed as causes of disease [29]. Here, the alternative flipon conformations locked in by these diseases nucleate condensates that are only formed transiently within normal cells. The condensates persist. The proteins involved may be critical to switching one cellular response to another or may be involved in a separate pathway that is disrupted by their sequestration. The outcomes vary with the functions of the flipons involved.

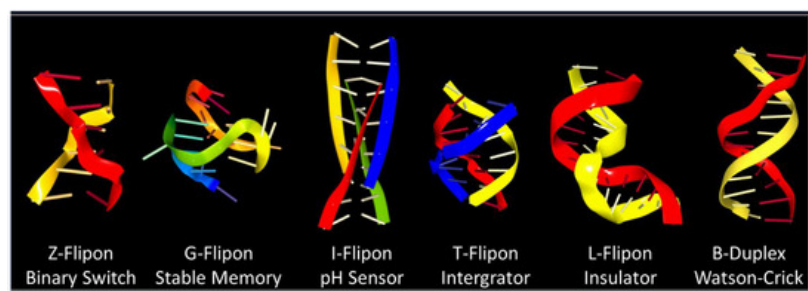


Figure 2. Different flipon conformations can enable different outcomes in a cell. Simple repeat sequences adopt alternative conformations, including left -handed Z-DNA and Z-RNA, G4 quadruplexes, I-motifs, and triplexes, like H-DNA formed locally by fold-back of DNA onto itself. The machines assembled on these higher-energy alternative structures differ from those made with lower-energy A-RNA, B-DNA, or on single-stranded nucleic acids. The change in flipon conformation switches outcomes. Currently, the best biologically characterized alternative conformations are Z- and G-flipons [30][31].

The question arises whether flipon sequences also encode peptide repeats that bind to the flipon that encodes them. The CGG repeat expansion that causes fragile X-related tremor/ataxia syndrome (FXTAS) does generate a polyglycine peptide (PPG) that appears to stabilize a G4 quartet formed by the RNA transcribed [32], with a PPG longer than nine residues by itself being insoluble [33]. In this case, the simple repeat codes for an alternative RNA conformation and a single amino acid repeat peptide that also forms higher-order protein structures. The interaction of the peptide with the RNA leads to disease. In cells, both PPG and G4 partition together into granules that stain with a G4-specific antibody. The granules also accumulate many other proteins, potentially interfering with the formation of alternative complexes that are essential for normal cell function. Here, codons and flipons clash. The two different schema for the encoding of

genetic information, one that specifies the amino acid sequence and the other that influences nuclear acid conformation, directly target each other, producing a negative outcome. RNA from a simple repeat expansion causes disease in FXTAS by undergoing RAN translation to produce peptides that locks the flippin sequence encoding the peptide into an alternative conformation, preventing the production of more peptide, which leads to the accumulation of more RNA that then is translated into new peptide. The self-referential and self-sustaining nature of this system produces a futile cycle, leading to system failure.

References

1. Herbert, A. To “Z” or not to “Z”: Z-RNA, self-recognition, and the MDA5 helicase. *PLoS Genet.* 2021, 17, e1009513.
2. Zhang, T.; Yin, C.; Boyd, D.F.; Quarato, G.; Ingram, J.P.; Shubina, M.; Ragan, K.B.; Ishizuka, T.; Crawford, J.C.; Tummers, B.; et al. Influenza Virus Z-RNAs Induce ZBP1-Mediated Necroptosis. *Cell* 2020, 180, 1115–1129.e13.
3. Schwartz, T.; Rould, M.A.; Lowenhaupt, K.; Herbert, A.; Rich, A. Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* 1999, 284, 1841–1845.
4. Placido, D.; Brown, B.A., 2nd; Lowenhaupt, K.; Rich, A.; Athanasiadis, A. A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* 2007, 15, 395–404.
5. Kim, J.I.; Nakahama, T.; Yamasaki, R.; Costa Cruz, P.H.; Vongpipatana, T.; Inoue, M.; Kanou, N.; Xing, Y.; Todo, H.; Shibuya, T.; et al. RNA editing at a limited number of sites is sufficient to prevent MDA5 activation in the mouse brain. *PLoS Genet.* 2021, 17, e1009516.
6. Takeuchi, H.; Hanamura, N.; Hayasaka, H.; Harada, I. B-Z transition of poly(dG-m5dC) induced by binding of Lys-containing peptides. *FEBS Lett.* 1991, 279, 253–255.
7. Li, J.; Wang, R.; Jin, J.; Han, M.; Chen, Z.; Gao, Y.; Hu, X.; Zhu, H.; Gao, H.; Lu, K.; et al. USP7 negatively controls global DNA methylation by attenuating ubiquitinated histone-dependent DNMT1 recruitment. *Cell Discov.* 2020, 6, 58.
8. Liu, H.; Mulholland, N.; Fu, H.; Zhao, K. Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling. *Mol. Cell. Biol.* 2006, 26, 2550–2559.
9. Zhang, J.; Ohta, T.; Maruyama, A.; Hosoya, T.; Nishikawa, K.; Maher, J.M.; Shibahara, S.; Itoh, K.; Yamamoto, M. BRG1 interacts with Nrf2 to selectively mediate HO-1 induction in response to oxidative stress. *Mol. Cell. Biol.* 2006, 26, 7942–7952.
10. Mulholland, N.; Xu, Y.; Sugiyama, H.; Zhao, K. SWI/SNF-mediated chromatin remodeling induces Z-DNA formation on a nucleosome. *Cell Biosci.* 2012, 2, 3.
11. McRae, E.K.S.; Booy, E.P.; Padilla-Meier, G.P.; McKenna, S.A. On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. *J. Nucleic Acids* 2017, 2017, 9675348.
12. Chen, M.C.; Tippiana, R.; Demeshkina, N.A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferre-D’Amare, A.R. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* 2018, 558, 465–469.
13. Traczyk, A.; Liew, C.W.; Gill, D.J.; Rhodes, D. Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic Acids Res.* 2020, 48, 4562–4571.
14. Masuzawa, T.; Oyoshi, T. Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Omega* 2020, 5, 5202–5208.
15. Gallo, A.; Lo Sterzo, C.; Mori, M.; Di Matteo, A.; Bertini, I.; Banci, L.; Brunori, M.; Federici, L. Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter. *J. Biol. Chem.* 2012, 287, 26539–26548.
16. Chong, P.A.; Vernon, R.M.; Forman-Kay, J.D. RGG/RG Motif Regions in RNA Binding and Phase Separation. *J. Mol. Biol.* 2018, 430, 4650–4665.
17. Zanotti, K.J.; Lackey, P.E.; Evans, G.L.; Mihailescu, M.-R. Thermodynamics of the Fragile X Mental Retardation Protein RGG Box Interactions with G Quartet Forming RNA. *Biochemistry* 2006, 45, 8319–8330.
18. Kondo, K.; Mashima, T.; Oyoshi, T.; Yagi, R.; Kurokawa, R.; Kobayashi, N.; Nagata, T.; Katahira, M. Plastic roles of phenylalanine and tyrosine residues of TLS/FUS in complex formation with the G-quadruplexes of telomeric DNA and TERRA. *Sci. Rep.* 2018, 8, 2864.
19. Vasilyev, N.; Polonskaia, A.; Darnell, J.C.; Darnell, R.B.; Patel, D.J.; Serganov, A. Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. USA* 2015, 112, E5391–E5400.

20. Lerner, L.K.; Sale, J.E. Replication of G Quadruplex DNA. *Genes* 2019, 10, 95.
21. Pipier, A.; Devaux, A.; Lavergne, T.; Adrait, A.; Couté, Y.; Britton, S.; Calsou, P.; Riou, J.F.; Defrancq, E.; Gomez, D. Constrained G4 structures unveil topology specificity of known and new G4 binding proteins. *bioRxiv* 2021.
22. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 2020, 21, 459–474.
23. Zheng, K.W.; Zhang, J.Y.; He, Y.D.; Gong, J.Y.; Wen, C.J.; Chen, J.N.; Hao, Y.H.; Zhao, Y.; Tan, Z. Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res.* 2020, 48, 11706–11720.
24. Wang, E.; Thombre, R.; Shah, Y.; Latanich, R.; Wang, J. G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Res.* 2021, 49, 4816–4830.
25. Khristich, A.N.; Mirkin, S.M. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* 2020, 295, 4134–4170.
26. Conlon, E.G.; Lu, L.; Sharma, A.; Yamazaki, T.; Tang, T.; Shneider, N.A.; Manley, J.L. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *eLife* 2016, 5, e17820.
27. Kovanda, A.; Zalar, M.; Sket, P.; Plavec, J.; Rogelj, B. Anti-sense DNA d(GGCCCC)n expansions in C9ORF72 form i-motifs and protonated hairpins. *Sci. Rep.* 2015, 5, 17944.
28. Freibaum, B.D.; Lu, Y.; Lopez-Gonzalez, R.; Kim, N.C.; Almeida, S.; Lee, K.H.; Badders, N.; Valentine, M.; Miller, B.L.; Wong, P.C.; et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* 2015, 525, 129–133.
29. Babić Leko, M.; Župunski, V.; Kirincich, J.; Smilović, D.; Hortobágyi, T.; Hof, P.R.; Šimić, G. Molecular Mechanisms of Neurodegeneration Related to C9orf72 Hexanucleotide Repeat Expansion. *Behav. Neurol.* 2019, 2019, 1–18.
30. Herbert, A. ALU non-B-DNA conformations, flipons, binary codes and evolution. *R. Soc. Open Sci.* 2020, 7, 200222.
31. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* 2020, 2, 123–136.
32. Asamitsu, S.; Yabuki, Y.; Ikenoshita, S.; Kawakubo, K.; Kawasaki, M.; Usuki, S.; Nakayama, Y.; Adachi, K.; Kugoh, H.; Ishii, K.; et al. CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci. Adv.* 2021, 7, eabd9440.
33. Ohnishi, S.; Kamikubo, H.; Onitsuka, M.; Kataoka, M.; Shortle, D. Conformational Preference of Polyglycine in Solution to Elongated Structure. *J. Am. Chem. Soc.* 2006, 128, 16338–16344.

Retrieved from <https://encyclopedia.pub/entry/history/show/33731>