



Review



Cite this article: Herbert A. 2025 Flipons enable genomes to learn by intermediating the exchange of energy for information. *J. R. Soc. Interface* **22**: 20250049.

<https://doi.org/10.1098/rsif.2025.0049>

Received: 14 January 2025

Accepted: 26 February 2025

Subject Category:

Life Sciences—Chemistry interface

Subject Areas:

bioinformatics, biomathematics, bioenergetics

Keywords:

DNA topology, flipons, entropy, learning, genome, evolution

Author for correspondence:

Alan Herbert

e-mails: alan.herbert@insideoutbio.com;

dr.alan.herbert@gmail.com

Flipons enable genomes to learn by intermediating the exchange of energy for information

Alan Herbert

Discovery, InsideOutBio Inc, Charlestown, MA, USA

AH, 0000-0002-0093-1572

Recent findings have confirmed the long-held belief that alternative DNA conformations encoded by genetic elements called flipons have important biological roles. Many of these alternative structures are formed by sequences originally spread throughout the human genome by endogenous retroelements (ERE) that captured 50% of the territory before being disarmed. Only 2.6% of the remaining DNA codes for proteins. Other organisms have instead streamlined their genomes by eliminating invasive retroelements and other repeat elements. The question arises, why retain any ERE at all? A new synthesis suggests that flipons enable genomes to learn and programme the context-specific readout of information by altering the transcripts produced. The exchange of energy for information is mediated through changes in DNA topology. Here I provide a formulation for how genomes learn and describe the underlying p-bit algorithm through which flipons are tuned. The framework suggests new strategies for the therapeutic reprogramming of cells.

1. Introduction

The role in biology of alternative DNA conformation like Z-DNA, G-quadruplexes (GQs), triple helices, i-motifs and others less well defined has until recently been controversial. As recently as 2007, searching for a biological Z-DNA was dismissed as ‘a dead end’ pursuit [1]. A host of recent findings have confirmed essential roles for Z-DNA and Z-RNA (collectively called ZNAs) in regulating innate immunity against viruses and cancers, and autoimmune responses [2–6]. The supporting evidence derives from many disciplines, including the detailed studies by structural, physical and biological chemists, the functional mapping of pathways by cellular biologists, and their genetic validation in both human and rodent studies [2,7–14]. Other roles for Z-DNA in transcription are also supported by computational and biological studies [15–20]. It is now possible to trace the Z α domain that binds Z-DNA [21], Z-RNA and GQ [22] back to the earliest eukaryotic unicellular precursor *Capsaspora owczarzaki* [15].

This work has led to the concept of a flipon, a genetically encoded element that regulates outcomes by its shape rather than by sequence [23,24]. The switch to an alternative fold requires energy that can arise from changes in local supercoiling due to the action of processive enzymes such as helicases and polymerases or the release of proteins such as histones that wrap DNA around themselves [25]. By tuning flipon conformation, the phase transitions associated with the readout of genetic information can operate close to criticality to vary the transcripts produced. In this review, I provide a formulation of how genomes learn to optimize flipon states. The loss function is based on well-established statistical principles that minimize the free energy costs incurred. The framework suggests new strategies for therapeutically reprogramming cells.

2. Flipons come in different shapes

There are many different types of flipon, each with a characteristic repeat sequence. For example, runs of guanosine can hydrogen-bond with each other to form tetrads that stack into GQs [26]. Alternating purine, pyrimidine repeats can invert the base pairs to form left-handed Z-DNA [27], while direct purine and pyrimidine repeats can form triplexes [28–30]. Cytosine base pairs can also intercalate to form i-motifs [31]. The energy required to initiate the transition can be generated within the cell by polymerases and helicases, or by other sources of helical stress, such as stretching and twisting [32,33].

The spread of flipons throughout genomes has largely been driven by transposons, with retrotransposition playing an important role in vertebrates. In humans, the ALU elements are usually dimeric and ~280 bases long. These repeats were named for the *Arthrobacter luteus* restriction endonuclease that cuts ALUs in genomic DNA at positions 168 and 228 to produce 49 base pair fragments. ALUs comprise about 11% of the total genomic sequence. They activate innate immunity by forming Z-RNA s. ALUs also contain GQ and triplex sequence motifs with biological activity [34], as do the long interspersed nuclear elements and long terminal repeat (LTRs) comprising another 20% of the human genome [35]. The high frequency of flipons in the genome renders them uninformative as B-DNA. However, in their alternative conformation, flipons flag active parts of the genome for the cellular machinery to engage. As flipon sequences are embedded in DNA and therefore heritable, they are subject to natural selection. In the simplest form, the encoding is binary, with flipons existing in either one shape or another. The flip occurs without any alteration to the underlying DNA sequence and without requiring strand breakage. For many situations, this outcome works well, with the switching on or off of innate immune responses by Z-DNA or Z-RNA providing an example [3–6]. However, many repeat sequences can adopt more than one alternative conformation. For example, the conformation adopted by a d(G:C)_n repeat B-DNA varies with its length: short duplex repeats are capable of forming Z-DNA [36], while longer stretches not only fold into GQs in single-stranded regions of DNA [37], but can accommodate a third strand to become triplex [38]. Further, the complementary cytosine strand can assemble into an i-motif [31]. The conformation adopted by d(G:C)_n repeat B DNA can vary with the location: triplex formation is highest in ribosomal arrays at sites where the meiotic recombination protein PRDM9 binds [39], with Z-DNA favoured downstream of a promoter [40,41] and GQ formation by oxidative damage to the guanine base [42]. Base modifications also alter the energetic cost of flipping other flipon types to an alternative conformation, with the d(5meC-G)_n forming Z-DNA under physiological salt conditions [43].

The rather dynamic formation of ZNA compared with the more stable nature of GQ and other flipons accounts for differences in the biology of these different flipon types. For example, during transcription, Z-flipons allow for the rapid reset of promoters. By contrast, G-flipons depend more on a lock and load mechanism, where GQ resolution is triggered by specific events. This mechanism times the onset of transcription, ensuring that promoters only fire when tripped [44–50]. The various roles for Z- and G-flipons in modulating transcription are well known and have been recently reviewed [50–53].

3. The evolutionary framework

Questions arise about the origins of flipons and their fit with the current concepts of gene regulation. It is likely that flipon-based mechanisms arose before codons evolved (figure 1). At an early stage, simple nucleic repeats capable of self-replication emerged from the primordial soup. Given the harsh conditions, these polymers probably adopted many alternative conformations. Informational roles arose from their interactions with peptide polymers and metals, a subset of which were catalytic [54]. I have called these entities ‘tinkers’ in reference to François Jacob’s description of Nature as a tinkerer [55]. The shape and sequence-specific interactions essential for the self-replication of tinkers probably contributed to the modern genetic code. The base-specific interactions with other nucleic acids embedded catalytic tinkers and flipons within modern genomes [53,56].

The information content of the first tinkers was low due to their repetitive nature. Those sequences able to form non-B-DNA conformations were able to nucleate different complexes in response to environmental change. The structure-specific outcomes increased their adaptability. The subsequent evolution by tinkers of sequence-specific proteins exploited the much greater dimensionality of a four-base code compared with one based on the different flipon structures. Coding by base rather than by backbone yielded a greater diversity of outcomes. As genomes developed, duplication of genes exploited adaptations that were successful in the past, with mutations giving rise to new functionalities [57]. Nevertheless, the new system was built on the flipons that already existed. Clearly, adaptations that relied on flipons to protect against pathogens still persist until this day, although in some organisms they have been supplanted by small interfering RNAs that bind to pathogen nucleic acids in a sequence-dependent manner to disrupt their function [15]. The retention of flipons also proved advantageous in other ways. By varying how coding transcripts are expressed and processed, flipons increased the algorithmic complexity of the RNAs produced [58]. The fliponware enabled the encoding of alternative gene products without a change in the codonware [59]. The transcripts generated varied dynamically, allowing rapid adaptation as selection pressures changed.

Both flipons and codons therefore play an important but unique role in the evolution of new phenotypes. The two forms of genetic encoding optimize outcomes in a different manner. With codons, the need to maintain function restricts the sequence space that can be explored. The limitations are most severe for proteins that perform multiple functions, where a mutation that optimizes a particular outcome could compromise its other roles. As Ohno noted, one way to escape such severe strictures is to duplicate protein coding genes and evolve each copy separately, ensuring that the different functionalities are conserved in the genome [57]. Flipons evolve in different ways. They can explore a larger sequence space than codons and vary the protein products made by altering gene expression and RNA processing. By flipping to an alternative conformation, flipons can just simply capture the energy needed to maintain an open chromatin region, enabling a sequence-specific transcription factor (TF) to dock to a promoter or an enhancer element. Alternatively, any effect that prevents the flip can simply prevent

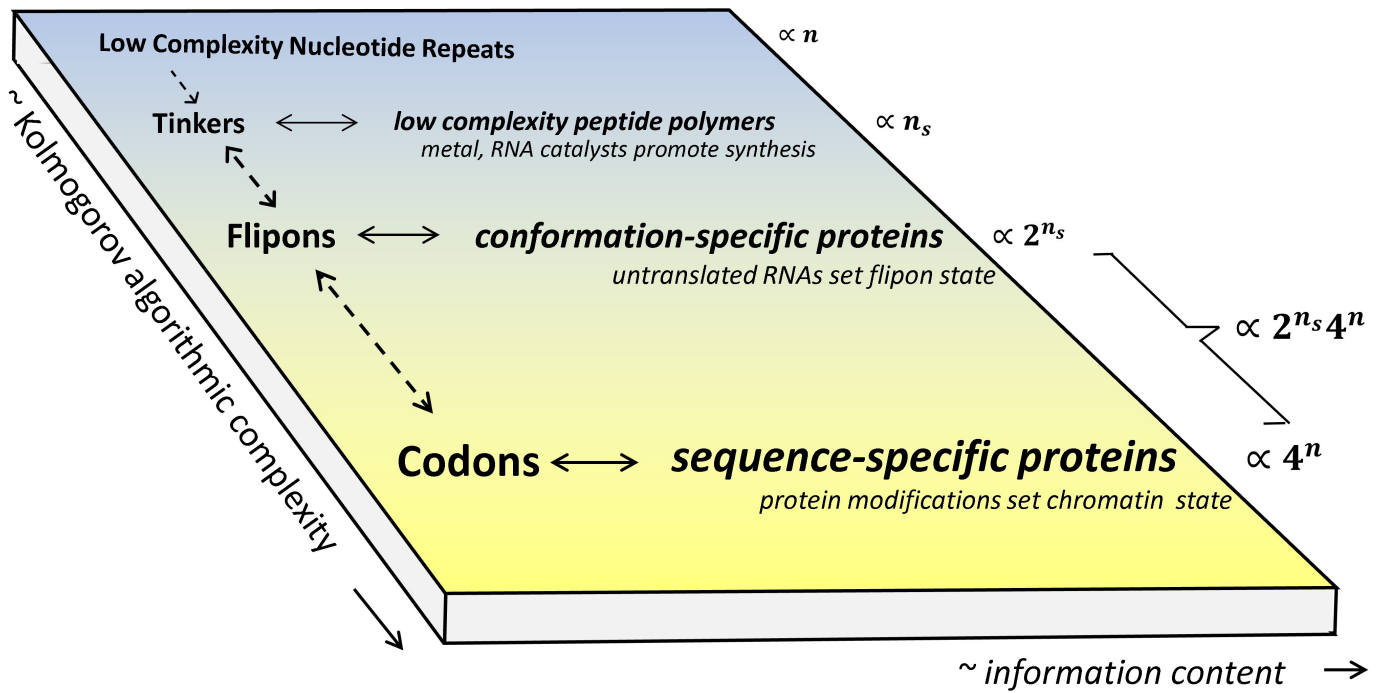


Figure 1. The evolutionary landscape of genetic regulation. A two-dimensional view of the increase in informational content and computational complexity over time, with tinkers as the first self-replicating agents that probably contributed to the development of the genetic code. The different flipons denoted by n_s switch between two or more states with probability p_s , with the conformation adopted depending on the length of the flipon, as discussed in the text. Flipons code by their shape, which, at an early stage, was probably regulated by sequence-specific interactions with other nucleic acids, rather than by the sequence-specific binding of proteins, which were a later innovation. The evolution of sequence-specific proteins provided another layer of regulation. Their synthesis was enabled once tinkers elaborated the genetic encoding of amino acids. The combination of codons and flipons greatly increased the coding potential of genomes as indicated by the scale on the right.

expression of that RNA by impeding the maintenance of an open chromatin region. The spread of flipons throughout the genome by transposable elements also generates phenotypic variation by changing the readout of genetic information in the neighbourhood where the insertion occurs. When integrated into pathways that perform core cellular functions, flipons will exhibit phylogenetic sequence conservation, as recently shown [16]. Interestingly, Z-DNA is not, as once thought, mutagenic, and GQs are only mildly so, suggesting that mutation is not the primary way that flipons evolve [60].

Flipon conformation is programmable with RNAs, both long and short. For example, the formation of triplexes induced by long non-coding RNAs (lncRNAs) has been experimentally demonstrated in a number of systems. Evidence also exists for the programmability of G- and Z-flipon conformation by RNAs ([30,61–63] and [56] plus reference within) that bind flipons in a sequence-specific manner (figure 2). Such interactions allow the use of miRNAs and other small RNAs such as tRNA fragments to regulate tissue development and to target flipons throughout the genome selectively. The interactions constrain the sequence variation of both miRNAs and flipons.

Selection pressure on flipons can also arise from transgenerational effects. Notably, suppression of flipon containing retroelements in the embryo by small sequence-specific piRNAs and t-RNA fragments is well known, as well as proteins such as ZBTB43 that target specifically a subset of Z-flipons composed of d(A–C) repeats [67–69]. However, small RNAs transmitted from the parents also potentially play a role in bootstrapping embryonic development [70–72]. By programming flipon conformation, the small RNAs help define the regions of open chromatin that are formed later in development. They probably act by placing epigenetic tags throughout the genome at the bivalent promoters later involved with tissue specification [56].

Exactly how flipon conformations were optimized during evolution is an interesting question, as it also relates to how genomes learn.

4. Flipons and codons: topology versus typography

The flipon-based systems work differently from the classical models of genetic programming, in which input sets a cell state that drives the response. This scheme is exemplified by the bacterial studies of Jacob and Monod, where an operon clusters the genes needed for the response to a particular environmental change [73]. By contrast, flipons settings are tasked with producing a particular state, either maintaining the current circuitry or switching to another programme. The system is extremely plastic, generating a variety of solutions, with some more robust than others. Interestingly, triplexes, Z- and G-flipons are clustered throughout the genome, particularly in promoter regions of genes where there are chromatin-free regions detectable by DNase footprinting [20,40,74–79]. This arrangement allows the conformation of flipons to vary, with many different combinations of flipon states possible. The exact promoter configuration depends on the local level of supercoiling (figure 2) and its partitioning between the different flipon types in the neighbourhood. These outcomes have been studied *in vitro* and *in vivo* [16,40,41,53,80,81].

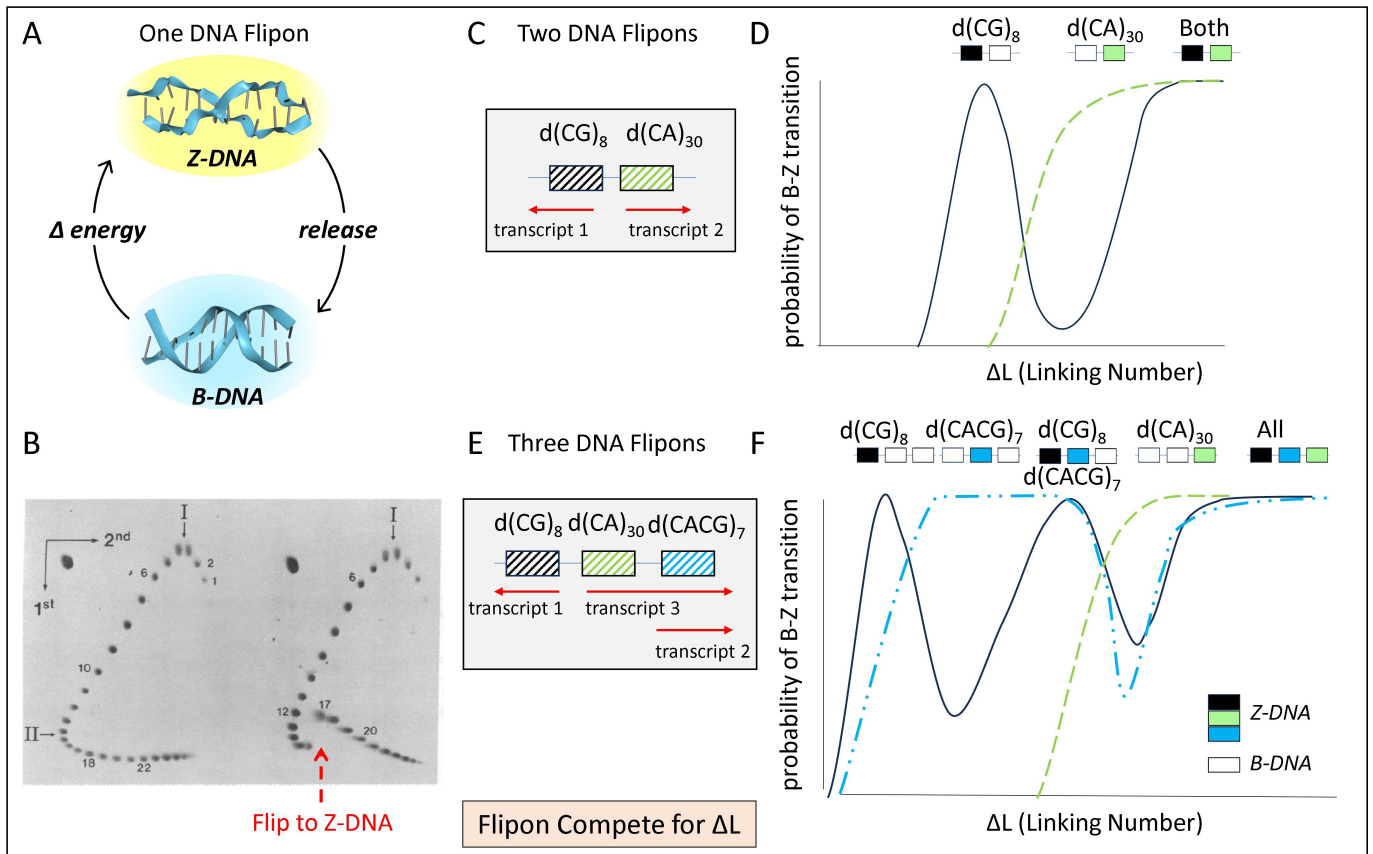


Figure 2. Flipons compete for energy. (A) The cycling of a Z-flipon between B-DNA and Z-DNA requires an input of energy that can be used to power different biological outcomes. (B) The change in linking number can be determined experimentally by separating topoisomers on two-dimensional agarose gels. A Z-DNA forming plasmid is shown on the right with the parental plasmid on the left. The change in twist associated with Z-DNA formation by the $d(CG)_{16}$ insert can be determined by comparing topoisomers 12 and 17 (as labelled in panel (B)), which have equivalent writhe, i.e. $\Delta L = \Delta T$ (from [64] with permission of James Wang). (C) Two adjacent flipons (indicated by the striped boxes) that are transcribed in opposite directions as indicated by the red arrows (adapted from [65]). (D) The conformation of each flipon depends on the level of negative supercoiling and the conformation of the other. The black shading indicates Z-DNA formation while the white and striped boxes indicates B-DNA, showing how $d(CG)_8$ flips to Z-DNA, back to B-DNA and then again to Z-DNA, then back to B-DNA and finally to Z-DNA again. (E,F) The three adjacent flipons that are transcribed in opposite directions. The set of active flipons changes as the negative supercoiling increases, with five different combinations shown (adapted from [66]).

In vitro, flipons are easily studied in isolation, allowing the determination of various parameters that determine their conformation (figure 2A). The literature for Z-DNA, G-quadruplexes, triplexes, and more recently, for i-motifs is quite extensive with the cellular biology of Z-DNA and G-quadruplexes best understood [45,82–84]. In the case of a single Z-flipon, the transition is quite straightforward as shown in figure 2B. Here, supercoiling is defined by the linking number L , which is the number of times the two DNA strands of the double helix twist around each other (T is positive for right-handed B-DNA and negative for left-handed Z-DNA), plus the writhe, which is the number of times a DNA crosses itself when imaged on a flat surface (W is positive for a right-handed cross). Defined this way, $L = T + W$. Normally, it is sufficient to measure the difference in L ($\Delta L = \Delta T + \Delta W$) relative to a reference DNA segment. The values are ascertained using agarose gels to separate the different plasmid topoisomers at sufficient resolution to detect each unit change in L . Topoisomers that have a difference in L but the same W , are used to determine the change in T due to the formation of an alternative flipon conformation: since $\Delta W = 0$ in this case, then $\Delta L = \Delta T$ [64–66,85]. This is shown in figure 2B, where the topoisomers labelled 12 and 17 have the same writhe after the flip to Z-DNA as indicated by the red arrow, giving the ΔT for the $d(CG)_{16}$ insert as equal to -5 (corresponding to a change in helix winding of $-(1/12 + 1/10)/\text{base}$ with 0.4 turns per B–Z junction). By comparing Z-DNA forming inserts of different sizes, it is possible to estimate the energetic cost of flipping (ΔE) and of forming B–Z junctions [86].

Techniques for estimating ΔL in cells after perturbation also exist [87–91]. The superhelical density is given by $\sigma = \Delta L/N$ (n = length of the closed circular DNA segment). In one system, the negative supercoiling upstream of a divergent, bidirectional, actively transcribing promoter can generate $\sigma \geq -0.07$ [76,88]. This level is more than enough to flip a $d(CG)_{16}$ insert to Z-DNA where a $\sigma = -0.011$ is sufficient. By comparison, a nucleosome reduces σ by approximately -0.0086 ($(\Delta L = -1.26)/147$ bases) [90]. The critical σ value to flip a sequence can vary with modifications to the DNA bases. For example, methylation of cytosine lowers the energy needed to propagate the change in conformation to adjacent sequences [43], as do polyamines such as spermine [92]. Topoisomerases and polymerases can also alter the local superhelical density to regulate the energy available to power the flip [20,32].

The situation is more complex when neighbouring flipons compete for the available supercoiling. An example is proved by the *in vitro* analysis of a closed circular DNA segment that contains two Z-DNA forming segments that differ in length and in their propensity to flip to Z-DNA (figure 2C,D) [65]. The longer $d(CA)_n$ flipon is less likely to form Z-DNA than

the shorter $d(CG)_n$ flipon, as more energy is required to initiate the initial transition. As negative supercoiling increases, the $d(CG)_n$ sequence first forms Z-DNA. As the supercoiling increases, the $d(CA)_n$ then flips. Due to the cooperativity of the flip, the $d(CA)_n$ flipon absorbs all the available energy, eventually causing the $d(CG)_n$ segment to flip back to B-DNA as there is insufficient $-\sigma$ available to maintain the Z-DNA conformation. The situation is even more complicated when three flipons are connected, as in figure 2E,F. In the case of $d(CG)_n$, the flip is reversed when $d(CACG)_n$ forms Z-DNA and then again when $d(CA)_n$ flips [66]. Each time $d(CG)_n$ forms Z-DNA, a region of open chromatin forms (dark boxes in figure 2C–F indicate Z-DNA formation and active promoters, while striped boxes represent flipons in the B-DNA conformation), potentially facilitating transcription (as indicated in the figure by the direction of the red arrows). The promoter associated with $d(CG)_n$ is then active either alone or in conjunction with one or both of the other two promoters, producing a different combination of transcripts.

The competition for energy from other flipons in the neighbourhood enables the reprogramming of cells, with multiple transitions between active and idle states as the level of free supercoiling increases. The resetting of flipons increases phenotypic diversity. The most extreme example of plasticity is the reversion of a mature cell to a stem cell state [93]. Interestingly, these processes appear to involve miRNAs and are inhibited in culture by knockout of either the *DICER1* or *DGCR8* genes [94]. Intriguingly, conserved miRNAs also have seed sequence binding sites in promoters that overlap flipons and the potential to regulate their conformation [56].

5. The statistical framework

Flipon-dependent outcomes are determined by how energy is partitioned between the different flipon states and each flipon type (figure 3). The transitions can be modelled using Jaynes's principle of maximal entropy [95], which subsumes the Shannon generalization of the Gibbs formulation [96–98]. The approach selects the probability distribution with the greatest entropy by partitioning the available free energy of supercoiling between the alternative DNA conformations. The elaboration is information theoretic and assumes only the measurable features of a system.

The following description is a general formulation for the thermodynamic equilibria involved and is currently experimentally validated for simple systems based on covalently closed, circular plasmids, which have DNA ends ligated together to constrain supercoiling. The approach measures the effects of DNA sequence variants on the partitioning of superhelical stress within the plasmid and the energetic cost of flipping these DNA variants to an alternative conformation. In more complex systems, such as in genomes and cells, ends are usually fixed by anchoring to a scaffold, leading to the creation of topological domains [99].

The analysis starts with the familiar definition of the Helmholtz Free Energy \mathcal{F} .

$$\mathcal{F} = \langle E \rangle - TS, \quad (5.1)$$

where $\langle E \rangle$ is the estimate of the internal energy, T is the temperature in K. The aim is to find a partition function Z that gives the best fit to the data (figure 3B), assuming a Boltzmann distribution of flipon states where n = the number of states S , with n_0 representing the count at the lowest energy state (defined as B-DNA for flipons), ΔE_s is the energy difference between the alternative flipon conformation and the ground state E_0 with $S \geq 1$, $s \in \mathbb{N}$, with

$$N = n_0 + n_0 e^{-\Delta E_1/K_B T} + n_0 e^{-\Delta E_2/K_B T} + \dots + n_0 e^{-\Delta E_s/K_B T}, \quad (5.2)$$

$$Z = \frac{N}{n_0} = 1 + e^{-\Delta E_1/K_B T} + e^{-\Delta E_2/K_B T} + \dots + e^{-\Delta E_s/K_B T} \quad (5.3)$$

and the probability of each state after normalization by Z given by

$$p_s = \frac{e^{-E_s/K_B T}}{Z}, \quad p_s \geq 0, \quad (5.4)$$

with the expected value of E and S after normalization given by

$$\langle E \rangle = \sum_s E_s p_s, \quad (5.5)$$

$$\langle S \rangle = \sum_s p_s \ln(p_s), \quad \text{with } \sum_s p_s = \sum_s \frac{e^{-(E_s/K_B T)}}{Z} = 1. \quad (5.6)$$

Substitution of p_s yields the relationship between entropy and the partition function

$$\langle S \rangle = \frac{\langle E \rangle}{T} + K_B \ln(Z). \quad (5.7)$$

Given the change in linking number $\Delta L = \sum_s (\Delta L_s)$ is the upper bound of supercoiling, and with T the twist and W the writhe, then

$$\Delta L = \Delta T + \Delta W. \quad (5.8)$$

The measured ΔL is numerically related to the ΔE by

$$\Delta E_s = w(\Delta L_s)^2, \quad (5.9)$$

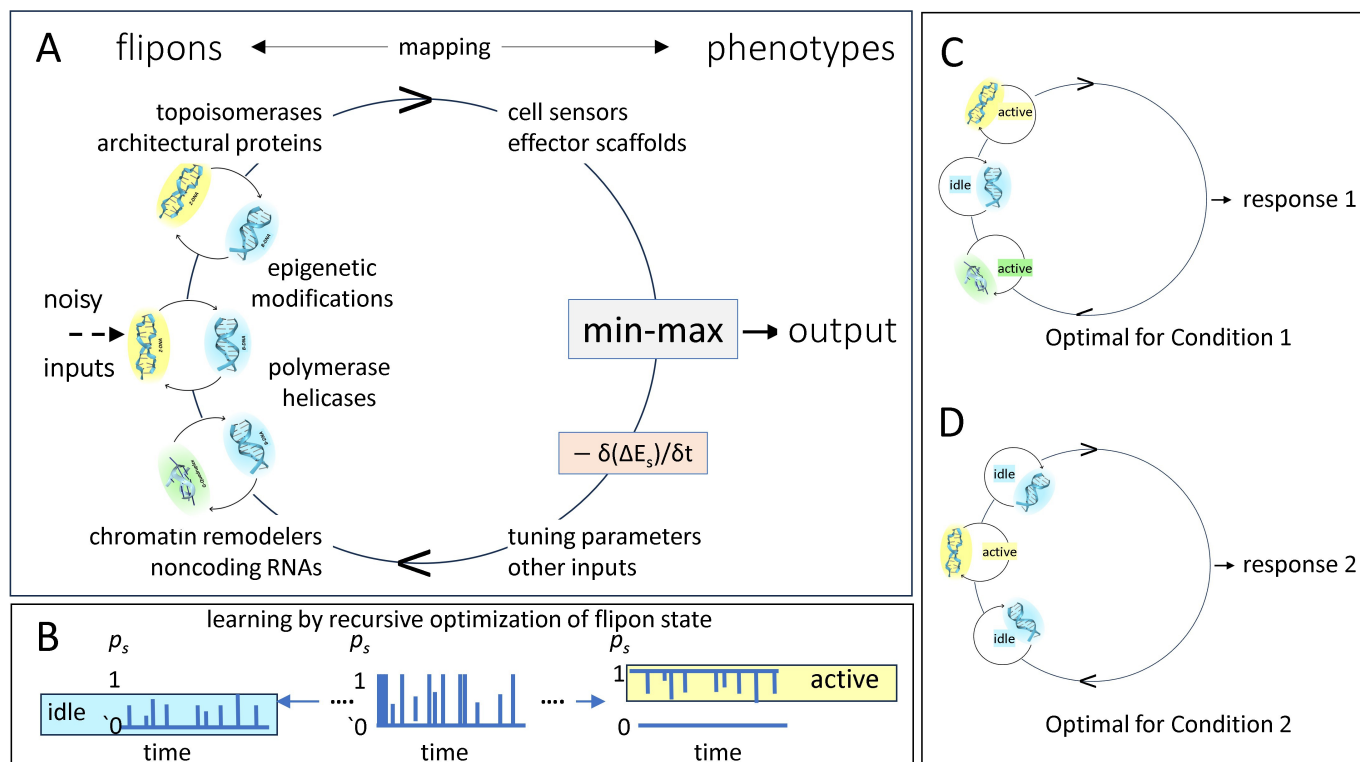


Figure 3. Flipon calibration. (A) Flipon settings are tuned to minimize or maximize a particular output in a variety of ways that impact the amount of energy necessary to induce the flip, and through pathways that control the local level of supercoiling. The process is iterative (as indicated by the directed cycle) and optimizes the mapping of the flipon state to output. Initially, flipons cycle stochastically between B-DNA (coloured blue) and the non-B-DNA active conformation (coloured yellow for Z-DNA or green for GQ). (B) Learning alters the probability p_s that a flipon will either be idle or active, as discussed in the text. (C,D) The phenotypes produced vary by condition with only some flipons cycling to the active state while the others remain idle.

with $w = 1100RT/\mathcal{N}$ [64,100], where R is universal gas constant, and $\mathcal{N} > 2000$ bases [101]. The probability distribution of flipon states and their internal energy can then be estimated using the Lagrange multipliers λ , μ and ν using the specified constraints,

$$\mathcal{L} = \sum_s p_s \ln(p_s) - \lambda \sum_s p_s e^{-(E_s/K_B T)} + \mu (\sum_s p_s - 1) + \nu (\Delta L - \sum_s (\Delta L_s)). \quad (5.10)$$

Solving \mathcal{L} in the usual way yields a solution where all partial derivatives of \mathcal{L} vanish. The first term for \mathcal{L} equates to the Shannon entropy, the second term to the energy distribution across states, the third term to the probability of each structure and the last term provides the upper bound for supercoiling of the system. The optimal solution for \mathcal{L} is usually found by enumeration. *In vitro* estimates of the energetic cost of nucleating the flip by creating a junction and for propagating a conformation to adjacent bases speed the calculation (figure 3B). Other costs vary by flipon type and arise from the concomitant compression, stretching and twisting of adjacent B-DNA regions that depend on the torsional rigidity of the fold and on changes in the axial orientation of the flipon relative to other DNA segments [64–66,85,102].

6. The computational framework and a practical guide

By setting the energy required for a flipon to change conformation, the system is capable of learning from the output generated. As a result, a lower level of energy (i.e. less $\Delta E = w(\Delta L)^2$) is required to switch flipon conformations, while the probability of other alternative flipon structures is greatly diminished. Figure 3A gives one example of an assembly containing three fully connected flipons that initially fluctuate between the B- and Z-DNA conformations, with system noise affecting the overall probability of a flip. The noise is useful as it allows exploration of different combinations of flipon states, and the escape from suboptimal, local energy minima. Over time, the system is tuned to either minimize or maximize a particular output in response to intra- and extra-cellular inputs (setting flipons at each time point to either an idled ‘off’ ($p_s = 0$) or an active ‘on’ state ($p_s = 1$) (figure 3B). Conceptually, the tuning is a Bayesian process, with the current probability distribution of flipon states updated as new information becomes available. Overall, learning minimizes the sum of these transitions as represented by the min-max function of figure 3A. The learning rate is then given by $\delta(\Delta E_s)/\delta t$ for each flipon in an ensemble. The adjustments improve future performance by resetting the threshold for assembling the appropriate cellular machinery at the locus where it is required. The tuning can involve only local flipons or incorporate others recruited from nearby or distant loci and can be summarized by how strongly the flipons states are coupled with each other. An increased number of active flipons in an ensemble allows both the response and the complexity of condensates to scale. The underlying phase transitions can then be set to operate close to criticality where a change in state propagates over a cellular domain [103].

Practically, the question arises of how to model such a computational framework? The implementation would take existing genomic maps of nucleosome-free regions defined by DNase I hypersensitivity sites (DHS). Greater than 94.5% of TF map to DHS [77], which are also enriched for experimentally validated G-, T- and Z-flipons [16,30,79]. The placement of the +1 nucleosome at transcription start sites and at the 3' end of the DHS is strictly defined by the H3K4me3 (histone H3, trimethylated lysine 4) epigenetic mark [77]. Even a limited analysis based on the DHS maps reveals the power of this approach. Using just self-organizing maps, patterns spanning across multiple chromosomes and shared by different cell types and cell states were identifiable [77]. The inclusion of other measures will help better map nucleosome phasing and identify the open TF binding sites and those flipons free to cycle conformation. The available maps include those for protein-free (i.e. direct) RNA interactions with DNA and the different measures of chromatin contacts made with nucleosome resolution. The maps are based on techniques such as RADICL-seq to identify the small RNAs and long non-coding RNAs that potentially set flipon state [104], and Micro-C to capture changes in chromatin conformation and composition as cells switch state [105,106].

Perturbation data to train models is available from resources such as DepMap. The measures include those for gene, drug and microRNA disruptions [107–109]. Available validation sets include the single cell and tissue atlases of RNA expression, with measures for coding and non-coding transcripts and their modifications [110–112]. Models with a focus on how changes in flipon conformation alter RNA outputs are a good place to start. These models would take into account the interactions of flipons with RNAs, their impact on promoter usage, RNA expression and the epigenetic modifications that alter the subsequent processing of transcripts. Here the learning algorithm will optimize flipon settings that predict the active chromatin states involved. These studies could be initially restricted to a local topological domain where a perturbation creates a large effect (e.g. a heat shock response), or to contexts where cells transition to a new state (e.g. activation of a dormant cell, or induction of differentiation).

One interesting example to evaluate would be the sequential activation of genes during development within a *HOX* (homeobox protein) gene cluster [113,114]. Another equally interesting example is the initiation of the transition from maternal transcripts to zygote gene activation in 2-cell zygotes that depends on MTa (mouse transposon a) retroviral promoters. In this pathway, the MTa promoters initiate embryonic gene transcription, producing a chimeric transcript that drives the production of key cellular proteins. The transcript is made by splicing the retroviral and downstream RNAs together. Later, the MTa transcript undergoes cleavage rather than splicing, with the downstream gene promoter then driving cellular gene transcription [115]. Both the MTa retroelement splice and polyadenylation motifs are adjacent to each other, and both are embedded within a cluster of potential triplex-forming flipons. The triplezes set by RNAs can position nucleosomes through their interactions with chromatin remodellers [116]. It is likely that the initial nucleosome placement favours splicing of the MTa transcript by ensuring that the splice site is in an open DNA region. Later, the nucleosomes are positioned to inhibit splicing and promote transcript cleavage at the polyadenylation site. In such an arrangement, inhibition of splicing and truncation of the retroviral transcript switches the transcription start site from the MTa promoter to the gene promoter. The transition is probably initiated by the spliced transcript made from the MTa promoter. This transcript folds back to form a triplex that changes the nucleosomal placement and the promoter used. Other examples of triplex-based promoter switches are known. For example, the SPHK1 downstream promoter produces an antisense KHPS1 RNA that initiates triplex formation at an upstream promoter, leading to enhanced expression of the sense transcript [117]. The human β -globin locus also uses a triplex switch to downregulate ϵ - and γ -globin upstream promoters and enhance expression of the downstream β -globin triplex-forming transcript [62]. These different use cases provide simple models for the evaluation of genomic learning models. In such examples, the probability that G- and Z-flipons in promoters adopt an alternative conformation is moderated by triplexes formed with long transcripts. Small RNAs targeting flipons and transposons, such as miRNA and piRNAs, or those induced by perturbation or sequestered at other sites, collectively adjust the probability that a particular flipon is active.

7. Flipons and p-bit computing

This overall schema represents an implementation of p-bit computing, a recently described architecture that offers greatly improved performance in estimating probability distributions compared with strictly deterministic digital computers, incurring a significantly lower energy cost and an easier implementation than possible with quantum bit systems [118]. An example given by Camsari and colleagues describes factorization of the number 'N'. The approach uses an adder circuit composed of inherently noisy logic gates that randomly switch between 'on' and 'off' states. The adder is run in reverse to calculate the factors of 'N'. The 'off/on' probability distribution of the adder gates undergoes tuning to minimize the difference between the current adder output and the target value. The iteration terminates with the discovery of adder settings that generate N. The factors of N are then read from the adder registers [118]. The Camsari approach requires very fast clock times to ensure rapid convergence on the most likely solution. Further, these gates must be updated asynchronously to prevent resonance effects where gates cycle between a small number states without converging on a meaningful answer. Neither fast times nor precise wiring are features of biological systems.

The genomic implementations of p-bit computation also differ in a number of other ways from the Camsari implementation. There is a difference in scale, as genomes can exchange information over large distances. They communicate by directly altering the topology of chromosomal domains [119,120], or by local changes to DNA twist and writhe [121]. The outcomes may be less direct and mediated by the structure-specific docking of RNAs or involve architectural proteins that align distant DNA segments [122]. There are also limits to how fast flipons can switch states. The rate for Z-flipons is of the order of 100 ms [123], while G-flipons often represent energy minima that potentially trap the GQ conformation, with helicases and energy

expenditure required to reform B-DNA [44,124]. The different kinetics are reflected in flipon biology; Z-flipons act as actuators, while G-flipons can serve as memory elements to preserve cell state during differentiation and cell division [53]. Flipons capable of adopting different alternative conformations increase the range of outcomes. The learning rate $\delta(\Delta E_s)/\delta t$ then varies with the flipon involved.

8. Error correction by genomic systems

Another notable contrast between flipons, p-bit computers and Turing machines [125] is that halting by a biological system usually signifies failure, not success. To ensure continued operation, genomes implement robust error correction protocols (ECP) to triage and reduce the complexity of the output. Such ECP are necessary as circuits like those in figure 2C–F can produce a multitude of defective transcripts. Eliminating low-quality RNA transcripts that have splicing errors or premature stop codons by nonsense-mediated decay pathways improves system performance [126–128]. Other pathways make base modifications to nascent transcripts that direct the destruction of the RNA if they are not subsequently removed from the transcript. These marks, like methy-6-adenosine, are frequently found in introns. A different set of ECPs use sequence-specific RNAs to guide the destruction of endogenous retroelements (ERE) and pathogen transcripts by piwi and argonaute proteins [67,129,130]. Other ECPs prevent flipons from permanently freezing in one state or another [131,132]. These pathways resolve persistent R-loops, facilitating the resolution of GQ by rehybridization of the G- and C-rich strands, or repair flipons when they break [84,133]. All these ECPs come at a metabolic cost as many of the RNA produced are never used. Instead, they are junked.

With the exception of *Monocercomonoides* species, eukaryotes depend upon mitochondria to power the setting and resetting of flipons that enable genomes to learn [134]. The strategy enables the rapid rewiring of genetic programmes without any change to the underlying genomic sequence while allowing genomes to learn those settings that are optimal for a particular set of conditions. The design provides a large canvas for evolutionary innovations that goes beyond the simple canalization of phenotypes [135]. Indeed, the evidence now suggests that canalization is easily reversed to regenerate stem cells [93]. The repetitive genome has a role to play in such processes: as the number of flipons increases, it becomes less likely that any two cells share the same exact flipon settings or have the same identical phenotype. In such cases, some cells are better adapted to the current context than others. The most attuned cells then are more likely to undergo selection and clonally populate tissues [136]. The outcomes may not represent the best of all possible solutions, but they allow individuals to survive the moment in ways that the selfish gene described by Dawkins does not [137].

9. Future

As with other disciplines, biology is moving to a more dynamic conceptualization of the processes occurring within the cell. As regulatory processes evolve, new adaptations are layered on top of pre-existing ones, with each subsequent innovation dependent on those that came before. Flipons probably represent an early form of genetic encoding that is still with us today and now provides a foundation for the more elaborate codon-based specification of the modern-day protein machinery (figure 1). Likewise, our understanding of such events is under constant revision. The changes are reflected in the way we currently conceptualize information flow within a cell. At the dawn of the electronic computer age, we looked for wiring diagrams that enabled the biology. Now the focus is on macromolecular condensates that coordinate critical cellular chemistries. The switches involved are the phase transitions that initiate condensate formation. The critical point reflects when and how the condensate formation is triggered. The current understanding is that the cell sets switches close to criticality. Similarly, the role of non-genetically encoded phase transitions in setting chromatin organization is more extensive than expected from bacterial genetic studies. The simpler genetic systems explored in the past were based on organisms with repetitive elements stripped from their genomes [138]. The use of host-based, rather than pathogen-targeted innate immunity is but one example we have learned of recently. In a subset of phyla, the rather haphazard retention of EREs evolved a highly programmable genome fashioned with flipons.

The Bayesian optimization of genomes is indeed empowered by the probabilistic nature of flipons. Despite their dependence on repetitive sequences, flipons increase the algorithmic complexity of genomes, all by tinkering with how the codon ciphers are parsed into programmes. Many of the concepts developed to further artificial intelligence by deep learning approaches, also have application to training genomes. This new knowledge will certainly improve our ability to engineer flipons and to reprogramme cells [139]. The models will empower us to design small molecules and RNA therapeutics that reset flipon conformations to prevent disease initiation and progression. The encoding of genetic information by flipons, while unexpected, is another example of how Nature exceeds our expectations.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All data and methods are available or described in the main text or the references.

Declaration of AI use. I have used AI-assisted technologies in creating this article.

Authors' contributions. A.H.: conceptualization, formal analysis, funding acquisition, methodology, resources, supervision, validation, visualization, writing—original draft, writing—review and editing.

Conflict of interest declaration. I declare I have no competing interests.

Funding. No funding has been received for this article.

References

1. Morange M. 2007 What history tells us IX. Z-DNA: when nature is not opportunistic. *J. Biosci.* **32**, 657–661. (doi:10.1007/s12038-007-0065-5)
2. Herbert A. 2020 Mendelian disease caused by variants affecting recognition of Z-DNA and Z-RNA by the Za domain of the double-stranded RNA editing enzyme ADAR. *Eur. J. Hum. Genet.* **28**, 114–117. (doi:10.1038/s41431-019-0458-6)
3. Zhang T *et al.* 2022 ADAR1 masks the cancer immunotherapeutic promise of ZBP1-driven necroptosis. *Nature* **606**, 594–602. (doi:10.1038/s41586-022-04753-7)
4. de Reuver R *et al.* 2022 ADAR1 prevents autoinflammation by suppressing spontaneous ZBP1 activation. *Nature* **607**, 784–789. (doi:10.1038/s41586-022-04974-w)
5. Jiao H *et al.* 2022 ADAR1 averts fatal type I interferon induction by ZBP1. *Nature* **607**, 776–783. (doi:10.1038/s41586-022-04878-9)
6. Hubbard NW *et al.* 2022 ADAR1 mutation causes ZBP1-dependent immunopathology. *Nature* **607**, 769–775. (doi:10.1038/s41586-022-04896-7)
7. Subramani VK, Ravichandran S, Bansal V, Kim KK. 2019 Chemical-induced formation of BZ-junction with base extrusion. *Biochem. Biophys. Res. Commun.* **508**, 1215–1220. (doi:10.1016/j.bbrc.2018.12.045)
8. Lee AR, Hwang J, Hur JH, Ryu KS, Kim KK, Choi BS, Kim NK, Lee JH. 2019 NMR dynamics study reveals the Za domain of human ADAR1 associates with and dissociates from Z-RNA more slowly than Z-DNA. *ACS Chem. Biol.* **14**, 245–255. (doi:10.1021/acscchembio.8b00914)
9. Nichols PJ, Bevers S, Henen M, Kieft JS, Vicens Q, Vögeli B. 2021 Recognition of non-CpG repeats in Alu and ribosomal RNAs by the Z-RNA binding domain of ADAR1 induces A-Z junctions. *Nat. Commun.* **12**, 793. (doi:10.1038/s41467-021-21039-0)
10. Balachandran S, Mocarski ES. 2021 Viral Z-RNA triggers ZBP1-dependent cell death. *Curr. Opin. Virol.* **51**, 134–140. (doi:10.1016/j.coviro.2021.10.004)
11. Koehler H, Cotsmire S, Zhang T, Balachandran S, Upton JW, Langland J, Kalman D, Jacobs BL, Mocarski ES. 2021 Vaccinia virus E3 prevents sensing of Z-RNA to block ZBP1-dependent necroptosis. *Cell Host Microbe* **29**, 1266–1276. (doi:10.1016/j.chom.2021.05.009)
12. Krall JB, Nichols PJ, Henen MA, Vicens Q, Vögeli B. 2023 Structure and formation of Z-DNA and Z-RNA. *Molecules* **28**, 843. (doi:10.3390/molecules28020843)
13. Guo X, Liu S, Sheng Y, Zenati M, Billiar T, Herbert A, Wang Q. 2023 ADAR1 Za domain P195A mutation activates the MDA5-dependent RNA-sensing signaling pathway in brain without decreasing overall RNA editing. *Cell Rep.* **42**, 112733. (doi:10.1016/j.celrep.2023.112733)
14. Liang Z, Chalk AM, Taylor S, Goradia A, Heraud-Farlow JE, Walkley CR. 2023 The phenotype of the most common human ADAR1p150 Za mutation P193A in mice is partially penetrant. *EMBO Rep.* **24**, e55835. (doi:10.15252/embr.202255835)
15. Herbert A. 2024 The ancient Z-DNA and Z-RNA specific Za fold has evolved modern roles in immunity and transcription through the natural selection of flipons. *R. Soc. Open Sci.* **11**, 240080. (doi:10.1098/rsos.240080)
16. Beknazarov N, Konovalov D, Herbert A, Poptsova M. 2024 Z-DNA formation in promoters conserved between human and mouse are associated with increased transcription reinitiation rates. *Sci. Rep.* **14**, 17786. (doi:10.1038/s41598-024-68439-y)
17. Mulholland N, Xu Y, Sugiyama H, Zhao K. 2012 SWI/SNF-mediated chromatin remodeling induces Z-DNA formation on a nucleosome. *Cell Biosci.* **2**, 3. (doi:10.1186/2045-3701-2-3)
18. Zhang J, Ohta T, Maruyama A, Hosoya T, Nishikawa K, Maher JM, Shibahara S, Itoh K, Yamamoto M. 2006 BRG1 interacts with Nrf2 to selectively mediate HO-1 induction in response to oxidative stress. *Mol. Cell. Biol.* **26**, 7942–7952. (doi:10.1128/mcb.00700-06)
19. Kwon JA, Rich A. 2005 Biological function of the vaccinia virus Z-DNA-binding protein E3L: gene transactivation and antiapoptotic activity in HeLa cells. *Proc. Natl Acad. Sci. USA* **102**, 12759–12764. (doi:10.1073/pnas.0506011102)
20. Wittig B, Dorbic T, Rich A. 1989 The level of Z-DNA in metabolically active, permeabilized mammalian cell nuclei is regulated by torsional strain. *J. Cell Biol.* **108**, 755–764. (doi:10.1083/jcb.108.3.755)
21. Herbert A, Alfken J, Kim YG, Mian IS, Nishikura K, Rich A. 1997 A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc. Natl Acad. Sci. USA* **94**, 8421–8426. (doi:10.1073/pnas.94.16.8421)
22. Kang HJ, Le TVT, Kim K, Hur J, Kim KK, Park HJ. 2014 Novel interaction of the Z-DNA binding domain of human ADAR1 with the oncogenic c-Myc promoter G-quadruplex. *J. Mol. Biol.* **426**, 2594–2604. (doi:10.1016/j.jmb.2014.05.001)
23. Herbert A. 2019 A genetic instruction code based on DNA conformation. *Trends Genet.* **35**, 887–890. (doi:10.1016/j.tig.2019.09.007)
24. Herbert A. 2024 *Flipons and the logic of soft-wired genomes*, 1st edn. Boca Raton, FL: CRC Press. (doi:10.1201/9781003463535)
25. Herbert A. 2022 Nucleosomes and flipons exchange energy to alter chromatin conformation, the readout of genomic information, and cell fate. *BioEssays* **44**, e2200166. (doi:10.1002/bies.202200166)
26. Arnott S, Chandrasekaran R, Marttila CM. 1974 Structures for polyinosinic acid and polyguanylic acid. *Biochem. J.* **141**, 537–543. (doi:10.1042/bj1410537)
27. Wang AHJ, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G, Rich A. 1979 Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686. (doi:10.1038/282680a0)
28. Felsenfeld G, Davies DR, Rich A. 1957 Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023–2024. (doi:10.1021/ja01565a074)
29. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. 1987 DNA H form requires a homopurine–homopyrimidine mirror repeat. *Nature* **330**, 495–497. (doi:10.1038/330495a0)
30. Sentürk Cetin N, Kuo CC, Ribarska T, Li R, Costa IG, Grummt I. 2019 Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res.* **47**, 2306–2321. (doi:10.1093/nar/gky1305)
31. Gehring K, Leroy JL, Guéron M. 1993 A tetrameric DNA structure with protonated cytosine–cytosine base pairs. *Nature* **363**, 561–565. (doi:10.1038/363561a0)
32. Liu LF, Wang JC. 1987 Supercoiling of the DNA template during transcription. *Proc. Natl Acad. Sci. USA* **84**, 7024–7027. (doi:10.1073/pnas.84.20.7024)
33. Herbert A. 2021 To 'Z' or not to 'Z': Z-RNA, self-recognition, and the MDA5 helicase. *PLoS Genet.* **17**, e1009513. (doi:10.1371/journal.pgen.1009513)
34. Herbert A. 2020 ALU non-B-DNA conformations, flipons, binary codes and evolution. *R. Soc. Open Sci.* **7**, 200222. (doi:10.1098/rsos.200222)
35. Kojima KK. 2018 Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob. DNA* **9**, 2. (doi:10.1186/s13100-017-0107-y)
36. Ellison MJ, Kelleher RJ 3rd, Wang AH, Habener JF, Rich A. 1985 Sequence-dependent energetics of the B-Z transition in supercoiled DNA containing nonalternating purine–pyrimidine sequences. *Proc. Natl Acad. Sci. USA* **82**, 8320–8324. (doi:10.1073/pnas.82.24.8320)
37. Sen D, Gilbert W. 1988 Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364–366. (doi:10.1038/334364a0)
38. Hisey JA, Masnovo C, Mirkin SM. 2024 Triplex H-DNA structure: the long and winding road from the discovery to its role in human disease. *NAR Mol. Med.* **1**, ugae024. (doi:10.1093/narmme/ugae024)

39. Chantzi N, Patsakis M, Nayak A, Montgomery A, Mouratidis I, Georgakopoulos-Soares I. 2024 Ribosomal DNA arrays are the most H-DNA rich element in the human genome. *BioRxiv* (doi:10.1101/2024.07.12.602585)
40. Kouzine F *et al.* 2017 Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a mammalian genome. *Cell Syst.* **4**, 344–356. (doi:10.1016/j.cels.2017.01.013)
41. Umerenkov D, Herbert A, Konovalov D, Danilova A, Beknazarov N, Koh V, Fedorov A, Poptsova M. 2023 Z-flipon variants reveal the many roles of Z-DNA and Z-RNA in health and disease. *Life Sci. Alliance* **6**, e202301962. (doi:10.26508/lsa.202301962)
42. Fleming AM, Zhou J, Wallace SS, Burrows CJ. 2015 A role for the fifth G-track in G-quadruplex forming oncogene promoter sequences during oxidative stress: do these 'spare tires' have an evolved function? *ACS Cent. Sci.* **1**, 226–233. (doi:10.1021/acscentsci.5b00202)
43. Behe M, Felsenfeld G. 1981 Effects of methylation on a synthetic polynucleotide: the B–Z transition in poly(dG–m5dC).poly(dG–m5dC). *Proc. Natl Acad. Sci. USA* **78**, 1619–1623. (doi:10.1073/pnas.78.3.1619)
44. Sauer M, Paeschke K. 2017 G-quadruplex unwinding helicases and their function *in vivo*. *Biochem. Soc. Trans.* **45**, 1173–1182. (doi:10.1042/bst20170097)
45. Spiegel J, Adhikari S, Balasubramanian S. 2020 The structure and function of DNA G-quadruplexes. *Trends Chem.* **2**, 123–136. (doi:10.1016/j.trechm.2019.07.002)
46. Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. 2020 The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474. (doi:10.1038/s41580-020-0236-x)
47. Banco MT, Ferré-D'Amaré AR. 2021 The emerging structural complexity of G-quadruplex RNAs. *RNA* **27**, 390–402. (doi:10.1261/rna.078238.120)
48. Esain-García I *et al.* 2024 G-quadruplex DNA structure is a positive regulator of MYC transcription. *Proc. Natl Acad. Sci. USA* **121**, e2320240121. (doi:10.1073/pnas.2320240121)
49. Troisi R, Sica F. 2024 Structural overview of DNA and RNA G-quadruplexes in their interaction with proteins. *Curr. Opin. Struct. Biol.* **87**, 102846. (doi:10.1016/j.sbi.2024.102846)
50. Herbert A. 2024 A compendium of G-flipon biological functions that have experimental validation. *Int. J. Mol. Sci.* **25**, 10299. (doi:10.3390/ijms251910299)
51. Mao SQ *et al.* 2018 DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957. (doi:10.1038/s41594-018-0131-8)
52. Spiegel J, Cuesta SM, Adhikari S, Hänsel-Hertsch R, Tannahill D, Balasubramanian S. 2021 G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **22**, 117. (doi:10.1186/s13059-021-02324-z)
53. Herbert A. 2023 Flipons and small RNAs accentuate the asymmetries of pervasive transcription by the reset and sequence-specific microcoding of promoter conformation. *J. Biol. Chem.* **299**, 105140. (doi:10.1016/j.jbc.2023.105140)
54. Herbert A. 2025 Flipons and the origin of the genetic code. *Biol. Lett.* **21**, 20240635. (doi:10.1098/rsbl.2024.0635)
55. Jacob F. 1977 Evolution and tinkering. *Science* **196**, 1161–1166. (doi:10.1126/science.860134)
56. Herbert A, Pavlov F, Konovalov D, Poptsova M. 2023 Conserved microRNAs and flipons shape gene expression during development by altering promoter conformations. *Int. J. Mol. Sci.* **24**, 4884. (doi:10.3390/ijms24054884)
57. Ohno S. 1970 *Evolution by gene duplication*, 1st edn. Heidelberg, Germany: Springer-Verlag Berlin.
58. Herbert A. 2023 The intrinsitive logic of directed cycles and flipons enhances the evolution of molecular computers by augmenting the Kolmogorov complexity of genomes. *Int. J. Mol. Sci.* **24**, 16482. (doi:10.3390/ijms242216482)
59. Herbert A. 2020 Simple repeats as building blocks for genetic computers. *Trends Genet.* **36**, 739–750. (doi:10.1016/j.tig.2020.06.012)
60. McGinty RJ, Sunyaev SR. 2023 Revisiting mutagenesis at non-B DNA motifs in the human genome. *Nat. Struct. Mol. Biol.* **30**, 417–424. (doi:10.1038/s41594-023-00936-6)
61. Mondal T *et al.* 2015 MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA–DNA triplex structures. *Nat. Commun.* **6**, 7743. (doi:10.1038/ncomms8743)
62. Zhou Z, Giles KE, Felsenfeld G. 2019 DNA–RNA triple helix formation can function as a *cis*-acting regulatory mechanism at the human β -globin locus. *Proc. Natl Acad. Sci. USA* **116**, 6130–6139. (doi:10.1073/pnas.1900107116)
63. Cordero J *et al.* 2024 Nuclear microRNA 9 mediates G-quadruplex formation and 3D genome organization during TGF- β -induced transcription. *Nat. Commun.* **15**, 10711. (doi:10.1038/s41467-024-54740-x)
64. Peck LJ, Wang JC. 1983 Energetics of B-to-Z transition in DNA. *Proc. Natl Acad. Sci. USA* **80**, 6206–6210. (doi:10.1073/pnas.80.20.6206)
65. Kelleher RJ 3rd, Ellison MJ, Ho PS, Rich A. 1986 Competitive behavior of multiple, discrete B–Z transitions in supercoiled DNA. *Proc. Natl Acad. Sci. USA* **83**, 6342–6346. (doi:10.1073/pnas.83.17.6342)
66. Ellison MJ, Fenton MJ, Ho PS, Rich A. 1987 Long-range interactions of multiple DNA structural transitions within a common topological domain. *EMBO J.* **6**, 1513–1522. (doi:10.1002/j.1460-2075.1987.tb02394.x)
67. Höck J, Meister G. 2008 The Argonaute protein family. *Genome Biol.* **9**, 210. (doi:10.1186/gb-2008-9-2-210)
68. Ohnishi Y, Totoki Y, Toyoda A, Watanabe T, Yamamoto Y, Tokunaga K, Sakaki Y, Sasaki H, Hohjoh H. 2010 Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Res.* **38**, 5141–5151. (doi:10.1093/nar/gkq229)
69. Meng Y *et al.* 2022 Z-DNA is remodelled by ZBTB43 in prospermatogonia to safeguard the germline genome and epigenome. *Nat. Cell Biol.* **24**, 1141–1153. (doi:10.1038/s41556-022-00941-9)
70. Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. 2006 RNA-mediated non-Mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474. (doi:10.1038/nature04674)
71. Gapp K, Bohacek J. 2018 Epigenetic germline inheritance in mammals: looking to the past to understand the future. *Genes Brain Behav.* **17**, e12407. (doi:10.1111/gbb.12407)
72. Bohacek J, Rassoulzadegan M. 2020 Sperm RNA: *Quo vadis?* *Semin. Cell Dev. Biol.* **97**, 123–130. (doi:10.1016/j.semcdb.2019.07.005)
73. Jacob F, Monod J. 1961 Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356. (doi:10.1016/s0022-2836(61)80072-7)
74. Avigan MI, Strober B, Levens D. 1990 A far upstream element stimulates *c-myc* expression in undifferentiated leukemia cells. *J. Biol. Chem.* **265**, 18538–18545. (doi:10.1016/s0021-9258(17)44785-5)
75. Wittig B, Wölfl S, Dorbic T, Vahrson W, Rich A. 1992 Transcription of human *c-myc* in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.* **11**, 4653–4663. (doi:10.1002/j.1460-2075.1992.tb05567.x)
76. Kouzine F, Liu J, Sanford S, Chung HJ, Levens D. 2004 The dynamic response of upstream DNA to transcription-generated torsional stress. *Nat. Struct. Mol. Biol.* **11**, 1092–1100. (doi:10.1038/nsmb848)
77. Thurman RE *et al.* 2012 The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82. (doi:10.1038/nature11232)
78. Meuleman W *et al.* 2020 Index and biological spectrum of human DNase I hypersensitive sites. *Nature* **584**, 244–251. (doi:10.1038/s41586-020-2559-3)
79. Qian SH, Shi MW, Xiong YL, Zhang Y, Zhang ZH, Song XM, Deng XY, Chen ZX. 2024 EndoQuad: a comprehensive genome-wide experimentally validated endogenous G-quadruplex database. *Nucleic Acids Res.* **52**, D72–D80. (doi:10.1093/nar/gkad966)

80. Cer RZ *et al.* 2013 Non-B DNA v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* **41**, D94–D100. (doi:10.1093/nar/gks955)
81. Georgakopoulos-Soares I *et al.* 2022 High-throughput characterization of the role of non-B DNA motifs on promoter function. *Cell Genom.* **2**, 100111. (doi:10.1016/j.xgen.2022.100111)
82. Herbert A. 2019 Z-DNA and Z-RNA in human disease. *Commun. Biol.* **2**, 7. (doi:10.1038/s42003-018-0237-x)
83. Abou Assi H, Garavis M, González C, Damha MJ. 2018 i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **46**, 8038–8056. (doi:10.1093/nar/gky735)
84. Matos-Rodrigues G, Hisey JA, Nussenzweig A, Mirkín SM. 2023 Detection of alternative DNA structures and its implications for human disease. *Mol. Cell* **83**, 3622–3641. (doi:10.1016/j.molcel.2023.08.018)
85. Zhabinskaya D, Benham CJ. 2011 Theoretical analysis of the stress induced B-Z transition in superhelical DNA. *PLoS Comput. Biol.* **7**, e1001051. (doi:10.1371/journal.pcbi.1001051)
86. Ho PS. 2009 Thermogenomics: thermodynamic-based approaches to genomic analyses of DNA structure. *Methods* **47**, 159–167. (doi:10.1016/j.ymeth.2008.09.007)
87. Drabik CE, Nicita CA, Lutter LC. 1997 Measurement of the linking number change in transcribing chromatin. *J. Mol. Biol.* **267**, 794–806. (doi:10.1006/jmbi.1997.0917)
88. Kouzine F, Sanford S, Elisha-Feil Z, Levens D. 2008 The functional response of upstream DNA to dynamic supercoiling *in vivo*. *Nat. Struct. Mol. Biol.* **15**, 146–154. (doi:10.1038/nsmb.1372)
89. Teves SS, Henikoff S. 2014 DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* **5**, 211–218. (doi:10.4161/nucl.29086)
90. Segura J, Joshi RS, Díaz-Ingelmo O, Valdés A, Dyson S, Martínez-García B, Roca J. 2018 Intracellular nucleosomes constrain a DNA linking number difference of -1.26 that reconciles the Lk paradox. *Nat. Commun.* **9**, 3989. (doi:10.1038/s41467-018-06547-w)
91. Gilbert N, Marenduzzo D. 2024 Topological epigenetics: the biophysics of DNA supercoiling and its relation to transcription and genome instability. *Curr. Opin. Cell Biol.* **92**, 102448. (doi:10.1016/j.cob.2024.102448)
92. Thomas TJ, Meryhew NL, Messner RP. 1990 Enhanced binding of lupus sera to the polyamine-induced left-handed z-DNA form of polynucleotides. *Arthritis Rheum.* **33**, 356–365. (doi:10.1002/art.1780330308)
93. Takahashi K, Yamanaka S. 2006 Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676. (doi:10.1016/j.cell.2006.07.024)
94. Lüningschrör P, Hauser S, Kaltschmidt B, Kaltschmidt C. 2013 MicroRNAs in pluripotency, reprogramming and cell fate induction. *Biochim. Biophys. Acta* **1833**, 1894–1903. (doi:10.1016/j.bbamcr.2013.03.025)
95. Jaynes ET. 1957 Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630. (doi:10.1103/physrev.106.620)
96. Gibbs JW. 1960 *Elementary principles in statistical mechanics, developed with especial reference to the rational foundation of thermodynamics*, p. 207. New York, NY, USA: Dover Publications.
97. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb00917.x)
98. Jaynes ET. 1965 Gibbs vs Boltzmann entropies. *Am. J. Phys.* **33**, 391–398. (doi:10.1119/1.1971557)
99. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380. (doi:10.1038/nature11082)
100. Depew DE, Wang JC. 1975 Conformational fluctuations of DNA helix. *Proc. Natl Acad. Sci. USA* **72**, 4275–4279. (doi:10.1073/pnas.72.11.4275)
101. Ellison MJ, Feigon J, Kelleher RJ III, Wang AHJ, Habener JF, Rich A. 1986 An assessment of the Z-DNA forming potential of alternating dA-dT stretches in supercoiled plasmids. *Biochemistry* **25**, 3648–3655. (doi:10.1021/bi00360a026)
102. White JH, Bauer WR. 1989 The helical repeat of nucleosome-wrapped DNA. *Cell* **56**, 9–10. (doi:10.1016/0092-8674(89)90977-x)
103. Muñoz MA. 2018 Colloquium: criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* **90**, 031001. (doi:10.1103/revmodphys.90.031001)
104. Bonetti A *et al.* 2020 RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* **11**, 1018. (doi:10.1038/s41467-020-14337-6)
105. Krietenstein N *et al.* 2020 Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* **78**, 554–565. (doi:10.1016/j.molcel.2020.03.003)
106. Hsieh THS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X. 2020 Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553. (doi:10.1016/j.molcel.2020.03.002)
107. Iorio F *et al.* 2016 A landscape of pharmacogenomic interactions in cancer. *Cell* **166**, 740–754. (doi:10.1016/j.cell.2016.06.017)
108. Tsherniak A *et al.* 2017 Defining a cancer dependency map. *Cell* **170**, 564–576. (doi:10.1016/j.cell.2017.06.010)
109. Merk DJ *et al.* 2024 CRISPR-Cas9 screens reveal common essential miRNAs in human cancer cell lines. *Genome Med.* **16**, 82. (doi:10.1186/s13073-024-01341-4)
110. Karlsson M *et al.* 2021 A single-cell type transcriptomics map of human tissues. *Sci. Adv* **7**, eabh2169. (doi:10.1126/sciadv.abh2169)
111. Rishik S, Hirsch P, Grandke F, Fehlmann T, Keller A. 2025 miRNATissueAtlas 2025: an update to the uniformly processed and annotated human and mouse non-coding RNA tissue atlas. *Nucleic Acids Res.* **53**, D129–D137. (doi:10.1093/nar/gkae1036)
112. Coorens THH *et al.* 2025 The human and non-human primate developmental GTEx projects. *Nature* **637**, 557–564. (doi:10.1038/s41586-024-08244-9)
113. Dollé P, Izpisua-Belmonte JC, Falkenstein H, Renucci A, Duboule D. 1989 Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. *Nature* **342**, 767–772. (doi:10.1038/342767a0)
114. Izpisua-Belmonte JC, Falkenstein H, Dollé P, Renucci A, Duboule D. 1991 Murine genes related to the *Drosophila* AbdB homeotic genes are sequentially expressed during development of the posterior part of the body. *EMBO J.* **10**, 2279–2289. (doi:10.1002/j.1460-2075.1991.tb07764.x)
115. Peaston AE, Esvikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004 Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606. (doi:10.1016/j.devcel.2004.09.004)
116. Zhao Z, Sentürk N, Song C, Grummt I. 2018 lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. *Genes Dev.* **32**, 836–848. (doi:10.1101/gad.311688.118)
117. Blank-Giwojna A, Postepska-Igielska A, Grummt I. 2019 lncRNA KHPS1 activates a poised enhancer by triplex-dependent recruitment of epigenomic regulators. *Cell Rep.* **26**, 2904–2915. (doi:10.1016/j.celrep.2019.02.059)
118. Camsari KY, Faria R, Sutton BM, Datta S. 2017 Stochastic p-bits for invertible logic. *Phys. Rev. X* **7**, 031014. (doi:10.1103/PhysRevX.7.031014)
119. Hemphill WO, Fenske R, Gooding AR, Cech TR. 2023 PRC2 direct transfer from G-quadruplex RNA to dsDNA has implications for RNA-binding chromatin modifiers. *Proc. Natl Acad. Sci. USA* **120**, e2220528120. (doi:10.1073/pnas.2220528120)
120. Hänsel-Hertsch R *et al.* 2016 G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272. (doi:10.1038/ng.3662)
121. Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, Przytycka TM, Levens D. 2013 Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* **20**, 396–403. (doi:10.1038/nsmb.2517)

122. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016 Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049. (doi:10.1016/j.celrep.2016.04.085)
123. Jovin TM, Soumpasis DM, McIntosh LP. 1987 The transition between B-DNA and Z-DNA. *Annu. Rev. Phys. Chem.* **38**, 521–558. (doi:10.1146/annurev.pc.38.100187.002513)
124. You H, Lattmann S, Rhodes D, Yan J. 2017 RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res.* **45**, 206–214. (doi:10.1093/nar/gkw881)
125. Turing AM. 1937 On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **s2-42**, 230–265. (doi:10.1112/plms/s2-42.1.230)
126. Grabski DF, Broseus L, Kumari B, Rekosh D, Hammarskjöld ML, Ritchie W. 2021 Intron retention and its impact on gene expression and protein diversity: a review and a practical guide. *Wiley Interdiscip. Rev. RNA* **12**, e1631. (doi:10.1002/wrna.1631)
127. Powers KT, Szeto JYA, Schaffitzel C. 2020 New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes. *Curr. Opin. Struct. Biol.* **65**, 110–118. (doi:10.1016/j.sbi.2020.06.011)
128. Kurosaki T, Popp MW, Maquat LE. 2019 Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **20**, 406–420. (doi:10.1038/s41580-019-0126-2)
129. Wang X, Ramat A, Simonelig M, Liu MF. 2022 Emerging roles and functional mechanisms of PIWI-interacting RNAs. *Nat. Rev. Mol. Cell Biol.* **24**, 123–141. (doi:10.1038/s41580-022-00528-0)
130. Watanabe T *et al.* 2008 Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539–543. (doi:10.1038/nature06908)
131. Wei G, Almeida M, Pintacuda G, Coker H, Bowness JS, Ule J, Brockdorff N. 2021 Acute depletion of METTL3 implicates *N*⁶-methyladenosine in alternative intron/exon inclusion in the nascent transcriptome. *Genome Res* **31**, 1395–1408. (doi:10.1101/gr.271635.120)
132. Zaccara S, Jaffrey SR. 2020 A unified model for the function of YTHDF proteins in regulating m⁶A-modified mRNA. *Cell* **181**, 1582–1595. (doi:10.1016/j.cell.2020.05.012)
133. Sato K, Knipscheer P. 2023 G-quadruplex resolution: from molecular mechanisms to physiological relevance. *DNA Repair* **130**, 103552. (doi:10.1016/j.dnarep.2023.103552)
134. Karnkowska A *et al.* 2016 A eukaryote without a mitochondrial organelle. *Curr. Biol.* **26**, 1274–1284. (doi:10.1016/j.cub.2016.03.053)
135. Waddington CH. 1942 Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565. (doi:10.1038/150563a0)
136. Wiktor-Brown DM, Hendricks CA, Olipitz W, Engelward BP. 2006 Age-dependent accumulation of recombinant cells in the mouse pancreas revealed by *in situ* fluorescence imaging. *Proc. Natl Acad. Sci. USA* **103**, 11862–11867. (doi:10.1073/pnas.0604943103)
137. Dawkins R. 2006 *The selfish gene*. 30th anniversary edn. Oxford, UK: Oxford University Press.
138. Herbert A. 2024 Osteogenesis imperfecta type 10 and the cellular scaffolds underlying common immunological diseases. *Genes Immun.* **25**, 265–276. (doi:10.1038/s41435-024-00277-4)
139. Doluca O, Withers JM, Filichev VV. 2013 Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.* **113**, 3044–3083. (doi:10.1021/cr300225q)