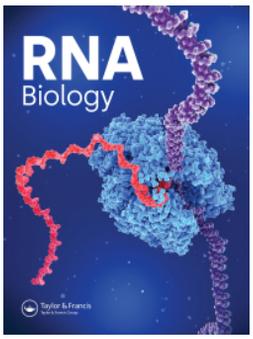

RNA Biology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/krnb20

Nucleic acid strand displacement – from DNA nanotechnology to translational regulation

Friedrich C. Simmel

To cite this article: Friedrich C. Simmel (2023) Nucleic acid strand displacement – from DNA nanotechnology to translational regulation, RNA Biology, 20:1, 154-163, DOI: 10.1080/15476286.2023.2204565

To link to this article: https://doi.org/10.1080/15476286.2023.2204565

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 24 Apr 2023.

Submit your article to this journal

Article views: 407

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=krnb20

RNA BIOLOGY
2023, VOL. 20, NO. 1, 154–163
https://doi.org/10.1080/15476286.2023.2204565

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Nucleic acid strand displacement – from DNA nanotechnology to translational regulation

Friedrich C. Simmel

TU Munich, School of Natural Sciences, Department of Bioscience, Garching, Germany

ABSTRACT

Nucleic acid strand displacement reactions involve the competition of two or more DNA or RNA strands of similar sequence for binding to a complementary strand, and facilitate the isothermal replacement of an incumbent strand by an invader. The process can be biased by augmenting the duplex comprising the incumbent with a single-stranded extension, which can act as a toehold for a complementary invader. The toehold gives the invader a thermodynamic advantage over the incumbent, and can be programmed as a unique label to activate a specific strand displacement process. Toehold-mediated strand displacement processes have been extensively utilized for the operation of DNA-based molecular machines and devices as well as for the design of DNA-based chemical reaction networks. More recently, principles developed initially in the context of DNA nanotechnology have been applied for the de novo design of gene regulatory switches that can operate inside living cells. The article specifically focuses on the design of RNA-based translational regulators termed toehold switches. Toehold switches utilize toehold-mediated strand invasion to either activate or repress translation of an mRNA in response to the binding of a trigger RNA molecule. The basic operation principles of toehold switches will be discussed as well as their applications in sensing and biocomputing. Finally, strategies for their optimization will be described as well as challenges for their operation in vivo.

ARTICLE HISTORY
Revised 19 March 2023
Accepted 13 April 2023

KEYWORDS
Strand displacement; riboregulators; RNA synthetic biology

1. Introduction

Nucleic acid strand displacement processes occur whenever several DNA or RNA molecules with the same or similar sequences compete for binding to another sequence, which is partly or fully complementary. In such a situation, one strand may displace the other, or vice versa, in a process termed 'branch migration' (Figure 1a). Strand displacement processes occur in various situations in biology, e.g. during homologous recombination, but have recently also been extensively utilized in DNA nanotechnology for the realization of synthetic molecular switches and devices [1]. Notably, the principles developed in this context have been increasingly applied also for the realization of gene regulatory switches that are operated in vivo. In this review, a brief discussion of biological and biophysical aspects of strand displacement will be given, followed by an overview of the use of SD processes in nucleic nanotechnology at its intersection with synthetic biology. Here, we will focus mainly on riboregulators for the control of translation. We conclude the article with a discussion of some of the challenges associated with the utilization of nucleic acid interactions *in vivo*.

2. Strand displacement processes in biology and nanotechnology

Strand displacement processes have been studied in a variety of biological contexts. Early studies of the process were conducted with the double-stranded genomes of bacteriophages [2,3]. Electron-microscopic studies revealed branched structures, which were interpreted as intermediates of reassociation or recombination reactions involving more than two DNA strands bound to each other (cf. Figure 2a). The kinetics of the displacement process was later studied by various researchers and found to be consistent with a random walk of the branch point along the length of the participating DNA molecules [6]. Notably, compared to three-way branch migration (two strands competing for one complementary strand), four-way branch migration (two homologous duplexes 'exchanging' strands) was found to be considerably slower and sensitive to the concentration of magnesium ions [7].

Apart from strand displacement involving 'naked' RNA or DNA molecules, many biological strand displacement or invasion processes involve the action of dedicated proteins. For instance, homologous DNA recombination events – occuring in DNA repair or during meiosis – are supported by proteins such as RecA in bacteria [8] or Rad51 and Dmc1 in eukaryotes [9].

RecA forms a helical nucleoprotein filament with ssDNA, which can invade a double-stranded DNA containing a homologous sequence (Figure 2b). Upon ATP consumption, the homologous strand is exchanged by the nucleoprotein filament, which is more tightly bound to the complementary strand. One of the challenges for the invasion of genomic double-stranded DNA is the search for a homologous sequence, which is thought to proceed in several steps [10]. During homology search, the

CONTACT Friedrich C. Simmel simmel@tum.de TU Munich, School of Natural Sciences, Department of Bioscience, Garching, Germany

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

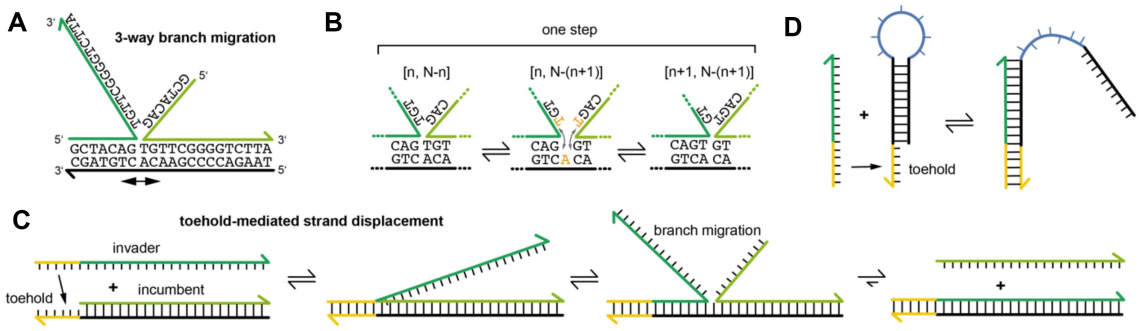

Figure 1. Nucleic acid strand displacement. a) Three-way branch migration occurs, when two identical sequences (green strands, example sequences are given for clarity) compete for binding to a complement. Shown is one of many states explored during branch migration. b) a single branch migration step consists of dissociation of one of the base-pairs at the branch point, followed by reformation of the base-pair with the alternative nucleotide on the competing strand. The process will proceed as an unbiased random walk. In the scheme, N is the total length of the strands and [n,m] denotes a state in which the strands form n and m base-pairs, respectively. c) in toehold-mediated strand displacement (TMSD), the invader contains an additional sequence complementary to the "toehold" on the incumbent-complement complex. This allows the invader to initiate a branch migration process more efficiently, ultimately resulting in displacement of the incumbent strand. d) a similar situation – relevant also to applications in riboregulation - arises when an invader attaches to and invades a nucleic acid secondary structure. Opening of the hairpin loop can be used to activate a sequence domain (blue) for binding.

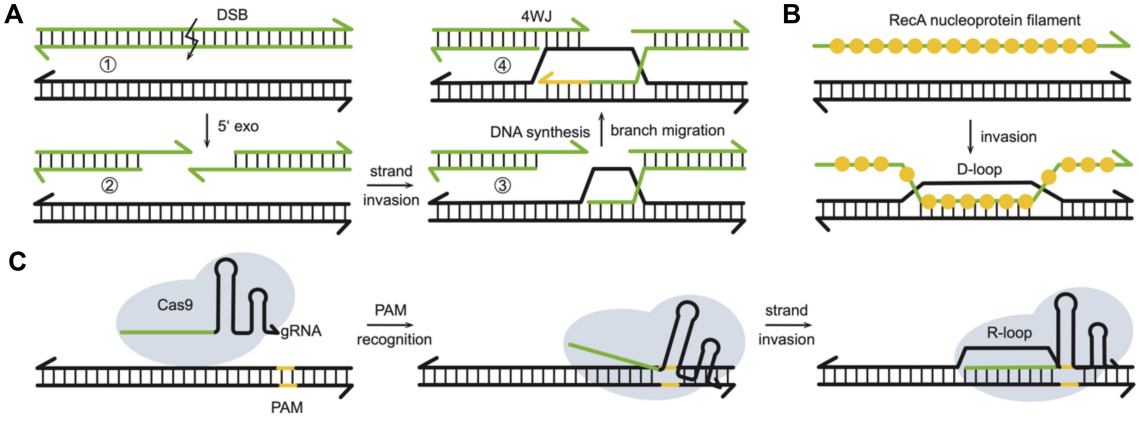

Figure 2. Biological processes involving strand displacement. a) Homologous recombination occurring after a double strand break (DSB). The broken strands (green) are trimmed at their 5′ ends with an exonuclease which allows invasion of the intact homologous duplex shown in black. Template-directed DNA polymerization is accompanied by a branch migration process (the newly synthesized sequence is shown in yellow, 4WJ = four-way junction). b) Recombination protein RecA forms a nucleoprotein filament with single-stranded DNA. RecA helps to invade a duplex with a homologous sequence, resulting in the formation of a D-loop. c) a Cas9-guide RNA complex binds to a PAM sequence on a target duplex, from where a strand invasion process is initiated. The RNA spacer sequence (green) displaces one of the DNA strands of the duplex and thereby forms an R-loop [4,5]. .

nucleoprotein filament repeatedly binds to the target dsDNA via a few initial base-pairs, until a homologous sequence region is detected and strand invasion is initiated.

Nucleic acid strand displacement also occurs during CRISPR/Cas processes [4,11]. In this case, a Cas nuclease such as Cas9 or Cas12a bound to a guide RNA (gRNA) 'scans' the double-stranded target for PAM (protospacer adjacent motif) sequences, followed by the tentative invasion of the double strand by the spacer region of the gRNA [12]. When the dsDNA happens to contain a sequence complementary to the spacer next to the PAM, the Cas/gRNA invades and tightly binds to the dsDNA upon the formation of a so-called 'R-loop' (Figure 2c).

3. Biophysical aspects

During branch migration, base-pairs repeatedly open and re-form (Figure 1b). When, driven by thermal fluctuations, one of the base-pairs at the branch point breaks, it may either immediately re-form, or the competing strand may take its chance and form an additional base-pair (while the other strand will have lost one base-pair). The free energy difference of either process is zero, and thus there is no thermodynamic bias for the strand displacement process to proceed in one direction or the other. The step times for a single branch migration step (defined as one strand losing, the other gaining one base-pair) have been measured to be approximately 10–20 μs for three-stranded branch migration [6,13], while they were≈0.3 ms for four-way branch migration in the absence of magnesium [7]. When one of the strands happens to displace the competitor strand completely, it will be fully bound to its complementary strand and thus form an ordinary duplex. The reverse process requires opening a few base-pairs at the terminus of the duplex, enabling a new strand displacement process to start.

Mathematically, the unbiased movement of the branch point can be described as a random walk process, and thus the overall time it takes for branch migration over a length of N base-pairs is expected to scale as $\sim N^2$. In the language of stochastic processes, given a complex of three strands – two competing for binding to their complement –, the time it takes for one of the strands to completely displace the other is obtained as the 'first passage time' of the corresponding random walk. In the symmetric situation, the process has two 'absorbing states', i.e. the process terminates when either of the two strands wins. The rate for strand displacement is then given as the inverse of the 'mean first passage time' (MFPT) of the process. The problem is related to the classic gambler's ruin problem [14] and has been treated, e.g. in Ref [15].

Researchers in DNA nanotechnology regularly utilize a simple trick to bias the strand displacement process: addition of a short single-stranded overhang to a duplex structure allows an 'invader strand' containing a sequence complementary to this overhang to attach to this so-called 'toehold' segment and start to invade the duplex from there (Figure 1c) [16,17]. The toehold gives the invader an advantage over the initially bound 'incumbent' strand. Thermodynamically, the invader can simply make more base-pairs with the complement than the incumbent and thus bind more stably. Kinetically, the process is biased as the incumbent can never completely displace the invader. For long enough toeholds the strand displacement process thus proceeds as a random walk on a domain with a reflecting and an adsorbing boundary. When the random walk returns to the starting point with the invader stably bound to the toehold, the invader can start another attempt to displace the incumbent. For shorter toeholds, it is also possible for the invader to spontaneously dissociate from the toehold. The overall speed of the strand displacement process is thus influenced by the stability (or length) of the toehold [17–19]. Under ideal conditions in vitro, TMSD kinetics has been found to depend exponentially on the toehold length for toeholds shorter than 8 nt, and remain constant for longer toeholds [18]. Empirically, researchers found a different behaviour in vivo, where typically longer toeholds are required to speed up strand displacement processes (as discussed further below) [20,21].

Rather than assuming uniform step times during TMSD, one can model the underlying process in greater detail by taking sequence-specific information into account. Each individual step has to fulfil detailed balance, and thus the ratio of opening and closing of a base-pair is given by $k_+/k_- = exp(-\Delta G/k_B T)$, where ΔG is the difference in free energy of the corresponding states. Based on these individual rates, the branch migration process can be simulated using a stochastic modelling approach such as the Gillespie algorithm [22], which has been implemented, e.g. in the program Multistrand [23]. Potentially, the branch migration process is also influenced by the detailed molecular structure at the branch point, e.g. via steric interactions between the invader and incumbent, which can be modelled using molecular dynamics programs such as OxDNA [19,24]. Using a combination of experimental and computational studies, also the influence of mismatches on the kinetics was investigated [25,26], and also the kinetics of RNA strand displacement was modelled [27].

There are many variations of the toehold principle, which result in processes with slightly different kinetics. For instance, so-called remote toeholds – resulting in lower effective concentrations of the invader sequence for the initiation of branch migration – can be used to further tune the kinetics of the TMSD process [28]. Further, invasion of a hairpin stem (or some other secondary structure formed by a single nucleic acid strand) differs from the displacement of an incumbent strand from a duplex, as in the case of the hairpin the displaced sequence stays bound to the complex (Figure 1d) and can thus re-invade the stem, making the process less efficient.

4. DNA-based molecular devices

Using the toehold-mediated strand displacement mechanism, a wide range of DNA-based molecular switches and devices have been developed over the past two decades (reviewed in [1]). A molecular switch is characterized by two or more distinct low free energy conformations, which are separated by a sufficiently high activation barrier (otherwise the conformations would too quickly interconvert in one another), and a mechanism to switch between these conformations. Such a mechanism typically lowers the free energy of the desired conformation with respect to the others and potentially also reduces the activation barrier for the transition.

In the case of DNA, many molecular switches are based on the fact that single-stranded DNA is much more flexible than double-stranded DNA. Hence, already the switching of a DNA molecule between duplex and single-stranded form is associated with a change in mechanical properties, and also geometrical size. More complex devices can be created by defining rigid mechanical elements from double-strands (or even DNA origami [29]), and flexible, hinge-like connections from single-strands. Inducing movements in such devices often involves the hybridization with 'fuel' or 'effector' strands. In order to revert the movements generated by these effectors, they have to be removed again, which can be accomplished by toehold-mediated strand displacement. A prototypical example for a DNA-driven DNA device is given by the DNA tweezers [16], which were composed of two rigid DNA arms connected by a flexible hinge. The DNA arms can be pulled together ('closed') by hybridization with a fuel strand, and opened by removal of the fuel strand by an anti-fuel strand (using, of course, TMSD).

In its simplest form, the cyclic operation of DNA nanodevices requires the manual addition of fuel and anti-fuel strands. Other devices have been operated in the context of chemical reaction cycles, e.g. using RNA fuels generated via transcription from a corresponding template, and degradation of the fuels using RNase H [30]. For instance, an RNA fuel can be used to remove a DNA strand from a DNA-based device via TMSD, resulting in a conformational change of the device and the concomitant formation of an RNA-DNA waste duplex. RNase H specifically degrades RNA in such hybrid duplexes via endonucleolytic cleavage, which releases the DNA strand and enables its re-binding to the DNA device [31].

Next to their role as chemical fuels, an important aspect of nucleic acid effectors and TMSD is that they also enable sequence-specific control over molecular devices. Toehold sequences can act as unique molecular 'addresses' for DNA devices. In addition, sequestration or exposure of toeholds can be utilized to temporally and conditionally control the movement of DNA devices [32,33].

Motivated by the goal to generate continuously running, autonomous molecular devices, ingenious concepts have been developed that involve the hybridization of so-called hairpin-fuels [34]. Sequences can be sequestered in the stem or loop of a hairpin structure, which makes them inaccessible for hybridization to their complementary sequences. However, opening such a hairpin structure using a toehold-mediated strand invasion process will expose these sequences and thus allow hybridization to proceed. In this way, DNA hairpins can be designed that are almost completely complementary to each other, but unable to react due to kinetic inhibition. Addition of a short invader strand, however, can promote hybridization between the hairpins. Similar systems can be designed, in which an invader triggers whole reaction cascades, which can be utilized as amplification or polymerization reactions. A wide variety of such catalytic reaction cycles have been developed for biosensing, for instance the so-called hybridization chain reaction (HCR) [35] and catalysed hairpin assembly (CHA) [33,36].

Notably, the basic concept of hiding sequences inside hairpin structures, and their activation via strand invasion has inspired the realization of a variety of synthetic RNA-based translational regulators that were shown to be functional also in a cellular context.

5. Translational regulation using strand displacement

Biological systems utilize various types of RNA molecules and RNA-RNA interactions for the control or modulation of gene expression processes. In particular, riboswitches, mainly found in bacteria, are an inspiring natural example of how RNA conformational changes can be used in this context [37,38]. Post-transcriptional riboswitches – regulating translation – are comprised of complex RNA secondary structures in the 5' untranslated region (UTR) of mRNA molecules that contain a ribosome binding site (RBS) and can change their conformation upon binding of a small molecule (typically, a metabolite such as guanine, adenine, vitamin B12, theophylline, etc.). In riboswitches, the so-called 'recognition domain' contains an RNA aptamer region for binding the metabolite. Binding of the ligand leads to a conformational change that either sequesters a previously accessible RBS (resulting in a translational OFF switch), or releases it from a secondary structure (in the case of an ON switch) [39].

5.1. Toehold switch activators and repressors

Rather than switching riboregulators via small molecules as in riboswitches, conformational changes can also be brought about by RNA hybridization. One of the first examples of a synthetic translational regulator based on RNA-RNA interactions was created by putting a designed cis-repressing RNA hairpin motif into the 5' UTR of an mRNA, sequestering the RBS and thus inhibiting translation [40]. Translation could be activated by the expression of a short RNA activator molecule that could open the repressing RNA hairpin, making the RBS accessible to the ribosomes and thus enabling translation of the downstream sequence (coding for GFP).

Using the principle of toehold-mediated strand invasion, an improved version of such a synthetic riboregulator termed 'toehold switch' was demonstrated later [21]. In the OFF state of a translational toehold activator, the RBS is initially sequestered in a hairpin loop, while the translation start codon (AUG) resides within a bulge in the stem of the hairpin (Figure 3a). The hairpin is extended at its 5' end with a 12 nt long unstructured sequence, which acts as a toehold for a trigger RNA molecule. Upon invasion of the toehold switch stem, the hairpin is opened and both RBS and start codon are released. The specific design of the toehold switch allows to circumvent any sequence constraints for the trigger RNA, and thus riboregulators can be designed to respond to arbitrary input sequences, at least in principle.

The toehold switch principle can also be used to realize translational repressors rather than activators [44]. In this case, in the absence of the trigger, the RBS is initially accessible to the ribosome. Toehold-mediated strand invasion by the trigger RNA leads to a conformational change, which sequesters the RBS in a stable hairpin, and thus represses translation. Alternatively, the hairpin containing the RBS can be made weak enough (and thus thermally fluctuating) to allow for binding of ribosomes. Stabilizing the hairpin by locking it with a trigger switches translation OFF (Figure 3c).

RNA switchable riboregulators have been realized that more closely resemble the structure of naturally occurring translational riboswitches (Figure 3b). In natural riboswitches, the RBS is often found sequestered within a duplex stem by an anti-RBS sequence. Using a similar principle as for the toehold switches, an anti-anti-RBS sequence can be placed inside the loop of a toehold hairpin upstream of the anti-RBS-RBS duplex. Upon activation by a trigger RNA, the toehold hairpin unfolds and thus allows the anti-anti-RBS to bind to the anti-RBS. This in turn releases the RBS and activates translation [41]. A similar strategy can be used to generate translational toehold repressors.

5.2. Applications as sensors

Toehold switches and related riboregulators have been successfully utilized as biosensors, as the free choice of the input sequence allows to adapt the switches to essentially arbitrary RNA analytes. In several examples, toehold switches were applied for the realization of paper-based biosensors for Ebola [45] or Zika [46], for which a cell-free expression system containing toehold switches was freeze-dried on a paper strip. The analyte RNA (the viral genome) was first amplified using a nucleic acid amplification method such as NASBA and then added to the re-hydrated paper strips (Figure 4a). The advantage of this workflow is that cell-free expression of a fluorescent protein provides an additional

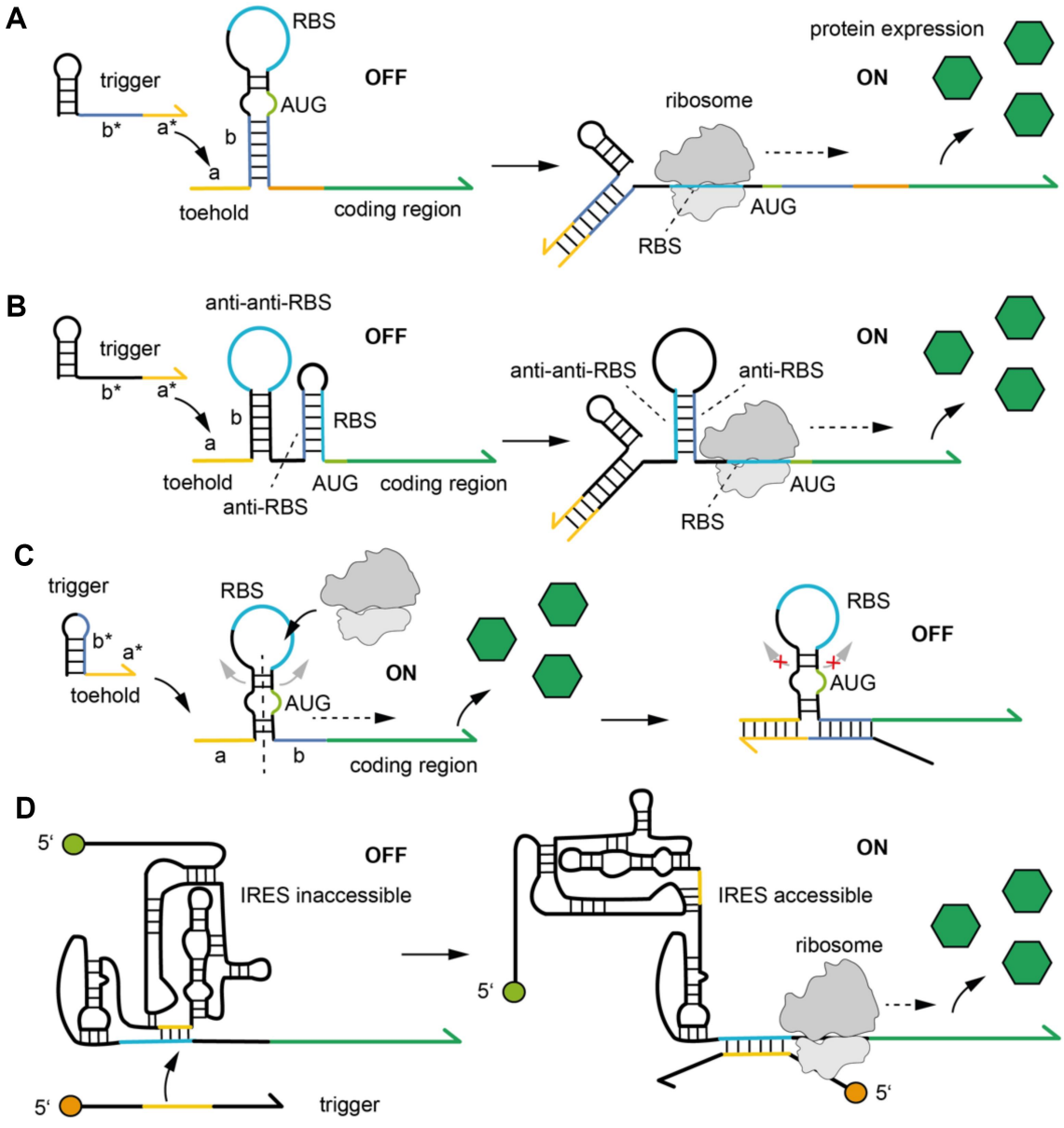

Figure 3. Translational toehold switches. a) in a toehold switch, the ribosome binding site (RBS) and translation start codon (AUG) are sequestered in a stable secondary structure and are therefore not accessible for a ribosome. Addition of a trigger RNA results in toehold-mediated strand invasion and opening of the toehold hairpin, which in turn activates translation of the downstream coding region. The hairpin at the 5′ end of the trigger is a measure against RNA degradation [21]. b) Toehold switches can also be created based on the architecture of translational riboswitches. Here the toehold hairpin sequesters an anti-anti-RBS, while the RBS is initially bound to an anti-RBS. Trigger RNA can induce refolding of the structure, which releases the RBS and thus activates translation [41]. c) a translational off switch can be constructed by sequestering the RBS with the help of a trigger. In this case, the hairpin containing the RBS loop is designed to be weak enough to be invaded by the ribosome (opening of the hairpin is indicated by the gray arrows). Trigger RNA locks the hairpin with the RBS and thus switches translation off [42]. d) Eukaryotic mRnas are capped at the 5′ end and do not have an equivalent of the ribosome binding site. Shown is an example of a eukaryotic toehold switch that is based on the accessibility of an internal ribosome entry site (IRES) on an mRNA, which was derived from viral RNA [43]. .

amplification of the signal, and is potentially more cost-effective than the use of fluorescently labelled sensors.

More recent examples for low-cost diagnostics based on toehold switches comprise the development of electrochemical and bioelectronic sensors. For electrochemical transduction, toehold switches were designed that coded for restriction nucleases rather than GFP [48]. Upon detection of RNA analytes (amplified RNA of a bacterial antibiotic resistance gene), different types of nuclease were expressed, which cleaved off electrochemical labels from DNA transducer molecules that were immobilized on a detector gold electrode. A different principle involved the expression of a protease in response to amplified SARS-CoV-2 RNA [49]. The protease was used to degrade a gelatin layer in a bioelectronic sensor based on an LC circuit, resulting in a measurable change in its resonance frequency.

Finally, a variant of the toehold switches termed SNIPRs (single-nucleotide-specific programmable riboregulators) was demonstrated to detect single nucleotide variations in the input sequence [50]. The SNIPR principle utilizes the fact that toehold-mediated strand displacement is extremely slowed down when the invading strand is mismatched. Single nucleotide sensitivity was achieved by careful tuning of the

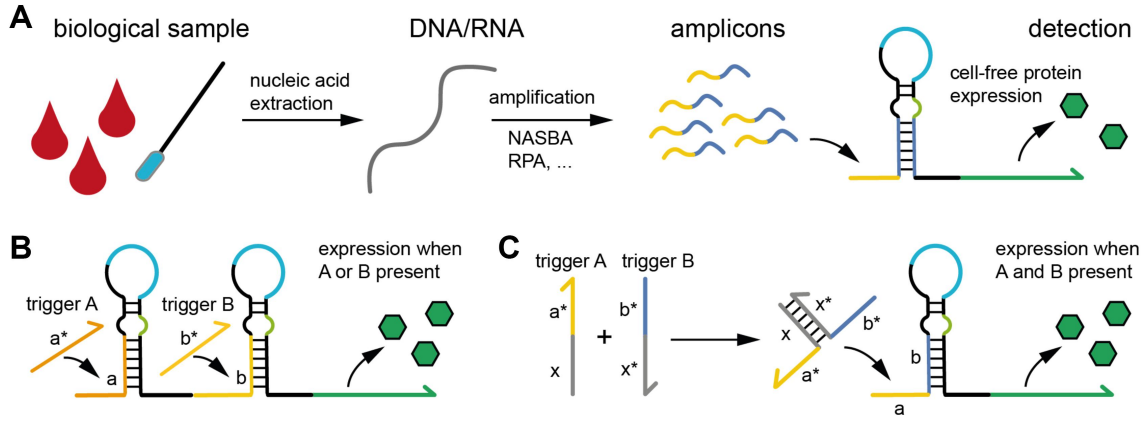

Figure 4. Applications in sensing and biocomputing. a) Toehold switches have been used in various biosensing applications. For nucleic acid sensing, the analyte contained in the initial sample typically has to be amplified first, the toehold switch then serves as a sequence-specific sensor for the amplicons. The toehold switch can be operated in a cell-free expression system, which can be utilized for low cost sensors [45,46]. b) Several toehold switches can be combined in series, in parallel, or in more complex configurations. Shown is an or gate based on two switches in series. In the presence of trigger A, the first RBS becomes available and the ribosome then opens the second hairpin [47]. c) a trigger for a toehold switch can be split into two parts (trigger a and B), resulting in an and gate configuration. As TMSD requires both a sequence complementary to the toehold and to the sequence to be invaded, translation is only activated in the presence of both triggers [47]. .

structure of the riboregulator and the position of the mismatch.

Compared to in vitro detection of nucleic acids, which often involve several signal amplification steps, detection of endogenous RNA *in vivo* is a major challenge. Typically, endogenous transcripts are present at only small copy numbers. RNA hybridization reactions at the correspondingly low concentrations are rather slow and compete with RNA degradation processes. Further, in the cellular context, RNA may interact with other RNA molecules or proteins, which further influences hybridization kinetics and RNA stability. Accordingly, toehold switches were so far only shown to detect mRNA molecules transcribed from high copy number plasmids in vivo (resulting in rather high mRNA concentrations). Notably, however, the detection of an endogenous small RNA (sRNA) by a toehold switch was successfully demonstrated for RyhB, which is an sRNA whose expression in *E.coli* depends on the intracellular concentration of iron. Influencing [Fe^{2+}] with an iron chelator could be shown to result in a measurable change in GFP expression from a toehold switch triggered by RyhB [21].

As indicated, there are two main future directions for applications of translational riboregulators in biosensing: low cost *in vitro* biosensors, and *in vivo* nucleic acid detection. Low-cost sensors leverage the potential for isothermal signal amplification in cell-free expression systems that is based on transcription and translation reactions, and the properties of in situ expressed enzymes and fluorescent proteins. As such sensor schemes compete against a variety of more established detection methods – most notably PCR-based approaches – it will be crucial to demonstrate their superiority in relevant use cases – e.g. in terms of costs, quick availability, engineerability, simplicity or robustness – and convince practitioners to actually adopt these schemes.

Successful *in vivo* nucleic acid detection would open up completely new possibilities for in vivo diagnostics as well as for synthetic biology applications. In a diagnostic context, detection of disease-associated endogenous mRNA, long non-coding RNA (lncRNA) or micro RNA (miRNA) by eukaryotic translational regulators could be coupled to the production of appropriate reporter or effector proteins. The ability to detect endogenous mRNA levels would be of great use to determine specific transcriptomic cell states or cell state transitions, and use these as inputs for downstream gene regulatory circuitry.

5.3. Logic gates and circuits

RNA-based regulatory mechanisms are particularly interesting for biocomputing applications. As RNA secondary structures and RNA hybridization are governed by sequence-programmable RNA interactions, they allow for a relatively straightforward 'rational' design of biocomputing components and circuits. It has been shown that several toehold switches can be simply connected in series to allow for the simultaneous regulation of the expression of several genes [21].

Alternatively, the same gene can be controlled by putting several toehold switches in series (on a single mRNA transcript) resulting in multi-input OR logic (Figure 4b) [42,47]. In the case of such a serial connection, ribosomes binding to any of the upstream RBS sequences have to be able to open all downstream secondary structure for the circuit to function. AND gate behaviour can be realized by utilizing composite trigger RNA molecules that are assembled from multiple input strands, which relies on the fact that for TMSD an invader has to be complementary both to the toehold and to the sequence bound by the incumbent (Figure 4c) [47]. Thus, input molecules containing either only the toehold or the incumbent sequence cannot initiate TMSD. Accordingly, assembling the RNA inputs into a complex that contains all the required sequence domains will result in an active trigger. Logical negation (often termed NOT gate) can be achieved by utilizing (excessive) anti-trigger RNAs that are completely complementary to the input triggers. Duplex formation between triggers and anti-triggers is favoured over binding to a toehold switch, thus preventing translational activation (or repression) by the trigger. Using a combination of

repressors, activators, AND, OR and NOT logic, more complex logical expressions can be realized, that control gene expression depending on the presence of a specific input pattern [42,47,51].

A different approach to realize AND and OR logic functions in the input domain of translational riboregulators is based on multi-arm junctions [52] and was recently demonstrated for applications in cell-free diagnostics. In contrast to toehold switches, strand invasion is initiated within the loop region of a hairpin loop (termed LIRA for loop-initiated RNA activator).

Rather than only utilizing translational riboregulators, a combination with transcriptional switching (via modulation of transcriptional termination) can also be used to realize logic functions. For instance, combining small transcription activators (STARS) [53] with toehold switches was used to realize effective AND gates for the control of gene expression [54]. Similarly, rho-dependent termination/anti-termination was combined with a riboswitch-inspired toehold repressor to realize NOR logic [41].

5.4. Toehold switching in eukaryotes

Due to the manifold differences between prokaryotic and eukaryotic gene expression, the realization of eukaryotic translational switches requires a different approach than that outlined above. In eukaryotes, mRNA is transcribed by RNA polymerase II, which results in an RNA molecule with a 7-methylguanylate cap at the 5' end and a poly A tail at the 3' end. In many cases, the tail is linked to the cap via the poly A binding protein PABP. Translation is initiated in the cap region, but there is no dedicated ribosome binding site, which could be utilized for the construction of a toehold switch analogous to the prokaryotic case. The so-called Kozak sequence [55] found close to the start codon of eukaryotic mRNAs is important, but not critical for translation initiation, and therefore switches based on the Kozak sequences show low performance. An alternative possibility for the realization of eukaryotic toehold switches is offered by cap-independent initiation processes, which start at secondary structural motifs such as the IRES (internal ribosome entry site). IRES modules are found in viral RNA molecules and are used to recruit the eukaryotic ribosome for translation. In an approach termed eToehold, indeed successful translational switching was demonstrated based on IRES modules from cricket paralysis virus, kashmir bee virus, and acute bee paralysis virus (Figure 3d) [43]. When T7 RNA polymerase is used to transcribe mRNA in eukaryotes it will not be translated due to the lack of a 5' cap, but inclusion of an IRES facilitates translation from such transcripts. In the eToehold concept, switching the IRES structure with an RNA trigger was successfully applied for the detection of viral RNA or endogenous mRNAs in eukaryotic cells.

6. Design and optimization

Among the major potential advantages of using RNA-based regulation mechanisms for synthetic biology are the wide availability of tools for the design of RNA sequences [56,57], the prediction of RNA secondary structures and thermodynamics [56,58,59], and the (superficially) straightforward sequence–function relationship. Furthermore, RNA-based components can be combined and connected into more complex systems or circuits based on sequence-programmable RNA–RNA interactions. Despite these features, rational design of riboregulators does not always yield the desired results and often has to be complemented by screening and machine-learning approaches – or simple trial and error.

6.1. Rational design, screening & deep learning

The first generation of toehold switches was designed based on the general concept of hiding an RBS within a hairpin loop and the start codon in a bulge in the hairpin stem (see above). A set of 168 toehold switches exhibited ON/OFF ratios for gene expression ranging from below 10 up to 300 [21]. Thermodynamic analysis of the structures was used to derive a set of predictors based on which the toehold switches could be further improved by forward-design – e.g. using a longer toehold, stem, or loop region. Similar as for riboswitches, achieving high ON/OFF ratios requires translational riboregulators to be tight in the OFF state (low leak activity), but not 'too tight' to maintain switchability [60]. One potential issue is the binding of ribosomes to the ribosome binding site – if the toehold hairpin stem is too weak, the hairpin loop (containing the RBS) might be invaded by the ribosome (and thus a translational activator is always ON). If it is too strong, strand invasion by the trigger RNA might be ineffective and the switch always remains in its OFF state). Other factors play a role such RNA stability and interactions with RNA binding proteins that are partly unknown.

In a cellular context, some of the constraints for the operation of synthetic components are difficult to account for by software-supported rational design, in which case experimenters often have to resort to screening approaches. Importantly, in the case of nucleic acid-based components this can be achieved by screening of a corresponding sequence library. Well-performing library members can be extracted by cell or droplet sorting and then identified by sequencing.

Based on the analysis of such a library, a machine-learning pipeline was established and successfully used to increase the dynamic range of tetracycline riboswitches from 8.5 to 40-fold [61]. In the case of translational toehold switches, a library with more than 90,000 members was generated, which contained sequence elements from viruses, the human genome, and also random sequences [62,63]. The library was experimentally characterized for their ON/OFF switch behaviour and the results were used to feed different types of deep-learning models (convolutional neural networks, but also a language model based on a recurrent neural network) to identify features that are associated with good or bad performance. The results indicated nucleotide preferences at certain positions (e.g. associated with sequence constraints such as the absence of stop codons), but also showed that biophysical properties alone were not capable of sufficiently predicting the performance of the toehold switches. Optimization of toehold switch sequences using the machine-learning models resulted

in improved ON/OFF ratios in the majority of the sequences tested.

6.2. Other challenges for in vivo operation

The rational design of *in vivo* RNA riboregulators is hampered, among others, by unknown interactions of RNA with cellular components such as RNA-binding proteins or other RNA molecules, and potential deviations between the predicted structure and the structure assumed in the cellular milieu. Techniques such as SHAPE-seq have been used to study the in vivo structure of RNA switches, and demonstrated deviations from the predicted minimum free energy structure [64,65]. As indicated above, some of these unknowns can potentially be accounted for when using machine learning approaches for the design of RNA regulators.

Another major problem for the in vivo operation of RNA-based regulators is the efficacy of RNA hybridization and TMSD reactions themselves. As a second-order reaction, binding of an input RNA to a toehold depends on the concentration of the species and the accessibility of the sequences. In order to improve translational switching based on RNA hybridization, one can attempt to increase the cellular concentrations of the involved RNA species by increasing their expression level or decreasing their degradation. For instance, in vivo stability of RNA species can be improved by adding RNA hairpins to 5' and 3' ends, or utilizing other types of secondary structure [66–68] or RNA-binding proteins such as Hfq [69] that protect from degradation by nucleases. Potentially, co-localization of RNA molecules can be used to increase local concentration and thus promote interaction. Interactions of non-coding RNA with nascent mRNA molecules appear to play role in various regulatory contexts and can alter the RNA structure during transcription [70,71].

Finally, it is known that the accessibility of a toehold sequence has a major influence on the kinetics of TMSD reactions. While in an *in vitro* context, toeholds are usually designed to be completely secondary structure free, this may actually be detrimental in a more complex environment as unstructured RNAs are more prone to degradation and interaction with non-cognate RNA molecules. It has been shown in an in vitro context that undesired interactions with a random sequence background can be prevented by using three-letter codes or designed secondary structures [72]. Furthermore, naturally occurring RNA molecules may be highly structured and bound to RNA-binding proteins, and the accessibility of sequences for hybridization may thus vary strongly. For instance, RNA accessibility has been systematically studied for sRNA-mRNA interactions in bacteria using an RNA-seq-based high throughput method [73].

7. Conclusion

Nucleic acid strand displacement reactions have made an interesting career over the past decades. Initially of interest in basic molecular biology studies of DNA hybridization and recombination events, the invention of the 'toehold' concept opened up a completely new field of application in the context of DNA-based molecular machines and devices. Further developments in this field were later 're-imported' into life science research and used for the realization of biosensors and amplification schemes, and more recently also for the implementation of synthetic gene regulatory mechanisms.

Synthetic translational regulators such as the so-called 'toehold switches' are based on an RNA conformational change that is triggered by an RNA input molecule via toehold-mediated strand invasion. Similar as in riboswitches, this conformational change is designed to expose or hide a critical nucleotide sequence such as a ribosome binding site. Rational design of these structures – sometimes combined with screening and recently also machine learning – has led to the development of translational switches with extraordinary dynamic ranges, and these devices have already found applications as low-cost biosensors for RNA and other analytes, and also as components for in vivo computing systems. A future challenge will be the development of specific use cases that will lead to a more widespread application in practice.

Despite the spectacular success of the application of the toehold concept in this context, there are still many challenges to overcome, in particular when endogenous RNA molecules are used as triggers inside live cells, e.g. due to rapid RNA degradation and correspondingly low analyte concentrations. In order to realize even better performing in vivo devices, ultimately a better understanding and control of RNA stability, RNA localization, RNA–RNA and RNA–protein interactions in the complex cellular environment will be required.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft through grant no. DFG SI 761/5-1. The author gratefully acknowledges funding via the Excellence Strategy of the Federal Government and the Länder through the TUM Innovation Network "RISE".

Data availability statement

This review article does not contain any data.

ORCID

Friedrich C. Simmel http://orcid.org/0000-0003-3829-3446

References

[1] Simmel FC, Yurke B, Singh HR. Principles and applications of nucleic acid strand displacement reactions. Chem Rev 2019 May 22;119(10):6326–6369.
[2] Lee CS, Davis RW, Davidson N. A physical study by electron microscopy of the terminally repetitious, circularly permuted DNA from the coliphage particles of Escherichia coli 15. J Mol Biol. 1970;48(1):1–22.
[3] Broker TR, Lehman IR. Branched Dna molecules - intermediates in T4 recombination. J Mol Biol. 1971;60(1):131- &.
[4] Klein M, Eslami-Mossallam B, Arroyo DG, et al. Hybridization kinetics explains CRISPR-Cas off-targeting rules. Cell Rep. 2018;22(6):1413–1423. DOI:10.1016/j.celrep.2018.01.045
[5] Rutkauskas M, Songailiene I, Irmisch P, et al. A quantitative model for the dynamics of target recognition and off-target

rejection by the CRISPR-Cas Cascade complex. Nat Commun. 2022 Dec 3;13(1):7460.
[6] Green C, Tibbetts C. Reassociation rate limited displacement of DNA-Strands by branch migration. Nucleic Acids Res. 1981;9 (8):1905–1918.
[7] Panyutin I, Hsieh P. The kinetics of spontaneous DNA branch migration. Proc Natl Acad Sci USA. 1994;91(6):2021–2025.
[8] Cox MM. Motoring along with the bacterial RecA protein. Nat Rev Mol Cell Bio. 2007;8(2):127–138.
[9] Murayama Y, Kurokawa Y, Mayanagi K, et al. Formation and branch migration Of holliday junctions mediated by eukaryotic recombinases. Nature. 2008;451(7181):1018–1021. DOI:10.1038/nature06609
[10] Neale MJ, Keeney S. Clarifying the mechanics of DNA strand exchange in meiotic recombination. Nature. 2006;442(7099):153–158.
[11] Szczelkun MD, Tikhomirova MS, Sinkunas T, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA. 2014;111 (27):9798–9803. DOI:10.1073/pnas.1402597111
[12] Jones DL, Leroy P, Unoson C, et al. Kinetics of dCas9 target search in Escherichia coli. Science. 2017;357(6358):1420–1424. DOI:10.1126/science.aah7084
[13] Radding CM, Beattie KL, Holloman WK, et al. Uptake of homologous single-stranded fragments by superhelical DNA. IV. branch migration. J Mol Biol. 1977;116(4):825–839. DOI:10.1016/0022-2836(77)90273-X
[14] Feller. An introduction to probability theory and itsApplications. Vol. 1. New York: Wiley and Sons; 1968.
[15] Reynaldo LP, Vologodskii AV, Neri BP, et al. The kinetics of oligonucleotide replacements11Edited by I. tinoco. J Mol Biol. 2000;297(2):511–520. DOI:10.1006/jmbi.2000.3573
[16] Yurke B, Turberfield AJ, Mills AP, et al. A DNA-fuelled molecular machine made of DNA. Nature. 2000;406(6796):605–608. DOI:10.1038/35020524
[17] Yurke B, Mills AP. Using DNA to power nanostructures. Genet Program Evolvable Mach. 2003;4(2):111–122.
[18] Zhang DY, Winfree E. Control of DNA strand displacement kinetics using toehold exchange. J Am Chem Soc. 2009;131 (47):17303–17314.
[19] Srinivas N, Ouldridge TE, Sulc P, et al. On the biophysics and kinetics of toehold-mediated DNA strand displacement. Nucleic Acids Res. 2013 Dec;41(22):10641–10658.
[20] Choi HM, Beck VA, Pierce NA. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano 2014 May 27;8(5):4284–4294.
[21] Green AA, Silver PA, Collins JJ, et al. Toehold switches: de-novo-designed regulators of gene expression. Cell. 2014 Nov 6;159 (4):925–939.
[22] Gillespie DT. EXACT STOCHASTIC SIMULATION of COUPLED CHEMICAL-REACTIONS. J Phys Chem. 1977;81 (25):2340–2361.
[23] Schaeffer JM, Thachuk C, Winfree E, et al. Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands. In: Phillips A, and Yin P editor. DNA Computing and Molecular Programming. DNA 2015. Lecture Notes in Computer Science. Vol. 9211. Cham: Springer; 2015.
[24] Doye JPK, Ouldridge TE, Louis AA, et al. Coarse-graining DNA for simulations of DNA nanotechnology. Phys Chem Chem Phys. 2013;15(47):20395–20414. DOI:10.1039/c3cp53545b
[25] Machinek RR, Ouldridge TE, Haley NE, et al. Programmable energy landscapes for kinetic control of DNA strand displacement. Nat Commun. 2014 Nov 10;5(1):5324.
[26] Irmisch P, Ouldridge TE, Seidel R. Modeling DNA-Strand displacement reactions in the presence of base-pair mismatches. J Am Chem Soc. 2020;142(26):11451–11463.
[27] Sulc P, Ouldridge TE, Romano F, et al. Modelling toehold-mediated RNA strand displacement. Biophys J. 2015;108 (5):1238–1247. DOI:10.1016/j.bpj.2015.01.023
[28] Genot AJ, Zhang DY, Bath J, et al. Remote toehold: a mechanism for flexible control of DNA hybridization kinetics. 2011;133 (7):2177–2182. DOI:10.1021/ja1073239
[29] Dey S, Fan C, Gothelf KV, et al. DNA origami. Nat Rev Methods Primers. 2021;1(1):13. DOI:10.1038/s43586-020-00009-8
[30] Grosso ED, Franco E, Prins LJ, et al. Dissipative DNA nanotechnology. Nat Chem. 2022;14(6):600–613. DOI:10.1038/s41557-022-00957-6
[31] Franco E, Friedrichs E, Kim J, et al. Timing molecular motion and production with a synthetic transcriptional clock_SI_appendix. PNAS. 2011;108(40):1–95. DOI:10.1073/pnas.1100060108
[32] Seelig G, Soloveichik D, Zhang DY, et al. Enzyme-free nucleic acid logic circuits. Science. 2006 Dec 8;314(5805):1585–1588.
[33] Yin P, Choi HMT, Calvert CR, et al. Programming biomolecular self-assembly pathways. Nature. 2008;451(7176):318–322. DOI:10.1038/nature06451
[34] Green SJ, Lubrich D, Turberfield AJ. DNA hairpins: fuel for autonomous DNA devices. Biophys J. 2006;91(8):2966–2975.
[35] Dirks RM, Pierce NA. Triggered amplification by hybridization chain reaction. Proc Natl Acad Sci 2004 Oct 26;101 (43):15275–15278.
[36] Li B, Ellington AD, Chen X. Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods. Nucleic Acids Res. 2011;39(16):e110.
[37] Tucker BJ, Breaker RR. Riboswitches as versatile gene control elements. Curr Opin Struct Biol. 2005;15(3):342–348.
[38] Sherwood AV, Henkin TM. Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. Annu Rev Microbiol. 2016;70(1):361–374.
[39] Mandal M, Breaker RR. Gene regulation by riboswitches. Nat Rev Mol Cell Biol. 2004 Jun;5(6):451–463.
[40] Isaacs FJ, Dwyer DJ, Ding C, et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol. 2004 Jul;22(7):841–847.
[41] Wang T, Simmel FC. Riboswitch-inspired toehold riboregulators for gene regulation in Escherichia coli. Nucleic Acids Res 2022 Apr 21;50(8):4784–4798.
[42] Kim J, Zhou Y, Carlson PD, et al. De novo-designed translation-repressing Riboregulators for multi-input cellular logic. Nat Chem Biol. 2019 Dec;15(12):1173–1182.
[43] Zhao EM, Mao AS, Puig H, et al. RNA-responsive elements for eukaryotic translational control. Nature Biotechnol. 2022;40 (4):539–545. DOI:10.1038/s41587-021-01068-2
[44] Kim J, Zhou Y, Carlson P, et al. De-novo-designed translational repressors for multi-input cellular logic. biorxivorg. 2018.
[45] Pardee K, Green AA, Ferrante T, et al. Paper-based synthetic gene networks. Cell. 2014 Nov 6;159(4):940–954.
[46] Pardee K, Green AA, Takahashi MK, et al. Rapid, low-cost detection of zika virus using programmable biomolecular components. Cell. 2016;165(5):1255–1266. DOI:10.1016/j.cell.2016.04.059
[47] Green AA, Kim J, Ma D, et al. Complex cellular logic computation using ribocomputing devices. Nature. 2017;548(7665):117–121. DOI:10.1038/nature23271
[48] Mousavi PS, Smith SJ, Chen JB, et al. A multiplexed, electrochemical interface for gene-circuit-based sensors. Nat Chem. 2019;12 (1):48–55. DOI:10.1038/s41557-019-0366-y
[49] Carr AR, Dopp JL, Wu K, et al. Toward mail-in-sensors for SARS-CoV-2 detection: interfacing gel switch resonators with cell-free toehold switches. ACS Sens. 2022;7(3):806–815. DOI:10.1021/acssensors.1c02450
[50] Hong F, Ma D, Wu K, et al. Precise and programmable detection of mutations using ultraspecific riboregulators. Cell. 2020;180 (5):1018–1032.e16. DOI:10.1016/j.cell.2020.02.011
[51] Falgenhauer E, Muckl A, Schwarz-Schilling M, et al. Transcriptional interference in toehold switch-based RNA circuits. ACS Synth Biol. 2022 May 20;11(5):1735–1745.
[52] Ma D, Li Y, Wu K, et al. Multi-arm RNA junctions encoding molecular logic unconstrained by input sequence for versatile cell-free diagnostics. Nat Biomed Eng. 2022;6(3):298–309. DOI:10.1038/s41551-022-00857-7
[53] Chappell J, Takahashi MK, Lucks JB. Creating small transcription activating RNAs. Nat Chem Biol. 2015 Mar;11 (3):214–220.

[54] Lehr FX, Hanst M, Vogel M, et al. Cell-free prototyping of AND-Logic gates based on heterogeneous RNA activators. ACS Synth Biol. 2019 Sep 20;8(9):2163–2173.
[55] Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Nat Acad Sci. 1990;87(21):8301–8305.
[56] Zadeh JN, Steenberg CD, Bois JS, et al. NUPACK: analysis and design of nucleic acid systems. J Comput Chem. 2010;32 (1):170–173. DOI:10.1002/jcc.21596
[57] Fornace NJP, Niles APME, Pierce NA. A unified dynamic programming framework for the analysis of interacting nucleic acid strands: enhanced models, scalability, and speed. ACS Synth Biol. 2020;9 (10):1–14.
[58] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003 Jul 1;31(13):3406–3415.
[59] Lorenz R, Bernhart SH, Honer Z, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011 Nov 24;6(1):26.
[60] Boussebayle A, Torka D, Ollivaud S, et al. Next-level riboswitch development—implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. Nucleic Acids Res. 2019;47(9):4883–4895. DOI:10.1093/nar/gkz216
[61] Groher A-C, Jager S, Schneider C, et al. Tuning the performance of synthetic riboswitches using machine learning. ACS Synth Biol. 2019;8(1):34–44. DOI:10.1021/acssynbio.8b00207
[62] Angenent-Mari NM, Garruss AS, Soenksen LR, et al. A deep learning approach to programmable RNA switches. Nat Commun. 2020 Oct 7;11(1):5057.
[63] Valeri JA, Collins KM, Ramesh P, et al. Sequence-to-function deep learning frameworks for engineered riboregulators. Nat Commun. 2020 Oct 7;11(1):5058.
[64] Loughrey D, Watters KE, Settle AH, et al. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. Nucleic Acids Res. 2014;42(21):e165. DOI:10.1093/nar/gku909
[65] Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res. 2015;44(2):e12.
[66] Wilusz JE, JnBaptiste CK, Lu LY, et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. 2012;26(21):2392–2407. DOI:10.1101/gad.204438.112
[67] Zhang Q, Ma D, Wu F, et al. Predictable control of RNA lifetime using engineered degradation-tuning RNAs. Nat Chem Biol. 2021;17(7):828–836. DOI:10.1038/s41589-021-00816-4
[68] Cetnar DP, Salis HM. Systematic quantification of sequence and structural determinants controlling mRNA stability in bacterial operons. ACS Synth Biol. 2021;10(2):318–332.
[69] Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol 2011 Aug 15;9(8):578–589.
[70] Lee N, Moss Walter N, Yario Therese A, et al. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell. 2015;160(4):607–618. DOI:10.1016/j.cell.2015.01.015
[71] Statello L, Guo C-J, Chen L-L, et al. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Bio. 2021;22(2):96–118. DOI:10.1038/s41580-020-00315-9
[72] Mayer T, Oesinghaus L, Simmel FC. Toehold-mediated strand displacement in random sequence pools. J. Am. Chem. Soc. 2023;146:634–644 .
[73] Mihailovic MK, Vazquez-Anderson J, Li Y, et al. High-throughput in vivo mapping of RNA accessible interfaces to identify functional sRNA binding sites. Nat Commun. 2018;9(1):4084. DOI:10.1038/s41467-018-06207-z