

Julia Pitzer, Bob Van Hove, Aaron M. Love,
Parayil Kumaran Ajikumar, Marjan De Mey, and Anton Glieder

Contents

2.1	Introduction	66
2.2	Synthetic Promoters	67
2.2.1	Hybrid Promoters	68
2.2.2	Common Strategies for the Engineering of Prokaryotic Promoters	70
2.2.3	Common Methods for the Engineering of Eukaryotic Promoters	73
2.3	Terminators	77
2.4	Ribozymes	78
2.5	Riboswitches	80
2.6	Small RNAs	81
2.6.1	Detection, Prediction, and Classification of Small RNAs	81
2.6.2	Small RNA Processing Pathways	82
2.6.3	Functions and Applications of Small RNAs	84
2.7	Long Noncoding RNAs	85
2.8	Aptamers and Adaptamers	86
2.9	DNA Barcodes	88
2.10	DNA Machines	89
2.11	DNA Walker	91
	References	94

J. Pitzer • A. Glieder (✉)

Institute of Molecular Biotechnology, Graz University of Technology, Petersgasse 14, 8010 Graz,
Austria

e-mail: a.glieder@tugraz.at

B. Van Hove • M. De Mey

Centre for Industrial Biotechnology and Biocatalysis, Ghent University, Coupure Links 653,
9000 Ghent, Belgium

A.M. Love • P.K. Ajikumar

Manus Biosynthesis, 1030 Massachusetts Avenue, Suite 300, Cambridge, MA 02138, USA

Abstract

Impressive advances in the field of synthetic biology go hand in hand with the discovery, design, and use of novel DNA and RNA elements. Efficient synthesis of large oligonucleotides and double-stranded DNA parts, chip-based synthesis of DNA libraries, and a detailed understanding of fundamental biological mechanisms and increased capacities in bioinformatics enable new findings and applications.

In this chapter, reengineered and model-based designed DNA parts such as promoters, terminators, and aptamers, which can be used for controlled gene expression, are discussed. Furthermore, recent advances in the RNA world concerning small RNAs, ribozymes, and riboswitches and novel applications of synthetic nucleotide sequences in the form of DNA walkers and DNA machines are emphasized. The knowledge and experiences gained in the investigations of large numbers of natural and novel DNA and RNA parts have built the foundation for the design of new functions of polynucleotides and model-based regulation of gene and pathway expression.

2.1 Introduction

DNA and RNA elements provide common and some of the most efficient tools to regulate the expression of genes and pathways. In contrast to other factors, such as the genetic background and cell physiology of host strains, the engineering and use of DNA and RNA elements are also more simple and knowledge-based approaches are frequently applied. The use of such synthetic nucleic acid parts also facilitates approaches, which build on simple design and therefore can be made more or less free of unknown natural regulatory effects. This makes the design and generation of novel DNA and RNA parts a key step toward model-based regulation of protein expression and regulatory circuit design. The most studied DNA and RNA elements include promoters, ribosome binding sites, terminators, ribozymes, riboswitches, small RNAs, aptamers, as well as DNA machines and walker elements, and recently also DNA sequence elements are used for protein scaffolding (see Fig. 2.1).

Synthetic control of protein expression can occur at different levels. Several elements influence the transcription of genes such as promoters, synthetic transcriptional amplifiers (Blazek et al. 2011, 2012), 5' untranslated regions (UTRs), and multiple cloning sites (Crook et al. 2011), as well as terminators and 3' UTRs (Chen et al. 2013; Curran et al. 2013), either by a direct influence on transcription or termination efficiency or by differences in transcript stability. Furthermore, the selection markers, the vectors, and the genetic context of individual systems effect expression.

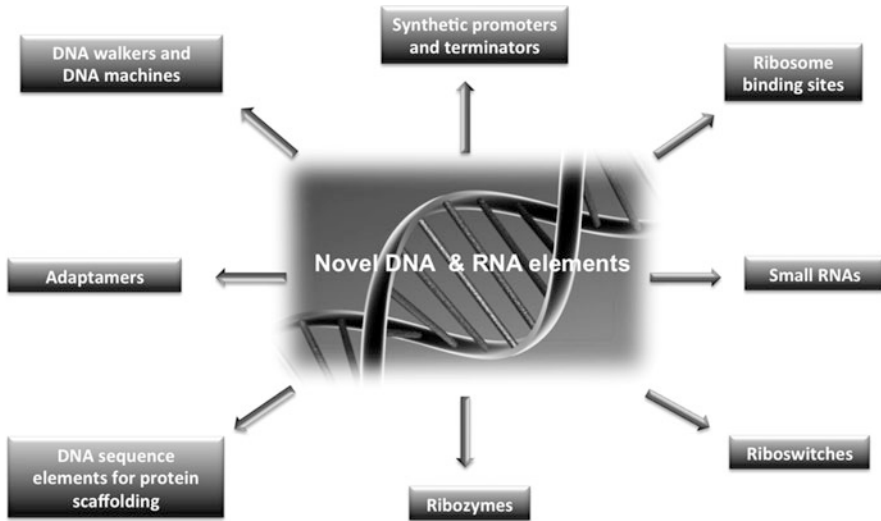


Fig. 2.1 Novel DNA and RNA elements. The most studied DNA and RNA elements comprise synthetic promoters and terminators, ribosome binding sites, small RNAs, ribozymes, riboswitches, DNA sequences for protein scaffolding, adaptamers, DNA walkers, and DNA machines. A picture of DNA molecule by Caroline Davis from <https://www.flickr.com/photos/53416677@N08/4973532326>

2.2 Synthetic Promoters

Engineering of promoters is a popular tool with high impact on protein expression. In order to fully understand transcriptional regulation, many different elements have to be considered, which often interact. However, in spite of such possible interactions, the modular design and construction based on the combinations of different synthetic nucleic acid parts are feasible.

Efficient and controllable promoters and knowledge about the involved transcriptional regulatory systems are essential for the optimization of protein expression (Vogl et al. 2013). There is increasing interest in using redesigned and synthetic promoters, since they broaden the natural biodiversity, facilitate the individual fine-tuning of the expression of the target gene (Ruth and Glieder 2010), and provide opportunities to overcome unknown or unexpected intrinsic regulation effects from natural promoters. In addition difficulties in respect to strain stabilities due to their sequence similarity and tendency to homologous recombination events can be avoided.

Although our knowledge about gene expression is constantly increasing, today we are able to predict transcript levels and translation initiation to some extent but not the optimal strength to express a maximum amount of biologically active recombinant protein, which differs between specific targets. Synthetic promoters

can span a wide range of expression levels and can therefore be used for many different purposes. They are especially useful for applications like the optimization of metabolic pathways. Traditional strategies focused either on gene knockout or on strong overexpression but in several cases none of these two extreme approaches lead to the desired results. Libraries of synthetic promoters provide a continuous set of different expression levels and allow fine-tuned control of gene expression (Hammer et al. 2006). However, one important aspect that has to be kept in mind concerning all different published studies about promoter engineering is the fact that the finally measured differences in the expression due to different promoter variants can have several reasons. They may be due to changed transcription, but they could also be caused by differences in mRNA stability or translation initiation (Vogl et al. 2014), since mostly reporter proteins are used to evaluate such variants rather than direct quantification of differences in transcript levels.

Promoter engineering was classified into four main strategies by Blazeck and Alper (Blazeck and Alper 2013), indicating a strong emphasis on top-down approaches rather than *ab initio* design of unprecedented promoter sequences so far:

1. Saturation mutagenesis of spacer regions
2. Random mutagenesis by ep-PCR
3. Hybrid promoter engineering
4. Direct modification of transcription factor binding sites (TFBSs)

Secondary structure elements probably have an important impact in successful strategies, too. More recently even the design of fully synthetic eukaryotic core promoter sequences including 5' UTRs was demonstrated (Vogl et al. 2014), and the existence of large datasets about yeast promoter regulation (Sharon et al. 2012) with repetitive regulatory patterns paved the way toward model-based bottom-up approaches in near future. The direct modification of computationally predicted and experimentally proven TFBSs can be applied for synthetic circuit design (Blazeck and Alper 2013) as well as controlled protein expression (Hartner et al. 2008).

While for the engineering of bacterial promoters, the first of the four strategies discussed above is the most popular one for fine-tuned expression of pathways. For eukaryotes, on the other hand, ep-PCR is very common for the same purpose. Hybrid promoters provided efficient novel prokaryotic promoters already in the early days of genetic engineering and recombinant protein production and are similarly useful for strong overexpression in eukaryotic hosts.

2.2.1 Hybrid Promoters

Common methods used to express heterologous and recombinant proteins in yeast or bacteria often involve the use of hybrid promoters (see Fig. 2.2). These constructs represent some of the first means to control transcription by merging important elements such as operators and consensus sequences of multiple promoters. They are advantageous in that select characteristics relating to

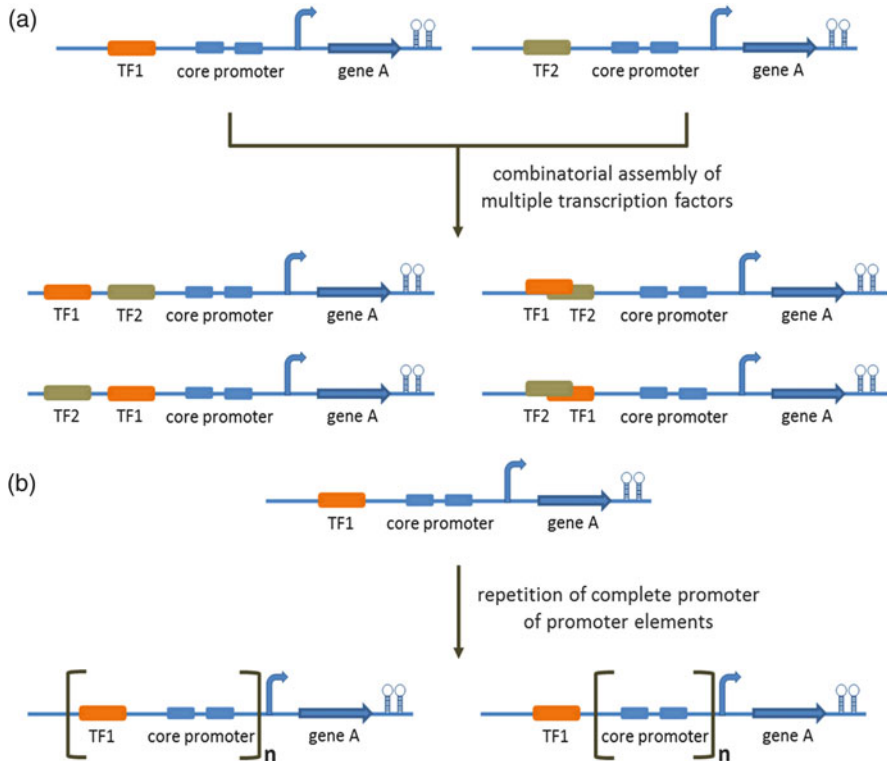


Fig. 2.2 Hybrid promoters. Construction of chimeric promoters by (a) merging of promoter elements and (b) replication of promoter (elements). n = number of tandem repeats

transcriptional regulation and/or strength can be combined in novel ways. Some of the oldest and best characterized hybrid promoters are the *tac/trc* promoters, which both consist of the -35 region from the *trp* promoter and the -10 region from the *lacUV5* promoter (Comstock 1983). The two promoters effectively merge the strength of the *trp* promoter with the lactose- or IPTG-induced regulation of the *lac* promoter, yielding a tightly regulated strong hybrid promoter that is functional in *E. coli*.

Years of empirical data gathering have brought to light a set of general design rules for hybrid promoters. It has been shown, for example, that placing activators upstream relative to the core promoter enhances RNAP binding by recruiting activating TFs, and placing repressor operators such that they overlap with the core promoter often hinders RNAP binding (Guazzaroni and Silva-Rocha 2014). Hybrid promoters have been proven to be able to amplify genetic expression over several orders of magnitude using activating operators and/or upstream activating sequences UASs and furthermore are capable of robust repression with just one repression operator in the core promoter (Cox et al. 2007). Upstream promoter (UP) elements have been shown to play a critical role in some prokaryotic systems

by increasing expression up to 90-fold, similar to homologous enhancer elements present in higher-order eukaryotic systems (Ross et al. 1998). Once a UAS and core promoter of interest are identified, expression can be enhanced further by creating tandem repeats of the complete promoter (Li et al. 2012) or one or several UASs, which enhances TF recruitment toward the core promoter (Blazeck and Alper 2013).

More complex hybrid promoters can be generated by combining operators capable of responding to different TFs, each with their own respective ligand. This type of behavior is an essential part in creating novel transcription control elements to program gene expression. Transactivation of genes in plants, for example, has been made possible by integrating hybrid promoters alongside orthogonal TFs, and by using TFs and corresponding promoters that respond to tissue-specific ligands, gene expression can be localized to individual cellular compartments (Moore et al. 2006). Beyond rational constructs, there has been much success generating combinatorial libraries of hybrid promoters, which can be used to investigate how different regulatory elements drive or inhibit transcription as well as interact with one another (Cox et al. 2007). It has been demonstrated empirically that no strong physical constraints exist for associations between different TFs having overlapping or adjacent binding sites with respect to a core promoter (Guazzaroni and Silva-Rocha 2014). Currently, hybrid promoters find use in transcription engineering because various well-characterized regulatory regions can often be recombined to yield novel control mechanisms in a predictable manner. Their historical validation and robustness make hybrid promoters frequently used devices and provide scientists a method to investigate new functionalities inspired by natural parts.

2.2.2 Common Strategies for the Engineering of Prokaryotic Promoters

Similar to protein engineering, different methods from two main categories can be selected for the engineering of promoter sequences. The first approach is based on random mutagenesis, whereas the second one relies on rational engineering strategies. Additionally, methods that combine both strategies can be applied.

A survey of consensus promoter sequences obtained from genomic sequencing data can provide the engineer with more rational insight into the preferred bases at different positions, from which targets for randomization can be deduced (see Fig. 2.3). In *E. coli*, these include the -35 (TTGACA) and -10 (TATAAT) regions relative to the transcription start site (Mutalik et al. 2013), nucleotides surrounding the boxes (Blazeck and Alper 2013), and the spacer in between (Jensen and Hammer 1998; Solem and Jensen 2002; Hammer et al. 2006; De Mey et al. 2007). Site-specific variability is generally introduced by PCR with degenerate oligonucleotide primers, which provide a much richer library than error-prone PCR. Coussement et al. (2014) recently reported a simple and efficient method to

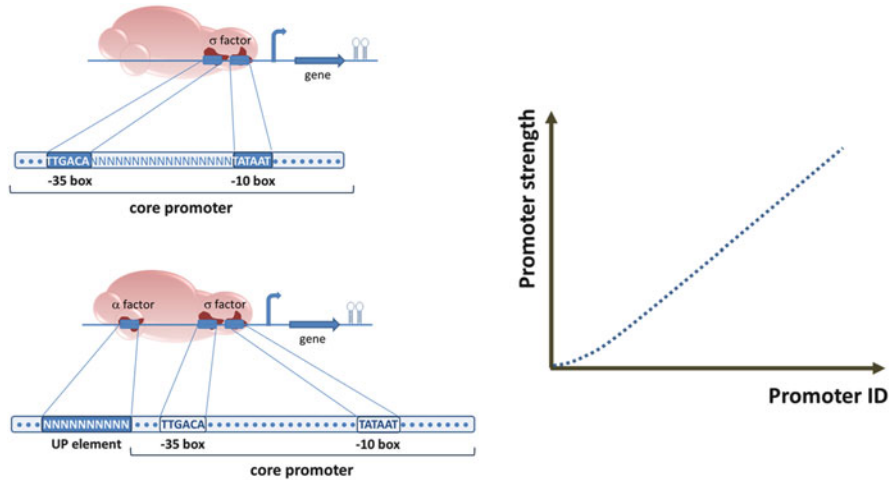


Fig. 2.3 Construction of promoter libraries. In prokaryotes the spacer sequence between the -10 box and -35 box (*top left*) is typically randomized yielding a promoter library with a broad range of expression strength (*right*). Sequences outside of the core promoter can also influence promoter strength via upstream promoter (UP) elements (*bottom left*)

assemble promoter libraries using degenerate oligonucleotides directly via Gibson assembly.

Strategies such as these have enabled the engineering of promoters over wide ranges of strength varying by several thousandfold in relative expression levels. Starting with a consensus promoter of high strength is often ideal, as the engineering process is typically more prone to reducing promoter strength than increasing it. The use of a constitutive promoter as a template for library generation is often preferable with a view toward later applications, since inducible systems tend to not be economical on an industrial scale. In addition, one can use an exogenous promoter template if a more orthogonal system with high expression is desired (Tyo et al. 2011).

Engineering of promoters frequently aims for the creation of very strong promoters. However, even for recombinant protein production but especially for balancing of (synthetic) metabolic pathways, it is often not desirable to use the strongest promoters available. Instead, it can be much more helpful to have a library of promoters with continuously increasing strength on hand frequently obtained by the randomization of spacer sequences which can be attained by a single PCR step and leading to synthetic promoter libraries with a large percentage (50–90 %) of variants with different expression levels (Alper et al. 2005; Hammer et al. 2006).

Studies of Hammer et al. showed that randomization of the spacer sequences of bacterial promoters can lead to 400-fold changes in promoter activity, making the spacer regions attractive targets for mutagenesis (Jensen and Hammer 1998). One option for spacer modifications is to use PCR primers with randomized spacer sequences and homology regions to the target gene (Solem and Jensen 2002).

Advances in the technologies for chemical DNA synthesis provide highly diversified oligonucleotide sequences, which can be especially useful for the simple and fast generation of prokaryotic promoter libraries (Ruth and Glieder 2010). Cheap double-stranded DNA blocks and even long single-stranded oligonucleotides provided by several companies worldwide easily cover whole bacterial promoter regions or operons.

Alternatively, mutagenic PCR can be performed, resulting in promoters with a wide, e.g., 200-fold range of expression levels. Drawbacks of mutagenesis by PCR are the low percentage of functional promoters (around 0.1 %), which implies tedious screening processes, and the high homology of obtained promoter variants, which may reduce the genetic stability (Alper et al. 2005; Hammer et al. 2006).

Alper et al. (2005) also demonstrated the utility of the red-colored compound lycopene for engineering promoters driving expression of upstream enzymes in the methyl erythritol phosphate (MEP) pathway used in terpenoid biosynthesis in *E. coli*. In this case, the productivity of an entire pathway is easily measured using a colorimetric reporter, and furthermore by utilizing an accurate contextual screening system, they were able to engineer a range of different strength promoters while also optimizing a metabolic pathway.

The bias of all available random mutagenesis methods is a serious disadvantage for short DNA stretches such as bacterial promoters in comparison to DNA synthesis and saturation mutagenesis. This was also demonstrated in a combined approach at the MIT. Random mutagenesis by error-prone PCR (ep-PCR) was performed for the construction of synthetic libraries of the P_L - λ promoter. Stephanopoulos et al. developed a statistical method to predict the effect of a single mutation in library variants, which contain several different mutations and tested the applicability of the method on the P_L - λ promoter. The promoter variants were analyzed by fluorescence measurements using flow cytometry and revealed positions, which correlate significantly with promoter strength. Site-directed mutagenesis was performed to target these positions as well as statistically insignificant sites and combinations thereof. Seven of eight mutants showed the expected phenotype, which was predicted by the statistical method. This technology can ease the identification of targets for rational mutagenesis of biomolecules (Jensen et al. 2006).

Mutations within TFBSs and in the consensus sequence of the -35 and -10 regions as well as changes in the length of the spacer between them often lead to reduced promoter activity (Jensen and Hammer 1998; Hammer et al. 2006). The low expression of these variants can give insight into the structure of the promoter, since their mutations can reveal regions, which are responsible for efficient expression or binding of transcription factors (Remans et al. 2005; Blazek and Alper 2013).

Since it is the arrangement and type of TFBSs that essentially define the architecture of a promoter, a systematic study was performed by Cox et al. to analyze the effect of different transcriptional regulators on expression. A combinatorial library of *E. coli* promoters was created by dividing the sequence of the promoters in three units: the distal region upstream of the -35 box, the core region

between the -10 and -35 box, and the proximal region downstream of the -10 box. Various combinations of four selected operators for transcriptional activators and repressors were incorporated in the units and they were randomly assembled by complementary overlaps. Analysis of these combinatorial variants allowed the authors to identify heuristic rules for the engineering of promoters, concerning the limits of regulation, number, and location of operators (Cox et al. 2007).

In a pioneering study by Kagiya et al., the generation of a prokaryotic promoter was achieved by random mutagenesis even though the starting eukaryotic 212-bp piece of genomic DNA from HeLa cells displayed no promoter activity at all. Within four rounds of ep-PCR, they obtained a strong bacterial promoter and demonstrated that synthetic prokaryotic promoters can be not only improved but also created in a relatively simple and fast way (Kagiya et al. 2005). However, it remains unknown if an enrichment of -10 and -35 like motifs or a change in DNA structure caused the activation of this eukaryotic genomic DNA fragment into a strong *E. coli* promoter.

No promoter is useful without reliable and reproducible expression characteristics in different contexts. This is often challenging when changing background expression strains or incorporating different genes upstream or downstream of a given promoter. These unwanted effects can be mitigated by properly insulating a promoter using buffering sequences 5' and 3' to the core promoter region. It is well known that UP elements can increase promoter expression several hundredfold by recruitment of core polymerase subunits, but the 20–30 nucleotides past the transcription start sites can also have a significant impact on promoter clearance and are thus important to consider when engineering an insulated promoter. Davis et al. (2011) have shown that by incorporating insulating DNA sequences from -105 up to $+55$ bp, they were able to negate the influence of UP elements as well as downstream inhibitory sequences on transcription efficiency, thus demonstrating the ability to engineer promoters with robust expression in a variety of contexts. Similarly, it can be of equal importance to keep the translation initiation rate constant for the resulting transcript, which can be achieved by using bicistronic domains, an architecture that couples translation of a gene of interest to an upstream miniature cistron, effectively normalizing the amount of gene expression regardless of variations in 5' secondary structure (Mutalik et al. 2013). The wealth of mutagenesis data and library information has made it possible to begin rationally predicting promoter sequences using empirical data alongside thermodynamic models (Brewster et al. 2012). Taken together, the state of the art not only offers predictive models for engineering new promoters but also provides existing expression constructs capable of reproducibly driving gene expression over a wide range.

2.2.3 Common Methods for the Engineering of Eukaryotic Promoters

The structure of eukaryotic promoters is much more complex compared to prokaryotic promoters, making their rational engineering more challenging (Ruth and

Glieder 2010) and random mutagenesis was applied successfully (in combination with screening). One way to circumvent this problem is the application of random mutagenesis. The group of G. Stephanopoulos created mutants of the strong and constitutive TEF1 promoter from *S. cerevisiae* by ep-PCR and obtained activities ranging from 8 % to 120 %. The authors confirmed by real-time PCR that the variations are caused by different transcript levels and that they are independent of the integration mode (plasmid or promoter replacement cassettes) and the carbon source (Nevoigt et al. 2006).

However, also for eukaryotic promoters, random mutagenesis approaches by ep-PCR and saturation mutagenesis predominantly resulted in mutants with decreased expression levels. Rational methods are more likely to facilitate the increase of promoter activity (Blazeck and Alper 2013). Especially the assembly of synthetic hybrid promoters has proven to be very successful. The design of these hybrid promoters takes advantage of the architecture of eukaryotic promoters, which consist of a core promoter and various upstream activating or repressing sequences (UAS/URS). Examples for the engineering of hybrid promoters among others can be found in *Pichia pastoris* (Hartner et al. 2008), *Y. lipolytica* (Blazeck et al. 2011), and *S. cerevisiae* (Blazeck et al. 2012) as well as mammalian promoters (Gehrke et al. 2003).

By adding UAS sites from different genes to the core promoter of P_{GPD} , Alper et al. created new promoter variants, which showed up to 2.5-fold higher mRNA levels compared to the native P_{GPD} , the strongest constitutive yeast promoter (Blazeck et al. 2012). These results show that it is possible to create synthetic promoters, which clearly exceed the strength of the strongest native yeast promoters.

For the design of synthetic hybrid promoters, the specific characteristics of different organisms have to be taken into account. In *Yarrowia lipolytica* it is, for example, possible to insert more than 20 copies of tandem UAS (Blazeck et al. 2011), whereas this would be hardly possible in *S. cerevisiae*. The efficient homologous recombination machinery of this yeast limits the number of identical UAS that can be maintained in a genetically stable manner (Blazeck et al. 2012).

Eukaryotic promoter libraries, which span a wide range of expression levels, were also achieved by fusion of Gal promoter-derived binding sites for transcription of UAS from P_{GAL1} to the minimal promoters of P_{LEU} and P_{CYC} . Therefore, binding sites of the transcriptional activator Gal4p were fused to constitutive core promoters in different combinations. Additional fine-tuning was achieved by adapting the distance between the UAS and the transcriptional start site. As a result, a galactose-inducible promoter library was created with continuously increasing strength. Furthermore, addition of UAS_{CLB} and UAS_{CIT} elements (from the mitotic cyclin CLB2 gene and the mitochondrial citrate synthase CIT1) led to a linear derepression of P_{GAL} under glucose-repressive conditions. Thereby, leaky hybrid promoters were created, which show low levels of protein expression in glucose-containing media. Compared to the strong increases seen for weak promoters, addition of UAS to the strong P_{GAL} increased the transcript level “just” by 15 %. With their work Blazeck et al. could demonstrate that synthetic hybrid promoters can not only be

used to obtain a wide dynamic range of expression but also for the establishment of new synthetic regulation mechanisms (Blazcek et al. 2012).

Although the methods described here are based on diverse engineering approaches, they are eventually all influencing transcription by TFBS effects. This is achieved through addition, removal, or modification of TFBSs either as individual elements or as larger parts in case of promoter fusions. As a consequence, the direct and systematic modification of TFBSs seems to represent the most straightforward and efficient approach and a very interesting target for further studies (Blazcek and Alper 2013). Applications of this approach had been demonstrated before, for example, by Hartner et al. (2008). Since TFBS are most times only known for a few intensively studied model promoters, the authors used sequence homology of putative TFBSs for targeted deletions and insertions. Co-occurring point mutations lead to additional unexpected effects.

Hartner et al. determined putative TFBSs within the strong, methanol-inducible P_{AOX1} promoter of *Pichia pastoris* (*Komagataella phaffi*). Their localization was achieved through computational predictions followed by deletion studies. By duplication and deletion of these TFBSs, a first-generation promoter library with activities between 6 % and 160 % of the native promoter was created. The new toolbox was not only tested with a reporter protein but also applied for expression of industrial enzymes. Repressing, derepressing, and inducible conditions were tested and at least 12 *cis*-acting elements were identified, which influence transcriptional regulation. Fusing these elements to basal promoters allowed the construction of short, synthetic promoters with different activities and regulation profiles. Promoter variants were constructed, with increased activity without the need of methanol addition. Short, synthetic promoters with such a regulation profile (Ruth et al. 2010) are well suited for conditions with carbon starvations in batch, fed-batch, or continuous cultivations. This approach, based on mutations within TFBSs, turned out to be highly successful for the generation of promoter variants with different strengths (Hartner et al. 2008). Some of the identified potential TFBS have been experimentally verified by others later (Kranthi et al. 2009).

Recently, Vogl et al. (2014) designed the first fully synthetic core promoter in *P. pastoris* and applied it for engineering and characterization of the P_{AOX1} core promoter. Since core promoters provide no or only very low basal transcription levels, they were fused to the upstream activating sequence of the *P. pastoris* AOX1 promoter (UAS_{AOX1}) to obtain high and easily quantifiable eGFP expression. The approach was based on a minimal consensus sequence that was obtained from the alignment of four different randomly chosen core sequences from natural *P. pastoris* promoters, which showed almost no sequence similarity. This first-generation synthetic core promoter sequence was functional in yeast but only to a very low degree. It was used as the basis for a second-generation core promoter, which was obtained by further incorporation of common core promoter sequence elements. The resulting synthetic core promoter showed at least some significant activity if fused to UAS_{AOX1} . Subsequently the native P_{AOX1} core promoter was engineered by replacing certain stretches with the sequence of the synthetic core promoter. The resulting library of synthetic variants spanned a range of 10–117 %

of the wild-type P_{AOX1} and can be used for the fine-tuning of protein expression (Vogl et al. 2014).

In Norway a different approach for the engineering of P_{AOX1} , based on the random introduction of point mutations, was used. Berg et al. selected for altered Zeocin resistance and were able to develop promoter variants with drastically increased tolerance under glucose-repressed conditions as well as under methanol-induced conditions (Berg et al. 2013). However, the effect of plasmid copy amplification of the ARS-based multicopy plasmids in *Pichia pastoris* was not discussed. The possible increase in promoter activity is much higher for the ep-PCR-based approach compared to the more rational engineering method by Vogl et al., but it requires the screening of a much higher amount of variants. The P_{AOX1} core promoters used in the two approaches differ in their length and reflect the versatile guiding principles according to which core promoters are defined. Due to the diverse promoter architectures, it is challenging to find universal rules for the definition of the exact length of core promoters. If a TATA box is present within the sequence, the 5' end of the core promoter is usually adjusted to this motif; alternatively the length of the core promoter can be determined experimentally.

A major limitation that hampers the de novo design of fully synthetic promoters is the incomplete understanding of how *cis*-regulatory motifs effect gene expression. Detailed and systematic analysis of thousands of designed promoters revealed the influence of several parameters on expression. The effect of changes in the number, affinity, orientation, position, and organization of TFBSs and nucleosome-disfavoring sequences was assessed and measured. It turned out that the orientation of the TFBSs influenced only 8 % of the tested TFs and that the effect of sequence context can be substantial but is not as important as, e.g., single base-pair mutations in TFBSs. As it is intuitively expected, increasing the distance between the transcriptional start site and the TFBSs decreases the effect of activators and repressors. Interestingly, a 10-bp periodic relationship between the position of the TFBS and expression was identified, so that even small changes in the location of binding sites can have large effects (Sharon et al. 2012) and reflect the 10–12 bases, which correspond to one helical turn. This confirmed the importance of the three-dimensional orientation of bound TFBS in relation to other binding factors, which are needed for transcription.

In a recent study, the group of Segal et al. aimed to unravel the connection between core promoter sequences and promoter activities in yeast and humans. They thereby identified k-mer and base content sequence features, which are predictive for highly active yeast promoters. These sequences are positioned within close proximity of the transcription start site, i.e., 75 bp upstream and 50 bp downstream (Lubliner et al. 2013). The findings of these studies can prove to be very useful for future promoter engineering and can provide an important basis for the design of fully synthetic promoter variants.

A completely different approach can be applied for the generation of synthetic promoters in higher eukaryotes. Instead of known TFBSs a library of random 18-bp DNAs was fused to a minimal promoter with a TATA box and an initiator element. Thereby, over 100 DNA sequences with functional *cis*-acting motifs were

identified, which enhance expression of the minimal promoter in neuroblastoma cell line Neuro2A. Database searches led to the identification of several known as well as novel sequence motifs (Edelman et al. 2000). In metazoans very high transcription levels can be successfully reached with a synthetic super core promoter, which consists of four core promoter motifs: the TATA box, initiator, motif ten, and downstream promoter element (Juven-Gershon et al. 2006).

In summary these studies demonstrated the broad applicability and high value of synthetic promoters for engineering of gene expression throughout different classes of organisms and how close we are right now toward the challenging goal of computer-based ab initio design of fully functional synthetic, strong, and tunable promoter parts for prokaryotes as well as for eukaryotic hosts.

2.3 Terminators

While non-intuitive, the termination of transcription can act as yet another important regulatory control point. In prokaryotes, termination is triggered by sequences that cause the RNAP to release the template and nascent RNA by means of hairpin formation or the recruitment of a Rho factor protein that races toward the RNAP (Platt 1986). The engineer should not underestimate the importance of transcription termination, as read-through transcription may disrupt the careful regulation of downstream systems, which could include plasmid copy number control elements or other ORFs (Mairhofer et al. 2013). For example, a combination of multiple terminators is required to efficiently halt the T7 RNAP and prevent read through (Mairhofer et al. 2015). Libraries of both natural and synthetic terminator sequences of varying strength have been reported and are easily incorporated downstream of a target gene (Chen et al. 2013).

In bacteria, where transcription and translation occur simultaneously, terminators can also be used to program transcription through attenuation mechanisms. Best known in the context of the *E. coli* trp operon, attenuation is the process during which a stretch of RNA conditionally forms either a terminator or an anti-terminator depending on environmental conditions (Yanofsky 1981; Naville and Gautheret 2010). In this case, the trp attenuator displays the terminator conformation as a result of ribosomes stalling due to tryptophan starvation.

Various studies (Pfleger et al. 2006; Cambray et al. 2013; Chen et al. 2013) provided more insight into the principles underlying transcriptional termination and the influence of 3'UTRs in prokaryotes. In eukaryotes and fungal systems, on the other hand, these mechanisms are not completely understood. However, the substantial influence of terminators on gene expression and their applicability for metabolic engineering was recently demonstrated for transcriptional terminators in *S. cerevisiae*. Depending on the terminator, a 13-fold dynamic range of expression levels of a fluorescent reporter gene was obtained compared to the construct lacking a terminator. The authors found out that the variations in the transcript and protein levels were mainly caused by changes in mRNA half-life (Curran et al. 2013). The results so far indicate that synthetic terminators and 3'UTRs are

a so far almost untouched but very promising tuning knob for transcriptional regulation of gene expression.

2.4 Ribozymes

Enzymes have been mainly seen as those biomolecules responsible for catalysis, but the discovery of RNA catalysts, so-called ribozymes, has opened a new view (Ramesh and Winkler 2014).

Ribozymes are RNA molecules with enzyme-like characteristics and activities, which are capable of breaking and forming covalent bonds. They have been identified for the first time in 1982 in *Tetrahymena thermophila*, where autocatalytic RNA rearrangements have been described. The authors discovered intrinsic RNA splicing activity, which occurs without the help of enzymes or small nuclear RNAs (Kruger et al. 1982).

Several features characteristic for enzymes are also true for ribozymes. First of all, both of them are able to accelerate reaction rates, they can use cofactors, and they can be regulated by the binding of allosteric effectors. Furthermore, the formation of specific tertiary structures and active sites is important for catalysis by enzymes as well as ribozymes (Doudna and Cech 2002).

Ribozymes such as self-splicing introns play an essential role in the RNA world hypothesis. This theory describes RNA molecules, which are capable of their own assembly and self-replication by recombination and mutation, as the starting point of evolution. After developing enzymatic functions through RNA cofactors, the synthesis of enzymes started based on RNA templates and the RNA core of the ribosome. The created proteins would then outperform ribozymes and predominate. Eventually DNA was constructed to provide a double-stranded, stable, linear form of information storage (Walter 1986). Despite several objections, such as the low stability and high complexity of RNA molecules as well as the rarity and small repertoire of reactions catalyzed by RNAs, the RNA world hypothesis kept high relevance (Bernhardt 2012).

The central role of RNA molecules is still illustrated, e.g., by the ribosome where rRNA is responsible for the catalytic peptidyl transferase reaction (Nissen 2000; Doudna and Cech 2002).

The most abundant and very well-studied small endonucleolytic ribozyme is the hammerhead ribozyme. It was the first ribozyme discovered and was found in subviral plant pathogens for cleavage of multimeric replication intermediates (Prody et al. 1986). Later on, the hammerhead ribozyme was found to occur in over 50 eukaryotic genomes, mainly in repetitive DNA sequences or introns. The ribozymes of the various organisms differ greatly in their sequences and length and seem to have evolved independently (Seehafer et al. 2011). Although self-cleaving ribozymes can vary largely in their structures and catalytic strategies, they can perform the same self-cleaving reaction of 5'-3' phosphodiester bonds or the reverse ligation (Fedor 2009). Their architectures and active sites are unique and

allow efficient general acid-base and electrostatic catalysis (Ferré-D'Amaré and Scott 2010).

Other important ribozymes that perform site-specific RNA self-cleavage are the hepatitis delta virus (HDV), hairpin, *Neurospora* Varkud satellite (VS), and *glmS* ribozymes. For cleavage they utilize base-pairing and alignment interactions between the target sequence and the cleavage site in the active center. In contrast to that, members of group I and II self-splicing introns use different mechanisms involving nucleophilic attacks and metal-ion catalysis to form mature transcripts by self-cleavage and ligation (Doudna and Cech 2002).

Interestingly, Bartel et al. found out that it is also possible for a single RNA sequence to assume two completely different ribozyme folds and consequently catalyze two different reactions (Schultes and Bartel 2000).

The mechanisms and characteristics of the different small self-cleaving ribozymes have been reviewed by Ferré-D'Amaré and Scott (2010). Recently, the list has been extended by the discovery of another member called twister RNA in many bacteria and eukaryotes (Roth et al. 2014; reviewed in Ramesh and Winkler 2014).

Additional to the natural function of cleavage of phosphodiester bonds, ribozymes can catalyze an impressive variety of reactions and they can do so even without the help of cofactors. In vitro-selected ribozymes can catalyze the formation of amide bonds, Michael adducts, nucleotides or coenzyme A, and so on (Doudna and Cech 2002). Ribozymes can furthermore catalyze the transfer of activated amino acids to tRNA. A covalent aminoacyl-ribosome intermediate is involved in charging of the tRNA (Jäschke 2001).

Recently, the research group of C. Voigt applied ribozymes as “insulator” parts in synthetic circuits to reduce the effect of the genetic context. The ribozymes do so by cleaving the 5'UTR of the mRNA, thereby generating a constant 5' mRNA context. In this way, ribozymes can reduce the coupling effects between the promoters and 5'UTRs and improve the predictability of layered circuits with mathematical models (Lou et al. 2012; Nielsen et al. 2013).

Lately, the computational design of highly specific small-molecule-sensing allosteric ribozymes was reported. The ribozymes can be created by fusing an aptamer for the desired target molecule to an extended or minimal version of the hammerhead ribozyme. The aptamer modules are tunable and provide therefore the possibility to design tailored functions. Conservation of important tertiary interactions between the stems I and II of the hammerhead ribozyme allowed to create high-speed molecular switches, which are very specific for their ligands and serve as YES or NOT logic gates. There are several potential applications of such ribozymes as molecular sensors for regulation of gene expression, high-throughput screening arrays, or antibacterial drug discovery (Penchovsky 2013).

Very recently flanking ribozyme sequences next to gRNAs were also successfully applied as an alternative to RNA polymerase III-driven expression of gRNAs for CRISPR/Cas9-mediated genome engineering. This specific processing allowed to use strong polymerase II-dependent promoters for gRNA expression (Gao and Zhao 2014).

In Chap. 3 about genome engineering methods by A. Weningen, M. Killinger, and T. Vogl, the applications of guide RNAs, essential for genome modifications based on the CRISPR/Cas9 system (Haurwitz et al. 2010), are described in more detail. The use of guide RNAs is gaining increasing popularity due to their versatile applications (Künne et al. 2014) and their convenient availability on gBlocks and on vectors in combination with the T7 promoter or as ready-to-use building blocks for direct expression.

2.5 Riboswitches

Ribozymes are involved in essential cellular functions such as translation and RNA processing. A different class of RNAs, which regulate downstream gene expression, are so-called riboswitches. They do so mainly by influencing translation initiation or premature termination of transcription. The class of metabolite-sensing riboswitches couples the detection of specific ligands to ribozyme activity (Ramesh and Winkler 2014; Winkler 2005). A well-studied member of this group is the *glmS* riboswitch, which triggers self-cleavage upon binding of glucosamine-6-phosphate. The following degradation of the cleaved products by RNases turned out to be different in *E. coli* compared to other bacteria. In the end, expression of the *glmS* gene is reduced by regulation of mRNA stability (Collins et al. 2007; Ramesh and Winkler 2014; Winkler et al. 2004).

Riboswitches are *cis*-acting regulatory RNAs, which are binding to intracellular metabolites and thereby regulating gene expression. They are structural elements, which are usually occurring in the 5' UTRs of mRNAs (Tucker and Breaker 2005). Research efforts have been mainly focused on riboswitches in bacteria, although they are also occurring in other organisms, e.g., thiamine pyrophosphate binding riboswitches in plants and fungi (Tucker and Breaker 2005; Kubodera et al. 2003; Sudarsan et al. 2003).

A detailed review about bacterial riboswitches and their role in regulation of gene expression and possible applications has been published by Winkler and Breaker (2005).

The structure of riboswitches can be divided into two main parts, namely, the aptamer and the expression platform, which is located downstream of the aptamer. The aptamer domain contains the sequence that binds to the metabolite and shows a very high degree of sequence conservation even within diverse organisms. Binding between the target metabolite and the aptamer causes a conformational change in the expression platform domain. These changes in conformation result then in different expression levels. The expression platforms show a high diversity with regard to their sequence, length, and structure. The fact that riboswitches are most of the time located upstream of the genes, which are coding for the synthesis or transport of the metabolite they are binding, can be exploited for the identification of target metabolites and the function of new genes (Winkler and Breaker 2005).

Based on the principle of riboswitches, Durand et al. have recently applied aptamers as biosensors for the detection of small ligands. These so-called

aptaswitches fold into a hairpin structure upon binding of the ligand. When no ligand is present, the aptamer is in its unfolded state. This structural switch, which depends on the absence or presence of the ligand, allows the application of aptaswitches as biosensors. In the folded state, a second hairpin recognizes the formed, apical loop and a kissing complex is formed through loop-loop interactions. The quantitative and specific detection of ligands by aptakiss-aptaswitch complexes was demonstrated successfully for GTP and adenosine. The development of such sensors based on hairpin aptamers can potentially be applied for any molecule with known hairpin aptamers, provided that the apical loop is not responsible for binding the ligand. The rationally designed, synthetic kissing loops could be combined with natural occurring kissing loops, which are involved in the regulation of different biological processes, and may in future also be useful for multiplexed analysis (Durand et al. 2014).

After the successful construction of riboswitches for translational regulation, they have also been engineered for transcriptional regulation. The theophylline aptamer was employed as sensor and the actuator part consisted of RNA sequences, which fold into functional intrinsic terminator structures. This concept allowed ligand-dependent regulation of gene expression by de novo design of synthetic riboswitches which influence transcriptional termination (Wachsmuth et al. 2013).

Recently, a novel approach was developed for the prediction of riboswitches in DNA sequences by a computational tool with high sensitivity and specificity called Denison Riboswitch Detector (Havill et al. 2014).

The knowledge and results of all these studies about ribozymes and riboswitches can provide the foundation for future intentions toward the design of synthetic, tailor-made riboswitches.

2.6 Small RNAs

2.6.1 Detection, Prediction, and Classification of Small RNAs

Small RNAs (sRNAs) play important and multifaceted roles in prokaryotes as well as eukaryotes. In prokaryotes, sRNAs are involved in the tagging of proteins destined for degradation and influence the activity of RNA polymerase and translation. The first bacterial sRNAs have been detected unintentionally by direct analysis of highly abundant sRNAs or during analysis of proteins or activities related to overexpression of genomic fragments. In order to get a more detailed and comprehensive insight into the role of sRNAs, studies for their systematic prediction were performed. A major challenge concerning the detection of sRNAs is that they lack conserved, characteristic sequence motifs that allow their identification (Wassarman et al. 2001). Furthermore, the discrimination between small, non-translated RNAs and random sequences is not possible solely based on secondary structural elements (Rivas and Eddy 2000; Argaman et al. 2001).

Three different approaches for sRNA prediction were applied simultaneously. Computational predictions based on transcription signals and genomic features of

already known sRNAs were used by Argaman et al. (Argaman et al. 2001). Wassarman and coworkers combined genome-wide computer searches using parameters identified in known sRNAs, genomic microarrays, and isolation of sRNAs associated with RNA-binding proteins (Wassarman et al. 2001). The identified sRNAs were experimentally confirmed and overlap with the small, noncoding RNAs predicted by Rivas et al. using a computational comparative genomic screen. Intergenic sequences of *E. coli* were analyzed and sequence data from four related bacterial strains were compared. Noncoding RNAs can have regulatory, structural, or catalytic roles, but in contrast to protein coding sequences, they lack inherent statistical biases and are therefore harder to predict. Position-specific mutational models have been applied to discriminate between probable coding regions, structural RNAs, and “other” sequences in pairwise alignments (Rivas et al. 2001). The sRNAs identified by the different methods varied in their length from 50 to several hundred nucleotides.

Apart from the noncoding, small RNAs described above, the term sRNAs is frequently used to refer to very short, usually 20–30-nucleotide-long RNAs, which are important for the regulation of gene expression and genome stability. Small RNAs can be divided into at least three classes. Depending on their mechanism, their localization within the cell, and the origin of the involved RNA molecule, they can be classified as short interfering RNAs (siRNAs), microRNAs (miRNAs), or PIWI-interacting RNAs (piRNAs) (Moazed 2009).

siRNAs and miRNAs are typically 21–25 nucleotides long, whereas piRNAs are with 24–31 nucleotides on average a bit longer. piRNAs are important components for defense mechanisms against parasitic DNA sequences and may also play a role in silencing of homologous genes. All three of them seem to be involved in posttranscriptional gene silencing (PTGS) as well as chromatin-dependent gene silencing (CDGS), which can be further divided into transcriptional and co-transcriptional gene silencing. These mechanisms illustrate that gene silencing by sRNAs can occur on the level of mRNA translation or stability and on the chromatin and DNA level. Interestingly, these mechanisms and their effect on chromatin regulation seem to be highly conserved among eukaryotes, except from *S. cerevisiae* (Moazed 2009). A very detailed description of the RNA processing pathways, the origin of different small RNA classes, and their role in chromatin silencing can be found in the review of D. Moazed.

2.6.2 Small RNA Processing Pathways

Initially, RNA interference (RNAi) was described for *C. elegans*, where injected dsRNA caused silencing of homologous host genes in the animal as well as their progeny. Notably, the effect of single-stranded sense and antisense RNA on gene expression was much lower compared to dsRNA (Fire et al. 1998). Injection of the dsRNA into the extracellular body cavity led to the spreading of the interference throughout a broad region of the organism. The interference effect was also observed when *C. elegans* larvae fed on *E. coli* bacteria, which express the

corresponding dsRNA (Timmons and Fire 1998). In later studies with HeLa cells, it was shown that short single-stranded 5'-phosphorylated antisense siRNAs can trigger gene silencing as well, since they are able to enter the mammalian RNAi pathway in vitro and in vivo (Martinez et al. 2002).

A ribonuclease III enzyme, called Dicer, is responsible for the generation of siRNAs and miRNAs by cleavage of precursor double-stranded RNAs (dsRNAs). Effector complexes, the so-called RNA-induced silencing complex (RISC) and RNA-induced transcriptional silencing complex (RITS), are involved in base pairing of the sRNA with the homologous target sequence (Hammond et al. 2000; Verdel et al. 2004; reviewed in Moazed 2009). Base pairing with the target mRNA can result in cleavage or degradation of that mRNA, so-called RNAi (Fire et al. 1998; Martinez et al. 2002).

The essential components of the silencing complexes are the so-called Argonaute proteins, which are binding to the guide sRNAs. The family of Argonaute proteins can be split into two clades, those resembling *Arabidopsis* AGO1 and those more similar to *Drosophila* Piwi. The Argonaute proteins are highly basic and contain two different domains, the C-terminal PIWI domain and PAZ domains, named after the Piwi, Argo, and Zwiller/Pinhead proteins containing this domain (Cerutti et al. 2000; Carmell et al. 2002). The PIWI domain, which resembles ribonuclease H, is responsible for binding the sRNA at its 5' end and provides the endonuclease activity. The PAZ domain, on the other hand, is involved in binding the 3' end of the sRNA and probably in positioning the recognition and cleavage of the mRNA target (Song et al. 2004; Zamore and Haley 2005). The mechanism of target RNA cleavage is not only seen as the siRNA and RNAi mode but also as an important way of gene silencing by plant miRNAs as well as sometimes animal and viral miRNAs (reviewed in Zamore and Haley 2005).

piRNAs, on the other hand, associate with Argonaute proteins of the Piwi clade and are mostly targeting transposable elements of metazoan genomes. piRNAs recognize and repress these transposable elements and also memorize them. They are much more diverse than miRNAs and it seems that they can originate from any sequence that is located within a piRNA cluster region and processed via multiple enzymatic steps. The piRNA clusters provide information about foreign genes that have to be silenced and explain why piRNAs can be seen in a way as an immune system (Stuwe et al. 2014).

More information about the RNAi mechanism and its application for targeted gene knockdown can be found in Chap. 3 about genome engineering methods by A. Weninger, M. Killinger, and T. Vogl.

The general mechanism of the miRNA pathway of plants and animals is conserved and involves the RNaseIII enzymes Dicer and Drosha, although the latter of them is not occurring in plants (Moazed 2009). These dsRNA-specific endonucleases are responsible for processing the long, widely unstructured precursor RNAs (pre-miRNAs) to mature, single-stranded miRNAs. They do so by cutting out ~70-nt-long hairpin structures of the pre-miRNA. There are certain criteria by which miRNAs can be identified and distinguished from other small RNAs, such as confirmation of their expression by hybridization assays, their

structure and phylogenetic conservation, and their accumulation due to reduced Dicer function. In contrast to miRNAs, siRNAs originate from dsRNAs with hundreds or thousands of nucleotides in length and are created by successive cleavage (Ambros et al. 2003; Kim 2005).

miRNAs and siRNAs can cause gene silencing by suppressing mRNA translation or cleavage of the mRNA of the target gene (Zeng et al. 2003; reviewed in Rana 2007). Cleavage of fully complementary mRNA target sites was previously seen as a characteristic of siRNA-induced RNAi. Downregulation of expression, on the other hand, was seen as a characteristic of miRNAs. Following studies however indicated that miRNAs and siRNAs are functionally interchangeable and able to use the same mechanisms for mRNA degradation and mRNA translation inhibition. It turned out that the main feature that determines which mechanism is carried out is the degree of complementarity with the target mRNA. Fully complementary sequences cause mRNA cleavage, whereas mismatches result in the formation of central bulges and consequently in translational inhibition (Zeng et al. 2003).

An alternative mechanism for gene silencing by miRNAs is based on enhancing mRNA degradation. It is independent of slicer activity and requires only partial base pairing. This mechanism emphasizes the importance of mRNA stability in miRNA pathways (Bagga et al. 2005).

Interestingly, only six or seven nucleotides of the sRNAs are decisive for the main binding specificity of an sRNA. This part is therefore called “seed sequence” (Yekta et al. 2004). The 5' end of the sRNA is contributing disproportionately to the binding of the target RNA, whereas the first nucleotide of the sRNA seems to remain unpaired (reviewed in Zamore and Haley 2005).

The interaction of these tiny sRNAs with their target relies on binding by Argonaute family proteins and is different from the mechanism of antisense oligonucleotide-target RNA pairing. The recognition sites of the sRNAs occur randomly every ~4000–65,000 nt. Upon binding of the target, the sRNA directs cleavage of a phosphodiester bond in the target RNA between the nucleotides corresponding to the middle of the guide sRNA. This cleavage requires binding of the appropriate Argonaute protein and of most of the sRNA nucleotides to the target RNA as well as the formation of at least one turn of an A-form helix. As a result, cleavage is more specific than sRNA binding itself (reviewed in Zamore and Haley 2005).

miRNAs are frequently targeting key transcription factors important for cellular identities. Due to the fact that expression of the miRNAs is regulated by transcription factors, they can be used to design diverse feedback loops (Stuwe et al. 2014).

2.6.3 Functions and Applications of Small RNAs

The wide range of functions of miRNAs in various regulatory pathways is outstanding, especially considering their tiny length of ~22 nt (Kim 2005). Their ability to act as posttranscriptional repressors by specific binding to the 3'UTRs of their target mRNA is just one example (Reinhart et al. 2000; reviewed in Ambros 2004). A

miRNA of *Drosophila* was found to be involved in regulation of apoptosis, cell proliferation, and tissue formation in a temporally and spatially regulated manner (Brennecke et al. 2003). Animal miRNAs are furthermore involved in hematopoiesis and neuronal patterning (reviewed in Kim 2005 and Ambros 2004). A miRNA from *C. elegans* named *lisy-6* has been shown to control the left/right asymmetric expression of genes in two chemosensory neurons. The miRNAs regulate the laterality of the chemosensory system of the nematode in a sequential and asymmetrical way. This sensory system enables the worm to discriminate between different attractive and repellent external, chemical stimuli. The miRNA produced from the *lisy-6* gene functions by repression of a downstream transcription factor, the so-called COG-1 transcription factor, through binding to a partially complementary sequence within the 3'UTR sequence of the *cog-1* mRNA (Johnston and Hobert 2003; Chang et al. 2004; reviewed in Ambros 2004).

Small RNAs are involved in regulation of gene expression and also genome stability. They have been shown to direct chromatin-modifying complexes to specific chromosome regions through interactions with nascent chromatin-bound ncRNAs (Moazed 2009).

Recent studies with flies and worms demonstrated that small RNAs can be involved in cellular memory and transgenerational inheritance, either in cooperation with chromatin modifications or independently (Stuwe et al. 2014).

Furthermore, it was shown recently that the plant RNAi machinery can be exploited by *Botrytis cinerea* to transfer “virulent” sRNA into the host cells. This fungal pathogen causes the gray mold disease, which can lead to severe impairments of many important agricultural crops. Bc-sRNAs can bind to the AGO proteins of the *Arabidopsis* or tomato plants and cause gene silencing of host genes with complementary target sequences. Detailed analysis of the effected genes revealed that the Bc-sRNAs target predominantly host immunity genes. Host gene silencing was not observed when the complementary sequences of the target genes were mutated and when AGO1 of the plant was knocked out. Suppression of host immunity genes was also abolished when the DCL genes of *B. cinerea*, which are involved in sRNA processing, were knocked out. These results support the suggested hijacking mechanism, by which sRNAs of the pathogen can achieve infection through suppression of host immunity genes (Weiberg et al. 2013).

2.7 Long Noncoding RNAs

In contrast to short- and mid-sized RNAs, long noncoding RNAs (lncRNAs) are more than 200 bp long. lncRNAs include transcribed ultraconserved regions (T-UCRs) as well as large intergenic noncoding RNAs (lincRNAs). They are involved in up- and downregulation of gene expression and chromatin architecture and in tumorigenesis and different neurological and cardiovascular diseases. The

locations, functions, and characteristics of the different ncRNA classes are described in more detail in several reviews (Esteller 2011; Wahlestedt 2013).

2.8 Aptamers and Adaptamers

In order to create a generic way for the formation of aptamers, which are binding to two target proteins, James et al. mixed two engineered aptamers. These two aptamers efficiently formed hybrid molecules, so-called adaptamers, which are able to bind two ligands simultaneously. The system was tested for the binding of streptavidin and a second target protein and widens the applicability of streptavidin-biotin-based detection systems (Tahiri-Alaoui et al. 2002).

Aptamers, the basic building blocks of adaptamers, are DNA and RNA molecules that are very selectively binding their target molecules. Their name is derived from the Latin word “aptus,” fitting, referring to a nucleotide polymer, which fits to its target (Ellington and Szostak 1990). The development of aptamers was achieved by *in vitro* selection studies where random sequence pools were evolved for high binding affinities to target ligands, using the so-called SELEX (systematic evolution of ligands by exponential enrichment) procedure (Tuerk and Gold 1990; Ellington and Szostak 1990; Hermann and Patel 2000).

Aptamers display an outstanding versatility regarding possible target molecules, which include proteins, drugs, whole cells, or small organic molecules and metal ions. A major advantage of aptamers is their high affinity, which permits their use for biomedical applications like targeted drug delivery or analytics. A very interesting research field focuses on the combination of aptamers with nanoparticles, which are frequently used for bioimaging in cancer diagnostics and treatment. Thereby, the specific binding of the aptamer to the target molecule improves the binding of the nanoparticle. Nevertheless, aptamers for target molecules in medicine are rare and their field of application is therefore restricted (Reinemann and Strehlitz 2014).

The prominent specificity of aptamers is grounded on their highly optimized three-dimensional structures for recognition of their target molecule. A single methyl group difference is enough for theophylline-binding RNA aptamers to bind their target theophylline 10,000 times stronger than caffeine (Jenison et al. 1994). Several different types of interactions contribute to the molecular recognition. Stacking and hydrogen-bonding interactions are, e.g., involved in the complex formation between aptamers and flat, aromatic ligands (Hermann and Patel 2000). Further interactions, which are important for the high specificity of aptamer binding, are based on molecular shape complementarity. Structural electrostatic complementarity arises from positively charged ligands and negatively charged RNA molecules (Hermann and Patel 2000; Tor et al. 1998).

Small molecules and their RNA aptamers have been used successfully for the regulation of eukaryotic gene expression in living cells. To this end, small-molecule

aptamers were inserted into the 5' untranslated region of a mammalian β -galactosidase mRNA and expressed in Chinese hamster ovary cells. In absence of the corresponding drug, no effect on expression was observed, whereas addition of the drug binding the aptamer inhibited β -galactosidase activity by more than 90 % (Werstuck and Green 1998).

The ability of aptamers to bind to bacterial cell surfaces was exploited in combination with quantum dots for the detection of bacteria. Therefore, the fluorescence emission of quantum dots was measured, which shifts upon binding to bacterial surfaces via DNA aptamers. The aptamers accomplished the role of antibodies, which can be used for the same application as well but they are significantly larger (Dwarakanath et al. 2004).

Research focused on aptamers allowed insight into intermolecular recognition and showed that they are very valuable and promising tools for molecular sensors and switches (Hermann and Patel 2000).

Furthermore, another kind of adaptamers can be an extremely useful tool for genome engineering, as it was shown, e.g., for the disruption of genes in *S. cerevisiae*. Here, the term adaptamer is used for primers with specific 5' fusion tags, which allow the generic combination of DNA elements by PCR, due to the annealing of the attached adaptamers. A set of intergenic adaptamers is commercially available from Research Genetics containing primers with such 5' sequence tags, which are not homologous to endogenous yeast DNA. The method (see Fig. 2.4) starts with the PCR amplification of the intergenic regions flanking the gene, which should be knocked out, with intergenic adaptamers. An appropriate selectable marker is PCR amplified in form of two overlapping fragments by adding the complement, reverse adaptamer tags. The two intergenic fragments and the marker fragments are fused by PCR. Thereby, two fusion segments are obtained which are co-transformed in yeast, where they recombine with genomic DNA and consequently disrupt the selected gene. Direct repeats, which are flanking the selectable marker, facilitate the removal and future reuse of the marker by recombination. The disruption of genes based on PCR and adaptamers provides a fast, efficient, and versatile tool, which can be used to study any gene disruption of interest and to increase the knowledge about gene functions in yeast (Erdeniz et al. 1997; Reid et al. 2002).

Genome modifications based on adaptamers and bipartite gene-targeting substrates were successfully implemented also in *Aspergillus nidulans*, a filamentous fungus, which shows mainly random integration of foreign DNA. The method applied by Mortensen et al. is very flexible and can reduce the amount of primers and PCRs needed for genome modifications and therefore the costs. Other advantages are a low number of false positives and the possibility to recycle the selectable marker, so that multiple genome modifications can be performed (Nielsen et al. 2006).

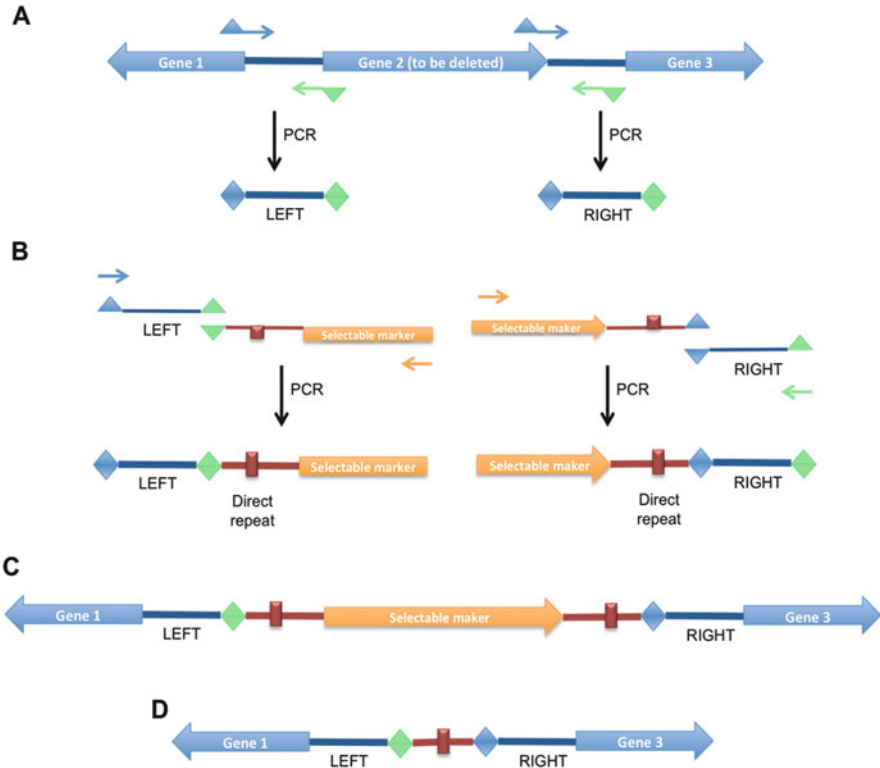


Fig. 2.4 Adaptamer-directed gene disruptions. In (a) the PCR, amplification of the intergenic regions flanking the gene, which is going to be disrupted in the genome, is shown. Adaptamers, depicted as *blue* and *green arrows with triangles*, are used to add adaptamer tags to the intergenic regions. The obtained PCR products are combined with a selectable marker (e.g., *K. lactis* URA3) by overlap extension PCR as illustrated in (b). Transformation of the two fusion DNA fragments results in recombination with genomic DNA and gene disruption as illustrated in (c). The original gene is replaced by the selectable marker and flanking direct repeats. Upon recombination of the direct repeats, the selectable marker is kicked out, resulting in the genome structure shown in (d). In the case of *K. lactis* URA3, marker-free constructs can be selected on 5-FOA medium, allowing the reuse of the marker for further genetic modifications. Figure adapted from Reid et al. (2002)

2.9 DNA Barcodes

DNA sequences can be employed as “barcodes,” which facilitate on the one hand the assignment of unknown specimens to species and on the other hand the discovery and identification of new, otherwise inaccessible species. The mitochondrial cytochrome *c* oxidase I (COI) gene turned out to be a suitable reference for species identification based on COI profiles (Hebert et al. 2003; Frézal and Leblois 2008).

The great capability of DNA synthesis, far beyond the size of expression cassettes or plasmids, was demonstrated in 2008 when the chemical synthesis of a whole genome was published by the J. Craig Venter Institute. In the course of the synthesis, assembly and cloning of the 582,970-bp genome of the bacteria *Mycoplasma genitalium* short “watermark” sequences were inserted. These watermarks were inserted at intergenic sites to minimize biological effects and they enabled the clear differentiation between the synthetic and the native genome (Gibson et al. 2008). The first complete chemical synthesis of a bacterial genome represented an important milestone in synthetic biology (Gibson et al. 2008). Two years later the genomes of *Mycoplasma genitalium* and two other bacteria were cloned in the yeast *S. cerevisiae* as single-DNA molecules (Benders et al. 2010). Recently, the first total synthesis of a functional designer eukaryotic chromosome was achieved (Annaluru et al. 2014).

In addition to their use the labeling or identification of chromosomes or genomes watermarks or barcodes can also be employed on a smaller scale, e.g., for the labeling of plasmids and DNA sequences, e.g., in next-generation sequencing experiments. For example, unique 20-bp-long “molecular barcodes” have been furthermore employed for the identification of *S. cerevisiae* deletion strains (Giaever 2002). Barcodes can be added by primers and used to identify, e.g., hits of a promoter library by 454 pyrosequencing (Kinney et al. 2010) or for deep sequencing of barcoded mRNAs (Patwardhan et al. 2009; Melnikov et al. 2012).

2.10 DNA Machines

The reason why it is possible to build machines made from DNA lies in the highly specific interactions between complementary nucleotides. As a consequence, two-dimensional and complex three-dimensional DNA structures can be constructed based on the base sequences and the formation of branched helices (Seeman 2003; Bath and Turberfield 2007; Seeman 2010). An important characteristic of these nanoscale architectures is their self-assembling nature. DNA molecules can be used as scaffolds for the periodic assembly of molecules with possible applications for memory devices and DNA-based computing (Seeman 1998). The structures, which can be formed, are becoming more and more complex and advanced from cubes (Chen and Seeman 1991) and octahedrons (Shih et al. 2004) to multifaceted DNA origami structures such as five-pointed stars (Rothemund 2006).

The next step toward nanorobotics was the development of dynamic nanodevices from static DNA structures. These include, e.g., boxes and pinching devices, which can be used to detect molecules with an extremely large-sized range from metal ions to whole proteins (Kuzuya and Ohya 2014).

It is important to keep in mind that the nanomechanical movements of the devices are defined by their nucleotide sequence. As a consequence, DNA nanomachines are programmable and useful for highly diverse applications. Very interesting examples are sequence-dependent rotatory devices which function in a

cyclic manner (Yan et al. 2002), DNA walkers (Tian et al. 2005; Sherman and Seeman 2004; Shin and Pierce 2004; Yin et al. 2004), and DNA tweezers (Landon et al. 2012; Yurke et al. 2000). Their movements range from relatively simple conformational changes like opening/closure or rotation to complex walking step sequences (Tian et al. 2005).

However, a major limitation of early nanomachines was that in contrast to macroscale machines, they required human interference after each step (Sherman and Seeman 2004; Shin and Pierce 2004). Subsequently, autonomous machines, in the sense of self-contained devices, which are independent of human interference, have been established. An example for such a device is the nanomotor from Mao et al., which consumes chemical energy for autonomous motion. It can walk in two directions, thereby destroying its track. Compared to protein-based motors, which move along straight tracks, their DNA counterparts are very slow but more versatile (Tian et al. 2005).

The applications of nanomachines are highly diverse and range from sensors to optoelectronic devices and biopharmaceutical purposes. DNA origami “sheath,” which imitates transcriptional suppressors, can be applied for controlling expression, whereas clamshell-like nanodevices allow the differentiation of cell lines by logic gates (Endo et al. 2012; Douglas et al. 2012; Kuzuya and Ohya 2014).

Recently, a DNA nanorobot was developed, which can transport cargo loads to specific cells and unload its charge after conditional, triggered activation and structural reconfiguration. Its function is controlled by different logical AND gates. The nanorobot is shaped as a hexagonal barrel and has two pairs of partially complementary lock strands. These lock strands contain an aptamer, which recognizes targets, such as cell line-specific antigens. Selective strand displacement causes the release of the loading at the target site. The applicability of these DNA nanorobots was demonstrated by the transport of fluorescent antibody fragments to the antigens on human cells. Unloading of the robot led to the fluorescent labeling of the specific cells (Douglas et al. 2012; Kuzuya and Ohya 2014).

DNA pliers have been shown to be some of the most versatile instruments of the DNA origami toolbox. They contain two 170-nm-long levers with a Holliday junction in between. These single-molecule beacons can be used for the detection of biomolecules by three different mechanisms. The first mechanism is based on pinching for the detection of target molecules, which are binding to ligands in the jaw. This method was demonstrated with biotin molecules serving as ligands and closing of the plier upon binding of streptavidin. In order to be able to detect molecules with weaker interactions compared to very strong protein-ligand interactions, a second zipping mechanism was developed. Zipping involves several elements in the levers, which are collectively binding together upon target addition, and allows the detection of, e.g., Na^+ ions. The reverse reaction is called unzipping and represents the third mechanism. Thereby, the initially closed plier is opened when target molecules, such as human microRNAs, are present (Kuzuya et al. 2011; Kuzuya and Ohya 2014).

Interestingly, this unzipping mechanism can also be used for the detection of specific binding modes such as the invasive binding of peptide nucleic acids in

DNA duplexes (Yamazaki et al. 2012). The transition between the opened and closed state of the pliers can be monitored in real time by labeling with fluorescent dyes or simply by agarose gel electrophoresis (Kuzuya and Ohya 2014).

DNA origami technology seems to have started a new epoch in structural DNA nanotechnology. In 2009, the first three-dimensional, hollow structures, comprising boxes, tetrahedrons, and prisms, were created. Advantages of DNA origami structures compared to conventional DNA nanomachines are their increased assembly yield and their ability to precisely assemble molecules with different functional groups. Furthermore, they are large enough to be detected by atomic force microscopy or transmission electron microscopy (Kuzuya and Ohya 2014).

At first sight, it does not seem to be logic to choose DNA as building material for machines since their catalytic capacity and structural versatility are lower compared to proteins or RNA. However, it is exactly this simplicity of DNA structures and interactions that facilitates researchers to predict their assembly and behavior and enables their use for nanomachines (Bath and Turberfield 2007).

2.11 DNA Walker

Precise intracellular transport along nanostructures represents a substantial difficulty, which was addressed by the construction of synthetic DNA walkers. The first DNA walkers have been designed in 2004 by the groups of Pierce et al., Seeman et al., and Reif et al. in parallel. The walkers, which were constructed at the beginning, moved in an inchworm-type gait, with one leg trailing the other (Sherman and Seeman 2004).

The group of Reif et al. designed a unidirectional and autonomous DNA motor, powered by ATP hydrolysis. The walker consists of a six-nucleotide DNA fragment, which is ligated to anchorages on a track and then released by a restriction endonuclease. Thereby, the walker may serve not only as a carrier of information but also of matter, such as nanoparticles (Yin et al. 2004).

The next step in the development of walking nanodevices consisted in bipedal DNA walkers, which are capable of moving forward by putting 1 f. in front of the other (Shin and Pierce 2004; Yin et al. 2004). The approach by Pierce et al. (see Fig. 2.5) consists of a walker, made of two partially complementary DNA strands with a double-stranded helix and two single-stranded legs. The legs can bind in an alternating manner to the protruding single-stranded branches of the track. Therefore, attachment fuel strands are used, which facilitate the anchoring by helix formation. After binding of both legs, the trailing leg is released from the track via displacement by the detachment fuel strand. The movement of the walker can be monitored by fluorescence measurements since the legs of the walker are marked with quenchers, whereas the ends of the branches are marked with various dyes. This allows real-time monitoring by multiplexed fluorescence quenching measurements (Shin and Pierce 2004).

Robots on single-molecule level represent an innovative and fascinating research area. A major challenge thereby is finding a way to store complex

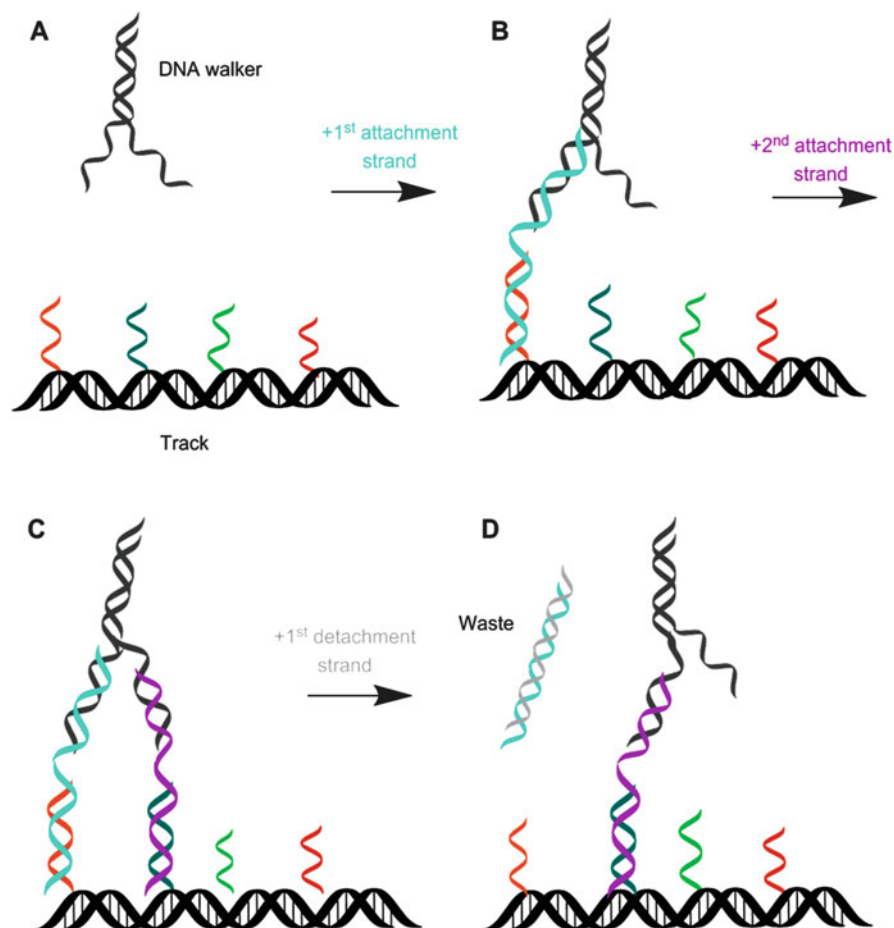


Fig. 2.5 Schematic drawing of the movement of a DNA walker. The *orange*, *dark green*, *green*, and *red* single-stranded branches represent dyes, and the *dark gray* strands of the walker represent quenchers for detection of walker locomotion. In (a), the unbound DNA walker is shown. Addition of the first attachment strand (*light blue*) results in the attachment of the walker to the first branch on the track, as depicted in (b). Upon addition of the second attachment strand (*pink*), the walker attaches to the first two branches with both legs before the first branch is released in form of duplex waste through addition of a detachment strand (*light gray*). Figure adapted from Shin and Pierce (2004)

information in individual molecules and to do a programming. In the examples mentioned before, the interaction of simple robots with their environment was utilized to create devices, which travel in a directional way along short, one-dimensional tracks. Lund et al. could show robotic behavior for so-called molecular spiders, made of an inert streptavidin molecule, which represents the body, and three legs consisting of deoxyribozymes, adapted from the 8-17 DNA enzyme. In contrast to the previously described one-dimensional movement,

spiders are able to move across two-dimensional DNA origami landscapes (Lund et al. 2010). These origami landscapes are self-assembling and consist of a long single-stranded scaffold and short oligonucleotide staple strands, which hold the scaffold in place (Rothenmund 2006). The landscapes can be shaped as desired. Thus, they can be designed in such a way that the molecular spiders move across it, thereby performing a series of actions such as “start,” “follow,” “turn,” and “stop.” The movement of individual spiders was monitored in real time by super-resolution fluorescence video microscopy. The spider is positioned on a start site by a 20-base single-stranded DNA oligonucleotide and released by a single-stranded DNA trigger. Furthermore, the cofactor Zn^{2+} is added to facilitate the cleavage by the 8-17 deoxyribozyme. This enzyme cleaves at an RNA base within the substrate and leads to the formation of two shorter product fragments and the release of a leg, which can then bind to the next substrate. A crucial factor, which is essential for the concept of the molecular spider and provides a simple memory mechanism, is the lower enzyme affinity of the product compared to the substrate. When the deoxyribozyme of the leg of a spider binds to a place where it has been before, it dissociates faster from it than from a new substrate, where it stays bound longer before it finally cleaves it. Consequently, a spider, released at the boundary between products and substrates, moves toward the substrates and follows a linear, directional track during substrate cleavage (Lund et al. 2010).

Previous nanomotors have been mostly based on burn-the-bridge methods, which provide directionality by chemically damaging the traversed track, for example, a nanomotor driven by a nicking enzyme for the transport of DNA cargo (Bath et al. 2005). In contrast to this DNA motor and the DNA walkers described before (Sherman and Seeman 2004; Shin and Pierce 2004), spiders can take Brownian walks across already visited product sites until they run into the next substrates.

As another alternative to burn-the-bridge methods, a DNA motor, based on a bioinspired concept, was recently established, using mechanics-mediated symmetry breaking. The technology relies on local alignment with the track through binding of a pedal and achieves directionality by adjusting the size of the motor. A single action of leg dissociation is enough to drive the motor. The symmetric bipedal nanomotor is able to move continuously along a track with only two different footholds. The concept is designed to be generally applicable for DNA molecules, peptides, or synthetic polymers (Cheng et al. 2014).

The average step size of DNA walkers is around 2–5 nm (Sherman and Seeman 2004; Shin and Pierce 2004). The DNA nanomotor powered by nicking enzymes is moving with a speed of 0.1 nm s^{-1} (Bath et al. 2005). In comparison to that, molecular spiders have been shown to travel around 100 nm and exhibited mean speeds of $1\text{--}6 \text{ nm s}^{-1}$. Although a lot of progress was already made on the field of DNA walkers, there are still several factors, which limit their performance. The traveling distance of molecular spiders is restricted by dissociation and backtracking. Other shortcomings of this concept are that the substrate has to be recharged and that molecular spiders are slower and not as efficient as protein-based walkers. However, the programmability and predictability of DNA walkers make

them attractive research targets for nanoscale robotics with defined interactions with their environment (Lund et al. 2010).

References

- Alper H et al (2005) Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* 102(36):12678–12683
- Ambros V (2004) The functions of animal microRNAs. *Nature* 431(7006):350–355
- Ambros V et al (2003) A uniform system for microRNA annotation. *RNA (New York)* 9(3):277–279
- Annaluru N et al (2014) Total synthesis of a functional designer eukaryotic chromosome. *Science* 344(6179):55–58
- Argaman L et al (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11(12):941–950
- Bagga S et al (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 122(4):553–563
- Bath J, Turberfield AJ (2007) DNA nanomachines. *Nat Nanotechnol* 2(5):275–284
- Bath J, Green SJ, Turberfield AJ (2005) A free-running DNA motor powered by a nicking enzyme. *Angewandte Chemie* 117(28):4432–4435
- Benders GA et al (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res* 38(8):2558–2569
- Berg L et al (2013) Combinatorial mutagenesis and selection to understand and improve yeast promoters. *BioMed Res Int* 2013:926985
- Bernhardt HS (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). *Biol Direct* 7:23
- Blazek J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* 8(1):46–58
- Blazek J et al (2011) Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl Environ Microbiol* 77(22):7905–7914
- Blazek J et al (2012) Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol Bioeng* 109(11):2884–2895
- Brenneke J et al (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113(1):25–36
- Brewster RC, Jones DL, Phillips R (2012) Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *PLoS Comput Biol* 8(12), e1002811
- Cambray G et al (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res* 41(9):5139–5148
- Carmell MA et al (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* 16(21):2733–2742
- Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci* 25(10):481–482
- Chang S et al (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* 430(7001):785–789
- Chen JH, Seeman NC (1991) Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* 350(6319):631–633
- Chen Y-J et al (2013) Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* 10(7):659–664
- Cheng J et al (2014) A bioinspired design principle for DNA nanomotors: mechanics-mediated symmetry breaking and experimental demonstration. *Methods (San Diego)* 67(2):227–233

- Collins JA et al (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev* 21 (24):3356–3368
- Comstock J (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U S A* 80:21–25
- Coussement P et al (2014) One step DNA assembly for combinatorial metabolic engineering. *Metab Eng* 23:70–77
- Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. *Mol Syst Biol* 3(145):145
- Crook NC, Freeman ES, Alper HS (2011) Re-engineering multicloning sites for function and convenience. *Nucleic Acids Res* 39(14), e92
- Curran KA et al (2013) Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab Eng* 19:88–97
- Davis JH, Rubin AJ, Sauer RT (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* 39(3):1131–1141
- De Mey M et al (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol* 7:34
- Doudna JA, Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* 418 (6894):222–228
- Douglas SM, Bachelet I, Church GM (2012) A logic-gated nanorobot for targeted transport of molecular payloads. *Science (New York)* 335(6070):831–834
- Durand G et al (2014) Riboswitches based on kissing complexes for the detection of small ligands. *Angewandte Chemie* 53(27):6942–6945 (International ed. in English)
- Dwarakanath S et al (2004) Quantum dot-antibody and aptamer conjugates shift fluorescence upon binding bacteria. *Biochem Biophys Res Commun* 325(3):739–743
- Edelman GM et al (2000) Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity. *Proc Natl Acad Sci* 97(7):3038–3043
- Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818–822
- Endo M et al (2012) Transcription regulation system mediated by mechanical operation of a DNA nanostructure. *J Am Chem Soc* 134(6):2852–2855
- Erdeniz N, Mortensen UH, Rothstein R (1997) Cloning-free PCR-based allele replacement methods. *Genome Res* 7(12):1174–1183
- Esteller M (2011) Non-coding RNAs in human disease. *Nat Rev Genet* 12(12):861–874
- Fedor MJ (2009) Comparative enzymology and structural biology of RNA self-cleavage. *Annu Rev Biophys* 38:271–299
- Ferré-D'Amaré AR, Scott WG (2010) Small self-cleaving ribozymes. *Cold Spring Harb Perspect Biol* 2(10):a003574
- Fire A et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806–811
- Frézal L, Leblois R (2008) Four years of DNA barcoding: Current advances and prospects. *Infect Genet Evol* 8(5):727–736
- Gao Y, Zhao Y (2014) Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* 56(4):343–349
- Gehrke S, Jérôme V, Müller R (2003) Chimeric transcriptional control units for improved liver-specific transgene expression. *Gene* 322:137–143
- Giaever EA (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418 (6896):387–391
- Gibson DG et al (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science (New York)* 319(5867):1215–1220
- Guazzaroni M-E, Silva-Rocha R (2014) Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators. *ACS Synth Biol* 3(9):666–675

- Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries—tuning of gene expression. *Trends Biotechnol* 24(2):53–55
- Hammond SM et al (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404(6775):293–296
- Hartner FS et al (2008) Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res* 36(12), e76
- Haurwitz RE et al (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York)* 329(5997):1355–1358
- Havill JT et al (2014) A new approach for detecting riboswitches in DNA sequences. *Bioinformatics (Oxford)* 30(21):3012–3019
- Hebert PDN et al (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* 270(1512):313–321
- Hermann T, Patel DJ (2000) Adaptive recognition by nucleic acid aptamers. *Science* 287(5454):820–825
- Jäschke A (2001) Artificial ribozymes and deoxyribozymes. *Curr Opin Struct Biol* 11(3):321–326
- Jenison RD et al (1994) High-resolution molecular discrimination by RNA. *Science (New York)* 263(5152):1425–1429
- Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* 64(1):82–87
- Jensen K et al (2006) Identifying functionally important mutations from phenotypically diverse sequence data. *Appl Environ Microbiol* 72(5):3696–3701
- Johnston RJ, Hobert O (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* 426(6968):845–849
- Juven-Gershon T, Cheng S, Kadonaga JT (2006) Rational design of a super core promoter that enhances gene expression. *Nat Methods* 3(11):917–922
- Kagiya G et al (2005) Generation of a strong promoter for *Escherichia coli* from eukaryotic genome DNA. *J Biotechnol* 115(3):239–248
- Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 6(5):376–385
- Kinney JB et al (2010) Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci U S A* 107(20):9158–9163
- Kranthi BV et al (2009) Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*. *Biochim Biophys Acta* 1789(6-8):460–468
- Kruger K et al (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31(1):147–157
- Kubodera T et al (2003) Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett* 555(3):516–520
- Künne T, Swarts DC, Brouns SJJ (2014) Planting the seed: target recognition of short guide RNAs. *Trends Microbiol* 22(2):74–83
- Kuzuya A, Ohya Y (2014) Nanomechanical molecular devices made of DNA origami. *Acc Chem Res* 47(6):1742–1749
- Kuzuya A et al (2011) Nanomechanical DNA origami “single-molecule beacons” directly imaged by atomic force microscopy. *Nat Commun* 2:449
- Landon PB et al (2012) DNA zipper-based tweezers. *Langmuir* 28(1):534–540
- Li M et al (2012) A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. *Microb Cell Fact* 11:19
- Lou C et al (2012) Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotechnol* 30(11):1137–1142
- Lubliner S, Keren L, Segal E (2013) Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res* 41(11):5569–5581

- Lund K et al (2010) Molecular robots guided by prescriptive landscapes. *Nature* 465 (7295):206–210
- Mairhofer J et al (2013) Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions. *Appl Environ Microbiol* 79(12):3802–3812
- Mairhofer J et al (2015) Preventing T7 RNA polymerase read-through transcription – a synthetic termination signal capable of improving bioprocess stability. *ACS Synth Biol* 4(3):265–273
- Martinez J et al (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110(5):563–574
- Melnikov A et al (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* 30(3):271–277
- Moazed D (2009) Small RNAs in transcriptional gene silencing and genome defence. *Nature* 457 (7228):413–420
- Moore I, Samalova M, Kurup S (2006) Transactivated and chemically inducible gene expression in plants. *Plant J* 45(4):651–683
- Mutalik VK et al (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* 10(4):354–360
- Naville M, Gautheret D (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biol* 11(9):R97
- Nevoigt E et al (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 72(8):5266–5273
- Nielsen ML et al (2006) Efficient PCR-based gene targeting with a recyclable marker for *Aspergillus nidulans*. *Fungal Genet Biol* 43(1):54–64
- Nielsen AAK, Segall-Shapiro TH, Voigt CA (2013) Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr Opin Chem Biol* 17 (6):878–892
- Nissen P (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289 (5481):920–930
- Patwardhan RP et al (2009) High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* 27(12):1173–1175
- Penchovsky R (2013) Computational design and biosensor applications of small molecule-sensing allosteric ribozymes. *Biomacromolecules* 14(4):1240–1249
- Pfleger BF et al (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* 24(8):1027–1032
- Platt T (1986) Transcription termination and the regulation of gene expression. *Annu Rev Biochem* 55:339–372
- Prody GA et al (1986) Autolytic processing of dimeric plant virus satellite RNA. *Science (New York)* 231(4745):1577–1580
- Ramesh A, Winkler WC (2014) Metabolite-binding ribozymes. *Biochim Biophys Acta* 1839 (10):989–994
- Rana TM (2007) Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol* 8(1):23–36
- Reid RJD et al (2002) Efficient PCR-based gene disruption in *Saccharomyces* strains using intergenic primers. *Yeast (Chichester)* 19(4):319–328
- Reinemann C, Strehlitz B (2014) Aptamer-modified nanoparticles and their use in cancer diagnostics and treatment. *Swiss Medical Weekly* 144(January):w13908
- Reinhart BJ et al (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403(6772):901–906
- Remans T et al (2005) Functional promoter analysis using an approach based on an in vitro evolution strategy. *BioTechniques* 38(2):209–10, 212, 214–216
- Rivas E, Eddy SR (2000) Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics* 16(7):583–605

- Rivas E et al (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 11(17):1369–1373
- Ross W et al (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J Bacteriol* 180(20):5375–5383
- Roth A et al (2014) A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* 10(1):56–60
- Rothmund PWK (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440(7082):297–302
- Ruth C, Glieder A (2010) Perspectives on synthetic promoters for biocatalysis and biotransformation. *Chembiochem* 11(6):761–765
- Ruth C et al (2010) Variable production windows for porcine trypsinogen employing synthetic inducible promoter variants in *Pichia pastoris*. *Syst Synth Biol* 4(3):181–191
- Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science (New York)* 289(5478):448–452
- Seehafer C et al (2011) From alpaca to zebrafish: hammerhead ribozymes wherever you look. *RNA (New York)* 17(1):21–26
- Seeman NC (1998) DNA nanotechnology: novel DNA constructions. *Annu Rev Biophys Biomol Struct* 27:225–248
- Seeman NC (2003) DNA in a material world. *Nature* 421(6921):427–431
- Seeman NC (2010) Nanomaterials based on DNA. *Annu Rev Biochem* 79:65–87
- Sharon E et al (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat Biotechnol* 30(6):521–530
- Sherman WB, Seeman NC (2004) A precisely controlled DNA biped walking device. *Nano Lett* 4(7):1203–1207, Available at: <http://pubs.acs.org/doi/abs/10.1021/nl049527q>
- Shih WM, Quispe JD, Joyce GF (2004) A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* 427(6975):618–621
- Shin JS, Pierce NA (2004) A synthetic DNA walker for molecular transport. *J Am Chem Soc* 126(35):10834–10835
- Solem C, Jensen PR (2002) Modulation of gene expression made easy. *Appl Environ Microbiol* 68(5):2397–2403
- Song J-J et al (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science (New York)* 305(5689):1434–1437
- Stuwe E, Tóth KF, Aravin AA (2014) Small but sturdy: small RNAs in cellular memory and epigenetics. *Genes Dev* 28(5):423–431
- Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA (New York)* 9(6):644–647
- Tahiri-Alaoui A et al (2002) High affinity nucleic acid aptamers for streptavidin incorporated into bi-specific capture ligands. *Nucleic Acids Res* 30(10), e45
- Tian Y et al (2005) A DNzyme that walks processively and autonomously along a one-dimensional track. *Angewandte Chemie* 44(28):4355–4358 (International ed. in English)
- Timmons L, Fire A (1998) Specific interference by ingested dsRNA. *Nature* 395(6705):854
- Tor Y, Hermann T, Westhof E (1998) Deciphering RNA recognition: aminoglycoside binding to the hammerhead ribozyme. *Chem Biol* 5(11):R277–R283
- Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* 15(3):342–348
- Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505–510
- Tyo KEJ, Nevoigt E, Stephanopoulos G (2011) Directed evolution of promoters and tandem gene arrays for customizing RNA synthesis rates and regulation. *Methods Enzymol* 497:135–155
- Verdel A et al (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science (New York)* 303(5658):672–676
- Vogl T, Hartner FS, Glieder A (2013) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. *Curr Opin Biotechnol* 24(6):1094–1101

- Vogl T et al (2014) Synthetic core promoters for *Pichia pastoris*. *ACS Synth Biol* 3(3):188–191
- Wachsmuth M et al (2013) De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res* 41(4):2541–2551
- Wahlestedt C (2013) Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nat Rev Drug Discov* 12(6):433–446
- Walter G (1986) Origin of life: the RNA world. *Nature* 319(20):618
- Wassarman KM et al (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15:1637–1651
- Weiberg A et al (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science (New York)* 342(6154):118–123
- Werstuck G, Green MR (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science* 282(5387):296–298
- Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol* 9(6):594–602
- Winkler WC, Breaker RR (2005) Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* 59:487–517
- Winkler WC et al (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428(6980):281–286
- Yamazaki T et al (2012) Clear-cut observation of PNA invasion using nanomechanical DNA origami devices. *Chem Commun (Camb)* 48(92):11361–11363
- Yan H et al (2002) A robust DNA mechanical device controlled by hybridization topology. *Nature* 415(6867):62–65
- Yanofsky C (1981) Attenuation in the control of expression of bacterial operons. *Nature* 289(26):751–758
- Yekta S, Shih I-H, Bartel DP (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science (New York)* 304(5670):594–596
- Yin P et al (2004) A unidirectional DNA walker that moves autonomously along a track. *Angewandte Chemie* 43(37):4906–4911 (International ed. in English)
- Yurke B et al (2000) A DNA-fuelled molecular machine made of DNA. *Nature* 406(6796):605–608
- Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. *Science (New York)* 309(5740):1519–1524
- Zeng Y, Yi R, Cullen BR (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* 100(17):9779–9784