RNA-based engineering of inducible CRISPR-Cas9 transcription factors for *de novo* assembly of eukaryotic gene circuits

Quentin R.V. Ferry

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

RNA-based engineering of inducible CRISPR-Cas9 transcription factors for *de novo* assembly of eukaryotic gene circuits

Quentin R.V. Ferry, St Cross College, DPhil in Genomic Medicine and Statistics, Michaelmas 2017

Abstract

Synthetic biology in mammalian cells holds great promise for reverse engineering biological processes and rewiring cellular behaviors for therapeutic purpose. An essential aspect in our ability to reprogram the cellular code is the availability of highly orthogonal, inducible transcriptional regulators. CRISPR-based strategies employing effector-domain tethering to the single guide RNA (sgRNA)-dCas9 complex have greatly advanced this field by allowing for precise activation or repression of any gene via simple sgRNA reprograming. However, the implementation of inducible CRISPR-based transcriptional regulators (CRISPR-TRs) has so far been restricted to dCas9 protein engineering and conditional effector tethering. Although elegant, these approaches are limited by dCas9 promiscuous loading of sgRNAs, which hinders their use for the creation of independent multi-gene transcriptional programs.

To address this limitation, I have developed a modular framework for the rational design of inducible CRISPR-TR, based on simple and reversible modifications of the sgRNA sequence. At the core of this conceptual framework lies the ability to inactivate native sgRNAs by appending on their 5'-end a short RNA segment, which folds to form a spacer-blocking hairpin (SBH). Base-pairing between the extension and the sgRNA spacer prevents docking of the CRISPR-TR on-target, fully abrogating its activity. Subsequently, I have created inducible SBH variants (iSBH) by replacing the hairpin loop with conditional RNA cleaving units. Using a variety of sensing-loops, I was able to engineer a panel of switchable iSBH-sgRNAs, designed to activate specifically in the presence of protein, oligonucleotide, and small molecule inducers. Leveraging the versatility of this method, I demonstrate that iSBH-sgRNAs expression can be multiplexed to assemble synthetic gene circuits implementing parallel and orthogonal regulation of multiple endogenous gene targets. Finally, I have distilled the design principles derived throughout this project to develop a web tool that automates the creation of iSBH-sgRNAs. Already a valuable addition to the synthetic biology toolkit, iSBH-based inducibility should in theory also be applicate to all CRISPR-Cas9 derivatives (genome editing, epigenetic alteration, DNA labelling, etc.) as well as other newly characterized RNA-guide nucleases from the CRISPR family.

Declaration of Authorship

I hereby declare that the submitted Thesis, as well as the entirety of the research it contains, is the product of my own original work unless stated otherwise. I am well aware of the Universtiy's regulations concerning plagiarism and certify that the rightful authors for any use of text, data, figure, or idea, which was not mine, are acknowledged. All experiments reported in this document were conducted at the Weatherall Institute of Molecular Medicine (University of Oxford) as part of the DPhil in Genomic Medicine and Statistics. None of the material included in this Thesis has been submitted as part of any other degree or professional application.

Quentin R.V. Ferry

Acknowledgements

First and foremost, I would like to thank my supervisor, Tudor Fulga, for inviting me in his laboratory and giving me – an engineer – a chance at tinkering with life. His ability to celebrate my “ups” and get me through the “downs” have certainly made the journey worth remembering. Thank you for inspiring me, instilling in me values that helped me grow as a scientist, and giving me a mentor to work with rather than work for.

For providing me with a stimulating working environment and a family for four years, I would like to express my sincere gratitude to all members of the Fulga lab: Tim Rajakumar, for his contagious enthusiasm and teaching me most of the methods used to complete this work; Qianxin Wu, for giving me the opportunity to contribute to the fascinating GenERA project; Toni Bäumler and Bruno Steinkraus, for driving the Circulus Therapeutics team; Yale Michaels, Markus Toegel, and David Knapp for their pertinent questions and valuable feedbacks; Aron Szabo for nurturing my passion for the neurosciences; Alice Lightowlers for her strength and positive attitude; and finally, Max Jamilly for the occasional French lesson.

Additionally, I would also like to thank Radostina Lyutova, who helped me with cloning and transfections for a summer, as well as all the WIMM staff for facilitating this work. I am particularly grateful to Paul Sopp and Kevin Clark, for patiently answering all my questions regarding flow cytometry, and Simon McGowan for kindly helping me upload the iSBHfold web tool onto the institute’s server. Thanks also to the TSS and Milne’s lab, especially Tatjana and Tom, for their guidance throughout this project and giving wise advices for my scientific career. Finally, I am much obliged to Jonathan Flint and Richard Mott at the WTCHG (Oxford), who took me aboard the DTC in Genomic Medicine and Statistics, as well as the Wellcome Trust for financing this entire DPhil.

I would like to finish by acknowledging my family, friends, and partner to whom I owe my resilience. Your support has been invaluable throughout. I notably owe a debt of gratitude to Ioanna Rota for disguising free counselling into “fruit breaks”, and Angela Lee for all the sleepless nights spent proofreading this Thesis.

Publications

Ferry, Q.R.V., Lyutova, R. & Fulga, T.A., 2017. *Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs*. **Nature communications**, 8, p.14633.

This manuscript reports the work presented in this Thesis. Accordingly, the material presented here has for the most part been presented in this publication. As a general note to the reader, I would like to acknowledge the fact that, despite my best efforts to generate original figures, the data as well as some graphical elements are shared between the two documents.

Wu, Q.[#], **Ferry, Q.R.V.**[#], Baeumler, T.A., Michaels, Y.S., Vitsios, D.M., Habib, O., Arnold, R., Jiang, X., Maio, S., Steinkraus, B.R., Tapia, M., Piazza, P., Xu, N., Holländer, G.A., Milne, T.A., Kim, J.S., Enright, A.J., Bassett, A.R., Fulga, T.A., 2017*. In situ functional dissection of a native miRNA target-network by multiplex genome engineering*. **Nature communications**, 8(1), p.2109.

In this paper, I have collaborated with postdoctoral scientist Qianxin Wu to develop GenERA, a CRISPR/Cas9-based platform for high-content phenotypic investigation of intragenic non-coding RNA cis-regulatory elements (RREs). GenERA represents a unique addition to the repertoire of genome editing applications and the first account of a CRISPR-based multiplex technology that enables direct coupling of genome editing events (NHEJ-based mutagenesis) to phenotypic variations in gene expression levels. Using this technology, we were able to provide unprecedented insights into the regulation of a complete miRNA target network, dissect the entire post-transcriptional regulatory landscape encoded within a candidate 3'UTR, and analyze at near-single nucleotide resolution the sequence determinants underlying RRE functionality.

Patent

Inducible guide RNAs for genome editing (iSBH). United Kingdom Patent Application No. 1701180.0, Oxford University Innovation Limited.

List of Abbreviations

4OHT	4-hydroxytamoxifen
a-sgRNA	allosteric sgRNA
aHHRz	allosteric hammerhead ribozyme
ASL	ASO sensing-loop
ASO	antisense oligonucleotide
asRNA	antisense RNA
bp	base-pair
C-terminus	carboxyl-terminus
Cas	CRISPR-associated [protein]
cDNA	complementary DNA
ChiP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation followed by sequencing
CMV	cytomegalovirus
CMVp	CMV promoter
CRISPR	clustered regularly interspaced short palindromic repeats
CRISPR-TR	CRISPR-based transcriptional regulator
CRISPRa	CRISPR-mediated transcriptional activation
CRISPRe	CRISPR-based genome editing
CRISPRi	CRISPR interference
CRISPRr	CRISPR-mediated transcriptional repression
CRISPRr	CRISPR-mediated gene regulation
crRNA	CRISPR RNA
crRNP	CRISPR ribonucleoprotein complex
CRY2	light-sensitive cryptochrome 2
CTS	CRISPR target site
dCas9	catalytically inactive (dead) Cas9
DD	dimerization domain
DNA	Deoxyribonucleic acid
dsRNA	double-stranded RNA
E. Coli	Escherichia coli
ECFP	enhanced cyan fluorescent protein

EGFP	enhanced green fluorescent protein
ER-LBD	ligand-binding domain of the human estrogen receptor-alpha
EXFP	any of EYFP, ECFP
EYFP	enhanced yellow fluorescent protein
FKBP	FK506 binding protein 12
FRB	FKBP rapamycin binding [domain]
GFP	green fluorescent protein
GOI	gene of interest
gRNA	[CRISPR] guide RNA
HEK	human embryonic kidney
HHRz	hammerhead ribozyme
HSF1	human heat-shock factor 1
IDT	Integrated DNA technology
IPTG	isopropyl-β-d-thiogalactoside
iSBH	inducible spacer-blocking hairpin
KRAB	Kruppel-associated box [domain]
Kt	kink turn
LacI	lac repressor
LNA	locked nucleic acids
mCMVp	minimal CMV promoter
MCP	MS2 coat proteins
miRNA	micro RNA
MRE	micro RNA response element
mRNA	messenger RNA
N-terminus	amino-terminus
NEB	New England Biolabs
NES	nuclear export signal
NHEJ	non-homologous end joining
NLS	nuclear localization signal
nt	nucleotide
nv	native
p-sgRNA	protected sgRNA
PAM	protospacer adjacent motif

PBAD	arabinose-induced promoter
PCR	polymerase chain reaction
PD-L1	programmed death-ligand 1
pDNA	plasmid DNA
PFS	protospacer flanking sequence
phage	bacteriophage
PI	PAM interaction [domain]
Pol	RNA polymerase
pre-crRNA	precursor crRNA
PTR	programmable transcriptional regulator
PTS	protein target site
RNA	ribonucleic acid
RNP	[CRISPR-] ribonucleoproteins
RRE	RNA cis-regulatory elements
RT-qPCR	quantitative reverse transcription PCR
Rta	Epstein-Barr virus R transactivator
rtTA	reverse tetracycline transactivator
RVD	repeat variable diresidue
SAM	synergistic activation mediator
SBH	spacer-blocking hairpin
sgRNA	single guide RNA
Sp	*Streptococcus pyogenes*
*Sp*Cas9	*Streptococcus pyogenes* Cas9
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
st-sgRNA	self-targeting sgRNA
SVG	scalable vector graphics
T7E1	T7 Endonuclease I
TALE	transcription activator-like effector
tetO	tet operator /operon
tetR	tet repressor
tracrRNA	trans-activating crRNA
TRE	tetracycline response element

TSS	transcription start site
tTA	tetracycline transactivator
U6p	U6 promoter
UAS	upstream activation sequence
UTR	untranslated region [of a gene]
VPR	VP64-p65-Rta
ZF	zinc-finger

Table of Contents

List of figures

Chapter 1 – Introduction

1.1 – Synthetic biology

Following the development of molecular cloning techniques in the 1970s and 1980s, and the rise of high-throughput biology ("omic" era) in the 1980s onwards, the field of molecular biology has progressively shifted from discovery-driven science to hypothesis-driven research. Alongside more traditional approaches to biology, entirely new fields have emerged based on the overarching idea that life could be understood by attempting to re-engineer it (Elowitz & Lim 2010; Khalil & Collins 2010). As such, those new areas of research were together labelled as "synthetic biology". While the term has been used in various ways since its first appearance in the literature in 1980 (Hobom 1980), synthetic biology now describes all disciplines aimed at building artificial biological systems for research purposes, as well as industrial and medical applications (Khalil et al. 2012). As of today, efforts in the field of synthetic biology can be broadly categorised as either (i) synthetic genomics, (ii) synthetic protocell biology, (iii) bioengineering, or (iv) unnatural molecular biology (Deplazes 2009).

Synthetic genomics, as a discipline, aims at engineering, using a reductionist approach, "minimal" cells that could be used as "chassis" organism to create new life forms. This top-down approach is best exemplified by the work of Craig Venter and colleagues who recently created a new *Mycoplasma mycoides* strain running on a fully chemically synthesised "minimal" genome (Hutchison et al. 2016). This genome was designed to only contain what the group identified as being the minimal set of genes required for the bacteria to be viable.

On the other hand, the field of synthetic protocell biology is following a bottom-up strategy to explain the origin of life. Led by the pioneering work of Szostak and colleagues, scientists are encapsulating various types of biomolecules in self-organising lipid vesicles with the aim of creating a protocell capable of self-replication, self-maintenance, and evolution (Solé et al. 2007).

Assuming that we can one day understand how life started, and we successfully delineate the minimal set of genes required for a unicellular organism to exist, one then has to explain how individual biological parts like genes, gene products, and regulatory elements, come together to form complex molecular pathways. In order to provide answers to this question, the field of bioengineering is cataloguing and characterizing the function of small biological parts isolated from various organisms (Lienert et al. 2014). The rationale behind this effort is that one can gain valuable insights into the function of these parts by assembling then into synthetic gene circuits, and comparing the predicted behaviour of such circuits against empirical measurements (Kobayashi et al. 2004). As it will transpire in the next section, the field of bioengineering not only provide methodologies to understand biology but is also assembling a growing toolbox of biological parts that is being used to reprogram cellular behaviours (Lienert et al. 2014).

Finally, the field of unnatural molecular biology seeks to engineer enhanced life forms capable of using either unnatural DNA/RNA nucleotides and/or unnatural amino acids (Chin 2017). The aim of this discipline is to endow cells with an extended and/or orthogonal genetic code that allows for an expansion of their proteome with designer proteins that perform unnatural functions (Davis & Chin 2012). Inherent to the four fields aforementioned, a fifth branch of synthetic biology, referred to as *in silico* synthetic biology, seeks to establish computational models that both explain and predict the behaviours of synthetic biological systems (Deplazes 2009).

Given that the work presented in this thesis falls into the bioengineering category, I present in the next section a brief history of gene circuit design as a field, its applications, as well as an overview of the toolbox of characterized biological parts available for mammalian systems. Of note, the term synthetic biology will be interchangeably used to refer to bioengineering in the remainder of this document.

1.2 – Synthetic gene circuit

Akin to computers following a set the instructions written in code, prokaryotic and eukaryotic cells run complex biological programs encoded in the interactions between the DNA, RNAs, and the proteins they contain. In an effort to understand these biological processes and ultimately be able to reprogram cellular behaviour, synthetic biologists are taking apart molecular pathways, isolating their building blocks, and characterising the function of these individual biological parts in a simpler and more controlled exogenous context (Elowitz & Leibler 2000). As eluded to in the previous section, the methodology entails assembling several parts into a synthetic gene circuit, whose behaviour can be predicted via *in silico* modelling. The theoretical behaviour is then tested experimentally by delivering the gene circuit to cells as a plasmid DNA (pDNA) and assaying for phenotypes. Subsequently, divergences between the predicted and the observed behaviour helps refine our understanding of the biological parts involved (Elowitz & Leibler 2000).

The first synthetic gene circuits were published in 2000 (D. E. Cameron et al. 2014). That year, Gardner et al. reported the construction of a “toggle switch” in bacteria, whereby the expression of a reporter gene could be toggled between active and inactive transcriptional states based on two distinct stimuli (Gardner et al. 2000) (Fig. 1.1a). The authors showed that the engineered cells could maintain a memory of the switching event across multiple divisions, even if the stimulus had been removed. This circuit was built using a rather simple biological part, borrowed from the bacterial kingdom: an inducible promoter whose transcriptional output is controlled by the conditional binding of a repressor protein. In the specific case of Gardner’s

work, two inducible promoters were created by flanking the TATA box of a constitutive promoter with docking DNA sequences for either the *lacI* or *lc* transcriptional repressors (see section 1.3 for details, Fig. 1.1a). Using the delivery of the small molecule Isopropyl-β-d-thiogalactoside (IPTG) or heat, which inhibit *lacI* and *cI* binding respectively, the authors were able to trigger activation of the corresponding promoter. The toggle switch was then assembled by placing *lacI* under the control of the *lc*-responsive promoter and expressing *lc* from the *lacI*-responsive promoter, so as to create a situation of mutually exclusive expression (Fig. 1.1a).

Using similar parts, Elowitz and Leibler devised a more complex circuit to generate oscillatory expression of a reporter gene (Elowitz & Leibler 2000) (Fig. 1.1b), while Becskei and Serrano created an autoregulatory negative-feedback circuit (Becskei & Serrano 2000) (Fig. 1.1c). As exemplified by these three studies, most of the pioneering work in synthetic biology focused on reprogramming micro-organisms or unicellular organisms, with *Escherichia coli* (*E. Coli*) being the most common model system. This choice was mainly motivated by the fact that the molecular biology of *E.Coli* had been fairly well characterized and technologies to manipulate its genetic makeup were readily available (Blount 2015).

As the field progressed, a growing repertoire of biological parts also became available for reprogramming mammalian cells (Lienert et al. 2014; Black et al. 2017). These devices were created to control and interfere with a broad range of biological processes including cellular signaling (Fig. 1.2a), protein activity and turn over (Fig. 1.2b, c), translation (Fig. 1.2d), mRNA stability (Fig. 1.2e), alternative splicing (Fig. 1.2f), and transcription (Fig. 1.2g, see sections 1.3, 1.5). Additionally, methods were developed to allow precise editing of the genome (Fig. 1.2h, see sections 1.3, 1.4).

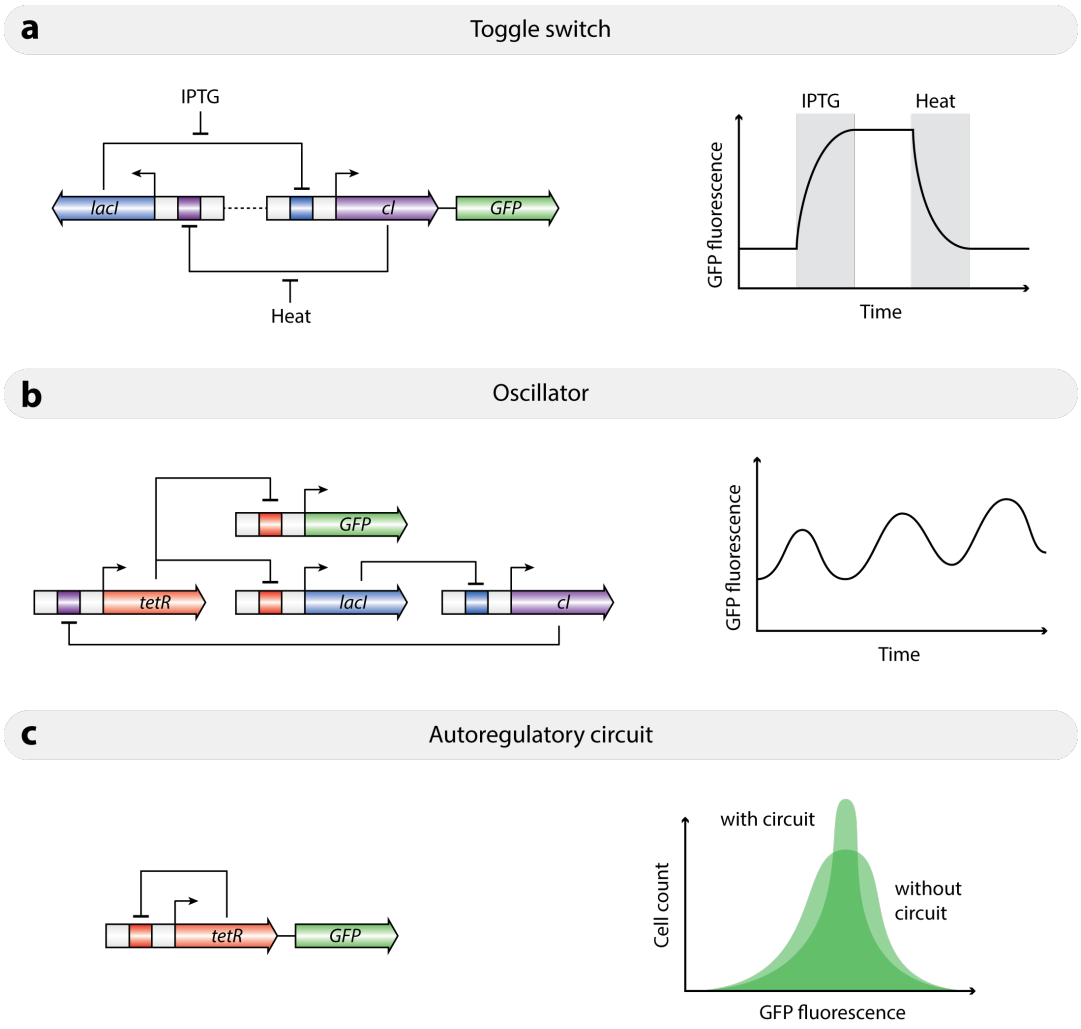

Figure 1.1 – Birth of synthetic biology, the first synthetic circuits. *Figure adapted from (D. E. Cameron et al. 2014)*. **(a)** Schematic of a gene circuit implementing a toggle switch. Target sequences for the *lacI* and *cI* transcriptional repressors are used to create responsive promoters which drive the expression of the opposite repressor. Consequently, only one of the two proteins can be expressed in the cell at a given time (mutually exclusive expression). Cell state can be toggled between *lacI* dominant and *cI* dominant (GFP expression) by providing external stimuli which temporarily inhibit *lacI* (IPTG = isopropyl-β-d-thiogalactoside) or *cI* (heat) binding. **(b)** Oscillatory circuit: three repressor:promoter pairs (*tetR*, *lacI*, *cI*) are organised in a loop with *tetR* controlling reporter expression (GFP). This cascade leads to the formation of oscillations in GFP fluorescence which can be observed at the single cell level using time-lapse fluorescence microscopy (right). **(c)** Autoregulatory circuit: Expression of GFP is tightly controlled using a negative feedback loop (tetR). This has for effect to reduce cell to cell variation in GFP expression as depicted by the histogram of GFP fluorescence (right).

Notably, synthetic receptors were engineered to reprogram cellular signaling by modifying either the input-sensing extracellular domain or the intracellular signaling modules of naturally existing receptors (Dong et al. 2010; Struhl & Adachi 1998; Gross et al. 1989) (Fig. 1.2a). Various strategies were also developed to control the activity of proteins of interest. These included creating split versions of enzymes conditionally re-assembled via the use of ligand-dependent dimerization domains (DD) (Hathaway et al. 2012; Konermann et al. 2013) or self-splicing protein fragments called inteins (Buskirk et al. 2004) (Fig. 1.2b). Protein availability was also successfully regulated by fusing the enzyme of interest with protein fragments, referred to as degrons, which conditionally favour proteasomal degradation (Nishimura et al. 2009; Bonger et al. 2011; Banaszynski et al. 2006) (Fig. 1.2c). Upstream, regulation of protein translation was achieved by incorporating protein binding sites in the 5'-untranslated region (UTR) of the corresponding messenger RNA (mRNA)(Ausländer et al. 2012). In this case, the authors showed that the presence of the cognate RNA binding protein would efficiently silence translation of the transgene (Ausländer et al. 2012) (Fig. 1.2d).

At the RNA level, various research groups were able to conditionally alter the levels of synthetic transcripts bearing in their 3'-UTR specific micro-RNA response elements (Xie et al. 2011) (see section 1.6.2) or ligand-controlled allosteric self-cleaving ribozymes (Win & Smolke 2008) (see chapter 7) (Fig. 1.2e). Culler and colleagues also shown that mRNA sequences could be altered to perform preferential alternative splicing (Culler et al. 2010) (Fig. 1.2f). Transcript levels can also be regulated at the transcriptional level using synthetic transcriptional regulators capable of recruiting the RNA polymerase (pol)-II machinery or repressing transcription initiation (Brown et al. 1987; Gossen & Bujard 1992; Maeder et al. 2008; Morbitzer et al. 2010; Hsu et al. 2014) (Fig. 1.2g). Full detail on this class of devices is given in the rest of the introduction (see section 1.3). Finally, precise alteration of the genome has been made possible through the adaption of bacterial site-specific recombinases (García-Otín & Guillou 2006) and the development of programmable nucleases (H. Kim & J.-S. Kim 2014) (see sections 1.3, 1.4) (Fig. 1.2g).

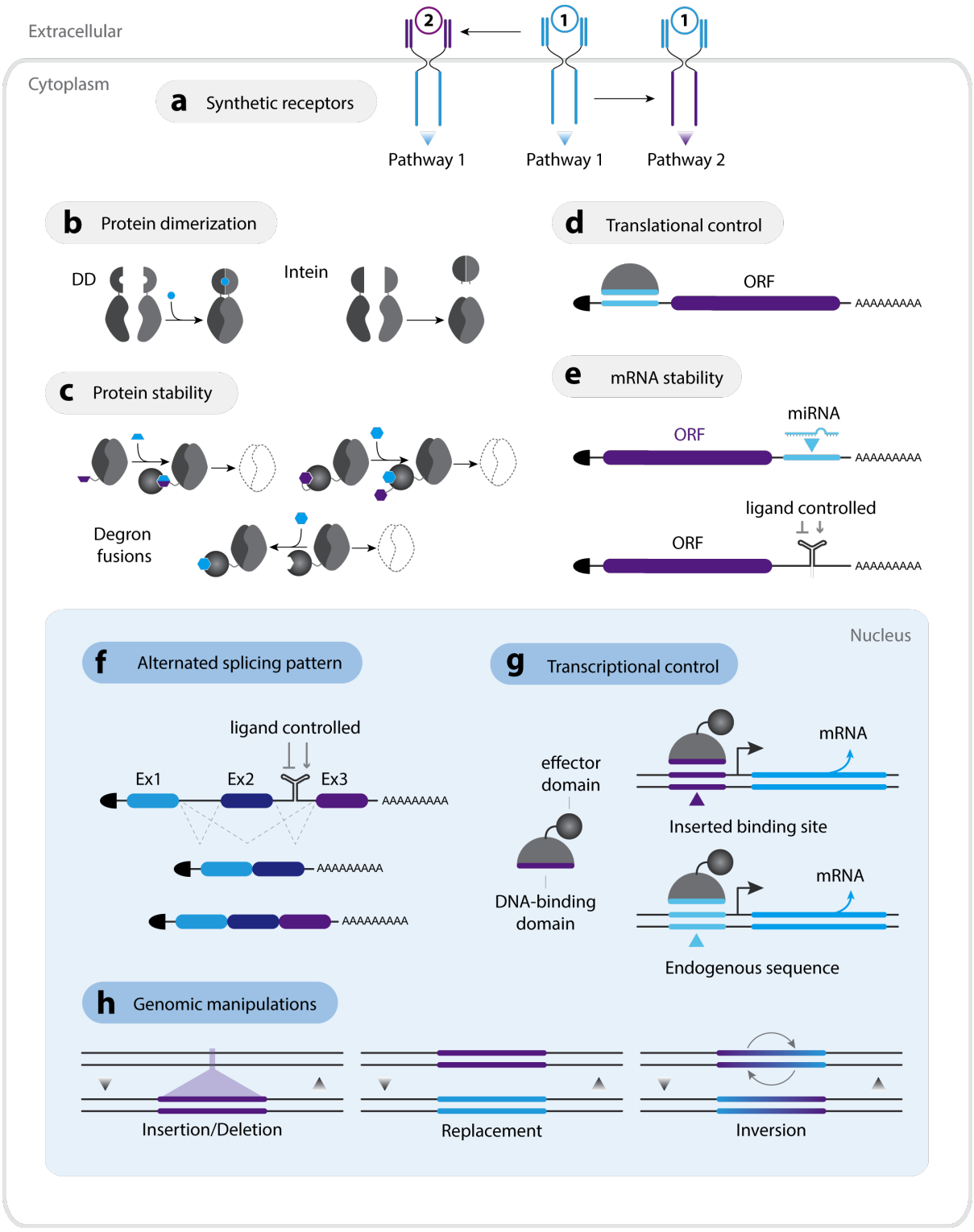

Figure 1.2 – Mammalian synthetic biology, biological parts overview.(a-h) Schematic representation of the various biological parts created to interfere with signaling pathways (a), enzymatic activity (b), protein availability (c), translation (d), mRNA stability (e), mRNA production, (g) alternative splicing, (h) genetic information. DD: dimerization domain; Ex: exon; ORF: open reading frame; miRNA micro RNA.

These biological parts are now being routinely used in laboratories to answer a wide range of biological questions (Lienert et al. 2014). The general approach entails generating perturbations in cellular homeostasis and studying the ways in which cells react to them. Notably, most of the "devices" presented in figure 1.2 have been used to conditionally remove a cellular component, such as a DNA sequence with regulatory potential, mRNA, specific splicing variant, protein, etc., while assessing the resulting phenotype. Moreover, this toolkit also provides means to analyze complex cellular processes by breaking them down and re-assembling subparts in isolation. To understand gene regulatory networks[1] for example, motifs from native networks can be recreated with synthetic parts and easily characterized using reporter genes (Riccione et al. 2012). Similarly, signaling pathways can be analyzed and better understood by isolating a few critical components and building a minimal version in an exogenous context (e.g. reconstitution of the T-cell receptor (TCR) signaling pathway in non-immune cells) (James & Vale 2012). Finally, epigenetics studies have relied on DNA binding proteins fused with chromatin re-modelers to unravel the function of various markers and to delineate enhancers and their targets (Kungulovski & Jeltsch 2016).

In parallel, synthetic biology approaches are also being used on the medical front to improve existing therapies and devise new ones (Kis et al. 2015; Ye & Fussenegger 2014; Black et al. 2017). A few examples chosen to reflect the breadth of such applications include: (i) Engineering the micro-organism *Saccharomyces cerevisiae*, better known as the baker's yeast, to produce the antimalaria drug Artemisinin (Paddon et al. 2013); (ii) Designing chimeric antigen receptor-based T-cell (CAR T-cells) therapy (B. Lim et al. 2017), in which T cells are engineered *ex vivo* before being delivered to the patient to express a synthetic T-cell receptor that target a desired cancer antigen; (iii) Improving gene therapy by providing better ways to target the delivery of exogenous genetic material to a particular cell type (e.g. virus engineering to specifically target retinal cells (Dalkara et al. 2013)) as well as control the tissue specificity

[1] Set of molecular regulators interacting with each other to control gene expression levels of mRNA and proteins (Thompson et al. 2015).

of transgene expression (Whitfield et al. 2012); (iv) Maintaining uric acid homeostasis in the bloodstream by delivering designer cells engineered with a prosthetic network capable of sensing uric acids levels and secreting urate oxidase to clear the excess (Kemmer et al. 2010; Ausländer & Fussenegger 2016).

1.3 – Programmable transcriptional regulators

Transcriptional regulation lies at the center of most cellular processes, from gene replication and repair, to cell division and differentiation (Thompson et al. 2015). Additionally, aberrant gene expression has been shown to play a role in the onset and progression of several diseases including cancers (Jusiak et al. 2016; Lelli et al. 2012). Accordingly, in an attempt to reprogram cellular behavior, the field of synthetic biology has mainly focused on developing ways to modulate the transcriptional output of genes (Boettcher & McManus 2015) (Fig. 1.2g).

The first generations of synthetic transcriptional regulators used by synthetic biologists were naturally evolved DNA-binding proteins which conditionally repress cognate promoters in bacteria/bacteriophage. As mentioned earlier, such regulators include the lac (*LacI*) and the *cI* repressors. In nature, *LacI* inhibits the expression of genes that code for proteins involved in the bacterial metabolism of lactose, by binding a specific operator sequence flanking the target promoters (Cronin et al. 2001). While its repressive activity is conditioned in the absence of allolactose, a byproduct of lactose, isopropyl β-D-1-thiogalactopyranoside (IPTG) has been shown to successfully prevent binding of the repressor on target (Daber et al. 2007). Similarly, the transcriptional repressor *cI* encoded in the *Escherichia phage lambda* genome, conditionally (temperature-sensitive) binds to and silences the phage pR and pL promoters consequently allowing the virus to establish and maintain latency (Burz et al. 1994). From this discovery, synthetic inducible promoters were engineered by modifying existing promoters to insert either the *LacI* or *cI* operators next to their TATA box (Wedler & Wambutt 1995; Hannig & Makrides 1998). As previously mentioned, such inducible promoter:repressor pairs were notably used to implement the first toggle switch in *E.Coli* (Gardner et al. 2000) (Fig. 1.2a).

Based on the same strategy, Bujard and Gossen later developed the TetR system (Gossen & Bujard 1992), which employs the Tet repressor (tetR) DNA binding protein found in gram-negative bacteria (Ramos et al. 2005), to engineer inducible promoters whose activity can be modulated using the antibiotic, tetracycline, or its derivatives. Both tet-OFF and tet-ON systems were engineered to silence or respectively activate gene expression depending on the presence of the drug. To do so, the team constructed synthetic promoter by coupling a minimal promoter, which does not mediate transcription on its own, with multiple repeats of the tetR binding motif (tet operator, tetO) to create a tetracycline response element (TRE). Additionally, TetR was fused with the herpes simplex virus protein VP16 (O'Reilly et al. 1997; Wysocka & Herr 2003) to engineer a tetracycline transactivator (tTA). In the tet-OFF system, tTA is recruited to the promoter and the VP16 effector domain mediates transcription of the downstream gene. Addition of the antibiotic prevents tTA binding and silences transgene expression. Conversely, in the tet-ON system, tTA has been modified to only bind to tetO in the presence of tetracycline hence producing the opposite behavior. Finally, the transcription factor GAL4, which specifically binds to a DNA sequence known as upstream activation sequence (UAS), was adapted from *Saccharomyces cerevisiae* to create synthetic promoters in eukaryotic cells (Kakidani & Ptashne 1988; Webster et al. 1988; Brand & Perrimon 1993).

While such solutions are still being used in specific cases, the scope of their applications is limited by the fact that the transcriptional regulators only bind specific target sequences. As such, these methods can only be used to control the expression of exogenously delivered or stably integrated transgenes placed downstream of one of these inducible promoters. This limitation was later mitigated by the advent of programmable transcriptional regulators (PTR), which can be engineered to bind various DNA sequences, including endogenous promoters (Carroll 2014). Before being adapted to design the next generation of transcriptional regulators, the versatility of such programmable DNA-binding proteins was demonstrated by the creation of “designer” nucleases used to perform targeted genome engineering (Gaj et al. 2016).

The first of such programmable nucleases was based on the most abundant class of DNA binding domains found in eukaryotic cells, namely Cys2-His2 zinc fingers (Tupler et al. 2001; Venter et al. 2001). Zinc fingers (ZFs) are small protein structural motifs of about 30 amino acids, which recognize specific DNA nucleotide triplets (Klug & Rhodes 1987). By fusing several ZF together using protein linkers it became possible to engineer polydactyl ZF proteins or ZF arrays to target specific DNA sequences along the genome (Q. Liu et al. 1997). In fact, assuming a random base distribution in the 3.24 billion nucleotides human genome, Liu and colleague have estimated that an array of as little as six fingers, targeting an asymmetrical 18nt DNA sequence, should be able to uniquely address any genomic locus (Q. Liu et al. 1997). Programmable nucleases employing ZF were created by fusing to different ZF-arrays (~3 to 4 ZFs) the *Flavobacterium okeanokoites* FokI endonuclease (Wah et al. 1997). Such fusion proteins were then used as heterodimers to precisely generate DNA double strand breaks along the genome (Y. G. Kim et al. 1996; Urnov et al. 2010) (Fig 1.3a). Subsequently, the first PTRs were created by tethering to a ZF-array transactivating domains such as VP16 or subunit p65 of the transcription factor NF-κB (Beerli & Barbas 2002; L. Zhang et al. 2000). Using either three fingers ZF-arrays (Beerli et al. 1998; P. Q. Liu et al. 2001) or six fingers (Dreier et al. 2001; Q. Liu et al. 1997) ZF-arrays, several group were able to demonstrate transcriptional activation of endogenous genes in mammalian cells. Additionally, transcriptional repression was also achieved by replacing transactivator for repressor domains such as the Krüppel-associated box (KRAB), ERD, or SID (Beerli et al. 2000; Beerli et al. 1998).

Despite the identification of a large ZF repertoire, some DNA triplets (5'-CNN-3', 5'-TNN-3') still could not be targeted (Wolfe et al. 2000). Furthermore, the method also suffered from off-target and context-dependent effects due to the binding properties of the different fingers within the array (Wolfe et al. 2000). Nevertheless, the most prominent obstacle for the democratization of the method was the extremely laborious, time consuming and expensive assembly protocols: the *in vitro* selection step (phage display), often required to evolve potent

and specific polydactyl ZF proteins , made its use prohibitive to non-specialized laboratories (Wolfe et al. 2000).

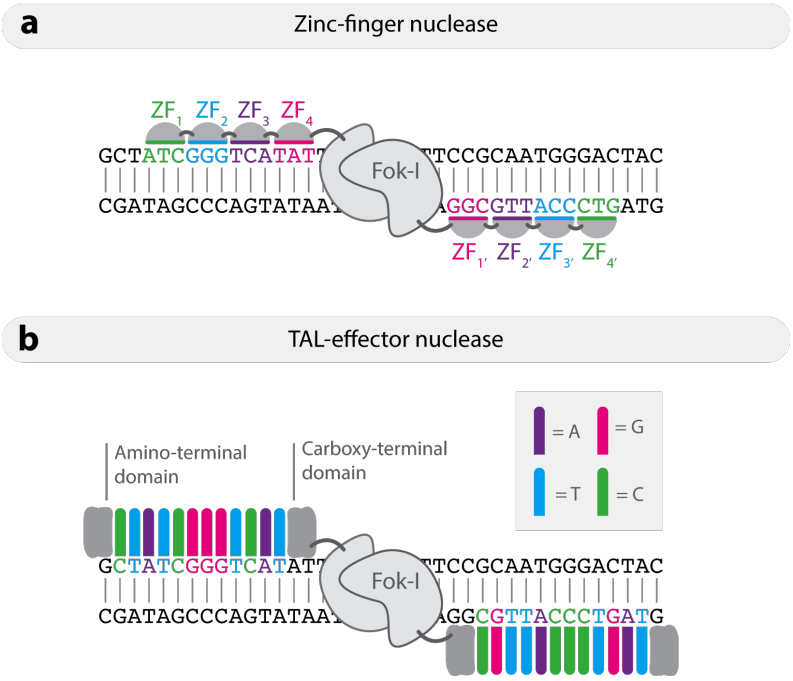

Figure 1.3 – ZF and TALE programmable nucleases.(a) A ZF nuclease dimerizes at a user defined location along the genome. Two ZF arrays (4 fingers) both fused to the endonuclease FokI are targeted to the transcribed and non-transcribe DNA strand, respectively. Each FokI generates a nick on the targeted strand resulting in a DNA double strand break. **(b)** TAL-effector nuclease dimerizes at a user defined location along the genome. Two TAL-effector array (12 units flanked by terminal domains, each recognizes a specific nucleotide) fused to the endonuclease FokI are targeted to the transcribed and non-transcribe strand of the DNA respectively. Each FokI generates a nick on the targeted strand, which result in a DNA double strand break. Gray box shows the TALE "code" indicating correspondence between each TALE element and the nucleotide it contacts.

Transcription activator-like effectors (TALEs) constitute the second generation of "designer" DNA-binding proteins. Secreted by the phytopathogenic *Xanthomonas* bacteria, TALEs mimic eukaryotic transcription factor and bind specific promoters in the host plant genome to drive the expression of genes that aid bacterial infection (Boch & Bonas 2010; Gu et al. 2005). The protein is composed of a transcriptional activator domain, nuclear localization signals (NLSs), and a central domain of tandem repeats which specifically bind DNA. The binding domain

comprises 17.5 repeats that share a common 34-35 amino-acids sequence, except for two residues (positions 12th and 13th) known as repeat variable diresidue (RVD) (Boch et al. 2009). Bioinformatic analyses of TALEs and their targets, as well as *in vitro* experiments have helped identify a one to one correspondence between RVDs and the DNA nucleotide they contact (Moscou & Bogdanove 2009; Boch et al. 2009). Using this code, TALE-nucleases targeting between 7 to 34 base-pair DNA segments have been engineered by coupling custom TALE arrays with FokI (Mahfouz et al. 2011; Miller et al. 2011) (Fig. 1.3b). As done previously with ZF nucleases, genome editing was performed by delivering two TALE-FokI fusion proteins programmed to bind the template and non-template DNA strands, respectively. Regarding transcriptional regulation, TALEs were first successfully reprogrammed to control different genes in plants (Morbitzer et al. 2010), before being later adapted to work in mammalian cells (Miller et al. 2011; F. Zhang et al. 2011). Using TALE arrays of 17.5 and 12.5 repeats respectively, coupled with VP16 or VP64 (4 copies of VP16 connected by Gly-Ser residues), two groups first demonstrated transcriptional activation of endogenous genes in human cells (Miller et al. 2011; F. Zhang et al. 2011). While proven to be more specific than ZF (reduced off-targets) (Mussolino et al. 2014) and easier to assemble (Reyon et al. 2012), the size of TALE arrays and their repetitive nature render them difficult to deliver using viral vectors, such as lentiviruses (Gaj et al. 2016; Holkers et al. 2013).

More recently, a third generation of programmable nucleases and associated transcriptional regulators has been developed which offers a more versatile and easy to implement counterpart to ZFs and TALEs (Carroll 2014; Laufer & S. M. Singh 2015). This new class of PTRs, which have been adapted from the prokaryotic immune system CRISPR (see below), largely differs from second generation binding proteins in that its DNA targeting specificity is dictated by a guide RNA rather than an array of proteins (Mali, Esvelt, et al. 2013). In the next two sections of this introduction, I present a detailed overview of the CRISPR system from its discovery in bacteria to its use in mammalian systems. I then focus on the development of

CRISPR-based transcriptional regulators (CRISPR-TR) before finishing the section on a comparison between this strategy and ZFs/TALE-based approaches (see sections 1.4, 1.5).

1.4 – CRISPR-Cas system, RNA-guided nucleases

1.4.1 – Discovery and classification

In 1987, Ishino et al. were the first to report on a curious set of short, repetitive DNA sequences (~29 nucleotides long) interspaced with variable palindromic sequences in the *E.coli* genome (Ishino et al. 1987). Ten years later, similar structures were identified in multiple bacterial and archeal genomes (Mojica et al. 2000), and in 2002 Jansen et al. coined the term CRISPR, which stands for Clustered Regularly Interspaced Short Palindromic Repeats, to describe these arrays (Jansen et al. 2002). Bioinformatics analysis of CRISPR-arrays led to the realization that the sequences intercalating these palindromic repeats, also known as "spacers", were matching bacteriophage (phage) genomes (Bolotin et al. 2005). The discovery that spacers are from foreign origins, along with the recognition that CRISPR-arrays are flanked by CRISPR-associated (Cas) proteins containing nuclease and helicase domains (Jansen et al. 2002; Haft et al. 2005), prompted Mojica and colleagues to propose in 2005 that the CRISPR-Cas system represents a form of bacterial adaptive immunity against phage infections (Mojica et al. 2005). In 2007, the first experimental evidence of a CRISPR-Cas immune function was reported when researchers showed that the infection of the gram-positive bacterium, *Streptococcus thermophilus,* with various bacteriophages led to spacer acquisition from the invader's DNA (Barrangou et al. 2007). Additionally, removal of these spacers from the bacteria's CRISPR-array led to a loss of immunity (see Fig. 1.4 for a timeline of milestones in the CRISPR field).

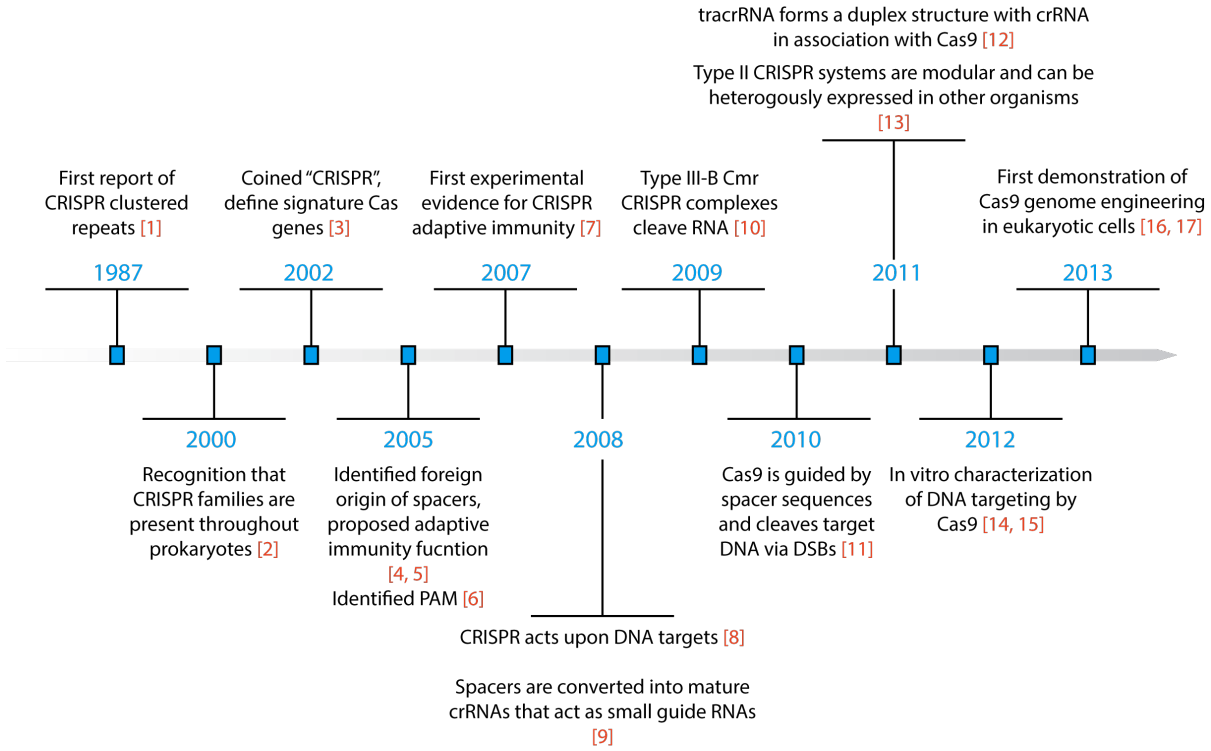

Figure 1.4 – Research milestones in the CRISPR field. *Figure adapted from (Hsu et al. 2014)*. Orange numbering shows references associated with each discovery. Correspondence between numbers and references is as follow: [1] (Ishino et al. 1987); [2] (Mojica et al. 2000); [3] (Jansen et al. 2002); [4] (Mojica et al. 2005); [5] (Pourcel et al. 2005); [6] (Bolotin et al. 2005); [7] (Barrangou et al. 2007); [8] (Marraffini & Sontheimer 2008); [9] (Brouns et al. 2008); [10] (Hale et al. 2009); [11] (Garneau et al. 2010); [12] (Deltcheva et al. 2011); [13] (Sapranauskas et al. 2011); [14] (Jinek et al. 2012); [15] (Gasiunas et al. 2012); [16] (Cong et al. 2013); [17] (Mali, Yang, et al. 2013). CRISPR: clustered regularly interspaced short palindromic repeats; Cas: CRISPR-associated genes; PAM: protospacer adjacent motif; crRNA: CRISPR RNA; DNA DBS: double strand break; tracrRNA: trans-activating crRNA.

Subsequently, multiple CRISPR-Cas systems were characterized across the bacterial and archaeal kingdoms (Mohanraju et al. 2016; Koonin et al. 2017). In all cases, the CRISPR locus is invariably composed of a CRISPR array of spacers, encoding a memory of past infections, and a set of Cas proteins (Mohanraju et al. 2016) (Fig. 1.5a). We now know that CRISPR-Cas systems have three main functions (Fig. 1.5b): (i) Immunization or adaptation: the ability of a system to recognize foreign DNA, generate new spacers, and store them in the CRISPR-array. The Cas1 and Cas2 proteins have been invariably associated with CRISPR immunization and

particularly spacer acquisition (Datsenko et al. 2012); (ii) Expression or biogenesis of a CRISPR RNA (crRNA): upon new phage infection the CRISPR-array is transcribed and processed to yield one crRNA per spacer. Proteins involved in the process vary between CRISPR types (see below); (iii) Interference: each mature crRNA is loaded in either one or a complex of CRISPR-effector protein(s), some of which possess endoribonuclease activity. Using the crRNA as a guide, the CRISPR-ribonucleoprotein complex (crRNP) finds, binds, and cleaves (single or double stranded breaks) invading DNA or RNA whose sequences are complementary to the crRNA (Fig. 1.5b).

As of 2017, the CRISPR family, whose evolution is the result of an arms race against viruses, is divided into two main classes depending on whether the CRISPR-system uses multiple effector Cas proteins (Class 1, found in both archaea and bacteria) or a single protein (Class 2, found mainly in bacteria) for interference (Koonin & Wolf 2015). Classes were further divided into 6 types I, III, IV in Class 1, and II, V, VI in Class 2 (Fig. 1.6). This classification was motivated by differences mainly in the type of target, whether DNA or RNA, as well as the number and function of respective Cas proteins (Makarova et al. 2015).

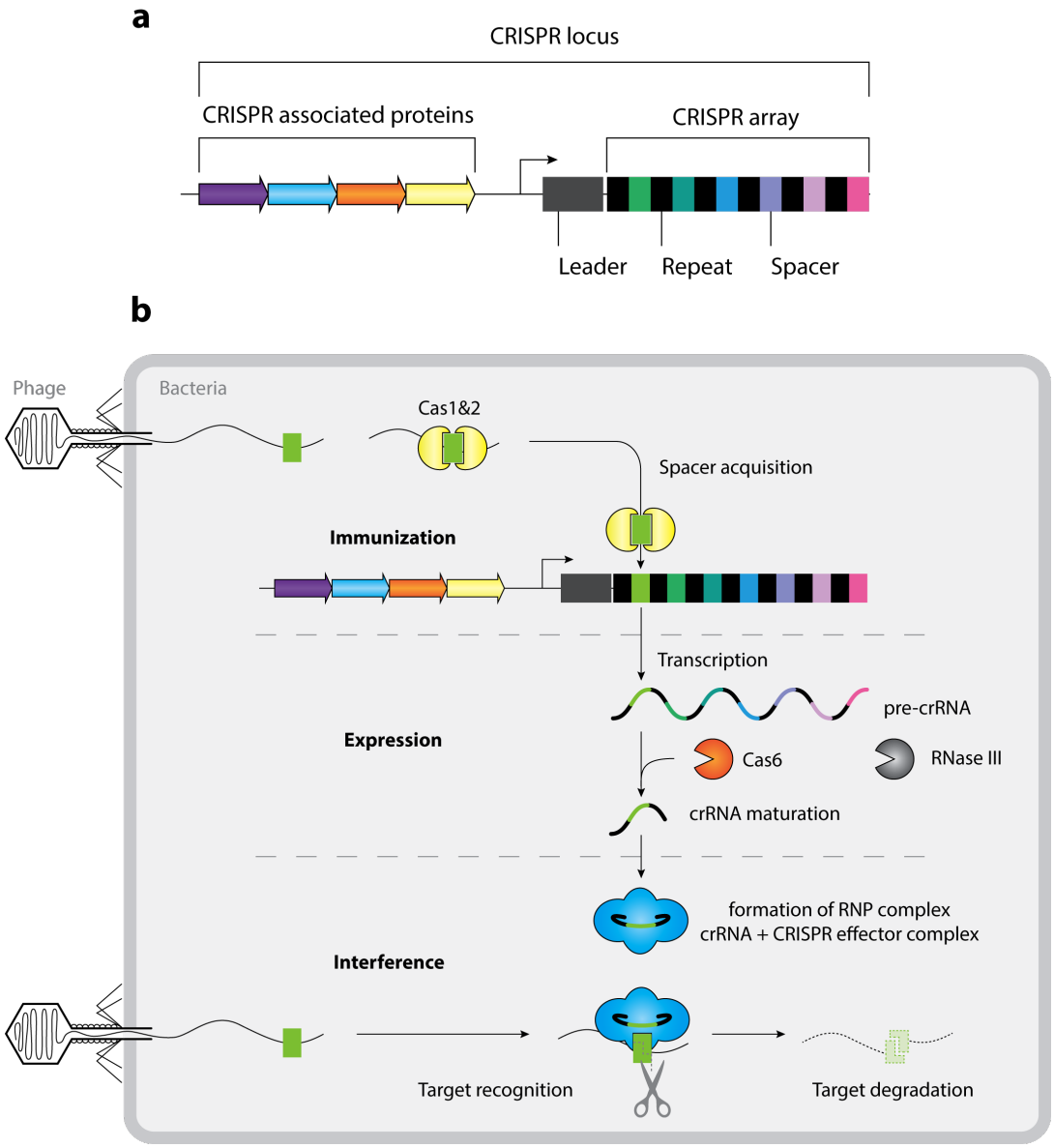

Figure 1.5 – Architecture and function of the CRISPR-Cas locus. *Figure adapted from (van der Oost et al. 2014)*. **(a)** General organization of a typical CRISPR-Cas locus. The number and identity of Cas proteins varies depending on the type of CRISPR system. **(b)** Mechanisms of action of CRISPR-based immunity. (i) Invading phage DNA is fragmented and processed to yield new spacers (green) integrated on the leading end (downstream the leader) of the CRISPR-array. This phase, known as immunization, involves the CRISPR-associated Cas1 and Cas2 proteins (yellow); (ii) During the expression phase, the CRISPR array is transcribed into a precursor CRISPR RNA (pre-crRNA) from which several mature CRISPR RNAs (crRNAs) are produced. Protein from the Cas6 family (orange) and housekeeping RNase III (grey) are involved in this process. Each crRNA is made of a spacer segment (green) and parts of the palindromic repeat (black). (iii) Finally the mature crRNA and a set of CRISPR effector proteins (blue) form the CRISPR-ribonucleoprotein complex (crRNP), which scans new invading DNA and cleaves sequences complementary to the crRNA spacer segment.

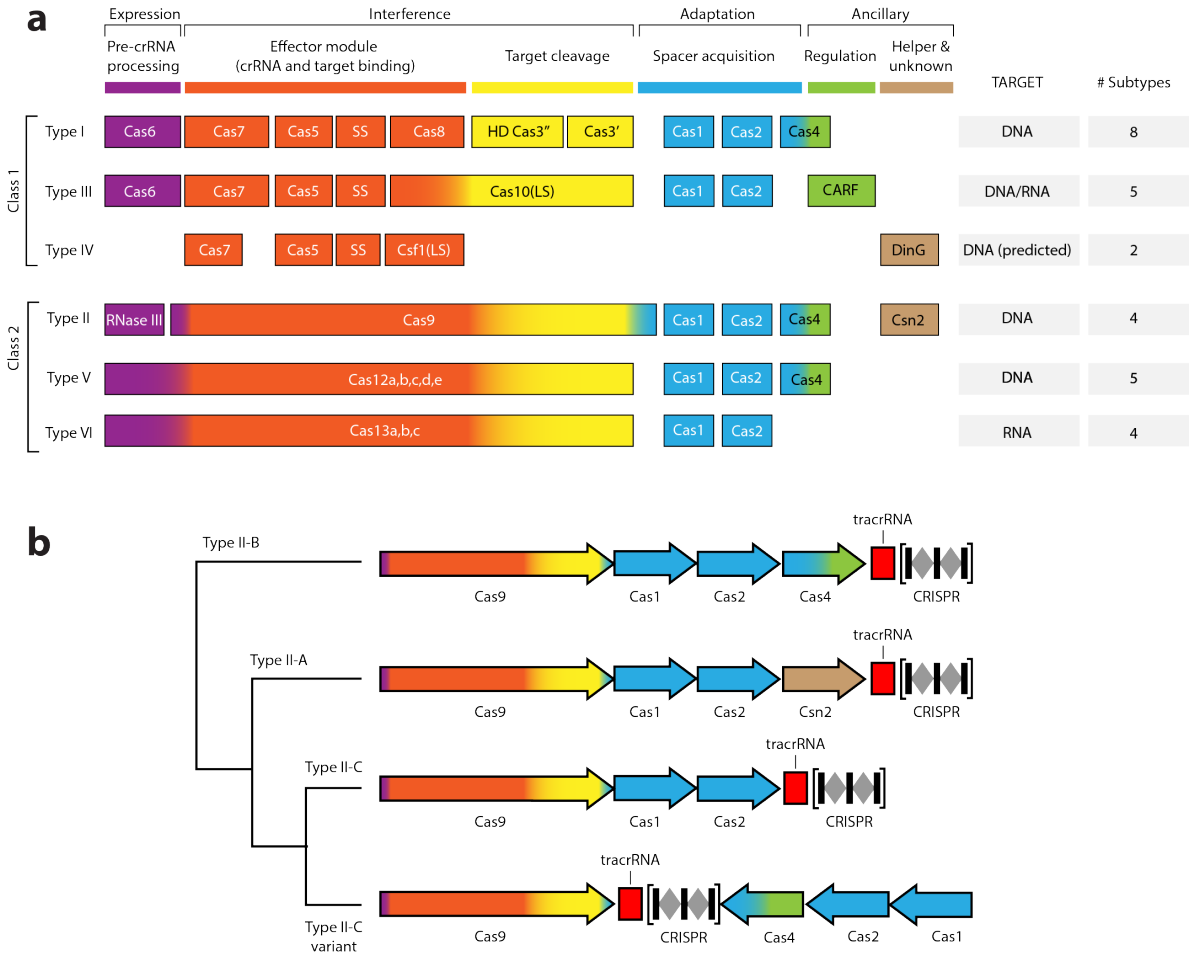

Figure 1.6 – Classification of CRISPR systems. *Figure adapted from (Makarova et al. 2015; Koonin et al. 2017)*. (a) CRISPR systems are classified into two classes and six types. Diagram shows the various CRISPR-associated (Cas) proteins found in each type, along with their specific role in CRISPR-based immunity. Note that each type is further divided in subtypes (see right panel) whose locus feature various subsets of the indicated Cas proteins. (b) Phylogenetic tree of type II CRISPR-Cas9 systems along with their characteristic locus architecture. SS, small subunit; LS, large subunit, CARF, CRISPR-associated Rossman fold.

1.4.2 – Type-II CRISPR-Cas9 system

The type II-A CRISPR-Cas9 system from *Streptococcus pyogenes* (*Sp*) was found to only rely on a single CRISPR-effector protein, Cas9 (also called *Sp*Cas9 when clarification is needed), for interference (Jinek et al. 2012; Gasiunas et al. 2012). Owing to its simplicity, the system was rapidly characterized by the scientific community and adapted as a programmable RNA-guided nuclease (Nishimasu et al. 2014). The *Sp* CRISPR-Cas9 locus is composed of a relatively simple Cas operon encoding proteins Cas9, Cas1, Cas2, and Csn2, as well as a CRISPR-array, and a short non-coding RNA, termed trans-activating crRNA (tracrRNA) (Deltcheva et al. 2011; Doudna & Charpentier 2014) (Fig. 1.7a). Interestingly and in contrast with other CRISPR systems, the Cas9 effector protein was found to take part in all three aspects of CRISPR immunity: adaptation/immunization (Cas9 is required for spacer selection), maturation (helps generate mature crRNAs), and interference in recognizing and cleaving DNA target (Ratner et al. 2016; Heler et al. 2015). Similar to all CRISPR systems, the *Sp* CRISPR-array is transcribed into a precursor crRNA (pre-crRNA) containing several direct repeats interspaced with 20nt (nucleotides) spacers (Fig. 1.7b). The pre-crRNA is then sliced into mature crRNAs in a process that requires both the Cas9 and the tracrRNA. Cas9 loads the tracrRNA, whose 5'-end hybridizes with one of the repeats in the pre-crRNA (Jinek et al. 2012) (Fig. 1.7b). Once the tracrRNA-Cas9 is bound, the pre-crRNA is cleaved by the host's RNase-III to yield functional guide RNAs (gRNA, tracrRNA:crRNA hybrid) (Fig. 1.7b). Further trimming of the gRNA is then performed by unknown host nucleases (Jiang & Doudna 2017; Deltcheva et al. 2011; Gasiunas et al. 2012). Once assembled, the crRNP scans DNA molecules for targets via 3D diffusion (Sternberg et al. 2014) (Fig. 1.7b). The complex will then recognize, bind, and cleave any 5'-N_{20}-NGG-3' sequence along the double stranded DNA where the first 20 nucleotides, also known as protospacer, match the 20nt gRNA spacer (Fig. 1.7c). Additionally, target recognition requires the protospacer to be flanked on its 3'-end by a particular sequence referred to as a protospacer adjacent motif (PAM). While the PAM requirement is shared by all CRISPR-systems as a mean to discriminate between self and non-self (Jinek et al. 2012; Leenay & Beisel 2016), the sequence and length of each PAM

varies between CRISPR-effectors. *Sp*Cas9 recognizes targets flanked by NGG, and to a lesser extend NAG PAMs (N = any base) (Shah et al. 2013).

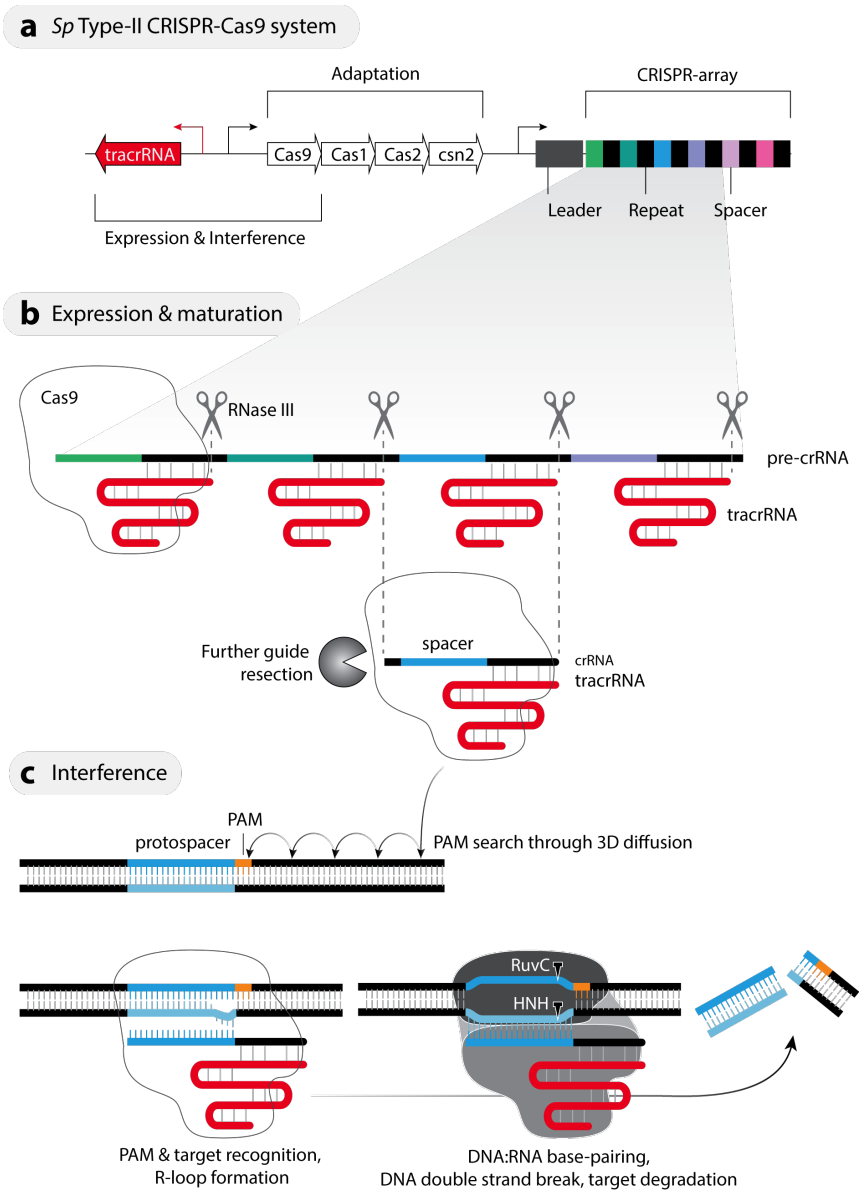

Figure 1.7 – *Sp* type-II CRISPR-Cas9, locus and mechanisms of immunity. *Figure adapted from (Doudna & Charpentier 2014)*. **(a)** Architecture of the type-II CRISPR-Cas9 locus from *Streptococcus pyogenes*. **(b)** Expression and maturation phase of CRISPR-Cas9 immunity: a pre-crRNA transcript is transcribed from the CRISPR-array. The tracrRNA-Cas9 complex binds a repeat on the pre-crRNA and engages the host RNase-III to generate mature guide RNAs (gRNA, tracrRNA:crRNA). Further resection of the loaded sgRNA is performed by host proteins (unidentified). **(c)** The crRNP scans the DNA for NGG or NAG protospacer adjacent motifs (PAM). Sequence complementarity between the sgRNA spacer and DNA nucleotides upstream of the PAM leads to the formation of a three stranded R-loop structure. Sufficient base-pairing between gRNA and DNA target activates Cas9's RuvC and HNH nuclease domains causing a blunt end DNA double strand break 3 nucleotides upstream the PAM.

X-ray crystallography and electron microscopy have revealed that the 1,368 amino acids long *Sp*Cas9 is divided into two distinct domains; a recognition domain necessary for binding the gRNA, and a nuclease domain responsible for introducing double stranded breaks into the DNA target (Fig. 1.8a) (Nishimasu et al. 2014). The latter features a PAM interaction (PI) domain and two nuclease domains, named HNH and RuvC, which cut the targeted and non-targeted DNA strands, respectively. Upon gRNA loading, Cas9 is switched from an inactive state, in which the nuclease domains are facing away from each other and PI is disordered, to a recognition competent state (Jinek et al. 2014; Jiang et al. 2015) (Fig. 1.8b). This conformational change is required for Cas9 to be able to probe DNA molecules for targets (Jinek et al. 2014).

The crRNP then scans DNA for PAMs using Cas9's PI domain (Sternberg et al. 2014). When a PAM is encountered, the interaction between this domain and the NGG/NAG nucleotides on the non-targeted strand causes a partial melting of the DNA double helix directly upstream of the PAM (Anders et al. 2014) (Fig. 1.8c). Subsequently, the spacer segment of the gRNA progressively invades the DNA target, leading to the formation of a three-stranded structure known as an R-loop (Szczelkun et al. 2014) (Fig. 1.8d). Studies employing degenerate protospacers or mutated gRNAs have shown that a high degree of complementarity between the target and the first 10-12 nucleotides of the spacer, termed the seed sequence[2], is required for the crRNP to bind on-target (Wiedenheft et al. 2011; Semenova et al. 2011; D. Singh et al. 2016). Further complementarity (>15 base-pairs) is necessary to trigger a second conformational change in Cas9, which aligns the nuclease domains and activates the catalytic activity of the protein (Sternberg et al. 2015). The enzyme then creates a blunt double stranded break in the DNA target 3 nucleotides upstream from the PAM (Jinek et al. 2012) (Fig. 1.8e).

[2] Numbering starts from the 3'-end of the spacer and increases towards the 5'-end of the guide.

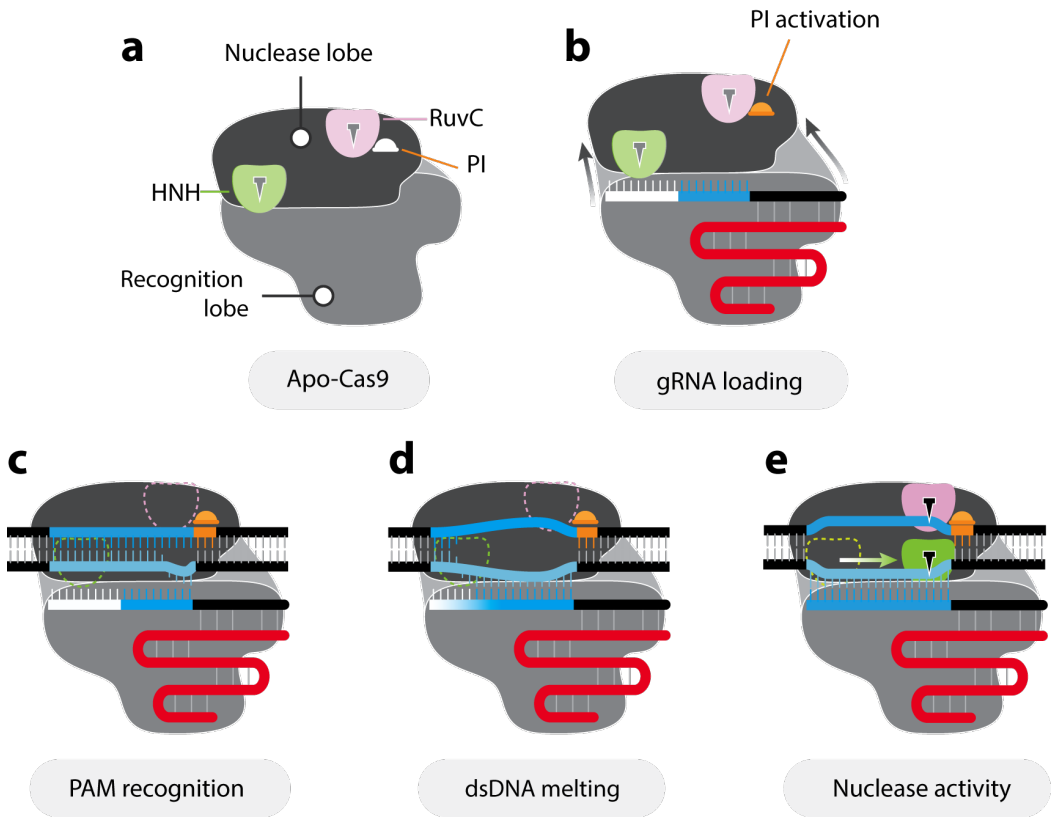

Figure 1.8 – Conformational changes required for Cas9 nuclease activity. *Figure adapted from (Doudna & Charpentier 2014)*. Several conformational changes separate the inactive apo-Cas9 from the nuclease competent crRNP: **(a)** In the absence of guide RNA (gRNA), Cas9 exists in an inactive state which prevents it from scanning, binding, and cleaving DNA. Nuclease domains RuvC and HNH as well as the PAM interacting domain (PI) are inactive. **(b)** Loading of the gDNA (tracrRNA is red, crRNA repeat is black, 20nt spacer is white (unstructured nucleotides) and blue (structured nucleotides)) repositions the nuclease and recognition lobes of the protein creating a positively charged channel which can accommodate the DNA:gDNA heteroduplex. PI gets activated allowing the complex to scan the DNA for targets. **(c)** Interactions between PI and a PAM on the DNA locally melts the DNA double helix and presents the targeted strand to the gDNA seed sequence. **(d)** The spacer segment on the gRNA then invades the DNA via base-pairing. Progressive hybridization stabilizes the positioning of the spacer 5'-end nucleotides (blue to white gradient). **(e)** Sufficient base paring (>15bp) reposition the HNH domains for DNA double strand break.

1.4.3 – From prokaryotic to eukaryotic cells

After studies successfully demonstrated reprogramming of the crRNP to precisely and specifically cut various DNA templates *in vitro* (Jinek et al. 2012; Gasiunas et al. 2012), four groups reported in 2013 CRISPR-based genome editing in human cells (Cho et al. 2013; Cong et al. 2013; Jinek et al. 2013; Mali, Yang, et al. 2013). Adaptation of the naturally evolved *Sp*CRISPR-Cas9 in eukaryotic cells required codon optimization of the Cas9 protein, as well as the addition of NLSs to its sequence to enable its translocation from the cytoplasm to the nucleus. It was also shown that the system could be simplified by fusing the crRNA and tracrRNA into a chimeric single guide RNA (sgRNA) (Jinek et al. 2012; Cong et al. 2013) (Fig. 1.9). In all four studies, it was demonstrated that the crRNP can be easily reprogrammed to target any 5'-N_{20}-NGG-3' sequence in the genome by simply replacing the spacer segment of the sgRNA. Precise CRISPR-induced DNA double strand breaks could then be leveraged for genome editing purposes to cause insertions or deletions in the DNA target sequence through the non-homologous end joining pathway, or to insert heterologous DNA sequences encoded in co-delivered donor DNA template via homology directed repair (Cho et al. 2013; Cong et al. 2013; Jinek et al. 2013; Mali, Yang, et al. 2013).

Since then, CRIPSR-Cas9 has been used in most plant and animal species with an increasing number of publications every year (Fig. 1.10). Its ease of programming and multiplexing potential led to the broad acceptance of the technology (Lebar & Jerala 2016). Additionally, the generation of large sgRNA libraries enabled highly parallel genome-scale knockout screens, which combined with next generation sequencing methodologies have been used to unbiasedly discover and functionally characterize specific genetic elements associated with various phenotypes (e.g. decipher the origin and susceptibility to drug resistance) (Shalem et al. 2014; T. Wang et al. 2014; Joung et al. 2017). The nuclease activity of Cas9 has also been recently exploited to create "memory loci" for lineage tracing that I will further discussed in chapter 8 (Guernet et al. 2016; McKenna et al. 2016; Perli et al. 2016; Kalhor et al. 2016).

Finally, gene editing with CRISPR-Cas9 has been instrumental in the development of new medical applications and drug discovery pipelines (Fellmann et al. 2017).

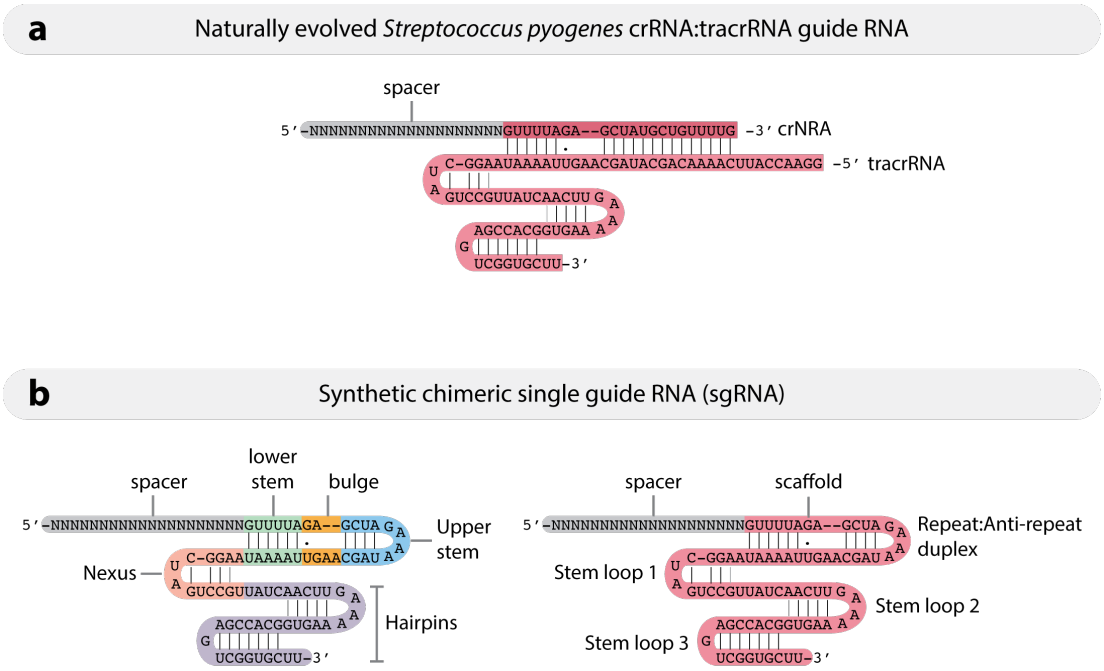

Figure 1.9 – Re-engineering the *Streptococcus pyogenes* guide RNA. (a,b) RNA sequence and predicted secondary structure of the naturally evolved *Streptococcus pyogenes* guide RNA (crRNA:tracrRNA) (a), and engineered chimeric single guide RNA (sgRNA) (b). Two different sgRNA segments nomenclatures as introduced in (Briner et al. 2014) (left) and (Nishimasu et al. 2014) (right). "N" denotes any nucleotide; "•" represents a G-U wobble pair.

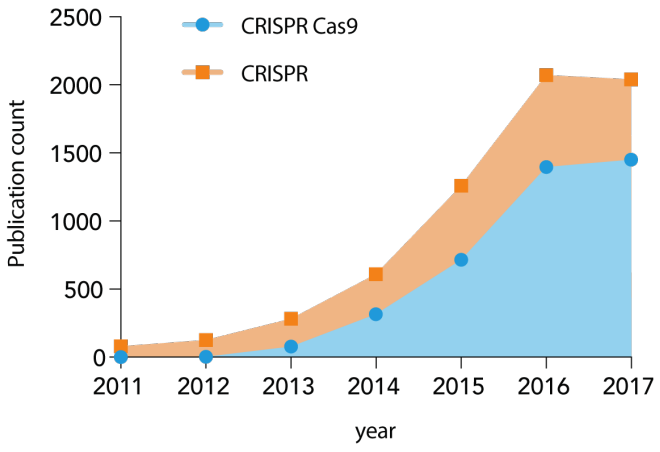

Figure 1.10 – CRISPR publication count from 2011 to 2017.
Number of Pubmed references matching the search term "CRIPSR" (orange) or "CRISPR Cas9" (blue) per year. Figures for 2017 only take into account paper published prior August 30th.

Thereafter, a catalytically inactive version of Cas9, also referred to as dead Cas9 (dCas9), was engineered by mutations of its nuclease domains, HNH (D10A mutation) and RuvC (H841A mutation) (Gilbert et al. 2013; Mali, Aach, et al. 2013) (Fig. 1.11a). Since these point mutations preserve the DNA-binding properties of Cas9, dCas9 was used as a means to direct various effectors tethered to the crRNP to specific DNA loci. As such, the CRISPR-Cas9 system was rapidly adapted to a new range of applications including chromosome labeling (fluorophore tether) (B. Chen et al. 2013; Hanhui Ma et al. 2016), epigenetic modifications (chromatin modifiers tether) (Hilton et al. 2015; Klann et al. 2017), base editing (deaminase tether) (Komor, Badran, et al. 2016; Nishida et al. 2016), and of particular interest for this Thesis, a new generation of programmable transcriptional regulators (Chavez et al. 2016; Dominguez et al. 2016).

1.5 – CRISPR-based transcriptional regulators

1.5.1 – First generation repressors and activators

In 2013, multiple studies were published showing transcriptional repression of gene of interests using dCas9 by programming the sgRNA to anchor the CRISPR-effector in the vicinity of transcription start sites (TSS) (Qi et al. 2013; Gilbert et al. 2013; Bikard et al. 2013) (Fig. 1.11b). It was proposed that dCas9 silences mRNA production by sterically blocking RNA polymerase and/or transcription factor binding, and the methodology was accordingly termed CRISPRi for CRISPR interference. The modest repression mediated by dCas9 alone was later improved by tethering transcriptional silencing effectors such as KRAB (Gilbert et al. 2013; Lawhorn et al. 2014) or SID4x (4 copies of the mSin3 interaction domain) (Konermann et al. 2013) to either the amino (N-) or carboxyl (C-) terminus of the protein (Fig. 1.11b).

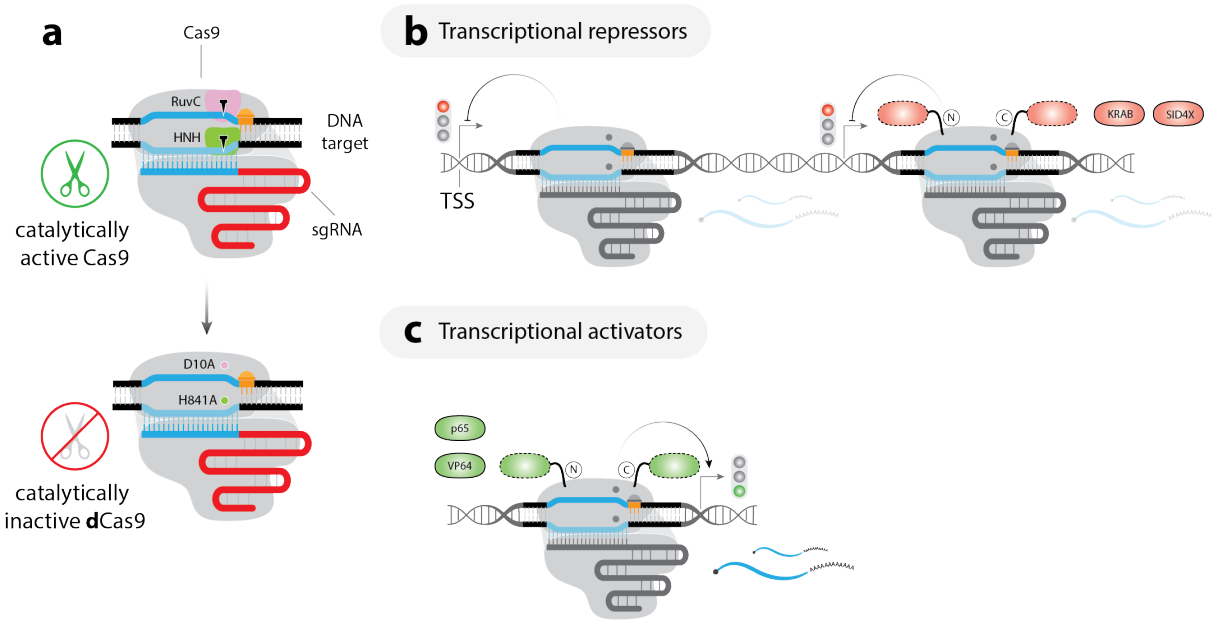

Figure 1.11 – First generation CRISPR-based transcriptional regulators. (a) A nuclease deficient Cas9 (dCas9, bottom) is obtained by mutating the RuvC (D10A mutation) and HNH (H841A mutation) nuclease domains of the protein. While abrogating nuclease activity, these mutations do not affect the DNA binding properties of the CRISPR-effector. Top: sgRNA-Cas9 complex cleaving the DNA target; Bottom: sgRNA-dCas9 complex bound on target. **(b)** CRISPR-based transcriptional repressors using dCas9 only (CRISPRi, left) or dCas9 fused to a repressor domain (right). The complex, sent to the vicinity of the target gene transcription start site (TSS) by reprogramming the sgRNA, blocks transcription elongation. **(c)** CRISPR-based transcriptional activators are created by tethering transactivator domains to dCas9. Docking of the complex upstream of the TSS recruits pol-II and increases the target gene transcriptional output.

Similarly, CRISPR-based transcriptional activation (CRISPRa) of target genes was achieved by tethering to dCas9 various transactivation domains known to stimulate the initiation and elongation of pol-II transcription (Gilbert et al. 2013; Mali, Aach, et al. 2013; A. W. Cheng et al. 2013; Maeder et al. 2013; Perez-Pinera et al. 2013) (Fig. 1.11c). Notably, Gilbert et al. successfully reported activation of a reporter gene by fusing dCas9 to the viral VP64 or human p65 effector domains, and programing this complex with an sgRNA targeting repetitive UAS sequences upstream of a minimal promoter (Gilbert et al. 2013). Using the same method several groups have subsequently demonstrated upregulation of endogenous target genes in human cells (Mali, Aach, et al. 2013; A. W. Cheng et al. 2013; Maeder et al. 2013; Perez-Pinera et al. 2013). Of importance for the rest of this Thesis, these studies showed that tiling

of multiple sgRNAs around the TSS, to increase the number of dCas9-VP64 at the target locus, was required to mediate strong gene activation (Maeder et al. 2013; Perez-Pinera et al. 2013). This insight prompted the development of a second generation of CRISPR-based transcriptional regulators (CRISPR-TR) which aimed to maximize the number of effector domains attached to the crRNP, allowing activation of target genes with a single sgRNA (Chavez et al. 2016; Jusiak et al. 2016).

1.5.2 – Second generation CRISPR-TRs

Strategies like VP64-dCas9-VP64 featuring VP64 on both the N- and C- termini (X. Gao et al. 2014) and dCas9-VP160 created by tethering 10 copies of VP16 (A. W. Cheng et al. 2013) to the CRISPR-effector, spearheaded the second generation of CRISPR-TRs. While more potent than dCas9-VP64, the authors found that these approaches still required sgRNA tiling. Next, Tanenbaum et al. developed a system, called Suntag, designed to recruit multiple "tagging" protein copies to a polypeptide scaffold (Tanenbaum et al. 2014). The method works by fusing a single chain variable fragment (scFv) antibody to a protein, enabling it to bind an array of peptide epitopes. Initially designed to tag proteins with multiple copies of a fluorescent protein, the Suntag system was adapted to create CRISPR-TR with enhanced potency. By attaching the peptide epitope array to the C-terminus of dCas9, the authors were able to demonstrate the recruitment of up to 24 copies of VP64 to a single crRNP (Fig. 1.12a). This CRISPR-TR displayed increased potency over previous generations and made it possible to achieve significant transcriptional activation of endogenous genes using a single sgRNA per target (Tanenbaum et al. 2014).

Two subsequent studies demonstrated that CRISPRa can also be drastically improved by co-recruiting different classes of trans-activator domains to the crRNP (Konermann et al. 2014; Chavez et al. 2015). Chavez et al. created an alternative system by fusing dCas9 with a protein tail, combining single copies of three different effectors: VP64, p65, and Rta[3] (Chavez et al.

[3] Epstein-Barr virus R transactivator (Hardwick et al. 1992)

2015) (Fig. 1.12a). The order of effector domains and linker parameters were optimized to generate dCas9-VPR (VP64-p65-Rta).

Differently, Konermann and colleagues at MIT circumvented the problem of sgRNA tiling by using protein-binding RNA structures, known as aptamers, to recruit additional effector domains directly to the sgRNA (Konermann et al. 2014). The crystal structure of the crRNP (Nishimasu et al. 2014) and mutational analysis of the sgRNA (Briner et al. 2014) have shown that both the repeat:anti-repeat and stem-loop 2 segments of the sgRNA (Fig. 1.9b) protrude out of dCas9 and, as such, can be mutated without affecting DNA binding. Based on these insights, Konermann and colleagues adapted these sequences to incorporate two RNA aptamers (MS2-loops), each capable of recruiting two phage MS2 coat proteins (MCP), to the sgRNA (Chao et al. 2008). Two effector domains, p65 and the human heat-shock factor 1 (HSF1) were then tethered to MCP. In the end, the assembled synergistic activation mediator, or SAM system, contained one copy of VP64 fused to dCas9, and four copies of the MCP-p65-HSF1 attached to sgRNA-2xMS2 (Fig. 1.12b). Following a similar idea, another group employed several RNA aptamers to construct scaffold-sgRNAs capable of recruiting specific effector domains (Zalatan et al. 2014). In this work, the MS2 (Chao et al. 2008), PP7 (F. Lim & Peabody 2002), and com (Zalatan et al. 2014) aptamers were added separately or in combination to the 3'-end of the sgRNA. By creating fusion proteins between the corresponding RNA-binding proteins and effector domains, the authors assembled several functional CRISPR-TRs each composed of dCas9, a scaffold-sgRNA, and the corresponding effector module (Zalatan et al. 2014) (Fig. 1.12b). Of particular interest, because scaffold-sgRNAs with distinct aptamers can recruit different effectors, Zalatan and colleague were able to respectively activate using VP64, and repress with the KRAB domain, two endogenous gene targets simultaneously.

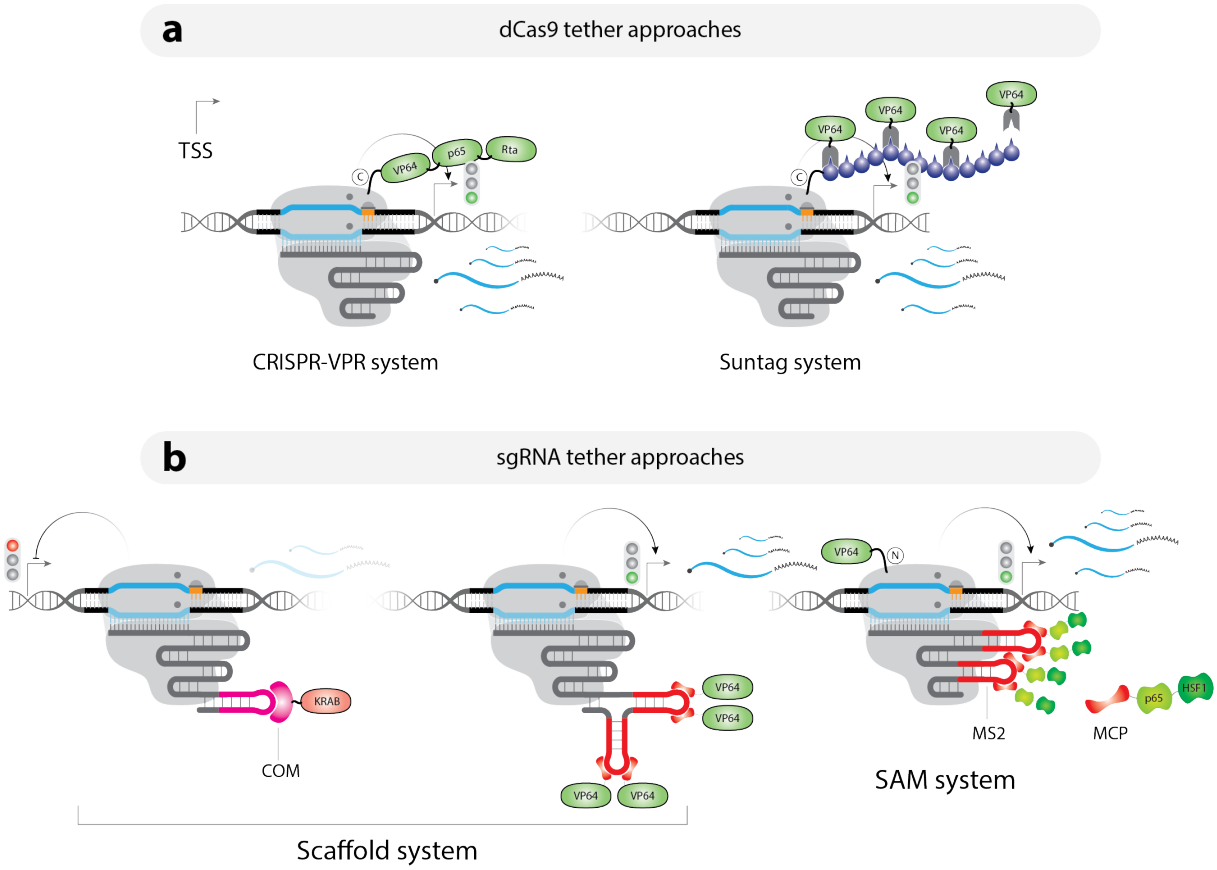

Figure 1.12 – Second generation CRISPR-based transcriptional regulators. (a, b) Potent activation of endogenous target genes using a single sgRNA is achieved by increasing the number of effectors tethered to the crRNP. Additional domains can either be attached to dCas9 (a), or the sgRNA (b).

Chavez et al. recently reported a comparative analysis of all main CRISPR-based transcriptional activators (Chavez et al. 2016). The different systems were benchmarked on their ability to activate, using a single sgRNA per target, a range of endogenous genes targets in human embryonic kidney (HEK)-293T cells. This analysis revealed that the Suntag, VPR, and SAM solutions were the most potent activators. Additionally, in an extended comparison across several human, mouse, and *Drosophila* cell lines, a slight advantage was attributed to the SAM system for the robustness of its effect. Experiments with multiple sgRNAs also showed that, in contrast to dCas9-VP64, all three systems displayed no significant drop in their

performances when simultaneously activating up to six distinct gene targets[4]. With regard to off-target effects, defined by spurious activation of non-targeted genes, total RNA-sequencing of cells transfected with either Suntag, VPR or SAM systems revealed a high degree of specificity in all cases, reflected by minimal perturbation of the transcriptome with the exception of the intended gene target that was found upregulated (Konermann et al. 2014; Gilbert et al. 2014; Chavez et al. 2016).

All second generation CRISPR-TR systems rely on decorating the crRNP with multiple transcriptional activators (e.g. VP64, p65, HSF1, etc.), with the aim to increase the type and number of pol-II transcriptional initiation factors they recruit on target. In order to maximise the recruitment of such factors, the combination of effectors used as well as their relative positioning on the crRNP was optimised in each system aforementioned. Tanenbaum et al. optimised the spacing between epitopes in the array to guarantee that all 24 copies of scFv-VP64 would be recruited (Tanenbaum et al. 2014). Chavez et. al tried varying the order of VP64, p65, and Rta as well as optimised the length of the linker connecting them (Chavez et al. 2015). Finally, Konermann et. al tested various combination of various effectors as well as the effect of the MS2 loops position in the sgRNA sequence on the strength of the activation (Konermann et al. 2014). Given that Suntag, VPR, and SAM rely on a different set of effectors or architecture, one might argue that greater activation could be achieved by creating a hybrid between these strategies. Nevertheless, Chavez and co-authors who conducted these experiments found no additional gain in fold activation when combining these systems in various ways (Chavez et al. 2016).

1.5.3 – Comparison with ZF and TALE approaches

CRISPR-TRs have several substantial advantages over previously reported transcriptional regulators based on ZFs and TALEs (Lo & Qi 2017). (i) With regard to the time and cost required for assembly and implementation, CRISPR-based approaches are far superior over

[4] Konermann et al. show simultaneous activation of up to 10 gene targets using the SAM system (Konermann et al. 2014).

previous platforms. Both ZF and TALE-based transcriptional regulators require expensive and cumbersome protein engineering, which are only accessible to specialized laboratories. In contrast, CRISPR-TR can be reprogrammed to target any new 5'-N_{20}-NGG-3' sequence by performing a simple 20 base-pairs (bp) oligonucleotide cloning (see methods). (ii) While all three methods rely on the use of similar effector domains (VP64, p65, KRAB, etc.), CRISPR-based gene silencing can be achieved with dCas9 alone (no effector domain tether). To the best of my knowledge, this was not demonstrated with ZFs and TALEs arrays. (iii) Regarding the choice of targets, I have mentioned earlier that ZF-based PTRs are limited to DNA sequences composed of base-pair triplets for which a corresponding ZF exists. In contrast, TALE-based systems have for unique requirement that the target must start with a thymine (Lebar & Jerala 2016). As such, TALE-arrays are, in theory, more versatile than *Sp*-dCas9 whose targets need to be flanked by a NGG or NAG PAM. Nevertheless, PAM limitations are being alleviated by re-engineering Cas9's PI domain to relax the PAM requirement (Kleinstiver, Prew, Tsai, Nguyen, et al. 2015; Kleinstiver, Prew, Tsai, Topkar, et al. 2015; Anders et al. 2016). Additionally, the discovery of new CRIPSR systems amenable to CRISPR-TR, like Cpf1 (Zetsche, Gootenberg, et al. 2015), and the adaptation of dCas9 orthologues to work in human (Esvelt et al. 2013), are broadening the repertoire of sequences targetable by CRISPR-TR. (iv) All three methods have minimal OFF-target effects compared to other approaches, such as RNAi (Lo & Qi 2017). (v) Studies have shown that while ZF and TALEs-based approaches are blocked by DNA methylation[5], dCas9 is reportedly only minimally affected by the chromatin state (Hsu et al. 2013; Perez-Pinera et al. 2013). (vi) Finally, owing to the simplicity of its implementation and the possibility to deliver multiple sgRNAs at once, CRISPR-based approaches are readily amenable to multiplex gene targeting, as well as genome scale screens (Konermann et al. 2014; Gilbert et al. 2014).

[5] Note that, nevertheless, TALEs can be programmed to specifically target methylation marks (Deng et al. 2012).

1.6 – Inducible CRISPR-TR strategies

In addition to be easily reprogrammable and highly specific (Chavez et al. 2016; Dahlman et al. 2015), CRISPR-TRs can be multiplexed to simultaneously activate/repress several gene targets, by simply delivering multiple sgRNAs. Additionally, studies conducting genome-scale CRISPRa screens have generated sgRNA libraries to control the expression of every isoform in the human genome (Gilbert et al. 2014; Konermann et al. 2014). As such, CRISPR-TRs offer great potential for the assembly of synthetic gene circuits (Jusiak et al. 2016). Amongst others, CRIPSR-TRs have already been used to create gene circuits for the detection of cancerous bladder cells (Y. Liu et al. 2014), implement various Boolean logic gates (Nielsen & Voigt 2014; Gander et al. 2017), redirected metabolic pathways in bacteria (Zalatan et al. 2014; Cress et al. 2016), and assemble layered circuits where the expression of one sgRNA specifically induces expression of an output sgRNA (Kiani et al. 2014). Nevertheless, as of 2014, this biological part remained "stand alone" as no inducible solutions were available to control its activity. Consequently, while CRISPR-TRs could be used to activate or repress the expression of a gene of interest, it was not possible to create a circuit whereby the transcriptional output of the gene would be conditioned on a particular biological event.

By 2014, there was a clear demand for inducible CRISPR methodologies that would allow spatial-temporal control of Cas9 or dCas9-effector functions (Nuñez et al. 2016; Zhou & Deiters 2016; Hilton & Gersbach 2016). In the field of genome editing, the need for inducible system was mainly motivated by several studies showing that a prolonged exposure to CRISPR-Cas9 was both toxic to the cell, due to transient crRNP binding at all PAM sites along the genome[6], and increased the rate of OFF-target editing events (Hsu et al. 2013; Pattanayak et al. 2013; Fu et al. 2014; Zuris et al. 2015). With regard to CRISPR-based transcriptional regulation and the creation of synthetic circuits, the development of inducible strategies was driven by the fact that most biological processes are both dynamic and conditional. This is particularly true of

[6] Also observed with dCas9.

gene expression as illustrated by the complex orchestration of transcriptional programs at play during embryogenesis (Biase et al. 2014) or oncogenesis (Hu et al. 2013). Accordingly, inducible CRISPR-TR were needed to help dissect, study, and re-assemble this biological processes.

Answering to this demand, a plethora of papers have since then been published, which proposed different inducible CRISPR-Cas9 strategies. The first account of inducibility was reported by Nissim et al. in 2014 (Nissim et al. 2014). This study showed that the Cas protein Csy4 (Haurwitz et al. 2010), naturally involved in the generation of crRNAs from pre-crRNAs, could be used to splice *Sp* sgRNAs out of a larger pol-II transcript in mammalian cells. The sgRNA was programmed to direct dCas9-VP64 to the TSS of a reporter gene and mediate transcription activation. While not intended as an inducible system, the authors found that their sgRNAs drove stronger reporter expression when excised from the parent transcript (presence of Csy4) compared to the condition lacking Csy4 (Nissim et al. 2014). Throughout the rest of this section, I review the different inducible CRISPR-TR strategies that have been proposed in the literature[7]. The methods have been grouped by type or mechanism of induction.

1.6.1 – Transcriptional control with inducible promoters

Perhaps the most obvious approach to creating conditional CRISPR-Cas9 systems is through the use of inducible promoters. Expressing either the Cas9/dCas9 protein or the sgRNA under the control of such promoters, several groups were able to condition CRISPR-Cas9 activity on both small molecules or light inducers (Fig. 1.13a-c). Employing a variant of the tetR system[8], several groups have notably created inducible CRISPR strategies whereby expression of Cas9 (Dow et al. 2015; Cao et al. 2016), dCas9-KRAB (Mandegar et al. 2016), or dCas9-VPR (J. Guo et al. 2017), was conditioned on doxycycline delivery (Fig. 1.13a). Similarly, Li et al. used an arabinose-induced promoter to drive conditional expression of dCas9 in *E. coli* (Li et al.

[7] Prior July 2017
[8] rtTA-TRE system: reverse tetracycline transactivator (rtTA) binds to a DNA segment upstream of a minimal promoter (TRE) in the presence of the antibiotic doxycycline (Gossen & Bujard 1992).

2016). At the same time, inducible promoters have also been employed to conditionally express sgRNAs. For example, the tetR systems adapted on the H1 pol-III promoter (Mi et al. 2006) was used to control sgRNA expression with doxycycline in mouse and human cells (Aubrey et al. 2015; de Solis et al. 2016) (Fig. 1.13b). Finally, a light-based approach was designed whereby the YF1, a blue-light-sensitive histidine kinase, phosphorylates the protein FixJ, who in turn triggers sgRNA expression. Using this approach in *E. coli*, the authors successfully demonstrated light-dependent silencing of CRISPRi activity (Hongyi Wu et al. 2014) (Fig. 1.13c).

1.6.2 – Post-transcriptional control of Cas9 production

Post-transcriptional control of Cas9 mRNA can also be employed to create inducible systems. For example, Hirosawa and colleagues have created two gene circuits in which Cas9 production was turned ON or OFF contingent upon the presence of a specific micro RNA (miRNA) (Hirosawa et al. 2017) (Fig. 1.13d). This was achieved by incorporating a miRNA response element (MRE) in the 5'-untranslated region (UTR) of the Cas9 mRNA.

Cas9-based genome editing activity, reflected by the decrease in GFP reporter expression, was largely reduced in the presence of a cognate miRNA (OFF-switch). In order to create an ON-switch, to specifically increase Cas9 availability when the miRNA is expressed, a Kink-turn RNA motif (Klein et al. 2001) was cloned in the Cas9 5'-UTR (Fig. 1.13d). Binding of the L7Ae protein to the Kink-turn motif (Caban et al. 2007; Huang & Lilley 2013) prevented Cas9 translation. By introducing an MRE in L7Ae 5'-UTR, the authors created a double negative gate that turned on Cas9 production in the presence of the miRNA. Using this method, the group demonstrated cell specific genome editing: transfection of a mixed cell population with the ON-switch circuit designed to conditionally knock out GFP, led to the loss of the transgene only in the miRNA positive cells (Hirosawa et al. 2017).

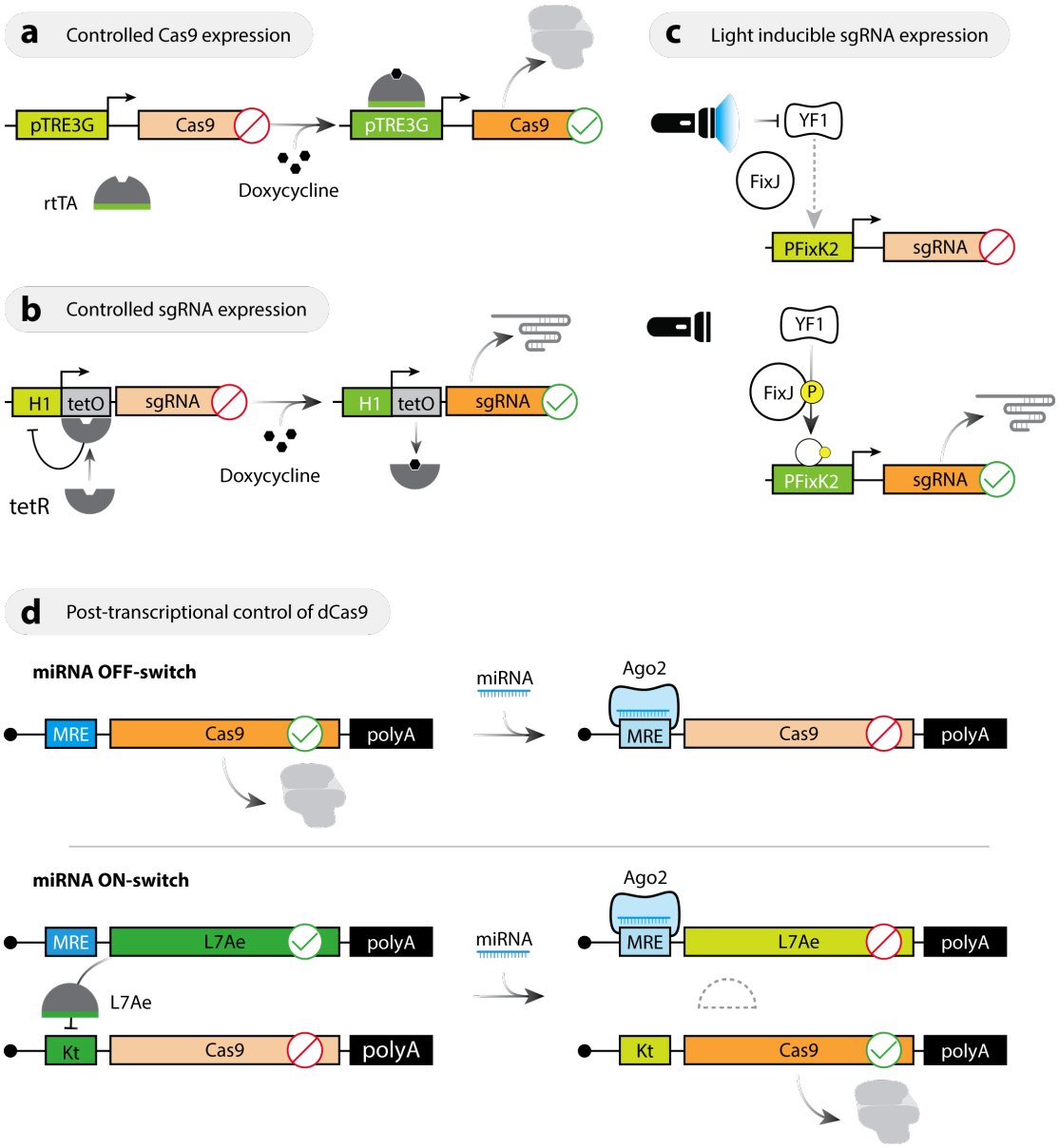

Figure 1.13 – Transcriptional and post-transcriptional control of CRISPR-TR. (a-b) Inducible promoters driving gene expression only in the presence of doxycycline, are used to control Cas9 mRNA (a) or sgRNA (b) production. **(c)** Photo-inducible system with a light-responsive promoter. In the absence of blue light, blue light-sensitive histidine kinase (YF1) phosphorylates FixJ, which in turn activates the PFixK2 promoter driving sgRNA expression (top). When cells are exposed to blue light, YF1 no longer phosphorylates FixJ and sgRNA production is halted (bottom). **(d)** Control of Cas9 mRNA levels with microRNA (miRNA)-responsive synthetic circuits. Cas9 transcript with a miRNA response element in its 5'-UTR get sliced and degraded in the presence of Argonaute Ago2 proteins loaded with the miRNA inducer (miRNA OFF-switch, left). Presence of a kink-turn (Kt) in Cas9 5'-UTR blocks translation of the CRISPR-effector in the presence of the binding partner L7Ae. Expression of the miRNA silences L7Ae and resumes Cas9 production (miRNA ON-switch, right).

1.6.3 – Switchable Cas9/dCas9 proteins

Several groups have engineered Cas9 variants whose nuclease activity can be directly controlled by small molecule or light inducers. Davis et al. incorporated into the structure of Cas9 a drug-responsive intein, a protein splicing fragment, which abrogated nuclease activity of the CRISPR-effector (Davis et al. 2015) (Fig 1.14a). Delivery of the small molecule 4-hydroxytamoxifen (4OHT) induced intein splicing and triggered Cas9 activation. Others created an inducible Cas9 responsive to UV-light by inserting a photocaged lysine residue into its amino acid sequence (Hemphill et al. 2015) (Fig 1.14b). This method, which required an expansion of the host cell codon repertoire, enabled Hemphill and colleagues to conditionally generate indels in the human genome upon exposing cells to UV light.

Yet another switchable Cas9 was engineered by fusing the protein with the ligand-binding domain of the human estrogen receptor-alpha (ER-LBD) (Oakes et al. 2016). Binding of ER-LBD to its agonist 4OHT, induced conformational changes that were transduced to Cas9, resuming its nuclease activity (Fig 1.14c).

Finally, Richter et al. created both inducible Cas9 and dCas9 variants by fusing the CRISPR-effector to the RsLOV photoreceptor (Richter et al. 2016) (Fig 1.14d). This protein found in the bacterium *Rhodobacter sphaeroides* adopts a homodimeric form in the dark and dissociates upon blue light exposure (Metz et al. 2012). After demonstrating that Cas9 and dCas9 dimers were not binding their DNA targets in the absence of UV light, the authors were able to achieve light-induced genome editing and CRISPRi (Richter et al. 2016).

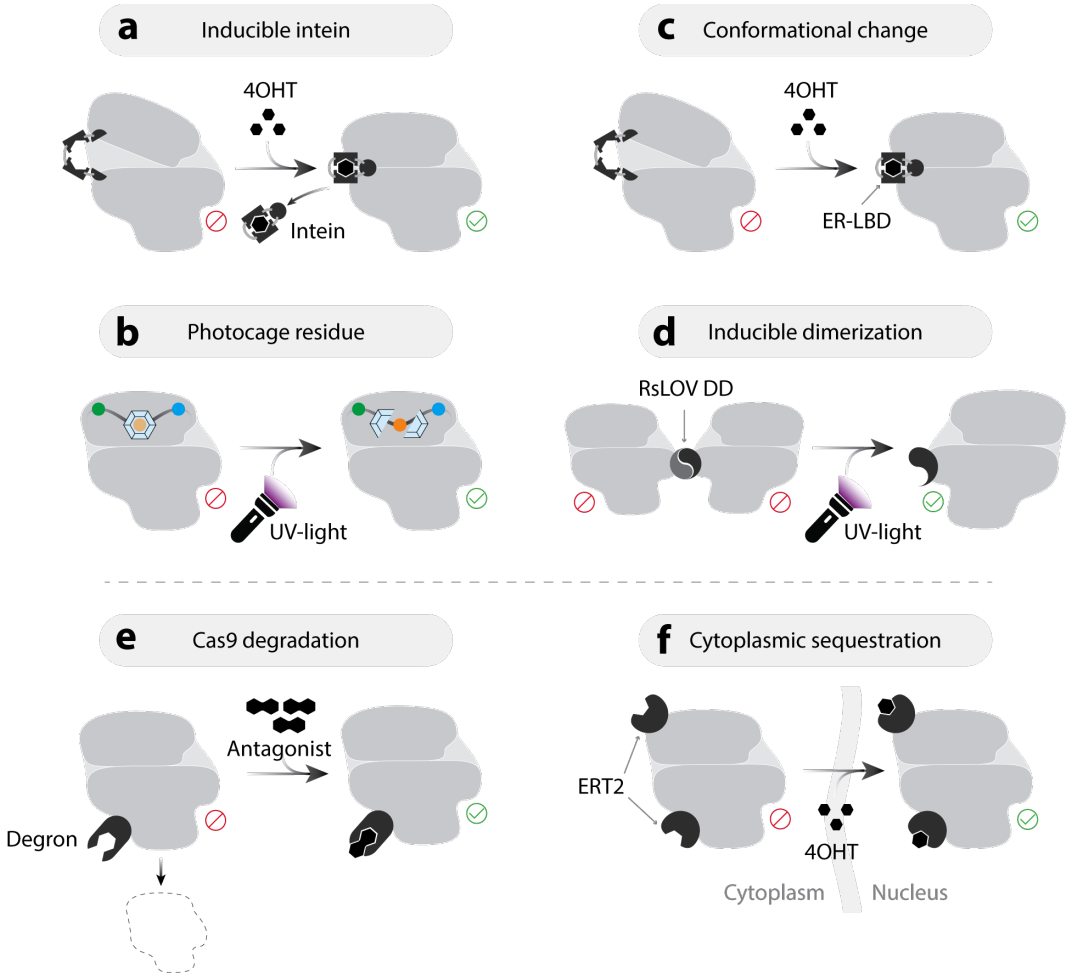

Figure 1.14 – Switchable Cas9 and control over protein availability. (a) Cas9 is maintained in an OFF-state by the presence of an inducible intein (self-splicing peptide) in its structure, which splices itself out in the presence of 4-hydroxytamoxifen (4OHT). **(b)** The codon anticodon code of HEK293-T cells was modified to incorporate a photocaged residue in the peptide chain of Cas9. Exposure to UV-light releases the amino-acid and activates an otherwise silent CRISPR-effector. **(c)** Cas9 was fused to the ligand-binding domain of the human estrogen receptor-alpha (ER-LBD), whose conformation change in the presence of 4OHT activates Cas9. **(d)** Using RsLOV dimerization domains (DD) which complex in the dark, Cas9-Cas9 dimers are formed that prevent binding and nuclease activity. The crRNP is then activated with UV-light. **(e)** Cas9 is fused to "degrons" (degradation domains) which mediate its proteolysis. Delivery of an antagonist resumes Cas9 availability. **(f)** Cas9 is sequestered to the cytoplasm by fusion with the hormone binding domain of the estrogen receptor (ERT2). Delivery of the antagonist (4OHT) resumes CRISPR activity.

1.6.4 – Control of Cas9/dCas9 turn over and localization

Several groups proposed engineering inducible CRISPR systems by controlling the stability of the Cas9 or dCas9 proteins (Feng et al. 2015; Senturk et al. 2017; Maji et al. 2016) (Fig. 1.14e). This was achieved by grafting degrons to the CRISPR-effector, which favour protein degradation via the ubiquitin proteasome system (OFF-state). The fusion protein could then be rescued by the delivery of an antagonist that protects Cas9 against proteolysis. Using this strategy, inducible Cas9 proteins were made responsive to digoxin, progesterone (Feng et al. 2015), Shield-1 (Senturk et al. 2017), trimethoprim, and 4OHT (Maji et al. 2016). Since sgRNA and Cas9/dCas9 have to co-localize in the cell nucleus for the crRNP to exert its function, it was suggested that CRISPR activity could also be controlled by segregating its components into distinct cellular compartments. Accordingly, Liu and colleagues have achieved conditional cytoplasmic sequestering of Cas9 by fusing it to several copies of the estrogen receptor ERT2 hormone binding domain. They showed that the CRISPR-effector is able to translocate to the nucleus and resume CRISPR activity upon addition of the ligand (4OHT) (K. I. Liu et al. 2016) (Fig. 1.14f).

1.6.5 – Inducible split Cas9/dCas9 variants.

Yet, another category of inducible Cas9/dCas9 was created by engineering split variants of the CRISPR-effector, whose two halves are re-assembled in the presence of a particular inducer (Zetsche, Volz, et al. 2015; Wright et al. 2015; Nihongaki, Kawano, et al. 2015; Nguyen et al. 2016) (Fig. 1.15a-c). Doudna and colleagues were the first to show that a split Cas9 variant could dimerize in the presence of a sgRNA and resume catalytic activity (Wright et al. 2015). Following this work, the Zhang's lab at MIT introduced a rapamycin-inducible split Cas9 variant (Zetsche, Volz, et al. 2015), created by attaching to the C- and N- termini halves of Cas9 to the FK506 binding protein 12 (FKBP) and FKBP rapamycin binding (FRB) domains of mTOR, respectively (Fig. 1.15a). Using this design, the group was able to demonstrate rapamycin-induced CRISPR-based genome editing (split Cas9) and transcriptional activation (split dCas9-VP64) in mammalian cells. This design, which suffered from a certain degree of

CRISPR activity in the absence of rapamycin (OFF-state leakage), was later improved by the work of Nguyen et al. (Nguyen et al. 2016) who added the ligand-binding domain ERT to each fragment, forcing both halves into the cytoplasm in the absence of 4OHT (Fig. 1.15b). As a result, the system was only activated when both 4OHT (translocation to the nucleus) and rapamycin (dimerization) were delivered to the cells. Finally, light-induced split Cas9 dimerization was demonstrated by Nihongaki and colleagues, who used the Magnet system to create photoactivatable Cas9 and dCas9 proteins (Nihongaki, Kawano, et al. 2015) (Fig. 1.15c). Adapted from photoreceptors found in the fungus *Neurospora crassa* (Zoltowski et al. 2007), the Magnet system features two protein domains, referred to as positive (pMag) and negative Magnets (nMag), which dimerize via electrostatic interactions when exposed to blue light (Kawano et al. 2015). By fusing each Magnet to the C- and N- terminus of split Cas9 and dCas9, the authors were able to show light-dependent genome editing and transcriptional regulation in human cell lines (Nihongaki, Kawano, et al. 2015).

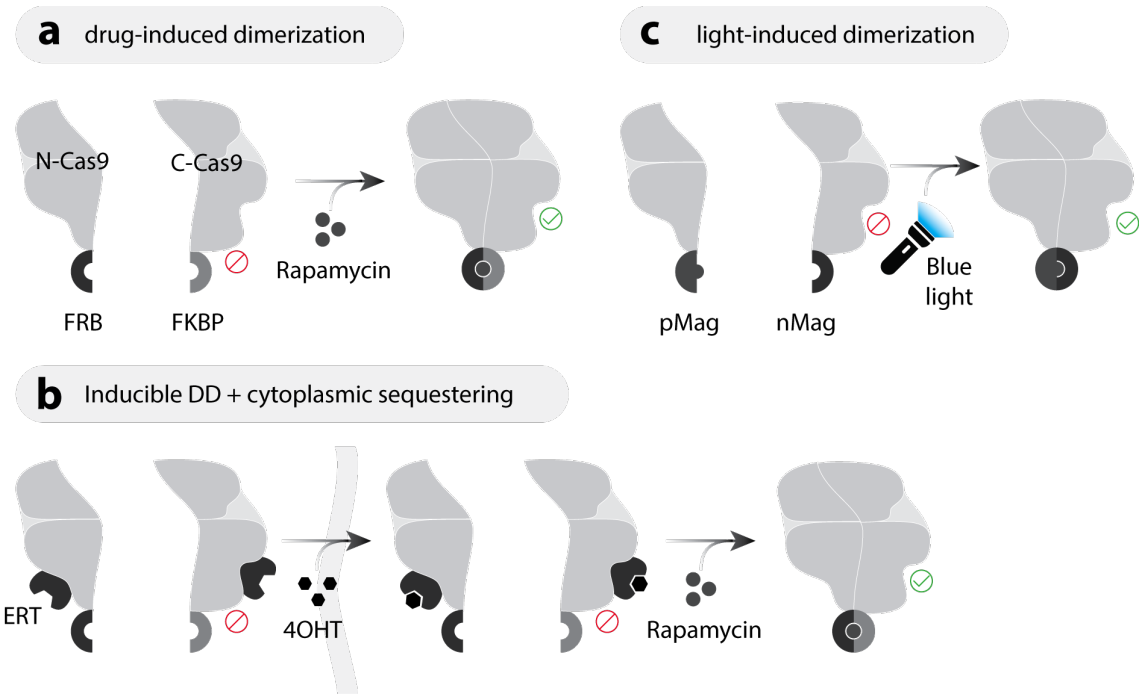

Figure 1.15 – Inducible split Cas9/dCas9. (a-c) The CRISPR-effector is split into a N-terminus Cas9 (N-Cas9) and a C-terminus Cas9 (C-Cas9), which are then fused to various dimerization domains. Dimerization is either chemically induced (a, b) of triggered by exposure to blue light (c).

1.6.6 – Conditional effector domain tethering.

As explained in section 1.5, CRISPR-TR are constructed by fusing effector domains to the crRNP[9]. Therefore, inducible CRISPR-TR solutions can be engineered by using ligand-dependent dimerization domains (DD) to conditionally recruit an effector of interest to the sgRNA-dCas9 complex (Fig. 1.16a-c). Following this principle, Nihongaki et. al. reported conditional gene activation in mammalian cells by using light-responsive dimerization domains (Nihongaki, Yamamoto, et al. 2015). The light-sensitive cryptochrome 2 (CRY2) and its binding partner, CIB1, from *Arabidopsis thaliana* (Kennedy et al. 2010), were respectively fused to dCas9 and VP64 so that they would bind to one another upon exposure to blue light (Fig. 1.16a). This approach was later improved by fusing CIB1 to both the N- and C- termini of dCas9 in order to recruit two CRY2-VP64 complexes to the crRNP (Polstein et al. 2015). In both papers, the authors showed, by tracking reporter expression over several days, that the resulting light-responsive CRISPR-TR could be reversibly and repeatedly activated (Polstein et al. 2015; Nihongaki, Yamamoto, et al. 2015). Additionally, spatial control over the system was exemplified using light patterns covering only a small portion of the cell population. In these experiments, reporter activation was solely observed in the irradiated subpopulation (Polstein et al. 2015; Nihongaki, Yamamoto, et al. 2015).

Several alternative designs were subsequently published, which employed chemically-induced dimerization. By using the dimerization domain pairs ABI-PYL1 and GID1-GAI24 to tether either KRAB or the VPR tail to dCas9, one group demonstrated repression and activation of endogenous loci in mammalian cells based on the delivery of abscisic acid and gibberellin, respectively (Y. Gao et al. 2016) (Fig. 1.16b). The authors were also able to simultaneously and independently control the transcriptional output of two target genes by coupling these pairs to distinct Cas9 orthologues. Furthermore, an AND gate was created by fusing the two pairs in series such that dCas9-DD1 recruited DD1*-DD2 fusion, which in turn recruited DD2*-effector fusion. In this way, activation of the target gene was conditioned on the simultaneous

[9] Note that CRISPRi is an exception as it only relies on dCas9

delivery of two inducers (Fig. 1.16b). Similarly, Bao et al. used the FKBP-FRB DD pair, introduced in the last section, to achieve conditional dCas9-effector fusion upon rapamycin delivery (Bao et al. 2017). Additionally, Maji and colleagues proposed a system whereby the effector domain is conditionally recruited to the sgRNA (Maji et al. 2016). The authors used modified scaffold-sgRNAs to accommodate MS2 or pp7 aptamers, which recruit effector domains fused with the DHFR or ER50 degrons. As such, the resulting CRISPR-TRs were maintained silent through degradation of the effector, and conditionally activated by delivery of specific antagonists. Using this design, the team notably created inducible CRISPR-TRs responsive to trimethoprim and 4OHT (Fig. 1.16c).

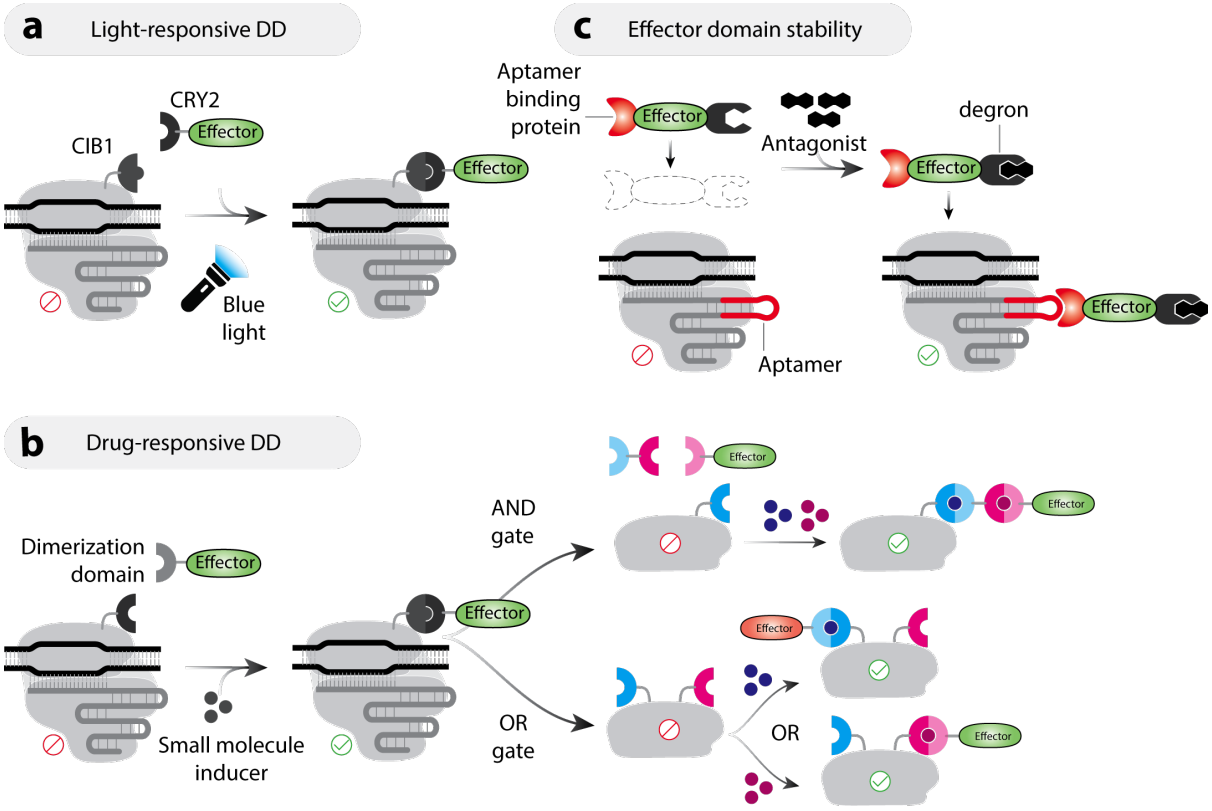

Figure 1.16 – Conditional effector domain binding to the crRNP. (a) Light-responsive dimerization domains (DD) are fused to dCas9 and the effector respectively. When blue light is shined, the effector gets recruited to the DNA locus targeted by the crRNP. **(b)** Similar to (a) except that DD dimerization is triggered by a small molecule inducer. Boolean AND or OR gates can be created by using DD in series or parallel. **(c)** Scaffold-sgRNA featuring an RNA aptamer recruits the cognate binding protein (red) along with a fused effector domain. Tethering a degron to the aptamer binding protein-effector complex makes it possible to regulate its availability and hence control CRISPR-TR activity.

1.6.7 – Anti-CRISPR proteins

Further analysis of the CRISPR system led to the discovery of phage-encoded proteins that can naturally inhibit Cas9 activity (Pawluk et al. 2014; Rauch et al. 2017). These proteins are now known as anti-CRISPRs. Discovered by *in silico* mining of the phages infecting bacteria with type-II C (Pawluk et al. 2014) and type-II A (Rauch et al. 2017) CRISPR systems, anti-CRISPR proteins have been shown to prevent both the nuclease activity and the binding capacity of several CRISPR-effectors, including the *Sp*Cas9. These studies have demonstrated that when delivered as plasmids along with the CRISPR-Cas9 system in HEK293-T cells, these small proteins could prevent CRISPR-based genome editing as well as CRISPR-TR activity.

1.7 – Aims, rationales, and significance of the work

The aforementioned methods employed to attain CRISPR inducibility have almost exclusively relied on protein engineering. While elegant, I argue below that such approaches, including switchable Cas9 and split variants, conditional effector tethers, and the use of anti-CRISPR, suffer from several limitations. For example, some of the inducible Cas9 strategies that have been successful at controlling genome editing might not be applicable to the development of inducible CRISPR-TRs. Indeed, while these studies have demonstrated impaired nuclease activity in the OFF-state due to conformational change, residue caging, etc., there is no indication that similar modifications on dCas9 might prevent on-target binding in the absence of the inducer. On the other hand, approaches that have been successful at controlling transcriptional activation/repression using conditional effector tethers also have caveats. In these cases, the DNA-binding properties of the crRNP stay unchanged between ON- and OFF-state. As such, dCas9 is expected to stay bound on-target even in the absence of the inducer. In addition to potentially have undesired effects on the transcriptome due to the fact that dCas9 can repress gene expression, the crRNP is expected to cause a certain degree of toxicity to the cells associated with its PAM probing activity (Hsu et al. 2013; Pattanayak et al. 2013; Fu et al. 2014; Zuris et al. 2015).

More importantly, I would argue that strategies achieving CRISPR inducibility through engineering of its protein component cannot be scaled up for the assembly of complex transcriptional programs. Because dCas9 loads all expressed sgRNAs in the cell regardless of their spacer sequence (i.e. target gene), programmed induction of CRISPR-effectors would lead to simultaneous activation at all targeted loci. For that reason, it is impossible to use such approaches to independently modulate the gene output of multiple targets.

A similar bottleneck was encountered by the field when trying to simultaneously activate and repress the expression of two distinct genes (*Gene*$_1$, *Gene*$_2$). In early 2014, this type of circuit would have required co-expression of dCas9-VP64 for activation, dCas9-KRAB for repression, and two sgRNAs programmed to target *Gene*$_1$ and *Gene*$_2$, respectively (Fig. 1.17a). In this instance, because dCas9 promiscuously binds both sgRNAs, dCas9-VP64 could have been regulating both *Gene*$_1$ and *Gene*$_2$, resulting in an outcome different from the intended one (Fig. 1.17a). Several studies having shown that dCas9-VP64 can act as a repressor depending on its position relative to the gene TSS (Mali, Aach, et al. 2013), *Gene*$_1$ and *Gene*$_2$ could in theory be activated and silenced, respectively, using the same dCas9-effector protein and carefully designed sgRNAs. Nevertheless, more elegant and versatile solutions were proposed by tethering the effector domains directly to scaffold-sgRNAs carrying RNA aptamers, thus linking the desired effect (activation or repression) to a specific target gene (Mali, Aach, et al. 2013; Zalatan et al. 2014). This was notably demonstrated by Zalatan and colleagues, who engineered two distinct scaffold-RNAs capable of recruiting the activator VP64 and repressor KRAB domain, respectively (Zalatan et al. 2014). The team was then able to assemble the gene circuit mentioned above, demonstrating concomitant activation and repression of two gene targets (Fig. 1.17b).

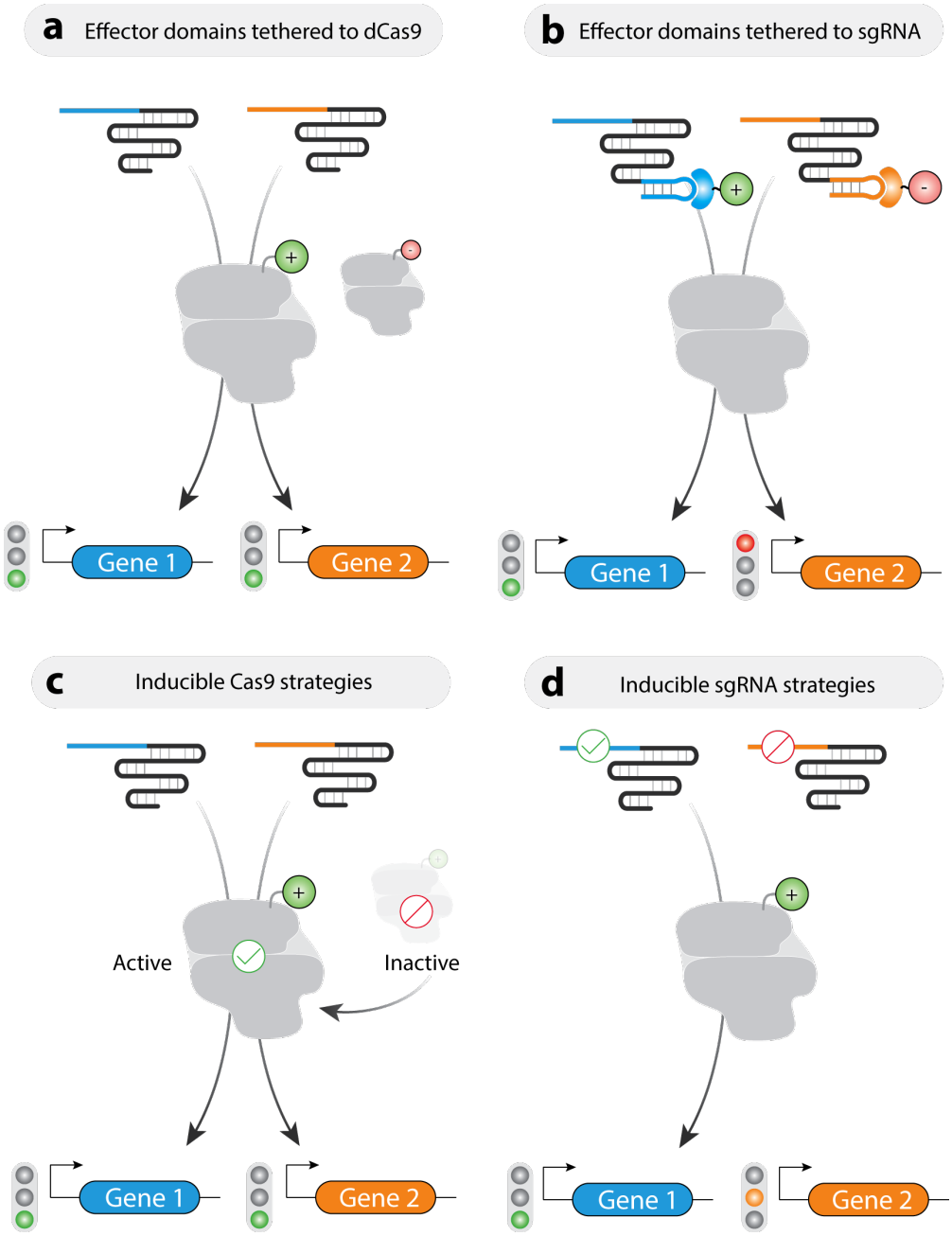

Figure 1.17 – Rationale for the creation of inducible sgRNAs. (a) Two sgRNAs with spacers programmed to target *Gene*$_1$ (blue) and *Gene*$_2$ (orange) are co-expressed along with dCas9(+) (transcriptional activator) and dCas9(-) (transcriptional repressor). Promiscuous sgRNA loading makes it impossible to guarantee that *Gene*$_1$ will be activated while *Gene*$_2$ is been repressed. **(b)** Two scaffold-sgRNAs targeting *Gene*$_1$ and *Gene*$_2$ recruits via RNA aptamers embedded in their sequence an activator and repressor domain, respectively. Using this approach, *Gene*$_1$ gets overexpressed while *Gene*$_2$ is silenced. **(c)** Two sgRNAs targeting *Gene*$_1$ and *Gene*$_2$ are expressed along with a switchable dCas9(+). Activation of the CRISPR-effector leads to the simultaneous upregulation of both gene targets. **(d)** Using inducible sgRNAs targeting *Gene*$_1$ and *Gene*$_2$ respectively, expressed along dCas9(+), makes it possible to independently activate the gene targets.

The same reasoning and conclusions applies when using inducible CRISPR-TRs to independently control the transcriptional output of two gene targets. In fact, promiscuous sgRNA loading by dCas9 make it impossible to sequentially activate *Gene*$_1$ and *Gene*$_2$ using a switchable Cas9 (Fig 1.17c). While solutions have been found by engineering inducible versions of Cas9 orthologues known to bind different gRNAs (Esvelt et al. 2013; Anders et al. 2016; Y. Gao et al. 2016), these approaches require extensive protein engineering and bear the non-negligible metabolic costs associated with the delivery of multiple CRISPR-effectors. Instead, I reasoned that a more elegant solution could be found in the creation of inducible sgRNAs designed to respond to orthogonal inducers. In that case, the sgRNA targeting *Gene*$_1$ could be first activated by delivery of its cognate trigger without effecting transcription of *Gene*$_2$ (Fig 1.17d).

Accordingly, I set out to re-engineer the *Sp* sgRNA to develop a range of inducible guides designed to activate in the presence of specific inducers. Using a previously reported CRISPRa assay, I first demonstrate how a simple modification of the sgRNA sequence can completely abrogate CRISPR-TR activity (chapter 2). Such “dormant” sgRNAs were created by appending their 5’-end with a short RNA extension bearing sequence complementarity to the spacer of the guide. Post-transcription, this segment base-pairs with the spacer to form a “spacer-blocking hairpin” (SBH), which prevents interactions between the sgRNA and its DNA target. After showing that SBH-sgRNAs are fully inert (chapter 2), I generate a wide range of inducible hairpins (iSBH) by replacing the loop of the structure with various sensing-loops, which get cleaved in the presence of particular inducers (chapter 3, 4, 7). Using this modular strategy, I then demonstrate conditional CRISPR-TR activity using switchable iSBH-sgRNAs designed to specifically respond to protein (chapter 3), oligonucleotide (chapter 4), and small molecules (chapter 7) inducers. As such, iSBH-sgRNAs act as transducers each encoding a causal relationship between the presence of a particular trigger and the modulation of downstream gene targets. Demonstrating the relevance of the approach for the design of

synthetic gene circuits, I show in chapter 5 that iSBH-sgRNAs can easily be reprogramed and multiplexed to achieve simultaneous or independent control of multiple reporter genes.

In order to help democratize the use of the iSBH platform, I have distilled the design principles derived throughout the project to develop a web tool that automates the construction of iSBH-sgRNA (chapter 6). Additionally, I show in the same chapter that the iSBH approach can be adapted to create inducible versions of the state of the art SAM system, and use the resulting iSBH-SAM CRISPR-TRs to control the transcriptional output of human genes in their endogenous context. Lastly, I present proof-of-principle data showing that aptazymes can be employed as RNA cleaving units in the design of iSBH-sgRNAs, hence opening up the scope of potential iSBH inducers (chapter 7).

Chapter 2 – Spacer-Blocking Hairpin

The goal of the research presented in this chapter was to identify modifications of the *Sp* sgRNA sequence that effectively abrogate CRISPR-TR activity. I notably aimed at creating a strong OFF-state from which the system could then be turned ON. As such, the alterations proposed to maintain the sgRNA silent had to be potentially reversible in order to allow the engineering of inducible strategies. The chapter starts with the implementation of a CRISPRa assay, that will be used throughout the project as the main experimental framework for testing switchable sgRNA designs. I then introduce and test the concept of using a spacer-blocking hairpin (SBH) appended to the 5'-end of the sgRNA as a mean of silencing its activity. After showing that this alteration is indeed able to fully halt CRISPRa, a series of experiments are presented which shed light onto the underlying mechanism of SBH-mediated silencing. Finally, the chapter ends with a refined analysis of the SBH structure and its potential use to modulate CRISPR-TR activity.

2.1 – CRISPR-based transcriptional activation assay

The first goal of this project was to set up a robust assay for engineering inducible sgRNAs, and demonstrating their relevance in the creation of conditional CRISPR-TR. Accordingly, I searched for established assays entailing activation or repression of reporter genes, that would satisfy the following criteria: (i) The assay should provide a clear read out of CRISPR-TR activity allowing to distinguish between a functional ON-state and a silenced OFF-state; (ii) The extent of transcriptional output variation observed between ON- and OFF-state should be

consequential enough to rank different design iterations based on their potency; (iii) The assay should preferentially be simplistic, easily reproducible, high-resolution, and high-throughput, enabling iterative testing of many hypotheses in a cost and time effective manner.

Two basic conceptual framework could be employed for these experiments: CRISPR-mediated activation (CRISPRa) or repression. Since my aim was to demonstrate activation of quiescent sgRNAs, I reasoned that CRISPRa would provide a more appropriate readout, because the system should transition from the absence of reporter gene expression to its induction. With a repression assay on the other hand, the reporter gene is constitutively expressed in the OFF-state and the transition to the ON-state is reflected by a reduction in protein levels. In this instance, minute changes in transcriptional output might not be easily detectable, especially if the system operates in a condition where transgene translation is saturated. This would notably be a problem when trying to quantify spurious CRISPR-TR activation in the OFF-state. In addition to having a higher sensitivity to system leakage, CRISPRa assays have consistently displayed a greater dynamic range than their silencing counterparts, as measured by fold change in transgene expression (Qi et al. 2013; Gilbert et al. 2013; Bikard et al. 2013). Additionally, as illustrated in the introduction of this Thesis, a larger body of work has been geared towards improving systems for transcriptional activation over transcriptional repression (see section 1.5). Regarding the choice of transgene type and the methodology used to assess its expression, I opted for the use of flow cytometry to quantify reporter fluorescent protein production due to obvious advantages including the ability to monitor expression at the resolution of a single cell, as well as time and cost effective nature of the readout (Givan 2011; Mali, Aach, et al. 2013; Gilbert et al. 2013; A. W. Cheng et al. 2013).

2.1.1 – Implementation

When the work presented in this Thesis was initiated (2013), the state of the art CRISPR-TR for CRISPRa made use of the catalytically inactive dCas9 fused to the effector domain VP64 (Gilbert et al. 2013; Mali, Aach, et al. 2013; A. W. Cheng et al. 2013; Maeder et al. 2013; Perez-

Pinera et al. 2013). At that time, it was shown that to mediate strong CRISPRa of a gene target one should co-localize multiple copies of dCas9-VP64 to its TSS (see section 1.5.1). This was achieved by delivering multiple sgRNAs, programmed with distinct spacer sequences, to cover the genomic region adjacent to the TSS of interest. In order to observe strong transgene activation while overcoming the need for sgRNA tiling in CRISPRa assays, synthetic promoters were engineered by cloning artificial target sequences upstream of a minimal promoter. Lacking binding sites for endogenous regulatory factors in the upstream activating sequence, these promoters instead featured an "operator" sequence comprising multiple repeats of the same 5'-N_{20}NGG-3' CRISPR target sites (CTS) (Gilbert et al. 2013; A. W. Cheng et al. 2013).

For the purpose of this work, I decided to adopt the synthetic promoter developed by Farzadfard et al., which featured a minimal cytomegalovirus (CMV) promoter, downstream of an operator sequence containing eight repeats of the same CTS (8xCTS-mCMVp) (Farzadfard et al. 2013; Nissim et al. 2014) (Fig. 2.1). Throughout the remainder of this thesis, two reporter plasmids featuring distinct operator sequences referred to as 8xCTS1 and 8xCTS2, are being used to drive the expression of the enhanced yellow fluorescent protein (EYFP) (Nagai et al. 2002) and the enhanced cyan fluorescent protein (ECFP) (Rizzo et al. 2004), respectively. I will refer to these reporter constructs as 8xCTS1-EYFP-pA and 8xCTS2-ECFP-pA. In order to test the feasibility of using such an assay for this research, HEK293-T (Graham et al. 1977) cells were transfected with the three following DNA plasmids (pDNA) (Fig 2.1): (i) a reporter construct, either 8xCTS1-EYFP-pA or 8xCTS2-ECFP-pA; (ii) a pDNA encoding the CRISPR-effector dCas9-VP64 fused to NLS to ensure its translocation into the nucleus; (iii) a sgRNA expression vector (U6p-sgRNA_SV40-iBlue). In the U6p-sgRNA_SV40-iBlue plasmid, the sgRNA was expressed under a pol-III U6 promoter, commonly used to produce short RNAs lacking a 5' cap and a poly-A tail (Murphy et al. 1986). Additionally, a constitutively expressed iBlue fluorescent protein was added to this vector in order to provide a readout for system delivery. Assuming that cells transfected with the sgRNA construct also received the other pDNAs, the expression of iBlue helped ruling out poor transfection efficiency in the case where

one of my systems failed to activate reporter expression. Furthermore, gating between iBlue positive and negative cells provided a means to characterize system performance only in the population that have received the CRISPR-TR.

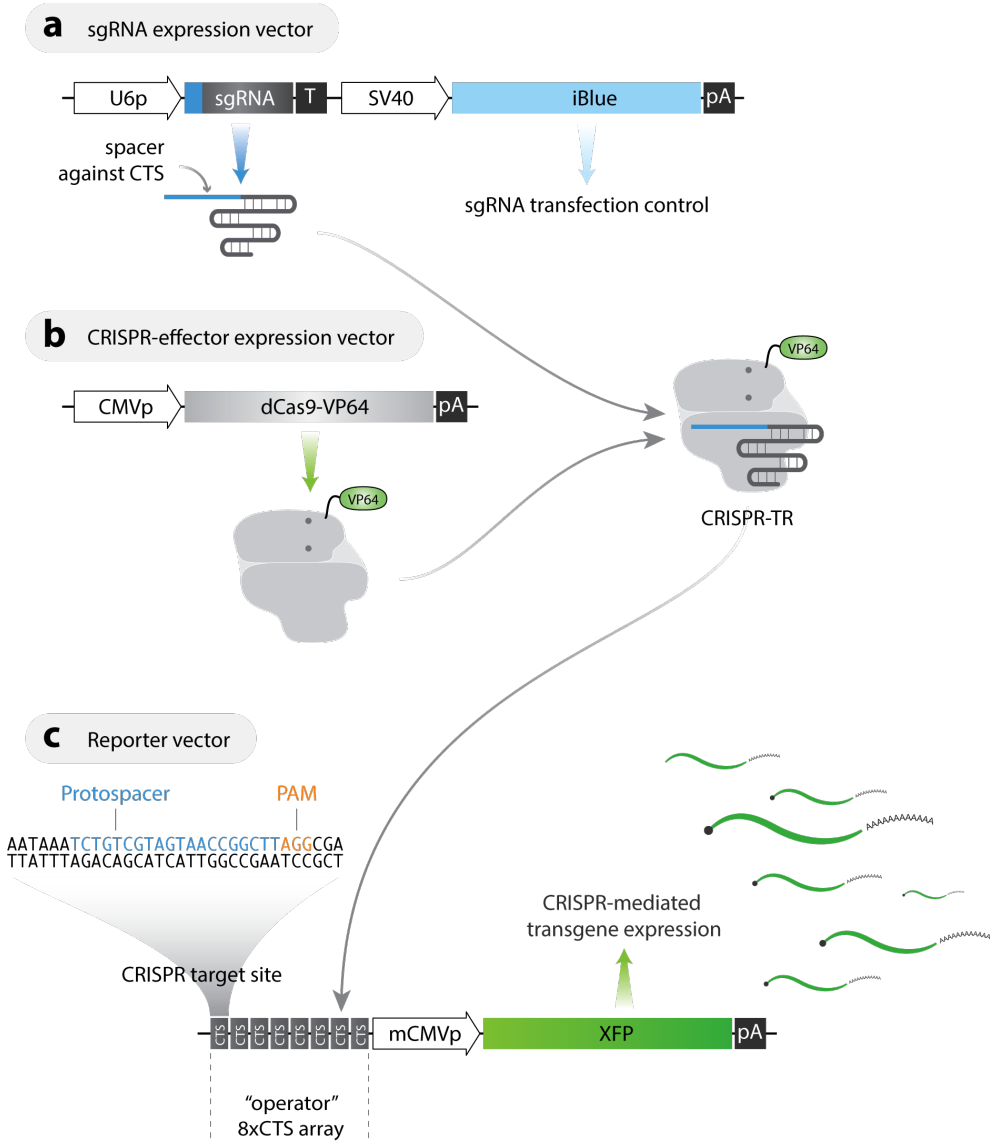

Figure 2.1 – CRISPRa assay. Overview of the components used to achieve CRISPRa of a reporter gene in human cells. The following three DNA plasmids are co-transfected: **(a)** sgRNA expression plasmid: the sgRNA, whose spacer segment has been reprogrammed to target the synthetic reporter on the reporter plasmid (c) is expressed under a pol-III U6 promoter (U6p). The fluorescent protein iBlue is constitutively expressed from a SV40 promoter and used as a transfection marker. **(b)** Vector expressing under a CMV promoter (CMVp) the protein fusion dCas9-VP64. **(c)** Reporter plasmid: An enhanced fluorescent protein (EYFP or ECFP) is expressed under a synthetic promoter featuring a minimal CMV promoter (mCMVp) flanked upstream by eight repeats of a given CTS (operator). The promoter drives reporter expression only when the CRISPR-TR binds the operator region.

As expected, when cells were transfected with a native "scrambled" sgRNA (nv-SCR)[10], whose spacer sequence was randomized, flow cytometry analysis 48h post-transfection revealed no reporter activation (Fig. 2.2). Conversely, when the sgRNA was reprogrammed to target the operator segment of the synthetic promoter (nv-CTS1 or nv-CTS2), strong EYFP or ECFP expression was observed, demonstrating CRISPR-TR specificity (Fig. 2.2c, d). Additionally, no crosstalk between sgRNA:promoter pairs was found in that, the delivery of nv-CTS1 sgRNA along with dCas9-VP64 did not drive expression of 8xCTS2-ECFP-pA, and conversely, nv-CTS2 sgRNA failed to activate 8xCTS1-EYFP-pA (data not shown).

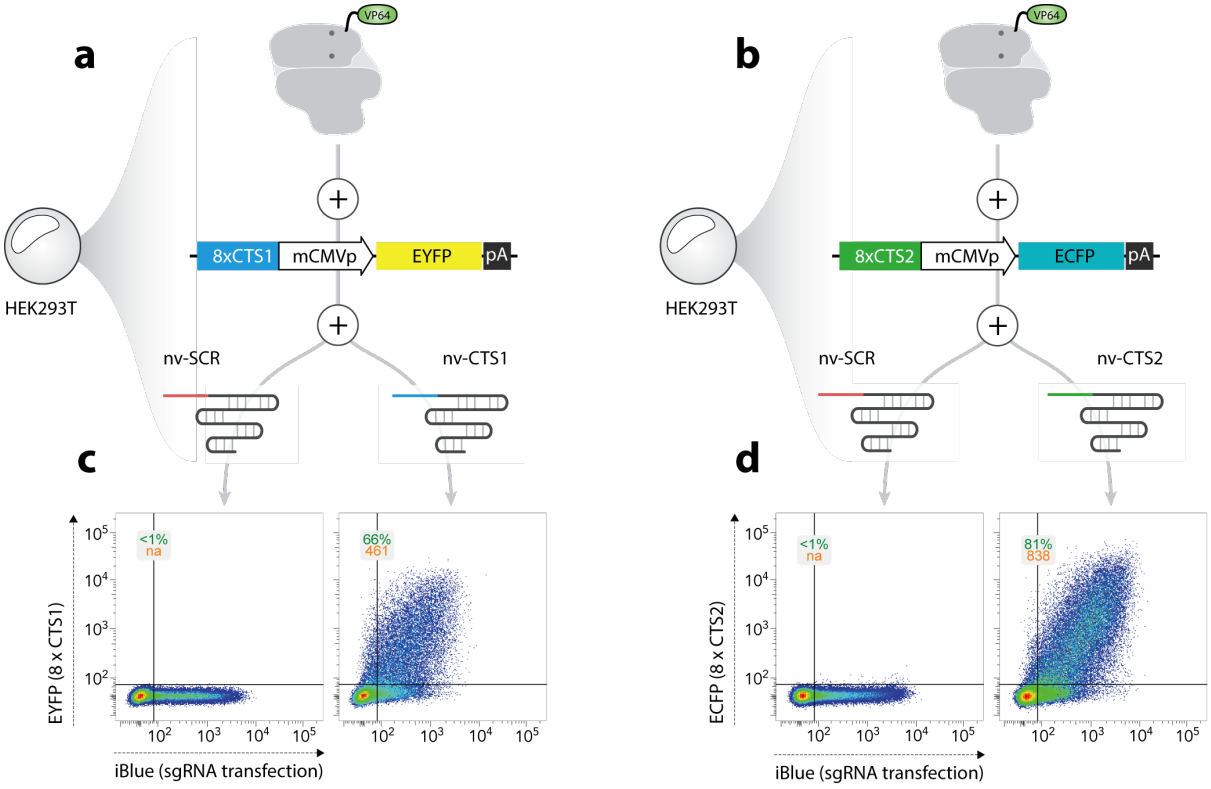

Figure 2.2 – Gene specific CRISPRa. (a,b) Transfection scheme for the CRISPRa assay using two sgRNA-reporter pairs: sgRNA is programmed to recognize the CTS1 (a) or CTS2 (b) operator. In each case either a native sgRNA with scramble spacer sequence (nv-SCR) or spacer for CTS (nv-CTS) is delivered. **(c,d)** Representative flow cytometry dot plots show reporter expression (y-axis, EXFP) against sgRNA transfection (x-axis, iBlue). Inset box show % of activated cells (double positive for EXFP and iBlue, green) and median reporter expression for this subpopulation (orange).

[10] Here native refers to the normal *Sp* sgRNA sequence and is used later to make the distinction with modified sgRNAs. The notation uses nv-X to refer to a native sgRNA with spacer targeting X.

2.1.2 – Activation score calculations

In order to rank the various iterations of switchable sgRNAs designed in this Thesis, a reporter activation score was defined, which provides a single measurement quantifying the extent of CRISPRa in a given condition. When creating an inducible CRISPR-TR system, one is interested in characterizing both ON- and OFF-state in terms of (i) the fraction of cells in which the system gets activated, and (ii) the potency of the system in this "activated" cell population. A comparison of these two parameters between ON- and OFF-states informs on the penetrance of the inducer as well as the strength of the induction. Additionally, both measurements provide valuable statistics to describe the potential OFF-state leakiness of the system (activation in the absence of the inducer).

In the experimental set up used for this work, the "% of activated cells" was defined as the fraction of cell displaying reporter expression ($EXFP^{+ve}$[11]) among the $iBlue^{+ve}$ population (Fig 2.3). Note that this measurement could be biased by the fact that not all cells transfected with the sgRNA expression plasmid ($iBlue^{+ve}$) also contained both the reporter and dCas9-VP64 constructs. Nevertheless, data presented in figure 2.2 show that activated cells account for a majority of the $iBlue^{+ve}$ population (66% for EYFP; 81% for ECFP), suggesting that most cells receiving the sgRNA pDNA also receive reporter and dCas9-VP64 plasmids (Fig 2.2c, d). On the other hand, the strength of activation was quantified by measuring the median of reporter fluorescence within the double positive cell population ($iBlue^{+ve}/EXFP^{+ve}$) (Fig 2.3). Finally, these two measurements were multiplied together, as previously done by others (Xie et al. 2011; Nissim et al. 2014; Ausländer et al. 2012), to derive an activation score reflecting both the strength and penetrance of the activation (Fig 2.3).

[11] EXFP refers to either EYFP or ECFP

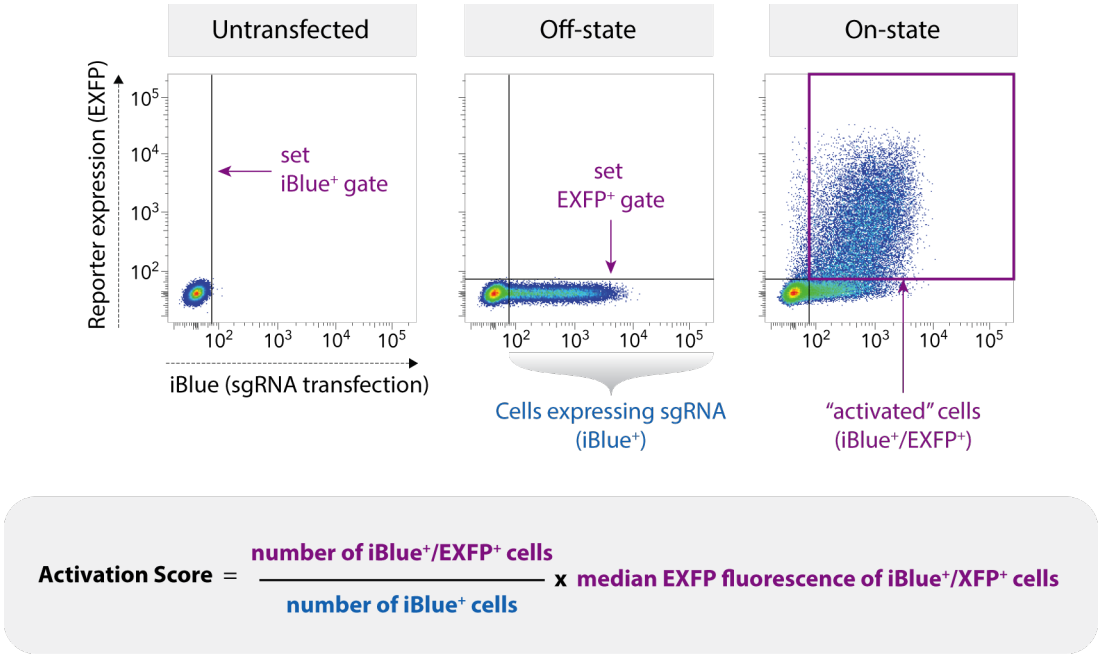

Figure 2.3 – Activation score calculation. Representative flow cytometry dot plots for three conditions showing transgene expression (y-axis, EXFP) against sgRNA transfection (x-axis, iBlue). Cell expressing sgRNAs (iBlue^{+ve}) and "activated" cells (double iBlue^{+ve} /EXFP^{+ve}) subpopulations are delineated (top). An activation score is derived for OFF-state and ON-state conditions by multiplying the % of activated cells with the median EXFP fluorescence for this subpopulation.

2.2 – sgRNA specification and function: How to silence a guide?

2.2.1 – Structural requirements of the sgRNA scaffold: allosteric sgRNAs.

The *Sp* sgRNA, as engineered by the Zhang's laboratory (Cong et al. 2013), can be separated into four distinct sections (Fig. 1.9b). From 5'- to 3'-end, the guide starts with a 20nt spacer segment. This variable part of the sgRNAs, providing the specificity of the system, can be modified to target the crRNP to different DNA loci. The following segment, known as repeat:antirepeat stem loop or *tetraloop*, is the hybridization point between the naturally occurring crRNA and tracrRNA. It is comprised of a lower and an upper stem, separated by a bulge. After that comes stem loop 1, also known as *nexus*, followed by the tracrRNA tail, made of stem loop 2 and 3 (Fig. 1.9b). The entire tetraloop-nexus-tail segment is commonly referred to as sgRNA scaffold.

While the role of the spacer has been known for a long time, recent mutational analyses of the guide scaffold have helped delineating sequence and structural requirements essential for sgRNA functionality, and consequently, CRISPR-Cas9 activity (Nishimasu et al. 2014; Briner et al. 2014). These analyses have notably shown that structural alterations in the tetraloop bulge, as well as sequence and structural modifications in the nexus segment, cause complete ablation of Cas9-mediated cleavage both *in vitro* and *in vivo*. Crystallography analysis of the crRNP structure later revealed that these sgRNA segments make extensive interactions with the Cas9 recognition and bridge helix domains, and as such, are key in inducing conformational changes required for PAM searching and target identification (Nishimasu et al. 2014; Jinek et al. 2014; Anders et al. 2014). Additionally, stem 2 and 3 of the scaffold were also found to help stabilize the sgRNA-Cas9 complex, and therefore to be essential for CRISPR activity.

These functional hotspots in the scaffold provide potential targets for the creation of switchable sgRNAs. One could imagine creating allosteric sgRNAs that can be toggled between active and inactive conformations by delivering a ligand that reversibly bind to the guide to alter its structure. Such allosteric structures have been engineered in the past by linking an RNA of interest with RNA aptamers capable of binding various classes of ligands (Kang & Y.-S. Lee 2013; Ausländer & Fussenegger 2013)[12]. Amongst many applications, this strategy has been used to create allosteric self-cleaving ribozymes (Wieland et al. 2012; Y. Y. Chen et al. 2010; Ausländer et al. 2010), ligand-responsive pri-microRNA hairpins (An et al. 2006; Kumar et al. 2011; Beisel et al. 2011); and conditional exon-splicing (D.-S. Kim et al. 2008; Culler et al. 2010). Such allosteric RNAs were designed based on the idea that a ligand-dependent conformational change in the aptamer segment can be propagated to the RNA of interest to alter its structure, and consequently its function (J. Tang & Breaker 1997).

[12] See chapter 7 for more details on the use of aptamers.

Similarly, allosteric sgRNAs could in theory be constructed by modifying the sequence of the tetra-loop segment to incorporate an RNA aptamer, whose conformational change would disrupt the bulge structure. However, when considering such solutions, I anticipated that the implementation of allosteric sgRNAs might be compromised by the following: (i) while disruption of the tetra-loop bulge structure has been shown to abrogate Cas9 nuclease activity (Briner et al. 2014), there is no evidence suggesting that it would affect dCas9 binding. Consequently, this approach might not help in the construction of inducible CRISPR-TRs. (ii) dCas9 binding to the OFF-state sgRNA, or inactive sgRNA conformation, might force the later into its active conformation despite the absence of a ligand. (iii) Finally, as I will show in chapter 7 and others have reported in the literature, allosteric RNAs can display significant leakiness in their OFF-state (Wieland et al. 2012; Y. Y. Chen et al. 2010; Ausländer et al. 2010). Based on these considerations, I sought to use a different approach for the creation of inducible sgRNAs.

2.2.2 – Blocking sgRNA:DNA target interactions, spacer sequestering

The spacer sequence of the sgRNA is central to the process of protospacer recognition, crRNP anchoring on-target, and Cas9-mediated double stranded DNA breaks (Jiang & Doudna 2017). Studies focusing on *Sp*Cas9 specificity have shown the importance of the first 10-12 spacer nucleotides (PAM proximal), termed the “seed” region, in distinguishing between ON and OFF-targets (Jinek et al. 2012; Cong et al. 2013; Fu et al. 2013; Hsu et al. 2013). Additionally, genome-wide binding analysis of the sgRNA-dCas9 complex using chromatin immunoprecipitation followed by sequencing (ChIP-seq), reported that as little as five PAM proximal nucleotides was sufficient to mediate strong dCas9 binding, and a longer seed sequence was required when the spacer contained a stretch of U nucleotides (Xuebing Wu et al. 2014). Subsequent structural analysis of the crRNP later revealed that while PAM distal spacer nucleotides are usually disordered, interactions with the CRISPR-effector maintain the seed nucleotides in an helical conformation, thermodynamically favorable for RNA:DNA pairing (Nishimasu et al. 2014; Anders et al. 2014; Jiang et al. 2015). Once the PAM flanking

the DNA target has been recognized by the PI domain, Cas9/dCas9 locally opens up the DNA double helix and presents the opposite DNA strand to the seed sequence of the guide (Sternberg et al. 2015) (Fig. 1.8). Sequence complementarity between the seed and the DNA target further denatures the double helix and leads to the progressive formation of an R-loop structure required to stabilize the crRNP on-target. Given the critical role of the sgRNA spacer segment, I reasoned that inducible sgRNAs could be engineered by reversibly interfering with its availability.

2.3 – Spacer-Blocking Hairpins

To prevent base-pairing between the sgRNA spacer and the DNA target, I envisioned extending the sgRNA on its 5'-end with an RNA sequence bearing complementarity to the spacer. The idea here was that this extension, referred to as "back-fold", would post-transcriptionally hybridize with the spacer to form a "spacer-blocking hairpin" (SBH) (Fig. 2.4). The resulting SBH-sgRNA should not be able to direct the crRNP to its DNA targets hence halting CRISPR-TR activity.

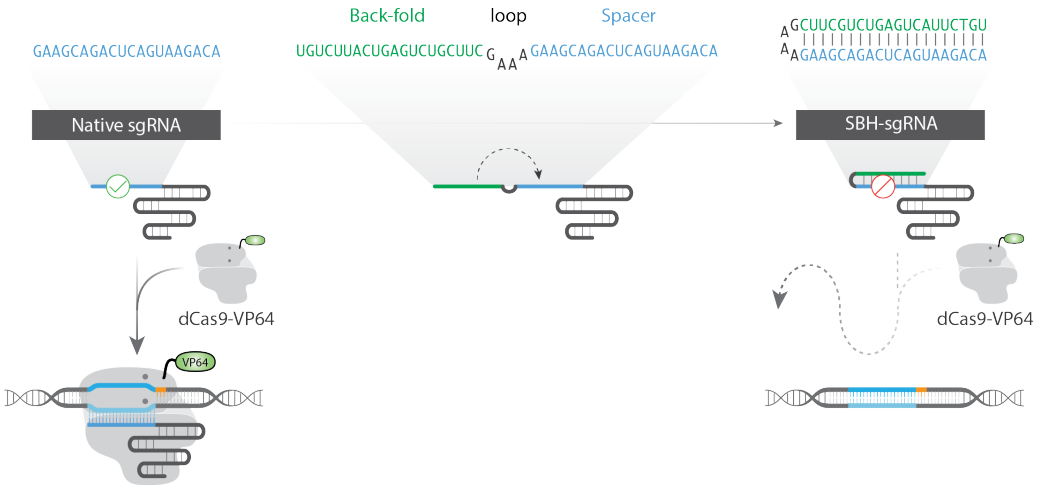

Figure 2.4 – SBH-mediated silencing of CRISPRa: in theory. A native sgRNA (left) is appended on it 5'-end with an RNA segment complementary to its spacer (back-fold) through a GAAA loop (middle). Post-transcriptionally the back-fold hybridizes with the sgRNA spacer to form a stem-loop structure referred to as spacer-blocking hairpin (SBH). The resulting SBH-sgRNA (right) should not be able to guide dCas9-VP64 on-target and consequently abrogate CRISPRa.

2.3.1 – SBH fully silences CRISPR-TR activity

To test the hypothesis that sgRNAs can effectively be silenced using SBHs, a full length back-fold was appended to the 5'-end of the nv-CTS1 sgRNA through a "GAAA" linker (Fig. 2.5a). The length and sequence of the back-fold were chosen such that the resulting SBH would sequester all spacer nucleotides. To further increase stem stability, the back-fold was extended to also base-pair with the first nucleotide of the sgRNA scaffold. The resulting SBH$^{(0)}$CTS1-sgRNA, where (0) stands for zero free spacer nucleotides, was then transfected into HEK293-T cells along with the dCas9-VP64 and reporter plasmid. Flow cytometry analysis 48h post-transfection revealed no detectable reporter expression, suggesting complete abrogation of CRISPRa (Fig. 2.5b, c). In fact, the result was comparable to what was observed when nv-CTS1 was replaced for a scramble control nv-SCR (Fig. 2.5b, c).

To confirm that the observed silencing was due to SBH-mediated sequestration of the spacer, we designed several control sgRNAs featuring 5' extensions recapitulating the length and/or structure of the SBH without base-pairing with the spacer nucleotides (Fig. 2.5d): SBH$^{(ctr-1)}$CTS1-sgRNA was designed such that the 20nt extension would fold on itself to form a 10bp stem-loop in front of the guide. SBH$^{(ctr-2)}$CTS1-sgRNA included an off-set 20bp stem-loop in front of the sgRNA sequence. Finally, the back-fold sequence of SBH$^{(0)}$CTS1-sgRNA was scrambled to create SBH$^{(ctr-3)}$CTS1-sgRNA. All three control sgRNAs were transfected separately in HEK293-T along with the CRISPR-effector and reporter constructs, and tested for their ability to mediate CRISPRa. The experiments revealed that all three degenerate SBH-sgRNAs could drive strong reporter expression, albeit weaker than the activation observed when using the native guide (nv-CTS1, Fig. 2.5c, d). Together these results suggest that SBH-mediated silencing of CRISPR-TR is not solely imparted by the hairpin structure but rather requires base-pairing between the back-fold and the spacer.

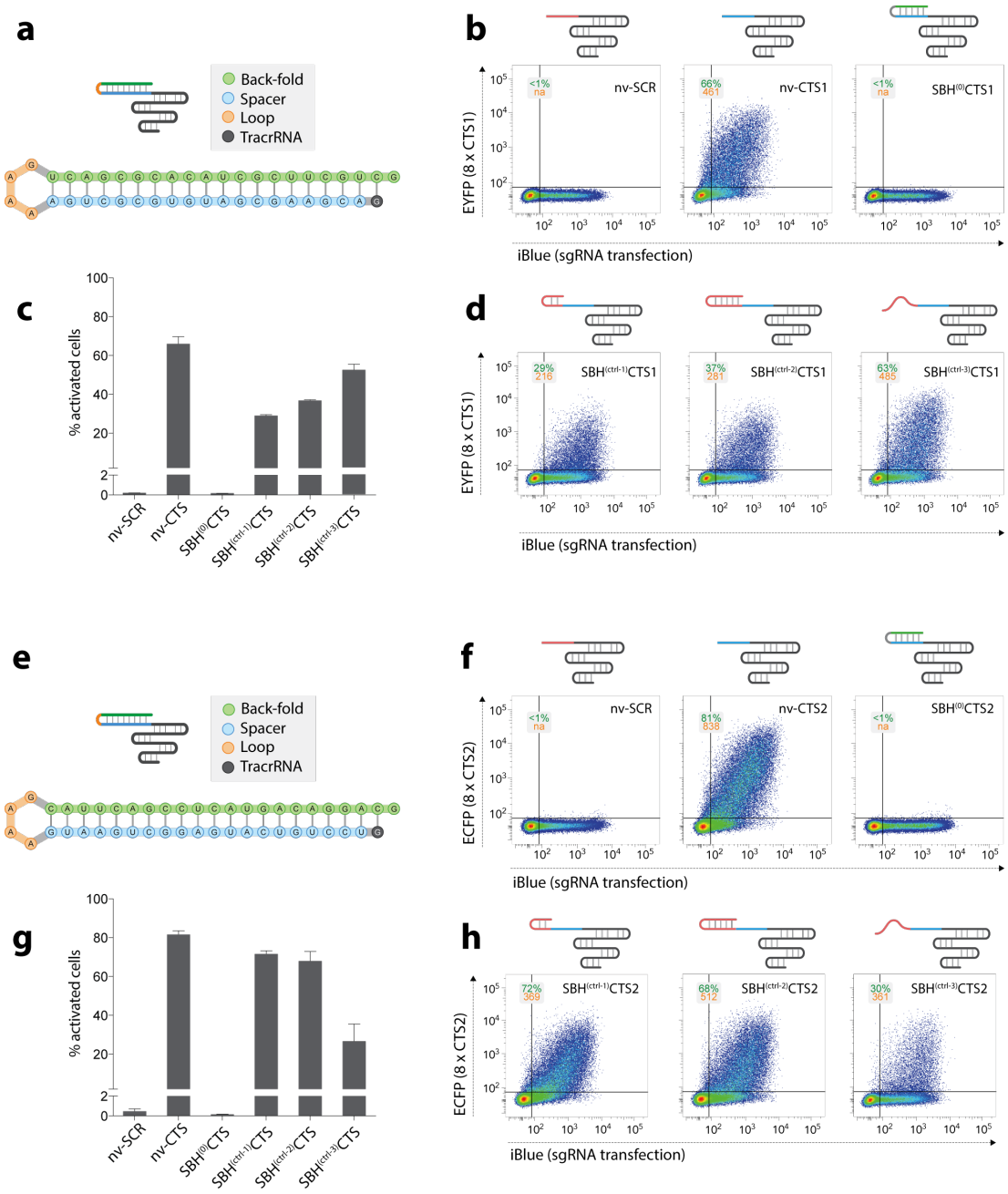

Figure 2.5 – SBH-mediated silencing of CRISPRa: in practice. (a) Sequence and RNA secondary structure of the SBH$^{(0)}$CTS1 hairpin. **(b)** HEK293-T cells were co-transfected with dCas9-VP64, 8xCTS1-EYFP-pA and either native sgRNA with scramble spacer (nv-SCR), native sgRNA targeting CTS1 (nv-CTS1), or SBH-sgRNA with the SBH$^{(0)}$CTS1 hairpin. Representative flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue). % of activated cell (green) and median reporter fluorescence for this population (orange) are given in the inset. **(c)** Quantification of % activated cells for conditions in (b) and (d) (n = 3, mean +/- s.d.). **(d)** HEK293T cells were co-transfected with dCas9-VP64, 8xCTS1-EYFP-pA reporter and various control SBH-sgRNAs: SBH$^{(ctr-1)}$CTS1 (10p off-set hairpin), SBH$^{(ctr-2)}$CTS1 (20bp off-set hairpin), SBH$^{(ctr-3)}$CTS1 (floating 20nt scramble back-fold). **(e-h)** Repeat of (a-d) with a different sgRNA:reporter pair (nv-CTS2 sgRNA and 8xCTS2-ECFP-pA reporter).

To show that SBH-mediated silencing of CRISPRa was independent of the sgRNA chosen, the experiments presented above were repeated using the nv-CTS2 sgRNA, which targets the 8xCTS2-ECFP-pA reporter (Fig. 2.5e-h). Applying the same design principles described above, the sequence of the nv-CTS2 sgRNA was modified to generate its silent counterpart SBH$^{(0)}$CTS2 (Fig. 2.5e). Again, it was found that full base-paring between the CTS2 spacer and the sequence complementary back-fold had for effect to completely silence CRISPR-TR activity (Fig. 2.5f, g). Additionally, as observed with the CTS1 sgRNA-reporter pair, experiments conducted with degenerate versions of the SBH$^{(0)}$CTS2-sgRNAs revealed that SBH-mediated silencing was contingent on spacer sequestering (Fig. 2.5g, h).

2.3.2 – Mechanism behind SBH-mediated silencing

As mentioned in the preamble of the chapter, spacer-blocking hairpins are expected to silence CRISPR-TR activity by hindering hybridization between the sgRNA spacer and the DNA target. Nevertheless, one could argue that SBH-mediated silencing might as well be caused by one or a combination of the following factors: (i) Appending the SBH to the guide could prevent pol-III transcription such that no SBH-sgRNA would be present in the cell; (ii) SBH-sgRNAs could be more prone to endo and exonuclease mediated degradation compared to native sgRNAs; or (iii) the SBH structure could prohibit loading of the SBH-sgRNA into the CRISPR-effector. Below I present some arguments as well as additional experiments to help pinpoint the origin of SBH-mediated silencing.

In section 2.3.1, I constructed and tested "degenerate" SBH-sgRNAs created by appending on the 5'-end of a native guide a back-fold which folds into a 20bp hairpin without base-paring with the spacer (SBH$^{(ctr-2)}$CTS1 and SBH$^{(ctr-2)}$CTS2). Such mutant SBH-sgRNAs are particularly suited to test whether or not the presence of an RNA hairpin impairs the transcription of SBH-sgRNAs and/or favour their degradation. Results presented in this chapter showed that both SBH$^{(ctr-2)}$CTS1 and SBH$^{(ctr2)}$CTS2 constructs were fully able to direct dCas9-VP64 on target and mediate transcriptional activation of their corresponding reporter (Fig. 2.5c, d, g, h). These findings suggest that the presence of a 20bp RNA hairpin on the 5'-end of the sgRNA does

not sufficiently affects the transcription and/or stability of SBH-sgRNA to explain SBH-mediated silencing. Consequently, I reasoned that the silencing of potent sgRNAs observed when using an SBH structure was due either to the incapacity of dCas9 to load SBH-sgRNAs or the inability of the sequestered spacer to hybridise with its complementary DNA target.

Accordingly, I next sought to test whether the dCas9-VP64 could load SBH-sgRNAs in spite of the 20-nt hairpin structure. To this end, an “sgRNA competition” experiment was devised, whereby the modified sgRNA under investigation would be tested for dCas9 loading by assessing its ability to titrate the CRISPR-effector away from a control sgRNA. It was shown in section 2.3.1 that co-transfection of dCas9-VP64, 8xCTS-EXFP-pA, and the native sgRNA targeting CTS (nv-CTS) leads to robust CRISPR-mediated reporter expression (Figure 2.2). In addition, delivery of a control sgRNA with a scramble spacer (nv-SCR), failed to stimulate EXFP expression (Figure 2.2). Consequently, I reasoned that the transfection of increasing concentrations of nv-SCR to a cell expressing a fix level of dCas9-VP64, reporter, and nv-CTS sgRNA, would lead to a progressive reduction in CRISPRa due to an accumulation of inactive nv-SCR:dCas9-VP64 complexes to the detriment of active nv-CTS:dCas9-VP64 (Fig. 2.6a). Similarly, the loading of SBH-sgRNAs in the CRISPR-effector could be tested by using an SBH-sgRNA in place of nv-SCR in this competition assay.

To first determine if dCas9-VP64 could be titrated away from nv-CTS using nv-SCR as a competitor, three conditions were tested where a fix amount of nv-CTS pDNA was transfected into HEK293-T along with either no (condition 1), 50-fold more (condition 2), or 100-fold more (condition 3) nv-SCR pDNA. As stated above, I expected to see a progressive reduction of CRISPRa as the concentration of the competitor increased. To rule out pol-III exhaustion as the cause of CRISPRa reduction, a decoy plasmid expressing a random RNA from a U6 promoter was supplemented in conditions 1 and 2, such that the DNA concentration of the decoy + nv-SCR constructs remained constant across all conditions (Fig. 2.6a).

Following optimization of dCas9-VP64, reporter 8xCTS2-ECFP-pA, and nv-CTS2 concentrations (data not shown), the competition assay was conducted using the nv-SCR guide as a competitor. Flow cytometry analysis of all three conditions revealed a progressive reduction in reporter expression as the ratio of nv-SCR to nv-CTS guides was increased (Fig. 2.6b, c). This was notably reflected in the quantification of the activation score for each condition (Fig. 2.6f).

In order to assess SBH-sgRNA loading into the CRISPR-effector, the competition assay was then repeated using the SBH$^{(0)}$CTS1-sgRNA as a competitor in place of nv-SCR. In this case, flow cytometry analysis also revealed a progressive loss in CRISPRa of the reporter gene as the concentration of SBH-sgRNAs was increased, suggesting that these guides were able to complex with dCas9-VP64 and titrate the effector away from nv-CTS2 guides (Fig. 2.6d, e, f). Interestingly, when comparing activation scores between experiments, it appeared that the SBH-sgRNA was less efficient at recruiting the CRISPR-effector than nv-SCR at matching plasmid concentrations (Fig. 2.6f). This effect could reflect a lowered binding affinity between dCas9 and SBH-sgRNA caused by the steric hindrance associated with the presence of the hairpin. Taken all together, these results suggest that SBH-sgRNAs are being loaded into dCas9-VP64 and consequently silence CRISPR-TR activity primarily by preventing crRNP docking at the DNA targets.

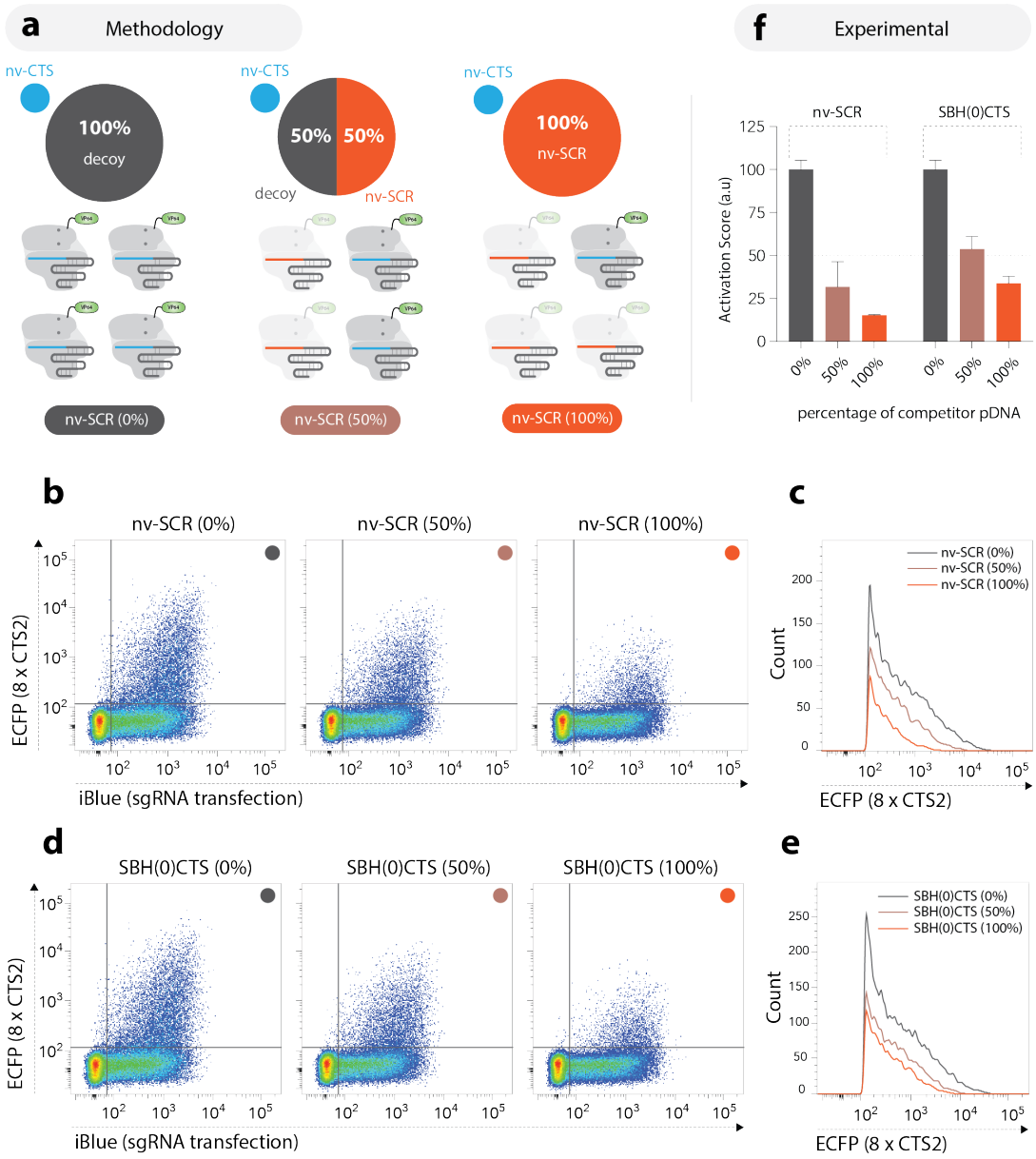

Figure 2.6 – Guide competition assay. (a) Overview of the competition assay designed to test loading of a modified sgRNA into dCas9. dCas9-VP64 is progressively titrated away from an active sgRNA (nv-CTS) by an inactive "competitor" sgRNA (nv-SCR or SBH$^{(0B)}$CTS) as the concentration of the later increases. Three conditions are tested whereby the competitor is not present (nv-SCR(0%)), present at 50% (nv-SCR(50%)), or 100% (nv-SCR(100%)) of the competitor pool. A decoy plasmid expressing a random RNA expressed under a U6 promoter is co-transfected to make up for the difference. **(b)** HEK293-T were co-transfected with dCas9-VP64, 8xCTS2-ECFP-pA reporter, 10ng of nv-CTS2 sgRNA and different competitor cocktails: (i) 1000ng of decoy (nv-SCR(0%)); (ii) 500ng of decoy, 500ng of nv-SCR (nv-SCR(50%)); (iii) 1000ng of nv-SCR (nv-SCR(100%)). **(c)** Histogram of ECFP expression for the double positive cell population (iBlue^{+ve}/ ECFP^{+ve}). **(d, e)** Similar to (b, c) with SBH$^{(0)}$CTS1 replacing nv-SCR as competitor. **(f)** Activation score calculation for all conditions in (b) and (d) (n = 3, mean +/- s.d., normalized to the nv-SCR(0%) condition).

2.3.3 – Back-fold requirements for effective silencing

Next, I sought out to investigate how varying the length of the back-fold segment affects CRISPR-TR activity. Accordingly, the RNA extension was shortened from full coverage of the spacer (SBH$^{(0)}$CTS1-sgRNA) to create SBH$^{(5)}$CTS1, SBH$^{(10)}$CTS1, and SBH$^{(15)}$CTS1-sgRNA having 5, 10 and 15 free spacer nucleotides, respectively (Figure 2.7a). HEK293-T cells were separately transfected with the CRISPRa system components and each of the three SBH-sgRNAs. Analysis 48h post-transfection showed that releasing 5 spacer nucleotides (SBH$^{(5)}$CTS1) was not sufficient to mediate reporter activation (Fig. 2.7b). Conversely, both transfections of SBH$^{(10)}$CTS1-sgRNA and SBH$^{(15)}$CTS1-sgRNA led to strong ECFP expression. Subsequent characterization of the activated population revealed that both the numbers of activated cells and the strength of reporter expression appeared to increase with the number of free spacer nucleotides (Fig. 2.7b). This inverse trend between the length of the back-fold and the degree of CRISPRa was also observed when repeating the same set of experiments on a different spacer:reporter pair (CTS2) (Fig. 2.8a-c).

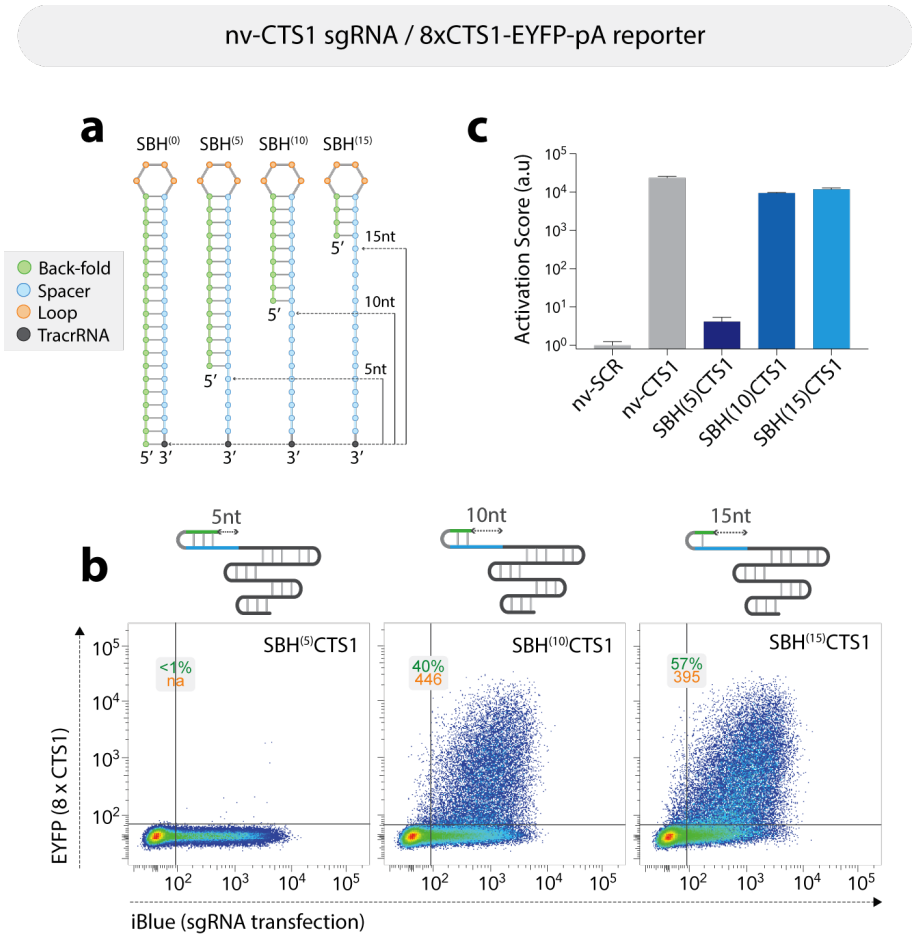

Figure 2.7 – Effect of back-fold shortening on CRISPRa, CTS1. (a) Schematic of SBH back-fold shortening from full coverage of the 19nt long CTS1 spacer ($SBH^{(0)}$CTS1) to 5 ($SBH^{(5)}$CTS1), 10 ($SBH^{(10)}$CTS1), and 15 ($SBH^{(15)}$CTS1) free spacer nucleotides. **(b)** HEK293-T were transfected with dCas9-VP64, reporter construct, and each SBH-sgRNAs in (a). Representative flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue). Inset shows % of activated cell (green) and median reporter fluorescence for this population (orange). **(c)** Activation score calculation for all conditions in (b) and additional negative control (scramble sgRNA, nv-SCR) and positive control (native sgRNA, nv-CTS1) (n = 3, mean +/- s.d. normalized to the nv-SCR condition).

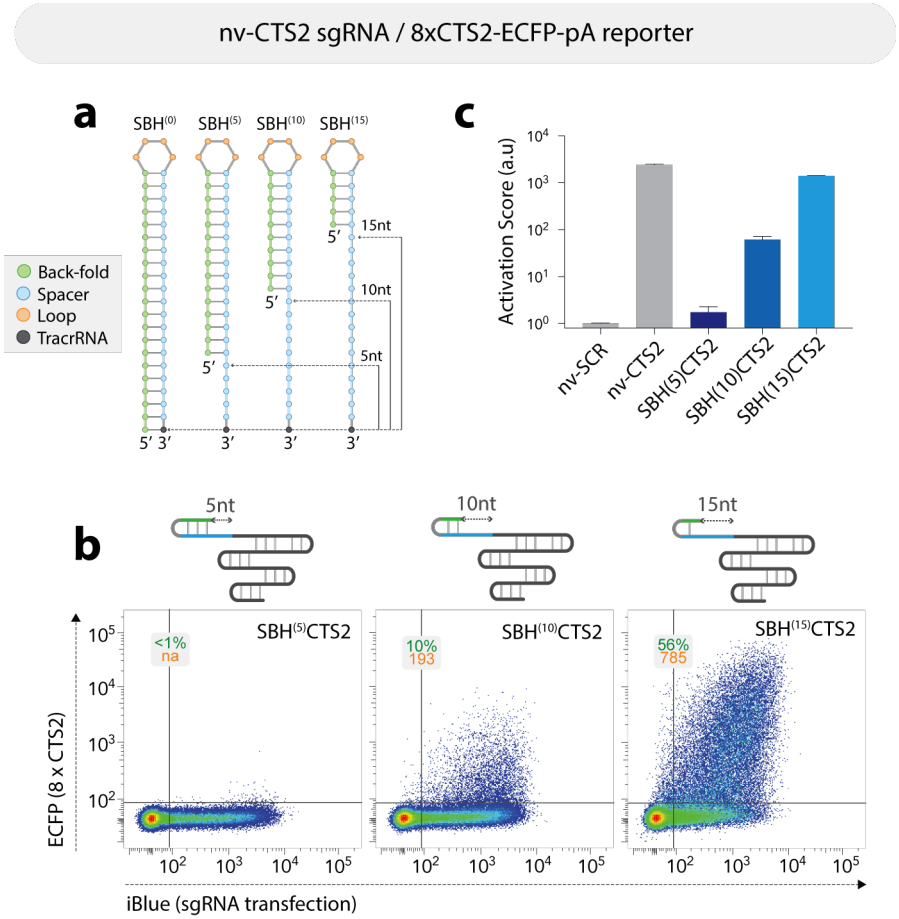

Figure 2.8 – Effect of back-fold shortening on CRISPRa, CTS2. (a) Schematic of SBH back-fold shortening from full coverage of the 20nt long CTS2 spacer ($SBH^{(0)}$CTS2) to 5 ($SBH^{(5)}$CTS2), 10 ($SBH^{(10)}$CTS2), and 15 ($SBH^{(15)}$CTS2) free spacer nucleotides. **(b)** HEK293-T were transfected with dCas9-VP64, reporter construct, and each SBH-sgRNAs in (a). Representative flow cytometry scatter plots show reporter expression (ECFP) against sgRNA transfection (iBlue). Inset shows % of activated cell (green) and median reporter fluorescence for this population (orange). **(c)** Activation score calculation for all conditions in (b) and additional negative control (scramble sgRNA, nv-SCR) and positive control (20nt sgRNA, nv-CTS2) (n = 3, mean +/- s.d. normalized to the nv-SCR condition).

To assess whether this correlation was related solely to the increase in the number of free spacer nucleotides or was also due to the SBH structure, shorter versions of the native sgRNA were created, with spacer sequences truncated to 5 ($nv^{(5)}$CTS), 10 ($nv^{(10)}$CTS), and 15 ($nv^{(15)}$CTS) nucleotides, respectively (Fig. 2.9a, Fig. 2.10a). Similar to $SBH^{(5)}$CTS-sgRNA, the truncated sgRNA with only a 5nt long spacer did not mediate any reporter activation irrespective of the spacer-reporter pair used (CTS1 or CTS2) (Fig. 2.9b, c, Fig. 2.10b,c). Genome-wide analyses of dCas9 binding have shown that the crRNP can bind DNA target sequences bearing as little as 5nt complementary to the 5nt PAM proximal spacer segment (Xuebing Wu et al. 2014). Accordingly, the results reported above showing that $nv^{(5)}$CTS and $SBH^{(5)}$CTS are not able to mediate CRISPRa point towards the fact that the SBH structure, and truncated spacer, might impair the sgRNA-dependent dCas9 conformational change required to convert the protein into an active state.

Furthermore, the positive trend between CRISPRa and back-fold shortening, observed when using $SBH^{(10)}$CTS and $SBH^{(15)}$CTS-sgRNAs, was not reproducible with $nv^{(10)}$CTS and $nv^{(15)}$CTS sgRNAs (Fig. 2.9b, c, Fig. 2.10b, c). Instead, for both spacer:reporter pairs, the $nv^{(10)}$CTS invariably drove higher CRISPRa than $nv^{(15)}$CTS. These results suggest that the hairpin structure was required to progressively increase reporter expression and that SBH-sgRNA with various back-fold length could potentially be used to modulate CRISPR-TR activity (see next section). Finally, the activation scores reported in figure 2.9c and 2.10c showed that native sgRNAs with truncated 10-nt spacers elicited similar or higher CRISPRa than their full-length (20-nt) counterparts. This finding was later corroborated by both *in vivo* and *in vitro* studies, showing that the binding properties of dCas9 were mainly unchanged when using sgRNAs with truncated spacers (>10-nt) (Dahlman et al. 2015; Josephs et al. 2015).

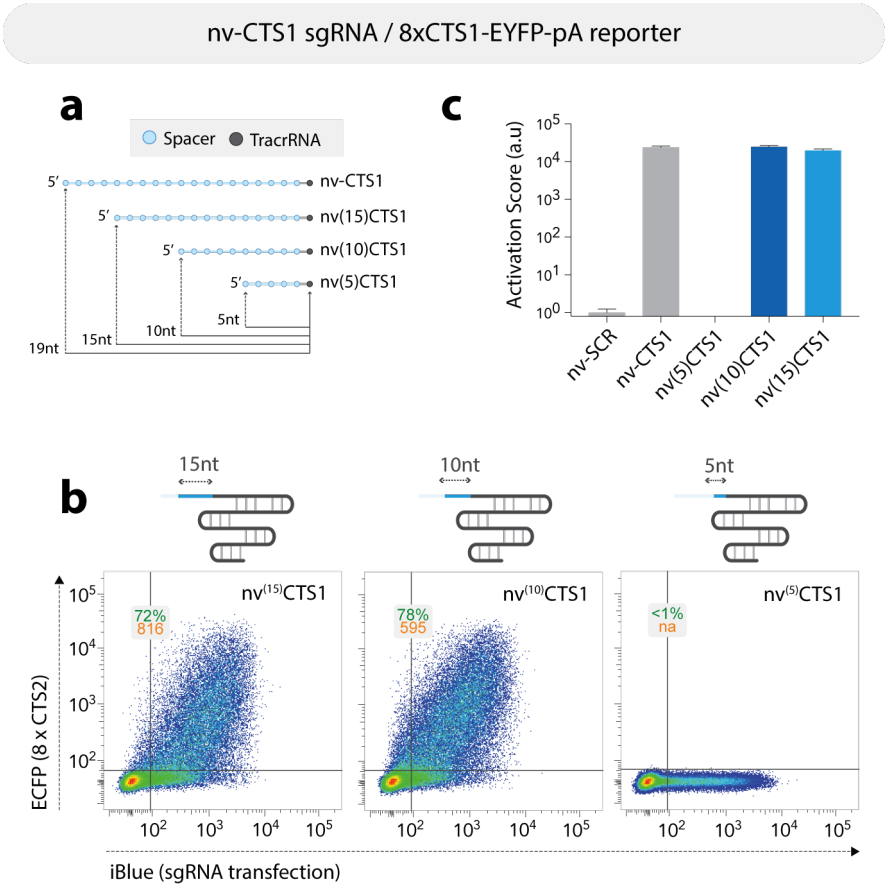

Figure 2.9 – Effect of spacer truncation on CRISPRa, CTS1. (a) Schematic of spacer truncation from the 19nt CTS1 spacer (nv-CTS1) to 15 (nv$^{(15)}$CTS1), 10 (nv$^{(10)}$CTS1), and 5 (nv$^{(5)}$CTS1) nucleotides. **(b)** HEK293-T were transfected with dCas9-VP64, reporter construct, and nv$^{(15)}$CTS1, nv$^{(10)}$CTS1, or nv$^{(5)}$CTS1 sgRNAs. Representative flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue). Inset reports % of activated cell (green) and median reporter fluorescence for the double positive cell population (iBlue^{+ve}/EYFP^{+ve}, orange). **(c)** Activation score calculation for all conditions in (b) and additional negative control (scramble sgRNA, nv-SCR) and positive control (native sgRNA, nv-CTS1) (n = 3, mean +/- s.d., normalized to the nv-SCR condition).

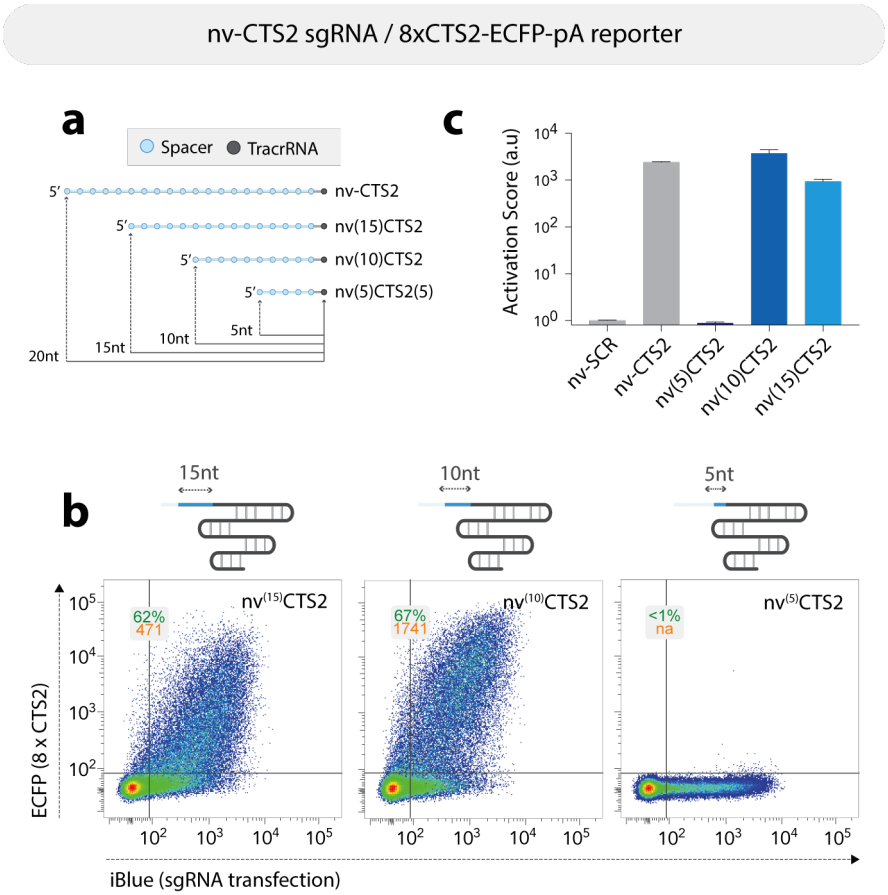

Figure 2.10 – Effect of spacer truncation on CRISPRa, CTS2. (a) Schematic of spacer truncation from the 20nt CTS2 spacer (nv-CTS2) to 15 (nv$^{(15)}$CTS2), 10 (nv$^{(10)}$CTS2), and 5 (nv$^{(5)}$CTS2) nucleotides. **(b)** HEK293-T were transfected with dCas9-VP64, reporter construct, and nv$^{(15)}$CTS2, nv$^{(10)}$CTS2, or nv$^{(5)}$CTS2 sgRNAs. Representative flow cytometry scatter plots show reporter expression (ECFP) against sgRNA transfection (iBlue). Inset reports % of activated cell (green) and median reporter fluorescence for the double positive cell population (iBlue^{+ve}/ECFP^{+ve}, orange). **(c)** Activation score calculation for all conditions in (b) and additional negative control (scramble sgRNA, nv-SCR) and positive control (native sgRNA, nv-CTS2) (n = 3, mean +/- s.d., normalized to the nv-SCR condition).

2.3.4 – SBH-based CRISPRa modulation

Prompted by previous results showing progressive silencing of CRISPRa when increasing the length of the back-fold (SBH$^{(x)}$CTS), I then asked if this effect could be used as means of modulating CRISPR-TR activity. To answer this question, a series of SBH-sgRNAs ranging from SBH$^{(6)}$CTS to SBH$^{(17)}$CTS were created for both spacer-reporter pairs (CTS1, CTS2) (Fig. 2.11a, b). Following transfection of each sgRNA variant along with matching reporter and dCas9-VP64 plasmids, the activation scores for each condition were calculated (Fig. 2.11c, d). Confirming the effect previously observed, back-fold shortening globally resulted in a progressive increase in CRISPRa (Fig. 2.11c, d): in most of the cases, SBH$^{(x)}$CTS-sgRNA drove stronger reporter expression than SBH$^{(y)}$CTS-sgRNA when x was greater than y. Amongst the few exceptions, the activation score for SBH$^{(17)}$CTS1 was found lower than the activation score for SBH$^{(15)}$CTS1, despite having two additional free spacer nucleotides. Predictions of SBH$^{(17)}$CTS1 RNA secondary structure suggested that the interactions between the PAM distal spacer nucleotide and complementary back-fold were too weak to form an hairpin (Fig. 2.11a). Consequently, the reduction in CRISPRa compared to SBH$^{(15)}$CTS1 could be due to an increase in exonuclease mediated degradation of the SBH$^{(17)}$CTS1 sgRNA caused by an unpaired back-fold protruding out of dCas9. In addition to these unexpected variations, the progressive increase in CRISPR-TR activity was not found to linearly follow the number of free spacer nucleotides. Instead, incremental steps were observed in both CTS1 and CTS2 profiles (Fig. 2.11c, d). Together, these results suggest that SBH-based modulation is likely to differ across spacer sequences.

a SBH(x)CTS1-sgRNA

NAME	Sequence + G1 tracrRNA	Predicted RNA secondary structure
SBH(17)CTS1	GCTGAAAA**GTCGCGTGTAGCGAAGCAG**	((((.....((((.....))))))))..
SBH(15G)CTS1	GACTGAAAAGTCGCGTGTAGCGAAGCAG	((((....)))).....((((.....)))).
SBH(15)CTS1	GGACTGAAAAGTCGCGTGTAGCGAAGCAG	.((((....)))).....((((.....)))).
SBH(13G)CTS1	GCGACTGAAAAGTCGCGTGTAGCGAAGCAG	((((((....)))))).((((.....)))).
SBH(13)CTS1	GGCGACTGAAAAGTCGCGTGTAGCGAAGCAG	.((((((....)))))).((((.....)))).
SBH(12)CTS1	GCGCGACTGAAAAGTCGCGTGTAGCGAAGCAG	.(((((((....)))))))....((...))..
SBH(10)CTS1	GCACGCGACTGAAAAGTCGCGTGTAGCGAAGCAG	.(((((((((....)))))))))..((...))..
SBH(6)CTS1	GCTACACGCGACTGAAAAGTCGCGTGTAGCGAAGCAG	((((((((((((....)))))))))))).....

b SBH(x)CTS2-sgRNA

NAME	Sequence + G1 tracrRNA	Predicted RNA secondary structure
SBH(17)CTS2	GTACGAAA**GTAAGTCGGAGTACTGTCCTG**	.(((....)))....(((.....))))..
SBH(15)CTS2	GCTTACGAAAGTAAGTCGGAGTACTGTCCTG	.(((((....)))))..(((.....))))..
SBH(13)CTS2	GGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	.(((((((....)))))))(((.....))))..
SBH(10)CTS2	GTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	.((((((((((....)))))))))).....
SBH(9)CTS2	GCTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	.(((((((((((....))))))))))).....
SBH(8)CTS2	GACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	.(((((((((((....))))))))))).....
SBH(7)CTS2	GCTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	..(((((((((((....))))))))))).....
SBH(6)CTS2	GGTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTG	.(((((((((((....))))))))))).....

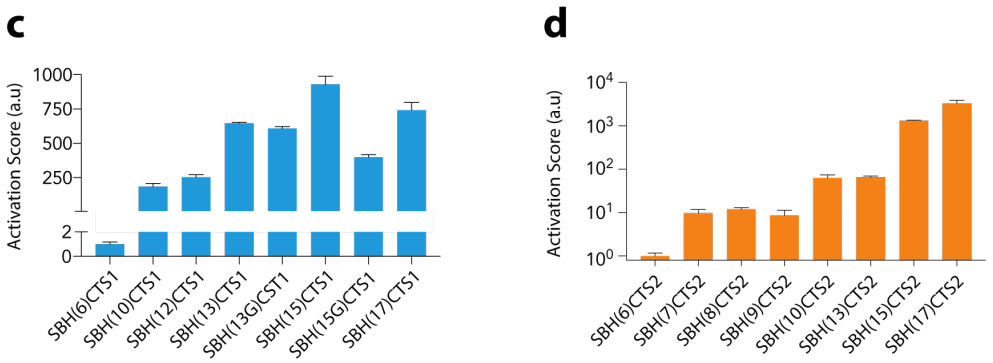

Figure 2.11 – SBH-based CRISPRa modulation. (a) Name, sequence, and predicted RNA secondary structure (dot bracket notation, "(" and ")" are used to indicate base-pairing, "." is used otherwise) of SBH$^{(x)}$CT1 hairpins (x stands for the number of free spacer nucleotides, G is added when the 5'-G required for U6 transcription base-pairs with the spacer, e.g. SBH$^{(13G)}$CTS1). Red and bold lettering indicate the predicted hairpin and the spacer segment, respectively. **(b)** Same as (a) for SBH$^{(x)}$CT2 hairpins. **(c)** HEK293-T were co-transfected with dCas9-VP64, the reporter 8xCTS1-EYFP-pA and each of the SBH-sgRNAs listed in (a). Activation scores quantifying CRISPRa for each guide are reported (n = 3, mean +/- s.d., normalized to SBH$^{(6)}$CTS1). **(d)** same as (c) for the CTS2 spacer:reporter pair.

2.4 – Closing remarks

In this chapter I have introduced the concept of spacer-blocking hairpin (SBH), a ~20bp RNA stem-loop structure, which forms co-transcriptionally as the result of base-pairing between the sgRNA spacer segment and a back-fold appended on the 5'-end of the guide. Using a CRISPRa assay, I confirmed that native sgRNAs can be efficiently silenced using this strategy to the extent where no expression of the target gene is detectable. In addition, mechanistic insights engendered from a competition assay suggested that SBH-mediated silencing is primarily due to spacer sequestration, as it was shown that SBH-sgRNA can be loaded into the CRISPR-effector.

Additional experiments will be required to establish if the crRNP loaded with a SBH-sgRNA still transiently interact with its DNA targets. On-target binding could notably be assayed using ChIP coupled with quantitative PCR (Shechner et al. 2015) and extended genome-wide via ChIP-seq (Xuebing Wu et al. 2014). Moreover, DNA curtain experiments could be employed to test whether SBH-sgRNAs loading into dCas9 is sufficient to drive the conformational change required for the CRISPR-effector to probe the genome for PAM sequences (Sternberg et al. 2014). This result may be of particular interest since previous studies have reported that ubiquitous PAM interrogation can be toxic for the host cell (Hsu et al. 2013; Pattanayak et al. 2013; Fu et al. 2014; Zuris et al. 2015).

In the second part of this chapter, I characterize the effect of progressively shortening the back-fold segment on CRISPR-TR activity. Notably, I have found a positive correlation between the number of free spacer nucleotides on the SBH-sgRNA and the extent of CRISPRa, as measured by activation score. Systematic characterization of the relationship between spacer coverage and CRISPR-TR activity across a large range of DNA targets will be required to derive a mathematical model capable to predict back-fold length versus gene target expression, based only on the identity of the spacer sequence.

Finally, a more precise characterisation of the system could be obtained by performing the CRISPR-based transcriptional activation assay in a stable cell line expressing both the reporter cassette (8xCTS-EXFP-pA) and the CRISPR effector. In its current form, the CRISPRa assay relies on the co-delivery of three distinct plasmid DNAs encoding the reporter, dCas9-VP64, and guide, respectively. Given that each cell receives a different amount of these plasmids, teasing out the cause of the variation in reporter expression observed between cells and across conditions is relatively difficult. Working with sable cell lines for which the copy number of reporter and CRISPR effector is known and constant, would not only increase the fraction of cells expressing all system components (reporter, dCas9-VP64, SBH-sgRNA), but also facilitate the interpretation of the data.

Chapter 3 – Protein-responsive guide RNAs

In the previous Chapter, I showed how native sgRNAs can be silenced by grafting on their 5'-end a spacer-blocking hairpin (SBH). Of particular interest for the design of switchable guides, I notably demonstrated that SBH-sgRNAs with full spacer coverage completely abrogated the activity of a CRISPR-based transcriptional regulator by preventing hybridization between the guide and the DNA target. From this strong OFF-state, I set out to re-engineer the hairpin to create inducible SBH variants (iSBHs), designed to conditionally release the sgRNA spacer upon sensing a specific inducer. I reasoned that such iSBH-sgRNA could be created by replacing the default "GAAA" loop of the SBH with sensing-loops encoding various conditional RNA cleaving units (Fig. 3.1). In this way, inducer-specific cleave of the iSBH sensing-loop is expected to initiate melting of the back-fold:spacer stem, thus converting the otherwise inert iSBH-sgRNA into an active guide capable of directing the CRISPR-TR to a specific target locus.

In this Chapter, I use this principle to engineer multiple iSBH-sgRNAs responsive to the CRISPR-associated endoribonuclease Csy4. First, I present proof-of-concept data showing that the RNA sequence targeted by Csy4 can be used as a sensing-loop to create highly specific iSBH-sgRNAs. I then employ a rational design process to further optimize the original design and improve the ON/OFF characteristics of this inducible CRISPR-TR system. Finally, the chapter ends on the demonstration that the ensuing design principles can be applied

across spacer sequences to engineer new Csy4-responsive iSBH-sgRNAs without the need for further optimization.

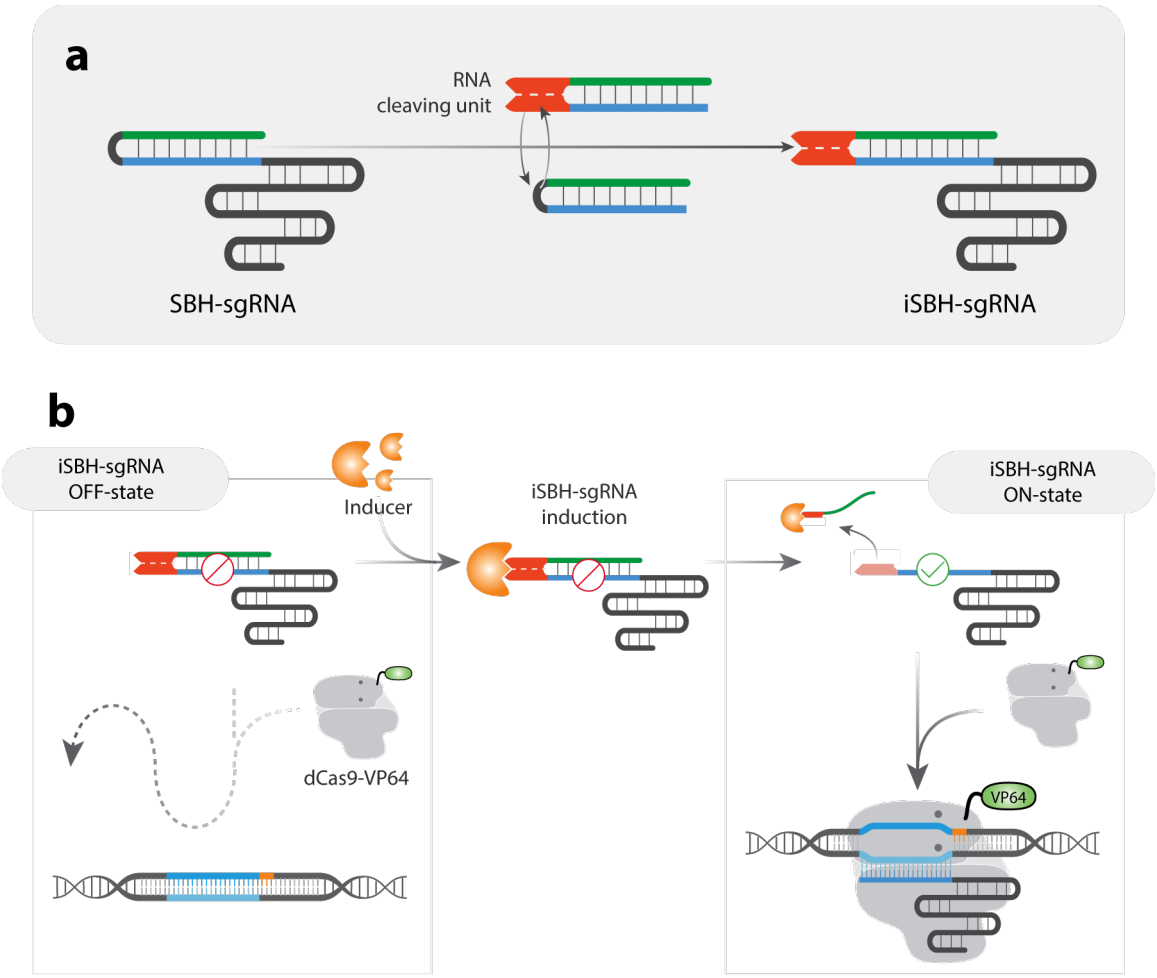

Figure 3.1 – Design of inducible SBH-sgRNAs. (a) An inducible space-blocking hairpin (iSBH) is created by replacing the SBH loop with a sensing-loop, which cleavage is conditioned on the presence of a particular inducer. **(b)** In the absence of the trigger (OFF-state), the resulting iSBH-sgRNA, and consequently CRISPR-TR are maintained silent. Delivery of the inducer (ON-state) triggers cleavage of the sensing-loop, melting of the back-fold:spacer stem, and release of an active sgRNA which direct the CRISPR-effector on-target.

3.1 – Introduction to the CRISPR-associated protein Csy4

All CRISPR systems transcribe their CRISPR array into a pre-crRNA transcript, which is then processed to yield mature crRNAs (see section 1.4). Both type I and III CRISPR systems rely on a single protein from the CRISPR associated (Cas) 6 protein family to perform this task (van der Oost et al. 2014; Hochstrasser & Doudna 2015). In the type-I F system from *Pseudomonas aeruginosa*, the Cas protein Csy4, also known as Cas6F, is responsible for splicing the pre-crRNA into mature crRNA (Haurwitz et al. 2010). This endoribonuclease recognizes, binds, and cleaves the 28nt RNA repeat interspacing the 32nt spacers of the pre-crRNA transcript (Grissa et al. 2007) (Fig. 3.2a). This RNA repeat, later referred to as Csy4-protein target site (PTS), forms a 15nt stem loop structure flanked by five upstream and eight downstream nucleotides.

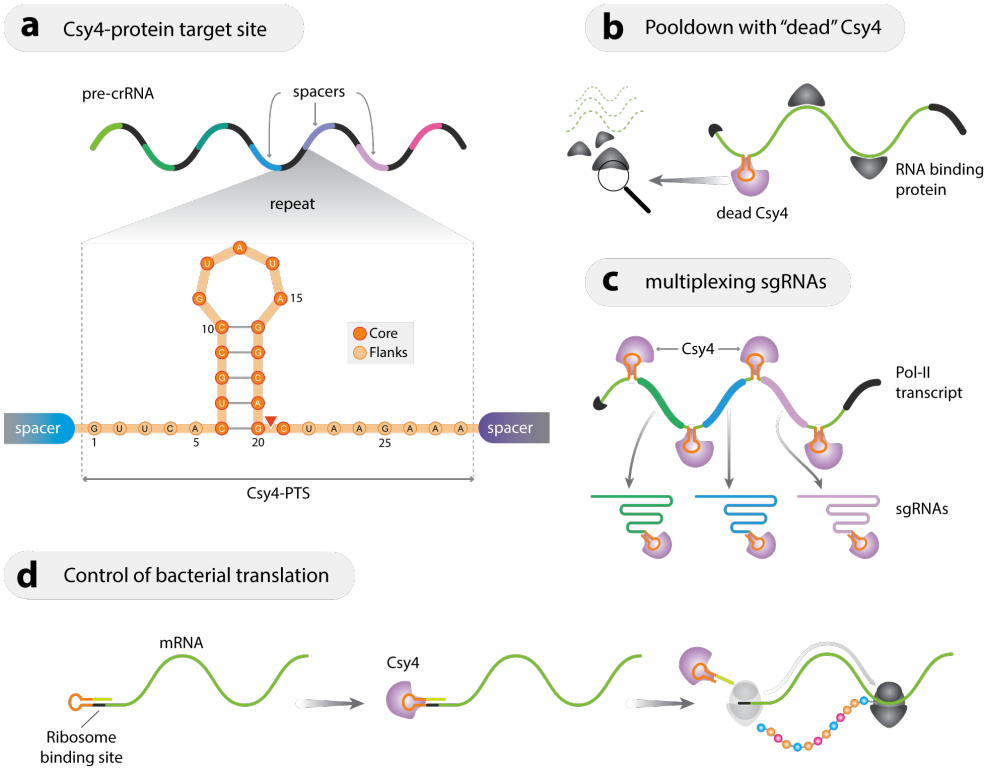

Figure 3.2 – Csy4 substrate and applications in synthetic biology. (a) *Pseudomonas aeruginosa* Csy4 protein target site (PTS). Sequence and RNA secondary structure of the Csy4-PTS is given with annotation of the core nucleotides required for cleavage. Orange arrowhead indicates the position of the cleavage site. **(b-d)** Applications based on catalytically active or inactive Csy4, see main text for details.

In vitro experiments with purified Csy4 proteins showed that the cognate PTS could be reduced to a 16nt core sequence (C6 to C21) without impairing Csy4-mediated cleavage (Haurwitz et al. 2010). The crystal structures of Csy4 complexed with this core stem revealed extensive interactions between the protein and the RNA. It was notably found that Csy4 contacts with the major groove and with the single-stranded RNA (ssRNA) to double-stranded RNA (dsRNA) junction at the base of the stem (C6-G20), and that these interactions are critical for Csy4 catalytic activity (Haurwitz et al. 2010). Although not necessary for cleavage, the work of Sternberg and colleagues also demonstrated that nucleotide A5 of the Csy4-PTS, as well as the GUAUA loop sequence, enhance Csy4 binding to the stem (Sternberg et al. 2012). Finally, *in vitro* electrophoretic mobility shift assay (EMSA) of variants containing structural and sequence alterations of the Csy4-PTS have shown that Csy4 binding and cleavage are extremely specific (Sternberg et al. 2012), making this endonuclease an attractive candidate for the design of iSBH-sgRNAs.

Since the publication of these structural insights, several groups have taken Csy4 out of its endogenous context and proposed applications for the endoribonuclease. Starting in 2013 with the work of Lee and colleagues, a catalytically inactive version of Csy4 was employed to specifically pull down mRNA species and profile RNA-binding proteins associated with these transcripts (H. Y. Lee et al. 2013) (Fig. 3.2b). Leveraging the exceptionally strong interaction between Csy4 and its cognate PTS[13] (Sternberg et al. 2012), this feat was made possible by modifying the mRNA to incorporate the protein biding site. The catalytically active Csy4 was later adapted to work in mammalian cells and was used to multiplex *Sp*Cas9 sgRNAs from a single Pol-II generated transcript (Nissim et al. 2014) (Fig. 3.2c). Very much like in the *Pseudomonas aeruginosa* pre-crRNA, sgRNAs were flanked on both site by Csy4-PTSs so as to be excised from the transcript in the presence of the endoribonuclease. This idea that Csy4 can be used physically separate genetic elements at the transcript level was also generalized by Qi et al. in an attempt to facilitate the modular programming of predictable genetic systems

[13] Dissociation constant equal to 50pM.

by alleviating unwanted contextual interferences (Qi et al. 2012). Finally, while the work that I present here was reaching completion, Du et al. have employed the endoribonuclease to achieve conditional translational control in *E. coli* (Du et al. 2015). In this instance, synthetic mRNAs were created whereby the ribosome binding site, required for the initiation of translation in bacteria, was capped with a stem which could then be cleaved off by Csy4 (Fig. 3.2d).

Similarly to Du's strategy (Du et al. 2015), I had envisioned that Csy4-responsive iSBH-sgRNAs could be engineered by using the Csy4-PTS as a sensing-loop in the iSBH designs (Fig. 3.3a). In addition to its specificity, the choice of using Csy4 as a spacer release mechanism was also motivated by studies showing that the protein works as a single turn over enzyme which, post-cleavage, remains bound to the 5'-end of the PTS product (Haurwitz et al. 2010; Sternberg et al. 2012). I anticipated that this feature would facilitate strand separation between back-fold and spacer thus improving production of active sgRNAs.

3.2 – Csy4-iSBH proof of concept

To test the feasibility of this concept, the default GAAA loop of SBH$^{(0)}$CTS1 (see section 2.3.1) was replaced with the full 28nt Csy4-PTS to create the Csy4-responsive hairpin iSBH$^{(0)}$Csy4$^{(full)}$CTS1 (Fig. 3.3b). No reporter expression was observed when the resulting iSBH$^{(0)}$Csy4$^{(full)}$CTS1-sgRNA was transfected along with dCas9-VP64 and the reporter plasmid (Fig. 3.3c). This finding suggested that the sequence and structure of the SBH loop can be altered without loss of the hairpin's repressive properties. To determine if this iSBH$^{(0)}$Csy4$^{(full)}$CTS1-sgRNA could be conditionally activated, the experiment was repeated while co-transfecting a pDNA constitutively expressing an NLS-tagged, human codon optimised Csy4 (Nissim et al. 2014). Flow cytometry analysis, conducted 48h post-transfection, revealed a weak but discernable activation of the reporter gene, suggesting that the guide had been activated by Csy4 expression, and successfully guided the crRNP to the promoter (Fig. 3.3c). To confirm that the observed reporter fluorescence was in fact due to iSBH-sgRNA

activation rather than a side effect of Csy4 expression, the iSBH-sgRNA was replaced in the experiment with its non-responsive SBH counterpart SBH$^{(0)}$CTS1-sgRNA, lacking the Csy4 sensing-loop (Fig. 3.3c). As expected, the CRISPR-TR did not drive reporter expression, despite expression of the protein inducer. Together these results demonstrated the feasibility of using the propose iSBH framework to create inducible transcriptional regulators.

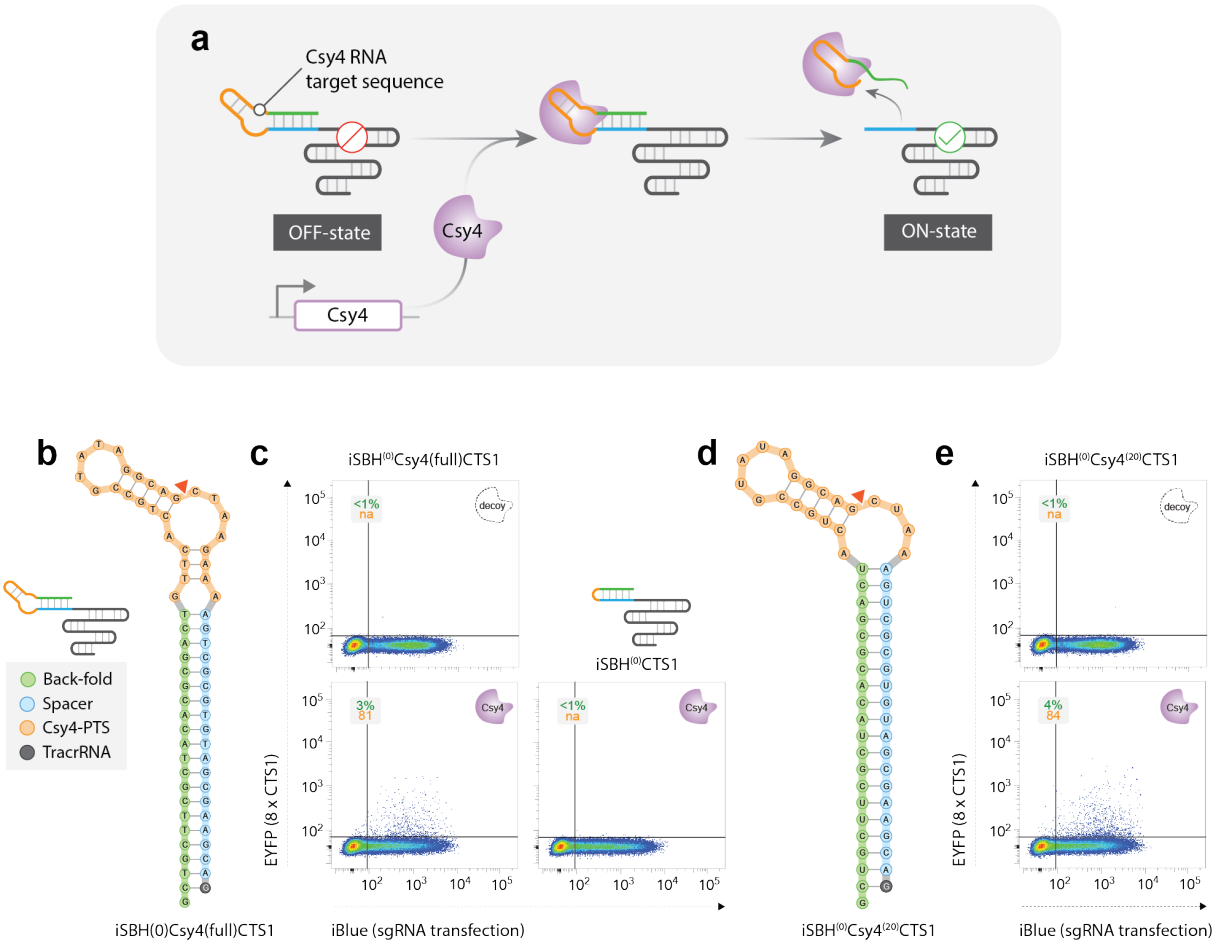

Figure 3.3 – Csy4-responsive iSBH-sgRNA: proof of concept. (a) A Csy4-responsive iSBH-sgRNA is created by replacing the SBH loop with a Csy4-proteing target site (PTS). The guide is expected to stay silent (OFF-state) until Csy4 is expressed. Upon expression, Csy4 binds, cleaves and remains bound to the 5' product of the substrate, leading to the release of an active sgRNA. **(b)** Sequence and RNA secondary structure of iSBH$^{(0)}$Csy4$^{(full)}$CTS1, orange arrowhead indicates the Csy4 cleavage site. **(c)** HEK293-T cells were transfected with dCas9-VP64, reporter construct, a Csy4-responsive iSBH-sgRNA (left) or SBH-sgRNA lacking the Csy4-PTS (right), and Csy4 expressing plasmid (bottom) or decoy (top). **(d)** Sequence and RNA secondary structure of iSBH$^{(0)}$Csy4$^{(20)}$CTS1. **(e)** The resulting iSBH-sgRNA (d) is tested in the presence or absence of Csy4.

I have previously shown that the presence of a floating tail on the 5'-end of the sgRNA could reduce CRISPR-TR efficiency (see SBH$^{(ctr-3)}$CTS1 in Fig. 2.5). Similarly, the iSBH$^{(0)}$Csy4$^{(full)}$CTS1-sgRNA tested above, is expected to have, post Csy4-cleave, a 8nt residual sequence on its 5'-end that might impact its ability to mediate CRISPRa (Fig. 3.3b). Also, I reasoned that the ON-state performances of the system could be improved by modifying the Csy4-PTS to reduce the length of the residual sequence left on the activated guide. Accordingly, a second Csy4-responsive iSBH was created by shortening the 28nt Csy4-PTS to a minimal 20-nt sequence (Fig. 3.3d). This reduced sensing-loop featured the 16nt core sequence as well as the nucleotide A5, which has been proved critical for Csy4 binding. Additionally, I made sure that a sufficient number of flanking nucleotides were conserved to form a bulge at the base of the Csy4-PTS stem (ssRNA:dsRNA junction is critical for Csy4 cleavage). Although the resulting iSBH$^{(0)}$Csy4$^{(20)}$CTS1-sgRNA was fully inert in the OFF-state and displayed Csy4-specific activation, it only provided a marginal 1.6 fold increase in ON-state activation score compared to the previous design (Fig. 3.3e).

Next, I hypothesized that the modest ON-state performances might be imputable to reduced enzymatic property of Csy4 in mammalian cells. To verify that endoribonuclease was indeed capable of binding and cleaving its substrate, the Csy4-PTS was cloned in the 5'-UTR of an enhanced green fluorescent protein (EGFP) reporter construct (Fig. 3.4a). Flow cytometry analysis of cells expressing the resulting plasmid showed a striking reduction in EGFP fluorescence when co-transfected with Csy4 versus a decoy DNA plasmid (Fig. 3.4a). Since previous studies have shown that RNA binding proteins targeted to the 5'-UTR of a genes can efficiently block translation and reduce protein output in human cells (Ausländer et al. 2012; Hirosawa et al. 2017; Wroblewska et al. 2015), a qPCR analysis with primers spanning the Csy4-PTS was also conducted to confirm Csy4-dependent slicing of the mRNA (Fig. 3.4b). Analysis of the data revealed that the mRNA transcript was highly depleted in the Csy4 condition compared to the negative control (>99% reduction) (Fig. 3.4b). Taken together, these

results confirmed the catalytic properties of Csy4 endonuclease, and ruled out reduced catalytic activity as a potential explanation to the low ON-state observed.

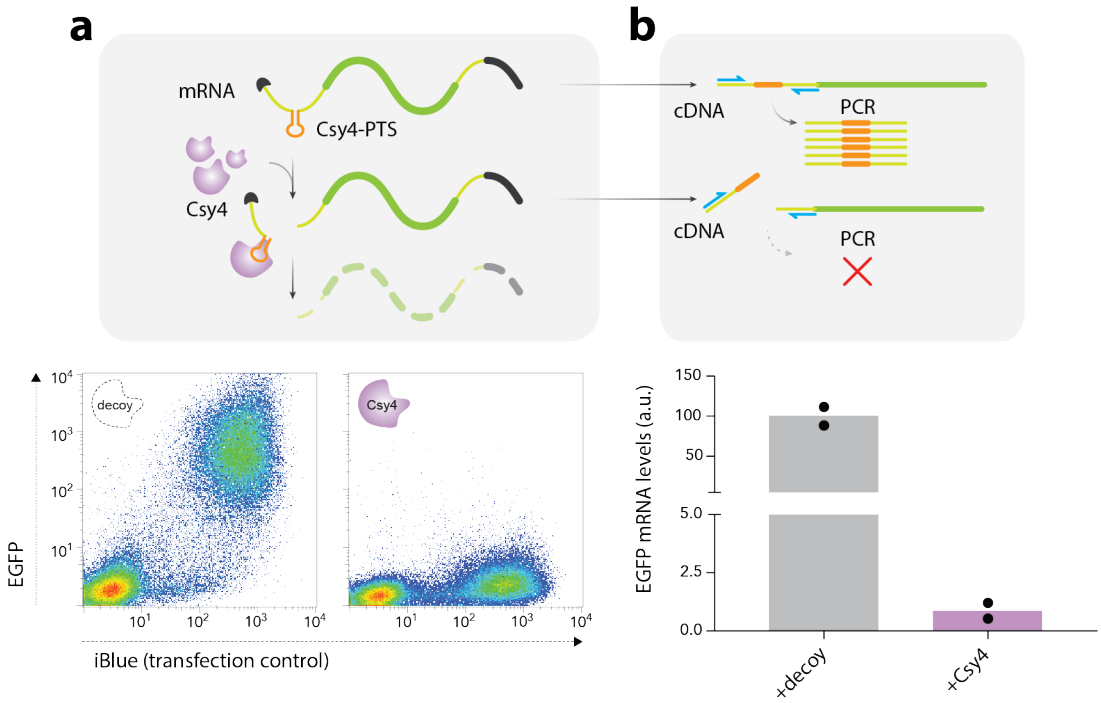

Figure 3.4 – Validation of Csy4-mediated RNA slicing. (a) HEK293-T cells were transfected with a EGFP construct modified to incorporate a Csy4-PTS in its 5'-UTR (top). Co-expression of Csy4 is expected to lead to EGFP mRNA degradation. (bottom) Representative flow cytometry scatter plots of cells transfected with the EGFP construct and either a Csy4 or decoy plasmid. **(b)** RT-qPCR analysis with a primer pair spanning the Csy4-PTS is performed on cells transfected with the EGFP construct and either a Csy4 or decoy plasmid (n = 2, bars show mean).

3.3 – Bulged SBH designs

After showing that Csy4 was indeed catalytically active, I speculated that the stability of the SBH might be a key determinant for ON-state activation. Indeed, the interaction between the back-fold and the spacer segments could have been too strong for the stem to melt, despite cleavage of the sensing-loop. Therefore, I sought for a way to lower the hairpin thermodynamic stability while maintaining the degree of silencing observed with full spacer coverage. Bulged stems have been previously used to reduce the melting temperature of RNA structures (G. F.

S. Lim et al. 2012). The solution of inserting bulges in the SBH to lower its stability was particularly fitting in my case, as it allowed to maintain base-pairing between the back-fold and the seed region of the spacer. Additionally, I reasoned that interspaced bulges in the iSBH would favor, post-cleavage of the sensing-loop, a stepwise strand separation process, requiring only small energy steps (Fig. 3.5a).

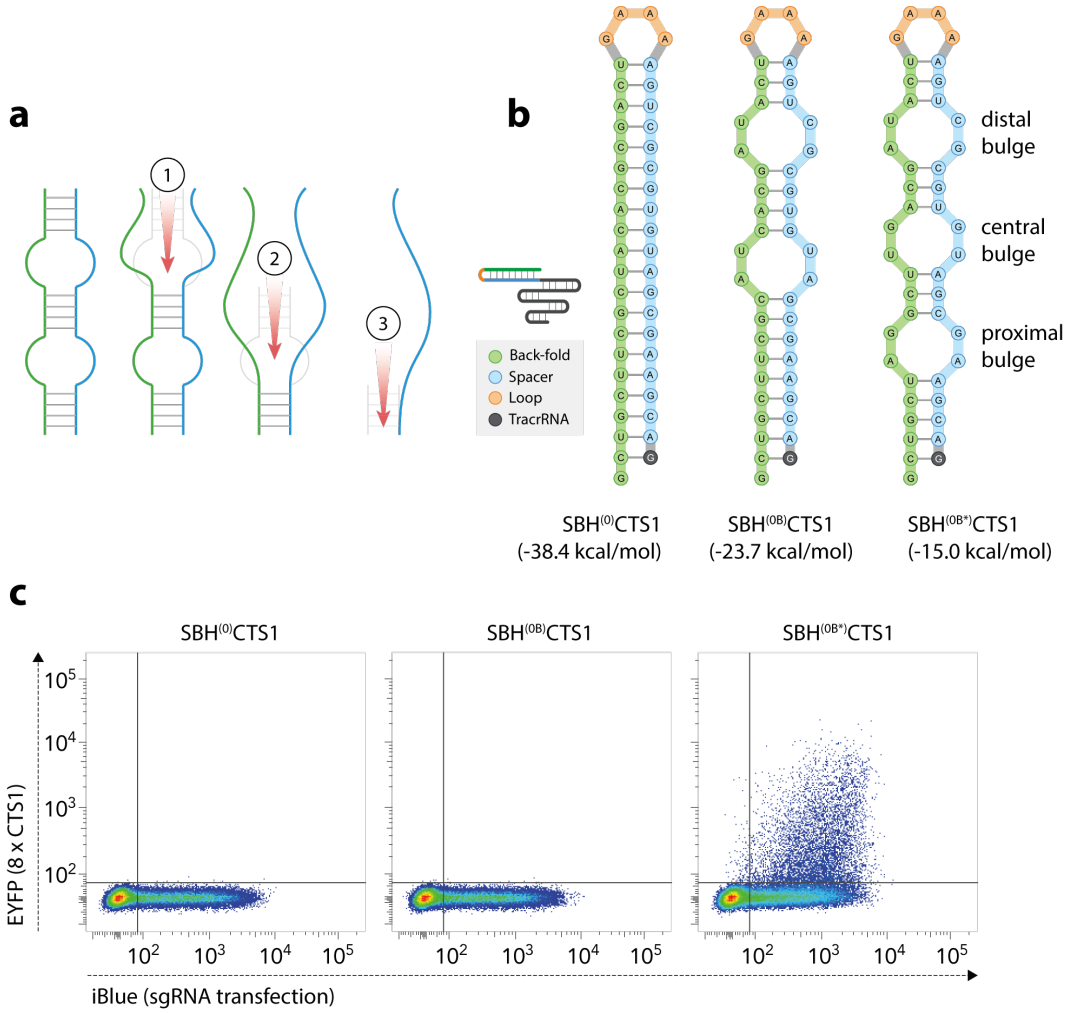

Figure 3.5 – Bulged SBHs and their silencing properties. (a) Schematic representation of the stepwise strand separation process (series of low-energy steps) expected from a bulged stem. **(b)** Sequence and RNA secondary structures of the original SBH$^{(0)}$CTS1 hairpin (full spacer coverage) and two variants modified to accommodate two (SBH$^{(0B)}$CTS1) and three (SBH$^{(0B*)}$CTS1) bulges, respectively. Bulges are created by mutating the back-fold sequence. Distal, central, and proximal bulges are named using the spacer seed sequence as reference. **(c)** HEK293-T cells were transfected with dCas9-VP64, reporter plasmid and each of the SBH-sgRNAs presented in (b). Flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue).

Instead of directly incorporating bulges in the Csy4-responsive iSBH design, I thought to optimize the number and positioning of the bulges on a non-responsive SBH-sgRNA, to verify that the alterations did not affect the hairpin ability to silence the guide. In the design of SBH, bulges can easily be incorporated in the stem by mutating the back-fold to create mismatches with the spacer sequence. Accordingly, the sequence of SBH$^{(0)}$CTS1 was modified to introduce three 2nt bulges equally interspaced along the length of the stem (SBH$^{(0B*)}$CTS1, Fig. 3.5b). Using the seed of the spacer as a reference, bulges were referred to as seed distal, central, and proximal, respectively. When testing this new SBH-sgRNA in the CRISPRa assay, strong reporter activation was observed, suggesting that either dCas9 was able to displace the back-fold upon loading the SBH-sgRNA, or the hairpin structure was not stable enough to outcompete hybridization between the spacer and the DNA target (Fig. 3.5c). Therefore, I reasoned that this leakage could be solved by removing the seed proximal bulge in order to increase the back-fold:spacer paring stability over the seed region (Fig. 3.5b, SBH$^{(0B)}$CTS1). Indeed, transfection of the corresponding SBH-sgRNA along with dCas9-VP64 and reporter constructs showed no sign of CRISPRa (Fig. 3.5c). Importantly however, this design increased the minimal free energy of the hairpin from -38.4 kcal/mol to -23.7 kcal/mol, making it a good candidate for the design of inducible hairpins.

In order to assess how this bulged stem would performed in the design of inducible sgRNAs, a new guide was created by grafting the full Csy4-PTS onto the SBH$^{(0B)}$CTS1 hairpin (Fig. 3.6a, iSBH$^{(0B)}$Csy4$^{(full)}$CTS1). The resulting iSBH$^{(0B)}$Csy4$^{(full)}$CTS1-sgRNA was transfected in HEK293-T along with dCas9-VP64 and the reporter pDNA and assayed for its ability to drive CRISPRa in the presence or absence of Csy4. Similar to its non-responsive counterpart, this guide was found fully inert in the absence of the protein trigger (Fig. 3.6b). More importantly, a drastic improvement in CRISPRa was observed compared to previous designs when Csy4 was co-transfected (Fig. 3.6b). The number of activated cells in the ON-state was notably increased from the 4%, observed with iSBH$^{(0)}$Csy4$^{(20)}$CTS1 (Fig. 3.3e), to 27%. To confirm that the activation of this iSBH-sgRNA was in fact Csy4-specific, a mutant hairpin was generated

by simply inversing the basal C6-G20 base-pair (becomes G6-C20) of the Csy4-PTS (Fig. 3.6c, iSBH$^{(0B)}$Csy4m$^{(full)}$CTS1). This sequence change was chosen for its ability to drastically impair Csy4 cleavage activity while having only minimal effect on the endoribonuclease binding properties (Sternberg et al. 2012; Haurwitz et al. 2010). No reporter expression was detected when testing the resulting SBH$^{(0B)}$Csy4m$^{(full)}$CTS1-sgRNA in the CRISPRa assay, including when Csy4 was co-expressed (Fig. 3.6d). This decoupling between the presence of the protein inducer and the activation of the target gene demonstrated that Csy4-mediated cleavage is required to activate the iSBH-sgRNA.

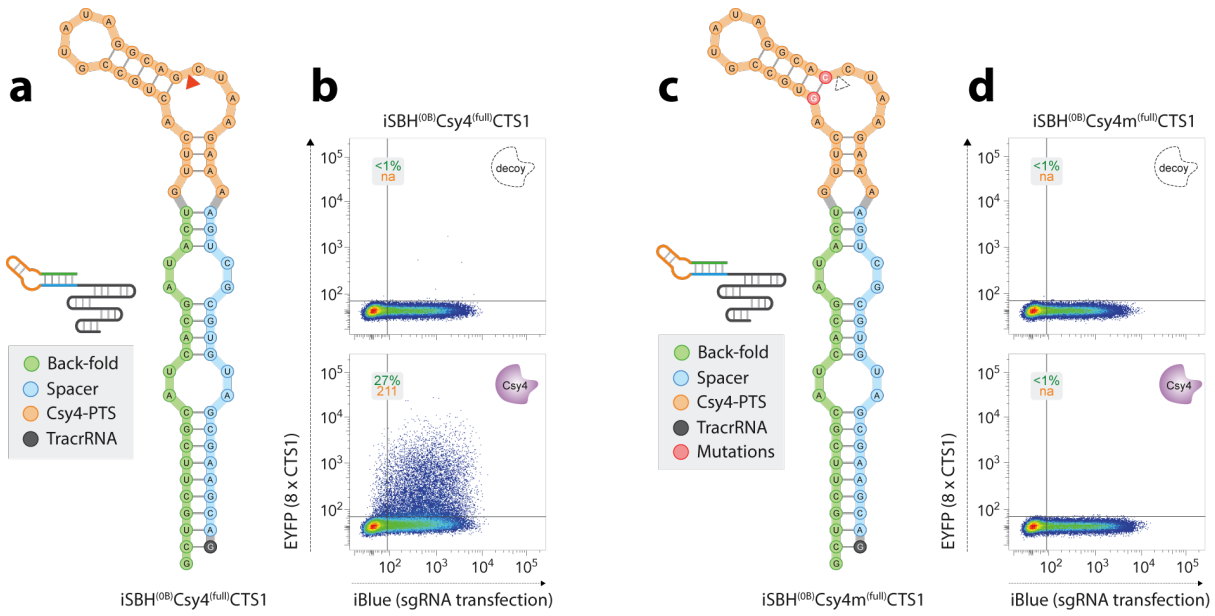

Figure 3.6 – Csy4-responsive iSBH-sgRNAs with bulged stem. (a) The full 28nt Csy4-protein target site (PTS) is grafted onto the bulged hairpin SBH$^{(0B)}$CTS1 to create iSBH$^{(0B)}$Csy4$^{(full)}$CTS1. **(b)** HEK293-T cells were transfected with dCas9-VP64, reporter, iSBH$^{(0B)}$Csy4$^{(full)}$CTS1-sgRNA, and either decoy plasmid (top) or Csy4 expression vector (bottom). Representative flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue). % of activated cells (green) and median reporter fluorescence for this population (orange) are shown in the inset. **(c)** The Csy4 sensing-loop of iSBH$^{(0B)}$Csy4$^{(full)}$CTS1 is modified to abrogate Csy4-mediated cleavage (base-pair inversion, red circles). **(d)** Repeat of (b) with the mutant iSBH$^{(0B)}$Csy4m$^{(full)}$CTS1-sgRNA (c).

3.4 – Second generation Csy4-responsive hairpins

The results presented so far suggested that both reducing the 5' residual post-cleavage as well as decreasing the spacer-blocking hairpin stability improved the ON-state system performance. Based on previous observations that sgRNAs with shortened spacers can still drive potent CRISPRa (see section 2.3.3), I wondered if the system could be further improved by shortening the 5'-end of the spacer and shifting down the sensing-loop.

With this idea in mind, I generated two new Csy4-reponsive iSBHs by fusing the 16nt core Csy4-PTS with either the seed distal (iSBH$^{(0B)}$Csy4$^{(medium)}$CTS1) or the seed central (iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1) bulge of SBH$^{(0B)}$CTS1 (Fig. 3.7a). These designs, referred to as *medium* and *nano*, increased the minimal free energy of the hairpin from -26.6 kcal.mol^{-1} (full-length stem) to -21.80 kcal.mol^{-1} (*medium*) and -16.6 kcal.mol^{-1} (*nano*), respectively[14]. Accordingly, a large increase in the ON-state performance was observed, both with regard to the fraction of activated cells, and the strength of reporter activation (Fig. 3.7b). Notably, the iSBH-sgRNA implementing the *nano* hairpin was the most potent with 54% of activated cells in the ON-state and a 4.8 fold-improvement in activation score compared to previous full length design (iSBH$^{(0B)}$Csy4$^{(full)}$CTS1) (Fig. 3.7b, c). Of importance, this increase in ON-state performance was achieved without compromising the OFF-state, as both iSBH-sgRNAs were found fully inert in the decoy plasmid condition (Fig. 3.7b, c).

[14] The free energy was calculated for the stem post-cleavage. Energy associated with the Csy4-PTS is not taken into consideration.

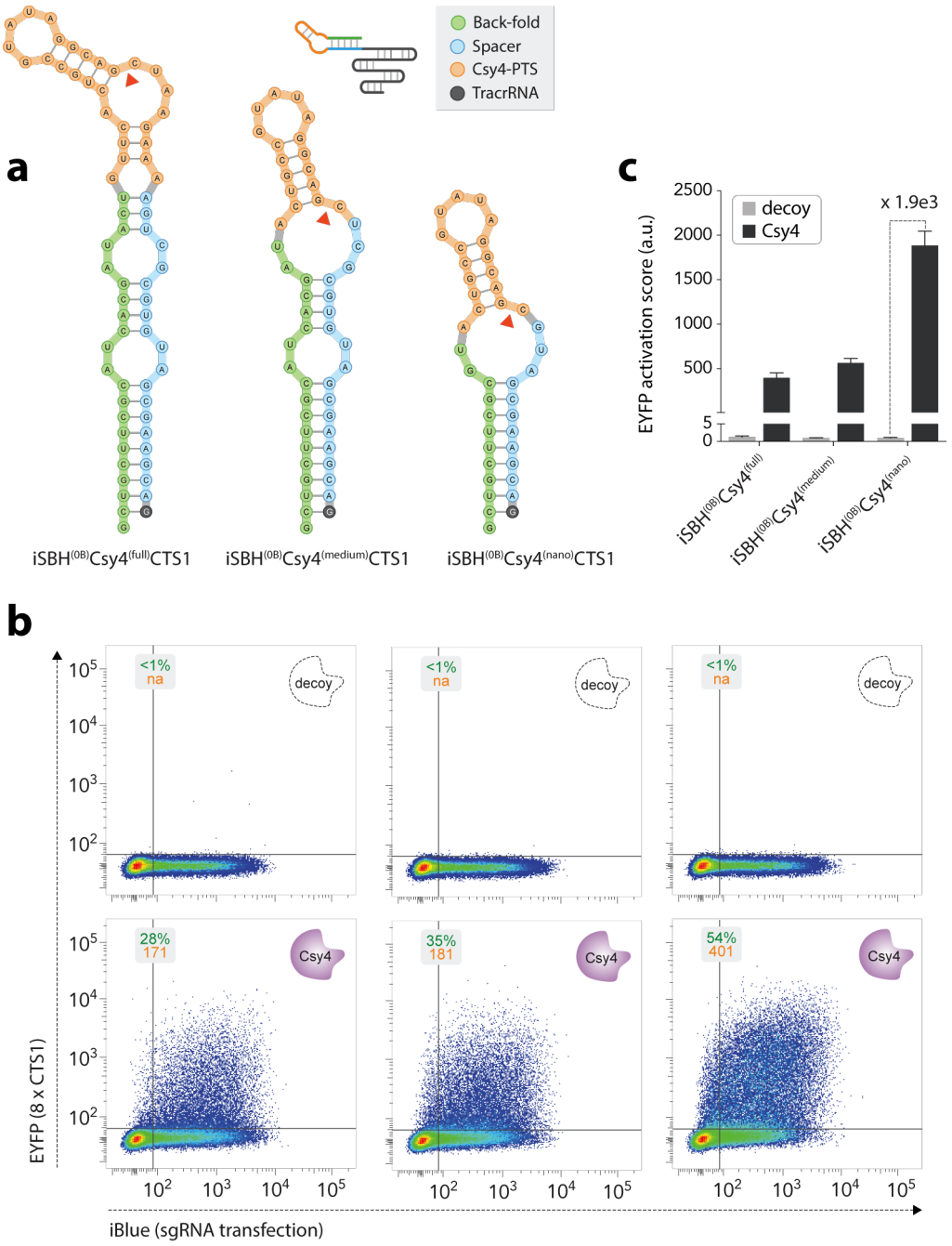

Figure 3.7 – Second generation Csy4-repsonsive iSBH-sgRNAs, CTS1 spacer. (a) Sequence and RNA secondary structure of Csy4-responsive hairpins featuring a CTS1 spacer. The Csy4-protein target site (PTS) is grafted onto the full SBH$^{(0B)}$CTS1 or fused with the seed distal, or proximal bulge of the stem to produce *medium* (iSBH$^{(0B)}$Csy4$^{(medium)}$CTS1) and *nano* (iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1) designs, respectively. **(b)** HEK293-T were co-transfected with dCas9-VP64, reporter construct, one of the iSBH-sgRNAs shown in (a), and a decoy or Csy4 expression plasmid. Flow cytometry dot plots show reporter expression (EYFP) against sgRNA transfection (iBlue). Insets display the percentage of activated cells (green) and median reporter fluorescence for this subpopulation (orange). **(c)** Activation score calculations for all conditions in (b) (n = 3, mean +/- s.d.). Data were normalized to a negative control condition employing a scramble sgRNA (data not shown).

I next sought to test if the hierarchy in ON/OFF-performances that was found between *full*, *medium*, and *nano* iSBH designs, would hold true when the sequence of the spacer was changed. I reasoned that if this was the case, the sequence of Csy4-responsive iSBH-sgRNAs could be automatically generated from new spacers by following a set of well-defined design principles. Correspondingly, I applied the design rationales derived above to create *medium* ($iSBH^{(0B)}Csy4^{(medium)}$CTS2) and *nano* ($iSBH^{(0B)}Csy4^{(nano)}$CTS2) Csy4-iSBHs for the CTS2 spacer (Fig. 3.8a). Consistent with the results obtained with CTS1 targeting guides, these iSBH-sgRNAs were fully silent in the OFF-state and mediated strong CRISPR-TR activity in the ON-state (Fig. 3.8b, c). Additionally, the shorter *nano* design was again found to outperformed its *medium* counterpart, with ON-state values approaching those of the native sgRNA (Fig. 3.8b, c).

3.5 – Closing remarks

In this chapter, I have reported on the construction of Csy4-responsive iSBH-sgRNAs. Demonstrating the modularity of the iSBH design, I have shown that inducible sgRNAs can simply be engineered by replacing the default GAAA loop of the hairpin with the Csy4-PTS. Results obtained using different iSBH-sgRNAs and a mutant hairpin, suggested that Csy4-mediated cleavage of the iSBH sensing-loop was required to activate the otherwise inert sgRNAs. Additionally, given that Csy4 is expected to stay bound to the 5' product of its substrate, it is likely that the endoribonuclease also facilitated melting of the back-fold:spacer hybrid post-cleavage. Together, these mechanisms led to the Csy4-dependent activation of silent iSBH-sgRNAs and downstream CRISPR-mediated transcriptional activation of target genes.

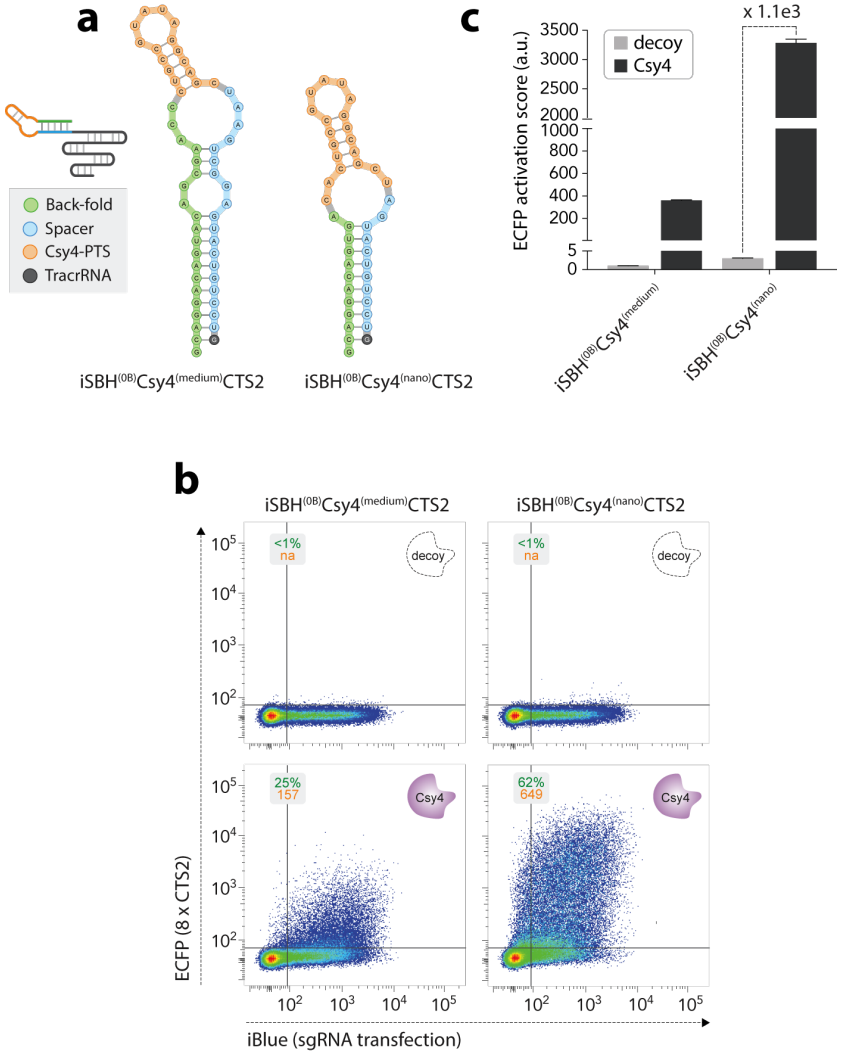

Figure 3.8 – Second generation Csy4-responsive iSBH-sgRNAs, CTS2 spacer. (a) Sequence and RNA secondary structure of Csy4-responsive hairpins featuring a CTS2 spacer. The Csy4-protein target site (PTS) is fused with the seed distal, or proximal bulge of SBH$^{(0B)}$CTS2 to produce *medium* (iSBH$^{(0B)}$Csy4$^{(medium)}$CTS2) and *nano* designs (iSBH$^{(0B)}$Csy4$^{(nano)}$CTS2). **(b)** HEK293-T were co-transfected with dCas9-VP64, reporter construct, one of the iSBH-sgRNAs shown in (a), and a decoy or Csy4 expression plasmid. Flow cytometry dot plots show reporter expression (ECFP) against sgRNA transfection (iBlue). Insets display the percentage of activated cells (green) and median reporter fluorescence for this subpopulation (orange). **(c)** Activation score calculations for all conditions in (b) (n = 3, mean +/- s.d.). Data were normalized to a scramble sgRNA control (data not shown).

In the second half, I applied rational design to incrementally increase reporter activation in the presence of the inducer, while maintaining full CRISPR-TR silencing in the OFF-state. Throughout this process, I notably found that the stability of the hairpin inversely correlated with the potency of the activation. Accordingly, the ON/OFF performances of the system were improved by incorporating 2nt bulges in the iSBH stem and further shortening the hairpin. Of interest for the bulk creation of iSBH-sgRNAs, I concluded the chapter with evidences suggesting that these design principles apply across spacer sequences, making it, in theory, possible to automate the design of Csy4-responsive iSBHs (see section 6.2).

As a general note for the reader interested in implementing inducible CRISPR-TR using Csy4 as an inducer, I provide below a series of guidelines and design principles that one can follow to create an Csy4-iSBH for any spacer sequence of choice: (i) Choose a gene target and determine the optimal CRISPR-target site (CTS, N20-NGG) for transcriptional activation. (ii) Clone the guide RNA spacer corresponding to the chosen CTS in a plasmid expressing the sgRNA under a U6 promoter (see methods). Test the efficiency of this guide by transfecting cells of interest with both sgRNA and dCas9-VP64 pDNAs, and assaying changes in gene expression 48 hours post transfection using reverse transcription quantitative PCR (see methods). Note: I recommend replacing dCas9-VP64 with a more potent second generation CRISPR-based transcriptional activators (see section 1.5.2). I notably show in section 6.1 that Csy4-responsive iSBH-sgRNAs are compatible with the synergistic activation mediator system (SAM) (Konermann et. al 2014). (iii) Once the sgRNA has been tested and has been found to efficiently increase the transcriptional output of the target gene, the corresponding Csy4-responsive iSBH-sgRNA can be designed using the iSBHfold web tool introduced in chapter 6 (see section 6.2). Note that this program creates by default a *nano* iSBH following well-defined designed rules presented in the figure 3.9 below. This choice is based on results presented in this chapter showing that the *nano* design consistently produced stronger ON-state activation compared to the *full* and *medium* hairpins (fig 3.7c and 3.8c). Nevertheless, I would like to emphasise that this trend was only demonstrated for two spacer sequences. Consequently,

one cannot exclude the possibility that for certain spacers not yet tested, the *full* or *medium* designs might display better ON/OFF properties than their *nano* counterpart. Design rules used to create these different stems are illustrated in figure 3.9 below.

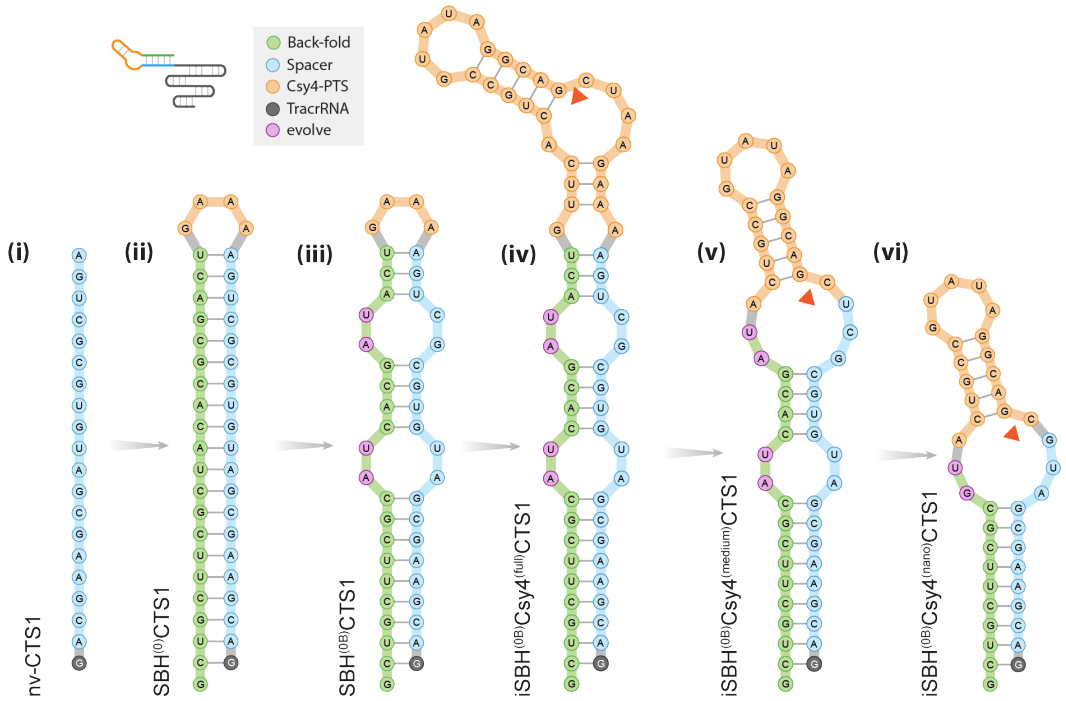

Figure 3.9 – Design principles used to create Csy4-responsive iSBHs. Starting from a given 19-20nt spacer sequence (i, here 19nt spacer targeting CTS1 is taken as example, nv-CTS1), the reverse complement is taken to obtain the sequence of the matching full length back-fold. The back-fold sequence is then fused to the spacer using a default GAAA loop to create SBH$^{(0)}$CTS1 (ii). Note that the SBH sequence starts with a 5' guanine required for U6 transcription. The stability of the full stem is then reduced by mutating the back-fold sequence to form two 2bp bulges over spacer nucleotides 4-5 and 10-11, respectively (iii, SBH$^{(0B)}$CTS1, spacer nucleotides are numbered from 1 to 19 from loop to tracrRNA). The *full* Csy4-iSBH is created by replacing the GAAA with the 28nt Csy4-PTS (iv, iSBH$^{(0B)}$Csy4$^{(full)}$CTS1). *Medium* (v, iSBH$^{(0B)}$Csy4$^{(medium)}$CTS1) and *nano* (vi, iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1) designs are obtained by fusing the minimal Csy4-PTS (orange) to the distal or central bulge of the stem. The newly formed bulge should be such that 4 or 5 consecutive nucleotides would assume an open conformation on the spacer side to improve Csy4 processing of the hairpin. This is achieved by evolving the sequence of the facing back-fold nucleotides (purple).

Chapter 4 – ASO-responsive guide RNAs

In the previous chapter, I showed that the iSBH methodology can be used to created inducible sgRNAs and conditionally activate CRISPR-TR based on the presence (ON-state) or absence (OFF-state) of a protein trigger. The ability to genetically encode both the inducer and the CRISPR-TR components (iSBH-sgRNA, CRISPR-effector) provide a means to pre-program cells with "IF/THEN"-like statements, which link upregulation/downregulation of target genes to the expression of a protein trigger. In an attempt to extend the iSBH toolkit, I sought to engineer new iSBH-sgRNAs to allow temporal exogenous control over CRISPR-TR activity.

Short, single-stranded DNA (ssDNA) antisense oligonucleotides (ASO) have been widely used to interfere with mRNA splicing, hasten transcript degradation, or block translation with the aim of reducing protein production in the cell (Bennett & Swayze 2010). Researchers have notably shown that 13 to 25nt ssDNA ASOs bearing sequence complementarity to an mRNA could selectively lead to its degradation by engaging the endoribonuclease RNase-H, present in the nucleus of mammalian cells (R. Y. Walder & J. A. Walder 1988; Zon 1995; Dias & Stein 2002). This enzyme, which had been implicated with the maintenance of genome stability (Sassa et al. 2016), detects DNA:RNA hybrids found in the cell nucleus and hydrolyses the phosphodiester bonds of the RNA strand (Donis-Keller 1979). Studies have shown that a minimum of 6 consecutive base-pairs is necessary for the DNA:RNA heterodimer to be processed (Dias & Stein 2002; Eder & J. A. Walder 1991). In contrary to the endoribonuclease used in the previous chapter, RNase-H does not cleave its substrate at specific locations.

Instead, upon binding, the enzyme proceeds to sporadically cleave the RNA strand of the DNA:RNA hybrid (Lima et al. 2007; Sassa et al. 2016).

I envisioned that RNase-H mediated RNA slicing could be used as a spacer release mechanism to create ASO-responsive iSBH-sgRNAs (Fig. 4.1a). Employing the same strategy used to create Csy4-responsive guides, I speculated that one could engineer ASO-responsive iSBH-sgRNAs by replacing the default GAAA loop of the SBH with a larger sensing-loop designed to be complementarity to an ASO trigger. In this scenario, delivery to the cell of an ASO complementary to the sensing-loop should lead to the formation of a DNA:RNA hybrid between the iSBH-sgRNA and the trigger. In turn, RNAse-H-mediated cleavage of the hairpin should yield an active sgRNA able to direct CRISPR-TR on-target (Fig. 4.1a).

I was particularly interested in using ASOs as trigger for the iSBH system because these molecules are easily synthesized and have been safely and efficiently delivered to cells and organisms in both laboratory settings and in the clinic (Khvorova & Watts 2017). Notably, the drug Vitravene, designed by Isis Pharmaceuticals Inc. to treat cytomegalovirus retinitis, has received FDA approval (Orr 2001; Roehr 1998), and many other compounds have entered clinical trials (Khvorova & Watts 2017; Agrawal & Kandimalla 2000). Additionally, I reasoned that the sequence diversity inherent to the design of ASOs should allow for the creation of a large repertoire of inducer:target pairs. Indeed, when using *N* nucleotide long ASOs, a total of 4^N distinct inducer sequences are available. Assuming that each ASO is only able to trigger iSBH-sgRNAs with a fully complementary sensing-loop, 4^N orthogonal ASO:iSBH pairs could, in principle, be created.

4.1 – ASO-responsive iSBH-sgRNA first generation

In order to create ASO-responsive sgRNAs, the loop segment of the bulged hairpin SBH$^{(0B)}$CTS2 was replaced with a longer 14nt ASO sensing-loop (ASL) to create iSBH$^{(0B)}$ASOα-CTS2 (Fig. 4.1b). Studies have shown that ASO-mediated mRNA degradation is highly dependent on target accessibility, and repeatedly found that targeting mRNA segments predicted to be in single-stranded conformation generally yield greater repression efficiency (Scherr et al. 2000; Ding & Lawrence 2001). Correspondingly, the design of the ASL was guided by the RNA secondary structure prediction algorithm Nupack (Zadeh et al. 2011), to guarantee that the iSBH would fold while leaving the loop nucleotides unpaired (Fig. 4.1b). Transfection of the resulting iSBH$^{(0B)}$ASOα-CTS2-sgRNA along with dCas9-VP64 and reporter plasmid showed that extending the hairpin loop from 4 to 14nt did not disrupt SBH-mediated silencing of CRISPR-TR (Fig. 4.1c).

To test if this iSBH-sgRNA could be specifically activated in the presence of the matching ASO, a 14nt DNA oligonucleotide (ASOα(14)), whose sequence was the reverse complement of the ASL, was synthesized. Studies have reported that simple DNA ASOs were short lived in mammalian cells due to exonuclease degradation (Urban & Noe 2003; Fisher et al. 1993; Dagle et al. 1991). As a solution to that problem, phosphorothioate modifications of the ASO, whereby a sulfur atom replaces a non-bridging oxygen in the oligo phosphate backbone, have been widely used to increase its bioavailability (Dias & Stein 2002). This modification has been found to increase half-life from one to ten hour(s) due to improved exonuclease resistance, while maintaining RNase-H-mediated cleavage (Dias & Stein 2002; Dagle et al. 1990; Stein et al. 1991; Srinivasan & Iversen 1995; Kurreck et al. 2002). Accordingly, ASOα(14) was synthesized with four adjacent phosphorothioate bonds on both 5-' and 3'-ends (Fig. 4.2a, b)[15].

[15] Replacing the entire ASO backbone with phosphorothioate bonds was found toxic to the cells (data not shown), an effect that was previously attributed to ASO OFF-target binding (Iannitti et al. 2014).

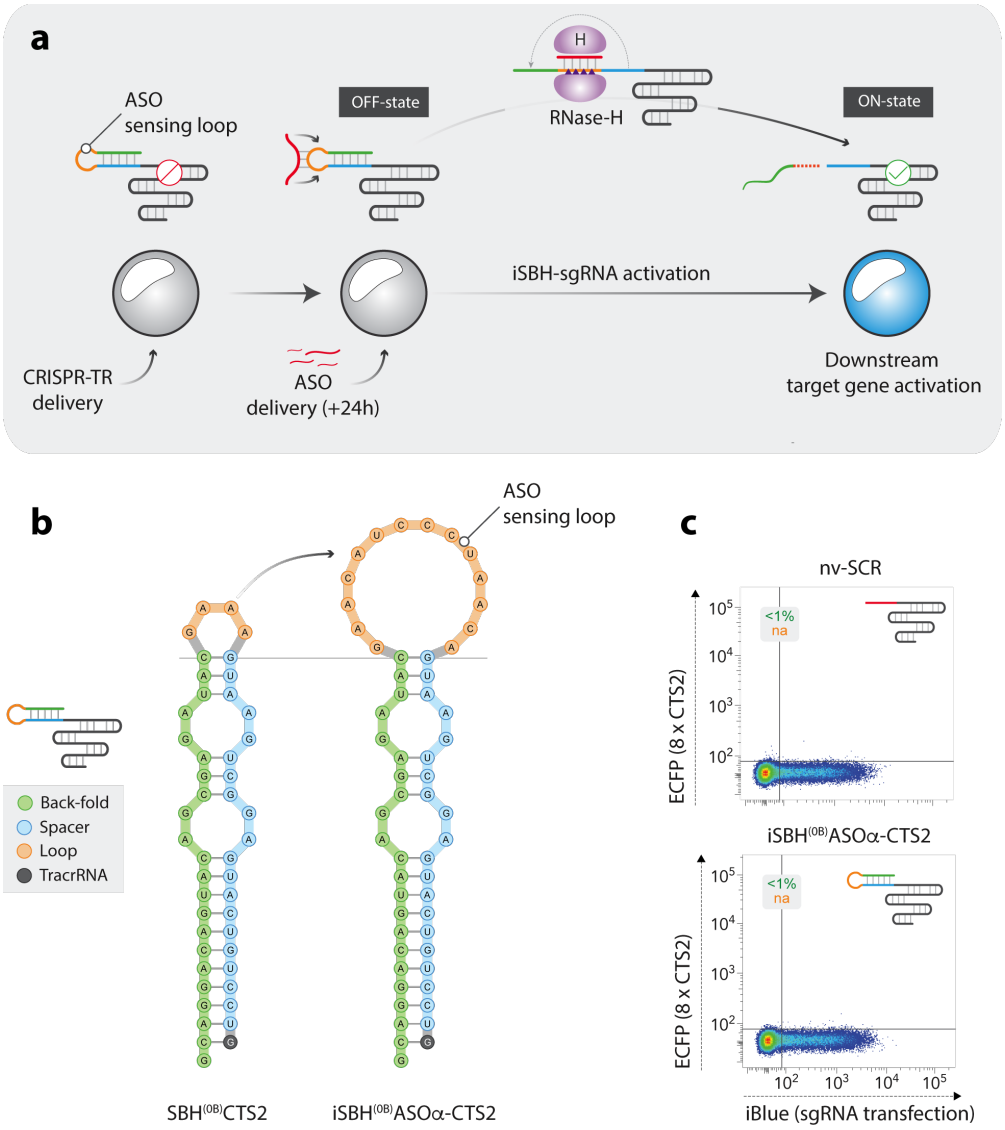

Figure 4.1 – ASO-responsive iSBH-sgRNA. (a) Implementation of iSBH-sgRNAs specifically triggered by DNA antisense oligonucleotides (ASO). The silent SBH-sgRNA is modified to incorporate an ASO sensing-loop. Upon delivery of the sequence complementary ASO, a DNA:RNA hybrid forms through Watson-Crick base-pairing. Follows RNase-H-mediated cleavage of the loop leading to spacer release and CRISPR-TR activation. **(b)** Sequence and RNA secondary structure of an ASO-responsive iSBH (iSBH$^{(0B)}$ASOα-CTS2) featuring a 14nt sensing-loop designed to assume an open conformation. **(c)** HEK293-T were transfected with dCas9-VP64, reporter plasmid and either a native sgRNA with scrambled spacer sequence (nv-SCR) or iSBH$^{(0B)}$ASOα-CTS2-sgRNA. Flow cytometry scatter plots show reporter fluorescence (ECFP) against sgRNA transfection (iBlue). Insets give the % of activated (green) and median reporter fluorescence for that population (orange).

In order to test and demonstrate temporal control over CRISPR-TR using ASO-responsive sgRNAs, the transfection scheme was designed so that the trigger would be delivered 24h after transfection of core system components: dCas9-VP64, 8xCTS2-ECFP-pA, and iSBH-sgRNA (see methods, Fig. 4.1a). This was intended to simulate delayed activation of CRISPR-TR in a situation whereby the CRISPR-effector and ASO-responsive sgRNAs had been stably integrated into a transgenic cell line or a model organism. Flow cytometry analysis 48h post-transfection of cells treated with ASOα(14) revealed CRISPR-mediated reporter activation in 7% of the iBlue^{+ve} cell population (Fig 4.2c). Demonstrating the importance of sequence complementarity between the trigger and the sensing-loop, no ECFP expression was observed when a decoy ASO[16] with a scramble DNA sequence was delivered instead of ASOα(14) (Fig 4.2c).

4.2 – Optimization of the ASO trigger

Next I reasoned that the ON-state performances of the system could be improved by creating ASOs designed to facilitate back-fold:spacer strand separation. To this end, two longer ASOs with increased complementarity to the iSBH were synthesized (Fig. 4.2a, b). The 20nt ASOα(20) was obtained by appending three additional nucleotides on both the 5'- and 3'-end of ASOα(14), to disrupt the three base-pairs of the stem separating the ASL from the seed distal bulge (Fig. 4.2a, b). The second one, ASOα(25), was a 25nt ASO similar to ASOα(20) but extended on the 3'-end with five additional nucleotides complementary to the iSBH back-fold (Fig. 4.2a, b). I hypothesized that extensive complementarity between this ASOα(25) and the back-fold would facilitate opening of the back-fold:spacer stem prior to ASL cleavage. Head to head comparison of all three ASOs revealed that both extended variants outperformed ASOα(14), in that both induced greater CRISPR-TR activity (Fig. 4.2c, d).

[16] The decoy ASO was also modified to incorporate flanking phosphorothioate bonds.

To my surprise however, ASOα(25) did not mediate stronger sgRNA activation than ASOα(20), a result that might be explained by a reduction in the overall accessibility of the RNA target (Fig. 4.2c, d).

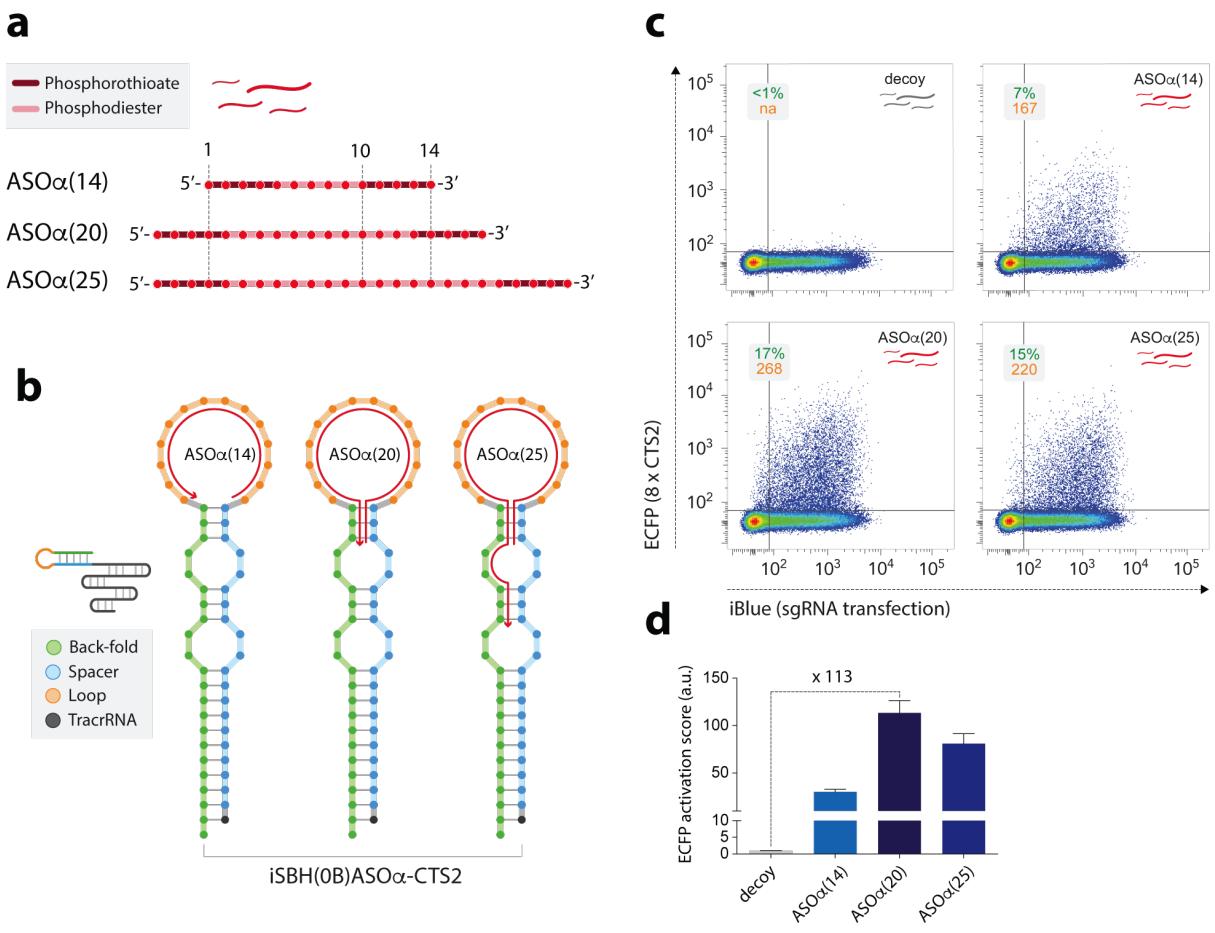

Figure 4.2 – Optimization of the ASO footprint. (a) Relative length and formulation of three antisense oligonucleotide (ASO) triggers. **(b)** Hybridization footprint for each ASO (red arrows) in (a) on the iSBH$^{(0B)}$ASOαCTS2 hairpin. **(c)** HEK293-T cells were transfected with dCas9-VP64, reporter plasmid, and iSBH$^{(0B)}$ASOαCTS2-sgRNA. ASOα(14), ASOα(20), ASOα(25), or a decoy ASO with scrambled sequence was delivered 24h post-transfection. Flow cytometry analysis scatter plots show reporter expression (ECFP) against sgRNA transfection (iBlue). % of activated cells (green) and median ECFP expression (orange) are given for each condition in the inset. **(d)** Activation score calculations for all conditions in (c) (n = 3, mean +/- s.d., normalized to the decoy condition).

I then went on to ask if the optimal design presented in figure 4.2 could be applied to the design of a new iSBH-sgRNA targeting a different CTS. Using the design principle listed above, I created the iSBH$^{(0B)}$ASOβ-CTS1 hairpin, programmed to target CTS1 and respond to a distinct 20nt ASOβ (Fig. 4.4a). As previously observed with iSBH$^{(0B)}$ASOα-CTS2-sgRNA, the resulting iSBH-sgRNA mediated reporter expression only in the presence of its cognate inducer (Fig. 4.3a, b). To demonstrate the specificity of this effect, both iSBH-sgRNAs were tested in two additional negative control conditions, whereby either the sequence of the inducer, or the sequence of the ASL was mutated (Fig. 4.3a). In both cases, transcriptional activation of the reporter gene was effectively decoupled from the delivery of the trigger ASO (Fig. 4.3a, b).

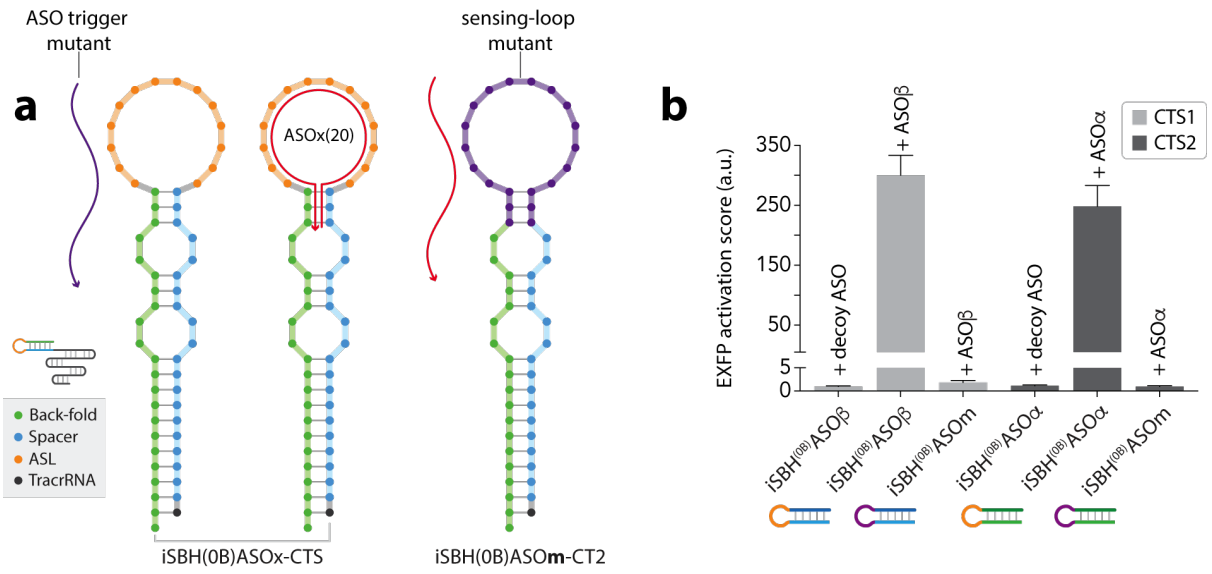

Figure 4.3 – Specificity of ASO-mediated CRISPR-TR activation. (a) Schematic representation of the three conditions designed to test the specificity of ASO-mediated sgRNA activation: (i) Cells were transfected with an ASO-responsive iSBH-sgRNA programmed to target a given CST and respond to ASOx (iSBH$^{(0B)}$ASOx-CTS) (left). A mutant ASO (purple arrow) whose sequence does not match the ASO sensing-loop (ASL) of the iSBH-sgRNA is delivered 24h later. (ii) Same as (i) but the trigger ASOx matching the ASL is delivered in place of the decoy (red arrow) (center). (iii) Same as (ii) but the ASL on the iSBH-sgRNA is scrambled (purple) (right). **(b)** The three conditions in (a) are tested for two iSBH-sgRNAs targeting CTS1, CTS2 and responding to ASOβ, ASOα, respectively. In all cases the iSBH-sgRNA is co-transfected in HEK293-T along with dCas9-VP64 and the reporter plasmid. The decoy or matching ASO is delivered 24h post-transfection. Bars show CRISRPa activation scores for each condition 48h post-transfection (n = 3, mean +/- s.d., value normalized to a scramble sgRNA condition).

4.3 – Stem shortening does not improve ON-state performances.

While optimizing Csy4-responsive iSBHs, I found that reducing the stability of the back-fold:spacer stem resulted in an increased ON-state performances. It was notably shown that iSBH-sgRNAs, in which the Csy4-PTS was directly fused to the seed distal or seed central bulge of SBH$^{(0B)}$CTS, performed better than their counterpart with a full 20-bp stem (see section 3.4). To determine if the same rule applied to ASO-responsive iSBHs, I modified the last sgRNA tested (iSBH$^{(0B)}$ASOβ-CTS1) to fuse the ASL with the seed distal bulge, thus increasing the minimal free energy of the stem from -23.9 kcal.mol^{-1} to -19.6 kcal.mol^{-1} (iSBH$^{(0B)}$ASO$^{\gamma(medium)}$-CTS1, Fig. 4.4a).

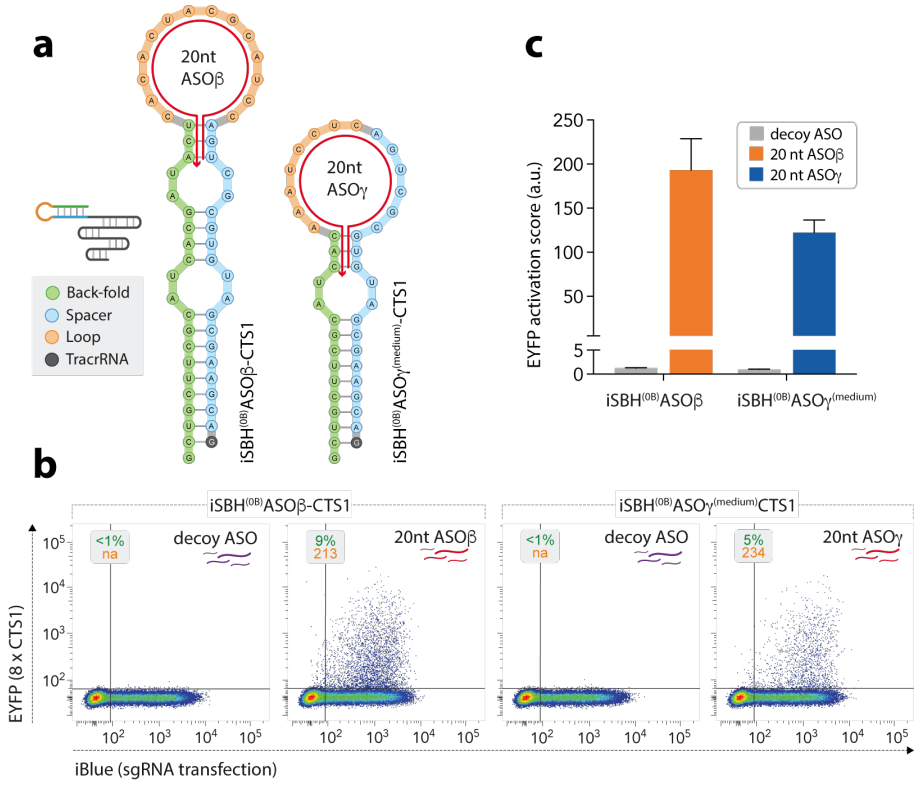

Figure 4.4 – Effect of stem shortening on the system's ON-state performances. (a) Sequence and RNA secondary structure of ASO-responsive iSBHs. A *medium* design is created by fusing the ASO sensing-loop (ASL) to the seed distal bulged of the stem. **(b)** HEK293-T were transfected with dCas9-VP64, reporter plasmid and each of the iSBH-sgRNAs in (a). A decoy (purple) or matching ASO (red) was delivered 24h post transfection. Flow cytometry scatter plots show reporter expression (EYFP) against sgRNA transfection (iBlue). Insets report % of activated cells (green) and their median reporter fluorescence (orange) for each condition. **(c)** Activation score calculations for each condition in (b) (n = 3, mean +/- s.d., normalized to a scramble sgRNA condition).

Direct comparison of iSBH$^{(0B)}$ASOβ-CTS1 and the shortened iSBH$^{(0B)}$ASOγ$^{(medium)}$-CTS1 in the presence their cognate ASO trigger, revealed that shortening the stem did not improve CRISPR-TR activity (Fig. 4.4b). Instead, the ON-state activation score of the shorter design was found to be lower than its full-length counterpart (Fig. 4.4c).

This surprising result might partly be explained by the fact that the ASL sequence were different between the two iSBHs. While conserving the same ASL across hairpins would have facilitated the comparison, I reasoned that ON-state performances of iSBH$^{(0B)}$ASOγ$^{(medium)}$-CTS1 could benefit from keeping as many spacer nucleotide as possible. Accordingly, iSBH$^{(0B)}$ASOγ$^{(medium)}$-CTS1 ASL was engineered such that the last 6 nucleotides of the new ASL (3'-end) would be part of the CTS1 spacer (Fig. 4.4a). The sequence of the remaining 8 nucleotides (5'-end) was evolved using a custom-made algorithm (see section 5.2.1) to favor open conformation of the ASL (Fig. 4.4a). Going back to the results presented in figure 4, one could argue that two distinct ASL sequences might have different affinity for their respective ASO trigger, leading to variation in the corresponding iSBH-sgRNA activation. While this reasoning could indeed explain why the shorter iSBH-sgRNA activated to a lesser extent than the original design, this possibility was mitigated by the fact that iSBH$^{(0B)}$ASOγ$^{(medium)}$-CTS1 ASL was chosen such that both loops would share identical GC content. An alternative, more plausible, explanation could be that the observed reduction in performance was due to ASL accessibility. Looking at the crystal structure of the crRNP bound to its DNA target, it appears that the 5'-end of the 20nt spacer just about exits the protein (Nishimasu et al. 2014). Therefore, while the ASL of iSBH$^{(0B)}$ASOβ-CTS1 is likely exposed to the solvent, thus facilitating ASO docking, the loop of the *medium* design could be buried into dCas9, reducing its accessibility for both ASO and RNase-H.

Finally, it was found in the previous chapter that *nano* hairpins, obtained by fusing the Csy4-PTS to the central bulge of SBH$^{(0B)}$CTS, drastically improve ON-state CRISPR-TR activity. Similarly, *nano* ASO-responsive hairpins could in theory be created by shorting the back-

fold:spacer stem and grafting the 14nt sensing-loop onto the central bulge. Nevertheless, I decided against testing such design for the following two reasons: (1) Results presented in figure 4.4 point towards the fact that shortening the stem in the case of ASO-iSBHs is detrimental for the system ON-state performances. (2) If 20nt ASO triggers are to be used on a *nano* ASO-iSBH, and considering that RNase-H cleaves off from the sgRNA all nucleotides base-pairing with the ASO, only 5 spacer nucleotides are expected be left post induction. In section 2.3.3 of this Thesis, I showed that native sgRNAs with spacer truncated to 5nt are not functional in that they cannot mediate transcriptional activation of their target gene (see figure 2.7 and 2.8).

4.4 – Closing remarks

In this section, I have introduced ASO-responsive sgRNAs, engineered by replacing the GAAA loops of the $SBH^{(0B)}$CTS hairpin with extended 14nt ASO sensing-loops. I showed that delivery of a sequence complementary 14nt ASO was able to trigger the otherwise silent iSBH-sgRNA, resulting in transcriptional activation of the targeted gene. While this effect has been attributed to RNase-H-mediated processing of the ASL, further experiments will be required to characterize the exact mechanism of activation. The exact role of RNase-H in this process could notably be tested through modulation of the nuclease's levels or rescue experiments in an RNase-H null cell line. If the mechanism proposed above is true, iSBH-sgRNA activation, and consequently reporter expression, should positively correlate with RNase-H concentration. Complementary experiments could also use chemical modification of the ASOs, such as incorporation of 2′-O-(2-methoxy)ethyl RNA nucleotides, that inhibit RNase-H catalytic activity without affecting its substrate binding affinity (Ebert et al. 2007; Kole et al. 2012).

Next, I showed that the ON-state performances of ASO-responsive iSBH-sgRNAs could be improved by using 20nt long ASOs, designed to locally disrupt back-fold:spacer base-pairing. As a next step, the system could be further optimized by trying new ASO formulation. Notably, chimeric ASOs, also known as gapmers, containing a phosphorothioate backbone and either

2′-O-methyl or locked nucleic acids (LNA), have been shown to mediate stronger antisense activity (Kurreck et al. 2002; Grünweller et al. 2003; Lennox et al. 2006). With regard to the design of the iSBH, future developments could include lengthening the ASL to increase the affinity with the inducer, extending the stem to maximize ASL accessibility, and identifying ASL sequences that favor RNase-H mediated cleavage.

In general, I remarked that the ASO triggered CRISPR-TR yielded weaker reporter activation than what was achieved with the previously presented Csy4-resonsive iSBH-sgRNAs. With regard to the strength of the activation, the median reporter fluorescence of the activated cell population was considerably greater with the *nano* Csy4-iSBH than the best ASO-iSBH design. In the activation of 8xCTS1-EYFP-pA, the ON-state median reporter fluorescence was almost doubled when using iSBH$^{(0B)}$Csy4$^{(nano)}$CT1 (401 a.u., Fig. 3.7b) compared to iSBH$^{(0B)}$ASOβ-CTS1 (213 a.u., Fig. 4.4b). Similarly, on the CTS2 reporter, a 2.4 fold increase in median ECFP fluorescence was observed between the *nano* Csy4-responsive iSBH (Fig. 3.8b) and iSBH$^{(0B)}$ASOα-CTS2 (Fig 4.2c). The difference in performances between Csy4- and ASO-responsive systems was even more striking when looking at the fraction of activated cells in the ON-state. Indeed, a maximum of 17% activated cells was observed when using iSBH$^{(0B)}$ASOα-CTS2 (Fig. 4.2c), while the corresponding Csy4-responsive hairpin reached up to 62% of cells expressing ECFP (Fig. 3.8b). Below I discuss the potential factors that may explain these differences in performance:

(i) iSBH residuals post-cleavage: As mentioned in the introduction of this chapter, RNase-H does not cleave its substrate at pre-defined locations (Lima et al. 2007; Sassa et al. 2016). Consequently, despite the fact that base-pairing between ASO and the hairpin is expected to extend into the stem, ASO-responsive iSBH-sgRNAs could have, post-cleavage, an ASL residual on their 5'-end. Data presented in chapter 2, suggest that the addition of a floating tail on the 5'-end of the sgRNA can impair to some degree CRISPRa (see section 2.3.1). In contrast, *nano* Csy4-resonsive iSBH-sgRNAs have spacer shorter than 20-nt post-cleavage.

(ii) Melting of the back-fold:spacer stem: It was found in the previous chapter that the thermodynamic stability of the stem was a major determinant of ON-state performance. Therefore, it was not surprising that the Csy4-iSBH *nano* design outperformed the best ASO-responsive sgRNA. Nevertheless, when comparing iSBH$^{(0B)}$Csy4$^{(full)}$CTS1 and iSBH$^{(0B)}$ASOβ-CTS1, both of which feature the same 20nt bulge stem, the Csy4-responsive hairpin still mediated stronger reporter activation. In addition to the other points listed in this discussion, this difference could partially be explained by the fact that Csy4 remains bound to its substrate post-cleavage, thus facilitating strand separation of the back-fold:spacer stem.

(iii) RNase-H kinetics and target accessibility: I have already suggested in this chapter that the ASL accessibility might be compromised by loading of the iSBH-sgRNA into dCas9. Additionally, studies using an antisense strategy to reduce mRNA levels have shown that RNase-H levels are often the limiting factor limiting the rate of mRNA degradation (Vickers & Crooke 2015; Hongjiang Wu et al. 2004; Vickers et al. 2014). While slow RNase-H kinetics should not affect the number the % activated cells in the ON-state, it might cap the extent of reporter expression by limiting the number of functional sgRNAs available.

(iv) Inducer availability: The Csy4 trigger was expressed under a constitutive promoter and co-transfected along with the iSBH-sgRNA and CRISPR-effector plasmids. Despite the fact that the protein is a single turn over enzyme, its constant expression engendered sustained production of newly activated iSBH-sgRNA, from the moment the inducer was translated until the analysis (~48h). In contrast, ASO triggers were delivered as a single dose 24h post transfection, which induced a transient wave of active iSBH-sgRNAs before being cleared from the cell. Therefore, it is expected that ASO-triggered CRISPR-TR would generate lower levels of reporter protein compared to the continuous Csy4 activation. Additionally, the time point chosen to analyze the cells might not have been optimal. A time course experiment would be required to monitor reporter expression following ASO delivery and precisely identify when the peak activation is reached.

As a general note for the reader interested in implementing inducible CRISPR-TR using ASO inducers, I provide below a series of guidelines and design principles that one can follow to create an ASO-iSBH for any spacer sequence of choice: (i) Choose a gene target and determine the optimal CRISPR-target site (CTS, N20-NGG) for transcriptional activation. (ii) Clone the guide RNA spacer corresponding to the chosen CTS in a plasmid expressing the sgRNA under a U6 promoter (see methods). Test the efficiency of this guide by transfecting cells of interest with the sgRNA and dCas9-VP64 pDNAs, and assaying changes in gene expression 48 hours post transfection using reverse transcription quantitative PCR (see methods). Note: I recommend replacing dCas9-VP64 with a more potent second generation CRISPR transcriptional activator (see section 1.5.2). I notably show in section 6.1 that ASO-responsive iSBH-sgRNAs are compatible with the synergistic activation mediator system (SAM) (Konermann et. al 2014). (iii) Once the sgRNA has been tested and has been found to efficiently increase the transcriptional output of the target gene, an ASO-iSBH for the corresponding spacer can be designed using the iSBHfold web tool presented in chapter 6. The design rules used to design an ASO-iSBH from a given spacer sequence are illustrated in figure 4.5 below. Note that several ASO-iSBHs can be created from the same spacer by selecting distinct ASO sensing-loops (ASL). The only requirement is to choose the ASL sequence such that it will assume an open conformation once the iSBH folds. Accordingly, one can synthesise several iSBH featuring distinct ASLs, test guide induction with their cognate ASO, and select the one that display the best ON-state activity.

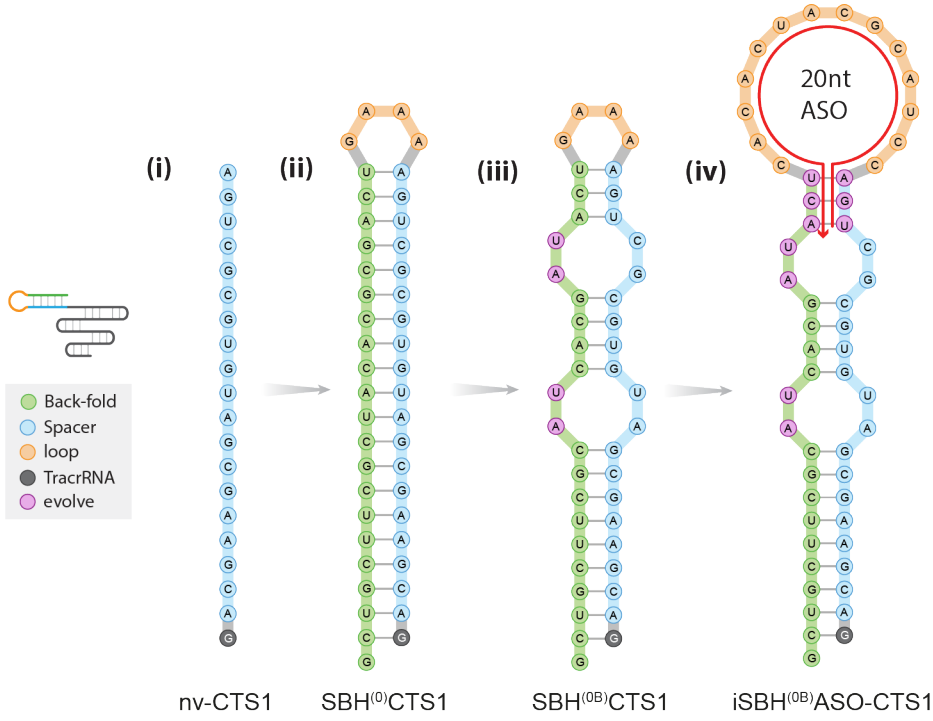

Figure 4.5 – Design principles used to create ASO-responsive iSBHs. Starting from a given 19-20nt spacer sequence (i, here 19nt spacer targeting CTS1 is taken as example, nv-CTS1), the reverse complement is taken to obtain the sequence of the matching full length back-fold. The back-fold sequence is then fused to the spacer using a default *GAAA* loop to create SBH$^{(0)}$CTS1 (ii). Note that the SBH sequence starts with a 5' guanine required for U6 transcription. The stability of the full stem is then reduced by mutating the back-fold sequence to form two 2bp bulges over spacer nucleotides 4-5 and 10-11, respectively (iii, SBH$^{(0B)}$CTS1, spacer nucleotides are numbered from 1 to 19 from loop to tracrRNA). Finally, a corresponding ASO-iSBH is created by replacing the default *GAAA* loop with a 14nt ASO sensing-loop whose sequence must be evolved to assume an open conformation upon iSBH folding (iv, iSBH$^{(0B)}$ASO-CTS1, evolution of the loop is automated in the iSBHfold web tool). Note that the three ASL proximal nucleotides on the back-fold and spacer segment (shown in purple) can be varied to sense a distinct ASO trigger without affecting the ability of the CRISPR-TR to find its DNA target.

Chapter 5 – Scaling up: assembly of gene circuit modules

Over the last three chapters, I have introduced the concept of inducible spacer-blocking hairpin and used the iSBH platform to create a series of switchable sgRNAs. Notably, I have shown that iSBH-sgRNAs can easily be reprogrammed to respond to different inducers, protein or ASOs, by interchanging their sensing-loops. In parallel, I have also demonstrated that the optimal design principles used to create an iSBH against a given trigger, can readily be applied across several spacer sequences to generate multiple iSBH-sgRNAs responding to the same trigger while driving the expression of distinct gene targets.

As such, the intrinsic modularity of the iSBH design and the simplicity of its implementation should make the platform a valuable addition to the synthetic biologist toolkit by facilitating the construction of complex synthetic gene circuits. To illustrate this point, I describe in this chapter the assembly of multi-genes transcriptional programs achieved by expressing, in the same cell, several iSBH-sgRNAs. Doing so, I demonstrate parallel and orthogonal control of distinct gene targets using both protein and ASO inducers.

5.1 – Gene circuits modules

Any complex gene circuits, involving a network of interacting biological parts (Fig. 5.1a), can be decomposed in two fundamental building blocks that I refer to as “branching” and “orthogonal” gene modules (Fig. 5.1b, c). The branching module corresponds to a situation

whereby a single upstream event triggers the activation of multiple downstream targets (Fig. 5.1b). In the particular case of conditional CRISPRa, the concomitant transcriptional activation of two genes, triggered by a specific intra- or extra-cellular cue, would qualify as such. On the other hand, the orthogonal module refers to a scenario whereby distinct upstream events independently trigger downstream targets (Fig. 5.1c). This would notably be the case if the expression of two genes was controlled by two different inducers without possible cross-talk between gene:inducer pairs.

5.1.1 – Implementation with iSBH-sgRNAs

Based on the considerations mentioned above, it is easy to see how the iSBH platform could be employed to assemble such gene modules. By delivering two iSBH-sgRNAs each targeting a given gene, branching and orthogonal modules are created by programming the two guides to respond either to the same inducer (Fig. 5.1d), or two distinct triggers (Fig. 5.1e). This reprogramming is achieved by grafting onto the iSBH-sgRNAs an identical, or two different sensing-loops, respectively. In the reminder of this chapter, I demonstrate construction of both gene circuits using either protein- or ASO-responsive inducible sgRNAs.

5.1.2 – Dual reporter plasmid

To demonstrate simultaneous and independent control of two output genes (EYFP, ECFP), I have created a dual reporter plasmid which alleviates the need to co-transfect both 8xCTS1-EYFP-pA and 8xCTS2-ECFP-pA vectors. The dual reporter was obtained by cloning, on the same vector, the two reporter cassettes separated by the pol-II transcriptional pause signal from the human alpha2 globin gene (see methods) (Fig. 5.2a). Since the aim was to measure concomitant or independent activation of both fluorophores, it was important to demonstrate that no crosstalk existed between the native nv-CTS1 and nv-CTS2 sgRNAs, and that any technical bleed-through between the two flow cytometry channels could be accounted for and corrected using a compensation matrix. To this end, the CTS1 and CTS2 sgRNA:reporter pairs were first transfected separately to calibrate the flow cytometer (Fig. 5.2b). A compensation

matrix was derived to correct for any technical bleed-through between the EYFP and ECFP channels.

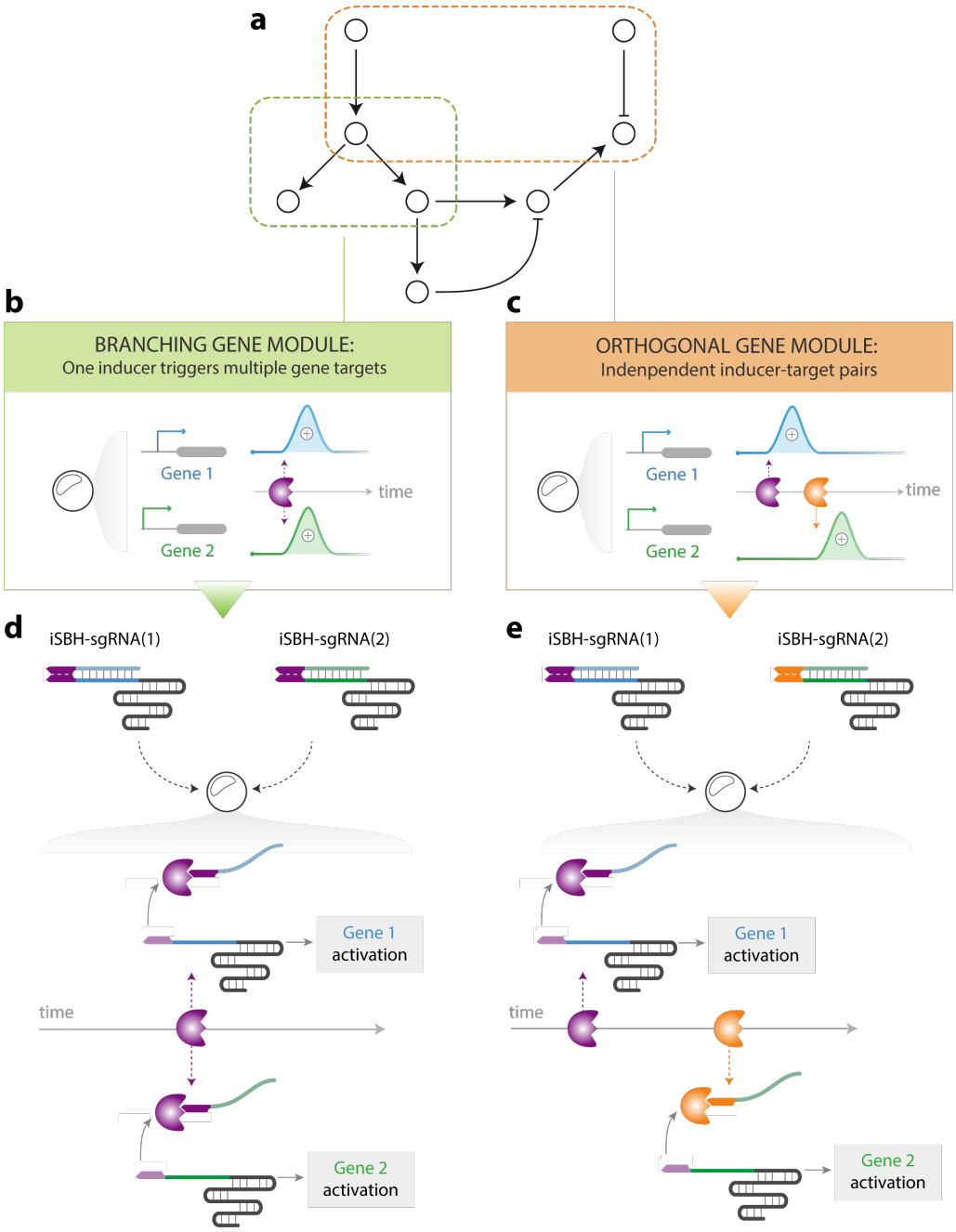

Figure 5.1 – Decomposition of complex synthetic gene circuits. **(a)** Schematic representation of a synthetic gene circuit represented as a network of interacting biological parts. Dashed line boxes outline examples of branching (green) and orthogonal (orange) gene modules, respectively. **(b)** Branching gene module: simultaneous activation of two target genes using a single inducer. **(c)** Orthogonal gene module: Two gene targets are controlled independently by two distinct inducers. **(d,e)** iSBH-based implementation of the branching (d) and orthogonal (e) gene modules.

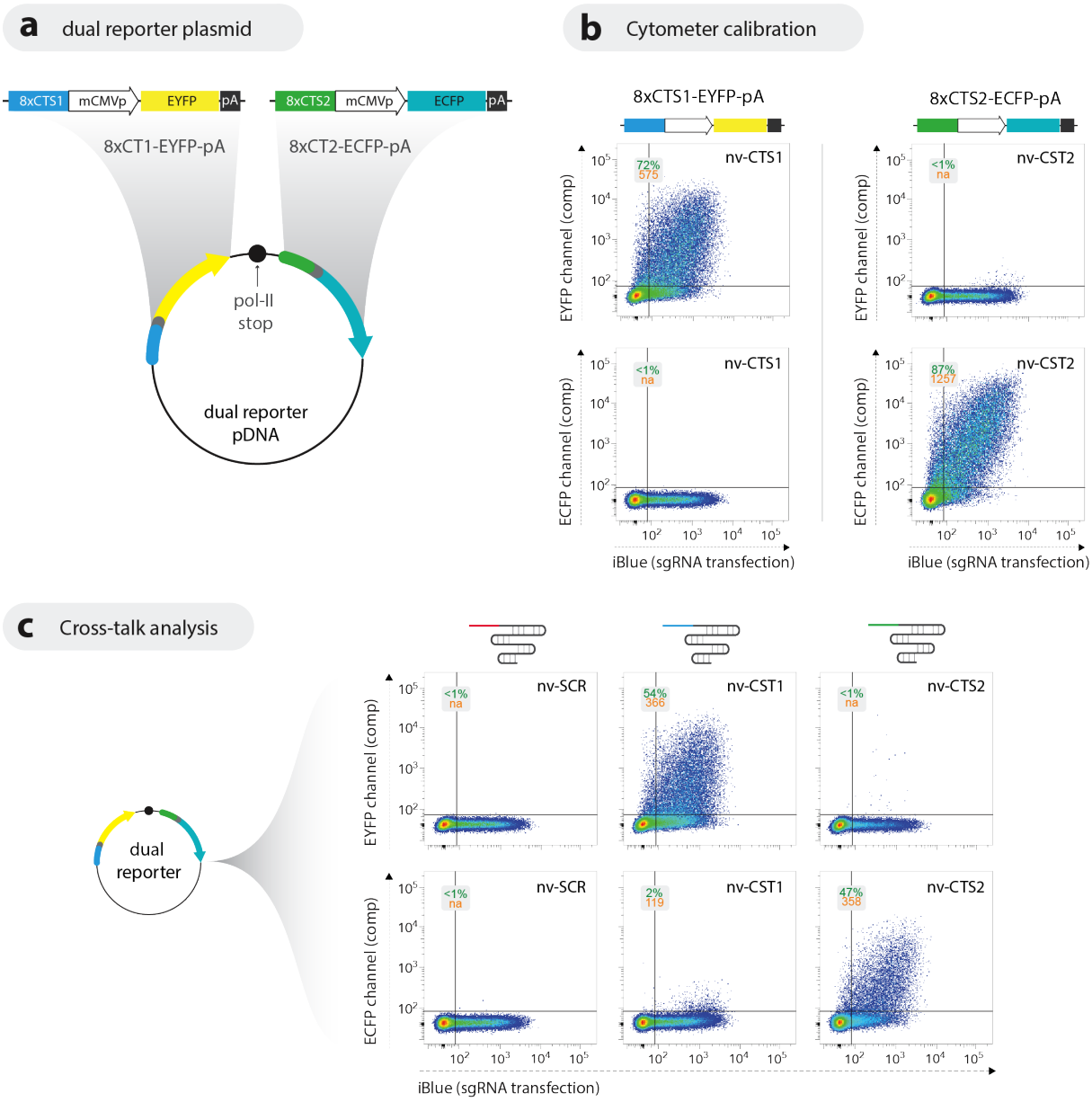

Figure 5.2 – Dual reporter plasmid and cross-talk analysis. (a) Schematic of the dual reporter plasmid. 8xCTS1-EYFP-pA and 8xCTS2-ECFP-pA cassettes are cloned in series, separated by a pol-II stop signal. **(b)** HEK293-T cells were transfected with dCas9-VP64, and each sgRNA:reporter pairs separately. The flow cytometer was calibrated so that no signal was picked up in the ECFP channel when EYFP was expressed and *vice versa*. Flow cytometry scatter plots show EYFP (top) or ECFP (bottom) expression against sgRNA transfection (iBlue). **(c)** HEK293-T cells were transfected with dCas9-VP64, dual reporter plasmid, and a sgRNA with either a scramble (nv-SCR), CTS1 (nv-CTS1), or CTS2 (nv-CTS2) spacer.

Once the compensation parameters were obtained, the CRISPRa experiments above were repeated using the dual reporter plasmid in place of 8xCTS1-EYFP-pA or 8xCTS2-ECFP-pA (Fig. 5.2c): HEK293-T cells were transfected with the dual reporter plasmid, dCas9-VP64, and either the nv-CTS1 or nv-CTS2 sgRNA. Results showed that, as expected, both reporter cassettes remained silent in the absence of their cognate sgRNA (delivery of sgRNA with scramble spacer, nv-SCR) (Fig. 5.2c). Additionally, both nv-CTS1 and nv-CTS2 sgRNAs were found to drive the expression of the corresponding fluorescent protein, albeit to lower levels than those observed when using the single-color reporter plasmids (Fig. 5.2b, c). Of note, it appeared that, while no activation of the CTS1 cassette was observed when nv-CTS2 was transfected, expression of the nv-CTS1 sgRNA led to a negligible activation of the ECFP (2% activated cells, Fig. 5.2c). Nevertheless, results presented in figure 5.2 suggested that this dual reporter plasmid was well suited for the assembly and analysis of branching and orthogonal gene modules.

5.2 – Assembly of branching gene modules

5.2.1 – Branching modules with protein-responsive iSBH-sgRNAs.

I reasoned that a branching gene module could be created by designing two iSBH-sgRNAs triggered by the endoribonuclease Csy4 (Fig. 5.1d). Accordingly, the optimized *nano* design was used to create Csy4-responsive iSBHs for both the CTS1 (iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1) and CTS2 (iSBH$^{(0B)}$Csy4$^{(nano)}$CTS2) spacers (see section 3.4). The resulting guides were transfected in HEK293-T cells along with dCas9-VP64, the dual reporter construct, and either a decoy pDNA or a vector expressing Csy4 (Fig. 5.3a). Flow cytometry analysis conducted 48h post-transfection revealed that neither EYFP nor ECFP were expressed in the absence of Csy4 (Fig. 5.3b, c, condition 1). In contrast, co-transfection of a plasmid encoding the inducer led to the simultaneous activation of both reporter genes with 51% and 65% of transfected cells (iBlue^{+ve}) showing EYFP and ECFP expression, respectively (Fig. 5.3b, condition 4).

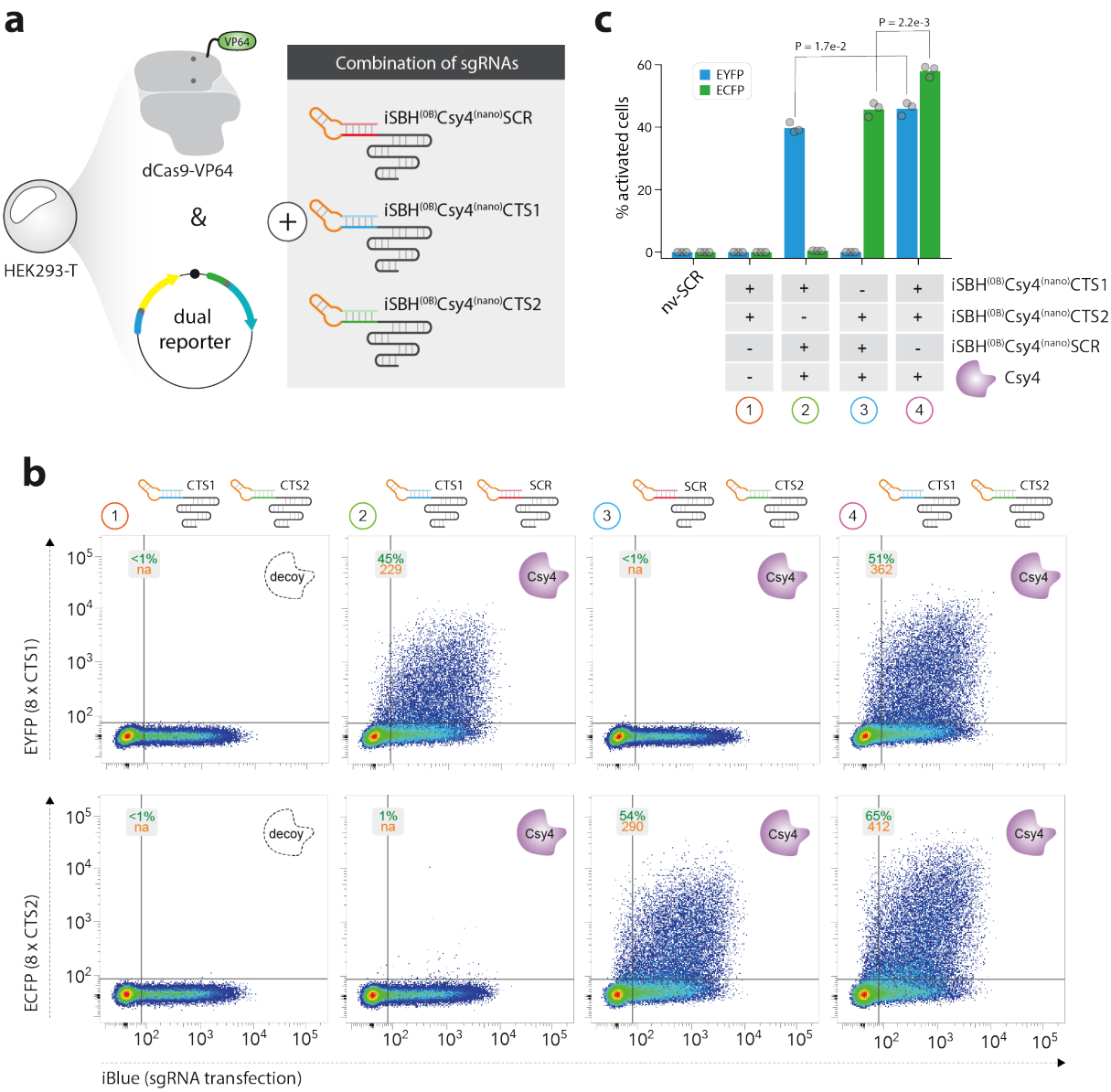

Figure 5.3 – Branching module with protein-responsive iSBH-sgRNAs. (a) Transfection scheme implementing the branching gene module: HEK293-T cells were transfected with dCas9-VP64, dual reporter plasmid, and a pair of Csy4-responsive iSBH-sgRNAs chosen among iSBH$^{(0B)}$Csy4$^{(nano)}$SCR (scramble spacer), iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1 (targets CTS1), and iSBH$^{(0B)}$Csy4$^{(nano)}$CTS2 (targets CTS2). Pairs chosen for each condition are shown in (b). Additionally, either a decoy inducer or a Csy4 expression vector was co-transfected (not represented). **(b)** Flow cytometry scatter plots show the expression pattern of EYFP (top) and ECFP (bottom) reporter genes against sgRNA transfection (iBlue) in all four conditions. Insets show fraction of activated cells (green) and median reporter fluorescence (orange) in both channels. **(c)** Bar graph gives the % of activated cells for all conditions in (b) (n = 3, bars show mean, t-test p-values are displayed).

In order to show that both iSBH-sgRNAs were in fact required for the concomitant activation of both reporters, two additional conditions were tested in which either iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1 or iSBH$^{(0B)}$Csy4$^{(nano)}$CTS2 was replaced by an iSBH-sgRNA featuring a functional Csy4-PTS but having a scramble spacer (iSBH$^{(0B)}$Csy4$^{(nano)}$SCR). In both conditions, it was found that removal of one of the original iSBH-sgRNAs effectively decoupled the activation of its gene target to the presence of the inducer, demonstrating their role as signal transducers (Fig. 5.3b, c, condition 2 and 3).

5.2.2 – Branching modules with ASO-responsive iSBH-sgRNAs.

Next, the iSBH-sgRNAs were reprogrammed to produce branched activation of EYFP and ECFP conditioned on the delivery of an ASO inducer. As explained in section 4.1, ASO-responsive iSBHs are designed to fold such that the ASL would assume an open conformation, rendering it accessible to its cognate ASO trigger. This was achieved by evolving the ASL sequence to minimize base-pairing between ASL nucleotides, as well as between the ASL and the rest of the stem (back-fold and spacer). As a consequence, an ASL compatible with SBH$^{(0B)}$CTS1 might not necessarily be compatible with SBH$^{(0B)}$CTS2, as they feature different spacer and back-fold sequences. To streamline the design process, I developed a custom-made computer program which combines RNA secondary structure predictions powered by Nupack (Zadeh et al. 2011) and an optimization approach known as a genetic algorithm to evolve ASLs displaying the desired conformation across multiple spacer sequences (Fig. 5.4).

The objective of the program is to derive from an input set of *n* spacer sequences (CTS1, ..., CTS*n*), a list of candidate ASLs which would assume an open conformation when grafted onto the corresponding SBH$^{(0B)}$CTS1,…, SBH$^{(0B)}$CTS*n*. Given that 20nt long ASO inducers are been used, those candidate ASLs would have to be isolated from a list of 4^{20} different sequences. As such, a “brute force” search for the best candidates, which involves testing every single sequence, would take years for a regular computer to complete. Instead, optimization methods have been developed to intelligently search the sequence space and find optimal solutions in fewer steps (Amaran et al. 2016).

The genetic algorithm mimics the natural selection process to evolve, from an original pool of solutions, incrementally fitter candidates. The system does so by iteratively adding to the population offspring generated by crossing together the solutions that best satisfy the optimization criterion, while discarding the ones that perform the poorest.

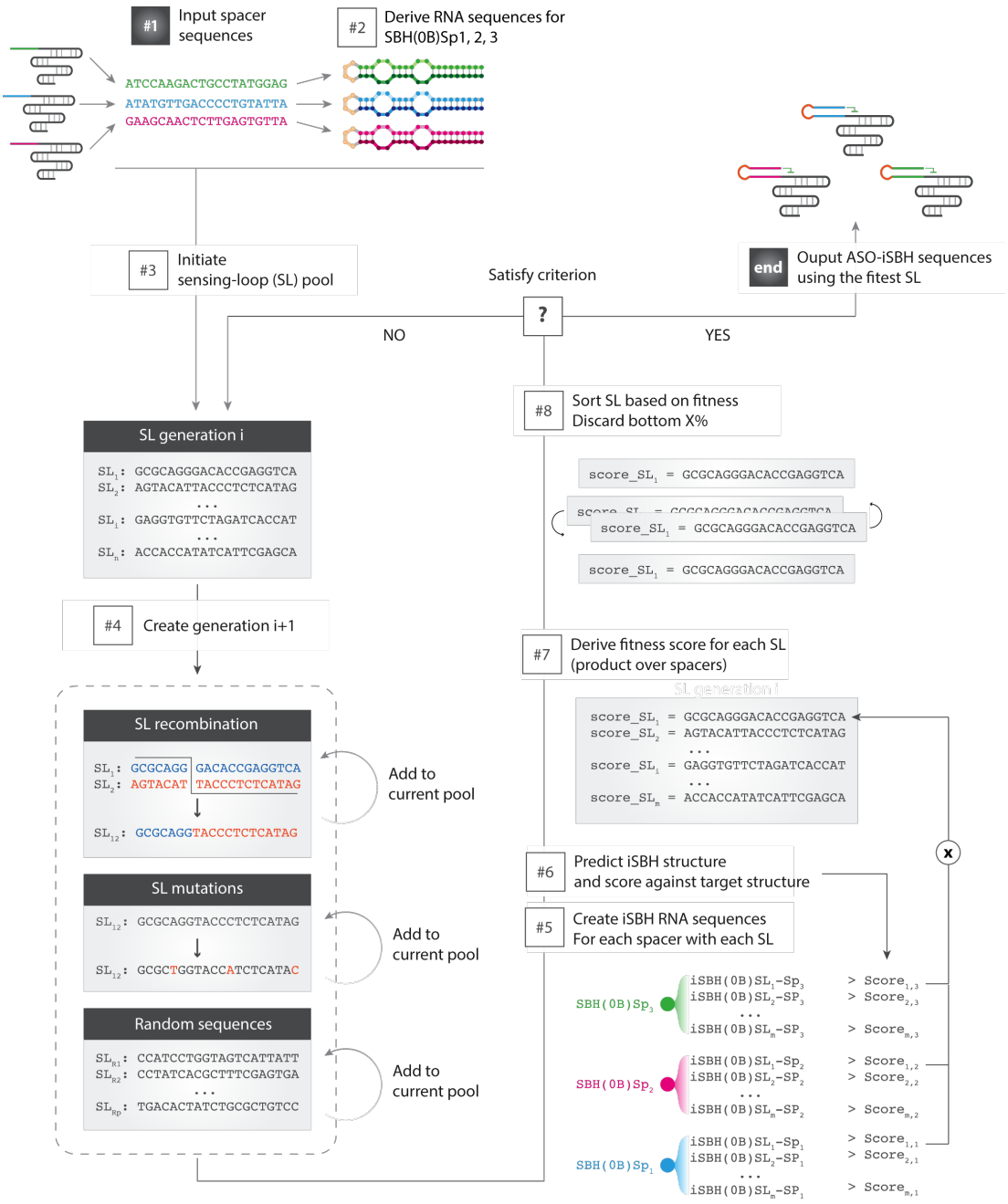

Figure 5.4 – Genetic algorithm used to evolve ASL sequences. Schematic representation of the custom made genetic algorithm used to evolve ASL sequences compatible with several spacer sequences. Each step is attributed a number and referenced in the main text explaining the algorithm.

Using this principle, I have created a computer program that iteratively evolves a list of favorable candidate ASLs from an original pool of randomly generated 20nt sequences (Fig. 5.4). As the starting point of the program, the user provides as input a list of *n* spacer sequences (Sp_1, …, Sp_n). For each spacer, the system then derives the DNA sequence of the corresponding back-fold required to create the corresponding full length $SBH^{(0B)}Sp_i$ (Fig. 5.4, step 1 and 2). A starting ASL pool (generation 0) is then assembled with randomly generated 20nt-long DNA sequences (Fig. 5.4, step 3). From there, the next generation is created by adding to the previous one "offspring" ASLs created by a sequence recombination between ASLs from the parent population (Fig. 5.4, step 4). Additionally, randomly mutated variants of the parent sequences, as well as new fully randomized sequences are injected into the pool to generate higher "genetic" diversity and avoid trapping in local optimum and (Fig. 5.4, step 4). For each ASL_j in the current population, the system then assembles the DNA sequences of $iSBH^{(0B)}ASL_j$-Sp_1, …, $iSBH^{(0B)}ASL_j$-Sp_n obtained by grafting the ASL_j onto the bulged SBH of each spacer (Fig. 5.4, step 5). These hairpins are then converted to RNA and the Nupack software is used to both predicts their RNA secondary structure and attribute a score to each iSBH ($Score_{j,1}$, …, $Score_{j,n}$), which reflects how well the predicted structure matches the target design (open conformation) (Fig. 5.4, step 6). The *n* scores $Score_{j,1}$, …, $Score_{j,n}$ are then multiplied together to generate a single "fitness" score for each ASL (Fig. 5.4, step 7). Finally, the ASLs of the current population are ranked based on their score from best to least fit candidate before discarding the bottom 50% of the pile (Fig. 5.4, step 8). Taking the resulting population as a new starting point, the process is then repeated all over again (Fig. 5.4, steps 4 to 8) until the resulting population satisfies a given set of criteria: (i) the maximum number of iteration is reached; (ii) The best score of the pool plateaus for several generations. Once the criteria are satisfied, the system outputs to the user a list of potential ASLs.

Using the aforementioned algorithm, an ASL sequence compatible with $SBH^{(0B)}$CTS1 and $SBH^{(0B)}$CTS2 was obtained and used to generate two ASO-responsive iSBH-sgRNAs, $iSBH^{(0B)}ASO\delta$-CTS1 and $iSBH^{(0B)}ASO\delta$-CTS2, designed to respond to the same 20-nt ASOδ

(Fig. 5.5a). Co-transfection of the two guides in HEK293-T cells, along with dCas9-VP64 and the dual reporter plasmid, showed that both designed iSBH-sgRNAs were efficiently silenced by their respective hairpin and did not activate when a decoy ASO (scramble sequence) was delivered 24h post-transfection (Fig. 5.5b, c, condition 1). In contrast, delayed adjunction of the cognate ASOδ led to the simultaneous activation of both CTS1 and CTS2 cassettes (Fig. 5.5b, c, condition 5). Finally, demonstrating the specificity of this effect and the role of each iSBH-sgRNA, EYFP and ECFP transcriptional activation could be decoupled from ASOδ delivery by mutating the ASL of the corresponding guide (Fig. 5.5b, c, conditions 2-4).

5.3 – Assembly of orthogonal gene modules

5.3.1 – Orthogonal modules with protein-responsive iSBH-sgRNAs

As explained in the introduction of this chapter, the iSBH platform can in theory be used to create several orthogonal inducer:target gene pairs that can be controlled independently. This is achieved by expressing in the same cell multiple iSBH-sgRNAs with distinct sensing-loops (Fig. 5.1e). In order to assemble an orthogonal module using protein-responsive iSBH-sgRNAs, I searched the family of CRISPR-associated endoribonuclease (Cas6) for potential candidates inducers that could be used alongside Csy4 (Hochstrasser & Doudna 2015).

The *Thermus thermophilus* Cse3, also known as CasE, has been shown to be involved in crRNA biogenesis by recognizing and cleaving the 29nt RNA repeat interspacing the 32nt spacer of the corresponding CRISPR-array (Sashital et al. 2011; Gesner et al. 2011) (Fig. 5.6a). *In vitro* experiments have shown that out of this 29 nucleotides, only the hairpin structure plus 2 additional downstream nucleotides (U5-G24) were required for substrate cleavage (Sashital et al. 2011). As in the case of Csy4, interaction between Cse3 and its cognate RNA target (Cse3-PTS) is mediated through contacts made with the base (ssRNA:dsRNA jucntion), core, and loop of the repeat stem (Sashital et al. 2011). These characteristics, along with *in vitro* work showing that the protein also remains bound to its substrate post-cleavage, made Cse3 an attractive candidate inducer.

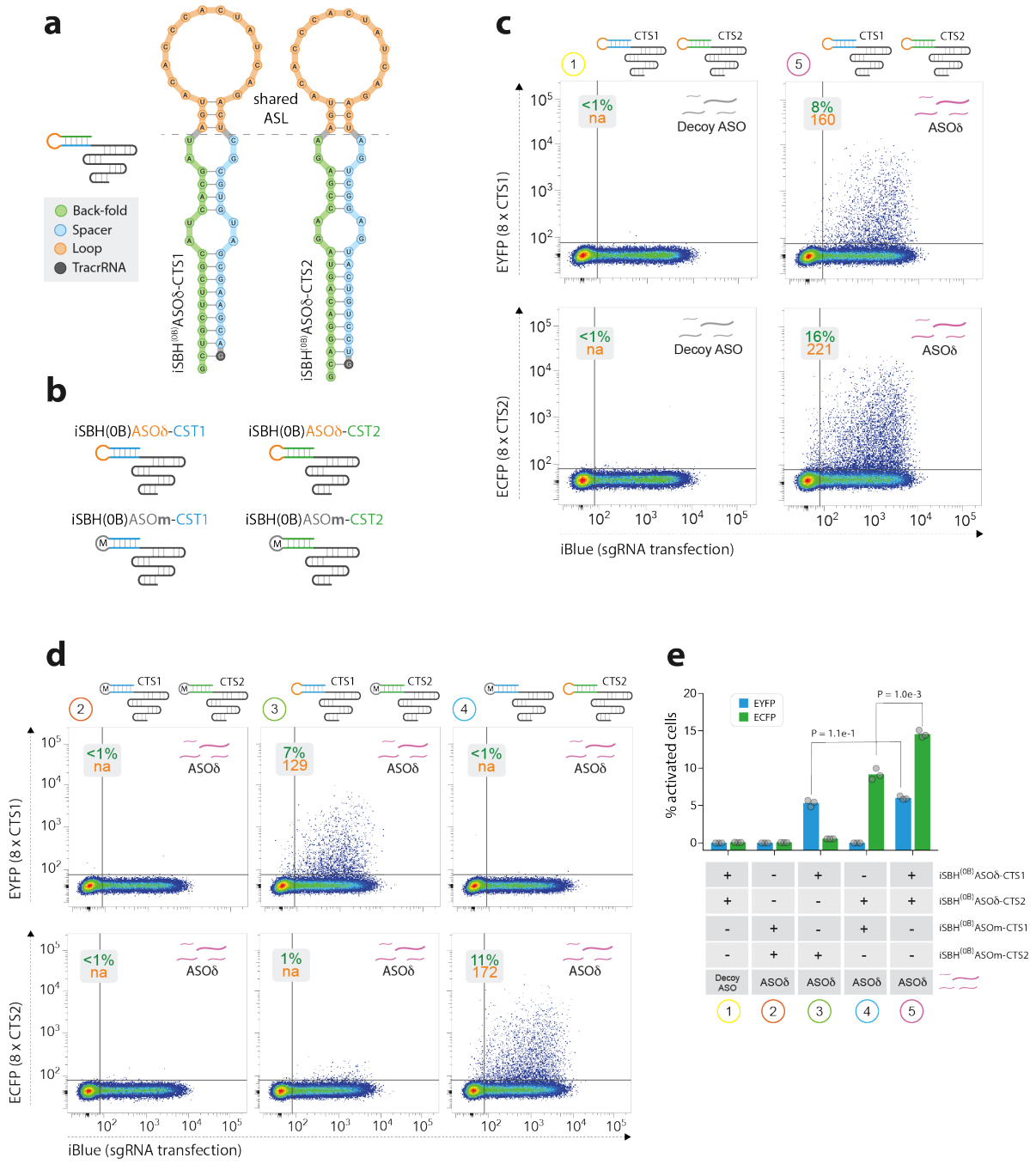

Figure 5.5 – Branching module with ASO-responsive iSBH-sgRNAs. (a) Sequence and RNA secondary structure of two ASO-responsive hairpins sharing the same ASO sensing-loop (ASL) and targeting CTS1 and CTS2, respectively. **(b)** Legend for (c) and (d): schematic representation of iSBH$^{(0B)}$ASOδ-CTS1 and iSBH$^{(0B)}$ASOδ-CTS2 guides along with mutated variants obtained by scrambling their ASL sequence. **(c)** HEK293-T cells were transfected with dCas9-VP64, dual reporter plasmid, and both iSBH$^{(0B)}$ASOδ-CTS1 and iSBH$^{(0B)}$ASOδ-CTS2 guides. Decoy ASO or cognate ASOδ was delivered 24h post-transfection. Flow cytometry scatter plots show reporter fluorescence on both EYFP and ECFP channels against sgRNA transfection (iBlue). Numbering matches conditions in (e). **(d)** Similar to (c) but using mutant ASLs (b). **(e)** % of activated for all conditions in (c) and (d) (n = 3, bars show mean, t-test p-values are displayed).

Based on these considerations, the iSBH$^{(0B)}$Cse3$^{(full)}$CTS2 hairpin, designed to be activated in the presence of Cse3, was created by replacing the default GAAA loop of SBH$^{(0B)}$CTS2 with the Cse3-PTS (A3-A26) (Fig. 5.6b). The design of this iSBH was guided by RNA secondary structure predictions, to ensure that the Cse3-PTS would fold properly and display the structural features that the endoribonuclease recognizes. To maximize Cse3 expression in HEK293-T cells, the protein was codon optimized and cloned in the same plasmid backbone used for Csy4 expression (see methods). Transfection of dCas9-VP64, 8xCTS2-ECFP-pA, and the iSBH$^{(0B)}$Cse3$^{(full)}$CTS2-sgRNA in the presence or absence of Cse3 showed that while the OFF-state behavior matched expectations, ON-state activation of the reporter was extremely weak (Fig. 5.6c). Based on previous design rationales derived when optimizing Csy4-iSBHs, the corresponding *medium* and *nano* Cse3- responsive iSBHs were generated by fusing the Cse3-PTS to the seed distal (iSBH$^{(0B)}$Cse3$^{(medium)}$CTS2) and proximal (iSBH$^{(0B)}$Cse3$^{(nano)}$CTS2) stem bulges (Fig. 5.6b). While these hairpins reduced stem stability from -21.7 to -19.1 and -15.5 kcal.mol^{-1} respectively, their implementation did not yield any improvement in the ON-state performance (Fig. 5.6c).

Thermus thermophiles lives in high temperature environments (~51 °C), which far exceeds the 37 °C at which HEK293-T cells are cultured. This difference in temperature could impair normal functioning of the enzyme and potentially explain the reduced ON-state reported above. To assess if Cse3 retains binding and processing activity in human cells, the Cse3-PTS was cloned into the 5'-UTR of EGFP (Fig. 5.7a). Transfection of this reporter into HEK293-T cells, along with Cse3, showed a clear reduction of EGFP levels compared to a control condition where the endoribonuclease was replaced by a decoy plasmid (Fig. 5.7a). Additionally, Cse3-mediated slicing of the EGFP mRNA was assayed by RT-qPCR using a primer pair spanning the Cse3-PTS (Fig. 5.7b). Comparison between +/- Cse3 conditions revealed a clear reduction in transcript levels, suggesting that the enzyme does in fact have nuclease activity under the tested conditions.

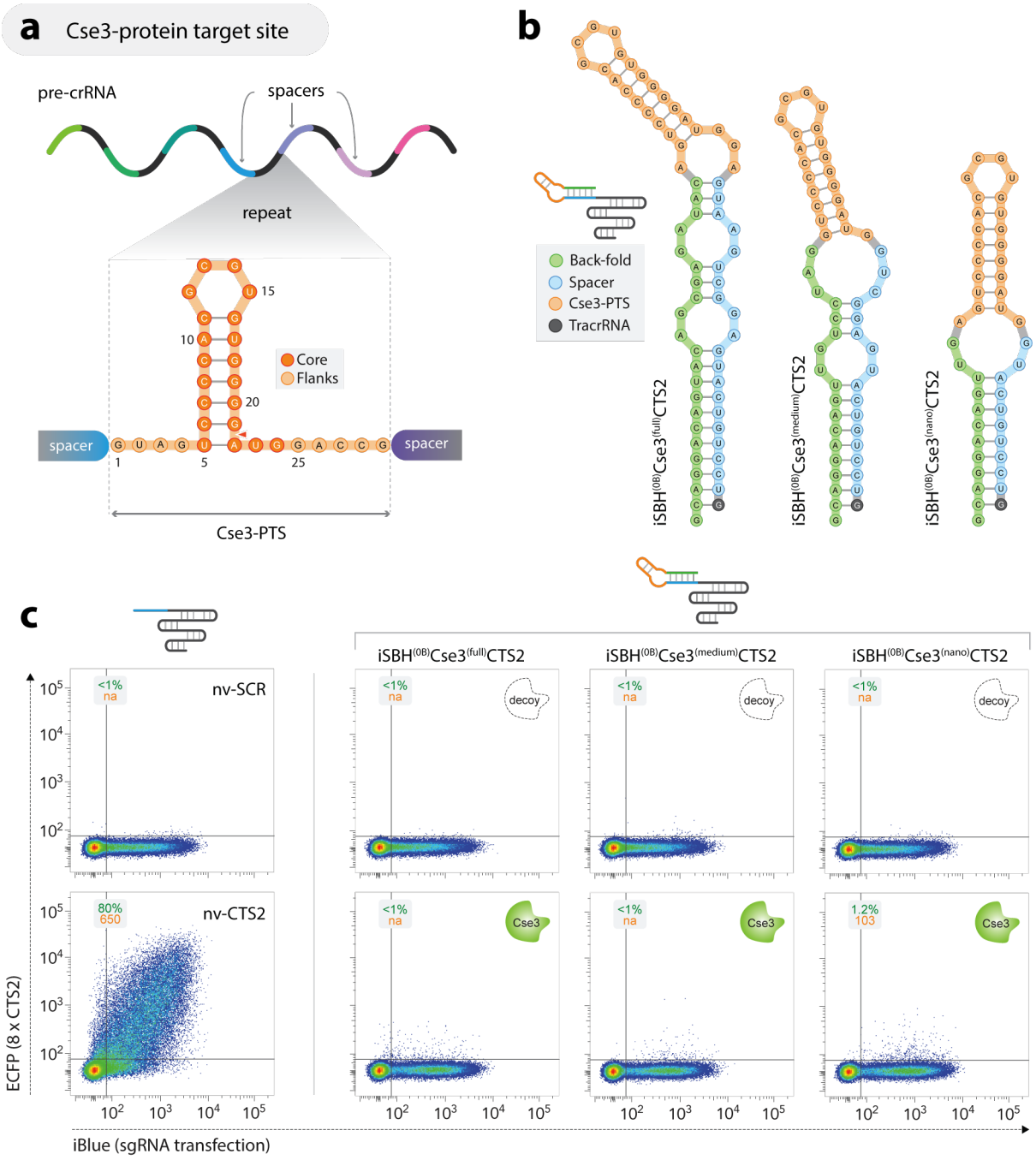

Figure 5.6 – Cse3-responsive iSBH-sgRNAs. (a) Sequence and structure of the RNA repeat targeted by the *Thermus thermophilus* endoribonuclease Cse3 (Cse3-PTS). **(b)** Sequence and RNA secondary structure of three Cse3-responsive iSBHs targeting CTS2. **(c)** HEK293-T cells were transfected with dCas9-VP64, reporter plasmid, and one of the following guides: (i) native sgRNA with scramble spacer (nv-SCR); (ii) native sgRNA targeting the reporter construct (nv-CTS2); (iii-v) Cse3-responsive iSBH-sgRNA of the *full*, *medium*, or *nano* design (see (b)). For the Cse3-responsive iSBH-sgRNA, cells were also transfected with either a decoy plasmid or a vector expressing Cse3. Flow cytometry scatter plots show reporter expression (ECFP) against sgRNA transfection (iBlue). % of activated cells (green) and median reporter fluorescence for those cells (orange) are given in the inset.

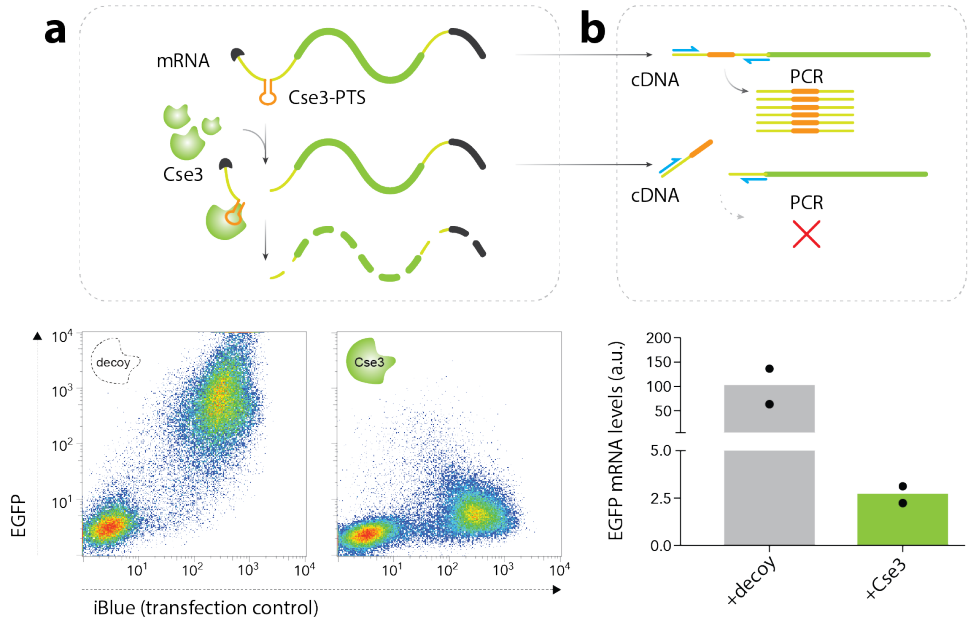

Figure 5.7 – Enzymatic properties of Cse3 in HEK293-T cells. (a) A Cse3 protein target site (PTS) is cloned into the 5'-UTR of a constitutively expressed EGFP reporter gene. HEK293-T cells were co-transfected with the EGFP construct and either a decoy vector (left) or pDNA encoding the endoribonuclease (right). Flow cytometry scatter plots show EGFP levels against iBlue expression (transfection control). **(b)** RT-qPCR analysis is performed on both conditions presented in (a) using a primer pair spanning the Cse3-PTS. Normalized EGFP levels are reported (n = 2, bar shows mean).

Based on this results and the fact that reducing the stability of Cse3-iSBHs did not improve ON-state performances, I then speculated that the endoribonuclease might have a repressive effect on CRISPRa. To test this, the enzyme was co-transfected with the native CRISPR-TR, which I know strongly activates the reporter gene (Fig. 5.8a, b). To my surprise, Cse3 drastically reduced the ability of CRISPR-TR to drive reporter expression irrespective of the spacer:gene target pair used (CTS1 and CTS2) (Fig. 5.8a, b). Bioinformatics analysis failed to uncover a core Cse3-PTS in the sequence of the sgRNA, reporter or in the dCas9-VP64 mRNAs that could have explained this result. Additionally, no Cse3-mediated repression was observed when co-transfecting Cse3 with constitutively expressed EYFP and ECFP (data not shown). Taken together these results suggest that Cse3 might exert a repressive effect on CRISPR-TR. Nevertheless, further experiments are needed to understand the nature of the interactions between the endoribonuclease and the crRNP. However, since this exceeds the scope of the project, I turned my attention to alternative protein inducers.

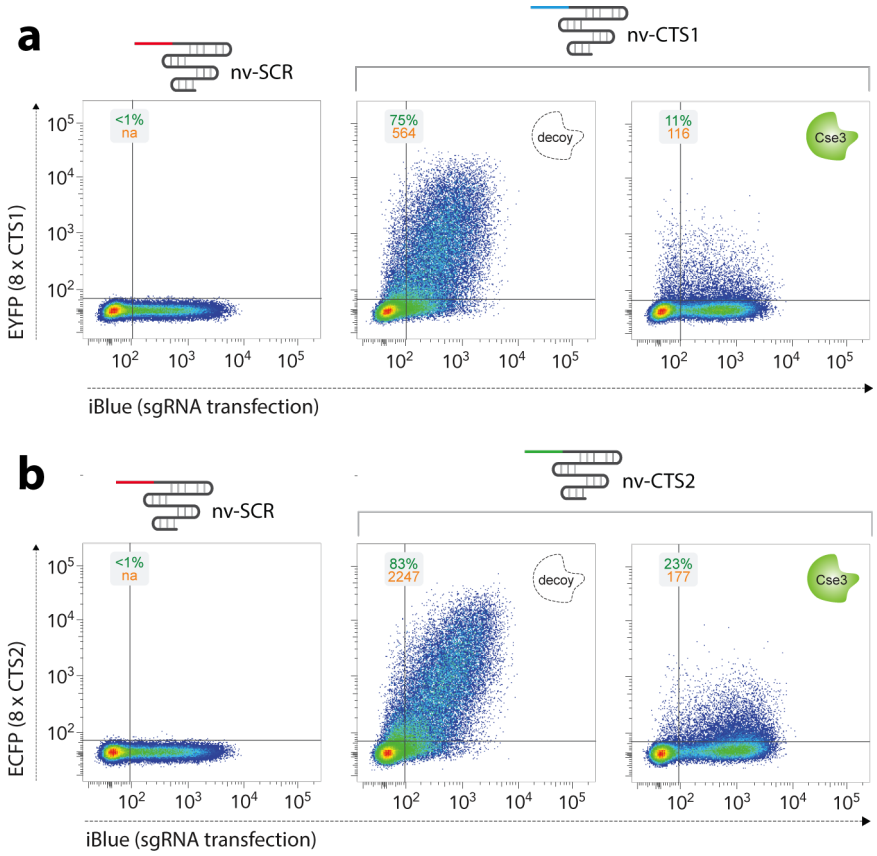

Figure 5.8 – Cse3 represses CRISPRa. (a, b) Effect of Cse3 expression on CRISPRa. HEK293-T cells were transfected with dCas9-VP64, reporter construct (8xCTS1-EYFP-pA or 8xCTS2-ECFP-pA), and the corresponding native sgRNA (nv-CTS1 or nv-CTS2). CRISPRa was tested in the absence (decoy plasmid) or presence of Cse3 and compared to a condition where a guide with scramble spacer is transfected (nv-SCR). Flow cytometry scatter plots show reporter expression (EYFP or ECFP) against sgRNA transfection (iBlue). % of activated cells (green) along with their median reporter expression (orange) is given in the inset.

Another endoribonuclease from the Cas6 clade, *Thermus thermophilus* Cas6A[17], recognizes, binds, and processes the 36-nt repeats of pre-crRNAs. The Cas6A-PTS forms a stem-loop structure flanked upstream and downstream by unstructured RNA segments of 16 and 8 nucleotides, respectively (Niewoehner et al. 2014) (Fig. 5.9a). Similar to Csy4 and Cse3, Cas6A is a single-turnover enzyme, which identifies its substrate via sequence and structure

[17] *Thermus thermophilus* HB8 harbors 11 CRISPR loci containing three distinct types of repeats, each processed by a distinct endoribonuclease (Niewoehner et al. 2014).

specific features. In addition, Cas6A makes specific interactions with the upstream RNA sequence (G8-G16), which are required for substrate binding (Niewoehner et al. 2014).

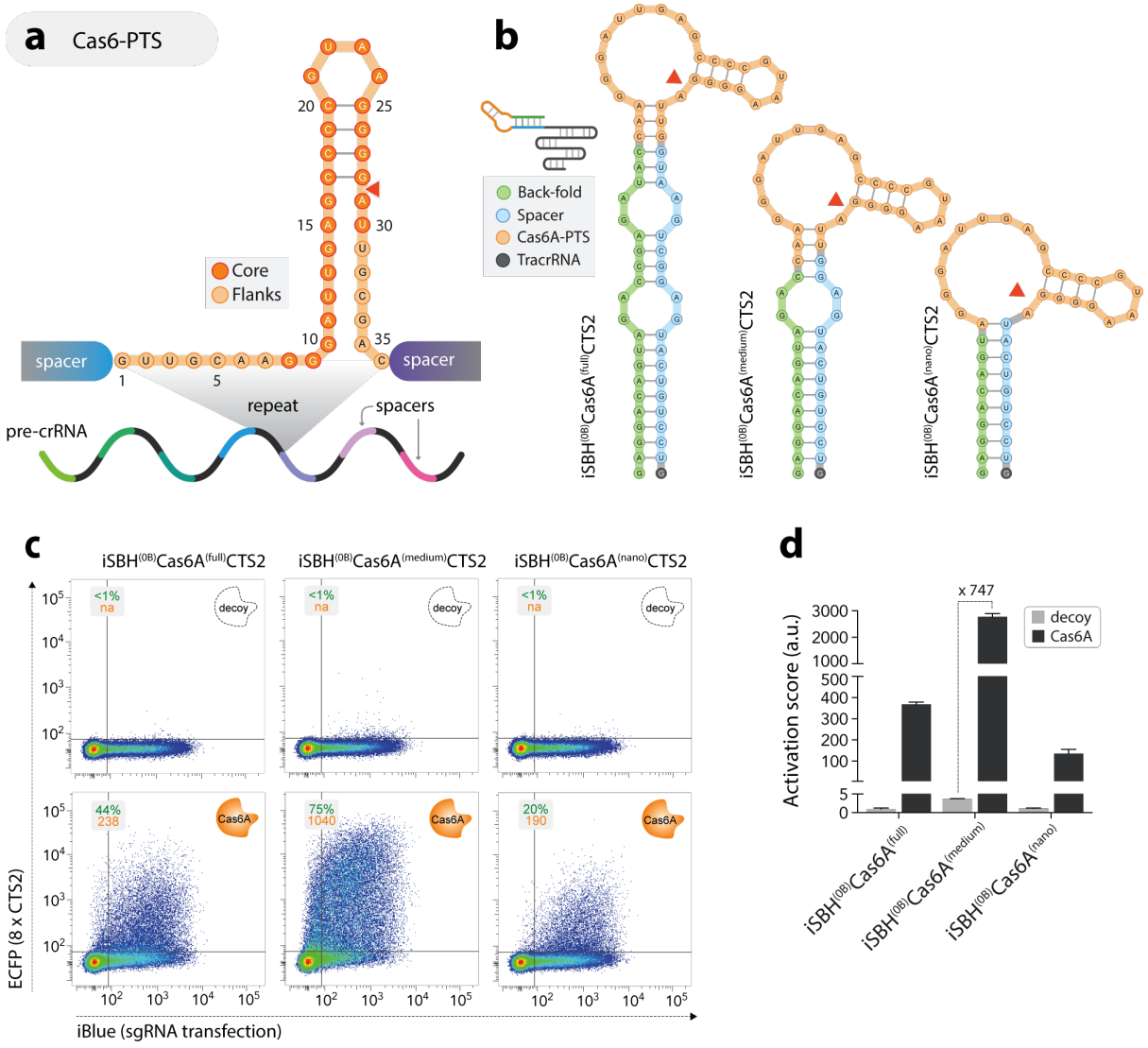

Figure 5.9 – Optimization of Cas6A-repsonsive iSBH-sgRNAs. (a) Schematic representation of the RNA substrate (protein target site, PTS) for the *Thermus thermophiles* Cas6A endoribonuclease. **(b)** Sequence and RNA secondary structure of the *full*, *medium*, and *nano* Cas6A-responsive iSBHs. **(c)** HEK293-T were transfected with dCas9-VP64, reporter construct, and one of the three Cas6A-responsive iSBH-sgRNA in (b). CRISPRa was assayed in the absence (decoy plasmid) and presence of the protein inducer (Cas6A) for each guide. Flow cytometry scatter plots show reporter expression (ECFP) against sgRNA transfection (iBlue). % of activated cells (green) and median reporter expression (orange) for this subpopulation are given in the inset. **(d)** Activation score calculations for each condition in (b) (n = 3, mean +/- s.d., normalized to scramble sgRNA).

To test if Cas6A could be used as a protein inducer, *full*, *medium*, and *nano* Cas6A-iSBHs were designed by fusing the Cas6A-PTS to the stem, distal bulge, and central bulge of the SBH$^{(0B)}$CTS2 hairpin, respectively (Fig. 5.9b). Direct comparison of these three corresponding iSBH-sgRNAs in the CRISPRa assay showed that all iSBHs were capable of maintaining complete silencing of the system in the OFF-state (decoy plasmid conditions) (Fig. 5.9c, d). Additionally, analysis of ON-state performances revealed that iSBH$^{(0B)}$Cas6A$^{(medium)}$CTS2 outperformed both *full* and *nano* designs, improving the activation score obtained with iSBH$^{(0B)}$Cas6A$^{(full)}$CTS2 by ~7.6-fold (Fig. 5.9c, d). Surprisingly, iSBH$^{(0B)}$Cas6A$^{(nano)}$CTS2 only reached 40% of iSBH$^{(0B)}$Cas6A$^{(full)}$CTS2 ON-state performance, as measured by activation score, making it the least potent of all three designs.

Having succeeded in identifying two functional protein inducers with distinct PTS requirements, I set out to assemble orthogonal gene modules using Csy4 and Cas6A to control the independent activation of EYFP and ECFP reporters, respectively. HEK293-T cells were co-transfected with dCas9-VP64, the dual reporter plasmid, and both iSBH$^{(0B)}$Csy4$^{(nano)}$CTS1 and iSBH$^{(0B)}$Cas6A$^{(medium)}$CTS2 iSBH-sgRNAs. Along these system components, four distinct combinations of inducers were delivered that were expected to trigger different patterns of gene activation (Fig. 5.10a) : (i) no protein trigger; (ii) only Csy4; (iii) only Cas6A; (iv) Csy4 and Cas6A together. In all cases expect (iv), a decoy plasmid was also added to match the quantity of pDNA delivered to the cells across conditions. Flow cytometry analysis 48h post-transfection revealed that both reporter cassettes remained silent when neither Csy4 nor Cas6A were expressed in the cells (Fig. 5.10b, c, condition 1). Conversely, delivery of Csy4 or Cas6A in isolation specifically triggered expression of the corresponding target gene (Csy4 triggered EYFP, Cas6A triggered ECFP) as encoded by the iSBH-sgRNAs (Fig. 5.10b, c, conditions 2 and 3). Importantly, no crosstalk was observed between branches, demonstrating that the iSBH-sgRNAs operated orthogonally. Finally, data showed that simultaneous activation of both reporter genes could be achieved by the co-delivery of both protein triggers (Fig. 5.10b, c, condition 4).

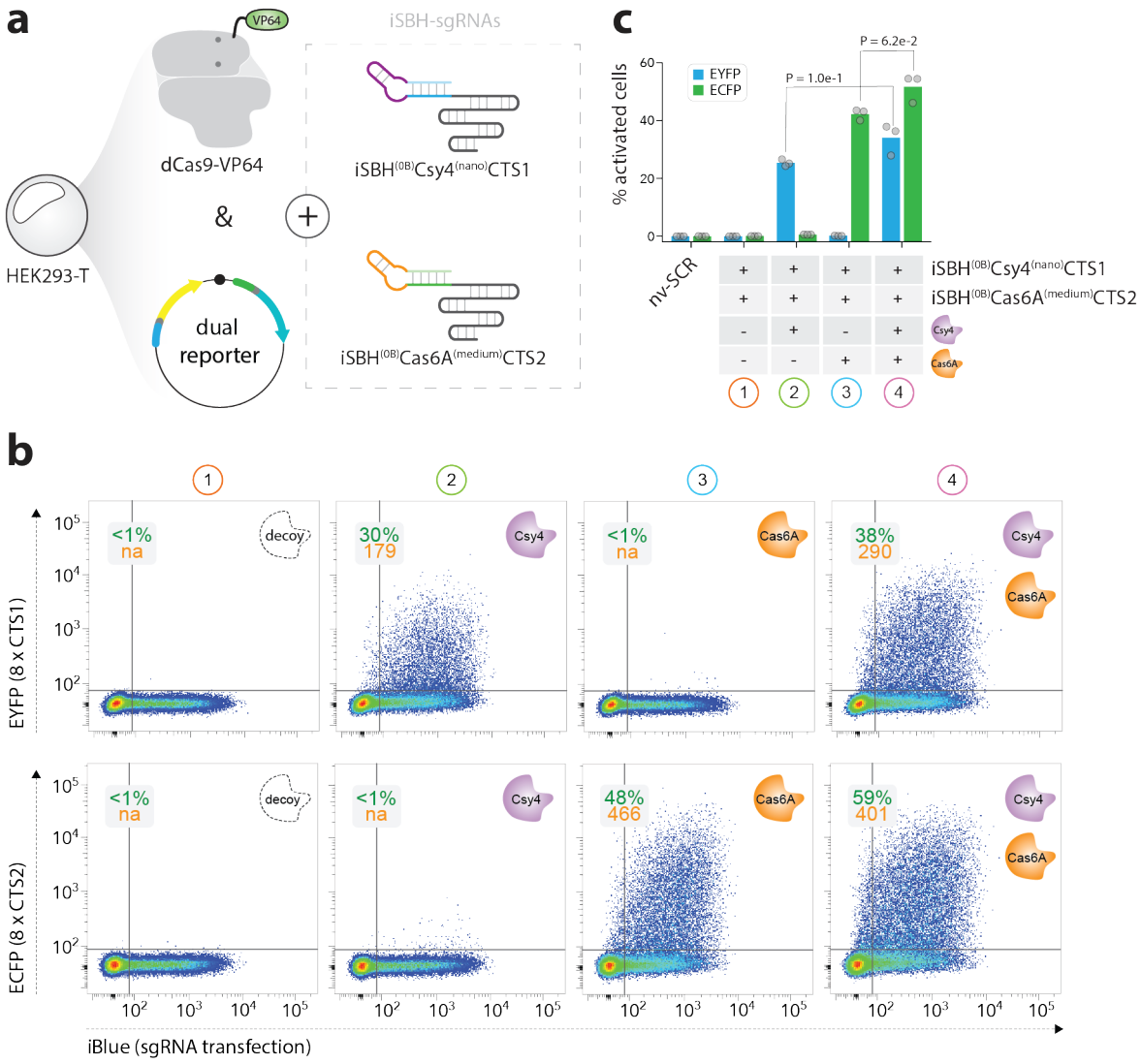

Figure 5.10 – Orthogonal module using protein-responsive iSBH-sgRNAs. (a) Transfection scheme required to implement the circuit: HEK293-T cells were transfected with dCas9-VP64, reporter plasmid, and two iSBH-sgRNAs targeting CT1 and CTS2, and responding to Csy4 and Cas6A respectively. Additionally, different protein inducer cocktails were co-transfected to control the expression of each reporter cassette (not shown). **(b)** Transfection shown in (a) was supplemented with either a decoy plasmid (condition 1), Csy4-expressing construct (condition 2), Cas6A-expressing vector (condition 3), or both inducer proteins (condition 4). Flow cytometry scatter plots show reporter expression (EYFP and ECFP) against sgRNA transfection (iBlue). Insets contain % of activated cells (green) along with median reporter fluorescence for this subpopulation (orange). **(c)** % activated cells for each condition in (b) (n = 3, bars show mean, t-test p-values are displayed).

5.3.2 – Orthogonal modules with ASO-responsive iSBH-sgRNAs.

Similarly, an orthogonal module was implemented using the ASO-responsive iSBH-sgRNAs, iSBH$^{(0B)}$ASOβ-CTS1 and iSBH$^{(0B)}$ASOα-CTS2, previously described in chapter 4. The former was activated by ASOβ and targeted the 8xCTS1-EYFP-pA cassette, while the latter was programmed to respond to ASOα and mediate transcription of 8xCTS2-ECFP-pA. The two guides were co-transfected in HEK293-T cells with the dual reporter plasmid, and dCas9-VP64, followed by delivery, 24h later, of various combination of ASOs (Fig. 5.11a). As illustrated by the absence of reporter expression, delivery of a decoy scramble ASO, lacking sequence similarity to either ASOβ or ASOα, failed to activate any of the two iSBH-sgRNAs (Fig. 5.11b, c, condition 1). In contrast, delivery of ASOβ or ASOα in isolation triggered specific activation of the iSBH-sgRNA carrying a matching ASL, which in turn drove the expression of the corresponding reporter gene (Fig. 5.11b, c, conditions 2, 3). As seen with protein-responsive iSBH-sgRNAs, no crosstalk was observed between inducer:gene target pairs, as ASOβ did not trigger ECFP activation and, respectively, ASOα did not trigger EYFP. Finally, co-delivery in the same cells of both ASO inducers led to the simultaneous activation of their associated target genes (Fig. 5.11b, c, condition 4).

5.4 – Closing remarks

In this chapter, the iSBH platform was used to assemble two fundamental synthetic gene modules. Using a range of iSBH-sgRNAs, it was possible to program transcriptional activation of two transgenes on the presence of either the same inducer (branching gene module) or distinct inducers (orthogonal gene module).

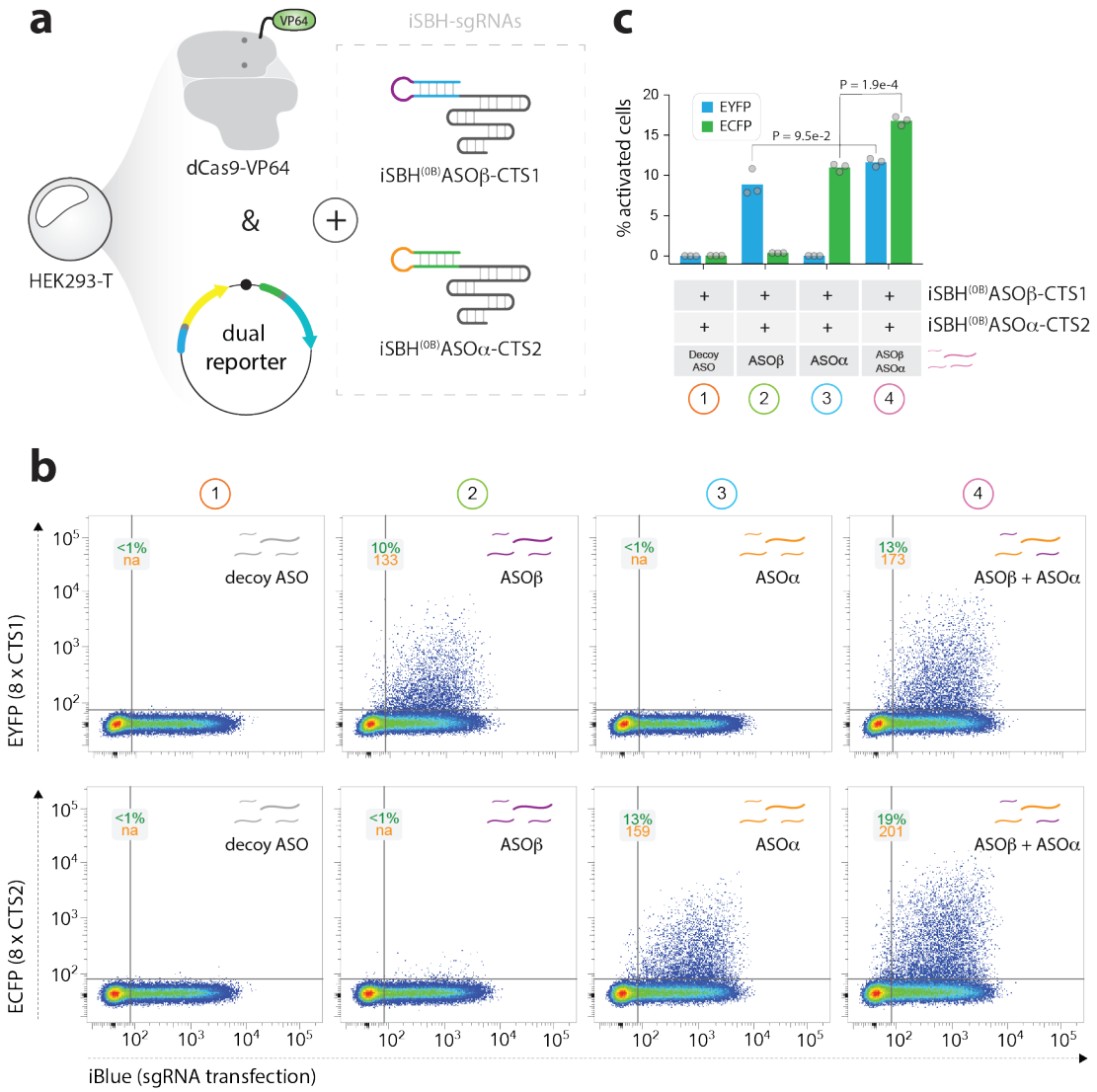

Figure 5.11 – Orthogonal module using ASO-responsive iSBH-sgRNAs. (a) Transfection scheme required to implement the circuit: HEK293-T cells are transfected with dCas9-VP64, reporter plasmid, and two iSBH-sgRNAs targeting CT1 and CTS2, and responding to ASOβ and ASOα respectively. **(b)** 24h post-transfection of the system components (a), cells were transfected with: decoy ASO with no sequence complementarity to the iSBH-sgRNAs sensing-loops (condition 1); ASOβ (condition 2); ASOα (condition 3); a combination of both trigger ASOs (condition 4). Flow cytometry scatter plots show reporter expression (EYFP and ECFP) against sgRNA transfection (iBlue). Insets contain % of activated cells (green) along with median reporter fluorescence for this subpopulation (orange). **(c)** % activated cells for each condition in (b) (n = 3, bars show mean, t-test p-values are displayed).

Throughout the chapter, several conditions were presented whereby both reporter genes were being simultaneously activated (Fig. 5.3, 5.5, 5.10, 5.11). An important consideration for these results was to provide evidence that the concomitant expression of EYFP and ECFP occurred in the same cells and not in two distinct subpopulations. To this end, I revisited data from all four gene modules, namely branched activation with protein- and ASO-responsive iSBH-sgRNAs, and orthogonal control with protein- and ASO-responsive iSBH-sgRNAs, to quantify the fraction of cells expressing both transgenes simultaneously (Fig. 5.12a-d). In all cases, analysis of flow cytometry EYFP/ECFP scatter plots, showed that the majority of cells expressing at least one of the two fluorophores were positive for both EYFP and ECFP. Together, these results suggest that iSBH-based CRISPRa inducibility can be multiplexed to implement complex transcriptional programs, linking a set of inducers with downstream gene targets, in a single cell. Additional experiments will be required to characterize how expression of multiple iSBH-sgRNAs sharing the same sensing-loop impacts the propensity of the inducer to activate each separate branch.

Additionally, I noted that, for all four module experiments, the fraction of cells expressing ECFP was consistently increased when EYFP was being simultaneously activated (Fig 5.3c condition 3 vs. 4, Fig 5.5e condition 4 vs. 5, Fig 5.10c condition 3 vs. 4, Fig 5.11c condition 3 vs. 4). This trend, found significant in three out of four experiments (p-value < 0.05), can most certainly be attributed to unspecific activation of the dual reporter construct. In fact, I have shown in figure 5.2 that targeting of the 8xCTS1-EYFP-pA cassette with a nv-CTS1 guide alone led to an increase in ECFP positive cells, suggesting spurious activation of 8xCTS2-ECFP-pA (see figure 5.2c).

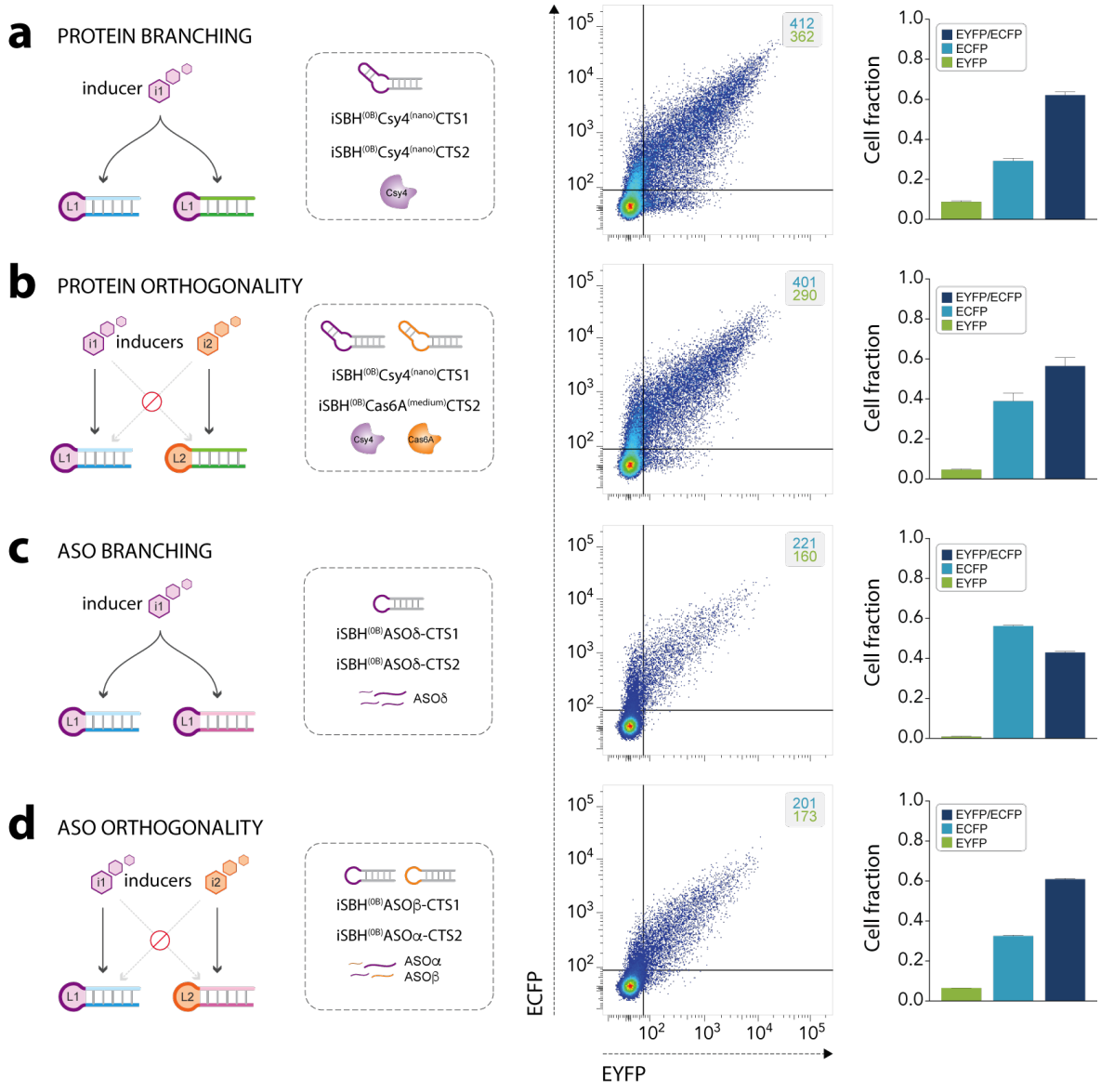

Figure 5.12 – Simultaneous transgene activation, single cell analysis. (a-d) From left to right: Name of the module and schematic representation of the experiment; iSBH-sgRNAs and inducers present in the experiment; Flow cytometry scatter plot showing ECFP expression against EYFP expression. Inset gives ECFP (blue) and EYFP (green) median fluorescence in the ECFP^{+ve}/EYFP^{+ve} cell populations; Fraction of cells expressing EYFP only, ECFP only, or both amongst the population of cells expressing at least one of the two fluorophores.

Chapter 6 – Control of endogenous gene targets & iSBHfold web tool

Proof-of-principle, optimization, and generalization of the iSBH methodology have all been demonstrated through the control of exogenous reporter genes. In all experiment, dCas9-VP64 was conditionally directed to the operator region of a synthetic promoter to drive the expression of a downstream fluorescent protein. While these results clearly demonstrate the successful implementation of inducible CRISPR-TRs using switchable sgRNAs, the regulators created so far might be of restricted used to the scientific community. In fact, others have shown that first generation CRISPR-TRs, such as dCas9-VP64, were quite limited in their capacity to upregulate the transcriptional output of endogenous target genes (Gilbert et al. 2013). Several studies have notably reported that tiling of multiple sgRNAs around the TSS of endogenous genes, and thus co-localization of multiple copies of dCas9-VP64, was required to achieve significant enrichment in transcript production (Gilbert et al. 2013; Mali, Aach, et al. 2013; A. W. Cheng et al. 2013; Maeder et al. 2013). While the expression of iSBH-sgRNAs can be multiplexed (chapter 5), the scientific community has, during the course of this project, moved away from dCas9-VP64 to use second generation CRISPR-TR solutions enabling potent activation of endogenous targets with a single sgRNA (Chavez et al. 2016).

The aim of the work presented in this chapter was to favor the democratization of the iSBH platform among the scientific community. To that end, I first sought to demonstrate that the iSBH methodology is also compatible with second generation CRISPR-TRs. After showing that

the approach can easily be used to control the state of the art synergistic activator modulator (SAM) system, I employ the resulting iSBH-SAM to assemble both branching and orthogonal gene modules on endogenous gene targets. Additionally, I showed in chapter 5 that the optimized designs derived in chapter 3 and 4, could readily be applied to rapidly construct new iSBH-sgRNAs, without the need for further optimization. Accordingly, I present in the second part of this chapter a new web tool, which I developed to automate the design of protein- and ASO-responsive iSBH-sgRNAs.

6.1 – Orchestrating endogenous gene expression

6.1.1 – SBH effectively silences the SAM system.

CRISPR-TRs with enhanced potency have been engineered by increasing the number of effector domains tethered to the crRNP (see section 1.5.2). A comparative analysis of these methods (Chavez et al. 2016) has shown that the SAM system (Konermann et al. 2014) outperforms other solutions when tested across a range of targets and cell lines. As explained in the introduction, this system employs modified sgRNAs featuring two MS2 stem-loops (sgRNA-2xMS2), which can each recruit a dimer of the MCP binding partner (sgRNA-2xMS2 Fig. 6.1a) (Konermann et al. 2014). By fusing MCP with the effector domains p65 and HSF1, the authors were able to create a crRNP coated with one copy of VP64, and 4 copies of MCP-p65-HSF1 (Konermann et al. 2014).

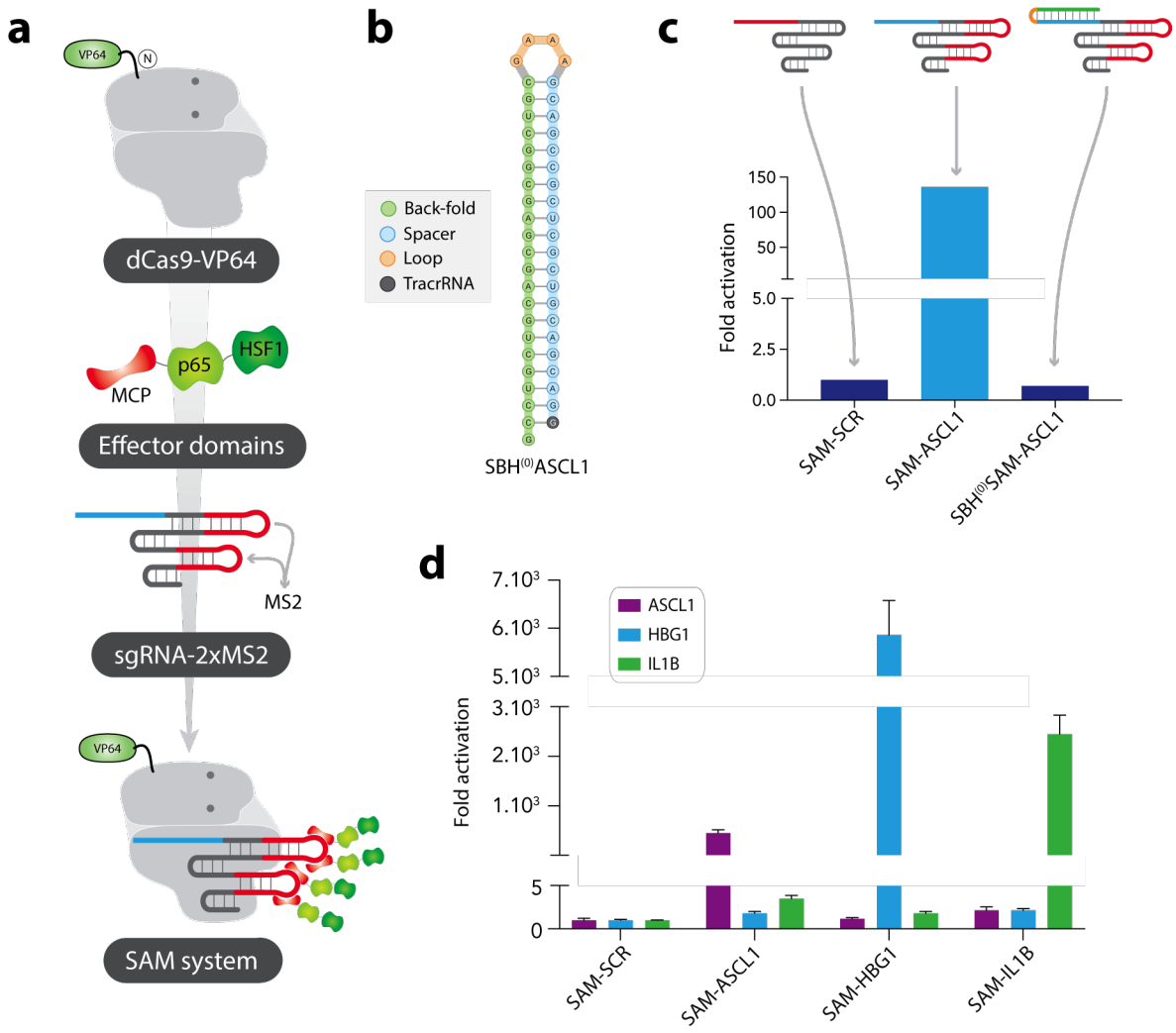

Figure 6.1 – SBH silences SAM-mediated transcriptional activation. (a) Schematic representation of the SAM system and its components. **(b)** Sequence and RNA secondary structure of a SBH designed for the *ASCL1* spacer (SBH$^{(0)}$ASCL1, 0 free spacer nucleotides). **(c)** HEK293-T cells were transfected with dCas9-VP64, MCP-p65-HSF1, and one of the following sgRNA-2xMS2: (i) SAM-sgRNA with scramble spacer sequence (SAM-SCR); (ii) SAM-sgRNA targeting *ASCL1* (SAM-ASCL1); (iii) SAM-sgRNA targeting *ASCL1* with the SBH shown in (b) (SBH$^{(0)}$SAM-ASCL1). RT-qPCR was conducted 48h post-transfection using a *ASCL1* specific primer pair. Fold-changes in transcript levels are given relative to the SAM-SCR condition (n = 1). **(d)** HEK293-T were transfected with SAM system components and a SAM-sgRNA targeting either nothing (SAM-SCR), *ASLC1* (SAM-ASCL1), *HBG1* (SAM-HBG1), or *IL1B* (SAM-IL1B). Transcript levels of all three genes were assayed by RT-qPCR for each condition. Bar plot shows fold-change in transcript levels normalized to the SAM-SCR condition (n = 3, mean +/- s.d.).

Since iSBH-based inducibility is based on minimal 5' modification of the sgRNA, I reasoned that the method should in theory be compatible with all CRISPR derivatives. To test if the SAM system could be silenced using iSBH, I modified one of the published sgRNA-2xMS2 to accommodate a 20nt SBH, sequestering the spacer for the *ASCL1* gene[18] (Fig. 6.1b). HEK293-T cells were co-transfected with pDNAs expressing dCas9-VP64, MCP-p65-HSF1, and either a scramble sgRNA-2xMS2 (SAM-SCR), an *ASL1* targeting sgRNA-2xMS2 (SAM-ASCL1), or the *ASCL1* guide fitted with an SBH ($SBH^{(0)}$SAM-ASCL1). RT-qPCR analysis of *ASCL1* mRNA levels 48h post-transfection revealed that the SAM CRISPR-TR was able to upregulate gene expression of its target by more than a 100-fold when the SAM-ASCL1 sgRNA was transfected (Fig. 6.1c). More importantly, $SBH^{(0)}$SAM-ASL1 was found to completely silenced SAM activity with transcript levels identical to what was observed when the spacer of the sgRNA-2xMS2 was scrambled to a random sequence (SAM-SCR) (Fig. 6.1c).

6.1.2 – protein-responsive iSBH-SAM.

To select a pair of potent SAM-sgRNAs for the assembly of branching and orthogonal gene modules, the SAM-ASCL1 guide and two additional SAM-sgRNAs targeting the *HBG1* and *IL1B* genes were compared for their ability to drive CRISPRa. HBG1 encodes the hemoglobin subunit gamma 1 (NCBI Gene ID: 3047), and IL1B codes for a protein member of the interleukin 1 cytokine family (NCBI Gene ID: 3553). The three guides were transfected separately along with the SAM system components (dCas9-VP64, MCP-p65-HSF1) and *ASCL1*, *HBG1*, and *IL1B* levels were assayed by RT-qPCR across all conditions (Fig. 6.1d). Based on the results, the SAM-HBG1 and SAM-IL1B guides were chosen for the implementation of gene circuits because they displayed the largest fold-changes in transcript levels compared to the negative control condition (SAM-SCR) and did not show any noticeable crosstalk with each other.

[18] Activation of the *ASCL1* gene in HEK293-T cells has been used across the literature to benchmark CRISPRa systems. ASCL1 encodes the achaete-scute family bHLH transcription factor 1 (NCBI Gene ID:429).

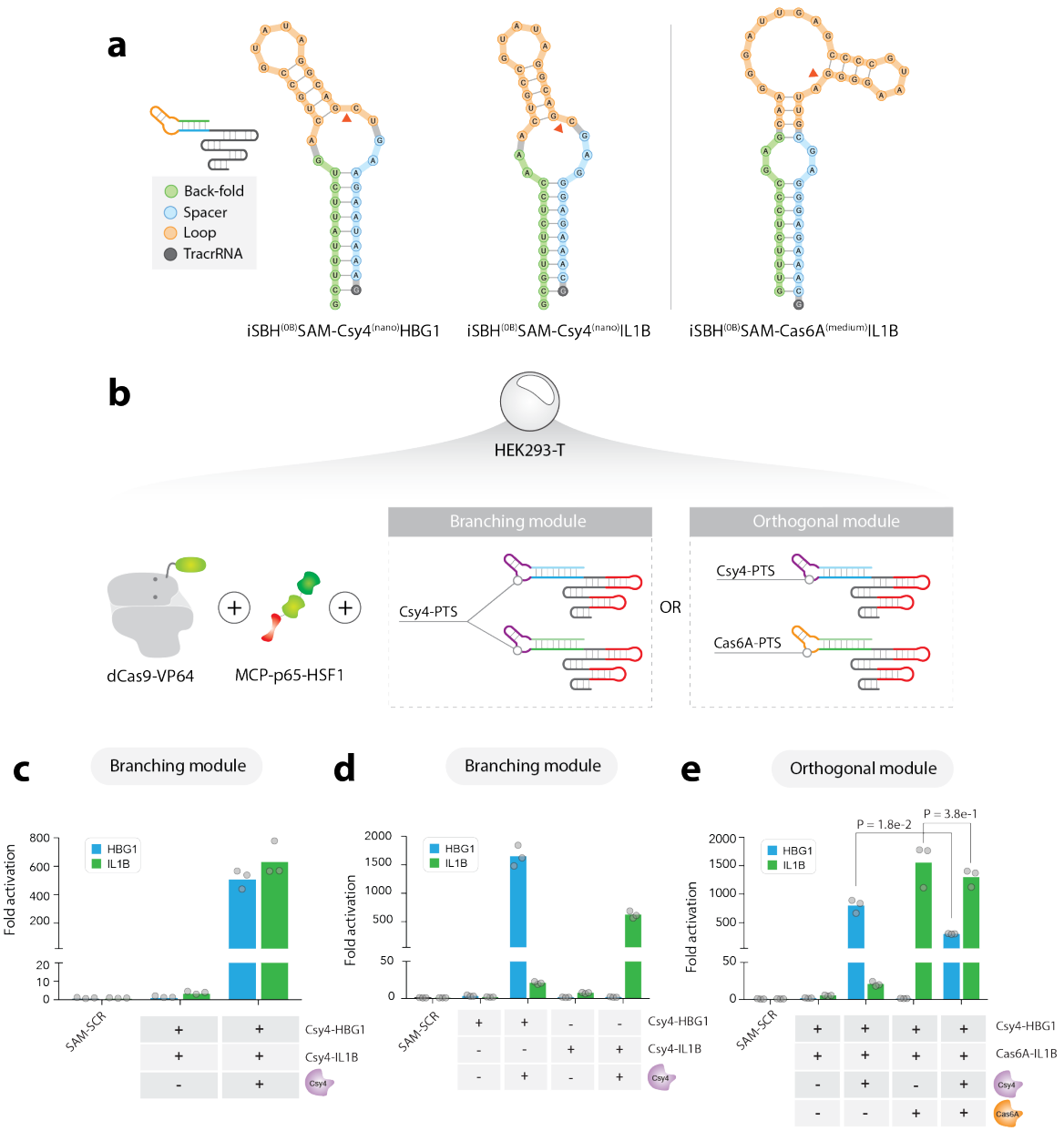

Figure 6.2 – Control of endogenous genes with protein-responsive iSBH-SAM. (a) Sequence and RNA secondary structure of Csy4- and Cas6A-responsive hairpins for the *HBG1* and *IL1B* spacers. **(b)** Transfection scheme for the implementation of branching and orthogonal gene modules. The branching module uses two iSBH-SAM sgRNAs with spacer for *HBG1* (blue) and *IL1B* (green) respectively, and the same Csy4-protein target site (PTS). For the orthogonal module, Cas6A-PTS is used in place of Csy4-PTS on the *IL1B* guide. **(c)** Branching module was implemented in HEK293-T as shown in (b). RT-qPCR analysis on both *HBG1* and *IL1B* transcript levels was conducted 48H post-transfection in the presence or absence of Csy4. Graph shows fold activation compared to a condition where only a SAM-sgRNA with scramble spacer (SAM-SCR) is transfected (n = 3, bars show mean, t-test p-values are displayed). **(d)** Similar to (c) but iSBH SAM-sgRNAs are being transfected in isolation. **(e)** Orthogonal module was implemented in HEK293-T as shown in (b). RT-qPCR analysis on both *HBG1* and *IL1B* transcript levels was conducted in the presence or absence of Csy4, Cas6A, or a combination of the two.

The SAM-sgRNAs were then modified to create a branching gene module whereby Csy4 triggers concomitant overexpression of *HBG1* and *IL1B*. This was achieved by generating *nano* Csy4-iSBHs (see section 3.4) for both *HBG1* and *IL1B* spacers (Fig 6.2a). The resulting iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$HBG1 and iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$IL1B sgRNAs were transfected in HEK293-T cells along with the SAM system components and either a decoy plasmid or a construct expressing Csy4 (Fig. 6.2b). As expected, RT-qPCR analysis conducted 48h post-transfection revealed that both gene targets remained expressed at background levels unless the protein inducer was expressed (Fig. 6.2c). In contrast, both genes were found upregulated in the presence of the protein inducer (Fig. 6.2c). To demonstrate that Csy4 was driving *HBG1* and *IL1B* overexpression via activation of the corresponding iSBH-sgRNAs, the experiment was repeated while transfecting only one guide at a time (Fig. 6.2d). To my surprise, *IL1B* was upregulated by ~20-fold in the presence of Csy4, even when iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$IL1B was not transfected (Fig. 6.2d). This suggests that this protein inducer might have an indirect effect on *IL1B*, perhaps associated with its role in immunity (Van Damme et al. 1985)[19]. Nevertheless, this effect was negligible compared to the >600 fold increase observed when iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$IL1B sgRNA and Csy4 were both co-transfected (Fig. 6.2d). Taken together, these results demonstrate that iSBH can easily be combined with the SAM system to condition the activation of multiple endogenous gene targets on the expression of a Cas endoribonuclease.

Next, the iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$IL1B-sgRNA was reprogrammed to sense Cas6A instead of Csy4. Based on insights gained in chapter 5 on the optimization of Cas6-responsive hairpins, the Cas6-PTS was fused to the seed central bulge of SBH$^{(0B)}$IL1B to create iSBH$^{(0B)}$SAM-Cas6$^{(medium)}$IL1B-sgRNA (Fig. 6.2a). Using this sgRNA along with iSBH$^{(0B)}$SAM-Csy4$^{(nano)}$HBG1, an orthogonal gene module was assembled to achieve independent control of both endogenous genes by distinct Cas endoribonucleases (Fig. 6.2b, e). HEK293-T cells were transfected with the SAM system components, both iSBH-sgRNAs, and each or both

[19] Note that IL1B background levels are extremely close to the limit of detection of our qPCR machine. Also, the variation observed here might be technical noise.

protein inducers (Fig. 6.2b, e). As expected, *HBG1* and *IL1B* were significantly upregulated at ~800 fold and ~1500 fold above background levels in the presence of Csy4 and Cas6A, respectively (Fig. 6.2e). Additionally, co-delivery of both protein inducers led to the simultaneous activation of both targets (~300 fold and ~1300 fold), albeit to a lower extent than what was observed when Csy4 or Cas6A were delivered in isolation (Fig. 6.2e). Matching the result reported above, *IL1B* transcript levels were found to be ~25 fold above background levels in the presence of Csy4, despite the absence of Cas6 (Fig. 6.2e). Since this effect has been previously observed in the absence of any *IL1B* sgRNA (Fig. 6.2d), and results from chapter 5 have shown perfect orthogonality between protein inducers, it is quite unlikely that the observed increase in *IL1B* levels was due to a cross-talk between inducer:target pairs.

6.1.2 – ASO-responsive iSBH-SAM

Branching and orthogonal gene modules were also implemented using ASO-responsive SAM-sgRNAs. iSBHs for the branching module were generated by submitting *HBG1* and *IL1B* spacer sequences to the genetic algorithm presented in chapter 5 (see section 5.2.2). The top ranking ASL outputted by this algorithm was then used to create iSBH$^{(0B)}$SAM-ASOε-HBG1 and iSBH$^{(0B)}$SAM-ASOε-IL1B sgRNAs designed to respond to a sequence complementary inducer ASOε (Fig. 6.3a). The two guides were transfected along with the SAM system components and RT-qPCR was used to assay the response of this circuit to the delivery, 24h post-transfection, of either the cognate ASOε or a decoy trigger (Fig. 6.3b, c). As expected, delivery of ASOε led to the upregulation of both target genes by ~80 fold and ~100 fold compared to baseline levels, an effect which was not observed with the decoy ASO (Fig. 6.3c). Consistent with what was reported on transgenes (see chapter 4), ASO-mediated induction of iSBH-sgRNA yielded weaker gene upregulation than what was observed with protein inducers. To confirm that ASOε did not cause activation on its own, the ON-state experiment was repeated while mutating the ASL of both guides. In this case, no overexpression of the gene targets was observed, demonstrating the efficient decoupling between the presence of the inducer and CRISPRa (Fig. 6.3c).

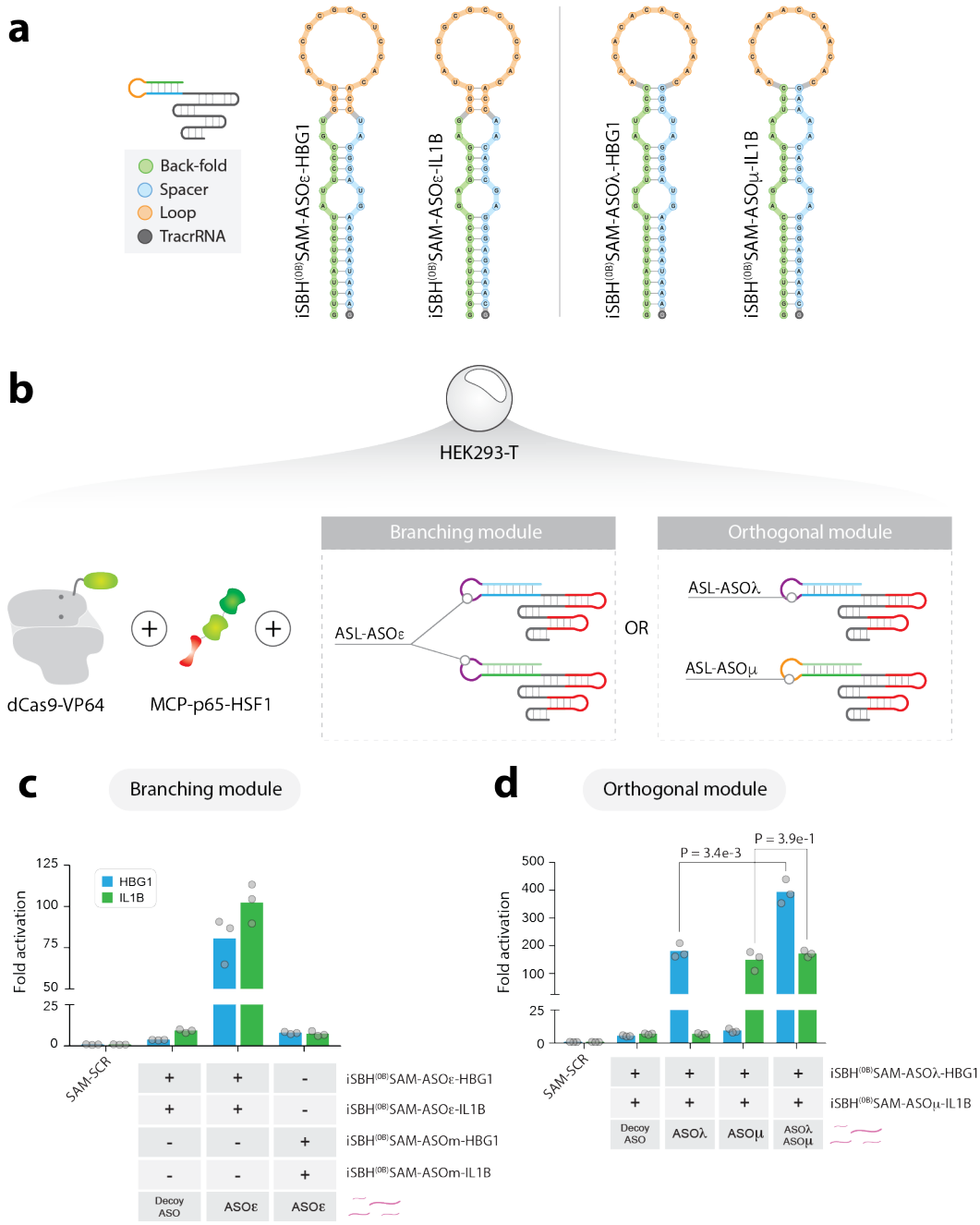

Figure 6.3 – Control of endogenous genes with ASO-responsive iSBH-SAM. (a) Sequence and RNA secondary structure of ASO-responsive hairpins for the *HBG1* and *IL1B* spacers. **(b)** Transfection scheme for the implementation of branching and orthogonal gene modules. Two iSBH-SAM sgRNAs, targeting *HBG1* (blue) and *IL1B* (green), either sense the same ASOε or distinct ASOλ and ASOμ triggers. **(c)** Branching module was implemented in HEK293-T as shown in (b). RT-qPCR analysis on both *HBG1* and *IL1B* transcript levels was conducted 48H post-transfection in the presence of a decoy ASO or ASOε. Graph shows fold activation for each gene compared to SAM-SCR (n = 3, bars show mean). **(d)** Orthogonal module was implemented in HEK293-T as shown in (b). RT-qPCR was conducted 48H post-transfection in the presence of a decoy inducer, ASOλ, ASOμ, or a combination of ASOλ and ASOμ. Graph shows fold activation compared to SAM-SCR (n = 3, bars show mean, t-test p-values are displayed).

Finally, the ASLs on both guides were replaced with new ones to create iSBH$^{(0B)}$SAM-ASOλ-HBG1 and iSBH$^{(0B)}$SAM-ASOμ-IL1B sgRNAs, programmed to respond to ASOλ and ASOμ, respectively (Fig. 6.3a). Transfection of both guides in HEK293-T cells (Fig. 6.3b) revealed successful coupling between each ASO trigger and the expression of its corresponding gene target (Fig. 6.3d). Notably, each branch could be activated independently without apparent crosstalk, suggesting perfect orthogonality between the iSBH-sgRNAs.

Of note, the ON-state HBG1 levels reported in figure 6.2e and 6.3d were found to be significantly affected by the concomitant activation of IL1B. Indeed, in the orthogonal module based on protein-responsive iSBHs, HBG1 levels were reduced by 60% when both Csy4 and Cas6A inducers were co-transfected versus Csy4 alone (Fig. 6.2e, p-value = 1.8e-2). On the other hand, an opposite trend was observed in the orthogonal module based on ASO-responsive iSBHs. HBG1 expression was increased by 2.2-fold when both ASO triggers (ASOλ and ASOμ) were delivered compared to ASOλ delivery alone (Fig. 6.3d, p-vlaue = 3.4e-3). The bi-directionality of the variation suggests that the observed effects are unlikely to be due to first or second order protein interactions between HBG1 and IL1B. In fact, no protein-protein interactions between HBG1 and IL1B have been catalogued in either the BioGRID (Chatr-Aryamontri et al. 2017) nor the String (Jensen et al. 2009) database. Further experiments involving different target genes will be required to determine if the observed effects are imputable to interactions between inducers.

6.2 – Introducing the iSBHfold web tool

In chapters 3, 4, and 5 of this thesis, I reported on the development and optimization of both protein- and ASO-responsive iSBH-sgRNAs. Out of several Csy4-iSBHs, the best ON/OFF characteristics were obtained when fusing the Csy4-PTS to the seed central bulge of SBH$^{(0B)}$CTS (*nano* design, see section 3.4). Subsequently, iSBH-sgRNAs were also designed to respond to the Cas6A endoribonuclease. Again, direct comparison of several candidate designs revealed that fusing the Cas6A-PTS to the distal bulge of SBH$^{(0B)}$CTS was the most potent solution (*medium* design, see section 5.3.1). Regarding the design of ASO-responsive hairpins, the full length SBH$^{(0B)}$CTS, modified to accommodate a 14-nt ASL, displayed the strongest response when activated with a 20nt ASO bearing sequence complementarity to the sensing-loop plus three flanking nucleotides on both sides (see section 4.2).

In chapters 5 and 6, these design principles were successfully applied to create multiple Csy4- and Cas6A-responsive guides as well as ASO-iSBHs across a large range of spacer sequences. The fact that all resulting iSBH-sgRNAs behaved as expected in both OFF-state, (i.e. full CRISPR-TR silencing) and ON-state (i.e inducer specific activation) without any need for further optimization, suggested that the design of new iSBHs could be automated.

Accordingly, I have developed a program called iSBHfold and made it freely available online[20]. This web tool was created to assist researchers in the creation of protein- and ASO-responsive iSBH-sgRNAs (Fig. 6.4). The core of this program was written using JAVA and handles evolution of ASLs and iSBH design in general. iSBHfold also features a custom-made algorithm (R script) created to render RNA secondary structures in scalable vector graphics (SVG) format. This code has notably been used to generate all RNA secondary structures presented in this thesis. Finally, HTML, CSS, JAVASCRIPT, and AJAX programming languages were used to build the iSBHfold user interface.

[20] iSBHfold can be found at http://apps.molbiol.ox.ac.uk/iSBHfold/cgi-bin/iSBHfold.cgi

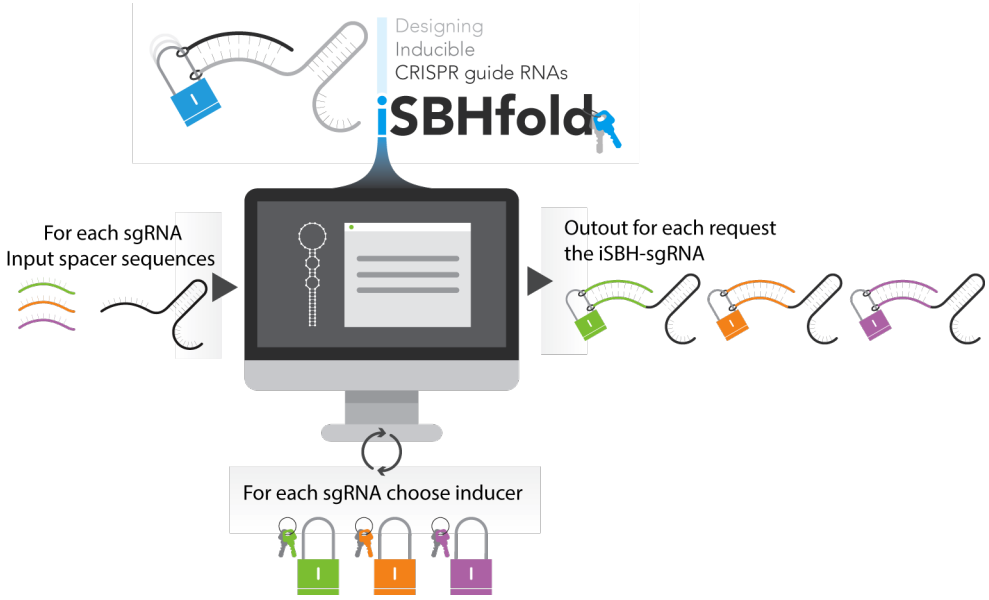

Figure 6.4 – iSBHfold webtool, automated iSBH-sgRNA design. Accessible online, this web application assists researchers in the creation of protein- or ASO-responsive iSBH-sgRNAs. The system takes as input a list of spacer sequences, asks the user to decide on a type of inducer for each guide, and automatically designs the corresponding iSBH-sgRNAs following the design rationales established in chapter 2, 4 and 5.

The application was divided in two main parts hosted on two different web pages and dedicated to the design of protein- and ASO-responsive hairpins, respectively. In both cases, the system first prompts its user to provide spacer sequences for the sgRNAs of interest (Fig. 6.5a, 6.6a). If designing protein-responsive guides, the user will then be given the opportunity to choose an inducer for each sgRNA (Fig. 6.5a). While this choice is currently restricted to the Csy4 and Cas6A endoribonucleases, additional protein triggers will be added to the system once shown to work as iSBH inducers. For ASO-iSBH designs, the user can choose among up to 10 different ASO triggers, whose sequences will be automatically derived by the genetic algorithm in function of the spacers inputted to ensure optimal iSBH folding (Fig. 6.6a). For example, if the experimental setup demands to wire activation of *Gene 1* and *Gene 2* on the presence of ASOx, while controlling transcription of *Gene 3* and *Gene 4* with ASOy, the user is asked to perform the following tasks: (i) create four iSBH-sgRNA entries; (ii) enter a 18 to 20nt spacer

sequence of for each sgRNA (spacer for *Gene1, Gene 2, Gene 3, Gene 4*); (iii) choose "ASO1" as inducer for the first two sgRNAs and "ASO2" for the others. Based on these parameters, iSBHfold will design two distinct ASLs, one for iSBH-sgRNA 1 and 2, and another one for iSBH-sgRNA 3 and 4.

Once all iSBH-sgRNA requests have been validated by the user and submitted (Fig. 6.5b 6.6b), the information is sent to the server where the core program will create the corresponding iSBHs. The process of generating new iSBHs involves evolution of certain RNA segments in order to guarantee that the resulting hairpin folds as intended. In the case of Csy4- and Cas6A-iSBHs, the sequence of the back-fold segment is not only optimized to ensure formation of the bulges along the stem, but also to guarantee that the PTS stem is flanked by ssRNA. Regarding the design of ASO-responsive sgRNAs, iSBHfold implements the genetic algorithm covered in Chapter 5 (see section 5.2.2), which allows users to create multiple iSBH-sgRNAs sharing the same ASL, and consequently respond to the same ASO trigger. In this instance, once the computation completed, the system outputs a list of potential candidate ASLs (Fig. 6.6c). The user-selected ASL is then employed to finalize ASO-iSBH sequences.

The results are then sent back to the user via the web interface, which will now display an entry for each hairpin with the following details (Fig. 6.5c, 6.6c): (i) name/sequence of the spacer and inducer as specified by the user; (ii) the computed iSBH sequence colored-coded to indicate back-fold, sensing-loop, and spacer segments; (iii) the DNA sequence of oligonucleotides that the user should use to clone the iSBH[21]; (iv) RNA secondary structure (bracket and dot notation) predicted from the output sequence and optimal target structure that the iSBH should match; (v) a downloadable rendering of the iSBH RNA secondary structure in .SVG format. Finally, all the results, including secondary structures and primer sequences for each iSBH, can be exported on the user computer (Fig. 6.5d, 6.6d).

[21] Note that the system has predefined overhangs which are compatible with the U6-sgRNA cassette used in this work (see methods)

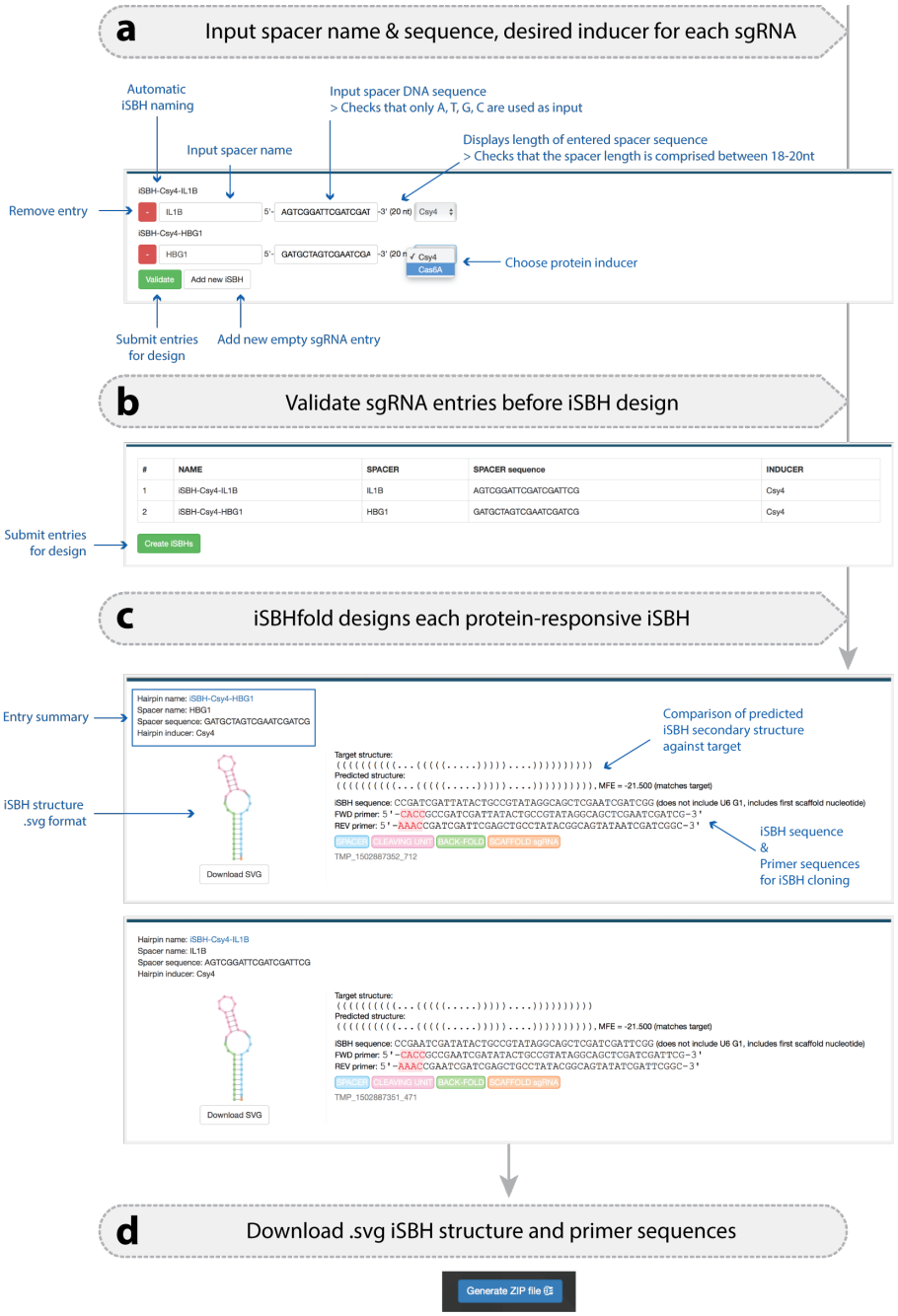

Figure 6.5 – iSBHfold: design of protein-responsive hairpins. Step by step overview of the user interface. **(a)** Input panel: iSBH entries can be created and deleted. User enters name and sequence for the spacer, and chooses the inducer (Csy4 or Cas6A). Each entry is given a name automatically derived from the spacer name and the inducer chosen. **(b)** Validation panel summarizes the iSBH requested by the user. After reviewing the information, the operator can either go back and modify the input panel or process and "create iSBHs". **(c)** The core program evolves the optimal sequence for each iSBH and returns the following for each iSBH-sgRNA: (i) Summary of the request; (ii) RNA secondary structure prediction along with downloadable SVG rendering; (iii) iSBH sequence and oligonucleotides required for cloning. **(d)** All information in (c) can be downloaded.

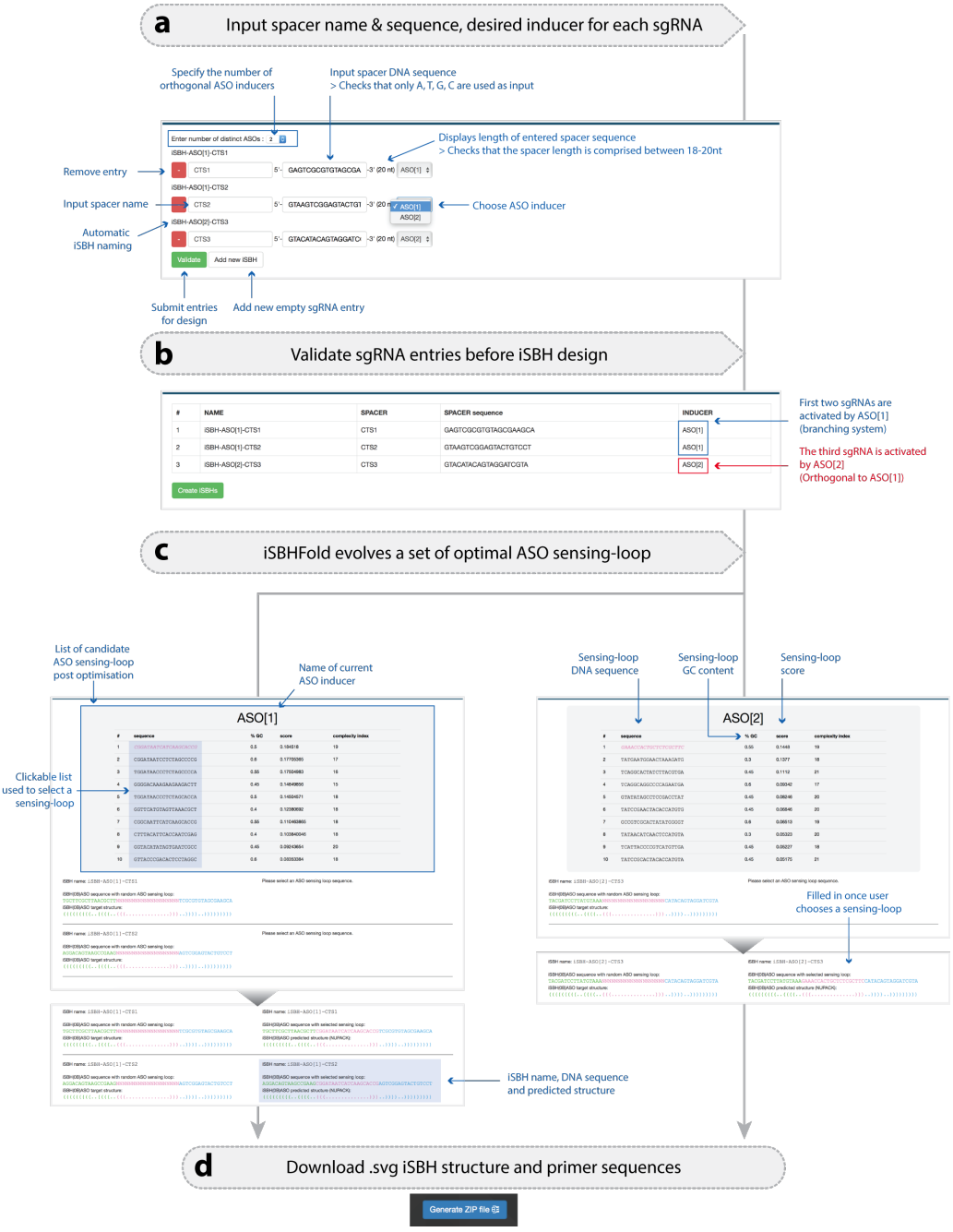

Figure 6.6 – iSBHfold: design of ASO-responsive hairpins. Step by step overview of the user interface. **(a)** Input panel: User enters name/sequence for the spacer, and choose the inducer (ASO_1, …, ASO_{10}). Each entry is automatically named. **(b)** Validation panel summarizes the requested iSBHs for the user to verify before proceeding to "create iSBHs". **(c)** The core program evolves a list of optimal ASO sensing-loops (ASLs) given the inputted spacers. Once the user chooses one ASL by clicking on it, results for each iSBH-sgRNA are displayed. **(d)** All information displayed in (c) as well as rendering of the RNA secondary structure of each ASO-iSBH can be downloaded.

6.3 – Closing remarks

This chapter has been dedicated to the democratization of the iSBH platform. To that end, it was demonstrated that the method can be combined with the SAM system to control the transcriptional output of human genes in their endogenous context. Additionally, a web tool that automates iSBH-sgRNA design was made available online.

It was also shown that using Csy4 and Cas6A or a pair of orthogonal ASOs, one can independently control multiple endogenous gene targets. An interesting extension of this work will be to demonstrate orthogonality between protein-responsive and ASO-responsive sgRNAs. The ability to use both classes of guides in the same cells may eventually enable the assembly of more complex synthetic gene networks, that could be used to restrict gene expression to both a given time point (ASO delivery) and a specific cell type (protein trigger under tissue-specific promoter).

Additionally, several new iSBH-sgRNAs were engineered by following design principles derived from previous work done using reporter transgenes. While the experiments presented above revealed that these guides behaved accordingly to their programming, one cannot rule out that further optimization of the iSBH in specific cases might yield stronger CRISPRa in the ON-state. Given that the thermodynamic stability of the stem is mostly decided by the spacer sequence of the guide, it is notably possible that a design found optimal for one spacer might not be best suited for another sequence. For example, spacers with a low GC content in their seed sequence might not be compatible with the *nano* Csy4-iSBH design, as the short stem might not be stable enough to fully silence the sgRNA in the OFF-state. Consequently, I plan to extend iSBHfold capabilities in order to allow the creation of *full*, *medium*, and *nano* iSBH for both Csy4 and Cas6A responsive hairpins[22].

[22] Current version only lets its user choose between Csy4 *nano* or Cas6A *medium* which were found optimal in this work.

Finally, while the current version of iSBHfold allows the creation of multiple sets of ASO-iSBHs programmed to sense distinct ASOs, the system does not take into account potential crosstalk between triggers (sequence overlap). Instead, the program deliberately allows the user to choose ASLs from the list of best candidates. Nevertheless, future versions of iSBHfold will supplement the current solution with an automated “suggested” ASL choice, which will take into account sequence similarities between supposedly orthogonal ASLs. Moreover, a function that can search the human transcriptome for potential off-targets of the proposed ASO triggers will be implemented in order to guarantee minimal perturbation of cellular homeostasis.

Chapter 7 – Going further: Aptazyme based iSBH-sgRNAs

iSBH-sgRNAs are constructed around the simple idea that a back-fold segment, sequestering the guide spacer, can be conditionally excised by replacing the loop of the spacer-blocking hairpin with inducer-dependent RNA cleaving units. So far, two types of RNA cleaving units have been characterized to create inducible sgRNAs responsive to genetically encoded (endoribonuclease) and externally delivered (ASO) triggers. Given that small molecules have been ubiquitously employed as triggers in the design of inducible systems (see section 1.6), I sought to establish if new iSBHs could also be raised against additional ligands, with a particular focus on the design of iSBH-sgRNAs responsive to small molecules. In this brief chapter, I present proof-of-principle data demonstrating the use of allosteric self-cleaving ribozymes (aptazymes) as RNA cleaving units for the design of iSBH-sgRNAs. This work has been motivated by the fact that aptazymes can, in theory, be evolved to sense and respond to a large range of ligands (Hermann & Patel 2000; Goler et al. 2014), and as such would considerably broaden the repertoire of iSBH inducers.

7.1 – Spacer release with hammerhead ribozymes

Ribozymes are short RNA structures with enzymatic properties, capable to catalyze intramolecular reactions (Ferré-D'Amaré & Scott 2010). The self-cleaving hammerhead ribozyme (HHRz) was first discovered by Prody et al. within the satellite RNA of the tobacco ringspot virus (Prody et al. 1986), and was subsequently identified as a potential regulator of

gene expression in other species (Birikh et al. 1997; Hammann et al. 2012). After the initial isolation of a minimal sequence which proved to be functional *in vitro* (Haseloff & Gerlach 1989; Hammann & Lilley 2002), other naturally occurring full-length HHRzs were characterized (Hammann et al. 2012). These variants displayed enhanced self-cleavage activity (1000-fold greater than their minimal counterpart) and were capable to function in low, physiological magnesium-concentrations (Canny et al. 2004; Khvorova et al. 2003). Subsequently, allosteric HHRzs (aHHRz), or aptazymes, were engineered by fusing ligand-binding RNA aptamers to the HHRz core sequence, so as to enable ligand-dependent activation (Soukup & Breaker 1999; J. Tang & Breaker 1997). Under this premise, the binding energy between ligand and aptamer was harvested to stabilize the linked ribozyme in a preferred active or inactive conformation, thus creating an RNA biosensor whose cleavage was conditioned on the presence of a specific ligand. As such, aHHRz have been extensively used to in synthetic biology to create post-transcriptional switches by incorporating these versatile regulators into the 5' or 3'-UTR of mRNAs of interest (Yen et al. 2004; Win & Smolke 2007; Win & Smolke 2008; Ausländer et al. 2010). Additionally, the availability of methodologies such as SELEX[23] (Ellington & Szostak 1990; Tuerk & Gold 1990), allowing to evolve new aptamers against a wide range of ligands, as yielded a set of apazymes responsive to small molecules, peptides and proteins (Hermann & Patel 2000; Goler et al. 2014).

Owing to its versatility, small size, self-sufficiency, and fast kinetics, I sought to leverage the sensor/actuator capabilities of aHHRz to conditionally activate iSBH-sgRNAs. I reasoned that, by replacing the default GAAA loop of SBH$^{(0B)}$CTS with an aptazyme, one could in theory control the excision of the back-fold in a ligand-dependent fashion (Fig. 7.1). To test the feasibility of this idea, I first sought to demonstrate that a plain HHRz (not allosteric) could be used as RNA cleaving unit to release the sgRNA from SBH silencing.

[23] Systematic evolution of ligands by exponential enrichment

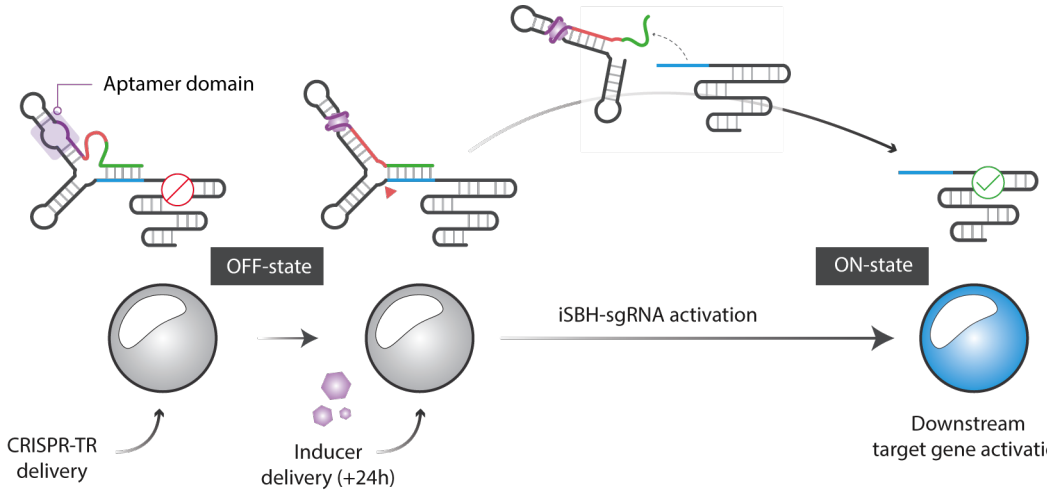

Figure 7.1 – Aptazyme based iSBH-sgRNAs. Schematic representation of iSBH-sgRNA built using allosteric self-cleaving hammerhead ribozymes (aHHRz). The aHHRz, featuring a ligand-binding aptamer domain, is used in place of the spacer-blocking hairpin (SBH) loop. In the absence of the ligand, the aHHRz is catalytically inactive and the sgRNA is silenced by the SBH (OFF-state). Once delivered, ligand binding to the aptamer triggers a conformational change which resumes the catalytic activity of the ribozyme. aHHRz self-cleavage releases a functional sgRNA that directs the CRISPR-effector to the gene of interest (ON-state).

The full-length type-I *Schistosoma mansoni* hammerhead ribozyme consists of a catalytic core flanked by three helices referred to as Stem-I, II, and III, respectively (Martick & Scott 2006) (Fig. 7.2a). While sequence conservation attests to the importance of the core nucleotides, crystallographic analysis have revealed that tertiary interactions between stem-I and stem-II bulges, are responsible for the enhanced catalytic properties of this ribozyme (Martick & Scott 2006). In order to minimize the residual sequence post-cleavage, the type-I *Schistosoma mansoni* HHRz was grafted onto SBH$^{(0B)}$CTS1 via Stem-I (Fig. 7.2a). This way, only 11 nucleotides from the ribozyme sequence should be left on the 5'-end of the sgRNA, once activated. The resulting SBH$^{(0B)}$HHRz-CTS1 sgRNA was transfected in HEK293-T cells, along with reporter plasmid 8xCTS1-EYFP-pA and dCas9-VP64. Flow cytometry analysis 48h post-transfection revealed a strong CRISPRa of the reporter with up to 25% of transfected cells being activated (Fig. 7.2b). To show that the observed EYFP expression was in fact due to HHRz self-cleavage rather than mis-folding of the SBH structure, a mutant hairpin was created

(SBH$^{(0B)}$mHHRz-CTS1), which harbored an adenine to guanine point mutation in the HHRz core (Fig. 7.2a). This specific mutation had been previously shown to abolish HHRz enzymatic property while maintaining its secondary and tertiary RNA structure (Martick & Scott 2006). Confirming that HHRz self-cleavage was indeed responsible for sgRNA activation, Transfection of the SBH$^{(0B)}$mHHRz-CTS1-sgRNA, in place of the catalytically active SBH$^{(0B)}$HHRz-CTS1-sgRNA, did not yield any reporter expression above background levels (Fig. 7.2b). Additionally, this result demonstrated that HHRz grafting was compatible with SBH-based silencing of the guide.

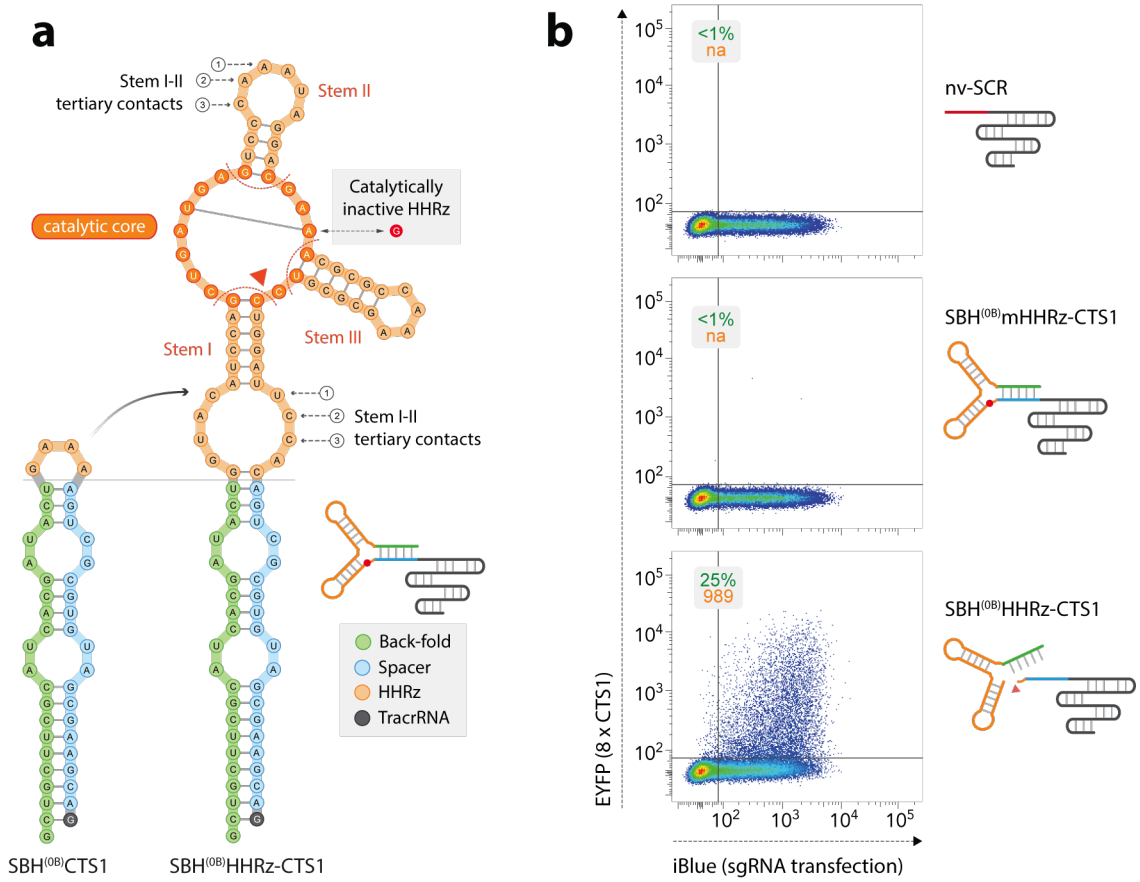

Figure 7.2– Self-cleaving HHRz as RNA cleaving unit for iSBH design. (a) Sequence and RNA secondary structure of iSBH$^{(0B)}$HHRz-CTS1. The hairpin features a type-I *Schistosoma mansoni* hammerhead ribozyme in place of the GAAA loop of SBH$^{(0B)}$CTS1. The HHRz comprises a catalytic core and three stem-loops (I, II, and III). Tertiary contacts between stem I and stem II bulges increases the enzymatic kinetic of the ribozyme. An A to G point mutation in the catalytic core abrogates HHRz self-cleavage. **(b)** HEK293-T cells were transfected with dCas9-VP64, reporter construct and a SBH-sgRNA featuring either a catalytically active (SBH$^{(0B)}$HHRz-CTS1), or inactive (SBH$^{(0B)}$mHHRz-CTS1) HHRz. Additionally, a native sgRNA with scramble spacer (nv-SCR) was used as negative control.

7.2 – Small molecule-responsive iSBH-sgRNA

Next I searched for small molecule aptamers that could be adapted onto the SBH$^{(0B)}$HHRz-CTS design to make it inducible. Amongst several candidates, the RNA aptamers evolved by Zimmermann et al. to bind the drug theophylline (Zimmermann et al. 2000) have been extensively used to create allosteric versions of the type I *Schistosoma mansoni* HHRz (Yen et al. 2004; Win & Smolke 2007; Win & Smolke 2008; Berschneider et al. 2009; Ausländer et al. 2010). Notably, Saragliadis et al. have evolved a theophylline aHHRz to control tRNA availability *in vivo* (Berschneider et al. 2009; Saragliadis & Hartig 2013). In this study, the authors engineered switchable tRNAs by modifying their native sequence to include an aHHRz, which upon activation cleaves off an RNA segment designed to interfere with tRNA folding via base-pairing. Since this framework was somehow reminiscent of the conditional sgRNA spacer release system, I engineered a theophylline-responsive iSBH by grafting the delta-33 theophylline aptamer (Zimmermann et al. 2000) to the HHRz stem III using their published optimized linker (Saragliadis & Hartig 2013) (Fig. 7.3).

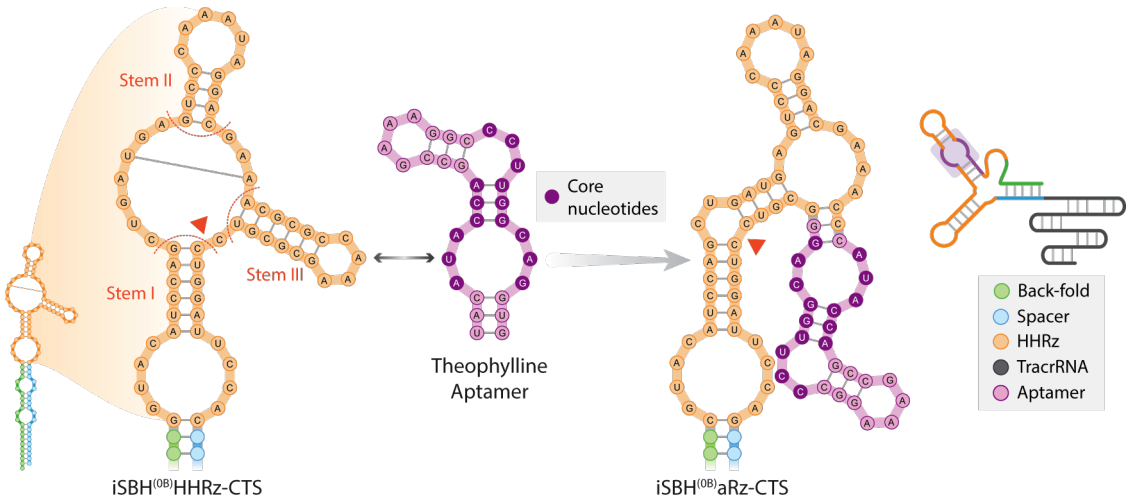

Figure 7.3– Design of a theophylline-responsive iSBH. The iSBH$^{(0B)}$HHRz-CTS hairpin (left) featuring a catalytically active HHRz is modified to achieve ligand-dependent back-fold excision. A theophylline binding aptamer (center) is grafted onto stem-III of the Type-I *Schistosoma mansoni* HHRz via an optimized linker sequence such that HHRz would adopt an inactive conformation (right). Upon theophylline delivery, interactions between ligand and aptamer cause a conformational change which restores HHRz's catalytic activity.

The resulting iSBH-sgRNA (iSBH$^{(0B)}$-aRz-CTS2) was transfected in HEK293-T cells along with the reporter construct and dCas9-VP64 to test its ON/OFF characteristics. in this case, the transfection media was replaced 24h post-transfection with either plain culture media (OFF-state) or media containing theophylline at a concentration of 1mM[24] (ON-state). This analysis revealed significant reporter expression in the absence of the drug with 12% of activated cells, suggesting that the aHHRz had a non-negligible catalytic activity in the OFF-state (leakiness) (Fig. 7.4a). Nevertheless, adjunction of theophylline led to a 3-fold increase in CRISPRa, as measured by activation score, with a notable increase in the number of activated cells (Fig. 7.4a, b, e). Similar results were also obtained when repeating the experiments with a different spacer:reporter pair (CTS3). In this case, a 2-fold increase in activation score was observed between theophylline positive and negative conditions (Fig. 7.4c, d, e).

Interestingly, experiments conducted to test the effect of the drug on the normal CRISPR-TR system (native sgRNA, dCas9-VP64), revealed that reporter expression was consistently lower in the theophylline positive condition, compared to the plain media control (flow cytometry analysis, data not shown). To assess if this effect was due to a reduction in CRISPRa rather than protein synthesis, the experiment was repeated and reporter mRNA levels were quantified by RT-qPCR 48h post-transcription (Fig. 7.4f). The results showed a 36% reduction in ECFP mRNA in the theophylline positive condition, confirming the observations made via flow cytometry analysis. While this finding seems to point towards the fact that theophylline represses CRISPP-TR activity, further experiments will be required to rule out post-transcriptional effects. Despite the effect of the drug on CRISPRa, I found that reporter levels were increased by 1.5-fold between OFF- and ON-state when using the theophylline-responsive iSBH-sgRNA targeting the ECFP cassette (Fig. 7.4f).

[24] Concentration commonly used for aHHRz experiments in HEK293-T cells (Win & Smolke 2007; Win & Smolke 2008; Ausländer et al. 2010; Nomura et al. 2013).

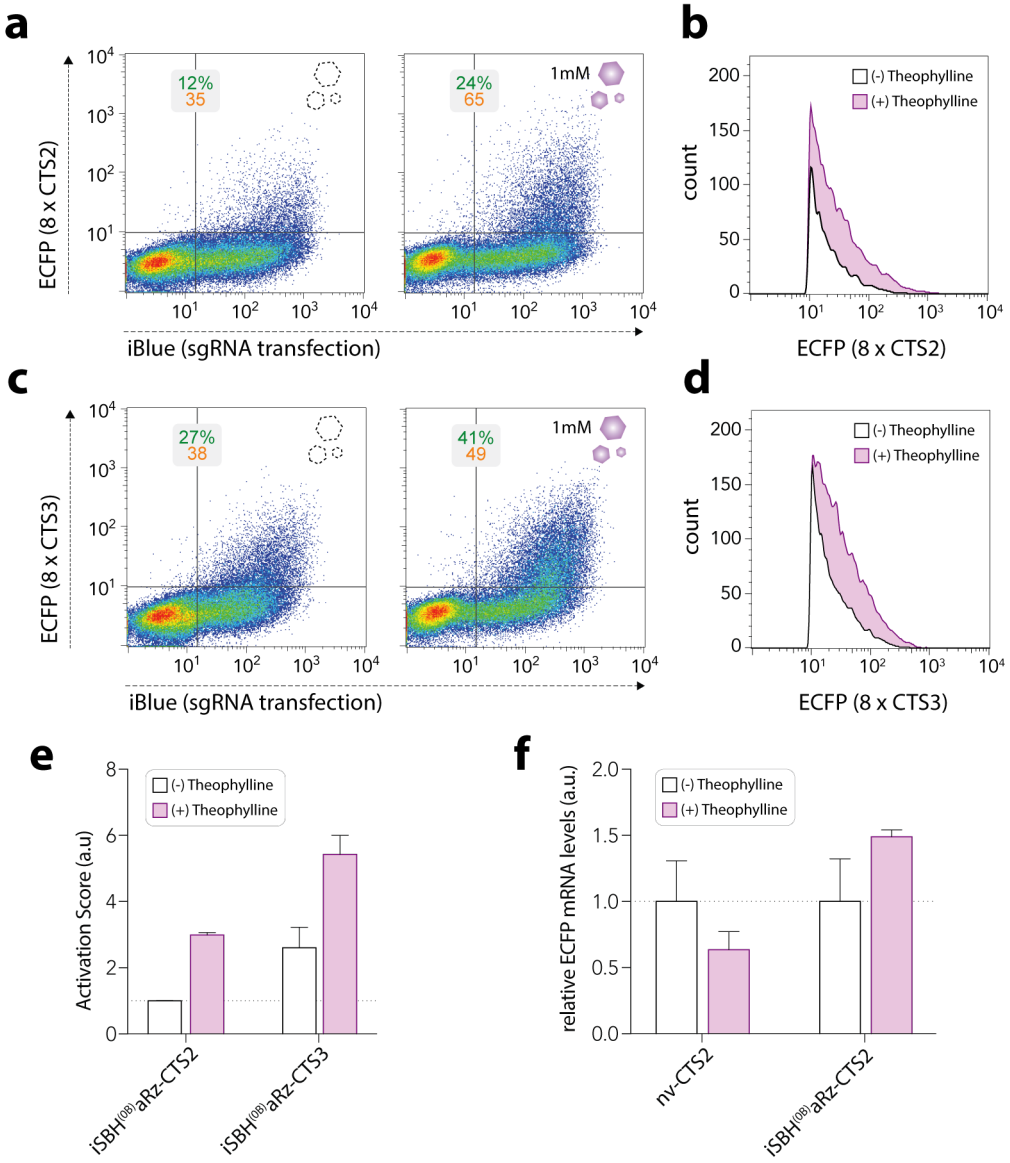

Figure 7.4 – Theophylline-responsive iSBH-sgRNA using allosteric ribozyme. (a) HEK293-T cells were transfected with dCas9-VP64, reporter construct, and the theophylline-responsive iSBH$^{(0B)}$aRz-CTS2 sgRNA. Transfection media was replaced 24h post-transfection with either plain culture media (left) or media supplemented with the drug at a concentration of 1mM (right). **(b)** Histograms of the reporter expression for the activated population (iBlue^{+ve}/ECFP^{+ve}) for each condition in (a). **(c, d)** Repeat of (a) and (b) with a different spacer:reporter pair: iSBH$^{(0B)}$aRz-CTS3 sgRNA is used in place of iSBH$^{(0B)}$aRz-CTS2. **(e)** Activation score quantification for all conditions in (a) and (c) (n = 2, mean +/- s.d., normalized to lowest value). **(f)** Repeat of experiment (a) with either iSB$^{(0B)}$aRz-CTS2 or a native sgRNA with spacer for CTS2 (nv-CTS2). RT-qPCR was performed 48h post transfection with a primer pair against the ECFP mRNA. Relative reporter mRNA levels normalized to the theophylline negative condition are reported (n = 2, mean +/- s.d.).

7.3 – Closing remark

In this chapter, I have reported on the construction of iSBH-sgRNAs responsive to the small molecule theophylline. This was achieved by using an allosteric hammerhead ribozyme, fitted with a theophylline-binding aptamer, as an RNA cleaving unit in the iSBH design. While an increase in CRISPR-TR activity was observed in the presence of the drug, demonstrating ligand-dependent CRISPRa, the usefulness of such iSBH-sgRNAs was mitigated by a non-negligible leakage in the OFF-state. The effect, which has been reported throughout the aHHRz literature, suggests that even in the absence of the small molecule trigger, a large fraction of the ribozyme pool was able to assume a catalytically active conformation (de Silva & Walter 2009). Others have shown than aHHRz ON/OFF performances could be improved by evolving (via *in vivo* or *in vitro* evolution methods) the sequence of the communication module linking the aptamer and the ribozyme (Nomura et al. 2013). This methodology was notably employed to render the catalytically inactive conformation more thermodynamically stable in the absence of the ligand. As a first step towards tightening up the system presented above, one could notably try replacing the linker used in this initial design, which was optimized for *E. coli* (Berschneider et al. 2009; Saragliadis & Hartig 2013), with published aHHRz sequences shown to promote allosteric regulation in mammalian cells (Win & Smolke 2007; Win & Smolke 2008; Ausländer et al. 2010; Nomura et al. 2013).

Finally, the adjunction of theophylline was found to reduced CRISPRa. While additional experiments using different target genes and dose response analysis will be needed to confirm this effect, it might be better for future iterations to select a different small molecule trigger. Suitable candidates, for which aHHRz sequences have already been evolved, notably include tetracycline (Win & Smolke 2007), toyocamycin (Yen et al. 2004), thiamine pyrophosphate (Saragliadis & Hartig 2013), and guanine (Nomura et al. 2012).

Chapter 8 – Discussion and Future Perspectives

Programmable transcriptional regulators adapted from the bacterial CRISPR-Cas9 system (CRISPR-TRs) are expected to play an instrumental role in the assembly of synthetic gene circuits (Jusiak et al. 2016). Astonishingly versatile, this biological part can easily be reprogrammed to target effector domains tethered to the crRNP to virtually any gene in a genome, thus controlling (activate or repress) its transcriptional output (Chavez et al. 2016). In contrast to previous generation of synthetic transcriptional regulators, whose implementation required cumbersome protein engineering, CRISPR-TRs specificity is encoded in the sequence of short RNA molecules known as sgRNAs. Simple modification of the sgRNA's first 20nt nucleotides, known as spacer, redirects the effector to DNA loci with complementary sequences. Accordingly, this method also offers the unique opportunity to target several genes in parallel by multiplexing sgRNA expression.

The work reported in this Thesis is part of a global effort aimed at creating inducible CRISPR-TR solutions to allow their integration in larger synthetic gene circuits. In particular, interfacing CRISPR-TR with other biological parts that can conditionally regulate its activity, will make it possible to wire intra- or extra-cellular cues to the downstream activation/repression of specific gene targets. Previous attempts at engineering inducible CRISPR-TRs have focused on the creation of switchable dCas9 and conditional effector tethering (see section 1.6). While elegant, dCas9 promiscuous sgRNA loading makes it hard to scale up such approaches to the independent control of multiple target genes. Instead I have proposed and demonstrated the

implementation of conditional CRISPR-TRs based on simple modifications of the sgRNA. Inducible sgRNAs were created by appending on their 5'-end inducible spacer-blocking hairpins (iSBH), which conditionally interfere with the guide's ability to hybridize with its DNA target. I showed that the resulting inducible iSBH-sgRNAs can easily be reprogrammed to respond to protein, nucleic acid, or small molecule inducers, and can be activated orthogonally providing a means to control their downstream targets independently.

After providing a quick overview of the main results presented in this thesis, I discuss below the advantages of the iSBH approach over previously reported inducible CRISPR-TRs based on protein engineering. Testifying to the relevance of the work presented in this Thesis, other groups have concomitantly reported the development of alternative inducible sgRNA strategies. These solutions are reviewed in the second part of this chapter alongside comparison with the iSBH platform. Finally, I end this discussion by listing a few potential applications that could emerge from this work, some of which I and others are currently pursuing in the Fulga laboratory.

8.1 – Results summary

In this Thesis, I have introduced the concept of inducible spacer blocking hairpin (iSBH) and showed how this platform can be employed to control CRISPR-based transcriptional activation. This work is based on the premise that CRISPR-TRs rely on a sgRNA to direct the catalytically inactive dCas9 and effector domains tethered to the sgRNA-dCas9 complex, to any desired genomic sequence. More specifically, target identification and CRISPR-TR docking on site requires hybridization between the first 20nt of the sgRNA, also known as spacer, and the complementary DNA.

Based on these mechanistic insights, I hypothesized that CRISPR-TR activity could be blocked by interfering with the interaction between the sgRNA and the matching DNA target. In the first result chapter of this Thesis, I showed that native sgRNAs can be fully silenced by appending on their 5'-end an RNA segment (back-fold) bearing sequence complementarity to their spacer.

In this case, base-pairing between the spacer and the complementary back-fold leads to the formation of an RNA stem loop which renders the spacer inaccessible for DNA target probing. Accordingly, I have chosen to name such structures spacer-blocking hairpins (SBH).

Building onto the strong OFF-state induced by the SBH, I then created inducible hairpin variants (iSBH) by replacing the default *GAAA* loop of the stem with RNA cleaving units, or sensing loops, whose cleavage is conditioned on the presence of specific inducers. Using this strategy, I successfully developed iSBH-sgRNAs responsive to CRISPR-associated endoribonucleases from the Cas6 family (protein-iSBH, chapter 3, 5), as well as single stranded DNA antisense oligonucleotides (ASO-iSBH, chapter 4). In both cases, it was shown that the iSBH-sgRNA is specifically activated by the presence of the cognate triggers. Importantly, no CRISPR-based activation of the gene targeted was observed with iSBH-sgRNAs, until the expression or delivery of the cognate inducer. As such, these switchable guides can be used as signal transducers to program a causal relationship between a particular biological event (trigger) and the activation of a user defined set of genes.

Inducible spacer-blocking hairpins are composed of two distinct parts, the stem (back-fold:spacer) and the sensing-loop, that specifies the gene to be controlled and the inducer of the iSBH-sgRNA, respectively. Accordingly, once an iSBH-sgRNA has been optimized to respond to a given inducer, the working sensing-loop can be combined with various stems to create a panel of iSBH-sgRNAs responding to the same trigger but targeting different genes (see chapter 5, 6). Conversely, sensing-loops can be interchanged while maintaining the sequence of the stem to reprogram an iSBH-sgRNA to respond to a distinct stimulus. Using this principle, I was able to generate a series of iSBH-sgRNAs used to demonstrate simultaneous and independent activation of two target genes (chapter 5, 6). It was notably showed that transfection of two inducible sgRNAs, sharing the same sensing-loop, could prime cells to concomitantly transcribe two reporters in the presence of a single specific trigger. Similarly, independent control of these two genes could be achieved by employing orthogonal inducer:target pairs encoded by iSBH-sgRNAs featuring distinct sensing-loops.

In order to promote the use of the iSBH platform, I then developed and made available to the scientific community, a web tool that automates the design of protein- and ASO-responsive sgRNAs for user defined spacer sequences (iSBHfold, chapter 6). In the same chapter, I also showed that the iSBH strategy is compatible with second generation CRISPR-TRs, and created inducible versions of the synergistic activator mediator (SAM) system (Konermann et al. 2014). These various iSBH-SAM constructs were then utilized to program conditional activation of human genes in their endogenous context.

Finally, in the optic of extending the repertoire of iSBH inducers, I concluded the result section of this Thesis by presenting proof-of-principle data for the implementation of aptazyme-based iSBH-sgRNAs. After showing that a self-cleaving ribozyme grafted onto the back-fold:spacer stem can be used as spacer release mechanism, I constructed a small molecule responsive iSBH-sgRNA by fitting the structure with a theophylline aptamer, which controls the catalytic activity of the ribozyme. While the resulting theophylline-responsive iSBH-sgRNA was only intended to demonstrate this concept, and suffered from OFF-state leakage, its implementation established the feasibility of using allosteric self-cleaving ribozyme in the design of iSBHs. This strategy could in theory allow the construction of new iSBH-sgRNAs capable of sensing any ligand for which an RNA aptamer has been evolved against.

8.2 – iSBH-sgRNAs versus protein-based approaches

8.2.1 – Compatibility with all CRISPR derivatives

Based on minimal RNA engineering of the sgRNA, the iSBH platform provides a “plug and play” approach to CRISPR-TR inducibility. Very much in line with what made the CRISPR-Cas9 system popular, namely the ease with which one can reprogram the system to target a different DNA locus, the creation of new iSBH-sgRNAs can be done in parallel and involves the same costs and cloning steps as a regular sgRNA. Additionally, because iSBH-based inducibility relies only on minimal alteration of the sgRNA sequence, I expect the method to be compatible with all CRISPR-Cas9 derivatives, and consequently allow spatiotemporal control

over genome editing (Hsu et al. 2014), genetic and epigenetic alterations (Dominguez et al. 2016; Thakore et al. 2016), base editing (Komor, Kim, et al. 2016), and labelling of genomic loci (B. Chen et al. 2013; Hanhui Ma et al. 2015). In this Thesis, It was notably shown that, while developed on the first generation CRISPR-TR dCas9-VP64, iSBH could also be used to control state of the art CRISPR-TRs such as the SAM system without any further optimization (chapter 6). Moreover, I expect that similar adaption for the Suntag and VPR systems will be even more straightforward, as those approaches are employing the original sgRNA sequence.

To further illustrate the universality of the approach, I sought to demonstrate that iSBH could be employed to achieve conditional control over genome editing in mammalian cells. This endeavor was primarily motivated by studies showing that a prolonged exposure of cells to a catalytically active sgRNA-Cas9 complex increases the incidence of undesired off-target editing events (Hsu et al. 2013; Pattanayak et al. 2013; Fu et al. 2014; Zuris et al. 2015). In this instance, I chose to create a Csy4-responsive sgRNA targeted to the therapeutically relevant programmed death-ligand 1 (*PD-L1*) gene, which encodes a transmembrane protein that transmits an inhibitory signal reducing T-cell function and proliferation (Park et al. 2010; Topalian et al. 2012). *PD-L1* has been linked tumor immune evasion (Stewart & Smyth 2011; D. S. Chen & Mellman 2013) and successfully targeted by checkpoint inhibitors to develop novel cancer immunotherapy treatments (Korman et al. 2006; D. S. Chen et al. 2012).

Studies have shown that extensive base-pairing between the spacer and the DNA target (>15bp) is required for Cas9-mediated DNA double strand break to occur (Jiang & Doudna 2017). Accordingly, the optimal *nano* Csy4-iSBH design presented in chapter 3, which yields a 10nt long spacer post-cleavage, was not expected to work for genome engineering applications. Instead, a full length iSBH$^{(0B)}$Csy4$^{(full)}$PDL1 hairpin was created from a previously reported *PD-L1* targeting sgRNA (Kataoka et al. 2016), by grafting the Csy4-PTS on the corresponding 19nt bulged stem (Fig. 8.1a). In order to confirm that the *PD-L1* locus could be successfully edited, a T7E1 assay (Tsuji & Niida 2008; Mashal et al. 1995) with primers asymmetrically flanking the DNA target (see methods, Fig. 8.1b) was conducted 48h after

transfecting HEK293-T cells with a pDNA expressing both the catalytically active Cas9 and the native sgRNA programmed against *PD-L1* (nv-PDL1). Analysis of the T7E1 reactions revealed the presence of editing events which were not observable when nv-PDL1 was replaced with a scramble sgRNA (nv-SCR, Fig. 8.1c). In contrast, and as expected, cell expressing Cas9 and the Csy4-responsive iSBH-sgRNA displayed no detectable traces of genomic alterations, unless co-transfected with the Csy4 endoribonuclease (Fig. 8.1c). Taken together, these results suggest that iSBH-sgRNA could also be used to control Cas9 nuclease activity. In addition to reducing cellular toxicity, this could potentially be used to achieve tissue specific editing *in vivo* by placing several Cas endoribonucleases under the control of tissue specific promoters. Moreover, it is highly probable that the same results, albeit weaker, could be obtained using ASO-responsive hairpins.

In addition to providing a means of controlling systems based on the *Sp* CRISPR-Cas9, I anticipate that sgRNA control via iSBH will also be applicable to other Cas9 orthologues (Esvelt et al. 2013; Braff et al. 2016) and newly characterized RNA-guided nucleases from the CRISPR family. These include the Type V-A Cas12A, also known as Cpf1 (Zetsche, Gootenberg, et al. 2015), Type V-B Cas12b, or C2c1 (Shmakov et al. 2015), and the RNA-guided RNA nuclease Type VI Cas13a, commonly referred to as C2c2 (Abudayyeh et al. 2016). All these CRISPR systems employ a guide RNA, whose hybridization with the DNA or RNA target is required for their catalytic function (Fig. 8.2a-c). Accordingly, inducible versions of the corresponding guide RNAs could in principle be created, using the iSBH strategy, to control the activity of such systems.

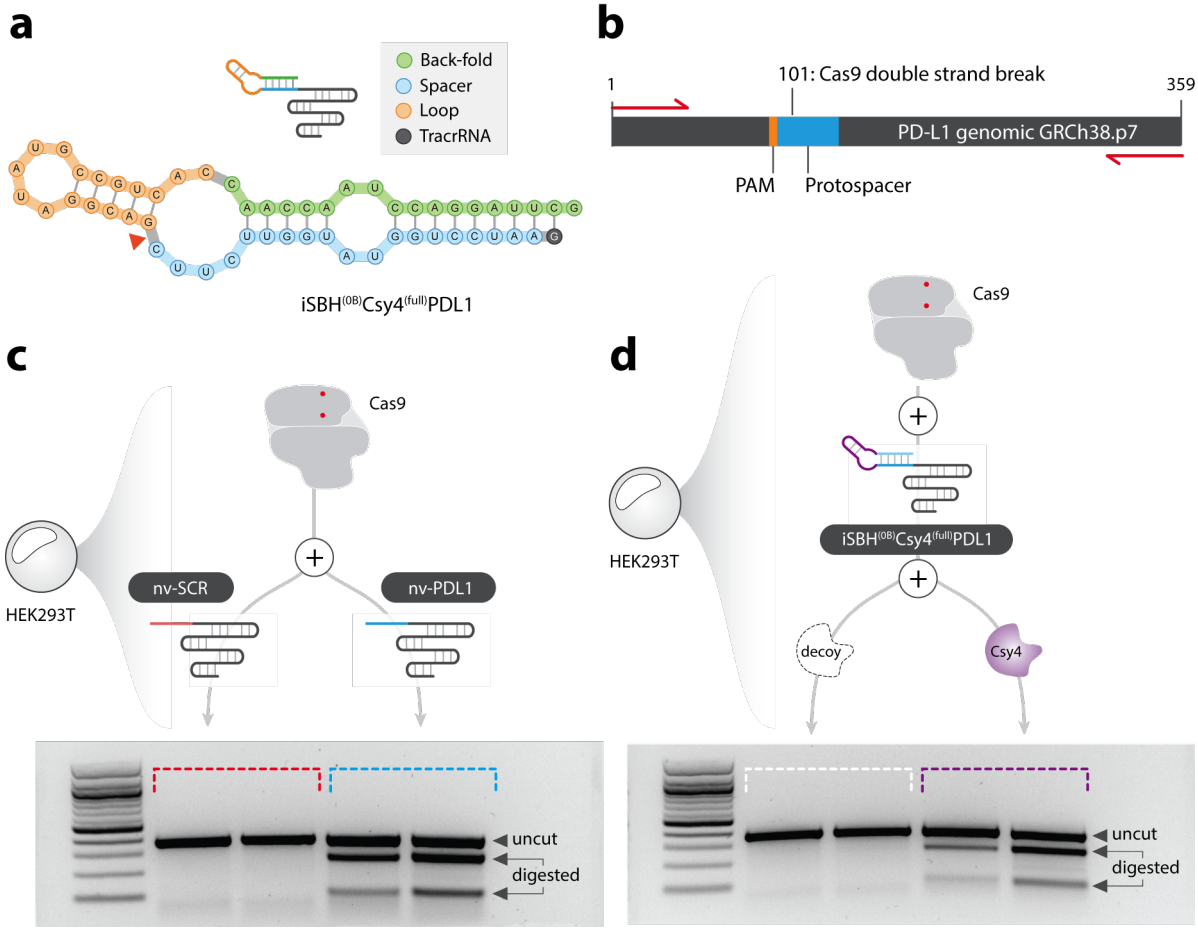

Figure 8.1 – Conditional CRISPR-based genome editing using iSBH. (a) Sequence and RNA secondary structure of a Csy4-responsive hairpin with spacer for the *PD-L1* gene. **(b)** Schematic representation of the target locus as PCR amplified for the T7E1 assay: A primer pair (red arrows) asymmetrically flanking the CRIPSR target site (protospacer + PAM) is chosen to amplify the region. PCR is performed on the genomic DNA extracted from the transfected cell population. Wild-type and mutant amplicons are then heated up and cooled down slowly to form double stranded DNA chimera with mismatches at the cut site. This pool is then incubated with the bacteriophage-derived T7 endonuclease I, which recognizes and cleaves non-perfectly matched DNA. The DNA pool is then run on an agarose gel. **(c)** HEK293-T were transfected with the catalytically active Cas9 and a native sgRNA with either a scramble spacer (nv-SCR) or the *PD-L1* spacer shown in (a) (nv-PDL1). T7E1 assay is performed 48h post-transfection. **(d)** Similar to (c) but iSBH$^{(0B)}$Csy4$^{(full)}$PDL1 iSBH-sgRNA is transfected in place of the native sgRNA, along with either a decoy or a Csy4-expressing vector.

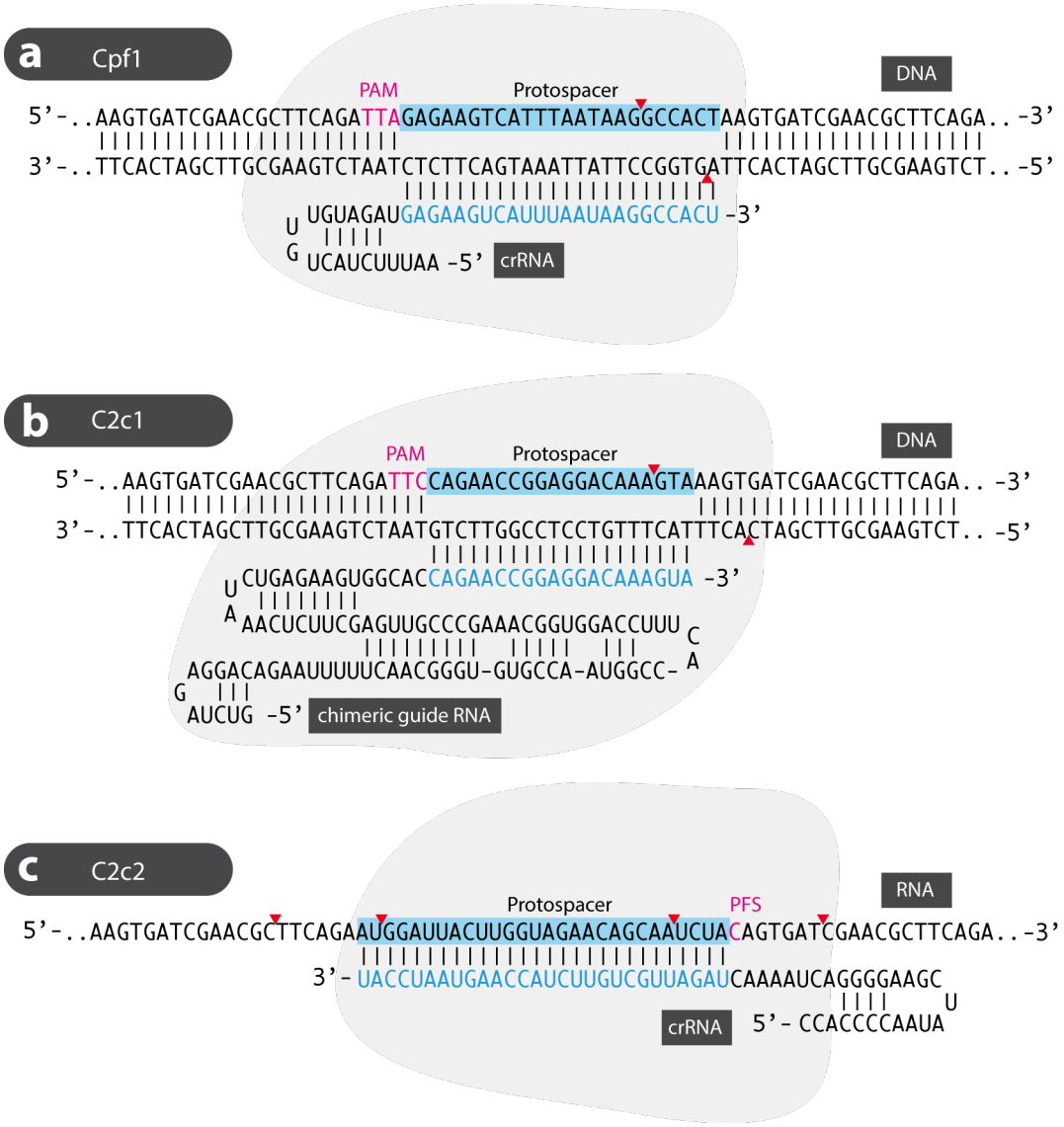

Figure 8.2 – Newly characterized CRISPR nucleases. (a-c) Schematic representation of the Cpf1 (a), C2c1 (b), and C2c2 (c) crRNP bound on-target. In each case, the CRISPR-effector is displayed as a light grey blob. Sequence and RNA secondary structure of either the natural crRNA or human engineered chimeric guide RNA is given. Spacer segments complementary to the DNA or RNA target are colored in blue. On the target side, protospacer, as well as protospacer adjacent motif (PAM) or protospacer flanking sequence (PFS) are highlighted. The position of cut on each DNA strand is indicated by a red arrow (a, b). In the case of C2c2, several arrows along the RNA sequence depict non-specific target cleavage.

8.2.2 – iSBH facilitates gene circuit assembly.

Regarding the assembly of complex synthetic gene circuits, the iSBH platform has several advantages over methodologies based on inducible dCas9. As alluded to earlier, the predominant benefit of working with inducible sgRNAs lies in the ability to independently control multiple target genes while using a single CRISPR-effector (e.g. dCas9-VP64). In contrast, achieving the same result with any switchable dCas9 strategy would necessitate re-engineering multiple dCas9 orthologues with distinct PAM requirements and different guide RNAs (Esvelt et al. 2013). In addition to the technical difficulties associated with designing such a collection of inducible dCas9 proteins, co-expression of multiple CRISPR-effectors will most likely constitute a non-negligible metabolic burden to the cell (Gang Wu et al. 2016). Moreover, it is to be expected that an extension of the PAM repertoire will also increase toxicity associated with DNA probing, as well as the risks for OFF-target effects.

Secondly, the availability of inducible sgRNAs strategies, such as the iSBH, makes it possible for the first time to encode an entire transcriptional program in a single molecule. In fact, by combining the iSBH technology with scaffold sgRNAs, able to recruit different effector domains (Zalatan et al. 2014; Konermann et al. 2014), one could construct a guide RNA that specifies when and where a specific target gene is to be activated or silenced (Fig. 8.3): the choice of the inducer, the gene targeted, and desired phenotype (up or downregulation), are programmed by changing the sensing-loop, spacer, and aptamer segments of the sgRNA, respectively. It was shown in chapter 6 of this Thesis that iSBH-based inducibility is compatible with the scaffold sgRNAs used in the SAM system. Engineered to recruit, via two MS2 loops, multiple copies of the protein fusion MCP-p65-HSF1, these sgRNAs could be easily modified to feature a distinct aptamer and tether repressors such as the KRAB domain. These two scaffold sgRNAs could be co-expressed to enable concomitant activation and repression of two gene targets in the same cell expressing dCas9. Alternatively, Zalatan et. al have already constructed such a pair of scaffold sgRNAs which could be enhanced with iSBHs (Zalatan et al. 2014).

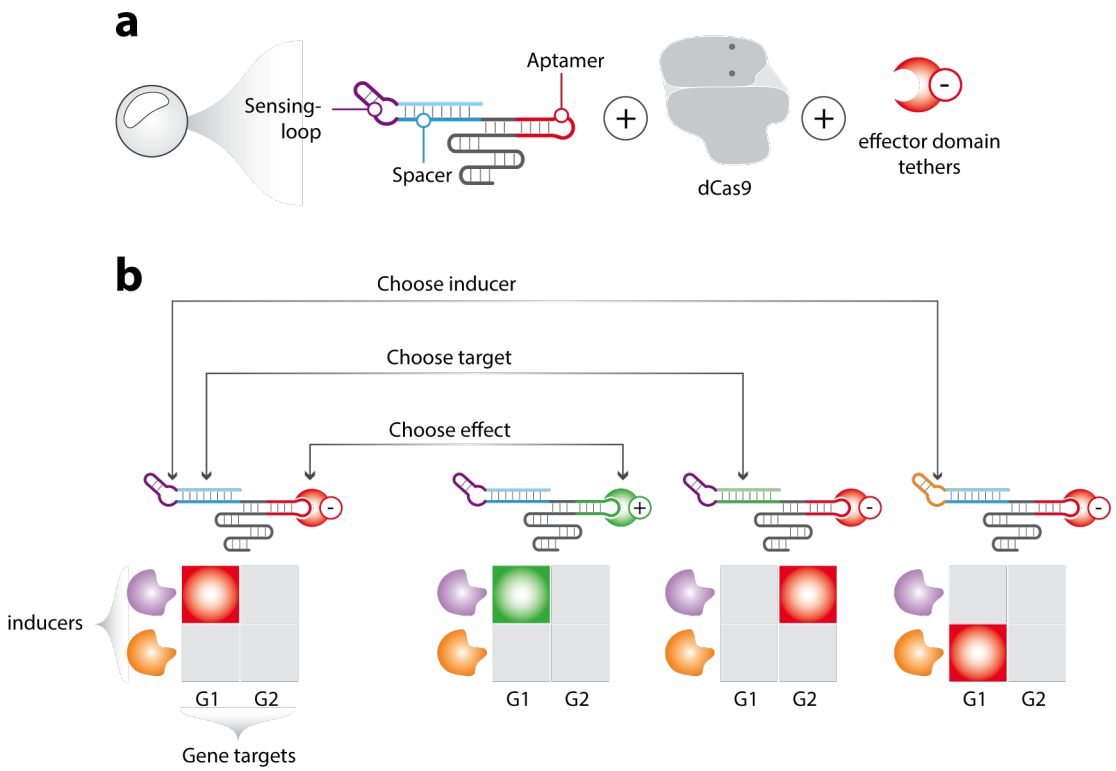

Figure 8.3 – Encoding an entire transcriptional program in single RNA molecule. (a) A scaffold sgRNA, fitted with a RNA aptamer capable of recruiting specific effector domains, is enhanced with iSBH-inducibility. Such iSBH-sgRNA-Apt can potentially encode an entire transcriptional program: researchers can decide when and where (iSBH), a particular gene (spacer), is to be overexpressed or silenced (RNA aptamer). Implementation of the program requires the co-transfection, along iSBH-sgRNA-Apt, of dCas9, and the effector domain tether. **(b)** Accordingly, a given transcriptional program (e.g. repression of G1 in the presence of the purple inducer) can be changed via simple modification of the sgRNA sequence to drive transcription instead of repressing it (activator to repressor domain), respond to a different inducer (purple to orange inducer), and target a different gene (G1 to G2).

Another important consideration and advantage of the iSBH system when trying to implement complex gene circuits, is the high specificity of iSBH-sgRNA induction. Indeed, all iSBH-sgRNAs created throughout this project solely activated in the presence of their designated trigger. In particular, it was shown in chapter 5 and 6 that iSBH-sgRNAs with distinct sensing-loops can be controlled orthogonally without cross-talk between the inducer:target pairs they encode. While enough evidence was provided to demonstrate the independence between Csy4 and Cas6A-responsive iSBHs, further experiments are required to fully characterize how

different degree of sequence complementarity between inducer and sensing-loop affects the activation of ASO-responsive iSBH-sgRNAs.

Finally, I reported that iSBH-sgRNAs are fully silent in the OFF-state as indicated by their incapacity to drive CRISPRa. This features contrasts with the leakiness often observed with inducible promoters and switchable dCas9 solutions (Chira et al. 2017). This tight OFF-state makes iSBH-based inducible CRISPR-TRs particularly attractive for the design of gene circuits implementing conditional kill switches involving downstream activation of pro-apoptotic genes in mammalian cells.

8.2.3 – iSBH-sgRNA, inducer selection and implementation.

Thorough this thesis, I have engineered multiple iSBH-sgRNAs and characterised their ability to direct CRISPR-TR on target in the absence (OFF-state) and presence (ON-state) of their cognate molecular trigger. Various design iterations were proposed and tested to create inducible hairpins responsive to endoribonucleases (chapters 3, 5) and antisense oligonucleotides (chapter 4). In order to help the reader select the best iSBH-sgRNA design for her/his application, I briefly list below the different stems developed throughout the project, and summarise the key findings that led me to privilege one design over another. Additionally, I discuss the advantages and limitations of using endoribonucleases or ASOs as inducers.

In chapter 3 and 5, I developed a series of iSBHs responsive to two CRISPR-associated endoribonucleases: *Pseudomonas aeruginosa* Csy4 and *Thermus thermophiles* Cas6A. For both inducers, an "entry" inducible hairpin (iSBH$^{(0B)}$Csy4$^{(full)}$CTS, iSBH$^{(0B)}$Cas6A$^{(full)}$CTS) was created by replacing the *GAAA* loop of the bulged SBH with the RNA sequence recognised by the endoribonuclease (protein target site, PTS). Based on results showing that the thermodynamic stability of the back-fold:spacer stem plays a critical role on the ON-state performance of the system, the hairpin was progressively shortened to create *medium* and *nano* inducible hairpins obtained by fusing the PTS to seed distal, or seed central bulge of SBH$^{(0B)}$CTS (iSBH$^{(0B)}$Csy4$^{(medium)}$CTS, iSBH$^{(0B)}$Csy4$^{(nano)}$CTS, and similar for Cas6A). The

design principles used to create these three different designs are explained in section 3.5. Head to head comparison of these hairpins revealed that, in the case of Csy4, *nano* iSBHs outperformed their *full* and *medium* counterparts. On the other hand, I found that the *medium* hairpin displayed the best ON/OFF-characteristic when using Cas6A as an inducer. As such, I would recommend using *nano* stems when sensing Csy4 and a *medium* iSBH-sgRNA when creating CRISPR-TR responsive to Cas6A. Nevertheless, I would like to emphasise that this hierarchy in design performances was only demonstrated for a few spacer sequences. Also, one cannot exclude the possibility that for certain spacers not yet tested, these privileged designs might not be optimal.

In chapter 4, I introduced and tested various designs of ASO-responsive sgRNAs as well as different ASO inducers with diverse base-pairing footprint on the iSBH. In contrast with protein-responsive hairpins, I found that shortening the back-fold:spacer stem from *full* to *medium* design does not improve ON-state activation. As such, the *nano* design was not tested and the preferred ASO-iSBH design is a full 20bp bulged stem featuring a 14nt sensing-loop, whose sequence has been evolved to assume an open conformation upon iSBH folding (iSBH$^{(0B)}$ASO-CTS). Regarding the characteristics of the oligonucleotide inducer, I found that the strongest guide induction was achieved when using a 20nt ASO designed to base-pair with the entire sensing-loop as well as the three adjacent nucleotides upstream and three adjacent nucleotides downstream.

Based on these considerations, the iSBHfold web tool introduced in chapter 6, will generate a *nano* iSBH when the Csy4 inducers is chosen, a *medium* hairpin for Cas6A-responsive sgRNAs, and a *full* stem for ASO-iSBHs. Is left to the user the choice of the spacer sequence and type of inducer. The sequence of the spacer will be determined by the choice of the gene that is to be regulated, while the choice of the inducer depends on the experimental set up and type of application. The use of endoribonuclease inducers makes it in theory possible to achieve tissue specific expression of target genes in an animal model (see section 8.4.3), something not possible with ASO-responsive guides. On the other hand, inducible CRISPR-

TRs responding to ASO triggers are preferred when it comes to achieve temporal control over gene expression (see section 8.4.4). While the later could be done with protein-responsive guides by placing Csy4 or Cas6A under the control of inducible promoters, these promoters tend to be leaky. ASO-responsive sgRNAs in the other hand are expected to remain fully silent until induction.

8.3 – Comparison with other inducible sgRNA strategies

Within the same time frame, several other groups have engineered inducible CRISPR-Cas9 systems based on controlling sgRNA activity or availability. In this section I briefly present, in chronological order, each of these strategies and discuss their pros and cons as well as how they compare, on a theoretical level, with iSBH-sgRNAs in term of their applications.

8.3.1 – Targeting antisense RNAs against the sgRNA

Young Je Lee et al. were the first to introduce the concept of interfering with sgRNA functioning to modulate CRISPR-TR (Y. J. Lee et al. 2016). In their 2016 paper, the authors demonstrated that CRISPRi activity could be silenced by directing an antisense RNA (asRNA) against the sgRNA (Fig. 8.4a, b). After establishing an assay whereby sgRNA guided dCas9 represses the expression of a reporter gene in *E.coli*, they tested several asRNA designs and reported that increase asRNA levels correlated with a progressive silencing of CRISPRi, as evidenced by the derepression of the target gene.

The asRNAs used in this study were single-stranded RNAs featuring a binding site for the native Hfq protein (chaperone protecting asRNA from degradation) on their 3'-end, and a segment complementary to the sgRNA sequence on their 5'-end. The first generation asRNAs, targeted to the sgRNA spacer (Fig. 8.4a), only showed marginal derepression (maximum 50%). A far better derepression (95%) was obtained by designing the asRNA to target a user defined synthetic linker sequence added on the 3'-end of the sgRNA (Fig. 8.4b). Finally, by

creating several sgRNAs featuring distinct linker segments, the authors were able to utilize this strategy to perform independent derepression of two target genes using orthogonal asRNA-sgRNA pairs.

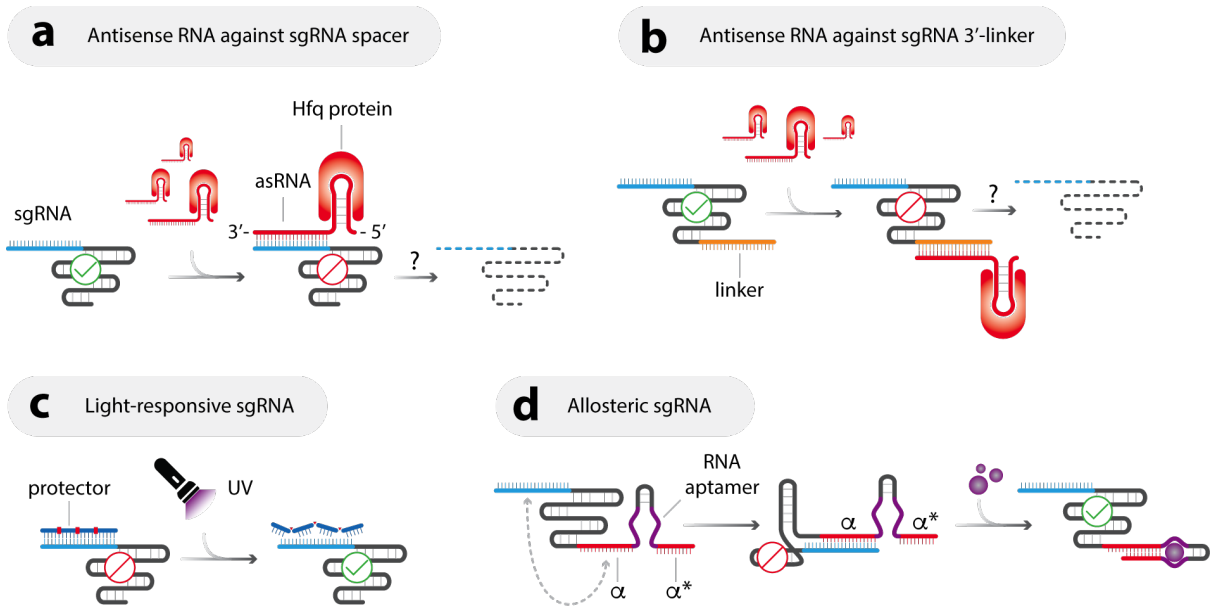

Figure 8.4 – Alternative inducible sgRNA strategies. (a) Silencing of CRISPRi activity by targeting an antisense RNA (asRNA) to the spacer segment of the sgRNA. The asRNA is composed of a 5' handle for the Hfq binding protein, and a 3' antisense segment bearing sequence complementarity to the spacer. Hybridization between asRNA and sgRNA is thought to lead to the degradation of the guide via RNase III endoribonuclease activity. **(b)** Similar to (a) except that the asRNA is programmed to target a linker sequence appended to the 3'-end of the sgRNA. **(c)** "Precise light-mediated unveiling of sgRNAs", CRISPR-plus. A ssDNA oligonucleotide "protector" complementary to the sgRNA spacer blocks CRISPR-Cas9 nuclease activity by preventing sgRNA:DNA target hybridization. The system is made responsive to UV-light by incorporation in the protector sequence of photo-cleavage groups (red). Exposure to UV-ligth induces lysis of the protector and resumes CRISPR-Cas9 nuclease activity. **(d)** Allosteric sgRNAs are created by adding on the 3'-end of the guide a ligand-binding RNA aptamer flanked by two flanking segments: α and α*. Segment α is designed complementary to the last 15nt nucleotides of the spacer forcing the guide to assume an inactive conformation in the absence of the ligand. Presence of the ligand brings α and sequence complementary α* segments in close proximity. Base-pairing between α and α* releases the spacer and activates the sgRNA.

This work presents the first account of CRISPR-TR modulation by repressing sgRNA availability. The best ON/OFF characteristics of their system was achieved using asRNAs designed to target a 13nt linker sequence appended on the 3' end of the sgRNA. Similar to the iSBH design, whereby the identity of the sensing-loop is chosen independently to the stem, the sgRNA spacer and linker sequences are unrelated in their design. Therefore, this strategy could also be used to implement the branching and orthogonal modules presented in chapter 5. Additionally, such design could in principle open the doors to sensing endogenous RNA species that would be highly beneficial to the synthetic biology field.

Nevertheless, as opposed to the full OFF-state silencing observed with iSBH-sgRNA, their data suggests that asRNAs expression was not able to completely derepress dCas9 targets (Y. J. Lee et al. 2016). The authors notably pointed out that derepression efficiency was highly dependent on the relative stoichiometry between asRNA and sgRNA, a feature which might impact the robustness of any gene circuits employing such strategy. Furthermore, additional work will be required to fully understand the mechanisms of asRNA-mediaded sgRNA silencing. In particular, the authors did not propose an explanation as to why asRNA targeted at the sgRNA spacer seem to be less efficient compared to those targeting a 3'-end linker. The research presented in this Thesis suggests that difference could be explained by the fact that once the sgRNA is loaded into dCas9, the spacer segment is out of reach for the asRNA. In the iSBH system, hairpin formation is expected to happen co-transcriptionally, meaning that the spacer has been caged prior to iSBH-sgRNA loading into the CRISPR-effector. However, in asRNA-sgRNA system, it is likely that a portion of the transcribed sgRNAs will be loaded into dCas9 before asRNA can find its target. As a consequence, the level of derepression is expected to plateau at a fixed value, which will be a function of the relative concentrations between dCas9, sgRNA, and asRNA. Conversely, when the linker is moved to the 3'-end of sgRNA, the available crRNP crystal structure (Nishimasu et al. 2014) suggests that it should be become accessible for asRNA targeting and thus increase derepression efficiency.

8.3.2 – CRISPR-plus: photo-cleavable spacer protectors

Jain et al. later engineered light-responsive inducible sgRNAs and published the technology under the name of CRISPR-plus, an acronym standing for 'precise light-mediated unveiling of sgRNAs' (Jain et al. 2016). Similar to the iSBH silencing mechanism, CRISPR-plus aims at blocking sgRNA-DNA target interaction by providing a ssDNA oligonucleotide, called "protector", which hybridizes with the sgRNA spacer. Inducibility is created by incorporation of photo-cleavable groups in the protector, which allow photolysis of the oligonucleotide in response to UV-light (Fig. 8.4c). Protectors used in this study were approximatively 25nt long, interspaced with photo-responsive groups every six nucleotides, and designed to base-pair with the entire spacer segment plus a portion of the sgRNA scaffold. Proof-of-concept was carried out *in vitro* where protected sgRNAs (p-sgRNA) were incubated with DNA targets and catalytically active Cas9. The authors were able to simultaneously activate multiple p-sgRNAs as confirmed by degradation of their respective DNA targets. Finally, p-sgRNA delivered to a HeLa cell line stably expressing EGFP and Cas9 showed loss of reporter expression in the presence of UV trigger. The authors also present data from a SURVEYOR nuclease assay (Qiu et al. 2004) confirming successful conditional editing of two endogenous gene targets using CRISPR-plus (Jain et al. 2016).

While the fast kinetics and non-invasive nature of optically controlled systems make CRISPR-plus attractive, the method comes with several caveats. First, its implementation is relatively cumbersome compared to the iSBH technology. In fact, p-sgRNAs need to be assembled prior to delivery in cells. In their protocol, the sgRNAs were *in vitro* transcribed before being annealed with their cognate protectors. In addition, the authors report that the system suffers from OFF-state leakage (loss of silencing) when p-sgRNAs are left more than 5 days in the cells. Finally, because CRISPR-plus is so far only compatible with UV-light, the approach cannot be employed to control multiple sgRNAs independently.

8.3.3 – CRISPR signal conductors, allosteric sgRNAs

A paper by Liu et al. was later released that introduced the concept of sgRNA signal conductors (Y. Liu et al. 2016). The team engineered a series of allosteric sgRNAs (a-sgRNAs), and termed them signal conductors due to their ability to reorient a particular signal in the cell to the activation or repression of a downstream gene target (CRISPR-TR). Comparably to the iSBH methodology, a native sgRNA is first silenced by appending to its sequence an RNA segment designed to hybridize with the spacer (Fig. 8.4d). Nevertheless, while iSBH is based on a 5'-end extension of the guide, a-sgRNA are modified on their 3'-end. However, the biggest differences between a-sgRNAs and iSBH-sgRNAs lies in their mode of activation. In the case of iSBH-sgRNAs, melting of the back-fold:spacer stem rely on cleavage of the iSBH sensing-loop. In contrast, the activation of a-sgRNAs relies on a conformational change induced by the specific binding of a ligand to an RNA aptamer incorporated in the sgRNA design (Fig. 8.4d). As depicted on figure 8.4d, the a-sgRNA 3'-extension is composed of three RNA segments that are referred to as α, Aptamer, and α*. From 5' to 3', segment α is designed to hybridise with the 15 seed distal nucleotides of the spacer. This is the equivalent to the iSBH back-fold. Downstream a ligand-binding RNA aptamer is attached, directly followed by segment α*, whose sequence is complementarity to α. Segments α and α* are designed such that base-pairing between the spacer and α is thermodynamically favored over α:α* interactions in the absence of the ligand. Binding of the ligand to the aptamer is accompanied by a conformational change that forces α to preferentially bind α*, leaving the spacer free for DNA target probing (Fig. 8.4d).

Leveraging this modular design, the authors constructed several a-sgRNAs with aptamers raised against both small molecules and protein inducers. These responsive guides were elegantly combined to demonstrate orthogonal and simultaneous control of multiple target genes, similar to the gene circuits modules presented in this Thesis (see chapter 5). Additionally, a-sgRNAs were employed to assemble Boolean logic gates and rewire endogenous signaling pathways to link oncogenic signal with activation of tumours suppressor

genes. Finally, the authors built an anti-tumor progression circuit that senses the presence of oncogenic markers NF-kB and β-catenin, and forces cells towards apoptosis by repressing *BCL2* and activating *BAX*, respectively. Speaking for the medical relevance of this circuit, Liu and colleagues notably presented data indicating that intra-tumoral delivery of such devices could halt the progression of tumours formed from subcutaneously injected cancer cells (Y. Liu et al. 2016).

As discussed in chapter 7, RNA aptamers can be evolved to bind a broad range of ligands, including proteins, peptides, and small molecules. Realizing this potential, the authors successfully generated inducible a-sgRNAs responsive to tetracycline, theophylline, p53, HSF1, NF-kB, and B-catenin, respectively (Y. Liu et al. 2016). Nevertheless, my and others work (see below) on the creation of aptazyme-based inducible sgRNAs, suggests that such allosteric sgRNA might be subject to strong OFF-state leakage (CRISPR-TR activity in the absence of ligand). While the OFF-state conformation should be favored in the absence of the ligand, it is expected that some RNA molecules would misfold into the active structure. Additionally, this effect could be exacerbated by the binding of dCas9 on the sgRNA, which according to the crystal structure is only compatible with the ON-state conformation of the a-sgRNA (Nishimasu et al. 2014). Surprisingly however, no sign of leakage was reported by the authors for all designs using α segments with at least 15nt complementarity to the sgRNA spacer.

To validate that a-sgRNAs are in fact capable of full CRISPR-TR silencing in the OFF state, I applied the design principle presented in their work to one of the native sgRNAs used in this Thesis (nv-CTS2, see methods). To be conservative and ensure maximum silencing of the sgRNA, the segment α was designed with 16nt complementarity to the CTS2 spacer. Additionally, the 3'-end Aptamer + α* segment, which is intended to stabilize the active conformation, was replaced by a shorter random RNA sequence that did not compete against α:spacer base-pairing (Fig. 8.5a). Transfection in HEK293-T cells of the resulting a-sgRNA,

along with dCas9-VP64 and the corresponding reporter construct, revealed that, in stark contrast to data published in Liu's manuscript, this a-sgRNA was able to mediate substantial CRISPRa (Fig. 8.5b). Nonetheless, further experiments using their exact a-sgRNA sequences will be required to conclude on the leakiness of their system.

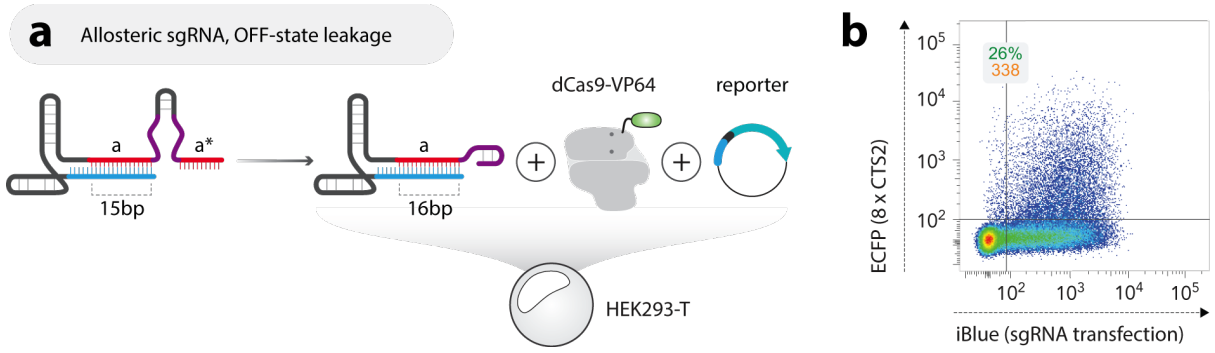

Figure 8.5 –OFF-state properties of allosteric sgRNAs. (a) An allosteric-sgRNA mimic is created to test the OFF-state leakage of the design. Segment α is extended to increase base-pairing with the spacer to 16nt, and the aptamer-α* section is replaced by a shorter 20nt tail followed on its 3'-end by a 3bp hairpin protecting the sgRNA from exonuclease degradation. **(b)** The sgRNA mimic was transfected in HEK293-T along with dCas9-VP64 and the reporter gene construct. Flow cytometry analysis 48h post-transfection shows strong reporter expression (ECFP). Inset display percentage of activated cells (green) and median reporter fluorescence for this population (orange).

8.3.4 – Aptazyme-embedded inducible sgRNAs

More recently, Tang et. al reported on an improved version of the work presented in chapter 7 of this Thesis (W. Tang et al. 2017). Using the same basic approach, the team has employed allosteric self-cleaving ribozymes to create inducible sgRNAs responsive to the small molecules theophylline and guanine. Reminiscent of the iSBH back-fold segment, cleavage of the aptazyme removed a "blocking" sequence, which otherwise sequesters the spacer. The team then demonstrated that these inducible sgRNAs could be used with Cas9, the CRISPR-TR dCas9-VPR, and the BE3 base editor (Komor, Kim, et al. 2016) to perform conditional genome editing, transcriptional activation, and single base editing, respectively. However, in line with what we have reported in chapter 7, the authors acknowledge that their system suffers from non-negligible OFF-state leakage in the absence of the small molecule inducer.

8.4 – Future prospects and applications

In this section, I first provide a list of future experiments that could be conducted to provide a more in depth characterization of the iSBH system. This list is not intended to be exhaustive but rather restricted to key analyses that will either help improve the system ON/OFF-performances or provide mechanistic insights that will benefit researchers using the iSBH technology. In the second half, a few potential applications of iSBH-sgRNAs are presented, which provide examples of their use for fundamental research and therapeutic purposes.

8.4.1 – Further characterization of iSBH-based inducibility

Silencing mechanism: As stated in chapter 2 of this Thesis, additional experiments are required to fully understand the mechanisms by which spacer-blocking hairpins repress CRISPR-TR activity. For example, a comparison analysis of sgRNA levels (RT-qPCR) between native guides and iSBH-sgRNAs would help ascertain that the presence of a 5'-end hairpin does not lead to a depletion of the sgRNA pool. In the case where the levels are found to differ, separate experiments using U6 driven iSBH-sgRNAs or the transfection of *in vitro* transcribed guides will help elucidate if the reduction is imputable to lower transcriptional rates or sgRNA degradation, respectively. Subsequently, it will be interesting to determine if silencing comes from the inability of dCas9 to load iSBH-sgRNAs, the inability of the iSBH-sgRNA:dCas9 complex to bind its DNA target, or a combination of both. In chapter 2, I have notably presented data suggesting that fully silent SBH-sgRNAs are being loaded into dCas9 (see section 2.3.2). Since it could be argued that replacing the default *GAAA* loop of SBH with longer sensing-loops might affect guide loading, those experiments will need to be repeated across several iSBH-sgRNAs (protein- and ASO-responsive) to determine if these results hold true. Additionally, more stringent and established experiments such as the electrophoretic mobility shift assay (Hellman & Fried 2007), could be used to further validate these findings. Finally, with regard to the ability of iSBH-sgRNAs to bind DNA targets, the fact that all variants tested in this work fully repress CRISPR-TR activity suggests that spacer-blocking hairpins do indeed prevent anchoring of the crRNP on-target. However, it would be interesting to also

investigate whether or not iSBH-sgRNA loading into dCas9 is sufficient to drive the conformational change in the protein structure required for the crRNP to probe for PAM sequences. This could notably be determined by comparing DNA binding patterns between native and iSBH-sgRNAs with a DNA curtain assay (Sternberg et al. 2014) or genome wide by performing ChiP-seq on dCas9 (Xuebing Wu et al. 2014).

Elucidate the mechanisms underlying iSBH-sgRNA-mediated activation: Current iSBH-sgRNAs designs could further be improved by gaining an understanding of the mechanism underlying iSBH-sgRNA activation. Notably, it will be beneficial to know whether sgRNA induction takes place co-transcriptionally, post-transcriptionally, prior or after being loaded into dCas9. For example, if only free iSBH-sgRNAs (non-loaded) were targeted by the inducer, one might want to revise the design to either increase targeting of the iSBH-sgRNAs complexed with dCas9 or prevent iSBH-sgRNAs loading in the OFF-state. Such insights could be gained by conducting *in vitro* experiments where the components (iSBH-sgRNA, inducer, and dCas9) are added to the mix at different time point, before determining the fraction of processed iSBH-sgRNAs by northern blot. In parallel, similar experiments could be carried out in cell culture by controlling the relative expression of the system components with inducible promoters, followed by RT-qPCR to quantify the proportion of cleaved iSBH-sgRNAs.

Dose response and reversibility: One might also want to determine if the extent of CRISPR-TR activity can be modulated by titrating inducer concentration. In the case of protein inducers, dose-response analyses could be conducted by generating stable cell lines expressing the trigger under the control of an inducible promoter. In the case of ASO-responsive iSBH-sgRNAs, further work will be required to improve oligo delivery before dose-response analyses can be performed. Likewise, it would also be important to perform time course experiment aimed at demonstrating the reversibility of the system and describing the dynamics of iSBH-sgRNA induction for each type of inducers. Since activated iSBH-sgRNAs get degraded overtime, CRISPR-TR activity is expected to peak after inducer delivery, before fading away progressively thus rendering the system reversible.

OFF-targets delineation: It is well established that CRISPR-Cas9 can tolerate mismatches between the DNA target and the sgRNA spacer, thus leading to potential OFF-target binding/editing events (Hsu et al. 2013; Fu et al. 2013; Pattanayak et al. 2013; Xuebing Wu et al. 2014). Accordingly, several methodologies have been developed to facilitate genome wide and unbiased mapping of spurious DNA double strand break along the genome: IDLV capture (Gabriel et al. 2011; X. Wang et al. 2015), GUIDE-seq (Tsai et al. 2014), HTGTS (Frock et al. 2015), BLESS (Ran et al. 2015; Crosetto et al. 2013), Digenome-seq (D. Kim et al. 2015), CIRCLE-seq (Tsai et al. 2017), SITE-seq(P. Cameron et al. 2017). In an attempt to reduce those events, both sgRNA and Cas9 were modified to increase the system specificity (tru-sgRNA (Fu et al. 2014), sgRNA extension (Cho et al. 2014), SpCas9-HF1 (Kleinstiver et al. 2016), eSpCas9 (Slaymaker et al. 2016)). In this works, both protein- and ASO-responsive iSBH-sgRNAs produce, post-cleavage, spacers that differ in length to the canonical 20nt-long sequence. In particular, the *nano* Csy4-iSBH design presented in chapter 3, generates active sgRNAs with a 10nt spacer. Therefore, it will be important to assess the specificity of different iSBH-sgRNA designs in the ON-state, as well as adapt iSBH-inducibility to work with more specific Cas9 variants.

8.4.2 – Toggle between binding and editing with a single CRISPR-effector

Two studies by Kiani et al. and Dahlman et al. respectively, have shown that CRISPR-based gene regulation (CRISPRr) and CRISPR-based genome editing (CRISPRe) can be simultaneously performed in the same cells using a single CRISPR-effector (Kiani et al. 2015; Dahlman et al. 2015). This was made possible by the discovery that truncated sgRNAs, with spacer segments ranging between 10 and 15nt, prevent Cas9 nuclease activity while maintaining its DNA binding capacity (Kiani et al. 2015; Dahlman et al. 2015). Accordingly, by delivering two sgRNA with 20nt and 15nt-long spacer along with the catalytically active Cas9, both studies were able to simultaneously control the transcriptional output of a specific gene while editing another DNA target. I propose that a single iSBH-sgRNA could be used to toggle between CRISPRr and CRISPRe based on the presence of a particular trigger.

In chapter 2, I presented data showing that SBH-sgRNAs featuring shorter back-folds are able to drive CRISPR-TR (see section 2.3.3). For both spacer:target pairs, it was found that transfection of SBH$^{(10)}$CTS or SBH$^{(15)}$CTS-sgRNAs along with dCas9-VP64 led to transcriptional activation of the reporter gene. Additionally, I successfully demonstrated conditional CRISPRe of an endogenous target using a Csy4-responsive iSBH-sgRNA (see section 8.2.1). Therefore, by fitting the Csy4-PTS, or any other sensing-loop, onto a spacer-blocking hairpin with reduced spacer coverage (e.g. iSBH$^{(15)}$Csy4$^{(full)}$CTS), one could in principle use the cognate trigger to toggle the resulting iSBH-sgRNAs from 15 to 20 available spacer nucleotides (Fig. 8.6a).

A potential application of such iSBH-sgRNAs could be to restrain the expression of a gene circuit to a specific cell population. Since targeting the delivery of genetic material to a specific cell subpopulation has remained relatively challenging, researchers have devised synthetic gene circuits that only activate in a particular cell type/population (Xie et al. 2011). Referred to as “cell classifiers”, such devices exploit divergence in biochemical signatures between cells to identify their origin. Nevertheless, while these circuits are only expected to “activate” in a subset of transfected cells, some components of the device still need to be expressed from the plasmid DNA (pDNA) in the non-targeted population, in order to maintain the system silent. Retention of the pDNA in the non-targeted population as remained problematic as most conditional system are leaky. As such, the presence of the vector in the non-targeted population could potentially be deleterious if, for example, the circuit is supposed to conditionally express pro-apoptotic genes.

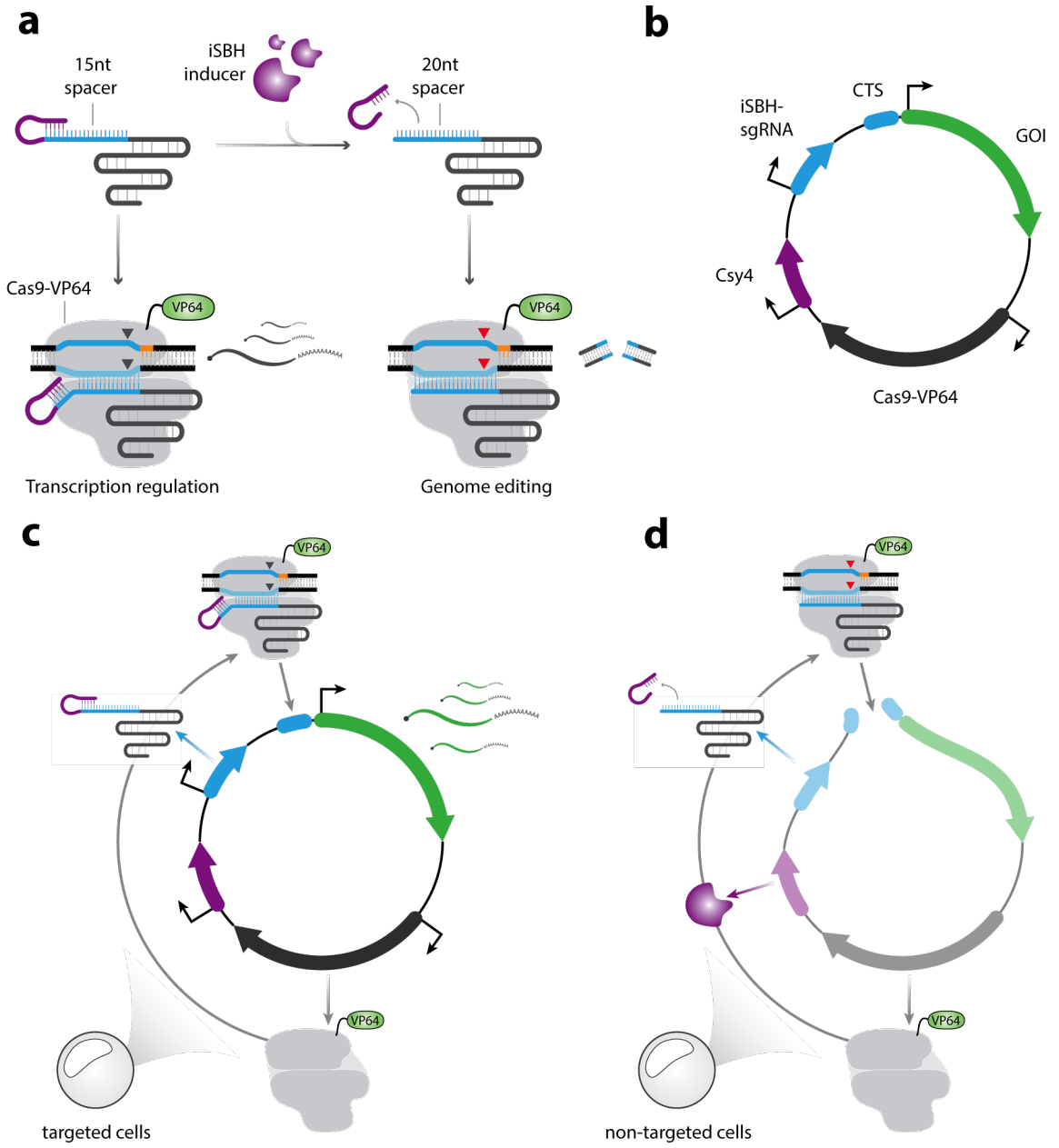

Figure 8.6 – CRISPRr/CRISPRe toggle with iSBH-sgRNA. (a) Conditional toggling between binding and editing is achieved using a iSBH-sgRNA. The iSBH-sgRNA back-fold is shortened so as to leave 15 free spacer nucleotides. In the OFF-state, the iSBH-sgRNA complexes with the catalytically active Cas9 fused to VP64 (Cas9-VP64) to drive transcriptional activation of a target gene. Delivery/Expression of the cognate inducer yields a sgRNA with 20nt-long spacer which resumes the catalytic activity of Cas9-VP64. **(b)** pDNA encoding a gene of interest (GOI) expressed under a CRISPR-responsive promoter, Cas9-VP64, Csy4 (expressed in the non-targeted population), and a toggle iSBH-sgRNA responsive to Csy4 as shown in (a). **(c)** pDNA presented in (b) is transfected in the targeted cell population: Cas9-VP64 complexes with the iSBH-sgRNA to drive GOI expression. **(d)** pDNA presented in (b) is transfected in the non-targeted cell population: Csy4 is expressed, processes the iSBH-sgRNA, which in turn complexes with Cas9-VP64 and cleaves the pDNA.

I reasoned that a solution to this problem could be provided by employing a CRISPRr/ CRISPRe toggle iSBH-sgRNA to create a gene circuit driving a gene of interest (GOI) in the target population and digesting the pDNA encoding the program in the non-target cells[25] (Fig. 8.6b). This would be implemented by encoding on a pDNA the GOI under a minimal promoter, the catalytically active Cas9 fused to VP64, an inducible sgRNA with a iSBH$^{(15)}$Csy4-GOI hairpin, and Csy4 whose expression is spatially controlled (tissue specific promoter, miRNA classifier, etc.). In the target population, the sgRNA only has a 15-nt spacer, and as such is expected to drive GOI expression (Fig. 8.6c). On the other hand, Csy4 would get activated in the non-targeted population, process the iSBH-sgRNA generating a 20nt spacer sgRNA, which in turn would lead to pDNA clearance (Fig. 8.6d).

8.4.3 – Tissue-specific and signal-specific CRISPRa

I also propose that protein-responsive iSBH-sgRNAs could be used to conditionally activate/silence distinct sets of genes in a tissue-specific manner. In theory, this could simply be achieved by placing the various Cas6 inducers under different tissue-specific promoters (Fig. 8.7a). In this case, all iSBH-sgRNAs designed to respond for example to Csy4, will activate/repress their cognate gene targets only in the tissues expressing the inducer (Fig. 8.7b). Other groups have previously demonstrated an alternative strategy for spatial regulation by directly placing the CRISPR-effector under the control of a tissue-specific promoter (Ablain et al. 2015). However, in contrast to the solution proposed here, this approach does not allow differential alteration of transcriptional output in distinct cell types of the same animal.

[25] Others have used pDNA encoding a Cas9 and sgRNA targeted at the pDNA itself, so as to clear the construct after expression of the system components (Caliando & Voigt 2015).

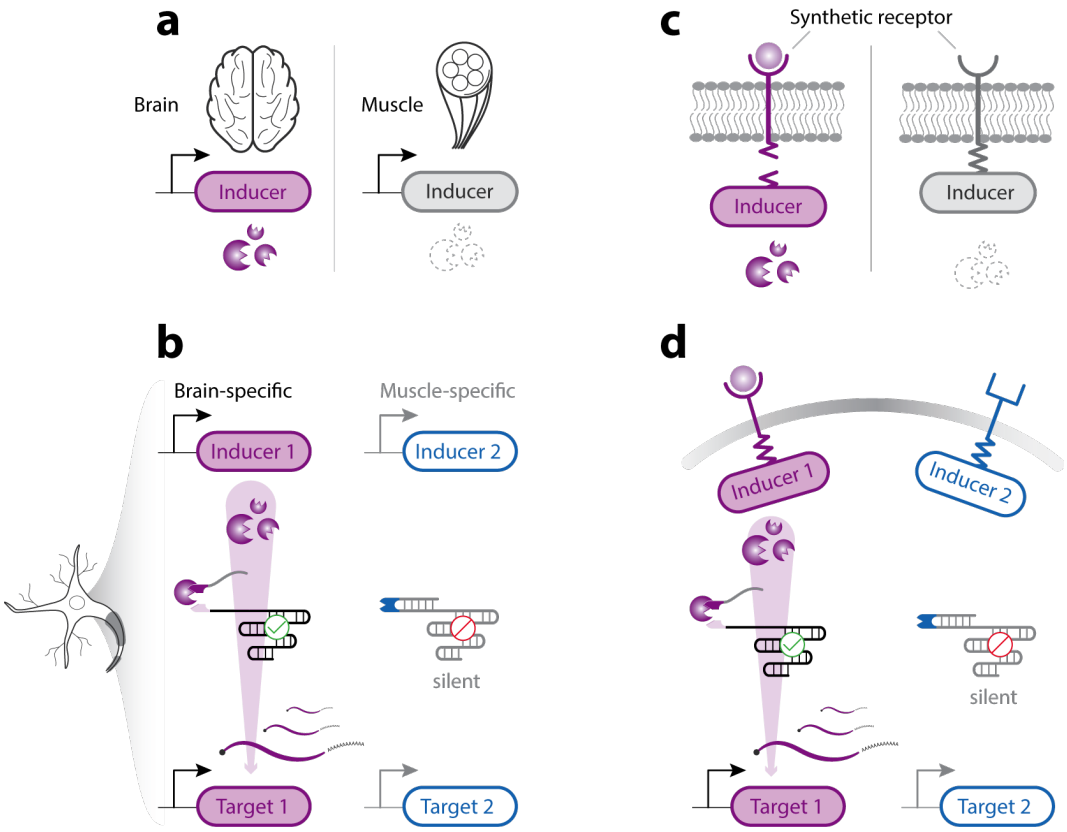

Figure 8.7 – Tissue- and signal-specific CRISPR-TR with protein-iSBHs. (a) Tissue-specific CRISPR-TR is achieved by placing the expression of Cas6 protein inducers under the control of a tissue-specific promoter (e.g. brain). **(b)** Two Cas6 protein inducers are expressed under a brain and muscle specific promoter, respectively. In a neuron population, inducer 1 expression leads to the activation of all cognate iSBH-sgRNAs bearing the corresponding sensing-loop. In turn, genes targeted by the iSBH-sgRNAs are being controlled. In contrast, inducer 2 is not expressed and the corresponding iSBH-sgRNAs remain inert. **(c)** Signal-specific CRISPR-TR is achieved by fusing a Cas6 protein inducer to a synthetic receptor. Presence of the corresponding ligand triggers release of the inducer. **(d)** Two Cas6 protein inducers are fused to two distinct synthetic receptors. In the presence of the extracellular signal activating receptor 1, inducer 1 is freed and imported in to the nucleus where it activates all cognate iSBH-sgRNAs. In turn, genes targeted by the iSBH-sgRNAs are being controlled.

Additionally, by using protein-responsive iSBH-sgRNAs in combination with newly developed synthetic receptors (W. A. Lim 2010; Brenner et al. 2017; Arber et al. 2017), one could achieve signal-specific regulation of downstream gene targets (Fig. 8.7c). Baeumler et al. have recently introduced dCas9-synR, a programmable synthetic receptor that conditionally releases dCas9-VP64 upon sensing a particular extracellular ligand (Baeumler et al. 2017). In this study, native

receptor tyrosine kinase and G-protein-coupled receptor were coupled with an intracellular signaling module composed of a split version of dCas9-VP64, which assembles into a functional CRISRP-TR upon receptor activation. Using this approach, the authors demonstrated activation of therapeutically relevant genes triggered by peptides, proteins, hormones, lipids, and sugars. Since all dCas9-synRs activate dCas9-VP64, it is not possible to have two receptors working orthogonally. I argue that this limitation could be addressed by adapting dCas9-synRs to release Cas6 protein inducers in place of the CRISPR-effector (Fig. 8.7c, d). In this scenario, two receptors can be engineered to respond to different ligands by intracellularly releasing Csy4 and Cas6A, respectively. Translocation of one or both proteins to the nucleus, would in turn lead to the activation of all targets for which an iSBH-sgRNA exists.

8.4.4 – Improved CRISPR-based lineage tracing

The nuclease property of CRISPR-Cas9, and more particularly its ability to precisely alter DNA loci has been recently used to perform lineage tracing in cell culture and model organisms (McKenna et al. 2016; Perli et al. 2016; Kalhor et al. 2016). Cellular labelling was achieved by stably integrating into cells a DNA barcode consisting of an array of CRISPR target sites (CTSs) (Fig. 8.8a). In such cells, expression of the catalytically active Cas9 and sgRNAs programmed to target the CTSs, lead to progressive and permanent editing of the DNA barcode, which mark cells and their descendants over several cellular divisions (Fig. 8.8b). In such endeavor, researchers are trying to increase the number of possible barcodes to differentially label as many cells as possible. To that end, some have created DNA barcodes targeted by as many as ten distinct sgRNAs (McKenna et al. 2016). Nevertheless, because these guides are all delivered simultaneously, and the CTSs cannot be retargeted once edited, tracing rapidly declines after a few generations. As a solution to this problem, other groups have employed self-targeting (or homing) sgRNAs (st-sgRNA), which continue editing themselves through time (Perli et al. 2016; Kalhor et al. 2016). While longer lived, such systems stop editing once the previous edit ablates the PAM embedded in the sgRNA sequence.

Additionally, co-expression of multiple sgRNAs targeting in proximity of the CTS can lead to partial or complete ablation of the DNA barcode.

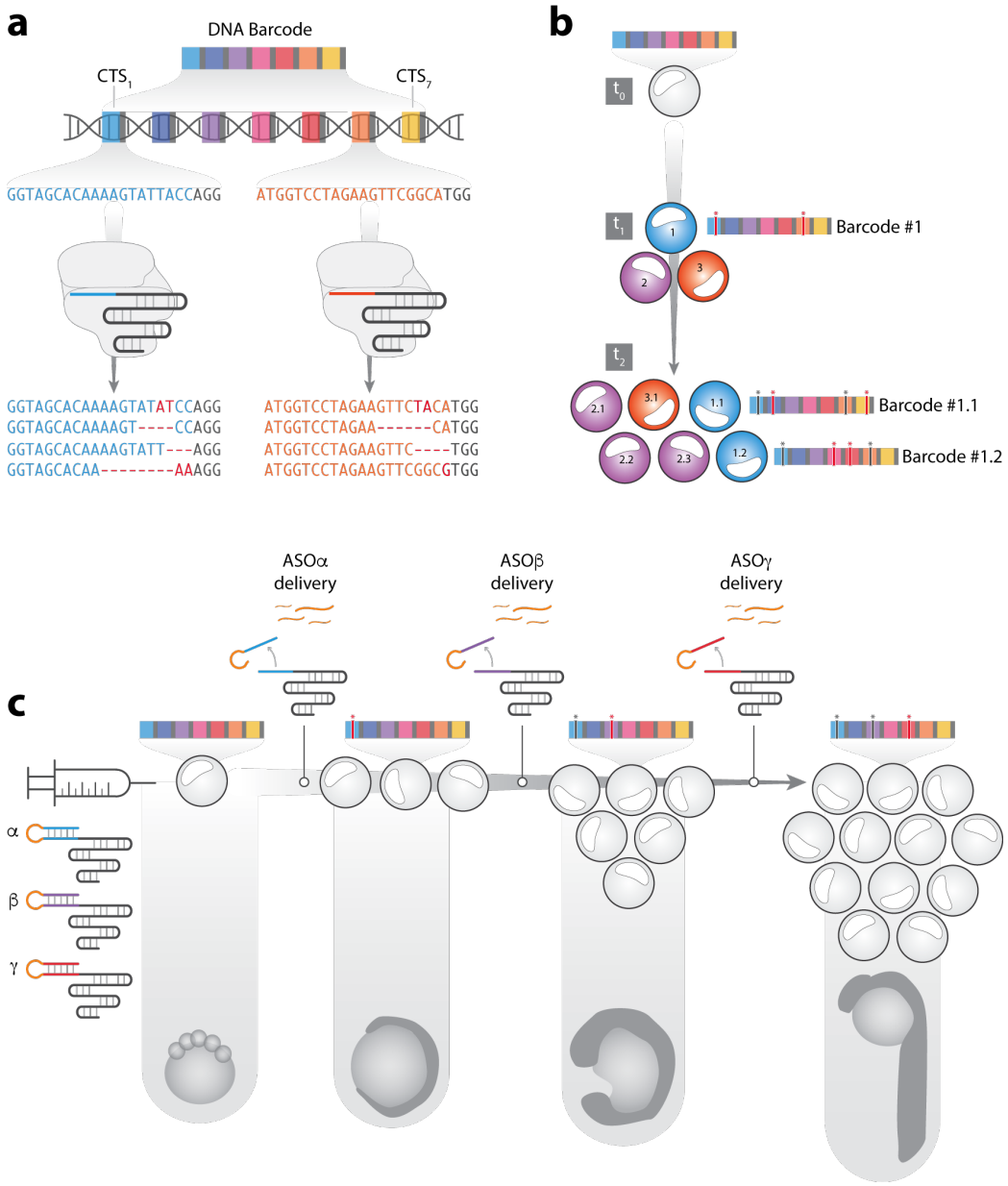

Figure 8.8 – Prolonged lineage tracing with iSBH-sgRNAs. (a) CRISPR-based lineage tracing. A DNA barcode made of distinct CRISPR-target sites (CTSs) is stably integrated into the genome. sgRNAs programmed to target the different CTSs complex with Cas9 to edit the barcode through time via the error prone non-homologous end joining pathway. **(b)** Schematic representation of the expected evolution through time (t0, t1, t2) of the DNA barcode for a given lineage. New editing event are represented as red line capped by an asterisk. **(c)** Use of ASO-responsive iSBH-sgRNAs to prolong lineage tracing throughout the development of a model organism (e.g. zebrafish embryo). Three iSBH-sgRNAs designed to respond to ASOα, ASOβ, and ASOγ, are injected during the first stages of embryogenesis. The guides are then sequentially activated and the DNA barcode in the cells edited accordingly.

Based on these considerations, it is evident that these methods would benefit from the ability to sequentially activate different sgRNAs throughout the duration of the experiment. Accordingly, I envision that iSBH-sgRNAs, and particularly ASO-responsive guides for which an extensive repertoire of inducer:sgRNA exist, could be used to control barcode editing through time (Fig. 8.8c). One could notably create transgenic model organisms expressing a set of ASO-responsive iSBH-sgRNAs each targeting a distinct CTS on the DNA barcode. These guides would then be sequentially activated throughout development by systemic delivery of their cognate ASO trigger.

8.4.5 – iSBH-based miRNA profiling for early cancer diagnosis

MicroRNAs (miRNAs) are short 22nt long non-coding RNAs involved in the post-transcriptional regulation of gene expression (H. Guo et al. 2010). Upon maturation, miRNAs are loaded in the miRNA-induced silencing complex to target, via Watson-Crick base-pairing, mRNAs containing one or several matching miRNA response elements (MREs). Hybridization between miRNA and MRE triggers a stepwise process leading to mRNA decay (Béthune et al. 2012; Bazzini et al. 2012; Valencia-Sanchez et al. 2006). Aside from various silencing mechanisms, miRNAs can also promote Argonaute protein Ago2-mediated slicing of target RNAs (Tuschl et al. 1999; Hammond et al. 2000; Allen et al. 2005). Directly relevant to the design of iSBH-sgRNAs, it was reported that perfect base pairing between the miRNA and target MRE tends to favor this pathway over other routes of repression (Meister et al. 2004; Yekta et al. 2004; H.-S. Guo et al. 2005).

Based on these mechanistic insights, I propose that miRNA-responsive iSBH-sgRNAs could be engineered by using a perfect MRE sequence as sensing-loop. Nevertheless, based on the fact that miRNA-mediated slicing mainly takes place in the cytoplasm (Leung 2015), it is unlikely that such miRNA-responsive sgRNAs would activate in eukaryotic cells due to cellular compartmentalization (U6 expressed sgRNAs mainly reside in the nucleus (Hongming Ma et al. 2014)). However, I envision that this strategy could be used to perform miRNA profiling *in vitro*.

In theory, one could test the presence of a given Ago-loaded miRNA X (miR-X) in a cell lysate by mixing it with a cell-free system[26] containing dCas9-VP64, reporter gene, and a miR-X responsive iSBH-sgRNA (Fig. 8.9a). In the presence of the miRNA, the guide should be activated and able to drive reporter expression, hence providing a read out indicating the presence of miR-X. Therefore, using an array of such iSBH-sgRNAs programmed to respond to a specific set of miRNAs, this strategy could in theory be used to profile the miRNA composition of a cellular lysate (Fig. 8.9b). Since oncogenesis is often accompanied by aberrant miRNA profiles that can be detected in the blood of patients (G. Cheng 2015), this strategy could notably be of interest to perform early cancer diagnosis from a single liquid biopsy.

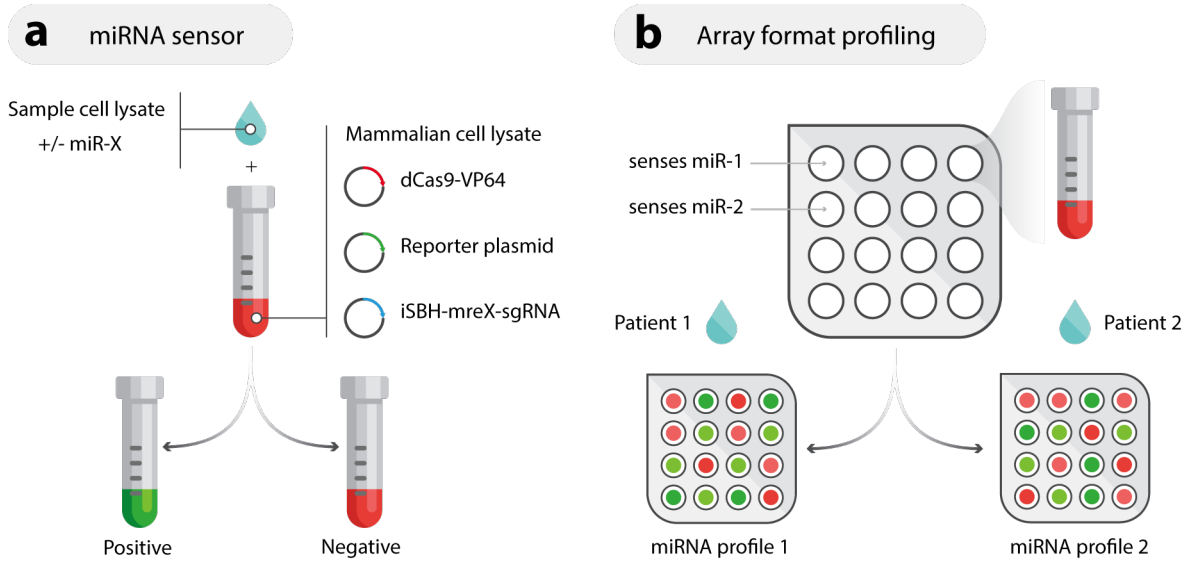

Figure 8.9 – Early cancer diagnosis based on miRNA profiling. (a) miRNA-responsive iSBH-sgRNAs are used to detect the presence of a miRNA of interest (miR-X) in a sample cell lysate. A cell free system is supplemented with dCas9-VP64, reporter construct, and an iSBH-sgRNA designed to sense miR-X (iSBH-mreX-sgRNA, miRNA responsive element of miR-X is used as sensing-loop). Mixing the sample with the reaction containing the sensor gives either a positive or negative outcome (reporter expression) depending on the presence of Ago bound miR-X in the sample. **(b)** Cell lysates containing dCas9-VP64, reporter gene, and iSBH-sgRNAs sensing distinct miRNAs (miR-1, miR-2, etc.) are organized in an array format. A patient sample is spiked in all lysates of a same plate producing a miRNA profile for the patient.

[26] Cell lysate containing all the cellular machinery necessary for transcription and translation. Such lysates are commercially available or can be made in-house.

8.5 – General conclusion

Synthetic transcriptional regulators that can be programmed to control the expression of any gene along the genome, constitute a key part of the synthetic biology toolbox. Gene regulation being at the center of most biological processes, these devices are expected to become instrumental in dissecting the intricate wiring of the cell, as well as developing the next generation of “smart” therapeutics.

First engineered in 2013 to work in mammalian cells, transcriptional regulators adapted from the bacterial CRISPR-Cas9 system have offered a simpler, cheaper, and more versatile alternative to zinc-finger and TALE-based approaches. Leveraging the DNA binding properties of the sgRNA-dCas9 complex, CRISPR-TRs were constructed by tethering to the crRNP various types and combinations of effector domains, that can enhance or reduce the transcriptional output of promoters. In that case, a simple reprogramming of the spacer segment of the sgRNA was required to direct the regulator to a sequence complementary DNA locus along the genome.

To provide the tools necessary to mimic the dynamics nature of native transcriptional programs, I have set out to develop inducible CRISPR-TR using minimal RNA engineering of the sgRNAs. In the search for a solution, I have created a new platform for the rational design of switchable guide RNAs, which exploits reversible spacer sequestering as a means to control CRISPR-TR activity. This was achieved by modifying the sgRNA to form inducible spacer-blocking hairpins on the 5’-end of its sequence. Using this strategy, I report in this Thesis the construction of a series of iSBH-sgRNAs, which were successfully employed to condition CRISPR-TR activity on the presence of protein, oligonucleotide, and small molecule inducers.

Demonstrating the relevance of this methodology, I combined iSBH-based inducibility with the state of the art SAM regulator to assemble fundamental gene circuits implementing the parallel and independent control of endogenous gene targets in human cells. Additionally, in an attempt to help the adoption of the technology by others, I developed a software which fully

automates the design of protein- and ASO-responsive iSBH-sgRNAs. Due to the modularity of its design, the low cost and ease of its implementation, and the ability to create and multiplex orthogonal systems, I believe that the iSBH will constitute a valuable research tool for biologists. Not only restricted to the modulation of gene expression, this platform could also be used to achieve spatio-temporal control over genome and epigenome editing, as well as all other applications based on CRISPR-Cas systems.

Chapter 9 – References

Ablain, J. et al., 2015. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Developmental cell*, 32(6), pp.756–764.

Abudayyeh, O.O. et al., 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (New York, N.Y.)*, p.aaf5573.

Agrawal, S. & Kandimalla, E.R., 2000. Antisense therapeutics: is it as simple as complementary base recognition? *Molecular medicine today*, 6(2), pp.72–81.

Allen, E. et al., 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*, 121(2), pp.207–221.

Amaran, S. et al., 2016. Simulation optimization: a review of algorithms and applications. *Annals of Operations Research*, 240(1), pp.351–380.

An, C.-I., Trinh, V.B. & Yokobayashi, Y., 2006. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA (New York, N.Y.)*, 12(5), pp.710–716.

Anders, C. et al., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513(7519), pp.569-573.

Anders, C., Bargsten, K. & Jinek, M., 2016. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Molecular cell*, 61(6), pp.895–902.

Arber, C., Young, M. & Barth, P., 2017. Reprogramming cellular functions with engineered membrane proteins. *Current opinion in biotechnology*, 47, pp.92–101.

Aricescu, A.R., Lu, W. & Jones, E.Y., 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 10), pp.1243–1250.

Atassi, M.Z. & Casali, P., 2008. Molecular mechanisms of autoimmunity. *Autoimmunity*, 41(2), pp.123–132.

Aubrey, B.J. et al., 2015. An Inducible Lentiviral Guide RNA Platform Enables the Identification of Tumor-Essential Genes and Tumor-Promoting Mutations In Vivo. *Cell reports*, 10(8), pp.1422–1432.

Ausländer, S. & Fussenegger, M., 2016. Engineering Gene Circuits for Mammalian Cell-Based Applications. *Cold Spring Harbor perspectives in biology*, 8(7). pii: a023895.

Ausländer, S. & Fussenegger, M., 2013. From gene switches to mammalian designer cells: present and future prospects. *Trends in biotechnology*, 31(3), pp.155–168.

Ausländer, S. et al., 2012. Programmable single-cell mammalian biocomputers. *Nature*, 487(7405), pp.123–127.

Ausländer, S., Ketzer, P. & Hartig, J.S., 2010. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Molecular bioSystems*, 6(5), pp.807–814.

Baeumler, T.A., Ahmed, A.A. & Fulga, T.A., 2017. Engineering Synthetic Signaling Pathways with Programmable dCas9-Based Chimeric Receptors. *Cell reports*, 20(11), pp.2639–2653.

Banaszynski, L.A. et al., 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, 126(5), pp.995–1004.

Bao, Z. et al., 2017. Orthogonal genetic regulation in human cells using chemically induced CRISPR/Cas9 activators. *ACS synthetic biology*, 6(4), pp.686-693.

Barrangou, R. et al., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819), pp.1709–1712.

Bazzini, A.A., Lee, M.T. & Giraldez, A.J., 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science (New York, N.Y.)*, 336(6078), pp.233–237.

Becskei, A. & Serrano, L., 2000. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786), pp.590–593.

Beerli, R.R. & Barbas, C.F., 2002. Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology*, 20(2), pp.135–141.

Beerli, R.R. et al., 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), pp.14628–14633.

Beerli, R.R., Dreier, B. & Barbas, C.F., 2000. Positive and negative regulation of endogenous genes by designed transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), pp.1495–1500.

Beisel, C.L. et al., 2011. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic acids research*, 39(7), pp.2981–2994.

Bennett, C.F. & Swayze, E.E., 2010. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annual review of pharmacology and toxicology*, 50(1), pp.259–293.

Berschneider, B. et al., 2009. Small-molecule-dependent regulation of transfer RNA in bacteria. *Angewandte Chemie (International ed. in English)*, 48(41), pp.7564–7567.

Béthune, J., Artus-Revel, C.G. & Filipowicz, W., 2012. Kinetic analysis reveals successive

steps leading to miRNA-mediated silencing in mammalian cells. *EMBO reports*, 13(8), pp.716–723.

Biase, F.H., Cao, X. & Zhong, S., 2014. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome research*, 24(11), pp.1787–1796.

Bikard, D. et al., 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research*, 41(15), pp.7429–7437.

Birikh, K.R., Heaton, P.A. & Eckstein, F., 1997. The structure, function and application of the hammerhead ribozyme. *European journal of biochemistry / FEBS*, 245(1), pp.1–16.

Black, J.B., Perez-Pinera, P. & Gersbach, C.A., 2017. Mammalian Synthetic Biology: Engineering Biological Systems. *Annual review of biomedical engineering*, 19(1), pp.249–277.

Blount, Z.D., 2015. The unexhausted potential of E. coli. *eLife*, 4, p.96.

Boch, J. & Bonas, U., 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*, 48(1), pp.419–436.

Boch, J. et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, 326(5959), pp.1509–1512.

Boettcher, M. & McManus, M.T., 2015. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular cell*, 58(4), pp.575–585.

Bolotin, A. et al., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)*, 151(Pt 8), pp.2551–2561.

Bonger, K.M. et al., 2011. Small-molecule displacement of a cryptic degron causes conditional protein degradation. *Nature chemical biology*, 7(8), pp.531–537.

Braff, J.L. et al., 2016. Characterization of Cas9-Guide RNA Orthologs. *Cold Spring Harbor protocols*, 2016(5), p.pdb.top086793.

Brand, A.H. & Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)*, 118(2), pp.401–415.

Brenner, M., Cho, J.H. & Wong, W.W., 2017. Synthetic biology: Sensing with modular receptors. *Nature chemical biology*, 13(2), pp.131–132.

Briner, A.E. et al., 2014. Guide RNA functional modules direct cas9 activity and orthogonality. *Molecular cell*, 56(2), pp.333–339.

Brouns, S.J.J. et al., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (New York, N.Y.)*, 321(5891), pp.960–964.

Brown, M. et al., 1987. lac repressor can regulate expression from a hybrid SV40 early promoter containing a lac operator in animal cells. *Cell*, 49(5), pp.603–612.

Burz, D.S. et al., 1994. Self-assembly of bacteriophage lambda cI repressor: effects of single-site mutations on the monomer-dimer equilibrium. *Biochemistry*, 33(28), pp.8399–8405.

Buskirk, A.R. et al., 2004. Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(29), pp.10505–10510.

Caban, K., Kinzy, S.A. & Copeland, P.R., 2007. The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2. *Molecular and cellular biology*, 27(18), pp.6350–6360.

Caliando, B.J. & Voigt, C.A., 2015. Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nature communications*, 6, p.6989.

Cameron, D.E., Bashor, C.J. & Collins, J.J., 2014. A brief history of synthetic biology. *Nature reviews. Microbiology*, 12(5), pp.381–390.

Cameron, P. et al., 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nature methods*, 14(6), pp.600-606.

Canny, M.D. et al., 2004. Fast cleavage kinetics of a natural hammerhead ribozyme. *Journal of the American Chemical Society*, 126(35), pp.10848–10849.

Cao, J. et al., 2016. An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic acids research*, 44(19), pp.e149.

Carroll, D., 2014. Genome engineering with targetable nucleases. *Annual review of biochemistry*, 83, pp.409–439.

Chatr-Aryamontri, A. et al., 2017. The BioGRID interaction database: 2017 update. *Nucleic acids research*, 45(D1), pp.D369–D379.

Chao, J.A. et al., 2008. Structural basis for the coevolution of a viral RNA-protein complex. *Nature structural & molecular biology*, 15(1), pp.103–105.

Chavez, A. et al., 2016. Comparison of Cas9 activators in multiple species. *Nature methods*, 13(7), pp.563–567.

Chavez, A. et al., 2015. Highly efficient Cas9-mediated transcriptional programming. *Nature methods*, 12(4), pp.326–328.

Chen, B. et al., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), pp.1479–1491.

Chen, D.S. & Mellman, I., 2013. Oncology meets immunology: the cancer-immunity cycle. *Immunity*, 39(1), pp.1–10.

Chen, D.S., Irving, B.A. & Hodi, F.S., 2012. Molecular pathways: next-generation immunotherapy--inhibiting programmed death-ligand 1 and programmed death-1. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 18(24), pp.6580–6587.

Chen, Y.Y., Jensen, M.C. & Smolke, C.D., 2010. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), pp.8531–8536.

Cheng, A.W. et al., 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research*, 23(10), pp.1163–1171.

Cheng, G., 2015. Circulating miRNAs: roles in cancer diagnosis, prognosis and therapy. *Advanced Drug Delivery Reviews*, 81, pp.75–93.

Chin, J.W., 2017. Expanding and reprogramming the genetic code. *Nature*, 550(7674), pp.53–60.

Chira, S. et al., 2017. CRISPR/Cas9: Transcending the Reality of Genome Editing. *Molecular therapy. Nucleic acids*, 7, pp.211–222.

Cho, S.W. et al., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), pp.132–141.

Cho, S.W. et al., 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31(3), pp.230–232.

Cong, L. & Zhang, F., 2015. Genome engineering using CRISPR-Cas9 system. *Methods in molecular biology (Clifton, N.J.)*, 1239(Chapter 10), pp.197–217.

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)*, 339(6121), pp.819–823.

Cress, B.F. et al., 2016. Rapid generation of CRISPR/dCas9-regulated, orthogonally repressible hybrid T7-lac promoters for modular, tuneable control of metabolic pathway fluxes in Escherichia coli. *Nucleic acids research*, 44(9), pp.gkw231–4485.

Cronin, C.A., Gluba, W. & Scrable, H., 2001. The lac operator-repressor system is functional in the mouse. *Genes & development*, 15(12), pp.1506–1517.

Crosetto, N. et al., 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods*, 10(4), pp.361–365.

Culler, S.J., Hoff, K.G. & Smolke, C.D., 2010. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science (New York, N.Y.)*, 330(6008), pp.1251–1255.

Daber, R. et al., 2007. Structural analysis of lac repressor bound to allosteric effectors. *Journal of molecular biology*, 370(4), pp.609–619.

Dagle, J.M., Walder, J.A. & Weeks, D.L., 1990. Targeted degradation of mRNA in Xenopus oocytes and embryos directed by modified oligonucleotides: studies of An2 and cyclin in embryogenesis. *Nucleic acids research*, 18(16), pp.4751–4757.

Dagle, J.M., Weeks, D.L. & Walder, J.A., 1991. Pathways of degradation and mechanism of action of antisense oligonucleotides in Xenopus laevis embryos. *Antisense research and development*, 1(1), pp.11–20.

Dahlman, J.E. et al., 2015. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nature biotechnology*, 33(11), pp.1159–1161.

Dalkara, D. et al., 2013. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Science translational medicine*, 5(189), pp.189ra76–189ra76.

Datsenko, K.A. et al., 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications*, 3, p.945.

Davis, K.M. et al., 2015. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology*, 11(5), pp.316-318.

Davis, L. & Chin, J.W., 2012. Designer proteins: applications of genetic code expansion in cell biology. *Nature reviews. Molecular cell biology*, 13(3), pp.168–182.

de Silva, C. & Walter, N.G., 2009. Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *RNA (New York, N.Y.)*, 15(1), pp.76–84.

de Solis, C.A. et al., 2016. The Development of a Viral Mediated CRISPR/Cas9 System with Doxycycline Dependent gRNA Expression for Inducible In vitro and In vivo Genome Editing. *Frontiers in molecular neuroscience*, 9, p.70.

Deltcheva, E. et al., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602–607.

Deng, D. et al., 2012. Recognition of methylated DNA by TAL effectors. *Cell research*, 22(10), pp.1502–1504.

Deplazes, A., 2009. Piecing together a puzzle. An exposition of synthetic biology. *EMBO reports*, 10(5), pp.428–432.

Dias, N. & Stein, C.A., 2002. Antisense oligonucleotides: basic concepts and mechanisms. *Molecular cancer therapeutics*, 1(5), pp.347–355.

Ding, Y. & Lawrence, C.E., 2001. Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic acids research*, 29(5), pp.1034–1046.

Dominguez, A.A., Lim, W.A. & Qi, L.S., 2016. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nature reviews. Molecular cell biology*, 17(1), pp.5–15.

Dong, S., Rogan, S.C. & Roth, B.L., 2010. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nature protocols*, 5(3), pp.561–573.

Donis-Keller, H., 1979. Site specific enzymatic cleavage of RNA. *Nucleic acids research*, 7(1), pp.179–192.

Doudna, J.A. & Charpentier, E., 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, 346(6213), pp.1258096–1258096.

Dow, L.E. et al., 2015. Inducible in vivo genome editing with CRISPR-Cas9. *Nature biotechnology*, 33(4), pp.390–394.

Dreier, B. et al., 2001. Development of zinc finger domains for recognition of the 5“-ANN-3” family of DNA sequences and their use in the construction of artificial transcription factors. *The Journal of biological chemistry*, 276(31), pp.29466–29478.

Du, P. et al., 2015. Engineering translational activators with CRISPR-Cas system. *ACS synthetic biology*, 5(1), pp.74-80.

Ebert, M.S., Neilson, J.R. & Sharp, P.A., 2007. MicroRNA sponges: competitive inhibitors of

small RNAs in mammalian cells. *Nature methods*, 4(9), pp.721–726.

Eder, P.S. & Walder, J.A., 1991. Ribonuclease H from K562 human erythroleukemia cells. Purification, characterization, and substrate specificity. *The Journal of biological chemistry*, 266(10), pp.6472–6479.

Ellington, A.D. & Szostak, J.W., 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287), pp.818–822.

Elowitz, M.B. & Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), pp.335–338.

Elowitz, M. & Lim, W.A., 2010. Build life to understand it. *Nature*, 468(7326), pp.889–890.

Esvelt, K.M. et al., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, 10(11), pp.1116–1121.

Farzadfard, F., Perli, S.D. & Lu, T.K., 2013. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS synthetic biology*, 2(10), pp.604–613.

Fellmann, C. et al., 2017. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nature reviews. Drug discovery*, 16(2), pp.89–100.

Feng, J. et al., 2015. A general strategy to construct small molecule biosensors in eukaryotes. J. W. Kelly, ed. *eLife*, 4, p.e10606.

Ferré-D'Amaré, A.R. & Scott, W.G., 2010. Small self-cleaving ribozymes. *Cold Spring Harbor perspectives in biology*, 2(10), pp.a003574–a003574.

Fisher, T.L. et al., 1993. Intracellular disposition and metabolism of fluorescently-labeled unmodified and modified oligonucleotides microinjected into mammalian cells. *Nucleic acids research*, 21(16), pp.3857–3865.

Frock, R.L. et al., 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature biotechnology*, 33(2), pp.179–186.

Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology*, 31(9), pp.822–826.

Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(3), pp.279–284.

Gabriel, R. et al., 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nature biotechnology*, 29(9), pp.816–823.

Gaj, T. et al., 2016. Genome-Editing Technologies: Principles and Applications. *Cold Spring Harbor perspectives in biology*, 8(12), p.a023754.

Gander, M.W. et al., 2017. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nature communications*, 8, p.15459.

Gao, X. et al., 2014. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic acids research*, 42(20), pp.e155–e155.

Gao, Y. et al., 2016. Complex transcriptional modulation with orthogonal and inducible dCas9

regulators. *Nature methods*, 13(12), pp.1043-1049.

García-Otín, A.L. & Guillou, F., 2006. Mammalian genome targeting using site-specific recombinases. *Frontiers in bioscience : a journal and virtual library*, 11, pp.1108–1136.

Gardner, T.S., Cantor, C.R. & Collins, J.J., 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767), pp.339–342.

Garneau, J.E. et al., 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), pp.67–71.

Gasiunas, G. et al., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp.E2579–86.

Gesner, E.M. et al., 2011. Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology*, 18(6), pp.688–692.

Gilbert, L.A. et al., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), pp.442–451.

Gilbert, L.A. et al., 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, 159(3), pp.647-61.

Givan, A.L., 2011. Flow cytometry: an introduction. *Methods in molecular biology (Clifton, N.J.)*, 699(Chapter 1), pp.1–29.

Goler, J.A., Carothers, J.M. & Keasling, J.D., 2014. Dual-selection for evolution of in vivo functional aptazymes as riboswitch parts. *Methods in molecular biology (Clifton, N.J.)*, 1111(Chapter 16), pp.221–235.

Goomer, R.S. & Kunkel, G.R., 1992. The transcriptional start site for a human U6 small nuclear RNA gene is dictated by a compound promoter element consisting of the PSE and the TATA box. *Nucleic acids research*, 20(18), pp.4903–4912.

Gossen, M. & Bujard, H., 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 89(12), pp.5547–5551.

Graham, F.L. et al., 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of general virology*, 36(1), pp.59–74.

Grissa, I., Vergnaud, G. & Pourcel, C., 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics*, 8(1), p.172.

Gross, G., Waks, T. & Eshhar, Z., 1989. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 86(24), pp.10024–10028.

Grünweller, A. et al., 2003. Comparison of different antisense strategies in mammalian cells using locked nucleic acids, 2'-O-methyl RNA, phosphorothioates and small interfering RNA. *Nucleic acids research*, 31(12), pp.3185–3193.

Gu, K. et al., 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, 435(7045), pp.1122–1125.

Guernet, A. et al., 2016. Modeling intratumor heterogeneity through CRISPR-barcodes. *Molecular & cellular oncology*, 3(6), p.e1227894.

Guo, H. et al., 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308), pp.835–840.

Guo, H.-S. et al., 2005. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *The Plant cell*, 17(5), pp.1376–1386.

Guo, J. et al., 2017. An inducible CRISPR-ON system for controllable gene activation in human pluripotent stem cells. *Protein & cell*, pp.1–15.

Haft, D.H. et al., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, 1(6), p.e60.

Hale, C.R. et al., 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, 139(5), pp.945–956.

Hammann, C. & Lilley, D.M.J., 2002. Folding and activity of the hammerhead ribozyme. *Chembiochem : a European journal of chemical biology*, 3(8), pp.690–700.

Hammann, C. et al., 2012. The ubiquitous hammerhead ribozyme. *RNA (New York, N.Y.)*, 18(5), pp.871–885.

Hammond, S.M. et al., 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature*, 404(6775), pp.293–296.

Hanahan, D. & Weinberg, R.A., 2000. The hallmarks of cancer. *Cell*, 100(1), pp.57–70.

Hannig, G. & Makrides, S.C., 1998. Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in biotechnology*, 16(2), pp.54–60.

Hardwick, J.M. et al., 1992. The Epstein-Barr virus R transactivator (Rta) contains a complex, potent activation domain with properties different from those of VP16. *Journal of virology*, 66(9), pp.5500–5508.

Haseloff, J. & Gerlach, W.L., 1989. Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. *Gene*, 82(1), pp.43–52.

Hathaway, N.A. et al., 2012. Dynamics and memory of heterochromatin in living cells. *Cell*, 149(7), pp.1447–1460.

Haurwitz, R.E. et al., 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York, N.Y.)*, 329(5997), pp.1355–1358.

Heler, R. et al., 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature*, 519(7542), pp.199-202.

Hellman, L.M. & Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols*, 2(8), pp.1849–1861.

Hemphill, J. et al., 2015. Optical Control of CRISPR/Cas9 Gene Editing. *Journal of the American Chemical Society*, 37(17), pp.5642-5645.

Hermann, T. & Patel, D.J., 2000. Adaptive recognition by nucleic acid aptamers. *Science (New York, N.Y.)*, 287(5454), pp.820–825.

Hilton, I.B. & Gersbach, C.A., 2016. Genetic engineering: Chemical control for CRISPR editing. *Nature chemical biology*, 13(1), pp.2-3.

Hilton, I.B. et al., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology*, 33(5), pp.510-517.

Hirosawa, M. et al., 2017. Cell-type-specific genome editing with a microRNA-responsive CRISPR-Cas9 switch. *Nucleic acids research*, 45(13), p.e118.

Hobom, B., 1980. Gene surgery: on the threshold of synthetic biology. *Medizinische Klinik*, 75(24), pp.834–841.

Hochstrasser, M.L. & Doudna, J.A., 2015. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in biochemical sciences*, 40(1), pp.58–66.

Holkers, M. et al., 2013. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic acids research*, 41(5), pp.e63–e63.

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31(9), pp.827–832.

Hsu, P.D., Lander, E.S. & Zhang, F., 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262–1278.

Hu, J. et al., 2013. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nature biotechnology*, 31(6), pp.522–529.

Huang, L. & Lilley, D.M.J., 2013. The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA (New York, N.Y.)*, 19(12), pp.1703–1710.

Hutchison, C.A. et al., 2016. Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)*, 351(6280), p.aad6253.

Iannitti, T., Morales-Medina, J.C. & Palmieri, B., 2014. Phosphorothioate oligonucleotides: effectiveness and toxicity. *Current drug targets*, 15(7), pp.663–673.

Ishino, Y. et al., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429–5433.

Jain, P.K. et al., 2016. Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors. *Angewandte Chemie (International ed. in English)*, 55(40), pp.12440-12444.

James, J.R. & Vale, R.D., 2012. Biophysical mechanism of T-cell receptor triggering in a reconstituted system. *Nature*, 487(7405), pp.64–69.

Jansen, R. et al., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, 43(6), pp.1565–1575.

Jensen, L.J. et al., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research*, 37(Database issue), pp.D412–6.

Jiang, F. & Doudna, J.A., 2017. CRISPR-Cas9 Structures and Mechanisms. *Annual review of biophysics*, 46(1), pp.505–529.

Jiang, F. et al., 2015. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science (New York, N.Y.)*, 348(6242), pp.1477–1481.

Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–821.

Jinek, M. et al., 2013. RNA-programmed genome editing in human cells. D. Weigel, ed. *eLife*, 2, p.e00471.

Jinek, M. et al., 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, 343(6176), pp.1247997–1247997.

Josephs, E.A. et al., 2015. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic acids research*, 43(18), pp.gkv892–8941.

Joung, J. et al., 2017. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature protocols*, 12(4), pp.828–863.

Jusiak, B. et al., 2016. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends in biotechnology*, 34(7), pp.535–547.

Kakidani, H. & Ptashne, M., 1988. GAL4 activates gene expression in mammalian cells. *Cell*, 52(2), pp.161–167.

Kalhor, R., Mali, P. & Church, G.M., 2016. Rapidly evolving homing CRISPR barcodes. *Nature methods*, 14(2), pp.195-200.

Kang, K.-N. & Lee, Y.-S., 2013. RNA aptamers: a review of recent trends and applications. *Advances in biochemical engineering/biotechnology*, 131(Chapter 136), pp.153–169.

Karr, J.R. et al., 2012. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2), pp.389–401.

Kataoka, K. et al., 2016. Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature*, 534(7607), pp.402–406.

Kawano, F. et al., 2015. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nature communications*, 6, p.6256.

Kemmer, C. et al., 2010. Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nature biotechnology*, 28(4), pp.355–360.

Kennedy, M.J. et al., 2010. Rapid blue-light-mediated induction of protein interactions in living cells. *Nature methods*, 7(12), pp.973–975.

Khalil, A.S. & Collins, J.J., 2010. Synthetic biology: applications come of age. *Nature reviews. Genetics*, 11(5), pp.367–379.

Khalil, A.S. et al., 2012. A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 150(3), pp.647–658.

Khvorova, A. & Watts, J.K., 2017. The chemical evolution of oligonucleotide therapies of

clinical utility. *Nature biotechnology*, 35(3), pp.238–248.

Khvorova, A. et al., 2003. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature structural biology*, 10(9), pp.708–712.

Kiani, S. et al., 2015. Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods*, 12(11), pp.1051–1054.

Kiani, S. et al., 2014. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature methods*, 11(7), pp.723–726.

Kim, D. et al., 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods*, 12(3), pp.237-243.

Kim, D.-S. et al., 2008. Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC molecular biology*, 9(1), p.23.

Kim, H. & Kim, J.-S., 2014. A guide to genome engineering with programmable nucleases. *Nature reviews. Genetics*, 15(5), pp.321–334.

Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), pp.1156–1160.

Kis, Z. et al., 2015. Mammalian synthetic biology: emerging medical applications. *Journal of the Royal Society, Interface / the Royal Society*, 12(106), pp.20141000–20141000.

Klann, T.S. et al., 2017. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nature biotechnology*, 7, p.46545.

Klein, D.J. et al., 2001. The kink-turn: a new RNA secondary structure motif. *The EMBO journal*, 20(15), pp.4214–4221.

Kleinstiver, B.P. et al., 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529(7587), pp.490–495.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., et al., 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature biotechnology*, 33(12), pp.1293-1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., et al., 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, 523(7561), pp.481-485.

Klug, A. & Rhodes, D., 1987. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harbor symposia on quantitative biology*, 52, pp.473–482.

Kobayashi, H. et al., 2004. Programmable cells: interfacing natural and engineered gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22), pp.8414–8419.

Kole, R., Krainer, A.R. & Altman, S., 2012. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature reviews. Drug discovery*, 11(2), pp.125–140.

Komor, A.C., Badran, A.H. & Liu, D.R., 2016. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, 169(3), p.559.

Komor, A.C., Kim, Y.B., et al., 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, 533(7603), pp.420–424.

Konermann, S. et al., 2014. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517(7536), pp.583-588.

Konermann, S. et al., 2013. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500(7463), pp.472–476.

Koonin, E.V. & Wolf, Y.I., 2015. Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus-host coevolution. *Molecular bioSystems*, 11(1), pp.20–27.

Koonin, E.V., Makarova, K.S. & Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology*, 37, pp.67–78.

Korman, A.J., Peggs, K.S. & Allison, J.P., 2006. Checkpoint blockade in cancer immunotherapy. *Advances in immunology*, 90, pp.297–339.

Kumar, D., Kim, S.H. & Yokobayashi, Y., 2011. Combinatorially inducible RNA interference triggered by chemically modified oligonucleotides. *Journal of the American Chemical Society*, 133(8), pp.2783–2788.

Kungulovski, G. & Jeltsch, A., 2016. Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends in genetics : TIG*, 32(2), pp.101–113.

Kurreck, J. et al., 2002. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic acids research*, 30(9), pp.1911–1918.

Laufer, B.I. & Singh, S.M., 2015. Strategies for precision modulation of gene expression by epigenome editing: an overview. *Epigenetics & chromatin*, 8(1), p.34.

Lawhorn, I.E.B., Ferreira, J.P. & Wang, C.L., 2014. Evaluation of sgRNA target sites for CRISPR-mediated repression of TP53. R. A. M. de Bruin, ed. *PloS one*, 9(11), p.e113232.

Lebar, T. & Jerala, R., 2016. Benchmarking of TALE- and CRISPR/dCas9-Based Transcriptional Regulators in Mammalian Cells for the Construction of Synthetic Genetic Circuits. *ACS synthetic biology*, p.acssynbio.5b00259.

Lee, H.Y. et al., 2013. RNA-protein analysis using a conditional CRISPR nuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), pp.5416–5421.

Lee, Y.J. et al., 2016. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic acids research*, 44(5), pp.gkw056–2473.

Leenay, R.T. & Beisel, C.L., 2016. Deciphering, communicating, and engineering the CRISPR PAM. *Journal of molecular biology*, 429(2), pp.177-191.

Lelli, K.M., Slattery, M. & Mann, R.S., 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annual review of genetics*, 46(1), pp.43–68.

Lennox, K.A. et al., 2006. Characterization of modified antisense oligonucleotides in Xenopus laevis embryos. *Oligonucleotides*, 16(1), pp.26–42.

Leung, A.K.L., 2015. The Whereabouts of microRNA Actions: Cytoplasm and Beyond. *Trends in cell biology*, 25(10), pp.601–610.

Li, X.-T. et al., 2016. tCRISPRi: tunable and reversible, one-step control of gene expression. *Scientific reports*, 6, p.39076.

Lienert, F. et al., 2014. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nature reviews. Molecular cell biology*, 15(2), pp.95–107.

Lim, B. et al., 2017. Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell*, 169(3), pp.547–558.e15.

Lim, F. & Peabody, D.S., 2002. RNA recognition site of PP7 coat protein. *Nucleic acids research*, 30(19), pp.4138–4144.

Lim, G.F.S. et al., 2012. Stability of single-nucleotide bulge loops embedded in a GAAA RNA hairpin stem. *RNA (New York, N.Y.)*, 18(4), pp.807–814.

Lim, W.A., 2010. Designing customized cell signalling circuits. *Nature reviews. Molecular cell biology*, 11(6), pp.393–403.

Lima, W.F. et al., 2007. Human RNase H1 discriminates between subtle variations in the structure of the heteroduplex substrate. *Molecular pharmacology*, 71(1), pp.83–91.

Liu, K.I. et al., 2016. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nature chemical biology*, 12(11), pp.980–987.

Liu, P.Q. et al., 2001. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *The Journal of biological chemistry*, 276(14), pp.11323–11334.

Liu, Q. et al., 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 94(11), pp.5525–5530.

Liu, Y. et al., 2016. Directing cellular information flow via CRISPR signal conductors. *Nature methods*, 66(4), pp.1173–1179.

Liu, Y. et al., 2014. Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nature communications*, 5, p.5393.

Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)*, 25(4), pp.402–408.

Lo, A. & Qi, L., 2017. Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Research*, 6, p.747.

Ma, Hanhui et al., 2015. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10), pp.3002–3007.

Ma, Hanhui et al., 2016. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature biotechnology*, 34(5), pp.528–530.

Ma, Hongming et al., 2014. Pol III Promoters to Express Small RNAs: Delineation of

Transcription Initiation. *Molecular therapy. Nucleic acids*, 3(5), p.e161.

Maeder, M.L. et al., 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature methods*, 10(10), pp.977–979.

Maeder, M.L. et al., 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Molecular cell*, 31(2), pp.294–301.

Mahfouz, M.M. et al., 2011. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), pp.2623–2628.

Maji, B. et al., 2016. Multidimensional chemical control of CRISPR-Cas9. *Nature chemical biology*, 13(1), p.9-11.

Makarova, K.S. et al., 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, 13(11), pp.722–736.

Mali, P., Aach, J., et al., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9), pp.833–838.

Mali, P., Esvelt, K.M. & Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nature methods*, 10(10), pp.957–963.

Mali, P., Yang, L., et al., 2013. RNA-guided human genome engineering via Cas9. *Science (New York, N.Y.)*, 339(6121), pp.823–826.

Mandegar, M.A. et al., 2016. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell stem cell*, 18(4), pp.541–553.

Marraffini, L.A. & Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (New York, N.Y.)*, 322(5909), pp.1843–1845.

Martick, M. & Scott, W.G., 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126(2), pp.309–320.

Mashal, R.D., Koontz, J. & Sklar, J., 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature genetics*, 9(2), pp.177–183.

Mazzarello, P., 1999. A unifying concept: the history of cell theory. *Nature cell biology*, 1(1), pp.E13–5.

McKenna, A. et al., 2016. Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science (New York, N.Y.)*, p.aaf7907.

Meister, G. et al., 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell*, 15(2), pp.185–197.

Metz, S., Jäger, A. & Klug, G., 2012. Role of a short light, oxygen, voltage (LOV) domain protein in blue light- and singlet oxygen-dependent gene regulation in Rhodobacter sphaeroides. *Microbiology (Reading, England)*, 158(Pt 2), pp.368–379.

Mi, J. et al., 2006. H1 RNA polymerase III promoter-driven expression of an RNA aptamer leads to high-level inhibition of intracellular protein activity. *Nucleic acids research*, 34(12),

pp.3577–3584.

Miller, J.C. et al., 2011. A TALE nuclease architecture for efficient genome editing. *Nature biotechnology*, 29(2), pp.143–148.

Mohanraju, P. et al., 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science (New York, N.Y.)*, 353(6299), pp.aad5147–aad5147.

Mojica, F.J. et al., 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, 36(1), pp.244–246.

Mojica, F.J.M. et al., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2), pp.174–182.

Morbitzer, R. et al., 2010. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21617–21622.

Moscou, M.J. & Bogdanove, A.J., 2009. A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)*, 326(5959), pp.1501–1501.

Murphy, S., Tripodi, M. & Melli, M., 1986. A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes. *Nucleic acids research*, 14(23), pp.9243–9260.

Mussolino, C. et al., 2014. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research*, 42(10), pp.6762–6773.

Nagai, T. et al., 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology*, 20(1), pp.87–90.

Nguyen, D.P. et al., 2016. Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nature communications*, 7, p.12009.

Nielsen, A.A. & Voigt, C.A., 2014. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular systems biology*, 10(11), pp.763–763.

Niewoehner, O., Jinek, M. & Doudna, J.A., 2014. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic acids research*, 42(2), pp.1341–1353.

Nihongaki, Y., Kawano, F., et al., 2015. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature biotechnology*, 33(7), pp.755-760.

Nihongaki, Y., Yamamoto, S., et al., 2015. CRISPR-Cas9-based Photoactivatable Transcription System. *Chemistry & biology*, 22(2), pp.169-174.

Nishida, K. et al., 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science (New York, N.Y.)*, p.aaf8729.

Nishimasu, H. et al., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), pp.935–949.

Nishimura, K. et al., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature methods*, 6(12), pp.917–922.

Nissim, L. et al., 2014. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular cell*, 54(4), pp.698–710.

Nomura, Y. et al., 2013. Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS synthetic biology*, 2(12), pp.684–689.

Nomura, Y., Kumar, D. & Yokobayashi, Y., 2012. Synthetic mammalian riboswitches based on guanine aptazyme. *Chemical communications (Cambridge, England)*, 48(57), pp.7215–7217.

Nuñez, J.K., Harrington, L.B. & Doudna, J.A., 2016. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS chemical biology*, 11(3), pp.acschembio.5b01019–688.

O'Reilly, D., Hanscombe, O. & O'Hare, P., 1997. A single serine residue at position 375 of VP16 is critical for complex assembly with Oct-1 and HCF and is a target of phosphorylation by casein kinase II. *The EMBO journal*, 16(9), pp.2420–2430.

Oakes, B.L. et al., 2016. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nature biotechnology*, 34(6), pp.646–651.

Orr, R.M., 2001. Technology evaluation: fomivirsen, Isis Pharmaceuticals Inc/CIBA vision. *Current opinion in molecular therapeutics*, 3(3), pp.288–294.

Paddon, C.J. et al., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446), pp.528–532.

Park, J.-J. et al., 2010. B7-H1/CD80 interaction is required for the induction and maintenance of peripheral T-cell tolerance. *Blood*, 116(8), pp.1291–1298.

Pattanayak, V. et al., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, 31(9), pp.839–843.

Pawluk, A. et al., 2014. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. *mBio*, 5(2), pp.e00896–e00896–14.

Perez-Pinera, P. et al., 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10(10), pp.973–976.

Perli, S.D., Cui, C.H. & Lu, T.K., 2016. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science (New York, N.Y.)*, p.aag0511.

Polstein, L.R. et al., 2015. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature chemical biology*, 11(3), pp.198–200.

Pourcel, C., Salvignol, G. & Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)*, 151(Pt 3), pp.653–663.

Prody, G.A. et al., 1986. Autolytic processing of dimeric plant virus satellite RNA. *Science (New York, N.Y.)*, 231(4745), pp.1577–1580.

Qi, L. et al., 2012. RNA processing enables predictable programming of gene expression. *Nature biotechnology*, 30(10), pp.1002–1006.

Qi, L.S. et al., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), pp.1173–1183.

Qiu, P. et al., 2004. Mutation detection using Surveyor nuclease. *BioTechniques*, 36(4), pp.702–707.

Ramos, J.L. et al., 2005. The TetR family of transcriptional repressors. *Microbiology and molecular biology reviews : MMBR*, 69(2), pp.326–356.

Ran, F.A. et al., 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, 520(7546), pp.186–191.

Ratner, H.K., Sampson, T.R. & Weiss, D.S., 2016. Overview of CRISPR-Cas9 Biology. *Cold Spring Harbor protocols*, 2016(12), p.pdb.top088849.

Rauch, B.J. et al., 2017. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell*, 168(1-2), pp.150–158.e10.

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, 30(5), pp.460–465.

Riccione, K.A. et al., 2012. A synthetic biology approach to understanding cellular information processing. *ACS synthetic biology*, 1(9), pp.389–402.

Richter, F. et al., 2016. Engineering of temperature- and light-switchable Cas9 variants. *Nucleic acids research*, p.gkw930.

Rizzo, M.A. et al., 2004. An improved cyan fluorescent protein variant useful for FRET. *Nature biotechnology*, 22(4), pp.445–449.

Roehr, B., 1998. Fomivirsen approved for CMV retinitis. *Journal of the International Association of Physicians in AIDS Care*, 4(10), pp.14–16.

Sapranauskas, R. et al., 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research*, 39(21), pp.9275–9282.

Saragliadis, A. & Hartig, J.S., 2013. Ribozyme-based transfer RNA switches for post-transcriptional control of amino acid identity in protein synthesis. *Journal of the American Chemical Society*, 135(22), pp.8222–8226.

Sashital, D.G., Jinek, M. & Doudna, J.A., 2011. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature structural & molecular biology*, 18(6), pp.680–687.

Sassa, A. et al., 2016. Impact of Ribonucleotide Backbone on Translesion Synthesis and Repair of 7,8-Dihydro-8-oxoguanine. *The Journal of biological chemistry*, 291(46), pp.24314–24323.

Scherr, M. et al., 2000. RNA accessibility prediction: a theoretical approach is consistent with experimental studies in cell extracts. *Nucleic acids research*, 28(13), pp.2455–2461.

Semenova, E. et al., 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp.10098–10103.

Senturk, S. et al., 2017. Rapid and tunable method to temporally control gene editing based

on conditional Cas9 stabilization. *Nature communications*, 8, p.14370.

Shah, S.A. et al., 2013. Protospacer recognition motifs: mixed identities and functional diversity. *RNA biology*, 10(5), pp.891–899.

Shalem, O. et al., 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, 343(6166), pp.84–87.

Shechner, D.M. et al., 2015. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nature methods*, 12(7), pp.664–670.

Shmakov, S. et al., 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell*, 60(3), pp.385–397.

Singh, D. et al., 2016. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature communications*, 7, p.12778.

Slaymaker, I.M. et al., 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science (New York, N.Y.)*, 351(6268), pp.84–88.

Solé, R.V. et al., 2007. Synthetic protocell biology: from reproduction to computation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1486), pp.1727–1739.

Soukup, G.A. & Breaker, R.R., 1999. Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp.3584–3589.

Srinivasan, S.K. & Iversen, P., 1995. Review of in vivo pharmacokinetics and toxicology of phosphorothioate oligonucleotides. *Journal of clinical laboratory analysis*, 9(2), pp.129–137.

Stein, C.A., Tonkinson, J.L. & Yakubov, L., 1991. Phosphorothioate oligodeoxynucleotides--anti-sense inhibitors of gene expression? *Pharmacology & therapeutics*, 52(3), pp.365–384.

Sternberg, S.H. et al., 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature*, 527(7576), pp.110–113.

Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), pp.62–67.

Sternberg, S.H., Haurwitz, R.E. & Doudna, J.A., 2012. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA (New York, N.Y.)*, 18(4), pp.661–672.

Stewart, T.J. & Smyth, M.J., 2011. Improving cancer immunotherapy by targeting tumor-induced immune suppression. *Cancer metastasis reviews*, 30(1), pp.125–140.

Struhl, G. & Adachi, A., 1998. Nuclear access and action of notch in vivo. *Cell*, 93(4), pp.649–660.

Szczelkun, M.D. et al., 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), pp.9798–9803.

Tanenbaum, M.E. et al., 2014. A Protein-Tagging System for Signal Amplification in Gene

Expression and Fluorescence Imaging. *Cell,* 159(3), pp.635-646.

Tang, J. & Breaker, R.R., 1997. Rational design of allosteric ribozymes. *Chemistry & biology*, 4(6), pp.453–459.

Tang, W., Hu, J.H. & Liu, D.R., 2017. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature communications*, 8, p.15939.

Thakore, P.I. et al., 2016. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nature methods*, 13(2), pp.127–137.

Thompson, D., Regev, A. & Roy, S., 2015. Comparative analysis of gene regulatory networks: from network reconstruction to evolution. *Annual review of cell and developmental biology*, 31(1), pp.399–428.

Topalian, S.L., Drake, C.G. & Pardoll, D.M., 2012. Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Current opinion in immunology*, 24(2), pp.207–212.

Tsai, S.Q. et al., 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature methods,* 14(6), pp.607-614.

Tsai, S.Q. et al., 2014. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology*, 33(2), pp.187-197.

Tsuji, T. & Niida, Y., 2008. Development of a simple and highly sensitive mutation screening system by enzyme mismatch cleavage with optimized conditions for standard laboratories. *Electrophoresis*, 29(7), pp.1473–1483.

Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968), pp.505–510.

Tupler, R., Perini, G. & Green, M.R., 2001. Expressing the human genome. *Nature*, 409(6822), pp.832–833.

Tuschl, T. et al., 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & development*, 13(24), pp.3191–3197.

Urban, E. & Noe, C.R., 2003. Structural modifications of antisense oligonucleotides. *Farmaco (Societa chimica italiana : 1989)*, 58(3), pp.243–258.

Urnov, F.D. et al., 2010. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11(9), pp.636–646.

Valencia-Sanchez, M.A. et al., 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & development*, 20(5), pp.515–524.

Van Damme, J. et al., 1985. Homogeneous interferon-inducing 22K factor is related to endogenous pyrogen and interleukin-1. *Nature*, 314(6008), pp.266–268.

van der Oost, J. et al., 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature reviews. Microbiology*, 12(7), pp.479–492.

Venter, J.C. et al., 2001. The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), pp.1304–1351.

Vickers, T.A. & Crooke, S.T., 2015. The rates of the major steps in the molecular mechanism of RNase H1-dependent antisense oligonucleotide induced degradation of RNA. *Nucleic acids research*, 43(18), pp.8955–8963.

Vickers, T.A. et al., 2014. Targeting of repeated sequences unique to a gene results in significant increases in antisense oligonucleotide potency. S. Maas, ed. *PloS one*, 9(10), p.e110615.

Wah, D.A. et al., 1997. Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, 388(6637), pp.97–100.

Walder, R.Y. & Walder, J.A., 1988. Role of RNase H in hybrid-arrested translation by antisense oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 85(14), pp.5011–5015.

Wang, T. et al., 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science (New York, N.Y.)*, 343(6166), pp.80–84.

Wang, X. et al., 2015. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nature biotechnology*, 33(2), pp.175–178.

Webster, N. et al., 1988. The yeast UASG is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 trans-activator. *Cell*, 52(2), pp.169–178.

Wedler, H. & Wambutt, R., 1995. A temperature-sensitive lambda cI repressor functions on a modified operator in yeast cells by masking the TATA element. *Molecular & general genetics : MGG*, 248(4), pp.499–505.

White, R.J., 2011. Transcription by RNA polymerase III: more complex than we thought. *Nature reviews. Genetics*, 12(7), pp.459–463.

Whitfield, T.W. et al., 2012. Functional analysis of transcription factor binding sites in human promoters. *Genome biology*, 13(9), p.R50.

Wiedenheft, B. et al., 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp.10092–10097.

Wieland, M., Ausländer, D. & Fussenegger, M., 2012. Engineering of ribozyme-based riboswitches for mammalian cells. *Methods (San Diego, Calif.)*, 56(3), pp.351–357.

Win, M.N. & Smolke, C.D., 2007. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences of the United States of America*, 104(36), pp.14283–14288.

Win, M.N. & Smolke, C.D., 2008. Higher-order cellular information processing with synthetic RNA devices. *Science (New York, N.Y.)*, 322(5900), pp.456–460.

Wolfe, S.A., Nekludova, L. & Pabo, C.O., 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure*, 29(1), pp.183–212.

Wright, A.V. et al., 2015. Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10), pp.201501698–2989.

Wroblewska, L. et al., 2015. Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nature biotechnology*, 33(8), pp.839–841.

Wu, Gang et al., 2016. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in biotechnology*, 34(8), pp.652–664.

Wu, Hongjiang et al., 2004. Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. *The Journal of biological chemistry*, 279(17), pp.17181–17189.

Wu, Hongyi et al., 2014. Quantitatively Relating Gene Expression to Light Intensity via the Serial Connection of Blue Light Sensor and CRISPRi. *ACS synthetic biology*, 3(12), pp.979–982.

Wu, Xuebing et al., 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology*, 32(7), pp.670-676.

Wysocka, J. & Herr, W., 2003. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in biochemical sciences*, 28(6), pp.294–304.

Xie, Z. et al., 2011. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science (New York, N.Y.)*, 333(6047), pp.1307–1311.

Ye, H. & Fussenegger, M., 2014. Synthetic therapeutic gene circuits in mammalian cells. *FEBS letters*, 588(15), pp.2537–2544.

Yekta, S., Shih, I.-H. & Bartel, D.P., 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science (New York, N.Y.)*, 304(5670), pp.594–596.

Yen, L. et al., 2004. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, 431(7007), pp.471–476.

Zadeh, J.N. et al., 2011. NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry*, 32(1), pp.170–173.

Zalatan, J.G. et al., 2014. Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell*, 160(1-2), pp.339-350.

Zetsche, B., Gootenberg, J.S., et al., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3), pp.759–771.

Zetsche, B., Volz, S.E. & Zhang, F., 2015. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature biotechnology*, 33(2), pp.139–142.

Zhang, F. et al., 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology*, 29(2), pp.149–153.

Zhang, L. et al., 2000. Synthetic zinc finger transcription factor action at an endogenous chromosomal site. Activation of the human erythropoietin gene. *The Journal of biological chemistry*, 275(43), pp.33850–33860.

Zhou, W. & Deiters, A., 2016. Conditional Control of CRISPR/Cas9 Function. *Angewandte Chemie (International ed. in English)*, 55(18), pp.5394-5399.

Zimmermann, G.R. et al., 2000. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA (New York, N.Y.)*, 6(5), pp.659–667.

Zoltowski, B.D. et al., 2007. Conformational switching in the fungal light sensor Vivid. *Science (New York, N.Y.)*, 316(5827), pp.1054–1057.

Zon, G., 1995. Antisense phosphorothioate oligodeoxynucleotides: introductory concepts and possible molecular mechanisms of toxicity. *Toxicology letters*, 82-83, pp.419–424.

Zuris, J.A. et al., 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature biotechnology*, 33(1), pp.73–80.

Chapter 10 – Materials and Methods

10.1 – System components, cloning and synthesis

10.1.1 – General molecular cloning

All DNA oligonucleotides used as polymerase chain reaction (PCR) primers to cloned the different plasmid DNA (pDNA) constructs presented in this work, were ordered from Integrated DNA technology (IDT). PCR amplifications were performed on the Bio-Rad C1000 Thermal Cycler using to the following protocol and cycling conditions: Unless otherwise specified, amplification reactions contained 1μl of forward primer (10μM stock), 1μl of reverse primer (10μM stock), 10μl of Phusion High-Fidelity DNA Polymerase bought from New England Biolabs (NEB), 0.2μl of template DNA (10ng/μl), and 7.8μl of nuclease free H_2O. Default cycling conditions were as follow: heat up to 98°C for 3min; perform 34 cycles of 98°C for 30s, primer melting temperature for 15s, 72°C for 30s; elongate at 72°C for 5min and cool down to 4°C. DNA backbones and inserts were both digested according to manufacturer guidelines with restriction enzymes bought from NEB. Only exception, BbsI was ordered from Thermo Fisher Scientific. Sequential digestions were carried out in the case of incompatibility between enzyme buffers: DNA digested with the first enzyme was passed through a PCR clean up column (Qiagen) before being mixed with the second enzyme. Post-digestion, DNA backbones and inserts were run on a 1% agarose gel to get rid of the enzyme and isolate the desired products. Specific bands were cut and the corresponding DNA was recovered using columns from a Gel extraction kit (Qiagen). Backbone DNA and insert(s) were then ligated together using T4-DNA ligase (NEB) for 1h at room temperature. Frozen DH5α-competent bacteria (in-

house made) were first thawed on ice for 30 minutes prior to being transformed, via heatshock (30 seconds, 42°C), with a fraction of the ligation mix. The cells were then incubated in 1ml of SOC medium (in-house made) for one hour at 37°C before being seeded on an agar plate (in-house made), prepared with the antibiotic matching the resistance gene on the DNA backbone. The plates were incubated for 16 hours at 37°C, after which single bacterial colonies were scraped out of the plate and grown overnight (16 hours) in LB medium (in-house made). The plasmid DNA was then recovered from each expanded bacterial population using pDNA purification columns (Miniprep kit, Qiagen). The obtain plasmids were used as such for subsequent transfections (see section 10.2).

10.1.2 – Native sgRNA and iSBH-sgRNA cloning

Cloning of the sgRNA backbone: Three different pDNA backbones were used throughout this work to express either native or iSBH-sgRNAs: A custom made *U6p-sgRNA-6xT_iBlue* vector, which cloning is detailed below, was used to express native and iSBH-sgRNAs in all experiments involving activation of a reporter gene (chapters 2, 3, 4, 5, 7). The *SAM_U6p-sgRNA-2xMS2* construct, used to express modified sgRNA-2xMS2 guides from the SAM system (Konermann et al. 2014), was a gift from the Feng Zhang's laboratory (Addgene plasmid 42230). This pDNA was used to express both native and iSBH-SAM-sgRNAs used for the control of endogenous targets (chapters 6). The *px330* plasmid, featuring an sgRNA cassette and expressing the catalytically active Cas9, was given to us by the Feng Zhang's laboratory (Addgene plasmid 42230). This plasmid was used to express both native and iSBH-sgRNAs in chapter 8 to demonstrate Csy4-dependent genome editing of the human *PD-L1* gene.

Cloning of *U6p-sgRNA-6xT_iBlue*: The U6p-sgRNA-6xT cassette from *px330* (Addgene plasmid 42230), featuring a pol-III U6 promoter, the sgRNA sequence, and U6 terminator (6xT), was PCR amplified using forward and reverse primers CP1 and CP2 (table 1). The resulting insert was digested with restriction enzymes SpeI and BcoDI, and cloned into a *pcDNA3.1* DNA backbone (gift from Timothy Rajakumar, Fulga laboratory) previously cut with

SpeI and BbsI. Additionally, the coding sequence of the iBlue fluorescent protein was amplified from the *iBlue-N1* vector (gift from Michael Davidson, Addgene plasmid 54781) using forward and reverse primers CP3 and CP4 (table 1). This DNA fragment was then concatenated by PCR assembly with an amplicon of the SV40 promoter from *pcDNA3.1* (amplified with CP5 and CP6, table 1). Both double stranded DNAs, designed to share an overlapping sequence, were mix at equimolar ratio in a PCR reaction without template. After 8 cycles, 1µl of the mix was used as template in a second PCR reaction containing flanking primer for the concatenated product (CP3, CP6, table 1). The resulting insert was digested with restriction enzymes NcoI and BsrGI before being cloned in *pcDNA3.1*, previously cut with the same enzymes.

Spacer cloning in sgRNA backbone: All ~20nt spacer segments were cloned as annealed DNA oligonucleotides in place of the 18 base-pairs "place holder", found directly upstream the sgRNA scaffold in all three sgRNA expression vectors, as previously described (Cong & F. Zhang 2015). Briefly, sequence complementary forward and reverse oligonucleotides encoding the spacer were first phosphorylated using T4 Polynucleotide Kinase (NEB) before being annealed together by heating up the reaction mix to 95°C, incubating at 95°C for 5 minutes and slowly cooled down to 4°C. The sgRNA expression vector (*U6p-sgRNA-6xT_iBlue*, *SAM_U6p-sgRNA-2xMS2*, or *px330*) was digested with BbsI to excise the spacer "place holder", dephosphorylated using Antarctic Phosphatase (NEB), and gel purified. Backbone and oligo duplex were then ligated together before proceeding with bacterial transformation. Note that the sequence of the oligonucleotides was chosen so that, once annealed, the complex would have overhangs compatible with the digested backbone: if for instance the spacer sequence to be cloned was GAGTCGCGTGTAGCGAAGCA, then the sequence complementary forward and reverse oligonucleotides would have been 5'-**CACC**GAGTCGCGTGTAGCGAAGCA-3' and 5'-**AAAC**TGCTTCGCTACACGCGACTC-3', respectively. Overhangs post BbsI digest were the same for all three sgRNA expression vectors (*U6p-sgRNA-6xT_iBlue*, *SAM_U6p-sgRNA-2xMS2*, *px330*). In the case of short

spacers (15, 10, 5nt) used in chapter 2, which could not be synthetized by IDT, the entire nv(x)CTS-sgRNAs (x <= 15) were PCR amplified from the corresponding *U6p-nv-CTS-sgRNA-6xT_iBlue* (either nv-CTS1 or nv-CTS2, table 2) using a forward primer encoding the spacer (CP7-1.5, 1.10, 1.15, 2.5, 2.10, 2.15, table 1) and a common reverse primer downstream the sgRNA scaffold (CP8). The amplicon was digested using BbsI restriction enzymes and clone into *U6p-sgRNA-6xT_iBlue*, previously cut with BbsI and NheI. A list of all spacer sequences featured in this work is given in table 2.

SBHs/iSBHs cloning in sgRNA backbone: (inducible) Spacer-blocking hairpins were cloned into the sgRNA expression vector (*U6p-sgRNA-6xT_iBlue*, *SAM_U6p-sgRNA-2xMS2*, or *px330*) as a single oligo duplex encoding the entire back-fold/sensing-loop/spacer, using the same protocol employed for normal spacer cloning (see above). Hairpins exceeding 60nt in length were cloned using two oligo duplexes sharing a compatible overhang. Both oligonucleotide pairs were annealed separately before being combined, along with the digested backbone, in the same ligation reaction. A list of all SBHs and iSBHs sequences features in this work is given in table 2. Csy4-, Cse3-, and Cas6A-responsive iSBHs were created by adapting the corresponding protein target site as explain in chapter 3, 5, and 6. ASO sensing-loops used to create ASO-iSBHs presented in chapter 4, 5, and 6 were evolved using the genetic algorithm described in chapter 5. In the case of SBHs and iSBHs featuring a hammerhead ribozyme (or allosteric counterpart), the hairpin (>100nt) was synthesized via assembly PCR before being cloned into the *U6p-sgRNA-6xT-iBlue* vector. Briefly, four overlapping oligonucleotides were chosen using the primer design software Primerize (https://primerize.stanford.edu/) to cover the hairpin sequence plus flanking BbsI restriction sites. A 20μl PCR reaction was assembled with 1μl of each primer (10μM stock) and 10μl Phusion High-Fidelity DNA Polymerase (NEB). The mixture was subjected to 8 cycles of amplification including a 95°C step (2 min, extended to 5 prior first cycle), primer melting temperature step (2 min), and a 72°C step (3 min, extended to 5 after completion of the last cycle). A second PCR reaction was then assembled using 1μl of the previous reaction as

template and flanking forward and reverse primers only amplifying the entire insert. The Amplicon was passed through a PCR clean-up column (Qiagen) before BbsI digest. The ensuing insert was gel purified and cloned in the *U6p-sgRNA-6xT-iBlue* backbone according to the spacer cloning protocol described above. Sequences for the type I *Schistosoma mansoni* HHRz and theophylline-responsive allosteric HHRz were adapted from (Saragliadis & Hartig 2013), and sequences of the corresponding SBHs/iSBHs are given in table 2 Note: all spacer, SBH, and iSBH sequences started on their 5'-end with a guanine required for successful U6 transcription (Goomer & Kunkel 1992) (table 2). Regarding the experiment presented in figure 8.5, the mimic allosteric sgRNA was cloned via PCR assembly. The sequence of the entire guide is given in the appendix (see sequence 2).

10.1.3 – CRISPR-effectors and effector domain tethers

The *dCas9m4-VP64* vector (Mali, Aach, et al. 2013), given to us by George Church (Addgene plasmid 47319), was used for all CRISPR-based activation of reporter gene expression. The plasmid encoded a human codon-optimized version of the *Streptococcus pyogenes* dCas9 (catalytically inactive) flanked by a nuclear localization sequence (NLS) on each terminus, and 4 copies of VP16 (VP64) on the C-terminus (Mali, Aach, et al. 2013). CRISPRa experiments on endogenous target genes (*ASLC1*, *HBG1*, *IL1B*) were conducted using the SAM system (Konermann et al. 2014). Vectors expressing human codon-optimized *dCas9-VP64* (NLS-dCas9-NLS-VP64) and *MCP-p65-HSF1* (MCP-NLS-p65-HSF1) were obtained from the Feng Zhang's laboratory (Addgene plasmid 61422 and 61423). Finally, the *px330* vector (Addgene plasmid 42230) expressing the catalytically active Cas9 (NLS-Cas9) was used to demonstrate Csy4-dependent genome editing of the human *PD-L1* gene (chapter 8).

10.1.4 – Reporter contructs

Reporter constructs *8xCTS1-mCMVp-EYFP-pA* and *8xCTS2-mCMVp-ECFP-pA*, expressing fluorescent proteins EYFP and ECFP under the control of a minimal CMV promoter, were gifts from Timothy Lu (Addgene plasmid 55197 and 55198) (Nissim et al. 2014). The dual reporter

combining the EYFP and the ECFP cassettes was created as follows: The EYFP cassette was PCR amplified out of *8xCTS1-mCMVp-EYFP-pA* using the forward and reverse primers CP9 and CP10 (table 1), and digested using BbsI. This insert was then ligated to *8xCTS2-mCMVp-ECFP-pA* previously digested with XhoI and BbsI restriction enzymes. *8xCTS3-mCMVp-ECFP-pA*, used in chapter 7, was made in-house by replacing the 8xCTS2 segment of *8xCTS2-mCMVp-ECFP-pA* with a DNA sequence encoding eight repeats of the CRISPR-target site GTACATACAGTAGGATCGTA**TGG** (CTS3, bold = PAM). The 8xCTS2 fragment was excised from *8xCTS2-mCMVp-ECFP-pA* using NheI. Repeats including the CTS3 sequence were then cloned two by two in four successive cloning steps: (i) Two oligonucleotide duplexes sharing a compatible overhand for the other duplex and an overhang matching the NheI digest backbone were generated by annealing together primers CP11-CP12, and CP13-CP14, respectively (table 1). Both oligo duplexes were ligated together with the digested *8xCTS2-mCMVp-ECFP-pA* to create *2xCTS3-mCMVp-ECFP-pA*. (ii) The 2xCTS3 insert was designed to feature a BstXI site (type-II restriction enzyme) used to clone into *2xCTS3-mCMVp-ECFP-pA* another two oligo duplexes (CP15-CP16 and CP17-CP18, table 1). Cloning of the insert was such that the original BstXI site was removed and a new one was inserted in the center of the CTS array. (iii-iv) The same operation was repeated twice using CP19-CP20 and CP21-CP22 (iii) followed by CP15-CP16 and CP17-CP18 (iv) (table 1), to produce the final *8xCTS3-mCMVp-ECFP-pA*.

10.1.5 – CRISPR-associated endoribonucleases

A plasmid expressing a human codon-optimized version of the *Pseudomonas aeruginosa* Csy4, *PGK1p-Csy4-pA*, was obtained from Timothy Lu (Addgene plasmid 55196). An NLS sequence from the *px330* vector (Addgene plasmid 42230) was PCR amplified with primer CP23 and CP24 (table 1), digested with HindIII and NheI, and cloned into *PGK1p-Csy4-pA* (HindIII, NheI digest) to create *PGK1p-NLS-Csy4-pA*. The resulting plasmid was used to expressed NLS-Csy4 in all experiments featuring the endoribonuclease. Protein sequences of *Thermus thermophillius* Cse3 (accession no. TTHB192) and Cas6A (accession no.

TTHA0078) were obtained from the National Center for Biotechnology Information (NCBI) server. Corresponding human codon-optimized DNA sequences were derived using the dedicated IDT web interface (https://www.idtdna.com/CodonOpt), synthesized as gBlock fragment (IDT), and cloned in place of Csy4 in *PGK1p-NLS-Csy4-pA* using NheI and NotI restriction enzymes. Cse3 and Cas6A sequences are given as appendix. Resulting vectors, *PGK1p-NLS-Cse3-pA* and *PGK1p-NLS-Cas6A-pA*, were used in chapter 3, 5, 6, 8 to express the corresponding endoribonucleases. EGFP constructs designed to test the binding and nuclease property of Csy4 and Cse3 were designed by cloning into the 5'-UTR of a CMV driven EGFP the Csy4- or Cse3-protein target site (PTS). Csy4- and Cse3-PTSs were cloned as oligonucleotide duplexes (CP23-CP24, CP25-CP26, table 1) into *pcDNA3.1-CMV-EGFP-pA* (gift from Timothy Rajakumar, Fulga laboratory, see appendix) previously digested with the NheI and BamHI restriction enzymes.

10.1.6 – Single-stranded DNA antisense oligonucleotides

All antisense oligonucleotide (ASO) sequences were designed by sequence complementarity to the ASO sensing-loop chosen to construct the different ASO-responsive iSBH-sgRNAs. Name, sequences, and iSBH-sgRNA targets of the different ASOs used in this study are listed in table 3. To increase the stability of the oligonucleotides in the cellular environment, ASOs were synthesized (IDT) with three or four terminal phosphorothioate bonds (indicated by an asterisk in table 3). ASOs were resuspended in nuclease-free water to a concentration of 100μM before being used for transfection (see section 10.2). The 100 μM ASO stocks were stored at -20°C in between experiments.

10.2 – Cell culture and transfections

All experiments were conducted in a human embryonic kidney (HEK293-T) cell line expressing a mutant version of the SV40 large T antigen, kindly donated to us by Professor Ahmed Ashour Ahmed (clone number ATCC-CRL-11268). Cells were cultured at 37 °C and 5% CO2 in Dulbecco's modified Eagle's medium (Thermo-Fisher Scientific) supplemented with 15% fetal bovine serum (Thermo-Fisher Scientific) and 1% penicillin–streptomycin (Thermo-Fisher Scientific) (full media). The 75 cm^2 Flasks (Cellstar) were passaged every 48h in a 1:6 ratio to a maximum of 30 times before being replaced with a new low passage batch, which had been previously frozen from the original stock. Before conducting any experiments, the transfection efficiency of each new batch was tested via flow cytometry one week after thawing (~ fourth passage) using a reference plasmid expressing EGFP (*pcDNA3.1-CMVp-EGFP-pA*, gift from Timothy Rajakumar, Fulga laboratory). Prior transfection, cells were seeded in 12-well plates (Cellstar) so as to reach 70% confluency 24h later. Full media was replaced with transfection media (Dulbecco's modified Eagle's medium, 2% fetal bovine serum, no antibiotics) and cells were left to rest for 15 minutes in the 37°C incubator before receiving the transfection mix (DNA plus transfection reagent). For each well (1ml well, 12-well plate, Cellstar), DNA plasmids (see details below) were mixed with polyethylenimine (PEI; Sigma-Aldrich, 1 $mg.ml^{-1}$) at a recommended 2:3 ratio (mg DNA/ml PEI) (Aricescu et al. 2006) in 100μl Opti-MEM (Thermo-Fisher Scientific). The reaction was vortexed for 10 seconds and left to rest for 15 minutes at room temperature before being delivered to the cells. Transfection media was replaced 24h post-transfection for full media, except in the case of delayed ASO delivery (see below).

Transfection conditions used for each type of experiment:

- CRISPRa of reporter gene targets with native sgRNAs or SBH-sgRNAs (no inducer): For each well (1ml well, 12-well plate, Cellstar), 500ng of sgRNA expressing plasmid (*U6p-sgRNA-6xT_iBlue)*, 250ng of *dCas9m4-VP64*, and 250ng of reporter construct were transfected. The reporter plasmid, either *8xCTS1-EYFP-pA* or *8xCTS2-ECFP-*

pA, was chosen to match the spacer segment on the sgRNA. Media was changed from transfection to full media 24h post-transfection and cells were harvested 24h after that for flow cytometry analysis.

- Conditional CRISPRa of reporter gene targets with protein-responsive iSBH-sgRNAs: Same as above with an additional 250ng of protein inducer expression plasmid (*PGK1p-NLS-Csy4-pA*, *PGK1p-NLS-Cse3-pA*, *PGK1p-NLS-Cas6A-pA*). 250ng of decoy plasmid *pcDNA3.1-Empty* (gift from Timothy Rajakumar, Fulga laboratory) was transfected in place of the inducer in all conditions where iSBH-sgRNAs are being tested "in the absence of the inducer". For experiments where simultaneous activations of both Csy4- and Cas6A-repsonsive iSBH-sgRNAs was performed, 125ng of each protein inducer was delivered. Media was changed from transfection to full media 24h post-transfection and cells were harvested 24h after that for flow cytometry analysis.

- Conditional CRISPRa of reporter gene targets with ASO-responsive iSBH-sgRNAs: 500ng of iSBH-sgRNA (*U6p-sgRNA-6xT_iBlue*), 250ng of *dCas9m4-VP64*, and 250ng of reporter construct, *8xCTS1-EYFP-pA* or *8xCTS2-ECFP-pA,* were transfect in each well (1ml well, 12 well plate, Cellstar). Transfection media was renewed 24h post-transfection and a new 100μl Opti-MEM transfection mix containing PEI, ASOs to a final concentration of 100nM (1ml well), and 500ng of carrier DNA (*pcDNA3.1-Empty*) was delivered to the well. The same ASO concentrations and amount of carrier were used when delivering two ASOs simultaneously. Cells were harvested 24h post ASO delivery for flow cytometry analysis.

- CRISPRa of endogenous gene targets with native, SBH-, and iSBH-sgRNA-2xMS2 (SAM system): For each 1ml well, a transfection mix was assembled in 100μl Opti-MEM containing all three SAM system vectors: 500ng *SAM_U6p-sgRNA-2xMS2* (native, SBH, or iSBH), 500ng *dCas9-VP64*, and 500ng *MCP-p65-HSF1*. Additionally, 250ng of *PGK1p-NLS-Csy4-pA*, *PGK1p-NLS-Cas6A-pA*, or *pcDNA3.1-Empty* was

delivered as explained above when assessing the ON/OFF properties of protein-responsive guides. For experiments involving ASO-responsive iSBH-sgRNAs, the oligonucleotide triggers were delivered as aforementioned. Cells were harvested 48h post-transfection and prepared for RT-qPCR analysis (see below).

- Conditional CRISPRa of reporter gene targets with small molecule-responsive iSBH-sgRNAs: 500ng of iSBH-sgRNA (fitted with an allosteric self-cleaving ribozyme, *U6p-sgRNA-6xT_iBlue*), 250ng of *dCas9m4-VP64*, and 250ng of reporter construct were transfected per well. 24h post-transfection, transfection media was replaced for either full media (OFF-state), or full media supplemented with theophylline (Sigma-Aldrich) to a concentration of 1mM (ON-state). Cells were harvested 48h post-transfection for flow cytometry analysis.

- Conditional CRISPR-based genome editing with Csy4-responsive iSBH-sgRNAs: For each well (1ml well, 12-well plate, Cellstar), 1000ng of *px330* plasmid expressing both sgRNA and catalytically active Cas9, and 250ng of either decoy *pcDNA3.1-Empty* or *PGK1p-NLS-Csy4-pA*, were delivered to the cells. Cells were then harvested 48h post-transfection and screened for editing events using a T7E1 assay (see below).

10.3 – Flow cytometry analysis

10.3.1 – Cell preparation and data acquisition

48h post-transfection, full (or transfection) media was removed from each well, cells were washed with 500μl of 1x phosphate-buffered saline (PBS, Thermo-Fisher Scientific), and detached using 200μl of 0.05% trypsin–EDTA (Thermo Fisher Scientific). The trypsin was then inactivated by adding 500μl of full media to the well, and the cells were collected by centrifugation (3 minutes at 500G). Cell palates were then resuspended in 1x PBS and kept on ice for a maximum of 60 minutes before being analyzed. Flow cytometry analyses were conducted on a BD LSR Fortessa Analyzer (BD Biosciences), with the exception of data presented in figures 2.11 and 7.4 which were acquired on a Cyan ADP (DakoCytomation). A

total of 1e5 total events were recorded for each condition, laser intensities and wavelength were identical between experiments and the following channels (excitation-emission) were used 488-530/30 (EYFP, EGFP), 405-450/50 (ECFP), and 640-670/14 (iBlue).

10.3.2 – Data analysis

All analyses were performed using FlowJo (v10.2). Forward and side scatter parameters were used to isolate healthy singleton cells from all events. The gate for the iBlue positive population ($iBlue^{+ve}$) was drawn for each experiment based on a non-transfected control condition were only PEI was delivered to the cells. Delineation between cells positive or negative for reporter expression (EYFP or ECFP = EXFP), was done on a negative control condition where the spacer sequence of the sgRNA was not matching the CRISPR-target site on the reporter construct (e.g. nv-CTS1 sgRNA with *8xCTS2-ECFP-pA* reporter gene). The gate was positioned to allow ~0.1% of false negative EXFP positive cells. Subsequently, the % of activated cells in each condition was measured as the fraction of cells expressing EXFP ($EXFP^{+ve}$) amongst the transfected $iBlue^{+ve}$ population. The activation score was computed by multiplying the % of activated cells by the median reporter fluorescence of the double positive population ($iBlue^{+ve}/EXFP^{+ve}$) as previously published (Nissim et al. 2014; Ausländer et al. 2012). This measure provided a weighted fluorescence value, which both integrates the penetrance of the inducer-mediated activation (% activated cells) and the strength of CRISPRa for the activated cell population. The $EXFP^{-ve}/EXFP^{+ve}$ gate was stringently positioned (~0.1% false positive) to detect minimal OFF-state leakage. For instance, if the number of activated cells in the OFF-state was as low as 1%, this would translate into a 10-fold fold change in activation score. A detail explanation of the activation score calculation is given in figure 2.3.

10.4 – Quantitative reverse transcription PCR (RT-qPCR)

Cells were harvested 48h post-transfection for total RNA extractions. Full (or transfection) media was removed, cells were washed twice with 500ul of 1x PBS, and total RNA of each well was extracted using the RNeasy Mini Kit (Qiagen), according to manufacturer protocol. Each RNA sample was then tested for purity using a NanoDrop 2000 (Thermo Fisher Scientific). For each condition 1μg of RNA was then reverse transcribed with random hexamers (non-specific) primers using the QuantiTect Reverse Transcription Kit (Qiagen). The volume of RT reactions was halved from the 20μl recommended in the manufacturer protocol. In each experiment, an additional non-RT reaction (no reverse transcriptase) was set up using RNA from a positive control to ensure that the signal was coming from reverse transcribed complementary DNA (cDNA) and not residual genomic or plasmid DNA. For each sample, *n* qPCR mixes, where *n* stands for the number of primer pairs, were prepared by adding to 12.25μl of nuclease-free water, 17.5μl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), 1.75μl of forward and 1.75μl of reverse primer (10μM, IDT), and 3.5μl of the reverse transcription reaction (cDNA). The content of each reaction was then split in three 10ul technical replicates. qPCR analyses were carried out on a CFX384 real-time system (Bio-Rad) using the following cycling protocol: heat up to 98 °C for 2min; perform 39 cycles of 98 °C 10s, 60 °C 30s, image; 65 °C for 5s and then perform a melt curve analysis. All primer pairs used in this work (see table 4 for sequences) were tested for efficiency using serial dilution on a positive sample (standard curves) prior experiments.

RT-qPCR results were analyzed using the ΔΔCt method (Livak & Schmittgen 2001). For each condition, transcript levels of dCas9-VP64 and target gene (*ASCL1*, *HBG1*, *IL1B*) mRNAs were normalized to the house keeping gene *GAPDH* to account for variations in total RNA between samples (mRNA levels for gene X = $2^{(Ct_{GAPDH} - Ct_X)}$). The estimated transcript levels of each gene of interest were further normalized to dCas9-VP64 mRNAs in order to account for differences in transfection efficiency between conditions. Finally, using the formula below, a

fold change increase in gene x transcript levels was computed for each sample (S) in reference to the negative control condition (Ref) where cells were transfected with a scramble sgRNA:

$$Fold\ change = \frac{2^{\left(Ct^{S}_{GAPDH} - Ct^{S}_{X}\right) - \left(Ct^{S}_{GAPDH} - Ct^{S}_{dCas9-VP64}\right)}}{2^{\left(Ct^{Ref}_{GAPDH} - Ct^{Ref}_{X}\right) - \left(Ct^{Ref}_{GAPDH} - Ct^{Ref}_{dCas9-VP64}\right)}}$$

For experiments presented in chapter 3 and 5, where slicing of the 5'-UTR of EGFP is assayed, the primer pair 5pCTS-EGFP was used (table 4). Calculation were performed as aforementioned minus the dCas9-VP64 normalization step.

10.5 – Detection of genome editing events, T7E1 assay

Cells were harvested 48h post transfection, washed twice with 500µl 1x PBS (per well), detached using 200µl of 0.05% trypsin–EDTA (inactivated with 500µl full media), and spun down for 3 minutes at 500G. After removing the supernatant, each cell palate was resuspended in 500µl NaOH (50mM), incubated for 10 minutes at 95°, before neutralizing the reaction with Tris-CL (50µl). Each sample was then centrifuged for 3 minutes at 13000G and 100µl of the supernatant was collected for later usage (PCR amplification). Using a *PD-L1* specific primer pair designed to asymmetrically flank the CRISPR cut site (same primer as (Kataoka et al. 2016)), genomic DNA was PCR amplified for 39 cycles in a 30µl reaction: 10µl nuclease-free water, 2µl genomic DNA, 1.5µl of each primer (10µM), 15µl Phusion High-Fidelity DNA Polymerase (NEB). PCR products were then purify using a PCR clean-up column (Qiagen) to a final volume of 10µl.

The amplicons were then screened for CRISPR-mediated editing events using the T7E1 assay (Tsuji & Niida 2008; Mashal et al. 1995). For each sample a reaction containing 2µl of genomic DNA (from 10µl elution), 2µl of T7 Endonuclease I buffer (T7E1, NEB), and 15µl nuclease free water, was assembled. The mix was incubated at 95°C and slowly cooled down to 4°C in order to form double stranded DNA hybrids. Subsequently, 1µl of T7 Endonuclease I (NEB) was

added to the reaction, followed by an incubation period of 30 minutes at 37°C. Finally, all samples were run on a 1% agarose gel to check for T7E1 products.

10.10 – Software

Several computer programs have been developed during the course of the project to assist SBH and iSBH design. Those were written mainly in Java and R with the exception of the iSBHfold web tool which also involved HTML, CSS, AJAX, and JAVASCRIPT programming. All RNA secondary structure predictions embedded in the algorithms are powered by Nupack 3.0.4 (http://www.nupack.org/downloads). Binaries for this release were obtained from the Pierce laboratory (Caltech). No additional package dependencies were used in this work. The majority of these programs (see below) are now embedded in the iSBHfold web tool (see section 6.2) and can be freely accessed online at http://apps.molbiol.ox.ac.uk/iSBHfold/cgi-bin/iSBHfold.cgi. Additionally, source codes will be made available upon request.

10.10.1 – SBH design

Full ~20nt back-fold sequences can be easily obtained by computing the reverse complement of the sgRNA spacer segment (SBH$^{(0)}$CTS, 0 free spacer nucleotides). From there, back-fold and spacer sequences need to be linked with a *GAAA* loop to form a SBH. The final sequence is flanked on both side with overhang fragments compatible with cloning into a U6-sgRNA cassette (see section 10.1.2). Sequences of the staggered forward and reverse oligonucleotides encoding the hairpin are then derived by taking 5' overhang + SBH, and reverse complement of SBH +3' overhang, respectively. While this operation can easily be done manually, repeating it across various spacer sequences can become tedious and error prone. To facilitate SBH design, I have developed a simple program which takes as input a spacer (e.g. CTS1) and outputs the sequence of the corresponding SBH$^{(0)}$CTS1, along with the oligonucleotides required to clone the structure into a sgRNA expression vector. Additionally, the system generates the same output for all intermediate stems SBH$^{(1)}$CTS1, …, SBH$^{(20)}$CTS1 (1 to 20 free spacer nucleotides), along with their respective predicted RNA

secondary structure and stability (Nupack). This program was notably used to create all stems used in chapter 2 to demonstrate SBH-based CRISPRa modulation. Finally, the same program was designed to let its user produce the sequence of bulged hairpins. Starting from a given fully base-paired SBH$^{(0)}$CTS hairpin, the user can slide along the stem a given bulge motif (e.g. 2 nucleotide mismatch flanked by 4 and 3 consecutive base-pairs on the 5'- and 3'-end, respectively). For each resulting bulged hairpin, the system outputs the sequence color-coded with the stability of the entire structure as well as details values for the seed proximal (important for OFF-state silencing) and loop proximal (important for iSBH melting post-cleavage) regions of the stem.

10.10.2 – iSBH sequence evolution (brute force)

iSBH sequences were designed so that the hairpin would fold into a user defined RNA secondary structure. This was achieved by using the RNA secondary structure prediction algorithm Nupack (Zadeh et al. 2011) and revising the iSBH sequence until the predicted structure matched the desired outcome. The process of evolving particular iSBH segments was automated using a code written in JAVA, which iterates through all possible sequences, folds the resulting iSBH, and scores it against the target structure. This type of optimization, referred to as "brute force" search, was only used to evolve the sequence of small segment (<5nt) for which the number of possible sequences was tractable. The program was notably employed to evolve the back-fold sequence to form bulges along the stem and guarantee ssRNA:dsRNA junction at the base of the Csy4-, Cse3-, and Cas6A-PTSs.

10.10.3 – ASO sensing-loop evolution (genetic algorithm)

Full details regarding the functioning of the genetic algorithm used to evolve the ASL are given in chapter 5 and figure 5.4. The program, written in Java, is included in the iSBHfold web tool and is automatically called for the design of ASO-responsive iSBH-sgRNAs. Briefly, the code takes as input a list of spacer sequence (SP_1, …, SP_n), and evolves 20nt ASO sensing-loops (ASLs) that would retained an open ssRNA conformation once fitted onto the bulged

$SBH^{(0B)}SP_1$, …, $SBH^{(0B)}SP_n$, to form the corresponding ASO-responsive $iSBH^{(0B)}ASO\text{-}SP_1$, …, $iSBH^{(0B)}ASO\text{-}SP_n$ sgRNAs. The list of ASLs is evolved from a starting population of 20nt random DNA sequences, which gets updated at each iteration by adding new loop composites created from ASLs that best satisfy the structural constrains, and discarding ASLs that score poorly. Scoring was performed using the Nupack algorithm (Zadeh et al. 2011): For each ASL_j, all *n* iSBHs constructed by putting the sensing loop onto $SBH^{(0B)}SP_i$ (i ranges from 1 to *n*) are attributed a score between 0 and 1 quantifying the probability for them to fold into a user defined structure (optimal ASO-iSBH design). Scoring is performed through the *prob* function of the Nupack package (Zadeh et al. 2011). A global score is then attributed to each ASL_j by multiplying all *n* individual scores together. Additionally, point mutations and new random sequences are injected in the population at each iteration to maintain a large genetic diversity. The evolution process stops when the score of the top ASL in the population plateaus for few iterations or reach a user defined number of iterations.

10.10.4 – iSBHfold web tool

Details on the function included in the iSBHfold web tool, along with a presentation of the user interface is given in chapter 6 and figure 6.5 and 6.6. The tool can be accessed at http://apps.molbiol.ox.ac.uk/iSBHfold/cgi-bin/iSBHfold.cgi, and videos have been uploaded on each design page, that provide a step by step instruction on how to use the program.

Appendixes

Table 1 – Cloning primers

PRIMER NAME	DNA sequence
CP1	TAGACTAGTGAGGGCCTATTTCCCATG
TCP2	CTAGTCTCTGCTAGCAAAAAAGCACCGACTCGG
CP3	GTATATCCATTTTCGGATCTGATCGCCACCATGTCGGTACC
CP4	CTACTATGTACATCATTTGGACTGAGACTGTGC
CP5	CTAGCCTAACTCCGCCCAGTTCC
CP6	GATCAGATCCGAAAATGGATATAC
CP7-1.5	TAGGAAGACTACACCGAAGCAGTTTTAGAGCTAGAAATAGCAAG
CP7-1.10	TAGGAAGACTACACCGTAGCGAAGCAGTTTTAGAGCTAGAAATAG
CP7-1.15	TAGGAAGACTACACCGCGTGTAGCGAAGCAGTTTTAG
CP7-2.5	TAGGAAGACTACACCGTCCTGTTTTAGAGCTAGAAATAGCAAG
CP7-2.10	TAGGAAGACTACACCGTACTGTCCTGTTTTAGAGCTAGAAATAGC
CP7-2.15	TAGGAAGACTACACCGTCGGAGTACTGTCCTGTTTTAGAG
CP8	CTAGAAGACTGCTAGCAAAAAAGCACCGACTC
CP9	CATTAGGAAGACTATCGAGTAAGTAATATTAAGGTACGGGAGG
CP10	CACCTGTCCTACGAGTTGC
CP11	CTAGTCCATACGATCCTACTGTATGTACTTAATACC
CP12	AATCAGTTGGTATTAAGTACATACAGTAGGATCGTATGGA
CP13	AACTGATTGGCCATACGATCCTACTGTATGTACTTAATATTTTCAGT
CP14	CTAGACTGAAAATATTAAGTACATACAGTAGGATCGTATGGCC
CP15	TTCGCCATACGATCCTACTGTATGTACTTAATACCAATACGTT
CP16	GGTATTAAGTACATACAGTAGGATCGTATGGCGAATCAG
CP17	GGCCATACGATCCTACTGTATGTACTTAATACTAACTGA
CP18	TTAGTATTAAGTACATACAGTAGGATCGTATGGCCAACGTATT
CP19	TTCGCCATACGATCCTACTGTATGTACTTAATACCAACTGATT
CP20	GGTATTAAGTACATACAGTAGGATCGTATGGCGAACGTA
CP21	GGCCATACGATCCTACTGTATGTACTTAATACTAATACG
CP22	TTAGTATTAAGTACATACAGTAGGATCGTATGGCCAATCAGTT
CP23	CTAGCGTTCACTGCCGTATAGGCAGCTAAGAAAG
CP24	GATCCTTTCTTAGCTGCCTATACGGCAGTGAACG
CP25	CTAGCGTAGTCCCCACGCGTGTGGGGATGGACCGACTAGTG
CP26	GATCCACTAGTCGGTCCATCCCCACACGCGTGGGGACTACG

Table 2 – SBH and iSBH sequences

NAME	Insert (5' overhang CACC, 3' overhang GTTT)
nv-SCR	GGGTCTTCGAGAAGACCT
nv-CTS1	GAGTCGCGTGTAGCGAAGCA
nv-CTS2	GTAAGTCGGAGTACTGTCCT
SBH(0)CTS1	GCTGCTTCGCTACACGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(0)CTS2	GCAGGACAGTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(ctrl-1)CTS1	GAAAGCAGCCGAAAGGCTGCTTTCAGTCGCGTGTAGCGAAGCA
SBH(ctrl-1)CTS2	GACAAGCAGCCGAAAGGCTGCTTGTCGTAAGTCGGAGTACTGTCCT
SBH(ctrl-2)CTS1	GCAGGACAGTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCTGAGTCGCGTGTAGC GAAGCA
SBH(ctrl-2)CTS2	GCTGCTTCGCTACACGCGACTGAAAAGTCGCGTGTAGCGAAGCAGGTAAGTCGGAGTACT GTCCT
SBH(ctrl-3)CTS1	GTAGCGATCAGTAGTCAGAATGAAAAGTCGCGTGTAGCGAAGCA
SBH(ctrl-3)CTS2	GACGTAGAGTTAGTTAGCTAGAGAAAGTAAGTCGGAGTACTGTCCT
SBH(5)CTS1	GCGCTACACGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(5)CTS2	GAGTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(10)CTS1	GCACGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(10)CTS2	GTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(15)CTS1	GGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(15)CTS2	GCTTACGAAAGTAAGTCGGAGTACTGTCCT
nv(5)CTS1	GAAGCA
nv(5)CTS2	GTCCT
nv(10)CTS1	GTAGCGAAGCA
nv(10)CTS2	GTACTGTCCT
nv(15)CTS1	GCGTGTAGCGAAGCA
nv(15)CTS2	GTCGGAGTACTGTCCT
SBH(17)CTS1	GCTGAAAAGTCGCGTGTAGCGAAGCA
SBH(15G)CTS1	GACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(15)CTS1	GGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(13G)CTS1	GCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(13)CTS1	GGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(12)CTS1	GCGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(10)CTS1	GCACGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(6)CTS1	GCTACACGCGACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(17)CTS2	GTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(15)CTS2	GCTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(13)CTS2	GGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(10)CTS2	GTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(9)CTS2	GCTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(8)CTS2	GACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(7)CTS2	GCTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT
SBH(6)CTS2	GGTACTCCGACTTACGAAAGTAAGTCGGAGTACTGTCCT

iSBH(0)Csy4(full)CTS1	GCTGCTTCGCTACACGCGACTGTTCACTGCCGTATAGGCAGCTAAGAAAAGTCGCGTGTA GCGAAGCA
iSBH(0)Csy4(20)CTS1	GCTGCTTCGCTACACGCGACTACTGCCGTATAGGCAGCTAAAGTCGCGTGTAGCGAAGCA
SBH(0B)CTS1	GCTGCTTCGCATCACGATACTGAAAAGTCGCGTGTAGCGAAGCA
SBH(0B*)CTS1	GCTGCTAGGCTTGACGATACTGAAAAGTCGCGTGTAGCGAAGCA
iSBH(0B)Csy4(full)CTS1	GCTGCTTCGCATCACGATACTGTTCACTGCCGTATAGGCAGCTAAGAAAAGTCGCGTGTA GCGAAGCA
iSBH(0B)Csy4m(full)CTS1	GCTGCTTCGCATCACGATACTGTTCAGTGCCGTATAGGCACCTAAGAAAAGTCGCGTGTA GCGAAGCA
iSBH(0B)Csy4(medium)CTS1	GCTGCTTCGCATCACGATACTGCCGTATAGGCAGCTCGCGTGTAGCGAAGCA
iSBH(0B)Csy4(medium)CTS2	GCAGGACAGTACAGCGAACCCTGCCGTATAGGCAGCTAAGTCGGAGTACTGTCCT
iSBH(0B)Csy4(nano)CTS1	GCTGCTTCGCGTACTGCCGTATAGGCAGCGTAGCGAAGCA
iSBH(0B)Csy4(nano)CTS2	GCAGGACAGTGACACTGCCGTATAGGCAGCTAGTACTGTCCT
iSBH(0B)ASOα-CTS2	GCAGGACAGTACAGCGAGATACGAACATCCCTAACAGTAAGTCGGAGTACTGTCCT
iSBH(0B)ASOm-CTS2	GCAGGACAGTAGACCGAGAAGTACACCCACTATCAGACTAGTCGGAGTACTGTCCT
iSBH(0B)ASOβ-CTS1	GCTGCTTCGCATCACGATACTCACACTACGCATCCAGTCGCGTGTAGCGAAGCA
iSBH(0B)ASOm-CTS1	GCTGCTTCGCATCACGATAGTACACCCACTATCAGACTCGCGTGTAGCGAAGCA
iSBH(0B)ASOγ(medium)CTS1	GCTGCTTCGCATCACAAATCCTCAGTCGCGTGTAGCGAAGCA
iSBH(0B)ASOδ-CTS1	GCTGCTTCGCATCACGATAGTACACCCACTATCAGACTCGCGTGTAGCGAAGCA
iSBH(0B)ASOδ-CTS2	GCAGGACAGTAGACCGAGAAGTACACCCACTATCAGACTAGTCGGAGTACTGTCCT
iSBH(0B)Cse3(full)CTS2	GCAGGACAGTACAGCGAGATACAGTCCCCACGCGTGTGGGGATGGAGTAAGTCGGAGTAC TGTCCT
iSBH(0B)Cse3(medium)CTS2	GCAGGACAGTTGTCCTAGGTCCCCACGCGTGTGGGGATGGTCGGAGTACTGTCCT
iSBH(0B)Cse3(nano)CTS2	GCAGGACAGTTGAGTCCCCACGCGTGTGGGGATGGTACTGTCCT
iSBH(0B)Cas6A(full)CTS2	GAGGACAGTAGACCGAGATACCAAGGGATTGAGCCCCGTAAGGGGATTGGTAAGTCGGAG TACTGTCCT
iSBH(0B)Cas6A(medium)CTS2	GAGGACAGTAGACCAAGGGATTGAGCCCCGTAAGGGGATTGGAGTACTGTCCT
iSBH(0B)Cas6A(nano)CTS2	GAGGACAGTAGGGATTGAGCCCCGTAAGGGGATACTGTCCT
SAM-SCR	GGGTCTTCGAGAAGACCT
SAM-ASCL1	GCAGCCGCTCGCTGCAGCAG
SAM-HBG1	GGCTAGGGATGAAGAATAAA
SAM-IL1B	GAAAACAGCGAGGGAGAAAC
SBH(0)SAM-ASCL1	GCCTGCTGCAGCGAGCGGCTGCGAAAGCAGCCGCTCGCTGCAGCAG
iSBH(0B)SAM-Csy4(nano)ASCL1	GCTGCTGCAGATACTGCCGTATAGGCAGCTCGCTGCAGCAG
iSBH(0B)SAM-Csy4(nano)HBG1	GCTTTATTCTGACTGCCGTATAGGCAGCTGAAGAATAAA
iSBH(0B)SAM-Csy4(nano)IL1B	GCGTTTCTCCAAACTGCCGTATAGGCAGCGAGGGAGAAAC
iSBH(0B)SAM-Cas6A(medium)HBG1	GTTTATTCTTATTCAAGGGATTGAGCCCCGTAAGGGGATTGATGAAGAATAAA
iSBH(0B)SAM-Cas6A(medium)IL1B	GTTTCTCCCGAGCAAGGGATTGAGCCCCGTAAGGGGATTGCGAGGGAGAAAC
iSBH(0B)SAM-ASOλ-HBG1	GTTTATTCTTGTTCCCATGCCAACACACACACAACGGCTAGGGATGAAGAATAAA
iSBH(0B)SAM-ASOμ-IL1B	GGTTTCTCCCAGGCTGAATTCAACCAAACCAACACGAAAACAGCGAGGGAGAAAC
iSBH(0B)SAM-ASOε-HBG1	GTTTATTCTTATTCCCGTGGTTACCGCGCCTCCACACCTAGGGATGAAGAATAAA
iSBH(0B)SAM-ASOε-IL1B	GGTTTCTCCCGAGCTGAGGGTTACCGCGCCTCCACACCAACAGCGAGGGAGAAAC
SBH(0B)HHRz-CTS1	GCTGCTTCGCATCACGATACTGGTACATCCAGCTGATGAGTCCCAAATAGGACGAAACGC GCCAAAGCGCGTCCTGGATTCCACAGTCGCGTGTAGCGAAGCA
SBH(0B)mHHRz-CTS1	GCTGCTTCGCATCACGATACTGGTACATCCAGCTGATGAGTCCCAAATAGGACGAGACGC GCCAAAGCGCGTCCTGGATTCCACAGTCGCGTGTAGCGAAGCA

iSBH(0B)aHHRz-CTS1	GCTGCTTCGCATCACGATACTCGTACATCCAGCTGATGAGTCCCAAATAGGACGAAACCC ATACCAGCCGAAAGGCCCTTGGCAGGGCGTCCTGGATTCCAGAGTCGCGTGTAGCGAAGC A
iSBH(0B)aHHRz-CTS2	GCAGGACAGTACAGCGAGATACGGTACATCCAGCTGATGAGTCCCAAATAGGACGAAACC CATACCAGCCGAAAGGCCCTTGGCAGGGCGTCCTGGATTCCACGTAAGTCGGAGTACTGT CCT
iSBH(0B)aHHRz-CTS3	GCTACGATCCTTGTGTAACTACGGTACATCCAGCTGATGAGTCCCAAATAGGACGAAACC CATACCAGCCGAAAGGCCCTTGGCAGGGCGTCCTGGATTCCACGTACATACAGTAGGATC GTA
nv-PDL1	GTCTTCTTGGTATGGTCCTAA
iSBH(0B)Csy4(full)PDL1	GCTTAGGACCTAACCAACCACTGCCGTATAGGCAGCTTCTTGGTATGGTCCTAA

Table 3 – Antisense oligonucleotide triggers

NAME	Sequence	Corresponding ASO-iSBH
ASO(14)α	T*G*T*TAGGGATGT*T*C*G	iSBH$^{(0B)}$ASOα-CTS2
ASO(25)α	T*A*C*T*GTTAGGGATGTTCGTAT*C*T*C*G	iSBH$^{(0B)}$ASOα-CTS2
ASOβ	A*C*T*G*GATGCGTAGTGT*G*A*G*T	iSBH$^{(0B)}$ASOβ-CTS1
ASOγ	C*A*C*G*CGACTGAGGATT*T*G*T*G	iSBH$^{(0B)}$ASOγ$^{(medium)}$CTS1
ASO(20)α	T*A*C*T*GTTAGGGATGTT*C*G*T*A	iSBH$^{(0B)}$ASOα-CTS2
ASOδ	A*G*T*C*TGATAGTGGGTG*T*A*C*T	iSBH(0B)ASOδ-CTS1/CTS2
ASOλ	G*C*C*G*TTGTGTGTGTGT*T*G*G*C	iSBH$^{(0B)}$ASOλ-HBG1
ASOτ	T*T*C*G*TGTTGGTTTGGT*T*G*A*A	iSBH$^{(0B)}$ASOτ-IL1B
ASOε	G*G*T*G*TGGAGGCGCGGT*A*A*C*C	iSBH$^{(0B)}$ASOε-HBG1/IL1B

Table 4 – qPCR primer pairs

Target mRNA	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
GAPDH	AACAGCGACACCCACTCCTC	CATACCAGGAAATGAGCTTGACAA
5pCTS-EGFP	CACTATAGGGAGACCCAAGC	TGAACTTGTGGCCGTTTAC
dCas9-VP64 (SAM)	AACCTATGCCCACCTGTTCG	ATCCAGGATTGTCTTGCCGG
ASCL1	CCCCAACTACTCCAACGACT	GGTGAAGTCGAGAAGCTCCT
HBG1	GTTGTCTACCCATGGACCCA	TCTCCCAAGGAAGTCAGCAC
IL1B	CGAATCTCCGACCACCACTA	AGGGAAAGAAGGTGCTCAGG

Sequence 1 – U6-sgRNA-6xT cassette

(U6 promoter, scramble spacer segment with 2x BbsI sites, sgRNAscaffold, terminator)

GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTGG
AATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTG
GGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTAT
TTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTT
AGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG
TGCTTTTTT

Sequence 2 – allosteric sgRNA mimic

(CTS2 spacer, scaffold sgRNA, α segment, stem loop)

GTAAGTCGGAGTACTGTCCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCCAGTACTCCGACTTACGTCCGGGAAGGGGTTCGCAAGCGG
AAACGC

Sequence 3 – Human codon optimized *Pseudomonas aeruginosa Csy4*

(Kozak sequence, 5' NLS, Csy4)

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGAAAACCTGTA
CTTCCAATCCAATGCAGCTAGCGACCACTATCTGGACATCAGACTGAGGCCCGATCCTGAGTTCCCTC
CCGCCCAGCTGATGAGCGTGCTGTTTGGCAAGCTGCATCAGGCTCTGGTCGCCCAAGGCGGAGACAGA
ATCGGCGTGTCCTTCCCCGACCTGGACGAGTCCCGGAGTCGCCTGGGCGAGCGGCTGAGAATCCACGC
CAGCGCAGACGATCTGCGCGCCCTGCTGGCCCGGCCTTGGCTGGAGGGCCTGCGGGATCATCTGCAGT
TTGGCGAGCCCGCCGTGGTGCCACACCCAACACCCTACCGCCAGGTGAGCCGCGTGCAGGCCAAGTCA
AATCCCGAGAGACTGCGGCGGAGGCTGATGAGGCGACATGATCTGAGCGAGGAGGAGGCCAGAAAGAG
AATCCCCGACACAGTGGCCAGAGCCCTGGATCTGCCATTTGTGACCCTGCGGAGCCAGAGCACTGGCC
AGCATTTCAGACTGTTCATCAGACACGGGCCCCTGCAGGTGACAGCCGAGGAGGGCGGATTTACATGC
TATGGCCTGTCTAAAGGCGGCTTCGTGCCCTGGTTCTGA

Sequence 4 – Human codon optimized *Streptococcus thermophilus* Cse3

(Kozak sequence, 5' NLS, Cse3)

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGAAAACCTGTA
CTTCCAATCCAATGCAGCTAGCTGGCTGACGAAGCTCGTGCTGAACCCCGCGTCACGCGCTGCCCGAC
GAGATCTCGCTAATCCCTACGAAATGCATCGAACACTGAGCAAGGCAGTTTCACGGGCCCTCGAGGAA
GGGCGCGAGCGACTGCTGTGGAGACTGGAGCCAGCTAGGGGGCTGGAGCCCCCTGTGGTCTTGGTGCA
AACACTTACTGAGCCTGATTGGAGCGTGCTTGATGAAGGCTACGCTCAGGTATTCCCCCCCAAGCCCT
TTCACCCTGCATTGAAACCCGGCCAGCGCCTGCGGTTTAGGCTCCGCGCAAATCCGGCAAAAAGGCTG
GCCGCAACAGGTAAGCGCGTTGCACTGAAGACACCTGCGGAAAAGGTGGCATGGCTGGAGAGGCGCTT
GGAGGAAGGGGGGTTCCGCCTGCTTGAGGGAGAACGAGGCCCGTGGGTACAGATCCTCCAGGATACTT
TCCTCGAAGTCCGGAGAAAGAAAGACGGCGAAGAGGCGGGAAAGCTGCTGCAGGTGCAGGCCGTGCTG
TTCGAGGGACGGCTGGAGGTGGTAGATCCCGAGCGCGCCCTGGCCACCCTGAGACGCGGAGTGGGGCC
TGGAAAAGCTCTCGGCCTTGGGCTGCTTTCCGTCGCCCCTTGA

Sequence 5 – Human codon optimized *Streptococcus thermophilus* Cas6A

(Kozak sequence, 5' NLS, Cas6A)

GCCACCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGAAAACCTGTA
CTTCCAATCCAATGCAGCTAGCGTGCTTGCCGCGCTTGTGCTCGTGTTGGAAGGGGAAGGCCTCCCGG
AGCCGTTGGGTCTTAGGGGTTTCTTTTACGGCCTTTTGCGCGAGGTAGCGCCAGAGGTACACGATCAA
GGAGAAAATCCTTTTGCTCTGGGATTCGGCGGGAGGGAAGGCGCAGCTTGGGCTAGGGTTAGTCTTCT
CGTAGAAGGGTTGTATGCGCGACTCGCACCTAGACTCTATGCGCTTGAAGGTGAGGAAGTTCGCCTTG
GACCTCCGTTCCGGGTACGCGCAGTCTTGCAAGAGGGCCATCCTTGGGCTGGAGTTTCAACGTATCCA
CGCTTGTTCCAGGGACCTCCCTCCAGAGATCTGGCCCTCCGATTTGCAAGTCCAACCTTCTTCCGCCG
GAAAGGCGTTCACTACCCCGTGCCGGAACCTCGCCTGGTTCTGGAGTCCCTCCTGCGGCGGTTGGAAG
CATTCGGTCCACTTAAAGCCCCGGAAGGAGTGAGAGAAGCTCTTCTGGAGAGAACCACGGTGCGATCA
TTGGAGGGACGCACACTTCCGGCGAGGACGGAAGTTGACACAGCTGGATTTGTAGGCAGGGTCGTGTA
CCACCTGCCTAGAGCGACAGAGGAAGAAGCACTCTGGTTGTCTGCGTTGGGCCGATTCGCGTTCTATT
CAGGAGTAGGTGCAAAGACCAGCTTGGGATATGGTAGAGCACGAGCAGAATCTGCGTAG

Sequence 6 –full plasmid sequence, *pcDNA-CMVp-EGFP-pA*

(CMVp, polyA signal, EGFP)

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGC
TACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGC
TTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATT
ACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCC
TGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAA
TAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA
GTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
TGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT
CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG
TAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAG
CTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAG
ACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTACCGAGCTCGGATCCCACAACCATGGTGAGCAA
GGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA
AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTT
CAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC
AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCA
AGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCA
CTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGAT
GATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTC
TGCTAGGATCAATGTGTAGATGCGCGGCGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGA
TCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGAC
CCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGC
AGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGG
GTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCG
CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCC

GGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCT
CGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTC
GCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAAC
CCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGA
GCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTC
CCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAA
AGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTC
CCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTG
ACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAG
GAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCT
GATCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTT
CGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAG
GAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTAT
CGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCT
GACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCG
CTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGG
TTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGC
TGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTC
TCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGC
TCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGG
GGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGA
CGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGC
ATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGG
TCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCG
CGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGT
CCGAGGGCAAAGGAATAGCACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGG
CTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCT
TCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTC
ACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCA
TGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAA
ATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCC
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTC
GTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCG
CTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAG
GCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCA
AAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGC
ATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTT
TCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCG
TTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAC
TATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAA
AAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCAC
GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGA
AGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAA
CTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCG
GCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTT
ATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTT
TGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTC
AGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTC
CTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCAC
TGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAG
TCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGC

GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTT
CCCCGAAAAGTGCCACCTGACGTC

Sequence 7 –full plasmid sequence, *U6p-sgRNA-6xT-iBlue*

(U6p, sgRNAscaffold, SV40p, iBlue, SV40 polyA signal)

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCC
GCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAG
CAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGG
TTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTG
ACTAGTGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAG
AGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTA
GAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCAT
ATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACG
AAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA
GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGCTAGCAGGCATGCTGG
GGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATC
CCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCAC
GTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTG
CTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCG
CCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTT
TGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAA
TTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGT
ATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGC
AGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTC
CGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATT
TTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGG
AGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCG
GATCTGATCGCCACCATGTCGGTACCGCTGACTACCTCAGCATTCGGCCACGCGTTTCTGGC
TAACTGTGAACGCGAGCAGATCCACCTGGCGGGCTCCATTCAGCCGCACGGTATCCTGCTGG
CTGTGAAAGAGCCGGACAACGTGGTGATCCAGGCTTCTATTAACGCTGCGGAGTTCCTGAAC
ACCAACTCTGTTGTTGGCCGTCCGCTGCGTGACCTGGGCGGCGATCTGCCTTTGCAGATCCT
GCCGCACCTGAACGGCCCGCTGCACCTGGCTCCGATGACCCTGCGTTGTACCGTGGGTTCTC
CGCCGCGTCGTGTGGACTGTACCATTCATCGTCCGTCTAACGGCGGCCTGATCGTAGAACTG
GAACCAGCAACCAAGACCACTAACATTGCGCCGGCTCTGGACGGTGCGTTTCATCGTATCAC
TTCTTCATCCTCCCTGATGGGCCTGTGTGACGAAACCGCGACTATTATCCGTGAGATTACTG
GCTACGACCGTGTGATGGTAGTACGTTTCGATGAAGAGGGTAATGGCCAAATTCTGTCCGAA
CGTCGTCGTGCGGACCTGGAAGCGTTCCTGGGTAACCGCTACCCGGCGTCTACTATTCCGCA
GATCGCTCGTCGCCTGTACGAACATAACCGTGTTCGCCTGCTGGTAGATGTGAACTATACTC
CGGTTCCGCTACAGCCGCGCATCAGCCCGCTGAACGGTCGTGATCTGGATATGTCCCTGTCT
TGCCTGCGCTCTATGTCCCCGTGCCACCAGAAATACATGCAGGACATGGGCGTTGGCGCGAC
CCTGGTTTGCTCTCTGATGGTGTCTGGTCGTCTGTGGGGTCTGATCGCTTGCCACCACTACG
AACCGCGCTTCGTTCCGTTCCACATTCGCGCTGCTGGCGAAGCGCTGGCGGAAACTTGTGCG
ATCCGCATCGCGACGCTGGAGAGCTTTGCACAGTCTCAGTCCAAATGATGTACAAGTAAAGC
ACGTGCTACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAATCGTT

TTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCA
CCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
CAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCT
TATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTT
TCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCC
GCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG
AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCG
TTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA
GGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGAC
GCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA
AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGG
TCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTA
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGC
CACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGT
GGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTT
ACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT
TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTA
AAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCT
CAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACG
ATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACC
GGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG
CCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTC
GTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCA
TGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCC
GCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGT
AAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGC
GACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA
AAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT
GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCA
CCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG
ACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
TTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC
CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC