Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

Tuning up Transcription Factors for Therapy

Attila Becskei [1,*]

[1] Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland
* Correspondence: attila.becskei@unibas.ch; Tel.: +41-61-207-22-22

Abstract: The recent developments in the delivery and design of transcription factors put their therapeutic applications within reach, exemplified by cell replacement, cancer differentiation and T-cell based cancer therapies. The success of such applications depends on the efficacy and precision in the action of transcription factors. The biophysical and genetic characterization of the paradigmatic prokaryotic repressors, LacI and TetR and the designer transcription factors, TALE and CRISPR-Cas9 revealed common rules, which can help the optimization of activators and repressors. Further studies will be required to analyze the linkage between dissociation constants and enzymatic activity, the role of phase separation and squelching in activation and repression, and the long-range interaction of transcription factors with epigenetic regulators in the context of the chromosomes. Understanding these mechanisms will help to tailor systematically optimized designer transcription factors to the needs of specific applications.

Keywords: LacI; Lac repressor; TetR; TAL-Effector; transcription activator-like effector; Cas9d; dissociation rate constant

1. Introduction

Transcription factors (TF) define what combination of genes is expressed in a given condition, in a given point of space and time and thus, they are highly appropriate to control cellular phenotypes. Indeed, TFs have been long known to be able to reprogram one cell type into another [1]. With appropriate combination of TFs it became possible to reprogram differentiated cell types even into embryonic stem cells [2], from which nearly any cell type can be obtained, opening the way to cell replacement therapies. This success refocused the attention to TFs as possible tools in medical therapy.

Two further discoveries increased the practical applicability of TFs, both of them revolve around the combinatorial principle. First, it turned out that eukaryotic TFs can be viewed as combinations of DNA binding and regulatory domains, which facilitated the design of various regulators targeted to a specific chromosomal location. Second, some natural and synthetic TFs follow a combinatorial principle in target recognition. Each element or domain in a TF uniquely define the nucleotide in the DNA it recognizes. Thus, a TF can be designed to recognize an arbitrary DNA sequence, simply by the permutation of these elements. This combinatorial principle is characteristic of the TALEs, CRISPR-Cas and zinc-finger TFs [3].

Having TFs that are able to target TFs arbitrary sequences, it has become possible to optimize their efficacy and specificity using kinetic-biophysical principles just as recent efforts to optimize the biophysical properties of the binding of small molecule drugs, such as affinity and the association rates, improved the drug discovery and design [4, 5].

Since many reviews have focused on the biotechnological optimization of the designer TFs, this review focuses on how their genetic-biochemical and biophysical properties affect the efficacy and specificity of designer and commonly used prototypical TFs. Following the brief introduction of these TFs, their possible applications are listed, in order to appreciate the needs of the TF optimization.

2. The modularity of transcriptional regulators

2.1. The modularity principle in sequence recognition

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

2.1.1. The TAL Effector

For a long time, pathogenic bacteria have been known as heavy force combatants, equipped with multiple weapons and defence shields. On the other hand, viruses possess only few genes which makes them rely on the host cell, often acting as agents that reprogram the cell. The plant pathogens from the genus Xanthomonas imitate in this sense the viruses, since they use an elaborate scheme to reprogram the host plant cell. They possess transcription activator-like effectors (TAL effectors, TALEs), proteins with the ability to directly bind the promoters of genes in the host; consequently, they induce their expression so that it helps bacterial colonization. Metaphorically, TALEs trick the plant into activating weak points that allow an invasion. This has been possible thanks to the remarkable DNA binding mechanism, known as the TALE code: each base pair is recognized by a specific repeat (protein domain) in the TALE [6]. Amino acids at positions 12 and 13, termed repeat variable di-residues (RVDs), determine the base preferences of a repeat.

2.1.2. The CRISPR-Cas system

The second well-known DNA-binding protein with a remarkable modularity in the sequence recognition are the Cas endonuceases, which are also of bacterial origin. The genomes of most Bacteria and Archaea harbour Clustered regularly interspaced short palindromic repeats (CRISPR), which are involved in resistance to bacteriophages. When bacteria encounter bacteriophages, they integrate sequences derived from phage genomic sequences. Removal or addition of such sequences modifies the phage-resistance phenotype of the cell [7]. These DNA sequences are transcribed into the CRISPR RNA, which binds to the Cas9 protein. This heterodimer binds to the target DNA, which is complementary to the CRISPR RNA, and cleaves it preventing new infection by the phages [8]. Cas9 from Streptococcus pyogenes is possibly the most well-characterized CRISPR-Cas system, which has been harnessed for genome editing in many eukaryotes. Cas9 has been also repurposed for transcriptional regulation, relying on the catalytically inactive Cas9 variant, dCas9.

2.1.3. Zinc finger proteins

The above examples may convey the idea that prokaryotes are the major source of TFs with a modular design. However, the third major class of modularly designed TFs relies on zinc fingers, which are widespread in eukaryotes, even though more recently bacterial zing fingers have been also identified [9]. Typically, two cysteines and the two histidines coordinate a zinc ion to form a compact structure that determines the DNA sequence to be recognized. Although the creation of zinc fingers was a critical advance in gene editing and the design of synthetic TFs, their use has grown less rapidly because of the engineering challenges associated with context- dependent assembly constraints [3].

2.1.4. Non-modular prototypic DNA binding proteins: the lac repressor (lacI) and the tet reperessor (tetR)

Two further TFs are included in this review that do not follow the modularity principle in sequence recognition but have played a major role in understanding transcriptional regulation and were widely used in a variety of organisms: lacI and tetR.

LacI is expressed as a low copy protein and regulates the genes in response to lactose. Upon complexing with lactose, lacI dissociates from the operator sequence in the lacZYA promoter, which leads to expression of proteins involved in lactose uptake and metabolism [10]. Besides being one of the first TF to be discovered, it can be conveniently controlled with lactose or lactose analogues, which contributed to its widespread use.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

The tetR repressor regulates the expression of the tetA tetracycline pump, a key determinant of bacterial resistance against tetracycline antibiotics. When a tetracycline permeates the cell membrane, it binds to the tetR, which then dissociates from the tet operator in the promoter of the tetA gene, enabling a high expression of the pump which then pumps out the antibiotic [11]. A few mutations in the amino acid sequence were sufficient to switch its behavior with respect to tetracycline binding: unlike the tetR, the reverse tetR mutant associates with the tet operator upon being complexed with the tetracycline [12].

2.2. The modularity principle in eukaryotic gene expression: the convenient conversion of a prokaryotic repressor into a eukaryotic activator

The eukaryotic transcriptional activators are modular consisting of a DNA binding domain and transcriptional activation domain, which stands in contrast to prokaryotes. Prokaryotic transcriptional activators, with few exceptions, are not modular [13].

The modularity was unveiled in the model eukaryotic organism, the budding yeast, by examining various chimeras of the potent transcriptional activator Gal4. By replacing the DNA binding domain of the Gal4 by the E. coli LexA repressor, the hybrid LexA-Gal4 fragment was fully capable of activating transcription in yeast [14]. In this case, the prokaryotic repressor was acting solely as a DNA binding domain that tethered the activation domain to the DNA. This convertibility of prokaryotic repressors into DNA binding domains in eukaryotes, relies on the fact that most prokaryotic repressors simply act as a roadblock in the regulation of prokaryotic gene expression, blocking the passage of the RNA polymerase (Figure 1). This logic is different from eukaryotic repressors that can interact with the polymerase even when they are bound upstream of the transcription initiation complex and do not block the passage of the polymerase [15].

A study aimed at the systematic identification of activation domains revealed that while modularity is widespread it is not an absolute rule, and activation domains can overlap with structured DNA binding domains [16] . The activation domains of different TFs are typically enriched in a specific type of amino acids, such as histidine, proline, acidic amino acids or glutamine.

The modularity of transcriptional factors permits a wide range of combinations by fusing a DNA binding domain with regulatory domains including activation and repression domains, such as the VP64 activation and KRAB repression domains [17]. Ongoing search for high activation potential across multiple cell lines has led to the design of novel activation domains. The VPR and SunTag has a higher activation potency than VP64 when fused to TALEs or Cas9d. [18, 19]. These activation domains can be modulated by fusing them to the estradiol receptor, which renders the activation inducible by the estradiol analogue 4-OH-tamoxifen [20]. For example, by controlling the expression level of CARs by tamoxifen, it was possible to modulate the killing activity of lymphocytes [21].

DNA binding domains can be also fused to endonucleases, which enable genome editing even in mammalian genomes. In addition to the DNA, RNA can be also targeted. Recent discoveries of the new Cas family members have extended the scope of the applications, exemplified by targeted RNA degradation and RNA-based manipulations, which rely on Cas13a, an RNA-guided RNA ribonuclease [22].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

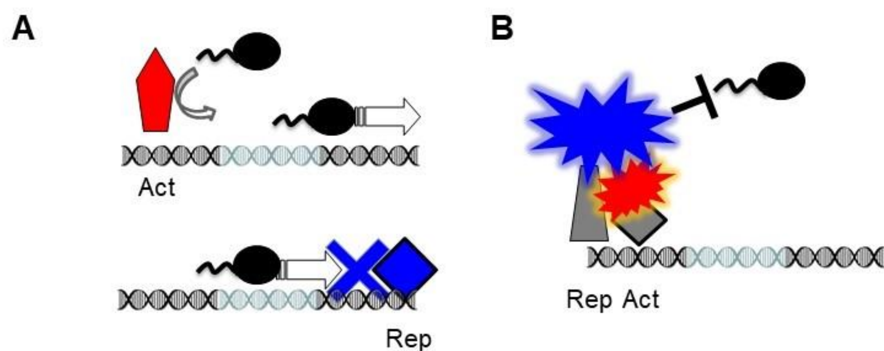

Figure 1. Basic mechanisms of transcriptional regulation. The light blue segment of the DNA denotes the sequence recognized by the RNA polymerase (e.g. TATA box). (A) In prokaryotes, the RNA polymerase binds to the core promoter directly, which can be enhanced by activators. Prokaryotic repressors are simply DNA binding proteins, which act by blocking the binding of the polymerase along the DNA or its progression as roadblock. Thus, it must bind at or downstream of the RNA polymerase binding sites. (B) In eukaryotes, most activators and repressors have two domains. The DNA binding domains (gray) tether the activation and repression domains to the DNA. Even a prokaryotic repressor can act as a DNA binding domain. An activation domain is required to recruit the RNA polymerase, whereas the repressor domain can shield the activation domain, even if it binds upstream of the activator.

3. Potential therapeutic applications of TFs

The therapeutic application of TFs depends on their delivery into the cells and the extent to which the expression of genes can be altered. For most genetically inherited diseases, the replacement of the defective gene, or the introduction of a healthy gene into cells that compensate the loss of gene function, provides the most direct solution. The first human gene therapy was based on such a curative gene expression: a variant of the lipoprotein lipase was expressed in muscle cells to compensate the defective gene (Table 1) [23].

A more direct application of TFs to correct gene expression is appropriate in cases, when the gene is not mutated but a related function declines in a pathophysiological condition. Such a goal was set when a zinc-finger TF was targeted to the VEGF promoter, in order to express the VEGF, which leads to the formation of new blood vessels, in order to boost the compromised microvasculature in diabetic neuropathy (Table 1) [24].

TF can be used also in theranostic devices, which can diagnose a disease state and trigger an autonomously regulated therapeutic response, exemplified by cells expressing a bacterial uric acid sensor. The uric acid concentration is converted by a TF to the appropriate production rate of urate oxidase, which then can control uric acid concentration in hyperuricemic mice [25]. Such an approach is relevant for the treatment of diseases associated with hyperuricemia, like gout.

TFs are expected to be employed in cell-based cancer therapies, as well. Patient T cells are harvested and engineered to express a cancer specific chimeric antigen receptor (CAR) (Table 1). In most T cell therapies, a single cancer antigen is expressed but it would be advantageous to detect specific combinations of presence or absence of antigens and other markers. In this case, the TFs can perform specific logical operations that decode multiple inputs and generate a single output to activate a T-cell response to kill a cell [26]. The reengineering of T cells may be even more thorough: to overcome

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at *Molecules* **2020**, *25*, 1902; doi:10.3390/molecules25081902

the effects of an immunosuppressive microenvironment, a frequent condition in cancers, the T cells can be modified to additionally express immune-modulatory proteins, including ligands and cytokines under the control of synthetic factors [27].

Table 1. Therapies or clinical trials employing control of gene expression

Disease	Therapeutic setting
Lipoprotein lipase (LPL) deficiency	
A rare autosomal recessive lipid disorder (1:1000 000). In female patients, the disease is manifested during pregnancy. The failure to produce active LPL protein causes severe hypertriglyceridemia, associated with a high incidence of life-threatening acute pancreatitis attacks. .	Alipogene tiparvovec (Glybera), the first human gene therapy, results in sustained expression of the naturally occurring, gain-of-function variant of the human LPL gene by muscle cells. The adenoassociated viruses that carry the LPL gene were injected intramuscularly. The number of pancreatitis attacks was reduced (but not eliminated) after the gene therapy [23, 28, 29].
Peripheral diabetic neuropathy	
	(Clinical trial)
A common complication of diabetes. The gradual decline of the functionality of the microvasculature leads to poorer neuronal signal conduction in affected extremities, causing pain and/or loss of sensation. Consequently, diabetic neuropathy sufferers are vulnerable to serious injury and infection.	To promote formation of new blood vessels (revascularization), a plasmid encoding three ZFPs that target a site in the vascular endothelial growth factor A (VEGFA) gene was injected intramuscularly. The ZFPs were linked to a p65 transcriptional activator. The therapy proved safe, with only minimal adverse effects, but with small, non-significant benefit relative to the placebo group [24, 30, 31].
Relapsed leukemia	
By the simultaneous introduction of the CAR and disruption of TCR and CD52 in T cells, functional CAR T cells were generated that could evade host immunity in the unmatched recipients. Such a combination is important for patients who do not have sufficient healthy T cells, which can occur in cases of relapsed leukemia.	Lentiviruses encoding a CAR19 were transduced into the cells, which were then subjected to electroporation of TALEN mRNA targeting TRAC and CD52. Thereafter, residual TCR-expressing cells were depleted [32].

New developments in the delivery methods (see below) may enable the application of TFs to differentiate cancer cells, which counteracts the progression of the dedifferentiated cancer cells [33]. The expression of appropriate TFs can induce cell differentiation in various cancer models (Table 2) [34, 35]. Targeting a TALE or CRISPR-Cas based transcriptional activator to the promoter of such a TF (e.g. Ascl, HNF-4α) could, in principle, induce the differentiation of cancers *in vivo*. It has to be noted that small molecules can also induce TFs that can differentiate cancer cells (Table 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

Table 2. Application of endogenous TFs

Cell type / TF	Outcome
Control of TFs by synthetic gene expression systems	
Expression of the proneuronal TF ASCL in glioblastoma stem cells under the control of the tetON promoter (stable transduction / piggyBac transposon) [36].	Activates neurogenic gene expression program and induces terminal differentiation, which may help the therapy of glioblastoma.
Control of TFs by small molecules	
The addition of the flavonoid Oroxylin A induces the expression of the TF HNF-4α (hepatocyte nuclear factor 4 alpha) [37].	The expression of HNF-4α-targeted genes leads to the differentiation of a model hepatome, blocking cancer progression.

4. Delivery and construction of TFs and endonucleases

The majority of pharmacological therapies rely on small molecules, ranging from antimicrobial antibiotics to antipsychotic drugs. However, macromolecules have been increasingly employed, exemplified by the antibodies and peptide or protein hormones. Since their primary point of action is in the extracellular space, they can be injected, which simplifies their delivery. TFs, as well as endonucleases targeting the genome, are exerting their effect intracellularly, which poses a major challenge for their delivery. However, recent technological advances have opened up new avenues for their delivery.

There are three main possibilities to introduce a TF into a cell: in the form of a protein, or in form of a DNA or RNA encoding the TF.

For the cell based therapies (Table 1), the cells can be transduced / transfected with DNA / RNA ex vivo with standard laboratory techniques. For example, the mRNA encoding the TALEN can be injected into the isolated cells or embryos [32, 38]. The chimeric antigen receptor is typically introduced by lentiviral vectors [22]. Patients that do not have sufficient healthy T cells require donor T-cells, in which the genes have to be deleted to prevent host response. A TALEN has been used to perform such a deletion in a clinically applied T-cell therapy [23]. If the introduced TALEN is encoded by DNA and not RNA, it is important to tightly regulate its expression because off-target effects, especially in the case of the CRISPR-Cas9 system may result in undesired cuts. The expression of the endonuclease can be controlled by another TF, for example, the tetR based rtTA [39].

The proteins can be also injected, which prevents long-term effects. For example, the TALEN protein injected by the bacterial type III secretion system is degraded and disappears 12h after the bacterial injection into human cells [40]. The Cas9 is a particularly stable protein. The reduction of the level by protein degradation may be advantageous in this case because it can attenuate undesired genome editing [41]. The TFs can be also tethered to protein transduction domains, which can cross the cell membranes [42].

In addition to the delivery, the construction of the designer TFs itself has been optimized because the speed of construction is a major determinant of the popularity of a technology. A major drawback of the TALE technology over the CRISPR is the relatively long time it takes to construct the long and repetitive DNA sequence. The ligation alone requires around 5 days. On the other hand, TALEs are can more easily act both as activators and repressors in comparison to CRISPR-Csa9 (see below). This is particularly relevant for therapeutic applications. Indeed, the first therapeutically applied chimeric antigen receptor (CAR) T cells were engineered with TALENs in combination with lentiviral transduction [32]. A recent study has reported that the assembly of the TALE sequence can be reduced to one day [18]. The limitation of this approach is that it is streamlined for TALEs recognizing a 18bpDNA sequence but this is adequate for most applications because it provides sufficient specificity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

5. Biotechnological applications

In addition to medical therapy, the designer DNA-binding proteins are of major interest in biotechnological applications. The DNA binding domains can be used, for example, to recruit enzymes to the DNA in order to increase the local concentration of enzymes. In this case, the DNA serves solely as a scaffold. When multiple enzymes that belong to one pathway are tethered to the DNA by TALEs, the production rate of the final metabolite can be enhanced [43].

The recruitment of GFP-TALE fusions can be also used as a DNA staining agent to monitor enzymatic reactions. A TALE-GFP fusion that recognized a 7-bp DNA was used to monitor the endonucleotic cleavage of single DNA molecules in real time in physiologically relevant conditions [44]. In this study, a binding to AT-rich sequences was observed in addition to the above 7-bp, which is likely to reflect the observation that TALE binding to the DNA is nonspecific in the absence of magnesium in the solution [45]. The advantage of TALE-fluorescent protein fusions over classical DNA stains is that most intercalating DNA dyes generate radical oxygen species or DNA strand breaks upon irradiation, which is used for their detection and can affect the interpretation of real-time observations of cleavage reactions.

6. Biophysical and molecular-genetic properties of the activator and repression domains

Recent biophysical studies have revealed that activation domains tend to form condensates, akin to equilibrium phase separation [46], which is supported by two types of evidence. First, these structures are disrupted by 1,6-hexanediol, which impairs hydrophobic interactions. Second, fluorescence recovery after photobleaching (FRAP) revealed that molecules move in and out of these condensates rapidly, indicating that the components that make up these structures are dynamic and not solid aggregates.

Interestingly, repressors that contain poly-glutamine (Poly-Q) repeats also have the propensity to form aggregates. The function of poly-Q-containing Ssn6 increases with its repeat number until a certain threshold where further expansion leads to aggregation [47]. In this case, the Ssn6 repressor can propagate as a prion [48]. Interestingly, the RNAs encoding proteins with polyQ-repeats also promote phase separation and form RNA droplets [49].

The expression of transcriptional activators, especially with potent transcriptional activation domains, can be toxic to the cells, by a mechanism described as squelching. It is unclear whether squelching is related to aggregation or other forms of phase separation. Since squelching acts through the sequestration of mediators of transcription [50], it is possible that, in principle, the sequestrated molecules form aggregates. It has been observed that expression of transcriptional activators comprising the potent VP16 activation domain often leads to a biphasic expression: they induce transcription shortly after their induction, followed by a decline [51, 52]. Thus, squelching seems to have delayed onset, which is reminiscent of the delayed onset of aggregation, since aggregates grow in size as a function of concentration after a long nucleating lag phase [53].

7. Biophysical and binding properties of TFs

This section focuses on binding properties of the TFs, by comparing kinetic measurements keeping in mind that the comparison of binding constants depends both on the fitting procedure (Box 1) and the employed experimental methods (Tables 3, 4).

Box. 1, Binding affinity, equilibrium dissociation constants and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

The formation of a complex (C) between a protein (P) and (DNA) is determined by the association (binding) and dissociation (unbinding) rate constants, k_a and k_d, respectively:

$$\frac{dC}{dt} = k_a[DNA][P] - k_d[C],$$

The binding affinity is typically expressed in terms of the equilibrium association constant or its inverse the equilibrium dissociation constant with molality units: $K_D = \frac{1}{K_A} = \frac{k_d}{k_a}$

The lower the value of K_D, the stronger the binding. Micromolar binding is considered weak, nanomolar is intermediate, whereas picomolar is strong. The residence time is the inverse of the dissociation rate constant, $\tau = 1/k_d$ which is expressed in seconds or minutes.

Most TFs are not monomeric (M) but form dimers or even tetramers. For a dimeric TF, the dimeric form (P) will represent only a proportion of the total protein amount: $[P] = [M]^2 / K_{D,Dim}$.

Therefore, the value of the estimated binding affinity of a dimeric TF to the DNA, depends on whether the dimerization reaction has been taken into account or not. Since not all studies measure or estimate the dimeric form explicitly, the equilibrium binding constants and the association rate constants, are not necessarily directly comparable when compared from different studies.

Weak dimerization does not result necessarily in a weak transcriptional activation or repression. For example, the yeast Gal4 is possibly the most portent activator in yeast when it binds to multiple sites in a promoter, yet its dimerization constant (K_D = 20 μM in vitro and K_D = 8.5 μM in vivo) reflects a rather weak binding [54, 55].

Dimerization of a TF can amplify the signal transmitted to gene expression when the protein concentration of the TF is relatively low. In this case, a small change in the TF concentration generates a large change in the output. This signal amplification due to dimerization (multimerization) can be used to generate a switch-like response to a sugar and can promote cellular memory [56]. The heterodimerization of the CRISPR guide RNA with the Cas9 also has the potential for signal amplification and nonlinear reaction response, which may explain why a synthetic CRISPR based gene regulatory networks display robust memory and oscillations [57].

7.1. TALE activators and repressors

TALEs bind double-stranded DNA, In addition, they can bind DNA:RNA hybrids, with a slightly lower affinity [58]. The binding affinity to the DNA can vary broadly. In a systematic study, designed TALEs were targeted to 18 bp long DNA sequences, and some of the TALEs displayed a high binding affinity for their target sequence (K_D = 0.16 nM) [59]. Kinetic studies indicate a relatively fast binding – unbinding reaction, with similar parameter values obtained from both in vitro and in vivo measurement: the residence time ranged from 3 to 16 s (Table 3) [60, 61]. When the respective

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

TALEs were fused to an activation domain (VP64), the TALEs with the highest affinity to the target sequence generated also the highest gene expression [59]. However, the correlation between in vitro binding affinity and gene expression is moderate: relatively strong expression can be seen both with strong and intermediate binding affinities.

A comparison between the TALE and tetR based activation system further underscored the high affinity and the potency of the TALE to control gene expression. When tet operator-specific TALEs, with an identical DNA-binding site as the Tet repressor (TetR), were created, the DNA-binding domain of tetTALE alone effectively counteracted trans-activation mediated by the potent tet trans-activator **[62].**

Table 3. TALE binding

Method of measurements	Results
Relation between binding affinity and transcriptional activation [59]	
The binding affinities were measured in vitro with Electrophoretic mobility shift assay (EMSA), while the transcriptional activation was measured with TALE-VP64 fusions.	The apparent K_D spanned four orders of magnitude, from 0.16 nM to 1800 nM.
***In vivo* residence time of TALEs with varying numbers of repeats [61]**	
The DNA residence time of the TF was quantified *in vivo* in U2-OS cells by single molecule imaging of the individual TFs labeled with an organic dye.	The residence times of TALEs comprising 5, 7, 9, 13, 16 and 21 repeats ranged from 3 to 16 s. The 21-repeat TALE had intermediate residence time while the shortest TALE (5 repeats) had the longest residence time.
The effect of increasing numbers of TALE repeats on the DNA binding specificity [63]	
The binding affinities were measured with Electrophoretic mobility shift assay (EMSA) in the presence of magnesium and with fluorescence anisotropy (FA) in the absence of magnesium (150 mM NaCl).	Target specific binding is around 30 times stronger than binding to random sequences (in the presence of magnesium). In the absence of magnesium, the nonspecific binding is ten times stronger.
***In vitro* binding kinetics of TALEs [60]**	
FRET was used to study the in vitro binding of TALEs to DNA, each of them being labelled with fluorescent dyes.	The bimolecular microscopic binding rate constant is 0.4 $nM^{-1}s^{-1}$ and the microscopic unbinding rate constant 0.3 s^{-1} for a 16-repeat TALE.

Importantly, the highest affinity TALE is not necessarily the optimal solution for all needs since the binding to nonspecific DNA correlates with the binding affinity to the target DNA (Figure 2). The relative binding to the non-specific sites depends on the number of TALE repeats; it reaches a minimum with an 18-repeat TALE, when it is around 30 times less than the binding to the specific sites. For shorter and longer arrays of TALE repeats, this ratio is around 10.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

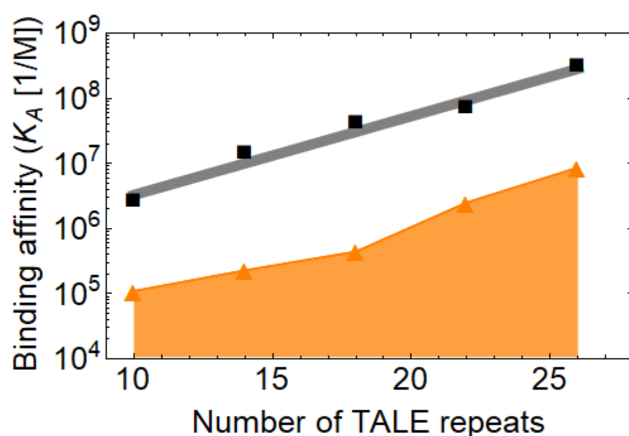

Figure 2. Specific binding and Non-specific binding of TALEs as a function of the number of repeats, plotted using data from Rinaldi et al [63].

The binding to nonspecific sites is affected by the ionic composition in the reaction solution, as well. TALEs demonstrate high sequence specificity only upon addition of small amounts of certain divalent cations (Mg2+, Ca2+). On the other hand, under purely monovalent salt conditions (K+, Na+), TALEs bind to target and random DNA sequences with nearly equal affinity. This effect was confirmed with TALEs having various numbers of repeats [45].

The specificity of the binding can be also enhanced if the positive charge of the protein is reduced, by mutating lysine and arginine residues to glutamine in the TALE protein, which decreases the nonspecific binding to the negatively charged DNA [64].

Even though the DNA binding part of TALE is an array of repeats, a polarity effect breaks the symmetry of the array: the N-terminal repeats recognizing the 5′ end of the target sequence contribute more to the affinity than C-terminal ones [59]. The TALE proteins are capable of rapid diffusion along DNA using a combination of sliding and hopping behavior [65]. The N-terminal region of TALEs is required for the initial non-specific binding and subsequent rapid diffusion along the DNA, whereas the central domain comprising the repeats is required to recognize the target sequence.

The binding to the target sequence can be also controlled by conformational stress. By inducing dimerization of the proteins connected to the N- and C-terminal domains of the TALE, the circularization of the protein locks a strained conformation of the protein, which leads to a reduction of transcriptional activation by TALE activators [66].

7.2. CRISPR-Cas

Three molecules interact to yield the Cas9-CRISPR RNA-DNA complex. Furthermore, the binding to the DNA is followed by an enzymatic reaction. Thus, in comparison to the TALE binding to the DNA, which is a bimolecular event, a more complex kinetics is expected. However, even this expectation was surpassed by the first kinetics analyses of the Cas9-CRISPR complex: the DNA cleavage failed to follow a Michaelis-Menten kinetics [67]. The cleavage reaction stopped soon after mixing the components well before all DNA would be cut, even though a usual enzymatic reaction is expected to proceed until completion. A higher proportion of DNA was cleaved only if the Cas9-RNA concentration was increased. These observations indicate that Cas9 is a single turnover enzyme that remains tightly bound to the DNA after the cleavage reaction [67].

Interestingly, so far only one enzyme has been identified among the Cas9 homologues that acts as multiple turnover enzyme, the Cas9 isolated form *Staphylococcus aureus* [68].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

It is difficult to measure the kinetic parameters of a single turnover enzyme. Gong et al tackled this problem by separating the binding and the enzymatic reactions, by adding magnesium only after the completion of the binding reaction. Cas9.gRNA was first incubated with radiolabeled DNA in the absence of Mg2+. After formation of the Cas9.gRNA.DNA complex, an excess of unlabeled DNA was added and the DNA cleavage was initiated by adding magnesium after various times of incubation. The dissociation rate constant was fitted to single-exponential decay function, which gave a mean residence time of 5 min [69]. This residence time is substantially longer than the few seconds measured for the TALEs (Table 3).

Table 4. CRISPR/Cas9 binding

Method of measurements	Results
Enzymatic and biophysical characterization of DNA cleavage by CRISPR/Cas9 [67]	
The binding affinities were measured with double tethered DNA curtains so that the binding events of quantum dot labelled Cas9-guide RNA were recorded with total internal reflection fluorescence microscopy.	The Cas9-guide RNA bound the target DNA with K_D = 0.5 nM. Even without guide RNA, the upper limit of binding was 25nM for the apo-Cas9–DNA complex.
Detailed kinetic characterization of the CRIPSR/Cas9 binding to the DNA [69]	
Quench flow experiments were performed by mixing Cas9.gRNA complex with 32P-labeled DNA substrate at 37°C. The reaction was stopped by the addition of EDTA at varying time points. The products were separated by polyacrylamide gel electrophoresis.	The dissociation rate constant of DNA from Cas9.gRNA.DNA (k_{off} = 0.0024 s^{-1}), giving a residence time of around 5 min. for the equilibrium DNA binding, K_D = 4 nM.
***In vivo* residence time as a function of varying guide RNA lengths [70]**	
Directly labeled guide RNA (with Broccoli aptamer) and mCherry labelled dCas9 were used to track their binding at their target-site using fluorescence recovery after photobleaching (FRAP) measurements. The guide RNA was targeted to unique sequence at the subtelomeric region.	The residence time and the off-rate of the dCas9/C3-11–guide RNA complex on the C3 target were estimated to be 206 min and 2.9 × 10^{-4} s^{-1}. When the guide RNA length was truncated from 11 to 8 nt, the residence time decreased from 206 to 25 min
***In vivo* binding kinetics of Cas9d in E.coli [71]**	
dCas9 was fused to a fluorescent protein and expressed only a few (about five) molecules per cell. DNA-bound fluorophores were detected as diffraction-limited spots with single-molecule fluorescence microscopy.	The association rate is 2.7 × 10^{-3} min^{-1} $molecule^{-1}$ while the dissociation time varied between 40 and 120 min, depending on the growth condition.

Even longer residence times were measured *in vivo*: more than 3 hours in mammalian cells and between 40 and 120 in Staphylococcus [70, 71]. The in vivo measurements also indicate a relatively short half-life (15 min) of the guide RNA [70], which suggests that the guide RNA may be a limiting factor if it is not expressed at sufficiently high level. The length of the guide RNA is also an important determinant of the binding: when the guide RNA length truncated from 11 to 8 nucleotides, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

residence time decreases 10 times [70], in agreement with an earlier study showing that guide RNAs 12 and 20 nt in length were generating comparably high gene expression, whereas those 8 nt or less in length had almost negligible activation [72].

7. 3. The prokaryotic repressors lacI and tetR

The lacI and tetR are classical prokaryotic DNA binding proteins widely used in eukaryotes, which leads to the question whether their biophysical properties explain their popularity.

The capacity of lacI and tetR to act as activators or repressors in eukaryotes was compared directly in yeast. The tetR-VP16 AD activated expression but the lacI-VP16 failed to do so [73]. When examined whether they can act as a roadblock to repress expression driven by a transcriptional activator, both lacI and tetR repressed transcription, with tetR displaying a somewhat higher efficiency.

The above finding is underscored by the fact that the tetR-VP16 is more frequently mentioned in publications than the lacI-VP16 (see Methods). The lacI-VP16 has been employed in mammalian cells and in tobacco, which suggests that at least some modifications can facilitate its application.
In mammalian cells, it was fused to a nuclear localization signal and activated gene expression when lac operators were positioned either upstream or downstream of the transcription unit **[74].** Promoters containing 14 or 21 operator sequences were induced around 1,000-fold. Activation was inhibited by isopropyl-beta-D-thiogalactoside, confirming the inducibility of the protein. The fusion protein was bifunctional, also acting as a repressor when the promoter contained an operator immediately downstream of the TATA box. In a more recent study, using mammalian embryonic stem cells, Gal4-VP16 activated gene expression while LacI-VP16 failed to do so **[75]**.

The lacI was fused to a Gal4 activation domain in tobacco because the VP16 was less effective [76]. For the DNA binding domain a lac repressor mutant, lacIhis, was used which contains a Y17H mutation because it was estimated to bind lac operator sequences with at least 100-fold greater affinity than wild-type lac repressor.

The above findings indicate that LacI and TetR have similar efficiencies to act as roadblock repressors but the LacI-activation domain fusions have a lower potential to activate gene expression and require a more thorough optimization by using for example mutant versions of LacI or alternative activation domains. The similar potency of the DNA binding domains is confirmed by *in vitro* measurements. K_D= 0.18 nM was estimated for TetR-DNA binding with surface plasmon resonance [12], and K_D = 2 nM was fitted from stop flow data based on fluorescence measurements [77]. In its active state, the reverse TetR binds the DNA with a lower affinity, K_D= 10 nM [12]. The control of TetR and reverse TetR by tetracycline still represents an advantage over TALE and CRISPR because the rapid dissociation permits the measurement of rapid post-transcriptional processes, such as the degradation of RNAs with very short half-lives [78, 79].

For LacI binding it is more difficult to fit the individual dissociation constants due to the linkage of the equilibria between protein monomers, dimers and tetramers [80]. In vitro binding data indicate a strong binding, with picomolar dissociation constants while in vivo data suggest a binding in the nanomolar range *in E. coli* [81]. An even higher discrepancy was found between the in vivo and in vitro data in yeast, where LacI-GFP fusions bound the DNA with micromolar dissociation rate constants [82]. This weak binding was shown to be sensitive to the nature of protein fusions [82], which possibly explains, why more careful design of the activation domain linkage is needed for LacI fusions.

8. Conclusions: optimization of transcriptional activation and repression.

The transcription factors TetR, LacI, CRISPR-Cas and TALE, covered in this review have been widely used in a range of organisms with varying efficiency and specificity. Their biophysical and genetic characterization makes it possible to formulate general suggestions how to design transcription

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

factors to activate or repress gene expression. There are three main aspects of optimization: the intracellular concentration, the binding affinity to the DNA and the integration of the TF activity into the chromosomal regulatory landscape.

While increasing the expression of a TF is an enticing option to boost its effect, the outcome can quickly turn into the opposite. This is particularly true for the activation domains due to squelching (see Section 6). Furthermore, chromatin, activation- and repression-domains, can promote phase separation and the formation of aggregates, which are known to occur at and above a critical concentration [83]. Above the critical point, unintended consequences can occur. Thus it can be useful to keep the intracellular concentration of the TF at intermediate or low level, which can be achieved by increasing reducing the translation by stem-loops, by increasing the degradation rate of the protein and by using appropriate promoters [52, 82]. An increase of the degradation rate affects the TF bound to the DNA, reducing its local effect; therefore, using a weaker promoter to drive the expression of the TF may be the preferable option [82]. At low expression, noise becomes prominent. A variety of promoters are available that can drive expression at a desired level and noise [84].

The comparison of the measurements indicates that all main families of TFs discussed in this review bind the target DNA with high affinity, with dissociation constants in the picomolar to low nanomolar range (Section 7). This strong binding was certainly a key factor that permitted their widespread use. In pharmacology, the affinity of a drug to its target strongly correlates with the biological effect [5]. In the case of TFs this correlation may be less strong due to the interaction of TFs with multiple molecules, in addition to their primary target, the DNA. Indeed, TALEs more efficiently activate gene expression, even when less physical binding to the DNA is observed in comparison to CRISPR-Cas9d-based activators [17, 85]. On the other hand CRISPR-Cas9d based repressors are highly efficient. Similarly, TetR based activators are more efficient than LacI-based activators even though their repression efficiency is similar (Section 7.3). This may reflect the more pronounced steric constrains in the design of the transcriptional activators or a more fundamental difference between the kinetics of activation and repression.

The kinetic parameters (binding and unbinding rate constants) drew a considerable attention in the design of targeted endonucleases. Kinetic models suggest that cleavage of off-target sequences can be reduced by increasing the dissociation from the DNA binding or by reducing the catalytic activity [86]. Such an enhancement of cleavage specificity has been realized in the meantime, by attenuating cleavage rate of the endonuclease or by fusing Cas9 to inhibitor proteins [87, 88]. In the above model, the specificity is affected by the dissociation rate and the enzymatic turnover because the binding and catalytic reactions are coupled. It is not clear whether such a coupling is present in transcription and thus similar principles apply to the specificity in transcriptional regulation. The binding of the TF to the DNA and the initiation of transcription by RNA polymerase represent separate molecular entities. However, there are some indications that transcription has some similarities to the single turnover enzyme. There is experimental evidence that TFs, especially activators, recruit the proteasome during transcriptional activation [89]. Thus, the initiation of transcription by an activator may be followed by its proteolysis, which is reminiscent of the single turnover enzymes. It is commonly thought that CRISPR-Cas9 based transcription elicits more off-target transcription than TALEs, which would follow from the above mechanism since the residence time of CRISPR-Cas9 is much longer (Table 3, 4). However, there are studies that suggest that the two TFs have similar propensities to affect the expression of off-target genes [85].

A systematic exploration is yet to come how binding kinetics affects gene expression but a study on the yeast Rap1 activator suggests that accelerating the binding-unbinding events reduces transcriptional activation, even if the equilibrium binding does not change, as evidenced by chromatin immunoprecipitation assay [90]. Variations in binding kinetics, even when masked by equal equilibrium binding, can affect the timing of the gene response [91].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at *Molecules* **2020**, *25*, 1902; doi:10.3390/molecules25081902

In addition to the classical off-target effect, caused by the binding of the TF to the off-target genes, alternative mechanisms, not involving direct binding, can play an important role, as well. Transcriptional control in higher eukaryotes is known to display long-range effects, and a transcriptional activator can induce gene expression even megabases away from its binding site. This off-target activation can vary with cellular differentiation, as evidenced by targeting a TALE-activator to a gene in the protocadherin gene array. The genes in this array are expressed in neurons, but not in embryonic stem cells and neuronal progenitors [92]. In stem cells, the TALE targeted to a specific protocadherin isoform activates primarily its target gene, and only minimally the adjacent genes. In neuronal progenitors, even more distant off-target genes are activated [93]. Most of this effect is mediated by epigenetic mechanism, such as DNA demethylation, induced by the strong activation domain. Thus, as cells differentiate from stem cells to neurons, the epigenetic activity gradient broadens.

Epigenetic control can be combined with transcriptional control to fortify the gene response. When targeting of a CRISPR-Cas9d activator was coupled with the recruitment of the TET1 enzyme, which catalyzes the first step in the DNA demethylation, the gene activation was substantially increased [94]. The VP16 (or its variants VP64, VP160) domain, which contains a large number of negatively charged amino acids, is likely to cause these changes, independent of transcription, since the recruitment of negatively charged peptides to the DNA can cause large-scale chromatin remodeling, without inducing transcription [95] .

With designer TF, the 3D structure of the chromatin can be also controlled. Two studies have shown that the repression efficiency of TALEs, which act as roadblocks to repress transcription, can be enhanced when the TALE fusion proteins are recruited to two sites so that the fusions contain to protein dimerization domains [96, 97]. Consequently, the intervening DNA segment is looped out. It will be interesting to assess whether dimerization of repressors and loop formation enhances the repression in eukaryotes too.

It is not clear which TFs have the propensity for the long-range effects, and understanding the underlying rules will be important to reduce these indirect off-target effects. A recent bioinformatic analysis suggests that there are two classes of TFs: with short- and long-range regulatory influence. These two classes differ in their chromatin-binding preferences and auto-regulatory properties [98]. The regulatory range is further affected by the 3D structure of the chromatin. Since many factors influencing transcription, activator domains, repressor domains, chromatin modifications can form bodies due to phase separation [83, 99], the models constructed to explain long-range effects may have to take these phenomena into consideration.

9. Methods

In order to compare the usage of LacI and TetR based activators, we searched for publications using the keywords LacI-VP16 and TetR-VP16. "TetR-VP16" retrieves 12 and 489 hits in PubMed and Google Scholar, respectively. LacI-VP16 retrieves 0 and 27 hits, respectively. Out of the 27 hits, 4 publication use lacI-VP16 as a TF (see Section 7) , while several other publications explore the role of VP16 in inducing epigenetic changes in the chromatin.

Funding: A part of this work was funded by the Swiss National Science Foundation (310030_185001).

Acknowledgments: We thank S. Voegeli for reading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

References

1. Jopling, C.; Boue, S.; Izpisua Belmonte, J. C., Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat Rev Mol Cell Biol* **2011,** 12, (2), 79-89.
2. Takahashi, K.; Yamanaka, S., A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol* **2016,** 17, (3), 183-93.
3. WareJoncas, Z.; Campbell, J. M.; Martinez-Galvez, G.; Gendron, W. A. C.; Barry, M. A.; Harris, P. C.; Sussman, C. R.; Ekker, S. C., Precision gene editing technology and applications in nephrology. *Nat Rev Nephrol* **2018,** 14, (11), 663-677.
4. AP, I. J.; Guo, D., Drug-Target Association Kinetics in Drug Discovery. *Trends Biochem Sci* **2019,** 44, (10), 861-871.
5. Folmer, R. H. A., Drug target residence time: a misleading concept. *Drug Discov Today* **2018,** 23, (1), 12-16.
6. Perez-Quintero, A. L.; Szurek, B., A Decade Decoded: Spies and Hackers in the History of TAL Effectors Research. *Annu Rev Phytopathol* **2019,** 57, 459-481.
7. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P., CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007,** 315, (5819), 1709-12.
8. Chen, J. S.; Doudna, J. A., The chemistry of Cas9 and its CRISPR colleagues. *Nat Rev Chem* **2017,** 1, (10).
9. Malgieri, G.; Palmieri, M.; Russo, L.; Fattorusso, R.; Pedone, P. V.; Isernia, C., The prokaryotic zinc-finger: structure, function and comparison with the eukaryotic counterpart. *FEBS J* **2015,** 282, (23), 4480-96.
10. Lewis, M., Allostery and the lac Operon. *J Mol Biol* **2013,** 425, (13), 2309-16.
11. Ramos, J. L.; Martinez-Bueno, M.; Molina-Henares, A. J.; Teran, W.; Watanabe, K.; Zhang, X.; Gallegos, M. T.; Brennan, R.; Tobes, R., The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **2005,** 69, (2), 326-56.
12. Kamionka, A.; Bogdanska-Urbaniak, J.; Scholz, O.; Hillen, W., Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Res* **2004,** 32, (2), 842-7.
13. North, A. K.; Klose, K. E.; Stedman, K. M.; Kustu, S., Prokaryotic enhancer-binding proteins reflect eukaryote-like modularity: the puzzle of nitrogen regulatory protein C. *J Bacteriol* **1993,** 175, (14), 4267-73.
14. Brent, R.; Ptashne, M., A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **1985,** 43, (3 Pt 2), 729-36.
15. Kelemen, J. Z.; Ratna, P.; Scherrer, S.; Becskei, A., Spatial epigenetic control of mono- and bistable gene expression. *PLoS Biol* **2010,** 8, (3), e1000332.
16. Arnold, C. D.; Nemcko, F.; Woodfin, A. R.; Wienerroither, S.; Vlasova, A.; Schleiffer, A.; Pagani, M.; Rath, M.; Stark, A., A high-throughput method to identify trans-activation domains within transcription factor sequences. *Embo J* **2018,** 37, (16).
17. Gao, X.; Tsang, J. C.; Gaba, F.; Wu, D.; Lu, L.; Liu, P., Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. *Nucleic Acids Res* **2014,** 42, (20), e155.
18. Zhang, S. Y.; Chen, H. T.; Wang, J. K., Generate TALE/TALEN as Easily and Rapidly as Generating CRISPR. *Mol Ther-Meth Clin D* **2019,** 13, 310-320.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at *Molecules* **2020**, *25*, 1902; doi:10.3390/molecules25081902

19. Chavez, A.; Tuttle, M.; Pruitt, B. W.; Ewen-Campen, B.; Chari, R.; Ter-Ovanesyan, D.; Haque, S. J.; Cecchi, R. J.; Kowal, E. J. K.; Buchthal, J.; Housden, B. E.; Perrimon, N.; Collins, J. J.; Church, G., Comparison of Cas9 activators in multiple species. *Nat Methods* **2016,** 13, (7), 563-567.
20. Zhao, C.; Zhang, Y.; Zhao, Y.; Ying, Y.; Ai, R.; Zhang, J.; Wang, Y., Multiple Chemical Inducible Tal Effectors for Genome Editing and Transcription Activation. *ACS Chem Biol* **2018,** 13, (3), 609-617.
21. Chakravarti, D.; Caraballo, L. D.; Weinberg, B. H.; Wong, W. W., Inducible Gene Switches with Memory in Human T Cells for Cellular Immunotherapy. *ACS Synth Biol* **2019,** 8, (8), 1744-1754.
22. Mahas, A.; Neal Stewart, C., Jr.; Mahfouz, M. M., Harnessing CRISPR/Cas systems for programmable transcriptional and post-transcriptional regulation. *Biotechnol Adv* **2018,** 36, (1), 295-310.
23. Gaudet, D.; Stroes, E. S.; Methot, J.; Brisson, D.; Tremblay, K.; Bernelot Moens, S. J.; Iotti, G.; Rastelletti, I.; Ardigo, D.; Corzo, D.; Meyer, C.; Andersen, M.; Ruszniewski, P.; Deakin, M.; Bruno, M. J., Long-Term Retrospective Analysis of Gene Therapy with Alipogene Tiparvovec and Its Effect on Lipoprotein Lipase Deficiency-Induced Pancreatitis. *Hum Gene Ther* **2016,** 27, (11), 916-925.
24. Eisenstein, M., Sangamo's lead zinc-finger therapy flops in diabetic neuropathy. *Nat Biotechnol* **2012,** 30, (2), 121-3.
25. Haellman, V.; Fussenegger, M., Synthetic Biology--Toward Therapeutic Solutions. *J Mol Biol* **2016,** 428, (5 Pt B), 945-62.
26. Wu, M. R.; Jusiak, B.; Lu, T. K., Engineering advanced cancer therapies with synthetic biology. *Nat Rev Cancer* **2019,** 19, (4), 187-195.
27. Khalil, D. N.; Smith, E. L.; Brentjens, R. J.; Wolchok, J. D., The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat Rev Clin Oncol* **2016,** 13, (5), 273-90.
28. Senior, M., After Glybera's withdrawal, what's next for gene therapy? *Nat Biotechnol* **2017,** 35, (6), 491-492.
29. Kassner, U.; Hollstein, T.; Grenkowitz, T.; Wuhle-Demuth, M.; Salewsky, B.; Demuth, I.; Dippel, M.; Steinhagen-Thiessen, E., Gene Therapy in Lipoprotein Lipase Deficiency: Case Report on the First Patient Treated with Alipogene Tiparvovec Under Daily Practice Conditions. *Hum Gene Ther* **2018,** 29, (4), 520-527.
30. Geel, T. M.; Ruiters, M. H. J.; Cool, R. H.; Halby, L.; Voshart, D. C.; Andrade Ruiz, L.; Niezen-Koning, K. E.; Arimondo, P. B.; Rots, M. G., The past and presence of gene targeting: from chemicals and DNA via proteins to RNA. *Philos Trans R Soc Lond B Biol Sci* **2018,** 373, (1748).
31. Klug, A., The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* **2010,** 79, 213-31.
32. Qasim, W.; Zhan, H.; Samarasinghe, S.; Adams, S.; Amrolia, P.; Stafford, S.; Butler, K.; Rivat, C.; Wright, G.; Somana, K.; Ghorashian, S.; Pinner, D.; Ahsan, G.; Gilmour, K.; Lucchini, G.; Inglott, S.; Mifsud, W.; Chiesa, R.; Peggs, K. S.; Chan, L.; Farzeneh, F.; Thrasher, A. J.; Vora, A.; Pule, M.; Veys, P., Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci Transl Med* **2017,** 9, (374).
33. Yan, M.; Liu, Q., Differentiation therapy: a promising strategy for cancer treatment. *Chin J Cancer* **2016,** 35, 3.
34. Sun, H.; Tang, H.; Xie, D.; Jia, Z.; Ma, Z.; Wei, D.; Mishra, L.; Gao, Y.; Zheng, S.; Xie, K.; Peng, Z., Kruppel-like Factor 4 Blocks Hepatocellular Carcinoma Dedifferentiation and Progression through Activation of Hepatocyte Nuclear Factor-6. *Clin Cancer Res* **2016,** 22, (2), 502-12.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

35. Cheng, X.; Tan, Z.; Huang, X.; Yuan, Y.; Qin, S.; Gu, Y.; Wang, D.; He, C.; Su, Z., Inhibition of Glioma Development by ASCL1-Mediated Direct Neuronal Reprogramming. *Cells* **2019,** 8, (6).
36. Park, N. I.; Guilhamon, P.; Desai, K.; McAdam, R. F.; Langille, E.; O'Connor, M.; Lan, X.; Whetstone, H.; Coutinho, F. J.; Vanner, R. J.; Ling, E.; Prinos, P.; Lee, L.; Selvadurai, H.; Atwal, G.; Kushida, M.; Clarke, I. D.; Voisin, V.; Cusimano, M. D.; Bernstein, M.; Das, S.; Bader, G.; Arrowsmith, C. H.; Angers, S.; Huang, X.; Lupien, M.; Dirks, P. B., ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* **2017,** 21, (3), 411.
37. Wei, L.; Dai, Y.; Zhou, Y.; He, Z.; Yao, J.; Zhao, L.; Guo, Q.; Yang, L., Oroxylin A activates PKM1/HNF4 alpha to induce hepatoma differentiation and block cancer progression. *Cell Death Dis* **2017,** 8, (7), e2944.
38. Chuang, H. C.; Tsai, C. Y.; Hsueh, C. H.; Tan, T. H., GLK-IKKbeta signaling induces dimerization and translocation of the AhR-RORgammat complex in IL-17A induction and autoimmune disease. *Sci Adv* **2018,** 4, (9), eaat5401.
39. Raikwar, S. P.; Thangavel, R.; Dubova, I.; Ahmed, M. E.; Selvakumar, P. G.; Kempuraj, D.; Zaheer, S.; Iyer, S.; Zaheer, A., Neuro-Immuno-Gene- and Genome-Editing-Therapy for Alzheimer's Disease: Are We There Yet? *J Alzheimers Dis* **2018,** 65, (2), 321-344.
40. Jia, J.; Jin, Y.; Bian, T.; Wu, D.; Yang, L.; Terada, N.; Wu, W.; Jin, S., Bacterial delivery of TALEN proteins for human genome editing. *PLoS One* **2014,** 9, (3), e91547.
41. Yang, S.; Li, S.; Li, X. J., Shortening the Half-Life of Cas9 Maintains Its Gene Editing Ability and Reduces Neuronal Toxicity. *Cell Rep* **2018,** 25, (10), 2653-2659 e3.
42. Zhang, Y.; Roise, J. J.; Lee, K.; Li, J.; Murthy, N., Recent developments in intracellular protein delivery. *Curr Opin Biotechnol* **2018,** 52, 25-31.
43. Xie, S. S.; Qiu, X. Y.; Zhu, L. Y.; Zhu, C. S.; Liu, C. Y.; Wu, X. M.; Zhu, L.; Zhang, D. Y., Assembly of TALE-based DNA scaffold for the enhancement of exogenous multi-enzymatic pathway. *J Biotechnol* **2019,** 296, 69-74.
44. Shin, E.; Kim, W.; Lee, S.; Bae, J.; Kim, S.; Ko, W.; Seo, H. S.; Lim, S.; Lee, H. S.; Jo, K., Truncated TALE-FP as DNA Staining Dye in a High-salt Buffer. *Sci Rep-Uk* **2019,** 9.
45. Cuculis, L.; Zhao, C.; Abil, Z.; Zhao, H.; Shukla, D.; Schroeder, C. M., Divalent cations promote TALE DNA-binding specificity. *Nucleic Acids Res* **2019**.
46. Plys, A. J.; Kingston, R. E., Dynamic condensates activate transcription. *Science* **2018,** 361, (6400), 329-330.
47. Gemayel, R.; Chavali, S.; Pougach, K.; Legendre, M.; Zhu, B.; Boeynaems, S.; van der Zande, E.; Gevaert, K.; Rousseau, F.; Schymkowitz, J.; Babu, M. M.; Verstrepen, K. J., Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol Cell* **2015,** 59, (4), 615-27.
48. Patel, B. K.; Gavin-Smyth, J.; Liebman, S. W., The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. *Nat Cell Biol* **2009,** 11, (3), 344-9.
49. Langdon, E. M.; Qiu, Y.; Ghanbari Niaki, A.; McLaughlin, G. A.; Weidmann, C. A.; Gerbich, T. M.; Smith, J. A.; Crutchley, J. M.; Termini, C. M.; Weeks, K. M.; Myong, S.; Gladfelter, A. S., mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **2018,** 360, (6391), 922-927.
50. Schmidt, S. F.; Larsen, B. D.; Loft, A.; Mandrup, S., Cofactor squelching: Artifact or fact? *Bioessays* **2016,** 38, (7), 618-26.
51. Ottoz, D. S.; Rudolf, F.; Stelling, J., Inducible, tightly regulated and growth condition-independent transcription factor in Saccharomyces cerevisiae. *Nucleic Acids Res* **2014,** 42, (17), e130.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

52. Hsu, C.; Jaquet, V.; Gencoglu, M.; Becskei, A., Protein Dimerization Generates Bistability in Positive Feedback Loops. *Cell Rep* **2016,** 16, (5), 1204-10.
53. Kaganovich, D., There Is an Inclusion for That: Material Properties of Protein Granules Provide a Platform for Building Diverse Cellular Functions. *Trends in Biochemical Sciences* **2017,** 42, (10), 765-776.
54. Gencoglu, M.; Schmidt, A.; Becskei, A., Measurement of In Vivo Protein Binding Affinities in a Signaling Network with Mass Spectrometry. *ACS Synth Biol* **2017,** 6, (7), 1305-1314.
55. Gadhavi, P.; Morgan, P. J.; Alefounder, P.; Harding, S. E., A physico-chemical investigation of the self-association of the DNA binding domain of the yeast transcriptional activator GAL4. *Eur Biophys J* **1996,** 24, (6), 405-12.
56. Jaquet, V.; Hsu, C.; Becskei, A., Measurement of bistability in a multidimensional parameter space. *Integr Biol (Camb)* **2017,** 9, (2), 167-177.
57. Santos-Moreno, J.; Schaerli, Y., Multistable and dynamic CRISPRi-based synthetic circuits in E. coli. In *BioArchive*, 2019.
58. Yin, P.; Deng, D.; Yan, C.; Pan, X.; Xi, J. J.; Yan, N.; Shi, Y., Specific DNA-RNA hybrid recognition by TAL effectors. *Cell Rep* **2012,** 2, (4), 707-13.
59. Meckler, J. F.; Bhakta, M. S.; Kim, M. S.; Ovadia, R.; Habrian, C. H.; Zykovich, A.; Yu, A.; Lockwood, S. H.; Morbitzer, R.; Elsaesser, J.; Lahaye, T.; Segal, D. J.; Baldwin, E. P., Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Research* **2013,** 41, (7), 4118-4128.
60. Geiger-Schuller, K.; Mitra, J.; Ha, T.; Barrick, D., Functional instability allows access to DNA in longer transcription Activator-Like effector (TALE) arrays. *Elife* **2019,** 8.
61. Clauss, K.; Popp, A. P.; Schulze, L.; Hettich, J.; Reisser, M.; Torres, L. E.; Uhlenhaut, N. H.; Gebhardt, J. C. M., DNA residence time is a regulatory factor of transcription repression. *Nucleic Acids Research* **2017,** 45, (19), 11121-11130.
62. Werner, J.; Gossen, M., Modes of TAL effector-mediated repression. *Nucleic Acids Res* **2014,** 42, (21), 13061-73.
63. Rinaldi, F. C.; Doyle, L. A.; Stoddard, B. L.; Bogdanove, A. J., The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Res* **2017,** 45, (11), 6960-6970.
64. Guilinger, J. P.; Pattanayak, V.; Reyon, D.; Tsai, S. Q.; Sander, J. D.; Joung, J. K.; Liu, D. R., Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **2014,** 11, (4), 429-35.
65. Cuculis, L.; Abil, Z.; Zhao, H.; Schroeder, C. M., Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat Commun* **2015,** 6, 7277.
66. Lonzaric, J.; Lebar, T.; Majerle, A.; Mancek-Keber, M.; Jerala, R., Locked and proteolysis-based transcription activator-like effector (TALE) regulation. *Nucleic Acids Res* **2016,** 44, (3), 1471-81.
67. Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014,** 507, (7490), 62-7.
68. Yourik, P.; Fuchs, R. T.; Mabuchi, M.; Curcuru, J. L.; Robb, G. B., Staphylococcus aureus Cas9 is a multiple-turnover enzyme. *RNA* **2019,** 25, (1), 35-44.
69. Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W., DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep* **2018,** 22, (2), 359-371.
70. Ma, H.; Tu, L. C.; Naseri, A.; Huisman, M.; Zhang, S.; Grunwald, D.; Pederson, T., CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol* **2016,** 214, (5), 529-37.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at Molecules 2020, 25, 1902; doi:10.3390/molecules25081902

71. Jones, D. L.; Leroy, P.; Unoson, C.; Fange, D.; Curic, V.; Lawson, M. J.; Elf, J., Kinetics of dCas9 target search in Escherichia coli. *Science* **2017,** 357, (6358), 1420-1424.
72. Kiani, S.; Chavez, A.; Tuttle, M.; Hall, R. N.; Chari, R.; Ter-Ovanesyan, D.; Qian, J.; Pruitt, B. W.; Beal, J.; Vora, S.; Buchthal, J.; Kowal, E. J.; Ebrahimkhani, M. R.; Collins, J. J.; Weiss, R.; Church, G., Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods* **2015,** 12, (11), 1051-4.
73. Rantasalo, A.; Kuivanen, J.; Penttila, M.; Jantti, J.; Mojzita, D., Synthetic Toolkit for Complex Genetic Circuit Engineering in Saccharomyces cerevisiae. *ACS Synth Biol* **2018,** 7, (6), 1573-1587.
74. Labow, M. A.; Baim, S. B.; Shenk, T.; Levine, A. J., Conversion of the Lac Repressor into an Allosterically Regulated Transcriptional Activator for Mammalian-Cells. *Molecular and Cellular Biology* **1990,** 10, (7), 3343-3356.
75. Guan, Z.; Hughes, C. M.; Kosiyatrakul, S.; Norio, P.; Sen, R.; Fiering, S.; Allis, C. D.; Bouhassira, E. E.; Schildkraut, C. L., Decreased replication origin activity in temporal transition regions. *J Cell Biol* **2009,** 187, (5), 623-35.
76. Moore, I.; Galweiler, L.; Grosskopf, D.; Schell, J.; Palme, K., A transcription activation system for regulated gene expression in transgenic plants. *Proc Natl Acad Sci U S A* **1998,** 95, (1), 376-81.
77. Berens, C.; Porschke, D., Recognition of operator DNA by Tet repressor. *J Phys Chem B* **2013,** 117, (6), 1880-5.
78. Baudrimont, A.; Voegeli, S.; Viloria, E. C.; Stritt, F.; Lenon, M.; Wada, T.; Jaquet, V.; Becskei, A., Multiplexed gene control reveals rapid mRNA turnover. *Sci Adv* **2017,** 3, (7), e1700006.
79. Baudrimont, A.; Jaquet, V.; Wallerich, S.; Voegeli, S.; Becskei, A., Contribution of RNA Degradation to Intrinsic and Extrinsic Noise in Gene Expression. *Cell Rep* **2019,** 26, (13), 3752-3761 e5.
80. Hsieh, M.; Brenowitz, M., Comparison of the DNA association kinetics of the Lac repressor tetramer, its dimeric mutant LacIadi, and the native dimeric Gal repressor. *J Biol Chem* **1997,** 272, (35), 22092-6.
81. Bintu, L.; Buchler, N. E.; Garcia, H. G.; Gerland, U.; Hwa, T.; Kondev, J.; Kuhlman, T.; Phillips, R., Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev* **2005,** 15, (2), 125-35.
82. Du, M.; Kodner, S.; Bai, L., Enhancement of LacI binding in vivo. *Nucleic Acids Res* **2019,** 47, (18), 9609-9618.
83. Erdel, F.; Rippe, K., Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* **2018,** 114, (10), 2262-2270.
84. Ali, M. Z.; Choubey, S., Decoding the grammar of transcriptional regulation from RNA polymerase measurements: models and their applications. *Phys Biol* **2019,** 16, (6), 061001.
85. Polstein, L. R.; Perez-Pinera, P.; Kocak, D. D.; Vockley, C. M.; Bledsoe, P.; Song, L.; Safi, A.; Crawford, G. E.; Reddy, T. E.; Gersbach, C. A., Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res* **2015,** 25, (8), 1158-69.
86. Bisaria, N.; Jarmoskaite, I.; Herschlag, D., Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. *Cell Syst* **2017,** 4, (1), 21-29.
87. Miller, J. C.; Patil, D. P.; Xia, D. F.; Paine, C. B.; Fauser, F.; Richards, H. W.; Shivak, D. A.; Bendana, Y. R.; Hinkley, S. J.; Scarlott, N. A.; Lam, S. C.; Reik, A.; Zhou, Y.; Paschon, D. E.; Li, P.; Wangzor, T.; Lee, G.; Zhang, L.; Rebar, E. J., Enhancing gene editing specificity by attenuating DNA cleavage kinetics. *Nat Biotechnol* **2019,** 37, (8), 945-952.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 February 2020 doi:10.20944/preprints202002.0458.v1

Peer-reviewed version available at *Molecules* **2020**, *25*, 1902; doi:10.3390/molecules25081902

88. Aschenbrenner, S.; Kallenberger, S. M.; Hoffmann, M. D.; Huck, A.; Eils, R.; Niopek, D., Coupling Cas9 to artificial inhibitory domains enhances CRISPR-Cas9 target specificity. *Sci Adv* **2020,** 6, (6), eaay0187.
89. Durairaj, G.; Kaiser, P., The 26S proteasome and initiation of gene transcription. *Biomolecules* **2014,** 4, (3), 827-47.
90. Lickwar, C. R.; Mueller, F.; Hanlon, S. E.; McNally, J. G.; Lieb, J. D., Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **2012,** 484, (7393), 251-5.
91. Culyba, M. J.; Kubiak, J. M.; Mo, C. Y.; Goulian, M.; Kohli, R. M., Non-equilibrium repressor binding kinetics link DNA damage dose to transcriptional timing within the SOS gene network. *PLoS Genet* **2018,** 14, (6), e1007405.
92. Wada, T.; Wallerich, S.; Becskei, A., Stochastic Gene Choice during Cellular Differentiation. *Cell Rep* **2018,** 24, (13), 3503-3512.
93. Wada, T.; Wallerich, S.; Becskei, A., Synthetic Transcription Factors Switch from Local to Long-Range Control during Cell Differentiation. *ACS Synth Biol* **2019,** 8, (2), 223-231.
94. Baumann, V.; Wiesbeck, M.; Breunig, C. T.; Braun, J. M.; Koferle, A.; Ninkovic, J.; Gotz, M.; Stricker, S. H., Targeted removal of epigenetic barriers during transcriptional reprogramming. *Nat Commun* **2019,** 10, (1), 2119.
95. Therizols, P.; Illingworth, R. S.; Courilleau, C.; Boyle, S.; Wood, A. J.; Bickmore, W. A., Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **2014,** 346, (6214), 1238-1242.
96. Smith, A. J.; Thomas, F.; Shoemark, D.; Woolfson, D. N.; Savery, N. J., Guiding Biomolecular Interactions in Cells Using de Novo Protein-Protein Interfaces. *Acs Synthetic Biology* **2019,** 8, (6), 1284-1293.
97. Becker, N. A.; Schwab, T. L.; Clark, K. J.; Maher, L. J., 3rd, Bacterial gene control by DNA looping using engineered dimeric transcription activator like effector (TALE) proteins. *Nucleic Acids Res* **2018,** 46, (5), 2690-2696.
98. Chen, C.; Zheng, R.; Fan, J.; Brown, M.; Liu, J. S.; Meyer, C. A.; Shirley Liu, X., Determinants of transcription factor regulatory range. *BioArchive* **2019**.
99. Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A., A Phase Separation Model for Transcriptional Control. *Cell* **2017,** 169, (1), 13-23.