ANNUAL REVIEWS Further
Click here to view this article's online features:
- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

A Single-Molecule View of Genome Editing Proteins: Biophysical Mechanisms for TALEs and CRISPR/Cas9

Luke Cuculis[1] and Charles M. Schroeder[1,2]

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801; email: cms@illinois.edu

[2]Department of Chemical & Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Annu. Rev. Chem. Biomol. Eng. 2017. 8:577–97

First published online as a Review in Advance on May 10, 2017

The *Annual Review of Chemical and Biomolecular Engineering* is online at chembioeng.annualreviews.org

https://doi.org/10.1146/annurev-chembioeng-060816-101603

Copyright © 2017 by Annual Reviews.
All rights reserved

Keywords

genome editing, TALEN, CRISPR/Cas9, single molecule, DNA search

Abstract

Exciting new advances in genome engineering have unlocked the potential to radically alter the treatment of human disease. In this review, we discuss the application of single-molecule techniques to uncover the mechanisms behind two premier classes of genome editing proteins: transcription activator-like effector nucleases (TALENs) and the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated system (Cas). These technologies have facilitated a striking number of gene editing applications in a variety of organisms; however, we are only beginning to understand the molecular mechanisms governing the DNA editing properties of these systems. Here, we discuss the DNA search and recognition process for TALEs and Cas9 that have been revealed by recent single-molecule experiments.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

INTRODUCTION

ZFN: zinc finger nuclease

TALE(N)s: transcription activator-like effector (nuclease)s

CRISPR/Cas: clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system

AFM: atomic force microscopy

In recent years, new techniques in genome engineering have inspired potentially revolutionary treatments for previously incurable diseases (1). At present, several therapeutic approaches are trickling into international markets (2), with many more advancing at various stages of clinical trials (3, 4). In the field of genome engineering, three major biotechnologies have emerged as efficient molecular tools for the precise editing of DNA in the context of large genomes (1): zinc finger nucleases (ZFNs) (5–8), transcription activator-like effector (nuclease)s [TALE(N)s] (9–11), and the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system (12, 13). These systems function as programmable nucleases that provide exquisite control over the editing of genetic information in live cells. In all cases, a nuclease is programmed to induce a double strand break at a desired position in the genome, after which the break is repaired via nonhomologous end joining or homology-directed repair, depending on the presence of exogenous template DNA and the cell cycle phase, among other factors (14). In this review, we discuss how single-molecule biophysical techniques have been utilized to reveal the fundamental processes behind the DNA binding and search behavior of TALE proteins and the CRISPR/Cas system.

ZFNs were the first of the three main classes of programmable nucleases to emerge as viable tools for genome engineering. Despite their use in gene editing applications (15, 16), ZFNs have not been extensively studied via single-molecule techniques and thus are not discussed further in this review. TALENs (**Figure 1*a***) consist of fusions between programmable DNA binding domains (transcription activator-like effectors, or TALEs) and nuclease domains (typically FokI). The TALE binding domain is characterized by an array of tandem repeats that bind DNA in a one-repeat-to-one-base-pair manner (17). TALEs offer an exceptional level of customization to essentially any target DNA sequence of interest, provided that a single thymidine is located at the 5′ terminus of the TALE binding site (18). The CRISPR/Cas9 system (**Figure 1*b***) consists of a combined protein/RNA element with sequence specificity imparted by a guide RNA, which is distinct from both TALENs and ZFNs (13). Considering that DNA sequence recognition is imparted by a short RNA rather than a custom protein, the CRISPR/Cas9 system is comparatively straightforward to implement for gene editing applications, which has resulted in explosive popularity for this method (19, 20). Despite major progress made in harnessing these systems for gene editing applications, however, there remains a strong need for improving specificity in DNA recognition and editing (21). To enable precise and accurate editing of large, intact genomes, a comprehensive understanding of the molecular details of DNA binding and search dynamics is needed for each system (22). Single-molecule techniques provide a direct window into observing the DNA binding and search mechanisms for these biomolecules.

Single-Molecule Techniques

Traditional or bulk biochemical methods report on ensemble average properties, such as an average enzymatic rate or average molecular conformation. However, single-molecule techniques can be used to study heterogeneity, distributions in molecular behavior, and real-time dynamics of single proteins and biomolecules. Single-molecule methods allow for the ability to follow individual chemical or biochemical reactions, thereby probing the stochastic nature of chemical processes in physiologically relevant conditions in real time. Broadly speaking, single-molecule techniques can be divided into two main categories: imaging-based approaches and force-based approaches. Within this classification, imaging-based methods are typically carried out via fluorescence microscopy, whereas force-based approaches generally rely on optical tweezers, magnetic tweezers, or atomic force microscopy (AFM) for precise mechanical control of single biomolecules.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

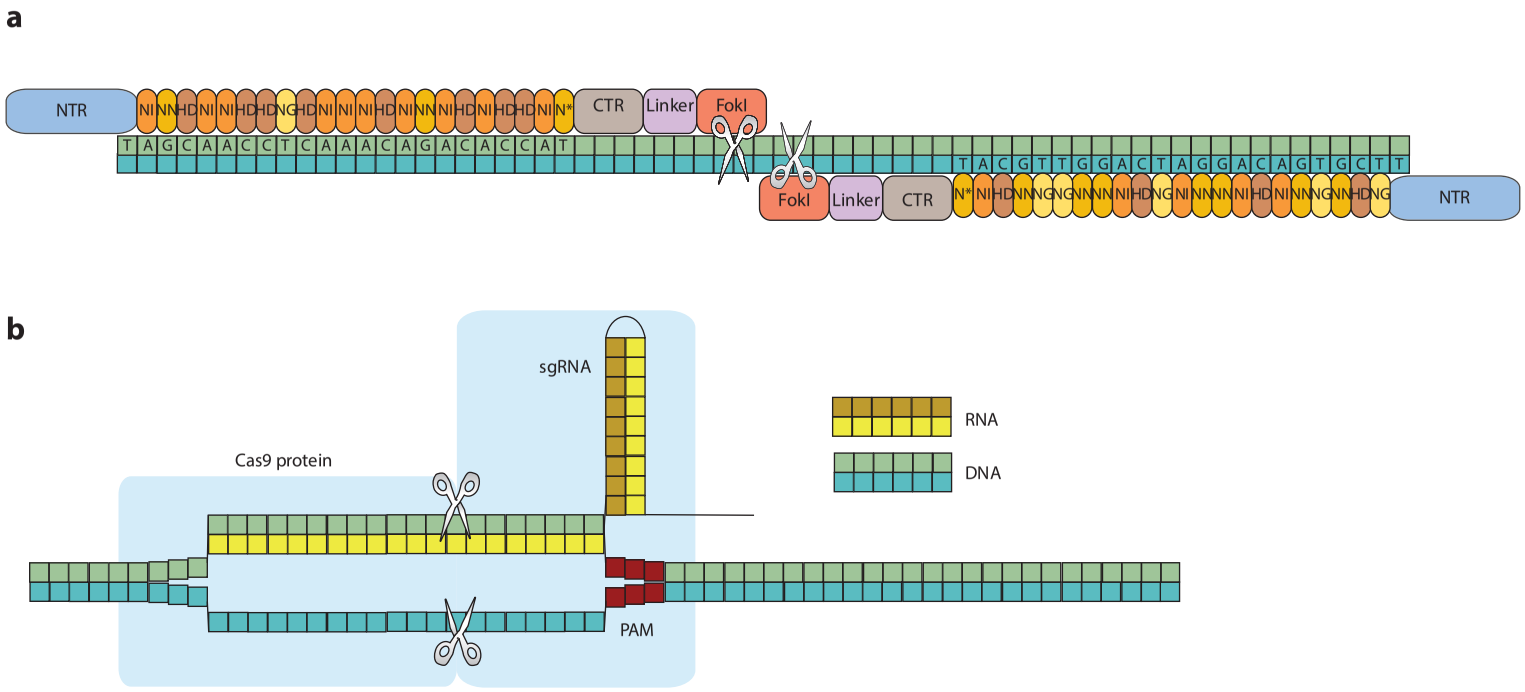


Figure 1

Overview of TALEN and CRISPR/Cas genome editing systems. (*a*) The TALEN system is composed of a TALE protein fused to the dimeric nuclease FokI. Here, the repeat region of the TALE is colored orange, and each repeat recognizes a single nucleotide. Only the repeat variable diresidues of the TALE are explicitly shown. (*b*) The CRISPR/Cas9 system consists of the Cas9 nuclease (*light blue*) and the sgRNA (*yellow/brown*), which acts to target the Cas9 nuclease to the DNA sequence of choice. Abbreviations: Cas, CRISPR-associated protein; CRISPR, clustered regularly interspaced palindromic repeat; CTR, C-terminal region; NTR, N-terminal region; PAM, protospacer adjacent motif; sgRNA, single guide RNA; TALEN, transcription activator-like effector nuclease.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

SMFM: single-molecule fluorescence microscopy

smFRET: single-molecule Förster resonance energy transfer

Single-molecule fluorescence microscopy (SMFM) can be used to detect single fluorescent probes (small-molecule organic dyes or genetically encoded fluorescent proteins) linked to biomolecules such as proteins or nucleic acids. By using this approach, single biomolecules can be localized with spatial precisions less than the diffraction limit of visible light (~300 nm) by collecting a sufficient number of photons from each emitter (23). From this perspective, molecular events can be directly observed at relevant length scales for proteins, DNA, or RNA via detection and localization of single biomolecules with spatial resolutions of 25–100 nm. In this way, SMFM has facilitated significant advances in our understanding of complex biological systems, including the hand-over-hand movement of myosin V (24) and kinesin (25) and the residence times for *lac* operator binding in vivo (26). In addition to methods based on single-molecule tracking, single-molecule Förster resonance energy transfer (smFRET) allows for direct observation of small-scale (1–10-nm) changes in intermolecular or intramolecular distances (27). In this way, smFRET can be considered as a powerful molecular ruler, and this method has been used to study myriad biological processes, including protein folding (28) and the repetitive shuttling mechanism of Rep helicase (29).

Single-molecule force-based measurements offer the ability to apply and/or measure subpiconewton to nanonewton forces, which are relevant for the study and manipulation of individual biomolecules (30). Optical tweezers and magnetic tweezers are based on the principle of particle trapping using an external force field. In the case of optical tweezers, a high-powered laser beam is focused through a high–numerical aperture objective lens, generating a tight focus and large optical gradient (31). Dielectric particles near the focus of the beam experience a restoring force toward the trap center, thereby enabling stable trapping. Magnetic tweezers function by applying an external magnetic field to manipulate ferromagnetic, paramagnetic, or superparamagnetic beads. Magnetic tweezers can also be used to generate torque on trapped particles by rotation of two permanent magnets, which is useful for inducing supercoiling in DNA (32). In all cases, these techniques require that the biomolecule of interest (nucleic acid or protein) be conjugated to a dielectric or magnetic bead (33), which is commonly a surface-functionalized micrometer-sized colloidal particle. Additional trapping techniques provide alternative methods for single molecule/particle manipulation, including the Stokes trap (34) and the electrokinetic trap (35).

AFM provides a method for force-based measurements or manipulation of single biomolecules as well as single-molecule imaging (36). AFM is a scanning probe method based on measuring the deflection of a microfabricated cantilever affixed with a sharp probe tip. Dragging or tapping the tip along biological samples enables acquisition of images with nanometer-level resolution. Moreover, AFM can be used to determine force-extension relations for individual DNA molecules or proteins (37).

Single-molecule techniques provide a powerful toolbox to investigate gene editing systems, with an overall goal of determining real-time dynamics, biophysical mechanisms, and interactions with DNA templates. In this review, we discuss recent applications of single-molecule imaging and force-based methods to study TALE proteins and the CRISPR/Cas9 system. In particular, we review recent experiments that have revealed the molecular mechanisms governing target site search and recognition on DNA and manipulation of DNA templates carried out by gene editing systems.

SINGLE-MOLECULE STUDIES OF TALE PROTEINS

DBP: DNA-binding protein

Discovery and Rapid Emergence of TALEs for Gene Editing

TALEs are a class of DNA-binding proteins (DBPs) naturally secreted by *Xanthomonas* bacteria to aid in their infection of plant cells. Early genetics studies identified a high number of repetitive

elements corresponding to repeats of 34 or 35 amino acids in the *avrBs3* gene, known to belong to a genome for an organism that causes deleterious infections in pepper plants (38). Shortly after the identification of the *avrBs3* repeats, several homologs in rice infections (39, 40), citrus cankers (41), and tomato blight (42) were identified. In vitro binding studies demonstrated that *avrBs3* strongly interacted with DNA, confirming that these proteins, with their nuclear localization signals, were modulators of transcription. Nevertheless, the repeat structures found in these proteins appeared to be unique among the broad class of previously identified site-specific DBPs.

RVDs: repeat variable diresidues

A breakthrough in understanding TALE-DNA interactions was made with the realization that the size of the UPA box (the region in the pepper genome upregulated by *avrBs3*) was roughly the same size as the number of repeat domains in *avrBs3*, which led to a hypothesis of a one-repeat-to-one-base-pair binding mode for TALEs. Moreover, each of the 34 to 35 amino acid repeats was highly conserved, albeit with differences in residue identities at positions 12 and 13 known as repeat variable diresidues (RVDs). Further work showed that the amino acid identities in the RVDs determined the DNA binding sequence, which led researchers to elucidate the binding code for TALEs (43). Following this advance (43, 44), researchers realized that nuclease domains could be fused to TALE proteins to generate site-specific gene editing tools (45), which spurred a rapid increase in applications beginning in 2011. TALENs were quickly adopted for genome editing in rice (46), zebrafish (11), mice (47), and human stem cells (9). Moreover, the commercial gene editing company Cellectis, along with Life Technologies, acquired the commercial rights to TALENs around this time (1) and began building a gene therapy method based on TALEN-modified off-the-shelf T cells. At the 2015 meeting of the American Society of Hematology, Cellectis announced clinical trial results for the first patient treated via its engineered CAR-T therapy for the treatment of leukemia.

Reconciling Dynamic DNA Search with a Superhelical Protein Structure

Despite the pervasive use of TALENs for gene editing applications, the DNA sequence search mechanism of these proteins was not well understood. The molecular structure of TALE proteins was determined by two independent labs in 2012 (48, 49). TALEs adopt a superhelical structure with no close structural analog; the mitochondrial transcription factors mTERF are perhaps the closest analogs in terms of structure (50). Indeed, structural studies confirmed earlier genomics work suggesting that TALE structure was highly unique among the broad class of DBPs (44). From a broad perspective, development of precise and efficient gene editing systems for human therapies based on TALE proteins requires a molecular-level understanding of the DNA search-and-recognition process for these proteins.

Most existing models for the sequence search mechanism of DBPs are based primarily on proteins that do not encircle (or only partially encircle) the DNA double helix during nonspecific search. However, these models are not immediately applicable to TALE proteins, given the distinct superhelical structure of TALEs in the specifically bound state. A crystal structure of a TALE protein in the absence of DNA showed an extended helical structure, out of phase with the normal pitch of natural B-form DNA (49). Interestingly, cocrystal structures of TALE proteins bound to target DNA showed a compressed helical structure that tightly tracks the major groove of DNA, without apparent distortion of the DNA structure (49). Recent molecular dynamics simulations revealed an inherent plasticity in TALE structure, providing evidence that the DNA-free and DNA-bound TALE structures could interconvert in a fluid helical extension/compression model (51). Despite these insights, however, the structure of TALE proteins bound to nonspecific DNA was unknown. From this perspective, single-molecule techniques can play a key role in revealing

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

the dynamic sequence search mechanism of TALEs along nonspecific DNA, which would uncover additional information beyond static crystal structures.

NTR: N-terminal region of TALE protein

CRD: central repeat domain of TALE protein

CTR: C-terminal region of TALE protein

Previous studies of TALE structure and binding suggested that TALEs might follow a 1D sliding mechanism during search, closely tracking the DNA backbone, given their superhelical structure (17, 52, 53). However, if TALEs adopted a tight superhelical conformation during nonspecific search (similar to that in the specifically bound state), then the energy landscape between random and target sequences might be too similar to allow for rapid search along a DNA template. The search speed-stability paradox (54) refers to the fact that stable target recognition (in the specifically bound state) and rapid search (in the nonspecifically bound state) cannot be reconciled, considering that natural DNA does not have sufficient chemical diversity to allow for these behaviors to realistically occur. A recent computational study on the energetics of TALE target binding revealed that TALE specificity arises primarily from negative discrimination by the RVDs, which supports this hypothesis (55). Indeed, the majority of TALE binding energy is conferred by nonspecific electrostatic interactions with DNA. Therefore, a TALE search process in which protein conformation was similar between search and bind modes would be prohibitively slow.

Herein lies a major perplexing question regarding the fundamental search mechanism of TALE proteins. On one hand, TALEs recognize target sites via RVDs by negative discrimination, which requires that the RVDs sample local DNA sequence. On the other hand, a dominant amount of binding energy conferred via nonspecific electrostatic interactions suggests that a TALE protein tightly associated with a DNA template would result in an incredibly slow search process. How do TALE proteins move along DNA in search of their target sites? This question can be addressed by using single-molecule techniques to study the DNA search process for TALEs.

Uncovering the DNA Search Mechanism for TALEs: Single-Molecule Tracking

Cuculis et al. (56) recently used SMFM to study the search process for TALEs along nonspecific DNA (**Figure 2**). In this work, TALE proteins were specifically labeled with a single organic dye by using a genetically encoded aldehyde tag, thereby enabling nonperturbative and stoichiometric fluorescent labeling TALEs for single-molecule imaging. Using this approach, the nonspecific search dynamics of single TALE proteins could be tracked on long DNA molecules. For these experiments, custom DNA templates (45 kbp) were stretched and end-tethered at both termini to a functionalized glass coverslip via biotin-NeutrAvidin linkages, thereby enabling direct visualization of TALE-DNA interactions (**Figure 2*a*,*b***). TALEs were observed to readily diffuse along extended DNA templates in a 1D manner. TALE diffusion was rapid and persisted for long times (on the order of several seconds) before unbinding (**Figure 2*c***). These experiments were the first to show that TALEs are readily capable of traversing DNA templates using a facilitated 1D search mechanism. Interestingly, TALE diffusion trajectories showed periods of rapid translocation interspersed with stagnant or attenuated diffusion (**Figure 2*d***). Interspersed periods of slow/fast diffusion were not expected a priori given the lack of discrete subunits in TALE proteins. Indeed, other DBPs with multiple subunits have displayed evidence of multimode search mechanisms, wherein different subunits within a protein are capable of executing a change of function during search (57, 58).

Although TALE proteins lack discrete subunits, their helical structure can be divided into three main subdomains: an N-terminal region (NTR) containing nonspecific repeats and a secretion (T3S) signal, the central repeat domain (CRD) containing tandem arrays of canonical repeats specifying target site binding, and a C-terminal region (CTR) containing the nuclear localization signal and an acidic activation domain. Gao et al. (52) recently provided a partial structure for the

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

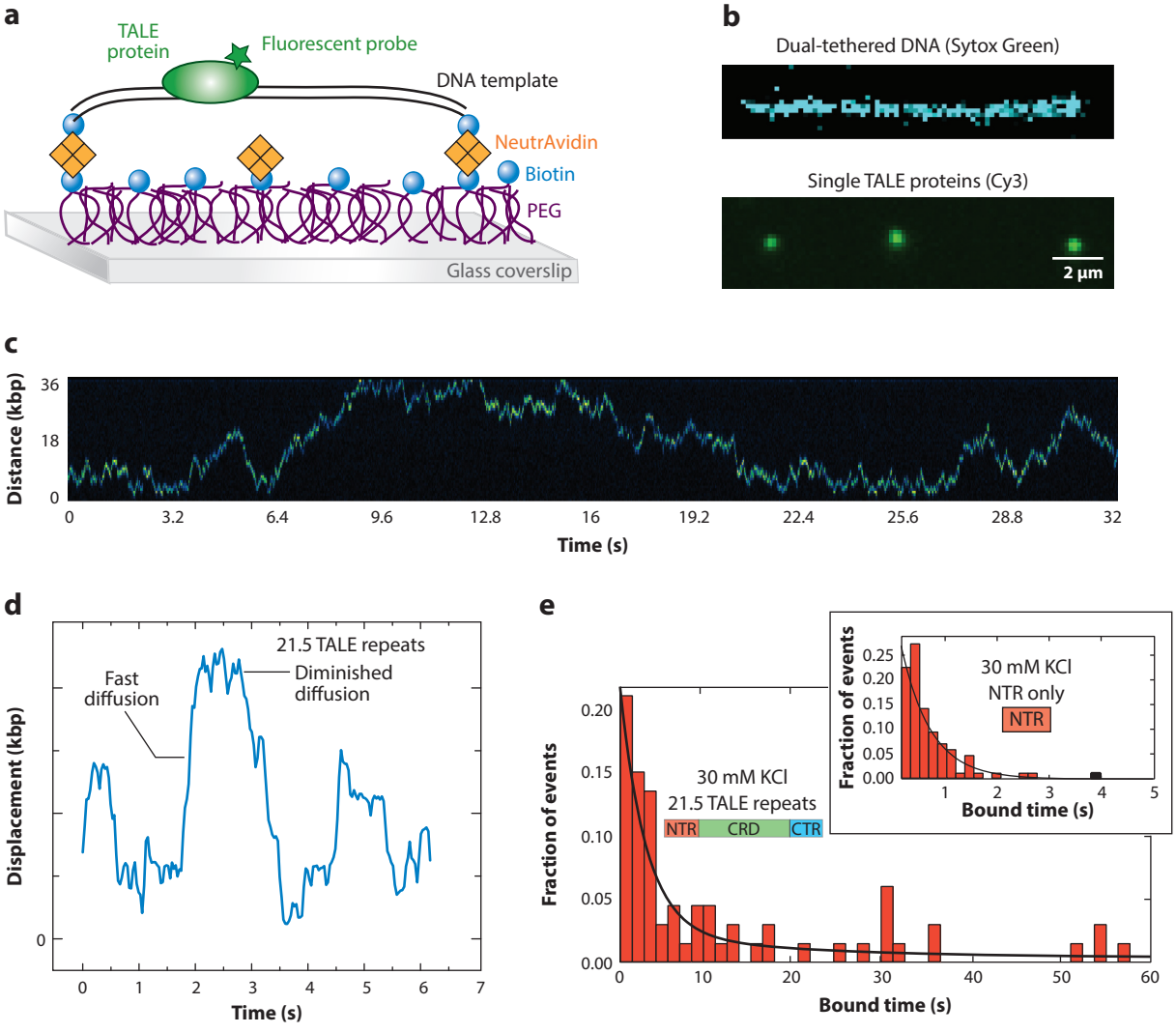


Figure 2

Single-molecule fluorescence microscopy experiments on TALE protein search dynamics (56). (*a*) Schematic of single-molecule assay showing dual-tethered DNA templates and fluorescently labeled TALE constructs. (*b*) Single-molecule images of nonspecific DNA templates (postlabeled with fluorescent dye following TALE diffusion experiments) and single TALE proteins bound to a DNA template. (*c*) Kymograph showing that TALEs readily diffuse along DNA templates (shown here at 500 mM KCl). (*d*) Individual molecular trajectory of TALE search along DNA at physiological ionic strength. Results show heterogeneous diffusive behavior, with periods of rapid diffusion interspersed with attenuated diffusion, suggesting a two-state mechanism for DNA search. (*e*) Histogram of binding lifetimes for full-length TALE proteins shows a multiexponential distribution, which suggests a two-state mechanism for binding. (*inset*) Histogram of binding lifetimes for NTR-only TALE truncation mutants, which shows a single exponential distribution of binding lifetimes. Adapted from References 56 and 61 with permission from Nature Publishing Group. Abbreviations: CRD, central repeat domain; CTR, C-terminal region; PEG, polyethylene glycol; NTR, N-terminal region; TALE, transcription activator-like effector.

NTR, which revealed that the NTR is structurally similar to the CRD. Key differences between the NTR and CRD include the absence of sequence-specific protein-DNA contacts within the NTR and an increase in the number of positively charged amino acids in the NTR, particularly at the far N terminus of the first nonspecific repeat. This study also confirmed earlier hypotheses that the NTR is necessary for TALE binding (59), as TALE mutants lacking the NTR displayed no binding via an isothermal titration calorimetry assay. However, the NTR (alone) was observed to bind DNA in a nonspecific manner, albeit with a lower affinity than a full-length TALE protein (NTR+CRD+CTR). Overall, these findings led researchers to propose that the NTR may be responsible for engaging the DNA and nucleating TALE binding, as well as possibly mediating the nonspecific search of TALEs along DNA (52).

Using SMFM, Cuculis et al. (56) observed the dynamics of TALE NTR-only truncation mutants lacking the CRD+CTR. Interestingly, their results showed that the NTR-only mutants are capable of rapid 1D diffusion along nonspecific DNA. Diffusion trajectories for NTR-only mutants were significantly more rapid compared with those for full-length TALEs at the same ionic strength. Importantly, NTR-only mutants lacked the characteristic periods of slow/fast heterogeneity observed in trajectories for full-length TALEs. In addition, full-length TALEs exhibited multimodal binding lifetimes along nonspecific DNA templates, whereas NTR-only mutants showed single-mode binding (**Figure 2*e***). Moreover, TALE mutants lacking the NTR did not bind to DNA. Taken together, these results support a model for TALE binding in which the NTR nucleates initial binding and mediates short, rapid nonspecific search. Following DNA binding by the NTR, the CRD is then able to engage in local sequence checking events interspersed with periods of rapid nonspecific search along DNA. This model was further supported by the observation that the 1D diffusion coefficient for TALE proteins decreases upon increasing the size of the CRD.

Although these experiments showed that TALE proteins exhibit rapid 1D search on DNA, the precise nature of the search mechanism was unclear. Seminal work by Winter et al. (60) laid out a framework for interpreting the salt dependence of protein diffusion on DNA, broadly classifying DBP search as sliding or hopping along DNA templates. A DBP with a salt-dependent diffusion coefficient is characterized as engaging, at least partially, in hopping events involving dissociation/reassociation of the protein during search along DNA. Conversely, a DBP with a salt-independent diffusion coefficient is characterized as searching DNA via a pure sliding mechanism, such that the DBP could track the major groove of DNA during diffusional search.

To further probe the DNA search mechanism of TALE proteins, Cuculis et al. (61) presented a second single-molecule study on TALE search dynamics. Here, TALE diffusion along DNA was studied while varying the ionic strength of the solution over more than an order of magnitude, which revealed a strong salt dependence on TALE 1D diffusion coefficients (**Figure 3*a–c***). These results suggest that TALEs search DNA by hopping behavior in the classical picture by Winter et al. (60), which may involve short dissociation/reassociation events. However, considering the crystal structure of a TALE protein specifically bound to DNA in a tight superhelical conformation, this result was unexpected.

Cuculis et al. (61) further quantified the magnitude of the ionic strength dependence for a series of TALE proteins, including an NTR-only mutant and full-length TALEs with 11.5, 15.5, and 21.5 repeats in the CRD. The number of repeats in the CRD (and hence the overall size of the TALE construct) was found to be directly related to the salt sensitivity of the 1D diffusion coefficient (**Figure 3*b,c***). These results suggested that TALE CRDs engage DNA during search, which could be accomplished if TALEs are fully wrapped around the DNA helix during nonspecific search. Interestingly, however, the NTR-only mutant showed no apparent

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

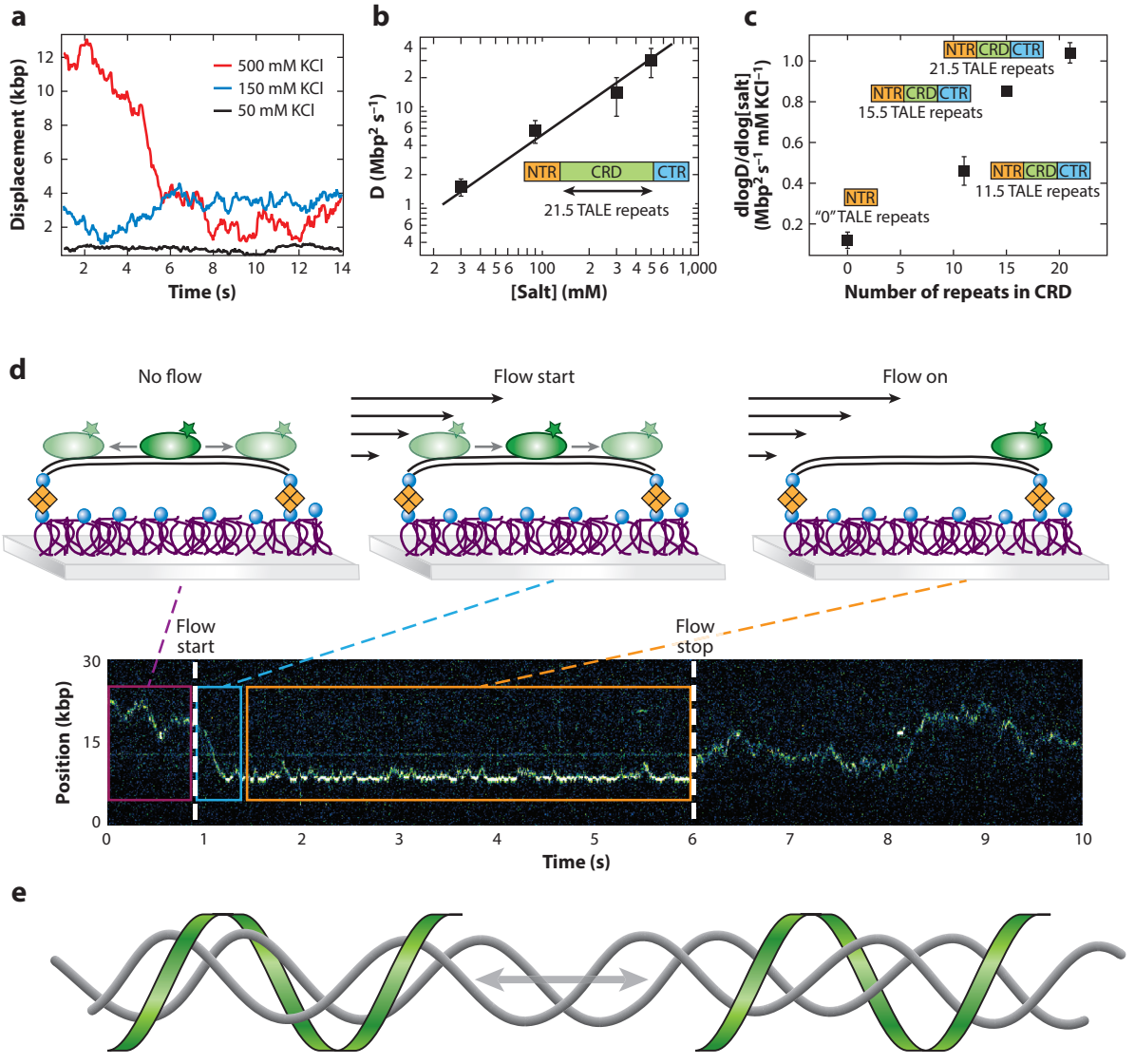


Figure 3

Single-molecule studies of TALE search along nonspecific DNA. (*a*) Individual molecular trajectories of TALE search along DNA at different ionic strengths. (*b*) One-dimensional diffusion coefficient for a TALE protein with 21.5 repeats in the CRD as a function of salt. (*c*) Salt sensitivity of the 1D diffusion coefficient for TALE proteins with different-sized CRDs. TALE diffusion was found to be strongly salt dependent, in a manner that was directly related to the length of the TALE CRD. (*d*) Single-molecule hydrodynamic flow assay showing that TALE diffusion could be strongly biased in the direction of an applied flow. (*e*) Schematic of proposed search mechanism for TALE proteins, wherein TALEs encircle DNA templates but remain loosely associated with the DNA backbone, which enables a rapid, rotationally decoupled search mechanism. Adapted from Reference 61 with permission from Nature Publishing Group. Abbreviations: CRD, central repeat domain; CTR, C-terminal region; NTR, N-terminal region; TALE, transcription activator-like effector.

ionic strength-dependent 1D diffusion. NTR-only diffusion events were observed to be short in duration and lacked the characteristic heterogeneity of full-length TALEs. These mechanistic insights suggested that the NTR is responsible for engaging in short, 1D sliding events that can transition into longer-duration nonspecific searching events for full-length TALEs.

crRNA: short CRISPR RNA

tracrRNA: *trans*-activating crRNA

sgRNA: single guide RNA

PAM: protospacer adjacent motif

These results suggest that TALE proteins undergo a hopping mechanism during nonspecific DNA search, yet the residues in the TALE CRD are engaged with the DNA backbone during the search process. These two observations present an apparent conundrum that was further reconciled by single-molecule experiments. Cuculis et al. (61) performed a series of additional experiments in which a hydrodynamic flow is applied in the direction parallel to DNA backbones during TALE search. In this way, a hydrodynamic force is applied to TALE proteins undergoing diffusion along DNA, thereby biasing TALE diffusion in the direction of fluid flow. It was observed that the diffusional bias is a function of the ionic strength of the imaging buffer. At supraphysiological ionic strengths, single TALE proteins can be pushed distances of ~10+ kbp along DNA templates, effectively pinning single TALEs at the DNA terminus near the chemical linkages to coverslip surfaces. During this process, TALEs remain bound to DNA templates despite the external flow force. These data suggest that TALEs adopt a loosely wrapped helical conformation that encircles DNA templates during nonspecific search (**Figure 3*d***). A third set of experiments examined the size dependence of TALE diffusion, which revealed that TALEs follow a rotationally decoupled trajectory during nonspecific search. Strikingly, this mechanism of nonspecific search is in stark contrast to the majority of other DBPs that use a 1D search mechanism and follow a pure rotationally coupled trajectory during diffusion along DNA.

Taken together, these findings provided support for an apparently distinct mechanism of nonspecific search, wherein TALE proteins are fully wrapped around the DNA backbone yet loosely associated with DNA template during search (**Figure 3*e***). TALEs primarily diffuse along DNA in a rotationally decoupled trajectory, essentially traversing DNA like a washer along a screw, as opposed to a nut along a screw. In this way, SMFM enabled a new picture of TALE sequence search along DNA templates.

SINGLE-MOLECULE STUDIES OF THE CRISPR/CAS9 SYSTEM

The CRISPR/Cas system is naturally found in prokaryotes, where it functions as an adaptable immune system capable of identifying and disabling invasive DNA (62). In this system, exogenous DNA is recognized and incorporated into the CRISPR loci by Cas proteins. Short CRISPR RNA molecules (crRNAs) are then generated, which allows the Cas nuclease to specifically identify and target invasive DNA harboring previously incorporated sequences. Moreover, RNA-DNA complementarity in DNA recognition ensures that the Cas nuclease is able to lock on to its target and proceed with nicking or progressive degradation (13).

Cas9 is a type II CRISPR-Cas protein that lies at the heart of CRISPR-mediated gene engineering applications. Cas9 is the only nuclease required for complete immunity-generation function and can act without additional protein partners, and this feature differentiates type II CRISPR systems from the type I and III systems that require an additional endoribonuclease (63). Early studies exploring the repurposing of Cas9 for creating targeted double stranded breaks in DNA identified the naturally defined three-part system: Cas9 protein, the target-defining crRNA, and the *trans*-activating crRNA (tracrRNA). These studies led to the design of a combined crRNA:tracrRNA (termed sgRNA), reducing the Cas9 programmable nuclease to just two components. A critical requirement for Cas9 targeting is the three-nucleotide protospacer adjacent motif (PAM) occurring downstream on the opposite strand from the DNA target, which places some limitations on genomic locations suitable for Cas9-mediated editing. Similar to TALEs, there have been reports

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

of Cas9 off-target binding (64), which underscores the need for a complete understanding of Cas9 binding and the factors influencing specificity.

DNA Search Mechanism for CRISPR/Cas9: In Vitro Single-Molecule Fluorescence Techniques

TALEs undergo rapid 1D diffusion along DNA during target site search. How does Cas9 locate target sites? The sequence search mechanism for Cas9 was recently investigated using SMFM, providing an intriguing answer to this question. Sternberg et al. (65) used arrays of long DNA molecules aligned side by side (known as DNA curtains) to directly visualize the dynamics of Cas9 interacting with DNA in search of target sites (**Figure 4*a–c***). This approach allows for multiplexed visualization of Cas9/DNA interactions in the presence or absence of applied flow, which facilitates data collection on several DNA templates in a single experiment (**Figure 4*c***). Using a nonperturbative, genetically encoded FLAG tag and anti-FLAG conjugated quantum dots, these researchers observed Cas9 binding to bacteriophage lambda DNA templates (**Figure 4*b,c***). Strikingly, Cas9 does not engage in any measurable 1D diffusion along DNA templates via sliding or apparent hopping. Instead, Cas9 randomly samples DNA sequences using only 3D collisions. This behavior is distinctly different than the DNA search mechanism for TALE proteins, which move along DNA using facilitated 1D diffusion.

Singh et al. (66) further studied Cas9 binding events using an smFRET assay with fluorescently labeled sgRNA and DNA templates (**Figure 5**). Here, Cas9 binding appeared to be a one-step event at target sites, at least within the 100-ms sampling rate for data acquisition. In addition, Cas9 binding events were observed to occur at nontarget sites characterized by lower FRET values. Indeed, Cas9 binding events revealed by intermediate FRET values suggest that Cas9 might be capable of local, rapid diffusion over a few base pairs during binding, and this behavior was termed as the sampling mode for Cas9 binding. Nevertheless, putative local diffusion by Cas9 would occur over significantly shorter distances compared with the long-range DNA diffusion events observed for TALEs.

Interestingly, neither the kinetics of Cas9 binding to target sites (long-lived) nor the kinetics of Cas9 binding to nontarget sequences (short-lived) were affected by solution ionic strength (65). Several DBPs (including TALEs) strongly depend on nonspecific electrostatic interactions with the DNA backbone to nucleate binding; however, Cas9-sgRNA interactions with DNA appear to be far less dependent on electrostatic effects. From this view, Cas9-sgRNA binding is thought to be highly sequence specific, driven primarily by the presence or absence of the PAM and the sequence complementarity proximal to the PAM region. Indeed, single-molecule studies have revealed that Cas9 preferentially samples PAM-enriched sequences (65, 66). This observation can be further explained by the directionality of the RNA-DNA heteroduplex that forms with the PAM sequence, which is further discussed below.

A recent crystal structure of Cas9-sgRNA bound to its DNA target revealed the molecular details of the PAM binding site, which is composed primarily of two conserved arginine residues that interact with the GG component of the PAM (67). Taken together with studies that underscore the necessity of the PAM for Cas9 activity, these findings suggest that Cas9 searches DNA rapidly via 3D diffusion. During search, Cas9 exhibits an increased dwell time near PAM regions owing to favorable binding interactions between the PAM and arginine residues, together with local sequence melting as the DNA-RNA heteroduplex initially begins to form. At PAM sites, Cas9 can either engage the local DNA sequence further via local sequence melting or unbind relatively quickly (on timescales of milliseconds or faster) without spending excess time at nontarget sites. In this way, Cas9 is able to strongly bias the tedious interrogation of local sequences to those sites

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

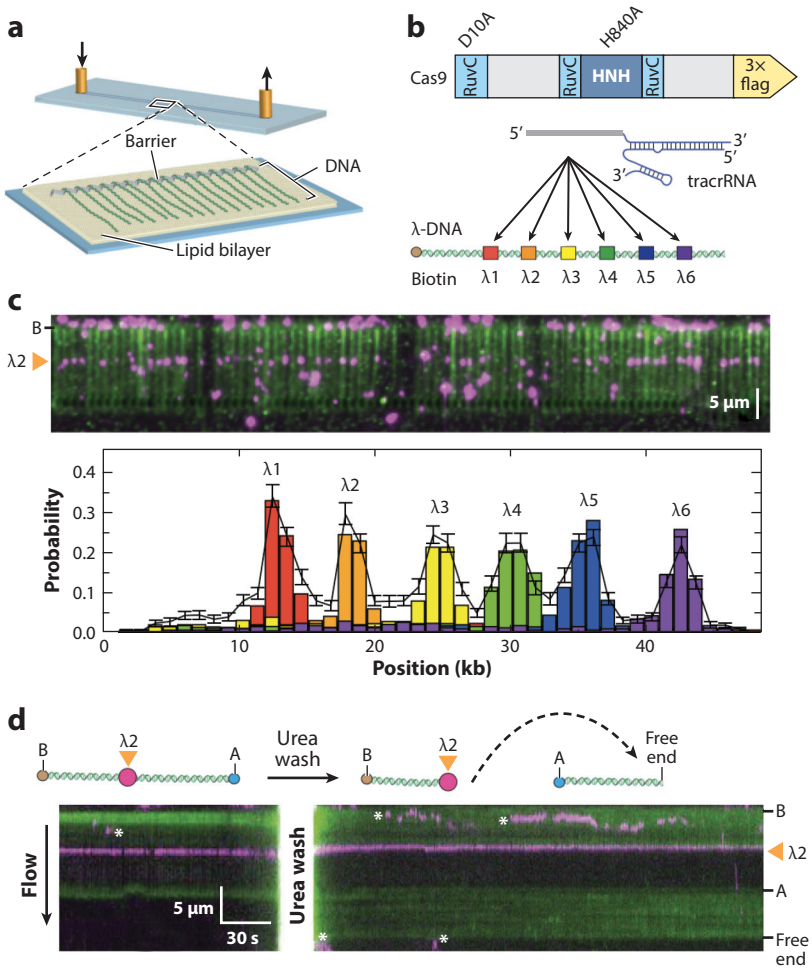


Figure 4

Single-molecule investigation of Cas9 binding and DNA cleavage in vitro (65). (*a*) Schematic of DNA curtains single-molecule assay. (*b*) Schematic of protein construct and binding sites along the DNA template. FLAG-tagged Cas9 and anti-FLAG-functionalized quantum dots were used to directly visualize Cas9 binding to bacteriophage lambda DNA templates. (*c*) Single-molecule images of Cas9 binding to DNA using the DNA curtains assay. By modifying the sgRNA in the Cas9-RNA complex, the fluorescently labeled Cas9 localized to different target locations on DNA templates. (*d*) Catalytically active Cas9 initially failed to cleave DNA templates at target sites; however, a stringent wash (7M urea) resulted in release of the two cleaved ends of DNA at the site of the double strand break. Adapted from Reference 65 with permission from Nature Publishing Group. Abbreviations: crRNA, short CRISPR RNA; tracrRNA, *trans*-activating crRNA.

containing a PAM. The inherent bias in DNA sequence search greatly aids efficiency and helps to explain the observed lack of 1D facilitated diffusion for Cas9.

Tracking Single Cas9 Proteins in Live Cells

Moving beyond in vitro single-molecule experiments, the dynamics of single Cas9 proteins were recently observed by in vivo imaging experiments by Knight et al. (68). Imaging single biomolecules

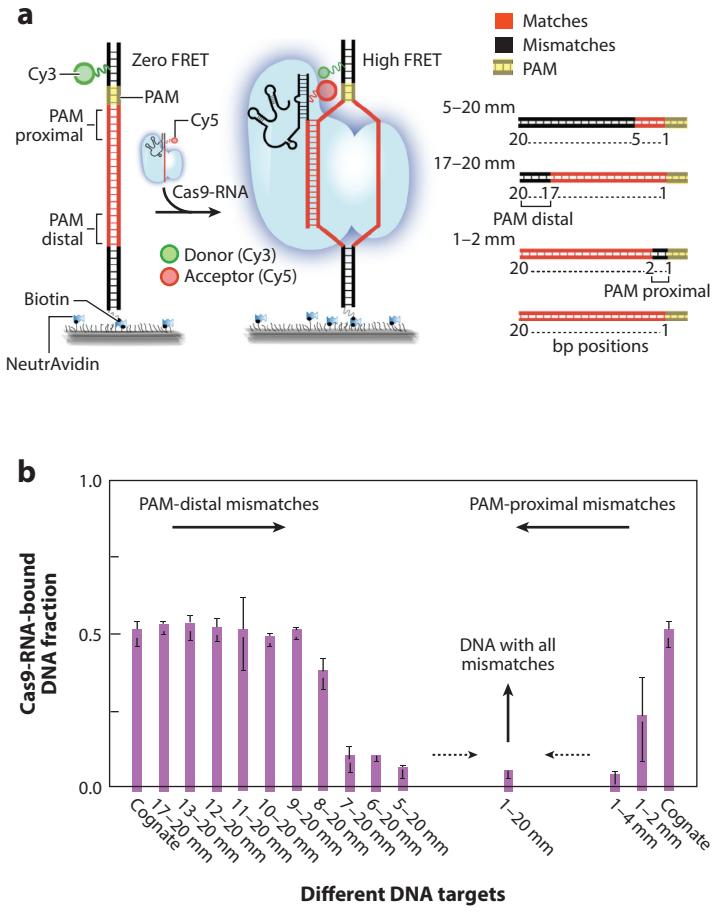


Figure 5

Application of single-molecule FRET (smFRET) to the study of Cas9 binding to target and mismatch DNA templates (70). (*a*) Schematic of smFRET assay. Custom-designed short DNA templates were labeled with a FRET donor (Cy3), and Cas9-sgRNA was labeled with a FRET acceptor (Cy5). DNA templates (*right*) were designed to probe target mismatches proximal and distal to the protospacer adjacent motif (PAM). (*b*) Fraction of Cas9 binding to different DNA sites showing that PAM-proximal mismatches strongly affected Cas9 binding, whereas multiple PAM-distal mismatches were relatively well tolerated. Adapted from Reference 70 with permission from Nature Publishing Group.

in living cells with high spatiotemporal resolution presents additional challenges that are not encountered in in vitro single-molecule imaging studies. In particular, it can be challenging to specifically conjugate sufficiently bright and photostable fluorophores to endogenously produced biomolecules. By using a genetically encoded HaloTag, which enabled covalent conjugation of a bright and photostable cell-permeable dye (69), 2D and 3D tracking of Cas9 was carried out in live mouse fibroblast cells (**Figure 6*a***).

To assess Cas9 dynamics in fibroblasts, fluorescently labeled H2B and Sox2 proteins were also imaged and tracked. By using this approach, the dynamic behavior of Cas9 could be compared to both binding-dominant nuclear proteins (H2B) and those displaying a mixture of binding and diffusion (Sox2) (**Figure 6*b*,*c***). Direct comparisons with H2B and Sox2 show that Cas9 spends the overwhelming majority of time engaged in 3D diffusion during sequence search when there

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

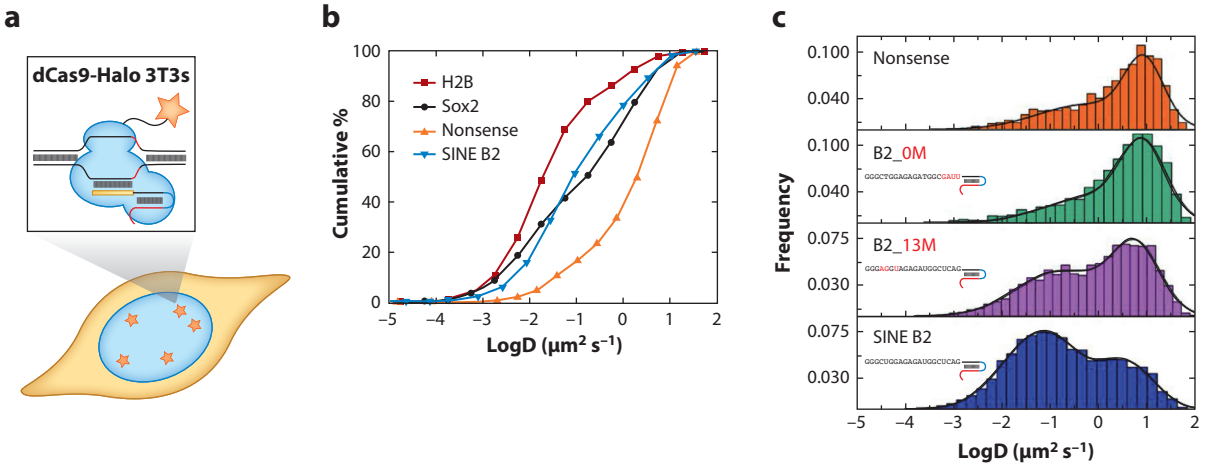


Figure 6

In vivo SMT of Cas9 binding (68). (*a*) Schematic of experiment showing a dCas9-HaloTag construct conjugated with cell-permeable dyes to enable real-time in vivo tracking of Cas9 in mouse 3T3 cells. (*b*) Cumulative distributions of diffusion coefficients for different proteins as determined by live-cell SMT. Comparing the diffusion of the Cas9 construct targeted to chromatin binding proteins (H2B, which remains stably bound, and Sox2, which both binds to and diffuses along DNA), Cas9 lacking a target site diffused more rapidly (*orange*, Nonsense) than both proteins, indicating that Cas9 primarily uses 3D diffusion. In contrast, Cas9 targeted to the SINE B2 repeats in the mouse genome (~300,000 sites), which appeared to follow the behavior of the stably bound H2B and Sox2. (*c*) Histograms showing distributions of diffusion coefficients for TALEs with PAM-proximal mismatches appeared similar to those of Cas9 with nontarget sgRNA, whereas distributions for TALEs with PAM-distal mismatches appeared similar to the distributions of diffusion coefficients of Cas9 targeted to the SINE B2 repeats. Adapted from Reference 68 with permission from the American Association for the Advancement of Science. Abbreviations: 3T3s, 3T3 cells; PAM, protospacer adjacent motif; SMT, single-molecule tracking; TALEs, transcription activator-like effectors.

is no endogenous target for the Cas9-bound sgRNA guide (**Figure 6*b*,*c***). These in vivo results are consistent with prior in vitro single-molecule experiments showing the apparent absence of 1D diffusion during DNA sequence search (65). Furthermore, the importance of the PAM in nucleating substantive Cas9-DNA binding events was further confirmed using in vivo SMFM experiments. Cas9 with sgRNA containing PAM-proximal mismatches to the SINE2 repeats (target sequence) behaved nearly identically to Cas9 with nonsense sgRNA, whereas PAM-distal mismatches appeared to be tolerated (**Figure 6*c***).

Although the complex higher-ordered structure of eukaryotic chromatin is challenging to recreate in vitro, live-cell in vivo imaging can be used to study the interactions between Cas9 and chromatin during different states of compaction. By selectively illuminating regions of heterochromatin via eGFP-tagged heterochromatin protein 1, DNA search trajectories of single Cas9 proteins could be superimposed onto the locations of heterochromatin regions (68). Here, Cas9 displays a clear decrease in search trajectories within regions of heterochromatin, which is likely due to the densely compacted higher-order structure in these regions compared with euchromatin. In addition, there was a marked slowdown in the rate of diffusion of Cas9 proteins within heterochromatin regions compared with whole-nucleus diffusion. Despite both a reduction in sampling of heterochromatin and a slowdown of target search in these regions, Cas9 was nevertheless readily able to localize to target sites within known heterochromatic regions. These results indicate that the search process is guided by mechanisms allowing for successful navigation through crowded macromolecular environments.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Effects of Target Mismatch on Cas9 Binding

Single-molecule techniques can be used to study the impact of target site mismatches on Cas9 binding, which is a particularly important aspect of gene editing technologies. Singh et al. (70) systematically studied Cas9 binding to an array of DNA templates with engineered mismatches using smFRET (**Figure 5**). PAM-distal mismatches are significantly better tolerated than PAM-proximal mismatches (**Figure 5*b***). In particular, these results showed that Cas9 binding was only minimally impacted when up to 13 mismatches from the PAM-distal end of the DNA binding site were introduced. In contrast, only two mismatches at the PAM-proximal end of the binding site were required to significantly reduce Cas9 binding. smFRET allows for a high spatiotemporal resolution, which further enabled a systematic study of the impact of target site mismatches on Cas9 binding kinetics. These results showed a sharp increase in Cas9 binding lifetimes when >7 base pairs of heteroduplex (PAM-proximal target match) formation were permitted. Here, Cas9 binding times increased from <1 to ~8 s and further to ~16 s as the heteroduplex formation was increased from 6 to 7 and further to 8 base pairs.

Cas9 Target Recognition and R-Loop Formation: Single-Molecule Fluorescence and Force Spectroscopy

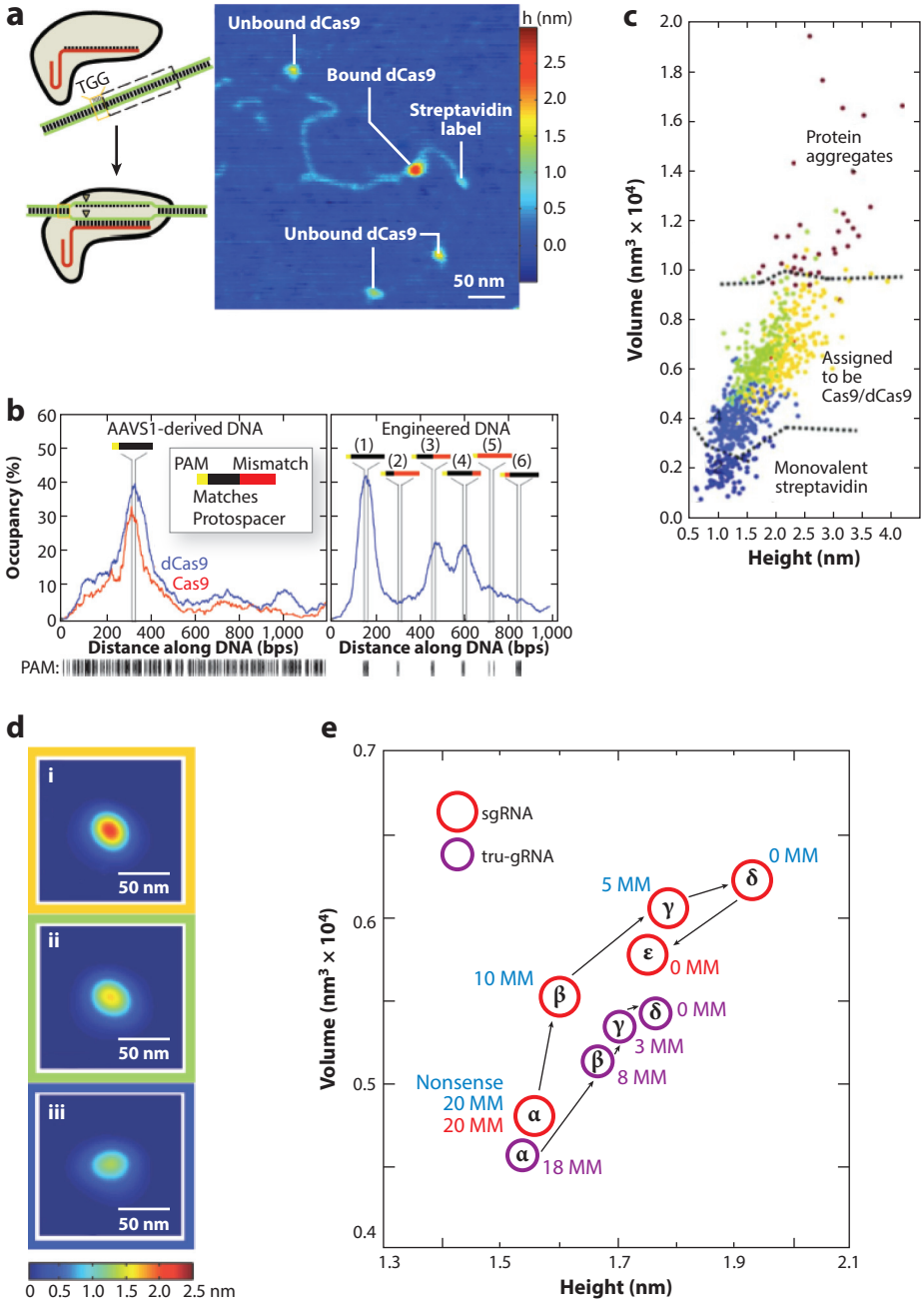
Moving beyond Cas9 target search, single-molecule imaging has also been used to study Cas9 binding to target DNA sites. Using the DNA curtains assay, Sternberg and colleagues (66) designed sgRNAs that localized fluorescently labeled nuclease-deficient Cas9 mutants (dCas9) to 6 different 20–base pair locations along the dual-tethered DNA substrate (**Figure 4*c***). As expected, dCas9 with these specific guide sequences overwhelmingly localized at the desired target locations. These experiments revealed an additional key insight regarding the nonspecific interactions between Cas9 and DNA. Here, the authors observed a bi-exponential distribution of Cas9 binding times to DNA templates, which was attributed to the formation of multiple intermediates preceding the formation of a stably bound complex.

Formation of an R-loop is a fundamental step in the CRISPR-mediated degradation of invasive DNA. Therefore, the molecular-level details affecting R-loop formation and stability may offer valuable insight into the rational design of improved CRISPR/Cas-mediated gene editing approaches. Szczelkun et al. (71) used a magnetic tweezers assay to observe Cas9-mediated R-loop formation on individual DNA templates in real time. The PAM region dictated the rate of formation of R-loops but not R-loop stability. By systematically mutating PAM elements on a target DNA template and probing for R-loop formation kinetics and stability, Cas9 was significantly more stringent in accepting PAM mutations leading to R-loop formation compared with Cascade, which is a type I CRISPR nuclease complex. Potential R-loop formation events at PAM-mutated sites, however, did not display a reduced stability. Furthermore, Szczelkun et al. (71) demonstrated that whereas Cas9 R-loop formation kinetics were not impacted by PAM-distal mismatches (protospacer mismatches), the stability of R-loops was closely linked to the extent of PAM-distal mismatches, in an apparent all-or-none threshold. From this perspective, this study reinforced the importance of the PAM in target search and initial DNA binding, while revealing the importance of PAM-distal sites on the stability of R-loop formation, which is critical for a cleavage event.

Conformational Changes upon Cas9 Binding: Single-Molecule Force Spectroscopy

Whereas single-molecule fluorescence-based tracking experiments can provide rich molecular-scale information regarding the dynamics of protein-DNA interactions, single-molecule force

spectroscopy can be used to resolve intramolecular changes with high levels of detail. Josephs et al. (72) recently used AFM to study Cas9/dCas9 binding to a variety of DNA templates. Using engineered target site mismatches in DNA templates, these authors were able to uncover several details of the underlying target recognition pathway (**Figure 7**). In particular, AFM was used to investigate the conformation of Cas9 bound to several different DNA templates. dCas9-sgRNA was



Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

observed to adopt an elongated, egg-shaped structure when bound to nontarget DNA. However, dCas9-sgRNA complexes were found to adopt a round conformation with a large central bulge when bound to target DNA, which is consistent with the open conformation for nuclease activity proposed by Jinek et al. (73) in a crystallographic/electron microscopy study (**Figure 7*d,e***). Josephs et al. (72) correlated the observed structures with their relative location along the engineered DNA templates. In this way, as the degree of sequence mismatch distal to the PAM increases, the structure of Cas9 transitions to the elongated, egg-shaped conformation (**Figure 7*e***). These results are consistent with the so-called conformational gating mechanism, whereby sgRNA base pairing with PAM-distal sites will act to stabilize the active conformation of the Cas9 complex, thereby allowing for DNA cleavage.

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Cas9-Mediated dsDNA Cleavage Events: Single-Molecule Fluorescence Imaging

The majority of single-molecule experiments on the Cas9 system have used the nuclease-deficient dCas9; however, a handful of studies have examined the dynamics of natural Cas9 nuclease activity. Using the DNA curtains single-molecule assay, Sternberg et al. (65) found that DNA templates bearing target sites for Cas9 appeared to remain intact despite the persistent binding of enzymatically active Cas9 for minutes or longer. Separation of cleaved DNA templates at the engineered Cas9 binding site was observed only under highly disruptive conditions (7M urea) (**Figure 4*d***). The insights provided from single-molecule imaging suggested that Cas9 is a single-turnover enzyme, which was further corroborated using bulk enzymatic cleavage assays. Singh et al. (66) repeated these experiments using an smFRET assay, and it was again found that highly denaturing conditions (7M urea) were required to observe a loss of fluorescence signal indicating DNA cleavage and release of the DNA template from the coverslip surface.

In related work, Josephs et al. (72) revealed a slight difference in the conformational change upon target binding of dCas9 versus Cas9 that may be attributed to additional conformational changes during cleavage. Moreover, Sternberg et al. (74) used bulk FRET experiments to probe the intramolecular conformational changes of Cas9 during cleavage. Importantly, this study revealed that a conformational change in the HNH nuclease domain of Cas9 is a critical layer of security in preventing off-target cleavage. Taken together, the collection of recent single-molecule studies has provided clear evidence for off-target Cas9 binding, even at highly mutated target sites. Interestingly, however, the bulk FRET experiments by Sternberg et al. (74) show that the critical

←

Figure 7

Direct visualization of Cas9 binding and subsequent intramolecular conformational changes (72). (*a*) Atomic force microscopy was used to observe Cas9 binding to natural and engineered DNA templates, wherein one end of DNA templates was labeled with streptavidin to correlate binding locations to sequence. (*b*) Both Cas9 and dCas9 localized to target sites within both DNA templates, but significant off-target binding at sites containing PAM-distal mismatches is also observed (*right*). (*c*) Analysis of the volume versus height of imaged Cas9 and DNA streptavidin tags reveals clusters within assigned Cas9 molecules. (*d*) The Cas9 in panel *c* can be roughly divided into (*i*) target-bound Cas9, (*ii*) partial mismatch-bound Cas9, and (*iii*) fully mismatch-bound Cas9, with the height and circularity of the Cas9 increasing with target complementarity. (*e*) Conformational changes upon increasing target complementarity are visualized for Cas9/dCas9. Mean volumes and heights of Cas9/dCas9 with sgRNAs (*red circles*, with *red labels* for Cas9 and *blue labels* for dCas9) or tru-gRNAs (*purple circles*) bound at each feature on the substrates. dCas9 with tru-gRNAs are expected to interact only with the first 3 or 8 PAM-distal mismatches of the 5 MM and 10 MM sites (labeled 3 MM and 8 MM here, respectively). Adapted from Reference 72 with permission from Oxford University Press. Abbreviations: PAM, protospacer adjacent motif; sgRNA, single guide RNA; tru-gRNA, truncated guide RNA.

HNH conformational change associated with nuclease activity is highly intolerant of sequence mismatches, even at PAM-distal sites.

OUTLOOK

Single-molecule techniques have provided fundamentally new insights into the search and recognition behavior of TALEs and CRISPR/Cas9, which have emerged as prominent gene editing systems. SMFM and single-molecule force techniques allow molecular-scale biological events to be interrogated at high spatiotemporal resolution, which provides real-time dynamic information on protein-DNA binding that extends beyond static crystal structures and bulk-scale biochemical assays. In the context of high-precision genome editing technologies, off-target cleavage events are of paramount concern for future clinical applications. Moving forward, a clear molecular-level picture of sequence search dynamics and binding may aid in the rational design of improved TALENs or Cas9 systems. From this view, SMFM and single-molecule force techniques are poised to play a key role in further elucidating the dynamic behavior and molecular mechanisms for these systems.

There is a need for further investigation of the temporal dynamics and molecular heterogeneity underlying individual Cas9 cleavage events. These experiments would further support Sternberg et al.'s (74) hypothesis that off-target cleavage of Cas9 may be attributed to rapid interconversion between protein conformational states. Moreover, these experiments could be invaluable in revealing how custom-designed Cas9 constructs would behave under different conditions in the context of practical gene editing applications.

Conformational changes and the associated plasticity of the TALE superhelical structure have also been investigated, but only using computational approaches (51). From this perspective, smFRET experiments could be leveraged to study the intramolecular conformational changes in TALE proteins to provide a real-time readout of the proposed two-state model for TALE search involving helical compression/relaxation (56). In addition, intermolecular smFRET studies of TALE binding to a systematic array of mismatch DNA templates could provide more clarity on TALE off-target binding, including the presence or absence of a threshold for stable binding and further investigation of the reported N-terminal polarity in TALE binding. Similar to studies of conformational changes of Cas9, these experiments hold the potential to reveal additional design rules for TALE proteins in minimizing deleterious off-target events.

In addition to single-molecule studies of intramolecular conformational changes, the bottom-up approach of in vitro single-molecule techniques can be significantly expanded to study additional aspects of protein search and binding. From this perspective, there have been no reported studies of search and binding for gene editing proteins interacting with additional chromatin-binding proteins. Prior single-molecule studies of other systems have explored how protein-protein interactions affect DBP search (75–77), and these experiments could be imported to study gene editing proteins to provide an increased understanding of the kinetics of TALE/Cas9 search and binding.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Gaj T, Gersbach CA, Barbas CF. 2013. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 31(7):397–405

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

2. Büning H. 2013. Gene therapy enters the pharma market: the short story of a long journey. *EMBO Mol. Med.* 5(1):1–3
3. Simonelli F, Maguire AM, Testa F, Pierce EA, Mingozzi F, et al. 2010. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. *Mol. Ther.* 18(3):643–50
4. Naldini L. 2015. Gene therapy returns to centre stage. *Nature* 526(7573):351–60
5. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 11(9):636–46
6. Holt N, Wang J, Kim K, Friedman G, Wang X, et al. 2010. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* 28(8):839–47
7. Urnov FD, Miller JC, Lee Y-L, Beausejour CM, Rock JM, et al. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435(7042):646–51
8. Torikai H, Reik A, Liu PQ, Zhou Y, Zhang L, et al. 2012. A foundation for universal T-cell based immunotherapy: T cells engineered to express a CD19-specific chimeric-antigen-receptor and eliminate expression of endogenous TCR. *Blood* 119(24):5697–705
9. Ding Q, Lee Y-K, Schaefer EAK, Peters DT, Veres A, et al. 2013. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12(2):238–51
10. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, et al. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39(12):e82
11. Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, et al. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature* 491(7422):114–18
12. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816–21
13. Jiang W, Marraffini LA. 2015. CRISPR-Cas: new tools for genetic manipulations from bacterial immunity systems. *Annu. Rev. Microbiol.* 69:209–28
14. Heyer W-D, Ehmsen KT, Liu J. 2010. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* 44(17):113–39
15. Beerli RR, Barbas CF. 2002. Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* 20(2):135–41
16. Maeder ML, Thibodeau-Beganny S, Osiak A, Wright DA, Anthony RM, et al. 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell* 31(2):294–301
17. Mak AN-S, Bradley P, Bogdanove AJ, Stoddard BL. 2013. TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* 23(1):93–99
18. Mussolino C, Cathomen T. 2012. TALE nucleases: tailored genome engineering made easy. *Curr. Opin. Biotechnol.* 23(5):644–50
19. Savić N, Schwank G. 2016. Advances in therapeutic CRISPR/Cas9 genome editing. *Transl. Res.* 168:15–21
20. Wright AV, Nunez JK, Doudna JA. 2016. Review biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 164(1–2):29–44
21. Stella S, Montoya G. 2015. The genome editing revolution: A CRISPR-Cas TALE off-target story. *Inside Cell* 1:7–15
22. Ledford H. 2015. CRISPR, the disruptor. *Nature* 522(7554):20–24
23. Thompson RE, Larson DR, Webb WW. 2002. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* 82(5):2775–83
24. Yildiz A. 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* 300(5628):2061–65
25. Yildiz A. 2004. Kinesin walks hand-over-hand. *Science* 303(5658):676–78
26. Elf J, Li G-W, Xie XS. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316(5828):1191–94
27. Roy R, Hohng S, Ha T. 2008. A practical guide to single-molecule FRET. *Nat. Methods* 5(6):507–16
28. Schuler B, Eaton WA. 2008. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* 18(1):16–26
29. Myong S, Rasnik I, Joo C, Lohman TM, Ha T. 2005. Repetitive shuttling of a motor protein on DNA. *Nature* 437(7063):1321–25

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

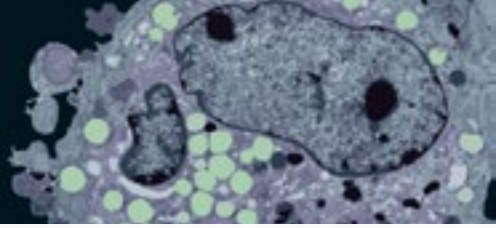
30. Neuman KC, Nagy A. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat. Methods* 5(6):491–505
31. Moffitt JR, Chemla YR, Smith SB, Bustamante C. 2008. Recent advances in optical tweezers. *Annu. Rev. Biochem.* 77:205–28
32. Mosconi F, Allemand JF, Bensimon D, Croquette V. 2009. Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers. *Phys. Rev. Lett.* 102(7):1–4
33. Wang MD, Yin H, Landick R, Gelles J, Block SM. 1997. Stretching DNA with optical tweezers. *Biophys. J.* 72(3):1335–46
34. Tanyeri M, Johnson-Chavarria EM, Schroeder CM. 2010. Hydrodynamic trap for single particles and cells. *Appl. Phys. Lett.* 96(22):22–24
35. Cohen AE, Moemer WE. 2005. Method for trapping and manipulating nanoscale objects in solution. *Appl. Phys. Lett.* 86(9):093109
36. Engel A, Müller DJ. 2000. Observing single biomolecules at work with the atomic force microscope. *Nat. Struct. Biol.* 7(9):715–18
37. Hansma HG, Pietrasanta LIAI, Auerbach ID. 2000. Probing biopolymers with the atomic force microscope: a review. *J. Biomater. Sci. Polym. Ed.* 11(7):675–83
38. Bonas U, Stall RE, Staskawicz B. 1989. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet*. 218(1):127–36
39. Gu K, Yang B, Tian D, Wu L, Wang D, et al. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435(7045):1122–25
40. Hopkins CM, White FF, Choi S-H, Guo A, Leach JE. 1993. Identification of a family of avirulence genes from *Xanthomonas oryzae* pv. *oryzae*. *Mol. Plant-Microbe Interact.* 5(6):451–59
41. Ishihara H, Ponciano G, Leach JE, Tsuyumu S. 2003. Functional analysis of the 3′ end of *AvrBs3*/PthA genes from two *Xanthomonas* species. *Physiol. Mol. Plant Pathol*. 63(6):329–38
42. Ballvora A, Pierre M, van den Ackerveken G, Schornack S, Rossier O, et al. 2001. Genetic mapping and functional analysis of the tomato *Bs4* locus governing recognition of the *Xanthomonas campestris* pv. *vesicatoria* AvrBs4 protein. *Mol. Plant-Microbe Interact*. 14(5):629–38
43. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, et al. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326(5959):1509–12
44. Moscou MJ, Bogdanove AJ. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science* 326(5959):1501
45. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, et al. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2):757–61
46. Li T, Liu B, Spalding MH, Weeks DP, Yang B. 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* 30(5):390–92
47. Sung YH, Baek I-J, Kim DH, Jeon J, Lee J, et al. 2013. Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.* 31(1):23–24
48. Mak AN-S, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* 335(6069):716–19
49. Deng D, Yan C, Pan X, Mahfouz M, Wang J, et al. 2012. Structural basis for sequence-specific recognition of TAL effectors. *Science* 335(6069):720–23
50. Jiménez-Menéndez N, Fernández-Millán P, Rubio-Cosials A, Arnan C, Montoya J, et al. 2010. Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat. *Nat. Struct. Mol. Biol.* 17(7):891–93
51. Wan H, Hu J-p, Li K-s, Tian X-h, Chang S. 2013. Molecular dynamics simulations of DNA-free and DNA-bound TAL effectors. *PLOS ONE* 8(10):e76045
52. Gao H, Wu X, Chai J, Han Z. 2012. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res*. 22(12):1716–20
53. Schreiber T, Bonas U. 2014. Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res*. 42(11):7160–69
54. Slutsky M, Mirny LA. 2004. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* 87(6):4021–35

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

55. Wicky BIM, Stenta M, Dal Peraro M. 2013. TAL effectors specificity stems from negative discrimination. *PLOS ONE* 8(11):e80261
56. Cuculis L, Abil Z, Zhao H, Schroeder CM. 2015. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* 6:7277
57. Lin J, Countryman P, Buncher N, Kaur P, E L, et al. 2014. TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres. *Nucleic Acids Res.* 42(4):2493–504
58. Tafvizi A, Huang F, Fersht AR, Mirny LA, van Oijen AM. 2011. A single-molecule characterization of p53 search on DNA. *PNAS* 108(2):563–68
59. Sun N, Liang J, Abil Z, Zhao H. 2012. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. Biosyst.* 8(4):1255
60. Winter RB, Berg OG, von Hippel PH. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor–operator interaction: Kinetic measurements and conclusions. *Biochemistry* 20(24):6961–77
61. Cuculis L, Abil Z, Zhao H, Schroeder CM. 2016. TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.* 12(10):831–37
62. Marraffini LA, Sontheimer EJ. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* 11(3):181–90
63. Chylinski K, Le Rhun A, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* 10(5):726–37
64. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31(9):822–26
65. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507(7490):62–67
66. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T. 2016. Real time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* 7:12778
67. Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513(7519):569–73
68. Knight SC, Xie L, Deng W, Guglielmi B, Witkowsky LB, et al. 2015. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* 350(6262):823–26
69. Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, et al. 2015. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* 12(3):244–50
70. Singh D, Sternberg SH, Fei J, Doudna JA, Ha T. 2016. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* 7:12778
71. Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, et al. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *PNAS* 111(27):9798–803
72. Josephs EA, Kocak DD, Fitzgibbon CJ, McMenemy J, Gersbach CA, Marszalek PE. 2015. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res.* 43(18):8924–41
73. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343(6176):1247997
74. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527(7576):1–14
75. Gorman J, Plys AJ, Visnapuu M-L, Alani E, Greene EC. 2010. Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* 17(8):932–38
76. Finkelstein IJ, Visnapuu M-L, Greene EC. 2010. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* 468(7326):983–87
77. Lee J, Finkelstein IJ, Arciszewska LK, Sherratt DJ, Greene EC. 2014. Single-molecule imaging of FtsK translocation reveals mechanistic features of protein-protein collisions on DNA. *Mol. Cell* 54(5):832–43

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

ANNUAL REVIEWS
Connect With Our Experts



New From Annual Reviews:

ONLINE NOW!

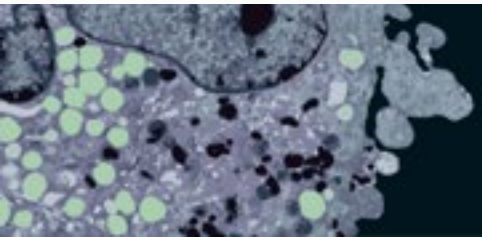
Annual Review of Cancer Biology
cancerbio.annualreviews.org • Volume 1 • March 2017

Co-Editors: **Tyler Jacks**, *Massachusetts Institute of Technology*
Charles L. Sawyers, *Memorial Sloan Kettering Cancer Center*

The *Annual Review of Cancer Biology* reviews a range of subjects representing important and emerging areas in the field of cancer research. The *Annual Review of Cancer Biology* includes three broad themes: Cancer Cell Biology, Tumorigenesis and Cancer Progression, and Translational Cancer Science.

TABLE OF CONTENTS FOR VOLUME 1:

- ***How Tumor Virology Evolved into Cancer Biology and Transformed Oncology***, Harold Varmus
- ***The Role of Autophagy in Cancer***, Naiara Santana-Codina, Joseph D. Mancias, Alec C. Kimmelman
- ***Cell Cycle–Targeted Cancer Therapies***, Charles J. Sherr, Jiri Bartek
- ***Ubiquitin in Cell-Cycle Regulation and Dysregulation in Cancer***, Natalie A. Borg, Vishva M. Dixit
- ***The Two Faces of Reactive Oxygen Species in Cancer***, Colleen R. Reczek, Navdeep S. Chandel
- ***Analyzing Tumor Metabolism In Vivo***, Brandon Faubert, Ralph J. DeBerardinis
- ***Stress-Induced Mutagenesis: Implications in Cancer and Drug Resistance***, Devon M. Fitzgerald, P.J. Hastings, Susan M. Rosenberg
- ***Synthetic Lethality in Cancer Therapeutics***, Roderick L. Beijersbergen, Lodewyk F.A. Wessels, René Bernards
- ***Noncoding RNAs in Cancer Development***, Chao-Po Lin, Lin He
- ***p53: Multiple Facets of a Rubik's Cube***, Yun Zhang, Guillermina Lozano
- ***Resisting Resistance***, Ivana Bozic, Martin A. Nowak
- ***Deciphering Genetic Intratumor Heterogeneity and Its Impact on Cancer Evolution***, Rachel Rosenthal, Nicholas McGranahan, Javier Herrero, Charles Swanton
- ***Immune-Suppressing Cellular Elements of the Tumor Microenvironment***, Douglas T. Fearon
- ***Overcoming On-Target Resistance to Tyrosine Kinase Inhibitors in Lung Cancer***, Ibiayi Dagogo-Jack, Jeffrey A. Engelman, Alice T. Shaw
- ***Apoptosis and Cancer***, Anthony Letai
- ***Chemical Carcinogenesis Models of Cancer: Back to the Future***, Melissa Q. McCreery, Allan Balmain
- ***Extracellular Matrix Remodeling and Stiffening Modulate Tumor Phenotype and Treatment Response***, Jennifer L. Leight, Allison P. Drain, Valerie M. Weaver
- ***Aneuploidy in Cancer: Seq-ing Answers to Old Questions***, Kristin A. Knouse, Teresa Davoli, Stephen J. Elledge, Angelika Amon
- ***The Role of Chromatin-Associated Proteins in Cancer***, Kristian Helin, Saverio Minucci
- ***Targeted Differentiation Therapy with Mutant IDH Inhibitors: Early Experiences and Parallels with Other Differentiation Agents***, Eytan Stein, Katharine Yen
- ***Determinants of Organotropic Metastasis***, Heath A. Smith, Yibin Kang
- ***Multiple Roles for the MLL/COMPASS Family in the Epigenetic Regulation of Gene Expression and in Cancer***, Joshua J. Meeks, Ali Shilatifard
- ***Chimeric Antigen Receptors: A Paradigm Shift in Immunotherapy***, Michel Sadelain



ANNUAL REVIEWS | CONNECT WITH OUR EXPERTS
650.493.4400/800.523.8635 (US/CAN)
www.annualreviews.org | service@annualreviews.org

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Annual Review of Chemical and Biomolecular Engineering

Volume 8, 2017

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.

Contents

A Conversation with John McKetta
John J. McKetta, Jr. and Thomas M. Truskett 1

At Light Speed: Advances in Optogenetic Systems for Regulating Cell Signaling and Behavior
Nicole A. Repina, Alyssa Rosenbloom, Abhirup Mukherjee, David V. Schaffer, and Ravi S. Kane 13

Nanoengineering Heterogeneous Catalysts by Atomic Layer Deposition
Joseph A. Singh, Nuoya Yang, and Stacey F. Bent 41

Big Data Analytics in Chemical Engineering
Leo Chiang, Bo Lu, and Ivan Castillo 63

Biocatalytic Nanocomposites for Combating Bacterial Pathogens
Xia Wu, Seok-Joon Kwon, Jungbae Kim, Ravi S. Kane, and Jonathan S. Dordick 87

A Review of Biorefinery Separations for Bioproduct Production via Thermocatalytic Processing
Hannah Nguyen, Robert F. DeJaco, Nitish Mittal, J. Ilja Siepmann, Michael Tsapatsis, Mark A. Snyder, Wei Fan, Basudeb Saha, and Dionisios G. Vlachos 115

Driving Forces for Nonnative Protein Aggregation and Approaches to Predict Aggregation-Prone Regions
Gulsum Meric, Anne S. Robinson, and Christopher J. Roberts 139

Progress in Brewing Science and Beer Production
C.W. Bamforth 161

Engineering Microneedle Patches for Vaccination and Drug Delivery to Skin
Mark R. Prausnitz 177

Process Principles for Large-Scale Nanomanufacturing
Sven H. Behrens, Victor Breedveld, Maritza Mujica, and Michael A. Filler 201

Magnetic Resonance Imaging and Velocity Mapping in Chemical Engineering Applications

Lynn F. Gladden and Andrew J. Sederman 227

Recent Developments and Challenges in Optimization-Based Process Synthesis
Qi Chen and I.E. Grossmann 249

Design and Scaling Up of Microchemical Systems: A Review
Jisong Zhang, Kai Wang, Andrew R. Teixeira, Klavs F. Jensen, and Guangsheng Luo 285

Aerogels in Chemical Engineering: Strategies Toward Tailor-Made Aerogels
Irina Smirnova and Pavel Gurikov 307

Algae to Economically Viable Low-Carbon-Footprint Oil
Ramesh Bhujade, Mandan Chidambaram, Avnish Kumar, and Ajit Sapre 335

Modular Chemical Process Intensification: A Review
Yong-ha Kim, Lydia K. Park, Sotira Yiacoumi, and Costas Tsouris 359

Thermophysical Properties and Phase Behavior of Fluids for Application in Carbon Capture and Storage Processes
J.P. Martin Trusler 381

Multivariate Analysis and Statistics in Pharmaceutical Process Research and Development
José E. Tabora and Nathan Domagalski 403

Atmospheric Aerosols: Clouds, Chemistry, and Climate
V. Faye McNeill 427

Hydrogen Storage Technologies for Future Energy Systems
Patrick Preuster, Alexander Alekseev, and Peter Wasserscheid 445

The Selectivity Challenge in Organic Solvent Nanofiltration: Membrane and Process Solutions
Patrizia Marchetti, Ludmila Peeva, and Andrew Livingston 473

Nanoscale Aggregation in Acid- and Ion-Containing Polymers
L. Robert Middleton and Karen I. Winey 499

High-Throughput Automation in Chemical Process Development
Joshua A. Selekman, Jun Qiu, Kristy Tran, Jason Stevens, Victor Rosso, Eric Simmons, Yi Xiao, and Jacob Janey 525

Artificially Engineered Protein Polymers
Yun Jung Yang, Angela L. Holmberg, and Bradley D. Olsen 549

A Single-Molecule View of Genome Editing Proteins: Biophysical Mechanisms for TALEs and CRISPR/Cas9
Luke Cuculis and Charles M. Schroeder 577

Annu. Rev. Chem. Biomol. Eng. 2017.8:577-597. Downloaded from www.annualreviews.org
Access provided by University of Illinois - Urbana Champaign on 09/10/17. For personal use only.