nature chemical biology

ARTICLE

PUBLISHED ONLINE: 15 AUGUST 2016 | DOI: 10.1038/NCHEMBIO.2152

TALE proteins search DNA using a rotationally decoupled mechanism

Luke Cuculis[1], Zhanar Abil[2], Huimin Zhao[1–5] & Charles M Schroeder[1,3–5]*

Transcription activator-like effector (TALE) proteins are a class of programmable DNA-binding proteins used extensively for gene editing. Despite recent progress, however, little is known about their sequence search mechanism. Here, we use single-molecule experiments to study TALE search along DNA. Our results show that TALEs utilize a rotationally decoupled mechanism for nonspecific search, despite remaining associated with DNA templates during the search process. Our results suggest that the protein helical structure enables TALEs to adopt a loosely wrapped conformation around DNA templates during nonspecific search, facilitating rapid one-dimensional (1D) diffusion under a range of solution conditions. Furthermore, this model is consistent with a previously reported two-state mechanism for TALE search that allows these proteins to overcome the search speed–stability paradox. Taken together, our results suggest that TALE search is unique among the broad class of sequence-specific DNA-binding proteins and supports efficient 1D search along DNA.

Recent advances in genome engineering offer the potential to dramatically alter the treatment of human disease. Achieving this potential, however, is a major challenge owing to the high degree of precision and accuracy required for modifying large, intact genomes. Genome-editing techniques based on programmable nucleases, including zinc-finger nucleases[1], the CRISPR–Cas9 system[2], and TALE nucleases[3–5], are finding widespread use for genomic editing in plants, bacteria, and mammalian cells. Despite recent progress, however, the molecular mechanisms underlying the DNA search process for TALEs are not fully understood.

Previous crystallographic studies have shown that TALEs adopt a characteristic compressed helical shape when bound to target DNA sequences, in which they are wrapped tightly around the DNA major groove in a superhelical conformation[6]. In the absence of DNA, TALEs show an extended helical conformation[7]. TALEs comprise three conserved regions: an N-terminal region (NTR) containing the type III translocation signal, a central repeat domain (CRD) that forms specific DNA contacts, and a C-terminal region (CTR) containing nuclear localization signals and an acidic activation domain[8]. The NTR harbors four noncanonical repeats (known as −3, −2, −1, and 0) that bind DNA nonspecifically via two large positively charged patches, where position 0 corresponds to the base immediately preceding the first base recognized by the canonical repeats[9]. NTR repeat −1 is thought to interact with a strictly conserved thymine (T) base at the 5′ end of binding targets of natural TALEs[7,10–12]. Sequence-specific DNA recognition is conferred by the CRD, which contains a series of tandem repeats that are typically 34 amino acids in length[8]. Repeat variable diresidues (RVDs) are located at residues 12 and 13 in each repeat element and confer specific binding to nucleobases. A 2009 study developed a DNA-recognition code for TALE proteins by deciphering the sequence specificity conferred by individual RVDs[11]. Despite recent progress, however, far less is known about the nonspecific search process and the transition from nonspecific to specifically bound states for TALEs.

Single-molecule techniques can be used to study the dynamic search process of DNA-binding proteins with high spatiotemporal resolution[13], offering a complementary approach to protein crystallography for probing DNA binding. Prior single-molecule studies have shown that a broad class of DNA-binding proteins track the major groove while translating along DNA, effectively rotating or 'spinning' via an apparent sliding mechanism[14]. We have previously shown that TALEs diffuse one dimensionally, and their DNA-interrogation activity is governed by a two-state search process in which periods of rapid 1D search are interspersed with periods of arrested motion that are thought to correspond to a recognition mode for DNA binding[15]. Interestingly, although the TALE NTR is essential for initial DNA binding, TALE mutants lacking the CRD and CTR (NTR-only mutants) do not have a recognition mode during search[9,15]. Despite these insights, key questions remain, including that of how the superhelical TALE structure mediates rapid nonspecific search on DNA. In this work we use single-molecule techniques to directly study the nonspecific search process for TALEs along DNA templates. Our results show that TALEs utilize a distinct molecular mode of nonspecific search, wherein they translate without rotating but remain closely associated with DNA templates in the process.

RESULTS

Single-molecule assay for TALE diffusion

We directly visualized TALE dynamics during nonspecific search using a dual-tethered DNA assay combined with total internal reflection fluorescence microscopy (TIRF-M) (**Fig. 1a–c**). In this assay, long DNA molecules are tethered to passivated glass coverslip surfaces at both termini using specific chemical linkages[16] (**Supplementary Results**, **Supplementary Figs. 1** and **2**). We studied the behavior of the 21.5-repeat TALE previously developed for the editing and correction of an *HBB* gene mutation associated with sickle cell disease[17] (**Fig. 1d**). In brief, we selected a TALE binding site with a minimal number of guanines, as recognition of guanine by RVDs arises from a compromise between specificity and strength of binding[18]. The parent TALE construct in these experiments contains two nonspecific and strong double-asparagine (NN) RVDs (**Fig. 1d**). The C-terminal repeat (the twenty-second repeat

[1]Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. [2]Department of Biochemistry, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. [3]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. [4]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. [5]Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, Illinois, USA. *e-mail: cms@illinois.edu

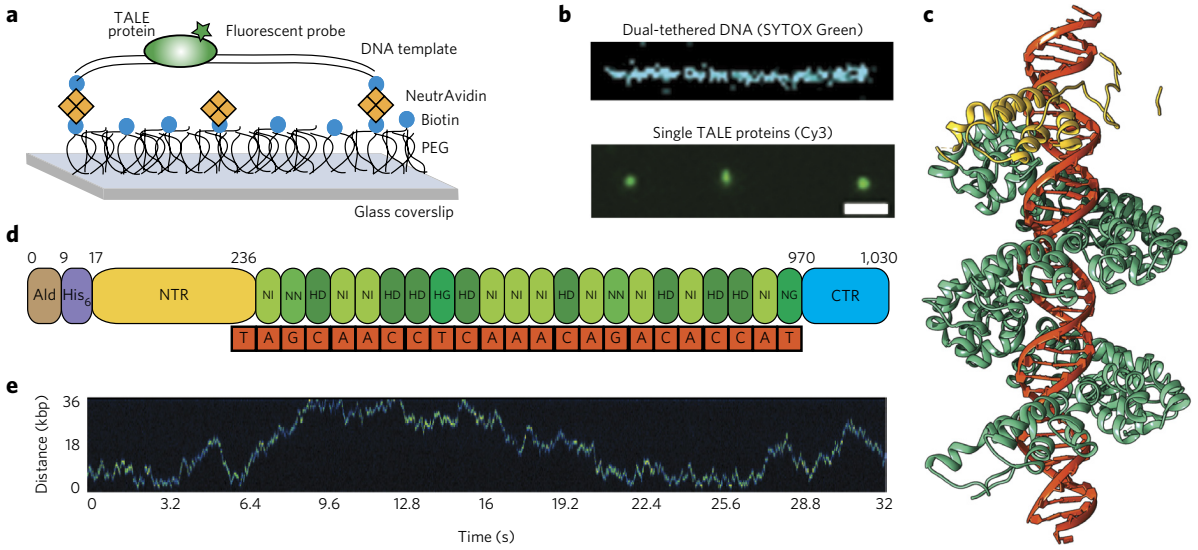

Figure 1 | Single-molecule imaging of TALEs during nonspecific search. (**a**) Schematic of single-molecule assay showing dual-tethered DNA templates with a fluorescently labeled TALE. Single proteins were imaged by TIRF-M. (**b**) Representative images of single DNA templates (top; post-labeled with SYTOX Green after TALE experiments) and individual Cy3-labeled TALEs bound to DNA (bottom). Scale bar, 2 µm. (**c**) Structure of the naturally occurring TAL effector PthXo1 rendered with the NTR in yellow and the CRD in teal. In this view, the CTR is not shown owing to lack of available crystal structures. (**d**) Schematic of the 21.5-repeat TALE construct used in this study, including the genetically encoded aldehyde (Ald) tag and the hexahistidine (His_6) tag. RVDs are shown in green, and their predicted DNA binding partners in orange. (**e**) Kymograph showing an individual TALE scanning >10^6 bp in ~32 s at 500 mM KCl. A total of 17 molecular trajectories were acquired under these conditions.

in the construct) is truncated to 20 residues in length (full-length TALE repeats contain 34 residues). A truncated C-terminal repeat is typical for TALEs and is thus commonly referred to as the 'last half-repeat'[19]. We truncated the N-terminal and C-terminal flanking regions to the last 208 and first 63 amino acids, respectively, as prior work has demonstrated that the adjacent terminal sequences are redundant for DNA binding[17,20], and added an N-terminal hexahistidine tag for affinity purification and a genetically encoded aldehyde moiety for fluorescent labeling (**Supplementary Fig. 3**). Last, we labeled the TALEs with small-molecule organic dyes[21,22] (or other fluorescent probes, as noted) using site-specific chemical conjugation (**Fig. 1b**) and verified that TALE labeling did not perturb function (**Supplementary Fig. 4**).

Using this approach, we studied the search behavior of fluorescently labeled TALE proteins using single-molecule fluorescence microscopy. We found that TALEs diffuse rapidly during nonspecific search on DNA templates in the absence of target sites (**Fig. 1e** and **Supplementary Fig. 5**). To quantify TALE motion, we observed the 1D diffusive behavior of TALEs using a high-sensitivity camera in order to determine the displacement of individual TALEs between successive frames (step sizes) during an acquisition event. Finally, we determined effective 1D diffusion coefficients using a covariance estimator scheme[23] (Online Methods).

Ionic-strength dependence of TALE 1D diffusion

We began by studying the ionic-strength dependence of TALE search dynamics for a series of TALEs with CRDs of various sizes and an NTR-only construct that lacks a CRD and the CTR domain and contains only the four nonspecific repeats that precede the CRD[9] (**Fig. 2**). In contrast to observations of many other DNA-binding proteins, TALE 1D diffusion was extremely sensitive to ionic strength and thus charge-screening effects, as demonstrated by diffusion trajectories for a TALE protein with 21.5 repeats in monovalent salt concentrations ranging from 0.05 to 0.5M KCl (**Fig. 2a** and **Supplementary Fig. 6**). We systematically quantified TALE diffusion as a function of salt concentration for TALEs containing 0 to 21.5 repeats, where 0 repeats corresponds to the TALE NTR-only construct[9] (**Fig. 2b–e**). Well-accepted models for facilitated search of DNA-binding proteins classify protein search behavior as either hopping or sliding, depending on whether 1D diffusion speeds are salt sensitive or invariant[24]. The salt sensitivity in our data suggests that the TALE search mechanism is dominated by hopping behavior, but the structure of target DNA-bound TALEs (**Fig. 1c**) clearly shows that TALEs track the major groove during specific binding. On the basis of these observations, we sought to reconcile how a DNA-wrapped TALE could follow a search mechanism involving hopping behavior, and we used single-molecule techniques to elucidate the search process.

To further probe the search mechanism, we determined the ionic-strength dependence of TALE protein diffusion by measuring the change in apparent 1D diffusion ($d\log D$) with respect to salt concentration ($d\log D/d\log[\text{salt}]$), for TALEs ranging in length from 0 to 21.5 repeats (**Fig. 2f**). The quantity $d\log D/d\log[\text{salt}]$ provides a measure of the salt dependence of TALE 1D diffusion. Previous single-molecule studies of DNA-binding proteins have used this quantity to estimate roughly the number of charged residues contacting the DNA backbone during nonspecific search[25]. Moreover, the dependence of protein binding affinity (K_a) on salt concentration ($d\log(K_a)/d\log[\text{salt}]$) has been used in the context of counterion condensation theory to provide an estimate for the number of charged residues contacting the DNA backbone during specific binding of proteins to DNA[26,27]. In this way, the ionic-strength dependence of TALE protein diffusion serves as a proxy to describe the number of charged residues engaged in the search process, thereby elucidating the role of TALE CRD repeats in nonspecific search. If the residues within the TALE repeat region (CRD) are not involved in the search process, a change in the number of repeats within the CRD should not affect $d\log D/d\log[\text{salt}]$. But if TALE repeats are involved in nonspecific search, then an increase in the number of charged residues via longer CRDs would induce larger changes in ionic-strength dependence, thereby resulting in larger values of $d\log D/d\log[\text{salt}]$. Our results show a nearly linear increase in the magnitude of $d\log D/d\log[\text{salt}]$ upon increase of the CRD size (**Fig. 2f**). These results support a model in which TALEs are wrapped around or in close contact with DNA during nonspecific search.

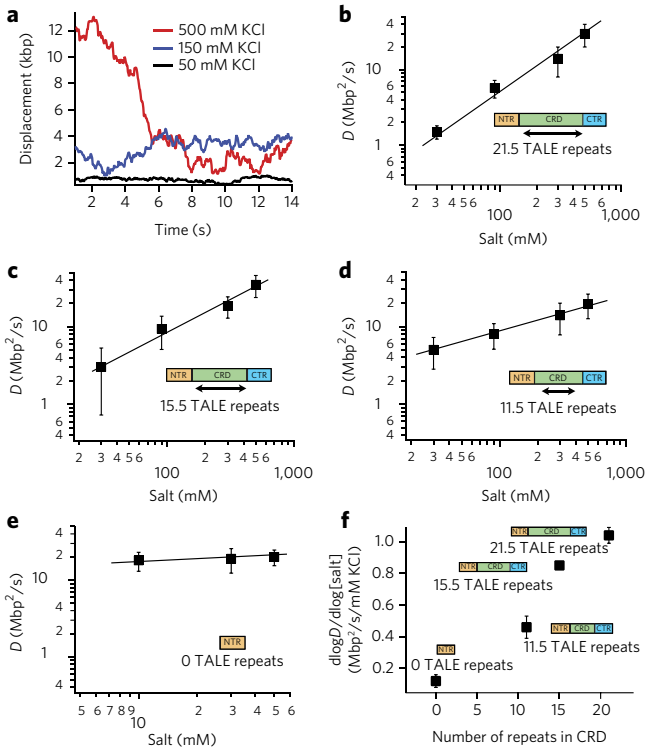

Figure 2 | Effects of ionic strength on TALE diffusion. (**a**) Sample trajectories of a 21.5-repeat TALE at 50 mM KCl (black), 150 mM (blue) and 500 mM KCl (red). Experiments included 22, 19, and 17 molecular trajectories recorded for these conditions, respectively. (**b**) Salt-dependent change in diffusion for the 21.5-repeat TALE. The slope of the linear fit is approximately equal to unity for this TALE variant. Data are mean ± s.d. for n = 24, 38, 41, and 17 measurements at 30, 90, 300, and 500 mM KCl, respectively. (**c**–**e**) Salt-dependent diffusion for TALEs with 15.5 (**c**), 11.5 (**d**), and 0 repeats (corresponding to NTR-only mutant) (**e**). In **c**, n = 36, 20, 18, and 34 for data points at 30, 90, 300, and 500 mM KCl, respectively; in **d**, n = 11, 33, 25, and 20 for data points at 30, 90, 300, and 500 mM KCl, respectively; in **e**, n = 28, 25, and 21 for data points at 10, 30, and 50 mM KCl, respectively. Error is reported as the s.d. between all measured diffusion coefficients for each condition (**c**–**e**). (**f**) Salt-dependent diffusion for TALEs with variable repeats, plotted as the slope of data in **b**–**e**. Error bars in **f** represent the standard error of regression.

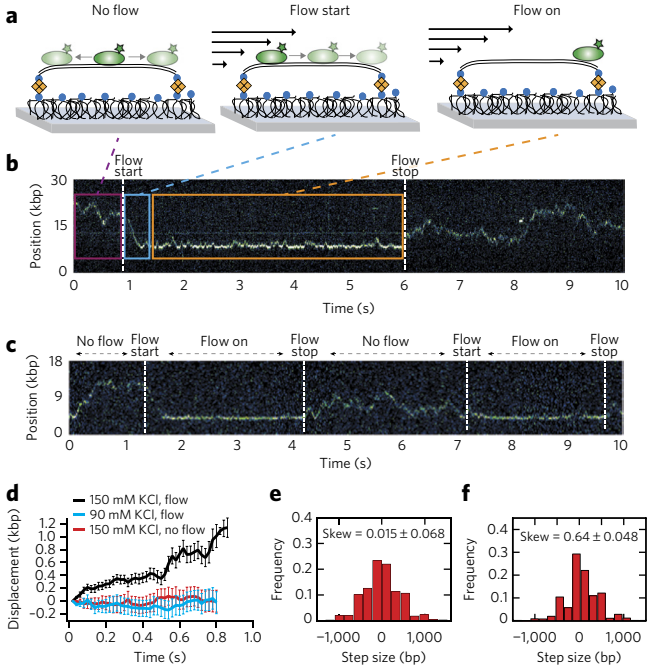

Figure 3 | Hydrodynamic flow assay for probing TALE diffusion along DNA. (**a**) Schematic of assay. Fluid flow is applied, corresponding to a linear flow rate of ~25 µm/s in the vicinity of the DNA (**Supplementary Fig. 7**). Green ovals represent diffusing TALE proteins; green stars represent fluorescent labels; blue circles represent biotin molecules; orange diamonds represent NeutrAvidin molecules; arrows represent the applied hydrodynamic flow. (**b**,**c**) Kymographs of TALE motion show the ability to reversibly bias TALE 1D diffusion using flow for TALEs containing 21.5 (**b**) and 15.5 (**c**) repeats at 500 mM KCl. Here, a single TALE diffuses along DNA in the absence of flow. Upon application of flow, single TALEs can be pushed along DNA in the direction of flow. Upon halting of the flow, TALEs resume unbiased Brownian search. (**d**) Diffusional bias at lower ionic strength, plotted as the average displacement in the absence of flow at physiological ionic strength (red, 150 mM KCl, n = 31 trajectories) and in the presence of flow at low (blue, 90 mM KCl, n = 36 trajectories) or physiological (black, 150 mM KCl, n = 29 trajectories) ionic strength. Error bars represent mean ± s.e.m. (**e**,**f**) Distributions of frame-by-frame displacements (step sizes) of TALE trajectories in the absence (**e**) and presence (**f**) of flow at 150 mM KCl (n = 3,218 from 31 trajectories (**e**) or 2,734 steps from 29 trajectories (**f**)). Bias was quantified by calculating sample skewness and comparing it to the standard error of skewness.

TALEs encircle DNA during search

We next sought to reconcile how TALE diffusion could be described by hopping behavior. Hopping behavior is commonly associated with rapid dissociation–reassociation events in which the protein dissociates from the DNA then re-binds a short distance from the site of dissociation[28,29]; however, this mechanism seems to be inconsistent with the superhelical structure of TALE proteins and our finding that the CRD repeats engage DNA during search. To study this further, we used hydrodynamic flow to determine whether TALE diffusion can be directionally biased during nonspecific search (**Fig. 3**). We applied laminar fluid flow in microfluidic channels containing a field of dual-tethered DNA templates in the presence of fluorescently labeled TALEs (**Fig. 3a** and **Supplementary Fig. 7**). In the vicinity of the surface, TALEs bound to DNA were subject to a simple shear flow with a flow rate of ~25 µm/s (**Supplementary Fig. 7**). Notably, individual TALE proteins bound to DNA were 'pushed' to the distal end of DNA templates under the action of flow in high salt conditions (500 mM KCl). TALEs remained bound to DNA and did not dissociate while being pushed along DNA templates in flow. Upon stopping the flow, TALEs bound to DNA immediately resumed directionally unbiased 1D diffusion, and this process could be repeated multiple times for the same molecule (**Fig. 3b**,**c**). This behavior was observed for all TALEs except NTR-only mutants, which lack a CRD. We found that the TALE NTR rapidly dissociated from the DNA template upon application of fluid flow and could not be 'pushed' along DNA using flow.

Using the hydrodynamic flow assay, we observed occasional unbinding and release of the TALE construct containing 11.5 repeats (6 out of 18 observed events) (**Supplementary Fig. 8**); however, our experiments showed no evidence of flow-induced release of TALEs containing 15.5 or 21.5 repeats (0 out of 17 and 0 out of 20 events, respectively). These observations are consistent with a model in which TALEs are wrapped loosely around DNA in a helical conformation during nonspecific search. It is known that TALEs bind to DNA in a 'one-repeat-to-one-nucleotide' manner, at least in the context of specific binding. From this perspective, the TALE construct containing 11.5 repeats is just long enough to completely encircle DNA within one helical turn of the protein.

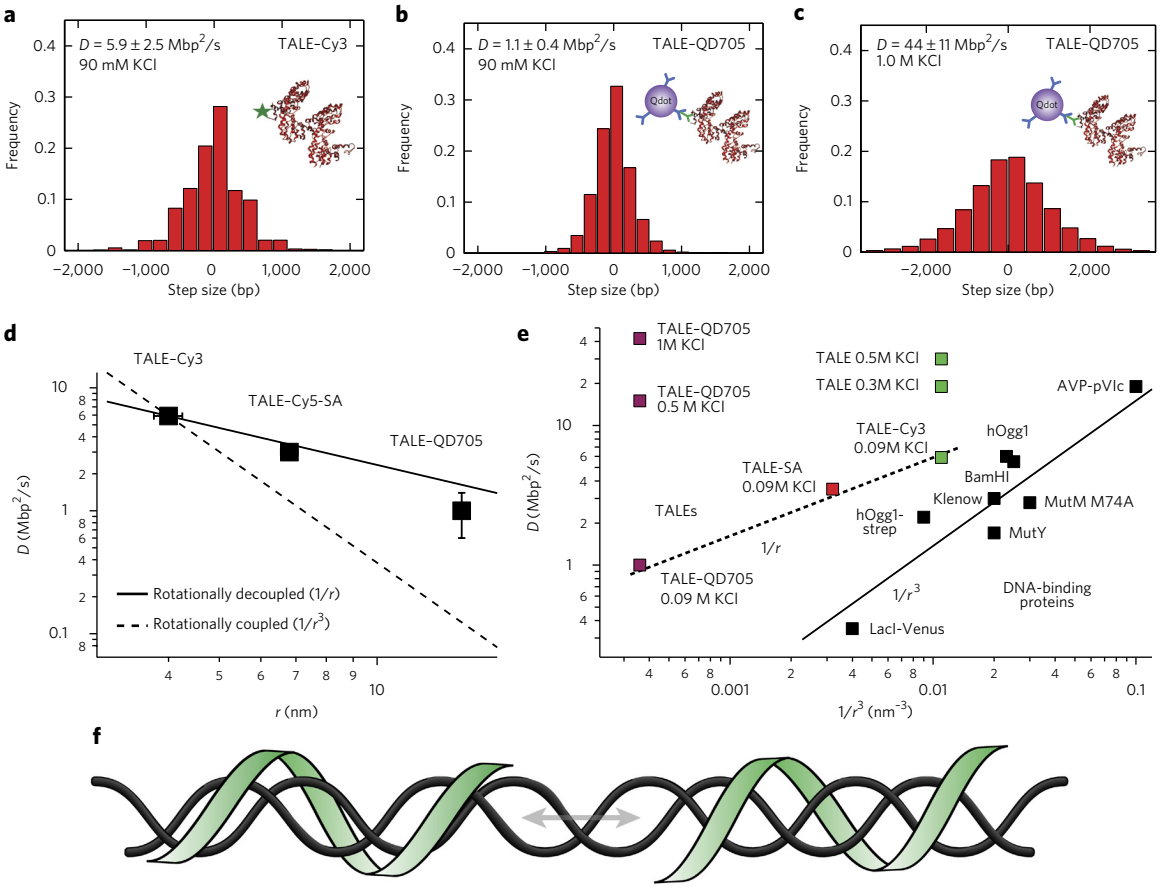

Figure 4 | Probe size dependence and mechanism of TALE diffusion along DNA. (**a**,**b**) Step size histograms for TALEs with 21.5 repeats, conjugated with Cy3 dye (**a**) or QD705 (**b**), at 90 mM KCl. n = 3,264 steps from 20 traces (**a**) or 5,530 steps from 16 traces (**b**). (**c**) Step size histograms for TALEs with 21.5 repeats at 1.0 M KCl conjugated with QD705 (n = 16,943 steps from 30 traces). Error is reported as the s.d. between all measured diffusion coefficients for each condition (**a**–**c**). (**d**) Scaling of the apparent 1D diffusion coefficients for TALEs as a function of probe size. The experimental scaling of the TALE diffusion coefficient is shown in comparison to scaling for rotationally coupled and rotationally decoupled diffusion. Error bars represent the s.d. between all measured diffusion coefficients for each condition (n = 20, 15, and 16 for Cy3, Cy5-SA, and QD705 conjugates, respectively). (**e**) Size-dependent scaling of diffusion coefficients for several DNA-binding proteins measured using single-molecule techniques at salt concentrations ranging 2 to 25 mM (refs. 14,33) overplotted with TALE diffusion coefficients measured over a range of salt concentrations. These data correspond to the data used in **Figures 2a**,**b** and **4a**–**d** and **Supplementary Figure 13**. (**f**) Schematic of rotationally decoupled TALE search mechanism. In this model, TALEs (green) remain associated with DNA templates (gray) in a loose-wrapped conformation around the DNA helix.

Therefore, although the 11.5-repeat TALE construct can be pushed along DNA templates in the presence of flow, it is more likely than the 15.5 and 21.5 repeat TALEs to be ejected during the application of flow.

We further investigated the impact of fluid flow on TALE dynamics by plotting the average displacement as a function of time for ensembles of single TALE trajectories at lower ionic strength (90 and 150 mM KCl) in the presence and absence of flow (**Fig. 3d**). Additionally, we quantified the presence of diffusional bias by determining the sample skewness of step size distributions (**Fig. 3e**,**f** and **Supplementary Fig. 9**). In the absence of flow, TALE motion was directionally unbiased (skew of 0.015 ± 0.068 (standard error of skewness (SES)) (**Fig. 3e**). Similarly, at low ionic strength (90 mM KCl) under applied flow, TALEs displayed no significant skew (0.010 ± 0.040) (**Supplementary Fig. 9**). In the presence of flow at physiological ionic strength (150 mM KCl), however, TALE diffusion was clearly biased in the direction of flow (skew of 0.64 ± 0.048) (**Fig. 3f**). Under these conditions, convection dominated over diffusion for unbound TALEs in solution near the surface, which can be quantified by the protein Peclet number (Pe = 10) for unbound TALEs in flow (Online Methods and **Supplementary Fig. 7**). For a model in which the protein spends some amount of time unbound from (or not tightly associated with) the DNA during its search, a convective flow would directionally bias the apparent 1D diffusion in proportion to the amount of time spent unbound from the DNA, which is consistent with our experiments[30]. Conversely, if the protein remained tightly coupled to DNA during search, there would not be a significant bias to its search process. Our findings from flow experiments support a search process in which TALEs remain wrapped around DNA yet are still able to dissociate effectively when the ionic contacts between positively charged residues and the DNA phosphate groups are temporarily broken. In this way, the TALEs are spatially confined to DNA templates but are able to 'hop' along DNA during 1D diffusion.

TALEs utilize a rotationally decoupled search process

With a clear picture of the TALE search process emerging, we sought to further understand the microscopic trajectories of TALEs during search. Given our findings, we hypothesized that the TALE search trajectory would not follow a helical path around DNA templates. To study this further, we generated a series of TALE proteins labeled with fluorescent probes of increasing size (**Fig. 4a–c** and **Supplementary Fig. 10**), a technique used to elucidate the search trajectories of other DNA-binding proteins[14,25,31]. For these

experiments, the parent TALE protein was designed to target a 22-bp DNA sequence (21.5 TALE repeats), and fluorescent probes were varied to include a single organic dye (Cy3), fluorescently labeled streptavidin (Cy5-SA), or a quantum dot (QD705) (**Fig. 4a–c** and **Supplementary Figs. 10–12**). Upon increasing probe size, TALEs showed a decrease in 1D diffusion that closely followed a $1/r$ dependence, where r is the protein–probe radius (**Fig. 4d**), and this trend persisted in multiple sets of solution conditions (**Supplementary Fig. 13**). According to previous models for diffusion of DNA-binding proteins, $1/r$ scaling suggests a rotationally decoupled trajectory, in which a protein traverses DNA without rotating, whereas $1/r^3$ scaling for 1D diffusion is consistent with a model in which a protein rotates around the DNA helix during its search[14,32]. These results suggest that TALEs do not tightly track the DNA helix during 1D search, which is consistent with our model.

TALE search is unique among DNA-binding proteins

We further compared the 1D diffusion behavior for TALE proteins to various DNA-binding proteins that are known to undergo target-site search along DNA. We found that TALE diffusion at high ionic strength showed extremely rapid 1D search with large diffusion coefficients (**Fig. 4c**). The QD705-labeled TALEs exceed the theoretical 1D rotationally coupled diffusion coefficients by nearly 200-fold at high ionic strength. These results further show that at high ionic strength, TALEs readily exceed the largest 1D diffusion coefficients reported for thermally driven DNA-binding proteins[33,34], which underscores the unique nature of the TALE search process. We compared the diffusive behavior of TALEs across a range of salt concentrations (0.09 to 1.0 M KCl) to that of several previously investigated[14,33] DNA-binding proteins (under conditions ranging from 0.002 to 0.025 M monovalent salt) undergoing nonspecific search (**Fig. 4e**). The majority of DNA-binding proteins studied previously diffuse along DNA with approximately $1/r^3$ scaling, in contrast to TALEs. Taken together, these data suggest that TALEs are loosely wrapped around DNA and do not fully track the DNA helix during search (**Fig. 4f**), in stark contrast to the broad class of transcription factors and DNA-binding proteins that rotate along the helical path of the DNA[35,36]. In this scenario, TALEs adopt a loose helical conformation in which they are wrapped around the DNA template but contact the phosphate backbone only briefly or transiently during search (**Fig. 4f**).

DISCUSSION

In recent years, researchers have probed the ionic-strength dependence of 1D diffusion rates to gain insight into the search mechanisms of DNA-binding proteins. The observation of an ionic-strength-dependent diffusion rate is generally cited as evidence of a hopping mechanism for DNA search, whereas the absence of salt-dependent diffusion is generally considered as evidence supporting a sliding mechanism. From a mechanistic perspective, protein hopping involves continual cycles of binding and unbinding to DNA templates during the search process. Hopping is thus thought to be salt dependent, because the binding of the protein after a 'hop' is strongly affected by the surrounding ion cloud and the number of ions that must be displaced for binding to occur. In this model, an increase in ionic strength would decrease the affinity of the protein for the DNA template, thereby increasing both the time the protein spends in free solution and the apparent 1D diffusion speed. Conversely, sliding behavior is thought to be independent of ionic strength, because the displacement of ions during sliding along the DNA helix is offset by the rebinding of ions at the site previously occupied by the protein. How can a single protein show evidence of both hopping and sliding during nonspecific search along DNA? A few studies have considered protein search behavior that can be described by both of these ideal mechanistic pictures[35,37–39]. Recent experimental[38] and molecular dynamics[39] studies have suggested that ionic-strength dependence may support a quasi-sliding mechanism that is not well described by the strict classifications of sliding or hopping behavior, and we show here that the TALE search process can be described by a hybrid of the two models.

Our data demonstrate a strong ionic-strength dependence of TALE diffusion; however, flow-based single-molecule experiments show that TALEs remain associated with DNA templates in the presence of strong convective flow during long-range search. These results cannot be reconciled with the classic hopping mechanism for nonspecific search. Moreover, the probe-size scaling of 1D diffusion coefficients for TALEs does not support a sliding mechanism for search. Taken together, our data indicate that TALE search behavior cannot be described by the traditional models for nonspecific search along DNA. Our results point to a superhelical protein structure that effectively wraps the protein around the DNA double helix while avoiding a pure sliding mechanism. Our results on the relationship between salt sensitivity and TALE CRD size suggest that the residues within the CRD are directly affected by charge-screening and electrostatic interactions in the system. These data are consistent with a superhelical protein wrapped around a DNA template without the tight threading associated with major groove binding. In this model, electrostatic interactions between individual residues in CRD repeats and DNA phosphate groups can be broken or reformed depending on the ionic strength of solution. At high ionic strength, several of these interactions may be temporarily disrupted, resulting in an effectively loosely bound TALE protein nevertheless associated with a DNA template due to the superhelical protein conformation. In general, the ionic strength of the solution in part determines the strength of TALE residue interactions with the DNA backbone, and thus acts in effect to smooth the energetic barriers to 1D diffusion. Finally, our results are consistent with prior studies suggesting commonalities between the sliding and hopping models of protein search[37–39]. These studies suggested that as electrostatic interactions between DNA and proteins are weakened by increased ionic strength, the energetic barrier to sliding decreases and the speed of apparent sliding thus increases. This argument is likely to be valid for TALEs, which appear to be wrapped around DNA during search and mediate electrostatic interactions with DNA that are susceptible to screening.

Rotationally decoupled search has been reported for other DNA-binding proteins undergoing nonspecific search, albeit in the context of vastly different protein function[25,40,41]. The eukaryotic proliferating cell nuclear antigen (PCNA) forms a ring-like structure around DNA and serves as a processivity factor for ε and δ polymerases among other functions, none of which, however, requires sequence-specific binding. Single-molecule investigation of PCNA dynamics revealed that this protein alternates between rotationally coupled and decoupled 1D diffusion, described by a rapid (D_{1D} values of 10 Mbp^2/s), rotationally decoupled sliding–1D hopping mechanism[25]. The eukaryotic mismatch repair (MMR) protein MutSα undergoes a conformational switch upon mismatch recognition and subsequent ATP-mediated release from mismatch sites[41]. Interestingly, the conformational change transitions MutSα from a rotationally coupled 1D trajectory along DNA to a rotationally decoupled trajectory. Similar behavior was also observed for Taq MutS[30]. Moreover, the type III restriction enzyme EcoP15I was also observed to diffuse rapidly in a rotationally decoupled trajectory (D_{1D} values of 8 Mbp^2/s) along extended DNA templates after target-site binding and a subsequent ATP-driven conformational switch[40]. Common among these DNA-binding proteins is the absence of sequence-specific binding after transitioning to rotationally decoupled diffusive paths, whereas TALEs, in contrast, appear to utilize this mechanism to locate target sites for sequence-specific binding.

Tumor suppressor p53 was recently studied via molecular dynamics simulations, and its search mechanism bears some similarities

to that of TALEs. Simulations of p53 nonspecific search on DNA under physiological ionic-strength conditions revealed rotationally decoupled trajectories mediated by the protein C-terminal domain (CTD), with tethered hopping events initiated by the core domain[39]. Nevertheless, the p53 core domain (and not the p53 CTD) is responsible for sequence-specific binding. TALEs thus appear to be unique in that the protein domains responsible for sequence-specific binding follow a rotationally decoupled trajectory during target search. However, although rotationally decoupled diffusion (which can be described as 1D hopping[41] and 2D sliding[42]) allows faster 1D search than rotationally coupled diffusion, this search mechanism results in a protein being situated out of phase with respect to the helical pitch of the DNA, thereby hindering the protein's ability to read the local sequence. In previous work we observed that TALEs utilize a two-state model for sequence search and conjectured that TALEs compress along their helical axis during a specific binding event, which enables them to 'check' the local sequence[15]. This model is consistent with a rotationally decoupled search mechanism and allows TALEs to retain their sequence-specific binding activity while enabling rapid translocation interspersed between sequence 'checking' events, thereby satisfying the search speed–stability paradox[43].

Our results support a model wherein TALEs adopt a loosely wrapped conformation around DNA templates during nonspecific search. These findings and the robust binding and search abilities of TALEs suggest that design of TALE fusion proteins could be expanded to other applications. For example, it might be possible to generate chimeric proteins by fusing TALEs with larger 'payloads' or active domains. Finally, the rotationally decoupled model is consistent with a two-state model for the TALE search process[15]. During search mode, the loosely wrapped conformation mediates rapid 1D nonspecific search. During recognition mode, TALEs are readily able to compress their superhelical pitch, which has been previously proposed as a mechanism of target-site binding[44,45]. Upon encountering a putative target site, TALEs can undergo a conformational change to compress and 'check' the local sequence, possibly resulting in stable binding. Additional molecular-based studies with higher spatial resolution will permit a dynamic view of the detailed conformational changes occurring during transitions between these modes.

Received 1 December 2015; accepted 27 May 2016;
published online 15 August 2016

METHODS

Methods and any associated references are available in the online version of the paper.

References

1. Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).
2. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
3. Bedell, V.M. *et al. In vivo* genome editing using a high-efficiency TALEN system. *Nature* **491**, 114–118 (2012).
4. Ma, N. *et al.* Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free β-thalassemia induced pluripotent stem cells. *J. Biol. Chem.* **288**, 34671–34679 (2013).
5. Berdien, B., Mock, U., Atanackovic, D. & Fehse, B. TALEN-mediated editing of endogenous T-cell receptors facilitates efficient reprogramming of T lymphocytes by lentiviral gene transfer. *Gene Ther.* **21**, 539–548 (2014).
6. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720–723 (2012).
7. Mak, A.N.-S., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716–719 (2012).
8. Boch, J. & Bonas, U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
9. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* **22**, 1716–1720 (2012).
10. Bogdanove, A.J., Schornack, S. & Lahaye, T. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* **13**, 394–401 (2010).
11. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
12. Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843–1846 (2011).
13. Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061–2065 (2003).
14. Blainey, P.C. *et al.* Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16**, 1224–1229 (2009).
15. Cuculis, L., Abil, Z., Zhao, H. & Schroeder, C.M. Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.* **6**, 7277 (2015).
16. Yardimci, H., Loveland, A.B., van Oijen, A.M. & Walter, J.C. Single-molecule analysis of DNA replication in *Xenopus* egg extracts. *Methods* **57**, 179–186 (2012).
17. Sun, N., Liang, J., Abil, Z. & Zhao, H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. Biosyst.* **8**, 1255–1263 (2012).
18. Richter, A., Streubel, J. & Boch, J. TAL effector DNA-binding principles and specificity. *Methods Mol. Biol.* **1338**, 9–25 (2016).
19. Mussolino, C. & Cathomen, T. TALE nucleases: tailored genome engineering made easy. *Curr. Opin. Biotechnol.* **23**, 644–650 (2012).
20. Miller, J.C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2011).
21. Carrico, I.S., Carlson, B.L. & Bertozzi, C.R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–322 (2007).
22. Shi, X. *et al.* Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. *Nat. Methods* **9**, 499–503 (2012).
23. Vestergaard, C.L., Blainey, P.C. & Flyvbjerg, H. Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E* **89**, 022726 (2014).
24. Halford, S.E. & Marko, J.F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
25. Kochaniak, A.B. *et al.* Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.* **284**, 17700–17710 (2009).
26. Cravens, S.L., Hobson, M. & Stivers, J.T. Electrostatic properties of complexes along a DNA glycosylase damage search pathway. *Biochemistry* **53**, 7680–7692 (2014).
27. Etson, C.M., Hamdan, S.M., Richardson, C.C. & van Oijen, A.M. Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. *Proc. Natl. Acad. Sci. USA* **107**, 1900–1905 (2010).
28. Komazin-Meredith, G., Mirchev, R., Golan, D.E., van Oijen, A.M. & Coen, D.M. Hopping of a processivity factor on DNA revealed by single-molecule assays of diffusion. *Proc. Natl. Acad. Sci. USA* **105**, 10721–10726 (2008).
29. Berg, O.G., Winter, R.B. & von Hippel, P.H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
30. Cho, W.K. *et al.* ATP alters the diffusion mechanics of MutS on mismatched DNA. *Structure* **20**, 1264–1274 (2012).
31. Dikić, J. *et al.* The rotation-coupled sliding of EcoRV. *Nucleic Acids Res.* **40**, 4064–4070 (2012).
32. Schurr, J.M. The one-dimensional diffusion coefficient of proteins absorbed on DNA hydrodynamic considerations. *Biophys. Chem.* **9**, 413–414 (1975).
33. Blainey, P.C. *et al.* Regulation of a viral proteinase by a peptide and DNA in one-dimensional space: IV. Viral proteinase slides along DNA to locate and process its substrates. *J. Biol. Chem.* **288**, 2092–2102 (2013).
34. Xiong, K. & Blainey, P.C. Molecular sled sequences are common in mammalian proteins. *Nucleic Acids Res.* **44**, 2266–2273 (2016).
35. Tafvizi, A., Huang, F., Fersht, A.R., Mirny, L.A. & van Oijen, A.M. A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. USA* **108**, 563–568 (2011).
36. Dunn, A.R., Kad, N.M., Nelson, S.R., Warshaw, D.M. & Wallace, S.S. Single Qdot-labeled glycosylase molecules use a wedge amino acid to probe for lesions while scanning along DNA. *Nucleic Acids Res.* **39**, 7487–7498 (2011).
37. Barbi, M. & Paillusson, F. in *Advances in Protein Chemistry and Structural Biology* Vol. 92 (ed. Karabencheva-Christova, T.) 253–297 (Elsevier, 2013).
38. Esadze, A., Kemme, C.A., Kolomeisky, A.B. & Iwahara, J. Positive and negative impacts of nonspecific sites during target location by a sequence-specific DNA-binding protein: origin of the optimal search at physiological ionic strength. *Nucleic Acids Res.* **42**, 7039–7046 (2014).

39. Terakawa, T., Kenzaki, H. & Takada, S. p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains. *J. Am. Chem. Soc.* **134**, 14555–14562 (2012).
40. Schwarz, F.W. *et al.* The helicase-like domains of type III restriction enzymes trigger long-range diffusion along DNA. *Science* **340**, 353–356 (2013).
41. Gorman, J. *et al.* Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc. Natl. Acad. Sci. USA* **109**, E3074–E3083 (2012).
42. Kampmann, M. Obstacle bypass in protein motion along DNA by two-dimensional rather than one-dimensional sliding. *J. Biol. Chem.* **279**, 38715–38720 (2004).
43. Slutsky, M. & Mirny, L.A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
44. Lei, H., Sun, J., Baldwin, E.P., Segal, D.J. & Duan, Y. in *Advances in Protein Chemistry and Structural Biology* Vol. 94 (ed. Donev, R.) 347–364 (Elsevier, 2014).
45. Schreiber, T. *et al.* Refined requirements for protein regions important for activity of the TALE AvrBs3. *PLoS One* **10**, e0120214 (2015).

Acknowledgments

We thank T. Ha (University of Illinois Urbana–Champaign) for providing the plasmid for the aldehyde labeling scheme and S. Li for assistance in acquiring transmission electron microscope images of quantum dots. C.M.S. is funded by the David and Lucile Packard Foundation. H.Z. and L.C. are funded by the Institute for Genomic Biology at the University of Illinois at Urbana–Champaign. L.C. is funded by the FMC Corporation.

Author contributions

L.C., Z.A., H.Z., and C.M.S. designed experiments. Z.A. generated protein and DNA samples. L.C. performed single-molecule experiments and data analysis. L.C., Z.A., H.Z., and C.M.S. prepared the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to C.M.S.

ONLINE METHODS

Preparation of DNA templates. The DNA template for nonspecific TALE search experiments (containing no TALE binding sites) is a 44,898-bp plasmid encoding the neoaureothin synthesis pathway from *Streptomyces orinoci*. This plasmid was purified using the Plasmid Midi Kit (Qiagen) and linearized with SnaBI enzyme (NEB) for 3 h, followed by additional purification using phenol–chloroform extraction and ethanol precipitation. The purified plasmid was subjected to the 3′–5′ exonuclease activity of T4 DNA polymerase (NEB) to generate 5′ overhangs at both ends of the linearized template. 10 µg DNA was treated with 5 units of T4 DNA polymerase in NEB buffer 2 (supplemented with BSA) at 25 °C for 1 h. The reaction was halted with 10 µL 20 mM dCTP, followed by heat inactivation of the enzyme at 75 °C for 20 min. The exposed 5′ overhangs were used to anneal 3′-biotinylated oligonucleotides. The oligonucleotide with the sequence 5′-CAGCAGTTCAACCTGTTGATAGTAC-Biotin TEG (triethylene glycol)-3′ (IDT) was added in 50× molar excess to the DNA (concentration of intact templates), annealed by heating at 90 °C for 5 min, then cooled to 4 °C. The second oligonucleotide, with sequence 5′-TACGTGAAACATGAGAGCTTAGTAC-Biotin TEG-3′, was then annealed in the microfluidic flow device, as described below.

Preparation and cloning of tSCA21.5 and and N-terminal aldehyde tagged (Naldt)-tSCA21.5. The gene encoding the TALE protein (tSCA21.5) lacking the aldehyde tag was assembled using the Golden Gate cloning method (Addgene TALEN Kit 1000000024) as described in prior work[46]. Here, repeats 1–10 of the TALE were assembled into the pFUS_A30A vector, repeats 11–20 into the pFUS_A30B vector, and repeat 21 into the pFUS_B1 vector using the BsaI-HF (NEB) and T4 DNA ligase (NEB) by cycling the temperature ten times between 5 min at 37 °C and 10 min at 16 °C. The reaction was then treated with Plasmid-Safe nuclease (Epicentre Biotechnologies). DH5α (Cell Media Facility at UIUC) transformants were assayed for correct assembly via restriction digestion of the purified (Qiagen) plasmids, and together with the plasmid encoding the last half-repeat pLR(NG), further assembled into the destination vector, pET28-GG-TALE using BsmBI (Fisher Scientific) and T4 DNA ligase. This vector contained an N-terminal hexahistidine (6× His) tag and flanking N-terminal (208 aa) and C-terminal (63 aa) regions of the TALE, as well as the BsmBI sites corresponding to the kit BsmBI sites (sequence below). In order to fluorescently label TALEs, we modified the original plasmid pET-tSCA21.5 with an oligonucleotide insert encoding an N-terminal LCPTSR hexapeptide (aldehyde tag)[21,22] upstream of the 6× His tag to generate Naldt-tSCA21.5. Here, the plasmid was amplified in fragments containing the insert, and assembled using the Gibson Assembly Kit (NEB).

Primary structure of Naldt-tSCA21. The amino acid sequence of the aldehyde-tagged 21.5 repeat TALE (Naldt-tSCA21) is as follows: MGPLCTPSR SSHHHHHHSSGLVPRGSHMLDTSLLDSMPAVGTPHTAAAPAECDEVQ SGLRAADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRT LGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALG TVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALLTEAGELRGPPL QLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASN IGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV VAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQAL ETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH GLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVL CQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLL PVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVL CQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDH LVALACLGGRPALDAVKKGLPHAPELIRRRINRRIPERTSHRVA.

Preparation and cloning Naldt-tSCA15.5 and Naldt-tSCA11.5. In order to generate TALEs with the first 15.5 and 11.5 repeats of the 21.5 repeat construct (Naldt-tSCA15.5 and Nald-tSCA11.5, respectively), we designed a destination vector (pET28-Naldt-GG-TALE) containing the LCTPSR aldehyde tag upstream of the 6× His tag, using the Gibson Assembly Master Mix (NEB). Naldt-tSCA15.5 and Nald-tSCA11.5 were assembled into the destination vector using Golden Gate cloning[46]. Here, the protocol for generating the 21.5-repeat TALE was followed as described above, but repeats 11–15 were inserted in place of repeats 11–20. For the 11.5-repeat TALE, repeat 11 was inserted into vector pFUS-B1 using the BsaI-HF and T4 DNA ligase by cycling the temperature ten times between 5 min at 37 °C and 10 min at 16 °C. The reactions were then treated with Plasmid-Safe nuclease. DH5α transformants were assayed for correct assembly using restriction digestion of purified plasmids. The vector (pFUSA) containing the first ten repeats was assembled with either pFUSB5 (for the 15.5-repeat TALE) or pFUSB1 (for the 11.5-repeat TALE) together with the plasmid encoding the last half-repeat into the pET28-GG-TALE vector using BsmBI and T4 DNA ligase.

Cloning of NTR-only mutant (Naldt-NTR). To generate Naldt-NTR, we first digested pET-Naldt-tSCA21.5 with SacII and HindIII and then inserted the following sequence: CAAAGCGTGGTGGCGTGACCGCGGTGGAAG CGGTCCATGCCTGGCGTAATGCGTTGACGGGCGCCCCCCTGAAC TAAGTCAGATAACCGGATACAGACAAGCTTGCGGCCGCACTCGAGC ACCAC.

This sequence was synthesized as a gBlock Gene Fragment (IDT) using Gibson Assembly Master Mix (NEB).

Primary structure of Naldt-NTR. The amino acid sequence of Naldt-NTR is as follows: MGPLCTPSRSSHHHHHHSSGLVPRGSHMLDTSLLDSMPAV GTPHTAAAPAECDEVQSGLRAADDPPPTVRVAVTAARPPRAKPAPRR RAAQPSDASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHHEALVGH GFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWS GARALEALLTEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRN ALTGAPLN.

Protein expression, purification, and labeling. BL21 (DE3) electrocompetent *Escherichia coli* cells were co-transformed with plasmids encoding TALEs and the pBAD-FGE plasmid (gift of T. Ha, University of Illinois at Urbana–Champaign). One colony was grown in 5 mL LB, spiked with 100 µg/mL ampicillin and 25 µg/mL kanamycin as a seeding culture until saturation, and subsequently in 500 ml Terrific Broth at 37 °C and 250 r.p.m. with the corresponding antibiotics to an OD_{600} of 0.3–0.4. FGE expression was induced with 0.2% L-arabinose (Sigma-Aldrich). The culture was grown at 37 °C 250 r.p.m. to OD_{600} 0.7–0.8. TALE expression was induced via 0.4 mM IPTG (Sigma-Aldrich). TALEs were expressed overnight in an incubated shaker, shaking at 250 r.p.m., while the temperature was held at 16 °C. Next, cell cultures were centrifuged at 4,000 r.p.m. for 15 min while the temperature was held at 4 °C. Resulting pellets were then resuspended in 10–20 mL lysis buffer (25 mM Tris-HCl (Fisher), pH 7.5, 300 mM NaCl (Fisher), 0.5% Triton X-100 (Sigma-Aldrich), 5% glycerol (Sigma-Aldrich), 4 unit/mL DNase I (NEB), 0.3 mM PMSF (Sigma-Aldrich), 1 mg/mL lysozyme (Sigma-Aldrich)). Cells were then lysed via sonication for 20 min. The resulting debris was then centrifuged at 13,000 r.p.m. for 20 min with temperature held at 4 °C. TALE proteins (with 6× His tag) were purified via AKTA chromatography system (GE Healthcare) with a 1 mL HisTrap column (GE Healthcare). The cell lysate was loaded onto the column and then washed with 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 20 mM imidazole (Sigma-Aldrich), followed by elution with 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 250 mM imidazole. Purified TALEs were dialyzed overnight (at 4 °C) against 2 L 50 mM phosphate buffer, pH 7–8.4 (where the pH varied depending on the estimated isoelectric point of the TALE construct in question) supplemented with 500 mM NaCl. If needed, TALEs were further purified using a Superdex 16/600 200 preparatory grade gel filtration column (GE Healthcare) using the phosphate buffer described above. Purified TALEs were assayed by SDS–PAGE (**Supplementary Fig. 3**). TALEs with 11.5, 15.5, and 21.5 repeats were stable for at least 12 months when stored in buffered

doi:10.1038/nchembio.2152

solutions in 15% glycerol at −80 °C. Purified NTR-only mutant was found to easily aggregate to form dimers and trimers (**Supplementary Fig. 3b**) and often precipitated from solution. NTR (25 kDa) was further separated and purified from the 50 kDa and 75 kDa aggregates via gel filtration. NTR was stored at low concentrations (<1 mg/mL) in 50 mM phosphate buffer, pH 7, 15% glycerol at −80 °C until use, and new aliquots were thawed for each experiment.

Protein labeling with Cy3. Purified TALEs (>90% purity as assayed via SDS–PAGE) were buffer exchanged into labeling buffer (250 mM potassium phosphate, pH 6, 500 mM KCl (Fisher), 5 mM dithiothreitol (Roche)) and concentrated (to 100–300 μM) using Amicon Ultra 0.5 mL centrifugal spin columns (EMD Millipore). Concentrated TALE solutions were used to resuspend 1 mg Cy3-hydrazide (GE Healthcare), and labeling reactions were allowed to proceed for 24 h at room temperature in the dark. Labeled TALEs were diluted with 400 μL fluorescence anisotropy buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.5 mM EDTA (Fisher)) and subsequently purified from unreacted Cy3 using Micro Bio-Spin 6 columns (Bio-Rad). The concentration of resulting Cy3-labeled TALE was determined spectrophotometrically via Bradford assay (Thermo) with a NanoDrop 1000 (Thermo).

Protein labeling with Cy5-SA and QD705. Biotin-functionalized TALEs were generated using the same protocol followed for Cy3 labeling, but the Cy3 hydrazide was replaced with (+)-biotinamidohexanoic acid hydrazide (Sigma-Aldrich). Cy5-SA was reacted with biotin–TALEs via incubation with twofold molar excess Cy5-SA for 15 min. Two separate methods were used for generation of QD705–TALE conjugates. In the first method, streptavidin-coated QD705 (Invitrogen) was linked to biotin–TALE (as generated above) by incubation with fivefold molar excess QD for 15 min. In the second method, goat anti-mouse QD705 conjugates (Invitrogen Q11062MP) were coated with mouse anti-His antibodies (Genscript A00186-100) in a 1:5 ratio. The resulting QD705 anti-His conjugates were incubated in a 1:1 ratio with TALEs for 15 min on ice before imaging.

Fluorescence anisotropy. Binding abilities of labeled and unlabeled TALEs were compared using a fluorescence anisotropy assay to ensure that the biological function of TALEs was not perturbed by dye labeling. A 29-bp oligonucleotide with the 23-bp TALE binding site was labeled at its 5′ end with 5,6-FAM-SE (fluorescein) (IDT) and annealed with its reverse-complementary sequence in annealing buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8, 50 mM NaCl). The sequence of the TALE-binding oligo is: 5,6-FAM-ATCTAGCAACCTCAAACAGACACCATACG-3′. Fluorescein-labeled DNA templates were mixed with different concentrations of proteins in fluorescence anisotropy buffer. Fluorescence polarization measurements were taken on Infinite 200 Pro microplate reader (Tecan) using excitation and emission wavelengths of 485 nm and 535 nm, respectively, in black 96-well plates (Corning). Fluorescence polarization (P) values were converted to anisotropy (A) values with the equation

$$A = \frac{2P}{3 - P} \quad (1)$$

Flow-cell preparation. Microfluidic flow cells were constructed by placing two pieces of double-sided tape parallel between a drilled quartz microscope slide and functionalized glass coverslip (No. 1.5, Ted Pella) to form a flow channel (approximately 50 mm × 4 mm × 0.5 mm). Coverslips were passivated with a 50:1 PEG/PEG–biotin mixture (Laysan BioSciences) and NeutrAvidin (Pierce) before formation of the flow cells, which allowed specific surface attachment of DNA and reduction of nonspecific protein adsorption. Here, glass coverslips were cleaned by successive KOH–ethanol sonication followed by functionalization with aminopropyltriethoxysilane (Alfa-Aesar) for 15 min, then washing and drying. Coverslips were then incubated with a 10% (w/v) solution of 100:1 methoxy-PEG5000/PEG5000–biotin for 4 h and washed to create a polymer brush[47]. Polyethylene tubing (PE60, Solomon Scientific) was affixed to ports (0.048-inch outer diameter) drilled in each end of the quartz microscope slides to allow exchange of buffer solutions and extension of tethered DNA templates.

Single-molecule assay and preparation of dual-tethered DNA templates. Flow-cell surfaces were initially incubated with blocking solution (50 mM MOPS, 50 mM KCl, 0.1 mM EDTA, 0.25 mg/mL BSA, pH 8.1) for 10 min and then incubated with 5 pM biotin-functionalized DNA for 45 min. Unbound DNA was removed from the flow cell by flushing the chamber with blocking buffer for 5 min under gentle flow. Next, double-tethered DNA was formed by flowing in 150 nM biotinylated oligo (IDT) at a rate of 120 μL/min. This oligonucleotide was complementary to the single-stranded overhangs on the untethered DNA ends and was added in the presence of 100 μM chloroquine (Sigma-Aldrich). Chloroquine intercalates between DNA bases and extends templates to ~90–95% of their contour length under flow, which reduces substrate fluctuation during single-molecule imaging[16]. Chloroquine was then removed by washing the flow cell with high-salt blocking solution (containing 40 mM $MgCl_2$ and 500 mM NaCl) for 5 min under gentle flow.

Single-molecule imaging. TALE–DNA interactions were imaged using an inverted microscope (Nikon IX70) with a 100× oil-immersion objective (1.4 NA, Olympus) coupled to an EMCCD camera (Andor iXon Ultra 897). Cy3-labeled proteins were illuminated via 532-nm diode-pumped solid state (DPSS) laser (CrystaLaser), and SYTOX Green–labeled DNA was illuminated via 488-nm DPSS laser (SpectraPhysics Excelsior). QD705- and Cy5-SA-labeled TALEs were illuminated via 637-nm DPSS laser (Coherent). Images were acquired by electron multiplying charge-coupled device (EMCCD) camera at 33–60 Hz. TALEs were added to the flow cell at concentrations ranging from 25 to 1,000 pM, always in buffers containing 150 mM or less KCl. For high-salt experiments, once initial TALE binding was achieved, high-salt buffers were titrated into the flow cells. 5 μM free biotin (Sigma-Aldrich) was also added for imaging of Cy5-SA or QD705 conjugates to minimize nonspecific binding of these TALE conjugates to the chamber surface. In addition, 7 mM β-mercaptoethanol (Sigma-Aldrich) and oxygen-scavenging system (glucose oxidase and bovine liver catalase, Sigma-Aldrich) along with 1% v/v glucose (Sigma-Aldrich) were added to the imaging buffer to minimize photobleaching. For Cy5-SA experiments, 1 mM methyl viologen (Sigma-Aldrich) and 1 mM ascorbic acid (Sigma-Aldrich) were also added to imaging solutions.

Single-molecule data analysis. Images were collected as 16-bit TIF stacks using the Andor Solis software and EMCCD camera. Single diffusing TALE proteins were isolated using ImageJ software, and the centroid locations of individual TALEs were determined using RapidStorm fitting software[48]. Single TALE diffusion traces were further analyzed using custom MATLAB scripts.

Localizing single TALE proteins via fluorescence microscopy. There are two main sources of error in localization of individual TALEs: (i) uncertainty resulting from the finite number of photons collected during image acquisition, which translates to localization precision, and (ii) thermal fluctuations in the DNA templates parallel to the direction of TALE diffusion, which translate to localization accuracy. In our experiments, we used an acquisition time of 20–33 ms per frame, during which we collected ~70–100 photons. Our imaging assay also showed an extremely low fluorescence background owing to total internal reflection (TIR) imaging. On the basis of these conditions, the localization precision of single TALE proteins can be estimated as ~25 nm. Considering the second source of error, the magnitude of DNA fluctuations parallel to the direction of TALE diffusion were estimated to be ~40 nm on the basis of the equipartition theorem and a Taylor series expansion of the DNA stretching force under experimental conditions. This calculation relies on the average extension of DNA templates in our assay (~90% to 95% of full contour length) and the Marko–Siggia force–extension relation[49]. Summing together error from both sources, we estimate the average maximum uncertainty in localization $<\delta z^2>$ to be ~65 nm.

Covariance-based estimator for 1D diffusion coefficients. We determined 1D diffusion coefficients using a covariance-based estimator (CVE) to avoid compounding errors associated with the localization uncertainty between consecutive frames[23]. Using an average displacement Δx_n of a TALE protein during

a single acquisition period (20 to 33 ms per frame) and the localization error $<\delta z^2>$ as described above, 1D diffusion coefficient D was determined as

$$D = \frac{\overline{(\Delta x_n)^2} - 2\langle \delta z^2 \rangle}{2(1-2R)\Delta t} \quad (2)$$

where Δt is the time between frame acquisitions, R is the motion blur coefficient (assumed to be 1/6 here to account for the fact that our camera shutter is open during the entirety of image acquisition), and Δx_n is the protein step size. In determining 1D diffusion coefficients, the protein step size Δx_n is directly related to the camera frame rate. We acquire images using an acquisition time of 20–33 ms, and an increased acquisition rate would decrease the magnitude of the apparent step sizes. This relationship was accounted for in the CVE analysis. Given limitations of probe photostability and brightness and the sensitivity of EMCCD cameras, however, an increased frame rate must be balanced with the need to acquire sufficient photons for single-molecule localization. We achieved this balance by ensuring that proteins were localized across several consecutive frames in a trajectory, maximizing the frame rate until we were not able to successfully localize proteins between successive frames. As in previous work[15], we collected several molecular trajectories (>12) and hundreds or thousands of values of Δx_n for each condition tested to establish a representative sample size of individual molecules.

Determination of theoretical limits for 1D diffusion. The two prevailing models for 1D diffusion of proteins along DNA consider the following effects: rotationally coupled and nonrotationally coupled (linear) motion. In the case of nonrotationally coupled motion, the diffusion coefficient for a protein translating along DNA is given by

$$D_{\text{linear}} = k_B T\left(\frac{1}{\xi_{\text{trans}}}\right) \quad (3)$$

where ξ_{trans} is the translational friction given as

$$\xi_{\text{trans}} = 6R\pi\eta \quad (4)$$

Here, R is the radius of the protein–probe complex and η is the viscosity of solution. In the case of rotationally coupled 1D diffusion, we adopted a previously developed model for a protein undergoing curvilinear motion along DNA[14,32,50], in which the diffusion coefficient is given by

$$D_{\text{linear}} = k_B T\left(\frac{1}{\xi_{\text{trans}} + \xi_{\text{rot}}}\right) \quad (5)$$

where ξ_{rot} is the rotational friction given as

$$\xi_{\text{rot}} = 6\pi\eta \mathrm{R} + \left(\frac{2\pi}{10BP}\right)^2 [8\pi\eta^3 + 6\pi\eta R(R_{\text{OC}})^2] \quad (6)$$

Here, BP is the distance between DNA base pairs, R is the radius of the protein–probe complex and R_{OC} is the distance between the center of the DNA axis and the center of the protein–probe complex. Owing to the apparent wrapped helical conformation of the TALE complexes, we took R and R_{OC} to be equivalent. We estimated the radii of the TALE proteins from previous structural characterizations carried out via NMR, SAXS and light-scattering measurements[51]. The radii of the QD antibody conjugates were determined via transmission electron microscopy (TEM) (**Supplementary Fig. 11**), and the radii of streptavidin tetramers were previously reported[52].

Hydrodynamic flow assay. In order to determine whether convection or diffusion dominates protein motion at the location of the DNA-bound TALEs in flow experiments, we used particle-tracking velocimetry to measure the flow speed. Fluid flow was driven via a computer-controlled syringe pump operating at 150 µL/min. A solution of fluorescent polystyrene microbeads (400 nm radius) was introduced into the microfluidic flow cell, and the particle velocities were measured at a distance of ~50–100 nm from the surface (**Supplementary Fig. 6**). Using the experimentally determined fluid velocity, the protein Peclet number Pe was determined by

$$Pe = \frac{Lu}{D} \quad (7)$$

where L is the characteristic length, D is the 3D diffusion coefficient, and u is the local flow velocity. Given the estimate for D, determined via the approximate radius of the TALEs discussed above, $Pe = 10$ for the hydrodynamic flow experiments conducted in this work. In order to quantify bias introduced by flow at low-salt (90–150 mM KCl) conditions, we calculated the sample skewness of the frame-by-frame displacements of TALEs (protein step sizes). We define skewness according to the sample skewness equation given by

$$\text{skew} = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^{n}(X_i - X_{\text{avg}})^3}{\sigma^3} \quad (8)$$

where n is the number of steps (frame-by-frame displacements), X_i and X_{avg} are the individual and average step values, and σ is the s.d. of the distribution. To determine whether skewness present in sample sets was significant, we compared values of skew to the standard error of skewness (SES) given by

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (9)$$

where n is the number of steps in the sample set.

46. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39**, e82 (2011).
47. Schroeder, C.M., Blainey, P.C., Kim, S. & Xie, X.S. in *Single-Molecule Technology: A Laboratory Manual* (eds. Selvin, P.R. & Ha, T.) 461–492 (2008).
48. Wolter, S. *et al.* rapidSTORM: accurate, fast open-source software for localization microscopy. *Nat. Methods* **9**, 1040–1041 (2012).
49. Marko, J.F. & Siggia, E.D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
50. Bagchi, B., Blainey, P.C. & Xie, X.S. Diffusion constant of a nonspecifically bound protein undergoing curvilinear motion along DNA. *J. Phys. Chem. B* **112**, 6282–6284 (2008).
51. Murakami, M.T. *et al.* The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. *Proteins* **78**, 3386–3395 (2010).
52. Desruisseaux, C., Long, D., Drouin, G. & Slater, G.W. Electrophoresis of composite molecular objects. 1. Relation between friction, charge, and ionic strength in free solution. *Macromolecules* **34**, 44–52 (2001).

doi:10.1038/nchembio.2152