*6960–6970* *Nucleic Acids Research, 2017, Vol. 45, No. 11* *Published online 29 April 2017*
*doi: 10.1093/nar/gkx342*

# The effect of increasing numbers of repeats on TAL effector DNA binding specificity

**Fabio C. Rinaldi[1], Lindsey A. Doyle[2], Barry L. Stoddard[2] and Adam J. Bogdanove[1,*]**

[1]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA and [2]Division of Basic Sciences, Fred Hutchinson Cancer Research, Seattle, WA 98019, USA

Received February 10, 2017; Revised April 11, 2017; Editorial Decision April 13, 2017; Accepted April 21, 2017

## ABSTRACT

**Transcription activator-like effectors (TALEs) recognize their DNA targets via tandem repeats, each specifying a single nucleotide base in a one-to-one sequential arrangement. Due to this modularity and their ability to bind long DNA sequences with high specificity, TALEs have been used in many applications. Contributions of individual repeat-nucleotide associations to affinity and specificity have been characterized. Here, using *in vitro* binding assays, we examined the relationship between the number of repeats in a TALE and its affinity, for both target and non-target DNA. Each additional repeat provides extra binding energy for the target DNA, with the gain decaying exponentially such that binding energy saturates. Affinity for non-target DNA also increases non-linearly with the number of repeats, but with a slower decay of gain. The difference between the effect of length on affinity for target versus non-target DNA manifests in specificity increasing then diminishing with increasing TALE length, peaking between 15 and 19 repeats. Modeling across different hypothetical saturation levels and rates of gain decay, reflecting different repeat compositions, yielded a similar range of specificity optima. This range encompasses the mean and median length of native TALEs, suggesting that these proteins as a group have evolved for maximum specificity.**

## INTRODUCTION

Transcription activator-like effectors (TALEs) are DNA-binding proteins that belong mainly, but not exclusively, to members of the phytopathogenic genus *Xanthomonas* (1–3). TALEs are injected into plant cells via the type III secretion system of these bacteria, bind to specific host gene promoter sequences, and function as transcriptional activators (1,2).

Structurally, TALEs can be divided into three domains: the N-terminal domain (NTD) containing a type III secretion signal; the central repeat region (CRR) responsible for sequence-specific DNA target recognition; and the C-terminal domain (CTD) harboring nuclear localization signals (NLS) and a transcriptional activation domain (AD) (1,2). The DNA recognition domain of TALEs is composed of tandem repeats, typically comprising 33–35 amino acids, that interact with individual nucleotide bases in a one-to-one correspondence. Although the repeats are very conserved, amino acids in positions 12 and 13 of each repeat, known together as the RVD (**r**epeat **v**ariable **di**-residue), are highly polymorphic. These two residues largely determine the nucleotide specificity of each repeat, such that the sequence of RVDs across the CRR defines the overall target sequence of the protein. This discovery revolutionized the study of TAL effector function by facilitating target identification (4–7).

Several crystal structures have revealed molecular details of the TALE–DNA interaction. TALEs form a super-helix that wraps around the major groove of the DNA target site, allowing each repeat to closely interact with the DNA backbone and the nucleotide bases on a single strand. Each repeat is composed of two alpha-helices connected by a loop that contains the RVD (residues 12 and 13). The binding of a TALE around the DNA precisely places the side chain of residue 13 (also known as the base specifying residue) in direct contact with a base, thereby contributing to DNA-binding specificity. The N-terminal domain of TALEs harbors four additional, cryptic repeats (repeats –3, –2, –1 and 0) that play a less specific but crucial role in the DNA interaction (8–11). In general, TALEs demonstrate preference for a thymine (T) nucleotide at the 5′ position immediately preceding the target site, though other nucleotide types at this '0th' position have been demonstrated to allow binding in certain contexts (12–15).

The mechanism of TALE-DNA recognition allows designer TALEs (dTALEs) to be assembled to target DNA sequences of choice. TALE fusion with the *Fok*I nuclease (TALENs) provided the first truly modular, readily customized platform for targeted genetic modification in many organisms (16–19). Fusions of TALEs with alternative transcriptional activator domains and repressor domains has yielded targetable, orthogonal regulators of transcription

*To whom correspondence should be addressed. Tel: +1 607 255 7831; Fax: +1 607 255 4471; Email: ajb7@cornell.edu

for synthetic biology (20–25). Other TALE fusions have also been developed and employed (3,14,26,27).

Across all these implementations, in-depth understanding and optimization of TALE specificity in order to eliminate off-targets is essential and has been the focus of several studies (28–32). Although in nature TALEs display a large diversity of RVDs, the application of TALEs in biotechnology has converged on the most common and most specific ones. These are HD, NI, NG, NN and NH, which individually specify C, A, T, G or A and G respectively. TALE–DNA interactions can tolerate mismatches between an RVD and a nucleotide at several positions; in fact mismatches are commonly observed in nature. The contribution of individual RVDs to overall binding affinities has been analysed in different contexts (30,33). Different RVDs contribute differently to binding affinity, and even the contribution of the same RVD type changes as a function of its position and context within the central repeat region (30,33). For example, Garg *et al.* observed that RVD-nucleotide mismatches positioned at the 3′ end of the target site of a dTALE had less negative impact on overall activity than mismatches at the middle or the 5′ end (25). Experiments by Meckler and colleagues demonstrated directly using the natural TALE AvrBs3 that RVD-nucleotide mismatches positioned at the 3′ end of the target site have less negative impact on overall affinity than mismatches at the 5′ end (30). This polarity of mismatch tolerance was corroborated by an analysis by Rogers *et al.* of position dependence for each of the four most common RVDs using synthetic TALEs containing from 9 to 19 RVDs (33). That study also showed that the magnitude of the polarity of mismatch tolerance varies with length and RVD composition (33). Indeed, in the Meckler study, two TALEs with the same RVDs distributed differently in the CRR displayed different binding affinities for their targets (30). The complex nature of these context effects was further underscored in the Rogers study by experiments showing that the identities of the N- and C-terminal neighbours affect the RVD-nucleotide specificity (33).

While much research has been concentrated on individual RVD-nucleotide contributions to specificity, less attention has been devoted to the fundamental question of how the total number of RVDs in a TALE affects the specificity of its interaction with DNA. A survey of the size of TALEs in nature found that the number of RVDs ranges from 2 to 34; however, the great majority of TALEs have between 16 and 20 RVDs (1).

In the work described below, we examined how the number of repeats of a TALE affects its overall specificity by measuring the effect of the number of repeats on binding affinity for target versus non-target DNA. Although much of the work done to date to evaluate properties of TALEs has relied on activity assays *in vivo*, we chose to use direct DNA binding assays because binding is the first and most important step of the chain of events that leads to activity; and, *in vivo* activity assays can be affected by variables like heterochromatin, competition, and target context, so results do not always correlate with binding affinity (30,34). First, we present evidence for two natural TALEs of different lengths that corroborates the reduction in specificity of individual RVD-nucleotide interactions at the 5′ end observed previously (30). Then, using a series of synthetic TALEs based on one of these proteins, but spanning a range of lengths, we demonstrate that each additional repeat incorporated in a TALE provides extra binding energy for the target DNA, but that the gain decays exponentially with increasing length. We also show that affinity for non-target DNA increases non-linearly with the number of repeats, but with a slower decay of gain. This results in maximum specificity (the ratio of affinity for target DNA to affinity for non-target DNA) at a length between 15 and 19 RVDs. To extend these results, we modeled the relationship between TALE specificity and the number of RVDs based on our data and varied the parameters of the model that might be expected to change as a function of RVD composition. This revealed that even with different hypothetical RVD compositions, TALEs reach an optimum length for specificity, between 16 and 25 repeats for the variations we modeled. Intriguingly, this range encompasses the mean and median length of TALEs found in nature.

## MATERIALS AND METHODS

### PthXo1 and AvrHah1 expression constructs

The PthXo1 construct used for the in-well assays, pAC-EXP31c, has been described (8). The AvrHah1 construct was made similarly, using AvrHah1-coding sequence from pB-Bam4 (35), and was designated as pAC-EXP41c. In each, the coding sequence spans from amino acid 127 at the N-terminal end (AARPP…) to the 62nd amino acid after the last (truncated) repeat (…TSHRV), and includes an N-terminal 6xHis tag and short linker (MASSHHHHHHSSGLVPRGSSM).

### TAL effector constructs containing different numbers of RVDs

Coding sequence for the dTALEs containing 10, 14, 18, 22 and 26 RVDs were assembled in the backbone vector pGEX6P2-TALE (13) by the Golden-Gate method using our previously described kit (36), with the last RVD residing within a truncated repeat as in native TALEs. Since this kit was not designed to assemble dTALE constructs encoding fewer than 12 RVDs, the construction of the 10-RVD dTALE was accomplished by mutating the sub-array vector pFUSB (36) to make it directly ligatable to a backbone vector, bypassing co-assembly with pFUSA. That was done by disrupting the 5′ *Esp*3I site using oligonucleotides FR5 (5′-gtcttaagcgtctcaccctctgaccccggaccaag-3′) and FR6 (5′-cttggtccggggtcagagggtgagacgcttaagac-3′), followed by insertion of four extra bases near the mutated site using oligonucleotides FR12 (5′-gtgccccccctgaatctgaccccgg-3′) and FR13 (5′-ccggggtcagattcagggggcac-3′) to position the construct in the correct frame. To complete the assembly of the 10-RVD construct, the first nine RVD modules were assembled into the modified pFUSB, then in the second Golden-Gate reaction the resulting plasmid was combined with the last repeat vector pLR-NG and the backbone vector. From pGEX6P2-TALE, the repeat region for each dTALE was used to replace the repeat region for PthXo1 in pAC_Exp31c using *Stu*I and *Bmg*BI (8).

### Protein expression and purification

Expression constructs were transformed into Rosetta pLysS (DE3) (Novagen, Madison, WI, USA) cells, and single colonies were grown overnight in 5 ml of LB supplemented with ampicillin (100 μg/ml and chloramphenicol 10 μg/ml). Cells were spun down, re-suspended in 5 ml of fresh media, and 0.5 ml was used to inoculate 50 ml cultures. Cultures were incubated at 37°C, and protein expression was induced with 100 μM IPTG when $OD_{600}$ reached ∼1.0. Protein expression was carried out for 16 h at 15°C. Cells were harvested at 3220 × g for 10 min, and the pellets were re-suspended in 5 ml of lysis buffer containing 20 mM Tris–HCl pH 8.0, 400 mM NaCl and 5% glycerol supplemented with protease inhibitors (ThermoFisher Scientific, Waltham, MA, USA). Samples were sonicated on ice, and lysate supernatants were applied onto a 250 μl bed volume of TALON resin (Clontech Laboratories, Inc, Palo Alto, CA, USA) pre-equilibrated with lysis buffer. The resin was then washed with 50 ml of lysis buffer, then 30 ml of high salt buffer (20 mM Tris–HCl pH 8.0, 2000 mM NaCl and 5% glycerol) for displacing pre-bound DNA molecules (30), and finally an additional 20 ml of lysis buffer. The proteins were eluted with 5 ml of elution buffer containing 20 mM Tris–HCl pH 8.0, 400 mM NaCl and 300 mM Imidazole. The eluates were concentrated down to 500 μl and dialyzed twice against 1 l of buffer containing 20 mM Tris–HCl pH 8.0, 400 mM KCl, 5 mM DTT and 2 mM EDTA. For fluorescence anisotropy, proteins were dialyzed against buffer containing 20 mM Tris–HCl pH 8.0 and 400 mM NaCl. Protein concentrations were determined by SDS-PAGE using BSA standards and three different dilutions of each protein. Proteins samples were freshly purified for each experiment.

### Binding specificity profiles

The position-dependent DNA binding specificity profiles of PthXo1 and AvrHah1 were determined using a fluorescence-based competition assay that has been previously described in detail (37,38). Briefly, for each protein, following purification, 200 μl of a 100 nM protein aliquot was added to each well of a 96-well Nickel-NTA 'HisSorb' plate (Qiagen Inc, Valencia, CA). The plates were then incubated at 4°C for 2 h and then washed four times with 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, and 0.05% Tween-20. For each protein, each well was then incubated with 200 μl of a DNA mixture containing 100 nM concentration of a fluorescently labeled 30 base pair DNA duplex that harbored the 'optimal' binding target for that protein (corresponding to the sequence 5′-TGCATCTCCCCCTACTGTACACCAC-3′ for PthXo1 and 5′-TTAAACCTAACCAT-3′ for AvrHah1) and a 3 μM concentration of a single unlabeled competitor DNA duplex. Each such competitor harbored a single base pair substitution relative to the labelled, optimal target site sequence. The plates were again incubated for 2 h at 4°C and then washed four times with 200 μl of TBS (50 mM Tris–HCl, pH 7.5, 150 mM NaCl).

After the final wash step, 200 μl of TBS were added to each well, and the retained fluorescence in each well was measured using a SpectraMax M5/M5e microplate reader (excitation: 510 nm, emission: 565 nm, cutoff: 550). The extent to which the competitor DNA construct reduces the fluorescent signal in each well corresponds to the ability of that DNA sequence to compete for protein binding with the labeled DNA target. Strong retention of the labeled DNA duplex after incubation with the competitor is indicative of a corresponding preference for the original DNA sequence, whereas reduction in the fluorescent signal is indicative of a reduced preference for the original sequence (indicating a lower level of target discrimination between the labeled and unlabeled DNA duplexes). Using this method, the effects of all three possible mismatches (i.e. a G, C or A in place of a wild-type T at position 0, etc.) were assayed at each position across the target site. The entire assay was replicated five times in one parallelized experimental run, using a single common stock of purified protein and a common stock of the fluorescently labeled DNA target site duplex. The averaged fluorescent signal for each competitor DNA target sequence was then converted into a relative association constant ($K_a$) to the 'optimal' binding site as described (38). In order to perform that calculation, the affinity of the effector was independently determined using a quantitative electrophoretic mobility shift analysis (EMSA), and the fluorescence competition experiment was augmented with two sets of additional control experiments in which the bound 'optimal' DNA complex competed against unlabeled versions of itself and then with a randomized DNA sequence (to generate limiting values for sequence-specific competition effects).

For each protein, the experiment was repeated a total of three times using independent protein preparations.

### Electrophoretic mobility shift assays

The EMSA protocol was adapted from one described previously (30). The oligonucleotides (Supplementary Table S2) were purchased (Integrated DNA Technologies, Coralville, IA, USA) as DNA duplexes tagged with a biotin molecule at the 5′ end of the reverse strand. The reactions were performed according to the LightShift Chemiluminescent EMSA kit protocol (ThermoFisher Scientific). Briefly, protein–DNA reactions were prepared by mixing 50 pM DNA probe and varying concentrations of protein (0–2000 nM), with the final buffer containing 10 mM Tris–HCl pH 7.5, 100 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM $MgCl_2$, 50 ng/ul Poly-dIdC, 0.05% NP-40, 0.4 mM EDTA and 0.1 mg/ml BSA. The binding reactions were incubated at room temperature for 50 min, mixed 9:1 with 10× sample buffer (25% Ficoll and 2.5 mg/ml xylene cyanol in TE buffer) and 10 μl of the total mixture was applied to a 1.3% agarose gel. The gel was run at 110 V for ∼45 min at 4°C. The separated DNA and protein–DNA complexes were transferred to a positively charged Nylon membrane (Sigma-Aldrich, St. Louis, MO, USA) pre-soaked in 0.5× TBE buffer at 100 V for 30 min at 4°C. The DNA was cross-linked to the nylon membrane immediately after transfer by placing the membrane facing down on a ChemiDoc imaging system (Bio-Rad Laboratories, Hercules, CA, USA) and exposing to UV for 12 min. The (biotin-labeled) DNA was detected by chemiluminescence following the manufacturer's protocol. All experiments were replicated at least twice using inde-

pendent protein preparations. Data were analysed using the imagelab software (Bio-Rad Laboratories), and the DNA bound fraction (FB) was calculated using the equation FB = DNAB/(DNAB + DNAF), where DNAB is the intensity of the band of DNA bound by protein and DNAF is the intensity of the band of DNA free of protein. The apparent $K_D$ ($K_{D,app}$) was calculated after fitting the data with single binding isotherms (Hill equation) using Prism 6 (GraphPad Software Inc., San Diego, CA, USA).

### Fluorescence anisotropy assays

Fluorescence anisotropy assays of TALE–DNA binding were carried out based on a previously described protocol (39). All probes (Supplementary Table S2) were purchased (Integrated DNA Technologies) as double strand DNA with a 6-FAM fluorescent molecule attached at the 5′ end of the forward strand. Experiments were performed on a FilterMax F5 plate reader (Molecular Devices, SunnyVale, CA, USA) mounted with 485 nm excitation/535 nm emission filters and polarizers. Measurements for the dTALE interactions with the target DNA were performed in buffer containing 50 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.1 mg/ml BSA and 1 mM TCEP. For each reaction, a different concentration of the purified protein was mixed in a solution containing 1 nM DNA probe and buffer to a final volume of 160 μl. Reactions were incubated for 60 min at 25°C before measurement. Measurements for the dTALE interactions with the non-target DNA were performed in the same way, and separately using a lower salt concentration (115 mM NaCl). All experiments were replicated at least three times using independent protein preparations. Fluorescence anisotropy data were normalized considering the protein-free reaction signal as zero and the values at the signal saturation as one. To calculate $K_{D,app}$, the normalized data were plotted against the protein concentration and fitted to a three-parameter dose-response curve using Prism 6 (GraphPad).

## RESULTS

### The magnitude of the polarity of mismatch tolerance can vary for different TALEs

We initiated our study of the effect of length on specificity by asking to what extent the previously observed polarity of mismatch tolerance (30,33) holds for the natural TALEs of different lengths, PthXo1 and AvrHah1, containing 24 and 14 RVDs, respectively. For each protein, we evaluated specificity of individual RVDs across the target site using a previously described, fluorescence-based, microtiter plate binding competition assay, in which the protein is immobilized in each well in complex with labeled perfect-match target DNA, and the ability of competitor DNA to displace the fluorescent target is measured. Competitor DNAs represent all possible single base pair substitutions at each position of the target (37,38). For PthXo1 (Figure 1A), the effects of individual base pair substitutions within the DNA target on overall binding affinity were quite small, in most cases causing <2-fold reduction in affinity, yet they were asymmetric in their amplitude across the target site (Figure 1A). Specificity of individual RVDs as a function of position starts

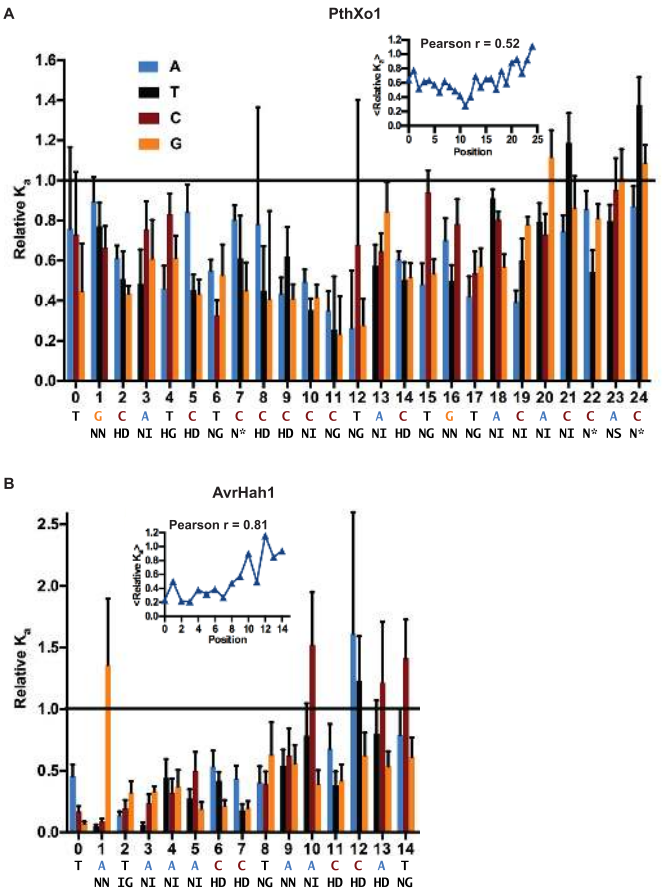


**Figure 1.** Individual base-RVD specificity in PthXo1 and AvrHah1 relaxes closer to the 3′ end of the target sites. (**A**) Relative binding affinities ($K_a$) of (A) a protein construct containing the DNA recognition domain of PthXo1 (24 RVDs) and (**B**) one carrying that of AvrHah1 (14 RVDs) for DNA targets with single nucleotide substitutions across the target site compared to the perfect match DNA targets. The plots show in both cases a significant decrease in binding specificity closer to the 3′ end of the target site, more readily observed when the relative $K_a$ are averaged for all three substitutions (figure insets). This effect is more pronounced for the smaller TALE AvrHah1 than for PthXo1. Error bars represent standard deviation of five technical replicates. Two additional replications of the experiment using independently purified protein preparations yielded similar results.

to decrease slightly starting approximately at position 13, and then decreases more substantially beyond position 18. The observed reduction in DNA sequence discrimination across the final positions approaching the 3′ end of the target site corresponds well with the crystallographic structure of PthXo1 bound to its DNA target site (8). In crystallography, B-factor is an expression of the degree to which the electron density for an atom is spread out, reflecting mobility. In the PthXo1 structure, the average B-factor observed for the backbone atoms in each repeat of the CRR increases in a linear fashion beyond position 19, with the final repeat completely disordered and unobservable in the electron density map (Supplementary Figure S1). This increase in the average *B*-factor at each repeat is also observable in two other TALE–DNA crystallographic structures, AvrBs3 and dHax3 (9,10). In both, the average *B*-factor increases

beyond the first half of the CRR independent of the total number of repeats.

For AvrHah1, a reduction in specificity of individual RVDs N- to C-terminally across the target site was also observed, but in contrast to PthXo1, in which the trend was moderate (correlation coefficient = 0.52), for AvrHah1 the polarity was more pronounced (correlation coefficient = 0.81): mismatches at the beginning of the CRR are much more deleterious to the overall binding of AvrHah1 than they are for PthXo1. We speculate that this difference is due to the greater relative contributions of individual RVDs to AvrHah1 binding affinity than to PthXo1 binding affinity due to AvrHah1 having fewer RVDs overall, such that the polarity effect is more readily apparent. However, the difference might also or instead relate to the different RVD compositions of the two proteins. Although the comparison between these two proteins does not alone resolve the relative contributions of RVD composition versus length, it further generalizes the polarity of mismatch tolerance observed in previous studies (25,30,33), and the variability in magnitude of that effect, as a fundamental property of TALEs. It also quantifies the effect for the PthXo1 protein, on which constructs for the experiments described following are based.

### A set of dTALEs and DNA probes to test the effect of repeat number on affinity and specificity

Irrespective of the possible influence of differences in RVD composition at the ends of PthXo1 and AvrHah1, the different trends for the reduction in specificity across the target sites for these proteins support a hypothesis in which the increase in affinity that results from an increase in the number of repeats is nonlinear. In order to quantitatively test this hypothesis, we designed five dTALEs with the number of RVDs systematically increasing from 10 to 26 (Figure 2A). For each, the initial 10 RVDs correspond to those of PthXo1 except for the 'NN' at position 1, for which 'NH' was substituted due to its greater specificity toward the nucleotide G (29,31). Thereafter, the dTALEs vary in length by blocks of four RVDs, one for each nucleotide, with each block kept constant ('NI HD NH NG') to control for differences in affinity contributed by different RVDs added individually. Target and non-target DNA probes were synthesized as double-stranded oligonucleotides. The target DNA probe contains the thymine (T) at position 0 followed by the perfect-match sequence for the longest dTALE, which because of the initial 10-RVD 'anchor' and subsequent 4-RVD blocks, also serves as target for each of the shorter dTALE constructs; the non-target DNA probe contains the target sequence, with the initial T, completely scrambled (Figure 2B). The probes were tagged with either a biotin molecule at the 5′ end of the reverse strand or a fluorescein molecule at the 5′ end of the forward strand (Supplementary Table S2). This set of dTALEs and DNA probes allowed us to assess the TALE-DNA binding affinity and specificity by electrophoretic mobility shift assay (EMSA, using the biotin-labeled probes) and fluorescence anisotropy (FA, using the fluorescein-labeled probes). The nucleotide sequences flanking the target and non-target DNA sequences were the same within but different across these two methods (Supplementary Table S2).

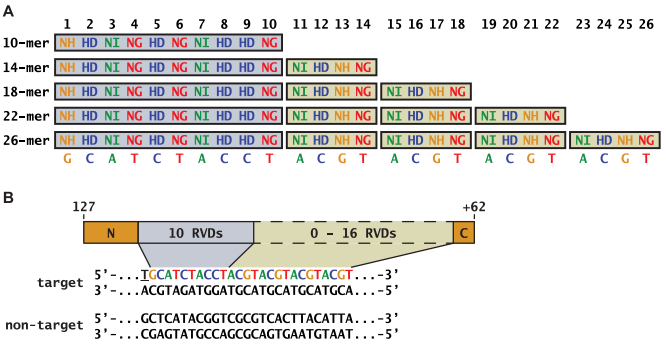


**Figure 2.** dTALEs and DNA probes used to measure the effect of length on specificity. (**A**) Five dTALEs with CRRs comprising from 10 to 26 RVDs are shown. The first 10 RVDs, shown in the blue box are based on the natural TALE PthXo1. The five dTALEs were constructed by adding 0 to 4 four-repeat blocks (yellow) each containing the RVDs NI-HD-NH-NG to the C-terminus of the first 10 RVDs. RVD-base associations are colour coded. (**B**) dTALEs span from residue 127 at the N-terminus to the 62nd residue following the CRR, relative to a full-length TALE. Sequences of the double strand DNAs used are shown. The thymine (T) at position 0 in the target is underlined; it was scrambled along with the RVD-specified bases to generate the non-target DNA sequence. The number of nucleotides flanking the target (top) and non-target (bottom) DNAs varies for different experiments (Supplementary Table S2).

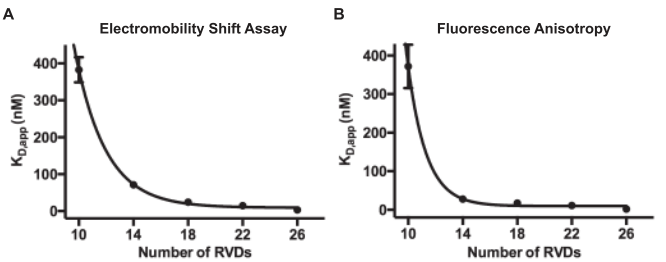


**Figure 3.** TAL effector affinity for the target DNA plateaus between 14 and 18 RVDs. Relationship between the apparent dissociation constant ($K_{D,app}$) and the number of RVDs measured by (**A**) electromobility shift assay (EMSA), and (**B**) fluorescence anisotropy (FA). All plots were fitted to one-phase exponential decay curves ($R^2 = 0.99$ for EMSA, $R^2 = 0.97$ for FA). The half-life of decay for EMSA is ~1.6 ± 0.2 RVDs while for FA it is ~1.1 ± 0.3 RVDs.

### Binding affinity for target DNA reaches a plateau with increasing number of RVDs

Binding of the dTALEs to target and non-target DNA probes was first analysed by EMSA. All five proteins showed clear differences in their affinities for target versus non-target DNA (Supplementary Figure S2). $K_{D,app}$ values calculated from band densitometry data span a wide range, increasing with length of the dTALE (Figure 3A and Table 1). Surprisingly, however, $K_{D,app}$ does not decrease at a linear rate with increasing number of RVDs. Rather, the relationship between $K_{D,app}$ and number of RVDs closely resembles an exponential decay, in which on average every 1.6 additional RVDs decreases $K_{D,app}$ by half, leveling off at close to 18 RVDs. To test this observed behavior further, we measured affinities by FA. For each dTALE-probe combination, $K_{D,app}$ was estimated by plotting the magnitude of anisotropic change due to complex formation against increasing protein concentration (Supplementary Figure S3).

The $K_{D,app}$ values obtained are in the same range as those obtained by EMSA (Table 1). Moreover, as observed in the EMSA experiments, the relationship between the number of RVDs and $K_{D,app}$ follows an exponential decay curve, although with a smaller half-life (1.1 RVDs), indicating an earlier saturation of affinity than observed by EMSA, at closer to 14 RVDs (Figure 3B). Taken together, the data consistently indicate that binding affinity indeed increases with the number of RVDs but levels off between 14 and 18 RVDs, i.e. beyond that size range the addition of RVDs does not provide substantial contribution to binding.

These results could explain the substantial difference in the relative impact of 5′ mismatches observed between PthXo1 and AvrHah1 (Figure 1). Based on the EMSA and FA data, one can predict that the $K_D$ for an interaction of a shorter TALE like AvrHah1 with DNA would fall within the steepest region of the exponential decay curve, such that any mismatch would be relatively deleterious. In line with this prediction, Streubel *et al.* showed that TALEs with smaller numbers of RVDs indeed rely more on the contributions of individual RVDs for binding (28). Given the polarity of mismatch tolerance inherent to TALEs, the deleterious effect would be more pronounced for 5′ mismatches. Conversely, for a larger TALE, like PthXo1, the $K_D$ would reside in the more level part of the curve, such that overall binding affinity would be influenced less by single mismatches at individual RVDs throughout, and this might dampen the observable polarity effect. To test these predictions further, we performed FA binding assays with the 14- and 26-RVD dTALEs using DNA probes containing two mismatches near either the 3′ end or the 5′ end of the target sequences (Supplementary Figure S4). As predicted, mismatches at either end reduced the binding affinity of the interaction with the 14-RVD dTALE to a greater relative extent than they did for the 26-RVD protein (Supplementary Figure S4C). Somewhat unexpectedly, while the 3′ mismatches reduced the affinity by 7-fold for the 14-RVD dTALE and only 2-fold for the 26-RVD dTALE, the difference in effect of the 5′ mismatches was not as pronounced, being relatively severe in both cases (15-fold for the 14-RVD dTALE and 12-fold for the 26-RVD one). The data thus actually show a dampening of the polarity effect not for the longer dTALE but for the shorter one, with mismatches at both ends in the latter causing substantial decreases in affinity. In contrast, for the longer dTALE, the marginal impact of the 3′ mismatches yet the strong impact of the 5′ mismatches amounts to stronger polarity. Here again, we are left to speculate that in the comparison with AvrHah1, the lesser polarity in the longer protein PthXo1 (Figure 1) results from differences in the RVD compositions of the two proteins. Yet overall, the data are consistent with the prediction based on the affinity saturation curve that shorter TALEs are impacted to a greater extent by mismatches, and in this context provide evidence that the degree of saturation influences the overall mismatch tolerance.

Regarding the observed affinity saturation itself, and the 5′-3′ polarization, which was consistently evident in our data, we surmise that both are consequences of a yet unknown, underlying structural property of TALEs. One possibility is that the binding affinity saturation originates from greater repeat flexibility nearer the protein C-terminus (Supplementary Figure S1), which in turn might increase the entropic cost of ordering each additional repeat on the 3′ end of the DNA target. Previous work showed that two regions of the protein mediate the interaction between TALEs and DNA; the cryptic repeats, located in the N-terminal domain, and the repeats found in the CRR (11,40). Interestingly, while the cryptic repeats can bind DNA independently, the CRR can not, suggesting that the cryptic repeats are responsible for the initiation of binding and provide an 'anchor' for the interaction of the CRR with DNA (11,40). It is possible that the influence of the binding energy provided by the cryptic repeats diminishes across the array such that the farther away a repeat is from the cryptic repeats the smaller its contribution to binding. As such, a repeat that is very far from the cryptic repeats might act just like the repeats in an 'unanchored' CRR by itself. Another, not mutually exclusive, possibility for explaining the binding affinity saturation is that N- to C-terminally across the CRR, structural associations between repeats and DNA nucleotides become increasingly misaligned, such that the contributions of RVD-nucleotide interactions to the overall binding decrease (41).

Based on our data, we would predict that increasing the number of RVDs for TALEs that would otherwise have already reached the saturation level for affinity would have negligible impact on activity. This is supported by the results of a reporter gene activation study using TALEs ranging from 15 to 25 RVDs, in which no significant differences in activation were observed (42). Similarly, in another study, TALENs containing from 10 to 16 RVDs showed no significant differences in activity (34).

### Modelling shows that the contributions of individual RVDs to binding energy vary depending on their position

The size-affinity relationship data prompted us to model individual RVD contributions to binding energy across the TAL effector repeat region. We used the exponential decay equations that describe the size versus affinity relationship (Figure 3 and Supplementary Tables S3 and S4) to extrapolate back to zero RVDs. Next, we calculated the binding free energy via the Gibbs free energy equation ($\Delta G^\circ = RT \ln K_D$; $R = 8.314$ J mol$^{-1}$ K$^{-1}$ and $T = 298$ K) and plotted the individual RVD contributions to binding energy ($\Delta G_n^\circ - \Delta G^\circ_{n-1}$) relative to the energy provided by the TALE N-terminal domain ($\Delta G_0^\circ$) as a function of the number of RVDs. This resulted in a sigmoidal curve with the inflection point at ∼21.8 RVDs for EMSA and 17.7 RVDs for FA (Supplementary Figure S5). This curve reflects the attenuation of the contribution of individual RVDs to binding affinity as distance from the beginning of the CRR increases; it yields an attenuation factor for each RVD position ($A_n$).

Defining the binding energy provided by the association of a certain type of RVD with its preferred base as $\Delta G^\circ_{\text{RVD-base}}$, the position effect can be modeled by multiplying that binding energy by the attenuation factor, i.e. $\Delta G^\circ_{\text{RVD-base}} \times A_n$. It is important to note that this attenuation factor is based solely on our select set of dTALEs and could be different for TALEs with different RVD compositions. Nonetheless, the model overall provides a use-

**Table 1.** Apparent dissociation constants obtained by EMSA and FA for the five dTALEs interacting with target and non-target DNA

| dTALEs | Specific binding | | Non-specific binding | |
|---|---|---|---|---|
| | $K_{D,app}$ (nM)$_{EMSA}$ | $K_{D,app}$ (nM)$_{FA}$ | $K_{D,app}$ (nM)$_{FA}$ | |
| | 100 nM Salt | 150 mM Salt | 150 mM Salt | 115 mM Salt |
| **10 RVDs** | 382.6 ± 24.0 | 372.1 ± 56.5 | 9144.0 ± 3431.1 | 659.6 ± 128.7 |
| **14 RVDs** | 71.0 ± 1.5 | 27.8 ± 4.1 | 4399.3 ± 614.6 | 368.1 ± 47.2 |
| **18 RVDs** | 24.3 ± 2.4 | 17.2 ± 7.0 | 2268.0 ± 618.0 | 219.4 ± 26.9 |
| **22 RVDs** | 14.4 ± 5.6 | 10.9 ± 2.7 | 398.4 ± 167.8 | 46.8 ± 8.9 |
| **26 RVDs** | 3.2 ± 1.4 | 1.8 ± 0.2 | 116.9 ± 39.4 | 14.5 ± 1.7 |

ful expression of the dampening of binding energy contribution that occurs as position in the CRR increases, and could serve as a basis for improving prediction and design of TALE specificity. The prediction programs Talvez and PROGNOS incorporate a position correction parameter empirically defined by relaxing individual RVD-nucleotide specificities beyond a certain number of RVDs (43,44). In line with our results, the performance of Talvez is improved substantially if the position correction parameter is applied between 15 and 19 RVDs (43).

### Binding affinity for non-target DNA also increases non-linearly with number of RVDs but with a slower decay of gain

Knowing the affinity of DNA binding proteins for both target and non-target sequences enables calculation of their binding specificity, defined as the ratio $K_D^{non-target}/K_D^{target}$ (45). Therefore, as a first step toward evaluating the effect of the number of RVDs on specificity of TALEs for their target sequence, we measured the affinity of our five dTALEs for the non-target probe. We used FA for this because it is more sensitive than EMSA; in EMSA, low affinity interactions are susceptible to dissociation during gel electrophoresis and transfer to the membrane. To be consistent with the FA experiments carried out for the target DNA (Figure 3B), reactions were performed in the same buffer with the same salt concentration (150 mM NaCl). As before, for each dTALE-probe combination, $K_D$ was estimated by plotting the magnitude of anisotropic change due to complex formation against increasing protein concentration (Supplementary Figure S6A). Similar to the relationship observed for the target DNA, the data show a decrease in $K_{D,app}$ with increasing number of RVDs (Figure 4A), resembling exponential decay, but with a slower decay than observed for the target DNA (Supplementary Table S5).

Positively charged, non-RVD amino acids normally found in TALE repeats have been suggested to provide a stronger contribution than the RVD-nucleotide association to the specific interaction with target DNA (46). TALE binding to non-target sequences depends highly on the ionic strength of the buffer, suggesting that these charged residues might play a role also in the non-specific interaction of TALEs with DNA (40,47). We therefore repeated the FA experiment using a lower salt concentration corresponding to 115 mM NaCl (Supplementary Figure S6B). Indeed, this change in ionic strength of the buffer resulted in a roughly 10-fold decrease in $K_{D,app}$ for each dTALE-non-target DNA interaction (Table 1; Supplementary Figure S7). However, the overall trend in the effect of number of RVDs on $K_{D,app}$

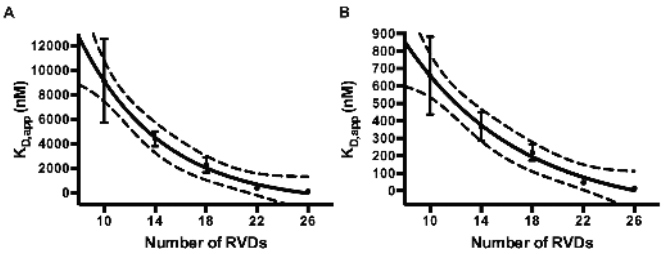


**Figure 4.** TAL effector affinity for non-target DNA increases with the number of RVDs. The relationship between the apparent dissociation constant ($K_{D,app}$) and the number of RVDs was determined by measuring the strength of the interaction of all five dTALEs for the non-target DNA by the change in fluorescence anisotropy on binding (Supplementary Figure S5). Experiments were performed using two different salt concentrations; (**A**) 150 mM NaCl and (**B**) 115 mM NaCl. The plots represent the average of at least three independent measurements for each protein. Error bars represent standard deviation. The 95% Confidence Intervals are displayed with dashed lines. The data were fitted to exponential decay curves with $R^2 = 0.86$ and 0.87, respectively (Supplementary Tables S5 and S6). The binding affinity for the non-target DNA increases with the number of RVDs indicating that each repeat contributes to the binding energy, likely via charged residues present in the repeats.

was the same, fitting an exponential decay curve with slower decay than for the target DNA (Figure 4B, and Supplementary Table S6). Thus, each repeat of the TALE CRR provides additional binding energy to the non-specific interaction, driven by charged residues on the repeats, and as a consequence, longer TALEs have higher non-specific affinity for DNA than shorter TALEs.

The ability of TALEs to bind non-target sequences has been analysed before (11,47,48). Cuculis *et al.* (47) demonstrated that TALEs search for the target sequence through a 1D diffusion mechanism. By analysing a series of truncations, the authors further showed that the N-terminal region is crucial for TALE nucleation on DNA and that increasing length of the CRR decreases the speed of diffusion. Their data support a model in which both the NTD and the CRR are loosely in contact with the DNA (47). Our data for the interaction of TALEs with non-target DNA agree with this model. Combined with our observation that larger TALEs bind to non-target sequences with higher affinities than smaller TALEs, the substantial effect of salt concentration on binding that we observed suggests that the interaction between protein and DNA is mediated mainly by electrostatic interactions provided by the additional repeats. In further agreement with this, molecular dynamics studies have suggested that most of the binding energy originates from the interaction of non-RVD residues on each repeat

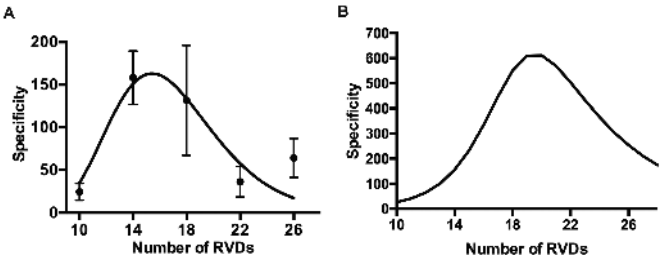


**Figure 5.** Binding specificity for the set of dTALEs analysed. (**A**) Specificity calculated by dividing the dissociation constants for non-target and target DNA shown in Table 1. Error bars represent the propagation of errors calculated using SD values. The plot was fitted to a Gaussian curve centered at 15.4 ± 1.1 RVDs. (**B**) Model for the change in specificity as a function of the number of RVDs calculated by directly dividing the exponential decay curves from Figures 4A and 3B. Maximum specificity in this case is centered at 19 RVDs. Both plots show that the specificity resembles a Gaussian distribution with a maximum between 15 and 19 RVDs.

with the DNA backbone (46). Taken altogether, these studies and our data provide compelling evidence that TALEs bind non-target DNA sequences through both the NTD and the CRR, and that the addition of repeats contributes positively to non-specific DNA binding energy.

### TALE specificity increases then declines with increasing number of RVDs

Having determined $K_{D,app}$ for target and non-target sequences for each of the dTALEs (Table 1), we next calculated the specificity for each using those experimentally determined values, $K_D^{non-target}/K_D^{target}$. Plotting the relationship between specificity and the number of RVDs yielded a roughly Gaussian curve along which specificity increases up to around 15–16 RVDs (15.4 ± 1.1), then diminishes (Figure 5A). The same type of relationship was observed when, instead of dividing individual values, the calculation was performed using the exponential decay curves for the non-target and target DNA (Supplementary Figure S8 and Figure 3B; equations shown in Supplementary Tables S4 and S7). In this case, the Gaussian distribution indicated a specificity maximum at closer to 19 RVDs (Figure 5B).

To examine whether the behavior exhibited by our dTALEs is a general feature of TALE–DNA binding, we expanded our analysis by modelling, using the following equation to express the exponential decay of the dissociation constant ($K_D$) of a TALE–DNA interaction,

$$K_D^N = (Y_0 - P)\ exp^{-kN} + P$$

where $N$ is the number of RVDs, $Y_0$ is the intersection of the curve with the y-axis, $P$ (for plateau) is the saturation level and $k$ is the rate of decay. We envisioned two possible scenarios. In the first, we assumed that the saturation in affinity observed in our experiments is caused by an intrinsic structural property of TALEs, such as the increasing flexibility or misalignment at increasing distance from the N-terminus discussed above, so that saturation always occurs in the range between 14 and 18 RVDs (as observed in Figure 3A and B). In this scenario, TALEs bearing different RVD arrays would follow a rate of decay ($k$) similar to that observed for the values obtained by FA and EMSA (Supplementary Tables S3 and S4), irrespective of RVD composition. Since different types of RVDs contribute with different binding energies, saturation ($P$) could occur at different levels of affinity, as a *function* of RVD composition (Figure 6A). For example, the maximum potential affinity of a TALE containing many weak RVDs at the beginning of the CRR would reach saturation at a higher $K_D$ than a TALE bearing many strong RVDs at the beginning of the CRR.

To examine the relationship between specificity and number of RVDs in this first scenario, we plotted the ratio of the dissociation constants for non-target and target DNA ($K_D^{non-target}/K_D^{target}$) as a function of both number of RVDs and $P$ (Figure 6B and Supplementary Table S8). For TALEs of the same length, affinity for non-target DNA can be assumed to be constant, irrespective of RVD composition, considering that the RVDs are not likely to participate significantly in the non-specific interaction. Therefore, the parameters for the curves fitting the FA data for target and non-target DNA (Supplementary Tables S4 and S7) were used directly to calculate specificity. The plot reflects a Gaussian distribution with the maximum varying according to both the saturation value ($P$) and the number of RVDs ($N$). Across a range of values for $P$, the specificity reaches its maximum between 16 and 21 RVDs. Also, the plot shows that the lower the saturation value, the higher the specificity.

In the second scenario, we assumed that the rate of decay depends on the RVD composition, such that different RVD compositions result in different points of saturation (Figure 6C). For example, strong RVDs at the beginning of the CRR would result in a higher $k$, with saturation of the affinity at a smaller number of RVDs, while weak RVDs at the beginning of the CRR would result in lower $k$, with saturation at a higher number of RVDs (Figure 6C). As in scenario one, to examine the relationship between specificity and number of RVDs, we plotted the ratio of the dissociation constants for non-target and target DNA ($K_D^{non-target}/K_D^{target}$), but in this case as a function of number of RVDs and $k$ (Figure 6D and Supplementary Table S8). As before, the relationship of $K_D^{non-target}$ to N was assumed to be independent of RVD composition. The results were similar to those for scenario one, with specificity reaching a maximum at between 16 and 25 RVDs across a wide range of rates of decay ($k$). In this scenario, however, the plot shows that a lower rate of decay causes the specificity to reach a maximum at a larger $N$ (Figure 6D).

Thus, in either scenario, each representing a broad expansion of our experimental data using a range of hypothetical values to represent possible effects of RVD composition, specificity increases then diminishes with increasing numbers of RVDs, peaking (using the outer bounds of both scenarios) between 16 and 25 RVDs. Binding assays using an exhaustive set of dTALEs, beyond the scope of this study, would be necessary to discern whether scenario one or two, or a combination, best describes the relationship of specificity to number of RVDs across TALEs of diverse RVD compositions, but since both scenarios demonstrate a specificity maximum within a similar range, we conclude that the behavior we observed using our five dTALEs is generalizable and reflects a fundamental property of TAL effector–DNA interactions.

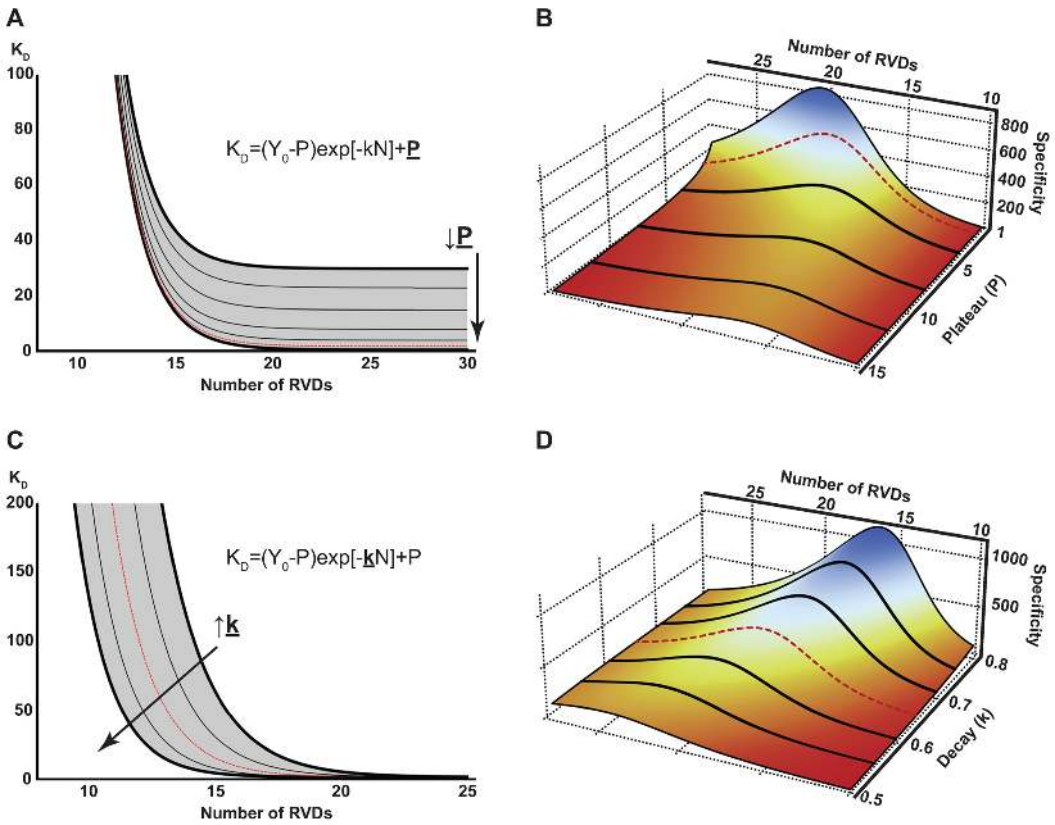


**Figure 6.** Generalized models of the relationship between number of RVDs and specificity**.** Calculations were performed based on our FA assays for binding of target (Figure 3B and Supplementary Table S4) and non-target DNA (Supplementary Figure S8 and Supplementary Table S7) by manipulating the parameters (rate of decay, *k*, and dissociation constant at the plateau or saturation level, *P*) of the exponential decay curves fitting the data for target DNA binding in Figure 3B (Supplementary Table S4). (**A**) The effect on the relationship of number of RVDs and affinity for target DNA of variation in saturation level, *P*, with rate of decay, *k*, held constant. The shaded area represents values of P between 1 and 15 nM (black lines). (**B**) Variation in TALE-DNA binding specificity deriving from the model in (A), calculated as the ratio of the dissociation constants for non-target and target DNA. For this, the equation fitting the relationship $K_D$ x #RVDs for non-target DNA in Supplementary Figure S7 (Supplementary Table S7) was divided by the equation fitting the data in Figure 3B (Supplementary Table S4). *N* (number of RVDs) and *P* (plateau) in the equation representing the target DNA binding shown in (A) were varied together with *N* in the equation for non-target DNA binding, resulting in a 3D plot. Red lines indicate the actual experimental curves. (**C**) The effect on the relationship of number of RVDs and affinity for target DNA of variation in *k*, with *P* kept constant. The shaded area represents values of *k* between 0.5 and 0.8. (black lines). (**D**) Variation in TALE–DNA binding specificity deriving from the model in (C), calculated as the ratio between the dissociation constants for non-target and target DNA. The equation fitting the relationship $K_D \times$ #RVDs for non-target DNA in Supplementary Figure S8 was divided by the equation fitting the data in Figure 3B (Supplementary Table S4). *N* (number of RVDs) and *k* (decay) in the equation representing the target DNA binding shown in (C) were varied together with N in the equation for non-target DNA binding. Red lines indicate the actual experimental curve values. Plots were generated using the software Mathematica v. 11.0 (Wolfram Research, Inc., Champaign, IL, USA).

### The length distribution of known TALEs in nature suggests selection for maximum specificity

Interestingly, the mean and median lengths of 114 native TALEs examined by Boch and Bonas in 2010 (1) (19 and 18 RVDs, respectively) fall within the range of lengths for maximum specificity derived from our results. Updating that analysis, we examined sequences now available for 213 *Xanthomonas* TALEs (Figure 7). The lengths are distributed in roughly Gaussian fashion with a mean of 18.5 RVDs and a median of 18 RVDs. This observation suggests that TALEs as a group are subject to length selection to achieve high specificity. The still relatively small number of validated, native TALE–target pairs all contain one or more RVD-nucleotide mismatches, in contrast with the target interactions we examined, which were perfectly matching. The presence of mismatches reduces the number of RVDs effectively contributing to the target DNA interaction and might in part explain the presence of natural TALEs with larger numbers of repeats, some as long as 34 RVDs.

## CONCLUSIONS

Our finding that TALE specificity peaks then declines with increasing length provides fundamental insight into the nature of TALE-DNA interactions. Also, this finding, and the distribution of TALE lengths in nature, suggest that natural selection can act not only on TALE RVD composition but also on CRR length as a pathogen adapts to activate host genes that promote infection (i.e. disease susceptibility genes), while avoiding any that impede it (e.g. disease resistance genes). Additionally, the results of our study inform the design of TALEs for biotechnology. Though refinements are needed in our understanding of the relationship of RVD composition to the length-dependence of specificity, particularly the effect of RVD composition on rate

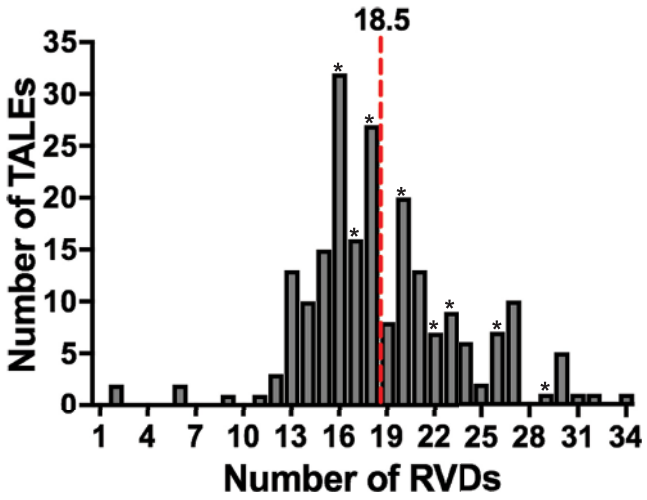


**Figure 7.** Distribution of the number of RVDs found in native TALEs. Two-hundred and thirteen distinct, native TALEs for which sequences were available are included (Supplementary Table S9). The red, vertical, dashed line indicates the mean. The median is 18 RVDs. Asterisks represent the number of RVDs of one or more native TALEs known to contribute to virulence.

of decay and plateau, our results suggest generally that it is important to design TALE-based DNA-targeting proteins with strong RVDs in the N-terminal part of the CRR, since the incorporation of many weak RVDs may not provide enough binding energy (with increasing CRR length) to reach the affinity threshold necessary for functional binding before affinity saturates and specificity decreases. An example in which this might be particularly important is TALE-based visualization of specific genomic loci, where a suitably low fluorescence background might only be achieved by the use of TALE-based probes with relatively short CRRs.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors thank A. Cernadas for making pAC-EXP41c.

## FUNDING

National Institutes of Health [GM9886 to A.B. and GM105691 to B.L.S.]. Funding for open access charge: National Institutes of Health [GM9886].

*Conflict of interest statement.* B.L.S. is a Senior Executive Editor for *Nucleic Acids Research*.

## REFERENCES

1. Boch,J. and Bonas,U. (2010) *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–436.
2. Bogdanove,A., Schornack,S. and Lahaye,T. (2010) TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.*, **13**, 394–401.
3. Bogdanove,A.J. (2014) Principles and applications of TAL effectors for plant physiology and metabolism. *Curr. Opin. Plant Biol.*, **19**, 99–104.
4. Cernadas,R.A., Doyle,E.L., Nino-Liu,D.O., Wilkins,K.E., Bancroft,T., Wang,L., Schmidt,C.L., Caldo,R., Yang,B., White,F.F. *et al.* (2014) Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.*, **10**, e1003972.
5. Boch,J., Bonas,U. and Lahaye,T. (2014) TAL effectors–pathogen strategies and plant resistance engineering. *New Phytol.*, **204**, 823–832.
6. Moscou,M. and Bogdanove,A. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science (New York, N.Y.)*, **326**, 1501.
7. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (New York, N.Y.)*, **326**, 1509–1512.
8. Mak,A., Bradley,P., Cernadas,R., Bogdanove,A. and Stoddard,B. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, **335**, 716–719.
9. Deng,D., Yan,C., Pan,X., Mahfouz,M., Wang,J., Zhu,J.-K., Shi,Y. and Yan,N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (New York, N.Y.)*, **335**, 720–723.
10. Stella,S., Molina,R., Yefimenko,I., Prieto,J., Silva,G., Bertonati,C., Juillerat,A., Duchateau,P. and Montoya,G. (2013) Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1707–1716.
11. Gao,H., Wu,X., Chai,J. and Han,Z. (2012) Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.*, **22**, 1716–1720.
12. Schreiber,T. and Bonas,U. (2014) Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res.*, **42**, 7160–7169.
13. Doyle,E.L., Hummel,A.W., Demorest,Z.L., Starker,C.G., Voytas,D.F., Bradley,P. and Bogdanove,A.J. (2013) TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1. *PLOS ONE*, **8**, e82120.
14. de Lange,O., Binder,A. and Lahaye,T. (2014) From dead leaf, to new life: TAL effectors as tools for synthetic biology. *Plant J.*, **78**, 753–771.
15. Yu,Y., Streubel,J., Balzergue,S., Champion,A., Boch,J., Koebnik,R., Feng,J., Verdier,V. and Szurek,B. (2011) Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv.oryzae depends on a new TAL effector that induces the rice nodulin-3 *Os11N3* gene. *Mol. Plant Microbe Interact.*, **24**, 1102–1113.
16. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
17. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
18. Sander,J.D., Cade,L., Khayter,C., Reyon,D., Peterson,R.T., Joung,J.K. and Yeh,J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.
19. Li,T., Liu,B., Spalding,M.H., Weeks,D.P. and Yang,B. (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.*, **30**, 390–392.
20. Morbitzer,R., Romer,P., Boch,J. and Lahaye,T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21617–21622.
21. Crocker,J. and Stern,D.L. (2013) TALE-mediated modulation of transcriptional enhancers in vivo. *Nat. Methods*, **10**, 762.
22. Blount,B.A., Weenink,T., Vasylechko,S. and Ellis,T. (2012) Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLOS ONE*, **7**, e33279.
23. Mahfouz,M.M., Li,L.X., Piatek,M., Fang,X.Y., Mansour,H., Bangarusamy,D.K. and Zhu,J.K. (2012) Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol. Biol.*, **78**, 311–321.
24. Politz,M.C., Copeland,M.F. and Pfleger,B.F. (2013) Artificial repressors for controlling gene expression in bacteria. *Chem. Commun.*, **49**, 4325–4327.

25. Garg,A., Lohmueller,J., Silver,P. and Armel,T. (2012) Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.*, **40**, 7584–7595.
26. Boissel,S., Jarjour,J., Astrakhan,A., Adey,A., Gouble,A., Duchateau,P., Shendure,J., Stoddard,B.L., Certo,M.T., Baker,D. *et al.* (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res.*, **42**, 2591–2601.
27. Doyle,E.L., Stoddard,B.L., Voytas,D.F. and Bogdanove,A.J. (2013) TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.*, **23**, 390–398.
28. Streubel,J., Blücher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593–595.
29. Cong,L., Zhou,R., Kuo,Y.-C., Cunniff,M. and Zhang,F. (2012) Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.*, **3**, 968.
30. Meckler,J.F., Bhakta,M.S., Kim,M.S., Ovadia,R., Habrian,C.H., Zykovich,A., Yu,A., Lockwood,S.H., Morbitzer,R., Elsaesser,J. *et al.* (2013) Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.*, **41**, 4118–4128.
31. Yang,J., Zhang,Y., Yuan,P., Zhou,Y., Cai,C., Ren,Q., Wen,D., Chu,C., Qi,H. and Wei,W. (2014) Complete decoding of TAL effectors for DNA recognition. *Cell Res.*, **24**, 628–631.
32. Miller,J.C., Zhang,L., Xia,D.F., Campo,J.J., Ankoudinova,I.V., Guschin,D.Y., Babiarz,J.E., Meng,X., Hinkley,S.J., Lam,S.C. *et al.* (2015) Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat. Methods*, **12**, 465–471.
33. Rogers,J.M., Barrera,L.A., Reyon,D., Sander,J.D., Kellis,M., Joung,J.K. and Bulyk,M.L. (2015) Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat. Commun.*, **6**, 7440.
34. Juillerat,A., Dubois,G., Valton,J., Thomas,S., Stella,S., Marechal,A., Langevin,S., Benomari,N., Bertonati,C., Silva,G.H. *et al.* (2014) Comprehensive analysis of the specificity of transcription activator-like effector nucleases. *Nucleic Acids Res.*, **42**, 5390–5402.
35. Schornack,S., Minsavage,G.V., Stall,R.E., Jones,J.B. and Lahaye,T. (2008) Characterization of AvrHah1, a novel AvrBs3-like effector from *Xanthomonas gardneri* with virulence and avirulence activity. *New Phytol.*, **179**, 546–556.
36. Cermak,T., Doyle,E., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J., Somia,N., Bogdanove,A. and Voytas,D. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
37. Zhao,L., Pellenz,S. and Stoddard,B.L. (2009) Activity and specificity of the bacterial PD-(D/E)XK homing endonuclease I-Ssp6803I. *J. Mol. Biol.*, **385**, 1498–1510.
38. Zhao,L. and Stoddard,B.L. (2014) Rapid determination of homing endonuclease DNA binding specificity profile. *Methods Mol. Biol.*, **1123**, 127–134.
39. Schreiber,T., Sorgatz,A., List,F., Bluher,D., Thieme,S., Wilmanns,M. and Bonas,U. (2015) Refined requirements for protein regions important for activity of the TALE AvrBs3. *PLOS ONE*, **10**, e0120214.
40. Cuculis,L., Abil,Z., Zhao,H. and Schroeder,C.M. (2016) TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.*, **12**, 831–837.
41. Jankele,R. and Svoboda,P. (2014) TAL effectors: tools for DNA targeting. *Brief. Funct. Genomics*, **13**, 409–419.
42. Maeder,M.L., Linder,S.J., Reyon,D., Angstman,J.F., Fu,Y.F., Sander,J.D. and Joung,J.K. (2013) Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods*, **10**, 243–245.
43. Perez-Quintero,A.L., Rodriguez,R.L., Dereeper,A., Lopez,C., Koebnik,R., Szurek,B. and Cunnac,S. (2013) An improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLOS ONE*, **8**, e68464.
44. Fine,E.J., Cradick,T.J., Zhao,C.L., Lin,Y. and Bao,G. (2014) An online bioinformatics tool predicts zinc finger and TALE nuclease off-target cleavage. *Nucleic Acids Res.*, **42**, e42.
45. Yang,Y., Sass,L.E., Du,C., Hsieh,P. and Erie,D.A. (2005) Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions. *Nucleic Acids Res.*, **33**, 4322–4334.
46. Wicky,B.I., Stenta,M. and Dal Peraro,M. (2013) TAL effectors specificity stems from negative discrimination. *PLOS ONE*, **8**, e80261.
47. Cuculis,L., Abil,Z., Zhao,H. and Schroeder,C.M. (2015) Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat. Commun.*, **6**, 7277.
48. Stella,S., Molina,R., Lopez-Mendez,B., Juillerat,A., Bertonati,C., Daboussi,F., Campos-Olivas,R., Duchateau,P. and Montoya,G. (2014) BuD, a helix-loop-helix DNA-binding domain for genome modification. *Acta Crystallogr. D Biol. Crystallogr.*, **70**, 2042–2052.