1406–1422 *Nucleic Acids Research, 2020, Vol. 48, No. 3* *Published online 21 December 2019*
doi: 10.1093/nar/gkz1174

Divalent cations promote TALE DNA-binding specificity

Luke Cuculis[1,†], Chuankai Zhao[2,†], Zhanar Abil[3], Huimin Zhao[1,2,3,4,5], Diwakar Shukla[2,5,6,7,8,*] and Charles M. Schroeder[1,2,4,5,8,*]

[1]Department of Chemistry, Urbana, IL 61801, USA, [2]Department of Chemical and Biomolecular Engineering, Urbana, IL 61801, USA, [3]Department of Biochemistry, Urbana, IL 61801, USA, [4]Carl R. Woese Institute for Genomic Biology, Urbana, IL 61801, USA, [5]Center for Biophysics and Quantitative Biology, Urbana, IL 61801, USA, [6]National Center for Supercomputing Applications, Urbana, IL 61801, USA, [7]NIH Center for Macromolecular Modeling and Bioinformatics, Urbana, IL 61801, USA and [8]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received August 28, 2019; Revised November 18, 2019; Editorial Decision December 04, 2019; Accepted December 06, 2019

ABSTRACT

Recent advances in gene editing have been enabled by programmable nucleases such as transcription activator-like effector nucleases (TALENs) and CRISPR–Cas9. However, several open questions remain regarding the molecular machinery in these systems, including fundamental search and binding behavior as well as role of off-target binding and specificity. In order to achieve efficient and specific cleavage at target sites, a high degree of target site discrimination must be demonstrated for gene editing applications. In this work, we studied the binding affinity and specificity for a series of TALE proteins under a variety of solution conditions using *in vitro* fluorescence methods and molecular dynamics (MD) simulations. Remarkably, we identified that TALEs demonstrate high sequence specificity only upon addition of small amounts of certain divalent cations (Mg^{2+}, Ca^{2+}). However, under purely monovalent salt conditions (K^+, Na^+), TALEs bind to specific and non-specific DNA with nearly equal affinity. Divalent cations preferentially bind to DNA over monovalent cations, which attenuates non-specific interactions between TALEs and DNA and further stabilizes specific interactions. Overall, these results uncover new mechanistic insights into the binding action of TALEs and further provide potential avenues for engineering and application of TALE- or TALEN-based systems for genome editing and regulation.

INTRODUCTION

Transcription activator-like effector (TALE) proteins are a unique class of DNA-binding proteins that have found broad use in genomic engineering (1). TALEs are naturally secreted by *Xanthomonas* bacteria as a means of facilitating infection of plant cells (2). In recent years, TALEs have garnered a keen interest due to their modular, programmable DNA-binding domain that allows for targeted binding to DNA (1,3,4). Natural TALE proteins consist of an N-terminal region (NTR) that facilitates DNA binding (5,6), a central repeat domain (CRD) that confers sequence specificity (7,8), and a C-terminal region (CTR) that contains nuclear localization signals (9,10) (Figure 1A, B). The CRD is comprised of a series of tandem repeats of conserved modules consisting of 33–35 residues. Each repeat in the CRD contains two residues located at positions 12 and 13, known as the repeat variable diresidue (RVD), which specify the nucleotide binding partner of the repeat (1). In 2009, the binding code of RVDs was elucidated, thereby enabling the facile engineering of TALEs for sequence-specific binding of DNA (7,8). For direct gene editing applications, TALE proteins can be fused to a nuclease domain such as *Fok*I, thereby generating a powerful molecular tool known as the TALEN system (7).

Despite the widespread use in a variety of gene editing and regulation applications (1,4,11), a complete understanding of TALE DNA-binding specificity has not yet been achieved. In prior work, there are a few studies focused on comparing the binding affinity of TALEs to specific versus non-specific DNA (5,12–16). In some cases, these studies have provided conflicting results, with some showing two or more orders of magnitude difference in specific versus non-specific binding affinity (12–15), while others show less

*To whom correspondence should be addressed. Tel: +1 2173333906; Fax: +1 2173335052; Email: cms@illinois.edu
Correspondence may also be addressed to Diwakar Shukla. Email: diwakar@illinois.edu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Present address: Zhanar Abil, Department of Biochemistry, University of Texas at Austin, Austin, TX 78712, USA.

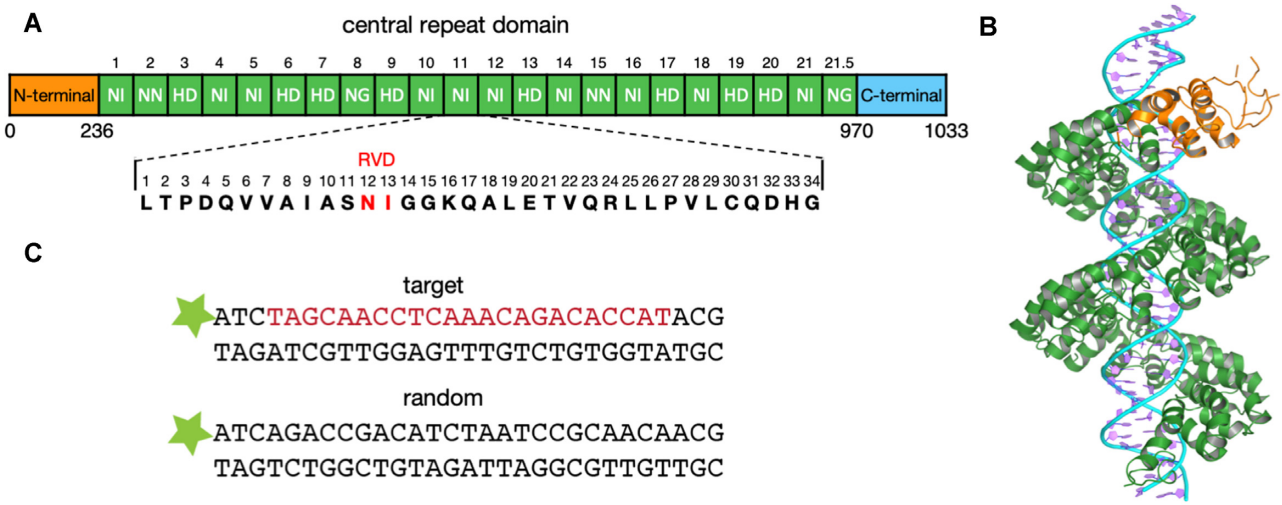

Figure 1. Crystal structure of TALE protein bound to target DNA and design of fluorescence anisotropy experiments. (**A**) General schematic of the 21.5 repeat TALE protein construct used in our work and an expanded view of one repeat sequence in the central repeat domain (CRD). The repeat variable diresidues in all repeats are shown in the boxes. The lengths of the N-terminal, the central and the C-terminal domains are specified. (**B**) TALE PthXo1 bound to a 36 bp target DNA sequence (PDB ID: 3UGM (23)), with the NTR shaded in orange and the CRD shaded in green. (**C**) DNA templates used in fluorescence anisotropy experiments. The binding site for the 21.5 repeat TALE construct used in this work is shown in red and the non-specific elements are shown in black. The green star represents the fluorescein dye molecule conjugated to the 5' terminus of DNA.

than an order of magnitude difference (5,16). Only in those studies showing the greatest difference between specific and non-specific binding does the biological activity and gene engineering utility of TALEs appear plausible. Similar to TALEs, there have been reports of off-target binding effects for the other cutting-edge gene editing tools, including the CRISPR–Cas9 system (17–20) and zinc finger nucleases (21,22). For precise gene editing, the optimal tool must introduce specific and intended modifications exclusively at the target site while leaving the rest of genome unmodified. From this perspective, it is essential to understand the factors that influence TALE binding affinity and specificity.

Prior experimental work has probed a few key factors contributing to TALE DNA-binding specificity. Recent work by Gao *et al.* showed that the NTR is critical for TALE binding, and that the NTR is capable of binding to DNA on its own in a non-specific manner (5). Schreiber *et al.* corroborated these results by demonstrating the decline in TALE activity as the NTR was truncated to shorter and shorter lengths (6). Several studies have shown that TALEs with longer CRDs have higher overall DNA-binding affinity and specificity in a genomic context (24–26). Meckler *et al.* highlighted that the choice of RVDs in the CRD can affect TALE specific binding affinity by several orders of magnitude (27). Recent single molecule fluorescence microscopy studies of TALE binding and non-specific search dynamics by Cuculis *et al.* confirmed both the necessity of the NTR for nucleation of TALE binding events, while also demonstrating the non-specific one-dimensional search abilities of NTR-only truncation mutants (28,29). Furthermore, it was found that TALEs utilize a rotationally decoupled mechanism for non-specific search along the DNA double helix, wherein the TALE CRD repeats engage DNA in a fully wrapped conformation during search, yet remain loosely associated with the DNA backbone during the process (29,30). Rinaldi *et al.* investigated the effect of increasing number of repeats on TALE affinity for target and non-target DNA in monovalent salt solutions, which showed that TALE specificity increases then diminishes with increasing TALE length (15). Geiger-Schuller *et al.* reported that TALEs undergo partial unfolding prior to DNA binding and increasing number of repeats slows down conformational change rate thereby affecting binding kinetics (31). Finally, the requirement of a thymine at the 5' end of the binding site was well characterized and found to significantly impact the binding of TALEs to target sequences (2,3,7,23).

A recent computational study on the energetics of TALE target binding suggested that TALE specificity arises mostly from negative discrimination by the RVDs (i.e. the 'least bad' binding site) (32). In this study, the major contributors to TALE binding to target sites are determined to be non-specific interactions between the lysine and glutamine residues at positions 16 and 17 within each repeat and the DNA backbone (32). The second largest contributor to binding energy is the NTR via an entirely nonspecific interaction. Specific interactions between the RVD (largely influenced by position 13 in each repeat, as the residue at position 12 is mainly a modulator of repeat structure) are suggested to be only a distant third contributor to overall TALE binding affinity (32). The authors postulated that TALE specificity is thus largely a result of steric clashes involving residue 13 of repeats in the RVD, and therefore described TALE specificity as arising from negative discrimination (32). Despite this understanding, however, several open questions still surround TALE specificity. Given the large amount of binding energy attributed to non-specific interactions, how do ionic strength and ionic species present in solution affect TALE specificity? How does the low contribution of specific binding to overall binding energy permit successful target search and localization? Given a negative discrimination model for TALE binding, how can ex-

perimental results showing similar binding activities for specific and non-specific binding be reconciled?

The effects of solution conditions (ionic strength and ion type) have been extensively investigated for several other sequence-specific DNA-binding proteins, while the effects of these factors on TALE specificity are not fully understood. One recurring theme in modulation of DNA-binding protein specificity is the suppression or enhancement of non-specific binding, which is mostly contributed by electrostatic interactions (33–39). Several sequence-specific DNA-binding proteins demonstrate minimal target site discrimination under low ionic strength conditions, when electrostatic interactions are not significantly screened (33–36). Indeed, these proteins display substantial target site specificity under elevated ionic strength conditions when non-specific binding affinity is suppressed. In some cases, sufficient electrostatic screening can be satisfied by higher concentrations of monovalent salts (33,34). For example, a fluorescence anisotropy study of tumor suppressor p53 binding to half and full target sites, as well as non-specific substrates, showed that a total ionic strength of 175 mM (NaCl and buffer agents) is required for apparent target discrimination (33). Several studies on basic leucine zipper proteins (B-ZIPs), including CREB and Fos-Jun (35), as well as the prokaryotic transcription factor NikR(36), also highlighted the roles of divalent cations in their binding specificity. In the case of B-ZIP proteins, no discernible sequence specificity is achieved even under elevated monovalent salt concentrations, and addition of a small amount of divalent salt nearly abolishes non-specific binding while only slightly impacting target binding (35,36). Several restriction enzymes, including BamHI (37), EcoRV (38) and EcoRI (39), have also been shown to require Mg^{2+} or Ca^{2+} for sequence-specific binding, though such divalent cations are also required for their enzymatic activity.

In this work, we presented an in-depth experimental and computational study of the effects of ionic strength and ionic species on TALE binding to DNA. Using an *in vitro* fluorescence anisotropy assay, we examined the binding affinities of a series of TALE proteins to both specific and non-specific DNA sequences under a variety of solution conditions. Under purely monovalent salt conditions, we found that TALE specificity is prohibitively minimal, especially for longer TALE proteins (here, a 21.5 repeat TALE construct). Strikingly, we observed a substantial degree of binding specificity upon addition of small amounts of certain divalent cations, such that non-specific binding is nearly abolished. The effects of divalent cations on binding specificity are enhanced in TALE constructs with larger CRDs. For the 21.5 repeat TALE protein, we also observed a slight enhancement of target binding, leading to an even greater target discrimination. Moreover, we found that the non-specific binding of NTR-only truncation mutants is significantly attenuated by divalent salts but not monovalent salts. To directly complement experimental results, we performed molecular dynamics (MD) simulations to quantify the preferential interactions between TALE and DNA in the presence of monovalent and divalent salts. Using MD simulations, we quantified the change in TALE binding free energy due to the presence of salts. Our results suggest that divalent cations preferentially bind to DNA, whereas monovalent cations are preferentially excluded from DNA. The significant attenuation of NTR binding by divalent cations is caused by the DNA-binding competition between the NTR and divalent cations, and the non-specific contacts between TALE CRD and DNA backbone are attenuated by divalent cations while some specific contacts may be stabilized by divalent cations. Taken together, these results elucidate the effects of divalent cations on TALE binding affinity and specificity, which can potentially be useful for future application of TALENs for gene editing and regulation.

MATERIALS AND METHODS

Preparation and cloning of tSCA21.5 and Naldt-tSCA21.5

In this work, we studied a series of AvrXa10 (40)-based TALE proteins, which were redesigned for editing the human β-globin gene containing a mutation associated with sickle cell disease. (41) We used this parent protein (containing 21.5 repeats) to design several TALEs with variable numbers of repeats in the CRD. The gene encoding for the untagged TALE protein (tSCA21.5) was assembled using the Golden Gate cloning method (Addgene TALEN Kit #1000000024, as described in our prior work (33)). For this purpose, repeats 1-10 of the TALE were assembled into the pFUS_A30A vector, repeats 11-20 into the pFUS_A30B vector, and repeat 21 was inserted into the pFUS_B1 vector using BsaI-HF (New England Biolas (NEB)) and T4 DNA ligase (NEB). In this reaction, the temperature was cycled 10 times between 37°C (5 minutes) and 16°C (10 minutes), after which the reaction was treated with Plasmid-Safe nuclease (Epicentre Biotechnologies). DH5α (Cell Media Facility at the University of Illinois at Urbana-Champaign) transformants were assayed for proper assembly using restriction digestion of the purified (Qiagen) plasmids. Next, together with the plasmid coding for the last half-repeat pLR(NG), the repeats were further assembled into a specifically engineered destination vector pET28-GG-TALE using BsmBI (Fisher Scientific) and T4 DNA ligase. The destination vector contained an N-terminal His-tag and flanking N- (208 aa) and C- (63 aa) terminal regions of the TALE, as well as the BsmBI sites corresponding to the kit BsmBI sites (see sequence below). For fluorescent labeling of TALEs, we modified the original plasmid pET-tSCA21.5 with an oligonucleotide inserted encoding an N-terminal LCPTSR hexapeptide (aldehyde tag (34,42)) upstream of the His-tag. To this end, the plasmid was amplified in fragments containing the insert and assembled using the Gibson Assembly Kit (NEB), thereby generating the construct pET-Naldt-tSCA21.5. Similar to prior single molecule experiments, TALEs used in fluorescence anisotropy experiments contained a genetically encoded aldehyde tag, which showed no difference in binding affinity compared to wild type TALEs (28).

Protein sequence of Naldt-tSCA21.5

MGPLCTPSRSSHHHHHHSSGLVPRGSHMLDTS
LLDSMPAVGTPHTAAAPAECDEVQSGLRAA
DDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQ
HHEALVGHGFTHAHIVALSQHPAALGTVAV

TYQDIIRALPEATHEDIVGVGKQWSGARALEA
LLTEAGELRGPPLQLDTGQLLKIAKRGGVTAV
EAVHAWRNALTGAPLNLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHG
LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNG GGKQALESIVAQLSRPDPALA
ALTNDHLVALACLGGRPALDAVKKGLPHAPELIR
RINRRIPERTSHRVA

Preparation and cloning of Naldt-tSCA15.5 and Naldt-tSCA11.5

In order to construct TALEs containing only the first 15.5 and 11.5 repeats of the parent TALE with 21.5 repeats, we first engineered a destination vector pET28-Naldt-GG-TALE containing the aldehyde tag upstream of the His-tag using the Gibson Assembly Master Mix (NEB). Naldt-tSCA15.5 and Nald-tSCA11.5 were assembled using the Golden Gate cloning method (33) into the destination vector pET28-Naldt-GG-TALE. To this end, repeats 1–10 were assembled into vector pFUS_A, repeats 11–15 were assembled into pFUS_B5, and repeat 11 was inserted into vector pFUS_B1 using the BsaI-HF and T4 DNA ligase by cycling the temperature 10 times between 37°C (5 min) and 16°C (10 min), after which the reactions were treated with Plasmid-Safe nuclease. DH5α transformants were assayed for proper assembly using restriction digestion of the purified plasmids. pFUSA containing the first 10 repeats was assembled with either pFUS-B5 (for TALE15.5) or pFUS-B1 (for TALE11.5) together with the plasmid coding for the last half-repeat into the pET28-Naldt-GG-TALE vector using BsmBI and T4 DNA ligase.

Cloning of Naldt-NTR

To generate the NTR truncation mutant, we digested pET-Naldt-tSCA21.5 with SacII and HindIII and inserted the sequence: CAAAGCGTGGTGGCGTGACCGCGG TGGAAGCGGTCCATGCCTGGCGTAATGCGTTG ACGGGCGCCCCCCTGAACTAAGTCAGATAACC GGATACAGACAAGCTTGCGGCCGCACTCGAGC ACCAC. This sequence was synthesized as a gBlock Gene Fragment (IDT), using the Gibson Assembly Master Mix (NEB).

Protein sequence of Naldt-NTR

MGPLCTPSRSSHHHHHHSSGLVPRGSHMLDTSLL
DSMPAVGTPHTAAAPAECDEVQSGLRAADDPPPT
VRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVD
LRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGF
THAHIVALSQHPAALGTVAVTYQDIIRALPEATH
EDIVGVGKQWSGARALEALLTEAGELRGPPLQLD
TGQLLKIAKRGGVTAVEAVHAWRNALTGAPLN

Protein expression and purification

BL21 (DE3) electrocompetent *Escherichia coli* cells were co-transformed with plasmids encoding TALE constructs and the pBAD-FGE plasmid (generous gift of Dr Taekjip Ha, University of Illinois at Urbana-Champaign). A single colony was grown in 5 ml lysogeny broth (LB) supplemented with 100 μg/ml ampicillin and 25 μg/ml kanamycin as a seeding culture until saturation, and subsequently cultured in 500 ml of Terrific Broth at 37°C and 250 revolutions per minute (RPM) with the corresponding antibiotics until the cell culture reached an optical density OD_{600} of 0.3–0.4. FGE expression was induced with 0.2% L-arabinose (Sigma), and the culture was grown further at 37°C 250 RPM until reaching OD_{600} 0.7–0.8. TAL effector expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma). The proteins were expressed overnight at 16°C in an incubated shaker at 250 RPM. Following expression, cell cultures were centrifuged at 4000 RPM at 4°C for 15 min, and the pellets were re-suspended in 10–20 ml of lysis buffer (25 mM Tris–HCl (Fisher) pH 7.5, 300 mM NaCl (Fisher), 0.5% Triton-X100 (Sigma), 5% glycerol (Sigma), 4 U/ml DNase I (NEB), 0.3 mM phenylmethanesulfonylfluoride (Sigma), 1 mg/ml lysozyme (Sigma)). The cells were lysed by sonication for 20 min total with alternating 5 s pulse/5 s rest, and cell debris were centrifuged at 13 000 RPM for 20 min at 4°C. The His-tagged TALEs were purified using AKTA pure chromatography system (GE Healthcare) with a 1 ml HisTrap column (GE Healthcare). The cleared lysate was loaded on the column, washed with 20 mM Tris–HCl pH 7.5, 500 mM NaCl and 20 mM imidazole (Sigma), and eluted with 20 mM Tris–HCl pH 7.5, 500 mM NaCl and 250 mM imidazole. The purified protein was dialysed in 2 L 50 mM phosphate buffer pH 7-8.4 (depending on the predicted protein isoelectric point), 500 mM NaCl at 4°C overnight. When necessary, the protein was further purified using a 16/600 Superdex 200 pg gel filtration column (GE Healthcare) using 50 mM phosphate buffer pH 7–8.4 (depending on the predicted protein isoelectric point), and 500 mM NaCl.

DNA templates

DNA templates were prepared using the following method. Single stranded oligonucleotides (Integrated DNA Technologies) were first prepared in stock solutions in low salt buffer (20 mM Tris–HCl, pH 7.5, 50 mM KCl). DNA concentrations were determined via UV/vis spectrophotometry

(NanoDrop, Thermo Fisher). Equivalent concentrations of a 5,6-FAM (fluorescein derivative) labeled oligonucleotides and their corresponding complimentary oligonucleotides were combined into solution in low salt buffer, placed in a covered heating block at 90°C for 2 min, followed by slow cooling to room temperature.

Fluorescence anisotropy experiments

Mixtures of 1 nM double stranded (ds) oligonucleotide templates and TALE protein were prepared in a standard buffer (20 mM Tris–HCl, pH 7.5) with varying amounts of monovalent salt (KCl or NaCl) and/or divalent salt ($MgCl_2$, $MgSO_4$, $ZnCl_2$, $MnSO_4$, $SrCl_2$, $CaCl_2$, where specified). Mixtures of TALE protein and DNA template (200 μl) were incubated for 10 min at room temperature (25°C) and then assayed in black 96-well plates (Corning) in duplicates. Fluorescence polarization measurements were performed on an Infinite 200 Pro microplate reader (Tecan) using excitation and emission wavelengths of 485 and 535 nm, respectively. Fluorescence polarization values were converted to fluorescence anisotropy values using the following relation:

$$A = \frac{2P}{3 - P} \tag{1}$$

where A is anisotropy and P is polarization. Fluorescence anisotropy values were determined by averaging over two repeated experiments and were plotted relative to anisotropy values in the absence of protein. The dissociation constant (K_d) was calculated using a non-linear least squares algorithm (Origin 8.5) using the following expression:

$$A = A_f + (A_b - A_f)\frac{(L_T + K_d + R_T) - \sqrt{(L_T + K_d + R_T)^2 - 4L_T R_T}}{2L_T} \tag{2}$$

where A is experimental anisotropy value, A_f is anisotropy of DNA without protein, A_b is anisotropy of protein-bound DNA, L_T is total ligand (DNA) concentration, and R_T is total receptor (protein) concentration. For experiments with unlabeled competitor DNA, 100 nM of DNA templates lacking the fluorescein dye (5,6-FAM) was added to solutions containing 1 nM labeled probe DNA.

Molecular dynamics simulation

We simulated the NTR, the CRD (containing 21 repeats), and DNA molecules (a 31 bp random DNA for the NTR and a 36 bp target DNA for the CRD) in their dissociated states, in addition to also simulating the DNA-bound NTR and CRD complexes in aqueous solutions with varying concentrations of KCl and/or $MgCl_2$. The crystal structure of TAL effector PthXo1 (PDB ID: 3UGM, Figure 1B) (23) from *Xanthomonas* was used for the molecular dynamics (MD) simulation studies. Missing parts of the NTR in the TALE crystal structure were added via SWISS-MODEL homology-modelling server (43). The NTR structure included the 1-240 residues. We aligned the modeled NTR structure and a 31 bp DNA (5'GACACCATTAGCAACC TCAAACAGACACCAT3') to the TALE crystal structure to get an initial structure of the NTR-DNA complex. The 1–21 repeats (the 289–1001 residues) in the CRD and the 36 bp DNA of the TALE crystal structure were used in MD simulations. The 36 bp DNA sequence in CRD-DNA complex was 5'TAGATATGCATCTCCCCCTACTGTACAC CACCAAAA3'.

All the simulations were setup in AmberTools16 (44) and performed using OpenMM 7.0 GPU platform (45). Each initial structure was solvated with TIP4P-EW water molecules in a cubic simulation box. The minimum distance from the boundary of the simulation box to the solute surface was at least 15 Å. Certain number of cations and anions were added to neutralize the system and match the specific salt solutions (Supplementary Table S1). Amber ff14SB force field parameters were used for the protein and DNA (with bsc1 modifications) (46). Monovalent ions parameters for TIP4P-EW water model by Joung *et al.* (47) were used for K^+ and Cl^-. 12–6–4 model with fine tuned C4 terms by Panteva *et al.* for divalent ions were used for Mg^{2+}, which were optimized to improve the approximation of interactions between divalent ions and nucleic acids in TIP4P-EW water box (48). After energy minimization, each system was running in NPT ensemble (300 K, 1 bar) for at least 50 ns with an integration time step of 2 ps. Monte Carlo barostat algorithm (49) was employed to maintain the constant pressure and Langevin integrator (50) was used to control the system temperature with a collision frequency of 1 ps^{-1}. Periodic boundary conditions were employed and the particle-mesh Ewald method (51) was used to treat the electrostatic interactions with a 10 Å cutoff distance. All the covalent bonds involving hydrogen atoms were constrained.

Preferential interaction coefficient calculation

The theory of computation of preferential interaction coefficient (Γ_{23}, Scatchard notation (52): 1—water, 2—solute, 3—salt) from MD simulations based on a statistical mechanical method with no adjustable parameters was developed by Baynes and Trout (53). The MD simulation data was used to calculate Γ_{23} using the following expression (53–58):

$$\Gamma_{23}(r, t) = n_3(r, t) - n_1(r, t)\left[\frac{n_3 - n_3(r, t)}{n_1 - n_1(r, t)}\right] \tag{3}$$

where $n_1(r, t)$, $n_3(r, t)$ are the number of water molecules and ions within distance r from solute surface at time step t, and n_1, n_3 are the total number of water molecules and ions in the simulation box. The average Γ_{23} over all time frames ($\langle\Gamma_{23}\rangle$) was determined as the final Γ_{23} value. To quantify the error bars on $\langle\Gamma_{23}\rangle$, we divided the individual MD trajectory into three blocks, and calculated the block average values of $\langle\Gamma_{23}\rangle$. The standard deviation of the block averages was determined as the error bar (54). For KCl or $MgCl_2$, the preferential interaction coefficient calculations require an estimate of Γ_{23} for both cation and anion, given by (54,59):

$$\Gamma_{2,KCl} = (\Gamma_{2,K} + \Gamma_{2,Cl} - |Z_2|)/2 \tag{4}$$

$$\Gamma_{2,MgCl_2} = (2\Gamma_{2,Mg} + \Gamma_{2,Cl} - |Z_2|)/2 \tag{5}$$

where $|Z_2|$ is the net charge of the solute.

Theory for quantifying the effect of salt on free energy of TALE binding

The theory is based on the previous method that estimates the effect of cosolutes on free energy of protein-protein binding (60). The protein–DNA binding free energy is defined by

$$\Delta G_b = G_{complex} - G_{DNA} - G_{protein} \tag{6}$$

Then the change in protein–DNA binding free energy ($\Delta\Delta G$) in the presence of salts in solution, compared to that in pure water, is given by

$$\Delta\Delta G_b = \Delta G_b^{salt} - \Delta G_b^{water} \tag{7}$$

Also, $\Delta\Delta G$ can be calculated by the following equation:

$$\Delta\Delta G_b = \Delta\mu_{complex}^{tr} - \Delta\mu_{DNA}^{tr} - \Delta\mu_{protein}^{tr} \tag{8}$$

where $\Delta\mu^{tr}$ is the transfer free energy that measures the change in interaction energy between the solute and water when the solute is transferred from pure water to a salt solution. The transfer free energy is equal to the change of chemical potential of the solute. Using Scatchard notation (1—water, 2—solute, 3—salt), the transfer free energy of protein is given by

$$\Delta\mu_2^{tr} = \mu_2^{salt} - \mu_2^{water} \tag{9}$$

where μ_2 is the chemical potential of the protein. Considering the process of adding a protein to a salt solution, the chemical potential of the salt is disturbed by the protein, $(\frac{\partial\mu_3}{\partial m_2})_{T,P,m_3}$, and the chemical potential of the protein is disturbed by the salt, $(\frac{\partial\mu_2}{\partial m_3})_{T,P,m_2}$. Here, m is molality, T and P are temperature and pressure, respectively. The total change in chemical potential of the protein due to the transfer to a salt solution is given by

$$\Delta\mu_2^{tr} = \int_0^{m_3} \left(\frac{\partial\mu_2}{\partial m_3}\right)_{T,P,m_2} dm_2 \tag{10}$$

$$= -\int_0^{m_3} \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \left(\frac{\partial m_3}{\partial m_2}\right)_{T,P,\mu_3} dm_3 \tag{11}$$

$$= -\int_0^{m_3} \left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} \Gamma_{23} dm_3 \tag{12}$$

where Γ_{23} is the preferential interaction coefficient. From a thermodynamic view, Γ_{23} is a measure of the number of salt molecules added to the solution per protein to keep the chemical potential of the salt constant. It is also a measure the excess number of salt molecules within the local domain around protein compared to the bulk. The derivative term $(\frac{\partial\mu_3}{\partial m_3})_{T,P,m_2}$ is the variation in the chemical potential of the salt as a function of its molality, which can be obtained from the experimental activity or osmolality data. Using the expression for chemical potential as a function of activity and the Gibbs–Duhem relationship to approximate the change in the activity of the salt in terms of the change in the activity of water:

$$\left(\frac{\partial\mu_3}{\partial m_3}\right)_{T,P,m_2} = RT\frac{\partial ln(a_3)}{\partial m_3} \tag{13}$$

$$\approx -RT\frac{m_1}{m_3}\frac{\partial ln(a_1)}{\partial m_3} \tag{14}$$

$$= -\frac{RT}{m_3}\frac{\partial Osm}{\partial m_3} \tag{15}$$

where Osm denotes the osmality of the salt solution. Further, assuming that Γ_{23} is a linear function of cosolute molality, the transfer free energy $\Delta\mu_2^{tr}$ can be approximated as

$$\Delta\mu_2^{tr} = -\int_0^{m_3} \frac{RT}{m_3}\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\left(\frac{\Gamma_{23}}{m_3}\right) m_3 dm_3 \tag{16}$$

$$= RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\left(\frac{\Gamma_{23}}{m_3}\right)\int_0^{m_3} dm_3 \tag{17}$$

$$= RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\Gamma_{23} \tag{18}$$

In the end, we can get the change in binding free energy $\Delta\Delta G_b$:

$$\Delta\Delta G_b = RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2} \times \left(\Gamma_{23}^{complex} - \Gamma_{23}^{DNA} - \Gamma_{23}^{protein}\right) \tag{19}$$

$$= RT\left(m_1\frac{\partial ln(a_1)}{\partial m_3}\right)_{T,P,m_2}\Delta\Gamma_{23} \tag{20}$$

Therefore, the experimental water activity or the osmolality data for the salt solution and the preferential interaction coefficients for the DNA, protein and complex are required to determine the $\Delta\Delta G_b$. The water activities in unsaturated KCl and $MgCl_2$ solutions have been previously reported (61,62). The water activity in salt solution is given by the following equation (61):

$$\ln(a_1) = -\phi M_1 \nu m_j \tag{21}$$

where ϕ is practical osmotic coefficient, M_1 is molar mass of water, ν is stoichiometric parameter. ν is the number of moles of ions produced by complete dissociation of one mole of salt, and for KCl ν equals 2 and for $MgCl_2$ ν equals 3.

The experimental osmotic coefficient and water activity data used in this study were summarized in Supplementary Table S2. The preferential interaction coefficients were calculated from MD simulation data.

RESULTS

TALEs show minimal target recognition in pure monovalent salt.

We utilized a fluorescence anisotropy assay in order to probe the role of monovalent and divalent salts on TALE binding specificity. To this end, we used both target and random DNA sequences to probe the affinity and specificity of TALEs containing 21.5, 15.5 and 11.5 repeats in the CRD

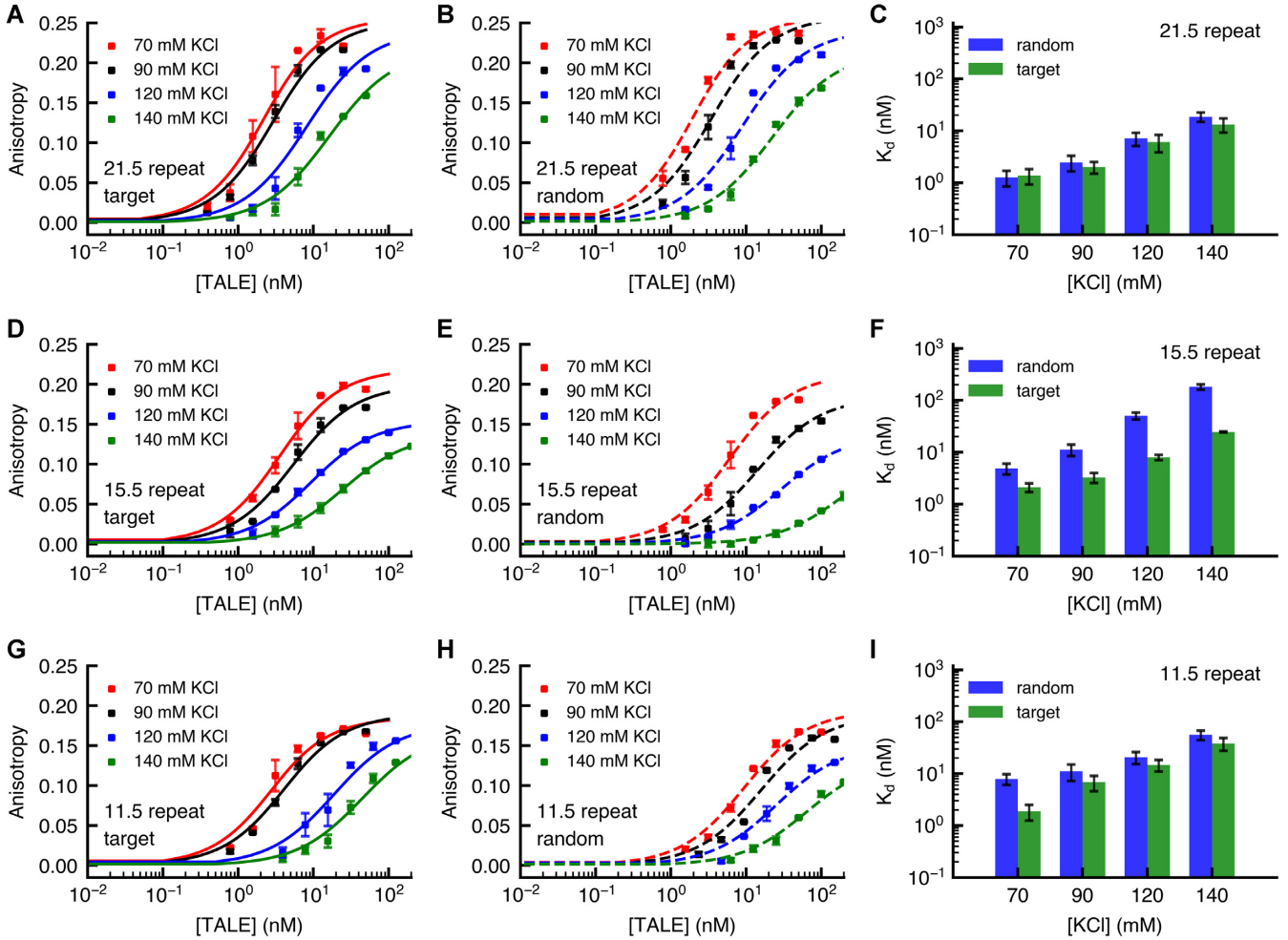

Figure 2. Effect of monovalent ions on TALE-DNA binding for 21.5 repeat, 15.5 repeat and 11.5 repeat TALEs. Binding of the 21.5 repeat TALE to (**A**) target (dots for raw data, solid lines for fitted curves) and (**B**) random (dots for raw data, dashed lines for fitted curves) DNA sequences measured by fluorescence anisotropy. (**C**) Equilibrium dissociation constants, $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$, determined for the 21.5 repeat TALE to random and target DNA sequences. Binding of the 15.5 repeat TALE to (**D**) target (dots for raw data, solid lines for fitted curves) and (**E**) random (dots for raw data, dashed lines for fitted curves) DNA sequences, and (**F**) $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ values. Binding of the 11.5 repeat TALE to (**G**) target (dots for raw data, solid lines for fitted curves) and (**H**) random (dots for raw data, dashed lines for fitted curves) DNA sequences, and (**I**) $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ values. TALE-DNA binding was measured in 20 mM Tris–Tris HCl buffer with different amounts of KCl. DNA templates were maintained at a constant concentration of 1 nM and TALE concentrations were adjusted. Duplicate experiments were performed to determine the mean and standard deviation of fluorescence anisotropy value.

(Figure 1C). We further compared these results to the binding data of an NTR-only TALE truncation mutant. In all cases, we characterized TALE binding affinities under a variety of solution conditions in terms of equilibrium dissociation constants, $K_{d,\text{non-specific}}$ for non-target binding and $K_{d,\text{specific}}$ for target binding (Supplementary Table S3), obtained from fluorescence anisotropy measurements. A small value of $K_{d,\text{non-specific}}$ or $K_{d,\text{specific}}$ suggests a strong association between TALEs and random or target DNA.

In pH-buffered aqueous solutions containing the monovalent salt KCl in the absence of divalent cations, TALE binding affinities to target and random DNA sequences appear to be similar (Figure 2). Strikingly, the apparent low degrees of specificity were observed for a series of different TALEs (21.5, 15.5 and 11.5 repeats) studied in this work and across a wide range of different concentrations of KCl and NaCl (Figure 2, Supplementary Figure S1). For the TALE containing 21.5 repeats in the CRD, we observed a significant lack of specificity between target and random DNA even under high concentrations of KCl (Figure 2A–C and Supplementary Figure S2). No difference in binding specificity was observed when the monovalent cation was exchanged for Na^+ in place of K^+ (Supplementary Figure S1). In the presence of monovalent salt, the 15.5 repeat TALE appears to show slightly greater discrimination between random and target sequences compared to both the 11.5 repeat and the 21.5 repeat TALEs (Figure 2). Our results are in agreement with the previous report that under pure monovalent salt solutions, TALE specificity increases and then decreases with increasing number of repeats, peaking between 15 and 19 repeats (15). Rinaldi *et al.* suggested that TALE binding affinities for both target and non-target DNA increases non-linearly as the number of repeats increases, with the gain decaying exponentially (15). How-

ever, the affinity gain for non-target DNA decays slower compared to that for target DNA (15). Nevertheless, there is only less than an order of magnitude preference for target DNA when comparing $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$, which is not enough for the specificity required for TALE and TALEN function *in vivo* (two or more orders of magnitude difference is required) (12–14).

Certain divalent cations enhance TALE binding specificity in an ionic strength dependent manner.

We further sought to understand the apparent lack of sequence specificity for TALE binding in monovalent salt by assessing the impact of ionic strength and solution conditions. Here, we first studied the 21.5 repeat TALE binding to target and random DNA in pH-buffered aqueous solutions containing 100 mM KCl and 10 mM various divalent salts ($MgCl_2$, $MgSO_4$, $CaCl_2$, $MnSO_4$, $SrCl_2$, $ZnCl_2$). Strikingly, specificity in TALE binding to target sites is restored in the presence of certain divalent cations (Figure 3A–E, Supplementary Figures S3 and S4). In particular, the divalent cations Mg^{2+} or Ca^{2+} substantially diminish non-target binding and simultaneously enhance the binding affinity to target sites (Figure 3E) compared to the binding in purely monovalent solutions containing 140 mM KCl (Figure 2C). We characterized TALE binding in the presence of a variety of divalent cations, and we observed the strongest effect with Mg^{2+} or Ca^{2+}, a moderate effect with Mn^{2+}, a slight effect with Sr^{2+}, and essentially no discrimination effect with Zn^{2+} (Figure 3A–E and Supplementary Figure S3). No differences in TALE binding were observed upon changing the monovalent anions or divalent anions (Supplementary Figure S4). Interestingly, these effects appear to be sensitive to the chemical identity of divalent cations, with an apparent correlation with the abundance of common cellular cations. However, this phenomenon is not entirely driven by ionic radius or any apparent particular periodic trend.

Next, we studied the effect of increasing the overall ionic strength on the binding behavior of the 21.5 repeat TALE (Supplementary Figure S5). Here, we maintained KCl concentration at 100 mM and increased divalent cation concentration from 5 to 10 mM. Our results show that TALE non-specific binding affinity generally decreases upon increasing the concentration of divalent cations, which is similar to TALE binding behavior in the presence of monovalent salt. Strikingly, TALE specific binding affinity is essentially unaffected by the increase in divalent cation concentration (Supplementary Figure S5). However, the decrease in DNA-binding affinity to off-target sites is significantly larger when the increase in total ionic strength results from a higher proportion of divalent cations (as opposed to pure monovalent cations).

Based on these results, we next studied the effect of increasing the proportion of divalent cations in solution while maintaining the total ionic strength at a constant value (90 or 120 mM). For all full length TALEs studied in this work (TALEs with 11.5, 15.5 and 21.5 repeats), we observed that TALE binding specificity increases for target DNA (corresponding to an increase in the difference between $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$) with increasing proportion of Mg^{2+} (Figure 3F–H, Supplementary Figures S6 and S7). For shorter TALEs (11.5 and 15.5 repeats), differences in binding affinity between random and target sites are minimal until approximately 40% of the total ionic strength is contributed by Mg^{2+} (Supplementary Figures S6 and S7). In contrast, the 21.5 repeat TALE shows higher specificity for target sites when the proportion of Mg^{2+} reaches similar levels (where the difference between $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ is >500-fold) (Figure 3H). In addition, the specific binding affinity for the 21.5 repeat TALE increases with increasing amount of divalent cations, in contrast to the slight decreases for both the 15.5 repeat and the 11.5 repeat TALEs. At higher total ionic strength with moderate proportion of Mg^{2+} (similar to physiological conditions), the 21.5 repeat TALE shows extremely high specificity (Supplementary Figure S8), consistent with prior studies that utilize TALEs for specific genomic edits (12–14). Interestingly, in buffered solutions containing only Mg^{2+} as the primary cation (40 mM $MgCl_2$), the 21.5 repeat TALE binds to target DNA with extremely high specificity (the ratio between $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ is approaching infinity, effectively as $K_{d,\text{non-specific}}$ is too large to measure) (Figure 3I–K). These results suggest that TALE binding specificity increases with higher concentrations of divalent cations, as opposed to proportional increases in ionic strength provided by monovalent cations only.

We further utilized a DNA competitor assay for fluorescence anisotropy in which a large molar excess (100-fold) of unlabelled random DNA was added to 1 nM of fluorescently labelled target probe DNA. We hypothesized that in the presence of a large excess of competitor (random) DNA, TALEs exhibiting relatively weak binding specificity would be sequestered away from target DNA, as they are unable to discriminate between target and random DNA. In purely monovalent salt conditions, the 21.5 repeat TALE shows substantially attenuated binding to both target and random DNA in the presence of random competitor DNA (Supplementary Figure S9). In the presence of only monovalent salt, therefore, there appears to be minimal differences in the behavior of TALE binding to target versus random DNA. Interestingly, the behavior of TALE binding in purely monovalent salt conditions is sharply contrasted by TALE binding to target DNA in the presence of 10 mM $MgCl_2$. Here, we note that TALE binding in either monovalent salt or monovalent and divalent salt conditions are compared under conditions of equivalent total solution ionic strength for consistency. Notably, TALE binding to target DNA is essentially unaffected by the large excess of random competitor DNA in the presence of Mg^{2+} (Supplementary Figure S9). Overall, our results suggest that the presence of small amounts of certain divalent cations generally attenuates non-specific binding and stabilizes specific binding, thereby enhancing TALE binding specificity.

TALE NTR binding to DNA is impaired by divalent cations.

In order to further elucidate the effect of divalent cations on target site discrimination for TALEs, we designed a TALE truncation mutant containing only the NTR (known as the NTR-only construct), which was used to probe the role of divalent cations on TALE NTR binding. In prior work, it has been clearly shown that the NTR is critical for nucle-

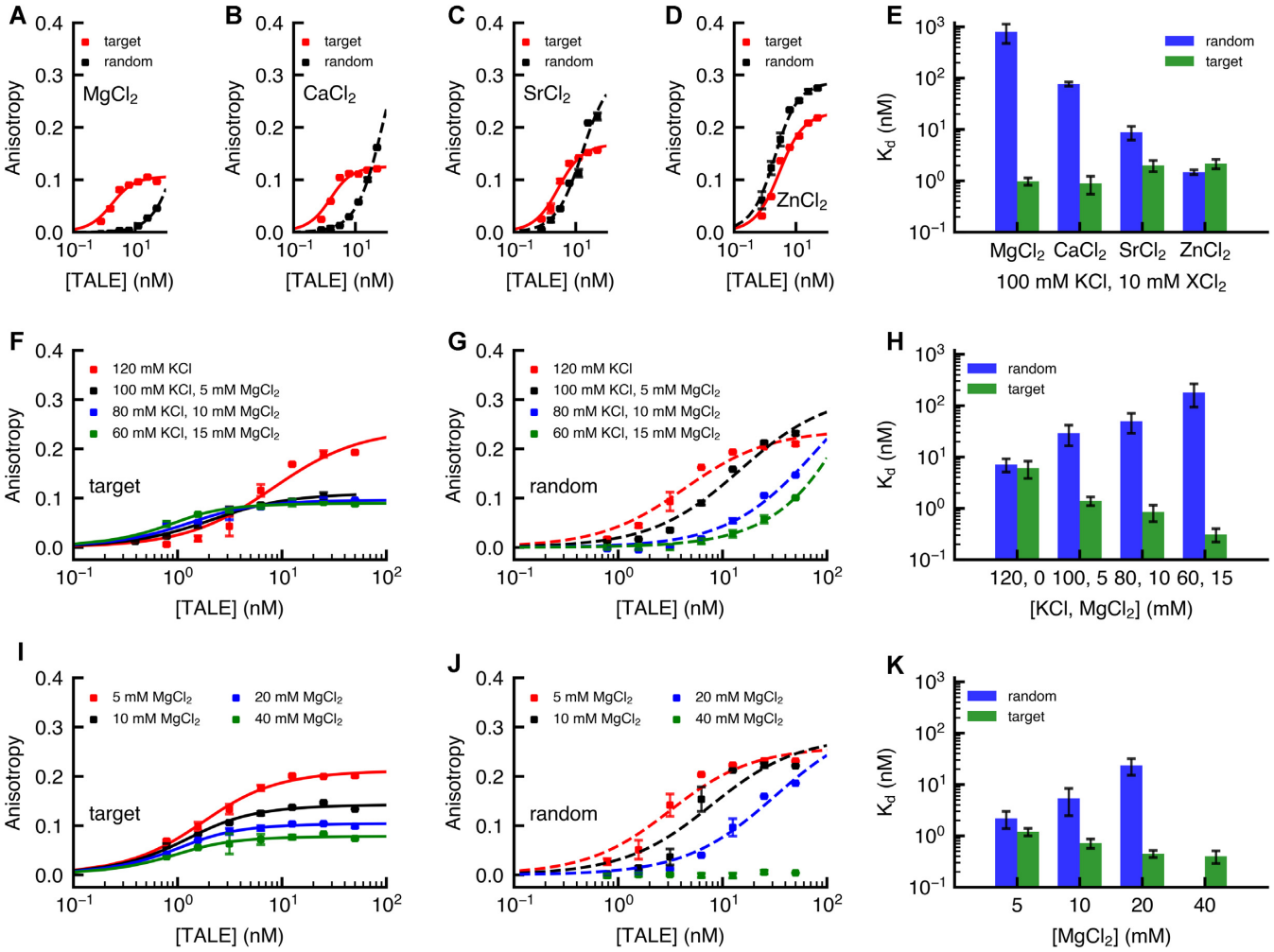

Figure 3. Effects of divalent cations on TALE-DNA binding for the 21.5 repeat TALE. Binding of the 21.5 repeat TALE to target (dots for raw data, solid lines for fitted curves) and random DNA (dots for raw data, dashed lines for fitted curves) measured by fluorescence anisotropy in 20 mM Tris–Tris HCl buffer containing 100 mM KCl and 10 mM (**A**) $MgCl_2$, (**B**) $CaCl_2$, (**C**) $SrCl_2$ or (**D**) $ZnCl_2$. (**E**) Equilibrium dissociation constants $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ for TALE binding in the presence of various divalent cations. Fluorescence anisotropy results for binding to (**F**) target DNA and (**G**) random DNA in 20 mM Tris–Tris HCl buffer with the total ionic strength of added salts held constant at 120 mM. (**H**) Equilibrium dissociation constants $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ for TALE–DNA binding in the presence of KCl and $MgCl_2$. Fluorescence anisotropy results for the binding to (**I**) target DNA and (**J**) random DNA in 20 mM Tris–Tris HCl buffer with increasing concentrations of $MgCl_2$ in the absence of any monovalent salt. (**K**) Equilibrium dissociation constants $K_{d,\text{non-specific}}$ and $K_{d,\text{specific}}$ for TALE–DNA binding in the presence of only $MgCl_2$. DNA concentration was held constant at 1 nM, and TALE concentrations were adjusted as noted. Duplicate experiments were performed to determine the mean and standard deviation of fluorescence anisotropy value.

ation of TALE binding events and NTR binding to DNA occurs in a non-specific fashion, driven primarily by electrostatic interactions (5,6,28). Furthermore, we have previously shown that the TALE NTR is capable of binding to and diffusing one-dimensionally along DNA in reduced ionic strength conditions in the presence of monovalent salt (28).

In the present work, our results show that Mg^{2+} entirely suppresses the ability of the NTR-only TALE mutant to bind to DNA, even in low concentrations of divalent cations (Figure 4). In particular, the binding affinity of the NTR-only TALE to DNA is significantly attenuated even at very low total ionic strength in the presence of 2.5 mM $MgCl_2$ compared to pure monovalent salt conditions (Figures 4D and E). As expected, a TALE construct lacking the NTR (CRD+CTR only) shows no binding affinity for target or random DNA even with addition of divalent cations (Supplementary Figure S10). Overall, our results suggest that addition of monovalent salt and/or an increase in overall ionic strength can weaken the NTR binding interactions, however, divalent cations greatly suppress NTR binding to DNA. Together, these data help to elucidate the effects of divalent cations on TALE binding. When the NTR binding is significantly attenuated by the presence of divalent cations, the overall non-specific binding affinity of TALEs decreases. TALEs with shorter CRDs have a greater proportion of their overall binding energy conferred via the NTR compared to longer TALEs. Thus, the overall binding affinities of shorter TALEs are expected to be more strongly impacted than the binding affinity of longer TALEs by the

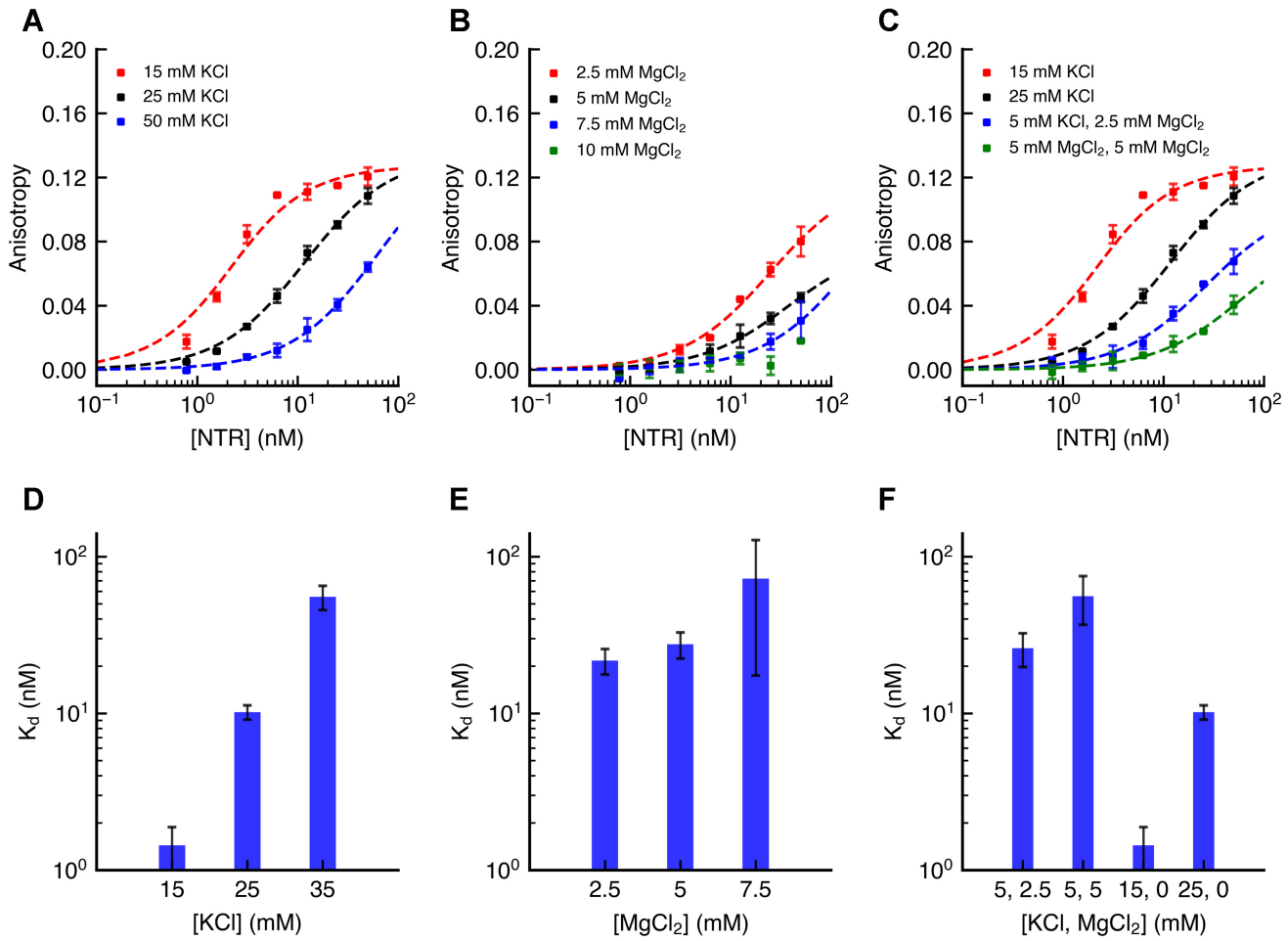

Figure 4. TALE NTR binding to DNA in a variety of solution conditions. Binding of the NTR-only TALE construct (without a CRD or CTR) to random DNA template was measured via fluorescence anisotropy in 20 mM Tris–Tris HCl buffer with increasing concentrations of (**A**) KCl only, (**B**) $MgCl_2$ only or (**C**) a combination of KCl and $MgCl_2$ with total ionic strength of either 15 or 25 mM (dots for raw data, dashed lines for fitted curves). Equilibrium dissociation constants for TALE NTR binding to DNA in the presence of (**D**) KCl, (**E**) $MgCl_2$ and (**F**) KCl and $MgCl_2$. Binding is rapidly diminished in the presence of $MgCl_2$, even when accounting for total ionic strength. DNA concentration was held constant at 1 nM. Duplicate experiments were performed to determine the mean and standard deviation of fluorescence anisotropy value.

presence of divalent cations, which is consistent with the results presented in this study (Figure 3, Supplementary Figures S6 and S7).

Divalent cations introduce significant changes in TALE binding free energy compared to monovalent cations.

On the basis of our experimental results, we hypothesized that divalent cations significantly attenuate TALE non-specific binding energy more than monovalent cations, thereby enhancing TALE target discrimination. To validate this hypothesis, we further sought to use MD simulations to characterize the effects of monovalent and divalent cations on TALE target binding free energy. Specifically, we aimed to estimate the change in free energy of TALE target binding ($\Delta\Delta G_b$) under salt conditions versus pure water environment. A positive $\Delta\Delta G_b$ therefore indicates destabilization of binding by the presence of salts, whereas a negative $\Delta\Delta G_b$ suggests a stabilizing effect. We tested our hypothesis by comparing $\Delta\Delta G_b$ under monovalent and divalent salt conditions at the same ionic strength.

We sought to evaluate $\Delta\Delta G_b$ for individual TALE components of the NTR and the CRD, respectively. Estimation of $\Delta\Delta G_b$ requires determination of the preferential interaction coefficients (Γ_{23}; 1—water, 2—protein or DNA (solute) and 3—ion (cosolute)) of the solutes (e.g. the NTR, the CRD, DNA and their complexes) in the presence of different ionic species (Materials and Methods) (55–58,60,63,64). Γ_{23} was determined from MD simulation data (53–58), which measures the excess number of ions in the local domain of a solute compared to the bulk (53,54,59). Positive values of Γ_{23} suggest that ions preferentially bind to the solute surface, whereas negative values of Γ_{23} suggest that ions are preferentially excluded from the solute surface.

For all simulations, the time-averaged Γ_{23} values converge after 50 ns (Supplementary Figures S11 and S12), and approach constant values for r >6 Å (Figure 5 and Supplementary Figure S13). Previous studies have also found that the extent of the local domain around the solute where Γ_{23} varies is ∼6 Å. (53–58,60,64) We observed that the TALE CRD relaxes to an extended superhelical conformation at

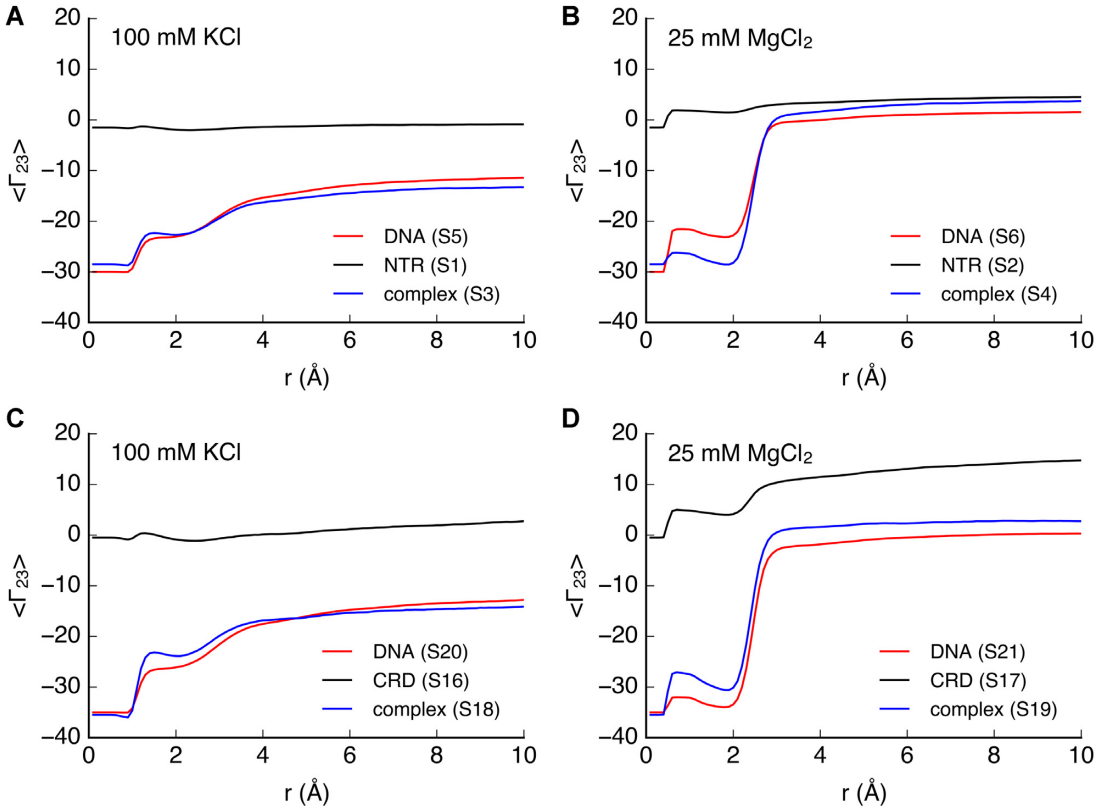

Figure 5. Variation in preferential interaction coefficients as a function of distance from the solute surface. For simulations involving DNA, TALE NTR and NTR-DNA complex in (**A**) 100 mM KCl and (**B**) 25 mM $MgCl_2$, and simulations involving DNA, TALE CRD and CRD–DNA complex in (**C**) 100 mM KCl and (**D**) 25 mM $MgCl_2$, the preferential interaction coefficients approach constant values beyond 6 Å from the molecule surface.

the dissociated state within 50 ns, which is consistent with the previously reported crystal structures of DNA-free and DNA-bound TALEs (65). Final Γ_{23} values were determined at $r = 8$ Å and $\Delta\Delta G_b$ values were then estimated using the calculated Γ_{23} (Table 1). Strikingly, our data show that divalent salts exhibit positive preferential interaction with all the solutes compared to monovalent salts. Moreover, divalent salts preferentially bind to all the solutes, and monovalent salts are preferentially excluded from all the solutes (except the CRD in 100 mM KCl).

Our results suggest that divalent salts introduce significant changes in TALE binding free energy compared to monovalent salts (Table 1). For both the NTR and the CRD, positive $\Delta\Delta G_b$ values suggest that the presence of either monovalent or divalent salts in solution destabilizes their DNA-bound complexes. For the TALE NTR, the presence of $MgCl_2$ results in an over 3-fold change in $\Delta\Delta G_b$ compared to KCl at the same ionic strength. Furthermore, $\Delta\Delta G_b$ is observed to increase upon increasing ionic strength. For example, the $\Delta\Delta G_b$ values in 200 mM KCl and 50 mM $MgCl_2$ are 2.3 and 6.5 kcal/mol, respectively. The large differences in $\Delta\Delta G_b$ suggest that the presence of divalent cations will strongly impact the non-specific NTR binding compared to monovalent cations. The predictions are in agreement with experimental observations that divalent cations greatly attenuate and even suppress the NTR binding at low concentrations. For the TALE CRD, the differences in $\Delta\Delta G_b$ caused by divalent and monovalent salts are even larger (more than 5-fold). Under conditions of 25 mM and 50 mM $MgCl_2$, $\Delta\Delta G_b$ values are found to be 18.6 and 14.5 kcal/mol, respectively, which could be large enough to suppress the TALE CRD binding to DNA. Again, these results are consistent with the experimental observations that the TALE CRD+CTR construct without the NTR cannot bind to DNA. Overall, these results demonstrate that divalent salts have clear effects on TALE binding free energy due to their strong preferential interactions with TALE or DNA compared to monovalent salts. Furthermore, the non-specific NTR binding affinity is significantly decreased by divalent salts compared to monovalent salts, which confirms our hypothesis.

Divalent cations attenuate non-specific binding and stabilize specific binding via preferential binding to DNA.

MD simulations provide detailed molecular descriptions of interactions between ionic species and TALE/DNA complexes. Using this approach, we further sought to identify the interactions of divalent cations with solutes (TALE or DNA) that are responsible for suppression of TALE non-specific binding, as well as possible stabilization of specific binding. DNA is a negatively charged polymer, and divalent cations will preferentially associate with DNA compared to monovalent cations due to stronger electrostatic interaction, as suggested by Γ_{23} (Figure 5).

Table 1. Change in free energy of binding ($\Delta\Delta G_b$) between TALE NTR/CRD and DNA in the presence of monovalent and divalent salts, and the preferential interaction coefficient (Γ_{23}) for the simulated systems. The error bars on $\Delta\Gamma_{23}$ and $\Delta\Delta G_b$ are given, and the error bars on Γ_{23} for the simulated systems are ∼1.

System	Concentration (mM)	$\Gamma_{23,\,complex}$	$\Gamma_{23,\,DNA}$	$\Gamma_{23,\,TALE}$	$\Delta\Gamma_{23}$	$\Delta\Delta G_b$ (kcal/mol)
NTR, KCl	100	− 13.5	− 11.9	− 1.0	− 0.6 ± 0.6	0.7 ± 0.7
NTR, $MgCl_2$	25	3.4	1.3	4.3	− 2.2 ± 1.1	3.6 ± 1.8
NTR, KCl	200	− 17.1	− 12.5	− 2.5	− 2.1 ± 0.7	2.3 ± 0.8
NTR, $MgCl_2$	50	3.6	2.0	5.6	− 4.0 ± 1.4	6.5 ± 2.3
CRD, KCl	100	− 14.6	− 13.5	1.9	− 3.0 ± 1.5	3.3 ± 1.6
CRD, $MgCl_2$	25	2.7	0.1	14.0	− 11.4 ± 1.3	18.6 ± 2.1
CRD, KCl	200	− 20.7	− 15.6	− 3.4	− 1.7 ± 1.1	1.8 ± 1.2
CRD, $MgCl_2$	50	7.0	2.8	13.1	− 8.9 ± 2.1	14.5 ± 3.4

We characterized the interaction strength between DNA and different ions by determining a contact coefficient, which is the ratio of the local ion concentration around each DNA nucleotide (within 8 Å from nucleotide surface) to the bulk ion concentration. We observed a >3-fold DNA-binding preference for Mg^{2+} compared to K^+, which is consistent with prior studies (Supplementary Figure S14) (66,67). A snapshot from MD simulations suggests that divalent cations can bind to the minor groove, the major groove, and the backbone of DNA (Supplementary Figure S14). As a result, when a TALE protein binds to DNA, divalent cations that are associated with DNA are excluded from its surface (Supplementary Figures S15 and S16), which leads to decreased binding affinity. Some divalent cations might still reside around the interface between TALE and DNA, which can influence TALE–DNA interactions. By comparing the distribution of ions near the TALE–DNA interface before and after a TALE protein binds to DNA, we can obtain structural insight into the effects of divalent cations on TALE binding. We therefore identified the TALE residues at the interfaces between the NTR/CRD and DNA and quantified their interactions with ions by determining the contact coefficient during a binding event.

Our results suggest that the suppression of NTR binding by divalent cations is caused by preferential binding of divalent cations to DNA (Figure 6). The residues in the NTR that interact with DNA were identified from the changes in solvent-accessible area (ΔSAA) of the NTR residues when the NTR is in the dissociated state compared to the associated state. Among the top ten residues with the largest ΔSAA values, seven are positively charged arginine or lysine residues (Figure 6A). This further highlights the significant contribution of electrostatic interaction to overall non-specific NTR binding energy. As expected, the NTR residues in contact with DNA do not have favorable interactions with divalent cations (Figure 6B, C). Therefore, divalent cations effectively abolish the electrostatic interactions between the NTR and DNA, which reduces the overall affinity of NTR to DNA backbones in the presence of divalent cations. In this way, divalent cations indirectly suppress NTR binding to DNA as opposed to acting by direct interactions with the NTR. In other words, for the NTR to bind DNA, the exclusion of divalent cations from the DNA interface results in an overall decreased binding affinity. In particular, in the presence of 25 mM $MgCl_2$, ∼4 Mg^{2+} are required to be excluded from the DNA surface to form a NTR–DNA complex (Supplementary Figure S15). Overall, these results show that divalent cations preferentially bind to DNA over monovalent cations, which in turn suppresses TALE NTR binding.

Our results also suggest that the presence of divalent cations attenuates the non-specific contacts between the CRD and DNA while stabilizing some specific contacts (Figure 7). From the co-crystal structure of a TALE–DNA complex, the interactions between the CRD and DNA include non-specific contacts mediated by a lysine and glutamine at positions 16 and 17 in each repeat, respectively, and specific contacts by the second residue (position 13) in RVDs (Figure 7A). When the CRD is dissociated from DNA, residues 16 and 17 in the CRD repeats do not have favorable interactions with monovalent and divalent cations; however, residue 13 in the RVD can have strong interactions with divalent cations compared to monovalent cations (Figure 7B). These strong interactions are mainly contributed by the negatively charged aspartic acid residue in the CRD repeats containing histidine/aspartic acid (HD) in the RVD (Figure 7C). When the CRD is bound to DNA, MD simulations suggest that the non-specific contacts between the CRD and DNA backbone are significantly attenuated by the presence of either monovalent or divalent cations (Supplementary Figure S17). Snapshots show that monovalent or divalent cations preferentially bind to DNA such that non-specific interactions are suppressed (Figure 7D, E). This is consistent with our previous single molecule observation and model for TALE non-specific binding in which the CRD only makes partial contacts with the DNA backbone (28). Strikingly, MD simulations suggest that Mg^{2+} can still bind to the major groove of DNA and interact with both DNA nitrogenous bases and the aspartic acid residues in RVDs containing the sequence HD (Figure 7F). We hypothesize that DNA-bound Mg^{2+} can stabilize the specific interactions between aspartic acid and cytosine, thereby contributing to overall specific interaction energy. The experimental observation that addition of small amounts of divalent cations enhances the specific binding affinity of the 21.5 repeat TALE can be explained by the stabilization of these specific contacts. Indeed, for the 21.5 repeat TALE construct used in our experiments, eight RVD pairs have the sequence HD and three of them are located in the last five CRD repeats. Compared to the 21.5 repeat TALE, the stabilizing effect is not as evident for TALEs with shorter CRDs (11.5 or 15.5 repeats), presumably due to the reduced number of RVDs bearing the sequence HD. As a result, for the 11.5 repeat and 15.5 repeat TALEs, the reduction

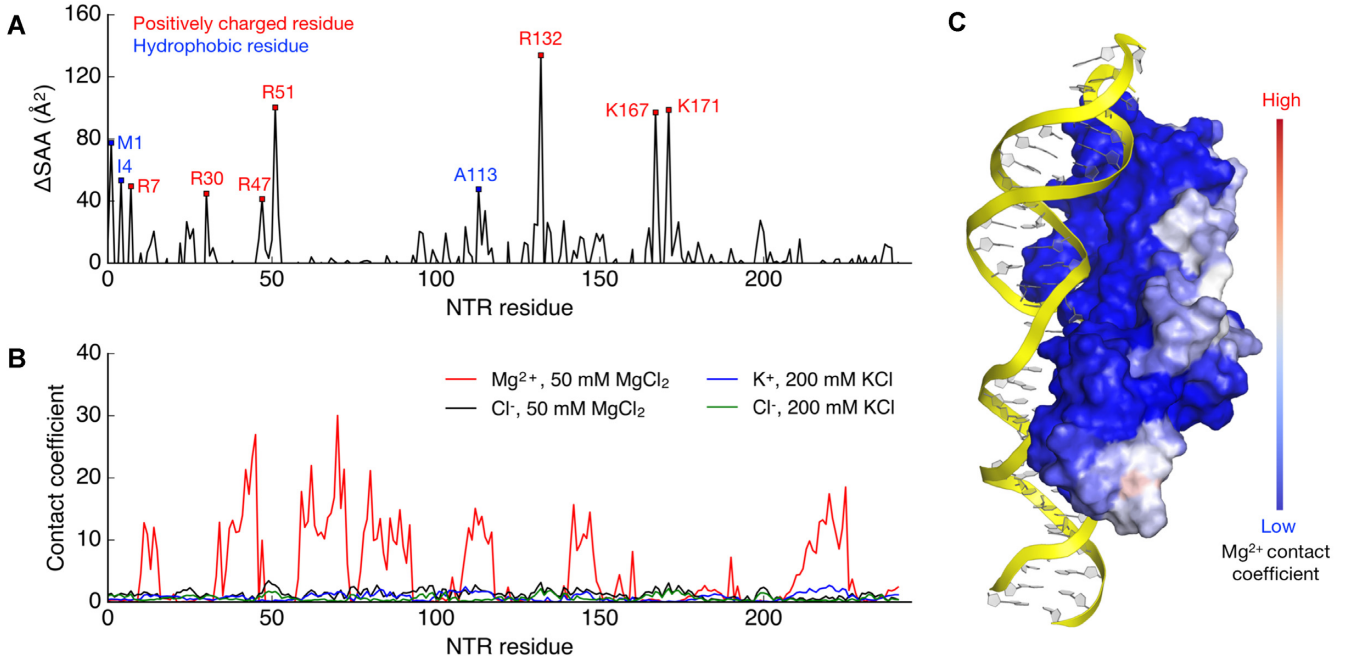

Figure 6. Preferential interactions of the NTR with ions. (**A**) Change in the solvent-accessible area ΔSAA of the residues in the TALE NTR when the NTR is in the dissociated state compared to the associated state. Ten residues with the largest ΔSAA values are labeled, with seven positively charged residues (red) and three hydrophobic residues (blue). (**B**) Contact coefficients for the residues in the NTR with the different ionic species. Mg^{2+} shows stronger interactions with parts of the NTR compared to K^+ and Cl^-. However, (**C**) the residues in the NTR that interact with DNA have little interactions with Mg^{2+}. The NTR surface is colored based on the contact coefficients of the NTR residues with Mg^{2+} at the dissociated state. Binding of the NTR to DNA requires the exclusion of cations from DNA. In the presence of Mg^{2+}, the TALE NTR binding is attenuated due to the strong associations between DNA and Mg^{2+}.

in non-specific interaction can not be compensated by the enhancement in specific interaction. Therefore, no enhancement in overall specific binding affinity is observed for these shorter TALE constructs. Overall, these results suggest that the presence of divalent cations attenuates the non-specific binding between the CRD and DNA backbone and also may contribute to specific binding by stabilizing contacts between HD residues in the RVD and cytosine.

DISCUSSION

Despite a large body of prior work on TALE proteins, a comprehensive view of the binding mechanism of TALEs to specific and non-specific DNA is generally lacking. In this study, we systematically characterized the effects of different ionic species over a range of concentrations on TALE specificity using both *in vitro* fluorescence anisotropy assay and MD simulations. Our results show that the presence of low concentrations of certain divalent cations (Mg^{2+} or Ca^{2+}) results in the ability of TALEs to discriminate between target and non-target sequences. This trend appears to be consistent with prior reports of TALEs binding to DNA; in particular, prior *in vitro* binding measurements carried out in the absence of Mg^{2+} (5,16) tend to show less sequence specificity when compared to those that include Mg^{2+} in binding buffers (12–14).

Our results show that even supra-physiological concentrations of monovalent cations are insufficient to account for the sequence specificity of TALEs. This behavior generally contrasts with many sequence-specific DNA-binding proteins, such as tumor suppressor p53 (33) and integration host factor (34), wherein non-specific protein binding to DNA can be adequately suppressed via sufficient concentrations of monovalent cations, thereby resulting in enhanced binding specificity. Surprisingly, our results suggest that the presence of Mg^{2+} or Ca^{2+} substantially suppresses off-target binding for all full length TALEs (11.5, 15.5, 21.5 repeats) and simultaneously enhances binding to target sites for the 21.5 repeat TALE, thereby enhancing TALE binding specificity. Furthermore, Mg^{2+} suppresses non-specific TALE NTR binding even at very low ionic strength. The divalent cation dependence for sequence-specific binding of TALEs is analogous to prior reports that Mg^{2+} is required to prevent non-specific binding of the B-ZIP proteins (CREB and Fos-Jun) (35) and the prokaryotic transcription factor NikR (36).

Our MD simulations provide energetic and structural insight into the effects of divalent cations on TALE binding. Compared to K^+, Mg^{2+} introduces significant changes in the free energy of binding for both the NTR and the CRD. Non-specific TALE binding energy includes contributions from both the NTR and the CRD. NTR binding is dominantly driven by the electrostatic interactions between the positively charged residues in the NTR and the negatively charged DNA backbone. Such electrostatic interactions are inhibited by competition with cations. As DNA preferentially interacts with divalent cations over monovalent cations, the effect of divalent cations on NTR bind-

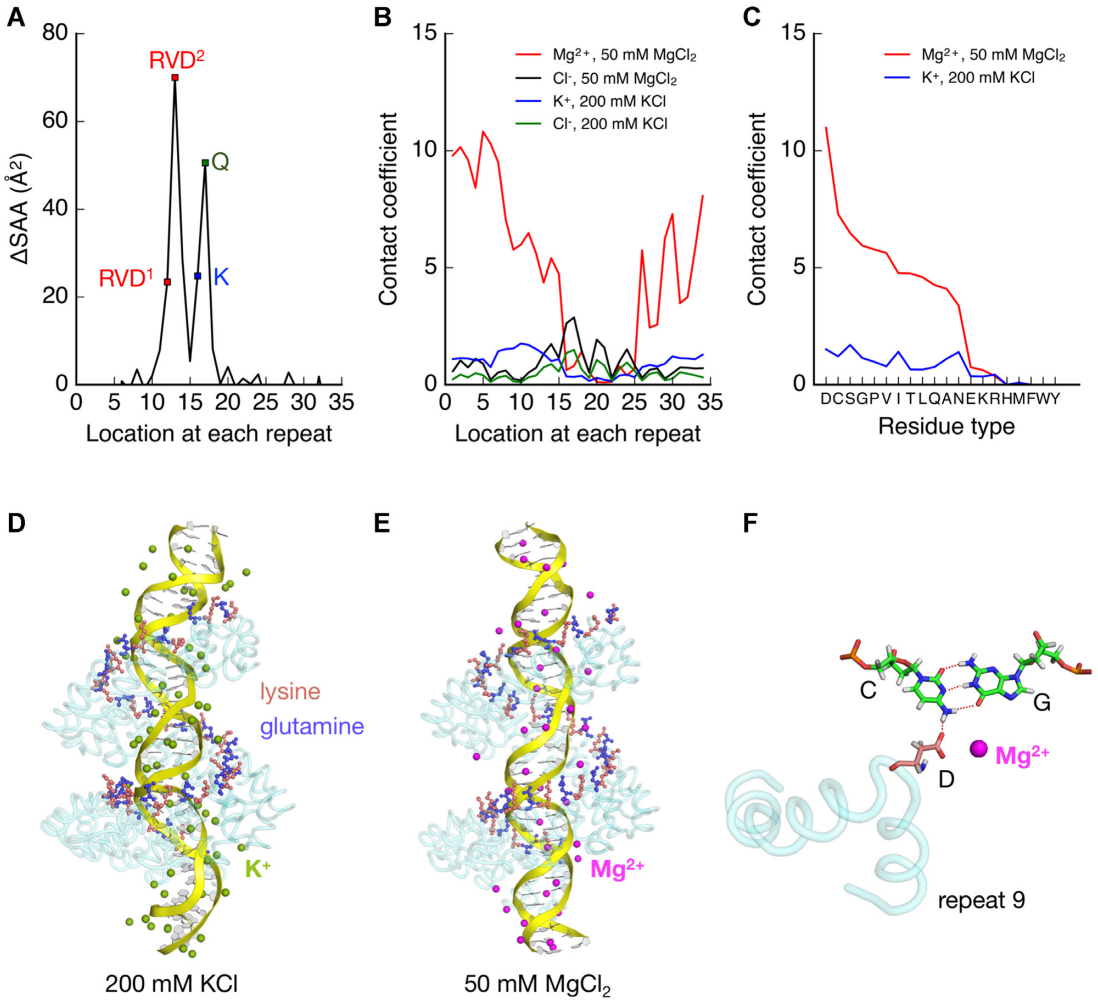

Figure 7. Preferential interactions of the CRD with ions. (**A**) The average change in the solvent-accessible area ΔSAA of the residues at each location within a repeat when the CRD is in the dissociated state compared to the associated state. (**B**) The average contact coefficients for (B) the residues at each position of the repeat and (**C**) different type of residues in the CRD at the dissociated state. Divalent cations show stronger interaction with the CRD residues compared to monovalent cations, which is mostly driven by the negatively charged aspartic acid residue at position 4 of each repeat or position 13 of the repeats with RVD of HD. Snapshots of the simulated CRD–DNA complex structures in the presence of (**D**) 200 mM KCl and (**E**) 50 mM $MgCl_2$. K^+ and Mg^{2+} preferentially bind to DNA that the CRD only partially contacts with DNA backbone. The 1–21 repeats in the CRD are shown in tube (cyan), and the lysine (pink) and glutamine (blue) residues are shown in stick. K^+ (green) and Mg^{2+} (magenta) are shown in spheres. (**F**) Mg^{2+} binds to the major groove of DNA, and interacts with DNA bases and the aspartic acid residue (the second residue in RVD of HD) in repeat 9, which may stabilize the specific interaction.

ing is more pronounced. Similar to the NTR, non-specific contacts between the CRD and the DNA backbone are suppressed by competition with monovalent and divalent cations. Overall, the presence of Mg^{2+} or Ca^{2+} significantly attenuates TALE non-specific binding.

TALE binding to target sites is driven by non-electrostatic interactions, including hydrogen bonding and van der Waals interactions, between the RVDs in the CRD and nitrogenous bases of target DNA sequences. Despite the exclusion of ions during TALE binding, MD simulations suggest that some Mg^{2+} specifically bind to the major groove of DNA and interact with both the aspartic acid in HD and nitrogenous bases. Prior studies have also reported that Mg^{2+} can interact favorably with π-systems of nucleic acid bases through the major groove (68). The diresidue HD is the most common RVD occurring in the 21.5 TALE construct used in our work, and the HD RVD is known to be one of the highest affinity RVDs (27). We conjecture that due to the presence of bound Mg^{2+} ions, the interactions between HD residues in the RVD and the cytosine base will be stabilized, thereby enhancing overall specific binding energy. This finding explains the experimental observation that divalent cations enhance the specific binding of the 21.5 repeat TALE construct to the target DNA sequence. For the 11.5 repeat and 15.5 repeat TALEs, the stabilization of specific binding is less pronounced likely due to the reduced number of HD RVD sequences. Taken together, these observations explain the enhanced overall non-electrostatic interactions between the 21.5 repeat TALE and target DNA sequences.

Compared to Mg^{2+} or Ca^{2+}, our results show that Mn^{2+} only moderately attenuates the non-specific TALE bind-

ing, while Sr^{2+} and Zn^{2+} have little effect. This trend appears to be correlated with the binding preferences of divalents cations to DNA bases relative to phosphates: $Zn^{2+}>Mn^{2+}>Ca^{2+}>Mg^{2+}$ (69). The pronounced attenuation of non-specific TALE binding in the presence of Mg^{2+} or Ca^{2+} could be attributed to their relative preferential binding to the DNA backbone. However, the overall effects of divalent cations on TALE specificity do not appear to be correlated with the strength of divalent cation interaction with DNA bases ($Mn^{2+}>Ca^{2+}>Mg^{2+}$, Sr^{2+}) (69) or the divalent cation Hofmeister series at the protein carboxylate sites ($Zn^{2+}>Ca^{2+}>Sr^{2+}>Mg^{2+}$) (70,71).

Similar to other DNA-binding proteins, the binding of TALEs to target DNA is governed by a balance of electrostatic and non-electrostatic interactions, which are in turn influenced by the presence of divalent cations and total ionic strength. In the presence of small amounts of certain divalent cations (Mg^{2+} or Ca^{2+}), the electrostatic interactions between the NTR or CRD and DNA backbone will be attenuated, and the non-electrostatic interactions between the RVDs and specific DNA sequences will be enhanced. In this way, the non-specific binding affinity generally decreases upon addition of divalent cations, as it is primarily driven by electrostatic interactions. On the other hand, changes in the specific binding affinity will depend on the length of the CRD in TALEs and the presence of HD RVD. For longer TALEs, our results suggest that attenuation of electrostatic binding interactions can be outweighed by enhanced non-electrostatic binding interactions in the presence of divalent cations, and therefore the specific binding affinity can increase or remain unchanged under these conditions. For shorter TALEs, the attenuated electrostatic binding interactions cannot be compensated by the enhanced non-electrostatic binding interactions, and therefore the specific binding affinity is impacted by divalent cations. These results are consistent with our observations that the effects of divalent cations are largest for the 21.5 repeat TALE, and generally decrease as the TALE CRD is shortened. We find that the 21.5 repeat TALE exhibits extremely high specificity under physiologically relevant salt concentrations, which suggests that physiological salt conditions may provide a near optimal balance between specific and non-specific TALE binding to DNA.

Finally, we propose a general model for the divalent cation effect on TALE specificity. Divalent cations strongly affect NTR binding, and the NTR contributes a significant proportion of the overall binding affinity for shorter TALEs than longer for TALEs. Meanwhile, divalent cations can stabilize some specific interactions between the CRD and DNA bases. Longer TALEs are able to overcome the reduction in non-specific binding affinity by their larger CRD regions. Thus, it holds that the effects of divalent cations can be thought of as a balancing act; they increase specificity by decreasing affinity for non-specific DNA. Our findings suggest that divalent cations in the cellular milieu are a critical component to enhance the binding specificity of TALEs and possibly other sequence-specific DNA-binding proteins. Overall, this work provides new insights into the fundamental nature of TALE–DNA binding process, which could potentially be useful for gene editing applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

D.S. and C.Z. acknowledge the support from the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois.

FUNDING

Carl R. Woese Institute for Genomic Biology (to H.Z.); FMC Fellowship (to L.W.C.); 3M Corporate Fellowship (to C.Z.); David and Lucile Packard Foundation (to C.M.S.). Funding for open access charge: University of Illinois.
Conflict of interest statement. None declared.

REFERENCES

1. Mak,A.N.-S., Bradley,P., Bogdanove,A.J. and Stoddard,B.L. (2013) TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.*, **23**, 93–99.
2. Bogdanove,A.J., 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. and Voytas,D.F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science*, **333**, 1843–1846.
4. Joung,J.K. and Sander,J.D. (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.*, **14**, 49.
5. 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.
6. Schreiber,T. *et al.* (2015) Refined requirements for protein regions important for activity of the TALE AvrBs3. *PLoS One*, **10**, e0120214.
7. Boch,J. *et al.* (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
8. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501–1501.
9. Yang,Y. (1995) Xanthomonas Avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Mol. Plant-Microbe Interact.*, **8**, 627.
10. Yang,B., Zhu,W., Johnson,L.B. and White,F.F. (2000) The virulence factor AvrXa7 of Xanthomonas oryzae pv. oryzae is a type III secretion pathway-dependent nuclear-localized double-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 9807–9812.
11. Sun,N. and Zhao,H. (2013) Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol. Bioeng.*, **110**, 1811–1821.
12. Yin,P. *et al.* (2012) Specific DNA-RNA hybrid recognition by TAL effectors. *Cell Rep.*, **2**, 707–713.
13. Doyle,E.L. *et al.* (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. Römer,P., Recht,S. and Lahaye,T. (2009) A single plant resistance gene promoter engineered to recognize multiple TAL effectors from disparate pathogens. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 20526–20531.
15. Rinaldi,F.C., Doyle,L.A., Stoddard,B.L. and Bogdanove,A.J. (2017) The effect of increasing numbers of repeats on TAL effector DNA binding specificity. *Nucleic Acids Res.*, **45**, 6960–6970.
16. Juillerat,A. *et al.* (2014) Comprehensive analysis of the specificity of transcription activator-like effector nucleases. *Nucleic Acids Res.*, **42**, 5390–5402.
17. Fu,Y. *et al.* (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822.
18. Pattanayak,V. *et al.* (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839.

19. Wang,X. *et al.* (2015) Unbiased detection of off-target cleavage by CRISPR–Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175.
20. Singh,D., Sternberg,S.H., Fei,J., Doudna,J.A. and Ha,T. (2016) Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.*, **7**, 12778.
21. Pattanayak,V., Ramirez,C.L., Joung,J.K. and Liu,D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods*, **8**, 765.
22. Gupta,A., Meng,X., Zhu,L.J., Lawson,N.D. and Wolfe,S.A. (2010) Zinc finger protein-dependent and-independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.*, **39**, 381–392.
23. Mak,A. N.-S., Bradley,P., Cernadas,R.A., Bogdanove,A.J. and Stoddard,B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
24. Streubel,J., Blücher,C., Landgraf,A. and Boch,J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593.
25. 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.
26. Guilinger,J.P. *et al.* (2014) Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods*, **11**, 429.
27. Meckler,J.F. *et al.* (2013) Quantitative analysis of TALE–DNA interactions suggests polarity effects. *Nucleic Acids Res.*, **41**, 4118–4128.
28. 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.
29. Cuculis,L. and Schroeder,C.M. (2017) A single-molecule view of genome editing proteins: biophysical mechanisms for TALEs and CRISPR/Cas9. *Annu. Rev. Chem. Biomol. Eng.*, **8**, 577–597.
30. 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.
31. Geiger-Schuller,K., Mitra,J., Ha,T. and Barrick,D. (2019) Functional instability allows access to DNA in longer transcription activator-like effector (TALE) arrays. *eLife*, **8**, e38298.
32. Wicky,B.I., Stenta,M. and Dal Peraro,M. (2013) TAL effectors specificity stems from negative discrimination. *PLoS One*, **8**, e80261.
33. Weinberg,R.L., Veprintsev,D.B. and Fersht,A.R. (2004) Cooperative binding of tetrameric p53 to DNA. *J. Mol. Biol.*, **341**, 1145–1159.
34. Holbrook,J.A., Tsodikov,O.V., Saecker,R.M. and Record,M.T. Jr (2001) Specific and non-specific interactions of integration host factor with DNA: thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J. Mol. Biol.*, **310**, 379–401.
35. Moll,J.R., Acharya,A., Gal,J., Mir,A.A. and Vinson,C. (2002) Magnesium is required for specific DNA binding of the CREB B-ZIP domain. *Nucleic Acids Res.*, **30**, 1240–1246.
36. Dosanjh,N.S., Hammerbacher,N.A. and Michel,S.L. (2007) Characterization of the Helicobacter pylori NikR- PureA DNA Interaction: Metal Ion Requirements and Sequence Specificity. *Biochemistry*, **46**, 2520–2529.
37. Viadiu,H. and Aggarwal,A.K. (2000) Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell*, **5**, 889–895.
38. Thielking,V. *et al.* (1992) Magnesium (2+) confers dna binding specificity to the EcoRV restriction endonuclease. *Biochemistry*, **31**, 3727–3732.
39. Terry,B., Jack,W., Rubin,R. and Modrich,P. (1983) Thermodynamic parameters governing interaction of EcoRI endonuclease with specific and nonspecific DNA sequences. *J. Biol. Chem.*, **258**, 9820–9825.
40. Hopkins,C.M., White,F.F., Choi,S.H., Guo,A. and Leach,J.E. (1992) Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae. *Mol. Plant-Microbe Interact.*, **5**, 451–459.
41. Sun,N., Liang,J., Abil,Z. and Zhao,H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. BioSyst.*, **8**, 1255–1263.
42. Eckert,K.A. and Kunkel,T.A. (1990) High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res.*, **18**, 3739–3744.
43. Biasini,M. *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.
44. Case,D.A. *et al.* (2014) *Amber 14*. University of California, San Francisco.
45. Eastman,P. *et al.* (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.*, **13**, e1005659.
46. Ivani,I. *et al.* (2016) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, **13**, 55.
47. Joung,I.S. and Cheatham,T.E. III (2009) Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*, **113**, 13279–13290.
48. Panteva,M.T., Giambaşu,G.M. and York,D.M. (2015) Force field for Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cd^{2+} ions that have balanced interactions with nucleic acids. *J. Phys. Chem. B*, **119**, 15460–15470.
49. Åqvist,J., Wennerström,P., Nervall,M., Bjelic,S. and Brandsdal,B.O. (2004) Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.*, **384**, 288–294.
50. Loncharich,R.J., Brooks,B.R. and Pastor,R.W. (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, **32**, 523–535.
51. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh ewald: an n log (n) method for ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
52. Scatchard,G. (1946) Physical chemistry of protein solutions. I. Derivation of the equations for the osmotic pressure. *J. Am. Chem. Soc.*, **68**, 2315–2319.
53. Baynes,B.M. and Trout,B.L. (2003) Proteins in mixed solvents: a molecular-level perspective. *J. Phys. Chem. B*, **107**, 14058–14067.
54. Shukla,D., Shinde,C. and Trout,B.L. (2009) Molecular computations of preferential interaction coefficients of proteins. *J. Phys. Chem. B*, **113**, 12546–12554.
55. Schneider,C.P., Shukla,D. and Trout,B.L. (2011) Arginine and the hofmeister series: the role of ion–ion interactions in protein aggregation suppression. *J. Phys. Chem. B*, **115**, 7447–7458.
56. Schneider,C.P., Shukla,D. and Trout,B.L. (2011) Effects of solute-solute interactions on protein stability studied using various counterions and dendrimers. *PLoS One*, **6**, e27665.
57. Shukla,D., Schneider,C.P. and Trout,B.L. (2011) Effects of PAMAM dendrimer salt solutions on protein stability. *J. Phys. Chem. Lett.*, **2**, 1782–1788.
58. Shukla,D., Schneider,C.P. and Trout,B.L. (2011) Complex interactions between molecular ions in solution and their effect on protein stability. *J. Am. Chem. Soc.*, **133**, 18713–18718.
59. Record,M.T. Jr and Anderson,C.F. (1995) Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a two-domain model. *Biophys. J.*, **68**, 786–794.
60. Shukla,D., Zamolo,L., Cavallotti,C. and Trout,B.L. (2011) Understanding the role of arginine as an eluent in affinity chromatography via molecular computations. *J. Phys. Chem. B*, **115**, 2645–2654.
61. Robinson,R.A., Stokes,R.H. and Wilson,J.M. (1940) A thermodynamic study of bivalent metal halides in aqueous solution. *Trans. Faraday Soc.*, **36**, 733–748.
62. Robinson,R.A. and Stokes,R.H. (1949) Tables of osmotic and activity coefficients of electrolytes in aqueous solution at 25°C. *Trans. Faraday Soc.*, **45**, 612–624.
63. Shukla,D. and Trout,B.L. (2010) Interaction of arginine with proteins and the mechanism by which it inhibits aggregation. *J. Phys. Chem. B*, **114**, 13426–13438.
64. Anand,G., Jamadagni,S.N., Garde,S. and Belfort,G. (2010) Self-assembly of TMAO at hydrophobic interfaces and its effect on protein adsorption: Insights from experiments and simulations. *Langmuir*, **26**, 9695–9702.
65. Deng,D. *et al.* (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720–723.
66. Anderson,C.F. and Record,M.T. Jr (1995) Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.*, **46**, 657–700.

67. Bai,Y. *et al.* (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.*, **129**, 14981–14988.
68. McFail-Isom,L., Shui,X. and Williams,L.D. (1998) Divalent cations stabilize unstacked conformations of DNA and RNA by interacting with base π systems. *Biochemistry*, **37**, 17105–17111.
69. Duguid,J., Bloomfield,V.A., Benevides,J. and Thomas,G.J. Jr (1993) Raman spectroscopy of DNA-metal complexes. I. Interactions and conformational effects of the divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. *Biophys. J.*, **65**, 1916–1928.
70. Kherb,J., Flores,S.C. and Cremer,P.S. (2012) Role of carboxylate side chains in the cation Hofmeister series. *J. Phys. Chem. B*, **116**, 7389–7397.
71. Okur,H.I. *et al.* (2017) Beyond the Hofmeister series: ion-specific effects on proteins and their biological functions. *J. Phys. Chem. B*, **121**, 1997–2014.