8860–8873 *Nucleic Acids Research, 2019, Vol. 47, No. 16* *Published online 16 July 2019*
doi: 10.1093/nar/gkz604

Specificity of protein–DNA interactions in hypersaline environment: structural studies on complexes of *Halobacterium salinarum* oxidative stress-dependent protein *hs*RosR

Nitzan Kutnowski[1], Fania Shmulevich[1], Geula Davidov[1,2], Anat Shahar[3], Dudy Bar-Zvi[1], Jerry Eichler[1], Raz Zarivach[1,2] and Boaz Shaanan[1,*]

[1]Department of Life Sciences, Ben-Gurion University of the Negev, Beer Sheva 8410510, Israel, [2]National Institute of Biotechnology in the Negev, Ben-Gurion University, Beer Sheva 8410510, Israel and [3]Macromolecular Crystallography Research Center, National Institute of Biotechnology in the Negev, Ben-Gurion University, Beer Sheva 8410510, Israel

Received March 06, 2019; Revised June 13, 2019; Editorial Decision July 01, 2019; Accepted July 02, 2019

ABSTRACT

Interactions between proteins and DNA are crucial for all biological systems. Many studies have shown the dependence of protein–DNA interactions on the surrounding salt concentration. How these interactions are maintained in the hypersaline environments that halophiles inhabit remains puzzling. Towards solving this enigma, we identified the DNA motif recognized by the *Halobactrium salinarum* ROS-dependent transcription factor (*hs*RosR), determined the structure of several *hs*RosR–DNA complexes and investigated the DNA-binding process under extreme high-salt conditions. The picture that emerges from this work contributes to our understanding of the principles underlying the interplay between electrostatic interactions and salt-mediated protein–DNA interactions in an ionic environment characterized by molar salt concentrations.

INTRODUCTION

The precise recognition and interaction of proteins with their cognate DNA sequences is essential for all living cells. Specific binding of DNA by protein is achieved through direct interactions with a DNA-binding motif and spatial complementarity to a unique DNA structure (1). Protein binding to DNA involves a strong electrostatic contribution, with the positively charged surface of the DNA-binding site of the protein promoting binding to the negatively charged DNA (2–10). However, as electrostatic interactions are weakened in the face of increasing salt concentrations (11), it is not clear how specific recognition of DNA targets by proteins takes place in the hypersaline cytoplasm of halophilic archaea, such as *Halobacterium* sp. NRC-1 (*Hbt. salinarum*).

Like other haloarchaea, *Hbt. salinarum* requires high-salt conditions for their survival. To cope with such harsh surroundings and avoid osmotic shock, the cytoplasm of haloarchaea is highly enriched in salt (12). Consequently, many cellular components in haloarchaea have adapted so as to properly function in such environments. For instance, halophilic proteins are enriched in acidic residues and fold such that the majority of these negatively charged residues are exposed on the protein surface, as seen first in the structures of haloarchaeal malate dehydrogenase and ferredoxin (13,14). This strategy enables halophilic proteins to remain soluble and active in molar salt concentration surroundings (15,16). Nonetheless, questions regarding the maintenance of the essential interactions between macromolecules in a hypersaline cellular environment, particularly protein–DNA interactions, remain unanswered.

Previous work has shown that high concentrations of hydrogen peroxide lead to increased expression of the *Hbt. salinarum* ROS-dependent regulator *hs*RosR (VNG0258H) (17). *hs*RosR plays a key role in the response to oxidative stress by directly activating or repressing genes encoding transcription factors (TFs) and other proteins that are required for adaptation to extremely oxidative conditions (18). While genomic target sites have been proposed for *hs*RosR under physiological conditions and during oxidative stress (17,18), the precise genomic binding sequences have yet to be defined.

Recently, we determined the structure of the free *hs*RosR protein (19), the only structure to be determined to date of a halophilic MarR/PadR DNA-binding protein. *hs*RosR is a winged-helix-turn-helix (wHTH) protein with unique

*To whom correspondence should be addressed. Tel: +97 25096 65880; Email: bshaanan@bgu.ac.il

features reflecting adaptation to the hypersaline environment. By using anomalous diffraction from crystals, we unambiguously located ions (bromide or chloride) in the first shell around the protein and in the DNA-binding site (19). To define the DNA motif recognized by *hs*RosR and gain insight into how its specific DNA targets are recognized, we determined the crystal structures of *hs*RosR in complex with different DNA target sequences, identified using SELEX (20,21) and genomic SELEX (22) (gSELEX). In addition, the effect of salt on such binding was examined. The structural data presented here, supported by biochemical and biophysical studies, shed light on the process of specific DNA-binding in an extremely high-salt environment.

MATERIALS AND METHODS

*hs*RosR expression and purification

*hs*RosR was expressed and purified as previously described (19)

SELEX

KCl is the dominant salt in the cytoplasm of *Hbt. salinarum* (23), hence all the selection and most binding experiments were performed in KCl throughout the work. To identify DNA sequences recognized by *hs*RosR by SELEX, a library of sequences was generated. The initial SELEX library was a pool of synthesized oligonucleotides (IDT, Israel) comprising 35 bp-long randomized DNA fragments flanked by 20 bp-long primers for PCR amplification. This initial library was prepared using seven PCR cycles. The random DNA fragments generated were separated on a polyacrylamide gel and extracted. As the classical electrophoretic mobility shift assay (EMSA) (24) routinely used in SELEX is not compatible with the high-salt concentrations required to keep the protein well folded (19) and hence able to bind DNA, the protocol described by Bouvet (25) was used with a few modifications. Following this protocol, purified *hs*RosR-His_6 (0.4 mg/ml) was bound to 50 μl Ni^{2+}-NTA beads, equilibrated and washed three times in 200 μl binding buffer (50 mM HEPES, pH 7.0, 1.75, 3 or 4 M KCl, 0.02% azide), prior to the addition of DNA sequences (1 μg) from the initial library. The *hs*RosR–DNA binding reaction was carried out at room temperature for 5 min by rocking the beads in an Eppendorf tube. After washing three times with 200 μl binding buffer, the protein and bound DNA were eluted and separated using 200 μl elution buffer (50 mM HEPES, pH 7.0, 300 mM imidazole, 0.02% azide). Aliquots (2 μl) from the eluted fractions were subjected to 14 cycles of PCR amplification and the products were separated on a polyacrylamide gel. After extraction from the gel, the DNA fragments served as input for another round of selection. Products of SELEX rounds 1, 4 and 7 in high-salt conditions (4 M KCl) and round 10 in all experiments were sequenced at the Technion Genome Center (Haifa, Israel) by deep sequencing on a MySeq or HySeq machine (Illumina). Bioinformatics and statistical analyses, as well as visualization of the results, were carried out using the Galaxy package and server (26,27).

gSELEX

The starting point for gSELEX experiments was a pool of oligonucleotides comprising ∼150 bp-long genomic DNA fragments flanked by known primers for PCR amplification, prepared as described (22). The quality of the initial library was tested using random primers from across the genome (Supplementary Table S1). gSELEX cycles and analysis were carried out as for SELEX described above. The enriched fragments were aligned to the *Hbt. salinarum* genome using Bowtie (28). MACS (29) was used to identify putative DNA-binding sites. Visualization of the results was carried out using the Galaxy package and server (27).

Fluorescence anisotropy (FA)

Candidate target double strand (ds)-DNA was labelled with 5 (6) carboxyfluorescein (5 (6)-FAM) at the 5′-end (IDT, Israel). FA (30) of the labelled DNA was performed using a Synergy 2 apparatus (BioTec, Winooski, VT). Parallel and perpendicular emissions were recorded at 528 nm at 15 sec intervals after excitation at 485 nm in 96-well plates for 20 min. Before measuring fluorescence intensity, 300 μl of samples containing 4 nM labelled DNA and increasing concentrations of *hs*RosR were incubated in binding buffer (50 mM HEPES, pH 7.0, 2, 2.5, 3, 3.5 or 4 M KCl or other salts, 0.02% azide) for 5 min at room temperature. The final K_d for binding was determined using SigmaPlot software by fitting the Michaelis–Menten equation to the data.

Isothermal titration calorimetry (ITC)

ITC (31) measurements of interactions of *hs*RosR with sequence S1 (25 bp) were performed at a constant temperature (25°C) by titrating the DNA sample into a solution containing *hs*RosR in binding buffer (50 mM HEPES, pH 7.0, 2 or 3 M KCl, 0.02% azide) in the sample cell of a low-volume Nano ITC unit (TA Instruments, New Castle, DE). As a control, a titration of sequence S1 (25 bp) into the same solution was performed. In an ITC competition assay, a solution containing bromide ions (20 mM HEPES, pH 7, 3 M KCl, 0.1 M KBr and 0.02% azide) was injected into a solution containing the *hs*RosR–S1 complex in binding buffer (20 mM HEPES, pH 7, 3 M KCl and 0.02% azide). As a control, titration of binding buffer into the same solution was performed. Heat changes were recorded and expressed as electrical power (J s^{-1}). The complete thermodynamic profile of the reaction, including the equilibrium constant (K_d) and binding stoichiometry (n), was also obtained (32). For controls, each DNA sequence was injected into buffer without protein. All measurements were compared to those obtained using a reference cell containing DDW.

Fluorescence spectrometry

Changes in intrinsic tryptophan fluorescence were recorded at ambient temperature using a Fluorolog-3 apparatus (Horiba Scientific, Edison, NJ) equipped with quartz cell with a 1-cm optical path length. A 1 ml sample of *hs*RosR at a 2 μM concentration in buffer containing 50 mM HEPES, pH 7.0, 2 or 3 M KCl and 0.02% azide was measured at λ_{ex} 280 nm, while the emission spectrum for each salt concentration was recorded at 300–450 nm.

Crystallization of protein–DNA complexes

Prior to crystallizing *hs*RosR–DNA complexes, complementary DNA sequences were annealed in 10 mM STE buffer (10 mM Tris–HCl, pH 8, 100 mM NaCl and 1 mM EDTA) for 5 min at 95°C and then gradually cooled overnight to 20°C. Purified protein was mixed with DNA sequences at different ratios (Supplementary Table S2) for 5 min before crystallization.

The complexes were crystallized by the sitting-drop vapour diffusion method. One microlitre complex solution (20 mM HEPES, pH 7, 2 M KCl and 0.02% azide) was mixed with 1 μl well solution (Supplementary Table S2) at 20°C. As in the case of the free protein (19), ammonium sulfate served as precipitant. Addition of minor quantities of $MnCl_2$, which does not affect the DNA binding (see Supplementary Figure S7), proved to be essential for producing a crystal form that accommodated the complexes (Supplementary Table S2). For data collection, crystals were cryo-protected using paratone oil (Hampton Research, Aliso Viejo, CA).

X-ray data recording, structure determination and refinement

Data were recorded at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and the Swiss Light Source (SLS, PSI, Switzerland) and processed using XDS (33,34) and Pointless/Aimless (35,36) from the CCP4 package (37). Datasets that showed anisotropic diffraction were further processed with the STARANISO server (http://Staraniso.globalphasing.org/cgi-bin/staraniso.cgi). The structures of complexes of *hs*RosR with the various sequences were solved by molecular replacement (MR) using CCP4 Phaser (38) by reading two structural models (assemblies) into the program in the same run. One model was that of the unbound protein, VNG0258H/RosR (19) (KCl form; PDB code: 6FDH) and the second was a DNA model derived from *Mycobacterium tuberculosis* MosR (39) (PDB code: 4FX4) but with the S1, S4, S5 or SG sequences (Supplementary Table S2). The correctness of the molecular replacement solution derived for each of the protein–DNA complexes was verified by running Phaser again with the model of the complete protein–DNA complex. The structures were refined using CCP4 Refmac (40) and Phenix (41), inspected and modified using Coot (42). Molecular models were rendered using UCSF-Chimera (43). Calculations of electrostatic properties were carried out using Delphi (44) and APBS (45). Analysis of buried surfaces was done using the Pisa server (http://pdbe.org/pisa/) (46).

RESULTS

SELEX detects the palindromic motif recognized by *hs*RosR

The SELEX experiment conducted in 4 M KCl revealed over-represented sequences (>0.1%) that all contained the same motif, namely **TGT**-N_{10}-**ACA**, although they differed substantially in terms of the bases within and outside this motif (Supplementary Table S3). The palindromic nature of the motif, found in all of the over-represented sequences, is typical of sequences recognized by MarR/PadR proteins (47). Sequence S1 and other most enriched sequences (S1–S5; Table 1) were further analysed. Sequence S1 showed no enrichment in low salt (1.75 M KCl) (Figure 1A), probably reflecting the poor folding state of the protein in such conditions (19,48). However, the same sequence (Table 1 and Supplementary Table S3) was enriched as a function of increasing salt concentration (Figure 1A). Enrichment of this sequence was enhanced in 3 M KCl (0.434%) and more so in 4 M KCl (0.934%) (Figure 1A). Sequence S1 was also exponentially enriched in high salt (4 M KCl) during successive rounds of enrichment (Figure 1B).

Table 1. SELEX results

	Sequence	K_d (nM)
S1	GCGAGG**TGT**AAATTGTCTG**ACA**TGTTCTTATTGGG	49.32±6.18
S2	GCGCTCA**TGT**CAACCTTCTT**ACA**CTTCCTCTCCCG	60.12±18.82
S3	GCACA**TGT**AAGTCTATTG**ACA**TGTTCTTTCCCTCG	68.03±11.34
S4	GCCCGAAG**TGT**AAACGCGTTG**ACA**TGTGGTGCTGG	91.59±13.24
S5	GCGAAG**TGT**CATCCCTCTT**ACA**TGACTTGTCTTGG	99.46±10.55

The five sequences most enriched in SELEX round 10 (4 M KCl) are listed. The DNA binding motif is marked in red. The sequences are ranked according to their percent enrichment over the 10 rounds of SELEX. The binding constants (K_d), as determined by FA, are also listed.

To assess DNA-binding affinity, sequence S1 (35 bp, Table 1) was labelled with a fluorophore and examined in an FA-based DNA-binding assay (30) at different salt concentrations (2–4 M). As MarR/PadR proteins usually bind 20–25 bp-long sequences (47), sequence S1 was trimmed by 5 bases at each end in an attempt to better define the recognized motif (Figure 1C and D). The binding affinities (K_d) obtained for both the complete and truncated S1 sequences were approximately the same (Figure 1C and D). These K_d values are in the range reported for DNA-binding by other MarR/PadR proteins (47,49). Therefore, we concluded that the sequence S1-bound DNA motif was not disrupted by trimming to 25 bp sequences; however, binding was completely abolished when the sequence was shortened to 20 bp. Finally, we measured the binding affinities of the next four sequences enriched in round 10 (4 M KCl) (Table 1). The ranking of the DNA-binding results correlated with the percent enrichment of the sequences (Table 1 and Supplementary Table S3).

The SELEX results combined with binding assays showed that *hs*RosR recognizes and binds different sequences containing exactly the same palindromic motif **TGT**-N_{10}-**ACA**, albeit with differing binding affinities. We thus concluded that this is the essential sequence needed for DNA recognition and defined it as the *hs*RosR DNA binding motif. The differences in the binding affinity observed between the DNA sequences probably stem from the different base composition within and outside this motif, among other factors.

The effect of salt on DNA binding to *hs*RosR

To gain insight into the binding process under high-salt conditions, the interaction of *hs*RosR with the most enriched DNA sequence (S1) was initially investigated biochemically. To explore the effect of salt on DNA binding by *hs*RosR, we

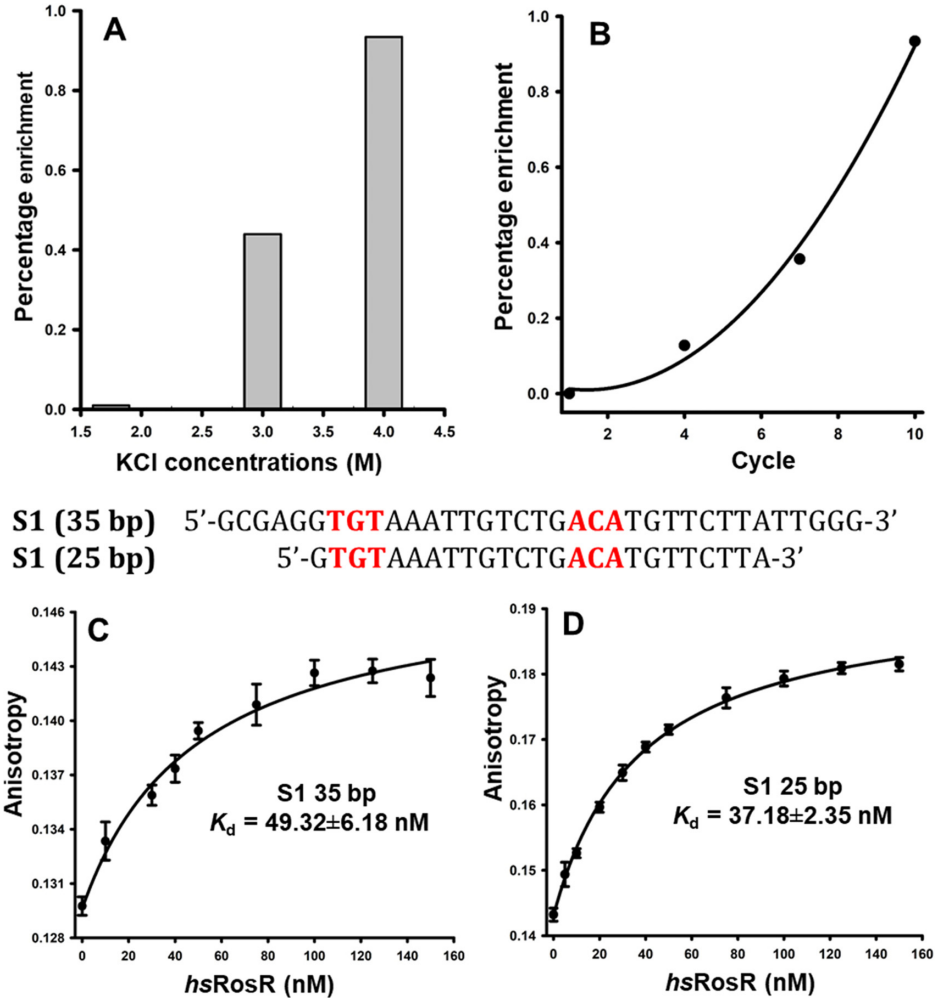

Figure 1. SELEX and DNA-binding results. (**A**) The percentage enrichment of sequence S1 (in the last SELEX round, round 10) at different KCl concentrations. (**B**) The percentage enrichment of sequence S1 upon performing SELEX in 4 M KCl. (**C**) Binding curve of the 35 bp-long S1 sequence. (**D**) Binding curve of the 25 bp-long S1 sequence. The putative motif recognized by *hs*RosR is coloured in red. Binding reactions were performed in buffer containing 50 mM HEPES, pH 7.0, 3 M KCl and 0.02% azide. The binding affinity (K_d) is shown for each sequence.

measured the binding affinities for sequence S1 (25 bp-long) at different KCl concentrations, as well as in the presence of different salts. The binding affinity increased as salt levels rose from 2 M (K_d = 168.67 ± 30.24 nM) to 3 M KCl (K_d = 37.19 ± 2.36 nM). Upon increasing the salt concentration to 3.5 and then to 4 M KCl, the affinity remained roughly the same (K_d = 23.96 ± 3.37 and 25.87 ± 2.02 nM, respectively) (Figure 2A, Supplementary Table S4). Upon measuring the *hs*RosR–DNA interaction in different salts (3 M), the highest affinity was observed in KCl (K_d = 37.19 ± 2.36 nM) (Figure 2B) in line with this being the dominant salt in the cytoplasm of *Hbt. salinarum* (23). The lower binding affinity in NaCl (K_d = 78.27 ± 8.34 nM) can be assigned to the lower preference of Na^+ over K^+ for the vicinity of DNA (50). The lowest affinity was in RbCl (K_d = 80.24 ± 10.24 nM), possibly as a result of the low affinity of Rb^+ towards carboxylates (51). Binding affinity was completely abolished in either KBr or NaBr, despite the protein being stable and soluble in these salts (Figure 2B) (19).

*hs*RosR–DNA interactions in high-salt conditions were further investigated using ITC. As with FA, the binding affinity dramatically increased as the salt level rose from 2 M (K_d = 258.75 ± 9.75) to 3 M KCl (K_d = 91.05 ± 3.25) (Figure 2C and D). The binding stoichiometry in both experiments was approximately one, indicating that one protein dimer bound a single sequence (Supplementary Table S5). The positive ΔS and ΔH values obtained by ITC suggest that the *hs*RosR–DNA interaction is unfavourable in enthalpy terms and is instead driven by an increase in entropy (Supplementary Table S5). However, several attempts to further establish the nature of the enthalpy–entropy relations by performing ITC experiments at higher temperature failed, possibly because of adverse effects of temperature on the protein at 2–3 M salt.

FA measurements of *hs*RosR with sequence S1 (25 bp-long) in buffer containing bromide ions (NaBr and KBr) yielded no significant anisotropy value differences, even though the unbound protein was stable and soluble in these salts (19), indicating that *hs*RosR does not bind DNA in

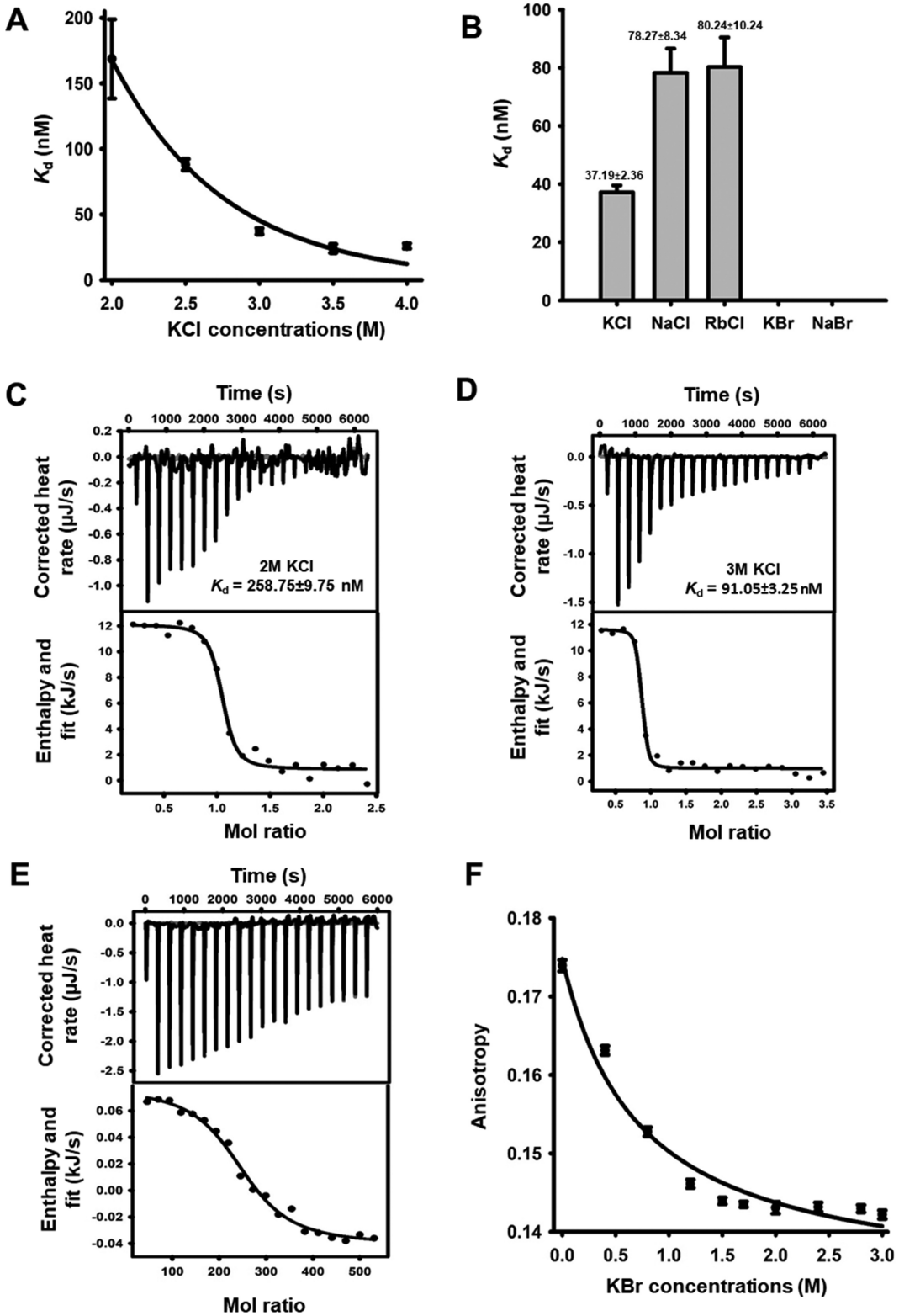

Figure 2. The effect of salt on DNA binding. (**A**) DNA binding at different KCl concentrations. The *X*-axis represents KCl concentrations, while the *Y*-axis represents the binding constant K_d. (**B**) The effect of different salts on binding affinity. The *X*-axis represents salt type, while the *Y*-axis represents the binding constant K_d. (**C**) Calorimetric titrations of sequence S1 (25 bp) into a solution containing *hs*RosR in buffer containing 20 mM HEPES, pH 7, 2 M KCl and 0.02% azide. Shown are the raw calorimetric data (top) and data fit to the integrated heat of the reaction (bottom). (**D**) The same as in panel (C) but the *hs*RosR was in buffer containing 20 mM HEPES, pH 7, 3 M KCl and 0.02% azide. (**E**) Calorimetric titrations of bromide ions (20 mM HEPES, pH 7, 3 M KCl, 100 mM KBr and 0.02% azide) into a solution containing the *hs*RosR–sequence S1 complex in buffer containing 20 mM HEPES, pH 7, 3 M KCl and 0.02% azide. Shown are the raw calorimetric data (top) and data fit to the integrated heat of the reaction (bottom). (**F**) The effect of bromide ions on sequence S1 (25 bp) binding. The *X*-axis represents KBr concentrations. The *Y*-axis represents anisotropic values.

presence of bromide ions. To gain insight into this phenomenon, we also investigated *hs*RosR–DNA complex separation using ITC and in a FA-based competition assay (Figure 2E and F). Titration of bromide ions into a solution containing the *hs*RosR–sequence S1 complex, over the control, led to separation of the complex (Figure 2E). These results can be correlated with the FA results showing that binding was completely abolished by gradually replacing chloride ions with bromide ions (Figure 2F). The same results were also observed for sequences S2 and S3 (data not shown).

The FA and ITC results clearly show the influence of KCl concentration and different salts on DNA binding by *hs*RosR. To better understand the effect of salt on such binding and the ability of *hs*RosR to bind DNA with specificity in high-salt conditions, crystallization of complexes comprising the protein and different DNA sequences was performed and their structures were determined.

Structure of the *hs*RosR–S1 complex

For initial *hs*RosR–DNA complex crystallization trials, the most enriched sequence, sequence S1 (Table 1) was used. This sequence was trimmed such that the binding motif was positioned exactly in the middle of the 28 bp-long DNA sequence (Supplementary Table S2). First, purified *hs*RosR was mixed with the DNA and complex formation was confirmed by FA (K_d = 53.46 ± 6.54 nM) and size exclusion chromatography (Supplementary Table S6). The elution volume of the complex indicated that a *hs*RosR dimer binds one DNA molecule and that the complex was soluble and stable in high salt-containing buffer (Supplementary Table S6). Inspection of the crystal packing of the free protein in the $P2_1$ form (19) suggested that DNA could not be incorporated into the lattice since the DNA-binding domain packed against neighbouring molecules and was hence blocked. Addition of $MnCl_2$, combined with a search for new crystallization conditions yielded crystals of protein–DNA complex in a new crystal form.

The structure of the complex was determined at a resolution of 2.0 Å (Supplementary Table S7). The crystal belongs to the P1 space group with two protein–DNA complexes in the asymmetric unit (i.e. P1 unit cell), with each complex consisting of one protein dimer and a 28-bp DNA duplex (Supplementary Figure S1). Interactions with *hs*RosR forced a 21.75° bend in the DNA (52). At the same time, no significant conformational changes were observed in the protein backbone relative to the free protein (PDB: 6FDH (19); RMSD = 0.531 Å for 214 atom pairs; Supplementary Figure S2). The recognition helices (α3) from both subunits penetrate and interact with the major groove of the DNA, in addition to the wing regions that interact with the minor groove (Figure 3). In both subunits, interactions with the major groove are mediated through helix α3 of the wHTH domain and involve His50, the first amino acid in the helix, which directly binds to the guanine base (G8) in the middle of the binding motif of the DNA (T**G**T). Lys68 from the wing also binds to the same guanine base via the phosphate to further stabilize the complex. Leu34, Gly33 and Asn76 interact with the phosphate of the T7 base (TG**T**) via their backbone amide groups, together with hydroxyl of Tyr54. The side chains of Asn49, Arg52 and Asn56 interact with the phosphates of bases T17 and C18 from the complementary strand (Figure 3). Leu34 in both subunits is also involved in van der Waals interactions with bases G6 and A6, through their aliphatic side-chains. Interactions with the minor groove are mediated by residues from the wing. Arg74 from both subunits penetrates between the DNA strands to form hydrogen bonds with bases from both strands in the minor groove (e.g. C5, G24 and T25). The amide group of Lys73 forms hydrogen bond with the phosphate base T26 in both subunits to further stabilize the interactions of the protein with the DNA minor groove (Figure 3).

Structures of *hs*RosR complexes with other enriched sequences containing the TGT-N_{10}-ACA motif

The sequence percentage enrichment obtained in the SELEX experiments and the FA experiments showed that *hs*RosR is able to bind different DNA sequences with affinities that parallel their ranking in terms of binding (Table 1 and Supplementary Table S3). While all the enriched sequences contain the same DNA motif, they differ substantially in terms of base composition within the variable region of the motif as well as in the flanking regions. We determined the crystal structures of *hs*RosR bound to different sequences presenting different binding affinities and analysed their relationships with *hs*RosR–S1. Specifically, crystallization of *hs*RosR with DNA sequences S2–S5 (Table 1) was attempted (see Supplementary Table S2). As in sequence S1, the binding motif recognized by *hs*RosR is also found exactly in the middle of the 28 bp-long DNA sequence in these sequences (Supplementary Table S2). Despite all attempts made, sequences S2 and S3 (Table 1) did not form crystals with *hs*RosR at any sequence length or under any crystallization conditions. However, sequences S4 and S5 (Table 1) did form crystals of complexes with *hs*RosR (Supplementary Table S2). These structures were solved at 2.13 and 2.43 Å resolution, respectively (Supplementary Table S7), using molecular replacement (as described for the *hs*RosR–S1 structure).

Superimposing the *hs*RosR–S1 and *hs*RosR–S4 structures shows high similarity between the structures (RMSD = 0.202 Å over 231 atom pairs of protein dimers; Figure 4A). Most of the *hs*RosR–S1 interactions were maintained in the *hs*RosR–S4 structure, except for a few alterations. The substitution of G5 (in sequence S1) by A5 (in sequence S4) reduced the ability of Arg74 in the *hs*RosR wing area to interact with the DNA (Supplementary Figure S3). In addition, the thymidine base in sequence S1 (T26) is changed to guanine in sequence S4 (G26), thus reducing the ability of Arg74 from the second *hs*RosR subunit to interact with the DNA (Supplementary Figure S3). Other interactions are also absent in the *hs*RosR–S4 and *hs*RosR–S5 structures compared to *hs*RosR–S1. Supplementary Table S10 lists the hydrogen bonds and electrostatic interactions in the four protein–DNA complexes studied here together with their binding constants. The number of specific protein–DNA interactions in the complexes seems to be somewhat inversely correlated with the binding affinities and could po-

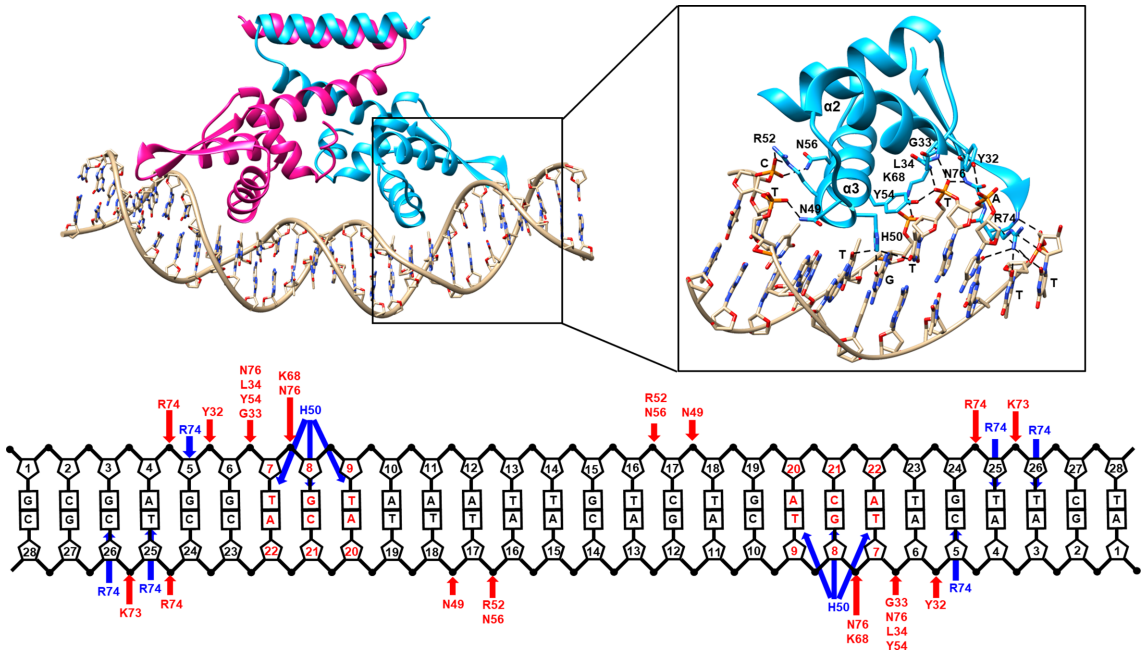

Figure 3. *hs*RosR–sequence S1 structure. The overall complex structure is shown as a ribbon diagram. Each subunit is coloured magenta or cyan. The interactions of the recognition helix and the wing are shown. A linear scheme of *hs*RosR–S1 interactions is also shown. Red arrows represent interactions of *hs*RosR with the phosphate backbone, while blue arrows represent interactions of *hs*RosR with DNA bases.

tentially contribute to the preference of *hs*RosR for different sequences (Supplementary Table S10).

Quite a pronounced difference between the *hs*RosR–S1 structure and the structures of the other enriched sequences is seen in the width of the DNA minor grooves. The high-affinity S1 sequence contains an AT-rich centre (T13-T16) that is not in direct contact with the protein but whose minor groove is severely compressed, relative to average B-form DNA (Figure 4B). The S4 sequence, for example, contains four central GC base pairs (C13-G16). These substitutions cause a wider minor groove in the latter structure (Figure 4A and B). The structure of *hs*RosR–S5 shows similar features in terms of interactions with the protein and width of the minor groove (Figure 4B). Thus, the structural data extracted from the S1-, S4- and S5-containing complexes, together with the FA results, suggest that the affinity of *hs*RosR for its DNA target is not simply dependent upon recognition of DNA functional groups by the protein side chains (i.e. direct readout) but also relies on an indirect readout mechanism of the minor groove width. The crystal structures of the *hs*RosR complexes together with their affinities support this notion (Table 1 and Figure 4).

Comparison of *hs*RosR–DNA interactions with mesophilic MarR/PadR–DNA interactions

The structures of *hs*RosR–DNA complexes described here show that binding of the *hs*RosR **TGT**-N_{10}-**ACA** motif is mediated by wHTH domains from both monomers such that the recognition helices insert into the major groove in two locations along the DNA helix. A similar binding pattern is employed by the PadR transcription factor from *Bacillus subtilis* (53) that binds a shorter palindromic **TGT**-N_8-**ACA** motif (Figure 5A) and also by the oxidation-sensing MosR from *Mycobacterium tuberculosis* (39), which binds an even shorter palindromic motif, **TGT**-N_4-**ACA** (Figure 5A). Since the three **TGT**-N_x-**ACA** motifs are of different lengths, the spatial organization of the wHTH domains around the DNA differs among the three proteins, as demonstrated schematically for the recognition helices in Figure 5B and D and which follows from the different dimeric structures of the proteins. Superposition of the complete $PadR^{dsDNA}$ structure (PDB code: 5X11) on the structure of *hs*RosR–S1 shows that only one wHTH-containing monomer of $PadR^{dsDNA}$ overlaps one monomer of *hs*RosR (Supplementary Figure S4A; RMSD = 0.84 Å over 58 atom pairs). The same holds true for superposition of the MosR structure (PDB code: 4FX4) on *hs*RosR (Supplementary Figure S4B; RMSD = 0.93 Å over 109 atom pairs). Despite the different dimeric organization, *hs*RosR–S1 and $PadR^{dsDNA}$ share common DNA-binding residues from their α3 helix; *hs*RosR His50, Tyr32, Tyr 54 and Arg52 and their PadR counterparts His38, Tyr20, Tyr42 and Gln40 appear to contribute to binding in similar manners (Figure 5C). In MosR, different residues from the recognition helix play the same role, namely Arg70, Thr71, Thr72, Arg75 and Asn76 together with Lys97 and Arg95 from the wing. Thus, the hypersaline environment in which *hs*RosR operates is not reflected in protein–DNA contacts of the complex. Rather, as explained below, the salinity effect manifests itself in the stages leading to formation of the protein–DNA complex.

Not surprisingly, MarR proteins with recognition motifs other than **TGT**-x-**ACA** (Supplementary Table S11) also show variations in the dimeric organization around the recognition motif, as well as in the residues participating in DNA binding from the recognition helix α3 and the wing.

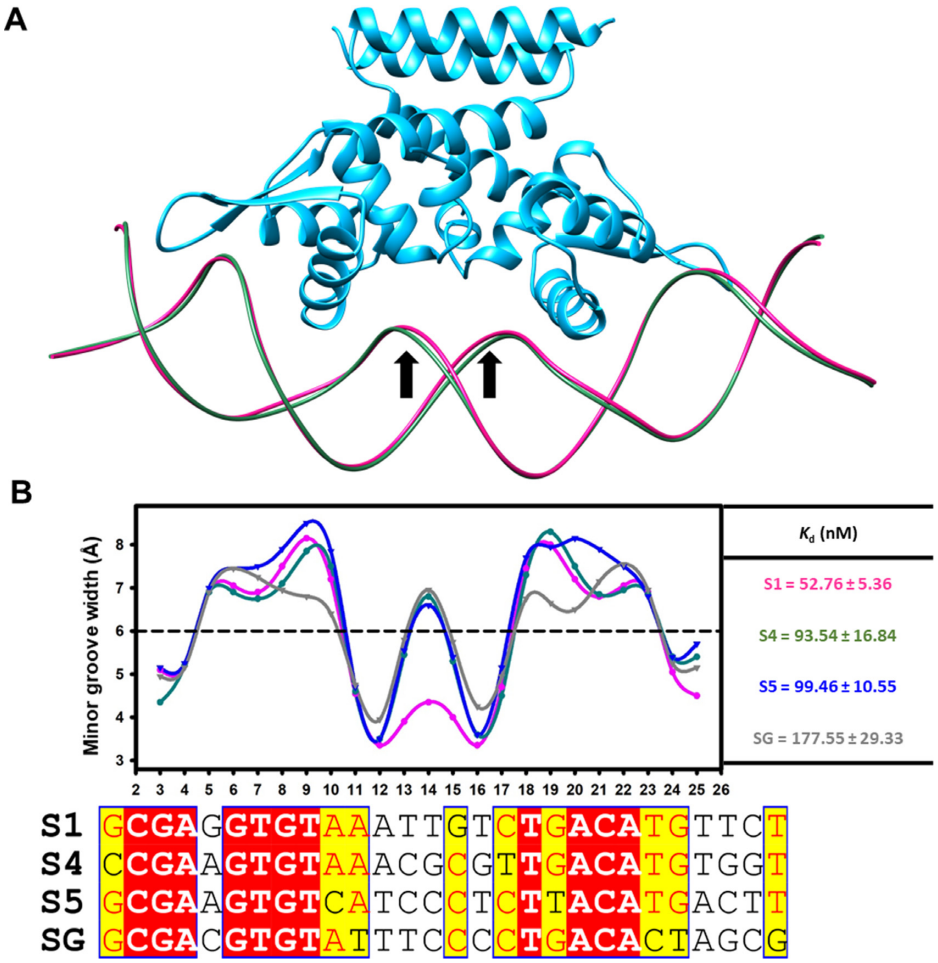

Figure 4. Minor groove widths within crystal structures of *hs*RosR–DNA complexes. (**A**) Superposition of the structures of *hs*RosR–S1 (pink) and *hs*RosR–S4 (green). (**B**) Minor groove width plots of sequences S1 (pink), S4 (green), S5 (blue) and SG (grey) DNA structures and sequence alignment. Values reflect interphosphate distances, minus the van der Waals surface (5.8 Å). The average minor groove width of B-form DNA (6 Å) is shown. Minor groove widths were calculated using the 3-DNA server (3). A table of binding affinity constants for each sequence is shown. The sequences were aligned using Tcoffee-Expresso (1) and are displayed so as to show secondary structure elements, as determined by ESPript3.0 (2). Bases in the red box are identical, while those in yellow box are conserved. The blue frame indicates similarity across proteins.

Supplementary Figure S8 shows these structural variations for two proteins from those listed in Supplementary Table S11. With the notable exception of helix α3 in MepR (54), residues involved in electrostatic interactions and hydrogen bonds dominate the composition of helix α3, as observed for *hs*RosR, PadR and MosR and other MarR proteins (see Supplementary Figures S8, S9 and Supplementary Table S11). Arginine in the wing domain is frequently encountered as a prominent feature in all the wHTH proteins presented here and probably also in other proteins assuming this topology (e.g. see Supplementary Figure S8).

Defining the genomic sequences bound by *hs*RosR

Having determined by SELEX the binding motif recognized by *hs*RosR, we subsequently asked whether this motif exists in the *Hbt. salinarum* genome and whether the interactions detected in the SELEX experiments are biologically relevant. While previous ChIP-ChIP studies proposed genomic binding sites for *hs*RosR under physiological conditions and during oxidative stress (17,18), the exact genomic *hs*RosR-binding sequences remain to be defined.

Scanning the *Hbt. salinarum* genome for the DNA motif recognized by *hs*RosR (**TGT**-N_{10}-**ACA**) yielded 852 candidate target sequences (only in the main chromosome). Since this is a relatively large number of targets, we performed gSELEX (22) experiments to define the exact genomic binding sites of *hs*RosR. The gSELEX experiments were conducted exactly as were the SELEX experiments, except for the initial library used. Here, the initial library comprised ∼150 bp-long genomic DNA fragments. Using gSELEX, we identified 22 peaks (score >1000) across the *Hbt. salinarum* genome (Supplementary Table S8). The five strongest peaks detected in experiments performed at either 3 or 4 M KCl were similar (data not shown). The genomic target sites identified are in agreement with previous results of Toner *et al.*, who investigated *hs*RosR genomic target sites using a ChIP-ChIP approach (Figure 6 and Table 2) (18). A similar motif as identified in the SELEX experiments (**TGT**-N_{10}-**ACA**) was observed in all five of the

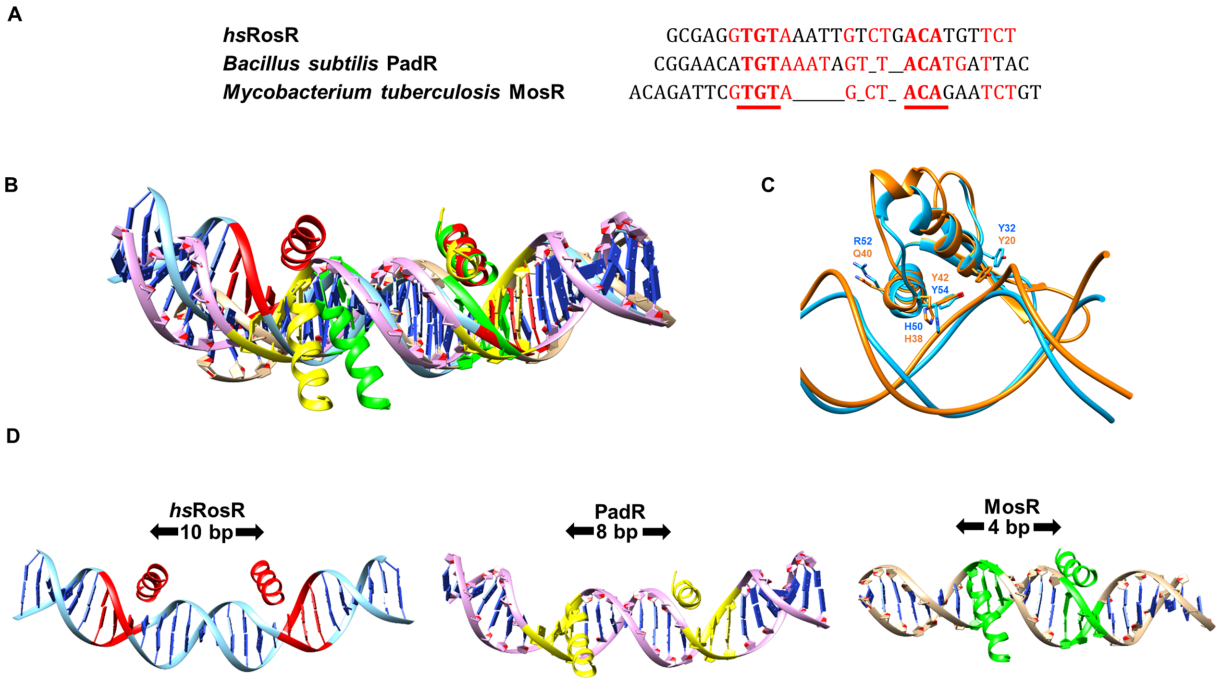

Figure 5. Comparison of DNA binding by *hs*RosR-S1with PadR^{dsDNA} and MosR. (**A**) Alignment of sequence S1, the sequence used in crystallization of PadR^{dsDNA} (PDB code: 5X11) and the sequence used in crystallization of MosR from *Mycobacterium tuberculosis* (PDB code: 4FX4). (**B**) Superposition of the DNA and the recognition helices of *hs*RosR–sequence S1 (cyan and red, respectively), PadR^{dsDNA} (pink and yellow, respectively) and MosR (grey and green, respectively). The complete structures were superimposed as described in the text. Only the relevant parts are shown. (**C**) Superposition of *hs*RosR–sequence S1 (cyan) and PadR^{dsDNA} (gold) complexes involving protein monomers. The common amino acids that participate in DNA binding are shown. (**D**) The DNA and the recognition helices of *hs*RosR–sequence S1 (cyan and red, respectively), PadR^{dsDNA} (pink and yellow, respectively) and MosR (grey and green, respectively) are shown.

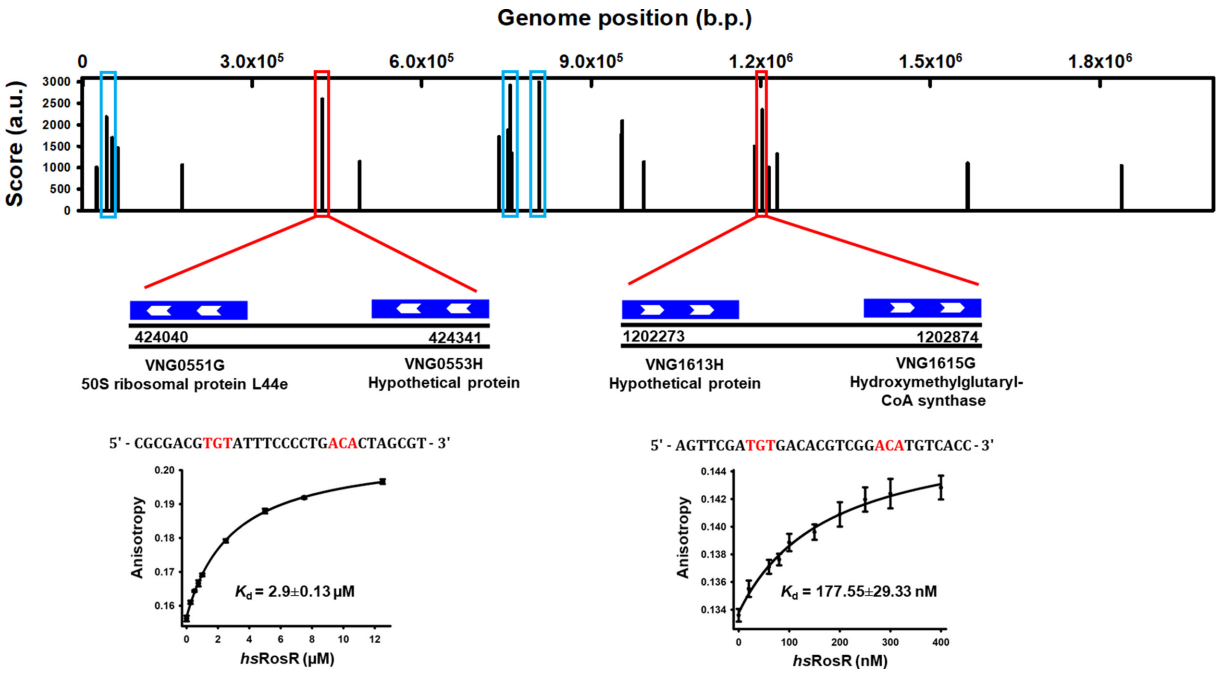

Figure 6. Genomic SELEX results. The 22 target sites of *hs*RosR (peak score > 1000) are plotted across the *Hbt. salinarum* genome (main chromosome). The five strongest peaks are marked as follows. The peaks marked in light blue are correlated to previously reported binding sites (15), while the peaks marked in red represent the new binding sites for *hs*RosR. Zoom-ins of the new binding locations (reflecting high and low affinity sites) are shown, as are the sequences, binding curves and K_d values.

strongest peaks detected in the gSELEX experiments, indicative of the agreement between the gSELEX and SELEX results (Figure 6 and Supplementary Table S9). The five strongest peaks areas identified by gSELEX are found in putative promoter regions of *vng1059c-vng1060h* (i.e. the site is between those two genes), *vng0991h, vng0551g, vng1615g* and *vng0050c* (Figure 6 and Table 2), as would be expected for a transcription factor like *hs*RosR (55,56). Using gSELEX, we were also able to identify new binding sites, namely in the promoters of *vng0551g* and *vng1615g* (the third and fourth strongest peaks, respectively). These sites were not previously reported as binding *hs*RosR (Figure 6 and Table 2).

To define more precisely the genomic target site(s), sequences containing the binding motif found in the five strongest peak areas identified by gSELEX (30 bp-long) were subjected to a FA-based binding assay (Figure 6 and Supplementary Table S9). Surprisingly, *hs*RosR bound only to the new binding sites, namely the third and fourth strongest peaks, corresponding to a high affinity site located in the promoter region of *vng1615g* and a low affinity site located in the promoter region of *vng0551g*, respectively (Figure 6, Table 2 and Supplementary Table S9).

Finally, we also determined the crystal structure of *hs*RosR bound to the sequence found in the high affinity site of *vng1615g* (termed sequence SG; Supplementary Table S2). A similar mode of DNA-binding, interactions with the protein, as well as correlation between DNA minor groove width and binding affinity were observed in the structure of *hs*RosR bound to sequence SG as seen in the structures of *hs*RosR–S1, *hs*RosR–S4 and *hs*RosR–S5 SELEX-identified sequences (i.e. sequences S1, S4 and S5) (Figure 4).

Investigating the effects of the hypersaline environment on the electrostatic and DNA-binding properties of *hs*RosR

It was shown by Elcock and McCammon (48) that the screening effect attributed to the high concentrations of salts in the hypersaline environment (~5 M) contributes significantly to the stability of halophilic proteins by causing a rise in the p*K*a values of acidic residues on the protein surface. This salt-related effect should also manifest itself as global changes in the surface charge distribution that is known to be critical for protein–DNA interactions (3,57). We, therefore, investigated the effect of salt concentration on the surface electrostatics of *hs*RosR and on the electrostatic contribution to the DNA-binding energy using the Poisson–Boltzmann approach as implemented in the Delphi (44,58,59) and APBS (45) programs. The dielectric constants used throughout the calculations were those recommended for halophilic proteins (48): $\varepsilon = 20$ for the protein and salt concentration-dependent ε values for the solvent (48,60). Calculations were performed on the *hs*RosR–S1 complex.

The results are shown in Supplementary Table S12 and Supplementary Figure S10. At zero salt concentration, the average electrostatic potential on the surface of *hs*RosR and DNA is extremely negative: −17.5 and −22.9 kT/e, respectively, with no evidence of positive potential on the protein surface (Supplementary Table S12A). Increasing the salt concentration significantly reduces the average negative potential while amplifying the positive potential on the protein surface. The effect of the increased salt concentration is also demonstrated by probing the electrostatic surface potential at several specific DNA contact points in the binding site (Supplementary Figure S10A). The dramatic difference with the electrostatic surface potential of the DNA-binding site of the mesophilic PadR is shown for comparison (Supplementary Figure S10B). As in other cases of formation of protein–DNA complexes (3,57), the electrostatic contribution to the binding free energy of the complex reflects mostly unfavourable electrostatic interactions that decreases with the salt concentration reaching a value of +35.9 kJ/mole at 5 M (Supplementary Table S12B). Additional factors contribute to binding free energy and formation of the protein–DNA complexes (see 'Discussion' section).

DISCUSSION

The effect of salts on the formation of protein–DNA complexes has been extensively studied (3,11,57). In mesophiles, increasing the salt concentration weakens binding (61), whereas in halophiles the salt effect is reversed. The protein-binding process is driven by long-range electrostatic interactions between the positively charged DNA-binding site on the protein and the negatively charged DNA (3,57). These favourable electrostatic interactions are, however, opposed by unfavourable changes in the solvation of both protein and DNA associated with removal of charged and polar groups from the solvent and displacement of salt ions around the DNA, together with conformational changes in the protein and DNA (3,62). The major force driving the protein–DNA binding process appears to be the release of water molecules from the surfaces of the protein and DNA that become buried upon complexation (3). Specificity of binding is achieved through attraction between the positively charged binding site and the cognate DNA. In the case of *hs*RosR, the encounter between the DNA-binding site and the DNA is hampered by the overall extreme negative charge on the protein surface (56 Glu/Asp per protein dimer versus 30 Arg/Lys) and moreover, by the negatively charged residues surrounding the DNA-binding site (see Figure 2 and Supplementary Figure S10A (19)). Screening by the high-salt concentration essential for stabilization of halophilic proteins (48) markedly reduces the negative charge on the protein surface. Our electrostatic calculations demonstrate the clear effect of high-salt concentrations on the overall surface potential and particularly in the binding site. Only at high-salt concentrations can the positive charge in the binding site exert sufficient attraction on the DNA, leading to the formation of a stable complex (Supplementary Table S12A). Probing the electrostatic potential at several DNA contact points in the binding site also reveals an increase in the positive potential as a function of the salt concentration (Supplementary Figure S10A). The sharp contrast with electrostatic potential in the DNA-binding site of the mesophilic PadR is demonstrated in Supplementary Figure S10B.

To get a rough estimate of the various contributions to the binding free energy, we followed steps described in the context of other protein–DNA complexes (3,57–59). We calculated using Delphi the electrostatic contribution to the

Table 2. Comparison of genomic SELEX results with previous results

Genomic SELEX position			Score	Known position	Protein encoded
vng1059c/vng1060h	GS1	807642-808110	2993.09	807561-807802	Colanic acid biosynthesis glycosyltransferase WcaL – hypothetical protein
vng0991h	GS2	756720-757226	2916.54	753242- 757251	Hypothetical protein
vng0551g	**GS3**	**424040-424341**	**2596.19**	**New binding site**	**50S ribosomal protein L44e**
vng1615g	**GS4**	**1202273-1202874**	**2353.24**	**New binding site**	**Hydroxymethylglutaryl-CoA synthase**
vng0050c	GS5	42458-42641	2181.65	41706 –48812	Lipopolysaccharide biosynthesis protein

The identities of the five strongest peaks were compared to *hs*RosR-binding sequences previous reported (15). Genes, genome positions, gSELEX scores and the proteins encoded by those genes are listed. The new genomic positions found in this work are marked in red.

binding free energy in various salt concentrations and derived values that are unfavourable to complex formation, as in other cases (3) (Supplementary Table S12B). When combined with the ΔG of −103.7 kJ/mole arising from burying surface residues at the protein–DNA interface and release of water molecules (46), the results differ from the experimental ΔG of binding (derived from the binding constants; see Supplementary Table S4) by +20 kJ/mole on average (Supplementary Table S12B), most likely due to neglecting unfavourable entropic interactions associated with conformational changes in the protein and DNA upon complex formation (3). Given the uncertainties in the model underlying the calculations, the estimate of the free energy of binding is acceptable, as also pointed out in other well-studied cases (3). Altogether, the Poisson–Boltzmann approach rationalizes quite well the DNA binding as a function of salt concentration observed in our experiments. It also emphasizes the contribution of the high-salt concentration to the early stages leading to formation of protein–DNA complex. The actual protein–DNA interactions observed in the final structures of *hs*RosR complexes do not involve salt ions (Figure 3 and Supplementary Table S10). Rather, they share similarities with such interactions observed in wHTH mesophilic proteins.

Our SELEX experiments identified the DNA motif recognized by *hs*RosR as **TGT**-N_{10}-**ACA** (Figure 1 and Table 1; Supplementary Table S3). The SELEX results showed that like other members of the MarR/PadR families, *hs*RosR binds a palindromic motif. gSELEX experiments yielded a list of 22 significant peaks that were assigned as the genomic targets bound by *hs*RosR. Measurable binding affinities could be recorded for only two of these sequences, namely a relatively high affinity site located in the promoter region of *vng1615g* (encoding hydroxymethylglutaryl-CoA synthase) and a lower affinity site located in the promoter region of *vng0551g* (encoding the 50S ribosomal protein L44e) (Figure 6 and Table 2; Supplementary Table S9). The biological significance of *hs*RosR binding to these novel genomic targets is yet unknown, including whether *hs*RosR activates or represses these genes. Compared to an earlier study (18), we identified three additional DNA target sites (the first, second and fifth most strongly gSELEX-bound sequences, namely the promoters to *vng1059c-vng1060h*, *vng0991h* and *vng0050c*; Table 2), reinforcing the idea that *hs*RosR is also involved in the regulation of these genes. The existence of binding sites recognized with different affinities could imply different modes of regulation, although further experiments are needed to clarify this point.

The SELEX and gSELEX results agree on the binding; however, the sequences recognized by *hs*RosR differ in terms of base composition within and bordering the binding motif, leading to differences in the binding affinity (Figure 6 and Table 1; Supplementary Tables S3 and S9). The structural data derived from the complex structures provided insight into the effect of DNA shape and sequence on the binding and suggested that direct and indirect readout mechanisms contribute to such binding (63,64). The direct mechanism involves specific interactions of *hs*RosR with the binding motif while, for example, in the case of the AT-rich N_{12}-N_{17} segment in sequence S1, the narrower minor DNA groove is also important for indirect recognition and binding affinity (Figure 4). The effects of base pair sequence composition on DNA shape and affinity have been extensively studied and addressed in studies of other DNA-binding proteins (1,64–69). This dual DNA-binding mechanism most likely explains the preference of *hs*RosR for different sequences. The same mechanism also provides rationale for the fact that not all the sequences identified by *in vitro* gSELEX showed measurable binding affinity to *hs*RosR, even though they contain the same recognition motif.

Both the SELEX and the gSELEX experiments showed how critical the salt concentration is for DNA binding. To more fully understand the effect of salt, we assessed *hs*RosR–S1 interactions in different KCl concentrations using FA. Unlike non-halophilic protein–DNA interactions, the binding affinity of *hs*RosR to its preferred DNA sequence increased as a function of salt concentration (Figure 2A). At 2 M salt concentration the protein is not well folded (19). Increasing the concentration from 2.5 to 4 M KCl augments the DNA-binding affinity, reflecting the stabilizing effect of higher salt concentration on *hs*RosR and on the electrostatic surface potential (Supplementary Table S12A (48,70)). The effect of salt concentration is also observed in the ITC measurements conducted in 2 and 3 M KCl (Figure 2C and D). Tryptophan emission scanning of *hs*RosR in different salt concentrations showed slightly different conformations when performed in 2 or 3 M KCl

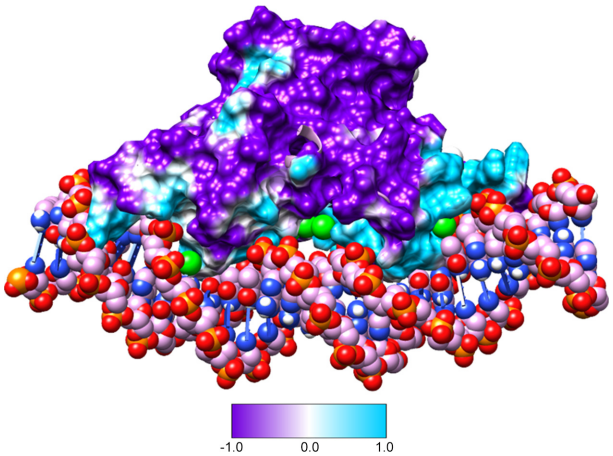

Figure 7. Superposition of the structures of the *hs*RosR–sequence S1 complex and of *hs*RosR solved from NaBr (PDB code: 6EZ1). The *hs*RosR protein structure is shown as electrostatic surface potential (units: $\pm 1\ kt/e$), while the DNA is shown as a sphere. Only the bromide ions of the *hs*RosR solved from NaBr are shown as green balls.

(Supplementary Figure S5). Once *hs*RosR attains its active conformation (in ∼3 M KCl), the binding affinity for DNA remains roughly the same, even as the salt concentration is raised to 4 M (Figure 2A and Supplementary Table S4).

In previous work, we used anomalous diffraction from bromides to locate a large number of anions in positively charged areas on the protein surface and in the DNA-binding site (see Figure 3 in (19)). As in other halophilic proteins, the presence of the ionic shells and their interaction with highly enriched acidic residues typical of such proteins rationalize their ability to remain soluble and stable in the hypersaline environment (19,48,71,72). In the current study, anomalous scattering from bromides could not be used for detecting ions directly in the crystals simply because, as explained above, bromides prevent formation of *hs*RosR–DNA complex. However, superposition of the electrostatic surface of *hs*RosR structure in NaBr (PDB code: 6EZ1 (19)) with the *hs*RosR–S1 structure clearly shows that bromide ions (or chlorides (19)) occupying the DNA-binding site would have to be released in order for DNA to bind to the site (Figure 7 and Supplementary Figure S6) and prevent electrostatic repulsion with DNA phosphates. Movement of ions upon formation of protein–DNA complex due to release from the protein and DNA surfaces is an important factor in theoretical and biophysical studies of DNA-binding processes (3). Here, we have a clear demonstration of movement of ions from the DNA-binding site.

The effect of salt composition on DNA binding to proteins is known (4,73,74). Most striking in the case of *hs*RosR is the effect of bromides that completely abolish formation of a protein–DNA complex. Furthermore, crystals of the complex dissolved upon addition of minute amounts of 2.5–3.0 M KBr solution (not shown). Both the ITC and FA experiments clearly demonstrated competition between DNA and bromide ions for binding the protein. In fact, the results suggested that bromide ions have higher affinity for the protein, as compared to the DNA. Hence, by binding to the protein, especially in the DNA-binding site, bromide ions essentially push out and replace the DNA, leading to complex separation.

The total protein and DNA surfaces area buried upon DNA binding is 1891 Å^2 contributing −106.3 kJ/mole (46) to the free energy of binding. Ions as well as water molecules in the DNA-binding site and around the DNA rearrange so as to physically allow binding to occur. As shown by ITC, the DNA-binding reaction is endothermic and could be entropy-driven (Figure 2C and D; Supplementary Table S4). The maintenance of interactions between the negatively charged surface of halophilic proteins and nucleic acids in a hypersaline environment requires specific structural adjustments (75). Comparison of the interactions of *hs*RosR and mesophilic PadR from *Bacillus subtilis* bound to their cognate DNA sequences identified common residues that participate in the binding, suggesting similar modes of DNA binding (Figure 5). The structures of *hs*RosR complexes showed that DNA binding is mediated by the wHTH motif, similar to what is seen with other MarR/PadR proteins. Therefore, we conclude that once the encounter between DNA and the DNA-binding site takes place in the hypersaline medium, the halophilic *hs*RosR protein recognizes its DNA targets much as do mesophilic proteins in a low salt environment. This conclusion is in accord with the preservation of function common to halophilic proteins on the background of the extensive substitution of surface residues necessary for adaptation to their hypersaline environment, as we previously discussed (19). In the context of DNA binding to *hs*RosR, the high-salt concentration in the *Hbt. salinarum* cytoplasm (∼5 M (23)) thus serves a dual role, namely stabilization of the protein and maintenance of surface electrostatic potential conducive to specific DNA-binding.

In summary, our study sheds light on the ability of *hs*RosR to specifically recognize and bind its DNA targets in a hypersaline environment where electrostatic interactions are in general weakened. A currently unresolved question asks how *hs*RosR functions in response to oxidative stress. Indeed, the fact that *hs*RosR lacks cysteine residues, known to be essential for the activity of other known RosR proteins for ROS signalling (47,76–78), raises questions as to the exact mechanism employed by *hsRosR*.

DATA AVAILABILITY

The atomic coordinates and structure factors for *hs*RosR–S1 (PDB code: 6QIL), *hs*RosR–S4 (PDB code: 6QFD), *hs*RosR–S5 (PDB code: 6QH0) and *hs*RosR–SG (PDB code: 6QUA) complexes have been deposited in the Worldwide Protein Data Bank (https://www.wwpdb.org)

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

At Ben Gurion University, we thank Eyal Gur for access to the Fluorolog-3 instrument and assistance with FA measurements and Sofia Kolusheva for assistance with ITC and fluorescence spectrometry measurements. Members of the

Technion Genome Center staff advised us and carried out the deep-sequencing analyses.
Diffraction data for *hs*RosR–S1 (PDB code: 6QIL), *hs*RosR–S5 (PDB code: 6QH0) and *hs*RosR–SG (PDB code: 6QUA) were recorded on beamline ID-29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Dr. David Flot and Dr. Antoine Royant for providing assistance with the experiments. We also acknowledge the Paul Scherrer Institute, Villigen, Switzerland for provision of synchrotron radiation beam time on beamline Xo6DA–PXIII of the SLS for recording diffraction data for *hs*RosR–S4 (PDB code: 6QFD) and would like to thank Dr. Basu Shibbom for assistance.
Author contributions: N.K. carried out the molecular biology and biochemical experiments. F.S. and G.D. provided technical assistance. D.B. provided guidance in the SELEX experiments. N.K. and A.S. recorded diffraction data. N.K. and B.S. solved, refined and interpreted the structures. N.K. and B.S. wrote the manuscript with assistance from J.E. and R.Z. B.S. designed and supervised the project.

FUNDING

Israel Science Foundation (ISF) [1383/13 to B.S.; 109/16 to J.E.]. Funding for open access by Ben Gurion University and ISF grants.
Conflict of interest statement. None declared.

REFERENCES

1. Rohs,R., Jin,X., West,S.M., Joshi,R., Honig,B. and Mann,R.S. (2010) Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.*, **79**, 233–269.
2. Bergqvist,S., Williams,M.A., O'Brien,R. and Ladbury,J.E. (2004) Heat capacity effects of water molecules and ions at a Protein–DNA interface. *J. Mol. Biol.*, **336**, 829–842.
3. Misra,V.K., Hecht,J.L., Yang,A.S. and Honig,B. (1998) Electrostatic contributions to the binding free energy of the lambdacI repressor to DNA. *Biophys. J.*, **75**, 2262–2273.
4. Hargreaves,V.V. and Schleif,R.F. (2008) Salt dependence of IRF1 DBD binding to DNA reveals ions are localized around protein and DNA. *Biochemistry*, **47**, 4119–4128.
5. Loregian,A., Sinigalia,E., Mercorelli,B., PalÃ[1],G. and Coen,D.M. (2007) Binding parameters and thermodynamics of the interaction of the human cytomegalovirus DNA polymerase accessory protein, UL44, with DNA: Implications for the processivity mechanism. *Nucleic Acids Res.*, **35**, 4779–4791.
6. Jen-Jacobson,L., Engler,L.E. and Jacobson,L.A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure*, **8**, 1015–1023.
7. Arosio,D., Costantini,S., Kong,Y. and Vindigni,A. (2004) Fluorescence anisotropy studies on the ku-DNA interaction: Anion and Cation Effects. *J. Biol. Chem.*, **279**, 42826–42835.
8. Lohman,T.M., Overman,L.B., Ferrari,M.E. and Kozlov,A.G. (1996) A highly salt-dependent enthalpy change for escherichia coli SSB Protein−Nucleic acid binding due to Ion−Protein interactions. *Biochemistry*, **35**, 5272–5279.
9. Fogolari,F., Elcock,A.H., Esposito,G., Viglino,P., Briggs,J.M. and McCammon,J.A. (1997) Electrostatic effects in homeodomain-DNA interactions. *J. Mol. Biol.*, **267**, 368–381.
10. O'Brien,R., DeDecker,B., Fleming,K.G., Sigler,P.B. and Ladbury,J.E. (1998) The effects of salt on the TATA binding protein-DNA interaction from a hyperthermophilic archaeon. *J. Mol. Biol.*, **279**, 117–125.
11. Record,M.T. Jr, Anderson,C.F. and Lohman,T.M. (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: The roles of ion association or release, screening, and ion effects on water activity. *Q. Rev. Biophys.*, **11**, 103–178.
12. Ginzburg,M., Sachs,L. and Ginzburg,B.Z. (1970) Ion metabolism in a *halobacterium*. *J. Gen. Physiol.*, **55**, 187–207.
13. Dym,O., Mevarech,M. and Sussman,J.L. (1995) Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. *Science*, **267**, 1344–1346.
14. Frolow,F., Harel,M., Sussman,J.L., Mevarech,M. and Shoham,M. (1996) Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.*, **3**, 452–458.
15. Kennedy,S.P., Ng,W.V., Salzberg,S.L., Hood,L. and DasSarma,S. (2001) Understanding the adaptation of halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.*, **11**, 1641–1650.
16. Siglioccolo,A., Paiardini,A., Piscitelli,M. and Pascarella,S. (2011) Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct. Biol.*, **11**, 50.
17. Sharma,K., Gillum,N., Boyd,J. and Schmid,A. (2012) The RosR transcription factor is required for gene expression dynamics in response to extreme oxidative stress in a hypersaline-adapted archaeon. *BMC Genomics*, **13**, 351.
18. Tonner,P.D., Pittman,A.M.C., Gulli,J.G., Sharma,K. and Schmid,A.K. (2015) A regulatory hierarchy controls the dynamic transcriptional response to extreme oxidative stress in archaea. *PLoS Genet*, **11**, 1–13.
19. Kutnowski,N., Shmuely,H., Dahan,I., Shmulevich,F., Davidov,G., Shahar,A., Eichler,J., Zarivach,R. and Shaanan,B. (2018) The 3-D structure of VNG0258H/RosR - A haloarchaeal DNA-binding protein in its ionic shell. *J. Struct. Biol.*, **204**, 191–198.
20. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
21. Gold,L., Polisky,B., Uhlenbeck,O. and Yarus,M. (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.*, **64**, 763–797.
22. Lorenz,C., von Pelchrzim,F. and Schroeder,R. (2006) Genomic systematic evolution of ligands by exponential enrichment (genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat.Protoc.*, **1**, 2204–2212.
23. Christian,J.H.B. and Waltho,J.A. (1962) Solute concentrations within cells of halophilic and non-halophilic bacteria. *Biochim. Et Biophys. Acta*, **65**, 506–508.
24. Stoltenburg,R., Reinemann,C. and Strehlitz,B. (2007) SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.*, **24**, 381–403.
25. Bouvet,P. (2015) Identification of nucleic acid high affinity binding sequences of proteins by SELEX. *Methods Mol. Biol.*, **1334**, 333–343.
26. Schütze,T., Wilhelm,B., Greiner,N., Braun,H., Peter,F., Mörl,M., Erdmann,V. A., Lehrach,H., Konthur,Z., Menger,M. *et al.* (2011) Probing the SELEX process with next-generation sequencing. *PLoS ONE*, **6**, 29604.
27. Goecks,J., Nekrutenko,A., Taylor,J. and The Galaxy Team. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, R86.
28. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
29. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-seq (MACS). *Genome Biol.*, **9**, R137.
30. Lundblad,J.R., Laurance,M. and Goodman,R.H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol. Endocrinol.*, **10**, 607–612.
31. Pierce,M.M., Raman,C.S. and Nall,B.T. (1999) Isothermal titration calorimetry of protein-protein interactions. *Methods*, **19**, 213–221.
32. Ilian,J. and Rudolf,B.H. (1990) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recog.*, **12**, 3–18.
33. Kabsch,W. (2010) Xds. *Acta Crystallogr. Sect. D*, **66**, 125–132.
34. Kabsch,W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D*, **66**, 133–144.

35. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D*, **69**, 1204–1214.
36. Evans,P.R. (2011) An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D*, **67**, 282–292.
37. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A. *et al.* (2010) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D: Biol. Crystallogr.*, **67**, 235–242.
38. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
39. Brugarolas,P., Movahedzadeh,F., Wang,Y., Zhang,N., Bartek,I.L., Gao,Y.N., Voskuil,M.I., Franzblau,S.G. and He,C. (2012) The oxidation-sensing regulator (MosR) is a new redox-dependent transcription factor in mycobacterium tuberculosis. *J. Biol. Chem.*, **287**, 37703–37712.
40. Murshudov,G.N., Skubák,P., Lebedev,A.A., Pannu,N.S., Steiner,R.A., Nicholls,R.A., Winn,M.D., Long,F. and Vagin,A.A. (2011) *REFMAC5* for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D*, **67**, 355–367.
41. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Echols,N., Headd,J.J., Hung,L., Jain,S., Kapral,G.J., Kunstleve,R.W.G. *et al.* (2011) The phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94–106.
42. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of coot. *Acta Crystallogr. Sect. D*, **66**, 486–501.
43. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
44. Li,L., Li,C., Sarkar,S., Zhang,J., Witham,S., Zhang,Z., Wang,L., Smith,N., Petukh,M. and Alexov,E. (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophysics*, **5**, 9.
45. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.
46. Krissinel,E. (2015) Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.*, **43**, W314–W319.
47. Perera,I.C. and Grove,A. (2010) Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. *J. Mol. Cell Biol.*, **2**, 243–254.
48. Elcock,A.H. and McCammon,J.A. (1998) Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.*, **280**, 731–748.
49. Davis,J.R., Brown,B.L., Page,R. and Sello,J.K. (2013) Study of PcaV from streptomyces coelicolor yields new insights into ligand-responsive MarR family transcription factors. *Nucleic Acids Res.*, **41**, 3888–3900.
50. Ha,J., Capp,M.W., Hohenwalter,M.D., Baskerville,M. and Record,M.T. (1992) Thermodynamic stoichiometries of participation of water, cations and anions in specific and non-specific binding of lac repressor to DNA: Possible thermodynamic origins of the "glutamate effect" on protein-DNA interactions. *J. Mol. Biol.*, **228**, 252–264.
51. 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.
52. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.
53. Park,S.C., Kwak,Y.M., Song,W.S., Hong,M. and Yoon,S. (2017) Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. *Nucleic Acids Res.*, **45**, 13080–13093.
54. Birukou,I., Seo,S.M., Schindler,B.D., Kaatz,G.W. and Brennan,R.G. (2014) Structural mechanism of transcription regulation of the staphylococcus aureus multidrug efflux operon mepRA by the MarR family repressor MepR. *Nucleic Acids Res.*, **42**, 2774–2788.
55. Gregor,D. and Pfeifer,F. (2005) In vivo analyses of constitutive and regulated promoters in halophilic archaea. *Microbiology*, **151**, 25–33.
56. Zeng,C., Zhao,Y., Cui,C., Zhang,H., Zhu,J., Tang,X., Shen,P., Huang,Y. and Chen,X. (2009) Characterization of the haloarcula hispanica amyH gene promoter, an archaeal promoter that confers promoter activity in escherichia coli. *Gene*, **442**, 1–7.
57. Misra,V.K., Hecht,J.L., Sharp,K.A., Friedman,R.A. and Honig,B. (1994) Salt effects on protein-DNA interactions: The λcI repressor and EcoRI endonuclease. *J. Mol. Biol.*, **238**, 264–280.
58. Chakravorty,A., Jia,Z., Li,L. and Alexov,E. (2017) A new DelPhi feature for modeling electrostatic potential around proteins: Role of bound ions and implications for zeta-potential. *Langmuir*, **33**, 2283–2295.
59. Jia,Z., Li,L., Chakravorty,A. and Alexov,E. (2017) Treating ion distribution with gaussian-based smooth dielectric function in DelPhi. *J. Comput. Chem.*, **38**, 1974–1979.
60. Hasted,J.B., Ritson,D.M. and Collie,C.H. (1948) Dielectric properties of aqueous ionic solutions. parts I and II. *J. Chem. Phys.*, **16**, 1–21.
61. Senear,D.F. and Batey,R. (1991) Comparison of operator-specific and nonspecific DNA binding of the .lambda. cI repressor: KCl] and pH effects. *Biochemistry*, **30**, 6677–6688.
62. Sharp,K.A. and Honig,B. (1990) Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 301–332.
63. Zhou,T., Shen,N., Yang,L., Abe,N., Horton,J., Mann,R.S., Bussemaker,H.J., Gordan,R. and Rohs,R. (2015) Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 4654–4659.
64. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
65. Hancock,S.P., Ghane,T., Cascio,D., Rohs,R., Di Felice,R. and Johnson,R.C. (2013) Control of DNA minor groove width and fis protein binding by the purine 2-amino group. *Nucleic Acids Res.*, **41**, 6750–6760.
66. Hizver,J., Rozenberg,H., Frolow,F., Rabinovich,D. and Shakked,Z. (2001) DNA bending by an adenine–textendashthymine tract and its role in gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 8490–8495.
67. Hancock,S.P., Stella,S., Cascio,D. and Johnson,R.C. (2016) DNA sequence determinants controlling affinity, stability and shape of DNA complexes bound by the nucleoid protein fis. *PLoS ONE*, **11**, 1–24.
68. Watkins,D., Hsiao,C., Woods,K.K., Koudelka,G.B. and Williams,L.D. (2008) P22 c2 Repressor−Operator complex: Mechanisms of direct and indirect readout. *Biochemistry*, **47**, 2325–2338.
69. Pryor,E.E. Jr, Waligora,E.A., Xu,B., Dellos-Nolan,S., Wozniak,D.J. and Hollis,T. (2012) The transcription factor AmrZ utilizes multiple DNA binding modes to recognize activator and repressor sequences of pseudomonas aeruginosa virulence genes. *PLOS Pathogens*, **8**, e1002648.
70. Winter,J.A., Patoli,B. and Bunting,K.A. (2012) DNA binding in high salt: Analysing the salt dependence of replication protein A3 from the halophile haloferax volcanii. *Archaea*, **2012**, 12.
71. Lanyi,J.K. (1974) Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.*, **38**, 272–290.
72. Mevarech,M., Frolow,F. and Gloss,L.M. (2000) Halophilic enzymes: Proteins with a grain of salt. *Biophys. Chem.*, **86**, 155–164.
73. Kozlov,A.G. and Lohman,T.M. (2012) SSB binding to ssDNA using isothermal titration calorimetry. In: Keck,JL (ed). *Single-Stranded DNA Binding Proteins: Methods and Protocols*. Humana Press, Totowa, pp. 37–54.
74. Kozlov,A.G. and Lohman,T.M. (1998) Calorimetric studies of E. coli SSB protein-single-stranded DNA interactions. effects of monovalent salts on binding enthalpy. *J. Mol. Biol.*, **278**, 999–1014.
75. Bergqvist,S., Williams,M.A., O'Brien,R. and Ladbury,J.E. (2003) Halophilic adaptation of protein-DNA interactions. *Biochem. Soc. Trans.*, **31**, 677–680.
76. Aslund,F., Zheng,M., Beckwith,J. and Storz,G. (1999) Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 6161–6165.
77. Hong,M., Fuangthong,M., Helmann,J.D. and Brennan,R.G. (2005) Structure of an OhrR-ohrA operator complex reveals the DNA binding mechanism of the MarR family. *Mol. Cell*, **20**, 131–141.
78. Newberry,K.J., Fuangthong,M., Panmanee,W., Mongkolsuk,S. and Brennan,R.G. (2007) Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. *Mol. Cell*, **28**, 652–664.