

RESEARCH ARTICLE

Ehrlichia chaffeensis TRP120 nucleomodulin binds DNA with disordered tandem repeat domain

Valerie J. Klema¹*, Krishna Mohan Sepuru¹*, Nadia Füllbrunn¹, Tierra R. Farris², Paige S. Dunphy², Jere W. McBride², Krishna Rajarathnam¹, Kyung H. Choi¹*

1 Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas, United States of America, **2** Department of Pathology, Center for Biodefense and Emerging Infectious Diseases, Sealy Center for Vaccine Development, University of Texas Medical Branch, Galveston, Texas, United States of America

* These authors contributed equally to this work.

* kychoi@utmb.edu



OPEN ACCESS

Citation: Klema VJ, Sepuru KM, Füllbrunn N, Farris TR, Dunphy PS, McBride JW, et al. (2018) *Ehrlichia chaffeensis* TRP120 nucleomodulin binds DNA with disordered tandem repeat domain. PLoS ONE 13(4): e0194891. <https://doi.org/10.1371/journal.pone.0194891>

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: December 19, 2017

Accepted: March 12, 2018

Published: April 11, 2018

Copyright: © 2018 Klema et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by a Sealy and Smith Foundation grant to the Sealy Center for Structural Biology and Molecular Biophysics, a University of Texas Medical Branch SCBMB-IHII pilot grant (to KHC and JWM) and National Institute of Allergy and Infectious Diseases grant AI106859 (JWM).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Ehrlichia chaffeensis, the causative agent of human monocytotropic ehrlichiosis, secretes several effector proteins that bind host DNA to modulate host gene expression. The tandem repeat protein 120 (TRP120), one of the largest effector proteins, has four nearly identical tandem repeat (TR) regions that each consists of 80 amino acids. In addition to playing a role in ehrlichial binding and internalization, TRP120 translocates to the host nucleus where it is thought to function as a transcription factor that modulates gene expression. However, sequence analysis of TRP120 does not identify the presence of DNA-binding or trans-activation domains typical of classical eukaryotic transcription factors. Thus, the mechanism by which TRP120 binds DNA and modulates gene expression remains elusive. Herein, we expressed the TR regions of the TRP120 protein, and characterized its solution structure and ability to bind DNA. TRP120, expressed as either a one or two TR repeat, is a monomer in solution, and is mostly disordered as determined by circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy. Using NMR spectroscopy, we further show that the 1 TR construct selectively binds GC-rich DNA. Although low pH was required for TRP120 TR-DNA interaction, acidic pH alone does not induce any significant structural changes in the TR region. This suggests that TRP120 folds into an ordered structure upon forming a protein-DNA complex, and thus folding of TRP120 TR is coupled with DNA binding.

Introduction

Ehrlichia chaffeensis, the etiologic agent of human monocytotropic ehrlichiosis (HME), is an obligate intracellular bacterium and a member of the *Anaplasmataceae* family [1, 2]. Transmitted by the lone star tick (*Amblyomma americanum*), *E. chaffeensis* infects human mononuclear phagocytes where it replicates forming microcolonies in cytoplasmic membrane-bound vacuoles called

morulae [3]. HME presents symptoms ranging from mild febrile illness, malaise, and headache, to life-threatening septic shock, respiratory distress, and renal or hepatic failure in more serious cases, with over half requiring hospitalization and 3% proving fatal [4, 5]. HME patients exhibit severe manifestations despite a low bacterial burden, suggesting a role for immunopathologic mechanism in disease pathogenesis [6, 7]. However, because the symptoms of *E. chaffeensis* infection are similar to other undifferentiated febrile illnesses and tick-borne infections endemic to the same geographic areas, HME is often misdiagnosed and underreported [8].

In order to survive intracellularly, *E. chaffeensis* reprograms the host cell by activating host cell pathways such as Wnt and Notch, and by secreting several tandem repeat protein (TRP) effectors into the host cell via a type I secretion system. Ehrlichial TRP effectors interact with an array of host cell proteins in order to exploit and modulate host cell processes. *E. chaffeensis* TRP120 and TRP32 are known to translocate to the nucleus where they bind genes associated with immune regulation, post-translational modifications, transcriptional and translational regulation, signal transduction, apoptosis, and metabolism [9]. In particular, TRP32 has been shown to bind promoter regions in the host DNA to repress or activate host gene transcription [9, 10]. Although ehrlichial TRP120 and TRP32 lack a classical nuclear localization signal sequence, nuclear translocation appears to be mediated in part by post-translation dependent modifications. Nuclear translocation of TRP32 has been shown to proceed via a tyrosine phosphorylation-dependent mechanism [10]; however, the mechanism of TRP120 nuclear translocation is unknown [9, 10]. Interestingly, *Ehrlichia* TRPs exhibit similarity to the *Xanthomonas* and *Ralstonia* species transcription activator-like (TAL) effectors, which utilize similar tandem repeat DNA binding domains to activate plant cell gene transcription [11, 12].

E. chaffeensis TRP120 consists of a small N-terminal domain of unknown function, a central TR repeat region, and a C-terminal domain (Fig 1A). The C-terminal domain is sumoylated at a canonical motif and also has HECT E3 ligase activity [13]. The central region consists of four 80 amino acid TR units and a fifth shortened repeat containing 19 residues. The TR units within TRP120 are 96% identical at the amino acid level, with only three variable amino acids at positions 15, 48, and 77 (Fig 1A). *E. chaffeensis* TRP120 binds a GC-rich DNA motif that is similar to GC-rich motifs bound by eukaryotic transcription factors, suggesting that TRP120 functions similarly [9]. Two full repeats of the TRP120 TR unit are sufficient to bind a double-stranded DNA (dsDNA) and oligonucleotide probes containing GC-rich motifs [9]. However, sequence analysis of TRP120 does not identify structural motifs characteristic of typical DNA-binding proteins. Thus, the mechanism by which TRP120 recognizes DNA and functions as a transcription factor has remained elusive. Neither the minimum TRP120 sequence nor the conditions required for DNA binding are well understood. Furthermore, TRP120 contains four copies of nearly identical repeats, but their influence on the structure and function of TRP120 has not been determined. In this study, we demonstrate that TRP120 containing, either a single or two TR units, exist mostly as intrinsically disordered monomers, and that a single TR unit is sufficient to bind DNA. The interaction between the TRP120 TR unit and GC-rich DNA is pH-dependent, but pH alone does not induce structural changes in the TR unit. Folding of TRP120 TR requires both low pH and interaction with DNA, and thus TRP120 undergoes a coupled folding and binding process.

Materials and methods

Generation of TRP120-1TR construct

Construction of the glutathione S-transferase (GST)-fused TRP120 containing two TR units (TRP120-2TR) is previously described [9]. TRP120 protein containing a single TR unit (TRP120-1TR) was generated as a GST-fusion protein. To produce TRP120-1TR, the

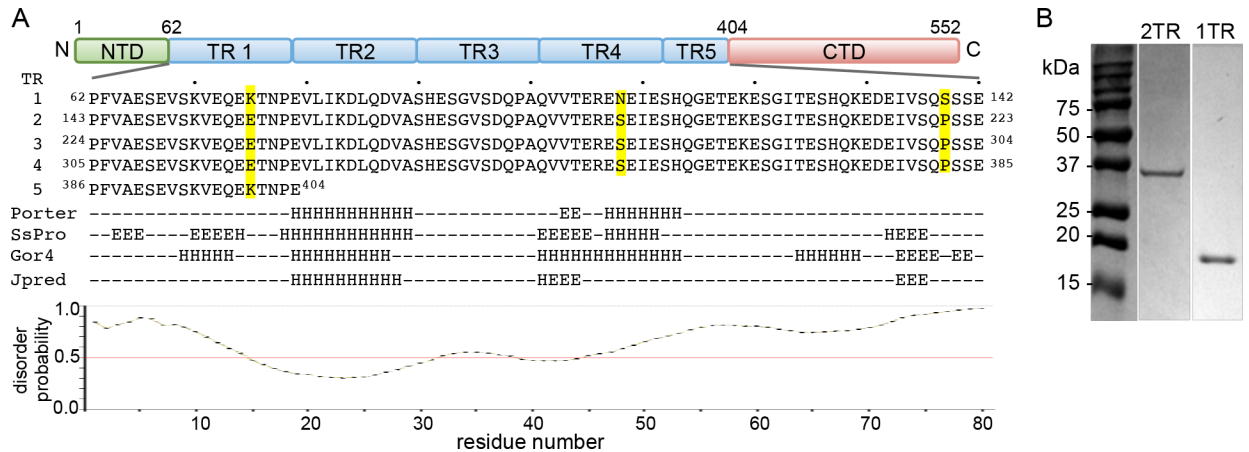


Fig 1. TR units of TRP120. (A) Domain organization of TRP120 and sequences of each TR unit. The primary sequences of TR units are shown with residues that differ between TR units highlighted in yellow (top). The sequence was analyzed by secondary structure (middle) and disorder (bottom) predictions. Predicted α -helical and β -strand regions are indicated by H and E, respectively. (B) SDS-PAGE of purified TRP120-1TR and TRP120-2TR proteins. Molecular weights for TRP120-1TR and -2TR proteins are 11.5 and 20.3 kDa, respectively, but the proteins migrate at twice their expected size.

<https://doi.org/10.1371/journal.pone.0194891.g001>

TRP120-2TR construct within the pGEX expression vector (TRP120-2TR) used in previous studies served as a template [9]. The oligonucleotide primers 5' -TTAGAACTTCACTTTC TGCAACAAATTACTCTGATGAAGATTGAGATACTATTTC-3' (forward) and 5' -GAAATA GTATCTCAATCTTCATCAGAGTAATTTGTTGCAGAAAGTGAAGTTTCTAA-3' (reverse) were used with the QuikChange II site-directed mutagenesis kit (Agilent Technologies) to isolate the second of the two TR units in the template. The sequence of the final TRP120-1TR construct was confirmed by DNA sequencing at the Molecular Genomics Core at the University of Texas Medical Branch. The final TRP120-1TR contains a single TR region and 22 residues from the thrombin cleavage site and N-terminal domain (Gln47-Glu61).

Expression and purification of TRP120 constructs

The TRP120-1TR and TRP120-2TR constructs were purified as previously described with minor changes [14, 15]. Briefly, 5 mL of overnight culture grown in Luria Broth were used to inoculate 4 L of Terrific Broth supplemented with 34 μ g/ml chloramphenicol and 100 μ g/ml ampicillin. The cells were grown at 37°C to an OD₆₀₀ of ~0.8, and protein expression was induced with 0.5 mM isopropyl- β -D-thiogalactopyranoside (IPTG) at 18°C overnight. The harvested cell pellet was sonicated in 60 mL lysis buffer (10 mM phosphate, pH 7.3, 138 mM NaCl, and 2.7 mM KCl) supplemented with 100 μ g/ml DNase A and 2 tablets of protease inhibitor cocktail (Roche Applied Science). The GST-fused TRP120 protein was bound to pre-equilibrated Glutathione Sepharose 4B resin (GE Healthcare) at 4°C overnight. The resin was washed three times in 60 mL wash buffer (20 mM Tris pH 7.0, 300 mM NaCl, and 1 mM DTT). The TRP120 protein was cleaved from the resin-bound GST tag by incubation with thrombin (1U/mg protein) from bovine pancreas (Sigma) at room temperature for 2 h. The eluted protein was buffer exchanged depending on the relevant experiment. Because the TRP120-1TR and -2TR constructs do not contain any Trp residues, the protein concentration was measured using the Bradford assay with lysozyme (MP Biomedicals) as the protein standard.

To express ¹⁵N-labeled TRP120-1TR, cells from 1 mL of overnight culture in Luria broth were harvested and resuspended in 50 mL M9 medium supplemented with 1 mg/L ¹⁵NH₄Cl

(Cambridge Isotope Laboratories), 34 µg/mL chloramphenicol and 100 µg/mL ampicillin. This intermediate culture was grown at 37°C for ~6 h before inoculation of 1 L M9 medium supplemented with 1 mg/L ¹⁵NH₄Cl, 34 µg/mL chloramphenicol and 100 µg/mL ampicillin. The expression and purification of ¹⁵N-labeled TRP120-1TR using affinity chromatography proceeded as described above.

Sedimentation velocity analytical ultracentrifugation

TRP120-1TR and TRP120-2TR constructs were subjected to sedimentation velocity analytical centrifugation using a Beckman Coulter XL-A analytical ultracentrifuge. To obtain an absorbance reading ~0.7 at a wavelength of 224 nm, proteins were diluted in 20 mM Tris (pH 7.0), 300 mM NaCl, and 1 mM DTT to a concentration of 27 µM. Protein sedimentation was carried out at 50,000 rpm at a temperature of 20°C and 350 absorbance scans were collected. Data were analyzed using the program SEDFIT [16]. During data processing, the size distribution model “*c(s)* continuous” was selected and the frictional ratio was allowed to float.

Sequence analysis. The sequence of TRP120 was analyzed by the ExPASy bioinformatics resource portal [17]. The program ProtParam [18] was used to determine the molecular weights and isoelectric points. The secondary structure prediction programs Jpred [19], SsPro [20], Porter [21], and Gor4 [22] were used to predict the secondary structure content of a single 80 amino acid TR unit. Disordered region prediction programs PrDOS [23], IUPred [24], DisEMBL [25], PredictProtein [26] were used to predict the disordered regions in TRP120.

Circular dichroism spectroscopy of TRP120 constructs

Far-UV circular dichroism (CD) spectroscopy was used to determine the secondary structure content of the TRP120-1TR and TRP120-2TR constructs. CD data were collected using 9 µM TRP120-1TR in 20 mM Tris buffer (pH 7.0) containing 100 mM NaCl. To monitor changes in TRP120-2TR secondary structure in different pH buffers, TRP120-2TR in 10 mM phosphate (pH 7.0) and 100 mM NaCl was buffer exchanged into 10 mM phosphate (pH 6.0) and 100 mM NaCl, or 10 mM citrate (pH 4.5) and 100 mM NaCl. CD spectra were measured at 20°C from 195 nm to 250 nm by a Jasco J-815 spectrometer using a data integration time of 8 sec and were averaged over 3 replicates. Data were converted from millidegrees to molar residue ellipticity (MRE) using the formula:

$$[\theta] = (\theta \times 10^6) / (c \times l \times n)$$

where $[\theta]$ is the molar residue ellipticity, θ is the CD signal in millidegrees, c is the protein concentration in µM, l is the pathlength in mm, and n is the number of residues. The CD data analysis program BestSel was used to analyze the data and estimate the secondary structure [27].

2D NMR spectroscopy of TRP120-1TR

Complementary DNA oligonucleotides containing the TRP120 GC-rich DNA binding motif 5′-GCGATTCTCCTGCCTCAGCCTCCCTAG-3′ and 5′-CTAGGGAGGCTGAGGCAGGAGAATCGC-3′ were synthesized by Integrated DNA Technologies (Coralville, IA). As a negative control, complementary AT-rich oligonucleotides, 5′-GTCACAAATTCTCATCTTTATATAAAGAT-3′ and 5′-ATCTTTATATAAAGATGAGAATTTGTGAC-3′ were also synthesized. The DNA probes were mixed in a 1:1 ratio before being denatured at 95°C for 10 minutes and annealed by slow cooling to room temperature. HSQC spectra of ¹⁵N-labeled TRP120-1TR were collected using the protein prepared at pH 7.0 (20 mM Tris), pH 6.5 (10 mM phosphate), pH 6.0 (10 mM phosphate), and pH 5.5 (10 mM phosphate) containing 100 mM NaCl. NMR

titrations of DNA to ^{15}N -TRP120-1TR were carried out in 20 mM Tris (pH 7.0) or 10 mM sodium phosphate (pH 5.5) containing 100 mM NaCl, 1 mM DSS (2,2-dimethyl-2 silapentane-sulfonic acid), 1 mM sodium azide, and 10% D_2O (v/v). NMR spectra were acquired at 25°C on a Bruker Avance III 800 MHz (equipped with a TXI cryoprobe) or 600 MHz (with QCI probe) spectrometers. Aliquots of DNA (~10 mM) prepared in the same buffer were added to ~60 μM TRP120-1TR at a 1, 5 or 10-fold molar excess, and a series of ^1H - ^{15}N HSQC spectra was collected.

Results

TRP120-1TR and TRP120-2TR exist as monomers in solution

TRP120 protein consists of an N-terminal domain, 4.3 TR units, and a C-terminal domain (Fig 1A). The 4.3 repeats of the TR unit (339 amino acids) form the majority of the protein sequence. At this time, little is known regarding the structural properties of the TR units, whether the TR units bind individually or if crosstalk between the domains is necessary for folding and DNA binding. To address these questions, we characterized two TRP120 TR constructs—one containing a single TR unit (80 residues, TRP120-1TR) and the second containing two TR units (TRP120-2TR). TRP120-1TR and TRP120-2TR were expressed in *E. coli* as GST-tagged fusion proteins and purified via affinity chromatography to a purity >95% as judged by SDS-PAGE (Fig 1B). Full-length TRP120 is a 60 kDa protein, but migrates as a ~120 kDa protein on SDS-PAGE, twice its predicted molecular weight. Similarly, although the predicted molecular weights of TRP120-1TR and -2TR based on their primary sequences are 11.5 kDa and 20.3 kDa, respectively, the constructs migrate with apparent molecular weights of ~19 kDa and ~36 kDa, respectively (Fig 1B). To determine if TR regions form a dimer that is stable during SDS-PAGE, the oligomeric states of the TRP120 constructs were examined by sedimentation velocity analytical ultracentrifugation. TRP120-1TR and -2TR constructs exhibited sedimentation coefficients of 1.1S and 1.4S, corresponding to molecular weights of 12.3 kDa and 22.3 kDa, respectively (Fig 2). Thus, both proteins are monomeric in solution. The frictional ratios determined during data processing were 2.0 and 2.2 for TRP120-1TR and TRP120-2TR, respectively, indicating a significantly elongated or asymmetric shape.

The anomalous migration of TRP120 in SDS-PAGE was determined to be related to the highly acidic pI of the protein [28]. TRP120 has an isoelectric point (pI) of 4.1 for the full-length protein, and 3.7, 4.3, and 4.1 for the individual N-terminal, central TR and C-terminal domains, respectively. We found that the TRP120-1TR and -2TR proteins expressed in *E. coli* also migrate to higher molecular weights than expected. The analytical ultracentrifugation clearly shows that both TRP120-1TR and -2TR proteins are monomers with expected molecular weights. Thus, aberrant migration of TRP120 TR units on SDS-PAGE is intrinsic to the nature of the domain. The single TRP120 TR unit contains 16–17 Glu and 4 Asp residues out of 80 amino acids (Fig 1A), and thus the highly acidic TRP120 TR proteins resist SDS binding and migrate as a higher molecular weight than expected [29] [30].

TRP120-1TR and TRP120-2TR proteins are mostly disordered in solution

Secondary structure prediction programs [19, 20, 22, 31] predict that the TRP120 TR region of 80 amino acids consists of mostly α -helical structure (Fig 1A). However, disorder prediction programs suggest that TR region may be disordered, consistent with the elongated shapes of the TRP120 TR constructs indicated by analytical ultracentrifugation (Fig 1A). We thus determined the secondary structure content of TRP120-1TR and -2TR by CD. The far-UV CD spectra of TRP120-1TR and TRP120-2TR at pH 7.0 showed similar profiles, with a minimum centered at 200 nm, indicating that TRP120 TR units are mostly intrinsically disordered (Fig 3). A shallow negative feature centered at 220 nm suggests a small percentage of β -turn/ β -sheet

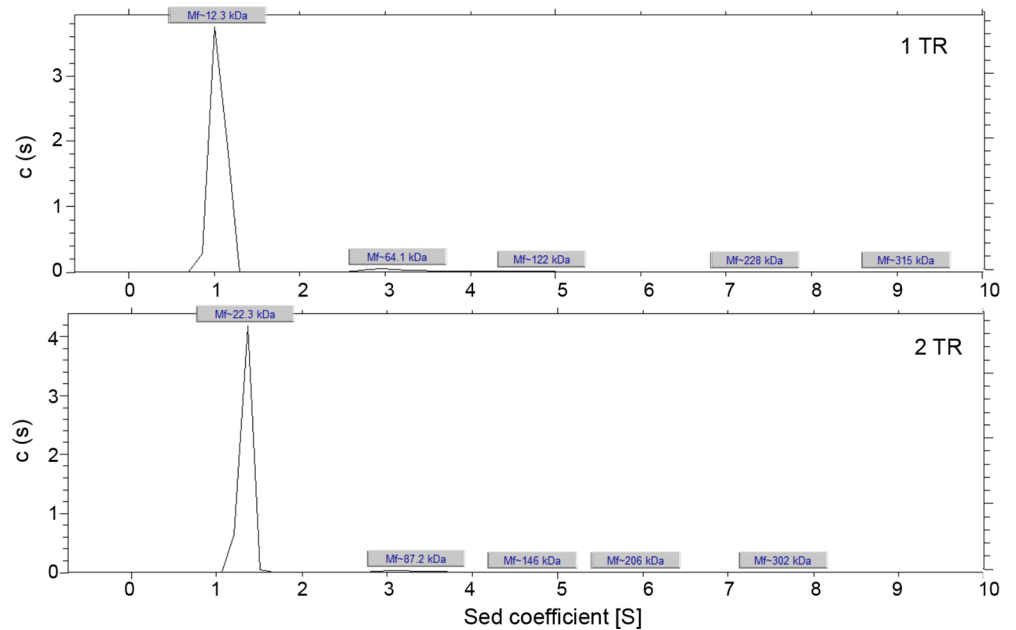


Fig 2. Sedimentation velocity profiles of TRP120-1TR and TRP120-2TR proteins. The molecular weights of TRP120-1TR and -2TR proteins were determined by sedimentation velocity experiments. Sedimentation coefficient distribution $c(s)$ profile for TRP120-1TR and TRP120-2TR show one major species, which corresponds to the molecular weight of 12.3 or 22.2 kDa, respectively. The observed molecular weights indicate that both TRP120 constructs are monomers in solution.

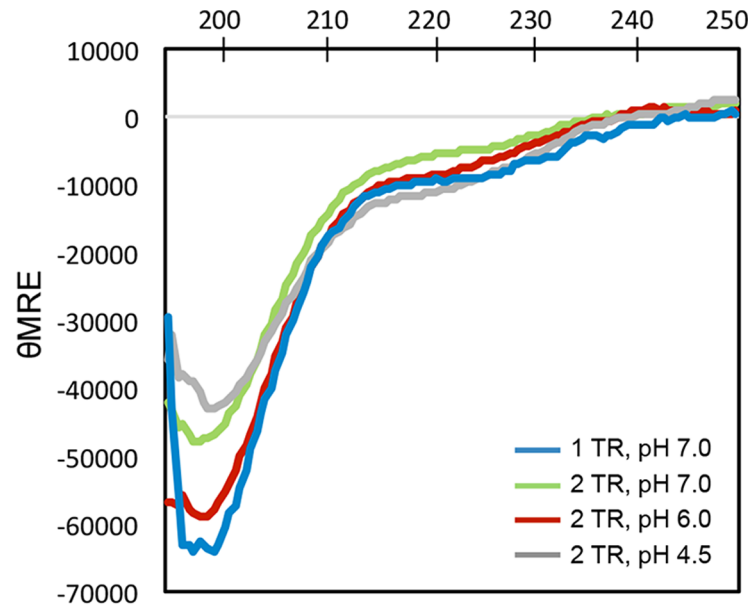
<https://doi.org/10.1371/journal.pone.0194891.g002>

secondary structure. The program BestSel was used to analyze the CD data, and the program estimated ~50% random coil and 31~36% antiparallel β -strand content for both proteins (Fig 3). In contrast to the secondary structure prediction, no α -helices were observed. Additionally, we measured the CD spectra of TRP120-2TR at lower pH ranges, because the pH of ehrlichial morulae where TRP120 is expressed is slightly acidic due to functional vacuolar H^+ ATPase that acidifies the compartments [32]. CD spectra of TRP120-2TR measured as a function of pH (7.0, 6.0 and 4.5) showed similar profiles except for small changes in the β -strand content (Fig 3). However, these changes are modest and the TRP120-1TR and -2TR proteins are mostly intrinsically disordered.

TRP120 tandem repeat binds GC-rich DNA

TRP120 is known to modulate gene expression in the host nucleus, and to bind 20–30 nt dsDNA containing a GC-rich region (Table 1) [9]. We characterized the binding of ^{15}N -labeled TRP120-1TR to dsDNA to determine if the TRP120 TR region can bind DNA and whether there is a structural change in the TR region upon DNA binding. We first collected 1H - ^{15}N NMR HSQC spectra of TRP120-1TR as a function of pH (5.5, 6.0, 6.5 and 7.0), and observed that the protein is essentially unstructured in solution at all pHs tested (Fig 4), consistent with the CD data. The NMR spectrum shows around 80 resonances between 6.7 and 8.6 ppm (Fig 4). Since the protein consists of 102 residues, most of the residues are accounted for. However, most of the chemical shifts correspond to that of a random coil and are not significantly downfield shifted (characteristic of β -sheet) or upfield shifted (characteristic of α -helix), and thus NMR data are consistent with a natively unfolded protein.

Interestingly, when we carried out binding measurements with 27 nt GC-rich DNA at pH 5.5 and 7.0, TRP120-1TR binds DNA at pH 5.5 but not at 7.0. At pH 7.0, the addition of DNA



Secondary structure (%)	1 TR	2 TR		
	pH 7.0	pH 7.0	pH 6.0	pH 4.5
α helix	0	0	0	2
antiparallel β strand	36	31	33	25
parallel β strand	0	0	0	0
β-turn	14	18	18	20
other	51	51	49	53

Fig 3. Far-UV CD spectra of TRP120-1TR and TRP120-2TR. The CD spectrum of TRP120-1TR was measured in Tris pH 7.0 containing 300 mM NaCl (blue). The CD spectra of TRP120-2TR were measured at three different pHs—phosphate pH 7.0 (green), phosphate pH 6.0 (red), and citrate pH 4.5 (gray) containing 100 mM NaCl. Secondary structure content of TRP120 constructs was estimated by BestSel analysis (bottom).

<https://doi.org/10.1371/journal.pone.0194891.g003>

to TRP120-1TR up to 10-fold molar equivalent did not show significant chemical shifts, indicating minimal binding and no induction of secondary structure formation (Fig 5A). However, addition of DNA to TRP120-1TR at pH 5.5 resulted in several new peaks with significant differences in chemical shifts compared to the parent peaks (Fig 5B). The presence of new downfield shifted peaks (>8.5 ppm) suggests an increase in β-strand content. Furthermore, the peaks corresponding to several side chain NH₂ of Asn/Gln residues were resolved in the DNA-bound spectrum, indicating that these residues become structured upon binding to

Table 1. DNA probes used in NMR titrations.

DNA probe	Sequence
GC-rich*	5' -GCGATTCTCCTGCTCAGCCTCCCTAG-3'
	3' -CGCTAAGAGGACGGAGTCGGAGGGATC-5'
AT-rich	5' -GTCACAAATTCTCATCTTTATATAAAGAT-3'
	3' -CAGTGGGAAGAGTAGAAATATATTTCTA-5'

* The GC-rich TRP120-binding motifs are underlined.

<https://doi.org/10.1371/journal.pone.0194891.t001>

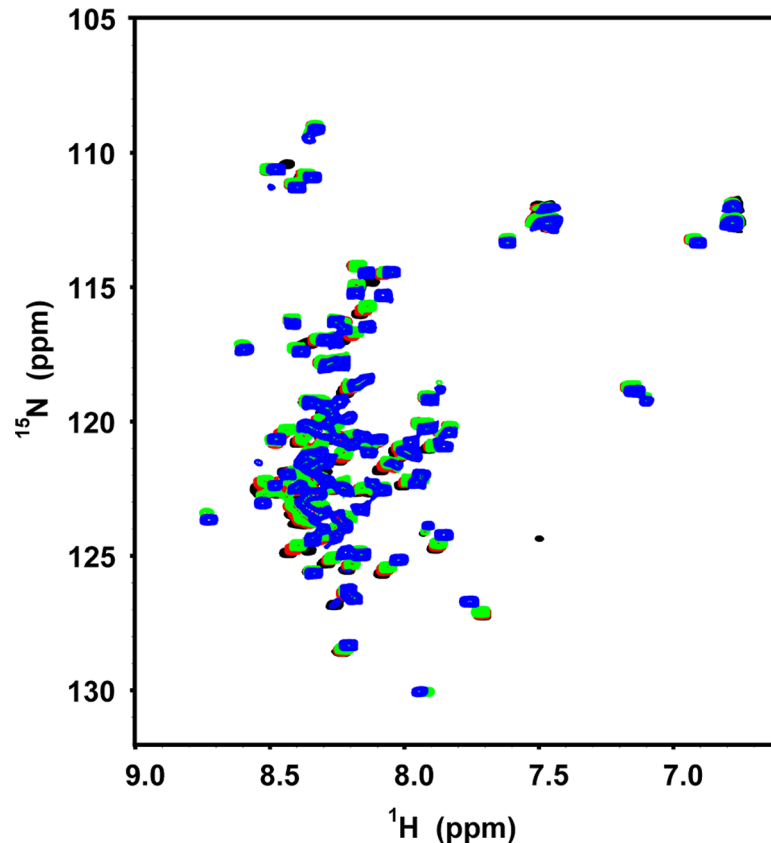


Fig 4. NMR spectra of TRP120-1TR at different pHs. HSQC spectra of ^{15}N -labeled TRP120-1TR were collected at four different pHs—Tris pH 7.0 (black), phosphate pH 6.5 (red), phosphate pH 6.0 (green), and phosphate pH 5.5 (blue) containing 100 mM NaCl.

<https://doi.org/10.1371/journal.pone.0194891.g004>

DNA. Overall, the significant chemical shift perturbation (CSP) and large dispersion in chemical shifts in the DNA-bound spectrum compared to the free TRP120-1TR indicate that the protein adopts a specific structure and conformation upon DNA interaction. In contrast, when the AT-rich oligonucleotide was used in TRP120-1TR binding, there were no significant chemical shifts at pH 7.0, and only a few minor chemical shifts at pH 5.5 (Fig 5C and 5D). Thus, the interaction between TRP120-1TR and the GC-rich oligonucleotide is specific, which would be necessary for observed transcription activities [9].

Discussion

E. chaffeensis TRP effectors are involved in the activation of host cell pathways, manipulation of gene transcription, and exploitation of post-translational modifications to reprogram the host cell. TRP120 and TRP32 translocate through the host nuclear membrane via mechanisms that do not involve a classic nuclear localization sequence. A previous study reported that TRP120 binds a GC-rich motif within the host DNA and alters gene expression [9]. However, many of the molecular and structural details of TRP120 DNA binding are unclear, including characterization of structural features, the minimum TRP120 sequence necessary for structural integrity, and conditions necessary for DNA binding. In this study, we demonstrate by CD and NMR spectroscopy that TRP120 TR regions are intrinsically disordered in the absence of DNA. However, in the presence of DNA, a single TR unit of TRP120 binds selectively to GC-rich DNA

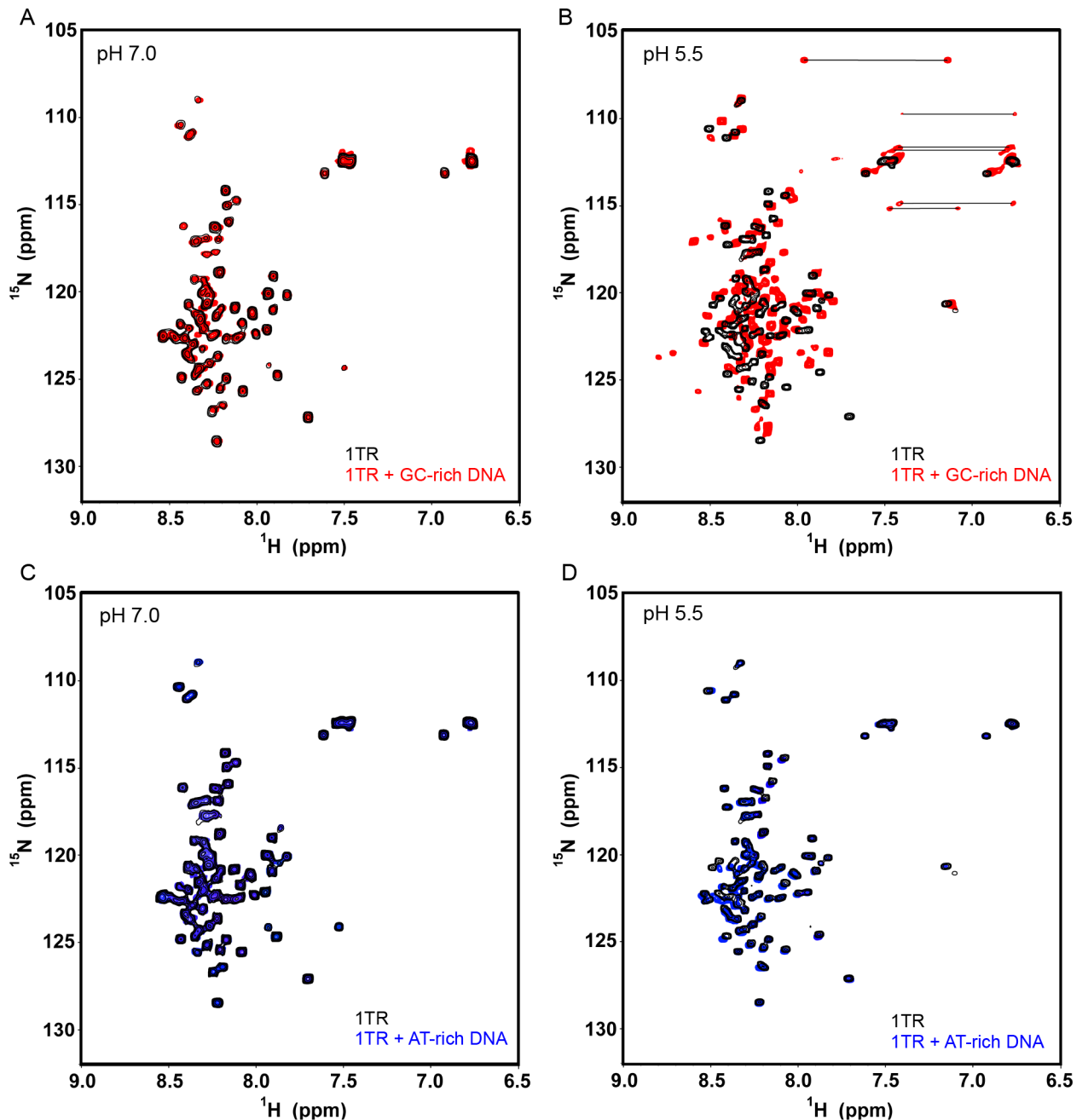


Fig 5. TRP120-1TR selectively binds to GC-rich DNA. (A-B) HSQC spectra of ^{15}N -labeled TRP120-1TR with GC-rich DNA at pH 7.0 or pH 5.5. TRP120-1TR (68 μM) alone (black) and with a 10-fold molar equivalent of unlabeled DNA (red, Table 1) in 20 mM Tris pH 7.0 containing 100 mM NaCl or in 10 mM phosphate pH 5.5 containing 100 mM NaCl are shown. The lines highlight Asn or Gln side chain NH_2 peaks that show distinctly different chemical shifts in the bound form. (C-D) HSQC spectra of ^{15}N -labeled TRP120-1TR with AT-rich DNA probe at pH 7.0 or pH 5.5. TRP120-1TR (30 μM) alone (black) and with a 10-fold molar equivalent of unlabeled DNA (blue, Table 1) in 20 mM Tris pH 7.0 containing 100 mM NaCl or in 10 mM phosphate pH 5.5 containing 100 mM NaCl are shown.

<https://doi.org/10.1371/journal.pone.0194891.g005>

over AT-rich DNA (Fig 5). Additionally, we show that DNA binding by TRP120 TR is pH-dependent, with virtually no binding at pH 7.0 but binding at pH 5.5. Notably, pH change alone does not trigger any structural changes in TRP120, since the secondary structure content of the protein does not change as a function of pH (Figs 3 and 4).

Intrinsically disordered proteins have been shown to be involved in regulation of transcription and translation, cellular signaling, protein phosphorylation, and self-assembly of large protein complexes [33]. TRP120 shares several features with intrinsically disordered proteins. First, intrinsically disordered proteins have characteristic sequences, typically containing a low proportion of order promoting amino acids (Cys and bulky hydrophobic amino acids) and a high proportion of particular polar and charged amino acids (Gln, Ser, Pro, and Glu) [33]. The TR region of TRP120 does not contain Cys, Met, Tyr, or Trp residues, and contains 17 Glu, 12 Ser, and 7 Gln residues, consistent with its disordered nature (Fig 1). The intrinsic plasticity typical of disordered proteins may enable TRP120 to interact with diverse host targets [34]. TRP120 interacts with an array of host proteins such as immunoglobulin light-chain lambda and polycarbonyl group ring finger 5, involved in immune and cellular signaling, metabolism, and protein synthesis/degradation to control over the immune response and facilitate infection [34]. Second, intrinsically disordered proteins often undergo a coupled folding and binding process, i.e., folding into an ordered structure upon complex formation with their targets [35]. Our data suggest that the conformational change in TRP120-1TR is coupled with DNA-binding, since low pH alone does not induce any structural changes in the TR regions of TRP120 (Figs 4 and 5). Third, intrinsically disordered protein regions generally participate in highly specific but low affinity interactions, which are required due to the transient nature of interactions occurring during regulation of transcription or cellular signaling [36]. Our NMR experiments suggest that TRP120-1TR binds in the micromolar range, much weaker than the low nanomolar dissociation constants (K_d) measured for zinc finger transcription factors [37]. Since TRP120 contains multiple TR units, determining whether the additional TR units increase the specificity of DNA targets or DNA binding affinity would require further examination.

Ehrlichial TRPs share several qualities with another group of pathogen effector proteins called TAL effectors, which are secreted by *Xanthomonas* species, and utilize similar transcriptional reprogramming strategies to hijack the host cell. Like the TRPs secreted by *E. chaffeensis*, TAL effectors translocate to the host cell nucleus and alter gene regulation to overcome host defenses [11, 38]. The TAL effectors also contain an N-terminal domain, central TR domain, and C-terminal domain. The TRs of ~34 residues contained within TAL effectors are nearly identical, with two varying residues conferring DNA binding specificity for each repeat [39]. Despite the similar domain organization and functional similarities between the TRPs secreted by *E. chaffeensis* and the TAL effectors of *Xanthomonas*, significant structural differences exist. Crystal structures of the TR regions from the TAL effectors dHax3 and PthXo1 bound to a DNA target showed that each repeat forms two short α -helices with a flanking loop, and multiple TR units form a right-handed superhelical assembly that wraps around the DNA helix [40, 41]. Notably, in the absence of a DNA target, TAL effectors retain their superhelical assembly [40]. This differs significantly from TRP120, which exists as an intrinsically disordered protein until a coupled DNA binding/folding process occurs.

TRP120 is also different from many DNA-binding proteins that contain a basic surface-exposed patch to complement their negatively charged cognate DNA at the site of interaction [42]. Although each TR unit in TRP120 contains 1 Arg and 5 Lys residues, the sequence of TRP120 is highly acidic overall with a predicted pI value of 4.3 for a single TR unit. TRP120-1TR binds DNA at pH 5.5, but not at pH 7.0, suggesting that acidic pH is required for interaction with DNA. The low pH would decrease the net negative charge on the protein, facilitating interaction with negatively charged DNA. Additionally, because the imidazole side chain has a pKa value within this range (~6.5), His residues are particularly sensitive to pH changes around physiological pH. In some DNA-binding proteins, His residues are known to act as a pH-sensitive trigger, thus this may be relevant because TRP120-1TR contains three His residues (Fig 1A). Such changes in the protonation state of critical His residues can enhance DNA binding affinity

without a change in the overall structure of the protein or secondary structure composition. For example, forkhead box P2 protein (FOXP2) is a transcription factor in which pH changes (acidification) increased DNA binding affinity without a change in the secondary structure of the protein in the presence or absence of DNA [43]. The DNA-binding activity can also be enhanced by deprotonation of a critical His residue. The transcription factor early growth response protein 1 (EGR1) undergoes an enhancement in its DNA-binding activity by more than an order of magnitude when a conserved His residue is deprotonated as the pH is raised from 5 to 8 [44]. Alternatively, TRP120 may recruit host or bacterial proteins for its transcriptional activator function to stabilize what would otherwise be a repulsive interaction between acidic TRP120 and DNA. Additional studies on the role of TR units in the context of the whole protein and crosstalk between the TR units and between TR units and the N-terminal and C-terminal domains are necessary to better understand the novel role of TRP120 in the pathology of *E. chaffeensis* infection.

Acknowledgments

This work was supported by a University of Texas Medical Branch SCBMB-IHII pilot grant (to KHC and JWM), National Institute of Allergy and Infectious Diseases grant AI106859 (JWM), and Sealy and Smith Foundation grant to the Sealy Center for Structural Biology and Molecular Biophysics. We thank Dr. Luis Holthausen for help with AUC experiments.

Author Contributions

Conceptualization: Valerie J. Klema, Jere W. McBride, Krishna Rajarathnam, Kyung H. Choi.

Formal analysis: Valerie J. Klema, Krishna Mohan Sepuru, Nadia Füllbrunn, Kyung H. Choi.

Investigation: Valerie J. Klema, Krishna Mohan Sepuru, Nadia Füllbrunn, Tierra R. Farris.

Resources: Paige S. Dunphy.

Writing – original draft: Valerie J. Klema, Krishna Rajarathnam, Kyung H. Choi.

Writing – review & editing: Valerie J. Klema, Jere W. McBride, Krishna Rajarathnam, Kyung H. Choi.

References

1. Paddock CD, Childs JE. Ehrlichia chaffeensis: a prototypical emerging pathogen. Clin Microbiol Rev. 2003; 16(1):37–64. <https://doi.org/10.1128/CMR.16.1.37-64.2003> PMID: 12525424
2. Anderson BE, Sumner JW, Dawson JE, Tzianabos T, Greene CR, Olson JG, et al. Detection of the etiologic agent of human ehrlichiosis by polymerase chain reaction. J Clin Microbiol. 1992; 30(4):775–80. PMID: 1374076
3. Anderson BE, Sims KG, Olson JG, Childs JE, Piesman JF, Happ CM, et al. Amblyomma americanum: a potential vector of human ehrlichiosis. Am J Trop Med Hyg. 1993; 49(2):239–44. PMID: 8357086
4. Ismail N, Bloch KC, McBride JW. Human ehrlichiosis and anaplasmosis. Clinics in laboratory medicine. 2010; 30(1):261–92. <https://doi.org/10.1016/j.cll.2009.10.004> PMID: 20513551
5. Nichols Heitman K, Dahlgren FS, Drexler NA, Massung RF, Behravesh CB. Increasing Incidence of Ehrlichiosis in the United States: A Summary of National Surveillance of Ehrlichia chaffeensis and Ehrlichia ewingii Infections in the United States, 2008–2012. Am J Trop Med Hyg. 2016; 94(1):52–60. <https://doi.org/10.4269/ajtmh.15-0540> PMID: 26621561
6. Dunphy PS, Luo T, McBride JW. Ehrlichia moonlighting effectors and interkingdom interactions with the mononuclear phagocyte. Microbes Infect. 2013; 15(14–15):1005–16. <https://doi.org/10.1016/j.micinf.2013.09.011> PMID: 24141087
7. Lina TT, Dunphy PS, Luo T, McBride JW. Ehrlichia chaffeensis TRP120 Activates Canonical Notch Signaling To Downregulate TLR2/4 Expression and Promote Intracellular Survival. mBio. 2016; 7(4): e00672–16. <https://doi.org/10.1128/mBio.00672-16> PMID: 27381289

8. Olano JP, Masters E, Hogrefe W, Walker DH. Human monocytotropic ehrlichiosis, Missouri. *Emerg Infect Dis.* 2003; 9(12):1579–86. <https://doi.org/10.3201/eid0912.020733> PMID: 14720399
9. Zhu B, Kuriakose JA, Luo T, Ballesteros E, Gupta S, Fofanov Y, et al. Ehrlichia chaffeensis TRP120 binds a G+C-rich motif in host cell DNA and exhibits eukaryotic transcriptional activator function. *Infection and immunity.* 2011; 79(11):4370–81. <https://doi.org/10.1128/IAI.05422-11> PMID: 21859854
10. Farris TR, Dunphy PS, Zhu B, Kibler CE, McBride JW. Ehrlichia chaffeensis TRP32 is a Nucleomodulin that Directly Regulates Expression of Host Genes Governing Differentiation and Proliferation. *Infect Immun.* 2016; 84(11):3182–3194.
11. Doyle EL, Stoddard BL, Voytas DF, Bogdanove AJ. TAL effectors: highly adaptable phyto-bacterial virulence factors and readily engineered DNA-targeting proteins. *Trends in cell biology.* 2013; 23(8):390–8. <https://doi.org/10.1016/j.tcb.2013.04.003> PMID: 23707478
12. Mak AN, Bradley P, Bogdanove AJ, Stoddard BL. TAL effectors: function, structure, engineering and applications. *Current opinion in structural biology.* 2013; 23(1):93–9. <https://doi.org/10.1016/j.sbi.2012.11.001> PMID: 23265998
13. Zhu B, Das S, Mitra S, Farris TR, McBride JW. Ehrlichia chaffeensis TRP120 Moonlights as a HECT E3 Ligase Involved in Self- and Host Ubiquitination To Influence Protein Interactions and Stability for Intracellular Survival. *Infect Immun.* 2017; 85(9):e00290–17. <https://doi.org/10.1128/IAI.00290-17> PMID: 28630068
14. Yu XJ, McBride JW, Diaz CM, Walker DH. Molecular cloning and characterization of the 120-kilodalton protein gene of Ehrlichia canis and application of the recombinant 120-kilodalton protein for serodiagnosis of canine ehrlichiosis. *J Clin Microbiol.* 2000; 38(1):369–74. PMID: 10618118
15. Yu XJ, Crocquet-Valdes P, Cullman LC, Walker DH. The recombinant 120-kilodalton protein of Ehrlichia chaffeensis, a potential diagnostic tool. *J Clin Microbiol.* 1996; 34(11):2853–5. PMID: 8897200
16. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J.* 2000; 78(3):1606–19. [https://doi.org/10.1016/S0006-3495\(00\)76713-0](https://doi.org/10.1016/S0006-3495(00)76713-0) PMID: 10692345
17. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research.* 2003; 31(13):3784–8. PMID: 12824418
18. Gasteiger E, Gattiker A, Duvaud S, Wilkins M.R, Appel R.D, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: Waler JM, editor. *The proteomics Protocols Handbook: Humana Press; 2005.* p. 571–607.
19. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* 2015; 43(W1):W389–94. <https://doi.org/10.1093/nar/gkv332> PMID: 25883141
20. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.* 2005; 33(Web Server issue):W72–6. <https://doi.org/10.1093/nar/gki396> PMID: 15980571
21. Pollastri G, McLysaght A. Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics.* 2005; 21(8):1719–20. <https://doi.org/10.1093/bioinformatics/bti203> PMID: 15585524
22. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* 1996; 266:540–53. PMID: 8743705
23. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic acids research.* 2007; 35(Web Server issue):W460–4. <https://doi.org/10.1093/nar/gkm363> PMID: 17567614
24. Dosztanyi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005; 21(16):3433–4. <https://doi.org/10.1093/bioinformatics/bti541> PMID: 15955779
25. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure.* 2003; 11(11):1453–9. PMID: 14604535
26. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic acids research.* 2014; 42(Web Server issue):W337–43. <https://doi.org/10.1093/nar/gku366> PMID: 24799431
27. Micsonai A, Wien F, Keryna L, Lee YH, Goto Y, Refregiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A.* 2015; 112(24):E3095–103. <https://doi.org/10.1073/pnas.1500851112> PMID: 26038575
28. Wakeel A, Zhang X, McBride JW. Mass spectrometric analysis of Ehrlichia chaffeensis tandem repeat proteins reveals evidence of phosphorylation and absence of glycosylation. *PLoS one.* 2010; 5(3):e9552. <https://doi.org/10.1371/journal.pone.0009552> PMID: 20209062

29. Armstrong DJ, Roman A. The anomalous electrophoretic behavior of the human papillomavirus type 16 E7 protein is due to the high content of acidic amino acid residues. *Biochemical and biophysical research communications*. 1993; 192(3):1380–7. <https://doi.org/10.1006/bbrc.1993.1569> PMID: 8389548
30. Tompa P. Intrinsically unstructured proteins. *Trends in biochemical sciences*. 2002; 27(10):527–33. PMID: 12368089
31. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999; 292(2):195–202. <https://doi.org/10.1006/jmbi.1999.3091> PMID: 10493868
32. Barnewall RE, Rikihisa Y, Lee EH. Ehrlichia chaffeensis inclusions are early endosomes which selectively accumulate transferrin receptor. *Infect Immun*. 1997; 65(4):1455–61. PMID: 9119487
33. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology*. 2005; 6(3):197–208. <https://doi.org/10.1038/nrm1589> PMID: 15738986
34. Luo T, Kuriakose JA, Zhu B, Wakeel A, McBride JW. Ehrlichia chaffeensis TRP120 interacts with a diverse array of eukaryotic proteins involved in transcription, signaling, and cytoskeleton organization. *Infection and immunity*. 2011; 79(11):4382–91. <https://doi.org/10.1128/IAI.05608-11> PMID: 21859857
35. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Current opinion in structural biology*. 2002; 12(1):54–60. PMID: 11839490
36. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. *Biochemistry*. 2006; 45(22):6873–88. <https://doi.org/10.1021/bi0602718> PMID: 16734424
37. Kriwacki RW, Schultz SC, Steitz TA, Caradonna JP. Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proceedings of the National Academy of Sciences of the United States of America*. 1992; 89(20):9759–63. PMID: 1329106
38. Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology*. 2010; 48:419–36. <https://doi.org/10.1146/annurev-phyto-080508-081936> PMID: 19400638
39. Mak ANS, Bradley P, Bogdanove AJ, Stoddard BL. TAL effectors: function, structure, engineering and applications. *Curr Opin Struct Biol*. 2013; 23(1):93–9. <https://doi.org/10.1016/j.sbi.2012.11.001> PMID: 23265998
40. Deng D, Yan C, Pan X, Mahfouz M, Wang J, Zhu JK, et al. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*. 2012; 335(6069):720–3. <https://doi.org/10.1126/science.1215670> PMID: 22223738
41. Mak AN, Bradley P, Cernadas RA, Bogdanove AJ, Stoddard BL. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*. 2012; 335(6069):716–9. <https://doi.org/10.1126/science.1216211> PMID: 22223736
42. Szilagyi A, Skolnick J. Efficient prediction of nucleic acid binding function from low-resolution protein structures. *J Mol Biol*. 2006; 358(3):922–33. <https://doi.org/10.1016/j.jmb.2006.02.053> PMID: 16551468
43. Blane A, Fanucchi S. Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry*. 2015; 54(25):4001–7. <https://doi.org/10.1021/acs.biochem.5b00155> PMID: 26055196
44. Mikles DC, Bhat V, Schuchardt BJ, Deegan BJ, Seldeen KL, McDonald CB, et al. pH modulates the binding of early growth response protein 1 transcription factor to DNA. *FEBS J*. 2013; 280(15):3669–84. <https://doi.org/10.1111/febs.12360> PMID: 23718776