JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

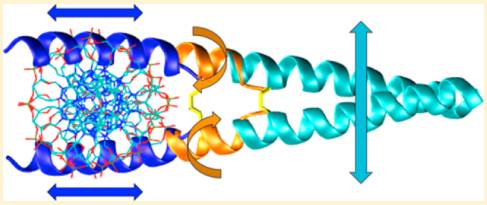
Truncated Variants of the GCN4 Transcription Activator Protein Bind DNA with Dramatically Different Dynamical Motifs

Danielle M. McHarris and Daniel A. Barr*

Department of Chemistry and Biochemistry, Utica College, 1600 Burrstone Road, Utica, New York 13502, United States

Supporting Information

ABSTRACT: The yeast protein GCN4 is a transcriptional activator in the basic leucine zipper (bZip) family, whose distinguishing feature is the "chopstick-like" homodimer of alpha helices formed at the DNA-binding interface. While experiments have shown that truncated versions of the protein retain biologically relevant DNA-binding affinity, we present the results of a computational study revealing that these variants show a wide variety of dynamical modes in their interaction with the target DNA sequence. We have performed all-atom molecular dynamics simulations of the full-length GCN4 protein as well as three truncated variants; our data indicate that the truncated mutants show dramatically different correlation patterns. We conclude that although the truncated mutants still retain DNA-binding ability, the bZip interface present in the full-length protein provides important stability for the protein−DNA complex.



INTRODUCTION

Sequence-specific DNA-binding proteins employ a wide variety of structural and thermodynamic strategies for forming tightly bound, site-specific complexes.[1−4] The sequence-specific binding of proteins to DNA results in the release of water and ions from the protein−DNA interface (a highly favorable entropic contribution to the binding free energy) as well as the formation of new electrostatic contacts between the protein and the DNA (a highly favorable enthalpic contribution to the binding free energy). In order to keep the binding free energy constant around −16 kcal/mol,[2] these highly favorable contributions to the binding thermodynamics must be offset by other (unfavorable) interactions. Proteins can employ, broadly speaking, two major strategies as part of this enthalpy−entropy compensation mechanism: regions of the protein can fold into ordered secondary structures (a dramatic decrease in entropy) or the protein−DNA complex can induce large bends or kinks in the DNA (a dramatic decrease in enthalpy due to broken π-stacking and hydrogen bonding). Comparative studies have shown that proteins which fold dramatically upon binding tend to bind straight (B-form) DNA, while proteins that strongly bend the DNA tend to have less local folding upon binding.[1,2] However, despite all of this work, it remains impossible to provide a single code to explain protein−DNA recognition and sequence-specific binding.[5−7]

Computational methods, and molecular dynamics simulations in particular, hold great promise for elucidating the details of sequence-specific protein−DNA interactions. Coarse grained and multiscale models are providing unprecedented insights into the behavior of large protein complexes and their cooperative role in DNA binding.[8,9] In combination with novel NMR experiments, molecular simulations are shedding considerable insight on the process by which proteins transition from non-specific to sequence-specific binding of DNA.[10−12] Docking analyses are even beginning to reach the point of being able to predict the binding sequences of DNA-binding proteins.[13−15] In the midst of all of this data, however, all-atom molecular dynamics simulations remain the backbone of the computational effort to elucidate the structures and properties of DNA that direct protein binding. Pioneering work is being done by the Asconsa B-DNA Consortium and others to characterize the dynamical properties of extended DNA sequences.[16−20] The insights provided by atomistic simulations remain unmatched by any other technique available. It is in light of these developments on DNA dynamics and flexibility that increased and careful attention be paid to the atomistic details of sequence-specific DNA-binding proteins.[21−23] A recent molecular dynamics study of the *lac* repressor protein indicated that the degree of dynamical flexibility of the DNA may be an integral part of the protein−DNA recognition mechanism.[24] In this work we use all-atom molecular dynamics simulations to examine the yeast protein GCN4, which folds dramatically upon binding to unbent DNA, to understand the dynamical nature of the protein−DNA complex and its implications for sequence-specific recognition.

The yeast transcriptional activator protein GCN4 regulates a network of approximately 35 genes responsible for controlling amino acid biosynthesis. The GCN4 protein binds DNA as a homodimer in which a basic leucine zipper (bZIP) domain facilitates the formation and stabilization of the dimerization interface.[25] Upon sequence-specific binding, the protein folds from 70% to 95% helical and the (positively charged) basic region of the zipper facilitates non-specific contacts between the protein and the DNA that help to position the recognition elements in the major groove.[26] Crystal structures of the

Received: July 24, 2014
Published: September 10, 2014

dx.doi.org/10.1021/ci500448e | J. Chem. Inf. Model. 2014, 54, 2869−2875

sequence-specific protein−DNA complex indicate that the DNA remains largely unbent in the canonical B form when the protein is bound (Figure 1A).[26]

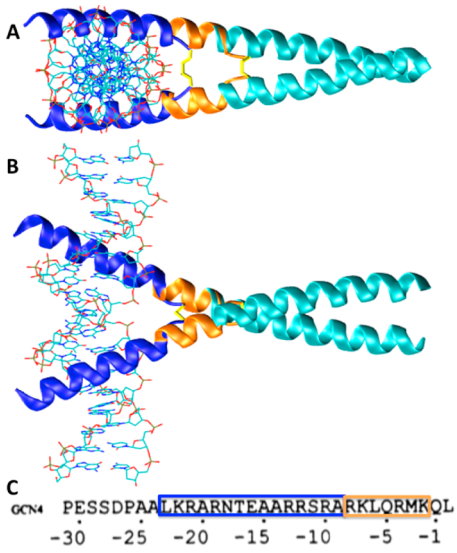


Figure 1. (A,B) Graphical depiction of the GCN4 variants used in this study looking along and perpendicular to the DNA axis, respectively; the 17-aa, 24-aa, and full-length DNA-binding protein are shown in blue, orange, and cyan, respectively. The disulfide bonds used in the truncated variants are shown in yellow. The DNA is shown as lines. (C) The amino acid sequences used for the 17-aa and 24-aa truncated variants; numbering follows the convention of Talanian et al.[31]

Mutation experiments have demonstrated that N235, A239, and R243 are critical for sequence recognition in GCN4.[27−29] The wild-type GCN4 protein shows a strong preference for the 5′-TGACTC-3′ sequence, with particular emphasis on the TGA triad; an A239V mutant binds preferentially to a TTG triad.[27] While the specific protein−DNA contacts are primarily located in the first three basepairs of the operator (TGA triad), experiments indicate that sequence preferences can extend as far as six or eight basepairs on either side of this triad.[29] It is our belief that these extended sequence preferences are dynamical in nature and arise from the mobility of the protein−DNA complex; we present evidence that variants of the GCN4 protein show significant differences in their mobility and dynamics in complex with DNA.

To date, the question of the minimum length required for a DNA-binding peptide or protein remains an open question; substantial work in this regard has been performed by Talanian and co-workers.[30,31] A particularly enlightening set of experiments on GCN4 revealed that 17 amino acids in each arm of the dimer are sufficient to confer sequence-specific binding ability, but the inclusion of an extra seven amino acids in the C-terminal (bZIP) domain significantly enhances the stability of the protein−DNA complex (Figure 1B).[31] In this work, we present the results of a molecular dynamics study showing that these truncated variants form sequence-specific complexes with the ATF/CREB operator DNA, though with dramatically different dynamical motifs than the full-length protein.

MATERIALS AND METHODS

The coordinates of the GCN4 protein bound to the ATF/CREB site DNA were obtained from the Protein Data Bank (access code 2DGC).[32] Truncated DNA-binding domains were prepared containing segments of 17 and 24 basepairs based on compounds GCN4-br9 and GCN4-br2 from Talanian et al.[31] As described in Talanian et al.,[31] each peptide was modeled with a GGC extension on the C-terminal end to allow the formation of a disulfide bond to hold the two peptide chains together. The GGC residues were modeled based on the backbone and CB atoms of the amino acids in the full protein and were subsequently minimized while keeping the rest of the protein and DNA atoms fixed. To investigate the importance of homodimerization for sequence-specific DNA-binding, a segment of GCN4 containing only one arm was also prepared by keeping one arm of the 17 amino acid system (without the GGC extension on the C-terminal end). To assess the distortion of DNA in the complex from canonical B-DNA, the ATF/CREB site was simulated in the absence of protein to establish a baseline for the DNA dynamics.

All five systems were solvated in a box of TIP3P water[33] sized to extend a minimum of 12 Å beyond the non-hydrogen atoms of the solute, with resulting system sizes of 41,231 atoms for the full-length protein, 26,664 atoms for the 24-amino acid (24-aa) variant, 22,769 atoms for the 17-aa variant, 20,033 atoms for the one-arm variant, and 19,028 atoms for the DNA only. A salt concentration of 100 mM KCl was used; the ions were placed using the cionize plugin in VMD,[34] with the constraint that ions not be initially placed closer than 5 Å to the solute or to one another. After minimization and heating, each system was equilibrated for 2 ns followed by 100 ns of production run. All simulations were performed using NAMD[35] in the NPT ensemble at 1 atm and 298 K using the CHARMM27 all-atom protein and nucleic acid force field[36] with the CMAP correction.[37] SHAKE[38] was applied to constrain bonds involving hydrogen, and a 2 fs time step was used. Langevin dynamics were used to control the temperature at 298 K with a damping coefficient of 1/ps applied to all non-hydrogen atoms. A Nose−Hoover Langevin piston was used for pressure control[39,40] with a target pressure of 1 atm at 298 K, an oscillation period of 0.2 ps, and a damping time scale of 0.1 ps. Electrostatic forces were calculated with the particle mesh Ewald method[41] using a real space cut off of 12 Å with a κ of 0.4 $Å^{-1}$. van der Waals forces were truncated with a switching function between 10 and 12 Å, and the interaction energies were calculated using the geometric mean of the well depths. Snapshots were saved every 0.2 ps and used at intervals of 1 ps for analysis. Structures were visualized with VMD,[34] and secondary structures were calculated with STRIDE.[42]

The backbone RMSD and the ion−ion radial distribution functions were computed using MDAnalysis.[43] Diffusion coefficients were calculated at lag times from 50 to 250 ps using the Diffusion Coefficient Tool for VMD,[44] and for all simulations were on the order of 10^{-5} cm/s, consistent with the literature values.[45−47] Hydrogen bonds between solute atoms were calculated in VMD using a distance cut off of 3 Å, an angle cutoff of 30°, and a minimum lifetime of 1 ps. The occupancy of hydrogen bonds between the protein and DNA was calculated as the percentage of the total simulation trajectory during which the hydrogen bond was present. Covariance matrices to analyze the collective motions in the proteins[48] were calculated using the Carma program.[49] To capture nonlinear correlations, the generalized correlation coefficient for each pair of atoms was calculated following the method of Lange and Grubmuller.[50] Structural analyses of the DNA were performed with 3DNA;[51,52] the energy penalty due to

distortion of the DNA from canonical B-DNA was calculated using the method of Olson et al.[53]

A block analysis[54] was performed to assess the convergence of the root-mean-square deviations (RMSD) of the CA atoms during the simulations. Convergence was calculated by dividing the trajectory into blocks (a range of values between 2 and 500 blocks was used); the RMSD was calculated for each block, and the standard deviation was calculated over all of the blocks. Convergence of the variance−covariance matrices was assessed by calculating $R(t) = (C(t) - C(t - N\tau))^2/N_{res}$, where N_{res} is the number of residues, $C(t)$ is the covariance matrix for the full trajectory, and $C(t - N\tau)$ is the covariance matrix calculated for a trajectory shortened by N multiples of τ (Figure S1 of the Supporting Information).[22] The value of $R(t)$ was obtained as the average of 75 different trajectory reshufflings using τ = 100 ps; upon convergence, $R(t) \rightarrow 0$.

RESULTS AND DISCUSSION

RMSF Indicates Sequence-Specific Binding Retained in All Protein Variants. Analysis of the root-mean-squared fluctuations (RMSF) of the backbone atoms demonstrates that all of the protein variants retain sequence-specific binding to the ATF/CREB site (Figure 2). The protein−DNA interface is stable, with large fluctuations only observed at the end of each chain (the peaks in the middle of Figure 2 arise from the contiguous numbering of the two protein arms and the two DNA strands). The protein variant containing only one DNA-binding arm shows slightly elevated RMSF compared with the two-arm systems, indicating that the protein dimerization confers extra stability to the protein−DNA complex. It is surprising to note that the 24-aa variant shows lower backbone RMSF than the full-length protein, though this by itself does not signify that this variant has the greatest binding affinity; the side chains in this variant are more mobile than those in the full-length protein (to be discussed in detail in the Analysis of Hydrogen Bonding Patterns Reveals Important Determinants of Specificity section). We believe that the low RMSF of the 24-aa variant is due to the nature of the dominant quasiharmonic (QH) motion discussed below.

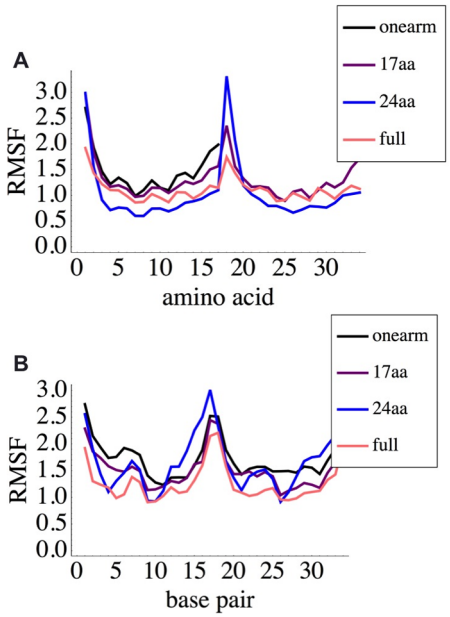


Figure 2. RMSFs for (A) the DNA-binding region of the protein (the first 17 amino acids in each arm, consisting of residues 23 through 6 in Figure 1, or in the case of the 17-aa variant, residues 23 through 9 and the GGC linker used to connect the two arms) and (B) the two 16-bp DNA strands.

Other methods to assess the stability of the simulations include the backbone RMSD and calculation of the radial distribution functions for the ions during each simulation (Figures S2−S3 of the Supporting Information). All of these methods show that our simulations are in good agreement with literature values and that the dynamics of the solute and ions are stable throughout the simulation time.

Correlation Patterns and QH Modes Show Dramatic Differences in Binding Dynamics. Perhaps the most surprising result of this study was the dramatic differences observed in the correlation patterns between the two DNA-binding arms of the protein among the various simulations (Figure 3A−C). To ensure that these differences were not an artifact of unconverged simulations, we assessed the convergence using a block analysis and find the error bars in the variance−covariance matrices to be 3 orders of magnitude smaller than the data (error <0.001 for all matrices; Figure S1 of the Supporting Information). The full-length protein shows by far the greatest amount of correlation between the two DNA-binding arms, with the magnitude of correlation decreasing as the arms are truncated. The correlation patterns show significant differences among the various systems as well, with the motion of the two DNA-binding arms being correlated in the full-length protein, anticorrelated in the 24-aa variant, and largely uncorrelated in the 17-aa variant.

Because the variance−covariance matrices only capture linear correlations, we also calculated the generalized correlation coefficient for these systems (Figure 3D−F). As expected, the full-length and 24-aa variant continue to show strong correlation in the generalized correlation matrices, indicating that the variance−covariance analysis captured the bulk of correlated motions in these systems. However, the 17-aa variant shows a surprising and dramatic increase in the magnitude of correlations captured in the generalized correlation matrix, indicating the presence of strong non-linear correlations in this system. These data clearly demonstrate that each protein variant shows strong correlation between its two DNA-binding arms, though the mechanisms by which this correlation is accomplished vary dramatically depending on the length of the arms.

While it is clear that the dramatic folding of the bZIP domain provides the chief entropy compensation for DNA-binding,[1,2] it is our belief that the significant correlation between the two DNA-binding arms represents a plausible case of entropy−entropy compensation in the sequence-specific DNA-binding of GCN4. The high degree of correlated motion may offset some of the entropy penalty of binding; a similar phenomenon was recently reported for the *lac* repressor DNA-binding headpiece as well.[24]

Diagonalization of the variance−covariance matrices and calculation of the QH modes for the three two-arm systems revealed a single dominant motion (Movies M1−M3 of the Supporting Information) for each system (the eigenvalue for the first mode was more than twice as large, and in some cases as much as 10 times as large, as that for the second mode). The dominant QH mode for the full-length protein involved the swinging of the C-terminal ends of the helices in a manner perpendicular to the long axis of the DNA; while the

dx.doi.org/10.1021/ci500448e | J. Chem. Inf. Model. 2014, 54, 2869−2875

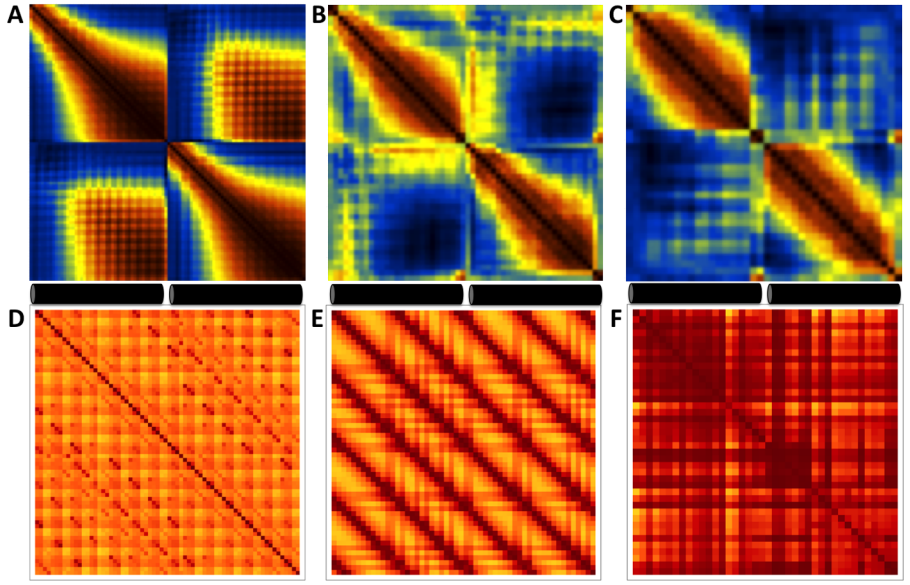


Figure 3. Normalized variance−covariance matrices and generalized correlation matrices for (A,D) the full-length protein, (B,E) the 24-aa protein, and (C,F) the 17-aa protein, respectively. Regions of correlated, anticorrelated, and uncorrelated motion are shown in red, blue, and yellow, respectively; by definition the generalized correlation coefficient is always positive and does not distinguish between correlated and anticorrelated motions. The two helices that compose the protein are shown by black cylinders.

physiological basis for this motion is unclear, it is possible that maintaining motion perpendicular to the DNA axis may facilitate the engagement of the recognition helices in the major groove of the DNA. The 24-aa variant showed a dominant QH motion that involved the twisting and stretching of the DNA-binding arms around/along the helical axes; this motion appears to explain the lower backbone RMSF observed in Figure 2, though the twisting motion may lead to disruption of some of the side-chain contacts between the protein and DNA, accounting for the lower DNA-binding affinity of this variant compared to the full-length protein. In the 17-aa variant, the dominant QH motion involved the helices swinging parallel to the long axis of the DNA; we also note significant rotational motion within the helices, giving rise to the non-linear correlations captured in the generalized correlation coefficient matrix for this system (Figure 3F). These rotations, as well as the motions parallel to the DNA axis, might tend toward slight disengagement of the helices from the major groove, accounting for the lower DNA-binding affinity of this variant.[31]

Analysis of Hydrogen-Bonding Patterns Reveals Important Determinants of Specificity. The hydrogen bonds between the protein and DNA are remarkably consistent across all of our simulations, demonstrating again the sequence-specific DNA-binding behavior of all four protein variants (Table 1). Additionally, the critical residues N235 and R243 make significant contacts to the DNA (Figure 4), consistent with mutation data on the DNA-binding interface[27−29] (A239 also makes strong, ubiquitous contacts with the T of the TGA operator, but as this is a hydrophobic interaction with the methyl group of the thymine this contact is not represented in Table 1). The full-length protein makes the most stable protein−DNA contacts at all positions except R232, which is strongest in the 24-aa variant; we believe this phenomenon is due to the dominant QH motion of this variant which may allow better contact for this residue. Not surprisingly, the shortest protein variants show deficiencies in the protein−DNA contacts at both ends of the (shortened) protein chain due to the enhanced mobility of these regions in the 17-aa arms.

Table 1. Hydrogen-Bond Occupancies for Residues That Are Important for Sequence Recognition[a]

	full	24-aa	17-aa	one arm
Lys231	50	16	1	3
Arg232	39	62	34	37
Arg234	53	48	30	20
Asn235	34	13	20	15
Arg241	57	52	37	46
Arg243	61	53	55	37
Arg245	66	50	34	23
Lys246	58	37	1	1
unique H-bonds	42	80	77	33

[a]All numbers are percent occupancies through the trajectory except the last row, which provides a count of the number of unique hydrogen bonds between the protein and DNA in each simulation (see text).

The number of unique hydrogen bonds observed between the protein and the DNA provides another measure of the stability of the protein−DNA interface. As the side chains move during the simulation, they often contact more than one base of the DNA; many of these contacts are short-lived (<1% of the trajectory) and thus are not reflected in Figure 4, however they provide useful insight into the stability of the protein−DNA interactions in each simulation. Of the three dimeric protein variants, the full-length protein shows the fewest unique hydrogen bonds, indicating a significantly more stable protein−DNA interface compared to the truncated variants. As expected, the one-arm variant shows slightly less than half the number of contacts as the dimeric protein variants due to the missing arm and the enhanced dynamical flexibility of the single DNA-binding arm (Figure 2A).

Unique Structural Features in the DNA Provide a Template for Recognition. Following the method of Olson et al.,[53] we observe that there are several base pairs within the DNA sequence that show greater than expected fluctuations relative to canonical B-DNA; these regions of enhanced

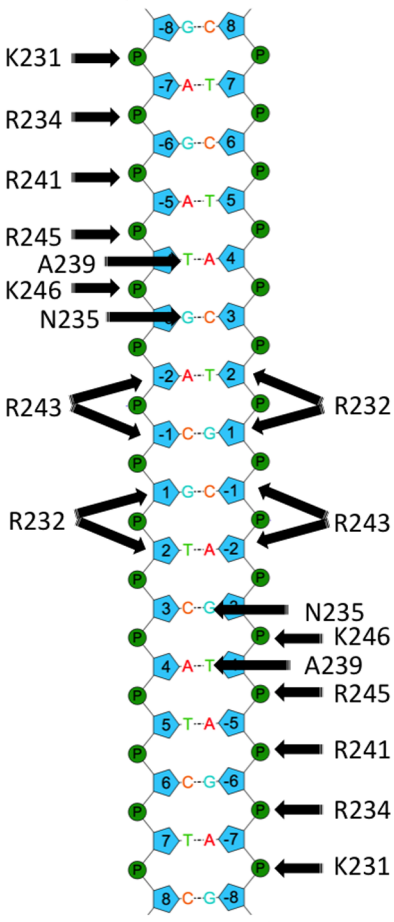


Figure 4. Schematic drawing of important hydrogen bonds.

fluctuation appear to provide a template for recognition (Figure 5). The strongest dynamical distortion from canonical B-DNA

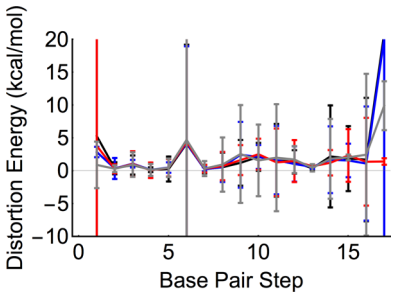


Figure 5. Distortion energy from canonical B-DNA for the simulations of the full-length protein (black), 24-aa variant (blue), 17-aa variant (red), and DNA only (gray).

occurs in basepair 6 of the operator (the G of the TGA triad), with additional enhanced fluctuations extending as far out as basepairs 13−15, consistent with experimental observations that sequence effects can extend as far as 8−10 basepairs away from the critical triad.[29] Base pair step parameter plots (Figures S4−S9 of the Supporting Information) confirm the minimal impact of the protein on the structure of the DNA.

Consistent with experimental observations,[26] the binding of GCN4 to DNA does not result in significant distortion of the DNA sequence. It is of particular interest, however, that the sequence-specific protein−DNA contacts occur at the sites that show the greatest dynamical flexibility (Figure 5). The critically conserved residues A239 and N235 contact the TG step of the TGA triad (basepairs 5 and 6). Many of the important basic residues in the bZIP domain, especially R232 and R243, make strong contacts with the CG(T)C sequence (bp 7−10) just downstream of the critical TGA. We do not observe any protein−DNA contacts as far out as basepairs 13−15, but the "neighbor effects" of these basepairs on the structure and dynamics of the operator likely contribute to the overall sequence recognition.[55]

CONCLUSIONS

We have performed a molecular dynamics study to assess the sequence-specific DNA-binding ability of truncated variants of the GCN4 protein. Consistent with experimental observations,[31] we find that 24-aa and 17-aa dimeric variants of the protein retain sequence-specific DNA-binding ability; however, we report for what we believe is the first time in the literature that a monomeric 17-aa variant also shows sequence-specific DNA-binding behavior.

Remarkably, we find that each of the dimeric protein variants show dramatically different correlation patterns and QH motions when bound to DNA. The most stable variants demonstrate a strong degree of correlation between the two protein arms, likely contributing to an entropy−entropy compensation mechanism to offset the free energy of binding. The presence of such strong correlations observed in all dimeric protein variants in this study leads us to believe that this is an important feature of many DNA-binding proteins that may be under-appreciated in the literature on the thermodynamics of protein−DNA binding. We anticipate that these types of entropy−entropy compensation strategies will become more apparent in the bZIP and other multimeric protein families as detailed studies of sequence-specific protein−DNA interactions continue.

We observe consistent sequence-specific protein−DNA contacts across all of our simulations, confirming the critical role of N235, A239, and R243 as identified by mutation experiments.[27] We also demonstrate the important DNA-binding role of many of the basic residues in the bZIP domain; long thought to be useful chiefly for the impact of their positive charge on non-specific binding, we show that several of these residues (most notably K231 and R234) contact basepairs that show enhanced dynamical flexibility relative to canonical B-DNA, indicating a plausible mechanism for the extended sequence preferences of the protein.[29] This is an exciting area of research that will require a much larger body of evidence before general conclusions can be drawn, but early signs are pointing to the importance of dynamical features of long DNA sequences in explaining the extended sequence preferences of many DNA-binding proteins.

ASSOCIATED CONTENT

Supporting Information

Movies of the principal quasiharmonic mode for each of the three dimeric protein variants and convergence data for the variance−covariance matrices. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dabarr@utica.edu. Tel: +1 315-792-3395.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Suliman Salman and Michael Convertino for helpful discussions. Computer time for this work was provided by NSF XSEDE grant number MCB110162 to D.B.

■ ABBREVIATIONS

MD, molecular dynamics; bZIP, basic leucine zipper; RMSF, root-mean-squared fluctuations; QH, quasiharmonic

■ REFERENCES

(1) Spolar, R. S.; Record, M. T. Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. *Science* **1994**, *263*, 777−784.
(2) Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structural and Thermodynamic Strategies for Site-Specific DNA Binding Proteins. *Structure* **2000**, *8*, 1015−1023.
(3) Jen-Jacobson, L.; Engler, L. E.; Ames, J. T.; Kurpiewski, M. R.; Grigorescu, A. Thermodynamic Parameters of Specific and Non-specific Protein-DNA Binding. *Supramol. Chem.* **2000**, *12*, 143−160.
(4) Hancock, S. P.; Hiller, D. A.; Perona, J. J.; Jen-Jacobson, L. The Energetic Contribution of Induced Electrostatic Asymmetry to DNA Bending by a Site-Specific Protein. *J. Mol. Biol.* **2011**, *406*, 285−312.
(5) Suzuki, M. A Framework for the DNA-Protein Recognition Code of the Probe Helix in Transcription Factors: The Chemical and Stereochemical Rules. *Structure* **1994**, *2*, 317−326.
(6) Milk, L.; Daber, R.; Lewis, M. Functional Rules for Lac Repressor−operator Associations and Implications for protein−DNA Interactions. *Protein Sci.* **2010**, *19*, 1162−1172.
(7) Pabo, C. O.; Nekludova, L. Geometric Analysis and Comparison of Protein-DNA Interfaces: Why Is There No Simple Code for Recognition? *J. Mol. Biol.* **2000**, *301*, 597−624.
(8) Villa, E.; Balaeff, A.; Schulten, K. Structural Dynamics of the Lac Repressor-DNA Complex Revealed by a Multiscale Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6783−6788.
(9) Kim, S.; Broströmer, E.; Xing, D.; Jin, J.; Chong, S.; Ge, H.; Wang, S.; Gu, C.; Yang, L.; Gao, Y. Q.; Su, X.; Sun, Y.; Xie, X. S. Probing Allostery Through DNA. *Science* **2013**, *339*, 816−819.
(10) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305*, 386−389.
(11) Givaty, O.; Levy, Y. Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* **2009**, *385*, 1087−1097.
(12) Furini, S.; Barbini, P.; Domene, C. DNA-Recognition Process Described by MD Simulations of the Lactose Repressor Protein on a Specific and a Non-Specific DNA Sequence. *Nucleic Acids Res.* **2013**, *41*, 3963−3972.
(13) Roberts, V. A.; Pique, M. E.; Hsu, S.; Li, S.; Slupphaug, G.; Rambo, R. P.; Jamison, J. W.; Liu, T.; Lee, J. H.; Tainer, J. A.; Eyck, L. F. T.; Woods, V. L. Combining H/D Exchange Mass Spectroscopy and Computational Docking Reveals Extended DNA-Binding Surface on Uracil-DNA Glycosylase. *Nucleic Acids Res.* **2012**, *40*, 6070−6081.
(14) Chen, C.-Y.; Chien, T.-Y.; Lin, C.-K.; Lin, C.-W.; Weng, Y.-Z.; Chang, D. T.-H. Predicting Target DNA Sequences of DNA-Binding Proteins Based on Unbound Structures. *PLoS One* **2012**, *7*, e30446.
(15) Mobarec, J. C.; Wolf, D.; Bischofs, I. B.; Kolb, P. Computational Modeling, Docking and Molecular Dynamics of the Transcriptional Activator ComA Bound to a Newly-Identified Functional DNA Binding Site. *J. Cheminf.* **2014**, *6*, P30.
(16) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. A Systematic Molecular Dynamics Study of Nearest-Neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299−313.
(17) Beveridge, D. L.; Barreiro, G.; Suzie Byun, K.; Case, D. A.; Cheatham Iii, T. E.; Dixit, S. B.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, Pè.; Young, M. A. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. I. Research Design and Results on d(CpG) Steps. *Biophys. J.* **2004**, *87*, 3799−3813.
(18) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E., III; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, Pè. Molecular Dynamics Simulations of the 136 Unique Tetranucleotide Sequences of DNA Oligonucleotides. II: Sequence Context Effects on the Dynamical Structures of the 10 Unique Dinucleotide Steps. *Biophys. J.* **2005**, *89*, 3721−3740.
(19) Spiriti, J.; van der Vaart, A. DNA Bending through Roll Angles Is Independent of Adjacent Base Pairs. *J. Phys. Chem. Lett.* **2012**, *3*, 3029−3033.
(20) Curuksu, J.; Zacharias, M.; Lavery, R.; Zakrzewska, K. Local and Global Effects of Strong DNA Bending Induced during Molecular Dynamics Simulations. *Nucleic Acids Res.* **2009**, *37*, 3766−3773.
(21) Araúzo-Bravo, M. J.; Fujii, S.; Kono, H.; Ahmad, S.; Sarai, A. Sequence-Dependent Conformational Energy of DNA Derived from Molecular Dynamics Simulations: Toward Understanding the Indirect Readout Mechanism in Protein-DNA Recognition. *J. Am. Chem. Soc.* **2005**, *127*, 16074−16089.
(22) Kamberaj, H.; van der Vaart, A. Correlated Motions and Interactions at the Onset of the DNA-Induced Partial Unfolding of Ets-1. *Biophys. J.* **2009**, *96*, 1307−1317.
(23) Dolenc, J.; Gerster, S.; van Gunsteren, W. F. Molecular Dynamics Simulations Shed Light on the Enthalpic and Entropic Driving Forces That Govern the Sequence Specific Recognition between Netropsin and DNA. *J. Phys. Chem. B* **2010**, *114*, 11164−11172.
(24) Barr, D.; Vaart, A. van der. The Natural DNA Bending Angle in the Lac Repressor Headpiece-O1 Operator Complex Is Determined by Protein-DNA Contacts and Water Release. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2070−2077.
(25) Hope, I. A.; Struhl, K. GCN4, a Eukaryotic Transcriptional Activator Protein, Binds as a Dimer to Target DNA. *EMBO J.* **1987**, *6*, 2781−2784.
(26) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Helices: Crystal Structure of the Protein-DNA Complex. *Cell* **1992**, *71*, 1223−1237.
(27) Suckow, M.; Von Wilcken-Bergmann, B.; Muller-Hill, B. Identification of Three Residues in the Basic Regions of the bZIP Proteins GCN4, C/EBP and TAF-1 That Are Involved in Specific DNA Binding. *EMBO J.* **1993**, *12*, 1193−1200.
(28) Suckow, M.; Von Wilcken-Bergmann, B.; Muller-Hill, B. The DNA Binding Specificity of the Basic Region of the Yeast Transcriptional Activator GCN4 Can Be Changed by Substitution of a Single Amino Acid. *Nucleic Acids Res.* **1993**, *21*, 2081−2086.
(29) Koldin, B.; Suckow, M.; Seydel, A.; Von Wilcken-Bergmann, B.; Muller-Hill, B. A Comparison of the Different DNA Binding Specificities of the bZip Proteins C/EBP and GCN4. *Nucleic Acids Res.* **1995**, *23*, 4162−4169.
(30) Talanian, R.; McKnight, C.; Kim, P. Sequence-Specific DNA Binding by a Short Peptide Dimer. *Science* **1990**, *249*, 769−771.
(31) Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S. Minimum Length of a Sequence-Specific DNA Binding Peptide. *Biochemistry (Moscow)* **1992**, *31*, 6871−6875.
(32) Keller, W.; Konig, P.; Richmond, T. J. Crystal Structure of a bZIP/DNA Complex at 2.2Å: Determinants of DNA Specific Recognition. *J. Mol. Biol.* **1995**, *254*, 657−667.
(33) Jorgensen, W.; Chandrasekar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(34) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(35) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(36) MacKerell, A. D.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(37) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(39) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.
(40) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(41) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(42) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins* **1995**, *23*, 566−579.
(43) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(44) Giorgino, T. *Computing Diffusion Coefficients in Macromolecular Simulations: The Diffusion Coefficient Tool for VMD*; GRIB: Barcelona, Spain, 2014; http://multiscalelab.org/utilities/DiffusionCoefficientTool (accessed September 1, 2014).
(45) Lyubartsev, A. P.; Laaksonen, A. Concentration Effects in Aqueous NaCl Solutions. A Molecular Dynamics Simulation. *J. Phys. Chem.* **1996**, *100*, 16410−16418.
(46) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. Ion Motions in Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14771−14775.
(47) Várnai, P.; Zakrzewska, K. DNA and Its Counterions: A Molecular Dynamics Study. *Nucleic Acids Res.* **2004**, *32*, 4269−4280.
(48) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 205−217.
(49) Glykos, N. M. Software News and Updates Carma: A Molecular Dynamics Analysis Program. *J. Comput. Chem.* **2006**, *27*, 1765−1768.
(50) Lange, O. F.; Grubmüller, H. Generalized Correlation for Biomolecular Dynamics. *Proteins* **2006**, *62*, 1053−1061.
(51) Lu, X.-J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31*, 5108−5121.
(52) Lu, X.; Olson, W. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3*, 1213−1227.
(53) Olson, W.; Gorin, A.; Lu, X.; Hock, L.; Zhurkin, V. DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163−11168.
(54) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461−466.
(55) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. A Systematic Molecular Dynamics Study of Nearest-Neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299−313.