Biophysical Journal, Volume *111*

Supplemental Information

Broken TALEs: Transcription Activator-like Effectors Populate Partly Folded States

Kathryn Geiger-Schuller and Doug Barrick

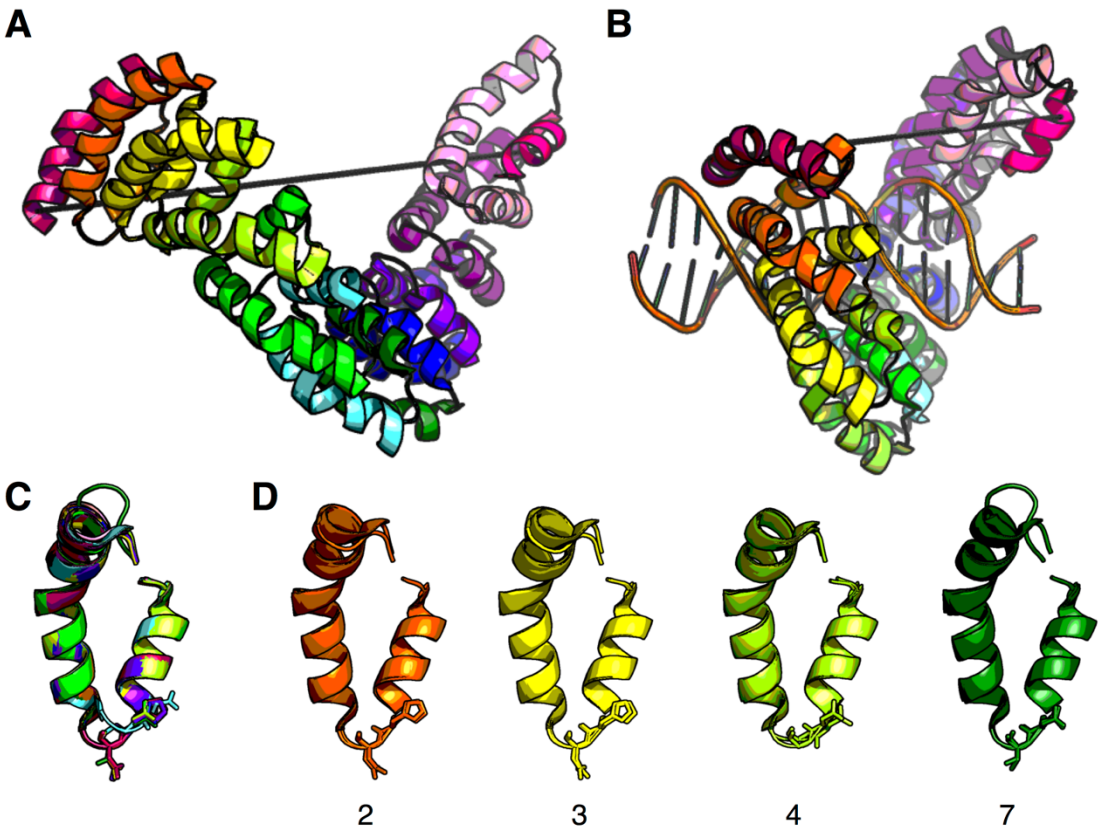


Figure S1. The TALE conformational change upon DNA binding is mediated by small changes propagated through many repeats. (A) DNA-free structure of dHax3 (colored by repeat, PDB:3V6P). The distance between C_α of residue 304 to C_α of residue 666 is 74 Å. (B) DNA-bound structure of dHax3 (PDB:3V6T). The distance between C_α of residue 304 to C_α of residue 666 is 50 Å. (C) Alignment of C_α s of all 11 repeats in the DNA-bound structure. RMSDs range from 0.3 to 0.6 $Å^2$. (D) Alignment of C_α s of repeats 2, 3, 4, and 7 in the DNA-free and DNA-bound structure. RMSDs range from 0.3 to 0.6 $Å^2$.

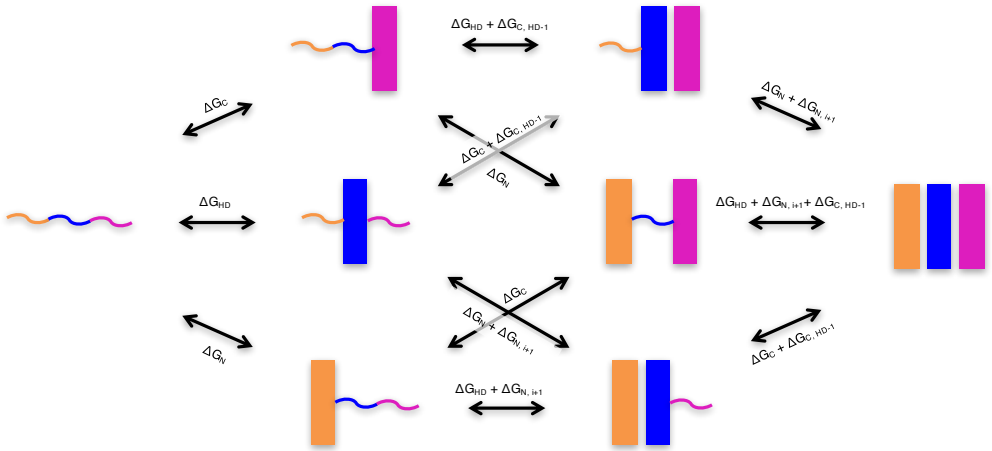


Figure S2. A 1-D Ising Model for a cTALE with three repeats. In this example, a construct with the one N-cap, one HD repeat, and one C-cap is shown. The partition function *q* is used to derive the dependence of the fraction folded repeats (f_{folded}) on urea concentration (See equations 1-4). The folding of each repeat results in an intrinsic free energy change, along with free energy changes from formation of one or two interfaces (for terminal repeat unfolding and internal repeat folding respectively).

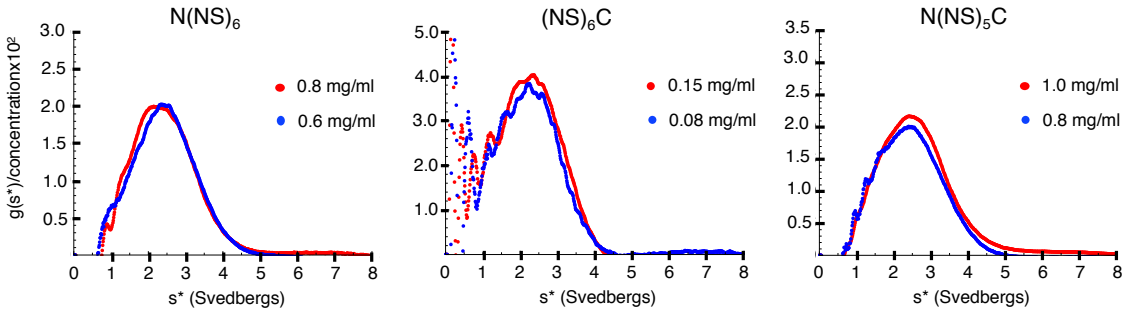


Figure S3. Sedimentation Velocity g(s*) plots for capped consensus TALE repeats. Sedimentation velocity experiments were performed with $N(NS)_6$, $(NS)_6C$, and $N(NS)_5C$, as described in Supporting Materials and Methods. The g(s*) distributions are consistent with monomers. Conditions: 300 mM NaCl, 10 mM Tris HCl pH 7.4, 5% glycerol, 25°C.

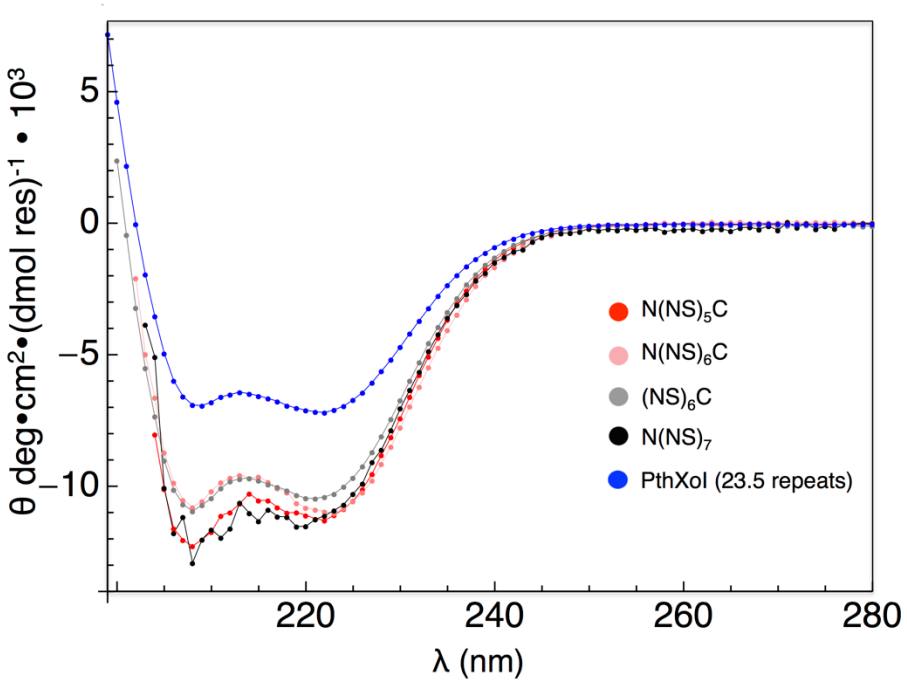


Figure S4. Consensus and natural TALEs and have α-helical secondary structure. FarUV CD of cTALEs and the repeat domain of PthXo1, a naturally-occurring TALE, show similar shape. PthXo1 has less helical structure than the cTALEs.

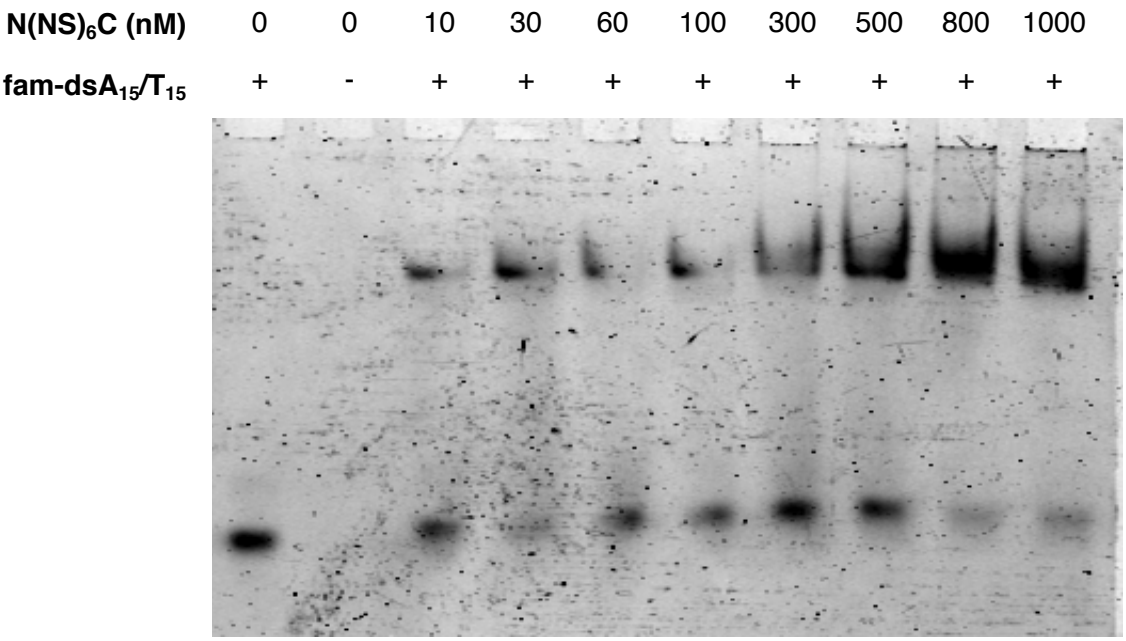


Figure S5. Consensus TALEs bind double stranded DNA. An electrophoretic mobility shift assay (EMSA) shows $N(NS)_6C$ binding 10 nM fam-dsA_{15}/T_{15} (see Supporting Materials and Methods). Conditions: 150 mM KCl, 0.1 mM DTT, 10 mM Tris, pH 7.4, 25% sucrose.

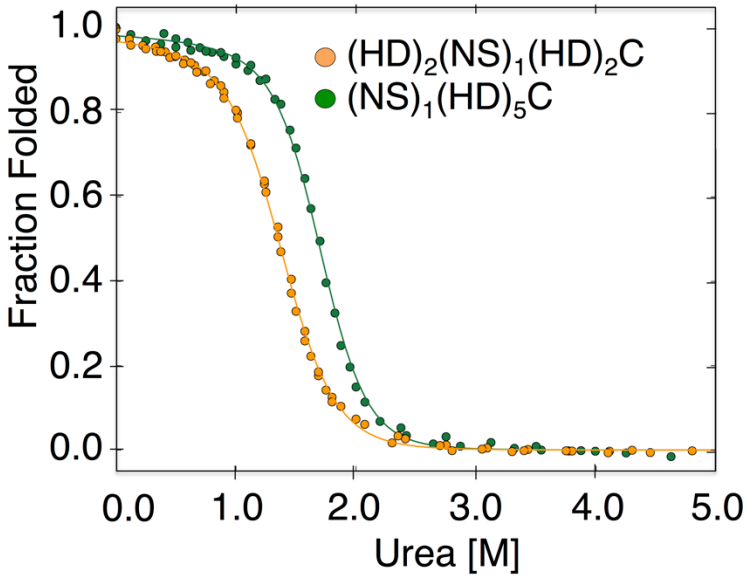


Figure S6. "Mixed RVD" constructs have single cooperative unfolding transition. Urea-induced unfolding transitions for constructs containing both NS- and HD-type repeats. Transitions are cooperative and slopes of $(HD)_2(NS)_1(HD)_2C$ and $(NS)_1(HD)_5C$ are similar to $(NS)_6C$, suggesting cooperative unfolding units of similar size.

	$p_{fully\ folded}$	$p_{end\ frayed}$	$p_{internally\ unfolded}$	$p_{interfacially\ fractured}$
cTALE (NS)	0.92	0.08	8.3×10^{-7}	1.9×10^{-5}
cTALE(HD)	0.85	0.15	1.7×10^{-4}	3.9×10^{-3}
cTALE(mix)	0.11	0.89	1.1×10^{-4}	3.1×10^{-3}
cANK	>0.999	<0.001	1.5×10^{-14}	8.4×10^{-9}

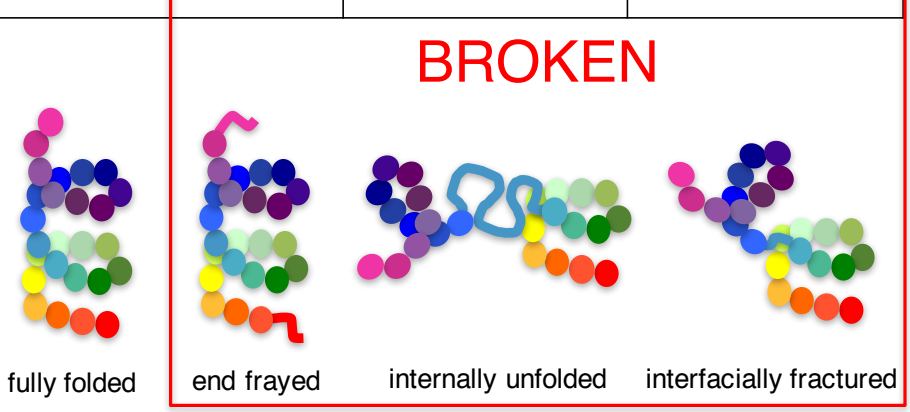


Figure S7. Calculated probabilities for fully folded and broken TALEs. Probabilities of fully folded, end frayed, internally unfolded, and interfacially fractured populations were calculated using free energies from Ising analysis of cTALEs and, for comparison, consensus ankyrin repeat proteins (cANK)(19). Calculations were performed on arrays of 20 repeats. Microstates containing unfolded repeats in the regions 1-5 or 16-20 were included in the calculation of $p_{end\text{-}frayed}$. Microstates containing unfolding in repeats 6-15 were included in the calculation of $p_{internally\ unfolded}$. To calculate $p_{interfacially\ fractured}$, a separate Ising-like model was generated that includes microstates with structural deformation between two consecutive uncoupled but folded repeats (Supporting Materials and Methods). There is a significant population of broken states for cTALEs, in contrast to the very low probabilities of broken states calculated for cANKs.

Supporting Materials and Methods

Sedimentation Velocity

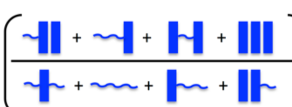
Analytical ultracentrifugation sedimentation velocity experiments were performed using a Beckman XL-1 analytical ultracentrifuge as previously described(22). Proteins were dialyzed in reference buffer for 24 hours prior to centrifugation. Concentrations ranged from 2 to 30 µM. AUC data were analyzed with SEDANAL. DCDT analysis was performed on scans spanning a one hour interval, generating s* value distributions (g(s*) versus s*)(23).

Electrophoretic Mobility Shift Assay (EMSA)

Flourescein (Fam)-labeled A_{15} single stranded DNA was annealed with unlabeled T_{15} single stranded DNA to prepare fam-dsA_{15}/T_{15}. Reactions with increasing concentrations of cTALEs were incubated at room temperature for 20 minutes. Samples were then loaded onto fresh 6% non-denaturing 0.5X TBE gels and were electrophoresed at 100 volts in the cold room for 60 minutes. Gels were imaged on the JHU Integrated Imaging Center Typhoon 9410 Variable Mode Imager and analyzed with ImageJ.

Calculating local folding free energy from intrinsic and interfacial stabilities

Local folding free energies (ΔG°_{local}) describe the free energy difference between all states where a particular repeat, *i*, is folded, and all states where the ith repeat is unfolded (Equation S1). For example, to calculate the ΔG°_{local} of the third repeat in a three repeat array, partition functions for each construct are created as illustrated in Figure S5. Next, q_3 sums all statistical weights of all conformations where the third repeat is folded. Placing zeros in the third matrix is equivalent to not counting statistical weights when the third repeat is unfolded (Equation S2). Dividing q_3 by the partition function, q (Equation S3), gives the probability that the third repeat is folded (Equation S4). The probability that the third repeat is unfolded is simply one subtracted by the probability of the third repeat being folded. The ΔG°_{Local} of a repeat is the log of the ratio of these probabilities (Equation S5).

Local Stability $= -RT \ln \left(\right.$  $\left. \right)$ (S1)

$$q_3 = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{(S2)}$$

$$q = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \kappa\tau & \kappa \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{(S3)}$$

$$p_{3,N} = \frac{q_3}{q} \quad \text{(S4)}$$

$$\Delta G^{\circ}_{Local} = -RT \ln \left(\frac{p_{3,N}}{p_{3,D}} \right) \quad \text{(S5)}$$

An extended Ising model to include fractured states

The 1-D Ising model we used to describe folding can be extended to include states where consecutive repeats are folded but uncoupled. The physical interpretation of such states is that a structural deformation occurs leaving repeats folded but preventing them from forming favorable interactions required for coupling. Such states should be accessible by opening a turn between helices, requiring changes to just a few backbone dihedral angles. Generation of partition functions for the folding Ising model is described in Equations 1-4 in the Materials and Methods. Equation 2 can be modified to include fractured states by replacing $\kappa\tau$ with $\kappa\tau + \kappa$ in the first column and row of the 2x2 matrices for each repeat.

$$\theta = \begin{bmatrix}1 & 1\end{bmatrix} \begin{bmatrix}\kappa_N\tau_N+\kappa_N & \kappa_N \\ 1 & 1\end{bmatrix} \begin{bmatrix}\kappa_R\tau_R+\kappa_R & \kappa_R \\ 1 & 1\end{bmatrix}^{n-2} \begin{bmatrix}\kappa_C\tau_C+\kappa_C & \kappa_C \\ 1 & 1\end{bmatrix} \begin{bmatrix}0 \\ 1\end{bmatrix} \quad \text{(S6)}$$

If the *i* and *i* + 1 repeats can be folded or unfolded, there are four possible states (FF, FU, UF, and UU). Each of the four elements in the 2x2 matrices represent one of these states. Row one, column one in each matrix represents states where the *i* and *i* + 1 repeats are folded (FF). In the folding Ising model, if *i* and *i* + 1 are folded, the statistical weight is $\kappa\tau$ because the interfacial energy is automatically associated with the folding of the i^{th} repeat (through τ). To allow for states where there is no coupling between consecutive repeats, we simply add a term κ to the position where both repeats are folded. This is saying that when the *i* and *i* + 1 repeats are folded (FF), either there is a coupling ($\kappa\tau$) or there is not (κ). Because these two options are mutually exclusive, their contributions to statistical weight sum.

Urea induced unfolding transitions fitted with this extended model returned best-fit values within 0.03% of the best-fit values in Table 1, with the exception of the interfacial energy of the N-cap and next repeat ($\Delta G_{N,\,N+1}$). Because the interface between the N-cap and the adjacent repeat is rather weak, the sum $\kappa\tau + \kappa$ in the modified Ising model is significantly larger than the Ising value $\kappa\tau$. As a result, the best-fit value for $\Delta G_{N,\,N+1}$ is -0.69 kcal/mol with the fracture Ising model as compared to -0.85 kcal/mol with the folding Ising model.

Supporting References

22. Marold, J.D., J.M. Kavran, G.D. Bowman, and D. Barrick. 2015. A Naturally Occurring Repeat Protein with High Internal Sequence Identity Defines a New Class of TPR-like Proteins. Struct. Lond. Engl. 1993. .

23. Stafford, W.F. 1992. Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. Anal. Biochem. 203: 295–301.