

Stability islands and the folding cooperativity of a seven-repeat array from Topoisomerase V

Mark Petersen¹, Rebecca Fang¹, Ananya Majumdar², and Doug Barrick^{1*}

¹The T.C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218

²The Johns Hopkins University Biomolecular NMR Center, Johns Hopkins University

*Email: barrick@jhu.edu

Keywords: Ising model, cooperativity, protein folding, Topoisomerase V, Helix-hairpin-helix motif, repeat protein.

Abstract: Cooperativity is a central feature of protein folding, but the thermodynamic and structural origins of cooperativity remain poorly understood. To quantify cooperativity, we measured guanidine-induced unfolding transitions of single helix-hairpin-helix (HhH)₂ repeats and tandem pairs from a seven-repeat segment of *Methanopyrus kandleri* Topoisomerase V (Topo V) to determine intrinsic repeat stability and interfacial free energies between repeats. Most single-repeat constructs are folded and stable; moreover, several pairs have unfolding midpoints that exceed midpoints of the single repeats they comprise, demonstrating favorable coupling between repeats. Analyzing unfolding transitions with a modified Ising model, we find a broad range of intrinsic and interfacial free energies. Surprisingly, the G repeat, which lacks density in crystal structure of Topo V without DNA, is the most stable repeat in the array. Using NMR spectroscopy we demonstrate that the isolated G repeat adopts a canonical (HhH)₂ fold, and forms an ordered interface with the F repeat but not with the H repeat. Using parameters from our paired Ising fit, we built a partition function for the seven-repeat array. The multistate unfolding transition predicted from this partition function is in excellent agreement with the experimental unfolding transition, providing strong justification for the nearest-neighbor model. The seven-repeat partition function predicts a native state in which three independent segments (“stability islands”) of interacting repeats are separated by two unstable interfaces. We confirm this segmented architecture by measuring the unfolding transition of an equimolar mixture of these three separate polypeptides. This segmented structural organization may facilitate wrapping around DNA.

INTRODUCTION

Despite the structural complexity of proteins, their folding is often quite simple, showing considerable cooperativity across distant structural elements. For most single-domain proteins, cooperativity leads to two-state folding. Although two-state transitions facilitate the quantification of protein stability, the all-or-none nature of these transitions prevents cooperativity from being quantified thermodynamically and resolved into its structural origins.

Linear repeat proteins have proven to be ideal experimental systems to quantify the thermodynamic and structural basis of cooperativity.^{1–5} Deleting individual repeats provides access to intrinsic folding free energies (which describe the stabilities of single repeats in isolation) and interfacial free energies (which describe the stability increment gained from forming an interface between adjacent folded repeats). These free energy values can be determined using a one-dimensional Ising model. This approach has been used to quantify cooperativity in a variety of non-autonomous repeat proteins (NARPs).^{1–6}

By definition, NARPs have intrinsically unstable repeats; because of this instability, repeats from NARPs cannot be studied in isolation. As a result, Ising analysis of NARPs has largely been restricted to arrays of identical repeats to reduce the number of unknown Ising parameters. This restriction precludes Ising analysis of naturally occurring repeat proteins, where there are substantial sequence differences from repeat to repeat. These sequence differences are expected to give rise to substantial differences in stability along tandem repeat arrays, which may be important for function.

Here we use an Ising model to analyze the stability of an array of naturally occurring helix-hairpin-helix ((HhH)₂) repeats from *Methanopyrus kandleri* Topoisomerase V (Topo V, Figure 1). (HhH)₂ repeats are found in a variety of proteins that bind DNA, such as helicases, polymerases, and glycosylases.⁷ (HhH)₂ repeats comprise two helix-hairpin-helix (HhH) motifs connected by a short helix. Whereas most

proteins contain only one or two (HhH)₂ repeats[†], Topo V contains an array of twelve tandem repeats. Both the larger size of the (HhH)₂ repeat compared to most NARP repeats (49-60 versus 25-40 residues) and the observation that many (HhH)₂ motifs are found as single copies suggests that these repeats may fold autonomously, facilitating Ising analysis of a sequence-variable repeat array.

In crystal structures of Topo V in the absence of DNA, the (HhH)₂ repeats extend out from the main topoisomerase domain in an arc.⁸⁻¹² Adjacent repeats typically make extensive contacts with each other, suggesting that neighboring repeats are stabilized by strong interfaces. These structural observations suggest that the Topo V (HhH)₂ array may be a semi-autonomous repeat protein (SARP), in which repeats can both fold independently and form stabilizing interactions between neighbors.⁶

SARPs are ideal candidates for Ising analysis because the intrinsic and interfacial stability variations resulting from the sequence variation among repeats can be accurately quantified by analysis of single- and paired-repeat constructs.⁶ Despite high sequence variability (an average pairwise identity around 25% among aligned repeats^{8,11,13}), a structural alignment of each of the nine resolved crystallographic Topo V (HhH)₂ repeats shows conserved folds.¹¹ The orientations of adjacent (HhH)₂ repeats are also structurally conserved, although the hydrogen bonding, electrostatic interactions, and buried hydrophobic surface vary from interface to interface. Thus, though the repeats and interfaces are structurally conserved, the low sequence conservation may lead to significant differences in intrinsic and interfacial energies, and thus, local stabilities within the larger Topo V array.

To quantify these intrinsic stabilities and interfacial free energies, we use a one-dimensional Ising model to analyze the equilibrium folding transitions of an array of seven (HhH)₂ repeats (D-J) from Topo V. We do this by globally fitting guanidine-induced unfolding and refolding transitions of the seven individual repeats along with transitions of six pairs of neighboring repeats. We find that the local stability is highly variable; surprisingly, the G repeat, which is only resolved in a crystal structure when bound to DNA^{11,12}, is the most stable part of the array. We use our parameters from the global Ising fit to build a partition function for the seven-repeat array. This partition function has 610 states, only some of which are populated to a significant degree in the predicted seven-repeat unfolding transition. We show that this predicted unfolding transition is in excellent agreement with an experimental unfolding transition, providing a strong validation of the nearest-neighbor approximation in the one-dimensional Ising model. Using multi-dimensional solution state NMR, we show that the isolated G repeat adopts a canonical (HhH)₂ fold without DNA, and that G forms an ordered and stabilizing N-terminal interface to the F repeat but not to the H repeat.

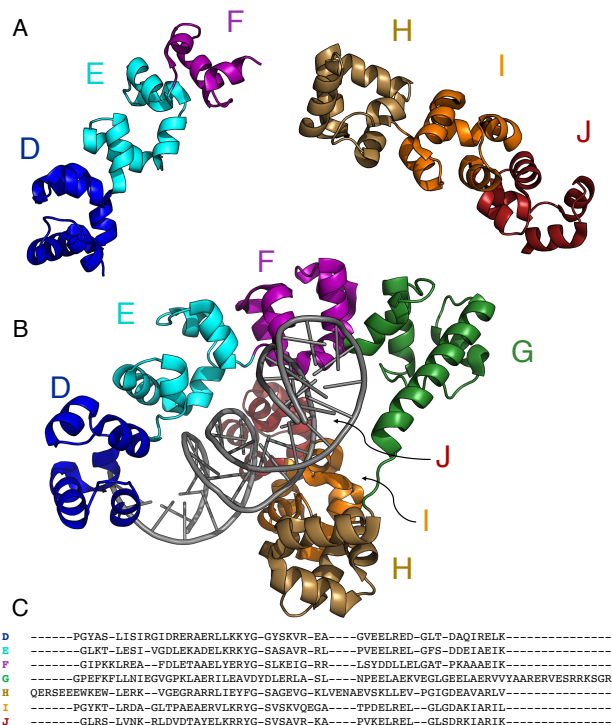


Figure 1. The Topo V (HhH)₂ repeats investigated in this study. (A) Structure of the fourth through tenth (HhH)₂ repeats from Topo V in the absence of DNA, colored by repeat, excised from the crystal structure 5HM5.¹¹ Half of F and all of G have poorly-resolved electron density, and could not be modeled. (B) Structure of the TopoV array excised from the DNA-bound crystal structure 8DF8.¹² (C) Structure-based multiple sequence alignment of the (HhH)₂ repeats. Repeat boundaries are from a previously published structure-based alignment,¹¹ with the exception of G and H, where the sequence QERSEE is included at the N-terminus of H instead of the C-terminus of G.

MATERIALS AND METHODS

Cloning and Expression. Genes encoding D, F, J, DE, EF, FG, GH, HI, IJ, and DEFGHIJ were synthesized by Thermofisher, and were codon optimized for *E. coli* expression. Genes were inserted into linearized expression plasmids by Gibson Assembly (New England Biolabs).¹⁴ DE was inserted into pET-24(+) with a cleavable N-terminal His tag, while the others were inserted into a modified pMal vector with cleavable His tag and maltose binding protein (MBP) solubility tag. Constructs expressing E, G, H, and I were cloned by adding a stop codon to pair constructs above using Quikchange (Agilent Technologies). EFG and HIJ constructs were synthesized and inserted into pET-29b+ by Twist Bioscience using NdeI and XhoI restriction sites.

Proteins were expressed in *E. coli* BL21[DE3] cells. Cells were grown in lysogeny broth at 37°C with either 50 µg/ml kanamycin (for DE, EFG, HIJ) or 100 µg/ml ampicillin (all

[†] See pfam 35.0 entry PF1420, *HHH_5*.

other constructs) to an OD₆₀₀ of 0.6-0.8. Protein expression was induced by adding IPTG to 0.5 mM and shaking at 20°C overnight. Cells were pelleted and stored at -80°C.

Protein Purification. Frozen cell pellets were resuspended in 25 mM Tris, 50 mM NaCl, 0.5 mM TCEP, pH 8 (20 ml per liter of cell culture); one cOmplete mini protease inhibitor tablet (Roche) was added per two liters of cell culture. Resuspended cells were lysed by sonication. Lysates were cleared by centrifugation at 15,500 RPM in a Beckman JA-20 rotor for one hour. DNaseI (Roche) and MgCl₂ were added to cleared lysates to final concentrations of 2.5-5 µg/ml and 2 mM, respectively, and were incubated at 4°C for one hour. Protein was purified on a gravity-flow Ni-NTA column equilibrated with 25mM Tris, 500mM NaCl, pH 8. The wash and elution contained 30mM and 500mM imidazole, respectively.

Eluted protein was cleaved overnight at 4 °C with 25 µg/ml TEV protease (prepared separately in our lab) during dialysis against 25 mM Tris, 150 mM NaCl, 0.5 mM TCEP, pH 8 (although none of the cleaved Topo V constructs contain cysteines, TCEP was added to prevent TEV inactivation). Cleaved protein was repassed over the Ni-NTA column, collected in the flow-through, and dialyzed overnight into 25 mM Tris pH 8. J-containing constructs (J, IJ, HIJ, DEFGHIJ) were prone to aggregation, which was reduced by adding 10% glycerol in cleavage, Ni-NTA repass, and ion exchange chromatography buffers, and by reducing the NaCl concentration in dialysis to 50 mM rather than 0 mM in preparation for ion exchange.

Constructs with low predicted pI values (E, F, G, H, EF, FG, GH, EFG, DEFGHIJ) were purified further with an anion exchange Q-FF column equilibrated with 25 mM Tris pH 8 (and 10% glycerol with 50 mM NaCl for J-containing constructs) and were eluted with a salt gradient with buffer containing 1M NaCl. Constructs with high predicted pI values (D, I, J, IJ, HIJ) were purified using a cation exchange SP-FF column equilibrated with the same Tris buffer described above. Constructs DE and HI were passed over the Q-FF column and collected in the flow-through to separate from cleaved MBP, which was retained on the column. All constructs were further purified on an S100 size-exclusion column equilibrated with PBS (50 mM sodium phosphate, 150 mM NaCl, pH 7).

To prepare isotopically labeled proteins for NMR spectroscopy, cells were grown in M9 minimal media with ¹⁵NH₄Cl and uniformly labeled ¹³C glucose (Cambridge Isotope Laboratories). Proteins were purified as described above except that the final size-exclusion column was run in 25 mM sodium phosphate, 50 mM NaCl, pH 6.4. The salt and pH were decreased to improve spectra quality.

Circular Dichroism Spectroscopy. Circular dichroism (CD) spectra and unfolding transitions of Topo V constructs were collected on either an Aviv model 420 or model 435 CD spectropolarimeter. For spectra, protein concentrations were between 30-70 µM in PBS. Spectra were acquired in a 1 mm quartz cuvette with 3s averaging time at each point.

Guanidine-induced unfolding transitions were acquired with 3-18 µM protein in 1 cm pathlength quartz cuvettes, averaging the CD signal at 222 nm for 30 seconds. Guanidine was titrated using a computer-controlled Hamilton Microlab syringe titrator, either injecting a solution of buffered protein in guanidine to a cuvette containing buffered protein for "forward" titrations, or buffered protein to a cuvette containing buffered protein with guanidine for "reverse" titrations. After each injection, samples were equilibrated with stirring for 5-6 min, or for 10 min for constructs containing proline residues.

Ultrapure guanidine (Invitrogen) was used to make denaturing PBS buffer. For titrations lacking TMAO, guanidine concentration was measured by refractometry.¹⁵ For titrations containing both guanidine and TMAO (Sigma), refractive index (RI) increments were assumed to be additive. The concentration of the titrant (e.g., guanidine) was determined refractometrically, assuming a known refractive index increment from the stationary component (e.g., TMAO).^{15,16}

Fractured-States Ising model and Fitting Equations for CD Melts. To globally analyze folding transitions of Topo V constructs, we generated fitting equations using a one-dimensional Ising model that includes fractured states as described previously.^{5,6} The equilibrium constant for an individual repeat *i* as a function of guanidine (*x*) is

$$\kappa_i = e^{-(\Delta G_{i,H2O} + m_i * x) / RT} \quad (1)$$

For transitions of J in the presence of TMAO, we include a second *m*-value for TMAO (*y*) sensitivity:

$$\kappa_j = e^{-(\Delta G_{j,H2O} + m_j * x + m_{j,TMAO} * y) / RT} \quad (2)$$

The equilibrium constant for interface formation between folded repeats *i* and *j* is

$$\tau_{ij} = e^{-(\Delta G_{ij,H2O}) / RT} \quad (3)$$

This expression assumes the interface free energy is independent of guanidine (and TMAO).

For one- and two-repeat constructs, the partition functions are simple enough to write out explicitly. The partition function for single-repeat constructs (D,E,F,G,H,I,J) is that of a simple two-state protein:

$$\rho_i = 1 + \kappa_i \quad (4)$$

The partition function for two-repeat constructs (DE, EF, FG, GH, HI, IJ) contains terms for five-states:

$$\rho_{ij} = 1 + \kappa_i + \kappa_j + \kappa_i \kappa_j + \kappa_i \kappa_j \tau_{ij} \quad (5)$$

The fourth term, $\kappa_i \kappa_j$, corresponds to a “fractured state” in which both repeats are folded but no interface is formed[‡].

The partition function for the seven-repeat construct DEFGHIJ is generated using seven 2x2 correlation matrices (W_j):⁶

$$\rho_{DEFGHIJ} = [0 \quad 1] \times W_D \times W_E \times W_F \times W_G \times W_H \times W_I \times W_J \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (6)$$

The left and right columns of W_j correspond to repeat j being folded and unfolded. The top and bottom rows correspond to the adjacent $j-1$ being folded and unfolded:

$$W_j = \begin{bmatrix} (\kappa_j \tau_{ij} + \kappa_j) & 1 \\ \kappa_j & 1 \end{bmatrix} \quad (7)$$

The zero in the row vector eliminates states where the D repeat forms an interface with the C repeat, which is not included in any of the constructs studied here. Multiplying out the full matrix product in equation 6 yields a scalar expression with 610 terms.⁶

For single-repeat constructs, the fraction of folded repeats (f_j) is

$$f_{f,i} = \frac{\kappa_i}{\rho_i} \quad (8)$$

For paired-repeat constructs, the fraction of folded repeats is

$$f_{f,ij} = \frac{0.5\kappa_i + 0.5\kappa_j + \kappa_i \kappa_j + \kappa_i \kappa_j \tau_{ij}}{\rho_{ij}} \quad (9)$$

For the DEFGHIJ construct the fraction of folded repeats is

$$f_{f,DEFGHIJ} = \frac{1}{7\rho_{DEFGHIJ}} \sum_{j \in \{D,E,F,G,H,I,J\}} \kappa_j \frac{\partial \rho_{DEFGHIJ}}{\partial \kappa_j} \quad (10)$$

These expressions for the fraction of folded repeats can be used to describe guanidine-induced unfolding transitions with the fitting equation

$$Y_{obs} = Y_N f_f + Y_D (1 - f_f) \quad (11)$$

Y_N and Y_D are linear baselines for the native and denatured species. Together, the two baselines for a single unfolding transition have four parameters (slope and intercept for each baseline). For fits of multiple transitions (either different constructs or replicates of the same construct), these four baseline parameters are local to each transition. For single-repeat constructs, the fitting function also has two global parameters, ΔG_{H2O} and m . For the two-repeat constructs DE, EF, and HI (equations 9, 11) the fitting function has five global parameters (two ΔG_{H2O} , two m -values, and one $\Delta G_{ij,H2O}$ for interface formation). J and IJ fitting functions contain an additional global parameter, $m_{J, TMAO}$. The fitting equation (11) assumes that the spectroscopic contribution of each repeat is the same.

For the two-repeat constructs FG and GH, we include an intermediate baseline (Y_I) for partly-folded species where only a single repeat is folded. Although our analysis below indicates that the G-repeat remains folded in the partly folded state, we allow contributions from both partly folded states for generality (Figure S3). With this modification, the fitting equation for FG and GH becomes

$$Y_{obs} = Y_N \frac{\kappa_i \kappa_j + \kappa_i \kappa_j \tau_{ij}}{\rho_{ij}} + Y_I \frac{\kappa_i + \kappa_j}{\rho_{ij}} + Y_D \frac{1}{\rho_{ij}} \quad (12)$$

The first term in equation 12 represents the states where both repeats are folded: one where the interface is formed and another where the interface is fractured. These states are treated as spectroscopically identical. The second term represents the two partially-folded intermediate states. Equation 12 has six local baseline parameters for each melt, along with five global parameters.

NMR Spectroscopy. A preliminary backbone ^1H - ^{15}N HSQC spectrum at 25°C for G, and ^1H - ^{15}N TROSY spectra for FG (40°C) and GH (25°C) were collected to assess chemical shift dispersion and peak intensities. TROSY spectra provided additional line narrowing for the FG and GH constructs even though the proteins were not ^2H labeled. Assignments for backbone ^{15}NH , $^{13}\text{C}\alpha$, and $^{13}\text{C}\beta$ groups in G, FG, and GH were made using standard triple resonance experiments including HNCACB, CBCA(CO)NH, and HN(CA)NH.^{17–19} Backbone carbonyl chemical shifts were assigned using HNCO spectra.²⁰ Spectra for G and FG were collected on a Bruker Avance II 600 MHz-spectrometer. Spectra for GH were collected on a Bruker Avance Neo 800 MHz spectrometer. Additional (H)CC(CO)NH-TOCSY and H(CCCO)NH-TOCSY TROSY spectra of G were acquired on the Neo 800 MHz spectrometer for side chain assignment.²¹

Two 3D NOESY spectra of G were acquired on the Neo 800 MHz spectrometer to generate ^1H - ^1H distance constraints for structure validation. The first was a ^{15}N resolved and ^{13}C selected $^1\text{H}\{\text{C}\}$ - $^1\text{H}\{\text{N}\}$ ^{13}C -HMQC NOESY ^{15}N -HSQC 3D spectrum with a 120 ms NOESY mixing time. The second was a band-selective ^{13}C resolved $^1\text{H}\{\text{C}\}$ - $^1\text{H}\{\text{C}\}$ NOESY spectrum that was centered on the methyl region, with a 150 ms NOESY mixing time. This spectrum was referenced with both band- and methyl-selective ^1H - ^{13}C gHSQC spectra that were assigned using side chain TOCSY spectra of G.

All NMR data were processed using NMRPipe on NMRbox.^{22,23} Resonances were assigned using either CARA (cara.nmr.ch) or CCPN.²⁴ Spectra were plotted using NMRFAM-Sparky.²⁵

RESULTS

Topo V Constructs and Repeat Definitions. To separate the Topo V (HhH)₂ array into a nested set of constructs for nearest-neighbor folding studies we selected a series of

[‡] The fractured state, which is not included in a conventional Ising model, is necessary to describe folded repeats separated by an unstable interface.

repeats that can be expressed and studied in isolation and as tandem pairs. In various crystal structures without bound DNA, (5HM5, 2CSB, and 4GFJ), the D-F and H-J repeats are well resolved, and lack contacts to the N-terminal topoisomerase domain. Thus, we expressed individual repeats starting with repeat D (the fourth repeat; Figure 1A) and ending with repeat J (the tenth repeat; Figure 1A, Table S1). The A-C repeats make extensive contacts with the N-terminal topoisomerase domain, and were excluded from this study. The K and L repeats expressed poorly and were relatively insoluble, and were also excluded from this study. We also made constructs of adjacent pairs of repeats to determine interfacial free energies (DE, EF, FG, GH, HI, and IJ; Table S1).

To define repeat boundaries, we used a previously described structure-based MSA as a starting point.¹¹ To avoid helix fraying, we include three residues at the C-terminus of each construct that come from the following repeat. Each construct also begins with the sequence GGSW as a result of TEV cleavage. The tryptophan allows for quantification of protein concentrations. To avoid potential helix fraying at the N-terminus, we tested the effects of including three residues from the preceding repeat on stability (data not shown). For some constructs, the additional three residues had no effect on stability, whereas for others (particularly for the H repeat) this N-terminal extension resulted in a modest change in stability; in such cases, we included these residues. Additionally, based on stability increase and TALOS+ analysis, we found that the N-terminal helix of the H repeat (QERSEE) extends into sequence that was previously attributed to the G repeat¹¹, and have included this sequence at the N-terminus of the H repeat (Figure 1C).

Circular Dichroism and Stability Measurements. To assess whether our single- and paired-repeat constructs are folded, we collected far-UV CD spectra. The spectra of all constructs are characteristic of folded α -helical proteins with minima at 222 and 208nm (Figure S1A). To determine equilibrium stabilities, we collected guanidine-induced unfolding transitions by monitoring CD at 222 nm as a function of guanidine hydrochloride (GdnHCl, hereafter “guanidine”) concentration (Figure S1B). For each construct, transitions were generated both by increasing and decreasing denaturant concentrations (“forward” and “reverse” titrations) to test for reversibility. Most single-repeat constructs displayed a single, well-resolved sigmoidal folding transition (Figure S1B). Because the unfolding midpoints of G-containing constructs (G, FG, GH) were high, they lacked well-resolved denatured baselines when titrated in the forward direction, so we only analyzed the reverse titrations of these constructs (Figure S1B,D). Because the midpoint of the J repeat was low, its native baseline was not well resolved even in forward titrations (Figure S1C). To better define the native baseline of the J repeat, we collected additional guanidine titrations in the presence of the stabilizing osmolyte TMAO, which better constrain ΔG_J and the corresponding m-value (Figure S2).

With the exception of GH, the paired repeat constructs also displayed a single well-resolved sigmoidal folding transition, although the transitions vary in steepness. The DE, FG, and IJ pairs have somewhat broad transitions compared to EF and HI (Figure S1B). In contrast, the GH pair shows clear deviation from a two-state folding mechanism, with separate transitions at 3.8 and 6.4 M guanidine (Figures S1D and S3).

To ensure that parameters obtained from denaturant unfolding report on thermodynamic stability despite the long extrapolation, we also collected temperature-induced unfolding transitions of six individual repeats (Figure S4A) and three pairs of repeats (Figure S4B,C,D). As expected for a protein from a thermophilic organism, the repeats from Topo V have high thermostability. The G repeat, which requires the highest concentration of guanidine to unfold, does not melt at 92°C with 2M guanidine. The least thermostable repeat, F, also unfolds at the lowest concentration of guanidine (C_m 2.5M guanidine, Figure S1B).

The DE thermal unfolding transition shows no stability increase compared to D and E, matching results from guanidine-induced transitions (Figures S1D and S4B). However, the EF and HI thermal transitions show significant stability increase compared to single-repeats, also matching guanidine-induced transitions. (Figure S4C,D). Overall, the matching stability profiles of the guanidine- and thermal unfolding transitions confirm that the guanidine-induced transitions faithfully report on intrinsic and interfacial stabilities despite the sometimes long extrapolations from high guanidine concentration.

Ising Analysis and Local Stabilities. To resolve folding free energies into intrinsic and interfacial terms, we globally fitted the guanidine-induced folding transitions of all single repeats and tandem pairs with a modified fractured-states one-dimensional Ising model (see Methods). The fractured-states model describes the folding transitions well, capturing both the cooperative and noncooperative transitions of the tandem constructs (Figures 2 and S2). By simultaneously fitting all the transitions of individual and paired-repeats, including TMAO-stabilized J transitions, we are able to tightly constrain both the intrinsic (ΔG_i , m_i) and interfacial (ΔG_{ij}) Ising parameters with narrow confidence intervals (Table 1).

From the global fit, we find that six of seven individual repeats are intrinsically stable, as expected from their CD spectra and temperature melts. The stabilities of these six repeats vary significantly, with folding free energies ranging from -2.46 (ΔG_E) to -11.52 kcal mol⁻¹ (ΔG_G). J is intrinsically unstable (ΔG_J of +0.24 kcal mol⁻¹ in the absence of TMAO). The observation that G is significantly more stable than the other repeats is very surprising given that G only displays interpretable electron density in a crystal structure complex with DNA, and is otherwise missing in the unbound structure.^{11,12} The high intrinsic stability of G in the absence of DNA indicates that G folding does not rely on a folding-upon-binding mechanism, and also suggests that the lack of electron density of G may be the result of orientational heterogeneity in the crystal lattice rather than unfolding of the G repeat.

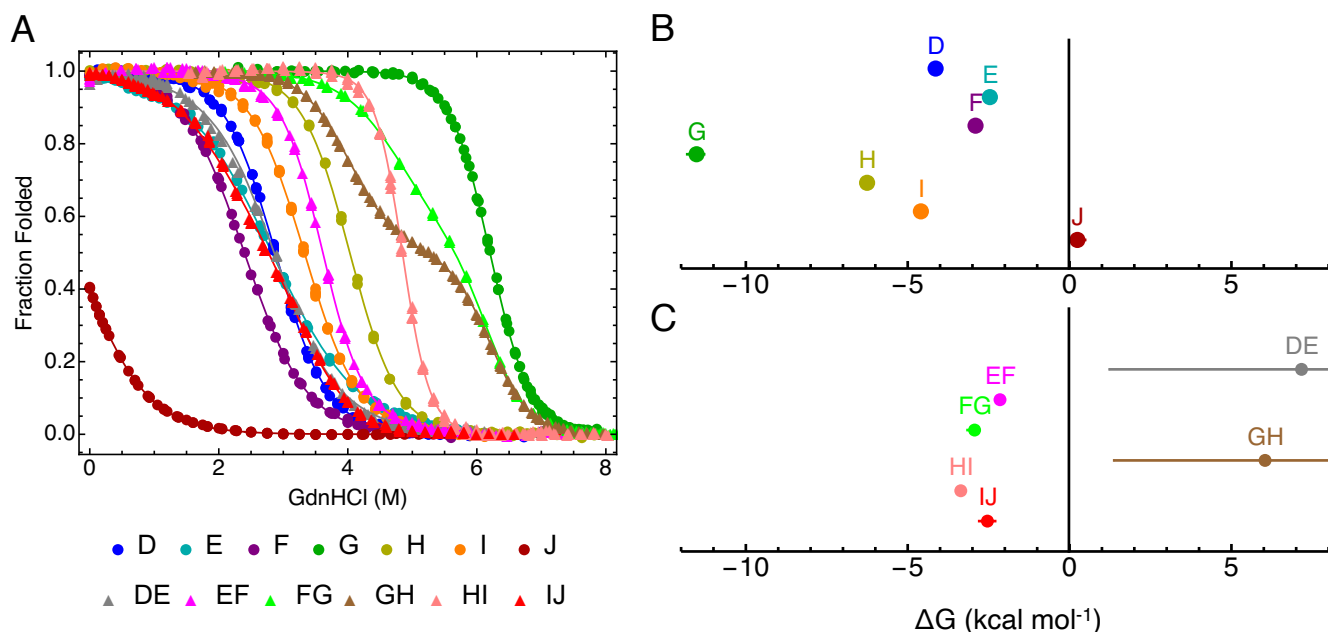


Figure 2. Stability and coupling of the Topo V (HhH)₂ repeats. (A) Guanidine-induced folding and unfolding transitions of the single and paired Topo V (HhH)₂ repeats monitored by CD at 222nm at 20°C. Data are fit globally with a fractured-states Ising model (see methods), and include guanidine-transitions of the J repeat stabilized with TMAO (Figure S2) to better constrain the parameters. (B) Intrinsic folding free energies and (C) interfacial free energies of the (HhH)₂ repeats from the global fit. Lines represent a 95% confidence interval from bootstrapping. Confidence intervals for DE and GH are unbounded from above, extending far beyond the right edge of the plot.

The six interfacial free energies also vary significantly. Four interfaces are stabilizing, with ΔG_{ij} values range from -3.36 to -2.15 kcal mol⁻¹ (Table 1). Two interfaces (DE and GH) are unstable, with positive ΔG_{ij} values. With the combination of intrinsically stable repeats and stabilizing interfaces between most repeats, the array of (HhH)₂ repeats in Topo V can indeed be classified as a semi-autonomous repeat protein (SARP).⁶

The instability of the DE and GH interfaces requires a “fractured state” partition function in which these two repeat pairs fold but do not interact (Figure S5B,C). Although the free energies of the stable interfaces are tightly bounded by the fractured states model (with 95% confidence intervals ranging from 0.10 to 0.49 kcal mol⁻¹), fitted free energies for the DE and GH interfaces can only be determined as lower bounds of around +1 kcal per mole. ΔG_{DE} and ΔG_{GH} values of +10 or even +100 kcal mol⁻¹ fit the folding transitions of DE and GH equally well (Figure S5E,F). Thus, the fitted values of ΔG_{DE} and ΔG_{GH} should not be interpreted beyond being positive.

The Folding Transition and Partly-Folded States of DEFGHIJ. If the nearest-neighbor model provides an accurate description of the Topo V (HhH)₂ array, the free energies and guanidine dependences determined by Ising analysis of repeat pairs should accurately predict the guanidine-induced unfolding transition of the seven-repeat DEFGHIJ (D-J) array. Using the matrix method (equations 6 and 10), we generated a seven-repeat partition function that is parameterized with free energies from the pair analysis. This partition function, which contains 610 distinct states, predicts a complex multistate guanidine-induced unfolding transition

(Figure 3A, dashed line). Analysis of the predicted unfolding of individual repeats from the D-J array (Figure 3A, colored lines) reveals that Topo V D-J is gradually unfolded from its two ends. Between 0.5 and 4 M guanidine, repeats J and D are predicted to unfold in an uncorrelated manner. Although these transitions are centered at different guanidine concentrations, they do not appear as separate transitions in the overall predicted unfolding of D-J. At 5 M guanidine, and H and I are predicted to unfold cooperatively; E is predicted to unfold independently at the same guanidine concentration, leaving F and G folded. At 6 M guanidine, F is predicted to unfold, leaving a lone folded G repeat.

To test our prediction for the global unfolding of the D-J array, we expressed and purified the seven-repeat D-J construct (Table S1). As predicted, the guanidine-induced unfolding transition of D-J is broad and multiphasic (Figure 3B, filled circles). To quantitatively test whether the predictions from the seven-repeat partition function match the experimentally observed unfolding transition of D-J, we fitted the data using a model based on our partition function where the only adjustable parameters are the four parameters describing the fully folded and unfolded baselines. This model assumes that each repeat has the same contribution to the overall CD signal, and uses fixed intrinsic and interfacial energies and m-values determined from analysis of paired repeats (Table 1). The predicted unfolding transition describes the data well (Figure 3B, fitted curve). The close agreement between the D-J folding transition and the baseline-fitted curve shows that the Ising model parameters derived

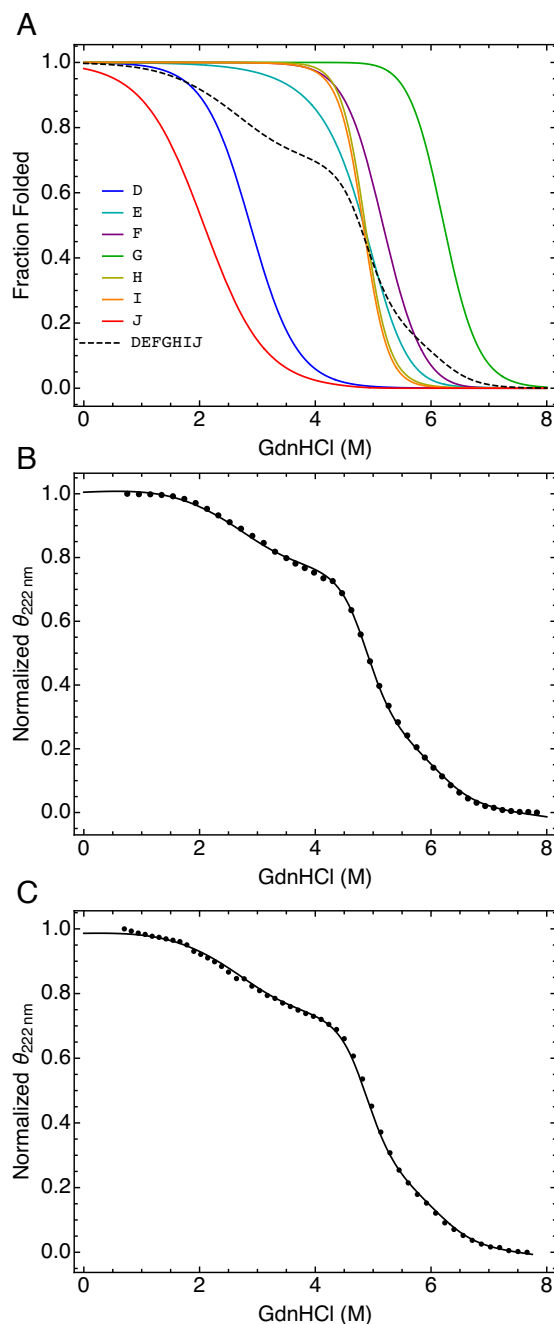


Figure 3. Global folding of the DEFGHIJ (HhH)₂ array. (A) Unfolding transitions predicted from the D-J partition function generated using parameters from the global Ising fit. The solid lines show the fraction of folded repeats in the seven-repeat construct. The dashed line shows the predicted folding transition of the seven-repeat DEFGHIJ construct using fitted Ising parameters (Table 1). (B) Guanidine-induced folding transition of DEFGHIJ monitored by CD at 20 °C. The solid line is generated from the D-J partition function, adjusting only the fully-folded and fully-unfolded baseline parameters. (C) Folding transition of a 1:1:1 mixture of D, EFG, and HIJ. Solid line is generated from the D-J partition function as in panel B.

Table 1. Global thermodynamic parameters from fractured-state Ising fit

Parameter	Best fit value	Bootstrap parameters ^c Lower CI, Upper CI
^a Intrinsic Folding Free Energies		
ΔG_D	-4.14	-4.22, -4.05
ΔG_E	-2.46	-2.57, -2.35
ΔG_F	-2.91	-3.00, -2.80
ΔG_G	-11.52	-11.81, -11.29
ΔG_H	-6.26	-6.38, -6.12
ΔG_I	-4.59	-4.69, -4.51
ΔG_J	+0.24	0.00, +0.48
^a Interfacial Folding Free Energies		
ΔG_{DE}	+7.17	+1.24, +41.60
ΔG_{EF}	-2.15	-2.20, -2.10
ΔG_{FG}	-2.93	-3.15, -2.86
ΔG_{GH}	+6.04	+1.38, +23.49
ΔG_{HI}	-3.36	-3.41, -3.31
ΔG_{IJ}	-2.54	-2.79, -2.30
^b Denaturant and Osmolyte Dependencies		
m_D	1.44	1.41, 1.46
m_E	0.87	0.84, 0.91
m_F	1.21	1.17, 1.25
m_G	1.86	1.82, 1.91
m_H	1.55	1.52, 1.58
m_I	1.38	1.35, 1.41
m_J	1.11	1.08, 1.14
$m_{J, \text{TMAO}}$	-1.08	-1.22, -0.95

^a ΔG values in kcal mol⁻¹, where negative values are stabilizing.

^b m values in kcal mol⁻¹ M⁻¹ GdnHCl or TMAO. ^cValues are from 2000 bootstrap iterations. CI, approximate lower and upper confidence intervals defined as the bootstrap mean minus and plus two standard deviations.

from single and paired repeat constructs, which lack non-nearest neighbor repeats, give a very good description of the seven-repeat construct, which includes numerous non-nearest neighbor repeats. This agreement provides a strong justification for the nearest-neighbor approximation.

The Structure of the G Repeat. The observation that G is the most stable repeat in the D-J (HhH)₂ array suggests that G is well-folded, despite being crystallographically unresolved in the absence of DNA.¹¹ To confirm this, we used heteronuclear NMR to determine the structure and probe the dynamics of the isolated G repeat. Because the structure of G may be influenced by its neighboring repeats, we also explored the structure and dynamics of the FG and GH constructs. For G and GH, we used constructs that did not contain EIK at the N-terminus (Table S1).

The cross-peaks in two dimensional ¹⁵N-¹H HSQC spectra of G, FG, and GH are well dispersed and have uniform intensity (Figure 4), suggesting that the G-repeat is folded. To map the secondary structural elements in the G repeat, we used standard triple-resonance experiments to make backbone resonance assignments.¹⁷⁻²¹ For isolated G we were able to assign 79 of 81 non-proline backbone resonances. For FG we assigned 124 of 130 non-proline backbone resonances, and for GH we assigned 132 of 134 non-proline backbone resonances. To determine the locations of helices, we used backbone H, CO, C_α, N, and sidechain C_β chemical shifts as inputs for the TALOS-N package (Figure S6).²⁶ For F and H, the TALOS-

N prediction agrees well with secondary structure in various crystal structures.^{10,11} For G, TALOS-N predicts five distinct helices, as expected for an (HhH)₂ repeat; these five helices span the same residues in all three constructs (G, FG, and GH). However, the fifth helix of G is significantly longer than the corresponding helices in the other (HhH)₂ repeats (D-F, H-J).

To model the packing of these five helices, we used chemical shift information to generate 3,000 models of G using CS-Rosetta.²⁷ We aligned the 10 lowest energy structures and found that they adopt a canonical (HhH)₂ fold, albeit with a C-terminally extended fifth helix (Figure 5). A similar structure is obtained using AlphaFold (Figure S7A,B).²⁸

To validate the CS-Rosetta model, we obtained two ¹H-¹H NOESY spectra using ¹³C- and ¹⁵N-edited NOE experiments with the isolated G repeat. An ¹⁵N-edited NOESY was used to identify NOEs between backbone amide and aliphatic protons. A ¹³C-edited NOESY was used to identify NOEs among methyl and aliphatic protons. From these spectra, we were able to assign 44 long-range NOEs (Table S2). We then examined the distances between heavy atoms with observed ¹H-¹H NOEs (Figure S8A) and found that the atoms in the large majority of these pairs are within NOE distance. The median distance of the NOE-detected pairs is 4.5 Å. Although five NOEs are observed from pairs with distances of 8 – 12 Å, four pairs all come from the same pair of residues (L639 and R679; L26 and R66 in our construct). For comparison, the median distance between all detectable heavy atom pairs in the CS-Rosetta models is 15.6 Å, emphasizing the close proximity of the observed NOE pairs (Figure S8C). Our CS-Rosetta model of G is quite similar to the DNA-bound crystal structure (including the extended C-terminal helix, Figure S7C), further suggesting that the lack of density in the absence of DNA is not a result of the G-repeat being unfolded in the unbound state.^{11,12}

The Structure of the FG and GH Interfaces. One possible explanation for the lack of electron density for the G-repeat in crystal structures lacking DNA is that G is folded but is in multiple orientations with respect to the rest of the Topo V array. This "folded but disordered" model would be promoted by disordered interfaces between the G-repeat and its neighbors, consistent with our observation that ΔG_{GH} is positive, but not with the negative value of ΔG_{FG} .

To examine the structural integrity of the FG and GH interfaces, we determined chemical shifts of the FG and GH pairs and compared them to those of the isolated G-repeat. (Figure S9, purple). The formation of a rigid interface should generate significant chemical shift changes involving interfacial residues. Indeed, addition of the F-repeat results in significant chemical shift perturbations (CSPs) in the G-repeat; these perturbations are localized primarily in helices 1 and 3, and to a lesser extent helix 5. In contrast, addition of the H-repeat results in comparatively minor CSPs in G, except at the C-terminus of G (Figure S8, yellow). These C-terminal residues include a QERSEE sequence, which extends the N-terminal helix of H in the crystal structure. Since deleting these residues from our H repeat construct increases ΔG_H by 2.5 kcal/mol, we consider them part of the H-repeat. Overall,

the CSP results are consistent with the interfacial free energies, which indicate that the FG interface is stable but the GH interface is not.

To further probe the integrity of the FG and GH interfaces, we measured backbone dynamics in the FG and GH pairs with three NMR experiments that access structural heterogeneity on different timescales: ZZ-exchange, CPMG, and Heteronuclear-NOE.²⁹⁻³¹ We find that the interfacial residues in GH show considerable dynamics on the heteronuclear ¹H, ¹⁵N-NOE timescale (ps-ns timescale), whereas those in the FG interface are rigid (Figure S10). In contrast, we see no evidence for dynamics in either construct on the slower ZZ-exchange and CPMG timescales, aside from some slow (>500ms) dynamics at the C-terminus of the GH construct which may result from *cis-trans* isomerization of P750 (P137 in our construct). These observations are consistent with a model in which the G repeat is folded but is orientationally disordered relative to the H repeat.

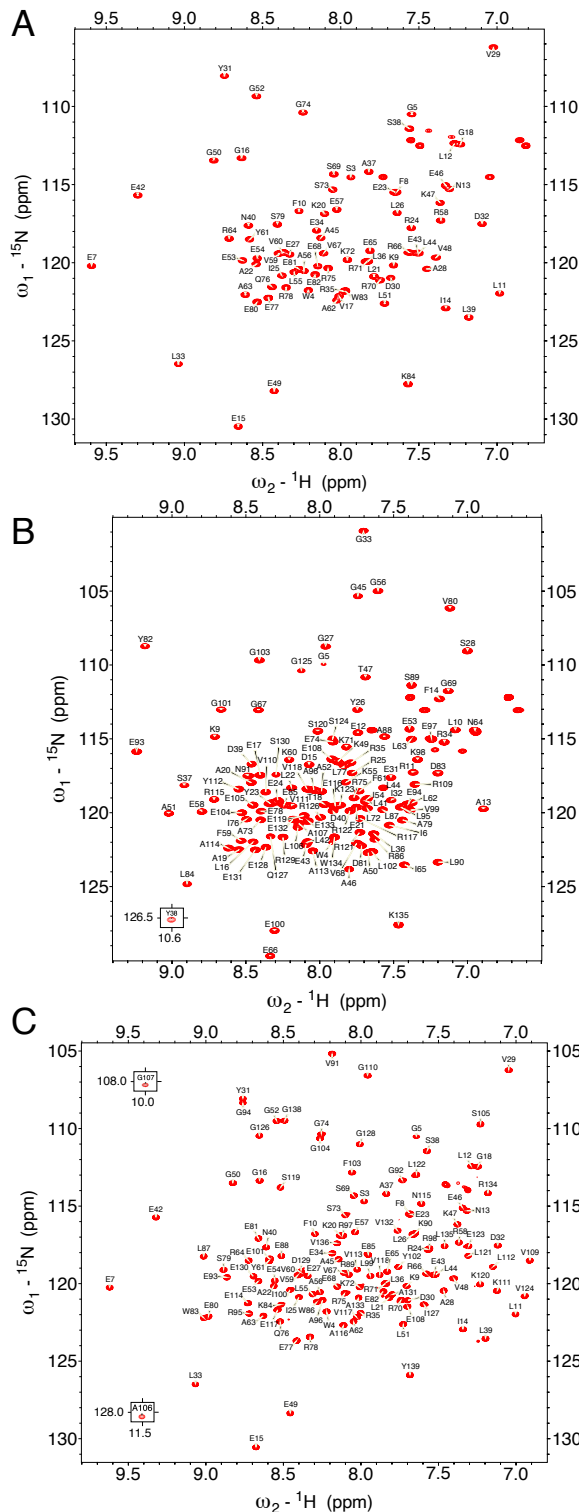


Figure 4. ^{15}N - ^1H NMR correlation spectra of G-repeat constructs (A) ^{15}N - ^1H HSQC spectrum of ^{15}N -labeled G at 600 MHz and 25°C. (B) ^{15}N - ^1H TROSY spectrum of ^{15}N -labeled FG at 600 MHz and 40°C. (C) ^{15}N - ^1H TROSY spectrum of ^{15}N -labeled GH at 800 MHz and 25°C. Inserts are three peaks that are downfield off the spectra in ^1H . Samples contained 50 mM sodium phosphate, 50 mM NaCl, pH 6.4.

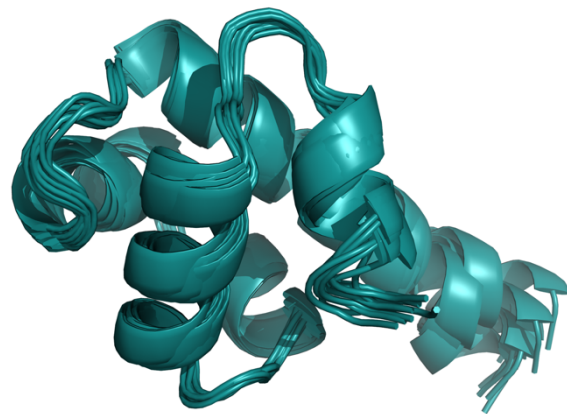


Figure 5. CS-Rosetta model of G. Alignment of the ten lowest energy models generated by CS-Rosetta using chemical shifts as constraints. Disordered tails have been trimmed from the models.

DISCUSSION

Cooperativity is a defining property in protein folding.^{3-6,30} Here we have used a modified Ising model to quantify the folding cooperativity of a series of $(\text{HhH})_2$ repeats from Topo V. Since these repeats remain folded in isolation, Ising analysis can be applied to single- and paired-repeat constructs strongly constraining the intrinsic and interfacial folding free energies of sequence-variable repeats. Intrinsic folding free energies vary significantly, ranging from +0.24 kcal mol⁻¹ (for the unstable J repeat) to -11.52 kcal mol⁻¹ (for the highly stable G-repeat). Free energies for the four stable interfaces are more tightly clustered, ranging from -2.15 kcal mol⁻¹ to -3.36 kcal mol⁻¹. Two interfaces (DE and GH) are unstable outliers in the otherwise cooperative $(\text{HhH})_2$ array.

In addition to providing a quantitative measure of the degree of folding cooperativity between repeats, the energy parameters from the Ising analysis can be used to predict a guanidine-induced unfolding transition of the D-J array, which agrees with experiment. These parameters also allow us to quantify the populations of partly folded states as a function of guanidine concentration (Figure 3A). As guanidine concentration increases, the repeats unfold from the outside inward. D and J unfold non-cooperatively at low guanidine concentrations, giving rise to a conformation where the E through I repeats are folded at 4M guanidine. At higher concentrations, repeats H and I unfold cooperatively, while other repeats fold non-cooperatively (E, then F, then G). By 6.5M guanidine, only the G repeat remains folded (with a population of about 50%). Overall, the DEFGHIJ unfolding transition is very multistate, especially compared to highly cooperative SARPs such as ankyrin.^{2,3}

Fractured-States Ising Model for SARPs. The classic version of an Ising model assumes that if two neighboring repeats i and j are both folded, they must form an interface regardless of the value of ΔG_{ij} . This assumption works well with NARPs because the interfaces are stabilizing ($\Delta G_{ij} < 0$),

and are thus expected to form between pairs of folded repeats. For SARPs, this assumption is not appropriate. If the repeats are intrinsically stable but don't form a stable interface ($\Delta G_{ij} > 0$ as with DE and GH), the lowest energy state will be that with both repeats folded but lacking an interface, a state that is not included in the classic Ising model. Analysis of the DE and GH pairs with a classic Ising model gives values of ΔG_{DE} and ΔG_{GH} of about $+0.5 \text{ kcal mol}^{-1}$ with confidence intervals of about $0.1 \text{ kcal mol}^{-1}$. The overall fit is slightly better (RMSD 2.79×10^{-5} vs 3.35×10^{-5}) than with the fractured states model. However, the fitted instability of the DE interface in the classic Ising model predicts unfolding of the E repeat below 1M guanidine, which is inconsistent with the unfolding transition of the isolated E repeat. In the fractured states fit, the E repeat remains folded.

The fact that the fractured states model can accommodate unstable interfaces is a strength, especially for the heterogeneous stability distribution in the Topo V (HhH)₂ array. However, a limitation of the model is that the confidence intervals associated with unstable interfaces are very broad. Here, the upper limits for ΔG_{DE} and ΔG_{GH} are unbounded, although the lower limits are well defined ($+1.2$ and $+1.4 \text{ kcal mol}^{-1}$, respectively). In contrast, the confidence intervals associated with stable interfaces are narrow and normally distributed, even though fractured states are included for all pairs of repeats. For those pairs, the fractured-states model is equivalent to the traditional Ising model.⁶

Stability Islands, Hinges, and Implications for DNA Binding. A recent crystal structure of Topo V in complex with DNA reveals that (HhH)₂ repeats A-J wrap around a full turn of duplex DNA (Figure 1B).¹² Wrapping ten (HhH)₂ domains around DNA would presumably be kinetically slow and energetically unfavorable, especially if the repeats and interfaces form an extended rigid array. Previously, the G repeat has been proposed as a possible hinge to facilitate wrapping based on the lack of electron density in the absence of DNA.¹¹ Indeed, the orientation of the HIJ three-repeat array to the rest of Topo V is significantly different between the two crystal structures (DNA-bound and DNA-unbound).^{11,12} From our thermodynamic studies, we propose that the hinge is not G itself, which is strongly folded and paired with the F-repeat in solution, but the junction between the G and H repeats. Consistent with this interpretation, AlphaFold models the FG interface as rigid and the GH interface as disordered, placing HIJ in a different orientation from the crystal structure (Figure S7B). These different orientations result from backbone heterogeneity in the loop between G and H, which our NMR relaxation studies show to be flexible on the ns-ps timescale (Figure S10C). The G and H repeats adopt a specific extended conformation in the DNA-bound crystal structure, suggesting that their relative orientations become fixed upon DNA binding.

Our thermodynamic studies identify a second hinge at the junction between the D and E repeats. This second hinge is not expected from the structure: both repeats are fully resolved, and form an interface that buries a similar amount of total surface area as the other interfaces in Topo V, including the highly cooperative HI interface. It is possible that, unlike

the junction between the G and H repeats, which contain a large helical insert that may prevent close approach of the G and H repeats, the very short linker between the D and E repeats limits the number of alternative conformations that the two repeats can adopt. In such a scenario, the DE interface would form in the crystal structure despite a low stability. One potential source of instability is electrostatic repulsion: the DE interface has three pairs of side chains of like charges within 5 Å (R495:K535, D509:E522, and K516:R536), whereas the EF and HI interfaces only have one such interaction. In addition, the DE interface has fewer interactions between charged side chains and uncharged polar groups (6 within 4.5 Å) than EF and HI (8 and 15, respectively).

Together, the DE and GH hinges define three or more independent “islands” of coupled (HhH)₂ repeats: (ABC)D, EFG, HIJ(KL). This model is thermodynamically equivalent to a system where the islands are not tethered by peptide bonds, but are on completely separate polypeptide chains. To test this three-island model, we prepared the three-repeat constructs EFG and HIJ, and mixed them with repeat D in a 1:1:1 molar ratio. This mixture of three proteins follows the same unfolding transition as the D-J seven-repeat construct (Figure 3C), demonstrating that in the seven-repeat folding transition these repeats behave as if they are on separate polypeptides. This highly heterogeneous view of cooperativity, where segments of internally coupled repeats are uncoupled from one another is novel, and has not been observed in previous Ising analyses of tandem repeat protein folding, which has been restricted to identical consensus repeats.

Though heterogeneous cooperativity may simply be a result of sequence drift among repeats, it is possible that heterogeneity may contribute to biological function. Since the DE and GH hinges in TopoV do not affect the folding stability of the molecule (Figure 3B, C), they may contribute to one or more aspects of topoisomerase and/or repair activity. We suspect that they facilitate rapid and stable DNA binding. Whereas a rigid, elongated array of 10 repeats (12 if K and L are included) would be energetically costly to distort, an array of 10 (or 12) noninteracting repeats would possess considerable conformational freedom, which may slow binding and entropically destabilize the bound state. Connecting three or four rigid multi-repeat segments with flexible hinges could provide the flexibility to wrap all the way around the DNA without sacrificing too much conformational entropy. This type of heterogeneous coupling in repeat proteins in which subsets of adjacent repeats act as independent structural units is novel, and may provide a general mechanism whereby long arrays of sequence-variable repeats can recognize extended substrates quickly and with high affinity.

SUPPORTING INFORMATION

The Supporting Information includes details on the constructs used in this study (Table S1), long range NOEs in G (Table S2), CD spectra and raw melt data of single and paired constructs (Figure S1), GdnHCl and TMAO transitions of J (Figure S2), transitions and fitted baselines of FG and GH (Figure S3),

temperature induced transitions of single and paired constructs (Figure S4), populations of partially folded states in GH, DE, and FG (Figure S5), TALOS-N predictions of helices in FG, G, and GH (Figure S6), AlphaFold2 models (Figure S7), distance histograms from CS-Rosetta and AlphaFold2 models (Figure S8), chemical shift perturbations in G (Figure S9), heteronuclear-NOE values for FG, G, and GH (Figure S10). Chemical shifts are in the process of deposition in the BMRB.

REFERENCES

- (1) Tripp, K. W.; Barrick, D. The Tolerance of a Modular Protein to Duplication and Deletion of Internal Repeats. *Journal of Molecular Biology* 2004, 344 (1), 169–178. <https://doi.org/10.1016/j.jmb.2004.09.038>.
- (2) Mello, C. C.; Barrick, D. An Experimentally Determined Protein Folding Energy Landscape. *Proc Natl Acad Sci U S A* 2004, 101 (39), 14102–14107. <https://doi.org/10.1073/pnas.0403386101>.
- (3) Aksel, T.; Majumdar, A.; Barrick, D. The Contribution of Entropy, Enthalpy, and Hydrophobic Desolvation to Cooperativity in Repeat-Protein Folding. *Structure* 2011, 19 (3), 349–360. <https://doi.org/10.1016/j.str.2010.12.018>.
- (4) Aksel, T.; Barrick, D. Direct Observation of Parallel Folding Pathways Revealed Using a Symmetric Repeat Protein System. *Biophysical Journal* 2014, 107 (1), 220–232. <https://doi.org/10.1016/j.bpj.2014.04.058>.
- (5) Geiger-Schuller, K.; Barrick, D. Broken TALEs: Transcription Activator-like Effectors Populate Partly Folded States. *Biophysical Journal* 2016, 111 (11), 2395–2403. <https://doi.org/10.1016/j.bpj.2016.10.013>.
- (6) Petersen, M.; Barrick, D. Analysis of Tandem Repeat Protein Folding Using Nearest-Neighbor Models. *Annual Review of Biophysics* 2021, 50, 245–265. <https://doi.org/10.1146/annurev-biophys-102220-083020>.
- (7) Shao, X.; Grishin, N. v. Common Fold in Helix-Hairpin-Helix Proteins. *Nucleic Acids Research* 2000, 28 (14), 2643–2650. <https://doi.org/10.1093/nar/28.14.2643>.
- (8) Taneja, B.; Patel, A.; Slesarev, A.; Mondragón, A. Structure of the N-Terminal Fragment of Topoisomerase V Reveals a New Family of Topoisomerases. *EMBO Journal* 2006, 25 (2), 398–408. <https://doi.org/10.1038/sj.emboj.7600922>.
- (9) Rajan, R.; Taneja, B.; Mondragón, A. Structures of Minimal Catalytic Fragments of Topoisomerase V Reveals Conformational Changes Relevant for DNA Binding. *Structure* 2010, 18 (7), 829–838. <https://doi.org/10.1016/j.str.2010.03.006>.
- (10) Rajan, R.; Prasad, R.; Taneja, B.; Wilson, S. H.; Mondragón, A. Identification of One of the Apurinic/Apyrimidinic Lyase Active Sites of Topoisomerase v by Structural and Functional Studies. *Nucleic Acids Research* 2013, 41 (1), 657–666. <https://doi.org/10.1093/nar/gks1017>.
- (11) Rajan, R.; Osterman, A.; Mondragón, A. Methanopyrus Kandleri Topoisomerase v Contains Three Distinct AP Lyase Active Sites in Addition to the Topoisomerase Active Site. *Nucleic Acids Research* 2016, 44 (7), 3464–3474. <https://doi.org/10.1093/nar/gkw122>.
- (12) Osterman, A.; Mondragon, A. Structures of Topoisomerase V in Complex with DNA Reveal Unusual DNA Binding Mode and Novel Relaxation Mechanism. *eLife* 11:e72702. DOI: <https://doi.org/10.7554/eLife.72702>.
- (13) Belova, G. I.; Prasad, R.; Kozyavkin, S. A.; Lake, J. A.; Wilson, S. H.; Slesarev, A. I. A Type IB Topoisomerase with DNA

ACKNOWLEDGMENTS

We thank the Johns Hopkins Center for Molecular Biophysics and the Johns Hopkins Biomolecular NMR Center for use of their facilities and instrumentation. This work was supported by National Science Foundation grant 1947561 and by National Institutes of Health grants 1R01-GM068462 and T32-GM008403

Repair Activities. *Proc Natl Acad Sci U S A* 2001, 98 (11), 6015–6020. <https://doi.org/10.1073/pnas.111040498>.

(14) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature Methods* 2009, 6 (5), 343–345. <https://doi.org/10.1038/nmeth.1318>.

(15) Nozaki, Y. The Preparation of Guanidine Hydrochloride. *Methods in Enzymology* 1972, 26 (C), 43–50. [https://doi.org/10.1016/S0076-6879\(72\)26005-0](https://doi.org/10.1016/S0076-6879(72)26005-0).

(16) Baskakov, I.; Wang, A.; Bolen, D. W. Trimethylamine-N-Oxide Counteracts Urea Effects on Rabbit Muscle Lactate Dehydrogenase Function: A Test of the Counteraction Hypothesis. *Biophysical Journal* 1998, 74 (5), 2666–2673. [https://doi.org/10.1016/S0006-3495\(98\)77972-X](https://doi.org/10.1016/S0006-3495(98)77972-X).

(17) Grzesiek, S.; Bax, A. Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J Am Chem Soc* 1992, 114 (16), 6291–6293. <https://doi.org/10.1021/ja00042a003>.

(18) Frueh, D. P.; Arthanari, H.; Koglin, A.; Walsh, C. T.; Wagner, G. A Double TROSY HNCAnH Experiment for Efficient Assignment of Large and Challenging Proteins. *J Am Chem Soc* 2009, 131 (36), 12880–12881. <https://doi.org/10.1021/ja9046685>.

(19) Grzesiek, S.; Bax, A. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *Journal of Magnetic Resonance* (1969) 1992, 99 (1), 201–207. [https://doi.org/10.1016/0022-2364\(92\)90169-8](https://doi.org/10.1016/0022-2364(92)90169-8).

(20) Lay, L.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *Journal of Magnetic Resonance* 1990, 89, 496–514. <https://doi.org/10.1016/j.jmr.2011.09.004>.

(21) Grzesiek, S.; Anglister, J.; Bax, A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in ¹³C/¹⁵N-Enriched Proteins by Isotropic Mixing of ¹³C Magnetization. *Journal of Magnetic Resonance, Series B*. 1993, pp 114–119. <https://doi.org/10.1006/jmr.1993.1019>.

(22) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *Journal of Biomolecular NMR* 1995, 6 (3), 277–293. <https://doi.org/10.1007/BF00197809>.

(23) Maciejewski, M. W.; Schuyler, A. D.; Gryk, M. R.; Moraru, I. I.; Romero, P. R.; Ulrich, E. L.; Eghbalian, H. R.; Livny, M.; Delaglio, F.; Hoch, J. C. NMRbox: A Resource for Biomolecular NMR Computation. *Biophysical Journal* 2017, 112 (8), 1529–1534. <https://doi.org/10.1016/j.bpj.2017.03.011>.

(24) Skinner, S. P.; Fogh, R. H.; Boucher, W.; Ragan, T. J.; Mureddu, L. G.; Vuister, G. W. CcpNmr AnalysisAssign: A Flexible Platform for Integrated NMR Analysis. *Journal of Biomolecular NMR* 2016, 66 (2), 111–124. <https://doi.org/10.1007/s10858-016-0060-y>.

(25) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* 2015, 31 (8), 1325–1327. <https://doi.org/10.1093/bioinformatics/btu830>.

(26) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *Journal of Biomolecular NMR* 2013, 56 (3), 227–241. <https://doi.org/10.1007/s10858-013-9741-y>.

(27) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. Consistent Blind Protein Structure Generation from NMR Chemical Shift Data. *Proc Natl Acad Sci U S A* 2008, 105 (12), 4685–4690. <https://doi.org/10.1073/pnas.0800256105>.

(28) Jumper, J.; Evans, R.; Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021, 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

(29) Li, Y.; Palmer, A. G. TROSY-Selected ZZ-Exchange Experiment for Characterizing Slow Chemical Exchange in Large Proteins. *Journal of Biomolecular NMR* 2009, 45 (4), 357–360. <https://doi.org/10.1007/s10858-009-9385-0>.

(30) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. An Improved ^{15}N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins. *Journal of Physical Chemistry B* 2008, 112 (19), 5898–5904. <https://doi.org/10.1021/jp074793o>.

(31) Kay, L. E.; Torchia, D. A.; Bax, A. Backbone Dynamics of Proteins As Studied by ^{15}N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* 1989, 28 (23), 8972–8979. <https://doi.org/10.1021/bi00449a003>.

For Table of Contents Only:

