Title

Correlating disordered activation domain ensembles with gene expression levels.

Journal

Biophysical Reports, 5(1)

Authors

Flores, Eduardo
Camacho, Aleah
Cuevas-Zepeda, Estefania
et al.

Publication Date

2025-03-12

DOI

10.1016/j.bpr.2024.100195

Peer reviewed

Biophysical Reports

REPORT

Correlating disordered activation domain ensembles with gene expression levels

Eduardo Flores,[1] Aleah R. Camacho,[1] Estefania Cuevas-Zepeda,[1] Mary B. McCoy,[1] Feng Yu,[1,2] Max V. Staller,[3,4,5] and Shahar Sukenik[1,6,*]

[1]Department of Chemistry and Biochemistry, University of California, Merced, Merced, California; [2]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California; [4]Center for Computational Biology, University of California, Berkeley, Berkeley, California; [5]Chan Zuckerberg Biohub–San Francisco, San Francisco, California; and [6]Department of Chemistry, Syracuse University, Syracuse, New York

ABSTRACT Transcription factor proteins bind to specific DNA promoter sequences and initiate gene transcription. These proteins often contain intrinsically disordered activation domains (ADs) that regulate their transcriptional activity. Like other disordered protein regions, ADs do not have a fixed three-dimensional structure and instead exist in an ensemble of conformations. Disordered ensembles contain sequence-encoded structural preferences that are often linked to their function. We hypothesize that this link exists between the structural preferences of AD ensembles and their ability to induce gene expression. To test this, we measured the ensemble dimensions of two ADs, HIF-1α and CITED2, in live cells using fluorescence resonance energy transfer microscopy and correlated this structural information with their transcriptional activity. We find that mutations that expanded the ensemble of HIF-1α increased transcriptional activity, while compacting mutations reduced it, highlighting the critical role of structural plasticity in regulating HIF-1α function. Conversely, CITED2 showed no correlation between ensemble dimensions and activity. Our results highlight a possible link between AD ensemble dimensions and their transcriptional activity, with implications for transcriptional regulation and dysfunction.

WHY IT MATTERS Transcription factors have activation domains (ADs) that bind to coactivator complexes to initiate gene transcription. Despite their key role, a comprehensive understanding of what drives their transcriptional activity has remained elusive. Efforts to understand AD activity have largely focused on their amino acid composition. In recent years, it is increasingly realized that the structural ensembles of disordered proteins can dictate their structural properties. For ADs, ensemble structures remain poorly explored, especially in relation to their activity. Here, we report a mutational study of two ADs, HIF-1α and CITED2, that examines how ensemble dimensions correlate with their gene expression activity. Our findings suggest that ensemble dimensions may drive activity in some ADs and that AD ensemble dimensions can be modulated not only through mutations but also through changes in the cellular environment.

INTRODUCTION

Transcription factors (TFs) bind to DNA and use their activation domains (ADs) to recruit transcriptional machinery to the site of transcription, initiating the process of gene expression (1). TFs have DNA-binding domains that facilitate DNA binding directly and have a well-defined structure. In contrast, ADs are intrinsically disordered (2–5). Like other intrinsically disordered regions (IDRs), ADs lack a well-defined tertiary structure and instead are composed of a dynamic collection of rapidly changing conformations, collectively referred to as an ensemble (6). Recent studies have used high-throughput experiments to link amino acid composition and the ability of ADs to induce gene expression (7–11). This body of work has provided evidence that in many ADs, acidic residues and intrinsic disorder work together to expose hydrophobic motifs, thereby enhancing their ability to bind to coactivators. However, recent evidence suggests that the acidic exposure model (AEM) may not fully account for all ADs, as some may have specific positions where single amino

Submitted October 19, 2024, and accepted for publication December 31, 2024.

*Correspondence: ssukenik@syr.edu

Editor: Rebecca Berlow.

https://doi.org/10.1016/j.bpr.2024.100195

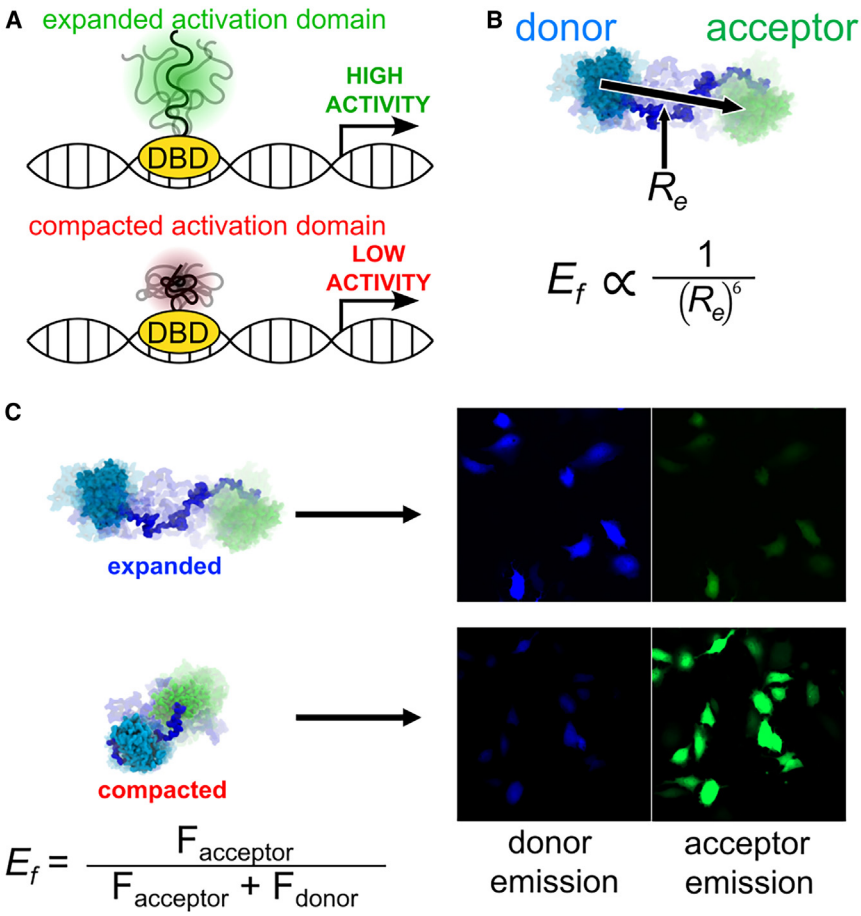

FIGURE 1 Ensemble:function relationship in disordered activation domains. (*A*) We hypothesize that that disordered activation domains (ADs; *green and red squiggly lines*) need an expanded ensemble to induce high levels of gene expression. (*B*) In the FRET constructs used here, AD sequences (in *dark blue*) are flanked by two fluorescent proteins—the donor, mTurquoise2 (in *cyan*), and the acceptor, mNeongreen (in *green*). The average distance between the fluorophore center of the fluorescent proteins, R_e, is inversely proportional to the FRET efficiency (E_f). (*C*) Constructs shown in (*B*) are transiently expressed in U-2 OS cells, which are then imaged using live cell FRET microscopy. Donor and acceptor fluorescence intensities (F_{donor} *and* $F_{acceptor}$, respectively) are quantified, corrected for artifacts, and used to calculate the average average FRET efficiency (E_f) for each cell.

acid substitutions, even beyond the residues proposed by the model, can fine-tune transcriptional activation (12). Despite significant experimental progress, most research has focused primarily on the amino acid sequence of ADs, with less attention given to the role that ensemble dimensions may play in regulating their activity (13,14).

We hypothesize that expanded ensemble dimensions in ADs can improve coactivator binding, resulting in higher gene expression (Fig. 1 *A*). In contrast, compact AD ensembles would hinder coactivator binding, resulting in low gene expression. The AEM could be seen as a specific example within the broader concept of expansion and compaction, connecting AD sequence, structure, and function.

MATERIALS AND METHODS

AD activity analysis

AD sequences and activity data for HIF-1α, CITED2, and their variants, as well as their standard deviations, and the *p* values adjusted using a 5% false discovery rate were obtained from Staller et al. (7). Activity fold change was calculated by dividing the activity in the mutant by the activity of the wild-type (WT) protein.

Sequence and ensemble feature analysis

R_e and asphericity prediction was performed using ALBATROSS (15) and sequence feature analysis performed using localCIDER (16) python packages. Statistical analysis of all sequence features was performed by comparing activating and deactivating sequences using a Mann-Whitney U test. Predicted end-to-end distance R_e was normalized by dividing the R_e value of each mutant by the R_e value of the WT (Fig. 1 *B*). To determine the fraction of residues that fall under the acidic exposure model (AEM), the fraction of aromatic, leucine, and acidic residues was calculated using the compute_residue_fractions() function within SPARROW (https://github.com/idptools/sparrow).

Helicity analysis

Helicity analysis was performed in two ways. First, AGADIR was used to determine the percentage of helix for each sequence. Specifically, the predict_alphahelix() function within PyAGADIR (https://github.com/reisalex/pyAGADIR) was used to predict the percentage of helical structure for each residue (17). Second, the structure of each mutant was predicted from sequence using ColabFold v.1.5.5, running AlphaFold 2.3 (18) coupled with MMseqs2 for sequence alignment and template generation. Protein sequences were input into the ColabFold notebook, and complex modeling

was enabled using paired and unpaired multiple sequence alignments derived from the UniRef100 database. Structural models were generated with the default model type (auto), which employs alphafold2_ptm for monomers. Amber relaxation was applied to refine the top-ranked structures, and model quality was assessed using pLDDT (predicted local distance difference test) scores. Secondary structure assignments and solvent-accessible surface area calculations were performed on the resulting PDB structure using the "define secondary structure of proteins" function within PyMOL. The percentage of helix was calculated by dividing "H"-assigned residues by the total number of residues.

Monte Carlo simulation for Pearson's correlation analysis

Pearson's r was calculated using scipy.stats.pearsonr() function. We employed a Monte Carlo approach to account for error bars when performing the correlation analysis. For this, Pearson's r was calculated 1000 times on simulated datasets obtained by adding normally distributed random numbers scaled by the error in the *x* and *y* axes for each data point. We then averaged all the r values and determined the average and standard deviation from these 1000 repeats. The code for this procedure is available at the GitHub repository accompanying this work (https://github.com/sukeniklab/Flores-et-al-2024).

Plasmid handling

Plasmids for mammalian expression were constructed by inserting the mTurquoise2 and mNeonGreen genes into a pcDNA3.1(+) plasmid. This was performed using the 5′ NdeI and 3′ XhoI restriction sites. The IDP regions were obtained from GenScript and inserted between the two fluorescent proteins (FPs) using the 5′ SacI and 3′ HindIII restriction sites. The resulting plasmids were then amplified in XL-1 Blue cells (Thermo Fisher Scientific, Waltham, Massachusetts) following the manufacturer's protocol. The amino acid sequences of all fluorescence resonance energy transfer (FRET) constructs, including their IDP inserts, are provided in Data S1.

Mammalian cell culture and plasmid transfection

U-2 OS cells (ATCC, Manassas, Virginia) were cultured in T25 (Thermo Scientific) flasks using Advanced DMEM (Gibco, Grand Island, New York) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco). For live-cell microscopy experiments, 10,000 U-2 OS cells were plated in each well of a μ-Plate 96-well black plastic, ibidi-treated imaging plate (ibidi, Fitchburg, Wisconsin), and allowed to adhere overnight ($\sim$16 h) before transfection. The cells were maintained at 37°C with 5% CO_2. Before transfection, the medium was replaced with fresh, warmed DMEM. XtremeGene HP (Sigma, St. Louis, Missouri) was used to transfect FRET construct plasmids into U-2 OS cells, following the manufacturer's protocol (for each well, 16 μL was prepared at a ratio of 2 μL XtremeGene HP to 1 ng of plasmid DNA). After transfection, the cells were incubated at 37°C with 5% CO_2 for 48 h.

Live-cell microscopy

Imaging was performed using a Zeiss epifluorescent microscope equipped with a 20$\times$ dry objective (0.9 NA). The samples were illuminated using a Colibri2 LED light engine (Zeiss, Oberkochen, Germany), and data were collected with a dual-camera setup, which included two linked ORCA Flash v.3 sCMOS cameras (Hamamatsu, Shizouka, Japan). The cells were imaged at an ambient temperature of 21°C, with exposure times of 150 ms and a LED intensity set to 7%. mTurquoise2 was excited at 430 nm (both donor and acceptor channels), and mNeonGreen was excited at 511 nm (direct acceptor channel). Emitted light was passed on to the cameras using a triple band-pass dichroic filter (467/24, 555/25, 687/145). For FRET measurements, the emitted light was split onto the two cameras using a downstream beam splitter with a 520 nm cutoff. Images were taken in PBS or HEPES buffers for mammalian cell culture.

For cell volume perturbation, cells were first imaged, followed by another round of imaging after the addition of the perturbation solution. To ensure thorough mixing, the solutions were pipetted up and down 10 times prior to imaging. The second image was captured within approximately 25 s. Perturbation solutions were created by either diluting the imaging medium (PBS) with autoclaved deionized water to achieve hypoosmotic conditions (100 mOsm final osmotic pressure) or adding 1 M NaCl stock solution to achieve hyperosmotic conditions (750 mOsm final osmotic pressure). Isosmotic conditions (300 mOsm) were obtained by using PBS (Gibco). Before imaging, cells were rinsed once with PBS and then left in 200 μL of PBS (300 mOsm) prior to imaging.

Image analysis

Images were analyzed using ImageJ (19). The images collected before and after the osmotic challenge, which contained three channels each, were stacked and aligned using the StackReg plugin with a rigid transformation (20). The aligned images were segmented based on the donor channel before perturbation, using a fixed threshold to select only pixels with intensities higher than 850 a.u. The resulting mask was processed using ImageJ's Watershed binary algorithm. Cells were selected using the "analyze particles" option in ImageJ, filtering for those with an area between 200 and 50,000 μm^2 and a circularity between 0 and 0.718. The average intensity of the selected regions of interest was calculated. Cells with average direct acceptor emission intensities ranging from 1,258.9 to 7,943.2 a.u. were selected. Values below this range could not be differentiated from camera noise on at least one channels, while higher intensities introduced artifacts and oligomerization effects (21). Cells with an absolute change in direct acceptor emission following osmotic change greater than 1,200 a.u. (typically cells that moved or lifted off the coverslip during measurement) were excluded.

To correct for donor bleedthrough and cross-excitation, cells were transfected with either the mTurquoise2 or mNeonGreen construct alone, as described previously (21). These cells were imaged and analyzed using the same protocol described above, and correlation plots were generated to determine the percentage of bleedthrough and cross-excitation. In the final filtering step, cells with a corrected donor/acceptor ratio that was negative or larger than 6 were removed, as anything outside of this range is indicative of artifacts. The code is available as an ImageJ macro at the accompanying GitHub repository for this paper (https://github.com/sukeniklab/Flores-et-al-2024).

FRET analysis

FRET efficiencies for perturbations and basal conditions were calculated for each segmented cell as follows:

$$E_f = \frac{F_A}{F_D + F_A}$$

where F_D represents the average donor emission and F_A represents the average acceptor emission corrected for bleedthrough and cross-excitation (Fig. 1 *C*). Thus, each E_f value represents the average FRET efficiency of a single cell. The E_f of all the cells in a single imaging well were used to construct a violin plot. Each well

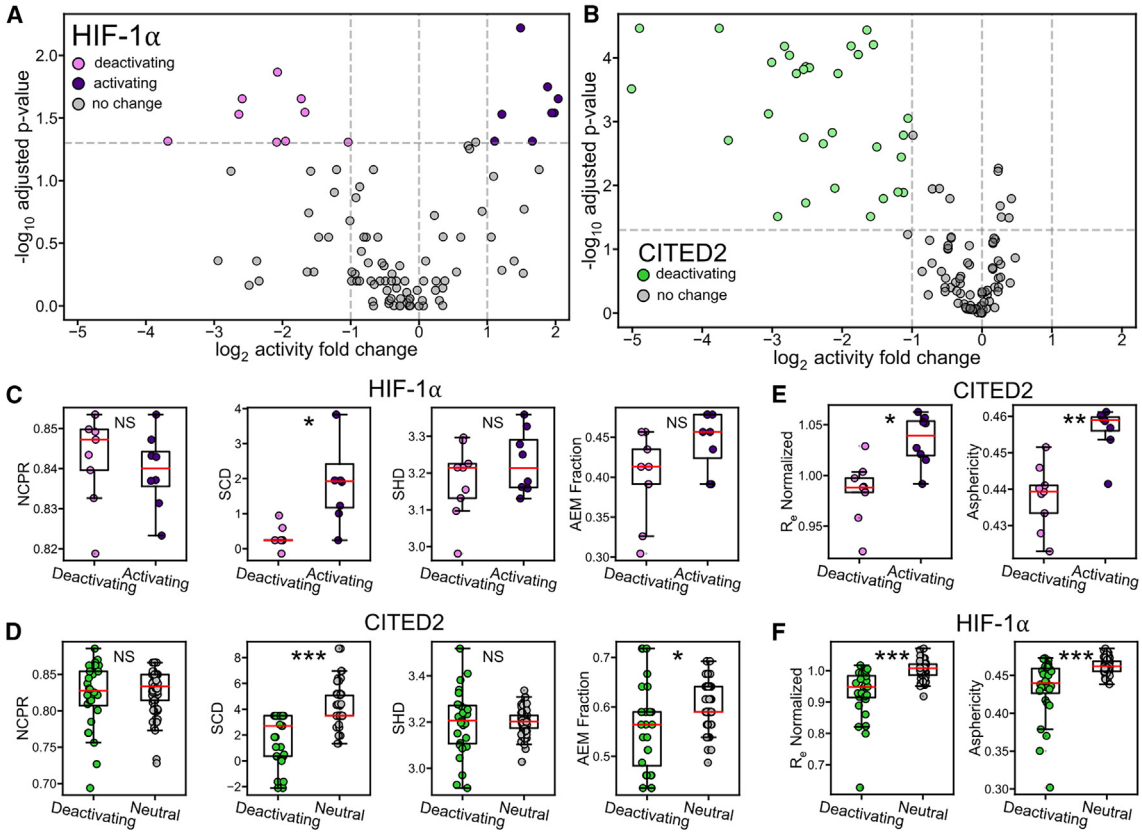

FIGURE 2 HIF-1α and CITED2 transcriptional activity levels and sequence properties. (*A*) Volcano plot of HIF-1α fold-change in gene expression activity compared to the WT sequence. Significantly activating and deactivating mutations are shown in purple and pink symbols, respectively (where the *p* value adjusted with a 5% false discovery rate is lower than 0.05). (*B*) Volcano plot of CITED2 fold-change in gene expression activity compared to the WT sequence. Mutants that significantly decrease activity are shown in green. (*C*) HIF-1α sequence features, including net charge per residue (NCPR), sequence charge decoration (SCD), and sequence hydropathy decoration (SHD), for significantly activating and deactivating HIF-1α mutants. (*D*) CITED2 sequence features (same as in (*C*)) significantly activating mutants and those with no significant change in activity. (*E* and *F*) Normalized R_e (end-to-end distance) ALBATROSS and asphericity predictions of HIF-1α and CITED2 mutants, respectively. Statistical significance in (*C*–*F*) is denoted by asterisks (*$p < 0.01$, **$p < 0.001$, and ***$p < 0.0001$, using a Mann-Whitney U test; NS denotes no significant difference).

contained at least 30 imaged, segmented, and filtered cells, with some wells holding hundreds of cells. The significance in E_f between any two FRET constructs was assessed based on the medians *n* imaged wells for two constructs using an independent, two-sided *t*-test. For osmotic perturbations, ΔE_f was calculated as $\Delta E_f = E_{f,after} - E_{f,before}$ for all osmotic perturbations, where E_f is again calculated per cell, with all cells from a single well used to create a single violin plot. Statistical analysis for these experiments also used an independent *t*-test to compare hypo- or hyperosmotic with isosmotic perturbation ΔE_f values. The number of cells measured for each construct and condition is summarized in Data S1.

RESULTS

Sequence and ensemble features for HIF-1α and CITED2 ADs and their mutant variants

We began by comparing activity measurements published previously for HIF-1α and CITED2 that included the activity of sequences with multiple mutations (7). Point mutations applied to HIF-1α were able to both increase or decrease activity significantly compared to the WT (Fig. 2 *A*). In contrast, CITED2 point mutations showed no significant increase in activity and, in many cases, only decreased activity (Fig. 2 *B*).

We next wanted to see if we could find sequence properties that would correlate with these changes in activity. To do this, we calculate the net charge per residue (NCPR), sequence charge decoration (SCD) (22), sequence hydropathy decoration (SHD) (23), and the fraction of residues participating in the AEM (tryptophan, aspartic acid, and glutamic acid, AEM fraction, see further explanation in Figs. S1 and S2) (Fig. 2, *C* and *D*). These were selected to explore whether charge-related (NCPR, SCD) or non-charge related (SHD) properties could explain the differences between activating and deactivating sequences. Lastly, we measured the fraction of amino acids that fall within the AEM to determine if this metric revealed any significant differences between the sequences

(Fig. S1). In HIF-1α, activating sequences had a significantly higher SCD when compared to deactivating sequences (Fig. 2 *C*). All other parameters we tested showed no significant difference (Figs. 2C and S1). CITED2 had no activating mutations, so instead we used mutations that had no significant effect on activity compared to WT ($|\log_2$ (fold change)$| < 1$ or $-\log_{10}$ adjusted $p < 1.3$). Here again, the only sequence feature that showed a significant difference between deactivating and neutral sequences was the SCD (Figs. 2 *D* and S2). This result led us to explore whether differences might be more significant in the structural features of the ensemble.

We predict ensemble features using ALBATROSS, a machine learning network trained to predict average ensemble dimensions of IDRs from sequence (15). Unlike sequence properties, predicted ensemble features showed significant correlations with activity. HIF-1α showed a statistically significant difference in the normalized predicted end-to-end distance (R_e) and asphericity between activating and deactivating sequences (Fig. 2 *E*). Specifically, mutations predicted to expand ensemble dimensions were associated with increased activity, while those predicted to cause compaction were linked to decreased activity. In contrast, CITED2 mutants showed either a loss of activity (deactivating) or no significant change in activity (neutral) (Fig. 2 *B*). Still, CITED2 displayed significant differences between deactivating and neutral sequences for the same ensemble features (Fig. 2 *F*). A growing body of work links the ensemble dimensions of disordered regions to their function (6,24,25). This, together with the correlations between ensemble structure and activity for ADs we found, motivated us to experimentally explore the link between AD ensemble structure and function directly.

Live-cell FRET reveals that HIF-1α, but not CITED2, displays structural plasticity

To establish the link between HIF-1α and CITED2 AD ensembles and their activity, we selected a subset of mutations whose activities were measured (Fig. 3 *A*). We reasoned that variants where hydrophilic and hydrophobic residues were mutated to aromatics would bring distal regions of the protein closer together through π-stacking interactions, compacting the ensemble (26,27). We also reasoned that replacing all positively charged residues with negatively charged residues would introduce electrostatic repulsive forces with existing negative charges that would expand the sequence (28). Lastly, we wanted to see what structural role positive charges, hydrophobics, and aromatics played when absent from the WT CITED2 (FL $>$ A_{NT} and RK $>$ A_{CT}).

To measure the ensemble dimensions of the selected mutants experimentally, we used live-cell FRET microscopy (21). Ensemble dimensions are measured through the cell-average FRET efficiency, E_f, between a donor and an acceptor flourescent protein (Fig. 1 *B*), where an E_f close to 1 denotes a compact ensemble and an E_f close to 0 denotes an expanded ensemble (Fig. 1 *C*). Our measurements reveal that the selected HIF-1α mutants behaved as predicted: sequences with aromatic mutations resulted in more compact ensembles than the WT sequence as indicated by a higher E_f. HIF-1α sequences that introduced negatively charged residues expanded the ensemble as shown by a decrease in E_f compared to the WT (Fig. 3 *C*). We also correlated the resulting E_f of mutations with the predicted R_e from ALBATROSS. All HIF-1α sequences showed a high correlation (Pearson's $r = -0.8 \pm 0.1$) between E_f and the predicted R_e. Since ALBATROSS predictions for R_e do not include the flourescent proteins, this correlation suggests that flourescent protein presence does not alter the trends of ensemble dimensions in our experiment (Fig. 3 *E*).

In contrast, CITED2 variants displayed some unexpected behavior with respect to ensemble dimensions. First, variants with insertion of aromatics (L $>$ W) did not compact the ensemble as we hypothesized. Instead, these variants maintained similar ensemble dimensions to the WT (L $>$ W) or, even more surprisingly, led to increased expansion (L $>$ F) (Fig. 3 *D*). Replacing positively charged lysine and arginine with negatively charged glutamate (RK $>$ E) led to an expansion of the ensemble, likely attributed to the introduction of repulsive electrostatic forces. Substituting aromatic phenylalanine and leucine residues with alanine (FL $>$ A_{NT}) led to the ensemble's expansion, consistent with the removal of hydrophobic and aromatic residues (Fig. 3 *D*) Lastly, substitution of positively charged residues with alanine in the C-terminus (RK $>$ A_{CT}) resulted in a small expansion in its ensemble dimensions. Unlike HIF-1α, E_f for CITED2 had poor correlation with R_e (Pearson's $r = -0.01 \pm 0.1$), which was predicted to be more variable, indicating the possibility of interactions between the IDR and FPs (Fig. 3 *F*).

To further investigate the structural basis for the observed activity changes, we assessed the helicity of both HIF-1α and CITED2 mutants using AGADIR and AlphaFold (17,18,29–31). Despite large changes in helicity between the two different predictors, we did not find a clear correlation between helicity and activity for either protein in either case (Fig. S3). This suggests that residual structure in the unbound ensemble may not be a dominant factor in modulating the activity of these proteins.

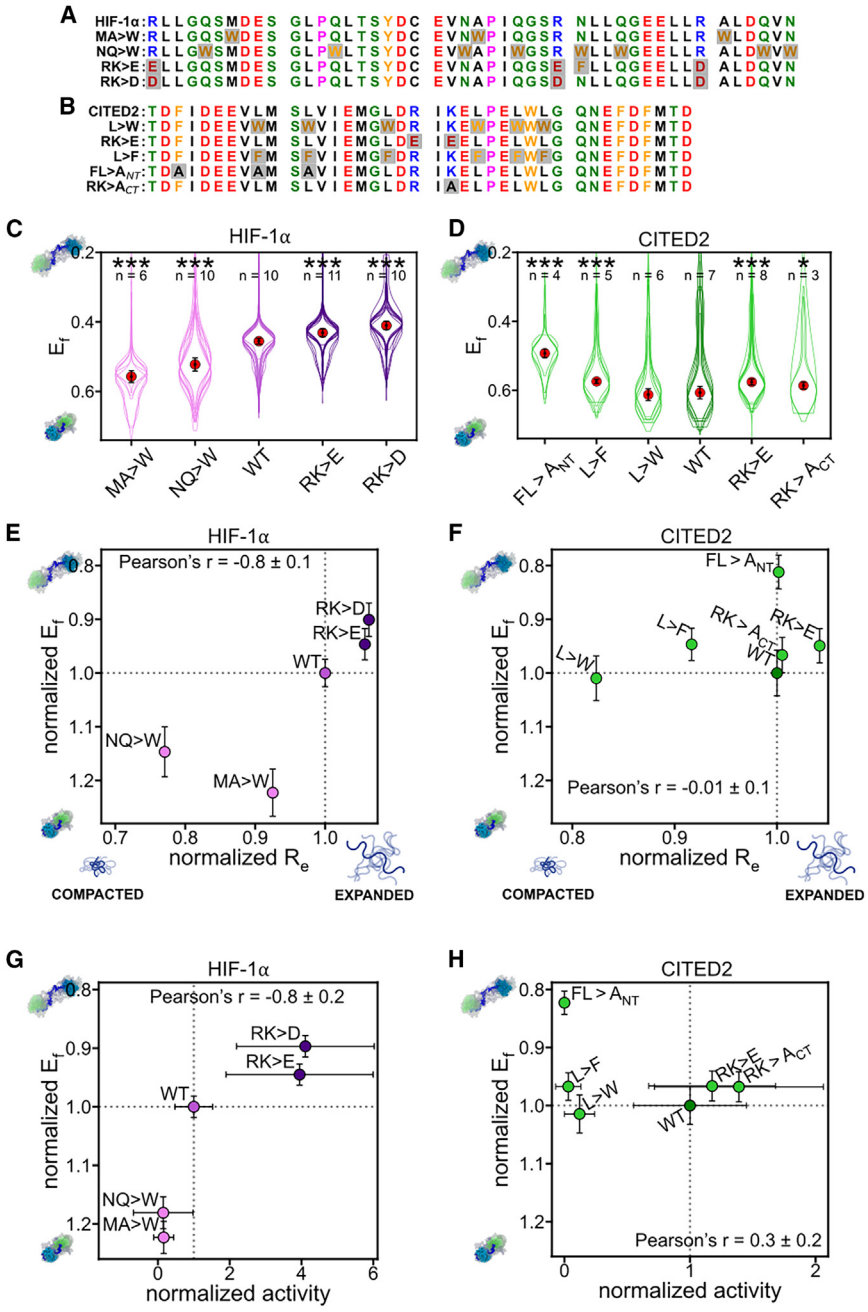

FIGURE 3 Ensemble dimensions of HIF-1α and CITED2 and their mutants measured in live cells and correlated with their gene expression activity. (*A*) Wildtype (WT) and mutant HIF-1α sequences. Mutated residues are highlighted in light gray. (*B*) Same as (*A*) but for CITED2. (*C*) Cell-averaged FRET efficiencies (E_f) in live cells for HIF-1α and its mutants. Each overlaid violin plot is obtained from an individual biological repeat done in a single imaging well, and contains at least 30 cells. Red markers are average of the medians of *n* wells, where *n* is shown at the top, and the black whiskers represent the standard deviation of the medians. Asterisks denote significant changes compared to WT (*$p < 0.01$ and ***$p < 0.0001$, using an independent, two-sided t-test; NS denotes not significant). (*D*) Same as (*C*) for CITED2. (*E*) E_f normalized to WT versus predicted end-to-end distance normalized to WT. (*F*) Same as (E) but for CITED2. (*G*) E_f normalized to WT of HIF-1α mutants plotted against their gene expression activity normalized to WT. (*H*) Same as (G) but for CITED2. In panels (*E–H*) dashed lines indicate the WT reference point.

To finally check if ensemble dimensions correlated with activity, we plotted the average median E_f against the activity measured for the same AD. Both the average median E_f and activity were normalized to the associated WT construct. We observed a strong correlation (Pearson's r $= -0.8 \pm 0.2$) between the normalized average median E_f and the normalized activity in HIF-1α sequences, indicating that an expanded ensemble facilitates higher HIF-1α activation, while a compact HIF-1α ensemble would hinder activity (Fig. 3 *G*). CITED2, on the other hand, showed a contrasting profile. Mutations either completely abolished transcriptional activity, even when ensemble dimensions remained unchanged or expanded, or maintained activity at levels similar to the WT, showing little correlation with ensemble dimensions (Pearson's r $= 0.3 \pm 0.2$) (Fig. 3 *H*).

The HIF-1α mutants support the AEM and link ensemble dimensions for AD function. The HIF-1α mutants that exhibit expanded ensemble dimensions often feature acidic residues, and the repulsion between those could drive ensemble expansion

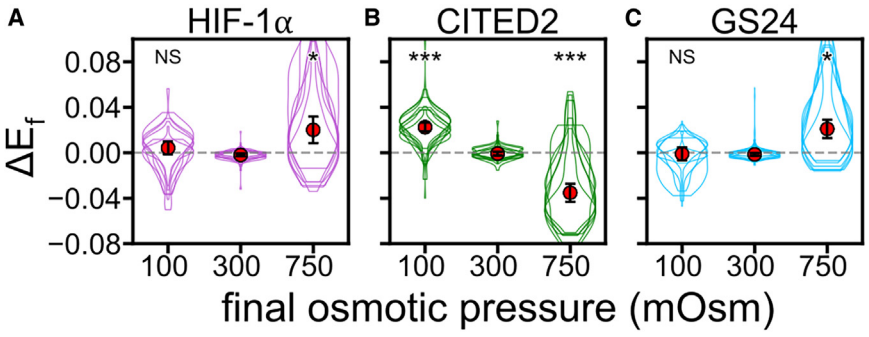

FIGURE 4 Characterization of ensemble sensitivity of AD constructs in response to osmotic stress in live cells. (*A–C*) Change in FRET efficiency following osmotic perturbations, ΔE_f, in response to osmotic stress for (*A*) HIF-1α, (*B*) CITED2, and (C) GS24. Violin plots show hypo-osmotic (100 mOsm), isosmotic (300 mOsm), and hyperosmotic (750 mOsm) stress. Each overlaid violin plot is composed from at least 30 individual cells before and after osmotic challenge, imaged in a single well. Red dot denotes the average ΔE_f median value of all overlaid violins along with its propagated error. Significant differences in ΔE_f are determined compared to the 300 mOsm control and denoted with an asterisk (*$p < 0.01$ and ***$p < 0.0001$, using an independent, two-sided *t*-test. NS denotes no significant changes).

(RK > E, RK > D). Conversely, HIF-1α mutants that are "collapsed" feature mutations to aromatic residues (NQ > W, MA > W), which are known to form attractive interactions and can thus collapse the ensemble. Importantly, acidic residues are key drivers of disorder, but they do not act alone. Instead, they work in synergy with other disorder-promoting residues to facilitate the exposure of hydrophobic motifs.

In contrast, we hypothesize that the WT CITED2 ensemble is structurally rigid, which makes further expansion through sequence modifications difficult. Structurally, only the FL > A_{NT} mutant resulted in a compact ensemble, while the other mutants maintained similar ensemble dimensions when compared to the WT. It is important to note that the FL > A_{NT} mutant occurs in a region previously characterized as activity driving (7). Thus, a decrease in activity may not be related to changes in ensemble dimensions but rather may result from changes in binding affinity or other mechanistic functions.

Compact, high-activity CITED2 shows structural sensitivity to osmotic perturbations

Our results show that ensemble dimensions can be modulated through changes in sequence, but past work indicates that they can also be modulated from changes in the cellular environment (21,32). If the link between ensemble dimensions and AD activity holds true, then in the case where changes to cellular environment could modulate AD ensembles they could also modulate their activity. This would be a previously unappreciated mechanism for regulating gene expression. To see if this hypothesis is a possibility, we decided to test the response of ADs to cellular volume change perturbations. Although these perturbations may not be directly relevant to the physiological activities of CITED2 or HIF-1α, they still provide insights into the structural plasticity of the sequence in a changing cellular environment.

Osmotic challenges act to rapidly and reproducibly modulate the volume of the cell (33). Cellular volume changes occur routinely during the cell cycle and change the composition of all solute molecules inside the cell (34,35). Our previous work has shown that some disordered regions could be sensitive to osmotic changes (21). Thus, their effect on ensemble structure could also be meaningful to AD activity. To test the structural sensitivity of AD ensembles to changes in the cellular environment, we measured changes in FRET efficiency ($\Delta E_f = E_f^{after} - E_f^{before}$) of the HIF-1$\alpha$ and CITED2 ADs inside the cell as it is subjected to hypo- or hyperosmotic challenge. We also measured the response of a control sequence composed of 24 Gly-Ser repeats (GS24), which acts as a homopolymeric point of reference (21). This point of reference allows us to gauge if the sequence displays sensitivity to changes in cell volume different than that of a homopolymer (21,32).

HIF-1α and CITED2 show opposite responses to our osmotic challenges. Hypo-osmotic pressure did not significantly impact the ensemble dimensions of HIF-1α when compared to the isosmotic control (Fig. 4 *A*). In contrast, a hyperosmotic challenge significantly increased FRET efficiency for HIF-1α ($\Delta E_f > 0$) compared to the isosmotic control, indicating ensemble compaction. CITED2 compacted under hypo-osmotic stress as indicated by an increase in FRET efficiency ($\Delta E_f > 0$) (Fig. 4 *B*). When exposed to the same hyperosmotic challenge, CITED2 exhibited ensemble expansion ($\Delta E_f < 0$). Here, we use a GS24 (48 total residues) glycine-serine repeat homopolymer for comparison (Fig. 4 *C*). We observe that the homopolymer responds similarly to HIF-1α. Overall, CITED2 is more sensitive to the cellular environment than either HIF-1α or the GS24 linker.

Our findings substantiate the elusive characterization of ADs: what is true for some might not be true for others. Given that ADs are largely disordered, we aimed to shed light on the role that ensemble structure plays in their activity. For HIF-1α, our study

shows that ensemble dimensions correlate with activity. Specifically, expanded ensembles showed higher HIF-1α gene expression activity, while compact ensembles showed lower activity levels. It has previously been suggested (36–38), though recently contended (39), that an expanded protein region enhances DNA binding by facilitating access through transient interactions. For ADs, this expanded ensemble could result in a high capture radius for the IDR, enhancing its ability to bind transcriptional coactivators necessary for transcription initiation.

For CITED2, the emerging picture is more complicated because mutations did little to change ensemble dimensions, and no mutations were found that increased the activity of the construct. We therefore propose that ADs like CITED2, in which the ensemble is less prone to changes through mutations, can be rather stable in the face of mutations. This stability highlights the flexibility and adaptability of IDRs, which are crucial for solving diverse biological problems and may offer evolutionary advantages by allowing functional robustness despite sequence variability. The robustness of CITED2's ensemble dimensions to mutation may provide selective advantages by enabling IDRs to maintain function while exploring diverse sequence spaces, thereby promoting the development of protein function over time. We emphasize that our findings are specific to the two ADs studied and may not be applicable to all ADs. Still, a common characteristic among ADs is that they are intrinsically disordered, which makes their sequence highly exposed to the cellular environment.

Our results contain some limitations worth considering. One key observation is the weak correlations between CITED2's ensemble dimension and predicted R_e in contrast to HIF-1α. To investigate this discrepancy, we analyzed the sequence and ensemble features of their respective variants. Of all the metrics, only the fraction of charged residues was significantly larger in CITED2 compared to HIF-1α. A higher concentration of charged residues could lead to stronger interactions with the flanking fluorophores, which, on their own, carry some charges (21). This could potentially affect the observed E_f. In this case, further studies of CITED2 mutations in the absence of the FP labels might show a more flexible ensemble. Nonetheless, activity measurements, which were done without the FPs, show that no mutant had an increase in activity.

Another important consideration is the length of the amino acid sequences used in both the FRET and activity assays. ADs are often embedded within larger IDRs, which means that structural or functional regulation may occur outside the specific regions studied. Therefore, our findings should be interpreted with the understanding that additional regulatory elements beyond the observed sequence window may interact with these ADs. We emphasize that while our study focuses on isolated ADs (7–10,40), future work incorporating full-length TFs will be essential for fully understanding the interplay between ADs and their sequence context. Additionally, the possibility that either AD exerts an autoinhibitory effect on DNA binding (41), and how this correlates with ensemble structure, should also be considered. Exploring such autoinhibitory effects in the context of full-length TFs may provide further insight into how ensemble dimensions are shaped by both structural constraints and functional requirements. Lastly, although the experiments measuring ensemble dimensions and transcriptional activity were conducted separately and with different constructs (but with the same AD), these experiments provide the first correlation linking AD ensembles and activity.

One additional finding in this work is that perturbations to cell volume had a much greater effect on CITED2 ensemble dimensions in comparison to HIF-1α or its GS24 control. Previous work has shown how changes in cell volume not only affect cell fate but also TF function (42,43). Further research is needed to determine whether the changes in ensemble dimensions caused by osmotic stress also correlate directly with activity. Based on our findings, we propose that the cellular environment may be a novel mechanism for regulating activity through changes in AD ensemble dimensions. It is interesting to note that the AD that displayed activity tuned by ensemble dimensions showed little sensitivity, while those whose activity was not easily tunable by mutations showed higher sensitivity. This may indicate mutually exclusive mechanisms by which AD function is carried out.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the NIH under awards R35GM150813 to M.V.S. and R35GM137926 to S.S. S.S. is supported in part by a Sloan Research Fellowship in Chemistry.

AUTHOR CONTRIBUTIONS

S.S., E.F., M.V.S., and F.Y. conceptualized the project. E.F. designed and performed all live-cell experiments, analyses, and figures, assisted by A.R.C., E.C.-Z., and M.B.M. All activity data were provided by M.V.S. S.S., M.V.S., and E.F. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING MATERIAL

Supplemental information can be found online at https://doi.org/10.1016/j.bpr.2024.100195. Raw data and scripts used to analyze images and numerical data, as well as scripts used for preperation of all figures are available at https://github.com/sukeniklab/Flores-et-al-2024.

REFERENCES

1. Latchman, D. S. 1990. Eukaryotic transcription factors. *Biochem. J.* 270:281–289.
2. Dyson, H. J., and P. E. Wright. 2016. Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300. *J. Biol. Chem.* 291:6714–6722.
3. Soto, L. F., Z. Li, …, J. I. Fuxman Bass. 2022. Compendium of human transcription factor effector domains. *Mol. Cell.* 82:514–526.
4. Már, M., K. Nitsenko, and P. O. Heidarsson. 2023. Multifunctional Intrinsically Disordered Regions in Transcription Factors. *Chemistry.* 29:e202203369.
5. Ferrie, J. J., J. P. Karr, …, X. Darzacq. 2022. "Structure"-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation. *Mol. Cell.* 82:3970–3984.
6. Holehouse, A. S., and B. B. Kragelund. 2023. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat. Rev. Mol. Cell Biol.* 25:187–211.
7. Staller, M. V., E. Ramirez, …, B. A. Cohen. 2022. Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst.* 13:334–345.e5.
8. DelRosso, N., J. Tycko, …, L. Bintu. 2023. Large-scale mapping and mutagenesis of human transcriptional effector domains. *Nature.* 616:365–372.
9. Staller, M. V., A. S. Holehouse, …, B. A. Cohen. 2018. A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst.* 6:444–455.e6.
10. Erijman, A., L. Kozlowski, …, S. Hahn. 2020. A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning. *Mol. Cell.* 79:1066.
11. Sanborn, A. L., B. T. Yeh, …, R. D. Kornberg. 2021. Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *Elife.* 10:e68068.
12. Sreenivasan, S., P. Heffren, …, L. Swint-Kruse. 2024. The intrinsically disordered transcriptional activation domain of CIITA is functionally tuneable by single substitutions: An exception or a new paradigm? *Protein Sci.* 33:e4863.
13. Vancraenenbroeck, R., Y. S. Harel, …, H. Hofmann. 2019. Polymer effects modulate binding affinities in disordered proteins. *Proc. Natl. Acad. Sci. USA.* 116:19506–19512.
14. Bjarnason, S., J. A. P. McIvor, …, P. O. Heidarsson. 2024. DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2. *Nat. Commun.* 15:1445.
15. Lotthammer, J. M., G. M. Ginell, …, A. S. Holehouse. 2024. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat. Methods.* 21:465–476.
16. Holehouse, A. S., R. K. Das, …, R. V. Pappu. 2017. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* 112:16–21.
17. Muñoz, V., and L. Serrano. 1994. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* 1:399–409.
18. Mirdita, M., K. Schütze, …, M. Steinegger. 2022. ColabFold: making protein folding accessible to all. *Nat. Methods.* 19:679–682.
19. Abramoff, M. D., P. J. Magelhaes, and S. J. Ram. 2004. Image Processing with ImageJ. *Biophot. Int.* 11:36–42.
20. Thévenaz, P., U. E. Ruttimann, and M. Unser. 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* 7:27–41.
21. Moses, D., K. Guadalupe, …, S. Sukenik. 2024. Structural biases in disordered proteins are prevalent in the cell. *Nat. Struct. Mol. Biol.* 31:283–292.
22. Sawle, L., and K. Ghosh. 2015. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* 143:085101.
23. Zheng, W., G. Dignon, …, J. Mittal. 2020. Hydropathy Patterning Complements Charge Patterning to Describe Conformational Preferences of Disordered Proteins. *J. Phys. Chem. Lett.* 11:3408–3415.
24. González-Foutel, N. S., J. Glavina, …, L. B. Chemes. 2022. Conformational buffering underlies functional selection in intrinsically disordered protein regions. *Nat. Struct. Mol. Biol.* 29:781–790.
25. Yu, F., and S. Sukenik. 2023. Structural Preferences Shape the Entropic Force of Disordered Protein Ensembles. *J. Phys. Chem. B.* 127:4235–4244.
26. Vernon, R. M., P. A. Chong, …, J. D. Forman-Kay. 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife.* 7:e31486.
27. Martin, E. W., A. S. Holehouse, …, T. Mittag. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science.* 367:694–699.
28. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA.* 110:13392–13397.
29. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
30. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* 245:275–296.
31. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* 245:297–308.
32. Moses, D., F. Yu, …, S. Sukenik. 2020. Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *J. Phys. Chem. Lett.* 11:10131–10136.
33. Sukenik, S., P. Ren, and M. Gruebele. 2017. Weak protein-protein interactions in live cells are quantified by cell-volume modulation. *Proc. Natl. Acad. Sci. USA.* 114:6776–6781.
34. Boucrot, E., and T. Kirchhausen. 2008. Mammalian cells change volume during mitosis. *PLoS One.* 3:e1477.
35. Bryan, A. K., V. C. Hecht, …, S. R. Manalis. 2014. Measuring single cell mass, volume, and density with dual suspended microchannel resonators. *Lab Chip.* 14:569–576.
36. Huang, Y., and Z. Liu. 2009. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J. Mol. Biol.* 393:1143–1159.
37. Levy, Y., J. N. Onuchic, and P. G. Wolynes. 2007. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* 129:738–739.

38. Trizac, E., Y. Levy, and P. G. Wolynes. 2010. Capillarity theory for the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 107:2746–2750.

39. Kuravsky, M., C. Kelly, …, S. L. Shammas. 2024. The transition state for coupled folding and binding of a disordered DNA binding domain resembles the unbound state. *Nucleic Acids Res.* 52:11822–11837.

40. Alerasool, N., H. Leng, …, M. Taipale. 2022. Identification and functional characterization of transcriptional activators in human cells. *Mol. Cell.* 82:677–695.e7.

41. He, F., W. Borcherds, …, J. Chen. 2019. Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding. *Proc. Natl. Acad. Sci. USA.* 116:8859–8868.

42. Wu, C.-Y., P. A. Rolfe, …, G. R. Fink. 2010. Control of transcription by cell size. *PLoS Biol.* 8:e1000523.

43. Loh, Y.-H., Q. Wu, …, H.-H. Ng. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* 38:431–440.