PAGE4 and Conformational Switching: Insights from Molecular Dynamics Simulations and Implications for Prostate Cancer

Xingcheng Lin[1,2], Susmita Roy[1], Mohit Kumar Jolly[1], Federico Bocci[1,3], Nicholas P. Schafer[1,3], Min-Yeh Tsai[1,3], Yihong Chen[4], Yanan He[4], Alexander Grishaev[4,5], Keith Weninger[6], John Orban[4,7], Prakash Kulkarni[4,8], Govindan Rangarajan[9,10], Herbert Levine[1,2], and José N. Onuchic[1,2,3,11†]

[1]Center for Theoretical Biological Physics, Rice University, Houston, TX
[2]Department of Physics and Astronomy, Rice University, Houston, TX
[3]Department of Chemistry, Rice University, Houston, TX
[4]Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD
[5]National Institute of Standards and Technology, Gaithersburg MD
[6]Department of Physics, North Carolina State University, Raleigh, NC
[7]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD
[8]Department of Medical Oncology and Therapeutics Research, City of Hope National Medical Center, Duarte, CA, United States
[9]Department of Mathematics, Indian Institute of Science, Bangalore
[10]Center for Neuroscience, Indian Institute of Science, Bangalore
[11]Department of BioSciences, Rice University, Houston, TX
[†]To whom correspondence should be addressed: jonuchic@rice.edu

Contents

1 Disordered nature of PAGE4 using Uversky's boundary line

The degree of disorder of PAGE4 can be measured by calculating its mean hydrophobicity [1] and mean net charge. The mean hydrophobicity was calculated using Kyte-Doolittle approximation with the ExPASy server (under normalized scale). We used a window size of 5 residues. The results, compared with the Uversky's boundary line (Eq. S1) [2], (Figure S1) indicates that all three forms of PAGE4 can be classified as disordered proteins.

$$\langle H \rangle = (|\langle q \rangle| + 1.151)/2.785 \tag{S1}$$

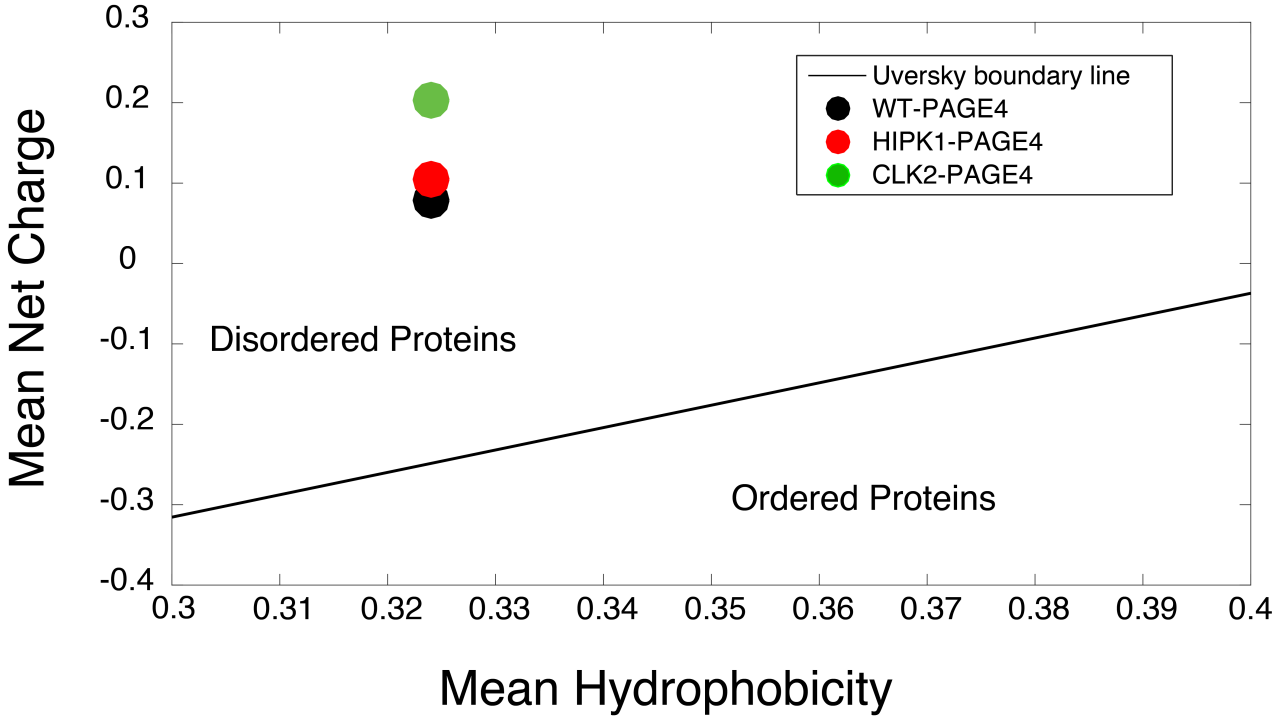

Figure S1: The mean hydrophobicity and mean net charge calculated for three phosphorylated forms of PAGE4. When compared with the Uversky's boundary line, all three forms PAGE4 fall into the category of disordered proteins.

2 Simulation Results with Different Scaling Factor C_{scaling}

The $V_{\text{elec}} = C_{\text{scaling}} K_{\text{elec}} \sum_{i<j} \frac{q_i q_j}{\epsilon_r r_{ij}} e^{-r_{ij}/l_D}$ used in our simulation potential has a constant scaling factor signifying the strength of the electrostatic potential compared with other parts of the AWSEM Hamiltonian. In the main text, we used $C_{\text{scaling}} = 4.0$ in our simulations. We show a comparison for other values of C_{scaling} below (Table S2). The RMS Dist 18_63 of WT-PAGE4 for both $C_{\text{scaling}} = 1.0$ and $C_{\text{scaling}} = 2.0$ are larger

than that in $C_{\text{scaling}} = 4.0$. Upon the reduction of electrostatic force (C_{scaling} changing from 4.0 to 1.0), the difference in sizes among different phoshporylated forms of PAGE4 decreases.

		$C_{\text{scaling}} = 1.0$	
	⟨Rg⟩ (Å)	RMS Dist 18_63 (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	38.4	67.0	55.9
HIPK-1 Form	38.4	66.8	57.5
CLK-2 Form	39.8	67.9	56.5
		$C_{\text{scaling}} = 2.0$	
	⟨Rg⟩ (Å)	RMS Dist 18_63 (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	37.4	66.4	54.1
HIPK-1 Form	37.4	66.5	55.8
CLK-2 Form	40.7	70.1	56.5
		$C_{\text{scaling}} = 4.0$	
	⟨Rg⟩ (Å)	RMS Dist 18_63 (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	32.9	57.4	51.2
HIPK-1 Form	32.1	55.6	51.4
CLK-2 Form	41.8	73.4	54.8

Table S1: A summary of the results from the simulations with different values of C_{scaling}.

To further our understanding in structural details. The structure ensemble and averaged contact map from the simulation with $C_{\text{scaling}} = 1.0$ are plotted in Figure S2. In contrast to Figure 2 of the main text, there are no N-terminal loop formation in either the structure ensemble or the averaged contact map. This further demonstrates that electrostatic force drives the N-terminal loop formation in our simulation with $C_{\text{scaling}} = 4.0$, the loss of this loop formation causes the expansion in the size of PAGE4 upon hyperphosphorylation.

3 Simulation Results without the shift of γ value

To test the simulation of PAGE4 without the shift of γ values, we performed the same simulations with $C_{\text{scaling}} = 1.0$ and $C_{\text{scaling}} = 4.0$ with the unshifted γ values. The results are shown below (Table S3 and Figure S3). It is clear that without the shift of γ parameters, the simulated proteins are over-collapsed.

		$C_{\text{scaling}} = 1.0$	
	⟨Rg⟩ (Å)	RMS Dist 18_63 (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	15.6	22.9	29.4
HIPK-1 Form	15.5	24.1	29.1
CLK-2 Form	15.4	24.7	30.5
		$C_{\text{scaling}} = 4.0$	
	⟨Rg⟩ (Å)	RMS Dist 18_63 (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	15.3	22.4	26.6
HIPK-1 Form	15.5	23.8	29.1
CLK-2 Form	18.1	30.5	30.8

Table S2: A summary of the results from the simulations of PAGE4 with unshifted γ parameters.

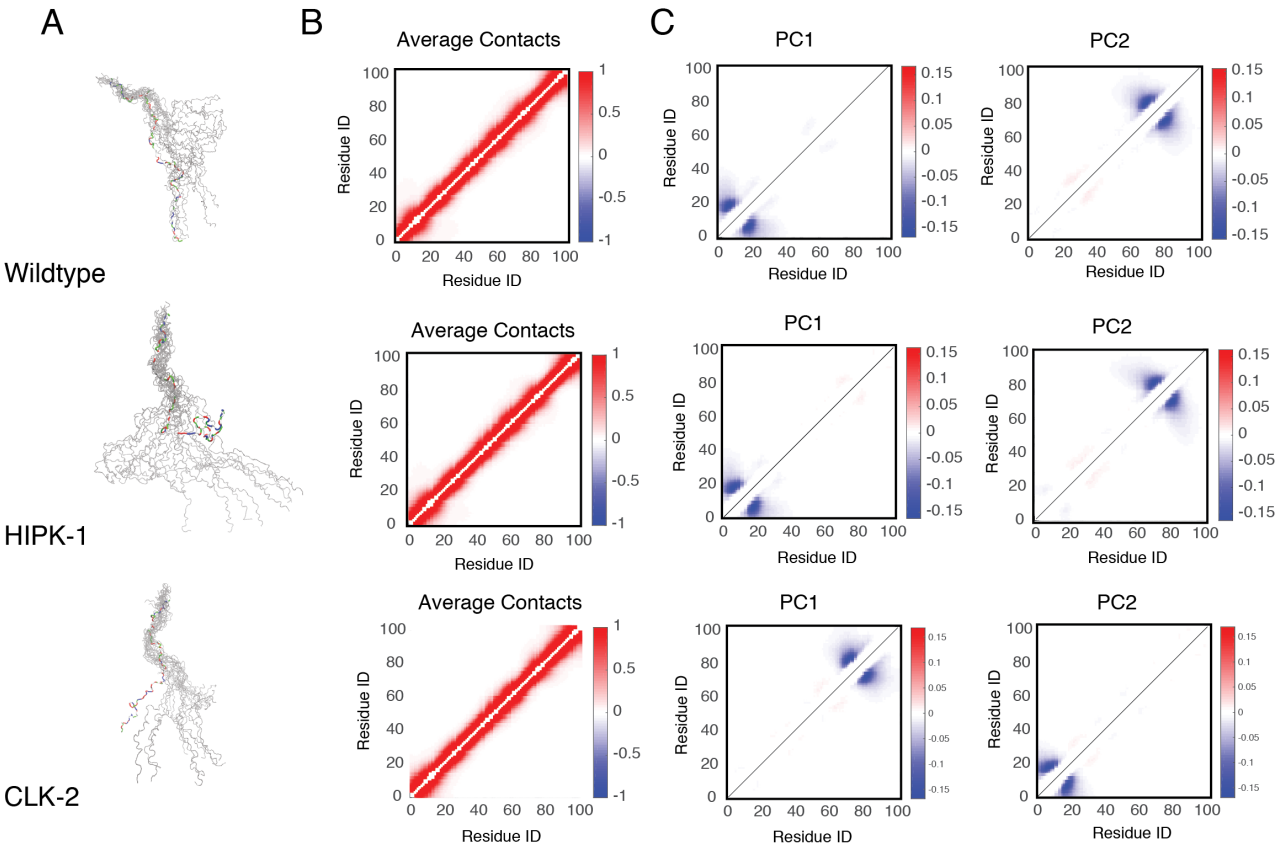

Figure S2: The structure ensemble of different types of PAGE4 collected from the simulations with $C_{scaling} = 1.0$. **A** The structural ensemble generated during our simulations. Randomly picked structures are RMSD aligned by the N-motif. In these simulations, the reduction in electrostatic force causes a lack of N-terminal loop formation. **B** The average contact map generated from the simulations. The color bar shows the probability of contacts formation during our simulations. **C** The top three principal modes generated from the contact-based principal component analysis. We plot the coefficients of the principal components, which corresponds to the change of each contact.

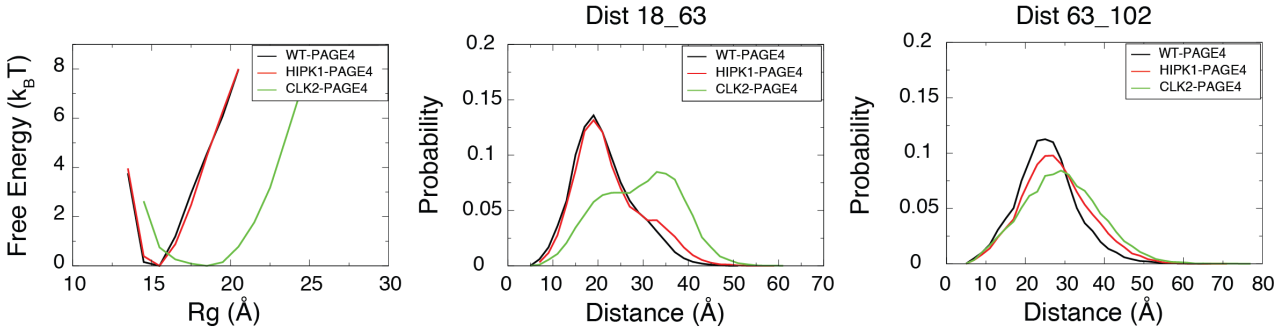

Figure S3: The ensemble averaged properties of PAGE4 from simulations with unshifted γ parameters. We plot here the results with $C_{Scaling} = 4.0$. **A** The free energy plot based on the Radius of Gyration (Rg). **B** The probability distribution of distance between two residue pairs that were previously measured in smFRET experiments.

4 Dimensionless model

The original set of equations for WT-PAGE4 (P_{U}), HIPK1-PAGE4 (P_{M}), CLK2-PAGE4 (P_{H}) and the CLK2 enzyme (C) is given by Equations (5-8) in the Methods section. To derive the dimensionless version of the model, time and variables are rescaled according to the following transformation: $t' = \tau t$, $(P'_{\mathrm{U}}, P'_{\mathrm{C}}, P'_{\mathrm{H}}) = P_0(P_{\mathrm{U}}, P_{\mathrm{M}}, P_{\mathrm{H}})$, $C' = C_0 C$ and $H' = H_0 H$. Here t', P'_{U}, P'_{C}, C' and H' are dimensionless and τ, P_0, C_0 and H_0 are scaling factors. Next, we choose $\tau = 1/k_{\mathrm{H}}$, $C_0 = k_{\mathrm{H}} P_0/g_{\mathrm{H}}$, $H_0 = k_{\mathrm{H}} P_0/g_{\mathrm{M}}$ and let P_0 be an arbitrary level of molecules. This choice leads to the dimensionless model (removing the '):

$$\frac{dP_{\mathrm{U}}}{dt} = \Gamma_{\mathrm{U}} - H\frac{P_{\mathrm{U}}}{P_{\mathrm{U}} + \alpha} - \gamma_{\mathrm{U}} P_{\mathrm{U}} \tag{S2}$$

$$\frac{dP_{\mathrm{M}}}{dt} = H\frac{P_{\mathrm{U}}}{P_{\mathrm{U}} + \alpha} - C\frac{P_{\mathrm{M}}}{P_{\mathrm{M}} + \beta} - \gamma_{\mathrm{M}} P_{\mathrm{M}} \tag{S3}$$

$$\frac{dP_{\mathrm{H}}}{dt} = C\frac{P_{\mathrm{M}}}{P_{\mathrm{M}} + \beta} - P_{\mathrm{H}} \tag{S4}$$

$$\frac{dC}{dt} = \Gamma_{\mathrm{C}} H^+(P_{\mathrm{M}}(t - \tau_{\mathrm{d}}), P_{\mathrm{M0}}, n_{\mathrm{C}}, \lambda_{\mathrm{C}}) H^+(U, U_0, n_{\mathrm{U}}, \lambda_{\mathrm{U}}) - \gamma_{\mathrm{C}} C \tag{S5}$$

where the new parameters are defined as follows: $\Gamma_{\mathrm{U}} = g_{\mathrm{U}}/P_0 k_{\mathrm{H}}$, $\alpha = A/P_0$, $\beta = B/P_0$, $\gamma_{\mathrm{U}} = k_{\mathrm{U}}/k_{\mathrm{H}}$, $\gamma_{\mathrm{M}} = k_{\mathrm{M}}/k_{\mathrm{H}}$, $\Gamma_{\mathrm{C}} = g_{\mathrm{H}} g_{\mathrm{C}}/k_{\mathrm{H}}^2 P_0$. The parameters of the shifted Hill functions were rescaled as follows: the delay term τ_{d} becomes $\tau_{\mathrm{d}}/k_{\mathrm{H}}$, the threshold levels P_{M0} becomes P_{M0}/P_0, and the Hill coefficients and fold-changes were already dimensionless. The parameters for androgen treatment U and U_0 were chosen in the dimensionless model, and therefore need no rescaling (see the "Parameters estimation" section). The set of equation S1-S4 was used for the numerical simulation of the Page4 circuit (see "Details on numerical simulation").

5 Parameters estimation

All the model's parameters are presented in Table S3. The parameters H, γ_{U}, α, β, γ_{M}, Γ_{C}, τ_{d}, P_{M0}, n_{C}, λ_{C} and γ_{C} were taken from the original model of the Page4 circuit[3]. Additionally, WT-PAGE4's production and degradation rates were chosen as $\Gamma_{\mathrm{U}} = 70$ and $\gamma_{\mathrm{U}} = 0.4$. $\Gamma_{\mathrm{U}} = 70$ was established phenomenologically to guarantee a level similar to that of HIPK1-PAGE4 and CLK2-PAGE4. Further, $\gamma_{\mathrm{U}} = 0.4$ considered that the half-life of the wild-type is 2-3 times larger than that of CLK2-PAGE4[4], therefore $k_{\mathrm{U}}/k_{\mathrm{H}} = 0.4$.

The parameters for ADT were estimated phenomenologically: since the treatment level U is a parameter, the strength of the therapy depends only on the ratio U/U_0: ratio of 1 represents no effect while a ratio > 1

represent an applied therapy (see the definition of the shifted Hill function in the Methods section). Therefore, we chose a level $U/U_0 = 3$ yielding a strong therapy (or, $U = 3$ and $U_0 = 1$, where both parameters are normalized by U_0). Additionally, $n_{\mathrm{U}} = 2$ and $\lambda_{\mathrm{U}} = 3$. To model the overexpression of AR during BAT therapy, we used a shifted Hill function with the same threshold and Hill coefficient but the new fold-change $\lambda_{\mathrm{BAT}} = 0.1$.

Role	Name	Value
HIPK1 Enzyme level	H	50
WT-PAGE4 degradation rate	γ_{U}	0.1
WT-PAGE4 threshold	α	5
HIPK-PAGE4 threshold	β	20
HIPK1-PAGE4 degradation rate	γ_{M}	0.25
CLK2 production rate	Γ_{C}	20
CLK2 activation Delay	τ_{d}	1.5
HIPK1 threshold for CLK2 activation	P_{M0}	20
Hill coefficient for CLK2 activation	n_{C}	4
Fold-change for CLK2 production	λ_{C}	10
CLK2 degradation rate	γ_{C}	1.0
WT-PAGE4 production rate	Γ_{U}	70
CLK2-PAGE4 degradation rate	γ_{H}	1.0
ADT/BAT level	U	3
Threshold for ADT/BAT	U_0	1
Hill coefficient for ADT/BAT	n_{U}	2
Fold-change of CLK2 production for ADT	λ_{U}	3
Fold-change of CLK2 production for BAT	λ_{BAT}	0.1

Table S3: A summary of the parameters used in the mathematical model.

6 Details on numerical simulation

The dimensionless system of Equations S1-S4 were solved numerically using a direct Eulerian integrator with a timestep $dt = 0.001$. All code was developed in Python by FB, and it is available online in the Github page of FB (https://github.com/federicobocci91). All figures regarding the Page4 circuit were realized with the Matplotlib Python library[5].

Addressing the timescale of oscillations requires converting the time t to the original physical measure. The physical time was obtained by multiplying the arbitrary simulation time by the factor $\tau = 1/k_{\mathrm{H}}$. In all figures, the time on the x-axis has been rescaled to its physical value to appreciate the timescale for oscillations.

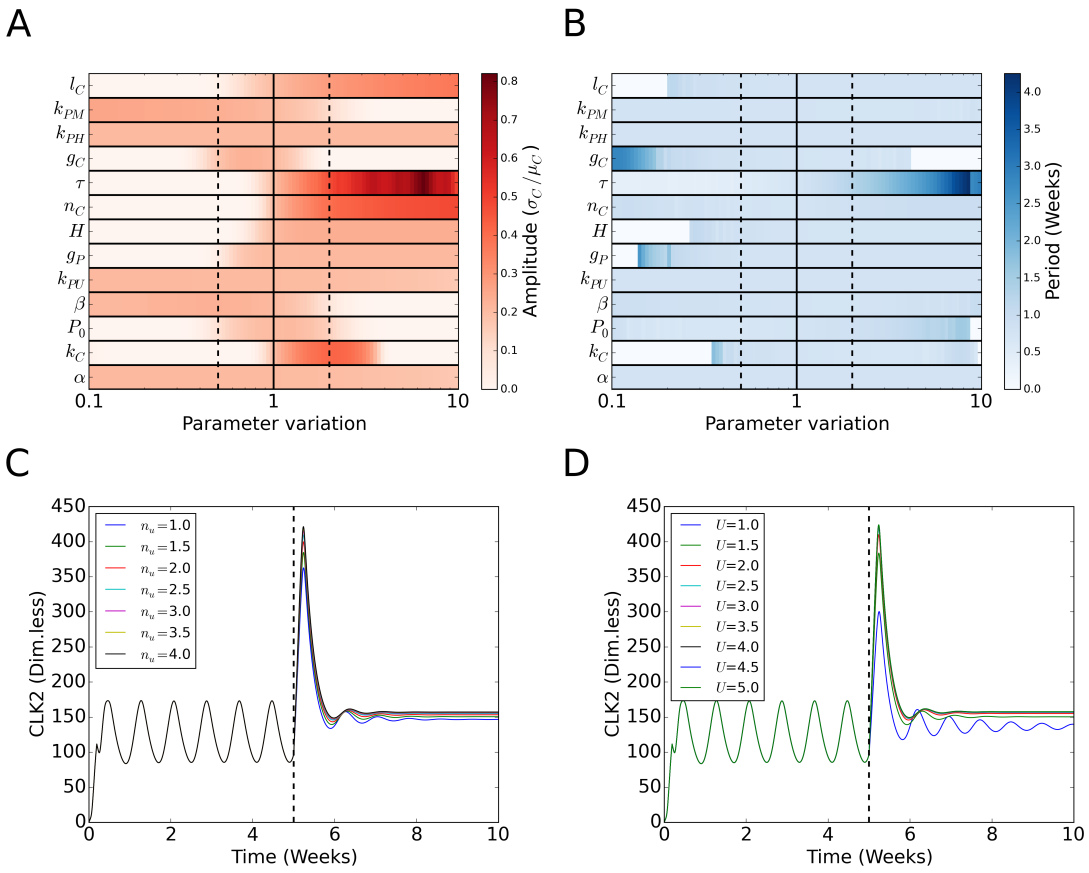

Figure S4: Sensitivity analysis of the model's parameters. **A.** Oscillation amplitude of the cellular level of CLK2 upon a 10-fold variation of the model's parameters. The oscillation amplitude is defined as the ratio between standard deviation and average of CLK2 over the temporal trajectory shown in Figure 5B. The continuous vertical black line depicts the original parameter value, while the dotted vertical black lines show the 2-fold variation interval. **B.** Oscillation period of the cellular level of CLK2 upon a 10-fold variation of the model's parameters. Plain white regions indicate the parameter regions where oscillations are absent (compare to panel A). **C-D** Temporal dynamics of CLK2 in the presence of Androgen deprivation for different values of the Hill coefficient for the treatment inhibition of CLK2 **C.** and treatment intensity **D.**. The vertical dotted black line indicates when ADT begins. The values used in Figure 5C-E are $n_{\mathrm{U}} = 2$, $U = 2$.

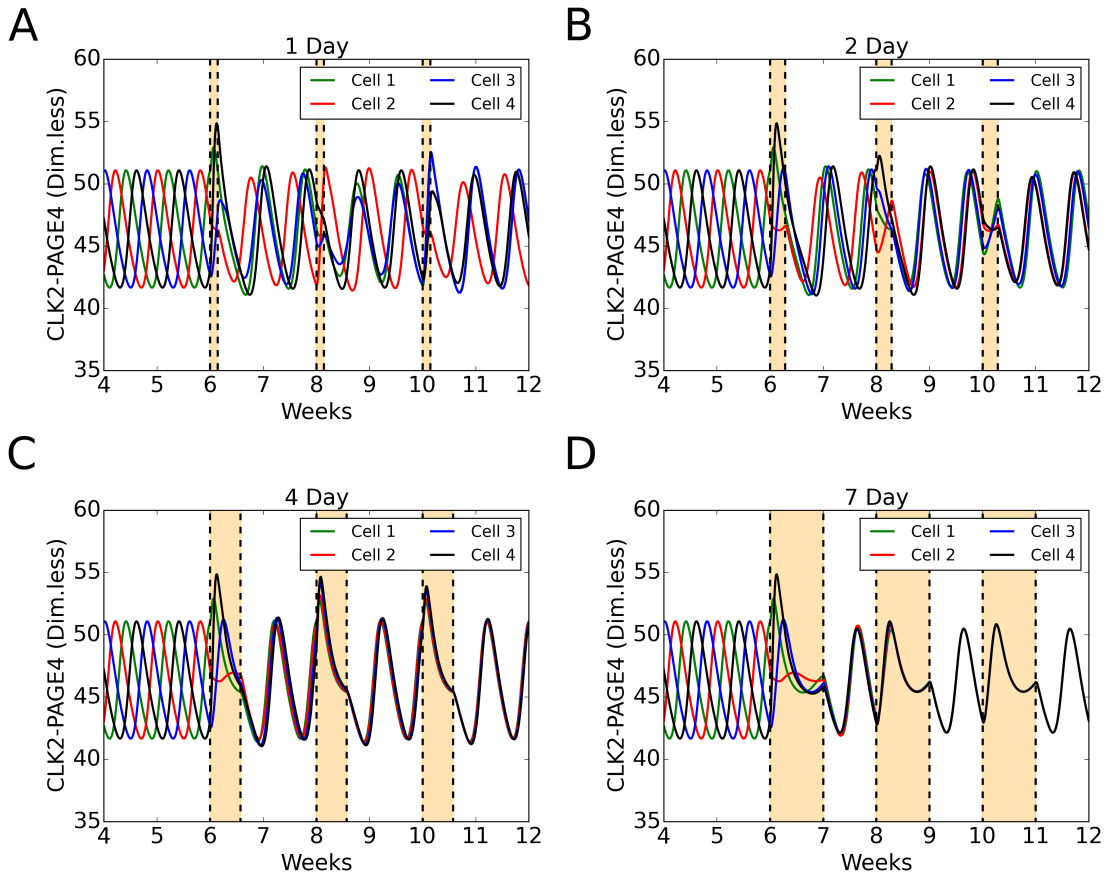

Figure S5: ADT for varying timescales. Simulations showing the dynamics of CLK2-PAGE4 as affected by varying timescales for inhibiting ADT (as shown in orange rectangles) — 1 day **A.**, 2 days **B.**, 4 days **C.**, and 7 days **D.**.

References

[1] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, pp. 105–132, May 1982.

[2] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?," *Proteins: Structure, Function, and Genetics*, vol. 41, pp. 415–427, Nov. 2000.

[3] P. Kulkarni, M. K. Jolly, D. Jia, S. M. Mooney, A. Bhargava, L. T. Kagohara, Y. Chen, P. Hao, Y. He, R. W. Veltri, A. Grishaev, K. Weninger, H. Levine, and J. Orban, "Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. E2644–E2653, Mar. 2017.

[4] S. M. Mooney, R. Qiu, J. J. Kim, E. J. Sacho, K. Rajagopalan, D. Johng, T. Shiraishi, P. Kulkarni, and K. R. Weninger, "Cancer/testis antigen PAGE4, a regulator of c-Jun transactivation, is phosphorylated by homeodomain-interacting protein kinase 1, a component of the stress-response pathway," *Biochemistry*, vol. 53, pp. 1670–1679, Mar. 2014.

[5] J. D. Hunter, "Matplotlib: A 2d Graphics Environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.