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 Schafer[1,3], Min-Yeh Tsai[1,3], Yihong Chen[4], Yanan He[4], Alexander Grishaev[4], Keith Weninger[6], John Orban[4,5], Prakash Kulkarni[4], Govindan Rangarajan[7,8], Herbert Levine[1,2], and José N. Onuchic[1,2,3,9†]

[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, College Park, MD
[5]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD
[6]Department of Physics, North Carolina State University, Raleigh, NC
[7]Department of Mathematics, Indian Institute of Science, Bangalore
[8]Center for Neuroscience, Indian Institute of Science, Bangalore
[9]Department of BioSciences, Rice University, Houston, TX
[†]To whom correspondence should be addressed: jonuchic@rice.edu

Abstract

Prostate-Associated Gene 4 (PAGE4) is a disordered protein implicated in the progression of prostate cancer. PAGE4 can be phosphorylated at two residue sites by Homeodomain-Interacting Protein Kinase 1 (HIPK1) to facilitate its binding to the Activator Protein-1 (AP-1) transcription factor. In contrast, a further hyperphosphorylation of PAGE4 by CDC-Like Kinase 2 (CLK2) reduces its binding affinity to AP-1, thus affecting the androgen receptor (AR) activity. Both SAXS and smFRET experiments have shown a structural expansion of PAGE4 upon hyperphosphorylation and a significant increase in size at its N-terminal half than that at its C-terminus. To understand the molecular mechanism underlying this structural transition, we performed a series of constant temperature molecular dynamics simulations using Atomistic AWSEM — a

multi-scale molecular model combining detailed atomistic and coarse-grained simulation approaches. Our simulations show that electrostatic interaction drives a transient formation of an N-terminal loop, which causes the change in size for different phosphorylated forms of PAGE4. Phosphorylation also changes the preference of secondary structure formation of PAGE4, which signifies a transition between states that display different degree of disorder. Finally, we construct a mechanism-based mathematical model that allows us to capture the interactions of different forms of PAGE4 with AP-1 and AR, a key therapeutic target in prostate cancer. Our model predicts intracellular oscillatory dynamics of HIPK1-PAGE4, CLK2-PAGE4 and AR activity, indicating phenotypic heterogeneity in an isogenic cell population. Thus, conformational switching among different forms of PAGE4 may potentially affect the efficiency of therapeutic targeting of AR.

1 Introduction

It is now evident that a large fraction of the human proteome comprises of proteins, or regions within proteins, that lack stable structure under physiological conditions [1, 2, 3, 4]. Such unstructured proteins exist in large ensemble of interconverting conformations in order to function and thus, they are referred to as intrinsically disordered proteins (IDPs). Despite lack of rigid structure, IDPs are involved in regulation, signaling, and control of information flow in the system, where binding to multiple partners and high-specificity/low-affinity interactions play a crucial role [5, 6]. Furthermore, disorder is a unique structural feature that enables IDPs to participate in both one-to-many and many-to-one signaling [3, 7]. But when overexpressed, or aberrantly expressed, IDPs are prone to engage in promiscuous interactions and lead to different pathological states [8, 9].

IDPs in general have remained intractable to classical experimental methods and therefore, mechanistic insight into how they function is limited to a few IDPs at best. PAGE4 is an IDP that has been well characterized both structurally and functionally. It has the hallmarks of a proto-oncogene; while it is highly expressed during its development in the fetal prostate [10, 11], it is aberrantly expressed in the diseased gland where it plays an important role in tumorigenesis [10].

PAGE4 is a stress-response factor [10]; it functions as transcriptional coactivator and potentiates transactivation by c-Jun [11], a component of the AP-1 transcription factor complex that can negatively regulate the activity of the androgen receptor (AR) in prostate cancer (PCa) cells [12, 13]. Furthermore, in PCa cells, PAGE4 is phosphorylated by the stress-response kinase Homeodomain-Interacting Protein Kinase 1 (HIPK1) predominantly at Thr51 and to a significantly lower level at Ser9 [14] but is hyperphosphorylated by a second kinase, namely CDC-Like Kinase 2 (CLK2) [15]. Cell-based studies reveal that phosphorylation of PAGE4 by the two kinases leads to opposing functions. HIPK1-phosphorylated PAGE4 (HIPK1-PAGE4) potentiates c-Jun but CLK2-phosphorylated PAGE4 (CLK2-PAGE4) attenuates c-Jun activity [15]. Biophysical measurements employing small-angle X-ray scattering (SAXS), single-molecule fluorescence resonance energy transfer (smFRET), and NMR indicate that HIPK1-PAGE4 exhibits a relatively compact conformational ensemble that binds AP-1, whereas CLK2-PAGE4 is more expanded and resembles a random coil with reduced affinity for AP-1 [15]. Although the data indicated different sizes and functions of PAGE4 with increasing phosphorylation, the atomistic details of critical interactions that are responsible for such conformational changes remain unclear.

To investigate the structural details causing the change of size of PAGE4 upon phosphorylation, and to overcome the time resolution limit associated with these biophysical experiments, molecular dynamics (MD) simulations with a physical-based model are needed as a complementary effort. One molecular dynamics approach based on the funneled energy landscape theory [16, 17, 18] is the associative memory, water mediated, structure and energy model (AWSEM) [19]. AWSEM has been

effective not only in the correct prediction of protein structures and binding interfaces [19, 20], but also in elucidating the molecular mechanisms underlying protein folding and regulation of cellular responses [21]. An important component of the AWSEM force field, called an associated memory term, utilizes local structural signals from external sources, such as protein fragments sharing similar sequence information in the protein database [22].

Recently, we developed a multi-scale atomistic AWSEM model (AAWSEM) [23], which harnesses the power of all-atom explicit-solvent simulations to directly generate atomistic structures for use as associative memory terms of an AWSEM simulation. AAWSEM has been shown to be able to reliably fold both α-exclusive proteins [23] and α/β proteins [24]. In particular, AAWSEM is especially useful for investigating dynamics of disordered proteins such as PAGE4 because the lack of known structure limits the utility of the AWSEM model. In addition, the coarse-grained nature of AAWSEM is suitable for fully sampling the conformational space of PAGE4 dynamics, which is significantly larger for disordered proteins compared to proteins in ordered states.

Here, we performed simulations with AAWSEM to investigate the mechanism(s) underlying structural transitions of the PAGE4 conformational ensemble. Consistent with the experimental findings, our simulations produce a change of size in PAGE4 upon phosphorylation. In addition, we suggest that formation of an N-terminal loop is the source of this variation. After gaining insights into conformational dynamics of different forms of PAGE4 (Wild type un-phosphorylated WT-PAGE4, double-phosphorylated HIPK1-PAGE4 and hyper-phosphorylated CLK2-PAGE4), we developed a mechanism-based mathematical model to describe how switching from one form to another (e.g. from HIPK1-PAGE4 to CLK2-PAGE4) can affect the dynamics of the PAGE4/AP-1/AR regulatory circuit. The oscillations observed in this circuit suggest that PAGE4 conformational dynamics may contribute to phenotypic plasticity, and thus potentially the efficacy of therapeutic targeting of AR in PCa.

2 Results

2.1 The sequence of PAGE4

PAGE4 is a 102-residue disordered protein. Most of the residues are predicted to have $> \%50$ chance to be a random coil by secondary structure prediction tools [25] (except for residue Cys63 and Gln64, which are predicted to be helical and located in the region experimentally found to go transiently into helix) (Figure 1). PAGE4 is also a highly charged protein, with a total of 16 positively charged residues and 24 negatively charged residues, resulting in a net charge of -8. The charged residues are not evenly distributed. The N- (residue ID 4 to 12) and C- (residue ID 82 to 95) motifs are positively charged, and there is a central acidic region (residue ID 43 to 62) that is negatively charged. Also,

there is a nine-residue region (residue ID 65 to 73) which can transiently turn into helix as shown by NMR experiment [26]. HIPK1 phosphorylates two sites in PAGE4, with a partial phosphorylation at Ser9 ($30\% \sim 40\%$ chance, within the N-motif) and a complete phosphorylation of Thr51 ($> 95\%$ chance, within the central acidic region) [26]. CLK2 phosphorylates 8 sites of PAGE4, with most of them completely phosphorylated: Ser7 ($> 95\%$), Ser9 ($> 95\%$), Thr51 ($> 95\%$), Thr71 (50%), Ser73 (75%), Ser79 (50%), Thr85 ($> 95\%$) and Thr94 (60%) [15]. Since both Ser7 and Ser9 reside in the N-motif, and Thr85 and Thr94 reside in the C-motif, the hyperphosphorylation neutralizes the N- and C motifs. On the other hand, Thr51 resides in the central acidic region, and Thr71 and Ser73 reside in the transient helix region. The phosphorylation of these amino acids will likely change the local arrangement of residues around this region, contributing to the shift of structural ensemble.

2.2 The Phosphorylation by HIPK1 and CLK2 Changes the Size of PAGE4

Our simulations indicate distinct changes in the size of PAGE4 upon different levels of phosphorylation. The size of the protein can be measured by the radius of gyration (Rg) of the protein during these simulations. The preference of the size of the protein structures in our simulation can be captured by a free energy plot (Figure 2 A). The free energy F was calculated as $F = -k_B T log(P)$ where P is the probability for the proteins to be at a specific Rg. Thus, the lowest free energy in this plot corresponds to the Rg of the most populated ensemble in the conformational space. It is clear that CLK2-PAGE4 is more expanded compared with HIPK1-PAGE4, which is also slightly more compact than WT-PAGE4.

In order to quantify the size of different forms of PAGE4 — WT-PAGE4, HIPK1-PAGE4 and CLK2-PAGE4 — and to compare them with experimental measurements [15], we calculated the average Rg of each form of PAGE4 following equation (1), as given below:

$$\langle Rg \rangle = \int p(Rg) Rg d(Rg) \tag{1}$$

where $p(Rg)$ is the probability for the protein to be at a specific Rg value. The calculated Rgs for three forms of PAGE4 are presented in the second column of Table 1. The simulations qualitatively reproduce the change of size upon phosphorylation, as seen in the SAXS experiments [15]. Moreover, the calculated $\langle Rg \rangle$ is also quantitatively similar to the experimental values (Table 1).

2.3 Phosphorylation expands the structure at the N-terminal end of PAGE4 more than that of the C-terminus

Aside from the SAXS experiments, the smFRET experiments helped us gain further insight into the expansion of PAGE4 upon phosphorylation [15]. The experiment found that the size of the

	$\langle \mathrm{Rg}(EXP) \rangle$ [a] (Å)	$\langle \mathrm{Rg}(SIM) \rangle$ [b] (Å)	RMS Dist 18_63 (EXP) (Å)	RMS Dist 18_63 (SIM) (Å)	RMS Dist 63_102 (EXP) (Å)	RMS Dist 63_102 (SIM) (Å)
Wildtype Form	36 ± 1.1	32.9	56	57.4	50	51.2
HIPK1 Form	34.7 ± 1.2	32.1	59	55.6	50	51.4
CLK2 Form	49.8 ± 1.9	41.8	75	73.4	55	54.8

[a] EXP stands for results from experiments [15]
[b] SIM stands for results from simulations

Table 1: A summary of the results from the simulations of different forms of PAGE4.

N-terminal half of the protein, manifested as the distance between the donor and acceptor dyes tagged at the residues 18 and 63 (Dist18_63), had a more significant increase of size in the hyperphosphorylated form (56 to 75 Å) than in its double-phosphorylated variant (56 to 59 Å). In comparison, the size of the C-terminal PAGE4 is less sensitive to phosphorylation. We find that the distance between residue 63 and 102 (Dist63_102) changes from 50 to 55 Å (Table 1) upon hyperphosphorylation, while we do not observe any noticeable change upon double-phosphorylation.

Our simulations reproduce distances that are similar to experimental observations (Table 1). Since our coarse-grained simulation model uses three beads (C_α, C_β and O atoms) to represent a residue, we approximated the distance between the probes of two residues as the distance between their C_α atoms. Consistent with the observations, our simulations show that Dist18_63 changes from 57.4 Å in wildtype to 73.4 Å in CLK2-PAGE4, while Dist63_102 changes from 51.4 Å to 54.8 Å.

2.4 The structural expansion in the N-terminus is associated with the loss of a loop formation

MD simulations enable us to probe the structural details of PAGE4 dynamics. The collected ensembles of simulated structures show that the expansion of PAGE4 upon hyperphosphorylation is caused by the loss of a N-terminal loop formation between the N-Motif and the central acidic region (Figure 3). Structurally, the N-motif has an excess of basic residues in both WT-PAGE4 and HIPK1-PAGE4 to allow the N-motif to easily interact with the central acidic region around residue 40. These specific interactions lead to frequent loop formation. In contrast, the N-motif is neutralized upon hyperphosphorylation, which causes a loss of this loop formation and ultimately, a more expanded CLK2-PAGE4. To visualize this loop formation, we plot the average contact maps of proteins during their simulations (Figure 3). Because the electrostatic interactions were simulated in a

Debye-Huckel form with a Debye length of 10.0 Å (see Method Section 4.1 for more details), we used the cutoff value of 20.0 Å between the C_β atoms for creating this contact map. This choice of cut-off value captures the electrostatic interactions between residues before they decay to 20% of its value at Debye length. Also, since we are mostly interested in non-local interactions, contacts are defined only for the residues that are at least three residues apart from each other in sequence. As shown in Figure 3B, in both WT-PAGE4 and HIPK1-PAGE4, the N-motif region has a non-zero probability to form contacts with the central acidic region, while no contacts are formed in CLK2-PAGE4. There are also interactions being formed between the C-motif and the central acidic region in WT-PAGE4 and HIPK1-PAGE4, though the probability is smaller.

Although the averaged contact map demonstrates a higher likelihood for the N-terminal loop that forms in WT-PAGE4 and HIPK1-PAGE4 than in CLK2-PAGE4, it lacks a direct description of any dynamic features. To understand more regarding collective conformational movements, we used a Principal Component Analysis (PCA) calculated based on the contacts to probe the dominant mode of dynamics, motivated by the use of similar approaches before [29, 30]. We plot the coefficients of the top two principal components, which correspond to the variation of each contact within the simulated ensembles. A larger value of coefficient means more variation of the corresponding contact movement in that principal mode. In WT-PAGE4 and HIPK1-PAGE4, the calculated PC1 shows a collective motion of the N-motif towards (or away from) the central acidic region (the blue halo between residues ~ 10 and residues ~ 50 in Figure 3C), indicating a frequent loop formation in the N-terminal end of the proteins. In contrast, in CLK2-PAGE4, regular motions only happen among residues local in sequence. Interestingly, we also observe collective motion between the C-motif and the central acidic region in HIPK1-PAGE4. The principal component coefficients of this motion have the opposite sign to that of the N-motif (the red halo between residues ~ 50 and residues ~ 90 in Figure 3C), albeit much weaker than that of the N-motif. The observed PCA patterns thus suggest some competitive motions of the N, C-motif, which take turns to form a loop with the central acidic region. The observation of loop formation also hints at a fly-casting motion frequently observed in the study of disordered proteins [31, 32, 33].

2.5 Simulations find changes in turn-like structures upon phosphorylation

In addition to indicating the expansion of protein size upon hyperphosphorylation, NMR experiments suggested that the more compact structure of HIPK1-PAGE4 is accompanied by an increase of turn-like structures in the central acidic region [26]. Our simulations indeed find an increase of turn structure in this region also (Figure 4). The most prominent increase in turn structure occurs between Arg48 and Thr51, where the probability of forming a turn structure increases from 40% to 80% (Figure 4A). This change is probably due to the phosphorylation of Thr51. The simulations also

display a more populated turn structure in the same region upon hyperphosphorylation. Because no other residues close to Thr51 are further phosphorylated upon hyperphosphorylation, CLK2-PAGE4 shares the same structural features as HIPK1-PAGE4 in the central acidic region.

Two other residues in the transient helix region (Thr71 and Ser73) are further phosphorylated by CLK2. Phosphorylation of these residues causes a reduction of turn structure in this region (Figure 4A). The simulation time trace of turn-like structure in this region also changes upon phosphorylation (Figure 4B). The change of turn-like structure signifies a transient order-disorder transition in this local region. It is hypothesized that the binding site for the AP-1 complex is located in this region of PAGE4 [26]. Therefore, our simulations suggest that the hyperphosphorylation of PAGE4 by CLK2 results in a loss of order in the transient helix, which may then be responsible for a decrease in binding affinity between PAGE4 and the AP-1 complex.

2.6 Implications of switching between CLK2-PAGE4 and HIPK1-PAGE4 for plasticity in prostate cancer cells

To investigate the consequences of PAGE4 structural dynamics in modulating cellular plasticity, we constructed a coarse-grained mathematical model for PAGE4 intracellular dynamics and its implications in interacting with the AP-1/AR signaling axis (Figure 5A). The model considers the three relevant PAGE4 forms (WT-PAGE4, HIPK1-PAGE4 and CLK2-PAGE4) and the enzymes catalyzing the reactions among them (HIPK1, CLK2). As mentioned above, HIPK1-PAGE4 and CLK2-PAGE4 have significant differences in their potentiation of c-Jun and binding to AP-1. Because c-Jun potentiation can indirectly increase the levels of CLK2, a negative feedback loop comes into play that can give rise to oscillatory behavior for the levels of HIPK1-PAGE4, CLK2-PAGE4 and consequently AR activity [15]. This PAGE4/AP-1/AR circuit can generate oscillations between an androgen-dependent phenotype (characterized by high levels of HIPK1-PAGE4) and an androgen-independent phenotype (characterized by high levels of CLK2-PAGE4) [15]. Here, we extended this model by including the temporal dynamics of WT-PAGE4 and modeling different hormone-deprivation treatments for prostate cancer (see Methods section 4.5 for further details). The extended model exhibited similar phenotypic oscillations between a (high HIPK1-PAGE4, low CLK2-PAGE4) state and a (low HIPK1-PAGE4, high CLK2-PAGE4) state when no treatment was applied (Figure 5B) in quite a robust manner; a 10-fold variation of model parameters also exhibited oscillations (Figure S3 A-B).

First, we establish the period of oscillation. The relatively large half-life of HIPK1-PAGE4 (≈ 150 hours) [14] enables an oscillation period of approximately a week (Figure 5B). This timescale is comparable with various hormone-deprivation treatments applied in the clinic [35, 36] and motivated investigating the effect of various therapies. As a first step, we considered a constant inhibitory signal acting on the Androgen Receptor (AR), to simulate the effect of continuous Androgen Depri-

vation Therapy (ADT) — the standard care of therapy for over 75 years for locally advanced and/or metastatic PCa [35]. ADT quenched the oscillations and pushed the cell to a resistant phenotype (Figure 5C). This result depended on the assumed fold-change in the decrease of AR activity levels, but depended weakly on the specific mathematical form chosen to model ADT (Figure 5D, S3 C-D).

Next, we simulated continuous ADT in a cohort of 10000 independent prostate cancer cells. Each cell displays oscillations with the same period, but they are not synchronized, i.e. their oscillations have a non-zero phase difference. Therefore, the distribution of intracellular levels of CLK2 is quite broad before ADT is applied (Figure 5E, Day 0). After a week of therapy, the heterogeneity is considerably decreased, and after 2 weeks, all cells present similar levels (compare the distribution of CLK2 levels in Day 0, Day 7 and Day 14 in Figure 5E). Thus, ADT for a period of 2 weeks can not only quench the oscillations, but also limit phenotypic heterogeneity significantly in a prostate cancer cell population. It should be noted that our model focuses on cells that are not killed by therapeutic treatment. Thus, on one hand, cells with an androgen-dependent phenotype are highly likely to be killed by ADT, and on the other, ADT limits phenotypic heterogeneity among the cells that are not killed by ADT.

Motivated by these results, we implemented an Intermittent ADT [35] where cells are periodically exposed to ADT for a limited time (2 weeks on — 2 weeks off, in the simulation). Oscillations disappeared during the cycles of ADT, but emerged once ADT was removed (Figure 6A). Intriguingly, intermittent ADT coupled the oscillations of unsynchronized cells. We set up a simulation of 4 unsynchronized cells and tracked their level of CLK2-PAGE4. The oscillations became synchronized after the first round of ADT because the treatment pushed all cells to a similar state (Figure 6B). Further, we varied the duration of ADT pulses, and discovered that a single duration as quick as 4 days could synchronize cells in the model (Figure S4). Such synchronization can limit the phenotypic heterogeneity in cells that survive androgen deprivation. Intermittent Androgen Deprivation (IAD) therapy has been studied both in xenografts and in the clinic, with varying time-periods of androgen deprivation and the drug holiday in between [37]. Our model simulations offer a plausible explanation for why the period of androgen deprivation should be at least on the order of a few weeks.

Last, we considered the Bipolar Androgen Treatment (BAT), another novel treatment for prostate cancer [36]. This therapy considered 4 weeks of ADT, while an AR overexpression signal was on for the first 2 weeks. BAT suppressed oscillations during both treatment phases, but the cell was pushed to different states (Figure 6C). The activation of AR and the consequent inhibition of CLK2 enforced a higher level of HIPK1-PAGE4. Conversely, the last 2 weeks of ADT led to the already observed resistant phenotype with low HIPK1-PAGE4 (compare pink-shaded and orange-shaded regions in Figure 6C). Similar to the intermittent ADT case, BAT synchronized phenotypic oscillations of otherwise unsynchronized cells by quenching the oscillatory dynamics of PAGE4/AP-1/AR circuit

(Figure 6D). Therefore, our results showed that similar to ADT, both BAT and IAD can also reduce phenotypic heterogeneity in a prostate cancer cell population, thus likely making the population more vulnerable to therapeutic strategies.

3 Discussion

It has become clear that a sizable fraction of eukaryotic proteins are unstructured before binding to a specific target [1]. The plasticity of these protein structures can facilitate the transcriptional and translational process, as well as the regulation of cellular functions. PAGE4 is a prototypical IDP with pleiotropic functions. Despite being considered almost entirely disordered based on data obtained using various biophysical measurements [38], in the present study, our simulations show that PAGE4 has functionally important structural features that are modulated by phosphorylation.

While the biophysical experimental techniques used previously to discern the various PAGE4 conformational ensembles are routinely used to study IDP, as far as we are aware, they have not been used in parallel to make measurements of the same IDP under identical experimental conditions. Furthermore, quite remarkably, the present data show a very high degree of similarity not only among the biophysical techniques used, but also between the theoretically computed and experimentally observed values.

Although the biophysical techniques employed can correctly capture some ensemble-based properties of PAGE4 such as the size and the distance in between two smFRET probes, there is still a gap between the highest time resolution that can be achieved experimentally and that of the structural dynamics of proteins. This problem is especially prominent for a protein that is highly disordered and exists as an ensemble of rapidly interconverting conformers. In this case, MD simulations with a physical force field can yield additional structural details and dynamical features of these proteins. In this study, our MD simulations accurately reproduce the ensemble averaged features of PAGE4 in both the unphosphorylated and phosphorylated states. This consistency provides good evidence that our model is able to correctly represent the structural ensemble seen experimentally.

Our studies also reveal the structural details underpinning the change of size of PAGE4 upon phosphorylation. Frequent N-terminal loop formation in both WT-PAGE4 and HIPK1-PAGE4, as reflected in our PCA analysis, is reminiscent of the "fly-casting" mechanism that has been proposed for other disordered proteins [31, 32, 33]. Here, a protein that typically has a relatively unstructured state can have a greater capture radius for a specific binding site than the folded state with its restricted conformational freedom. In this scenario of binding, it is envisioned that the unfolded state binds weakly at a relatively large distance followed by folding as the protein approaches the binding site, analogous to a fly-casting motion. The structural plasticity of the protein allows for a combination of binding and folding motions, which lowers the free-energy barrier for the protein-DNA

recognition. In this study, PAGE4 is a highly charged molecule and therefore, it is possible that it can bind to both C-Jun of the AP-1 protein complex and DNA (the AP-1 binding site on the promoter). The regular dynamic motion of the N-terminal portion and the frequent loop formation can enlarge the scope for PAGE4 to find its binding partner, which would possibly increase the chance for the c-Jun transactivation. In contrast, CLK2-PAGE4 loses this regular motion with the replacement of a more disordered N-terminal motion. The loss of specificity may lead to a reduced binding affinity of PAGE4 [15] to AP-1, which eventually results in the degradation of PAGE4 by itself.

In addition, our simulations indicate an increase of turn-like structure in the central acidic region upon double- and hyperphosphorylation, and the loss of turn-like structure in the transient helical region upon hyperphosphorylation. Consistent with the former observation, a recent NMR study [26] also suggested that an increase of turn-like structure in the central acidic region was the source of a slightly more compact structure of HIPK1-PAGE4. The loss of turn structure in the transient helix indicates a partial order-disorder transition. The order-disorder transition has been previously shown to be a part of protein function to lower the free energy barrier and increase the kinetic rate of conformational rearrangements [39, 40, 41, 42]. Here, this transition offers another way by which phosphorylation modulates the binding affinity of PAGE4 to the AP-1 protein complex. Because the transient helical region was previously suspected to be a potential binding site of the AP-1 complex [26], this order-disorder transition can serve to disrupt the binding site, leading to the release of AP-1 upon PAGE4 hyperphosphorylation and the attenuation in c-Jun potentiation. Of note, this conjecture is based entirely on the structural analysis of PAGE4. Further studies including simulations for the equilibrated binding of PAGE4 to the AP-1 complex along with the cognate DNA binding site are currently underway.

To gain insight into the functional implications of the conformational dynamics of PAGE4 insofar as its role in state switching is concerned, we designed a coarse-grained mathematical model. The main goal of the model was to discern how the different conformational ensembles of PAGE4 may modulate androgen insensitivity in PCa cells through their opposite interactions with the AP-1/AR axis. Our model predicted oscillations in the levels of the different PAGE4 conformational ensembles as well as HIPK1 and CLK2, the two kinases that are components of the PAGE/AP-1/AR regulatory circuit between an Androgen-Dependent (AD) and an Androgen-Independent (AI) phenotype. These oscillations correspond to a timescale of a week which corresponds approximately to the half-life of HIPK1-PAGE4. These observations support our previous experimental observations that cells in an isogenic population may exhibit phenotypes with varying androgen deprivation sensitivities [43].Therefore, it is plausible that, in addition to extrinsic signals, phenotypic plasticity can also arise from intracellular dynamics; in this case, by “conformational noise” [44] emerging from switching between different conformational states of PAGE4. Moreover, cells in a given population are highly likely to oscillate with some phase difference among each other, giving rise to non-genetic

heterogeneity, a burgeoning therapeutic challenge in PCa [45, 46].

According to our model, all therapies — continuous Androgen Deprivation Therapy (ADT), Intermittent Androgen Deprivation (IAD) and Bipolar Androgen Therapy (BAT) — eliminated the oscillations and pushed cells toward a resistant phenotype. Furthermore, the treatments largely diminished phenotypic heterogeneity in a cohort of prostate cancer cells within 2 weeks. A limitation of our model is that we do not consider cell death being induced by any of the therapies; in other words, the model predictions mentioned above are only for a subset of cells that are not killed during therapeutic treatment. Extending this mathematical formalism to a population-level approach that considers accelerated cell death in the presence of hormonal treatments would represent a step forward toward formulating predictions of potential clinical relevance, especially given that ADT remains the standard of care for metastatic and locally advanced PCa for over 75 years. Oscillations were rescued when the inhibition on AR during IAD was removed. However, due to limited phenotypic heterogeneity, the cell population oscillated in phase after the treatment was applied. The perfect synchronization observed here can potentially also be disrupted by noise and biological effects beyond what is considered in the model. Yet, the hormonal treatment [47] could introduce a coherence time or window where the intracellular temporal dynamics of different cells appear correlated. Such predictions that can be experimentally validated using live-cell reporters for AR activity are currently being explored in our laboratories.

4 Materials and Methods

4.1 The AAWSEM Force Field

We used a recently developed model, Atomistic AWSEM (AAWSEM), for our simulation. AAWSEM is an adapted version of AWSEM. Details of AAWSEM have been elaborated elsewhere [23, 24]. Here we briefly summarize this force field. The AWSEM Hamiltonian is a coarse-grained force field that represents the backbone structure of the simulated proteins using three atoms per residue: C_α, C_β and O atoms. The positions of all the other backbone atoms can be calculated by assuming an ideal planar peptide bond geometry. The Hamiltonian can be decomposed into four components:

$$V_{\text{total}} = V_{\text{backbone}} + V_{\text{non-backbone}} + V_{\text{FM}} + V_{\text{elec}} \tag{2}$$

where $V_{\text{non-backbone}} = V_{\text{contact}} + V_{\text{burial}} + V_{\text{HB}}$. The V_{backbone} term dictates the backbone geometry of the protein. The $V_{\text{non-backbone}}$ term describes the residue-residue interactions that are not related to backbone connectivity. The V_{burial} term is a many-body term that depends on the density of the immediate environs surrounding each residue. The V_{HB} hydrogen bonding term is governing the secondary structure of the protein. In practice, this term is correlated with the in-

put based on secondary structure prediction [48, 25]. The $V_{\text{contact}} = V_{\text{direct}} + V_{\text{water}}$ is a contact term describing the tertiary interaction between two residues. It is composed of two terms. $V_{\text{direct}} = -\lambda_{\text{direct}} \sum_{j-i>9}^{N} \gamma_{ij}(a_i, a_j)\Theta_{ij}^{I}$ captures the direct protein-protein interaction. $V_{\text{water}} = -\lambda_{water} \sum_{j-i>9}^{2} \Theta_{ij}^{II}(\sigma_{ij}^{wat}\gamma_{ij}^{wat}(a_i, a_j) + \sigma_{ij}^{prot}\gamma_{ij}^{prot}(a_i, a_j))$ describes the water-mediated interaction. The γ parameters were optimized based on the training structures culled from the PDB database. The optimization was performed to maximize the ratio of the folding temperature to the glass transition temperature, $T_{\text{f}}/T_{\text{g}}$ [49, 50, 51]. The fragment based associative memory term, V_{FM}, is in the form of $V_{FM} = -\lambda_{FM} \sum\sum exp\left[-(r_{ij} - r_{ij}^{m})^2/2\sigma_{ij}^2)\right]$, where $\sigma_{ij} = |i - j|^{0.15}$ is a separation in protein sequence. r_{ij} is a cartesian distance between the C_α atoms of two residues i and j. r_{ij}^{m} is the corresponding distance in the memory structures. V_{FM} guides the local-in-sequence interactions of the C_α and C_β atoms and can be based on fragment structures obtained in one of the two ways: In conventional AWSEM simulations, V_{FM} comes from existing PDB structures which share a similar local sequence with the simulated sequence. In AAWSEM, V_{FM} comes from exhaustive all-atom explicit-solvent simulations. Since disordered proteins do not usually have solved structures in the PDB database, here we choose to use atomistic simulations to obtain the fragment memory structures.

In the original version of AAWSEM, the full sequence of a protein was initially segmented in overlapping parts based on the output from secondary structure prediction tools such as PSSpred [25]. Next, the segments are simulated using an all-atom mode in explicit solvent. Here, since PAGE4 is predicted by PSSpred to be almost completely disordered, we segmented it according to the empirical motifs that have been annotated based on previous experiments [26]. The segmentation scheme is shown in Figure 1. Segments are simulated with an enhanced sampling method CST (Continuous Simulated Tempering) [27, 52] to achieve an equilibrated structural ensemble. The CST was performed in a temperature range of 293-350K and the structures that were sampled at temperatures below 330K were selected for a structural clustering. After the clustering of structures by the "single linkage" algorithm, the central structure was selected as the memory structure. The weight of the V_{FM} term is assigned according to the size of each structural cluster.

The V_{elec} electrostatic term is the Debye-Huckel term which approximates the screening effect of charge-charge interactions in solution. The details of this term were described in a separate paper [53]. Simply put, the form of V_{elec} is as follows:

$$V_{\text{elec}} = C_{\text{scaling}} K_{\text{elec}} \sum_{i<j} \frac{q_i q_j}{\epsilon_r r_{ij}} e^{-r_{ij}/l_D} \tag{3}$$

where $K_{\text{Elec}} = (4\pi\epsilon_0)^{-1} = 332.24\, kcal \cdot \mathring{A} \cdot mol^{-1} \cdot e^{-2}$ and ϵ_r is the dielectric constant of the media. q_i and q_j are the charges of the residues. $l_D = \sqrt{\epsilon_r \epsilon_0 k_B T/2e^2 I}$ is the Debye length where I is the ionic strength of the solutions. Following the previous study, we use $\epsilon_r = 80.0$ and $l_D =$

$10.0\AA$ throughout our simulations to achieve a typical physiological condition. C_{scaling} is a constant we introduced for tuning the relative strength of the electrostatic potential compared with other energy terms of the AWSEM Hamiltonian. In this study, we performed simulations with $C_{\text{scaling}} = 1.0$, 2.0 and 4.0. We found the simulations with $C_{\text{scaling}} = 4.0$ matched the experimental results quantitatively [15]. Smaller values of C_{scaling} resulted in qualitative agreement between simulations and experiments but poorer quantitative agreement (see the Supporting Information Figure S1, Table S1). Thus, the analyses throughout the main text are based on the $C_{\text{scaling}} = 4.0$ simulations. We used the same value of $C_{scaling}$ to study different phosphorylated forms of PAGE4. Thus the changes in the conformations of PAGE4 upon phosphorylation are true predictions of this model. Also, we performed simulations in a neutral pH environment by setting the charges of Glu (Glutamic Acid) and Asp (Asparic Acid) residues to -1, while those of Arg (Arginine) and Lys (Lysine) residues were set to +1.

4.2 Treatment of Phosphorylation

Approaches to simulate phosphorylation have been studied in previous work using the AMH model [29, 30], which was the predecessor of AWSEM model. In that approach, the phosphorylated residues were simulated as "supercharged" Glutamic acid (Glu) residues. We adopted the same method in this study, where both the phosphorylated Ser (serine) and phosphorylated Thr (Threonine) were simulated as a Glus. Since the pKAs of phospho-serine and phospho-threonine were determined to be less than 7 in experiments [54], we simulated the phosphorylated residues with charge -2. Furthermore, because the phosphorylation probability was found to be variable among the different forms of PAGE4, we multiplied the charge by its phosphorylated probability to reflect this variability. For example, the Ser9 was shown by experiment to have a $30\% \sim 40\%$ probability to be phosphorylated in HIPK1-PAGE4, so it was simulated as Glu with charge $35\% \times (-2) = -0.7$. In the all-atom explicit-solvent simulations, the phosphoserine and phosphothreonine were simulated by adding phosphoryl groups onto Ser and Thr and simulated under CHARMM36 potential [55].

4.3 The Change of water mediated interaction for disordered proteins

The γ parameters in V_{contact} of the AWSEM potential were optimized in a self-consistent way to maximize the foldability of proteins [49, 50, 51]. In practice, the optimization is done by maximizing the $T_{\text{f}}/T_{\text{g}}$, or the ratio between the energy gap $\delta E = A\gamma$ and the roughness of the landscape $\Delta E = \gamma B \gamma$. Here A and γ are vectors and B is a matrix. The elements of A and B are defined as:

$$
\begin{aligned}
A_i &= \langle \phi_i \rangle_{\mathrm{mg}} - \phi_{\mathrm{n}} \\
B_{i,j} &= \langle \phi_i \phi_j \rangle_{\mathrm{mg}} - \langle \phi_i \rangle_{\mathrm{mg}} \langle \phi_j \rangle_{\mathrm{mg}}
\end{aligned} \tag{4}
$$

where ϕ_i is the functional form of a particular type of interaction. The "mg" stands for "molten globule" and "n" stands for "native". The optimization objective is the functional $A\gamma - \lambda_1\sqrt{\gamma B \gamma}$, which is maximized with $\gamma \propto B^{-1}A$. In practice, γ parameters were determined in a self-consistent way by iterating through simulating to generate a set of molten-globule structural decoys and recalculating its value until γ converges. Additionally, The collapse temperature of the simulated proteins is related to γ parameters by $T_{\mathrm{c}} = A'\gamma$, whose element is $A'_{\mathrm{i}} = \langle \phi_i \rangle_{\mathrm{mg}}/N$ and N is the number of residues. The degree of collapse of proteins in a simulation with a fixed temperature can be controlled by changing T_{c}. Therefore, we can change the degree of collapse of simulated proteins by shifting the γ parameters.

Our simulation with the default model showed an over-compact structure of PAGE4 (Table S2 and Figure S2). To resolve this problem, we shifted the γ parameters of V_{contact} to make the AWSEM potential, on average, more repulsive. The shifting number used throughout our simulations was -0.7, which ensured the simulated ensemble of WT-PAGE4 has a similar Rg to that measured from the SAXS experiments [15]. We used the same γ values in all of our simulations. In this way, any difference of Rgs among three phosphoforms of PAGE4 from our simulations comes from the effect of phosphorylation.

4.4 Details of Atomistic Simulations

The atomistic simulations were performed for the segments of PAGE4 in an explicit-solvent environment (Figure 1). We use the CST [27] method to speed up the thermodynamic sampling. The temperature was continuously changing ranging from $T = 293 \sim 350K$, and the detailed balance was kept throughout the simulations. The atomistic simulation were carried out with a time step of 2.0 fs. The simulated system was neutralized in a physiological condition (ionic concentration of 0.15M) by Na^+ and Cl^- ions. For each segment, we performed a total of 1 μs of atomistic simulation with CHARMM36m force field, which was calibrated based on the latest forcefield of CHARMM36 for IDPs [56]. After the simulations, we used only data with $T < 330K$ to feed into the AWSEM V_{FM} potential. We also disregarded the first 25 ns of each simulation, which was required for convergence of the CST algorithm.

4.5 Mathematical Model for PAGE4/AP-1/AR regulartory axis

We modelled the PAGE4 circuit depicted in Figure 5A by extending the mathematical framework formulated by Kulkarni et al [15]. The temporal dynamics of WT-PAGE4 (P_{U}), HIPK1-PAGE4 (P_{M}), CLK2-PAGE4 (P_{H}) and the CLK2 enzyme (C) in the prostate cancer cell is described by:

$$\frac{dP_{\mathrm{U}}}{dt} = g_{\mathrm{U}} - g_{\mathrm{M}} H \frac{P_{\mathrm{U}}}{P_{\mathrm{U}} + A} - k_{\mathrm{U}} P_{\mathrm{U}} \tag{5}$$

$$\frac{dP_{\mathrm{M}}}{dt} = g_{\mathrm{M}} H \frac{P_{\mathrm{U}}}{P_{\mathrm{U}} + A} - g_{\mathrm{H}} C \frac{P_{\mathrm{M}}}{P_{\mathrm{M}} + B} - k_{\mathrm{M}} P_{\mathrm{M}} \tag{6}$$

$$\frac{dP_{\mathrm{H}}}{dt} = g_{\mathrm{H}} C \frac{P_{\mathrm{M}}}{P_{\mathrm{M}} + B} - k_{\mathrm{H}} P_{\mathrm{H}} \tag{7}$$

$$\frac{dC}{dt} = g_{\mathrm{c}} H^{+}(P_{\mathrm{M}}(t - \tau_{\mathrm{d}}), P_{\mathrm{M0}}, n_{\mathrm{C}}, \lambda_{\mathrm{C}}) H^{+}(U, U_0, n_{\mathrm{U}}, \lambda_{\mathrm{U}}) - k_{\mathrm{C}} C \tag{8}$$

where $(g_{\mathrm{U}}, g_{\mathrm{M}}, g_{\mathrm{H}}, g_{\mathrm{C}})$ and $(k_{\mathrm{U}}, k_{\mathrm{M}}, k_{\mathrm{H}}, k_{\mathrm{C}})$ are basal production and degradation rates. The term $g_{\mathrm{M}} H \frac{P_{\mathrm{U}}}{P_{\mathrm{U}}+A}$ represents the conversion of WT-PAGE4 (P_{U}) to HIPK1-PAGE4 (P_{M}) via enzymatic activity of Hipk1 (H). Here, A represents the threshold level for the catalytic reaction. Similarly, $g_{\mathrm{H}} C \frac{P_{\mathrm{M}}}{P_{\mathrm{M}}+B}$ represents the conversion of HIPK1-PAGE4 (P_{M}) to CLK2-PAGE4 (P_{H}) via the enzyme CLK2 (C) and threshold B. P_{M} regulates C via the intermediates C-Jun and AR activity. This interaction is modelled via the shifted Hill function [57] $H^{+}(P_{\mathrm{M}}(t-\tau_{\mathrm{d}}), P_{\mathrm{M0}}, n_{\mathrm{C}}, \lambda_{\mathrm{C}}) = \frac{1+\lambda_{\mathrm{C}}(\frac{P_{\mathrm{M}}(t-\tau)}{P_{\mathrm{M0}}})^{n_C}}{1+(\frac{P_{\mathrm{M}}(t-\tau)}{P_{\mathrm{M0}}})^{n_C}}$, where λ_{C} is the production rate fold-change of C in the presence of HIPK1-PAGE4, n_{C} is the Hill coefficient regulating the fold-change dependence on P_{M}, P_{M0} is a threshold concentration for C activation and $P_{\mathrm{M}}(t-\tau_{\mathrm{d}})$ is the level of P_{M} at time $t-\tau_{\mathrm{d}}$. The delay term τ_{d} considers the intermediate steps in the signaling.

The androgen treatments effectively regulate CLK2 via the modulation on AR activity (see Figure 5A), and were therefore modelled with the term $H^{+}(U, U_0, n_{\mathrm{U}}, \lambda_{\mathrm{U}})$. U, U_0, n_{U}, λ_{U} are the treatment level, the threshold for treatment activation, the Hill coefficient and the fold-change in production rate of CLK2. The fold-change for CLK2 production rate was $\lambda_{\mathrm{U}} > 1$ for ADT (inhibition of AR and, therefore, activation of CLK2) and $\lambda_{\mathrm{U}} < 1$ for the overexpression phase of BAT (overexpression of AR, inhibition of CLK2).

A dimensionless version of Equation (5-8) was used for the numerical solution of the model. The Supplementary Information provides details about the dimensionless model (Section "Dimensionless model"), the model's parameters (Section "Parameters estimation") and the numerical solution (Section "Details on numerical simulation").

5 Acknowledgement

This work is supported by the Center for Theoretical Biological Physics sponsored by the NSF (Grant PHY-1427654) and by NSF-CHE 1614101. XL was partially supported by the National Institute of Health Grant R01-GM110310. We also wish to acknowledge support from NIH grants GM062154 (JO) and CA181730 (JO and PK). GR was supported by JC Bose Fellowship and SERB Centre for Mathematical Biology Phase II grant. MKJ was supported by a Gulf Coast Consortia on the Computational Cancer Biology Training Program (CPRIT Grant No. RP170593).

References

[1] P. E. Wright and H. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal of Molecular Biology*, vol. 293, pp. 321–331, Oct. 1999.

[2] A. Dunker, J. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, and Z. Obradovic, "Intrinsically disordered protein," *Journal of Molecular Graphics and Modelling*, vol. 19, pp. 26–59, Feb. 2001.

[3] V. N. Uversky, "Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins," *Journal of Biological Chemistry*, vol. 291, pp. 6681–6688, Mar. 2016.

[4] S. DeForte and V. N. Uversky, "Order, Disorder, and Everything in Between," *Molecules (Basel, Switzerland)*, vol. 21, Aug. 2016.

[5] V. N. Uversky, "Unusual biophysics of intrinsically disordered proteins," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1834, pp. 932–951, May 2013.

[6] V. N. Uversky, "Intrinsic disorder-based protein interactions and their modulators," *Current Pharmaceutical Design*, vol. 19, no. 23, pp. 4191–4213, 2013.

[7] C. J. Oldfield and A. K. Dunker, "Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions," *Annual Review of Biochemistry*, vol. 83, pp. 553–584, June 2014.

[8] T. Vavouri, J. I. Semple, R. Garcia-Verdugo, and B. Lehner, "Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity," *Cell*, vol. 138, pp. 198–208, July 2009.

[9] E. M. Marcotte and M. Tsechansky, "Disorder, promiscuity, and toxic partnerships," *Cell*, vol. 138, pp. 16–18, July 2009.

[10] Y. Zeng, D. Gao, J. J. Kim, T. Shiraishi, N. Terada, Y. Kakehi, C. Kong, R. H. Getzenberg, and P. Kulkarni, "Prostate-associated gene 4 (PAGE4) protects cells against stress by elevating p21 and suppressing reactive oxygen species production," *American Journal of Clinical and Experimental Urology*, vol. 1, no. 1, pp. 39–52, 2013.

[11] K. Rajagopalan, R. Qiu, S. M. Mooney, S. Rao, T. Shiraishi, E. Sacho, H. Huang, E. Shapiro, K. R. Weninger, and P. Kulkarni, "The Stress-response protein prostate-associated gene 4, interacts with c-Jun and potentiates its transactivation," *Biochimica Et Biophysica Acta*, vol. 1842, pp. 154–163, Feb. 2014.

[12] N. Sato, M. D. Sadar, N. Bruchovsky, F. Saatcioglu, P. S. Rennie, S. Sato, P. H. Lange, and M. E. Gleave, "Androgenic induction of prostate-specific antigen gene is repressed by protein-protein interaction between the androgen receptor and AP-1/c-Jun in the human prostate cancer cell line LNCaP," *The Journal of Biological Chemistry*, vol. 272, pp. 17485–17494, July 1997.

[13] K. Tillman, J. L. Oberfield, X. Q. Shen, A. Bubulya, and L. Shemshedini, "c-Fos dimerization with c-Jun represses c-Jun enhancement of androgen receptor transactivation," *Endocrine*, vol. 9, pp. 193–200, Oct. 1998.

[14] 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.

[15] 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.

[16] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, pp. 7524–7528, Nov. 1987.

[17] P. E. Leopold, M. Montal, and J. N. Onuchic, "Protein folding funnels: a kinetic approach to the sequence-structure relationship.," *Proceedings of the National Academy of Sciences*, vol. 89, pp. 8721–8725, Sept. 1992.

[18] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, pathways, and the energy landscape of protein folding: a synthesis," *Proteins*, vol. 21, pp. 167–195, Mar. 1995.

[19] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, and G. A. Papoian, "AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing," *The Journal of Physical Chemistry B*, vol. 116, pp. 8494–8503, July 2012.

[20] W. Zheng, N. P. Schafer, A. Davtyan, G. A. Papoian, and P. G. Wolynes, "Predictive energy landscapes for protein-protein association," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 19244–19249, Nov. 2012.

[21] D. A. Potoyan, W. Zheng, E. A. Komives, and P. G. Wolynes, "Molecular stripping in the *NF-B/IB/DNA* genetic regulatory network," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 110–115, Jan. 2016.

[22] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, vol. 28, pp. 235–242, Jan. 2000.

[23] M. Chen, X. Lin, W. Zheng, J. N. Onuchic, and P. G. Wolynes, "Protein Folding and Structure Prediction from the Ground Up: The Atomistic Associative Memory, Water Mediated, Structure and Energy Model," *The Journal of Physical Chemistry B*, vol. 120, pp. 8557–8565, Aug. 2016.

[24] M. Chen, X. Lin, W. Lu, J. N. Onuchic, and P. G. Wolynes, "Protein Folding and Structure Prediction from the Ground Up II: AAWSEM for / Proteins," *The Journal of Physical Chemistry B*, vol. 121, pp. 3473–3482, Apr. 2017.

[25] R. Yan, D. Xu, J. Yang, S. Walker, and Y. Zhang, "A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction," *Scientific Reports*, vol. 3, Dec. 2013.

[26] Y. He, Y. Chen, S. M. Mooney, K. Rajagopalan, A. Bhargava, E. Sacho, K. Weninger, P. N. Bryan, P. Kulkarni, and J. Orban, "Phosphorylation-induced Conformational Ensemble Switching in an Intrinsically Disordered Cancer/Testis Antigen," *Journal of Biological Chemistry*, vol. 290, pp. 25090–25102, Oct. 2015.

[27] C. Zhang and J. Ma, "Enhanced sampling and applications in protein folding in explicit solvent," *The Journal of Chemical Physics*, vol. 132, p. 244101, June 2010.

[28] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, Feb. 1996.

[29] T. Shen, C. Zong, D. Hamelberg, J. A. McCammon, and P. G. Wolynes, "The folding energy landscape and phosphorylation: modeling the conformational switch of the NFAT regulatory domain," *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, vol. 19, pp. 1389–1395, Sept. 2005.

[30] J. Ltzer, T. Shen, and P. G. Wolynes, "Conformational Switching upon Phosphorylation: A Predictive Framework Based on Energy Landscape Principles," *Biochemistry*, vol. 47, pp. 2110–2122, Feb. 2008.

[31] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, "Speeding molecular recognition by using the folding funnel: The fly-casting mechanism," *Proceedings of the National Academy of Sciences*, vol. 97, pp. 8868–8873, Aug. 2000.

[32] Y. Levy, J. N. Onuchic, and P. G. Wolynes, "Fly-Casting in ProteinDNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition," *Journal of the American Chemical Society*, vol. 129, pp. 738–739, Jan. 2007.

[33] E. Trizac, Y. Levy, and P. G. Wolynes, "Capillarity theory for the fly-casting mechanism," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 2746–2750, Feb. 2010.

[34] D. Frishman and P. Argos, "Knowledge-based protein secondary structure assignment," *Proteins: Structure, Function, and Genetics*, vol. 23, pp. 566–579, Dec. 1995.

[35] Z. Kratiras, C. Konstantinidis, and K. Skriapas, "A review of continuous vs intermittent androgen deprivation therapy: redefining the gold standard in the treatment of advanced prostate cancer. Myths, facts and new data on a perpetual dispute," *International Braz J Urol: Official Journal of the Brazilian Society of Urology*, vol. 40, pp. 3–15; discussion 15, Feb. 2014.

[36] M. T. Schweizer, E. S. Antonarakis, H. Wang, A. S. Ajiboye, A. Spitz, H. Cao, J. Luo, M. C. Haffner, S. Yegnasubramanian, M. A. Carducci, M. A. Eisenberger, J. T. Isaacs, and S. R. Denmeade, "Effect of bipolar androgen therapy for asymptomatic men with castration-resistant prostate cancer: Results from a pilot clinical study," *Science Translational Medicine*, vol. 7, pp. 269ra2–269ra2, Jan. 2015.

[37] N. C. Buchan and S. L. Goldenberg, "Intermittent androgen suppression for prostate cancer," *Nature Reviews. Urology*, vol. 7, pp. 552–560, Oct. 2010.

[38] Y. Zeng, Y. He, F. Yang, S. M. Mooney, R. H. Getzenberg, J. Orban, and P. Kulkarni, "The cancer/testis antigen prostate-associated gene 4 (PAGE4) is a highly intrinsically disordered protein," *The Journal of Biological Chemistry*, vol. 286, pp. 13985–13994, Apr. 2011.

[39] O. Miyashita, J. N. Onuchic, and P. G. Wolynes, "Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 12570–12575, Oct. 2003.

[40] K.-i. Okazaki, N. Koga, S. Takada, J. N. Onuchic, and P. G. Wolynes, "Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 11844–11849, Aug. 2006.

[41] P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, "Biomolecular dynamics: orderdisorder transitions and energy landscapes," *Reports on Progress in Physics*, vol. 75, p. 076601, July 2012.

[42] X. Lin, N. R. Eddy, J. K. Noel, P. C. Whitford, Q. Wang, J. Ma, and J. N. Onuchic, "Order and disorder control the functional rearrangement of influenza hemagglutinin," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 12049–12054, Aug. 2014.

[43] N. Terada, T. Shiraishi, Y. Zeng, K.-M. Aw-Yong, S. M. Mooney, Z. Liu, S. Takahashi, J. Luo, S. E. Lupold, P. Kulkarni, and R. H. Getzenberg, "Correlation of Sprouty1 and Jagged1 with aggressive prostate cancer cells with different sensitivities to androgen deprivation," *Journal of Cellular Biochemistry*, vol. 115, pp. 1505–1515, Sept. 2014.

[44] N. Rangarajan, Z. Fox, A. Singh, P. Kulkarni, and G. Rangarajan, "Disorder, oscillatory dynamics and state switching: the role of c-Myc," *Journal of Theoretical Biology*, vol. 386, pp. 105–114, Dec. 2015.

[45] A. Brock and S. Huang, "Precision Oncology: Between Vaguely Right and Precisely Wrong," *Cancer Research*, vol. 77, pp. 6473–6479, Dec. 2017.

[46] D. Jia, M. K. Jolly, P. Kulkarni, and H. Levine, "Phenotypic Plasticity and Cell Fate Decisions in Cancer: Insights from Dynamical Systems Theory," *Cancers*, vol. 9, p. 70, June 2017.

[47] Y. Hirata, K. Akakura, C. S. Higano, N. Bruchovsky, and K. Aihara, "Quantitative mathematical modeling of PSA dynamics of prostate cancer patients treated with intermittent androgen suppression," *Journal of Molecular Cell Biology*, vol. 4, pp. 127–132, June 2012.

[48] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton, "JPred4: a protein secondary structure prediction server," *Nucleic Acids Research*, vol. 43, pp. W389–W394, July 2015.

[49] K. K. Koretke, Z. Luthey-Schulten, and P. G. Wolynes, "Self-consistently optimized energy functions for protein structure prediction by molecular dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 2932–2937, Mar. 1998.

[50] G. A. Papoian, J. Ulander, and P. G. Wolynes, "Role of Water Mediated Interactions in ProteinProtein Recognition Landscapes," *Journal of the American Chemical Society*, vol. 125, pp. 9170–9178, July 2003.

[51] N. P. Schafer, B. L. Kim, W. Zheng, and P. G. Wolynes, "Learning To Fold Proteins Using Energy Landscape Theory," *Israel Journal of Chemistry*, vol. 54, pp. 1311–1337, Aug. 2014.

[52] T. Zang, L. Yu, C. Zhang, and J. Ma, "Parallel continuous simulated tempering and its applications in large-scale molecular simulations," *The Journal of Chemical Physics*, vol. 141, p. 044113, July 2014.

[53] M.-Y. Tsai, W. Zheng, D. Balamurugan, N. P. Schafer, B. L. Kim, M. S. Cheung, and P. G. Wolynes, "Electrostatics, structure prediction, and the energy landscapes for protein folding and binding: Electrostatic Energy Landscapes for Folding and Binding," *Protein Science*, vol. 25, pp. 255–269, Jan. 2016.

[54] Y. Xie, Y. Jiang, and D. Ben-Amotz, "Detection of amino acid and peptide phosphate protonation using Raman spectroscopy," *Analytical Biochemistry*, vol. 343, pp. 223–230, Aug. 2005.

[55] S. S. Mallajosyula, O. Guvench, E. Hatcher, and A. D. MacKerell, "CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates," *Journal of Chemical Theory and Computation*, vol. 8, pp. 759–776, Feb. 2012.

[56] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmller, and A. D. MacKerell, "CHARMM36m: an improved force field for folded and intrinsically disordered proteins," *Nature Methods*, vol. 14, pp. 71–73, Jan. 2017.

[57] M. Lu, M. K. Jolly, H. Levine, J. N. Onuchic, and E. Ben-Jacob, "MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 18144–18149, Nov. 2013.

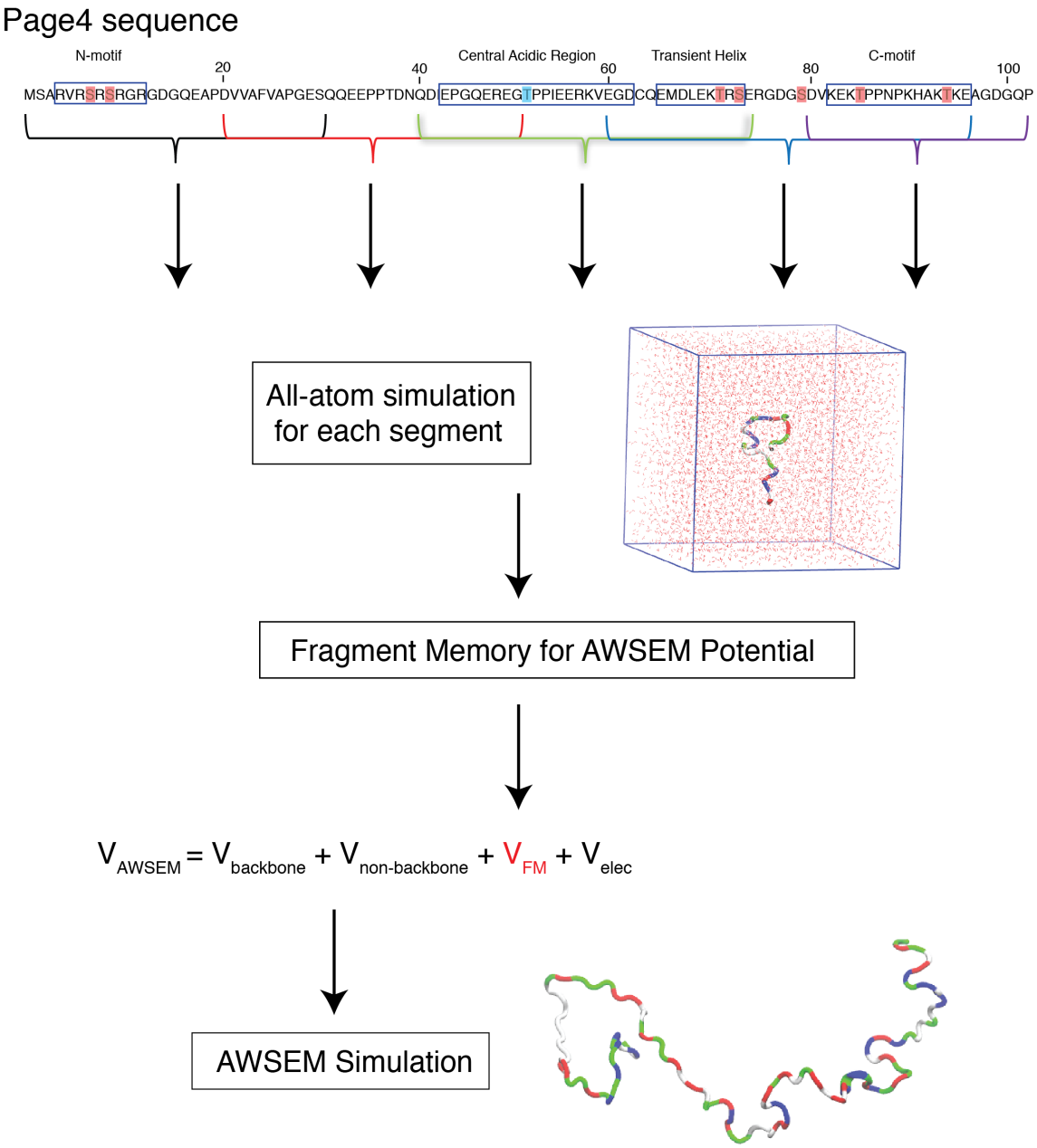

Figure 1: The sequence nature of PAGE4 and the flowchart for the AAWSEM simulation. The PAGE4 sequence was adapted from [15]. The phosphorylation sites in HIPK1-PAGE4 are highlighted in blue, and those in CLK2-PAGE4 are highlighted in red. Each phosphorylated form of PAGE4 protein was separated into 5 segments according to its sequence nature: Segment 1 (residue ID 1 to 30), Segment 2 (residue ID 20 to 50), Segment 3 (residue ID 40 to 74), Segment 4 (residue ID 60 to 96) and Segment 5 (residue ID 80 to 102). Each segment was simulated in a explicit-solvent environment with Continuous Simulated Tempering (CST) [27]. The structural ensemble from CST at T = 293 $\sim$ 330K was clustered into structural ensemble and passed as fragment memories into AWSEM model for the coarse-grained level simulations of the entire PAGE4 (a snapshot of which in simulation is colored by residue type in the right bottom).

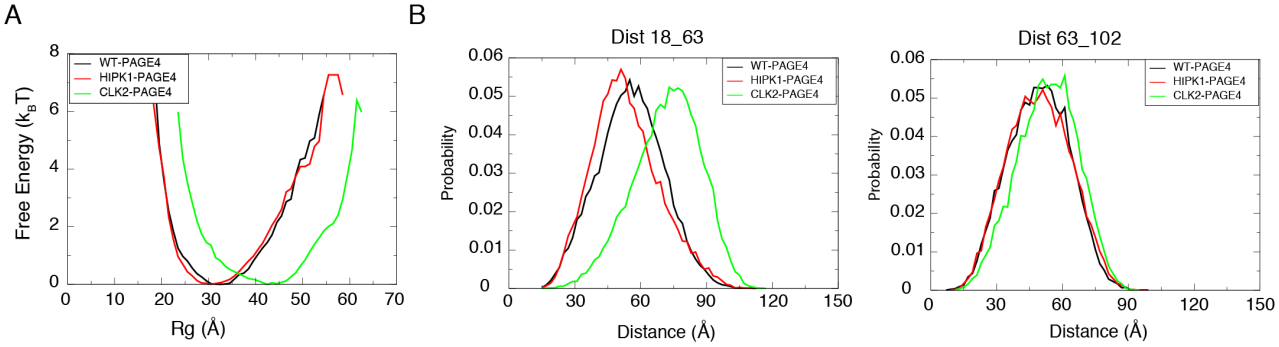

Figure 2: **A.** The free energy plot based on the Radius of Gyration (Rg) of the simulated proteins. **B.** The probability distribution of distance between two residue pairs that were previously measured in smFRET experiments.

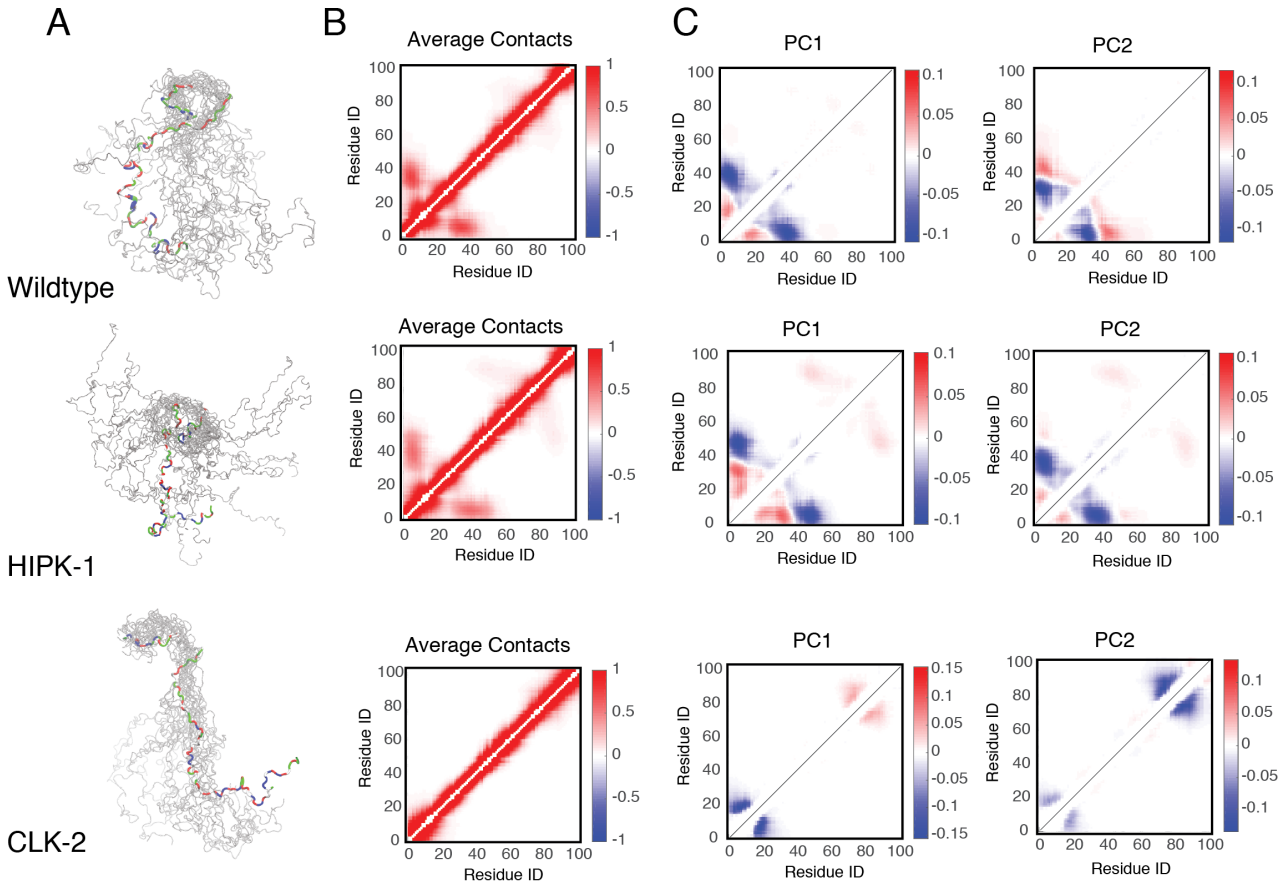

Figure 3: The structure ensemble of different forms of PAGE4 collected from the simulations. **A.** The structural ensemble generated during our simulations. Randomly picked structures are RMSD aligned by the N-motif [28]. It is clear that the N-terminal loop forms in the wildtype and HIPK1 form, while no loops are formed in the CLK2 form. **B.** The average contact maps generated based on the simulations. The color bar shows the probability of contact formation during our simulations. There are non-zero probabilities for interactions between the N-, C-motif and the central acidic region. **C.** The top three principal modes generated from the contact-based principal component analysis. We plot the coefficients of the first two principal components, which correspond to the collective change of each contact.

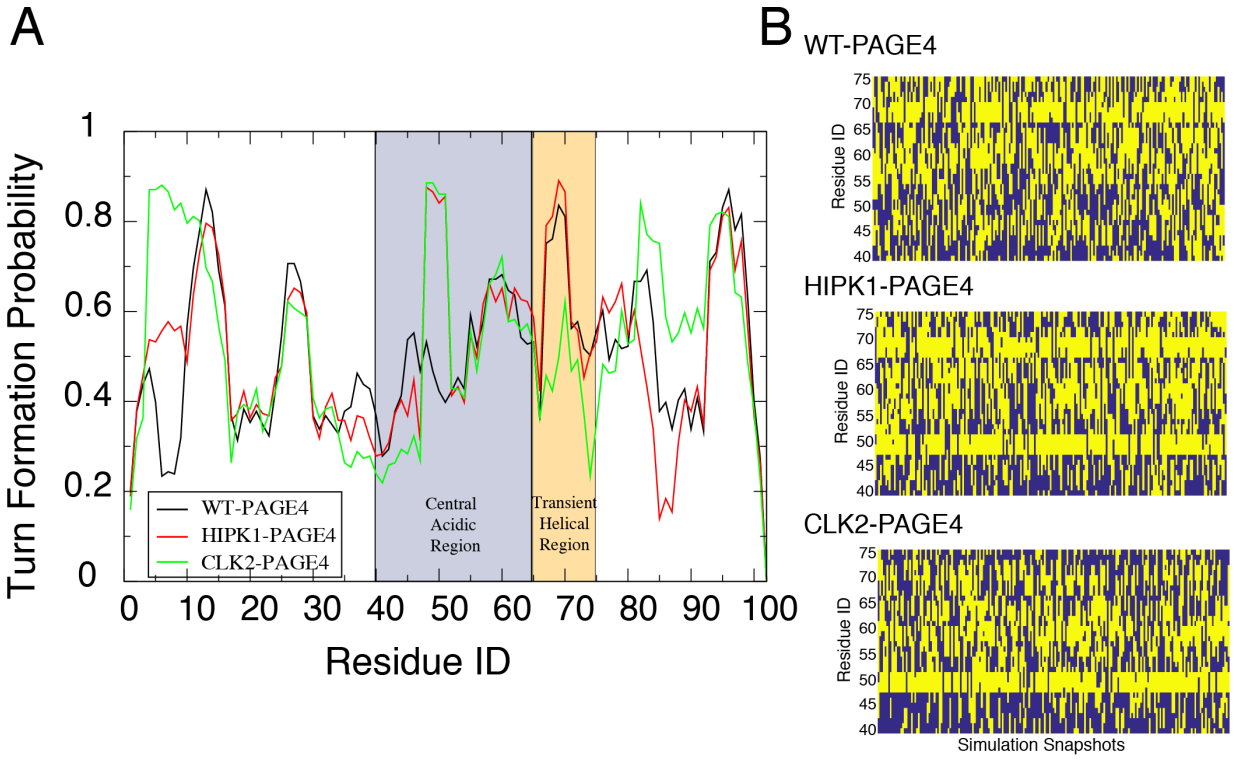

Figure 4: **A.** The probability for each residue to be a turn-like structure. The central acidic region and transient helical region are shaded in blue and orange respectively. The secondary structure was calculated using Stride algorithm [34]. **B.** The dynamics of turn formation per residue along the simulations of each phosphorylated form of PAGE4. We only show the residues in the central acidic region and the transient helical region. Yellow means the residue is identified by Stride algorithm to be a turn structure, while blue means it is not.

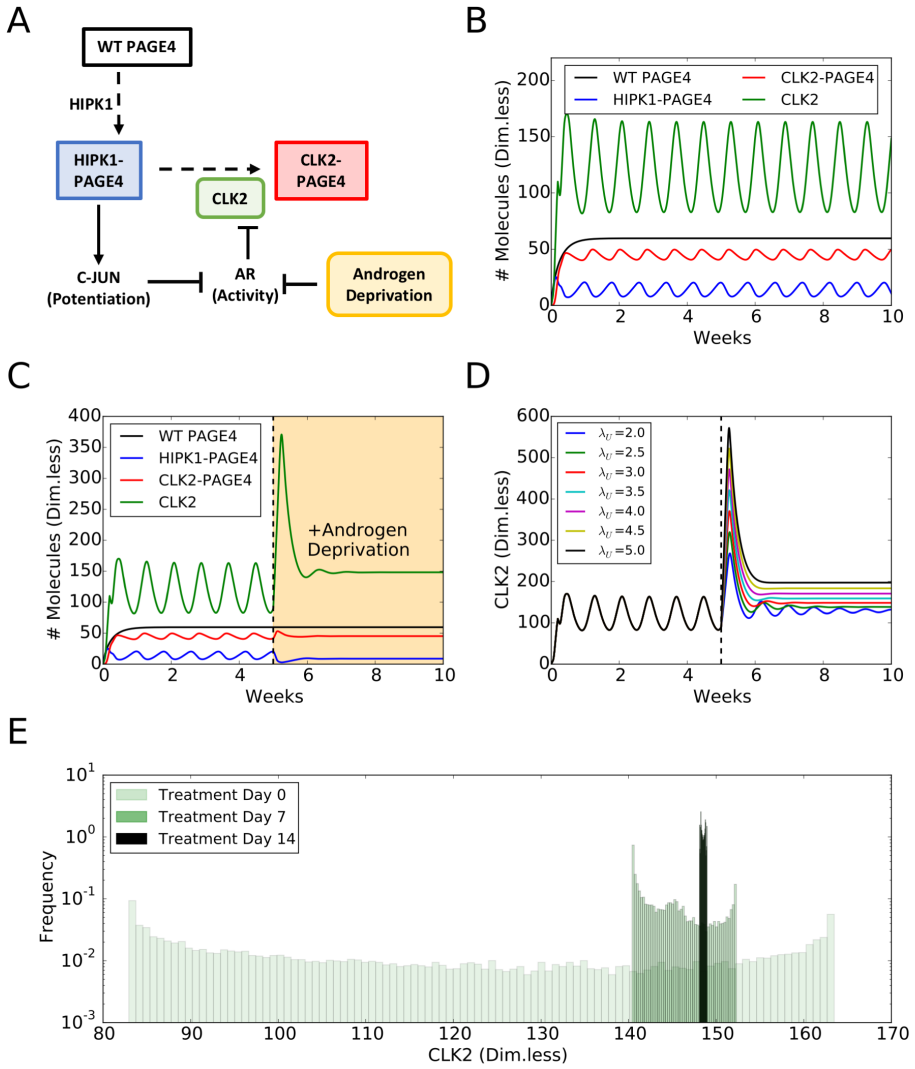

Figure 5: Androgen deprivation suppresses phenotypic oscillations and eliminates heterogeneity in a prostate cancer cell population, as predicted by a mathematical model. **A.** Schematic of PAGE4 phosphorylation circuit and its connection with androgen resistance. Wild-type PAGE4 is double phosphorylated by HIPK1 enzyme, and the HIPK1-PAGE4 complex is hyperphosphorylated via CLK2 enzyme. Further, HIPK1-PAGE4 regulates CLK2 via the intermediates C-Jun and Androgen Receptor (AR). The cell under hormonal therapy is deprived of androgen, thereby decreasing AR activity. **B.** Temporal dynamics of the cellular level of WT PAGE4, HIPK1-PAGE4, CLK2-PAGE4 and CLK2. Without ADT, the oscillatory behaviour exhibits a period of approximately 1 week. **C.** ADT (orange-shaded area) quenches oscillations within approximately 2 weeks. **D.** Temporal dynamics of the CLK2 cellular level for various CLK2 production rate fold-changes λ_{U} due to ADT. Any fold-change larger/equal than 2 is sufficient to quench the oscillations. The value used in (C) is 3. **E.** Distribution of CLK2 cellular level in a simulated cohort of 10000 prostate cancer cells. Before treatment (Day 0), the distribution spans over the whole range of CLK2 levels observed in the single cell simulation of (B). After one week of treatment (Day 7), the distribution considerably shrinks. At Day 14 of treatment, all cells have a similar level of CLK2. In all plots, WT PAGE4, HIPK1-PAGE4, CLK2-PAGE4 and CLK2 are represented in arbitrary dimensionless units.

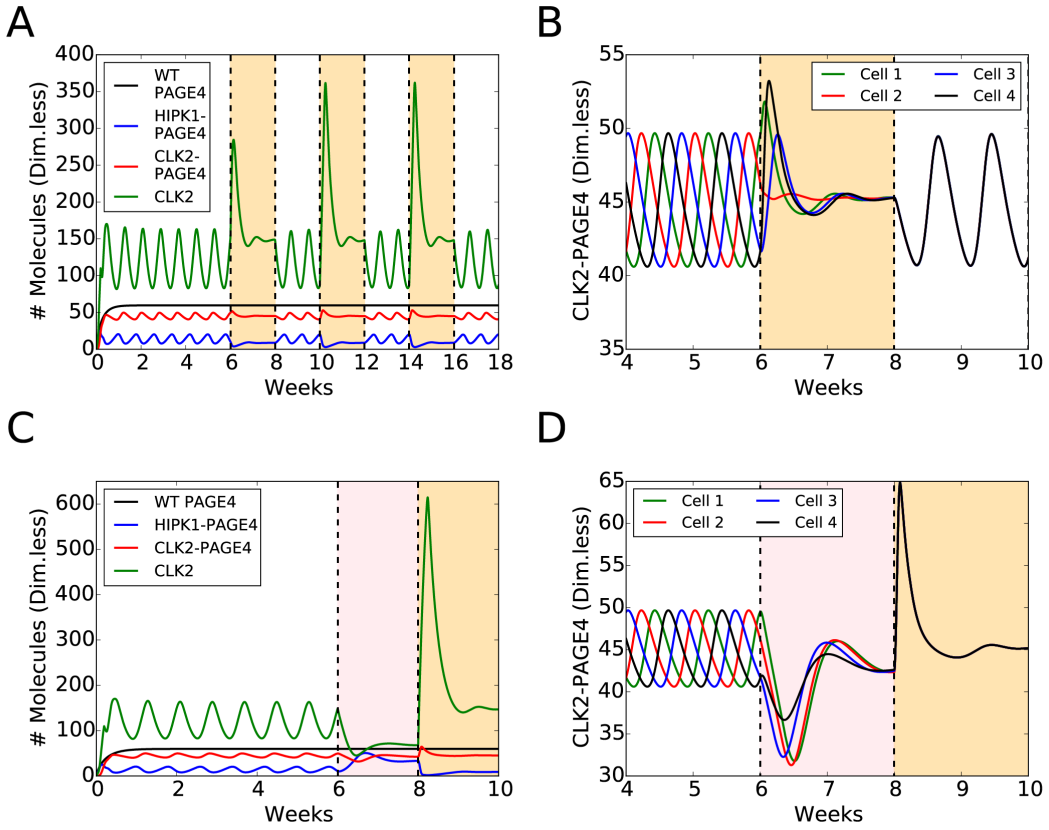

Figure 6: Intermittent ADT and BAT temporarily remove oscillations and synchronize the cells phenotypic dynamics. **A.** Temporal dynamics of WT PAGE4, HIPK1-PAGE4, CLK2-PAGE4 and CLK2 under Intermittent ADT. In the simulation, ADT was applied and removed at regular intervals of 2 weeks (the orange-shaded areas indicate when the therapy is on). **B.** Temporal dynamics of CLK2-PAGE4 in 4 different unsynchronized cells (i.e. the oscillations of CLK2-PAGE4 are not in phase). A 2-weeks pulse of ADT synchronizes the cells phenotypic dynamics. **C.** Temporal dynamics under BAT. The pink-shaded area indicates the 2 weeks of AR overexpression, while the orange-shaded area depicts the subsequent 2 weeks of ADT. Oscillations are quenched in 4 weeks of treatment, but the after 2 weeks, levels of HIPK1-PAGE4 decrease, while those of CLK2-PAGE4 increase. After 2 weeks of ADT, the levels are similar to those observed in continuous ADT. **D.** Similar to Intermittent ADT, BAT synchronizes the cell fates dynamics. In all plots, WT-PAGE4, HIPK1-PAGE4, CLK2-PAGE4 and CLK2 are represented in arbitrary dimensionless units.