Cryptic sequence features within the disordered protein p27^{Kip1} regulate cell cycle signaling

Rahul K. Das[a,1], Yongqi Huang[b,1], Aaron H. Phillips[b,1], Richard W. Kriwacki[b,c,2], and Rohit V. Pappu[a,2]

[a]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130; [b]Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN 38105; and [c]Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Sciences Center, Memphis, TN 38163

Edited by Alan R. Fersht, Medical Research Council Laboratory of Molecular Biology, Cambridge, United Kingdom, and approved April 7, 2016 (received for review August 15, 2015)

Peptide motifs embedded within intrinsically disordered regions (IDRs) of proteins are often the sites of posttranslational modifications that control cell-signaling pathways. How do IDR sequences modulate the functionalities of motifs? We answer this question using the polyampholytic C-terminal IDR of the cell cycle inhibitory protein p27^{Kip1} (p27). Phosphorylation of Thr-187 (T187) within the p27 IDR controls entry into S phase of the cell division cycle. Additionally, the conformational properties of polyampholytic sequences are predicted to be influenced by the linear patterning of oppositely charged residues. Therefore, we designed sequence variants of the p27 IDR to alter charge patterning outside the primary substrate motif containing T187. Computer simulations and biophysical measurements confirm predictions regarding the impact of charge patterning on the global dimensions of IDRs. Through functional studies, we uncover cryptic sequence features within the p27 IDR that influence the efficiency of T187 phosphorylation. Specifically, we find a positive correlation between T187 phosphorylation efficiency and the weighted net charge per residue of an auxiliary motif. We also find that accumulation of positive charges within the auxiliary motif can diminish the efficiency of T187 phosphorylation because this increases the likelihood of long-range intra-IDR interactions that involve both the primary and auxiliary motifs and inhibit their contributions to function. Importantly, our findings suggest that the cryptic sequence features of the WT p27 IDR negatively regulate T187 phosphorylation signaling. Our approaches provide a generalizable strategy for uncovering the influence of sequence contexts on the functionalities of primary motifs in other IDRs.

p27 | motif | disordered regions

Intrinsically disordered proteins and regions (IDPs/IDRs) serve as scaffolds for short linear motifs, which are highly conserved functional sequence modules that are typically 3- to 10-residues long (1, 2). Motifs often mediate protein–protein interactions and they encompass sites of specific posttranslational modifications that control a range of regulatory functions in cells (3). An example of cellular regulation that is achieved through posttranslational modifications of specific residues in motifs is the control of cell cycle arrest through Tyr and Thr phosphorylation in linear motifs within the cell cycle inhibitor protein p27^{Kip1}.

Progression through the mammalian cell cycle is controlled by the interplay between cyclin-dependent kinases (Cdks), cyclins, and the Cip/Kip family of Cdk inhibitors that includes p27^{Kip1}—referred to hereafter as p27 (4). In isolation, p27, which is disordered (5), encompasses two functional domains. The N-terminal domain ($p27_{1–95}$) folds upon binding to Cdk2/cyclin A, resulting in complete kinase inhibition. Conversely, whereas $p27_{1–95}$ folds upon binding to Cdk2/cyclin A, the C-terminal domain ($p27_{96–198}$) remains disordered and participates in a phosphorylation/ubiquitination-signaling cascade (6) (Fig. 1). Proliferative signals cause the activation of nonreceptor tyrosine kinases (NRTKs) that phosphorylate p27 within the Cdk2-bound N-terminal domain at Tyr-88 (Y88) (7). Phosphorylation of Y88 leads to partial restoration of the kinase activity of Cdk2 that is achieved without releasing p27 from the complex with Cdk2/cyclin A; this triggers intracomplex phosphorylation of Thr-187 (T187) near the C terminus of p27, creating a phosphodegron site (8). Subsequent recruitment of the E3 ubiquitin ligase, SCF^{Skp2} (9, 10), catalyzes polyubiquitination of three lysine-containing motifs within $p27_{96–198}$ (11). Polyubiquitinated p27 is then selectively degraded by the 26S proteasome, releasing fully active Cdk2/cyclin A to drive progression of the cell cycle into S phase (7).

T187 is part of the primary substrate motif T_{187}-P_{188}-K_{189}-K_{190}. Phosphorylation of T187 is the key signaling step that leads to p27 degradation and full activation of Cdk2/cyclin A, and this signaling reaction is enabled by disorder within $p27_{96–198}$ (6). Here, we investigate the impact of changes to sequence-encoded disorder on the conformational properties of $p27_{96–198}$ and the efficiency of T187 phosphorylation. This investigation was motivated by recent studies showing that the conformational properties and the degree of disorder of IDPs/IDRs are influenced by sequence attributes such as the net charge per residue (NCPR) and the fraction of charged residues (12–18). Importantly, a majority of IDPs/IDRs, including $p27_{96–198}$, are polyampholytes that are enriched in both positively and negatively charged residues (19–21). In polyampholytes, the sequence patterning of oppositely charged residues is predicted to modulate the degree of chain expansion/compaction and the amplitudes of conformational fluctuations (20). Alterations to the linear patterning of oppositely charged residues leads to changes in the sequence-specific NCPR profiles and these changes can be quantified using a single parameter κ (20). In sequences with low κ values, the oppositely charged residues

Significance

Intrinsically disordered regions (IDRs) of proteins are scaffolds for linear motifs that mediate protein–protein interactions and are the sites of posttranslational modifications. Using the cell cycle inhibitory protein p27 as an archetypal example, we show that the patterning of oppositely charged residues controls the conformational properties of IDRs. Charge patterning also encodes for auxiliary motifs within IDRs. We find that the functionalities of primary motifs are modulated by a combination of the net charge per residue within auxiliary motifs and their intra-IDR interactions that result from sequence-encoded charge patterning. These findings demonstrate that the sequences of IDRs are not just passive scaffolds for motifs. Instead, they encode features that regulate the functions of primary motifs.

Author contributions: R.K.D., A.H.P., R.W.K., and R.V.P. designed research; R.K.D., Y.H., and A.H.P. performed research; R.K.D., Y.H., and A.H.P. contributed new reagents/analytic tools; R.K.D., Y.H., A.H.P., R.W.K., and R.V.P. analyzed data; and R.K.D., A.H.P., R.W.K., and R.V.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]R.K.D., Y.H., and A.H.P. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pappu@wustl.edu or richard.kriwacki@stjude.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516277113/-/DCSupplemental.

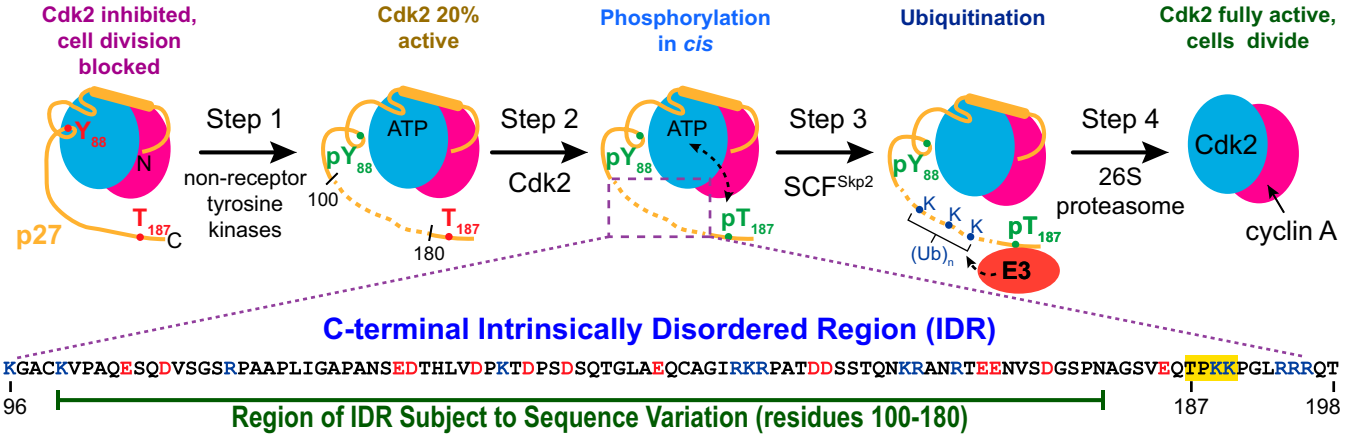

Fig. 1. The p27 IDR integrates proliferative signals from NRTKs to release active Cdk2/cyclin A via T187 phosphorylation. Proliferative signals activate NRTKs that phosphorylate p27 at Y88, partially restoring Cdk2 activity (step 1). This allows for intracomplex phosphorylation of p27 at T187 by Cdk2 and creates a phosphodegron (step 2). Recruitment of the SCF^{Skp2} E3 ubiquitin ligase leads to ubiquitination of lysine residues within p27-C (step 3) followed by selective degradation of p27 by the proteasome (step 4) and release of the fully active Cdk2/cyclin A. The primary sequence of the p27 IDR is shown with positively and negatively charged residues depicted in blue and red, respectively. The primary substrate motif is highlighted in yellow.

are well mixed. In such sequences, intrachain electrostatic repulsions and attractions are mutually screened and the result is a heterogeneous ensemble of expanded, random coil-like conformations (20). In contrast, in sequences with high κ values, the opposite charges are segregated within the sequence and electrostatic attractions dominate. The result is an ensemble of compact conformations.

We used sequence design to uncover the role of charge patterning on the conformations of $p27_{96–198}$ and the impact, if any, on the efficiency of T187 phosphorylation. Specifically, we generated distinct variants of $p27_{96–198}$ by manipulating the charge patterning of its sequence between residues 100 and 180 while keeping the amino acid composition fixed. The sequence of the primary motif and its immediate sequence context, namely, positions 181–198, were also held fixed in the designed variants. In a motif-centric model, the efficiency of T187 phosphorylation should depend only on the presence or absence of the primary substrate motif and be insensitive to sequence changes within the unconstrained region of p27. However, contrary to a purely motif-centric model, we have uncovered cryptic sequence features within $p27_{96–198}$ that regulate T187 phosphorylation. Our findings suggest that the $p27_{96–198}$ IDR is not just a passive tether between the primary motifs encompassing Y88 and T187. Instead, the sequence features within $p27_{96–198}$ actively regulate the coupling of Y88 and T187 phosphorylation.

Results

Design Strategy. For a given sequence stretch, the NCPR is calculated as $(f_+ - f_-)$, where f_+ and f_-, respectively, denote the fraction of positively and negatively charged residues. In our design strategy the amino acid composition of p27 between residues 100 and 180 was fixed, whereas the linear sequence was varied to yield p27 variants with different NCPR profiles and different values of κ (Fig. 2*A*). In a NCPR profile (Fig. 2*B*), the numeric value for each sequence position is the NCPR within a window of five residues that is centered on the residue in question (20). The primary substrate motif is embedded in a region of net positive charge, which is complementary to the charged Cdk2 surface near the binding groove for the substrate (*SI Appendix*, Fig. S1). Therefore, in all of the design variants we fixed the sequence of the C-terminal 18 residues of p27 (181–198). This constraint guards against disruptions to the primary substrate motif. We use the following conventions to refer to the p27 sequence variants: p27 refers to the full-length construct, whereas p27-C refers to a C-terminal construct containing residues 96–198 and a four-residue, nonnative, N-terminal extension. Within full-length p27, we refer to the disordered C-terminal region (residues 96–198) as the p27 IDR. Each variant of p27 is characterized by a distinct amino acid sequence corresponding to positions 100–180 within p27 and is referred to as p27-vXY. Here, XY is a two-digit number that corresponds to the κ value, multiplied by 100, which is calculated over all residues in the p27 IDR. As an illustration, p27-v14 and p27-C-v14 refer to the sequences of p27 and p27-C, respectively, where the sequence of the p27 IDR has a κ value of 0.14.

Our design constraints restrict our choices to sequences within the range $0.1 \leq \kappa \leq 0.8$. Even with this restriction, the number of conceivable sequences is astronomically large (*SI Appendix*, Fig. S2 *A* and *B*). To obtain a systematic assessment of the impact of varying the NCPR profile on the conformational properties and functions modulated by the p27 IDR, we generated 11 sequence variants of the p27 IDR sequence to be compared with the properties of the WT sequence. The κ values of these variants uniformly span the interval $0.1 \leq \kappa \leq 0.8$ (*SI Appendix*, Fig. S2*C*). A total of 5 of the 11 designed sequences were amenable to overexpression in *Escherichia coli* and purification for in vitro studies. These include the p27-C variants expressed as isolated domains and the full-length variants of p27. All biophysical characterization, computer simulations, and biochemical studies of T187 phosphorylation efficiencies were performed using five designed variants (Fig. 2*A*) and the results obtained were compared with those obtained using the WT sequence. In addition to these designs, we included 4 other sequences, and their design was motivated by our functional studies as discussed toward the end of *Results*.

Global Conformational Properties of p27 Variants. We performed atomistic temperature replica exchange Metropolis Monte Carlo simulations (22) using the ABSINTH implicit solvation model and force field paradigm (23) for each of the p27-C-vXY sequences shown in Fig. 2*A*. The results in Fig. 3*A* illustrate the inverse correlation between the ensemble averaged radii of gyration (R_g) and κ. Predictions from simulations were tested using small-angle X-ray scattering (SAXS) measurements to determine the R_g values for each of the p27-C variants. The numerical R_g values obtained from analysis of simulation results and SAXS data are shown in *SI Appendix*, Table S1. The R_g values from simulations and SAXS data are positively correlated with a Pearson r value of 0.97 (Fig. 3*B*). The influence of charge patterning on the dimensions of p27-C-vXY constructs was also reproduced in the context of the full-length p27-vXY constructs bound to Cdk2/cyclin A; this is shown using data obtained from SAXS and size-exclusion chromatography measurements (Fig. 3*C* and *SI Appendix*, Fig. S4 and Table S1).

The impact of designed changes to sequence-encoded NCPR profiles on the global conformational properties and the pattern

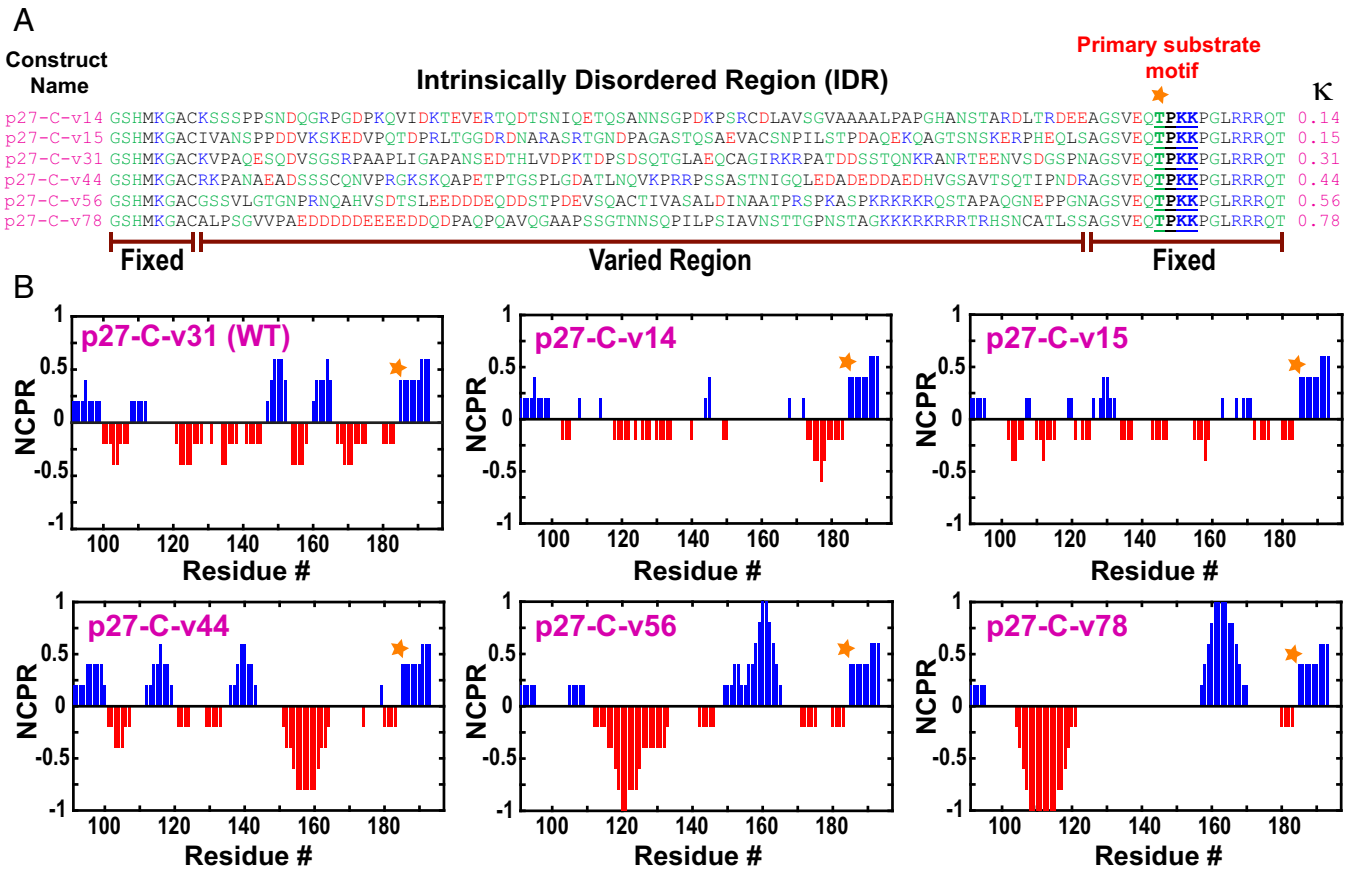

Fig. 2. Design of p27-C permutants to probe effects of charge patterning on conformational features and intracomplex phosphorylation. (*A*) Sequences of the variants of p27-C including the WT. Hydrophobic residues are shown in black, polar residues in green, positively and negatively charged residues in blue and red, respectively. The κ value of each variant is shown to the *Right* of each sequence. (*B*) Each panel shows the NCPR profile along the linear sequence for each p27-C variant. T187 in each NCPR profile is marked using an orange asterisk. The blue and red peaks denote blocks of positive and negative charge, respectively.

of ensemble-averaged intramolecular distances for different p27-C-vXY constructs are summarized in *SI Appendix*, Figs. S5–S9. Variants with low κ values are characterized by larger distances between pairs of residues that are far apart in the linear sequence. However, as the value of κ increases, there is increased likelihood of sampling conformations with long-range attractions as evidenced by the pattern of shorter distances between pairs of residues that are far apart in the linear sequence. The scaling of spatial separation as a function of sequence separation (*SI Appendix*, Fig. S7) suggests that the patterns of interresidue distances are consistent with those of canonical Flory random coils (24) for p27-C-v14, p27-C-v15, p27-C-v31 (WT), and p27-C-v44. For variants p27-C-v56 and p27-C-v78, the distance profiles deviate toward shorter distances for longer sequence separations.

Finer details regarding the conformational ensembles of p27-C-vXY constructs were also analyzed experimentally for a subset of the variants. NMR spectroscopy was used to study the conformational ensembles of p27-C-v14, p27-C-v31, and p27-C-v56. Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence spectra confirmed that the two sequences exhibit significant conformational heterogeneity as evidenced by the limited dispersion of $^{1}H_N$ chemical shifts (*SI Appendix*, Fig. S10*A*). These data are also consistent with previous observations for the WT sequence, which is designated here as p27-C-v31 (6). Analysis of ^{13}C secondary chemical shift values indicates a lack of persistent secondary structure. Similarly, ^{15}N relaxation data show that residues within the three constructs exhibit similar local correlation times (*SI Appendix*, Fig. S10*B*). We also performed measurements of paramagnetic relaxation enhancements (PREs) using constructs where Gly-192 was mutated to Cys for paramagnetic labeling. These data (*SI Appendix*, Fig. S11) are consistent with the simulation results (*SI Appendix*, Fig. S12). The PRE data for p27-C-v56 are indicative of long-range interactions, which is consistent with this construct being significantly more compact than the other two variants.

The collection of simulation results and experimental data demonstrate that the sequence-encoded NCPR profiles determine the global conformational properties of p27-C and p27. These results are consistent with previous predictions (20). To our knowledge, they also represent the first head-to-head comparisons between simulations and experiments with regard to the effects of charge patterning on the conformational properties of a naturally occurring IDR.

T187 Phosphorylation Is Influenced by the Sequence of p27 IDR. To measure the impact of designed changes to the p27 IDR on T187 phosphorylation we mutated Y88 to Glu (Y88E) in all of the constructs. This mimics Y88 phosphorylation and induces partial activation of Cdk2 within the p27-vXY/Cdk2/cyclin A complexes. In these experiments, the N-terminal domain of p27 remains bound to Cdk2 and we measure the efficiency of T187 phosphorylation within the ternary complex; this is designated as the "*in cis*" mechanism of T187 phosphorylation. Fig. 4 shows the relative efficiencies of T187 phosphorylation via the *in cis* mechanism for different p27-vXY variants. These data were obtained at a concentration of 2 μM of the p27-vXY/Cdk2/cyclin A ternary complex; data for two other concentrations of the ternary complex are shown in *SI Appendix*, Fig. S13. Surprisingly, the relative efficiency of T187 phosphorylation is lowest for the WT construct (p27-v31). Among the designed sequence variants, p27-v14 and

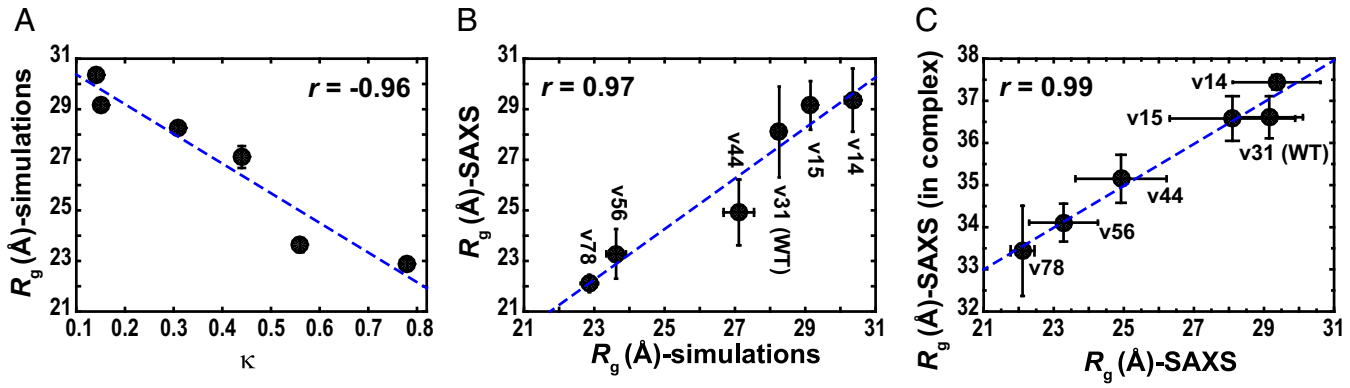

Fig. 3. Dimensions of the p27-C-vXY and p27-vXY variants. (*A*) R_g values for the p27-C-vXY variants are inversely correlated with κ as indicated by the Pearson *r* value –0.96. (*B*) R_g values determined from SAXS measurements for the p27-C variants are plotted against the R_g values estimated from the simulations. The two sets of values are positively correlated ($r = 0.97$). (*C*) The influence of charge patterning on dimensions of the p27-C-vXY variants in isolation was reproduced in the context of the full-length p27-vXY variants bound to Cdk2/cyclin A, as shown by strong positive correlation ($r = 0.99$) between the R_g values determined from SAXS measurements on both sets of constructs. In all plots, the blue dashed line denotes the linear fit to the quantity plotted in the respective ordinate. The error bars for the simulation results are SEs about the mean and for the SAXS they denote SDs based on experiments performed in triplicate.

p27-v56 exhibited the lowest and the highest relative efficiencies, respectively.

We also measured T187 phosphorylation efficiencies using p27-vXY constructs with WT Y88 bound to Cdk2/cyclin A with 1:1:1 stoichiometry. With this experimental design, Cdk2/cyclin A phosphorylates T187 within p27-vXY constructs through an intermolecular ("*in trans*") mechanism (7), which is due to the small fraction of catalytically active Cdk2/cyclin A that is dissociated at equilibrium from the p27-vXY constructs. We imagined two scenarios regarding the influence p27-IDR sequence features on the *in cis* and *in trans* T187 phosphorylation mechanisms (*SI Appendix*, Fig. S14). First, if the varied region of the IDR is an inert tether, then altering its compaction/expansion by varying κ should influence phosphorylation efficiency by the *in cis* mechanism (by indirectly affecting the accessibility of T187 within the primary motif to the Cdk2 active site) but not by the *in trans* mechanism. Second and in contrast, if segments within the IDR apart from the primary motif (e.g., an auxiliary motif) participate in some aspect of the phosphorylation mechanism, then κ-dependent sequence variations could affect both the *in cis* and *in trans* mechanisms. We observed that the relative efficiencies of T187 phosphorylation by the two mechanisms are positively correlated with each other (Fig. 5). This finding implies that specific sequence features that differ among the p27-vXY constructs variably contribute to the efficiency of T187 phosphorylation and that the varied region of the p27 IDR is not an inert tether.

Uncovering Cryptic Sequence Features That Govern T187 Phosphorylation Efficiencies. The preceding observations suggest that the measured differences in phosphorylation efficiencies are attributable to sequence-encoded differences among p27-vXY variants. Our analysis shows that the global conformational properties of p27-vXY variants are poorly correlated with the measured phosphorylation efficiencies (*SI Appendix*, Fig. S15), which rules out simple models that invoke a direct role for global expansion/compaction as the determinant of the observed sequence-specific changes to T187 phosphorylation efficiencies. Given the significance of electrostatic interactions between phosphorylation substrates and the active site region of Cdk2/cyclin A (25–28), we hypothesized that enhanced efficiency of T187 phosphorylation in p27-vXY variants with respect to the WT sequence might be modulated by the net charge per residue within sequence regions N-terminal to the primary substrate motif in p27.

We quantified the weighted NCPR in nonoverlapping sequence windows—auxiliary motifs—that are N-terminal to the primary substrate motif. The weighted NCPR is defined as: $\mathrm{wNCPR} = \sum_{i=1}^{n} w_i[\mathrm{NCPR}^{(i)}(L)]$. It is the weighted sum of NCPR values from *n* nonoverlapping sequence windows. Here, $\mathrm{NCPR}^{(i)}(L)$ is the NCPR of the i^{th} sequence window of length *L*. The parameter w_i is the weight associated with window *i* that satisfies the constraint $0 \le w_i \le 1$ and *i* is an adjacency index such that the larger the value of *i*, the larger the sequence separation between the window *i* and position 181 (*SI Appendix*, Fig. S16). We imposed three additional constraints, namely, $w_1 = 1$, $w_{i+1} < w_1$, and $w_{i+1} \le w_i$ for $i \ne 1$ on the feature selection procedure. As a result, the free parameters in the model are the window length *L*, which fixes the number of nearest-neighbor sequence windows *n* and the weights w_i for windows $i \ne 1$. Setting $n = 2$ leads to Pearson correlation coefficients between the *in cis* phosphorylation efficiencies and wNCPR that are larger than 0.9. Therefore, we restricted the feature selection to the minimalist model comprising two sequence windows. A flowchart outlining the algorithm is shown in *SI Appendix*, Fig. S17. It is worth emphasizing that wNCPR, which is calculated across *n* windows, each of size *L*, is distinct from κ, which summarizes the global patterning of oppositely charged residues across the entire sequence.

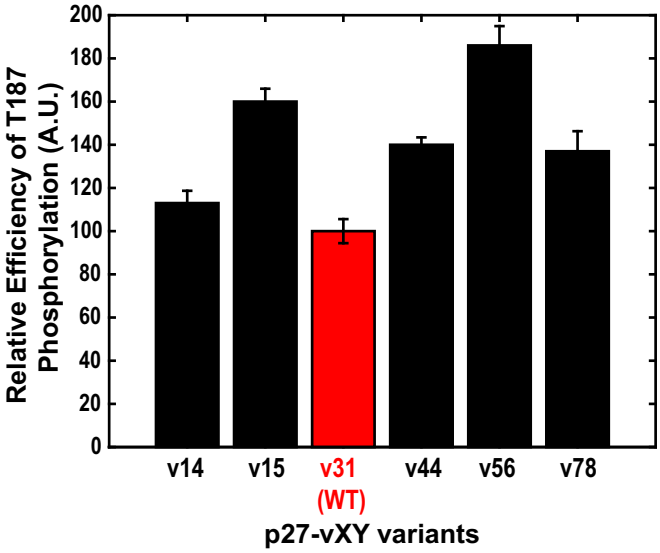

Fig. 4. Relative efficiencies of T187 phosphorylation via the *in cis* mechanisms. A.U., arbitrary units.The error bars denote SDs from three independent experiments. These data were obtained at a concentration of 2 μM for the p27-vXY/Cdk2/cyclin A ternary complex.

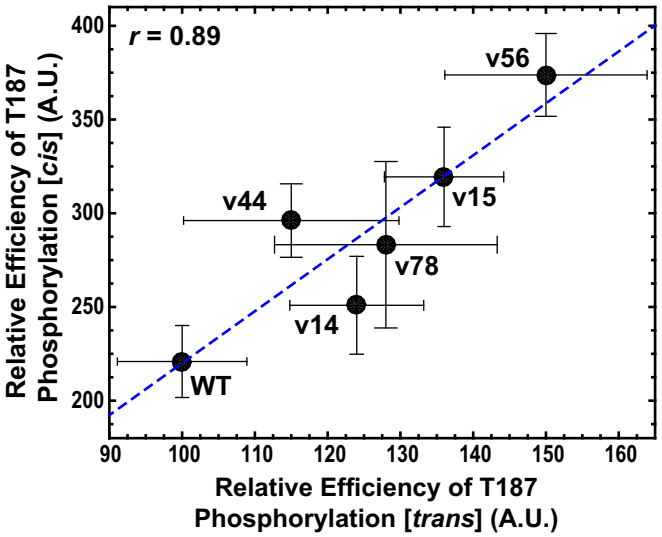

Fig. 5. Correlation between the relative efficiencies of T187 phosphorylation measured via the *in cis* and *in trans* mechanisms for different p27-vXY variants. The Pearson *r* value is 0.89. The blue dashed line denotes the linear fit of the *in cis* phosphorylation efficiency as a function of the *in trans* efficiency. Both *in cis* and *in trans* phosphorylation data used in this analysis were obtained at 4 μM for the ternary complex. The error bars denote SDs of the values for phosphorylation efficiency measured from three independent experiments.

We uncovered a strong positive linear correlation between data for the normalized relative efficiencies of T187 phosphorylation and wNCPR values in an auxiliary motif for wNCPR values less than +0.1 (Fig. 6). However, this analysis reveals one clear outlier (corresponding to p27-v78 with a wNCPR > +0.3). If data for this construct are excluded (as discussed below), the Pearson *r* value that quantifies the linear correlation between the normalized relative phosphorylation efficiencies and wNCPR exceeds 0.9. The parameters for the auxiliary motif in the model, which exhibits the largest correlation coefficient are $n = 2$, $L = 10$, $w_1 = 1.0$, and $w_2 = 0.26$. The auxiliary motif lies outside the primary substrate motif and spans at most 20 residues directly N-terminal to position 181. The prediction is that T187 phosphorylation efficiencies increase linearly as wNCPR values in the auxiliary motif become less negative because this should assist in steering and docking the primary substrate motif to the active site of Cdk2.

We addressed the issue of p27-v78 and possibly other constructs as outliers using a leave-one-out cross-validation procedure (*SI Appendix*, Figs. S18–S24). Leaving out data for four of the six variants, one at a time, did not improve the Pearson *r* values, whereas leaving out the data for p27-v56 improved the Pearson *r* values for a small range of *L* values (9–11). Strikingly, leaving out data for p27-v78 dramatically improved Pearson *r* values across the entire range of *L* values. The lowering of the phosphorylation efficiency of T187 for p27-v78 can be attributed to long-range intrachain interactions within the p27-v78 IDR. Indeed, simulation results showing the ensemble-averaged interresidue distance maps and interresidue contact maps show evidence of long-range interactions for p27-v56 and p27-v78 IDRs (*SI Appendix*, Figs. S6 and S25). Importantly, electrostatic attractions within p27-v78 sequester the primary and auxiliary motifs in intrachain contacts. As a result, p27-v78 has lower phosphorylation efficiencies despite having the most positive value of wNCPR among all of the variants.

Our analysis leads to the hypothesis that the wNCPR values of the auxiliary motif and the intra-IDR interactions that involve the auxiliary and primary motifs should compete with one another to influence the efficiency of T187 phosphorylation. To test our hypothesis, we measured the *in cis* T187 phosphorylation efficiencies for additional variants of p27. We designed new variants of the p27-IDR that enable a titration of the auxiliary motif wNCPR, while maintaining the WT κ value and achieving the highest possible sequence similarity to the WT sequence. These constraints were deployed to probe the effects of changes in auxiliary motif wNCPR values, while making minimal perturbations to the positions of neutral residues and the WT κ value. We designed 22 sequences in silico and selected seven of these for synthesis, expression, and purification. Four of the seven sequences were amenable to expression and purification and these were deployed in functional assays. The four new variants are designated as: p27-v31$^{(-0.30)}$, p27-v31$^{(-0.10)}$, p27-v31$^{(0.00)}$, and p27-v31$^{(+0.25)}$, where the superscript refers to the wNCPR value within the auxiliary motif. If the value of auxiliary motif wNCPR were sufficient as a determinant of the efficiency of T187 phosphorylation, then we would expect a linear increase in functional activity as the wNCPR value increases from –0.3 to +0.25. If there were a convolution between the wNCPR of the auxiliary motif and the sequence-encoded intra-IDR electrostatic interactions, then we would expect deviations from the linear correlation between functional activity and wNCPR. Based on the sequence-encoded NCPR profiles (*SI Appendix*, Fig. S26), we expect there to be inhibitory long-range intra-IDR interactions for the variants p27-v31$^{(0.00)}$ and p27-v31$^{(+0.25)}$.

The measured T187 phosphorylation efficiencies for the four new variants are shown in Fig. 6. The data for p27-v31$^{(-0.30)}$ and p27-v31$^{(-0.10)}$ are in line with the expected positive correlation between T187 phosphorylation efficiency and the wNCPR value of the auxiliary motif. However, this is not the case for p27-v31$^{(0.00)}$ and p27-v31$^{(+0.25)}$, which suggests that if the wNCPR exceeds a threshold positive value, then competing long-range

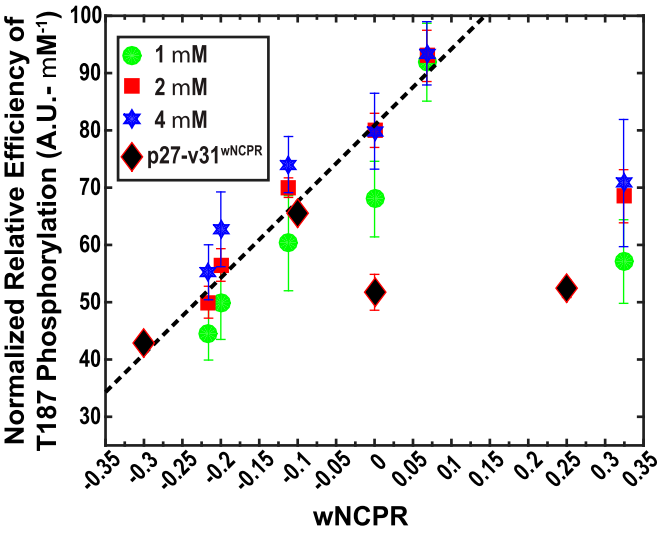

Fig. 6. Normalized relative efficiency of T187 phosphorylation plotted against the wNCPR values within auxiliary motifs for p27-vXY and p27-v31$^{(wNCPR)}$ variants. For p27-vXY variants, the datasets correspond to concentrations of 1, 2, and 4 μM of the 1:1:1 ternary complex of p27-vXY/Cdk2/cyclin A. For each dataset, shown using different colors, we calculated the normalized relative efficiency *e* by dividing the relative efficiencies by the corresponding concentrations of the ternary complex. Within error, the normalized relative efficiencies are coincident for each p27-vXY variant. The Pearson *r* value quantifying the positive correlation between *e* and wNCPR is 0.93. For the linear regression and correlation analysis we left out the data for p27-v78. The values along the ordinate can be written as $e = m(\text{wNPCR}) + c$. Here, *m* is the slope and *c* is the intercept. Because wNCPR is a dimensionless quantity the slope quantifies the increase in the normalized efficiency per unit rise in wNCPR and the intercept quantifies the normalized relative efficiency if the wNCPR is zero. The values for *m* and *c* are 133.1 A.U.·μM^{-1} and 80.9 A.U.·μM^{-1}, respectively. This model predicts the normalized relative efficiency of T187 phosphorylation as a function of the wNCPR within the auxiliary motif of p27-vXY. The solid diamonds are data for p27-v31$^{(-0.30)}$, p27-v31$^{(-0.10)}$, p27-v31$^{(0.00)}$, and p27-v31$^{(+0.25)}$, respectively, and test the generality of the linear model in sequences where the κ value is fixed and wNCPR is varied by changes to the NCPR profile.

intra-IDR interactions can diminish the enhancement of phosphorylation efficiencies that would be expected from the positive auxiliary motif wNCPR values. Simulations show that p27-C-v31$^{(+0.25)}$ is considerably more compact ($R_g \sim 24$ Å) than would be expected based on its κ value alone. The accumulation of positive charge within the auxiliary motif leads to the concomitant dispersion of negatively charged residues across the region that is N-terminal to the auxiliary motif. This promotes long-range intra-IDR interactions that involve the primary and auxiliary motifs (*SI Appendix*, Fig. S27).

Discussion

Our functional studies suggest that the efficiency of T187 phosphorylation depends on three factors: (*i*) the primary motif that encompasses the phosphosite; (*ii*) the interactions of the auxiliary motif with Cdk2, which appear to be governed by the wNCPR of the auxiliary motif; and (*iii*) the strength of long-range intramolecular interactions of the primary and auxiliary motifs with the N terminus of the p27 IDR. Factors (*ii*) and (*iii*) are intrinsically coupled to the wNCPR value of the auxiliary motif and charge patterning across the remainder of the sequence. The latter emerges as a consequence of achieving a particular wNCPR value within the auxiliary motif. If the wNCPR within the auxiliary motif becomes less negative without an accompanying increase in intra-IDR electrostatic interactions, then the efficiency of T187 phosphorylation will increase. Conversely, if making wNCPR less negative or increasingly positive leads to concomitant increases in intra-IDR electrostatic attractions that involve the primary and/or auxiliary motifs, then the efficiency of T187 phosphorylation will be reduced. Given these two opposing effects, the implication is that for a given κ value, there should be sequences with specific wNCPR values and charge patterns outside the auxiliary motif that either minimize or maximize the efficiency of T187 phosphorylation. Indeed, there exist a large number of sequences with a range of wNCPR values for a given κ value (*SI Appendix*, Fig. S28). Therefore, it should be feasible to perform systematic, high-throughput experiments to uncover, for a given κ value, the wNCPR value within the auxiliary motif that minimizes/maximizes the efficiency of T187 phosphorylation.

T187 phosphorylation is the essential step that controls progression of the cell cycle to the S-phase. Negative regulation by cryptic sequence features within the p27 IDR should slow its ubiquitination and degradation, as well as slow progression of the cell cycle to the S-phase. The fact that we were able to design p27-vXY constructs that signal more efficiently than WT p27 suggests that this critical cell cycle regulatory system may have evolved to suppress the efficiency of T187 phosphorylation and the kinetics of cell cycle progression. Multiple sequence alignments show that the amino acid composition of the p27 IDR is absolutely conserved across orthologs (*SI Appendix*, Fig. S29); this leads to absolute conservation of the wNCPR of the auxiliary motif and hence the pattern of intra-IDR interactions between primary/auxiliary motifs and their N-terminal contexts.

Cryptic sequence features do not appear to be unique to the p27 IDR. Indeed, their existence was previously postulated (29). Their impact has been established for the C-terminal region of the core subunit of the Polycomb repressive complex 1 that compacts chromatin and inhibits chromatin remodeling (16) and in the Nck system that undergoes phase separation via interactions with polyvalent partners (30). Our analysis clearly shows that the sequence context of an IDR can modulate the functionality of a primary motif. This analysis uncovers two plausible cryptic sequence features that confer a modulatory role on the IDR sequence context. Of course, there could be many more features that play further modulatory roles and high-throughput studies driven by sequence design should provide a powerful approach for a comprehensive uncovering of cryptic, functionally relevant sequence features in IDRs.

Materials and Methods

Atomistic Monte Carlo simulations were performed using the CAMPARI molecular modeling suite (campari.sourceforge.net) and were based on the ABSINTH implicit solvation model and forcefield paradigm (23). Details of the simulation procedures and analysis of conformational ensembles are presented in *SI Appendix, Section 2*. The SAXS experiments were conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory. Details of the experimental methods and analysis including the design, preparation, and purification of constructs, SAXS data collection and analysis, NMR spectroscopy, and kinase assays are described in detail in *SI Appendix, Section 3*.

ACKNOWLEDGMENTS. This work was supported by Grant R01CA082491 from the NIH (to R.W.K.); National Cancer Institute Cancer Center Support Grant P30CA21765 to the St. Jude Children's Research Hospital; and National Science Foundation Grant MCB 1121867 (to R.V.P.). Y.H. was a recipient of the Gephardt Named Fellowship provided by the St. Jude Children's Research Hospital. Additional support comes from the NIH project MINOS (R01GM105404).

1. Puntervoll P, et al. (2003) ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res* 31(13):3625–3630.
2. Neduva V, et al. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 3(12):e405.
3. Tompa P, Davey NE, Gibson TJ, Babu MM (2014) A million peptide motifs for the molecular biologist. *Mol Cell* 55(2):161–169.
4. Morgan DO (1995) Principles of CDK regulation. *Nature* 374(6518):131–134.
5. Lacy ER, et al. (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 11(4):358–364.
6. Galea CA, et al. (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J Mol Biol* 376(3):827–838.
7. Grimmler M, et al. (2007) Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* 128(2):269–280.
8. Orlicky S, Tang X, Willems A, Tyers M, Sicheri F (2003) Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* 112(2):243–256.
9. Tsvetkov LM, Yeh KH, Lee SJ, Sun H, Zhang H (1999) p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr Biol* 9(12):661–664.
10. Carrano AC, Eytan E, Hershko A, Pagano M (1999) SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nat Cell Biol* 1(4):193–199.
11. Shirane M, et al. (1999) Down-regulation of p27(Kip1) by two mechanisms, ubiquitin-mediated degradation and proteolytic processing. *J Biol Chem* 274(20):13886–13893.
12. Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J* 98(10):2383–2390.
13. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(18):8183–8188.
14. Müller-Späth S, et al. (2010) From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(33):14609–14614.
15. Vuzman D, Levy Y (2010) DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci USA* 107(49):21004–21009.
16. Beh LY, Colwell LJ, Francis NJ (2012) A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc Natl Acad Sci USA* 109(18):E1063–E1071.
17. Das RK, Ruff KM, Pappu RV (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* 32(0):102–112.
18. Mittal A, Lyle N, Harmon TS, Pappu RV (2014) Hamiltonian switch metropolis Monte Carlo simulations for improved conformational sampling of intrinsically disordered regions tethered to ordered domains of proteins. *J Chem Theory Comput* 10(8):3550–3562.
19. Sickmeier M, et al. (2007) DisProt: The database of disordered proteins. *Nucleic Acids Res* 35(Database issue):D786–D793.
20. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci USA* 110(33):13392–13397.
21. Bright JN, Woolf TB, Hoh JH (2001) Predicting properties of intrinsically unstructured proteins. *Prog Biophys Mol Biol* 76(3):131–173.
22. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314(1–2):141–151.
23. Vitalis A, Pappu RV (2009) ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* 30(5):673–699.
24. Flory PJ (1969) *Statistical Mechanics of Chain Molecules* (Oxford Univ Press, New York).
25. Stevenson-Lindert LM, Fowler P, Lew J (2003) Substrate specificity of CDK2-cyclin A. What is optimal? *J Biol Chem* 278(51):50956–50960.
26. Holmes JK, Solomon MJ (2001) The role of Thr160 phosphorylation of Cdk2 in substrate recognition. *Eur J Biochem* 268(17):4647–4652.
27. De Vivo M, Cavalli A, Bottegoni G, Carloni P, Recanatini M (2006) Role of phosphorylated Thr160 for the activation of the CDK2/Cyclin A complex. *Proteins* 62(1):89–98.
28. Russo AA, Jeffrey PD, Pavletich NP (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 3(8):696–700.
29. Das RK, Mao AH, Pappu RV (2012) Unmasking functional motifs within disordered regions of proteins. *Sci Signal* 5(220):pe17.
30. Banjade S, et al. (2015) Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. *Proc Natl Acad Sci USA* 112(47):E6426–E6435.