Section 1 – Results from simulations, experiments, and feature selection analysis that supplement the presentation in the main text

Table S1. Numerical R_g values of designed variants obtained from simulations and SAXS data.

p27-C-vXY variant	R_g calculated from simulations (T = 328 K) (Å)	R_g inferred from analysis of the pair distribution function extracted from SAXS data (Å)		p27-vXY variant	R_g inferred from analysis of the pair distribution function extracted from SAXS data for p27-vXY/Cdk2/cyclin A ternary complexes (Å)
p27-C-v14	30.35 ± 0.22	29.36 ± 1.25		p27-v14	37.44 ± 0.17
p27-C-v15	29.15 ± 0.19	29.15 ± 0.96		p27-v15	36.61 ± 0.50
p27-C-v31	28.26 ± 0.12	28.10 ± 1.79		p27-v31	36.58 ± 0.53
p27-C-v44	27.11 ± 0.43	24.92 ± 1.30		p27-v44	35.15 ± 0.57
p27-C-v56	23.62 ± 0.28	23.28 ± 0.98		p27-v56	34.11 ± 0.45
p27-C-v78	22.86 ± 0.24	22.11 ± 0.34		p27-v78	33.44 ± 1.07

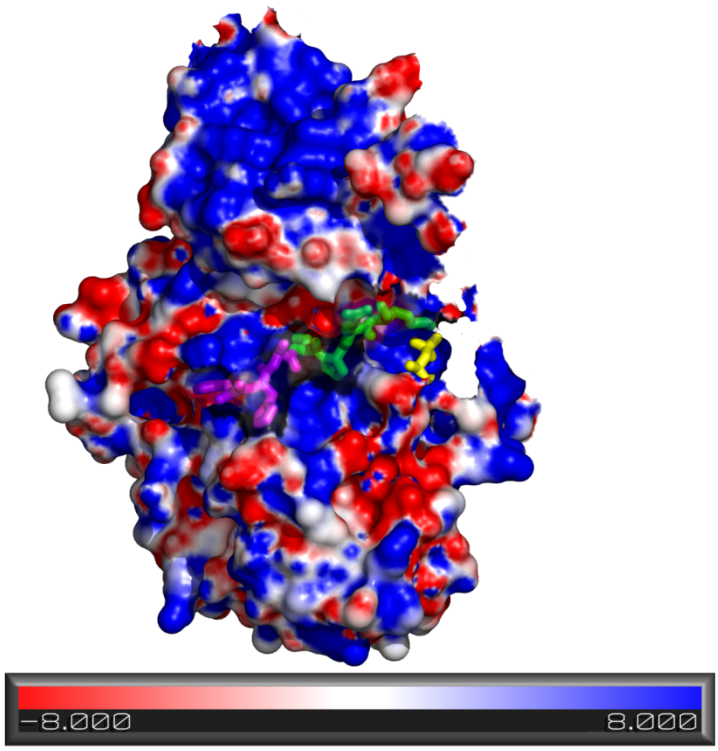

Fig. S1: Electrostatic surface potential of phosphorylated Cdk2 in its crystal structure with cyclin A and an optimal peptide substrate, HHASPRK (PDB id 1QMZ). The color bar on the bottom displays the color map and indicates the mapping of electrostatic potential in units of thermal energy, k_BT, where k_B is the Boltzmann constant and T denotes temperature in Kelvin. The colors change from hotter to cooler hues as electrostatic surface potential changes from negative to positive. The pThr160 of Cdk2 is shown in yellow sticks. The residues SPRK corresponding to a peptide mimicking the primary substrate motif within p27 are shown as green sticks. Other residues in the model peptide substrate are shown as magenta sticks. The red colored surface below pThr160 is the acidic patch on Cdk2 comprising Glu162, Glu208, and Asp235. The electrostatic potential was computed using the Adaptive Poisson Boltzmann Solver (APBS) (1, 2) and rendered using Pymol (3).

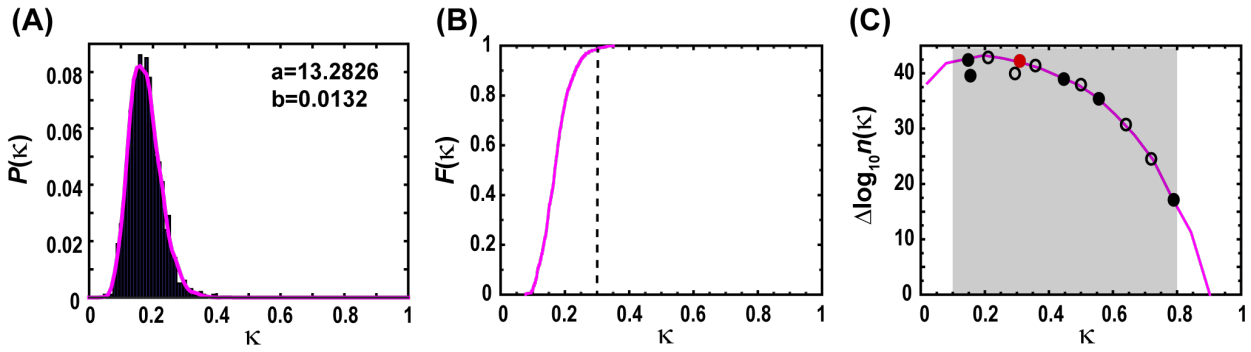

Fig. S2: Statistics that characterize the magnitude of the sequence design space for p27-C. We generated distinct sequences by shuffling residues to alter the NCPR profile (and hence κ) while keeping the amino acid composition fixed. The value of κ provides a concise summary of the patterning of oppositely charged residues. **A large number of sequences are compatible with a given value of κ.** Panel (A) shows the probability distribution function $P(\kappa)$. Here, $P(\kappa)\Delta\kappa$ quantifies the probability of finding a sequence with κ-value between κ and $\kappa+\Delta\kappa$. The actual distribution $P(\kappa)$ is shown in black bars. $P(\kappa)$ can be fit to a gamma distribution $P(\kappa) = f(\kappa \mid a,b) = \frac{1}{b^a \Gamma(a)} \kappa^{a-1} \exp\left(-\frac{a}{b}\right)$, which is shown as a magenta curve**.** The maximum likelihood estimates for the two parameters *a* and *b* are shown in the inset. Panel (B) shows the cumulative probability distribution that is derived by integrating the probability distribution function. Here, $F(\kappa=x)$ quantifies the probability of finding a sequence with κ value that is less than or equal to x. The vertical dashed line denotes the κ-value of the WT p27-C. **The inference is that given the composition of p27-C, the probability of generating sequences with κ ≤ 0.31 is essentially unity, implying that the probability of finding sequences with κ values larger than 0.31 is extremely low.** Panel (C) shows the number density of sequences with specific κ-values. The total number of conceivable sequences consistent with the composition of p27-C was computed using the multinomial theorem. In all, there are ~10^{113} distinct sequences. In panel (C) the ordinate $\Delta\log_{10}n(\kappa)$ quantifies the differences in the orders of magnitudes for number densities of conceivable sequences with different κ values relative to κ=0.9. Given the constraints on our designs, sequences corresponding to κ-values that lie outside the shaded region cannot be sampled. The plot is annotated by κ-values of the designed variants that could be expressed and purified (black filled circles) and those that could not be robustly expressed in *E.coli* (open circles). The red filled circle depicts the position of the WT sequence. Our chosen designs enable a systematic titration of sequences whereby we uniformly sample from the range of conceivable κ-values.

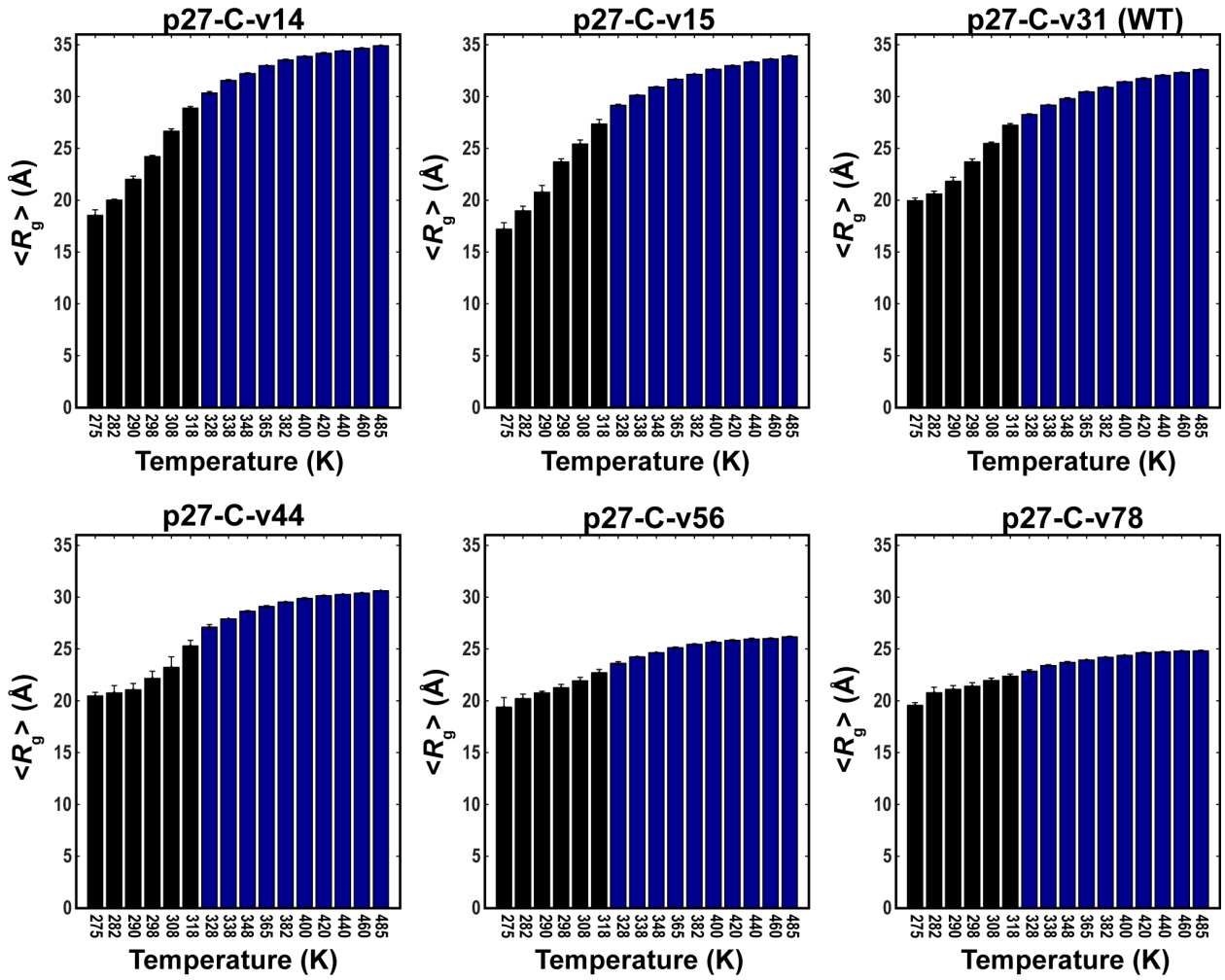

Fig. S3: Simulation results for the temperature dependence of ensemble-averaged radii of gyration R_g of p27-C-vXY variants. Going from low to high temperature, the magnitude of change in R_g is the lowest for p27-C-v78. The error bars denote standard error of means from three independent sets of temperature replica exchange simulations (see details in Section 2). In all cases, the simulation results are statistically robust for temperatures ≥ 328 K (see analysis in Section 2).

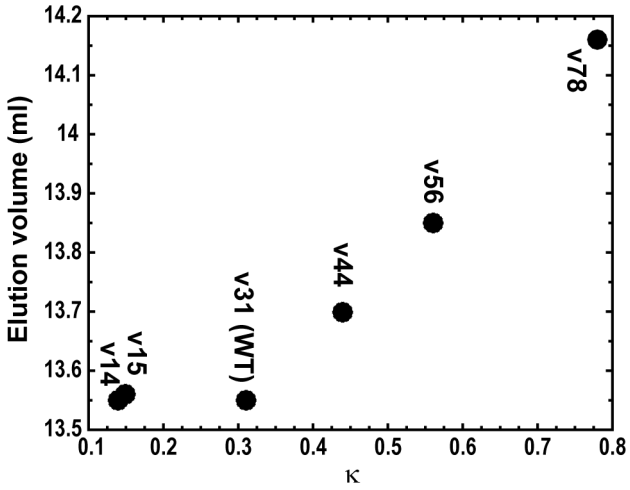

Fig. S4: Experimentally measured elution volumes from size exclusion chromatography of the p27-vXY variants bound to Cdk2/cyclin A. Compaction of p27-C-vXY, which accompanies an increase in κ, leads to larger elution volumes.

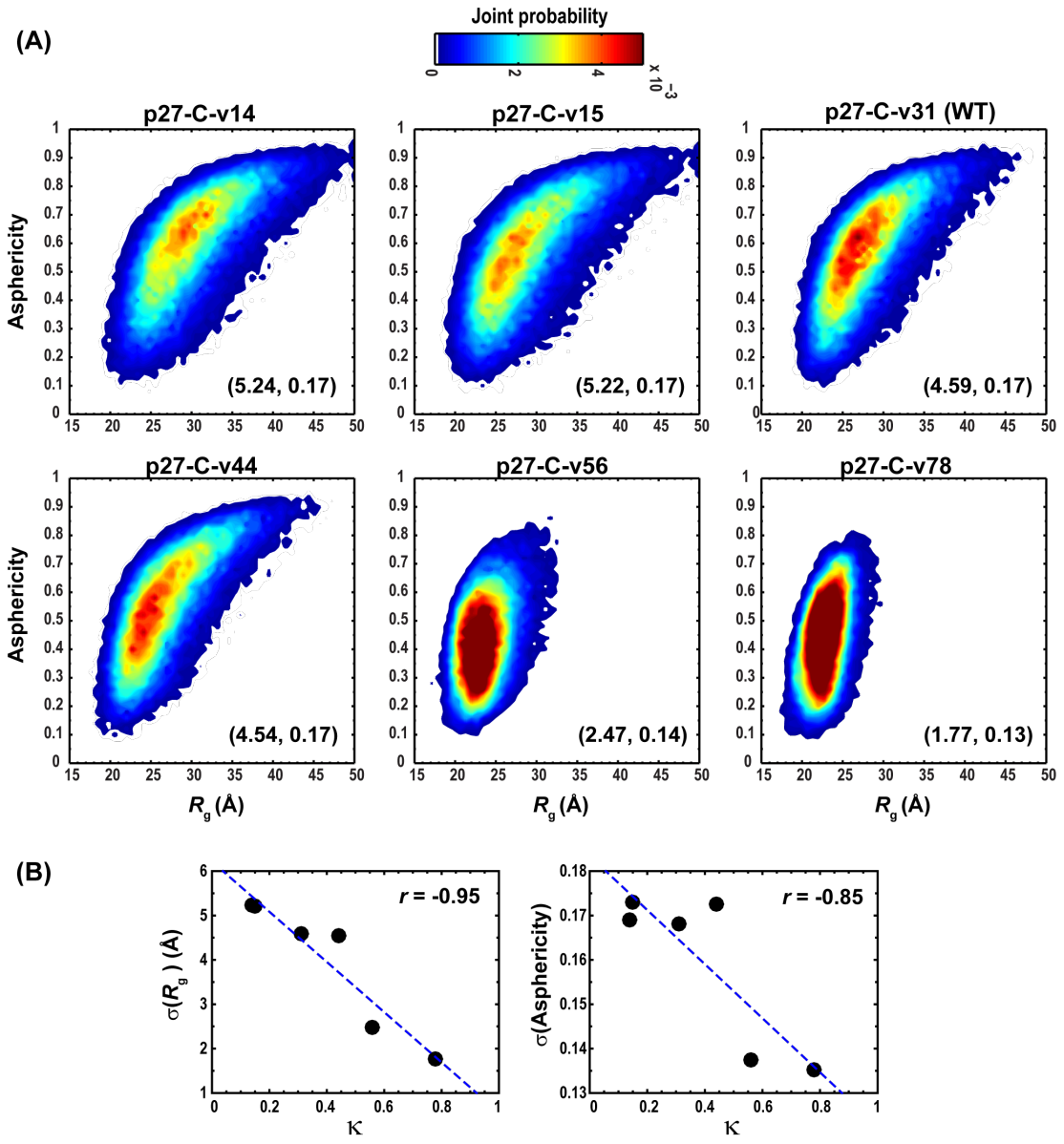

Fig. S5: (A) Simulation results for the joint probability distributions of instantaneous values of R_g and asphericity for the p27-C in p27-C-vXY variants. Asphericity quantifies the overall shape of a specific conformation and is calculated from eigenvalues of the gyration tensor. The smaller the asphericity values, the more globular the conformation and conversely, rod-like conformations have asphericity values that approach unity. The bin sizes for R_g and asphericity are 0.7 Å and 0.02, respectively. The color bar on the top displays the color map and indicates the mapping of joint probabilities onto the color map. The colors change from cooler to hotter hues as joint probabilities increase. The two numbers within the parentheses denote standard deviations of distributions for R_g (in Å) and asphericity values, respectively in the respective p27-C-vXY variant. (B) Inverse correlations between the amplitudes of conformational fluctuations and the κ values. The standard deviations of distributions for both R_g (in Å) and asphericity values [$\sigma(R_g)$ and σ(Asphericity), respectively] were used to quantify the amplitudes of the conformational fluctuations of the p27-C in p27-C-vXY variants. The Pearson *r*-values are –0.95 and –0.85, respectively.

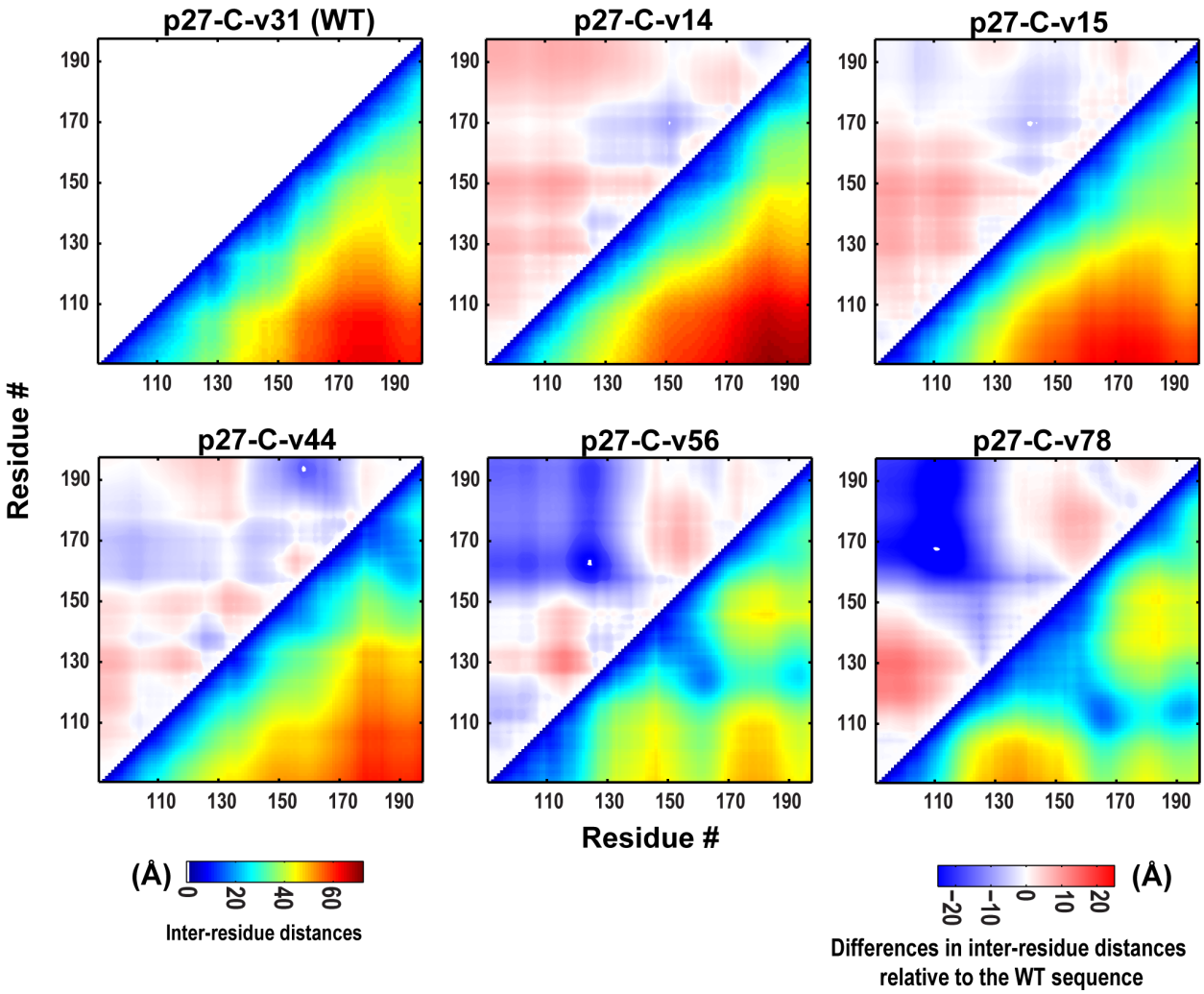

Fig. S6: Lower triangle: Ensemble-averaged inter-residue distances from simulation results for p27-C-vXY variants at 328K. The color bar on the bottom left displays the color map and indicates the mapping of distances onto the colors. The colors change from cooler to hotter hues as inter-residue distances increase. **Upper triangle: Difference between ensemble-averaged inter-residue distances of the designed κ-variants and the WT (p27-C-v31).** The color bar on the bottom right displays the color map and indicates the mapping of the differences in distances for a κ-variant (relative to the WT) onto the color map. If an inter-residue distance increases in the κ-variant relative to the WT, the difference in distances is positive (red). Similarly, if an inter-residue distance decreases in the κ-variant relative to the WT, the difference in distances is negative (blue).

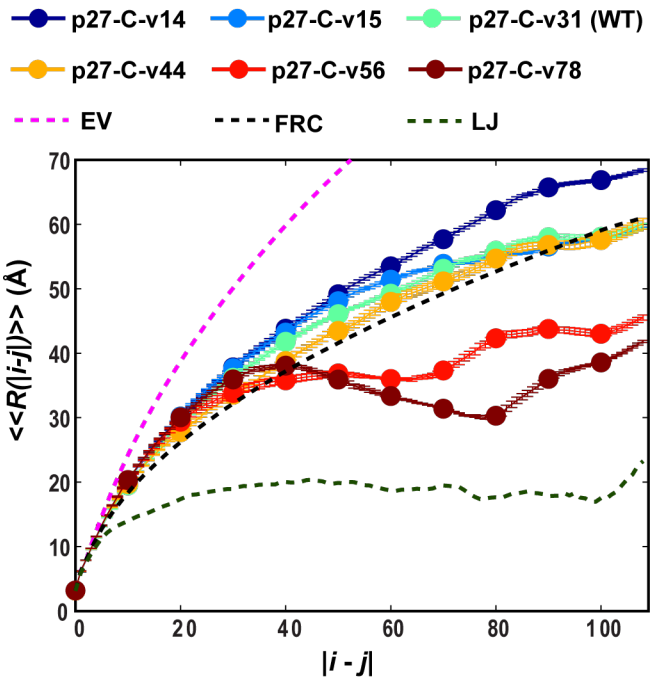

Fig. S7: Ensemble-averaged scaling of inter-residue distances <<*R*(|*i-j*|)>> plotted against sequence separation |*i – j*| at 328 K for the p27 IDR in p27-C-vXY variants. The three dashed curves in magenta, black, and green, respectively denote <<*R*(|*i – j*|)>> profiles for three reference models designated as Excluded Volume (EV), Flory Random Coil (FRC), and Lennard-Jones (LJ) globule, respectively. The details of these reference models are described in **Section 2**. The error bars denote standard errors of respective means. For p27-C-v14, p27-C-v15, p27-C-v31 (WT), and p27-C-v44, <<*R*(|*i – j*|)>> increases monotonically with increasing |*i-j*|. For p27-C-v56 and p27-C-v78, <<*R*(|*i – j*|)>> profiles indicate presence of long-range electrostatic attractions between strongly segregated oppositely charged domains. <*R*(|*i – j*|)> values from three sets of independent temperature replica exchange simulations (see details in Section 2) were averaged to give <<*R*(|*i – j*|)>> values.

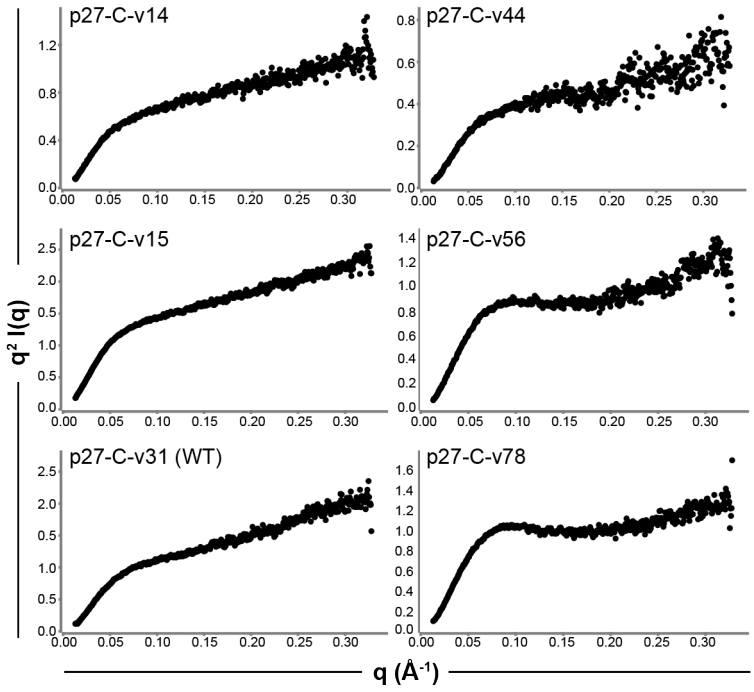

Fig. S8: Kratky plots obtained from the SAXS data for the scattering intensities I(q) as a function of the scattering vector q. With the exception of p27-C-v56 and p27-C-v78, the Kratky plots show characteristic random-coil-like profiles. The impact of long-range attractions is evident in the sequences where the segregation of opposite charges is most significant, viz., p27-C-v56 and p27-C-v78.

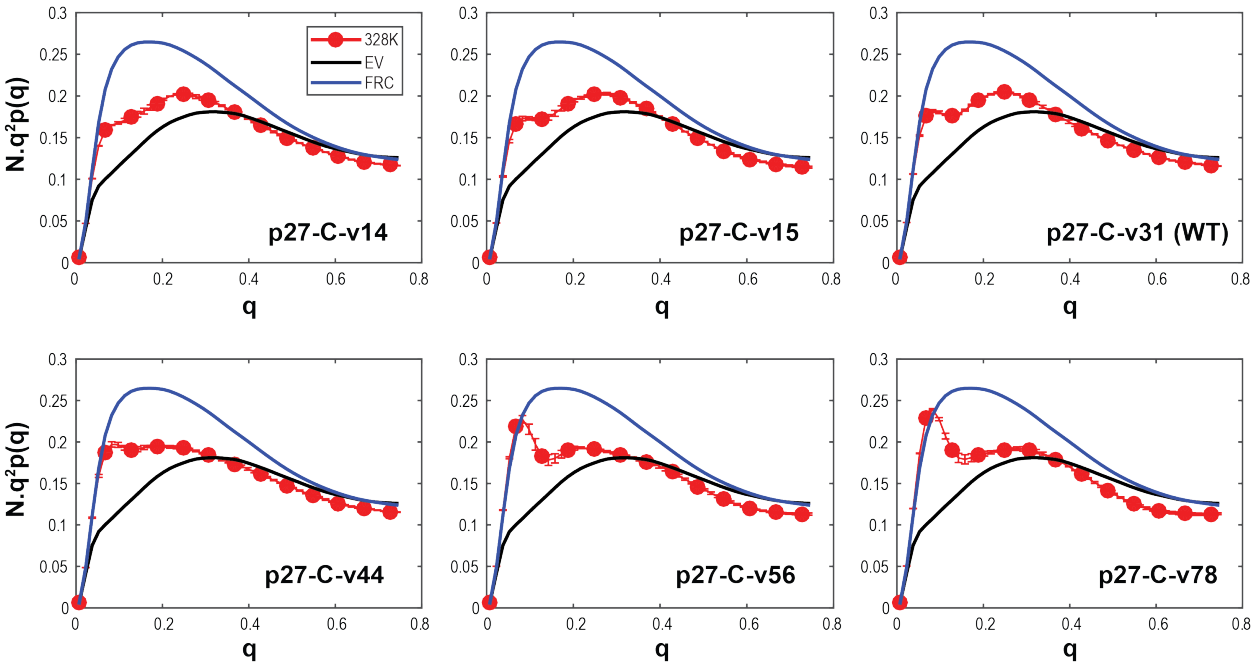

Fig S9: Kratky plots calculated from the ABSINTH-based ensembles for each of the p27-C-XY variants. For each sequence, the Kratky plots calculated from ensembles obtained at 328 K (red curves) are compared to profiles from two reference ensembles for the sequence in question. These reference ensembles are designated as EV (black curves) for excluded volume and FRC (blue curves) for the Flory random coil. Details of how the two references ensembles were derived are provided in section 2. With the exception of p27-C-v56 and p27-C-v78, which show clear evidence for long-range attractions – see the pronounced hump in the long-q regime – the profiles are essentially consistent with coil-like behavior for all variants. This implies that the conformational ensembles are, on average, well solvated, which is consistent with inferences from SAXS data as well. For each ensemble, the Kratky plot was calculated as follows: We first calculate the form factor $P(q)$, assuming uniform scattering cross-sections for all atoms using the formula: $P(q) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{\sin(qR_{ij})}{qR_{ij}}$. Here, N is the number of residues, R_{ij} refers to the distances between residues i and j quantified using the positions of alpha carbon atoms, and q is the scattering wave number with units of $Å^{-1}$. The form factor, calculated across the three ensembles for each sequence, is used to calculate the Kratky plot as $Nq^2P(q)$. The profiles for the two reference ensembles provide a way to calibrate our interpretations for the profiles calculated using the ABSINTH-based ensembles.

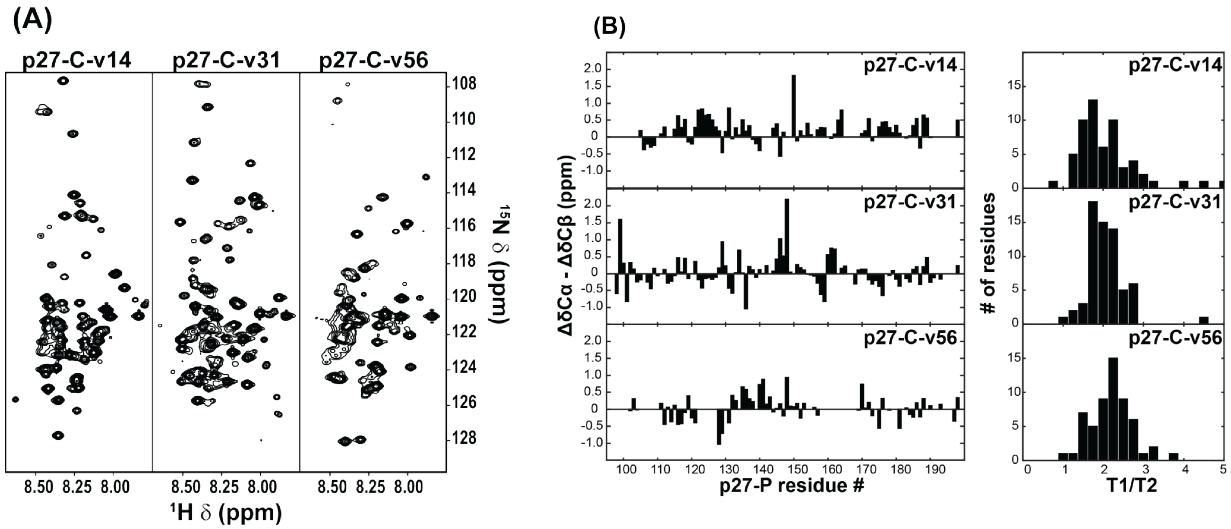

Fig. S10: Results from Nuclear Magnetic Resonance spectroscopy for p27-C-v14, p27-C-v31 (WT), and p27-C-v56 variants. (A) Two-dimensional ^{1}H-^{15}N HSQC (heteronuclear single quantum coherence) spectra show limited $^{1}H_N$ chemical shift dispersion and confirm that the two designed variants and the wild type sequence are highly disordered. (B) Analysis of ^{13}C secondary chemical shift values (left panel) indicated a lack of secondary structure in all three variants. ^{15}N relaxation data (right panel) indicated that residues within the three variants exhibited similar local correlation times.

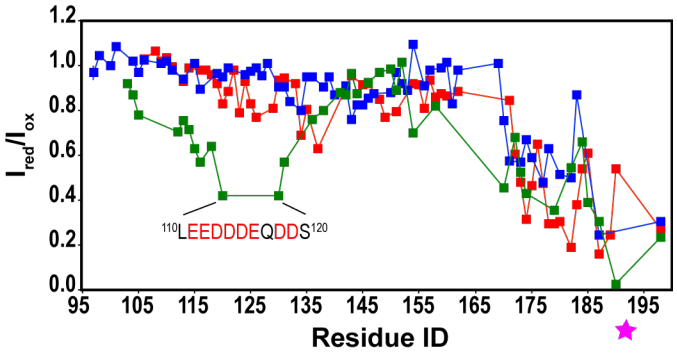

Fig. S11: Data from measurements of paramagnetic relaxation enhancements (PREs) for p27-C-vXY variants: In these p27-C-vXY constructs, Gly 192 (position is depicted by magenta star) was mutated to Cys for paramagnetic labeling. PRE data for both p27-C-v14 (blue profile) and p27-C-v31 (red profile) point to a lack of long-range interactions in these two variants. In contrast, the PRE data for p27-C-v56 (green profile) are indicative of long-range contacts between the sequence C-terminal to the primary substrate motif and the acidic block spanning residues 111-119. These PREs are consistent with p27-C-v56 being significantly more compact than the other two variants.

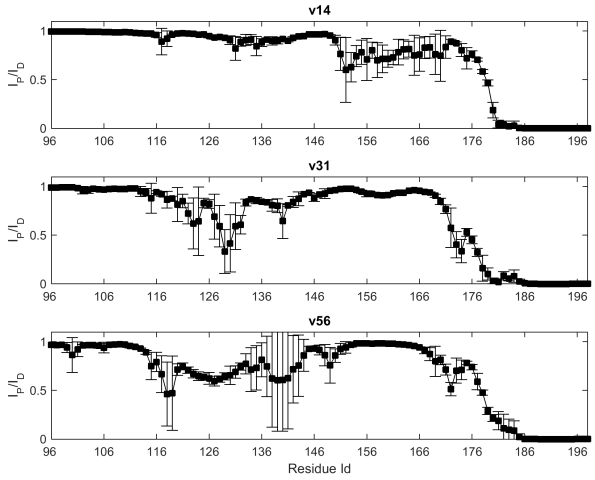

Fig. S12: Calculated PRE profiles for three p27-C-vXY variants. These profiles are to be compared, qualitatively, to the data shown in Fig. S11. The contribution of paramagnetic relaxation enhancement to the transverse relaxation rate, Γ_2 was calculated using the equation $\Gamma_2 = \frac{K}{\langle r^6 \rangle}(4\tau_c + \frac{3\tau_c}{1+\omega_H^2\tau_c^2})$. Here, K is 1.23×10^{-32} cm^6s^{-2}; r is the distance between the alpha carbon of Gly192 and the backbone amide nitrogen atom of a given residue. For all three variants the effective correlation time τ_C is 3 ns as determined by the backbone ^{15}N relaxation rates and ω_H, the proton Larmor frequency is 600.17 MHz. The peak intensity ratios between the paramagnetic and diamagnetic forms were calculated using $\frac{I_P}{I_D} = \frac{R_{2D}\exp(-\Gamma_2 t)}{R_{2D}+\Gamma_2}$; Here, R_{2D}, the transverse relaxation rate of the backbone amide protons in the diamagnetic form, is $17s^{-1}$ and t the total length of the INEPT and refocusing delays is 18 ms. The simulated ensembles were used to calculate the ensemble averaged value of r^{-6}. This was achieved using the ensembles obtained at 328 K. The inter-residue distances that enable the calculation of r^{-6} was recorded once every 4000 steps along the Monte Carlo simulations. The distances were binned into histograms with a bin size of 0.2Å. The calculated PRE profiles are shown here for three variants and are to be compared to the three profiles derived from experimental measurements as shown in **Fig. S11**. The profiles from the simulation results are considerably noisier than those derived from experimental data and highlight the sampling challenges we encounter when there is any level chain compaction. Therefore, the comparisons between calculated and experimental profiles can only be qualitative in nature. Within the scope of these caveats, the PRE profiles do show signatures of long-range contacts involving the block that spans a set of acidic residues 111-119 for p27-C-v56 (referred to as v56 in the figure). This is qualitatively similar to what we observe in the experiments. For p27-C-v14 (v14) there is little evidence of long-range interactions and the simulation results do point to some persistent long-range contacts for the wild type sequence (v31). It is worth noting that r^{-6} is very sensitive to small variations in r. For a heterogeneous ensemble, this leads to PRE profiles that are exceedingly noisy – a feature that is shared between the experimentally derived and calculated PRE profiles for all three variants. A more thorough and complete accounting of the inter-residue distance profiles that do not involve the weighting and hence masking of the true distances by the inverse sixth power calculation may be obtained from **Fig. S6**. Here, we show the unweighted mean inter-residue distances that enable detailed comparisons among the simulated ensembles for all variants.

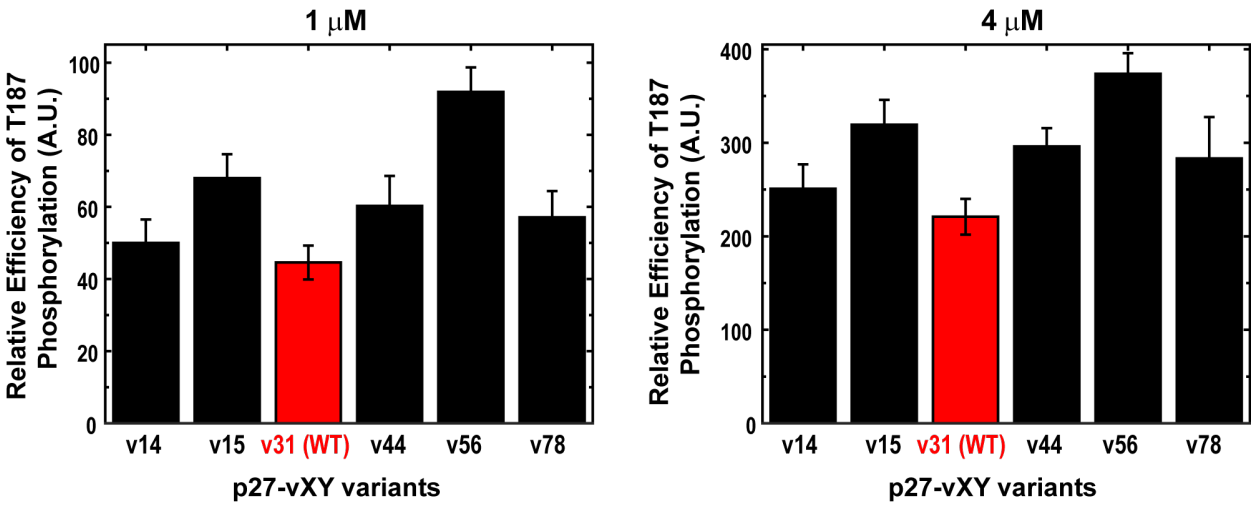

Fig. S13: In *cis* phosphorylation efficiency data for the full-length p27-vXY variants at concentrations of 1 and 4 μM of p27-vXY(Y88E)/Cdk2/cyclin A ternary complexes. In these experiments, Y88 was mutated to Glu in all p27-vXY variants (Y88E) to mimic Y88 phosphorylation and to partially activate Cdk2 within the p27-vXY/Cdk2/cyclin A complexes. Relative efficiencies were obtained by dividing the measured efficiencies for a given concentration of the ternary complex by the efficiency of in *cis* phosphorylation of T187 in the Y88E-p27-v31/Cdk2/cyclin A complex measured at a concentration of 2 μM for the ternary complex. The error bars denote standard deviations of the mean of the values for phosphorylation efficiency (in arbitrary units, A.U.) measured from three independent experiments.

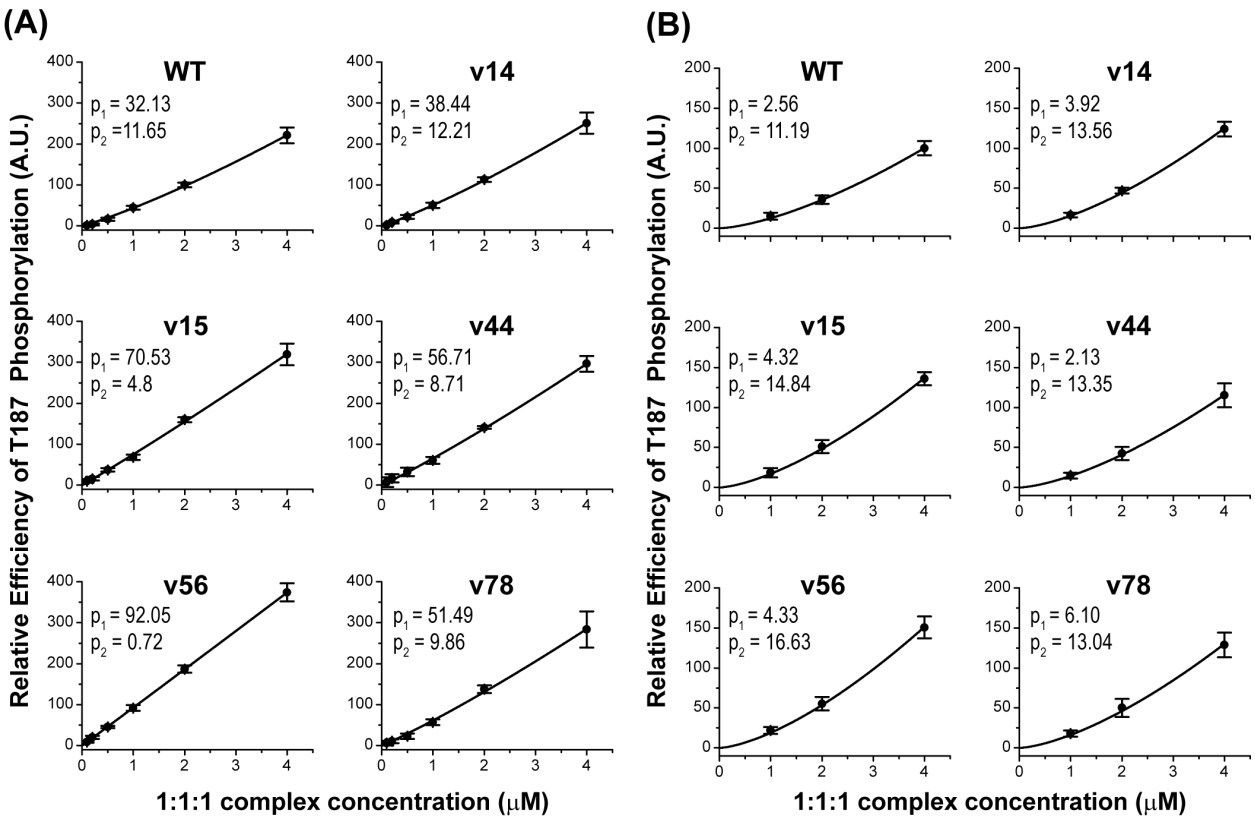

Fig. S14: Relative efficiencies of T187 phosphorylation measured via the in *cis* and in *trans* mechanisms for different p27-vXY constructs. These relative efficiencies (*E*) were normalized to the efficiency of in *cis* phosphorylation of T187 in the Y88E-p27-v31/Cdk2/cyclin A complex measured at a concentration of 2 μM for the ternary complex. The set of six plots in panel (A) shows the dependence of *E* on the concentration of the ternary complex of Y88E-p27-vXY/Cdk2/cyclin A. The relative efficiency for each p27-vXY construct was fit to the equation $E = p_1C+p_2C^{3/2}$. Here, *C* is the concentration of the 1:1:1 ternary complex. The coefficients p_1 and p_2 quantify the contributions from the in *cis* and in *trans* mechanisms, respectively. The values for p_1 and p_2, obtained for data in the six plots in panel (A), confirm the dominance of the in *cis* mechanism for T187 phosphorylation, although phosphorylation can also occur *in trans*, which leads to nonzero values for p_2. The set of six plots in panel (B) shows the concentration dependence of phosphorylation of T187 for all ternary complexes of variants in Y88-p27-vXY/Cdk2/cyclin A. The fit of *E* to the equation $E = p_1C+p_2C^{3/2}$ leads to smaller values for p_1 when compared to p_2 thus confirming the dominance of the in *trans* mechanism. **Fig. 5** in the main text shows the positive correlation between values of *E* measured via the in *cis* and in *trans* mechanisms.

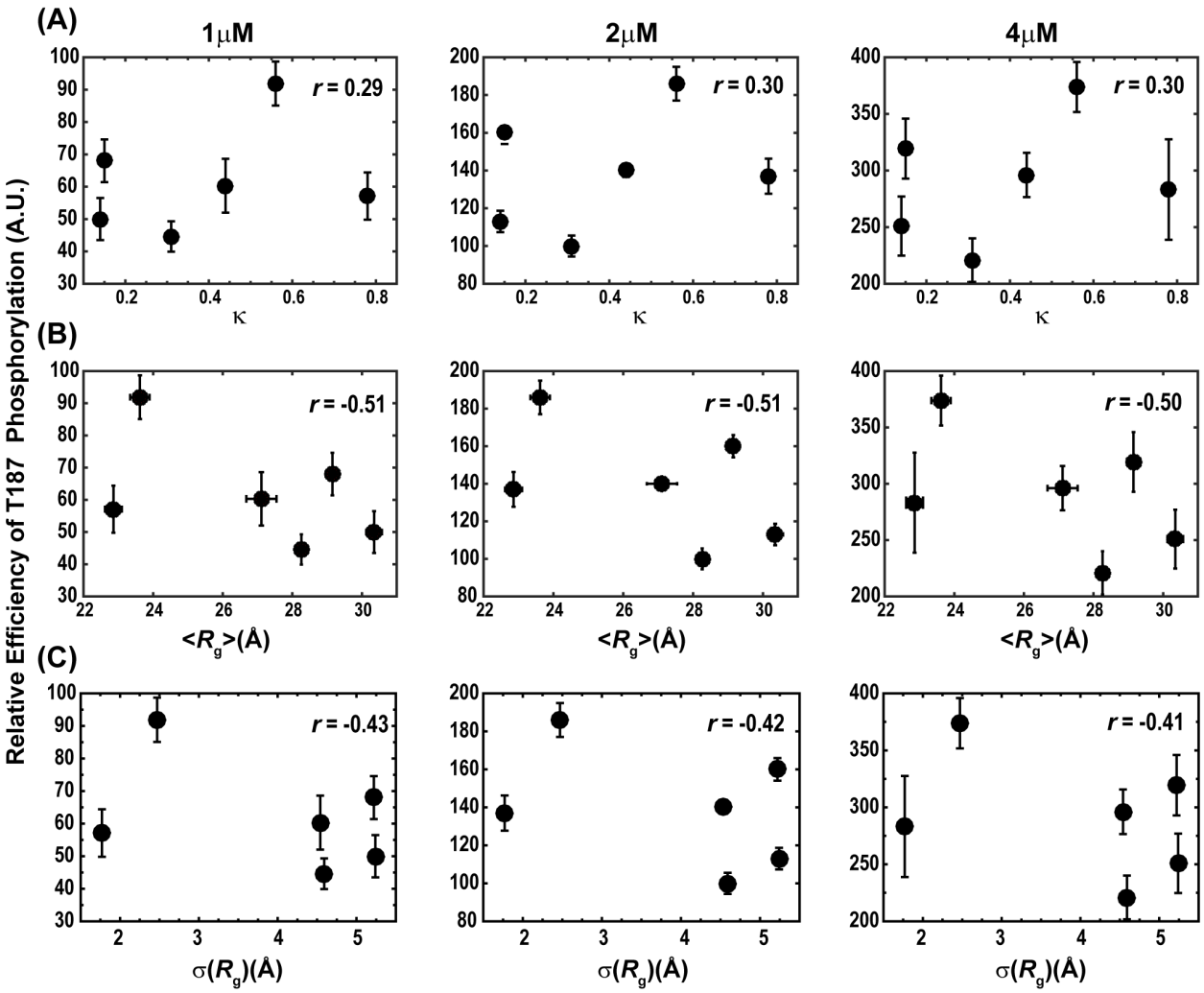

Fig. S15: Correlation analysis between experimentally observed in *cis* phosphorylation efficiencies and the charge patterning parameter or global conformational properties obtained from simulated trajectories. Each panel shows correlations obtained between data for in *cis* phosphorylation efficiencies and κ (panel A), R_g (panel B), and σ(R_g) (panel C) of p27-C-vXY variants. These correlations were calculated using data for phosphorylation efficiencies (in arbitrary units) measured at different concentrations of the ternary complex of p27-vXY/Cdk2/cyclin A. For each correlation analysis, we calculated the Pearson correlation coefficient *r*. Poor correlations of phosphorylation efficiencies with κ, R_g, and σ(R_g) suggest that changes in the global dimensions and the amplitudes of conformational fluctuations of the C-terminal tail are insufficient for explaining the observed trend in the functional data. The error bars in phosphorylation data denote standard deviations of the mean of the values for phosphorylation efficiency measured from three independent experiments.

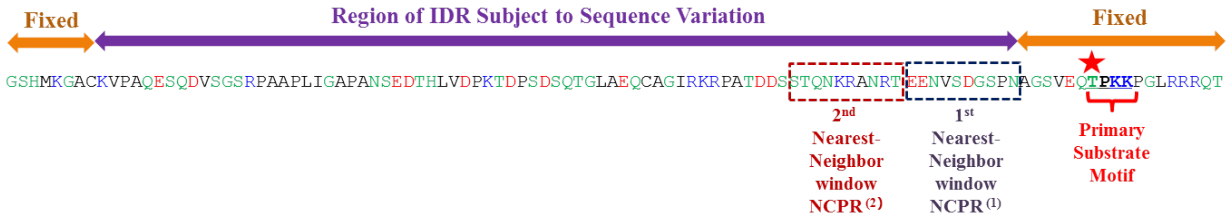

Fig. S16: Schematic showing how nearest neighbor windows (designated using boxes) directly N-terminal to the constrained region were defined to compute wNCPR. This illustration shows two windows with *L*=10. The primary substrate motif is highlighted by a curly parenthesis.

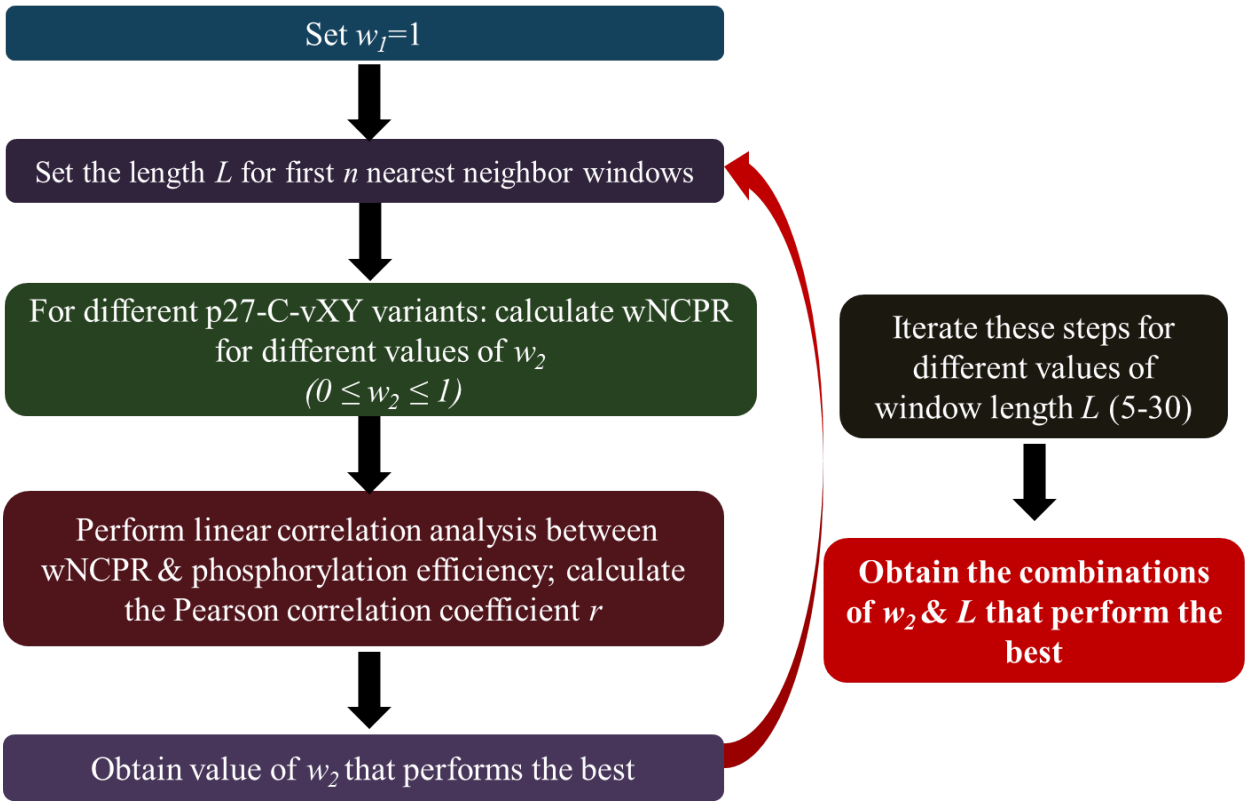

Fig. S17: Flowchart outlining the algorithm used for extracting the values of w_2 and L that yield the strongest correlation between wNCPR and the phosphorylation efficiency data.

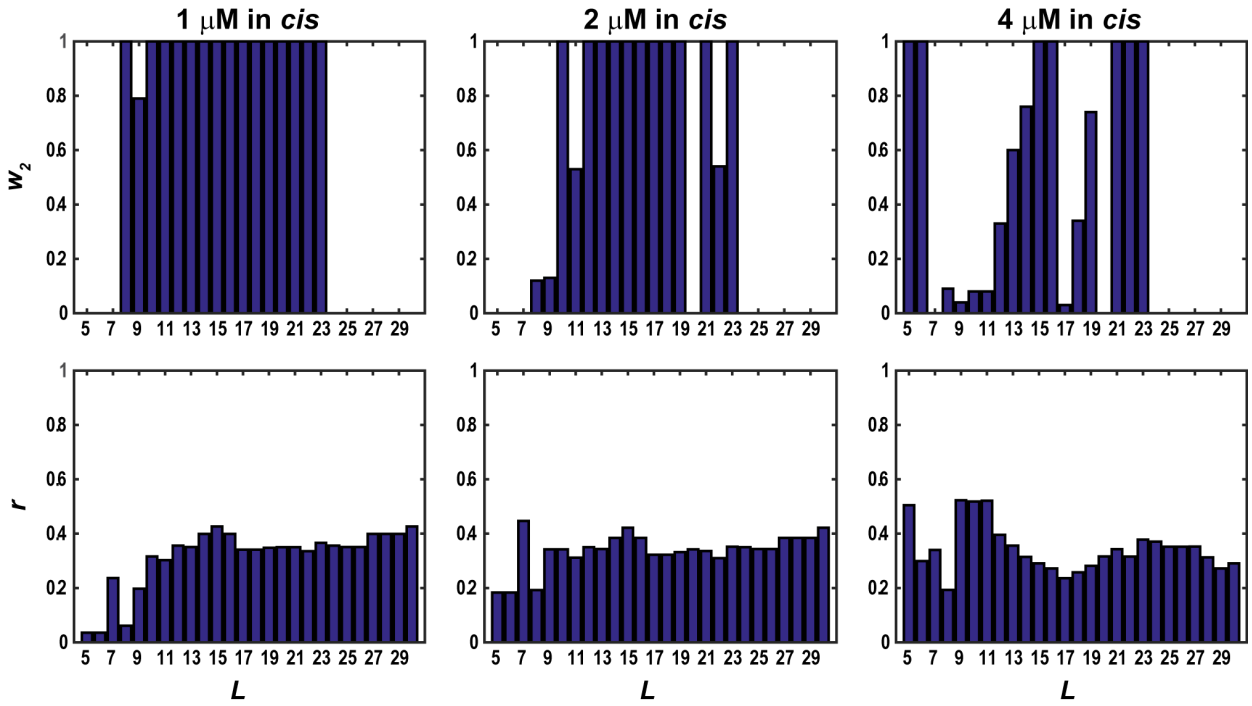

Fig. S18: Pearson correlation coefficient (*r*) quantifying the linear correlation between wNCPR and in *cis* phosphorylation efficiency when all sequences are included in the analysis. The analysis was performed for different values of *L*, where 5 ≤ *L* ≤30, and by setting *n*=2. For each of the *L* values, the best Pearson *r*-value and the corresponding w_2 are shown in bottom and top row, respectively. The columns show w_2 and *r*-values obtained using in *cis* phosphorylation efficiencies at different concentrations of the 1:1:1 p27-vXY variant/Cdk2/cyclin A ternary complexes are used (left to right 1 μM, 2 μM, and 4 μM, respectively) for the analysis.

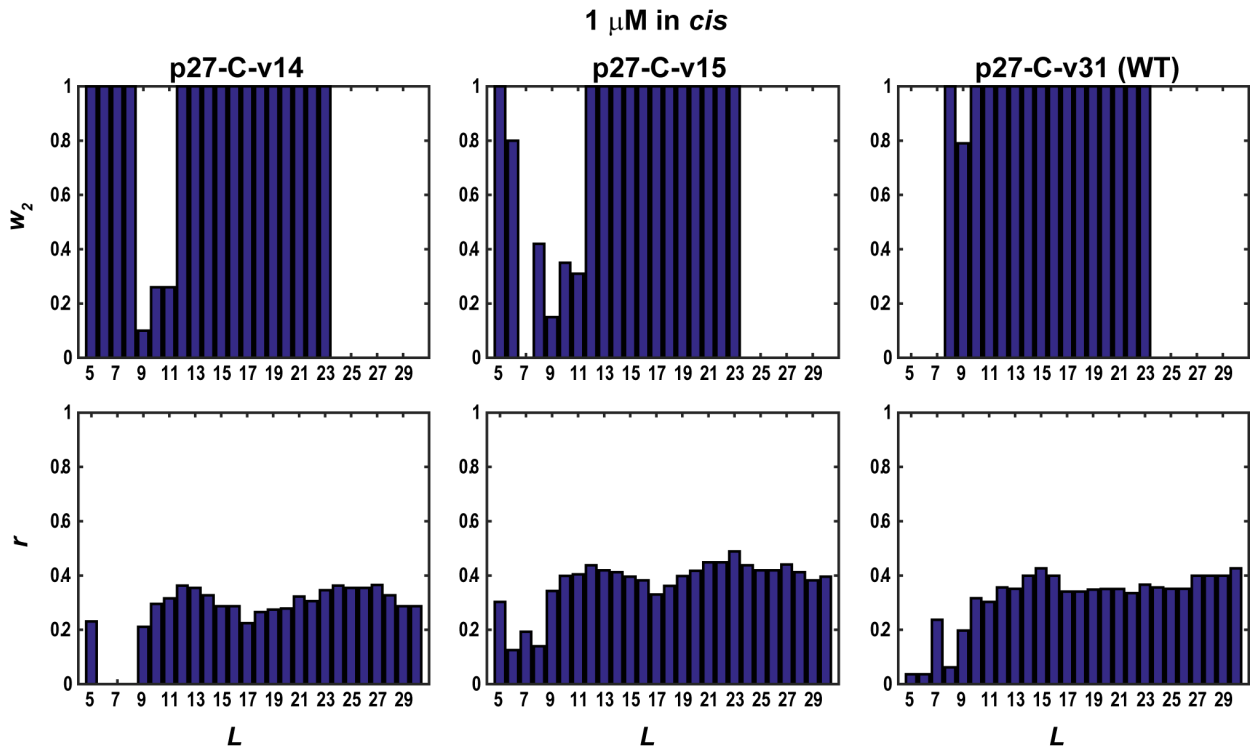

Fig. S19: Leave One Out Cross-Validation (LOOCV) Analysis: From left to right, the columns correspond to the statistics from analysis when p27-C-v14, p27-C-v15, and p27-C-v31, respectively, were left out. In all analyses, *n*=2. For different values of *L*, the best Pearson *r*-value and the corresponding w_2 value are shown in bottom and top rows, respectively. In all three cases, the best Pearson *r*-value never exceeded 0.5, indicating that p27-C-v14, p27-C-v15, and p27-C-v31 (WT) were not the sources of the observed poor correlation when all the sequences were used for analysis. In *cis* phosphorylation efficiency data at 1 μM were used for the analysis. The observations did not change for phosphorylation data at the two other concentrations (see following figures), showing the robustness of predictions from the model.

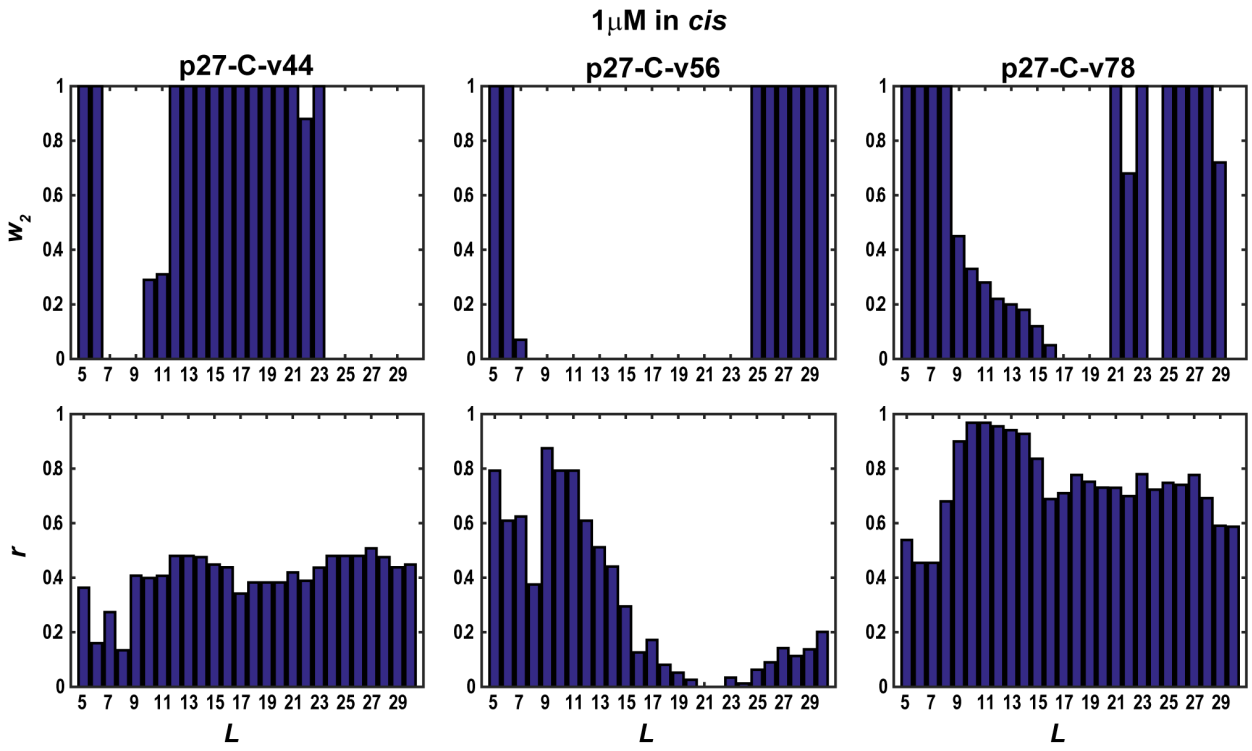

Fig. S20: LOOCV Analysis: From left to right, the columns correspond to the statistics from analysis when p27-C-v44, p27-C-v56, and p27-C-v78, respectively were left out. In all analyses, n=2. For different values of window length L, the best Pearson r-value and the corresponding w_2 value are shown in the bottom and top row, respectively. LOOCV analysis, when p27-C-v78 was left out, yielded the best correlation (r>0.95) over the longest range of L values (L=10-14), indicating that p27-C-v78 is the source of the observed poor correlation when all sequence variants are used for analysis. In *cis* phosphorylation efficiency data at 1 μM were used for the analysis. The observations did not change for phosphorylation data at two other concentrations (see following figures), showing the robustness of predictions from the model.

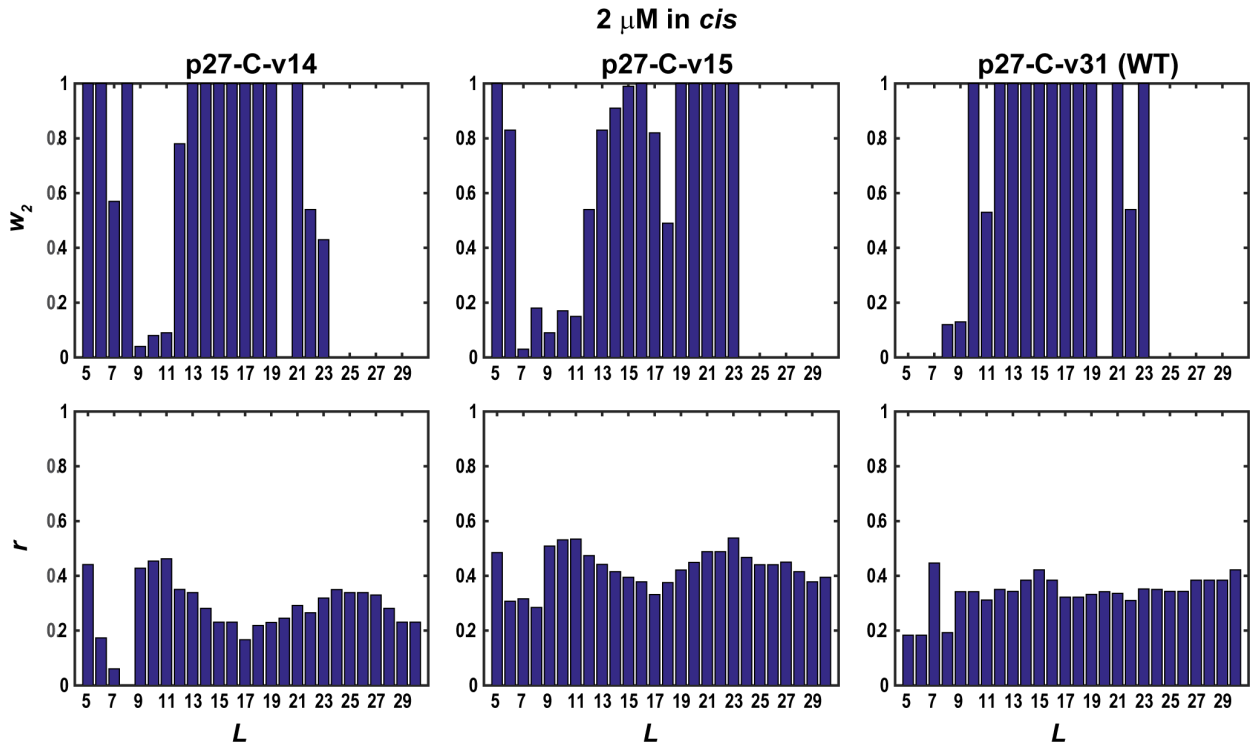

Fig. S21: LOOCV Analysis: From left to right, the column corresponds to the statistics from analysis when p27-C-v14, p27-C-v15, and p27-C-v31 (WT), respectively, were left out. In all analyses, *n*=2. For different values of *L*, the best *r*-value and the corresponding w_2 value are shown in bottom and top row, respectively. In all the three cases, the best *r*-value never exceeded 0.5, indicating that p27-C-v14, p27-C-v15, and p27-C-v31 were not the sources of the observed poor correlation when all the sequence variants were used for analysis. In *cis* phosphorylation efficiency data at 2 μM were used for the analysis.

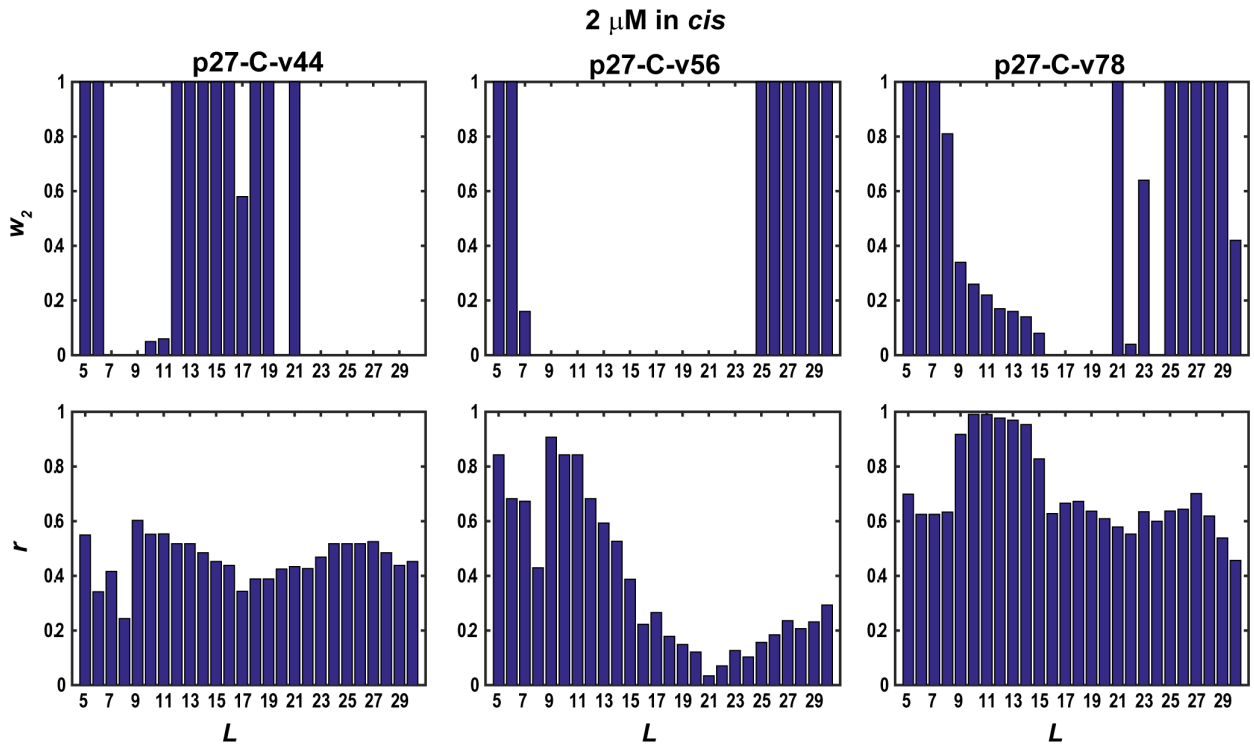

Fig. S22: LOOCV Analysis: From left to right, the column corresponds to the statistics from analysis when p27-C-v44, p27-C-v56, and p27-C-v78, respectively, were left out. For different values of *L*, the best *r*-value and the corresponding w_2 value are shown in bottom and top row, respectively. In *cis* phosphorylation efficiency data at 2 μM were used for the analysis. LOOCV analysis when p27-C-v78 were left out yielded the best correlation (r>0.95) over the largest range of *L* values (*L*=10-14).

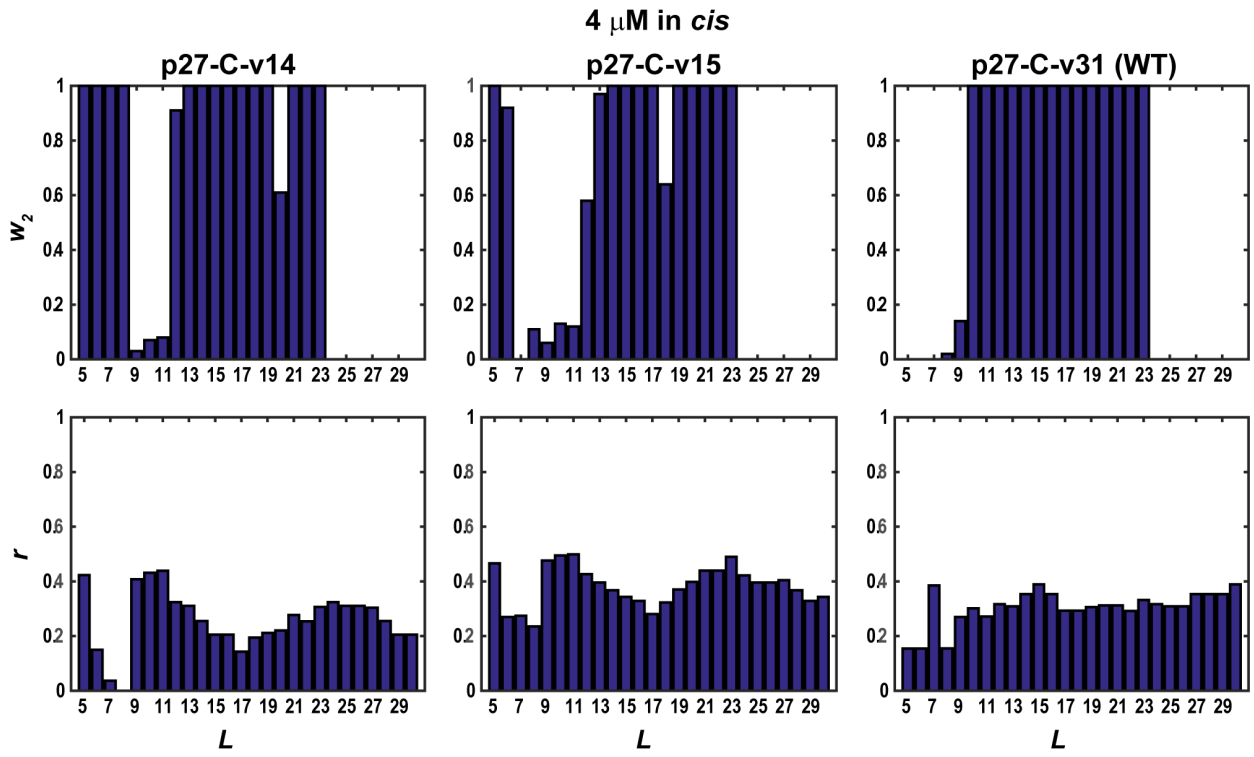

Fig. S23: LOOCV Analysis: From left to right, the column corresponds to the statistics from analysis when p27-C-v14, p27-C-v15, and p27-C-v31 (WT), respectively, were left out. In all analyses, n=2. For different values of L, the best r-value and the corresponding w_2 value are shown in bottom and top row, respectively. In all the three cases, the best r-value never exceeded 0.5, indicating none of p27-C-v14, p27-C-v15, and p27-C-v31 was the source of observed poor correlation when all the sequence variants were used for analysis. In *cis* phosphorylation efficiency data at 4 μM were used for the analysis.

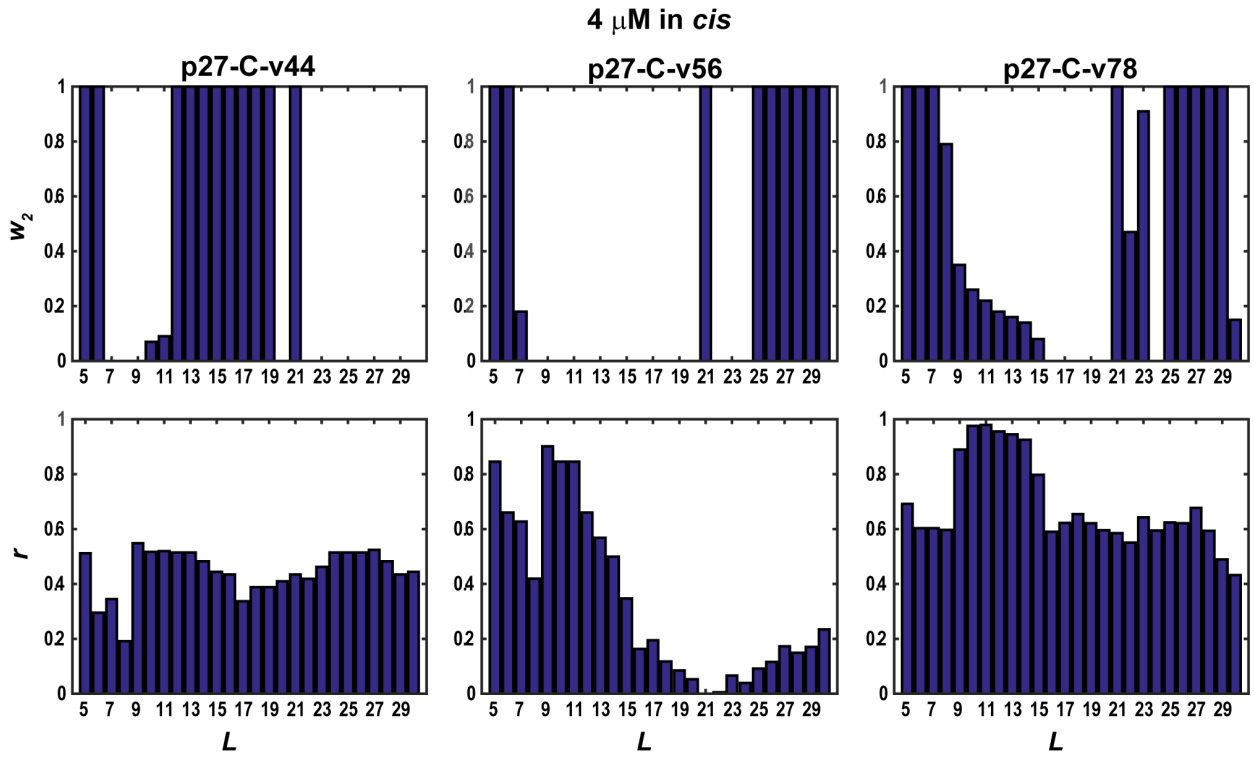

Fig. S24: LOOCV Analysis: From left to right, the column corresponds to the statistics from analysis when p27-C-v44, p27-C-v56, and p27-C-v78, respectively, were left out. In all analyses, n=2. For different values of L, the best r-value and the corresponding w_2 value are shown in bottom and top row, respectively. In *cis* phosphorylation efficiency data at 4 μM were used for the analysis. LOOCV analysis when p27-C-v78 was left out yielded the best correlation (r>0.95) over the largest range of L values (L=10-14).

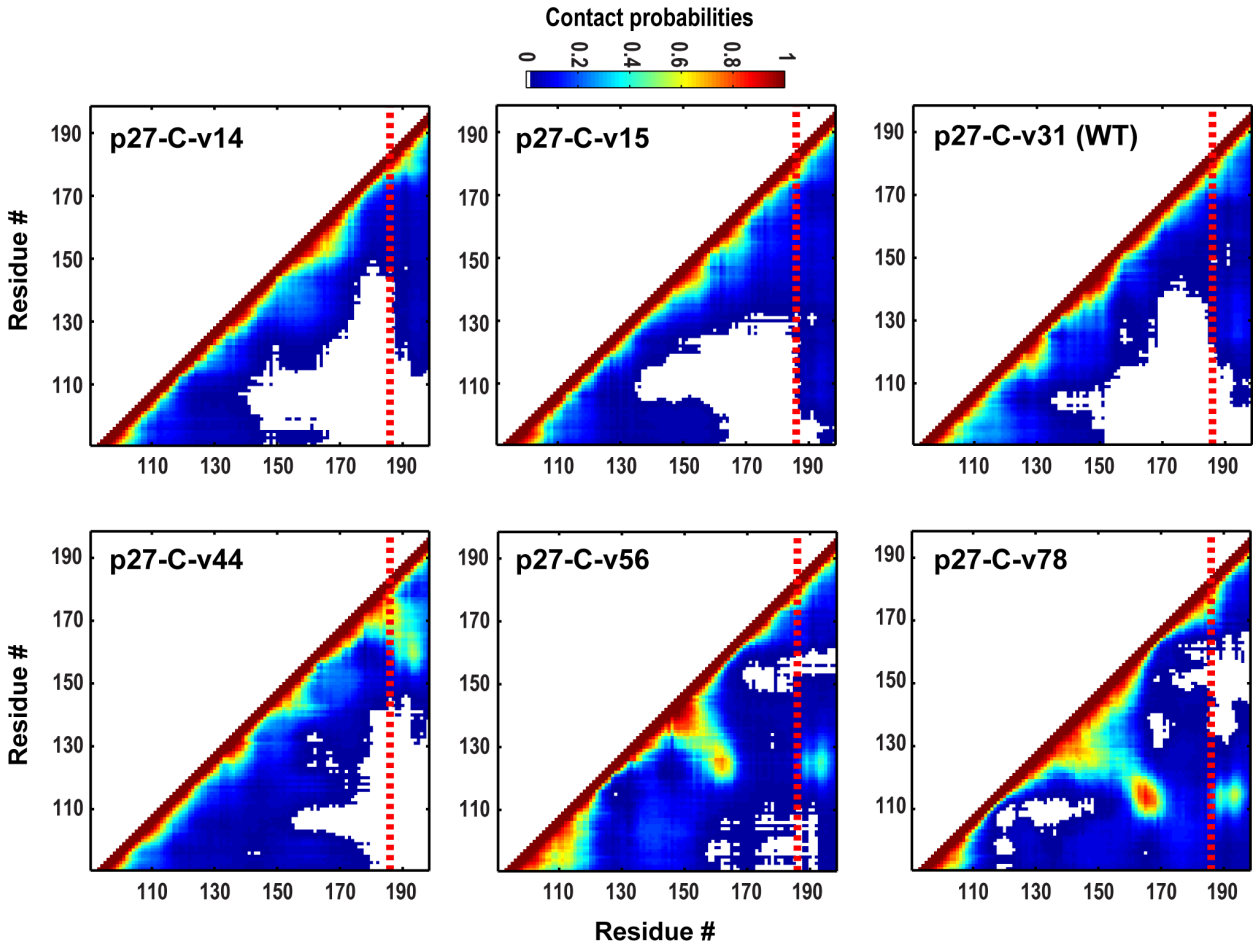

Fig. S25: Analysis of simulation results to quantify the ensemble-averaged inter-residue contact probabilities at 328 K for the p27 IDR in p27-C-vXY variants. A contact is defined between a pair of residues if there is at least one pair of atoms that is less than 15 Å apart. The vertical red dashed line denotes position of T187 in the sequences of p27-C-vXY variants. The color bar displays the color map and indicates the mapping of contact probabilities onto the color map. Long-range contacts between basic residues, C-terminal to the T187, with residues within the N-terminal acidic auxiliary motif are evident in p27-C-v56 and p27-C-v78 with the probability of the contacts being higher in p27-C-v78.

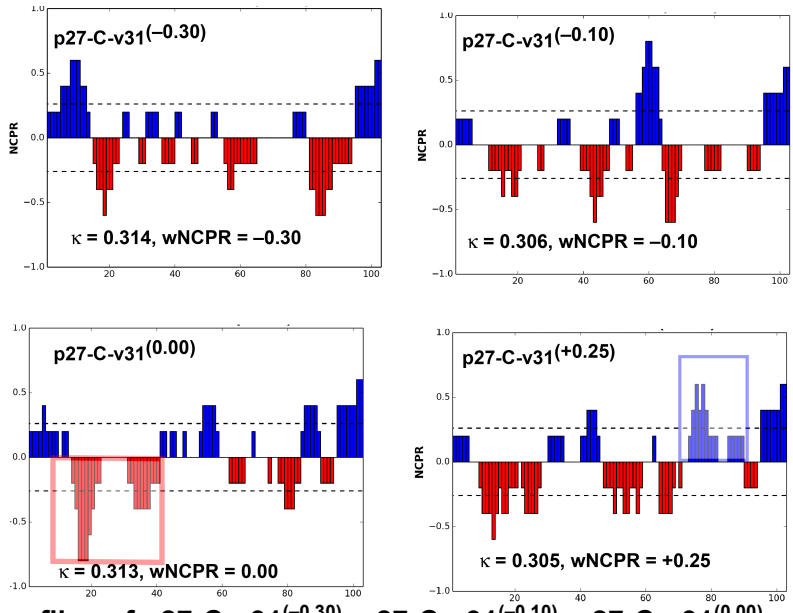

Fig. S26: NCPR profiles of p27-C-v31$^{(-0.30)}$, p27-C-v31$^{(-0.10)}$, p27-C-v31$^{(0.00)}$, and p27-C-v31$^{(+0.25)}$. This figure is analogous to the one shown in Fig. 2B. The amino acid sequences of these variants are shown in Table S2 below. The top row shows the NCPR profiles along the linear sequences of p27-C-v31$^{(-0.30)}$and p27-C-v31$^{(-0.10)}$, respectively. In these sequences, the positive and negative charges that lie outside the auxiliary motif are locally well mixed thus creating a situation where only the wNCPR of the auxiliary motif becomes a determinant of the T187 phosphorylation efficiency. In variant **p27-C-v31$^{(0.00)}$** accumulation of compensatory charges within the auxiliary motif leads to the accumulation of an acidic cluster near the N-terminal end in order to satisfy the constraint of realizing a sequence with a κ value of 0.31. This leads to a large, negatively charged attractive patch – see translucent box with red borders – that engenders long-range intra-IDR interactions that can sequester the positive charges within the primary and auxiliary motifs thus diminishing the T187 phosphorylation below what is expected based on the wNCPR value of the auxiliary motif. As a direct comparison, the functional data for p27-v15 has an auxiliary motif with wNCPR = 0 but the normalized relative efficiency of T187 phosphorylation is higher because of the lower κ value. This point is also made by comparing the NCPR profile of **p27-C-v31$^{(0.00)}$ to that of p27-C-v15 from Fig. 2 in the main text.** The accumulation of positive charge within the auxiliary motif of p27-C-v31$^{(+0.25)}$ – see translucent box with blue borders – engenders long-range attractions with acidic clusters that are situated near the N-terminal end in order to realize a κ value of 0.31. This leads to a sequestration of residues from the primary and auxiliary motifs in long-range attractions as shown in Fig. S27. To design the additional variants, we generated a set of 22 designs with essentially fixed κ values of 0.31 and wNCPR values for the auxiliary motif that range from –0.53 to +0.4. The set was reduced to seven representative sequences based on preliminary tests for expression of the full-length variants in *E. coli*. Of these seven, four expressed robustly and were amenable to further purification using the protocols described in Section 3.

Table S2: Sequences of p27-C-v31$^{(-0.30)}$, p27-C-v31$^{(-0.10)}$, p27-C-v31$^{(0.00)}$, and p27-C-v31$^{(+0.25)}$

ID	Amino Acid Sequence
p27-C-v31$^{(-0.30)}$	GSHMKGACKKNKIRSISDSEPEAGPANRPDAASSRASVDAVSQKQDLNAT GTLQRHESDSTVPEGPTTPTSANCPANLGRGVPDDQDQEAGSVEQTPKK PGLRRRQT
p27-C-v31$^{(-0.10)}$	GSHMKGACSPHSGAEALPDAAEQPKVDANGPLGVAKTSTDRGDSDADAT GSKTSEVQCRRVKRRPSEDENSTQSIPQLPNDNNSQTIPAAGSVEQTPKK PGLRRRQT
p27-C-v31$^{(0.00)}$	GSHMKGACKSAAGPKVEDDDSDPPPSSGATPSAAEHDPGEAAQQKLERA SNTNTVPKRTSTGQGQDVCVSNSRETQSANIDLDLNIPRRAGSVEQTPKKP GLRRRQT
p27-C-v31$^{(+0.25)}$	GSHMKGACPTSVDDVLDANETANAIDDSGVQTQKQCPPPTAIQRSRPLPE SAQDTAEPSSDHKGGSEDGALSTERNRRGVKNASSANPKAGSVEQTPKK PGLRRRQT

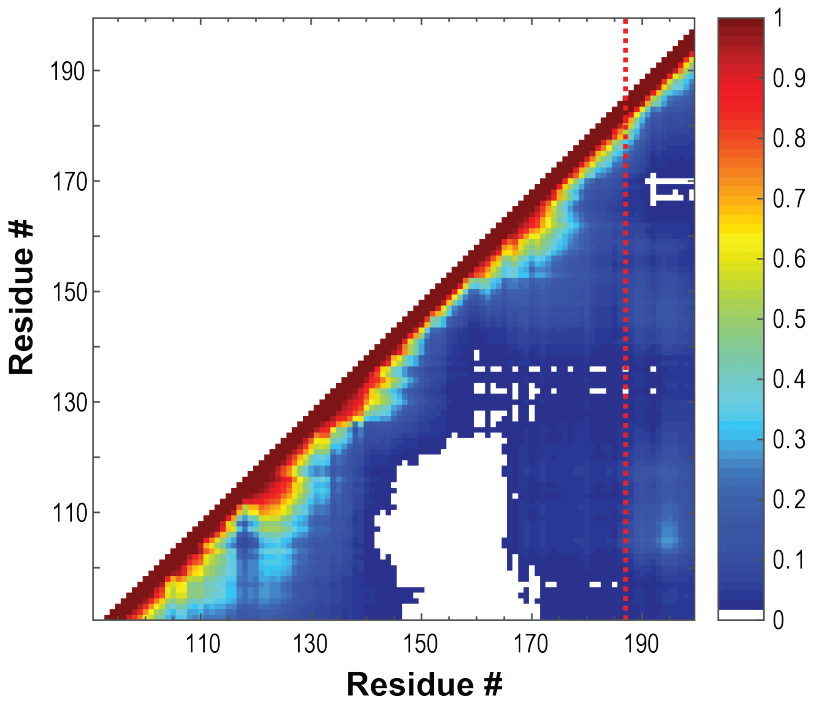

Fig. S27: Analysis of simulation results to quantify the ensemble-averaged inter-residue contact probabilities at 328 K for p27-C-v31$^{(+0.25)}$. The definition of a contact is in accord with the definition used in our analysis summarized in **Fig. S25**. The vertical red dashed line denotes position of T187 in the sequence. The color bar displays the color map and indicates the mapping of contact probabilities onto the color map. Long-range contacts between basic residues, C-terminal to the T187, with residues within the N-terminal acidic auxiliary motif are clear and are to be contrasted with the absence of such contacts for the wild type sequence – **Fig. S25**.

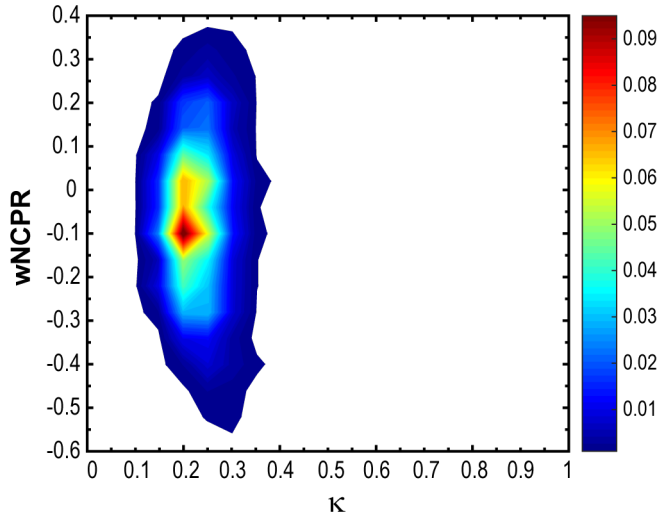

Fig. S28: Probability density map quantifying the probability of finding sequences with particular κ AND auxiliary motif wNCPR values. The color bar on the right provides an annotation for the density map such that $\rho(\kappa,w)\Delta\kappa\Delta w$ is the probability of finding a sequence with a κ value between κ and $\kappa+\Delta\kappa$ AND an auxiliary motif with wNCPR value between w and $w+\Delta w$. Here, $\rho(\kappa,w)$ is the value of the probability density for (κ,w). The sequences were enumerated using the amino acid composition of the p27 IDR and the design constraints that are summarized in the main text. The salient messages from this analysis are two-fold: First, a wide range of auxiliary motif wNCPR values are compatible with a given κ value and it therefore follows that whether or not the auxiliary motif is engaged in long-range intra-IDR interactions will be determined by the sequence-encoded NCPR profile – see **Fig. S26**. Second, the highest number density of sequences corresponds to **κ** and wNCPR values that are approximately 0.2 and –0.1, respectively. These values are distinct from the parameters for the wild type p27 IDR for which the values are 0.31 and –0.2. Hence, the inference is that the wild type parameters would have a very low likelihood of being chosen at random. This points to a strong selection pressure for the wild type parameters – a feature that is accentuated by the apparent negative regulation of signaling T187 phosphorylation that characterizes the wild type sequence.

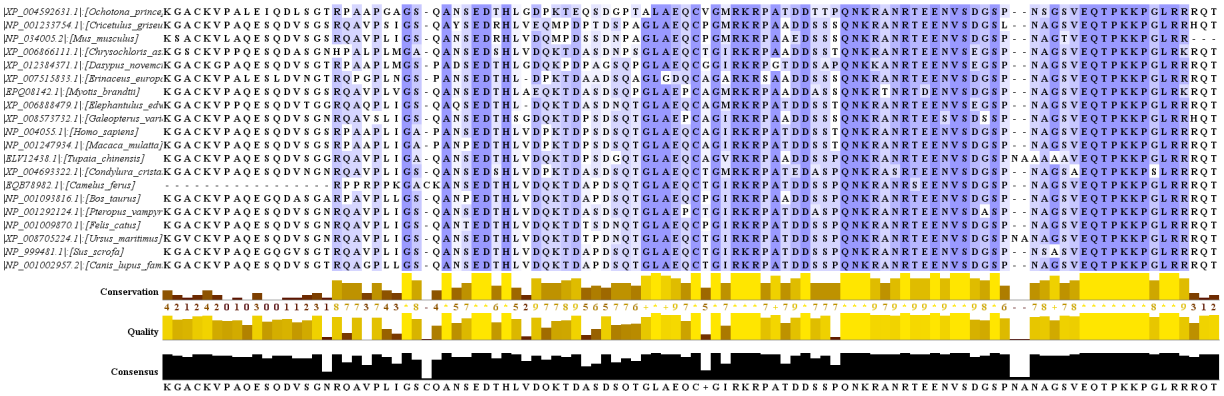

Fig. S29: Results from multiple sequence alignment analysis of the p27 IDRs across orthologs. A BLASTP (version 2.3.1) search was performed using the sequence of human p27-C as a template. This analysis was performed using the NCBI web interface. Sequences belonging to the same order in taxonomy were excluded from the top 100 hits. The final list had sequences from 20 species that belong to different orders. The alignment was performed using the Jalview software and was colored according to BLOSOM62 scores. In terms of percent identity, the conservation of the sequence context, including the region encompassing the auxiliary motif is slightly lower than the level of conservation for the primary substrate motif. However, the amino acid composition within the auxiliary motif is absolutely conserved. As a reminder, the auxiliary motif is defined as the first 20-residues directly N-terminal to the sequence window that encompasses the primary substrate motif.

Section 2 – Details of simulation methods, assessments of simulation quality, and analysis of simulation results

ABSINTH implicit solvation model and forcefield parameters: Details of the ABSINTH model and forcefield paradigm have been published previously (4-7). Parameters for all of the interaction terms in the forcefield were taken from the abs3.2_opls.prm parameter set. This set incorporates ABSINTH-based parameters for van der Waals interactions, torsional potentials, and reference free energies of solvation for solvation groups. The partial charges and the neutral group paradigm are adapted from the OPLS-AA/L forcefield (8). In the ABSINTH paradigm the effects of solvent-mediated interactions are captured using an implicit representation of the solvent that models the contributions from conformation-dependent dielectric inhomogeneities. The effects of mobile ions are simulated using explicit representations of Na^+ and Cl^- ions. For this, we used the parameters developed by Mao and Pappu (9). In accord with the empirical choices recommended in the original ABSINTH paper (5), we set the reference free energies of solvation for the charged sidechains of Asp, Glu, Arg, and Lys to be more favorable by 30 kcal/ mol vis-à-vis default values that are based on estimates from experiments. This choice, as explained in published work, ensures against artifacts due to the formation of spurious salt-bridges and the use of a fixed charge model (the charges are fixed by the pKa values of amino acids at neutral pH) rather than a constant pH simulation paradigm (10). For proline residues we utilize the modified parameters for bonded and Lennard-Jones interactions that have been developed by Radhakrishnan et al. (11).

Details regarding the MMC move sets: Conformational space is sampled using Metropolis Monte Carlo (MMC) simulations. The degrees of freedom for these MMC simulations include the backbone torsion angles ϕ, ψ, and ω, sidechain torsion angles χ, and the rigid body coordinates of polypeptides and solution ions (5). The move sets include translation of ions combined with proposals of small- and large-scale conformational changes of the polypeptide

degrees of freedom. The latter are achieved through a combination of local, pivot, and concerted moves and their frequencies are based on decision tree used in previous work (4). Conformations of sequences with proline residues utilize the improved move sets published by Radhakrishnan et al. (11). These move sets include the deformation of pyrrolidine rings, the modeling of ring puckering, and the coupling between ring puckering and backbone degrees of freedom.

Simulation set up: All simulations and analyses were performed using the CAMPARI molecular modeling software suite (http://campari.sourceforge.net). For simulations with the full ABSINTH model, all polypeptides and ions were modeled in atomic detail. The simulations were carried out using spherical boundary conditions. In each simulation, the system comprised the polypeptide chain, neutralizing Na^+, Cl^- ions plus 49 excess ion pairs to mimic 10 *m*M NaCl enclosed within a spherical droplet of radius 125 Å. The choice for the droplet radius was justified using the end-to-end distance distributions for all sequences simulated in the excluded volume (EV) limit. These simulations revealed that none of the sequences sampled conformations with end-to-end distances larger than 250 Å. This argues against artifacts due to confining effects posed by small droplet sizes. The cut-off distances for van der Waals interactions and for electrostatic interactions between charge-groups that are net neutral were set to 10 Å and 14 Å, respectively. No cut-offs were used for electrostatic interactions involving Na^+ and Cl^- ions and the charge-groups of Asp, Glu, Arg, and Lys sidechains.

We also performed three types of simulations that serve as atomistic reference models for all sequences. These simulations are denoted as EV, FRC, and LJ in the text and figure legends in the *SI Appendix*. The EV limit refers to the excluded volume limit. Two sets of reference simulations (EV and LJ) were performed using the ABSINTH model while zeroing out the mean field solvation and Coulomb terms of the potential. All other terms of the potential were used as prescribed by the ABSINTH model. The two reference potentials *viz*., EV and LJ are distinguished by the choice of λ in equation (1). In one set of reference simulations, λ=0 (EV) and in the other λ=1 (LJ).

$$U_{\text{ref}} = 4\sum_i \sum_{j<i} \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \lambda \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]; \qquad (1)$$

The summation runs over all unique pairs of non-bonded atoms as defined by the ABSINTH model. Metropolis Monte Carlo simulations were performed at a simulation temperature of 298 K. The temperature dependence of the reference ensembles is negligible. The parameters for ε_{ij}, σ_{ij} and other non-zero terms of the potential were taken from the abs3.2_opls.prm parameter file (see http://campari.sourceforge.net). In the excluded volume (EV) limit, quantities such as the mean radius of gyration R_g, the mean end-to-end distance R_{ee}, and the mean hydrodynamic radius R_h scale as $N^{0.59}$ with chain length N. These features are reproduced using ensembles obtained by performing simulations with λ=0 in equation (1) and zeroing out the mean field solvation and electrostatic terms of the ABSINTH potential. By pursuing a similar approach and setting λ=1 we obtain the reference globule or Lennard-Jones (LJ) limit wherein quantities such as R_g and R_h scale as $N^{0.33}$ with chain length. In this limit, the conformations are, on average, non-specifically compacted globules.

We also performed reference simulations using the rotational isomeric approximation to mimic the Flory random coil or FRC limit (12, 13). The ABSINTH model with λ=0 in equation (1) combined with the mean field solvation and electrostatic terms being zeroed out was used to

perform Metropolis Monte Carlo simulations of dipeptides i.e., Ac-Xaa-Nme for all twenty amino acids (Xaa) at 298 K. The distributions of ϕ, ψ, and χ angles from the dipeptide simulations were used to create libraries of rotational isomers for every amino acid. To generate FRC ensembles for longer chains, ϕ, ψ, and χ angles were randomly drawn from residue-specific libraries of rotational isomers. In these simulations all inter-residue interactions are explicitly ignored. The resultant ensembles conform to Flory's approach for mimicking conformational distributions that result from the counterbalancing of chain-chain and chain-solvent interactions in an indifferent or theta solvent (12, 13). Quantities such as R_g, R_{ee}, and R_h scale as $N^{0.5}$ as a function of chain length for all systems in the FRC limit. Similarly, distributions for a range of polymeric quantities match expectations from theory and simulation for chains in a theta solvent.

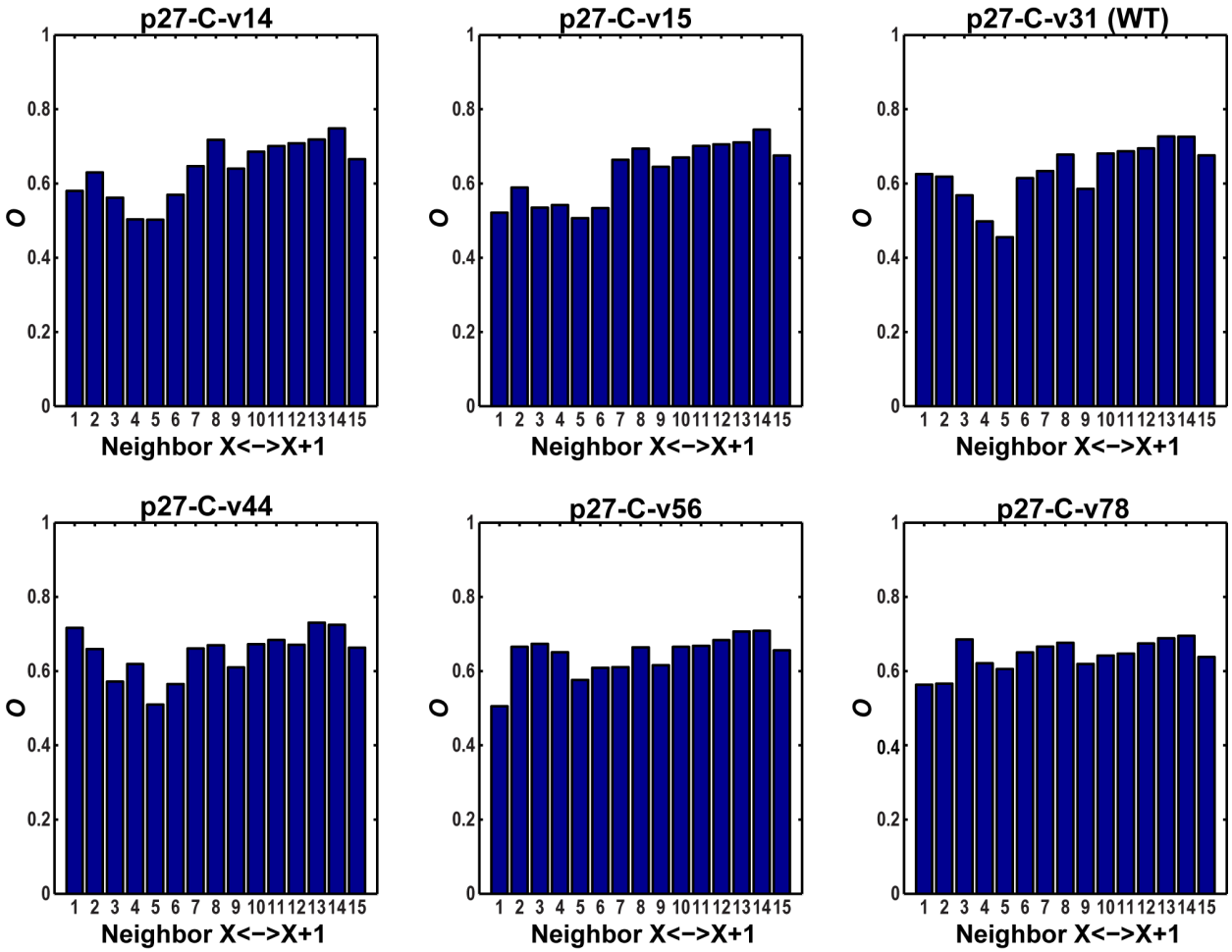

Fig. S30: Overlap statistics between nearest-neighbor temperature replicas in TREx simulations of different p27-C-vXY variants. The overlap statistics is defined as the overlap fraction (*O*) between energy distributions of nearest-neighbor temperature replicas. The bars show the mean overlap statistics over three independent TREx runs.

Temperature Replica Exchange simulations: To enhance the quality of sampling we deployed temperature replica exchange (TREx) simulations (14). For all the simulations, the temperature schedule comprised sixteen temperatures: [275K, 282K, 290K, 298K, 308K, 318K, 328K, 338K, 348K, 365K, 382K, 400K, 420K, 440K, 460K, and 485K]. The choice of the

temperature schedule was justified from the computed overlap statistics for specific observables between neighboring temperature replicas (**Fig. S30**). For a pair of temperature replicas X and X+1, the overlap fraction *O* between the sampled distributions was defined as shown in equation (2):

$$O = 1 - \frac{\int_{E=E_{min}}^{E=E_{max}} |P_X(E) - P_{X+1}(E)|\, dE}{2} \qquad (2)$$

In equation (2), E is the energy associated with a particular conformation, $P_X(E)$ is the probability density function in replica X, and E_{min}, E_{max} refer to a pair of prescribed lower and upper limits for the energies.

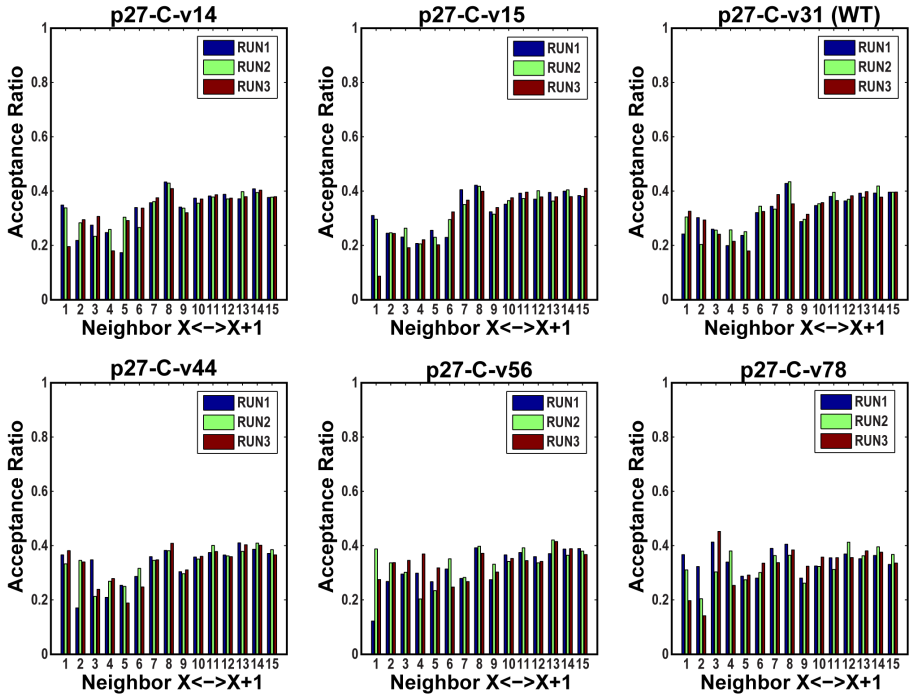

Fig. S31: Acceptance ratios of swaps between nearest-neighbor temperature replicas in TREx simulations of different p27-C-vXY variants. The three bars denote acceptance ratios for three independent TREx runs.

The ratio for accepting proposed swaps between neighboring temperature replicas was greater than 0.25 on average (see **Fig. S31**). For each temperature replica, the simulation was initiated using a conformation drawn at random from a prior simulation in the EV limit. To improve the overall statistics and to assess reproducibility, there were three independent TREx simulations for each sequence. For a given sequence, each TREx simulation comprises of 6.15×10^7 MMC steps. The first 4×10^6 steps were discarded as equilibration steps. Swaps between pairs of neighboring replicas were attempted every 5×10^4 steps. For each simulation, snapshots from the MMC trajectories were saved once every 4×10^3 steps. This yielded ~1.3×10^4 uncorrelated conformations per trajectory. All subsequent analyses were performed using these conformations.

Extracting conformational ensembles whose properties are characterized by high statistical reliability: Simulations based on importance sampling, which include MMC methods and molecular dynamics simulations, are characterized by intrinsic errors (i.e., errors due to inaccuracies of the molecular mechanics forcefields) and statistical errors, which are errors associated with the limitations of finite sampling and broken ergodicity. From the perspective of finite sampling, the statistical error will decrease as the inverse square root of the number of uncorrelated samples drawn from the canonical ensemble.

For a given p27-C variant, our goal was to identify the lowest simulation temperature, designated as $T_{L,R}$, within the temperature schedule that yields statistically reliable estimates of conformational properties. The protocol we deployed to identify the value of $T_{L,R}$ was guided by the observation that all p27-C variants adopt coil-like ensembles. Even variants that are compact are considerably more expanded than the reference LJ globules. This implies that the statistics of various conformational properties are concordant with expectations from the Flory random coil limit. Indeed, this convergence toward or fluctuation around the FRC limit is anticipated from the annotation of p27-C as a coil-forming sequence from the predictive diagram-of-states that has been developed for IDPs (15, 16). Convergence toward or fluctuation around the FRC limit implies that, on average, in the simulations with the full interaction model the sequence-encoded intra-chain interactions counterbalance the sequence-encoded chain-solvent interactions. This should lower the barriers for interconversion between distinct conformations at or above a suitable threshold sampling temperature. Below this threshold temperature, the overall sampling quality, quantified by the statistical reliability of observables, will be influenced by the ruggedness of the underlying energy landscape, which creates metastable traps. Therefore, for the collection of sequence variants that have identical amino acid compositions, we identified the threshold temperature $T_{L,R}$ above which the finite sampling leads to statistically reliable descriptions of the conformational properties. Conversely, below $T_{L,R}$, the statistical reliability of finite sampling is expected to be low on account of sequence-specific metastable traps that limit the efficiency of conformational sampling.

We developed an unbiased protocol to identify the value of $T_{L,R}$ beyond which the simulated ensembles of all variants show the lowest statistical error and hence the highest statistical reliability. For each p27-C sequence variant, we used the following protocol: (1) For each of the simulation temperatures, we used random sub-sampling from the set of three independent TREx trajectories to generate 45 sub-ensembles each comprising 10^3 conformations drawn from the corresponding ensembles. We used a procedure based on bootstrap sampling with replacement. (2) For each simulation temperature, the sub-ensembles enable the calculation of 990 distinct pairs of values for the overlap fractions (O) of energy distributions using the formula shown in equation (2). These values were used to calculate the minimum value of O i.e., $O_{min}(T)$ for each temperature T. Next, we quantified the value of O_{min} for the FRC model. In this model, the sampling is not limited by the presence of energy barriers or metastable traps. Accordingly, the value of O_{min} we calculate from this reference ensemble is governed exclusively by finite sample sizes, i.e., the only bottleneck is entropic in nature and this reference ensemble sets a natural upper bound on conformational heterogeneity (17).

For each of the p27-C variants in the FRC limit, we deployed the procedure of creating sub-ensembles using bootstrapping with replacement and quantified the value of O_{min}(FRC). The value of O_{min}(FRC) is approximately 0.8. For each sequence variant, we next compared the values of $O_{min}(T)$ for all the simulation temperatures to O_{min}(FRC). The lowest simulation temperature for which $O_{min}(T) \geq O_{min}$(FRC) is the value we choose for $T_{L,R}$. Using this strategy, we determined the value of $T_{L,R}$ to be 328 K for all of the p27-C sequence variants (see **Fig.**

S32). Therefore, we used the ensembles drawn from 328 K for all subsequent analysis including comparisons to experimental data.

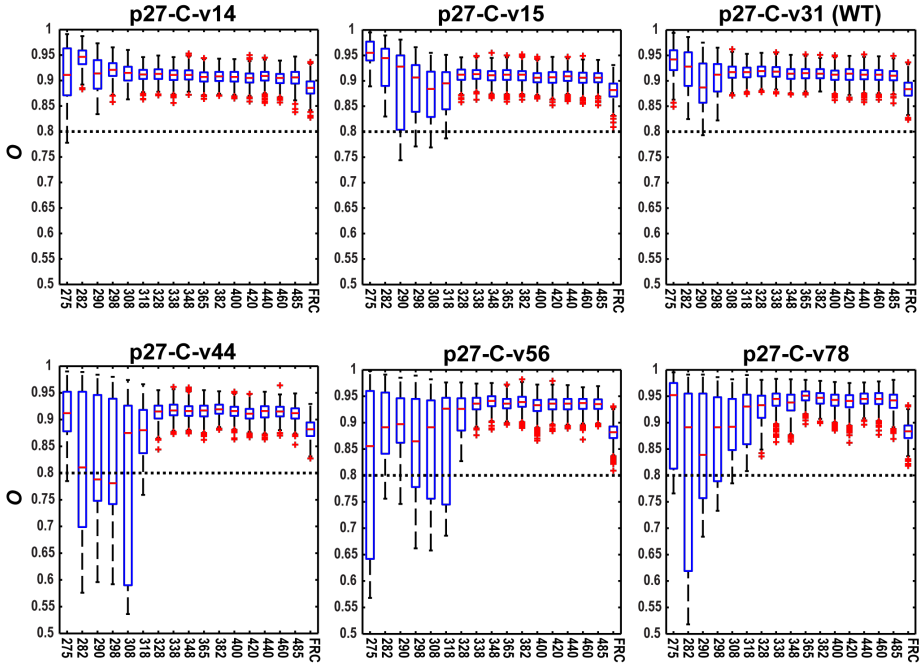

Figure S32: Box plots summarizing the results from unbiased analysis used to extract $T_{L,R}$. The box plots show the distributions of overlap fractions (O) highlighting the values of the minimum (O_{min}), first quartile, median, third quartile, maximum, and outliers for each temperature as well as the FRC limit. The numerical tick labels along the abscissae are the simulation temperatures in Kelvin. The horizontal dashed black line denotes the threshold value for O_{min}, which was set as 0.8. This value was used to select the value of $T_{L,R}$ as the lowest temperature that yields statistically reliable simulation results for subsequent analysis and comparisons to experimental data.

Section 3 – Details of experimental methods and analysis of experimental data

Constructs and Protein Purification: The generation of the p27, Cdk2, and cyclin A proteins as well as the preparation of the ternary complex have been described previously (18). p27-vXY (full-length) and p27-C-vXY (residues 96-198) constructs were generated by insertion of synthetic DNA sequences (Integrated DNA Technologies) into a pET28a vector (Novagen). The initial p27-vXY constructs bear the Y88E and Y89F mutations to mimic Y88 phosphorylation (19). For kinase assays probing the *trans* phosphorylation mechanism, the Y88E mutation was reverted to tyrosine with site-directed mutagenesis. The G192C constructs of p27-C-v14, p27-C-v31, and p27-C-v56 used in PRE experiments had all other cysteine residues mutated to alanine. All variants were generated using the QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene).

p27 variants were expressed in *E. coli*, purified by Ni^{2+} affinity chromatography, His-tags removed by cleavage with thrombin or TEV, and further purified by reverse phase HPLC. p27-v56 has an internal thrombin site and therefore the His-tag cleavage site was mutated to a TEV site. The p27-v44 and p27-v78 proteins were partially proteolysed during expression and required additional ion-exchange chromatography steps to obtain high purity. ^{13}C and ^{15}N labeled proteins for NMR were prepared by expression in minimal media and purified as described above (20, 21). The G192C proteins were labeled with (2,2,5,5-Tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) (Toronto Research Chemicals Inc.) by incubating the purified proteins at a concentration of 0.5mM in 20mM NaPi, pH 7.2, 50mM NaCl with a ten-fold molar excess of the spin label in the dark at 4 °C overnight. The MTSL was added from a 190 mM stock in acetonitrile resulting in a final concentration of < 3% acetonitrile in the reaction mixture. Excess MTSL was removed by repeated dilution and concentration steps resulting in a final dilution factor of over 10^4.

SAXS data collection and analysis: p27-C-vs and p27-v/Cdk2/cyclin A complexes were dialyzed in high salt buffer (20 mM NaPi, pH 7.2, 150 mM NaCl, 5 mM DTT) or low salt buffer (20 mM NaPi, pH 7.2, 50 mM NaCl, 5 mM DTT) for 48 hours at 4 °C and concentrated to 1, 2, and 4 mg/ml total final protein concentration. SAXS data were collected at the ALS beamline 12.3.1 (SIBYLS) at Lawrence Berkeley National Lab, Berkeley, California (22-24). The X-ray wavelength λ was 1.0 Å and the sample-to-detector distance was 1.5 m. The scattering vector q ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle) ranged from 0.01 $Å^{-1}$ to 0.32 $Å^{-1}$. All experiments were performed at 20 °C. Data were analyzed by the SIBYLS ScÅtter analysis package.

NMR Spectroscopy: NMR data sets were collected on Bruker spectrometers operating at 600 and 800 MHz on ca. 0.5 mM protein samples in 20 mM NaPi, pH 7.2, 50 mM NaCl, 5 mM DTT, 10% D_2O unless otherwise noted. Assignments were obtained from HNCA, HNCACB, CBCAcoNH, HNCO, and ^{13}C-detected CoN and CACON spectra. Secondary chemical shifts were determined relative to sequence corrected random coil chemical shifts (http://www1.bio.ku.dk/english/research/pv/sbin_lab/staff/MAK/randomcoil/script/) (25). Due to the much better chemical shift dispersion in the CoN spectra, ^{15}N relaxation data were detected via the carbonyl (26). Relaxation delays were 50, 200, 300, 400, 600, 800 ms for T1 and 16, 48, 96, 144, 192, 240 and 288 ms for T2. Relaxation data were collected in interleaved pseudo-3D data sets in order to minimize potential instrumental errors. Errors in the ^{15}N relaxation data were estimated by fitting errors. NMR data were processed and analyzed with nmrPipe (27) and CARA (http://cara.nmr.ch). PRE values were collected by the ratios of peak intensities in HSQC spectra of the MTSL-labeled proteins before and after the addition of a 5-fold molar excess of DTT to reduce the spin label. Before data collection on the reduced samples, the

protein was incubated at room temperature for 1 hour. Errors in PRE data are derived from the noise level in the spectra.

Kinase assays: The kinase buffer was 20 mM HEPES (pH 7.3), 25 mM Sodium β-glycerolphosphate, 15 mM $MgCl_2$, 16 mM EGTA, 0.5 mM Na_3VO_4, and 10 mM DTT. Size-exclusion purified p27-vXY/Cdk2/cyclin A complexes with concentrations ranging from 0.1 μM to 4.0 μM were incubated with 6 μCi [γ-^{32}P]-ATP and 40 μM non-radioactive ATP at 37 °C for 10 min (for p27-v-Y88E/Y89F) or 60 min (for p27-v-Y89F). The reactions were then terminated by SDS loading buffer and the products were separated by 10% SDS-PAGE. The ^{32}P incorporated p27 was detected using a phosphorimager (Typhoon 9200, GE Healthcare) and analyzed using the TotalLab Quant software (Totallab).

Analytical size-exclusion chromatography: p27-vXY and the Cdk2/cyclin A complex were incubated with 1:1:1 molar ratio at a concentration of approximately 5 μM at 4 °C for 1 hour in size-exclusion buffer (20 mM HEPES, pH 7.5, 300 mM NaCl, and 5 mM DTT). The samples were then concentrated to 0.2 mM and loaded into a Superdex 200 10/300 GL column (GE Healthcare). Samples were eluted at a flow rate of 0.5 ml/min.

References

1. Baker NA, Sept D, Joseph S, Holst MJ, & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98(18):10037-10041.
2. Dolinsky TJ, Nielsen JE, McCammon JA, & Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research* 32(Web Server issue):W665-667.
3. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
4. Das RK & Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America* 110(33):13392-13397.
5. Vitalis A & Pappu RV (2009) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of computational chemistry* 30(5):673-699.
6. Vitalis A & Pappu RV (2009) Methods for Monte Carlo simulations of biomacromolecules. *Annual reports in computational chemistry* 5:49-76.
7. Mao AH, Crick SL, Vitalis A, Chicoine CL, & Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* 107(18):8183-8188.
8. Kaminski GA, Friesner RA, Tirado-Rives J, & Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†. *The Journal of Physical Chemistry B* 105(28):6474-6487.
9. Mao AH & Pappu RV (2012) Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions. *The Journal of chemical physics* 137(6):064104.
10. Goh GB, Hulbert BS, Zhou H, & Brooks CL (2014) Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Proteins: Structure, Function, and Bioinformatics* 82(7):1319-1331.
11. Radhakrishnan A, Vitalis A, Mao AH, Steffen AT, & Pappu RV (2012) Improved atomistic Monte Carlo simulations demonstrate that poly-L-proline adopts heterogeneous

ensembles of conformations of semi-rigid segments interrupted by kinks. *The journal of physical chemistry. B* 116(23):6862-6871.
12. Flory PJ (1953) *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, NY).
13. Flory PJ (1969) *Statistical Mechanics of Chain Molecules* (Oxford University Press, New York).
14. Sugita Y & Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314(1-2):141-151.
15. Das RK & Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* 110(33):13392-13397.
16. 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.
17. Lyle N, Das RK, & Pappu RV (2013) A quantitative measure for protein conformational heterogeneity. *J. Chem. Phys.* 139(12):121907.
18. Lacy ER*, et al.* (2004) p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nature structural & molecular biology* 11(4):358-364.
19. Grimmler M*, et al.* (2007) Cdk-inhibitory activity and stability of p27(Kip1) are directly regulated by oncogenic tyrosine kinases. *Cell* 128(2):269-280.
20. Neidhardt FC, Bloch PL, & Smith DF (1974) Culture medium for enterobacteria. *Journal of bacteriology* 119(3):736-747.
21. Cai M*, et al.* (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in Escherichia coli. *J Biomol Nmr* 11(1):97-102.
22. Dyer KN*, et al.* (2014) High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods in molecular biology* 1091:245-258.
23. Hura GL*, et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature methods* 6(8):606-612.
24. Putnam CD, Hammel M, Hura GL, & Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics* 40(3):191-285.
25. Kjaergaard M & Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol Nmr* 50(2):157-165.
26. Lawrence CW & Showalter SA (2012) Carbon-Detected N-15 NMR Spin Relaxation of an Intrinsically Disordered Protein: FCP1 Dynamics Unbound and in Complex with RAP74. *J Phys Chem Lett* 3(10):1409-1413.
27. Delaglio F*, et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol Nmr* 6(3):277-293.