Published in final edited form as:

J Am Chem Soc. 2016 November 30; 138(47): 15323–15335. doi:10.1021/jacs.6b10272.

Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation

Erik W. Martin[†,§], **Alex S. Holehouse**[‡,§], **Christy R. Grace**[†], **Alex Hughes**[†], **Rohit V. Pappu**[‡,*], and **Tanja Mittag**[†,*]

[†]Department of Structural Biology, St. Jude Children's Research Hospital, 263 Danny Thomas Place, Memphis, TN 38105, USA

[‡]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, One Brookings Drive, Campus Box 1097, St. Louis, MO 63130, USA

Abstract

Many cell signaling events are coordinated by intrinsically disordered protein regions (IDRs) that undergo multisite Serine/Threonine phosphorylation. The conformational properties of these IDRs prior to and following multi-site phosphorylation are directly relevant to understanding their functions. Here, we present results from biophysical studies and molecular simulations that quantify the conformational properties of an 81-residue IDR from the *S. cerevisiae* transcription factor Ash1. We show that the unphosphorylated Ash1 IDR adopts coil-like conformations that are expanded and well-solvated. This result contradicts inferences regarding global compaction that are derived from heuristics based on amino acid compositions for IDRs with low proline contents. Upon phosphorylation at ten distinct sites, the global conformational properties of pAsh1 are indistinguishable from those of unphosphorylated Ash1. This insensitivity derives from compensatory changes to the pattern of local and long-range intra-chain contacts. We show that the conformational properties of Ash1 and pAsh1 can be explained in terms of the linear sequence patterning of proline and charged residues vis-à-vis all other residues. The sequence features of the Ash1 IDR are shared by many other IDRs that undergo multisite phosphorylation. Accordingly, we propose that our findings might be generalizable to other IDRs involved in cell signaling.

TOC image

[*]**Corresponding Authors:** pappu@wustl.edu (R.V.P.), tania.mittag@stiude.org (T.M.).

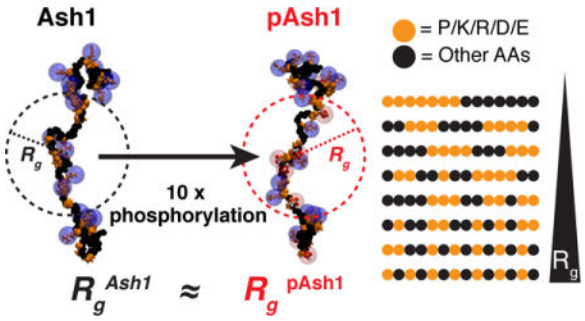
[§]**Author Contributions:** These authors contributed equally. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Supporting Information. Supporting Methods, Supplementary Tables S1 –S5, Supplementary Figures S1 – S17. Additional supplemental data are provided to quantify the quality and methods of SAXS data fitting, NMR chemical shift assignments for individual peptides, *cis*/*trans* proline equilibrium and the polyelectrolyte model. This material is available free of charge via the Internet at http://pubs.acs.org.

Database deposition. Ash1 and pAsh1 NMR chemical shifts were deposited in the BMRB data bank with identifiers 26719 and 26720.

Notes

The authors declare no competing financial interest.



INTRODUCTION

The downstream responses of cells to different cues are often controlled by signals that are initiated by post-translational modifications. These include multisite Ser/Thr/Tyr phosphorylation within intrinsically disordered regions (IDRs) of specific proteins.[1] Multisite phosphorylation is dynamic[2] and it enables rapid signal integration[3–4]. Many nonlinear downstream responses[5–10] such as transcriptional regulation[11–13], cell cycle control[4–5, 14–15], and cell proliferation[16] are coordinated by multisite phosphorylation of IDRs. Archetypal IDRs that undergo multisite phosphorylation include the C-terminal domain of RNA polymerase II[13], the C-terminal tail of the epidermal growth factor receptor[13], and sidearms of intermediate filaments in neurons[17].

Sites of phosphorylation are located within short linear motifs (SLiMs)[18] and these motifs are the substrates for kinases and phosphatases that catalyze site-specific phosphorylation and dephosphorylation, respectively. The accessibilities of substrate motifs to kinases (writers), downstream binding partners (readers), and phosphatases (erasers) are governed by sequence-encoded local and global conformational properties of IDRs prior to and following multisite phosphorylation. The overall fraction of charged residues (FCR) increases upon multisite phosphorylation. If the net charge per residue (NCPR) prior to phosphorylation is close to zero, then multisite phosphorylation will induce a polyampholyte to polyelectrolyte transition. Conversely, if the NPCR is larger than zero, then multisite phosphorylation will induce a transition from a polyelectrolyte to a polyampholyte, and the sequence patterning of oppositely charged residues[19] is expected to become an important determinant of conformational properties. Therefore, multisite phosphorylation has the potential to induce significant changes to the conformational properties of IDRs.

The sequence-encoded balance between protein-solvent and intra-protein interactions determines the conformational properties of IDRs.[20–21] Recent studies have combined results from all atom simulations and in vitro experiments to uncover how sequence encodes the balance between intra-chain and chain-solvent interactions. These findings support a grouping of IDRs into distinct conformational classes based on their amino acid compositions and the sequence patterning of oppositely charged residues.[19] This classification applies to sequences of IDRs that are deficient in hydrophobic and proline residues and are enriched in polar and/or charged residues. It is noteworthy that SLiMs encompassing phosphosites often include proline residues and proline-directed kinases regulate a larger number of proteins than non-proline directed kinases.[22–23] Consequently,

many IDRs that undergo multisite phosphorylation will include a moderately high fraction of proline residues. Proline is unique in being an imino acid, giving it distinct structural properties when compared to the amino acids. It disrupts the propagation of regular secondary structural elements, helps nucleate alpha helices[24], promotes turn formation[25], engenders local stiffening of the backbone[26], encodes a distinct preference for locally expanded polyproline II conformations when the peptide bond is in the *trans* configuration[27], engenders a preference for positive backbone ψ-angles for residues directly N-terminal to it, and can promote global compaction via *trans* to *cis* isomerization.[28–29]

Here, we go beyond previous descriptions of composition-to-conformation relationships for IDRs[20–21] to investigate the interplay amongst proline, charged, and post-translationally modifiable Ser/Thr residues as determinants of conformational properties of an archetypal IDR prior to and following multisite phosphorylation. The IDR of interest is derived from the protein Ash1, which is a transcription factor that regulates mating type switching in *Saccharomyces cerevisiae*.[30] In Ash1 a C-terminal zinc finger domain binds DNA while the remainder of the sequence is predicted to be disordered. It contains 23 Ser/Thr phosphosites that are part of distinct proline-containing SLiMs. These Ser/Thr residues are phosphorylated by the cyclin-dependent kinases Cln1,2/Cdc28.[31] Ash1$^{420–500}$, which is the object of our study, is an 81-residue section of the IDR. It encompasses ten Ser/Thr residues within phosphosites, sixteen arginine and lysine residues, and one aspartate residue (Figure 1). The FCR and NCPR of Ash1$^{420–500}$ prior to phosphorylation are 0.2 and +0.18, respectively. Stoichiometric multisite phosphorylation should change the NCPR to +0.06 while increasing the FCR to 0.35. This change converts the sequence of Ash1$^{420–500}$ from a weak polyelectrolyte to a well-mixed strong polyampholyte. Accordingly, heuristics that do not account for the contribution of proline residues suggest that multisite phosphorylation would convert Ash1$^{420–500}$ from a globule to a swollen, well-solvated coil (Figure S1). If this is valid, then there should be a substantial increase in the radius of gyration (R_g)[19, 32] upon multisite phosphorylation.

We quantified the conformational properties of Ash1$^{420–500}$ using small angle x-ray scattering (SAXS), nuclear magnetic resonance (NMR) spectroscopy, and all atom simulations. These studies reveal that unphosphorylated Ash1$^{420–500}$ adopts expanded coil-like conformations in aqueous solutions. These conformational preferences persist upon stoichiometric as well as sub-stoichiometric multisite phosphorylation. We identified sequence features within Ash1$^{420–500}$ that determine its intrinsic conformational properties. Specifically, we show that the apparent insensitivity of global dimensions upon phosphorylation derives from compensatory conformational changes along the sequence of multiply phosphorylated Ash1$^{420–500}$.

RESULTS AND DISCUSSION

Ash1$^{420–500}$ populates an expanded ensemble of conformations

SAXS data for unphosphorylated Ash1$^{420–500}$ (referred to hereafter as Ash1) were collected to probe the global conformational preferences of this IDR. Data were recorded at the SIBYLS beamline and the results were independently verified at the Advanced Photon Source at Argonne National Lab, where the data were collected immediately after samples

were processed using size-exclusion chromatography (SEC) (see Figure 2a for representative data collected in the presence of 150 mM NaCl). The features of the normalized Kratky plots are consistent with an expanded, coil-like ensemble for Ash1 (Figure 2b). The Guinier regions of the data did not show any indication of aggregation or intermolecular interactions at the concentrations used for SAXS measurements (Figure 2c).

A linear fit of the Guinier transformation yields an estimate of the ensemble averaged radius of gyration (R_g). For IDPs, the q-region available for a Guinier analysis is typically smaller than for folded proteins and this was optimized for each sample. We typically analyzed q-regions with $q \times R_g < 1$, in agreement with other reports.[33–34] For SAXS data collected immediately after processing by SEC, Guinier analysis yielded an R_g estimate of 28.5 ± 3.4 Å for Ash1 in aqueous solutions with 150 mM NaCl (complete data as a function of NaCl concentration available in Table S1). As a reference, a compact globule with the same number of residues would have an R_g of ~13 Å.

To further analyze the SAXS data, we used the ensemble optimization method (EOM;[35–36] details in the Supporting Information) to generate distributions of radii of gyration that are compatible with the SAXS data for Ash1. EOM is based on a genetic algorithm whereby a distribution of R_g values is chosen from a randomly generated pool of conformations to ensure that the linear combination of the SAXS profiles of all conformations in the ensemble regenerates the experimental data. Ensembles comprising of 20–30 conformations were typically needed to fit the measured SAXS data (Figure 3a and S2).[36] Additionally, the R_g distribution for Ash1 is shifted to larger sizes with respect to a random starting pool (Figure 3b).

Global dimensions of phosphorylated states of Ash1 are similar to those of unphosphorylated Ash1

The 10-fold phosphorylated version of Ash1 (referred to hereafter as pAsh1) was generated via overnight incubation of Ash1 with Cyclin A/Cdk2 (Figure S3). Analyses of pAsh1 were performed identically to Ash1. In aqueous solutions with 150 mM NaCl, Guinier analysis of SAXS data for pAsh1 yields a mean R_g value of 27.5 ± 1.2 Å. Within error, this value is similar to that of the unphosphorylated Ash1. The EOM analysis yielded R_g distributions for pAsh1 that were similar to those of Ash1 (Figure 3b). These results are surprising given the substantial increase in FCR and reduction in NCPR between Ash1 and pAsh1. To assess the robustness of the invariance of global conformational properties to phosphorylation, we generated an Ash1 mutant with only five intact phosphosites, while the Ser/Thr residues in other phosphosites were mutated to alanine. This construct is referred to as 5pAsh1. Additionally, in an alternative approach, we limited ATP in phosphorylation reactions to generate sub-stoichiometric phosphorylated variants of Ash1 while keeping its sequence intact. In all cases, the average global dimensions and R_g distributions were similar to those of Ash1 and pAsh1 (Figure 4a), indicating a robustness of the invariance of global dimensions to stoichiometric or sub-stoichiometric phosphorylation.

Preferences for expanded conformations are insensitive to screening of long-range electrostatic interactions

We reasoned that the net positive charge of Ash1 might engender intra-chain electrostatic repulsions leading to chain expansion. This would be true of archetypal polyelectrolytes. Accordingly, the addition of salt should induce a statistically significant chain compaction through the screening of electrostatic repulsions (see discussion and analysis in the Supplementary Information). To test for this possibility we collected SAXS data for Ash1 over NaCl concentrations ranging from 75 mM to 1500 mM. The overall dimensions, quantified in terms of EOM-generated.

R_g distributions, were essentially insensitive to changes in salt concentration (Figure 4b and Table S1). The broadening of the R_g distributions at higher salt concentration is likely due to increasingly poor data contrast in SAXS measurements. We find a similar weak sensitivity of R_g distributions to changes in salt concentration for pAsh1 (Figure 4c and Table S2). This is true despite a significant increase in the overall charge content (Figure 4c and Table S2). Taken together, these results suggest that in Ash1 and pAsh1 long-range electrostatic interactions are not the main drivers of chain expansion.

Preferential interactions with denaturants engender further expansion of Ash1 and pAsh1 ensembles, as evidenced by modest increases of R_g values in the presence of 4 M guanidinium hydrochloride (Figures 4b and 4c). This suggests the presence of weak local structural preferences within the Ash1 ensemble that are lost upon chemical denaturation. Overall, the Ash1 ensembles show a clear preference for coil-like properties and this is true irrespective of the presence or absence of denaturants.

NMR data show changes in local chemical environments upon phosphorylation

We used NMR spectroscopy to perform comparative assessments of site-specific conformational preferences of Ash1 and pAsh1. Figure 5a shows ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra for Ash1 and pAsh1. The HSQC spectra show poor chemical shift dispersions and sharp line widths for both sequences. This is consistent with averaging of the magnetic environment via interconversion amongst different conformations. Phosphorylation resulted in a downfield shift of ^{1}H resonances and upfield shift of ^{15}N resonances, especially for phosphorylated residues. This is consistent with the presence of phosphoryl-amide hydrogen bonds.[37] Less pronounced shifts were observed for residues that are proximal in the linear sequence (Figure 5a and Figure S4). Overall, the NMR data are consistent with specific conformational changes that are localized to phosphorylated residues that accompany multisite phosphorylation of Ash1.

The increase in FCR upon phosphorylation would be expected to promote expansion of pAsh1 when compared to Ash1. The invariance of global conformational properties to multisite phosphorylation suggests the possibility of a compensatory expansion and compaction within the ensemble. Given the possibility of *cis*/*trans* proline isomerization we asked if an increase in the population of *cis* proline isomers could counter the effects of multisite phosphorylation and explain the invariance of global dimensions.

Due to the repetitive nature of the Ash1 sequence and limited chemical shift dispersion, the proline resonances in ^{3}H-^{13}C HSQC spectra are degenerate, affording a global comparison of *cis* and *trans* proline populations (Figure 5b). When averaged over all proline residues in the protein, the *cis* populations in Ash1 and pAsh1 are highly similar (Table S4). Although the carbon-detect CON spectra showed chemical shift differences between Ash1 and pAsh1, the numbers and intensities of minor proline signals were qualitatively similar (Figure 5c). The low populations of *cis* proline isomers and the relatively low sensitivity of carbon-detect experiments did not allow for the assignment of all minor signals, which are a mixture of *cis* and *trans* proline resonances that are shifted by the sequence proximity of *cis* proline residues (Figure S5). Measurements directed at shorter peptides, which recapitulate the sequence-local effects on *cis*/*trans* proline isomerization, confirmed the insensitivity of *cis* proline contents to phosphorylation for each proline within individual phosphosites (Figures S6 and S7, Tables S3 and S4). Our data therefore suggest that the population of *cis* proline isomers is essentially insensitive to phosphorylation in Ash1. Therefore, changes to the overall content of *cis* proline isomers do not appear to provide compensatory compaction to offset the expected expansion from the increased FCR upon multisite phosphorylation.

We next asked if multisite phosphorylation leads to a set of new interactions that were not attainable in the unphosphorylated state, and if these interactions might afford compensatory compaction, assuming an expansion associated with an increase in FCR? One candidate for this type of interaction would be local pSer/Arg salt bridges, which are proposed to lead to compaction within shorter IDRs.[38] We do measure evidence for such local salt bridges in model peptides as indicated by changes in arginine side chain chemical shifts upon phosphorylation (Figure S8). However, we do not observe chemical shift differences of Arg sidechains between Ash1 and pAsh1. This suggests that if pSer/Arg salt bridges are present, they are less persistent than in short peptides and cannot provide compensatory compacting effects to offset any expansion derived from the increased FCR due to multisite phosphorylation.

Our results thus far suggest that while phosphorylation has no impact on the global dimensions of Ash1, it does lead to quantifiable local changes, especially in the chemical environments of phosphorylated residues. Our data argue against chain compaction upon multisite phosphorylation due to proline isomerization or persistent pSer/Arg or pThr/Arg salt bridges. The other alternative is that the degeneracy of local/non-local interactions along the chain of a long disordered protein might compete with and compensate the effects of one another. This type of intrachain screening of attractive and repulsive interactions, proposed by Flory, can lead to invariance of global dimensions even after multisite phosphorylation. Such effects are difficult to discern experimentally. Accordingly, we turned to all atom simulations to explore and understand the synergy between sequence-encoded global and local conformational preferences.

All atom simulations reproduce the experimentally observed, sequence-encoded expansion of Ash1

We performed all atom Metropolis Monte Carlo simulations using the ABSINTH implicit solvation model and forcefield paradigm that we combined with parameters from the OPLS-

AA/L molecular mechanics forcefield. These simulations were aided by the development of optimized parameters for proline residues that were made interoperable with the ABSINTH model and OPLS-AA/L forcefield.[39] Solution ions and all polypeptide atoms are modeled explicitly. Parameters for solution ions are interoperable with any solvation paradigm, including ABSINTH[40]. The explicit modeling of solution ions allows us to query the effects of changes to salt concentration on conformational properties.

In the presence of 50 mM NaCl, the ensembles generated by the all atom simulations yield a mean R_g value of 28.9 ± 1.2 Å (Figure 6a). We obtained similar R_g values from simulations in the presence of 150 mM NaCl. Within experimental error, these values are in agreement with inferences from the SAXS data for Ash1. In order to calibrate the pattern of intra-chain distances in simulation results we generated ensembles to reproduce two theoretical reference limits. These are designated as the Flory Random Coil (FRC) and Excluded Volume (EV) ensembles.[19] The mean R_g scales with chain length (N) as $N^{0.5}$ and $N^{0.59}$ for FRC and EV ensembles, respectively.[19] We have implemented a method to generate sequence-specific ensembles that conform to the FRC and EV limits.[41] Using this approach, we calculated the mean R_g values for Ash1 in the FRC and EV limits to be 23.4 ± 1.6 Å and 33.4 ± 1.8 Å, respectively. The mean R_g value and the distribution of R_g values calculated from the ABSINTH ensembles lie in between the FRC and EV limits (Figures 6a).

We also performed ABSINTH-based all atom simulations on a phosphomimetic version of Ash1, which we refer to as eAsh1. In these simulations, every phosphorylated Ser/Thr residue of Ash1 was replaced with Glu. The mean R_g value of eAsh1 in the presence of 50 mM NaCl was 27 ± 1.2 Å. The mean R_g values for Ash1 and eAsh1 from simulations are within error of one another, and within error of the R_g values determined by SAXS for Ash1 and pAsh1, respectively. The chemical structure of and local conformational properties engendered by Glu and pSer/pThr are distinct from one another. Despite these differences, the agreement between pAsh1 SAXS results and eAsh1 simulation results suggests that the global conformational properties of pAsh1 might be governed by generic features captured by the phosphomimetic eAsh1. We obtained similar R_g values from simulations of eAsh1 in the presence of 150 mM NaCl – an observation that is consistent with the negligible salt dependence observed from SAXS measurements for pAsh1.

We used the ABSINTH, FRC and EV ensembles to calculate internal scaling profiles. These profiles quantify the mean spatial separation between all pairs of residues that are |j–i| residues apart along the linear sequence[41] (Figure 6b). They are formal order parameters in polymer physics theories and are useful for quantifying the intramolecular density of chain atoms around one another and for making quantitative comparisons across different ensembles. A monotonic increase of the spatial separation with sequence separation is shown by the black and red curves in Figure 6b thus confirming the expanded, coil-like nature of the Ash1 and eAsh1 ensembles. The results in Figures 6a and 6b demonstrate that Ash1 and eAsh1 sample globally similar ensembles that lie between the FRC and EV limits. In addition, simulations of partial phosphomimetic constructs match SAXS results for partially phosphorylated Ash1 (Figure S9). Overall, the ABSINTH simulations recapitulate the general insensitivity of global dimensions to changes in the charge states of Ash1.

In order to place the comparison between ABSINTH ensembles and the scattering data on a quantitative footing, we used the ABSINTH all atom ensembles to calculate scattering curves using the CRYSOL package for Ash1. The results of the comparisons are shown in Figure 7a (see black curve) and Figure 7b. We also calculated scattering curves for eAsh1, and compared those results to the scattering curve from pAsh1 (see red curve in Figure 7a). In the interest of completeness we also calculated the scattering curves obtained using the Flexible Meccano model (see green curve in figure 7b).[42] The favorable comparisons between the Flexible Meccano and ABSINTH derived scattering curves as well as between the ABSINTH and experimental data suggest that, on a global scale, the ensembles of Ash1 and eAsh1 resemble that of an expanded random coil whose mean size lies in between two well-defined theoretical limits.

In order to compare the R_g distributions obtained from ABSINTH ensembles and those obtained from the EOM approach, we calculated the degree of overlap between the distributions (see Supporting Information for details, Figures S10 and S11). Ash1 (SAXS) and Ash1 (simulation) showed a high degree of overlap (~0.85, see Figure 7c), highlighting the congruence between simulated ensembles and distributions obtained using models that are designed to match the experimental data. Similarly, the overlap between the EOM and ABSINTH-derived R_g distributions for pAsh1 and eAsh1 is ~0.88 (see Figure 7d) indicating that the global conformational preferences measured by SAXS for pAsh1 are similar to those obtained from ABSINTH-based simulations of eAsh1. Favorable comparison between the measured scattering curve of pAsh1 and the calculated scattering curve of eAsh1 suggests that the phosphomimetic sequence captures the global conformational preferences of pAsh1. Accordingly, a detailed analysis of these ensembles should provide an explanation for the observed coil-like conformations of Ash1 and the invariance of global conformational properties to multisite phosphorylation.

Sequence determinants of Ash1 expansion

We examined sequence features of Ash1 to uncover the source of the intrinsic, sequence-encoded expansion. Ash1 has a proline content of 15%. This is relevant because published heuristics regarding composition-to-conformation relationships of IDRs were derived from simulation results and spectroscopic investigations of sequences with low proline contents.[20] In Ash1, 35% of the residues are either proline or charged. Since proline and charged residues respectively drive local and global expansion, we reasoned that the linear sequence distribution of proline and charged residues might explain the observed expansion of Ash1.

In Ash1, the proline and charged residues are uniformly distributed with respect to all other residue types along the linear sequence (Figure 1). We quantified this as the mixing or segregation of proline and charged residues (Pro, Lys, Arg, Asp, Glu) vis-à-vis all other residues (Xaa). Specifically, we computed a normalized patterning parameter designated as Ω where $0 \leq \Omega \leq 1$. Our definition of Ω is analogous to the definition of the κ parameter introduced by Das and Pappu[19] to quantify the mixing vs. segregation of oppositely charged residues. The calculation of Ω is described in the Supplementary Information. If proline and charged residues are well mixed with respect to all other residues, then the value of Ω for the sequence of interest approaches zero. Conversely, if proline and charged residues are

segregated with respect to all other residues in the sequence of interest, then Ω approaches unity. We find that Ω = 0.l for Ash1 and 0.13 for pAsh1/eAsh1. Therefore, we hypothesized that the uniform distribution of expansion-driving proline and charged residues along the Ash1/pAsh1/eAsh1 sequences give rise to a sequence-encoded preference for expanded conformations. To test this hypothesis, we used an unbiased sequence design algorithm to design a series of sequence permutants of Ash1. These permutants – all of which have an identical amino acid composition – were generated by shuffling the positions of proline residues (red symbols in Figure 8), charged residues (cyan), or both (dark blue). Using this approach we generated sequences corresponding to different values of Ω. The complete set of sequences can be found in Figure S13. We performed multiple independent atomistic simulations for each of the Ω-permutants. Figure 8 shows the calculated R_g for each Ω-permutant plotted against Ω. This analysis shows a strong negative correlation (Pearson's correlation coefficient = −0.81) between the degree of expansion and Ω (Figure 8), suggesting that the mixing or segregation of proline and charged residues vis-à-vis other residues engenders expansion versus compaction, respectively. This analysis provides a plausible explanation for the sequence-encoded preference for expanded Ash1 ensembles in aqueous solvents.

The invariance of global conformational properties upon multisite phosphorylation derives from compensatory conformational changes

We quantified secondary structure propensities by comparing NMR derived C_α, C_β, C' and N chemical shifts with random coil values. We used three different methods (SSP[43], ncSCP[44], and δ2D[45]) to convert the measured chemical shifts to estimates of local structural propensities. Although there is reasonable agreement among the estimates obtained using the three methods, there is also considerable variation suggesting the need for caution in extracting precise quantitative trends from the experimental data. We also used the simulated ensembles for Ash1 and eAsh1 to calculate local structural propensities. In order to avoid comparisons among metadata, we compared the results from our analysis of experimental data to analysis of simulation results that are based on backbone ϕ/ψ angles as implemented in the BBSEG algorithm that is part of the CAMPARI modeling suite (Figure 9). Overall, the local structural propensities calculated from simulations agree with the consensus interpretation that emerges from analysis of experimental data. All four methods point to an increase in alpha-helical propensities upon phosphorylation. This increase in helicity is around residue 430–435 and residue 470–480.

We also examined the propensities for polyproline II (PPII) conformations (Figure S13). Although there are modest changes in PPII propensities upon phosphorylation, both ensembles appeared to have a relatively high PPII propensity across the entire sequence. Analysis of the simulation results using BBSEG and of the experimental data using δ2D allow us to evaluate the PPII propensity for each residue. We also evaluated the simulated ensembles for persistent PPII preferences across consecutive stretches along the linear sequence. This analysis suggests that while individual residues have distinct preferences for the PPII basin of Ramachandran space, these local preferences derive from uncorrelated transitions into and out of the PPII basin. Accordingly, the expansion of Ash1 cannot be attributed to persistent preference for PPII helices, which require that at least three

consecutive residues simultaneously occupy the PPII basin. Instead, the overall expansion of Ash1 can be attributed to the synergistic combination of proline and charge contents, the uniform mixing of these residues, the local stiffening due to proline residues, and the favorable solvation of charged residues. Furthermore, we found no correlation between global PPII occupancy and the R_g values for the series of Ω–permutants of Ash1 (Figure S14), suggesting that sequences with similar PPII propensities can have very different R_g values.

We next asked if the invariance of global conformational properties between Ash1 and pAsh1/eAsh1 might derive from compensatory changes in the patterns of preferred intramolecular distances. Using the ABSINTH-based ensembles for Ash1 and eAsh1 we calculated the ensemble-averaged distances between the centers-of-mass of every unique pair of residues in the sequence. Figure 10a shows the raw data with the upper triangular portion corresponding to Ash1 and the lower triangular portion corresponding to eAsh1. Given the coil-like nature and the wide range of inter-residue distances within the ensembles for both sequences, it is difficult to uncover the important distinctions between the two ensembles. This is remedied by calculating normalized distances, whereby the distance for every pair of residues is normalized by the value we obtain for the sequences in the EV limit. These two-dimensional scaling maps are shown in Figure 10b. The scaling maps reveal the following insights: Residues 455–460 in eAsh1 make long-range contacts with spatial separations in the range of 25 – 35 Å with residues 474–490. Residues 455–460 contain four phosphosites and two Arg residues, whereas the region spanning 474–490 contains eight Lys/Arg residues and one phosphosite. This suggests the presence of non-local, intermediate-range electrostatic interactions between a cluster of positively charged residues near the C-terminal region and a cluster of negatively charged residues in the central region. These complementary, non-local electrostatic interactions engender a modest compaction vis-à-vis the EV limit that is not observed in the Ash1 ensemble. However, the effects of compaction are offset by expansion vis-à-vis the EV limit across the region spanning residues 450–470. An explicit example of the local compensatory changes is shown in Figure 10c, where the dimensions of two sub-peptides examined in the context of Ash1 and eAsh1 are compared. We attribute this expansion to enhanced electrostatic repulsions and the favorable free energy of solvation of the negatively charged residues within this region. Finally, we observe a modest local compaction and longer-range expansion for the region spanning residues 435–440, which is attributable to the increased alpha helix propensity upon phosphorylation (Figures 9 and 10b).

Importantly, the changes observed within the eAsh1 ensemble are mutually compensatory. This is an example of intra-chain screening that is a central tenet of Flory's theory for realizing unperturbed global dimensions.[46] Repulsive interactions that lead to local/non-local chain expansion are screened by the effects of attractive interactions that lead to local/non-local chain contraction. Since these compensatory interactions involve partially overlapping regions of the sequence, the intra-chain screening leads to unperturbed chain dimensions vis-à-vis the unphosphorylated ensemble. Additionally, the negligible salt dependence of the phosphorylated ensemble is explained by weak screening provided by solution ions when compared to the screening of repulsive interactions by attractive ones

that are encoded by the sequence, which also controls the effects of post-translational modifications.

Compensatory changes in conformational dynamics are amenable to scrutiny via NMR relaxation methods. Figure 11 shows a comparative analysis of the spin-lattice relaxation rates (R_1) and the spin-spin relaxation rates (R_2) for Ash1 versus pAsh1. While R_1 rates and heteronuclear NOE values (Figure S15) are similar for both Ash1 and pAsh1, there are discernible jumps in R_2 rates in clusters along the pAsh1 sequence.[47] These enhanced R2 rates are indicative of a slowdown in local dynamics upon phosphorylation caused by transient interactions, in agreement with the proposed model of competing local/nonlocal interactions. Specifically, enhanced R_2 rates in the central region (~450–460) and less pronounced clusters toward the C-terminus are consistent with the main regions identified from analysis of simulation results as being involved in long-range electrostatic interactions upon multisite phosphorylation.

Summary

Overall, the simulation results provide a nuanced description of how multisite phosphorylation might influence the conformational properties of Ash1. The effects of local/non-local expansion and compaction involving partially overlapping sequence regions leads to unperturbed global dimensions vis-à-vis the unphosphorylated Ash1. This Flory-like screening of intra-chain attractions by repulsions is encoded by the amino acid sequence of Ash1, which controls the overall conformational properties prior to and upon multisite phosphorylation. With regard to the latter, it is worth noting that the patterning of proline and charged residues with respect to all other residues changes only slightly upon multisite phosphorylation. This is quantified in terms of the value of Ω, which changes from 0.1 to 0.13 (Figure 7a) implying a uniform dispersion of proline and charged residues along both sequences.

Our NMR data do not directly report on weak, transient compensatory local/non-local interactions, which seem to be the driving forces of the expanded global dimensions of Ash1 and pAsh1. However, the lack of observable stable structural motifs such as persistent salt bridges, and the highly averaged chemical shifts are consistent with transient, competing interactions in Ash1 and pAsh1. In long IDRs, the balance of local/non-local interactions strongly depends on their patterning along the sequence and this determines whether interactions spanning distinct spatial scales reinforce or compete with each other. In shorter peptides, the competition from truly long-range interactions is absent. Hence, the effect of local interactions on the global conformational properties will be more direct. Although the 81 residue stretch is significantly larger than the 15–30 residue fragments often examined, even in Ash1 there remains a substantial sequence context that we have ignored. Therefore, understanding the hierarchical influence of sequence and structural contexts on the conformational properties of IDRs remains an open challenge.

Sequence features of Ash1 are shared by other IDRs that undergo multisite phosphorylation

We asked if the patterning of proline and charged residues vis-à-vis other residues is a feature that is shared by other proteins that undergo multisite phosphorylation. A conservative search through the human proteome for proline-rich regions that are predicted to be disordered and undergo multisite phosphorylation identified a number of putative candidate regions. A full list is included in the Supporting Information (Table S5).

The proline-rich region of the microtubule-associated protein tau shares many of the sequence features of Ash1; in a 90-residue stretch, it contains 13 phosphorylation sites and 22 proline residues. A recent Förster Resonance Energy Transfer study of a 14 residue peptide extracted from this region demonstrated an expansion upon phosphorylation.[48] Earlier studies showed that multisite phosphorylation causes local conformational changes, as determined by NMR,[49] while global dimensions measured by SAXS remain unperturbed in a phosphomimetic construct.[50] These observations – global insensitivity and local changes – are highly reminiscent of our results from Ash1. The *S. cerevisiae* cyclin-dependent kinase (CDK) inhibitor Sic1 undergoes multisite phosphorylation, triggering its degradation and subsequent cell cycle progression.[14, 51] The overall dimensions of the non-phosphorylated and phosphorylated states are highly similar as determined by SAXS and NMR, and yet there are extensive local conformational changes.[52–53]

A relatively uncharacterized protein, Chromosome alignment-maintaining phosphoprotein (CHAMP1), contains a 350 residue, proline-rich domain that undergoes extensive CDK1-dependent phosphorylation during mitosis. Given the results for Ash1, we may expect this proline-rich region to form a highly expanded ensemble, with the overall dimensions remaining unperturbed in response to varying degrees of phosphorylation.

We also found a number of disordered regions that undergo multisite phosphorylation that do not have the sequence characteristics of Ash1. These sequences are deficient in proline residues and they are weak/strong polyampholytes rather than polyelectrolytes. How these different regions respond to multisite phosphorylation will depend on their specific sequence contexts and the patterning of relevant residues therein. A key question is if all IDRs that undergo multisite phosphorylation will show a global conformational insensitivity to phosphorylation? Clearly, in some proteins, specific local/non-local interactions form efficiently, because of a lack of competing interactions along the chain. As an example, the protein 4E-BP2, which regulates the initiation of cap-dependent mRNA translation, folds into a stable structure upon multisite phosphorylation that is able to form a complex with its binding partner eIF4E.[54] In this case, multisite phosphorylation generates synergistic, long-range conformational changes. These must be encoded in the sequence as well, albeit by different sequence features.

CONCLUSION

Our findings for Ash1 and pAsh1 lead to the proposal of a synergistic relationship between proline and charged amino acids that results in expanded conformations of a disordered protein that undergoes multisite phosphorylation, irrespective of its phosphorylation state.

Importantly, proline residues appear to offer a mode of local expansion that is independent of the charged residues – a property that may be desirable in regions that undergo reversible changes in local charge density mediated by phosphorylation. IDRs that undergo multisite phosphorylation may in general utilize such proline-based conformational buffering to provide access to modifying enzymes and downstream signaling effectors. Further work is needed to determine the connections between sequence encoded global and local conformational properties and the functional consequences for IDRs prior to and upon multisite phosphorylation.

EXPERIMENTAL METHODS

Protein expression and purification

His-tagged Ash1$^{420–500}$ and 5pAsh1$^{420–500}$ was expressed in an *E. coli* BL21 GOLD (DE3) strain (Agilent) in LB or M9 media for isotope-labeled samples. Expression was induced at OD_{600}=0.8 with 0.6 mM IPTG and cells were cultured at 20 °C for an additional 18 hours. His_6-Ash1$^{420–500}$ was purified from inclusion bodies (for a detailed protocol see the Supporting Information). The polyhistidine tag was cleaved with a TEV protease, which left the protein with two additional N-terminal residues, i.e., Gly-Ala. We refer to this protein construct as Ash1$^{420–500}$. Final protein samples were generated by size exclusion chromatography on a Superdex 75 column (GE Life Sciences) into the desired buffer. Purified proteins were concentrated using Millipore centrifugal concentrators with 3000 Da cutoff. Their purity, integrity and identity were analyzed by SDS PAGE gel (Figure S3), MALDI-TOF and LC-MS/MS. The concentration was assessed via absorbance at 280 nm (ε = 2980 $M^{-1}cm^{-1}$).

Protein phosphorylation

Phosphorylated samples were prepared by treatment of Ash1 with Cyclin A/Cdk2 (prepared according to Huang et. al[55]) at a kinase/Ash1$^{420–500}$ ratio of 1:100 in the presence of 50 fold excess of ATP and 2.5 mM $MgCl_2$ overnight at 30 °C. Sub-stochiometric ratios of ATP to Ash1 of 12.5 and 5 were used to generate Ash1 populations with distributions centered around 5 and 2 phosphorylated sites, respectively. The yield of the phosphorylation reaction was determined by ESI-TOF mass spectrometry (Figure S3).

SAXS sample preparation and data collection

Samples of Ash1$^{420–500}$ were prepared in a buffer containing 50 mM Tris pH 7.5, 10 mM DTT and 2 mM TCEP. High concentrations of Tris and DTT were used to scavenge radicals and prevent radiation damage. The addition of TCEP served to stabilize the buffer reduction potential over the course of shipping and waiting for measurement. Purified protein samples were concentrated to approximately 2 mM and were then diluted into buffers to achieve the desired NaCl concentrations.

SAXS data collection

Solution SAXS data were collected at both the 12-ID-B beamline at the Argonne National Laboratory Advanced Photon Source and through the mail-in program at the SIBYLS beamline at the Lawrence Berkeley National Laboratory Advanced Light Source. SAXS

data were acquired manually at APS, where protein samples were loaded, then gently refreshed with a syringe pump to prevent x-ray damage. A Pilatus 2M detector provided q-range coverage from 0.015 $Å^{-1}$ to 1.0 $Å^{-1}$. Wide-angle x-ray scattering data were acquired with a Pilatus 300k detector and had a q range of 0.93 – 2.9 $Å^{-1}$. Calibration of the q-range calibration was performed with a silver behenate sample. Protein samples were freshly prepared using size exclusion chromatography (GE Life Sciences, Superdex 75 10/300 GL) in a buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 5 mM DTT, and 2 mM TCEP. Elution fractions were loaded without further manipulation. Buffer collected 1 column volume after protein elution from the column was used to record buffer data before and after each protein sample. Twenty sequential images were collected with 1 sec exposure time per image with each detector. Data were inspected for anomalous exposures and mean buffer data were subtracted from sample data using the WAXS water peak at q ~ 1.9 $Å^{-1}$ as a subtraction control.

Samples were sent to SIBYLS in 96 well plates (VWR). A pipetting robot automatically exchanged samples. SAXS data were measured for samples at protein concentrations of 450, 225, and 112 μM for each NaCl concentration. Matched buffers were collected from centrifugal concentrator filtrate and were included in wells before and after each dilution series. Data were collected in a q-range of 0.012–0.324 $Å^{-1}$ using 0.5, 1, 2 and 5-second exposures. Buffer-subtracted data from each exposure time were manually assayed for high noise and radiation damage. Data were then merged into a single data set using the program PRIMUS.[56]

SAXS data analysis

Basic analysis including raw data plotting, Kratky transformations to determine flexibility and Guinier transformations to estimate R_g were performed with the program ScÅtter[57] or in-house written MATLAB (MathWorks) scripts. Care was taken to limit the Guinier region to very low q values suitable to a disordered protein system. The form factors of IDP ensembles will span the range between rods and spheres implying that the appropriate q-range maximum for Guinier analysis should lie between $q \times R_g = 0.7–1.4$.[58–59] The best region was chosen by minimizing deviations in the calculated R_g due to either the removal of points near the beam stop or inclusion of higher q points. Ensemble modeling of SAXS data was done using the Ensemble Optimization Method (EOM2.0) in the ATSAS software package[35] in which a genetic algorithm is used to select an ensemble of conformations from a randomly generated pool. Pools of Ash1 conformations used were alpha carbon traces created by EOM. Ensemble R_g distributions obtained for all salt concentrations were fit to a function describing the R_g distribution of a non-intersecting chain in three dimensions (for detailed information see SI).[60–61]

NMR data collection

NMR data were acquired on Bruker Avance 600 and 800 MHz spectrometers equipped with TCI triple-resonance cryogenic probes and pulsed-field gradient units. All samples were prepared in an NMR buffer consisting of phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), 10 mM DTT pH 6.95 and 10% D_2O at 5°C. For assignment, approximately 0.7 mM ^{15}N,^{13}C Ash1 sample was used to acquire standard

triple-resonance backbone assignment experiments and carbon-detect triple resonance experiments. Standard assignment experiments were based on sensitivity enhanced ^{1}H-^{15}N HSQC (8 scans, 2048 × 320 complex data points, with 12 ppm and 25 ppm as ^{1}H and ^{15}N sweep widths). Carbon detect experiments were based on (HA Start) CON-IPAP (16 scans, 1024 × 512 complex data points, with 18 ppm and 36 ppm as ^{13}C and ^{15}N sweep widths).[62]

These included a HNCACB and CBCA(CO)NH (8 scans, 1024 (^{1}H) × 32 (^{15}N) × 128 (^{13}C) complex data points, with 12 ppm, 22 ppm, and 70 ppm as ^{1}H, ^{15}N and ^{13}C sweep width, respectively), a HN(CA)CO (8 scans, 1024 (^{1}H) × 24 (^{15}N) × 64 (^{13}C) complex data points, with 10 ppm, 22 ppm, and 22 ppm as ^{1}H, ^{15}N and ^{13}C sweep widths), a HNCO (8 scans, 1024 (^{1}H) × 32 (^{15}N) × 75 (^{13}C) complex data points, with 10 ppm, 22 ppm, and 22 ppm as ^{1}H, ^{15}N and ^{13}C sweep widths), and a (H)N(COCA)NH allowed a “backbone NH walk” (8 scans, 1024 (^{1}H) × 20 (^{15}N F1) × 64 (^{15}N F2) complex data points, with 10 ppm, 22 ppm, and 22 ppm as ^{1}H, ^{15}N F1 and ^{15}N F2 sweep widths).[63–64] Additionally a 3D HNCA (1024 (^{1}H) × 64 (^{15}N) × 32 (^{13}C) complex data points and 12 ppm (^{1}H) × 22 ppm (^{15}N) × 32 (^{13}C) sweep width) and carbon detect experiments based on 2D CON and 3D (HA)CANCO (512 (^{1}H) × 64 (^{15}N) × 32 (^{13}C) complex data points and 12 ppm (^{1}H) × 20 ppm (^{15}N) × 32 (^{13}C) sweep width) were needed to resolve chemical shift degeneracy in pAsh1 samples.[65] Additionally, CCCON-IPAP, a 3D ^{13}C TOCSY, was used to distinguish trans versus cis proline (32 scans, 1024 (^{13}C) × 48 (^{15}N) × 160 (^{13}C) complex data points, with 18 ppm, 36 ppm and 72 ppm as ^{13}C, ^{15}N and ^{13}C (TOCSY) sweep widths).[66]

^{15}N NMR relaxation experiments acquired on a Bruker Avance 800 MHz spectrometer at 278 K using standard pulse programs (16 scans, 2048 (^{1}H) × 150 (^{15}N) complex data points). The longitudinal R_1 spin-lattice relaxation rates were measured using relaxation delays of 20, 50, 200, 500, 1000, 1500, 2000 and 3000 ms. Transverse R_2 spin-spin relaxation rates were measured using relaxation delays of 92.5, 185, 277.5, 370, 462.5, 555, 740 and 925 ms. Relaxation rates were determined by integrating peak amplitudes and fitting to a single exponential decay. Error values are determined via 95% confidence intervals calculated using the residuals and Jacobian matrix from the nonlinear fit.

Data were processed using BRUKER Topspin version 3.2, NMRPipe (v.7.9)[67] and analyzed using CARA (v.1.8.4)[68]. All spectra were referenced directly using DSS for the ^{1}H dimension, ^{13}C and ^{15}N frequencies were referenced indirectly. Secondary structural propensities were calculated using ^{13}C’, ^{13}Cα and ^{13}Cβ chemical shifts and the SSP[43] and ncSPC[44]. PPII propensities were calculated from the same pool of chemical shifts with the addition of N, and $^{1}H^{N}$ shifts using δ2D.[45]

All atom Monte Carlo simulations

All simulations were performed using the CAMPARI Monte Carlo modeling suite (http://campari.sourceforge.net). The simulations deploy the ABSINTH implicit solvent model and forcefield paradigm.[39] The protein atoms and mobile solution ions are modeled in atomistic detail, while the solvent is treated using a mean field (implicit) model. Move sets combine pivots, concerted rotations, sidechain rotations, mutual reorientations, translations of the mobile ions, and a series of moves that enable the efficient sampling of the conformational degrees of freedom coupled to proline ring systems.[69] Multiple independent simulations for

each construct were run. Accurate modeling of Ash1 conformational equilibria requires the use of suitably optimized parameters for proline residues.[69] Without these parameters, details such as *cis-to-trans* isomerization, proper prolyl ring puckering, and the accurate coupling amongst ring puckering, peptide bond isomerization, and backbone phi angles cannot be reproduced. The optimized parameters for proline residues are interoperable with the abs_3.2_opls.prm parameter file in CAMPARI. However, since parameters for phosphorylated residues are currently unavailable for this parameter set, we pursued the route of replacing phosphorylated Ser and Thr residues with Glu. This strategy allowed us to investigate the impact of altering the charge distribution upon Ser/Thr phosphorylation, but it should not be viewed as a perfect mimic of Ser/Thr phosphorylation. For additional details see the Supporting Information. Simulation analysis was performed using CTraj, MDTraj[70] and routines built into CAMPARI. Sequence analysis and permutant design was performed using CIDER and localCIDER.[71]

Aggregate scattering curves were calculated from simulated ensembles using the program CRYSOL in the ATSAS package.[72] Scattering intensities were calculated for individual PDB files and were combined and scaled using MATLAB. In order to generate a random pool of all-atom conformations with statistically validated backbone angles, we used Flexible Meccano[42] and modeled amino acid sidechains using SCCOMP[73]. This pool was used to generate a SAXS curve for comparison with experimental data and ABSINTH ensembles.

Proteome-wide bioinformatics screen for Ash1-like regions

The human proteome was obtained from UniProt[74], and sequences were annotated for consensus disorder predictions using the D^2P^2 database.[75] Specifically, only regions for which five or more predictors indicated disorder were designated as disordered for further analysis; this is a relatively stringent threshold. Putative phosphosite data for proteome-wide screening were taken from the ProteomeScout database parsed via the ProteomeScoutAPI[76], although only regions where UniProt annotation also showed multisite phosphorylation were included in this analysis. To limit the analysis to regions equivalent in size to Ash1, we focused on disordered segments equal to or less than 100 residues. While this provides a distinctly conservative estimate of phosphorylation, it ensures that the identified regions come from high-fidelity data. For the final data set see the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Dr. Philipp Selenko for helpful discussions and critical inputs. We thank anonymous reviewers for their constructive comments and suggestions that improved our analysis. The mass spectrometry analysis was performed by K. Kavdia in the St. Jude Children's Research Hospital Proteomics Facility, partially supported by Cancer Center Support Grant (P30CA021765) from the National Cancer Institute and ALSAC (American Lebanese Syrian Associated Charities). The U.S. Department of Energy, under Contract DE-AC02-06CH11357, supported the use of the 12-id-B beamline at the Advanced Photon Source, an Office of Science User Facility operated for the DoE Office of Science by Argonne National Laboratory. Part of this work was conducted at the Advanced Light

Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health proiect MINOS (R01GM105404). The center for high performance computing at Washington University provided computational resources for the simulations.

Funding Sources

Grants from the V Foundation Scholar Grant (to T.M.), the American Lebanese Syrian Associated Charities (to T.M.), the National Science Foundation (MCB-1121867, MCB-1614766 to R.V.P), and the National Institutes of Health (5RO1NS056114 to R.V.P) funded this work.

ABBREVIATIONS

FCR	Fraction of Charged Residues
FRC	Flory Random Coil
NCPR	Net Charge Per Residue
EV	Excluded Volume
SLiMs	Short Linear Motifs

References

1. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. Nucleic Acids Res. 2004; 32(3):1037–1049. [PubMed: 14960716]
2. Kanshin E, Bergeron-Sandoval LP, Isik SS, Thibault P, Michnick SW. Cell reports. 2015; 10(7): 1202–14. [PubMed: 25704821]
3. Cohen P. Trends in biochemical sciences. 1992; 17(10):408–13. [PubMed: 1333658]
4. Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pryciak PM. Cell. 2007; 128(3): 519–31. [PubMed: 17289571]
5. Deshaies RJ, Ferrell JE Jr. Cell. 2001; 107(7):819–22. [PubMed: 11779457]
6. Gunawardena J. P Natl Acad Sci USA. 2005; 102(41):14617–14622.
7. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS. P Natl Acad Sci USA. 2007; 104(23):9650–9655.
8. Malleshaiah MK, Shahrezaei V, Swain PS, Michnick SW. Nature. 2010; 465(7294):101–5. [PubMed: 20400943]
9. Suwanmajo T, Krishnan J. J R Soc Interface. 2013; 10(89) web.
10. Suwanmajo T, Krishnan J. J R Soc Interface. 2015; 12(107) web.
11. Holmberg CI, Tran SE, Eriksson JE, Sistonen L. Trends in biochemical sciences. 2002; 27(12): 619–27. [PubMed: 12468231]
12. Dahmus ME. The Journal of biological chemistry. 1996; 271(32):19009–12. [PubMed: 8759772]
13. Phatnani HP, Greenleaf AL. Genes & development. 2006; 20(21):2922–36. [PubMed: 17079683]
14. Verma R, Annan RS, Huddleston MJ, Carr SA, Reynard G, Deshaies RJ. Science. 1997; 278(5337):455–60. [PubMed: 9334303]
15. Koivomagi M, Ord M, Iofik A, Valk E, Venta R, Faustova I, Kivi R, Balog ERM, Rubin SM, Loog M. Nat Struct Mol Biol. 2013; 20(12):1415–1424. [PubMed: 24186061]
16. Li X, Jiang S, Tapping RI. Cytokine. 2010; 49(1):1–9. [PubMed: 19775907]
17. Pant HC, Veeranna, Grant P. Current topics in cellular regulation. 2000; 36:133–50. [PubMed: 10842750]
18. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. Chemical reviews. 2014; 114(13):6589–631. [PubMed: 24773235]

19. Das RK, Pappu RV. P Natl Acad Sci USA. 2013; 110(33):13392–13397.
20. Das RK, Ruff KM, Pappu RV. Curr Opin Struc Biol. 2015; 32:102–112.
21. Mao AH, Lyle N, Pappu RV. The Biochemical journal. 2013; 449(2):307–18. [PubMed: 23240611]
22. Ubersax JA, Ferrell JE. Nat Rev Mol Cell Bio. 2007; 8(8):665–665.
23. Lu KP, Liou YC, Zhou XZ. Trends Cell Biol. 2002; 12(4):164–72. [PubMed: 11978535]
24. Aurora R, Rose GD. Protein science: a publication of the Protein Society. 1998; 7(1):21–38. [PubMed: 9514257]
25. Rose GD, Gierasch LM, Smith JA. Advances in protein chemistry. 1985; 37:1–109. [PubMed: 2865874]
26. Creamer TP, Campbell MN. Advances in protein chemistry. 2002; 62:263–82. [PubMed: 12418106]
27. Reiersen H, Rees AR. Trends in biochemical sciences. 2001; 26(11):679–84. [PubMed: 11701327]
28. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A, Eaton WA. Proc Natl Acad Sci U S A. 2007; 104(48):18964–9. [PubMed: 18029448]
29. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. Proc Natl Acad Sci U S A. 2005; 102(8):2754–9. [PubMed: 15699337]
30. Cosma MP. Embo Rep. 2004; 5(10):953–957. [PubMed: 15459746]
31. Liu Q, Larsen B, Ricicova M, Orlicky S, Tekotte H, Tang X, Craig K, Quiring A, Le Bihan T, Hansen C, Sicheri F, Tyers M. Molecular and cellular biology. 2011; 31(3):584–98. [PubMed: 21098119]
32. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. P Natl Acad Sci USA. 2010; 107(18): 8183–8188.
33. Borgia A, Zheng W, Buholzer K, Borgia MB, Schuler A, Hofmann H, Soranno A, Nettels D, Gast K, Grishaev A, Best RB, Schuler B. J Am Chem Soc. 2016; 138(36):11714–11726. [PubMed: 27583570]
34. Receveur-Brechot V, Durand D. Curr Protein Pept Sci. 2012; 13(1):55–75. [PubMed: 22044150]
35. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. J Am Chem Soc. 2007; 129(17):5656–5664. [PubMed: 17411046]
36. Tria G, Mertens HDT, Kachala M, Svergun DI. Iucrj. 2015; 2:207–217. [PubMed: 25866658]
37. Theillet FX, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon MK, Kriwacki RW, Landrieu I, Lippens G, Selenko P. J Biomol Nmr. 2012; 54(3):217–36. [PubMed: 23011410]
38. Kumar P, Chimenti MS, Pemble H, Schonichen A, Thompson O, Jacobson MP, Wittmann T. J Biol Chem. 2012; 287(21):17050–64. [PubMed: 22467876]
39. Vitalis A, Pappu RV. J Comput Chem. 2009; 30(5):673–99. [PubMed: 18506808]
40. Mao AH, Pappu RV. The Journal of chemical physics. 2012; 137(6):064104. [PubMed: 22897252]
41. Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV. J Am Chem Soc. 2015; 137(8):2984–95. [PubMed: 25664638]
42. Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Bernado P, Charavay C, Blackledge M. Bioinformatics. 2012; 28(11):1463–70. [PubMed: 22613562]
43. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Protein Sci. 2006; 15(12):2795–804. [PubMed: 17088319]
44. Tamiola K, Mulder FA. Biochem Soc Trans. 2012; 40(5):1014–20. [PubMed: 22988857]
45. Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Biochemistry. 2012; 51(11):2224–31. [PubMed: 22360139]
46. Flory, PJ. Principles of Polymer Chemistry. Cornell University Press; Ithaca, NY: 1953.
47. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H. Science. 2002; 295(5560):1719–22. [PubMed: 11872841]
48. Chin AF, Toptygin D, Elam WA, Schrank TP, Hilser VJ. Biophys J. 2016; 110(2):362–71. [PubMed: 26789759]
49. Schwalbe M, Kadavath H, Biernat J, Ozenne V, Blackledge M, Mandelkow E, Zweckstetter M. Structure. 2015; 23(8):1448–1458. [PubMed: 26165593]

50. Mylonas E, Hascher A, Bernado P, Blackledge M, Mandelkow E, Svergun DI. Biochemistry. 2008; 47(39):10345–53. [PubMed: 18771286]
51. Schwob E, Bohm T, Mendenhall MD, Nasmyth K. Cell. 1994; 79(2):233–44. [PubMed: 7954792]
52. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD. Structure. 2010; 18(4):494–506. [PubMed: 20399186]
53. Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, Kay LE, Tyers M, Forman-Kay JD. Proc Natl Acad Sci U S A. 2008; 105(46):17772–7. [PubMed: 19008353]
54. Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE, Forman-Kay JD. Nature. 2015; 519(7541):106–9. [PubMed: 25533957]
55. Huang YQ, Yoon MK, Otieno S, Lelli M, Kriwacki RW. J Mol Biol. 2015; 427(2):371–386. [PubMed: 25463440]
56. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. J Appl Crystallogr. 2003; 36:1277–1282.
57. BIOISIS BIOISIS.net. http://www.bioisis.net/
58. Svergun, D. Determination of the integral parameters of particles. In: Taylor, GW., editor. Structure analysis by small-angle x-ray and neutron scattering. Plenum; New York and London: 1987. p. 59-107.
59. Fornet, AGAG. Small-Angle Scattering of X-Rays. John Wiley and Sons; New York: 1955.
60. Lhuillier D. J Phys-Paris. 1988; 49(5):705–710.
61. Minton AP. Biophys J. 2005; 88(2):971–985. [PubMed: 15596487]
62. Bastidas M, Gibbs EB, Sahu D, Showalter SA. Concept Magn Reson A. 2015; 44(1):54–66.
63. Bracken C, Palmer AG, Cavanagh J. J Biomol Nmr. 1997; 9(1):94–100. [PubMed: 9081546]
64. Sun ZY, Frueh DP, Selenko P, Hoch JC, Wagner G. J Biomol Nmr. 2005; 33(1):43–50. [PubMed: 16222556]
65. Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. Angew Chem Int Edit. 2005; 44(20):3089–3092.
66. Bermel W, Bertini I, Felli IC, Kummerle R, Pierattelli R. Journal of Magnetic Resonance. 2006; 180(2):321–321.
67. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. J Biomol Nmr. 1995; 6(3):277–293. [PubMed: 8520220]
68. Keller, RLJ. The Computer Aided Resonance Assignment Tutorial. 1. CANTINA Verlag; Goldau, Swizterland: 2004.
69. Radhakrishnan A, Vitalis A, Mao AH, Steffen AT, Pappu RV. The journal of physical chemistry B. 2012; 116(23):6862–71. [PubMed: 22329658]
70. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernandez CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. Biophys J. 2015; 109(8):1528–32. [PubMed: 26488642]
71. Holehouse AS, Ahad J, Das RK, Pappu RV. CIDER: Classification of Intrinsically Disordered Ensemble Regions. Biophys J. 2015; 108(2):228a. [PubMed: 25606670]
72. Svergun D, Barberato C, Koch MHJ. J Appl Crystallogr. 1995; 28:768–773.
73. Eyal E, Najmanovich R, McConkey BJ, Edelman M, Sobolev V. Journal of computational chemistry. 2004; 25(5):712–724. [PubMed: 14978714]
74. UniProt C. Nucleic Acids Res. 2015; 43:D204–12. Database issue. [PubMed: 25348405]
75. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, Gough J. Nucleic Acids Res. 2013; 41:D508–16. Database issue. [PubMed: 23203878]
76. Holehouse AS, Naegle KM. PloS one. 2015; 10(12):e0144692. [PubMed: 26659599]

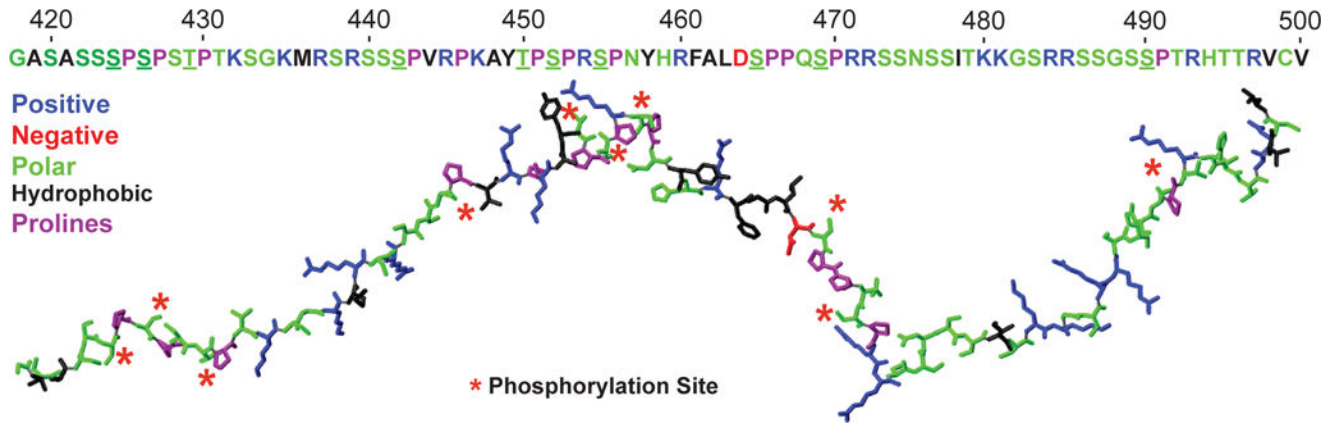


Figure 1. Primary sequence of Ash1$^{420–500}$

The sequence of Ash1$^{420–500}$ is shown at the top and the color-coding of residues is described below. The residues are also shown in a stick representation for a generic conformation of Ash1$^{420–500}$. Phosphorylation sites are underlined in the primary sequence and marked by red asterisks in the conformation. Positively and negatively charged residues, small polar residues, hydrophobic and proline residues are colored blue, red, green, black and purple, respectively. The Ash1 construct has two exogenous N-terminal residues, GA, and these remain after cleavage of the affinity tag.

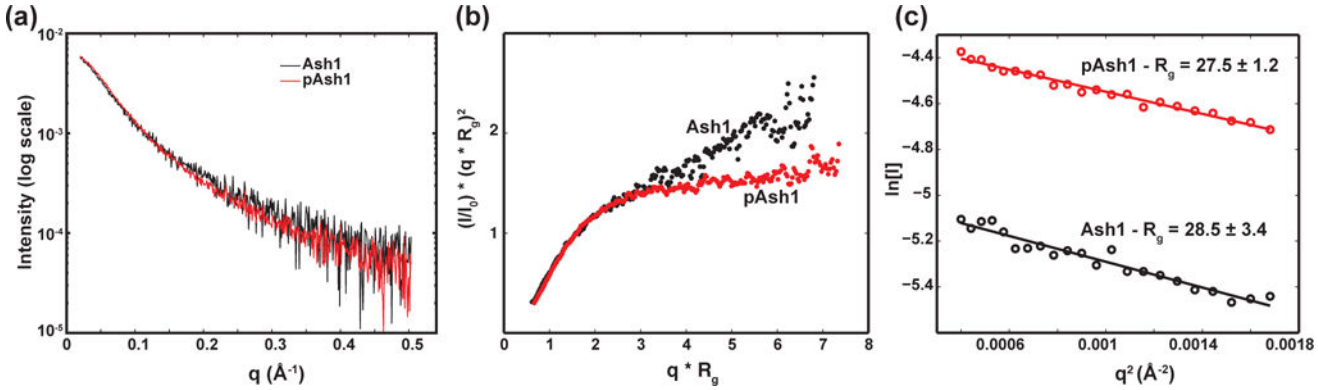


Figure 2. Experimental SAXS data indicates disordered nature of Ash1 and pAsh1

(a) Raw SAXS data truncated at q = 0.5, (b) dimensionless Kratky plots generated using R_g and I_0 that are calculated from the Guinier analysis and (c) Guinier plots for Ash1 (black) and pAsh1 (red) in aqueous solution and 150 mM NaCl.

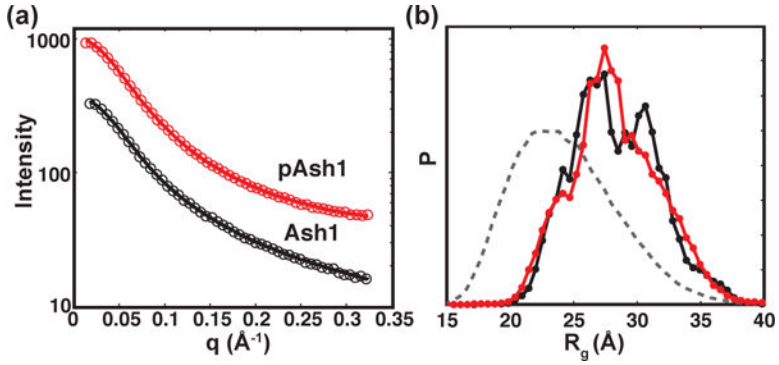


Figure 3. Ensemble modeling of Ash1 and pAsh1 SAXS data

(a) Fits of the scattering curves calculated from one representative EOM ensemble to experimental data. Final ensembles are the result of averaging 100 independent iterations. (b) Rg distributions of the random pool (grey, dashed line), and of Ash1 (black line, markers) and pAsh1 (red line, markers) calculated with EOM from SAXS data of samples in aqueous solution containing 150 mM NaCl. The EOM ensembles contain 20–30 conformers, resulting in rough Rg distributions.

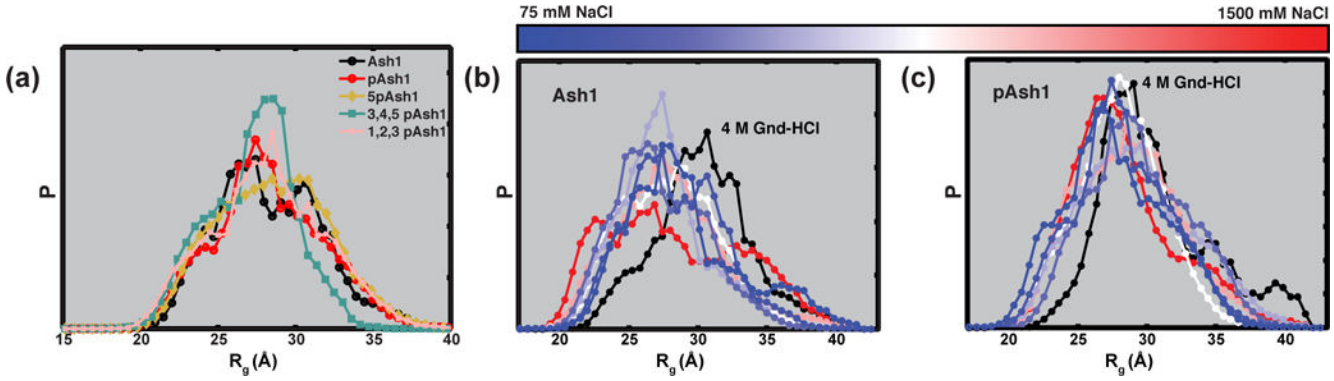


Figure 4. The global dimensions of Ash1 are largely insensitive to phosphorylation state and NaCl concentration

(a) Ensemble R_g distributions of Ash1 with different extent of phosphorylation. 5pAsh1 is a mutant with only 5 phosphorylation sites intact, (phosphorylation sites at residues 424, 429, 450, 455 and 469 mutated to Ala), 3,4,5pAsh1 and 1,2,3pAsh1 are generated by kinase treatment with sub-stoichiometric amounts of ATP. Ensemble R_g distributions of (b) Ash1 and (c) pAsh1, respectively, for NaCl concentrations ranging from 75 (blue) to 1500 mM (red) and 4 M Gnd-HCl (black line). SAXS curves can be found in Figure S3 and the means and standard deviations for all histograms can be found in Tables S1 and S2.

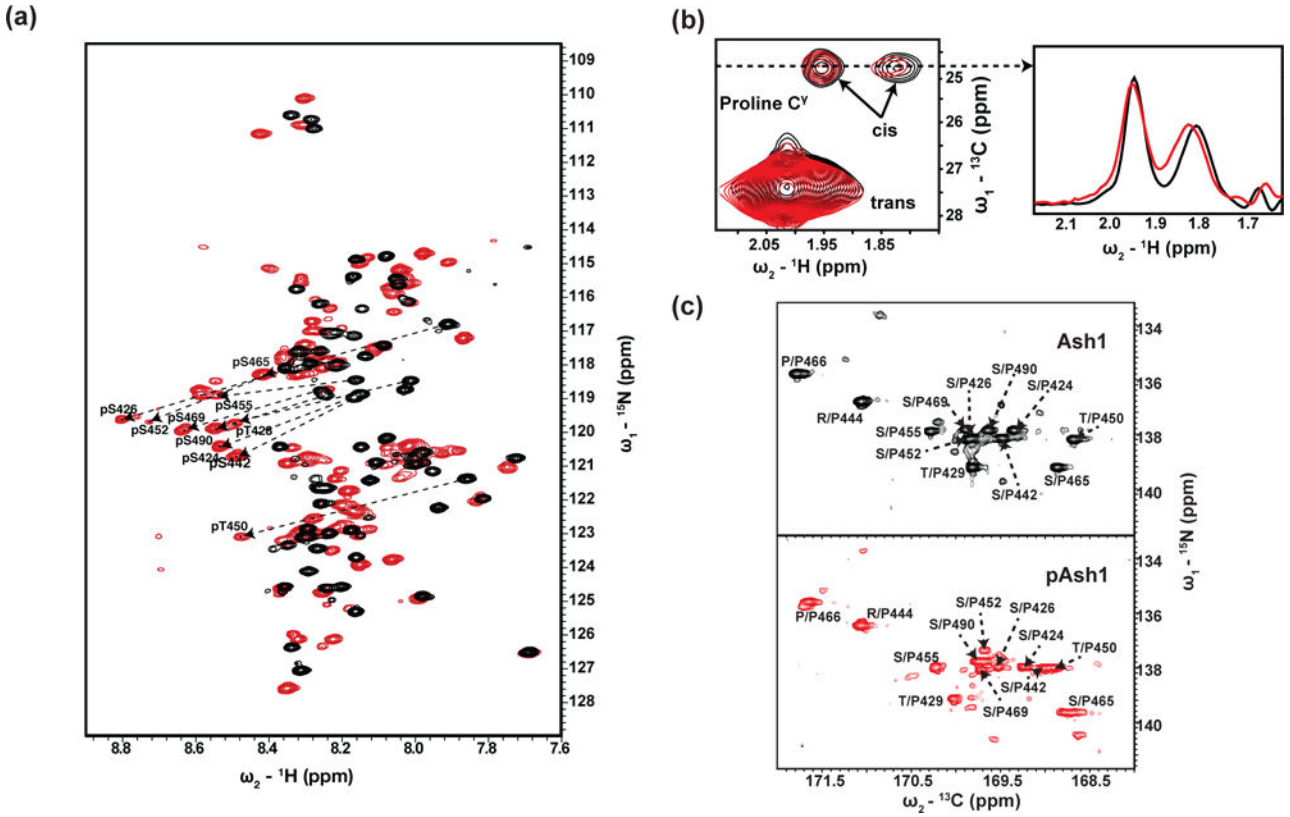


Figure 5. Phosphorylation of Ash1 preserves disorder but leads to local conformational changes

(a) Superposition of $^{1}H,^{15}N$ HSQC NMR spectra for Ash1 (black) and pAsh1 (red). Signals of phosphorylated residues are labeled and experience a ^{1}H downfield shift at pH 6.95. Additional chemical shift changes indicate local conformational changes. For spectra with full assignments see Figure S4. (b) Proline Cγ region of ^{1}H, ^{13}C HSQC spectra for Ash1 (black) and pAsh1 (red). One strong degenerate resonance from all trans proline Cγs is observed, while the cis proline Cγs result in two small signals, with an upfield shift of ~3 ppm. 1D slices through the cis proline signals for Ash1 (black) and pAsh1 (red), intensities normalized to the trans resonances, show similar global cis proline populations for Ash1 and pAsh1 of 9.7 ± 1.8 % and 8.4 ± 1.1 %, respectively (Table S4). (c) Proline region of CON spectra show extensive resonance splitting due to cis/trans proline isomerization and the extent of splitting is qualitatively similar for Ash1 and pAsh1. The major signals stem from trans proline residues, minor signals from cis proline residues and trans proline residues in sequence vicinity of a cis proline.

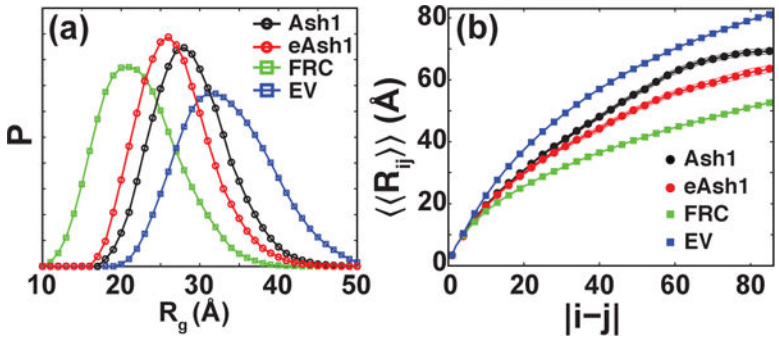


Figure 6. Ash1 and eAsh1 sample expanded coil-like ensembles

(a) The distribution of R_g values obtained from all atom simulations is shown for Ash1, the phosphomimetic eAsh1, and the two reference ensembles viz., the Flory Random Coil (FRC) and Excluded Volume (EV) limits. Simulations were carried out in the presence of 50 mM NaCl for Ash1 and eAsh1. Ash1 and eAsh1 show similar distributions, with global dimensions between the EV and FRC reference limits. (b) Internal scaling profiles for the four simulations from panel a. For every pair of residues at a given sequence separation (|i–j|) the average through-space distance between each pair of residues at that sequence separation, $\langle\langle \boldsymbol{R}_{i,j} \rangle\rangle$, is shown. This provides a summary description of the scaling of intra-chain distances of the polymer. The mean ± s.e.m. is shown as two thinner solid lines.

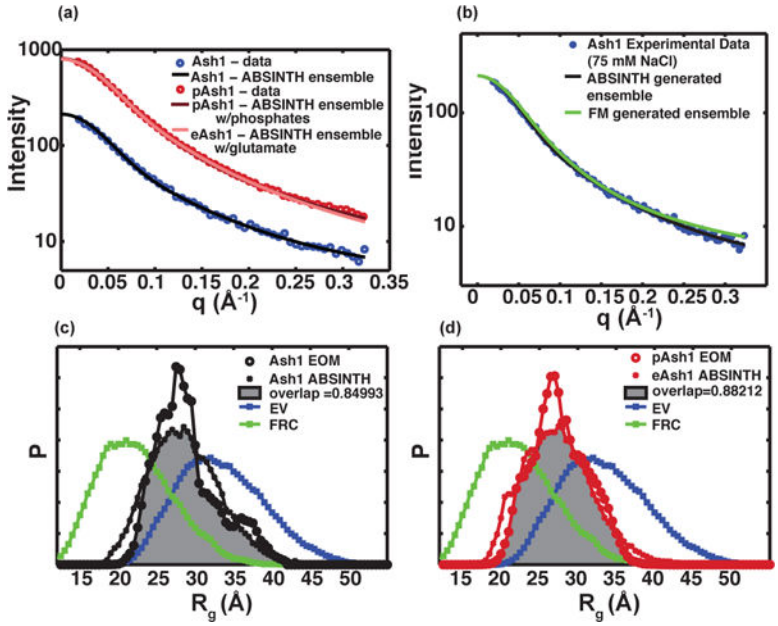


Figure 7. All atom simulations reproduce the expansion of Ash1 and pAsh1/eAsh1 observed by SAXS

(a) Simulation derived scattering curves for Ash1 and eAsh1 compared to the SAXS scattering curve for Ash1 and pAsh1. To generate a pAsh1 ensemble, all phosphomimetic glutamate residues in eAsh1 were substituted by pS or pT residues. (b) Comparison of Ash1 scattering profiles derived from ABSINTH simulations and a Flexible-Meccano ensemble with experimental scattering data. (c) and (d) Overlap of the R_g distributions for Ash1 (c) and e/pAsh1 (d) ensembles generated by EOM or all atom simulations. The overlap is best for the EOM and simulation-derived ensembles. The incomplete overlap is in part caused by the jagged size distributions of the EOM ensembles, which are caused by the fact that they consist of a small number of conformers that collectively agree with the experimental data, but do not explicitly have a coil-like size distribution.

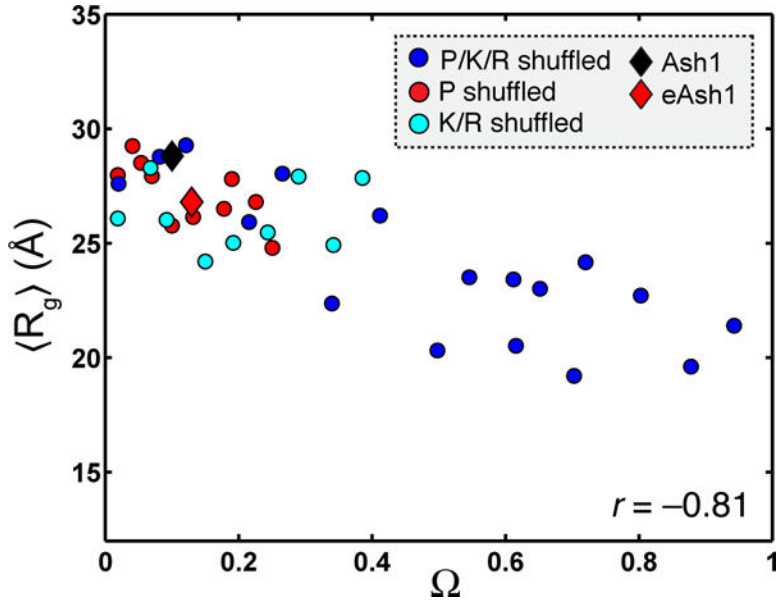


Figure 8. Summary of simulation results showing the variation of radii of gyration with Ω. A rational sequence design algorithm was deployed to generate 30 distinct sequence permutants by changing the patterning of proline and charged residues vis-à-vis all other residues. This was achieved by shuffling the positions of proline and charged residues and fixing the positions of all other residues. The complete set of sequences can be found in Figure S13. Three independent ABSINTH simulations were run for each permutant to determine the mean radius of gyration associated with the ensemble. The results show an inverse correlation between Ω and the R_g values.

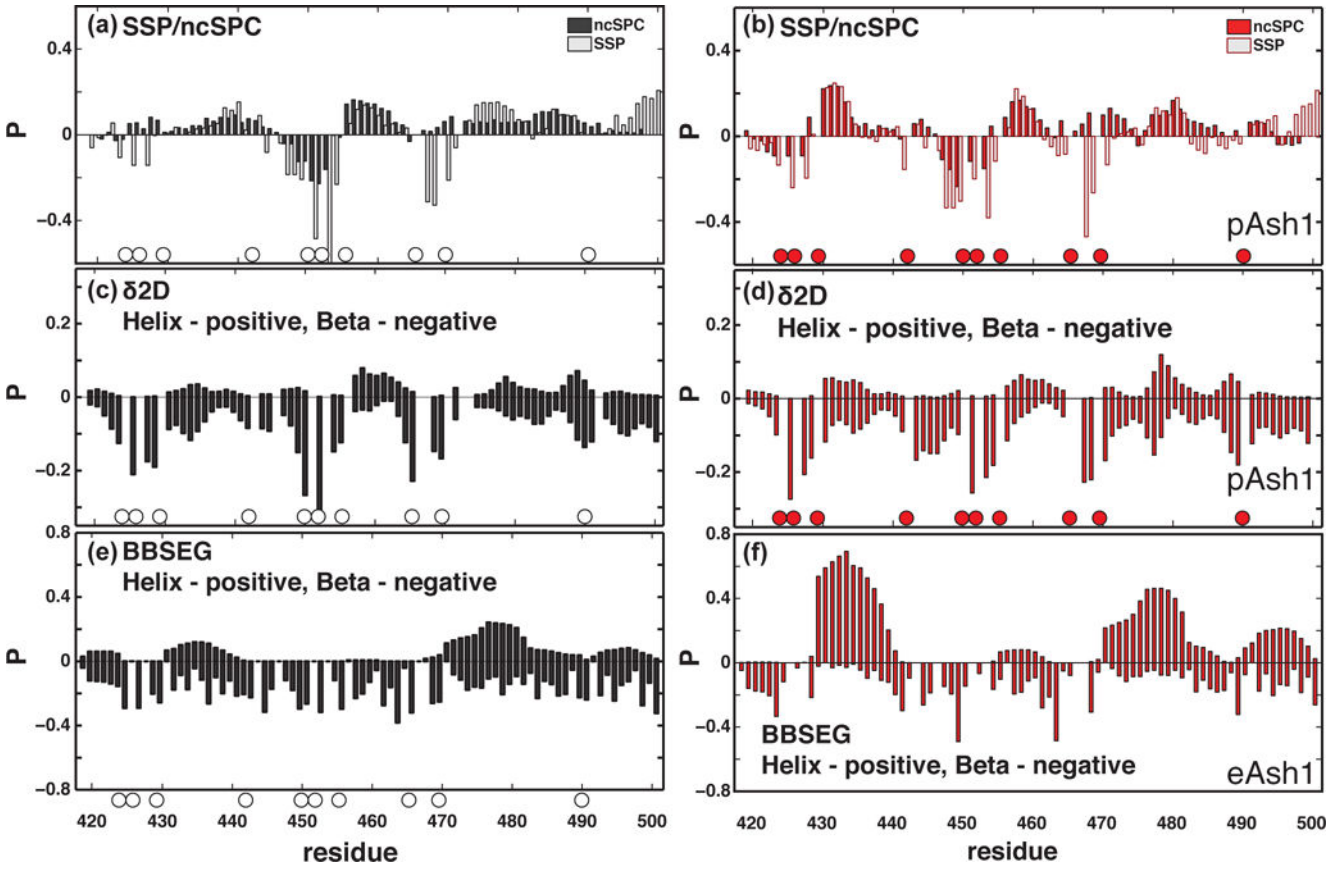


Figure 9. Secondary structure propensities from chemical shift data and simulation

The circles define the position of phosphosites. (a, b) Secondary structure propensities for Ash1 (black) and pAsh1 (red) calculated from C′, C_α and C_β chemical shifts using SSP[43] and ncSPC.[44] (c, d) Secondary structure propensities for Ash1 (black) and pAsh1 (red) calculated from C′, C_α, C_β, N and $^1H^N$ chemical shifts using δ2D.[45] (e, f) Secondary structure propensities for Ash1 (black) and pAsh1 (red) calculated from atomistic simulations using the distributions of backbone dihedral angles.

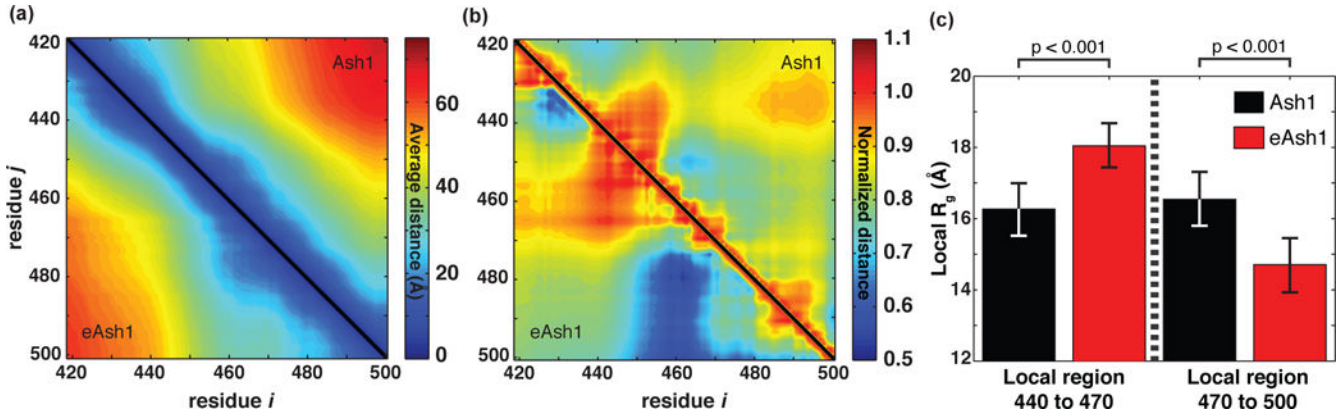


Figure 10. Scaling maps reveal compensatory compaction and expansion upon substituting Thr and Ser residues within phosphosites of Ash1 with Glu

(a) The distance map summarizes the average distance between each pair of residues in Ash1 (upper triangle) and eAsh1 (lower triangle). Both Ash1 and eAsh1 show apparently uniform expansion across all length scales, consistent with expanded, coil-like ensembles sampled by both sequences. (b) All inter-residue values in panel (a) were normalized using the inter-residue distances from an EV simulation for Ash1 (upper triangle) and eAsh1 (lower triangle). This leads to a normalized scaling map. Regional biases for compaction (scaled distances less than unity) or expansion (scaled distances greater than unity) become clearer when operating in a normalized distance space. (c) We calculated the local R_g associated with two sub-peptides in the context of the full chain to demonstrate the compensatory changes observed in Ash1 vs. eAsh1. The 30-residue stretch between residue 440 and 470 is more expanded in eAsh1 than in Ash1, while the 30-residue stretch between residues 470 and 500 is more compact. These regions were identified from the scaling maps in panel (b).

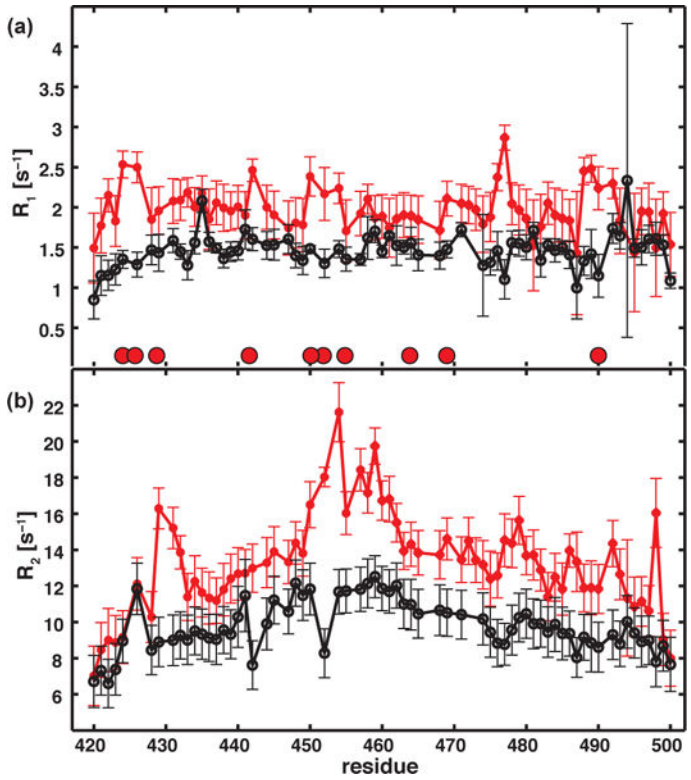


Figure 11. Experimental longitudinal and transverse relaxation rates

(a) Ash1 (black) and pAsh1 (red) R_1 rates and (b) R_2 rates. Phosphorylation sites are marked by red circles. Enhanced R_2 rates for pAsh1 are in agreement with competing transient interactions.