Received: 23 October 2023 | Revised: 13 March 2024 | Accepted: 26 March 2024

DOI: 10.1002/pro.4986

RESEARCH ARTICLE

THE PROTEIN SOCIETY WILEY

How hydrophobicity, side chains, and salt affect the dimensions of disordered proteins

Michael C. Baxa[1] | **Xiaoxuan Lin**[1] | **Cedrick D. Mukinay**[2] | **Srinivas Chakravarthy**[3] | **Joseph R. Sachleben**[4] | **Sarah Antilla**[5] | **Nina Hartrampf**[6] | **Joshua A. Riback**[7] | **Isabelle A. Gagnon**[1] | **Bradley L. Pentelute**[6] | **Patricia L. Clark**[2] | **Tobin R. Sosnick**[1]

[1]Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, Illinois, USA

[2]Department of Chemistry & Biochemistry, University of Notre Dame, Notre Dame, Indiana, USA

[3]Biophysics Collaborative Access Team (BioCAT), Center for Synchrotron Radiation Research and Instrumentation and Department of Biological and Chemical Sciences, Illinois Institute of Technology, Chicago, Illinois, USA

[4]Division of Biological Sciences, University of Chicago, Chicago, Illinois, USA

[5]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[6]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[7]Graduate Program in Biophysical Science, University of Chicago, Chicago, Illinois, USA

Correspondence
Patricia L. Clark, Department of Chemistry & Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA.
Email: pclark1@nd.edu

Tobin R. Sosnick, Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL 60637, USA.
Email: trsosnic@uchicago.edu

Present addresses
Srinivas Chakravarthy, Cytiva, Fast Trak, Marlborough, MA, USA; Nina Hartrampf, Department of Chemistry, University of Zurich, Switzerland; and
Joshua A. Riback, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA.

Funding information
National Institute of General Medical Sciences; National Institutes of Health, Grant/Award Numbers: GM055694, GM130122, GM103622, GM148233, 1S10OD018090-01, DP1-GM146256; W. M.

Abstract

Despite the generally accepted role of the hydrophobic effect as the driving force for folding, many intrinsically disordered proteins (IDPs), including those with hydrophobic content typical of foldable proteins, behave nearly as self-avoiding random walks (SARWs) under physiological conditions. Here, we tested how temperature and ionic conditions influence the dimensions of the N-terminal domain of pertactin (PNt), an IDP with an amino acid composition typical of folded proteins. While PNt contracts somewhat with temperature, it nevertheless remains expanded over 10–58°C, with a Flory exponent, ν, >0.50. Both low and high ionic strength also produce contraction in PNt, but this contraction is mitigated by reducing charge segregation. With 46% glycine and low hydrophobicity, the reduced form of snow flea anti-freeze protein (red-sfAFP) is unaffected by temperature and ionic strength and persists as a near-SARW, $\nu \sim 0.54$, arguing that the thermal contraction of PNt is due to stronger interactions between hydrophobic side chains. Additionally, red-sfAFP is a proxy for the polypeptide backbone, which has been thought to collapse in water. Increasing the glycine segregation in red-sfAFP had minimal effect on ν. Water remained a good solvent even with 21 consecutive glycine residues ($\nu > 0.5$), and red-sfAFP variants lacked stable backbone hydrogen bonds according to

Keck Foundation; U.S. Department of Energy, Grant/Award Number: DEAC02-06CH11357

Review Editor: Lynn Kamerlin

hydrogen exchange. Similarly, changing glycine segregation has little impact on ν in other glycine-rich proteins. These findings underscore the generality that many disordered states can be expanded and unstructured, and that the hydrophobic effect alone is insufficient to drive significant chain collapse for typical protein sequences.

KEYWORDS

hydrogen bonding, hydrogen exchange, hydrophobicity, ionic strength, NMR, SAXS, solvent quality, unfolded states

1 | INTRODUCTION

Understanding the molecular details of denatured state ensembles (DSEs) and intrinsically disordered proteins (IDPs) and how they respond to changes in the environment will be central to resolving outstanding questions in biomolecular thermodynamics (English et al., 2018; Kohn et al., 2004; Riback et al., 2017; Wuttke et al., 2014). For example, does the stability of a protein, as defined by the Gibbs–Helmholtz equation, depend on whether the thermodynamic reference state is the unfolded state ensemble under native conditions (DSE_{Native}), or the thermally or chemically denatured ensemble (DSE_{therm} or DSE_{chem})? What is the relative importance of backbone hydrogen bonding versus side chain hydrophobicity in determining the properties of the unstructured chain? Is solvent quality at room temperature the same as at 60°C?

We have extensively studied the passenger domain of pertactin, a 539-residue parallel β-helix from *Bordetella pertussis*, in our investigations of IDP properties (Bowman et al., 2020; Renn et al., 2012; Riback et al., 2017; Riback et al., 2019). Despite having a sequence that can fold in the context of the full-length protein, the N-terminal 334 amino acids of pertactin (PNt), behaves as an intrinsically disordered, highly expanded chain. PNt can therefore serve as a valuable model system for studying the properties of unstructured protein ensembles at equilibrium under physiologically relevant conditions. Moreover, the large size of PNt relative to other DSE models increases the accuracy of determining global properties of the DSE.

These properties include the radius of gyration, R_g, and the solvent quality, describable using the Flory scaling exponent, ν, in the relationship $R_g = R_0 N^{\nu}$, where N is the number of residues and R_0 is a pre-factor that depends on, among other things, the persistence length of the polymer (Flory, 1953; Riback et al., 2017). The Flory exponent provides a universal parameter for describing how polymers respond to their environment. In the Flory approximation of protein polymers, solvent quality typically ranges from $\nu = 1/3$ to 3/5, with a transition at $\nu = 1/2$, the so-called θ-solvent limit, below which intra-protein interactions are stronger than protein–solvent interactions (Flory, 1953). For $\nu < 1/2$, the solvent quality is poor, and protein–protein interactions promote chain compaction. By contrast, a good solvent exists when $\nu > 1/2$, and protein–solvent interactions are favored. The Flory scaling limit of a homopolymer is the ideal self-avoiding random walk (SARW), or $\nu = 3/5$, where protein–solvent interactions dominate (although interactions can affect the scaling limit; e.g., helical rods have $\nu = 1$ [Das and Pappu, 2013; Mao et al., 2010]). In 2 M guanidinium chloride (GdmCl), PNt and other proteins approach the SARW limit (Jacob et al., 2004; Jacob et al., 2007; Yoo et al., 2012).

For many protein sequences, especially those in the PDB, water is predicted to be a poor solvent for the chain, that is, $\nu < 1/2$ (Das et al., 2015; Hofmann et al., 2012). However, previous small-angle x-ray scattering (SAXS) measurements found that water is a good solvent for PNt, with $\nu = 0.54$. Nevertheless, the properties of the PNt conformational ensemble are sensitive to its amino acid sequence pattern. Disrupting the well-mixed hydrophobic/hydrophilic sequence pattern by rearranging only a small subset (2%–5%) of PNt's amino acids led to a measurable contraction and decrease in ν (Bowman et al., 2020).

Temperature and ionic strength are two environmental variables that impact protein stability and a polypeptide chain's physical properties. Far-UV circular dichroism (CD) measurements of disordered chains have reported ellipticity changes at elevated temperatures, which have been interpreted as the loss of secondary structure (English et al., 2018; Uversky, 2009). However, the interpretation of CD signal changes can be ambiguous, for example, reflecting differences in backbone dihedral angle propensities rather than stable hydrogen-bonded structure (Kjaergaard et al., 2010).

SAXS and Förster resonance energy transfer (FRET) studies have also reported chain contractions at elevated temperatures in both IDPs and DSEs of hydrophobic and

hydrophilic sequences, with interpretations including hydrogen bond formation, non-specific hydrophobic collapse, and temperature-dependent solvation energies (Nettels et al., 2009; Sadqi et al., 2003; Stenzoski et al., 2018; Wuttke et al., 2014). Multiple groups have also measured the effect of different salts and ionic strength on the conformation of disordered chains (Hofmann et al., 2008; Maity et al., 2022; Müller-Späth et al., 2010; Vancraenenbroeck et al., 2019; Wiggers et al., 2021) and the stabilities of proteins (Pegram et al., 2010; Pegram and Record Jr., 2008). However, there has not been an analysis of the effect of temperature and ionic strength on the solvent quality for disordered proteins.

In this work, we investigate the effect of temperature and salt on the conformational properties of polypeptides outside the native state energy basin using a variety of model IDPs, including PNt and the reduced, glycine-rich snow flea anti-freeze protein (red-sfAFP), which serve as mimics of the DSE of a foldable protein and the polypeptide backbone, respectively. Due to improved data quality and modern analytical tools, coupled with size-exclusion chromatography (SEC), SAXS can now be used to simultaneously determine both the R_g and solvent quality ν of DSEs in different environments, independent of aggregation effects (Riback et al., 2017). Both PNt and red-sfAFP are expanded IDPs under physiological conditions, yet the structural features of their DSEs vary with the denaturing agent and ionic strength. For PNt, we found that the solvent quality decreases monotonically with temperature, while changes to ionic strength elicit a more complex behavior. The effect of ionic strength on the solvent quality of PNt, however, appears to be independent of temperature. The dimensions of the red-sfAFP conformational ensemble are consistent with a near-SARW and are nearly temperature-independent, implying that the temperature-dependent contraction of PNt is due to the increased strength of side chain hydrophobic interactions at higher temperatures, rather than a change in backbone properties. Our findings have broad implications for interpreting biothermodynamic experiments, notably that the hydrophobic effect alone is insufficient to drive global collapse of protein chains below 60°C.

2 | RESULTS

2.1 | Hydrophobicity-driven temperature-dependent contraction

We performed SAXS using an in-line SEC column to ensure monodispersity of the PNt sample. By analyzing SAXS profiles using the molecular form factor (MFF) for disordered polymers (Riback et al., 2017), we obtained both the R_g and ν of PNt in an aqueous buffer (10 mM sodium phosphate, 150 mM KCl, pH 7.4) at temperatures from 10 to 58°C (Figure 1). As temperature increased, the R_g and ν of PNt decreased by ~11% (from 52.72 ± 0.14 Å to 46.98 ± 0.14 Å) and 9% (0.555 ± 0.002 to 0.505 ± 0.002), respectively, while the pre-factor R_0 increased from 2.10 ± 0.02 Å to 2.50 ± 0.02 Å as expected (Riback et al., 2017). Despite this temperature-dependent contraction, our data demonstrate that water remains a good solvent for PNt (e.g., $\nu > 1/2$) over a significant temperature range, well above the *Bordetella pertussis* growth temperature (37°C).

The mild contraction of PNt at higher temperatures is consistent with the known increase in the strength of hydrophobic interactions with increasing temperature (Kjaergaard et al., 2010; Langridge et al., 2014; Nettels et al., 2009; Sadqi et al., 2003; Wuttke et al., 2014). To better examine the origin of this effect, we also measured the temperature dependence for a model protein that serves as a proxy for the behavior of the protein backbone alone. This protein, reduced sfAFP (red-sfAFP), has a high glycine content (46%) and a low percentage of hydrophobic (6% I/L/V/F/Y/W) and charged (7% R/K/D/E) residues, and the reduction of the two disulfide bonds in sfAFP leads to a disordered conformation (Gates et al., 2017). At 10°C, red-sfAFP is in the good solvent regime, with $\nu = 0.530 \pm 0.008$ and $R_g = 23.69 \pm 0.12$ Å. In contrast to PNt, the dimensions of red-sfAFP increased slightly with temperature. The R_g and ν changed to 24.29 ± 0.08 Å and 0.554 ± 0.006, respectively, at 58°C (Figure 1). The lack of a temperature-dependent contraction of red-sfAFP, our backbone control, supports the interpretation that the mild temperature-dependent contraction observed for PNt is due to interactions between its hydrophobic side chains. Alternative hypotheses to explain this temperature-dependent contraction, including a change in polypeptide backbone properties, driven by changes to hydrogen bonding and/or Ramachandran angle distribution, are inconsistent with our other investigations, as discussed below.

2.2 | Changing dihedral angle propensities have a negligible effect on R_g

To investigate the relationship between dihedral angle distributions and the dimensions of an IDP, we simulated PNt as a SARW ensemble at multiple temperatures using replica exchange molecular dynamics (MD) simulations with the nearly-all-atom *Upside* MD algorithm (Jumper et al., 2018a; Jumper et al., 2018b). In *Upside*, the backbone dihedral angle (or Ramachandran) energy term is

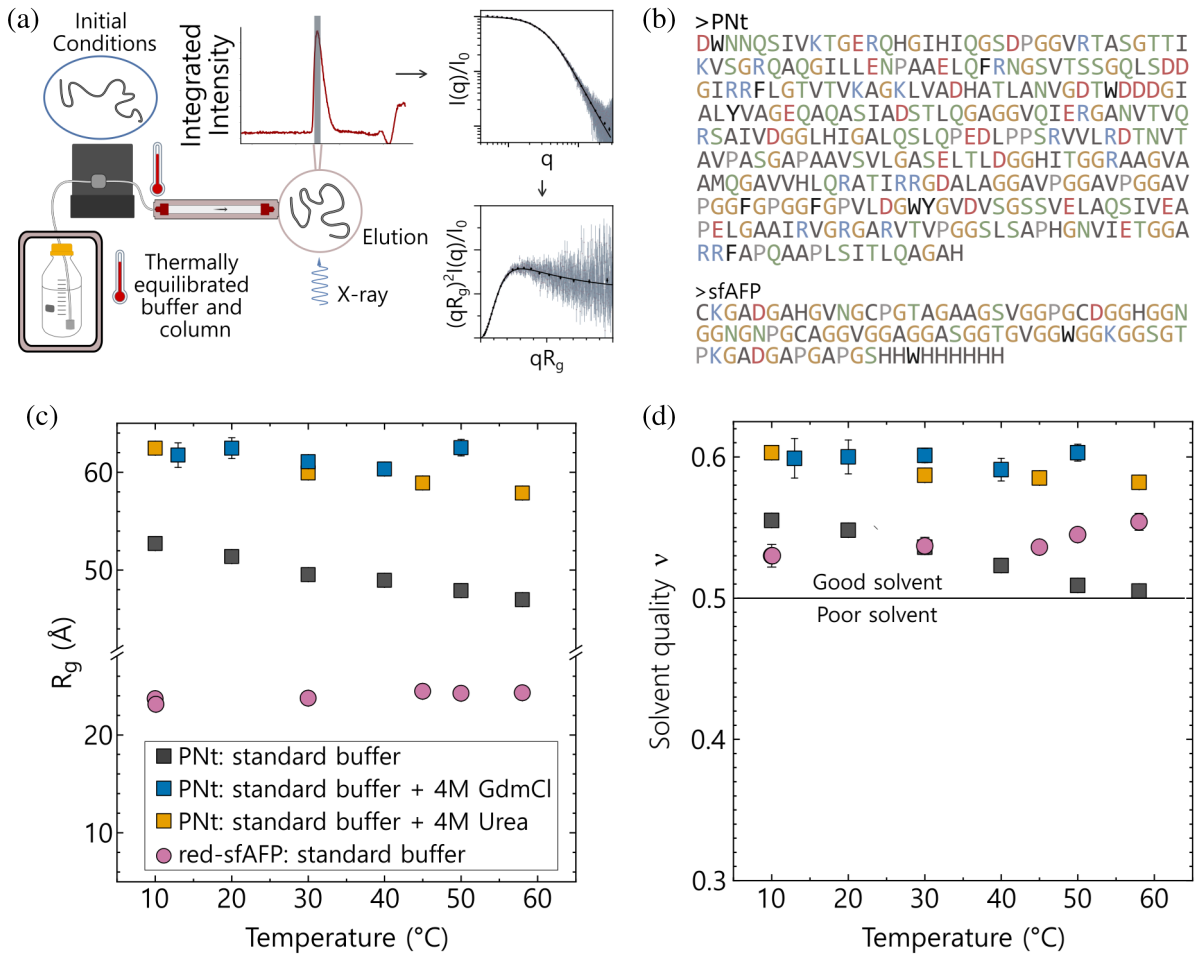

FIGURE 1 (a) Temperature-dependent SEC-SAXS of an IDP with a protein-like sequence (PNt) and a glycine-rich IDP (red-sfAFP). (b) Residues in PNt and sfAFP sequences are categorized as hydrophobic (gray), aromatic (black), polar (green), positive (blue), and negative (red). Glycine and proline residues are colored orange and light gray, respectively. (c) The R_g and (d) solvent quality, ν, of PNt and red-sfAFP are obtained using a molecular form factor (Riback et al., 2017) (MFF) from SAXS measurements collected across a temperature range from 10 to 58°C in our standard buffer condition: 10 mM sodium phosphate, 150 mM KCl, pH 7.5 (2 mM DTT is included to reduce disulfide bonds). By 58°C, the buffer is approaching the θ-solvent limit ($\nu = 1/2$), while in 4 M GdmCl, PNt behaves as a self-avoiding random walk ($\nu = 3/5$) at all temperatures. In contrast to PNt, the global properties of a glycine-rich IDP (reduced sfAFP) do not significantly change with temperature. Error bars from fitting with the MFF are present for all points but are often smaller than the symbols. IDP, intrinsically disordered protein; PNt, N-terminal domain of pertactin; SEC-SAXS, size-exclusion chromatography-small-angle x-ray scattering.

solely enthalpic and is constant with temperature. Nevertheless, the relative Ramachandran basin populations are governed by free energies and thus subject to entropic effects (i.e., the distribution becomes more uniform at higher temperatures as the importance of backbone entropy increases with temperature, leading to broader sampling of the Ramachandran map) (Figure 2 and Figure S1 in Data S1). For PNt ensembles, the populations of the poly-proline 2 (PPII) and β basins are anti-correlated with increasing temperature. While PPII is the dominant basin at low temperatures, the two populations nearly converge at higher temperatures as the β basin is broader and, hence, has higher entropy. An increased population of the β basin at higher temperatures is consistent with experimental measurements of peptides (Kjaergaard et al., 2010; Shi et al., 2002; Toal et al., 2014). Despite the temperature-dependent shifts in Ramachandran basin occupancies for PNt, the effect on R_g is minimal, changing by less than 1.9 Å (3%) across the entire simulated temperature range (Figure 2) and only 0.4 Å over the temperature range explored in our SAXS measurements ($T_{\text{Upside}} = 0.82$–0.97, corresponding to $T_{\text{exp}} = 11$–63°C). Similarly, across the simulated temperature range, the predicted scattering profiles yielded $\nu = 0.6$, as expected for a SARW (Figure S2 in Data S1). These results demonstrate that changes in backbone basin probabilities do not account for the temperature-dependent decrease in solvent quality for PNt.

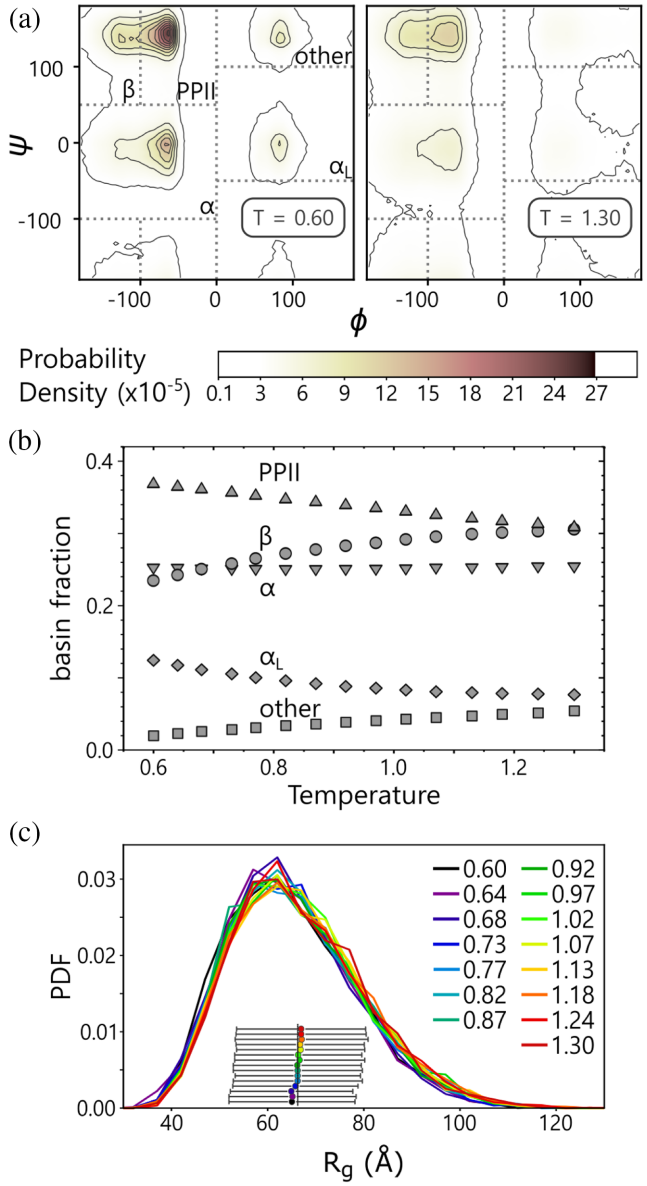

FIGURE 2 The R_g of a self-avoiding polypeptide is invariant as a function of temperature, despite changes in Ramachandran basin probabilities. (a) Ramachandran basins are defined as in Jha et al. (2005) and depicted by gray dashed lines. Replica exchange simulations of PNt were carried out using the *Upside* algorithm but with only the self-avoidance and dihedral angle energy terms turned on. The Ramachandran angle distributions change between $T = 0.60$ (*left*) and 1.3 (*right*) *Upside* temperature units as the shape of the individual basins are different. Contour line values are indicated on the color bar. (b) Some basin probabilities change with temperatures (e.g., PPII and β) while others do not (α-helix), even though the Ramachandran potential in *Upside* does not have an explicit temperature dependence. (c) Despite the changing dihedral angle propensities, the average R_g at each temperature (inset) is indistinguishable. Error bars represent the standard deviation of the R_g across the structural ensemble. PNt, N-terminal domain of pertactin.

2.3 | Dimensions of PNt in other buffers

In contrast to the temperature-dependent behavior of PNt described above, the dimensions of PNt in 4 M GdmCl were essentially constant from 10 to 58°C (R_g: 60.31–62.51 Å; ν: 0.591–0.603; Figure 1). An increase in temperature is normally equated with a loss of interactions and favors higher entropy states. However, for a SARW having no interactions other than hard-sphere repulsions, the ensemble of chain conformations is already maximally entropic, and hence, the dimensions of the chain are expected to be independent of temperature (Sneppen and Zocchi, 2005). Since chemically denatured proteins are known to behave as SARWs (Kohn et al., 2004), the observed lack of temperature dependence for PNt in 4 M GdmCl is expected. Overall, these results underscore that temperature changes do not inherently alter the dimensions of a polypeptide chain that lacks hydrophobic interactions. We also measured the temperature-dependent conformational ensemble of PNt in urea, a weaker chemical denaturant than GdmCl (Myers et al., 1995; Tischer et al., 2018). PNt behaved as a SARW only up to 20°C in 4 M urea, exhibiting a modest contraction at higher temperatures ($R_g = 62.44 \pm 0.57$ Å and 57.88 ± 0.34 Å at 10°C and 58°C, respectively), although the solvent quality of 4 M urea at 58°C remained high, with $\nu = 0.582 \pm 0.004$ (Figure 1). We interpret the slight contraction at higher temperatures to be the result of some degree of hydrophobic interactions that persist in 4 M urea.

PNt has a 24 amino acid low-complexity region with 12 glycine residues, including four PGG repeats (GGAVPGGAVPGGAVPGGFGPGGFG). To test whether this P/G-rich region affects the dimensions or salt-sensitivity of PNt, we created a truncated version of PNt with residues 232–255 removed (PNt-deltaPG). In our standard 10 mM sodium phosphate, 150 mM KCl buffer, removing the P/G region did not affect the solvent quality (Figure S3 in Data S1).

To test the effects of the cellular milieu on IDP collapse, we compared the dimensions of PNt in our standard buffer to a more physiologically relevant buffer: 50 mM potassium phosphate, 20 mM NaCl, 50 mM KCl, 100 mM glutamate, 0.5 mM $MgCl_2$, and 10 mM ATP (Bar-Even et al., 2011; Milo and Phillips, n.d.) (hereafter referred to as bio-buffer). Beyond testing the effects of a solvent more relevant to the cellular environment—an ongoing concern for in vitro studies using nonphysiological buffers—the addition of ATP was used to investigate the proposal that this cosolute is a hydrotrope that increases protein solubility (Patel et al., 2017), suggesting

it may affect solvent quality. The use of the bio-buffer has only a minimal impact on the chain dimensions of PNt-deltaPG relative to our standard buffer; the R_g and ν decrease slightly, that is, 47.58 ± 0.38 Å and 0.541 ± 0.010 compared to 49.53 ± 0.22 Å and 0.560 ± 0.005, respectively.

2.4 | Probing H-bond formation in glycine-rich variants

Our SEC-SAXS results indicated that a buffer containing 10 mM sodium phosphate, 150 mM KCl, 2 mM DTT, pH 7.5 is a good solvent for red-sfAFP. Nevertheless, the chain does not behave fully as a SARW as ν is lower than expected for a SARW, even at 10°C (Figure 1d). We considered whether this difference could be due to the glycine-rich sequence forming a low level of hydrogen bonds, which can be probed using NMR hydrogen exchange (HX) methods. Specifically, we measured the millisecond HX times between the protons on water molecules with amide backbone protons using the solvent saturation transfer method, CLEANEX-PM (Hwang et al., 1998).

Given the considerable challenge in assigning NMR peaks for a low-complexity IDP such as red-sfAFP with 46% glycine, we distinguish glycine residues from non-glycine residues using ^{15}N-TOCSY-HSQC experiments (Marion et al., 1989) and take advantage of the tendency for glycine amide NH peaks to occupy a distinct region in NMR ^{1}H^{15}N-HSQC spectra. Coupled with the observation that the peaks undergo minimal changes in chemical shift in different solvent conditions, we performed individual comparisons of the HX rates for 21 distinct glycine residues and 27 distinct non-glycine residues across three solvent conditions, despite the absence of specific residue assignments.

We compared HX rates for red-sfAFP in a typical in vitro buffer (50 mM potassium phosphate, 1 mM TCEP, pH 7.0) to our bio-buffer with and without 1 M urea. An sfAFP variant (red-sfAFP*: H8D/C13H/T16D/C43H/V59N/K72R), was used for these measurements, due to its availability. We found that the distribution of HX rates for both glycine and non-glycine residues does not significantly change in the three solvent conditions (Figure 3; R^2 of linear fits of ln k_{obs} were 0.88–0.94). The root mean square (RMS) deviations of ln k_{obs} between the bio-buffer and in vitro buffer are 0.10 and 0.12 for glycine and non-glycine residues, respectively. We did not observe any significant change in HX rates in urea (the RMS for ln k_{obs} between conditions is 0.16 and 0.17 for glycine and non-glycine residues, respectively). If there was significant hydrogen bond formation in our standard

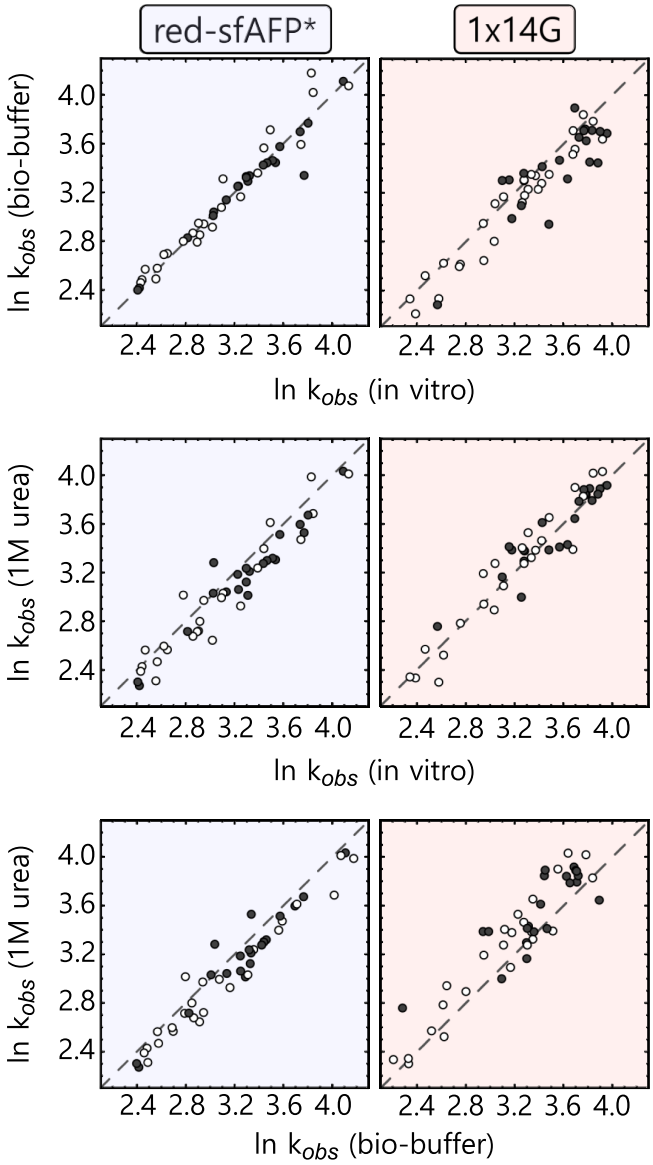

FIGURE 3 CLEANEX-PM hydrogen exchange rates for unfolded sfAFP variants are consistent across different buffer conditions. Red-sfAFP* (H8D/C13H/T16D/C43H/V59N/K72R), a variant of red-sfAFP (*Upper row*), is compared with a shuffled sequence having 14 consecutive glycines (1x14G, *Lower row*). Exchange rates are measured in an "in vitro buffer" (50 mM potassium phosphate, 1 mM TCEP, pH 7.0) and compared with rates from a "bio-buffer" (in vitro buffer + 20 mM NaCl, 50 mM KCl, 100 mM KGlu, 0.5 mM $MgCl_2$, 10 mM ATP), and "1 M urea" buffer (bio-buffer + 1 M urea). Filled and open symbols indicate glycine and non-glycine residues, respectively. The identity line is shown in each panel to guide the eye.

in vitro buffer, addition of urea should have reduced the hydrogen bond frequencies across the protein. In contrast, our results suggest that there is no significant hydrogen bonding in red-sfAFP. Comparing the measured HX rates with the predicted distributions of intrinsic exchange rates (Bai et al., 1993; Connelly et al., 1993;

Nguyen et al., 2018) suggests a maximum protection factor of ~2. The SAXS data provide no evidence that our bio-buffer alters solvent quality for PNt. Overall, our HX/NMR data are consistent with the SAXS measurements, describing the glycine-rich red-sfAFP as an unstructured IDP in a good solvent.

Our findings appear to differ from previous computational and experimental studies, which proposed that glycine-rich sequences undergo collapse in the absence of denaturant (Asthagiri et al., 2017; Karandur et al., 2014; Karandur et al., 2016; Tran et al., 2008). A possible resolution of this paradox is that red-sfAFP, while 46% glycine, has at most two consecutive glycine residues, whereas earlier studies of poly-glycine typically have stretches of 5+ glycines (Asthagiri et al., 2017; Holehouse et al., 2015; Karandur et al., 2014; Karandur et al., 2016; Teufel et al., 2011; Tran et al., 2008). Consequently, we measured HX rates for a variant of red-sfAFP, 1x14G, which has the same sequence composition as red-sfAFP but with 16 residues reordered to create a stretch of 14 consecutive glycine residues.

As with red-sfAFP, the HX rates for 1x14G are unchanged across the three buffer conditions (Figure 3) and we estimate that amide protection factors do not exceed 2. The RMS differences in ln k_{obs} for glycine and non-glycine residues are 0.23 and 0.14, respectively, in the bio-buffer and 0.13 and 0.15 in 1 M urea. Since we do not have specific peak assignments, it is not possible to determine which glycine peaks, if any, correspond to the poly-glycine stretch in the 1x14G variant. However, the invariance of the rates across the different buffers is consistent with a lack of stable hydrogen bonding across an expanded chain that is midway between a SARW and a θ-solvent, where chain–solvent interactions match the strength of intra-chain interactions.

2.5 | Dimensions of glycine-rich variants

In addition to the 1x14G variant, we also rearranged the sfAFP sequence to produce two other poly-glycine stretches, 3x7G and 1x21G. Recombinant expression of 1x21G proved difficult, so the construct was instead produced by chemical synthesis (Hartrampf et al., 2020). The dimensions of 1x14G and 3x7G are similar to WT red-sfAFP and show no signs of contraction at 20°C, despite their internal poly-glycine stretches (Figure S4 in Data S1). Although 1x21G was prone to aggregation, its monomeric form eluted from the SEC column with $\nu = 0.521 \pm 0.015$, consistent with a good solvent but somewhat contracted conformational ensemble. Taken together, the SEC-SAXS data indicate that water is not inherently a poor solvent for poly-glycine stretches.

We also tested the large isoform of AFP (lfAFP) (Mok et al., 2010), which has a slightly higher glycine content compared to sfAFP and in reducing conditions behaves as a SARW. Shuffling or increasing the glycine content in lfAFP had little impact on the dimensions of the conformational ensemble (Figure S5A–C in Data S1). Another glycine-rich protein, the RGG domain of LAF1, contains a significant amount of glycine (35%) and multiple stretches of six to seven consecutive glycine residues. Shuffling the sequence to restrict consecutive glycine residues to no more than two in a row had no effect on the solvent quality of the chain, which was just under the θ-solvent limit (Figure S5D,E in Data S1). Just as with the red-sfAFP variants, the chain dimensions of glycine-rich chains in our study are robust to the presence of consecutive glycine residues.

2.6 | Dimensions of other low complexity sequences

In addition to poly-glycine, poly-glutamine has also been reported to collapse in water, although this collapse has been reported to be sensitive to multiple factors including the length of glutamine stretch, the chemical composition of the flanking sequence, and pH (Crick et al., 2006; Singh and Lapidus, 2008; Vitalis et al., 2008; Walters and Murphy, 2009). Using our shuffled sfAFP sequences as templates, we replaced the glycine stretches with an equivalent number of glutamine residues. While these constructs were aggregation-prone, SEC-SAXS revealed that the monomeric form variants behaved as near SARWs (Figure S6 in Data S1). Similarly, the SAXS profile of poly-asparagine chains with a diversity of chain lengths is well-fit with a $\nu = 0.603 \pm 0.010$ (Figure S6 in Data S1). Collectively, our data reflect that in general, water tends to be a good solvent for a wide variety of polypeptides.

2.7 | Effects of low and high concentrations of salt

We extended previous studies of the ion dependence of IDP conformational ensembles (Hofmann et al., 2008; Müller-Späth et al., 2010; Vancraenenbroeck et al., 2019; Wiggers et al., 2021) by obtaining the solvent quality, which provides a universal parameter for comparison across systems, for PNt and red-sfAFP. We next measured the temperature dependence of R_g and ν for PNt across various ionic conditions, beginning first at 10°C with a low ionic buffer (10 mM sodium phosphate, pH 7.4). The removal of salt leads to an 11% and 9% reduction in R_g

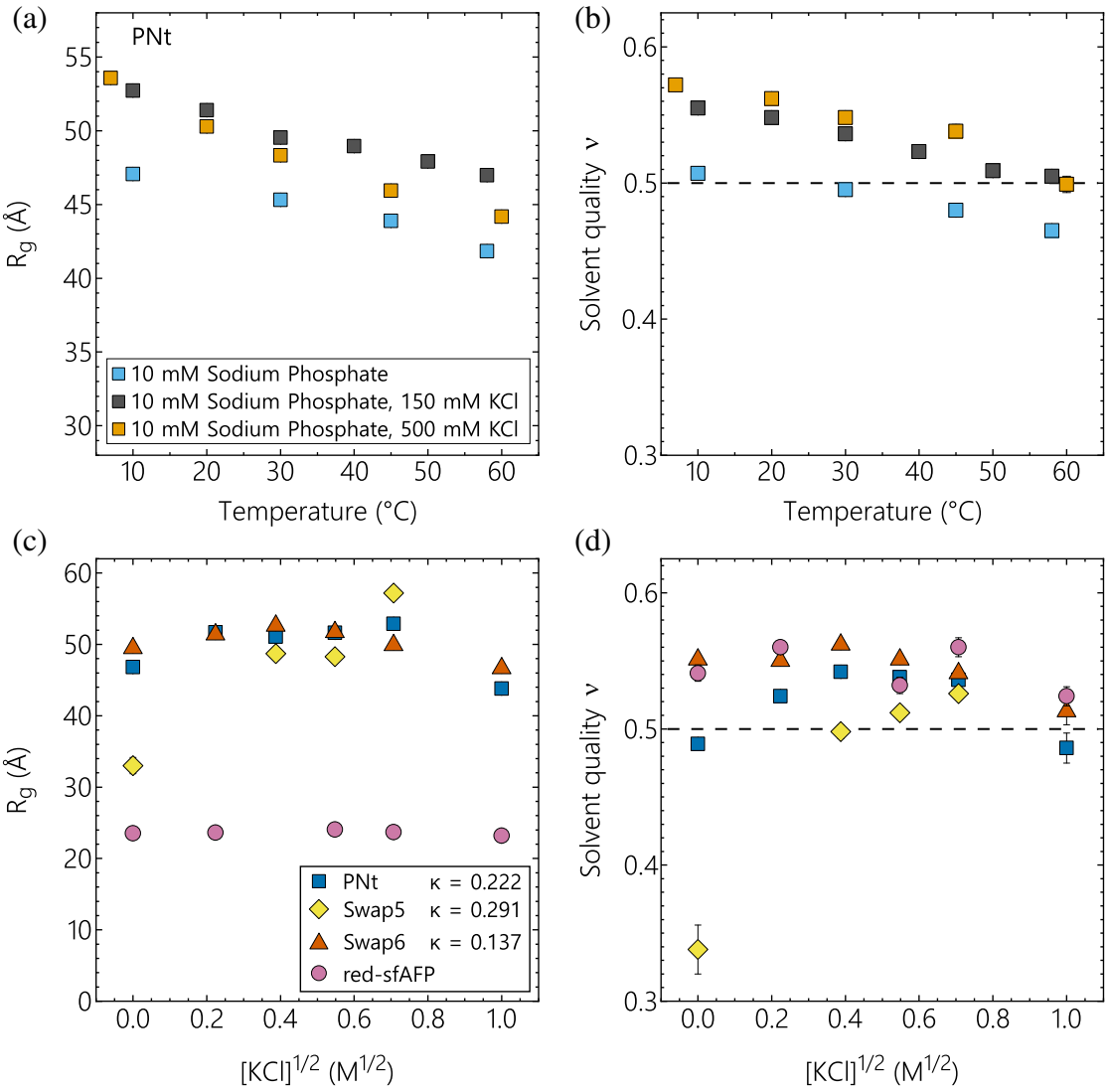

FIGURE 4 Salt-dependent SEC-SAXS of PNt variants. (a) The R_g and (b) solvent quality, ν, for PNt are obtained from SAXS measurements collected across a temperature range from 10 to 60°C in 0, 150, and 500 mM KCl. The temperature dependence of the buffers is similar across the different KCl concentrations, although the low ionic buffer becomes a poor solvent above 30°C ($\nu < 0.5$). (c) The R_g and (d) ν values of WT PNt and shuffled mutants, Swap5 and Swap6 (as described in Bowman et al., 2020), show a non-monotonic dependence on $[KCl]^{1/2}$. The baseline buffer for the PNt variants was 20 mM HEPES, pH 7.5, whereas for red-sfAFP, the baseline buffer was 20 mM sodium phosphate, 2 mM DTT, pH 7.5. The κ values are listed for each variant showing the degree of charge partitioning; see text (Holehouse et al., 2017). For red-sfAFP, the fraction of charged residues is low (0.07), hence its κ value is potentially unreliable. Fitting error bars for R_g and ν are present for all points but are often smaller than the symbols. PNt, N-terminal domain of pertactin; SEC-SAXS, size-exclusion chromatography-small-angle x-ray scattering.

and ν for PNt (47.07 ± 0.12 Å and 0.507 ± 0.002, respectively) at 10°C relative to the 150 mM KCl condition (Figure 4a,b). In this low (25 mM) ionic strength buffer, changing the temperature produces a similar reduction in R_g and ν as seen with 150 mM KCl. Specifically, at 58°C, R_g decreased by ~11% to 41.83 ± 0.21 Å, and ν decreased by 8% to 0.465 ± 0.004. The low ionic buffer becomes a poor solvent at temperatures above 30°C ($\nu = 0.495 \pm 0.003$). In 500 mM KCl, R_g and ν exhibit the same temperature dependence as in 150 mM KCl, suggesting that temperature and salt effects are effectively additive in this range of conditions.

To complement the temperature dependence of ν in different salt conditions, we measured R_g and ν of PNt at 20°C at six different salt concentrations: 0, 0.05, 0.15, 0.3, 0.5, and 1 M KCl (Figure 4c,d). From 0 to 0.15 M KCl, solvent quality increases, yet decreases above 0.5 M. Specifically, 20 mM HEPES alone at pH 7.5 leads to a conformational ensemble for PNt that is just below the θ-solvent limit ($\nu = 0.489 \pm 0.002$). However, the

addition of even modest amounts of KCl led to good solvent quality ($\nu = 0.524 \pm 0.002$ and 0.542 ± 0.001 in 50 mM and 150 mM KCl, respectively). This solvent quality is maintained up to 500 mM KCl, but increasing KCl further ultimately leads to a marginally poor solvent ($\nu = 0.486 \pm 0.011$ at 1 M KCl). These findings are consistent with previous observations on changes to the dimensions of IDPs in response to low and high ionic strength (Maity et al., 2022).

We also measured the effect of KCl titration on the dimensions of red-sfAFP, and two sequence-reordered variants of PNt, "Swap5" and "Swap6" where the positions of charged residues were reordered to alter charge segregation (Bowman et al., 2020), as quantified using the κ parameter, where lower κ values are associated with more well-mixed charges and therefore more expanded ensembles (Das and Pappu, 2013; Holehouse et al., 2017) (Figure 4c,d). The Swap5 and Swap6 sequences have very different charge-segregation patterns (and κ values) while maintaining the same amino acid composition as the WT sequence. Swap5 was constructed by swapping charged residues so that nearly all negative charges (Asp, Glu) are N-terminal of the positively charged ones (Arg, Lys), whereas charged residues exchanged positions in Swap6 so that the charge pattern alternates sign. For WT PNt, $\kappa = 0.222$, whereas $\kappa = 0.137$ and 0.291 for Swap6 and Swap5, respectively. The range of KCl concentration over which water is a good solvent varies for the three versions of PNt and is consistent with the relative values of κ.

Swap6 is a very well charge-mixed variant of the PNt sequence and is the least affected by changes in KCl concentration. Even in the absence of KCl, $\nu = 0.551 \pm 0.003$ for Swap6, and this solvent quality is maintained over a large range, dropping only to 0.513 ± 0.010 in 1 M KCl. In contrast to Swap6, the effect of KCl on the dimensions of Swap5 is very pronounced. In a low ionic buffer (20 mM HEPES buffer lacking KCl, pH 7.5), Swap5 is collapsed with $R_g = 33.02 \pm 1.15$ Å and $\nu = 0.338 \pm 0.018$. At 150 mM KCl, solvent quality dips under the θ-solvent limit for Swap5, with $R_g = 48.71 \pm 0.34$ Å and $\nu = 0.498 \pm 0.005$. WT PNt has an intermediate κ value relative to Swap5 and Swap6, and its R_g and ν likewise show an intermediary response to KCl.

For comparison, the dimensions of red-sfAFP, which has low charge content (7%) as well as low hydrophobicity, are nearly independent of ionic strength up to 1 M KCl, although ν slightly decreases at this high ionic condition ($\nu = 0.524 \pm 0.007$). These results are consistent with previous work, indicating that the effect of ionic strength at lower salt concentrations is primarily on the interactions between charged residues rather than the backbone, while chain contraction at higher ionic strengths can be attributed to exclusion of salt from hydrophobic surfaces (Sukenik et al., 2014; Vancraenenbroeck et al., 2019).

2.8 | Changes in surface area and denaturant *m*-values

To investigate the influence of changes in solvent quality on the physical properties of DSEs and stability, we ran simulations with various levels of side chain–side chain interactions to produce a range of values in ν, as was done to create an MFF (Riback et al., 2017). Specifically, we calculated the dependence of solvent accessible surface area (SASA) of PNt ensembles on ν (Figure S7 in Data S1). The SASA of DSE_{Native} is 10% less than that of DSE_{chem} at 8 M GdmCl. This difference in SASA translates to a decrease in the denaturant dependence of ΔG_{eq} (i.e., the *m*-value) from 2.25 to 1.81 kcal mol^{-1} M^{-1}. Of note, half of the decrease in the *m*-value occurs below 1 M GdmCl. As a result, the net effect on ΔG_{eq} is small, that is, only about 6% at 0 M, which may escape experimental identification. In summary, the structural effects on the DSE due to changes in solvent condition have a minimal effect on stability measurements.

3 | DISCUSSION

We used two distinctly different IDPs, one having 39% hydrophobic residues (PNt), and the other having only 6% hydrophobic residues but with 46% glycine (red-sfAFP), to test hypotheses regarding the molecular basis of the effects of the temperature, ionic strength, and sequence variation on the properties of DSEs. Consistent with the hydrophobic effect becoming stronger at higher temperatures, PNt continually contracts up to 60°C, our highest measured temperature, corroborating results for other proteins (Das and Pappu, 2013; Stenzoski et al., 2018; Wuttke et al., 2014; Zerze et al., 2015). The temperature dependence of the DSE for CTL9 was characterized by CD and SAXS using a destabilizing mutation although there remained a 10%–50% native population that had to be accounted for in the analysis (Stenzoski et al., 2018). Interestingly the R_g of CTL9 contracts significantly (~18%) from 10 to 30°C, which is greater than the 6% observed for PNt. Additionally, while PNt continues to contract, the R_g for the DSE of CTL9 remains invariant above 30°C (a trend also observed by FRET in denatured proteins but not in authentic, highly charged IDPs [Aznauryan et al., 2013; Nettels et al., 2009; Wuttke et al., 2014]). The origin of these differences for CTL9 and PNt is unknown as the chains have similar

hydrophobic content (41% and 39%, respectively) and patterning (Bowman et al., 2020).

A previous study by Wuttke et al. using FRET also reported temperature-induced compaction in a set of IDPs that varied in hydrophobicity, but with hydrophilic chains exhibiting the greatest contraction with temperature (Wuttke et al., 2014). Using SAXS, we find that the solvent quality for red-sfAFP, a peptide backbone analog, remains in the good solvent regime ($\nu \sim 0.54$) up to 60°C. Whereas red-sfAFP is quite soluble, it contains many fewer charged residues (7%) compared to the polypeptide chains studied by Wuttke et al. (25%–56% charged residues) (Wuttke et al., 2014), which may reconcile the lack of collapse in red-sfAFP with the compaction observed in the FRET study. Regardless, the temperature invariance of red-sfAFP and the rapid backbone HX support a model where the contraction of PNt at higher temperatures results from stronger interactions between hydrophobic side chains rather than interactions involving the backbone.

Water has been thought to be a poor solvent for the protein backbone and glycine-rich chains (Holehouse et al., 2015; Teufel et al., 2011). Yet, despite being glycine-rich, the red-sfAFP and other glycine-rich variants are expanded in a variety of buffer conditions, with no evidence of residual hydrogen bonding according to HX-NMR measurements. Earlier SAXS measurements on short poly-glycine stretches similarly identified extended conformations (Ohnishi et al., 2006). Furthermore, for their less-soluble longer glycine chains, x-ray diffraction measurements by Ohnishi et al. of the aggregates were consistent with poly-proline 2 helical structure (Ohnishi et al., 2006), similar to the structural elements of native sfAFP. We observed a contraction only for the 1x21G construct, which was also the only glycine variant to display a significant aggregation propensity. The Ohnishi et al. results suggest that the aggregation of 1x21G could be due to the glycine chain trying to form poly-proline 2 helix–helix interactions, which does not happen with the shorter glycine stretches. Even though our NMR analysis of red-sfAFP* and 1x14G did not find any significant hydrogen bond protection, it would be interesting to examine if there is a slowing in HX for the 1x21G monomer, given the contraction we observed.

Another study measured the effect of temperature on the radius of hydration, R_h, for a sequence-reversed analog of staphylococcal nuclease (Retro-nuclease) using SEC and dynamic light scattering (English et al., 2018). Whereas the SEC data indicated continual compaction with temperature, R_h was non-monotonic, with an initial increase in R_h with temperature followed by a decrease at higher temperatures. This behavior was attributed to temperature-dependent changes in the backbone dihedral angle propensities. Our simulations and SAXS measurements on systems where the hydrophobic effect should play a minimal role (e.g., PNt in 4 M GdmCl and red-sfAFP) were independent of temperature, which argues against temperature-dependent changes in dihedral angle propensities being responsible for the contraction we observed.

3.1 | Effects of buffer composition

For PNt, the increase in solvent quality at moderate ionic strength reflects small amounts of salt-shielding ionic interactions between side chains. At elevated salt concentrations, collapse is enhanced due to an exclusion of ions around hydrophobic sidechains (Pegram et al., 2010; Pegram and Record Jr., 2008; Vancraenenbroeck et al., 2019). For red-sfAFP, the solvent quality does not change at low salt concentrations, and only a modest contraction occurs at high salt concentrations. This finding is unexpected as both red-sfAFP and PNt have low charge content, although sfAFP is lower than PNt (7% vs. 15%), which may explain the different solvent quality at low salt. However, the different behaviors for the two proteins at high salt concentrations are likely due to the enhancement of the hydrophobic effect, which is more relevant for PNt than red-sfAFP.

Beyond the effects of charge *composition*, it is not altogether surprising that protein sequences with different charge *distributions* also respond differently to changes in ionic condition (Das and Pappu, 2013; Huihui et al., 2018; Maity et al., 2022; Müller-Späth et al., 2010). Using SAXS, we measured the impact of ionic strength on solvent quality for different sequences. The behavior of the PNt variants in response to KCl reflects the range of κ values of the sequences (although the actual fractional charge content of the sequence affects the specific mapping of κ to ν). Compared to our standard buffer, the dimensions of PNt-deltaPG were essentially unchanged in a more biologically relevant buffer (Bar-Even et al., 2011; Milo and Phillips, n.d.). Potentially the extra solutes, especially the proposed hydrotrope, ATP (Patel et al., 2017), have a greater effect on folded proteins than on IDPs.

3.2 | Implications for protein thermodynamics

Our SAXS data provide insight into the robustness of protein thermodynamics to the perturbant used. The thermodynamic reference state, the DSE, may be a function of the environment (buffer composition, temperature,

etc.), and, as a result, the extracted ΔG values could, in principle, depend on the mode of denaturation. For example, the DSE_{therm} at the midpoint of a temperature melt may differ from the DSE_{chem} at the chemical denaturation midpoint, for example, the DSE at 60°C versus the DSE at 4 M GdmCl. In addition, HX methods can measure stability under native conditions and report on a third pertinent ensemble, DSE_{Native}.

Nevertheless, denaturation by temperature, denaturant, and even pressure often produces the same values of ΔG as HX measurements when extrapolated to a common reference condition (Meisner and Sosnick, 2004; Roche et al., 2012; Skinner et al., 2014; Yu et al., 2016). Pace et al. found this to be the case within 1 kcal mol^{-1} for 19/20 comparisons (Huyghues-Despointes et al., 1999).

These findings argue that despite differences in their ν values, and by implication the amount of buried surface area, the DSE_{chem}, DSE_{therm}, and DSE_{Native} (and $DSE_{pressure}$) are thermodynamically equivalent and the extrapolation methods using either the Gibbs–Helmholtz equation or a linear denaturant model are reasonably accurate for comparing stabilities (although examples of multiple, distinct DSEs have been reported [Tischer et al., 2018]).

The agreement between HX and the other methods also implies that the DSE undergoes complete unfolding to facilitate the exchange of the most stable backbone amide hydrogens even under native conditions. This implication is supported by the lack of measurable HX protection for red-sfAFP.

Despite the thermodynamic equivalence of the different DSEs, their far-UV CD spectra can be different. Narayan et al. (2019) found that for 32 proteins, the ellipticity, $[\theta]$, at $\lambda = 222$ nm (often perceived as a signature of helical structure) was stronger for DSE_{therm} as compared to DSE_{chem}. Numerically, the $[\theta]_{222nm}$ value for DSE_{therm} is between -4000 and -6000 deg cm^2 $dmol^{-1}$, whereas for DSE_{chem}, $[\theta]_{222nm}$ is between -3000 and 0 deg cm^2 $dmol^{-1}$.

However, we have found the $[\theta]_{222nm}$ levels are nearly temperature-independent over a broad temperature range for DSE analogs of cytochrome c and RNase A (Qi et al., 1998; Sosnick et al., 1997), as well as $[\theta]_{218nm}$ for PNt (Figure S8 in Data S1). The similarity of the CD values for DSE_{Native} and DSE_{therm} implies that the residual CD signal need not reflect authentic, hydrogen-bonded α or β secondary structure. Consistent with this view, previous work from Kjaergaard et al. (2010) found that, despite a temperature-induced increase in $[\theta]_{222nm}$ for two IDPs, NMR measurements did not suggest any corresponding secondary structure formation in the chains. Hence, the residual CD values may reflect that the polypeptide chain backbone in DSE_{therm} and DSE_{Native} has, on average, more backbone dihedral angles in the α and β regions of the Ramachandran map compared to DSE_{chem}, which is dominated by PPII geometries and has near zero $[\theta]_{222nm}$ values (Chellgren and Creamer, 2004; Whittington et al., 2005). Intriguingly, isodichroic points are observed in some peptides (Kjaergaard et al., 2010; Shi et al., 2006; Toal et al., 2014), potentially indicating a residue-level two-state transition from the PPII to the β Ramachandran basin, rather than a gain in hydrogen-bonded secondary structure. Collectively, these results provide an explanation for how the three DSEs can be thermodynamically, but not spectroscopically, equivalent.

It is worth noting that a contrived "rigid-segment" model consisting of rigid native secondary structure elements linked by flexible residues has been put forth as an ensemble that can produce the R_0 and ν scaling observed for SARWs despite having extensive local structure (Fitzkee and Rose, 2004; Kohn et al., 2004). The implication is that the reproduction of the scaling behavior is insufficient to conclude that an ensemble is a SARW. While true, other data supports the view that our ensembles resemble SARWs and not the rigid-segment model. We observe scattering expected for a SARW whereas the contrived model would have a steeper scattering on a normalized Kratky plot at $qR_g > 3$ (Fitzkee and Rose, 2004; Receveur-Brechot and Durand, 2012). In addition, the CD spectra for DSE are that of disordered proteins rather than native-like (Narayan et al., 2019). In sum, the contrived model is not plausible for multiple reasons and emphasizes the importance of using multiple techniques.

3.3 | Concluding thoughts

While we share the view that the hydrophobic effect is the major stabilizing factor in protein folding, additional stabilizing factors are required, such as stronger hydrogen bonding or improved van der Waals packing, for a protein to be stable. As we have previously noted, the insufficiency of the hydrophobic effect to drive chain collapse may confer biological advantages (Clark et al., 2020). With a moderate hydrophobic effect in water, the level of promiscuous interactions and deleterious aggregation is reduced. Indeed, one can hypothesize that the 20 naturally occurring α-amino acids were selected in part as they provided sufficient chemical diversity with just enough affinity to support specific folding and binding without excessive misfolding and aggregation. Whereas a recent study investigating peptide libraries built with non-proteinogenic amino acids proposed that amino acid selection was based in large part on foldability (Makarov et al., 2023), our study emphasizes the importance of balancing the energetics, striking

a compromise between desirable structures and unwanted interactions. The current set of amino acids certainly is imperfect in this regard, as cells do exhibit some aggregation and misfolding (and hence produce folding chaperones as well to mitigate these issues).

Finally, our results raise questions about how thermophilic organisms cope with water being a poor solvent above 60°C. No obvious sequence differences have been found between meso- and thermophiles that would suggest a strategy to inhibit collapse of unfolded chains at high temperatures. To the contrary, thermophilic proteins tend to have a higher percentage of charged and hydrophobic residues than their mesophilic homologs (Kohli et al., 2020), which would further reduce solvent quality for the DSE. Potentially, thermophiles thrive in poor solvents by employing a higher level of chaperones, or some other factor(s) yet to be identified. Additionally, the effect of pressure on ν will be important to elucidate as many thermophiles inhabit high pressure environments (Ando et al., 2021).

4 | MATERIALS AND METHODS

4.1 | Polyasparagine

Poly-L-asparagine (5000–15,000 g/mol) was purchased from Sigma-Aldrich (P8137).

4.2 | Protein sequences

Protein sequences used in this analysis are listed in Table S1 in Data S1.

4.3 | Protein expression and purification

PNt, PNt-deltaPG, Swap5, and Swap6 were expressed and purified from inclusion bodies as previously described (Bowman et al., 2020). Following expression and lysis, inclusion bodies were isolated from the cell lysate via centrifugation and resuspended in 50 mM Tris, 6 M GdmCl, pH 7.5. GdmCl was removed by a combination of dilution and dialysis. The variants are purified using anion exchange chromatography followed by SEC and centrifugation to concentrate the final product for SEC-SAXS. Protein solubility prior to SEC-SAXS is accomplished by the re-addition of GdmCl (2 M final).

The AFP and LAF1 variants were cloned into either a pET21 expression vector with an N-terminal His8-tag coupled with a stable soluble domain and a tobacco etch virus (TEV) cleavage site or a pET28 vector with a C-terminal sequence-specific nickel-assisted cleavage (SNAC)-tag and His6-tag (Dang et al., 2019). All samples were expressed in *E. coli* BL21(DE3) that were first grown to an OD_{600nm} of 0.6 or 0.8 in M9 minimal media or LB media, respectively, before inducing protein expression with 1 mM IPTG. Cells were pelleted by centrifugation and lysed by sonication or the addition of 8 M urea. The proteins were first purified using metal ion affinity chromatography, followed by TEV cleavage when necessary. Proteins were further purified using reverse-phase chromatography coupled with an acetonitrile gradient in 0.1% TFA. The eluted samples were frozen at −80°C and then lyophilized.

4.4 | Protein synthesis

The 1x21G variant of sfAFP was produced using automated fast-flow solid-phase peptide synthesis (Hartrampf et al., 2020); 0.4 M fluorenylmethoxycarbonyl (Fmoc)-protected amino acids, 0.38 M HATU (hexafluorophosphate azabenzotriazole tetramethyl uranium), and 0.38 M PyAOP [(7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate] were dissolved in amine-free *N*,*N*-dimethylformamide (DMF). For each coupling, the desired amino acid was combined with either HATU or PyAOP and diisopropylethylamine and heated to 90°C or 60°C (cysteine and histidine only) then flowed over a bed of low-loading (0.18 mmol/g) RINK amide resin at 85°C. The growing peptide was deprotected using 20% piperidine with 1% formic acid in amine-free DMF to remove Fmoc from the N-terminus. Following synthesis, the sfAFP variant was cleaved from the resin by suspending the resin in Reagent K cleavage solution (82.5% TFA, 5% water, 5% thioanisole, 5% phenol, and 2.5% 1,2-ethanedithiol) at room temperature for 4 h with rocking. The peptide was precipitated with cold diethyl ether and centrifuged three times, then redissolved in 70% acetonitrile in water with 0.1% TFA. The sfAFP variant was purified by reverse phase chromatography on a semiprep Zorbax C3 column at 60°C with an acetonitrile gradient in 0.1% TFA (solvent A: water with 0.1% TFA, solvent B: acetonitrile with 0.1% TFA; gradient: 1% B for 2 min, linearly increase to 21% B over 20 min, linearly increase to 36% B over 78 min, linearly increase to 56% B over 20 min).

4.5 | Temperature-dependent SEC-SAXS experiments

SAXS data was collected at the BioCAT beamline at the Advanced Photon Source at Argonne National

Laboratory. A Superdex 200 gel-filtration column coupled with sheath co-flow (Kirby et al., 2016) is used to separate aggregates from monomer samples (Malaby et al., 2015; Mathew et al., 2004). The sheath co-flow buffer was identical to the running buffer for each measured condition.

The temperature is controlled at almost every stage of the sample trajectory. The running buffer reservoir sits in an incubator that can be controlled from 5 to 60°C, and the jacket on the SEC column ranges from 4 to 55°C. The PEEK tubing lines connecting the FPLC to the co-flow cell are jacketed and connected to a water bath that ranges from 10 to 60°C. The tubing that connects the co-flow pumps to the cell is also jacketed. The co-flow cell itself is connected to a water bath, and the temperature is monitored using a thermocouple which is in contact with the liquid at the entry point of the capillary.

The scattering data were processed using BioXTAS RAW (Hopkins et al., 2017; Nielsen et al., 2009) and the resulting intensity profiles were analyzed using the MFF server at sosnick.uchicago.edu/SAXSonIDPs. All the SAXS profile data analyzed in this paper are available at https://github.com/sosnicklab/SAXSonIDPs.

4.6 | SAXS profiles of Upside-generated ensembles

Ensembles of PNt in the SARW limit were generated via replica exchange simulations in *Upside* (Jumper et al., 2018a; Jumper et al., 2018b), where each amino acid is represented by three authentic backbone atoms along with an amide proton, carbonyl oxygen, and a side chain bead. The protein is modeled in its SARW form by only including Ramachandran potential energy, steric repulsions, and bond distance/angle terms in the simulations. The Ramachandran dihedral angle distributions are based on the distribution in a coil library and are both residue-type and neighbor dependent (Ting et al., 2010). Fifteen replicas were simulated across a temperature range of 0.600–1.300 Upside temperature units, or ∼208–450 K, which encompasses the experimental temperature range of 10–58°C. To expedite conformational sampling, an attempt to randomly change the Ramachandran angle of a randomly selected residue (i.e., a pivot) was made every 10 time units and evaluated using a Metropolis criterion. Temperature swaps between replicas were attempted every 10 units of time. In total, each replica was simulated for $\sim 3.5 \times 10^6$ time units each, and frames are saved every 500 time units. For each temperature, 1000 structures were randomly selected from the latter portion of the simulations (i.e., the first 4000 frames are dropped).

Following sidechain reconstruction using TreePack (Xu, 2005; Xu and Berger, 2006), the SAXS profiles were calculated for the randomly selected ensembles using FoXS (Schneidman-Duhovny et al., 2013; Schneidman-Duhovny et al., 2016). Specifically, the scattering intensity was calculated at 100 values of q ranging from 0.004 to 0.4 with the parameters c1 and c2 being fixed to 1.0 and 2.0, respectively. The predicted SAXS profiles at each temperature were averaged together and analyzed using the MFF server.

4.7 | NMR experiments

NMR spectra were acquired either on a Bruker AVANCE III 500 MHz or an AVANCE IIIHD 600 MHz NMR spectrometer equipped with a room temperature TXI probe. ^{15}N-TOCSY-HSQC (Marion et al., 1989) measurements of 500 μM sfAFP-biHis were carried out in 50 mM sodium phosphate, 1 mM TCEP, pH 6.4 at 5°C. Series of ^{1}H^{15}N CLEANEX-PM (Hwang et al., 1998) experiments measured the rate of H_N exchange with water. Exchange mixing times were varied from 0 to 200 ms. Typical acquisition parameters at 500 MHz were ^{1}H 90° time 9 μs, ^{15}N 37 μs, ^{1}H sweep width 12 ppm, ^{15}N sweep width 25 ppm, ^{1}H center frequency 4.695 ppm, ^{15}N center frequency 117.5 ppm, and a recycle delay of 1 s. 2D data arrays were 1024 complex points in ^{1}H by 128 complex points in ^{15}N.

All data were processed in NMRFx Analyst (Norris et al., 2016), with cosine apodization, in both dimensions, linear prediction in the ^{15}N dimension, and zero filling once in both dimensions, followed by Fourier transformation and phasing. The resulting spectra were integrated in NMRViewJ (Johnson and Blevins, 1994) using the rate analysis routine. The integrated time series for each resolved peak volume fraction V/V_0 were fit using the standard function (Hwang et al., 1998):

$$\frac{V(\tau_m)}{V_0} = \frac{k}{R_{1A,app} + k - R_{1B,app}} \left(e^{-R_{1B,app}\tau_m} - e^{-\left(R_{1A,app}+k\right)\tau_m} \right),$$

where k is the effective pseudo-first order rate of exchange between water and the amide hydrogen, $R_{1A,app}$ and $R_{1B,app}$ are the apparent relaxation rates of the amide and water, respectively, and V_0 is the reference peak volume of the amide group.

AUTHOR CONTRIBUTIONS

Michael Baxa: Conceptualization; writing – review and editing; writing – original draft; formal analysis; visualization; data curation; project administration; investigation. **Xiaoxuan Lin:** Investigation; writing – review and editing; visualization. **Cedrick D. Mukinay:** Investigation; writing – review and editing; visualization. **Srinivas**

Chakravarthy: Conceptualization; investigation; data curation; methodology; writing – review and editing. **Joseph R. Sachleben:** Methodology; investigation; formal analysis; data curation; writing – review and editing. **Sarah Antilla:** Resources; writing – review and editing. **Nina Hartrampf:** Resources; writing – review and editing. **Joshua A. Riback:** Conceptualization; formal analysis; data curation; writing – review and editing. **Isabelle A. Gagnon:** Resources; writing – review and editing. **Bradley Pentelute:** Writing – review and editing. **Patricia L. Clark:** Writing – review and editing; conceptualization; supervision; writing – original draft. **Tobin R. Sosnick:** Supervision; conceptualization; writing – original draft; writing – review and editing; methodology.

ACKNOWLEDGMENTS

We thank Gina Jozwiak for assistance with protein production and purification and Qing Luan for constructing the PNt-deltaPG expression plasmid. We thank members of our groups, S. Sukenik, and A. Holehouse for stimulating discussions and input. The lfAFP plasmid was generously provided by Peter Davies.

FUNDING INFORMATION

This work was supported by grants from the National Institutes of Health (GM055694 [TRS], GM148233 [TRS], GM130122 [TRS, PLC], DP1 GM146256 [PLC], GM103622, and 1S10OD018090-01 [T.C. Irving]) and the W. M. Keck Foundation (PLC). Use of the Advanced Photon Source, an Office of Science User Facility, operated for the Department of Energy Office of Science by Argonne National Laboratory, was supported by the U.S. Department of Energy (DOE) under Contract DEAC02-06CH11357.

ORCID

Michael C. Baxa https://orcid.org/0000-0001-8764-1262
Xiaoxuan Lin https://orcid.org/0000-0001-5356-9135
Patricia L. Clark https://orcid.org/0000-0001-5462-8248
Tobin R. Sosnick https://orcid.org/0000-0002-2871-7244

REFERENCES

Ando N, Barquera B, Bartlett DH, Boyd E, Burnim AA, Byer AS, et al. The molecular basis for life in extreme environments. Annu Rev Biophys. 2021;50(1):343–72. https://doi.org/10.1146/annurev-biophys-100120-072804

Asthagiri D, Karandur D, Tomar DS, Pettitt BM. Intramolecular interactions overcome hydration to drive the collapse transition of Gly_{15}. J Phys Chem B. 2017;121(34):8078–84. https://doi.org/10.1021/acs.jpcb.7b05469

Aznauryan M, Nettels D, Holla A, Hofmann H, Schuler B. Single-molecule spectroscopy of cold denaturation and the temperature-induced collapse of unfolded proteins. J Am Chem Soc. 2013;135(38):14040–3. https://doi.org/10.1021/ja407009w

Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins Struct Funct Genet. 1993;17(1):75–86. https://doi.org/10.1002/prot.340170110

Bar-Even A, Noor E, Flamholz A, Buescher JM, Milo R. Hydrophobicity and charge shape cellular metabolite concentrations. PLoS Comput Biol. 2011;7(10):e1002166. https://doi.org/10.1371/journal.pcbi.1002166

Bowman MA, Riback JA, Rodriguez A, Guo H, Li J, Sosnick TR, et al. Properties of protein unfolded states suggest broad selection for expanded conformational ensembles. Proc Natl Acad Sci. 2020;117(38):23356–64. https://doi.org/10.1073/pnas.2003773117

Chellgren BW, Creamer TP. Short sequences of non-proline residues can adopt the polyproline II helical conformation. Biochemistry. 2004;43(19):5864–9. https://doi.org/10.1021/bi049922v

Clark PL, Plaxco KW, Sosnick TR. Water as a good solvent for unfolded proteins: folding and collapse are fundamentally different. J Mol Biol. 2020;432(9):2882–9. https://doi.org/10.1016/j.jmb.2020.01.031

Connelly GP, Bai Y, Jeng M-F, Englander SW. Isotope effects in peptide group hydrogen exchange. Proteins Struct Funct Genet. 1993;17(1):87–92. https://doi.org/10.1002/prot.340170111

Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. Proc Natl Acad Sci. 2006;103(45):16764–9. https://doi.org/10.1073/pnas.0608175103

Dang B, Mravic M, Hu H, Schmidt N, Mensa B, DeGrado WF. SNAC-tag for sequence-specific chemical protein cleavage. Nat Methods. 2019;16(4):319–22. https://doi.org/10.1038/s41592-019-0357-3

Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci. 2013;110(33):13392–7. https://doi.org/10.1073/pnas.1304749110

Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102–12. https://doi.org/10.1016/j.sbi.2015.03.008

English LR, Tischer A, Demeler AK, Demeler B, Whitten ST. Sequence reversal prevents chain collapse and yields heat-sensitive intrinsic disorder. Biophys J. 2018;115(2):328–40. https://doi.org/10.1016/j.bpj.2018.06.006

Fitzkee NC, Rose GD. Reassessing random-coil statistics in unfolded proteins. Proc Natl Acad Sci. 2004;101(34):12497–502. https://doi.org/10.1073/pnas.0404236101

Flory PJ. Principles of polymer chemistry. Ithaca, New York: Cornell University Press; 1953.

Gates ZP, Baxa MC, Yu W, Riback JA, Li H, Roux B, et al. Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. Proc Natl Acad Sci USA. 2017;114(9):2241–6. https://doi.org/10.1073/pnas.1609579114

Hartrampf N, Saebi A, Poskus M, Gates ZP, Callahan AJ, Cowfer AE, et al. Synthesis of proteins by automated flow chemistry. Science. 2020;368(6494):980–7. https://doi.org/10.1126/science.abb2491

Hofmann H, Golbik RP, Ott M, Hübner CG, Ulbrich-Hofmann R. Coulomb forces control the density of the collapsed unfolded

state of Barstar. J Mol Biol. 2008;376(2):597–605. https://doi.org/10.1016/j.jmb.2007.11.083

Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc Natl Acad Sci USA. 2012;109(40):16155–60. https://doi.org/10.1073/pnas.1207719109

Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. Biophys J. 2017;112(1):16–21. https://doi.org/10.1016/j.bpj.2016.11.3200

Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. J Am Chem Soc. 2015;137(8):2984–95. https://doi.org/10.1021/ja512062h

Hopkins JB, Gillilan RE, Skou S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J Appl Crystallogr. 2017;50(5):1545–53. https://doi.org/10.1107/S1600576717011438

Huihui J, Firman T, Ghosh K. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. J Chem Phys. 2018;149(8):085101. https://doi.org/10.1063/1.5037727

Huyghues-Despointes BMP, Scholtz JM, Pace CN. Protein conformational stabilities can be determined from hydrogen exchange rates. Nat Struct Biol. 1999;6(10):910–2.

Hwang T-L, van Zijl PCM, Mori S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a fast-HSQC (FHSQC) detection scheme. J Biomol NMR. 1998;11(2):221–6. https://doi.org/10.1023/A:1008276004875

Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. J Mol Biol. 2007;367(3):609–15. https://doi.org/10.1016/j.jmb.2007.01.012

Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR. Early collapse is not an obligate step in protein folding. J Mol Biol. 2004;338(2):369–82. https://doi.org/10.1016/j.jmb.2004.02.065

Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. Proc Natl Acad Sci. 2005;102(37):13099–104. https://doi.org/10.1073/pnas.0506078102

Johnson BA, Blevins RA. NMR view: a computer program for the visualization and analysis of NMR data. J Biomol NMR. 1994;4(5):603–14. https://doi.org/10.1007/BF00404272

Jumper JM, Faruk NF, Freed KF, Sosnick TR. Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. PLoS Comput Biol. 2018a;14(12):e1006342. https://doi.org/10.1371/journal.pcbi.1006342

Jumper JM, Faruk NF, Freed KF, Sosnick TR. Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. PLoS Comput Biol. 2018b;14(12):e1006578. https://doi.org/10.1371/journal.pcbi.1006578

Karandur D, Harris RC, Pettitt BM. Protein collapse driven against solvation free energy without H-bonds. Protein Sci. 2016;25(1):103–10. https://doi.org/10.1002/pro.2749

Karandur D, Wong K-Y, Pettitt BM. Solubility and aggregation of Gly_5 in water. J Phys Chem B. 2014;118(32):9565–72. https://doi.org/10.1021/jp503358n

Kirby N, Cowieson N, Hawley AM, Mudie ST, McGillivray DJ, Kusel M, et al. Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. Acta Crystallogr D Struct Biol. 2016;72(12):1254–66. https://doi.org/10.1107/S2059798316017174

Kjaergaard M, Nørholm A, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB. Temperature-dependent structural changes in intrinsically disordered proteins: formation of α-helices or loss of polyproline II? Protein Sci. 2010;19(8):1555–64. https://doi.org/10.1002/pro.435

Kohli I, Joshi NC, Mohapatra S, Varma A. Extremophile – an adaptive strategy for extreme conditions and applications. Curr Genomics. 2020;21(2):96–110. https://doi.org/10.2174/1389202921666200401105908

Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, et al. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci. 2004;101(34):12491–6. https://doi.org/10.1073/pnas.0403643101

Langridge TD, Tarver MJ, Whitten ST. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. Proteins Struct Funct Genet. 2014;82(4):668–78. https://doi.org/10.1002/prot.24449

Maity H, Baidya L, Reddy G. Salt-induced transitions in the conformational ensembles of intrinsically disordered proteins. J Phys Chem B. 2022;126(32):5959–71. https://doi.org/10.1021/acs.jpcb.2c03476

Makarov M, Sanchez Rocha AC, Krystufek R, Cherepashuk I, Dzmitruk V, Charnavets T, et al. Early selection of the amino acid alphabet was adaptively shaped by biophysical constraints of foldability. J Am Chem Soc. 2023;145:5320–9. https://doi.org/10.1021/jacs.2c12987

Malaby AW, Chakravarthy S, Irving TC, Kathuria SV, Bilsel O, Lambright DG. Methods for analysis of size-exclusion chromatography–small-angle X-ray scattering and reconstruction of protein scattering. J Appl Crystallogr. 2015;48(4):1102–13. https://doi.org/10.1107/S1600576715010420

Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci. 2010;107(18):8183–8. https://doi.org/10.1073/pnas.0911107107

Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, et al. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry. 1989;28(15):6150–6. https://doi.org/10.1021/bi00441a004

Mathew E, Mirza A, Menhart N. Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. J Synchrotron Radiat. 2004;11(4):314–8. https://doi.org/10.1107/S0909049504014086

Meisner WK, Sosnick TR. Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: implications for downhill folding. Proc Natl Acad Sci USA. 2004;101(44):15639–44.

Milo R, Phillips R. Cell biology by the numbers. n.d. Available from: http://book.bionumbers.org/

Mok Y-F, Lin F-H, Graham LA, Celik Y, Braslavsky I, Davies PL. Structural basis for the superior activity of the large isoform of snow flea antifreeze protein. Biochemistry. 2010;49(11):2593–603. https://doi.org/10.1021/bi901929n

Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, et al. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci. 2010;107(33):14609–14. https://doi.org/10.1073/pnas.1001743107

Myers JK, Pace CN, Scholtz JM. Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci. 1995;4(10):2138–48. https://doi.org/10.1002/pro.5560041020

Narayan A, Bhattacharjee K, Naganathan AN. Thermally versus chemically denatured protein states. Biochemistry. 2019;58(21):2519–23. https://doi.org/10.1021/acs.biochem.9b00089

Nettels D, Müller-Späth S, Küster F, Hofmann H, Haenni D, Rüegger S, et al. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc Natl Acad Sci. 2009;106(49):20740–5. https://doi.org/10.1073/pnas.0900622106

Nguyen D, Mayne L, Phillips MC, Walter Englander S. Reference parameters for protein hydrogen exchange rates. J Am Soc Mass Spectrom. 2018;29(9):1936–9. https://doi.org/10.1007/s13361-018-2021-z

Nielsen SS, Toft KN, Snakenborg D, Jeppesen MG, Jacobsen JK, Vestergaard B, et al. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. J Appl Crystallogr. 2009;42(5):959–64. https://doi.org/10.1107/S0021889809023863

Norris M, Fetler B, Marchant J, Johnson BA. NMRFx processor: a cross-platform NMR data processing program. J Biomol NMR. 2016;65(3):205–16. https://doi.org/10.1007/s10858-016-0049-6

Ohnishi S, Kamikubo H, Onitsuka M, Kataoka M, Shortle D. Conformational preference of Polyglycine in solution to elongated structure. J Am Chem Soc. 2006;128(50):16338–44. https://doi.org/10.1021/ja066008b

Patel A, Malinovska L, Saha S, Wang J, Alberti S, Krishnan Y, et al. ATP as a biological hydrotrope. Science. 2017;356(6339):753–6. https://doi.org/10.1126/science.aaf6846

Pegram LM, Record MT Jr. Thermodynamic origin of Hofmeister ion effects. J Phys Chem B. 2008;112(31):9428–36. https://doi.org/10.1021/jp800816a

Pegram LM, Wendorff T, Erdmann R, Shkel I, Bellissimo D, Felitsky DJ, et al. Why Hofmeister effects of many salts favor protein folding but not DNA helix formation. Proc Natl Acad Sci. 2010;107(17):7716–21. https://doi.org/10.1073/pnas.0913376107

Qi PX, Sosnick TR, Englander SW. The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. Nat Struct Biol. 1998;5(10):882–4.

Receveur-Brechot V, Durand D. How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci. 2012;13(1):55–75. https://doi.org/10.2174/138920312799277901

Renn JP, Junker M, Besingi RN, Braselmann E, Clark PL. ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. Chem Biol. 2012;19(2):287–96. https://doi.org/10.1016/j.chembiol.2011.11.009

Riback JA, Bowman MA, Zmyslowski AM, Knoverek CR, Jumper JM, Hinshaw JR, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science. 2017;358(6360):238–41. https://doi.org/10.1126/science.aan5774

Riback JA, Bowman MA, Zmyslowski AM, Plaxco KW, Clark PL, Sosnick TR. Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. Proc Natl Acad Sci. 2019;116(18):8889–94. https://doi.org/10.1073/pnas.1813038116

Roche J, Dellarole M, Caro JA, Guca E, Norberto DR, Yang Y, et al. Remodeling of the folding free energy landscape of staphylococcal nuclease by cavity-creating mutations. Biochemistry. 2012;51(47):9535–46. https://doi.org/10.1021/bi301071z

Sadqi M, Lapidus LJ, Muñoz V. How fast is protein hydrophobic collapse? Proc Natl Acad Sci. 2003;100(21):12117–22. https://doi.org/10.1073/pnas.2033863100

Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J. 2013;105(4):962–74. https://doi.org/10.1016/j.bpj.2013.07.020

Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. FoXS, FoXSDock and MultiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. 2016;44(W1):W424–9. https://doi.org/10.1093/nar/gkw389

Shi Z, Chen K, Liu Z, Sosnick TR, Kallenbach NR. PII structure in the model peptides for unfolded proteins: studies on ubiquitin fragments and several alanine-rich peptides containing QQQ, SSS, FFF, and VVV. Proteins Struct Funct Genet. 2006;63(2):312–21. https://doi.org/10.1002/prot.20788

Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a sequence of seven alanine residues. Proc Natl Acad Sci. 2002;99(14):9190–5. https://doi.org/10.1073/pnas.112193999

Singh VR, Lapidus LJ. The intrinsic stiffness of Polyglutamine peptides. J Phys Chem B. 2008;112(42):13172–6. https://doi.org/10.1021/jp805636p

Skinner JJ, Yu W, Gichana EK, Baxa MC, Hinshaw JR, Freed KF, et al. Benchmarking all-atom simulations using hydrogen exchange. Proc Natl Acad Sci USA. 2014;111(45):15975–80. https://doi.org/10.1073/pnas.1404213111

Sneppen K, Zocchi G. Physics in molecular biology. Cambridge, UK; New York: Cambridge University Press; 2005.

Sosnick TR, Shtilerman MD, Mayne L, Englander SW. Ultrafast signals in protein folding and the polypeptide contracted state. Proc Natl Acad Sci USA. 1997;94(16):8545–50.

Stenzoski NE, Luan B, Holehouse AS, Raleigh DP. The unfolded state of the C-terminal domain of L9 expands at low but not at elevated temperatures. Biophys J. 2018;115(4):655–63. https://doi.org/10.1016/j.bpj.2018.07.013

Sukenik S, Boyarski Y, Harries D. Effect of salt on the formation of salt-bridges in β-hairpin peptides. Chem Commun. 2014;50(60):8193–6. https://doi.org/10.1039/C4CC03195D

Teufel DP, Johnson CM, Lum JK, Neuweiler H. Backbone-driven collapse in unfolded protein chains. J Mol Biol. 2011;409(2):250–62. https://doi.org/10.1016/j.jmb.2011.03.066

Ting D, Wang G, Shapovalov M, Mitra R, Jordan MI, Jr RLD. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. PLoS Comput Biol. 2010;6(4):e1000763. https://doi.org/10.1371/journal.pcbi.1000763

Tischer A, Machha VR, Rösgen J, Auton M. "Cooperative collapse" of the denatured state revealed through Clausius-Clapeyron analysis of protein denaturation phase diagrams. Biopolymers. 2018;109(8):e23106. https://doi.org/10.1002/bip.23106

Toal SE, Verbaro DJ, Schweitzer-Stenner R. Role of enthalpy–entropy compensation interactions in determining the conformational propensities of amino acid residues in unfolded peptides. J Phys Chem B. 2014;118(5):1309–18. https://doi.org/10.1021/jp500181d

Tran HT, Mao A, Pappu RV. Role of backbone−solvent interactions in determining conformational equilibria of intrinsically disordered proteins. J Am Chem Soc. 2008;130(23):7380–92. https://doi.org/10.1021/ja710446s

Uversky VN. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. Protein J. 2009;28(7):305–25. https://doi.org/10.1007/s10930-009-9201-4

Vancraenenbroeck R, Harel YS, Zheng W, Hofmann H. Polymer effects modulate binding affinities in disordered proteins. Proc Natl Acad Sci. 2019;116(39):19506–12. https://doi.org/10.1073/pnas.1904997116

Vitalis A, Wang X, Pappu RV. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. J Mol Biol. 2008;384(1):279–97. https://doi.org/10.1016/j.jmb.2008.09.026

Walters RH, Murphy RM. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. J Mol Biol. 2009;393(4):978–92. https://doi.org/10.1016/j.jmb.2009.08.034

Whittington SJ, Chellgren BW, Hermann VM, Creamer TP. Urea promotes polyproline II helix formation: implications for protein denatured states. Biochemistry. 2005;44(16):6269–75. https://doi.org/10.1021/bi050124u

Wiggers F, Wohl S, Dubovetskyi A, Rosenblum G, Zheng W, Hofmann H. Diffusion of a disordered protein on its folded ligand. Proc Natl Acad Sci. 2021;118(37):e2106690118. https://doi.org/10.1073/pnas.2106690118

Wuttke R, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, et al. Temperature-dependent solvation modulates the dimensions of disordered proteins. Proc Natl Acad Sci. 2014;111(14):5213–8. https://doi.org/10.1073/pnas.1313006111

Xu J. Rapid protein side-chain packing via tree decomposition. In: Miyano S, Mesirov J, Kasif S, Istrail S, Pevzner PA, Waterman M, editors. Research in computational molecular biology. Berlin, Heidelberg: Springer; 2005. p. 423–39. https://doi.org/10.1007/11415770_32

Xu J, Berger B. Fast and accurate algorithms for protein side-chain packing. J ACM. 2006;53(4):533–57. https://doi.org/10.1145/1162349.1162350

Yoo TY, Meisburger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, et al. Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. J Mol Biol. 2012;418(3):226–36. https://doi.org/10.1016/j.jmb.2012.01.016

Yu W, Baxa MC, Gagnon I, Freed KF, Sosnick TR. Cooperative folding near the downhill limit determined with amino acid resolution by hydrogen exchange. Proc Natl Acad Sci USA. 2016;27078098:4747–52. https://doi.org/10.1073/pnas.1522500113

Zerze GH, Best RB, Mittal J. Sequence- and temperature-dependent properties of unfolded and disordered proteins from atomistic simulations. J Phys Chem B. 2015;119(46):14622–30. https://doi.org/10.1021/acs.jpcb.5b08619

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Baxa MC, Lin X, Mukinay CD, Chakravarthy S, Sachleben JR, Antilla S, et al. How hydrophobicity, side chains, and salt affect the dimensions of disordered proteins. Protein Science. 2024;33(5):e4986. https://doi.org/10.1002/pro.4986