



Published in final edited form as:

Nat Mater. 2015 November ; 14(11): 1164–1171. doi:10.1038/nmat4418.

Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers

Felipe García Quiroz^{1,2} and Ashutosh Chilkoti¹

¹Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA

Abstract

Proteins and synthetic polymers that undergo aqueous phase transitions mediate self-assembly in nature and in man-made material systems. Yet little is known about how the phase behaviour of a protein is encoded in its amino acid sequence. Here, by synthesizing intrinsically disordered, repeat proteins to test motifs that we hypothesized would encode phase behaviour, we show that the proteins can be designed to exhibit tunable lower or upper critical solution temperature (LCST and UCST, respectively) transitions in physiological solutions. We also show that mutation of key residues at the repeat level abolishes phase behaviour or encodes an orthogonal transition. Furthermore, we provide heuristics to identify, at the proteome level, proteins that might exhibit phase behaviour and to design novel protein polymers consisting of biologically active peptide repeats that exhibit LCST or UCST transitions. These findings set the foundation for the prediction and encoding of phase behaviour at the sequence level.

Introduction

Proteins that undergo phase transitions, as the result of a stimulus-triggered change in water solubility, mediate important self-assembly events in nature^{1–5}. Prototypical examples range from elastin and collagen fibers that provide mechanical integrity to the extracellular matrix, intracellular complexes for nucleic acid storage and processing^{2,3} and transport barriers in nuclear pore complexes⁶. Analogously, polymers that exhibit phase behaviour⁷ in water enable innovative approaches to nanoparticle self-assembly^{8,9}, cancer therapy^{10,11}, regenerative medicine^{12–14} and protein purification^{15–17}.

Despite this widespread interest phase behaviour has thus far eluded the scope of sequence-level predictions. This is in contrast to the progress in understanding sequence-structure relationships in proteins to predict folding¹⁸ and intrinsic disorder¹⁹ by computational approaches. Part of the problem in developing a molecular understanding of phase behaviour

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence to: chilkoti@duke.edu.

²Present address: Howard Hughes Medical Institute, Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, New York 10065, USA.

Author contributions: F.G.Q designed and performed experiments, analyzed data and wrote the manuscript. A.C. analyzed data and wrote the manuscript.

Competing financial interests: The authors hold US patent No. 8,470,967 that covers many of the peptide sequences described in this manuscript.

by computational approaches is that it is a multi-body, cooperative phenomenon that is computationally intractable for current all-atom models with explicit solvation²⁰. While coarse-grained models can handle multi-body interactions, they fail to fully capture the molecular complexity of protein-protein and protein-water interactions to be of broad relevance to the study of this problem. These limitations have severely hampered a clear understanding of the sequence determinants of phase behaviour in proteins.

There are two types of soluble to insoluble phase transitions that are of interest: the first, lower critical solution temperature (LCST) transition, occurs upon heating above a critical solution temperature, while the second, upper critical solution temperature (UCST) transition, occurs upon cooling below a critical temperature. To understand the sequence determinants of these orthogonal phase transitions in proteins, herein we examine intrinsically disordered proteins (IDPs) with a repetitive polymer-like architecture as a model system to readily link their amino acid sequence, which we specify at the repeat level, to their phase behaviour in aqueous solution. The relationships between sequence and phase behaviour that emerge from this approach provide a set of heuristics to encode LCST or UCST phase behaviour in protein polymers as well as identify proteins that may exhibit phase behaviour.

Compositional analysis of Pro and Gly-rich IDPs

Because tropoelastin, collagen and resilin are Pro- and Gly-rich IDPs that exhibit temperature-triggered phase transitions^{21,22}, we began our search from this known point in the sequence–phase behaviour landscape by focusing on IDPs with a similarly high content of these two well-known structure-breaking residues²³ (Supplementary Fig. 1). Our initial goal was two-fold: first, to identify patterns of Pro and Gly in these proteins that would provide a generic —and minimal— IDP scaffold; and second, to identify candidate residues and residue interactions that when incorporated onto such a scaffold would encode phase behaviour.

We first mapped (see Supplementary Methods) Pro and Gly pairs spaced by up to 4 other amino acids (that is, P-X_n-G motifs, where *n* varies from 0 to 4) across prototypical Pro- and Gly-rich IDPs. Although a P-G dipeptide (i.e., *n*=0) is the predominant motif in these proteins (Fig. 1a), as in the canonical Val-**Pro-Gly**-Xaa-Gly motif that forms the basis of most known LCST peptide polymers, namely elastin-like polypeptides (ELPs)^{24,25}, we also identified a large fraction (30%) of recurring P-X₄-G motifs in resilin (Fig. 1a). These structure-breaking Pro/Gly pairs repeat every 4-9 residues (Supplementary Fig. 2), but the unusual high Gly content of these proteins (typically >30%), which is rare among the bulk of Pro-rich proteins (Supplementary Fig. 3), results in inter-motif segments that are also Gly-rich (Supplementary Fig. 4). To pinpoint the relevance of this excess Gly in defining a robust IDP-like scaffold, we broadened our search space to examine Pro and Gly-rich proteins with a moderate enrichment (at least 10%) for both Pro and Gly residues across several eukaryotic proteomes. Interestingly, a bias for P-G dipeptides persists at the proteome level (Fig. 1b). In contrast to prototypical Pro and Gly-rich IDPs, however, the top 9 human proteins with the most number of clustered P-X_n-G motifs from this expanded set of proteins have Pro and Gly residues arranged in a relatively uniform fashion, with *n* values

that range from 0 to 3 (Fig. 1c), and with the P-X_n-G motifs separated on average by up to 15 residues (inset of Fig. 1b and supplemental Fig. 2b) that follow a rather unbiased amino acid distribution (Supplementary Fig. 4). Because a similar picture emerges if the search is restricted to proteins that are classified as IDPs (Supplementary Fig. 4 and Supplementary Fig. 5), these observations suggested a simple heuristic to encode disorder into a protein scaffold: *the sequence space described by P-X_n-G-containing motifs (where n varies from 0 to 4 and X is any amino acid) that are separated by 3–15 residues with a promiscuous distribution of amino acids constitutes a generic scaffold for the synthesis of intrinsically disordered protein polymers (IDPPs).*

Next, to understand how LCST and UCST phase behaviour could be encoded into these IDPP scaffolds, we contrasted the amino acid composition of three known resilins with that of tropoelastins (Fig. 1d). Resilins are of interest because they exhibit intriguing dual UCST and LCST behaviour at extremes of temperature^{22,26}, whereas tropoelastins are canonical LCST exhibiting proteins. These proteins sample different regions of the hydropathy space (inset of Fig. 1d), as resilins are enriched in polar residues and tropoelastins are enriched in nonpolar residues. Resilins also exhibit two distinctive features: i) positively and negatively charged residues occur nearly at par (Arg + Lys are 49% ± 9% of all charged residues) and ii) Arg residues consistently account for the majority (88% ± 0.02) of all cationic residues (Fig. 1d). Tropoelastins, in contrast, are nearly devoid of negatively charged amino acids and are enriched for Lys as opposed to Arg. Interestingly, the UCST of rec1-resilin is undetectable at pH 10.9 upon Arg deprotonation²², suggesting a role for the nearly zwitterionic character of the protein on its UCST behaviour. Moreover, because all charged residues except Arg²⁷ are strong modulators of the hydropathy of ELPs and dramatically increase their LCST, we surmised that Arg residues are key determinants of the hydrophobicity and net charge of resilins. These observations led us to consider two additional sequence heuristics: *LCST behaviour is likely to be encoded by enriching IDPP scaffolds composed of P-X_n-G motifs, with residues (i.e., X_n and neighboring amino acids) that are largely nonpolar. In contrast, IDPP scaffolds are likely to exhibit UCST phase behaviour if populated with residues that contain zwitterionic pairs of residues, with Arg as a preferred cationic residue in these pairs.*

Sequence space of candidate LCST and UCST peptide motifs

To test the validity of these sequence heuristics, we designed a diverse set of IDPPs; these polymers explore a sequence space defined by 40 distinct peptide motifs of 5–9 residues in length that comprise the repeat units of these polymers. The motifs are characterized by a P-X_n-G unit, where *n* varies from 0 to 4 (Supplementary Fig. 6) and where X residues and neighboring amino acids were chosen to cluster into two distinct subspaces: 1) motifs that target the overall apolar hydropathy of tropoelastins to drive LCST phase behaviour, and 2) Arg-based, charged or zwitterionic motifs that approach the hydropathy of resilins to drive UCST phase behaviour (inset of Fig. 1d). We also included two Pro-devoid, Ala mutants to further study the role of Pro in these peptide motifs. Using a highly parallel method for the synthesis of DNA repeats²⁵, we then recombinantly synthesized (Supplementary Table 2 and Supplementary Table 3) 2–5 genes of various lengths for each peptide motif in this library. Subsequent expression of these genes in a T7 expression system in BL21(DE3) *E.*

coli typically resulted in purified yields (see Supplementary Methods) between 200 and 300 mg of IDPP per liter of culture without any optimization of the expression conditions.

LCST and UCST phase behaviour of IDPPs

Using temperature-dependent UV-Vis spectroscopy in PBS at a fixed polymer concentration (50 μM), we observed that IDPPs containing residues that target the overall hydrophathy of tropoelastins displayed a sharp soluble to insoluble transition upon heating above a critical temperature (LCST-type transition) (Fig. 2a). In contrast, Arg-based IDPPs with a zwitterionic repeat unit that target the hydrophathy of resilins displayed a sharp UCST-type phase transition in PBS as they became insoluble upon cooling below a range of temperatures (20–70 $^{\circ}\text{C}$) (Fig. 2b). Using circular dichroism (CD) spectroscopy at 25 $^{\circ}\text{C}$ — in their soluble state— on a subset of these polymers, we also confirmed that the P-X_n-G scaffold (for n=0–4) is disordered despite the diverse nature of the inserted residues as they all share a negative peak at ~197 nm that is characteristic of IDPs²¹ (Fig. 3a). The phase behaviour of this diverse family of IDPPs in PBS is the first demonstration that IDP scaffolds can be designed at the sequence level to undergo LCST or UCST phase behaviour over a wide range of temperatures (Fig. 2a–b). Notably, these sequences, which depart from repeat sequences found in canonical phase transition proteins, provide a diverse new family of stimulus responsive polymers with utility for a range of prospective applications.

Because the two Pro-devoid protein polymers that we synthesized exhibited UCST phase behaviour similar to Pro-containing IDPPs (Fig. 3b and Supplementary Fig. 7), we suggest that a well-mixed distribution of oppositely charged residues and Gly —as observed in our library of UCST motifs— is suitable for the design of IDPPs that exhibit phase transition behaviour. This suggestion is in agreement with the abundance of charged residues in IDPs¹⁶ and with results from recent simulations on peptides with well-mixed sequences of Glu and Lys²⁸. We do not know, however, the degree of Gly enrichment that is required to obtain a robust IDP scaffold with such a design as we are aware that short polymers composed of R-A-D-A repeats form β -sheet structures²⁹. Thus the inclusion of P-X_n-G motifs is a sufficient but not a necessary sequence constraint to design an IDPP that exhibits phase behaviour. We suggest that Pro devoid sequences that are intrinsically disordered can also serve as a scaffold for UCST IDPPs, although the exact sequence rules that lead to disorder are less transparently obvious for this class of protein polymers, and elucidating these rules is beyond the scope of this paper.

Although hydrophathy successfully clustered LCST and UCST motifs into separate categories (Fig. 3c), we observed that it failed to account for the effect of residue interactions on details of the phase behaviour of IDPPs. For instance, despite the predicted hydrophilic character of zwitterionic motifs in our library of protein polymers, the zwitterionic IDPPs that exhibit UCST phase behaviour are surprisingly hydrophobic as they are insoluble in PBS over a wide temperature range (Fig. 2b). The inability to capture the hydrophobicity of these polymers by averaging intrinsic residue hydrophathy is arguably the result of backbone and residue interactions (e.g., hydrogen bonds, salt bridges, cation- π and stacking interactions) that alter the hydration preferences for these otherwise highly polar residues^{28,30,31}. Similarly, motifs with a positive or negative net charge typically clustered

closely within the hydrophathy range of uncharged LCST or zwitterionic UCST motifs (Fig. 3c). The corresponding IDPPs, however, were soluble in PBS (pH 7.4) over the studied temperature range (gray rectangles in Fig. 2a[ii] and 2b) except when placed under conditions of high ionic strength or low pH that reduce their charged character and increase their hydrophobicity (Fig. 3d). This observation points to the responsiveness of IDPPs to multiple types of non-thermal stimuli and suggests that IDPs that do not exhibit phase behaviour under physiological conditions may exhibit pronounced phase transitions upon post-translational modifications (e.g., phosphorylation, methylation), under conditions of cellular stress or in subcellular compartments with reduced pH (e.g., lysosomes).

Tunable LCST and UCST phase transitions

Next we examined how different sequence parameters that are readily controlled at the sequence level in these genetically encoded polymers affect the LCST and UCST behaviour of IDPPs. First, we examined the effect of chain length on LCST and UCST and found that the number of repeat units (i.e., molecular weight) tuned the LCST (Fig. 4a) and UCST (Fig. 4b) of IDPPs. While this is consistent with previous studies for LCST peptide polymers³²; we note that these are the first examples of protein polymers whose UCST can be tuned by their molecular weight. Noteworthy, as few as 12 repeats of some motifs, equivalent to IDP domains of 100 residues in length, was sufficient to impart UCST phase behaviour in physiological solution conditions (Fig. 4b). Second, the overall hydrophobicity of the repeat unit modulated the UCST of IDPPs (Fig. 4c). For instance, substituting Val for the only strongly hydrophobic residue in RGDSPY₆G, Tyr, which is the second most hydrophobic natural amino acid according to Urry's hydrophathy scale²⁷, resulted in polymers without a measurable UCST in PBS (data not shown), while substituting Tyr with His — a residue that is more hydrophobic than Val, Ile and Leu when above its pK_a²⁷— led to polymers with a measurable UCST (Fig. 4c).

Concentration also modulated the UCST cloud point, as the phase transition temperature varied linearly with the logarithm of the IDPP concentration (Fig. 4d and Supplementary Fig. 8). As expected, these observed relationships between the UCST of IDPPs and their molecular weight, hydrophathy and concentration are the inverse of the behaviour commonly seen for ELP's³² and other LCST polymers²⁵. We also confirmed that the UCST phase behaviour of IDPPs is reversible over a wide range of temperatures (Fig. 4d), as shown by repeated cycles of cooling and heating. This ability to control the UCST via two orthogonal variables —sequence and chain length— that can be precisely specified by genetically encoded synthesis provides a powerful set of molecular parameters to manipulate UCST phase behaviour with a dexterity that is nearly impossible to achieve with synthetic UCST polymers, even despite recent progress in chemical synthesis of hydrogen bonding polymers with tunable UCST behaviour^{33,34,35}.

Phase behaviour of zwitterionic IDPPs

Although zwitterionic polymers (e.g., polysulfobetaines) are known to exhibit UCST behaviour in pure water, screening of charge-charge interactions in these synthetic polymers by counter ions in electrolyte solutions abrogates their UCST behaviour in a physiological

milieu³⁶. In contrast, the UCST phase behaviour exhibited by zwitterionic IDPPs is intriguing as it occurs under physiologically relevant conditions over a wide temperature range. Furthermore, because the “non-fouling” (i.e., protein resistance) and hemocompatibility of zwitterionic polymers is of technological and biomedical interest^{37,38}, this new class of UCST IDPPs may have applications that transcend their UCST phase behaviour.

Hence, we investigated the role of their zwitterionic character on the UCST phase transition by examining the phase behaviour of (GRGDSPYG)₂₀ and (GRPSDSYG)₂₀ as a function of solution pH (Fig. 4f). At pH 2.0, these zwitterionic IDPPs become cationic on protonation of Asp (D) residues (Supplementary Fig. 9). The resulting charge-charge repulsion reduced their UCST cloud point below the lowest accessible temperature (~ 2 °C) in our spectrophotometer (Fig. 4f). Upon supplementation of PBS with 0.5 M NaCl, both polymers again exhibited cloud points that approach those measured in PBS at pH 7.4. Because our UCST library contains at least one cationic motif (GRGNSPYG) that resulted in an IDPP with a measurable UCST cloud point in PBS (Fig. 2b[i]), while substituting Lys for Arg in (GRGDPSYG)₂₄ abrogated its UCST phase transition in PBS but not in water (Supplementary Fig. 10), we conclude that attractive electrostatic forces—already weakened in 140 mM NaCl—do not drive the UCST phase transition of IDPPs in PBS. Notably, the inability of Lys to drive UCST phase transitions of zwitterionic IDPPs reaffirms our hypothesis that the intriguingly poor hydration of guanidinium ions³⁹ in Arg is a major contributor to the hydrophathy of resilins and IDPPs enriched in charged residues. The dramatic decrease in the UCST upon disruption of the zwitterionic character, which is opposite to the response of rec1-resilin to low pH²², further suggests that oppositely charged residues in the repeating sequence facilitate bulk aggregation by minimizing the net charge of the polymer. In contrast, IDPPs with a large net charge may experience substantial repulsive electrostatic interactions that must be outcompeted by hydrophobic interactions. Hence, *a useful heuristic for the design of UCST IDPPs with a large fraction of charged residues (e.g., >15%), which we first recognized in the compositional biases of resilins, is that Arg must account for a large fraction (>85% based on resilins) of the positively charged residues in the polymer.*

While P-X_n-G motifs containing a zwitterionic pair (e.g., R/D and R/E) invariably led to IDPPs that exhibit UCST phase behaviour (Fig. 2b) in PBS at pH 7.4, we also consistently included an aromatic amino acid in the repeat motif as a means to tune their hydrophathy so as to impart UCST phase behaviour in a useful temperature range. To determine the sufficiency of the R/D pair in driving the observed behaviour, we designed a zwitterionic, P-X_n-G motif that incorporates the R/D pair and one other residue, Val, to match the hydrophathy of UCST IDPPs (Fig. 3c). Surprisingly, IDPPs composed of 30–50 repeats of such a motif, VPRDG (H_i = -1.3), did not show phase behaviour in PBS at pH 7.4 (Fig. 5a). Under acidic conditions, however, the UCST behaviour is observed, but depends on the number of VPRDG repeats (Fig. 5b). Moreover, mutation of either R or D to an aliphatic residue abrogated the UCST phase behaviour at all pH values and encoded orthogonal, LCST-type phase transitions (Fig. 5c). Because of these observations, while there are environmental conditions under which Arg and Asp encode UCST behaviour, we suggest

the following corollary to the sequence heuristic for UCST exhibiting IDPPs: *the incorporation of a significant fraction (>10% from IDPPs in Fig. 2b) of aromatic amino acids (e.g., Y, W, F and H) into IDPPs that are enriched for charged residues is a prerequisite to tune the UCST of the polymer to a physiologically relevant window of both temperature and ionic strength.* We believe that the molecular basis of this sequence heuristic is partly explained by the intrinsic hydrophobicity of aromatic residues. Kyte-Doolittle's scale, as used throughout this work, largely underestimates the hydrophobicity of Y, F, W and H, which is most evident when compared with Urry's scale that is derived from LCST values²⁷. However, because of the negative zeta potential (i.e., surface charge) of zwitterionic UCST IDPPs at neutral pH (Supplementary Fig. 9), we are intrigued by the possibility that aromatic residues in IDPPs engage in additional residue interactions with protonated Arg to limit its solvent exposure. This possibility is consistent with reported zeta potential data for rec1-resilin²², with mounting evidence of cation- π interactions between Arg and aromatic residues⁴⁰ and with the lack of phase behaviour of both rec1-resilin²² and UCST IDPPs at pH 12 (data not shown).

IDPPs built from biologically active peptide repeats

To utilize our ability to encode phase transition behaviour at the amino acid sequence level, we asked whether we could reprogram peptides that exhibit biological activity to also exhibit a given form of phase behaviour. In the first example, we designed a cationic, fibronectin-like polymer that consists of repeats of a 19-mer peptide wherein the GRGDSP peptide and its synergy motif (VPHSRN) from fibronectin are connected by a short Gly-rich linker that approximates the distance between these motifs in the folded protein⁴¹. As expected for a peptide motif that meets our sequence heuristics for UCST phase behaviour, this fibronectin-like IDPP exhibited UCST in PBS (Fig. 6a). In a second example, we designed a zwitterionic IDPP built from repeating a laminin-derived peptide drug (DPGYIGSR) that has been studied for over two decades for its anti-cancer activity⁴². As expected, this drug-like IDPP also showed UCST phase behaviour in PBS (Fig. 6b). Noteworthy, several of the UCST IDPPs already described are also built from biologically active peptides as they are polymers of the prototypical cell adhesion motif GRGDSP found in fibronectin^{43,44} (e.g., GRGDSPY, GRGDSPH and GRGDSPYQ in Fig. 2). We anticipate that for IDPPs built from peptide drugs, RGD motifs and other short linear motifs⁴⁵, phase behaviour may operate as a bioactivity switch. A similar mechanism could modulate the activity of IDPs in cellular environments in addition to observed conformational transitions^{28,30}. Preliminary evidence indeed suggests that the bioactivity of RGD-containing IDPPs is temperature switchable (Supplementary Fig. 11 and Supplementary text), but future work will explore this possibility in detail. We have also synthesized LCST-exhibiting IDPPs based on a short matrikine motif G-X-X-P-G⁴⁶ (Supplementary Fig. 12) and the 25–27 amino acid long bioactive peptide domains of endostatin from humans and mice (Supplementary Fig. 13 and Supplementary text). In Supplementary Fig. 14 we present additional candidate peptide hormones for the design of drug-like IDPPs.

We also explored whether LCST and UCST IDPPs could be fused to generate useful self-assembled structures with programmable behaviours. We hypothesized that a diblock copolymer consisting of a LCST block fused to an UCST block—wherein the UCST <

LCST— would self-assemble into nanoparticles at a temperature below the LCST and the UCST of the respective blocks but would disassemble at $UCST < T < LCST$. We hence synthesized a diblock copolymer composed of an RGD-containing UCST IDPP with a UCST of 60 °C in PBS, and a highly soluble LCST-exhibiting IDPP with an LCST of >80°C in PBS. Dynamic light scattering (DLS) shows that the core-forming UCST IDPP drives the formation of monodisperse particles of ~ 18 nm in radius that undergo reversible, temperature-triggered disassembly upon heating above a critical disassembly temperature (CDT) of ~19 °C (Fig. 6c). Unlike LCST-only diblock copolymers wherein the LCST of the core-forming block is a good predictor of the assembly temperature^{47,48}, here we observed that the CDT of the nanoparticles differs substantially from the UCST of the core-forming block prior to the C-terminal fusion of the LCST domain. This unexpected difference points to an intrinsic high sensitivity of UCST IDPPs to protein fusions that is absent in LCST IDPPs. Understanding this phenomenon will be important to inform the self-assembly of eukaryotic proteins that contain disordered N- or C-terminal domains fused to a ligand binding or functional domain, as in the low complexity regions of some RNA binding proteins that are reminiscent of UCST IDPPs (Supplementary Fig. 15) and in the triblock architecture of the exon1 of the huntingtin protein⁴⁹. Here, theoretical and computational advances on predicting the self-assembly of synthetic diblock copolymers may accelerate progress⁵⁰, but the interaction parameters in these models will require refinements to account for the unique properties and sequence complexity of IDPPs⁵¹. From an engineering perspective, however, because of the tunable UCST phase behaviour of IDPPs, the CDT of these nanostructures should be readily tunable to enable clinically-approved mild hyperthermia (~40–42 °C) to be used as an extrinsic trigger to drive disassembly of nanoscale drug carriers derived from these IDPPs, and hence address the issue of poor intratumoral penetration of nanocarriers^{52,53}. Interestingly, fusion of UCST and LCST motifs at the repeat level may offer a simple strategy to design IDPPs with dual LCST and UCST phase behaviour wherein the $UCST < LCST$. For instance, we suggest that rec1-resilin's dual phase behaviour stems from a peptide polymer design wherein the consensus repeat unit (GRPSDSYGAPGGN²²) fuses a UCST motif (GRPSDSYG; Fig. 2b[iii]) and a putative LCST motif (APGGN).

Prediction of phase behaviour at the proteome level

While we focused herein on the utility of these heuristics for the rational design of novel phase transition polymers that are multifunctional, these rules can also be used to identify proteins that may exploit LCST and UCST-type transitions in a biological milieu. Using 5 sequence parameters to specify a set of heuristics for UCST behaviour —Molecular weight: >100 residues, Charge content: >20%, Zwitterionic character: $\pm 5\%$, Aromatic content: >10%, and Arg-enrichment: >85%; see Supplementary text for details on threshold values —, a proteome-wide search identified 83 human proteins that exhibit sequence features reminiscent of UCST IDPPs (Supplementary Fig. 16). Among these proteins, we underline the presence of gamma-crystallin proteins that appear to coacervate in the cytoplasm of lens cells and exhibit UCST-type phase diagrams⁵⁴, RNA-binding proteins suspected to undergo cold-induced RNA regulation, and a protein (filaggrin, P20930) that is found as cytoplasmic granules in the skin and is required to establish a functional epidermal barrier⁵⁵. Because

more permissive search criteria with Lys comprising 25% of positively charged residues would increase the number of relevant candidates to 403 human proteins, these results call for a systematic study that investigates the phase behaviour of Lys-containing UCST IDPPs and sets the stage to study the potential role of phase behaviour in mediating the biological function of a growing number of human proteins in homeostasis and disease.

Outlook

In summary, this work presents guiding principles —sequence heuristics— to encode LCST or UCST phase behaviour in IDPPs. As a corollary, we also set the stage for *de novo*, sequence-level design of phase transition polymers and proteome-wide identification of proteins that exhibit phase behaviour. These heuristics and the accompanying library of the largest extant set of protein polymers that exhibit LCST and UCST phase behaviour are new tools to study phase behaviour in biology or to exploit phase transitions for applications in fields as diverse as materials science, biotechnology and medicine.

Methods

Compositional analysis of natural Pro and Gly-rich proteins

We implemented custom-made scripts (Scripts 1–5; see Supplementary information) for analysis of proteomes and protein sequences using MATLAB R2013a. The amino acid sequence of these proteins (Supplementary Table 1) was retrieved from the NCBI for analysis. Full proteomes were downloaded as FASTA files from UniProt. With Script 1, we quantified the fraction of P-X_n-G motifs (where X is any residue but Pro and Gly) for each *n* value with respect to the total number of motifs (i.e., *n*=0–4). We then calculated (Script 2) the average distance between the Gly residue in motif *i-1* and the Pro residue in motif *i*, as well as the amino acid composition of these inter-residue segments. With Scripts 3 and 4, we applied similar analyses at a proteome-wide level. For the compositional analysis of resilins and tropoelastins, Script 5 (Supplementary Methods) calculated the percent abundance and (Kyte-Doolittle) hydrophathy⁵⁶ of all amino acids — excluding Pro and Gly— in these proteins.

Genetically encoded synthesis of protein polymers

To enable the rapid screening of a large number of protein polymers, we utilized overlap-extension rolling circle amplification (OERCA). The large majority of polymers reported in this manuscript were synthesized using this methodology (Supplementary Table 2 and Supplementary Table 4) and following the instructions described in reference²⁵ and in Supplementary Information. For the synthesis of additional IDPPs and copolymers therefrom, we used the plasmid reconstruction variant (PRe-RDL) of the recursive directional ligation (RDL) method as originally described by McDaniel and collaborators⁵⁷. The oligomers used for the synthesis of these genes are reported in Supplementary Table 3 and Supplementary Table 5.

Expression and purification of IDPPs

Protein polymers were produced in *Escherichia coli* (BL21) from plasmid-borne genes after overnight induction with IPTG. LCST polymers were purified by inverse transition cycling (see Supplementary Information for details). Most UCST IDPPs accumulated in the insoluble fraction after cell disruption, but purification was easily accomplished by dissolving the pellets in fresh PBS or PBS supplemented with 2–4 M Urea followed by 2 cycles of thermally triggered phase separation (see Supplementary Methods for details).

Phase transition behaviour and secondary structure of IDPPs

The phase transition behaviour of LCST and UCST IDPPs was characterized by monitoring their optical density (at the concentrations indicated in the figures but most typically at 50 μM) in PBS as a function of temperature, with heating and cooling performed at a rate of 1 $^{\circ}\text{C min}^{-1}$, on a Cary 300 UV-visible spectrophotometer equipped with a multicell thermoelectric temperature controller. The secondary structure displayed by these protein polymers was studied by circular dichroism in water using an Aviv Model 202 instrument and 1 mm quartz cells (Hellma) by scanning from 260 nm to 180 nm with 1 nm steps and a 3 second averaging time at various temperatures.

Cell adhesiveness of RGD-based IDPPs

5×10^5 PC3-luc-C6 cells per well were seeded on a 24-well plate and immediately after, while at room temperature, the indicated amounts of filter-sterilized UCST IDPPs with cloud points below and above 37 $^{\circ}\text{C}$ were added to the media. The cultures were incubated for either 3 or 18 h at 37 $^{\circ}\text{C}$ in 5% CO_2 . The media was then removed and the number of adherent cells was measured using a conventional MTT assay.

Proteome-wide search for proteins with features of UCST IDPPs

Using a minor variant of Script 3 we scanned the human proteome for proteins that fulfill a set of 5 sequence parameters that control UCST phase behaviour: i) protein length, ii) charge content ($\text{R} + \text{K} + \text{D} + \text{E}$), iii) percent of positively charged residues ($\text{R} + \text{K}$) with respect to all ($\text{R} + \text{K} + \text{D} + \text{E}$) charged residues—as a measure of the zwitterionic character of the protein—, iv) aromatic content ($\text{Y} + \text{H} + \text{W} + \text{F}$) and v) percent of Arg residues with respect to all ($\text{R} + \text{K}$) positively charged residues. In the Supplementary information we provide a detailed description for the minimal threshold values that we chose for our proteome-level screen.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

F.G.Q. thanks Kevin Zhu for his assistance with the preparation of Fig. 4d. This work was funded by the NIH through grant # GM061232 to A.C. and by the NSF through the Research Triangle MRSEC (NSF DMR-11-21107).

References

1. Li P, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature*. 2012; 483:336–340. [PubMed: 22398450]
2. Kato M, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012; 149:753–767. [PubMed: 22579281]
3. Weber SC, Brangwynne CP. Getting RNA and protein in phase. *Cell*. 2012; 149:1188–1191. [PubMed: 22682242]
4. Vekilov PG. Phase transitions of folded proteins. *Soft Matter*. 2010; 6:5254–5272.
5. Toretzky JA, Wright PE. Assemblages: Functional units formed by cellular phase separation. *The Journal of cell biology*. 2014; 206:579–588. [PubMed: 25179628]
6. Frey S, Richter RP, Görlich D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*. 2006; 314:815–817. [PubMed: 17082456]
7. Roy D, Brooks WL, Sumerlin BS. New directions in thermoresponsive polymers. *Chemical Society Reviews*. 2013
8. Stuart MAC, et al. Emerging applications of stimuli-responsive polymer materials. *Nature Materials*. 2010; 9:101–113. [PubMed: 20094081]
9. McDaniel JR, et al. Self-Assembly of Thermally Responsive Nanoparticles of a Genetically Encoded Peptide Polymer by Drug Conjugation. *Angewandte Chemie International Edition*. 2013; 52:1683–1687.
10. MacKay JA, et al. Self-assembling chimeric polypeptide–doxorubicin conjugate nanoparticles that abolish tumours after a single injection. *Nature Materials*. 2009; 8:993–999. [PubMed: 19898461]
11. Liu W, et al. Brachytherapy Using Injectable Seeds That Are Self-Assembled from Genetically Encoded Polypeptides In Situ. *Cancer research*. 2012; 72:5956–5965. [PubMed: 23155121]
12. Nishida K, et al. Corneal reconstruction with tissue-engineered cell sheets composed of autologous oral mucosal epithelium. *New England Journal of Medicine*. 2004; 351:1187–1196. [PubMed: 15371576]
13. Caves JM, et al. Elastin-like protein matrix reinforced with collagen microfibers for soft tissue repair. *Biomaterials*. 2011; 32:5371–5379. [PubMed: 21550111]
14. Korja P, et al. Self-assembling elastin-like peptides growth factor chimeric nanoparticles for the treatment of chronic wounds. *Proceedings of the National Academy of Sciences*. 2011; 108:1034.
15. Meyer DE, Chilkoti A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nature biotechnology*. 1999; 17:1112–1115.
16. Bellucci JJ, Amiram M, Bhattacharyya J, McCafferty D, Chilkoti A. Three-in-One Chromatography-Free Purification, Tag Removal, and Site-Specific Modification of Recombinant Fusion Proteins Using Sortase A and Elastin-like Polypeptides. *Angewandte Chemie International Edition*. 2013
17. Stayton PS, et al. Control of protein–ligand recognition using a stimuli-responsive polymer. *Nature*. 1995; 378:472–474. [PubMed: 7477401]
18. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science*. 2012; 338:1042–1046. [PubMed: 23180855]
19. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chemical reviews*. 2014
20. Li NK, Quiroz FG, Hall CK, Chilkoti A, Yingling YG. Molecular description of the LCST behaviour of an elastin-like polypeptide. *Biomacromolecules*. 2014; 15:3522–3530. [PubMed: 25142785]
21. Muiznieks LD, Keeley FW. Proline periodicity modulates the self-assembly properties of elastin-like polypeptides. *Journal of Biological Chemistry*. 2010; 285:39779. [PubMed: 20947499]
22. Dutta NK, et al. A Genetically Engineered Protein Responsive to Multiple Stimuli. *Angewandte Chemie International Edition*. 2011; 50:4428–4431.
23. Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*. 2006; 14:1667–1676. [PubMed: 17098192]

24. De Las Heras Alarcon C, Pennadam S, Alexander C. Stimuli responsive polymers for biomedical applications. *Chem Soc Rev.* 2005; 34:276–285. [PubMed: 15726163]
25. Amiram M, Quiroz FG, Callahan DJ, Chilkoti A. A highly parallel method for synthesizing DNA repeats enables the discovery of ‘smart’ protein polymers. *Nature Materials.* 2011; 10:141–148. [PubMed: 21258353]
26. Balu R, et al. An16-resilin: An advanced multi-stimuli-responsive resilin-mimetic protein polymer. *Acta biomaterialia.* 2014; 10:4768–4777. [PubMed: 25107894]
27. Urry DW, et al. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers.* 1992; 32:1243–1250. [PubMed: 1420991]
28. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences.* 2013; 110:13392–13397.
29. Yokoi H, Kinoshita T, Zhang S. Dynamic reassembly of peptide RADA16 nanofiber scaffold. *Proceedings of the National Academy of Sciences of the United States of America.* 2005; 102:8414–8419. [PubMed: 15939888]
30. Müller-Spätth S, et al. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences.* 2010; 107:14609–14614.
31. Möglich A, Joder K, Kiefhaber T. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proceedings of the National Academy of Sciences.* 2006; 103:12394–12399.
32. Meyer DE, Chilkoti A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. *Biomacromolecules.* 2002; 3:357–367. [PubMed: 11888323]
33. Seuring J, Agarwal S. First example of a universal and cost-effective approach: polymers with tunable upper critical solution temperature in water and electrolyte solution. *Macromolecules.* 2012; 45:3910–3918.
34. Shimada N, et al. Ureido-derivatized polymers based on both poly (allylurea) and poly (L-citrulline) exhibit UCST-type phase transition behaviour under physiologically relevant conditions. *Biomacromolecules.* 2011
35. Seuring J, Agarwal S. Polymers with Upper Critical Solution Temperature in Aqueous Solution: Unexpected Properties from Known Building Blocks. *ACS Macro Letters.* 2013; 2:597–600.
36. Seuring J, Agarwal S. Polymers with upper critical solution temperature in aqueous solution. *Macromolecular rapid communications.* 2012; 33:1898–1920. [PubMed: 22961764]
37. Azzaroni O, Brown AA, Huck WT. UCST Wetting Transitions of Polyzwitterionic Brushes Driven by Self-Association. *Angewandte Chemie.* 2006; 118:1802–1806.
38. Lowe AB, McCormick CL. Synthesis and solution properties of zwitterionic polymers. *Chemical reviews.* 2002; 102:4177–4190. [PubMed: 12428987]
39. Mason P, Neilson G, Dempsey C, Barnes A, Cruickshank J. The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. *Proceedings of the National Academy of Sciences.* 2003; 100:4557–4561.
40. Mahadevi AS, Sastry GN. Cation– π interaction: Its role and relevance in chemistry, biology, and material science. *Chemical reviews.* 2013; 113:2100–2138. [PubMed: 23145968]
41. Kao WJ, Lee D, Schense JC, Hubbell JA. Fibronectin modulates macrophage adhesion and FBGC formation: the role of RGD, PHSRN, and PRRARV domains. *Journal of biomedical materials research.* 2001; 55:79–88. [PubMed: 11426401]
42. Iwamoto Y, et al. YIGSR, a synthetic laminin pentapeptide, inhibits experimental metastasis formation. *Science.* 1987; 238:1132–1134. [PubMed: 2961059]
43. Pierschbacher MD, Ruoslahti E. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature.* 1983; 309:30–33. [PubMed: 6325925]
44. Lee BW, et al. Strongly Binding Cell-Adhesive Polypeptides of Programmable Valencies. *Angewandte Chemie International Edition.* 2010; 49:1971–1975.
45. van der Lee R, et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews.* 2014; 114:6589–6631. [PubMed: 24773235]

46. Brassart B, et al. Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts. *Journal of Biological Chemistry*. 2001; 276:5222. [PubMed: 11084020]
47. Dreher MR, et al. Temperature triggered self-assembly of polypeptides into multivalent spherical micelles. *Journal of the American Chemical Society*. 2008; 130:687–694. [PubMed: 18085778]
48. Callahan DJ, et al. Triple stimulus-responsive polypeptide nanoparticles that enhance intratumoral spatial distribution. 2012
49. Crick SL, Ruff KM, Garai K, Frieden C, Pappu RV. Unmasking the roles of N-and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proceedings of the National Academy of Sciences*. 2013; 110:20075–20080.
50. Bates FS, et al. Multiblock polymers: panacea or Pandora's box? *Science*. 2012; 336:434–440. [PubMed: 22539713]
51. Hassouneh W, Zhulina EB, Chilkoti A, Rubinstein M. Elastin-like Polypeptide Diblock Copolymers Self-Assemble into Weak Micelles. *Macromolecules*. 2015
52. Tong R, Chiang HH, Kohane DS. Photoswitchable nanoparticles for in vivo cancer chemotherapy. *Proceedings of the National Academy of Sciences*. 2013; 110:19048–19053.
53. Wong C, et al. Multistage nanoparticle delivery system for deep penetration into tumor tissue. *Proceedings of the National Academy of Sciences*. 2011; 108:2426–2431.
54. Clark JI. Self-assembly of protein aggregates in ageing disorders: the lens and cataract model. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2013; 368:20120104.
55. Smith FJ, et al. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nature genetics*. 2006; 38:337–342. [PubMed: 16444271]
56. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*. 1982; 157:105–132. [PubMed: 7108955]
57. McDaniel JR, MacKay JA, Quiroz FG, Chilkoti A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules*. 2010; 11:944–952. [PubMed: 20184309]

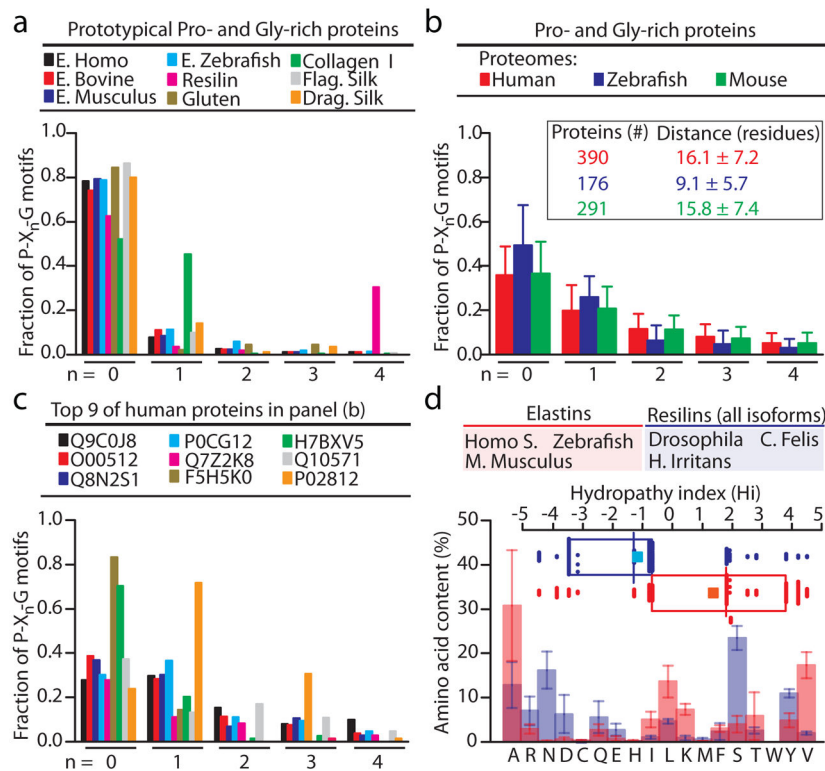


Figure 1. The amino acid sequence of Pro- and Gly-rich proteins is characterized by discrete arrangements of Pro and Gly residues

(a) Abundance of individual P-X_n-G motifs among prototypical Pro- and Gly-rich IDPs relative to the total number of P-X_n-G motifs per protein. X is any amino acid except Pro or Gly, and n varies from 0 to 4. (b) Overall abundance of individual P-X_n-G motifs relative to the total number of motifs (for n=0–4) per protein among human, mouse and zebrafish proteins with >10% Pro and >10% Gly. The columns represent the average values and the error bars are the standard deviations. We excluded collagens and elastins and only considered proteins with more than 50 residues and at least 15 P-X_n-G motifs. The inset shows the number of identified proteins per proteome and the corresponding average number of residues (i.e., distance) that separate consecutive P-X_n-G motifs (n=0–4). (c) Abundance of individual P-X_n-G motifs among the top 9 human proteins from panel (b) with the most P-X_n-G motifs. (d) Abundance of all amino acids different from Pro and Gly among resilins and elastins from several species. Error bars are the standard deviations. The inset shows the corresponding distribution of hydropathy values —according to the Kyte-Doolittle scale: hydrophilic < 0 < hydrophobic⁵⁶— where each box delimits the 25th and 75th percentile and the inner line and square indicate the median and mean, respectively. Individual protein names and accession numbers for proteins in (a) are reported in Table S1. Proteins in (c) are identified by their UniProt accession numbers.

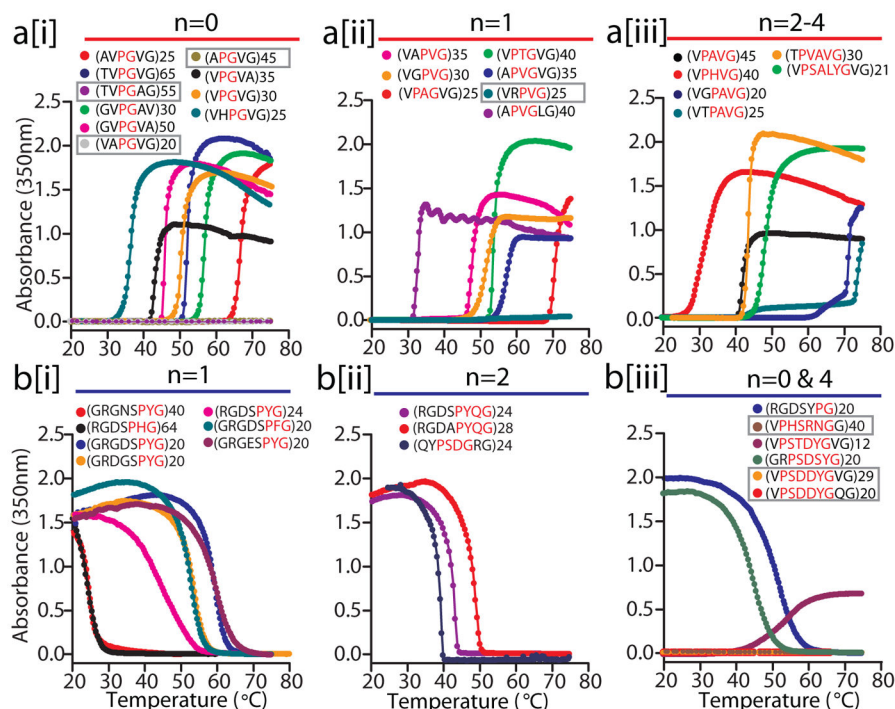


Figure 2. Protein polymers with repeating P-X_n-G motifs (n=0–4) can be designed to exhibit LCST or UCST phase behaviour under physiologically relevant conditions

(a) Temperature-dependent turbidimetry for protein polymers with periodic Pro and Gly residues arranged as P-X_n-G units, where n=0 (a-i), 1 (a-ii), and 2–4 (a-iii), and having pentapeptide, hexapeptide and nonapeptide repeat units whose amino acid composition is reminiscent of elastins (Fig. 1d). (b) Turbidimetry data on cooling for genetically encoded polymers with P-X_n-G units whose amino acid composition is reminiscent of resilins (Fig. 1d). Gray rectangles in (a–b) indicate protein polymers that did not undergo a phase transition (i.e., A_{350nm} = 0 over the studied temperature range). All turbidity measurements were performed in PBS (pH 7.4) at a polypeptide concentration of 50 μM, except for VPSALYGVG (+ 8 M urea) and RGDSPYQ (+1 M urea).

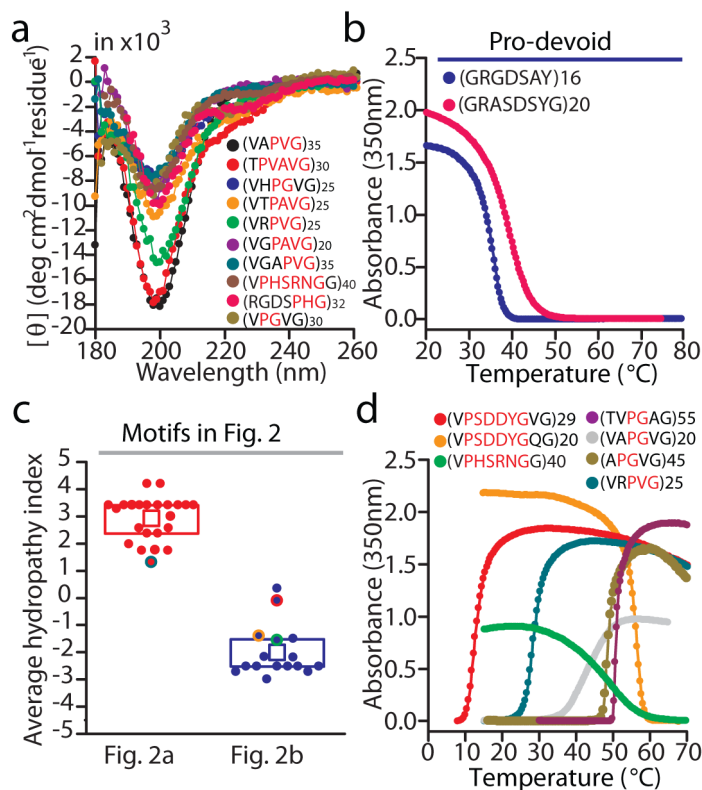


Figure 3. Intrinsically disordered protein polymers can be designed to exhibit LCST or UCST phase behaviour under a wide range of conditions

(a) Protein polymers with repeating P-X_n-G motifs lack ordered secondary structures as shown by their CD spectra (at 25 $^{\circ}\text{C}$) that are characteristic of IDPs. CD studies were conducted in water at a polypeptide concentration of 5 μM . (b) Turbidimetry data in PBS and on cooling for two protein polymers composed of Pro-devoid motifs. (c) Average hydropathy index for each amino acid motif studied in Fig. 2. This index corresponds to the average of the Kyte-Doolittle hydropathy indices⁵⁶ of all residues in the motif. Each box delimits the 25th and 75th percentile and the inner square indicates the mean index. Solid circles with a colored outline correspond to motifs in Fig. 2 that have a net charge and whose phase behaviour is revisited in panel (d). (d) Temperature-dependent turbidimetry in PBS supplemented with salt or in acidified PBS for polymers in Fig. 2 that did not undergo a phase transition behaviour in PBS over the studied temperature range: VRPVG (+ 1M NaCl), VAPGVG (+ 0.5 M NaCl), TVPGAG (+2 M NaCl), APGVG (+2 M NaCl), VPHSRNGG (+2 M NaCl), VPSDDYGVG (PBS pH 2 + 2 M Urea) and VPSDDYQGQ (PBS pH 2).

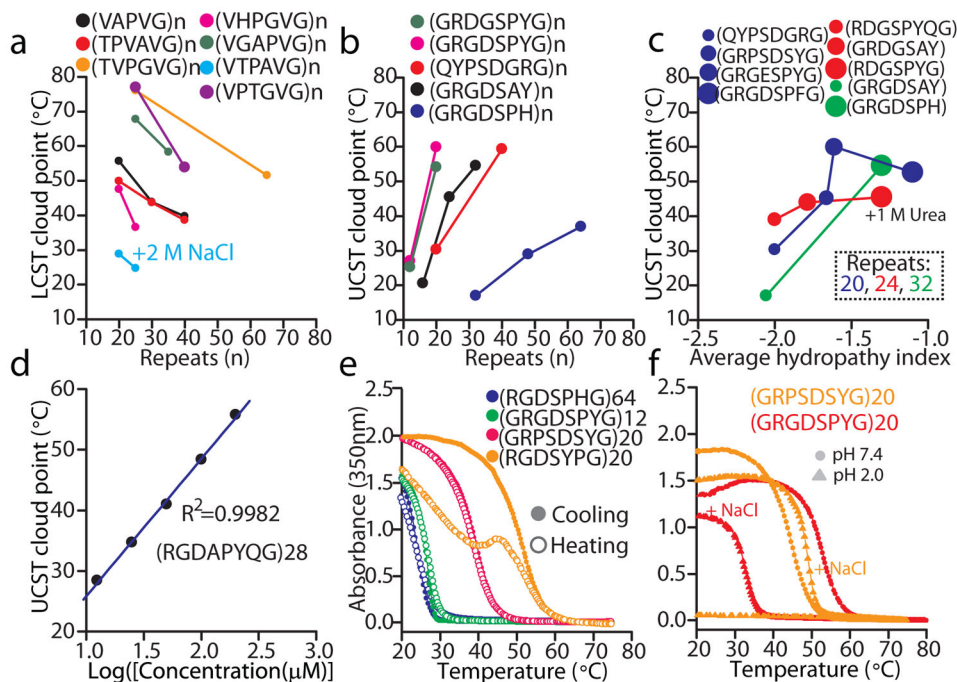


Figure 4. The phase transition behaviour of IDPPs is tunable

Cloud points for a subset of LCST (a) and UCST (b) protein polymers as a function of the number of repeats as extracted from turbidimetry data at a polymer concentration of 50 μM in PBS (unless otherwise stated). (c) UCST cloud points —extracted as in (a–b)— as a function of the average hydrophathy index of the amino acid motif for protein polymers grouped by repeat length. To guide the eye, the size of the solid circles increases with the hydrophathy index. (d) UCST cloud points for (RGDAPYQG)28 as a function of the concentration of the polymer in PBS (Supplementary Fig. 8a shows the raw turbidimetry data). (e) Turbidimetry data for representative UCST IDPPs upon cooling and subsequent heating. (f) The UCST behaviour of zwitterionic IDPPs is disrupted upon losing charge neutrality under acidic conditions (PBS pH 2), unless the acidified PBS is supplemented with salt (+0.5 M NaCl). Turbidimetry data was acquired in PBS (pH 7.4) at 50 μM unless otherwise stated.

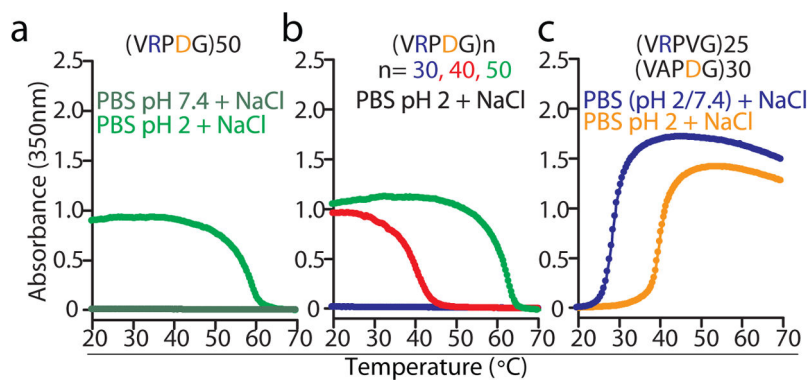


Figure 5. A zwitterionic IDPP that is devoid of aromatic residues exhibits UCST phase behaviour in acidic environments

(a) Turbidimetry data on cooling for (VRPDG)₅₀ in PBS (pH 7.4 and 2) supplemented with 1.5 M NaCl. (b) Turbidimetry data on cooling for polymers with an increasing number of VPRDG repeats. (c) Substitution of V for D and A for R in the VRPDG motif abolished the UCST phase behaviour of the resulting polymers at all pH values. Instead, turbidimetry data on heating revealed that IDPPs composed of these mutated motifs show LCST-type phase transitions.

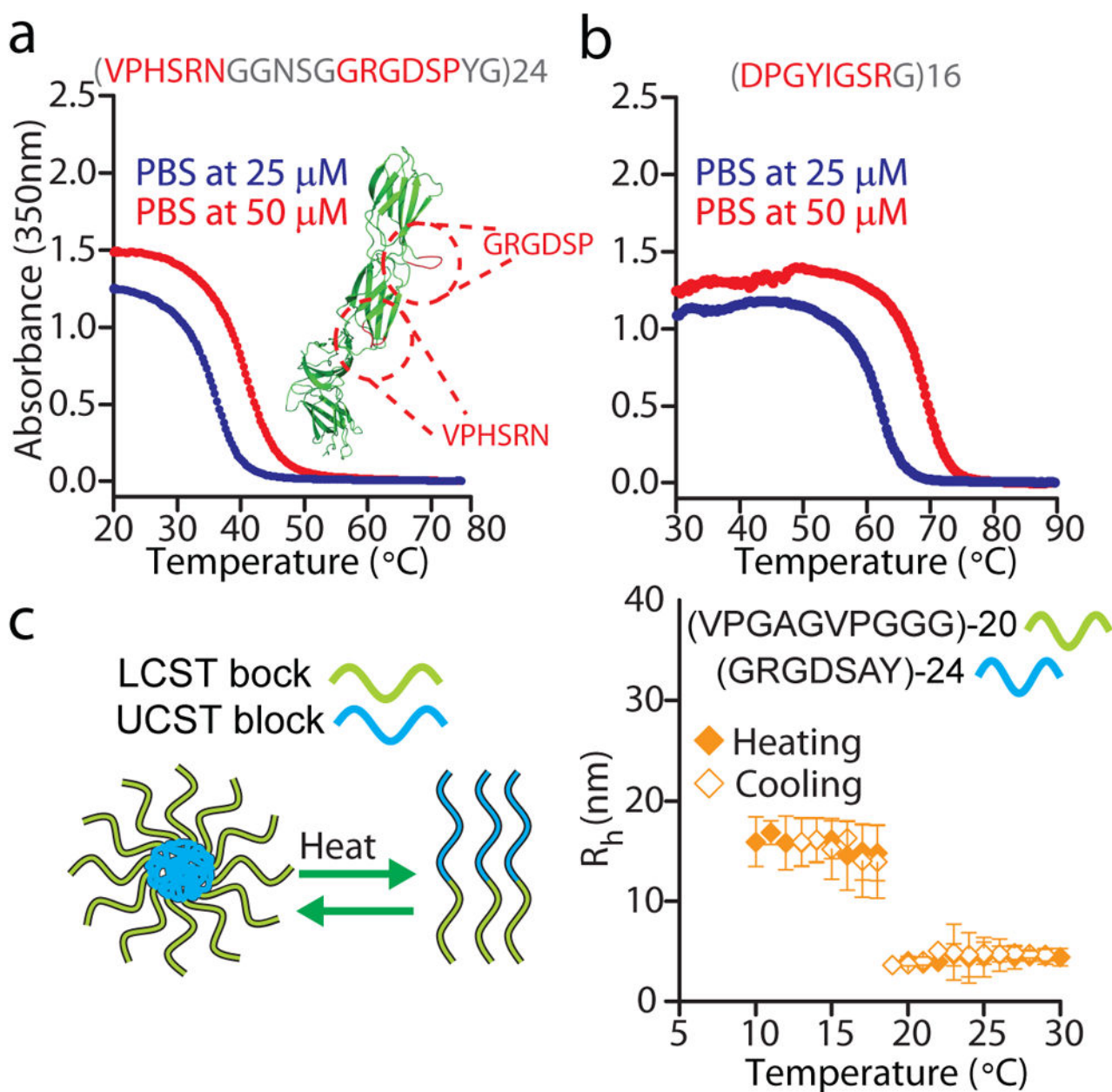


Figure 6. IDPPs built from biologically active peptide repeats

(a) A fibronectin-like polymer, which combines—in the repeat unit—the cell adhesion (GRGDSP) and the synergy (VPHSRN) motifs from fibronectin (inset in figure), displays a physiologically relevant UCST in PBS. The structure of fibronectin (PDB file: 1FNF) was rendered using PyMOL (<http://pymol.org/>). (b) The laminin-derived, bioactive peptide DPGYIGSR is programmed into a UCST IDPP upon polymerization, as shown by turbidimetry data in PBS. The repeat sequences are shown as red or gray amino acid letters. Red corresponds to amino acids in the native peptide, whereas gray denotes residues introduced to modulate the UCST of the resulting polymers. (c) Dynamic light scattering

shows the temperature-triggered disassembly of a block copolymer composed of a UCST and a LCST block.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript