Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/5/10/eaax5177/DC1

Supplementary Materials for

Intrinsically disordered proteins access a range of hysteretic phase separation behaviors

Felipe Garcia Quiroz, Nan K. Li, Stefan Roberts, Patrick Weber, Michael Dzuricky, Isaac Weitzhandler, Yaroslava G. Yingling, Ashutosh Chilkoti*

*Corresponding author. Email: chilkoti@duke.edu

Published 18 October 2019, *Sci. Adv.* **5**, eaax5177 (2019)
DOI: 10.1126/sciadv.aax5177

This PDF file includes:

Supplementary Materials

Supplementary Methods

Genetically-encoded synthesis of IDPPs

We relied on our recently published library (*10*) of genes encoding IDPPs that exhibit LCST and UCST phase behavior and which we generated using OERCA (*42*). All details of gene synthesis can be found in the supplementary information of ref (*10*). In addition, to study the hysteretic phase behavior of specific IDPPs in detail and with full control on repeat number, we generated new genes encoding these IDPPs by Pre-RDL (table S1). Briefly, we synthesized small dsDNA cassettes encoding 4 repeats of the motif of interest (table S1), which were further multimerized using a conventional Pre-RDL approach (*43*) to the length of interest. Using this approach, we typically added an N-terminal leader peptide, SKGP, and on occasion —to enable direct comparison with OERCA-generated genes— added a C-terminal His-tag (table S1). The Pre-RDL vector that we used is different from the vector originally reported but the same vector (JMD3) as the one in our most recent publication (*10*), which encodes a short C-terminal GWP tripeptide. For the synthesis of diblock IDPPs, we again took advantage of Pre-RDL for the synthesis of genes that combined our full-length IDPP genes with a widely reported ELP gene encoding 80 repeats of a VPGXG motif (where X=[A:G]) (*43*).

Expression and purification of IDPPs

Starter cultures (5 mL starter culture per 1 L of UltraBL21 cells form EdgeBio) of Terrific broth (TB) media (BioExpress) supplemented with 100 μg/mL ampicillin (for all IDPP-encoding plasmids synthesized by OERCA and reported in ref (*10*)) or 45 μg/mL Kanamycin (for IDPP-encoding plasmids synthesized by PRe-RDL and specifically for this manuscript; table S1) were inoculated with transformed cells from DMSO stocks stored at −80 °C, and incubated overnight at 37 °C while shaking at 250 rpm. The starter cultures were then centrifuged at 3000 g for 7 min and resuspended in 1 mL of fresh TB medium. Expression cultures (4 L flasks containing 1 L of TB media with ampicillin or kanamycin as in the starter cultures) were inoculated with the resuspended starter culture and incubated at 37 °C with shaking at 200 rpm. After 8-9 h of growth, expression was induced by the addition of IPTG (from Goldbio) to a final concentration of 1 mM. Cells were harvested 24 h after inoculation, and LCST IDPPs were purified by inverse

transition cycling (ITC) —with minor modifications— as described elsewhere (*44*). These IDPPs were then purified through regular ITC in PBS. UCST IDPPs were purified as previously reported (*10*). LCST IDPPs made of VGPVG repeats were purified from the insoluble fraction obtained after sonication and centrifugation. Briefly, cells were sonicated and centrifuged at 14000 rpm for 15 min at 4 °C. The supernatant was discarded and the pellet was resuspended in 15 ml 6M $GnCl_2$. These solutions were centrifuged at 14000 rpm for 30 min at 4 °C and the supernatants were extensively dialyzed against water at 4 °C.

Characterization of the phase transition behavior and secondary structure of IDPPs

To characterize the phase behavior of the synthesized peptide polymers, the optical density of IDPP solutions (at the concentrations indicated in the manuscript and figures) was monitored at a wavelength of 350 nm as a function of temperature, with heating and cooling performed at a rate of 1 °C min^{-1} unless otherwise indicated in the text or figures, on a Cary 300 UV-visible spectrophotometer equipped with a multicell thermoelectric temperature controller (Varian Instruments, Walnut Creek, CA). Most experiments were conducted in PBS, but the main manuscript indicates all relevant exceptions (e.g., supplementation with urea or NaCl, or water in the case of conformational studies and as noted in corresponding figure legends).

Temperature-dependent changes in the secondary structure displayed by IDPPs was studied by circular dichroism (CD) 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. IDPPs were typically dialyzed overnight against Milli-Q water, protein purity was assessed by SDS-PAGE and the polypeptides were diluted to 5 µM in water. Raw CD data in millidegrees was first corrected by subtracting the corresponding CD signal from water blanks and transformed into Mean Residue Ellipticity (θ) as reported elsewhere (*45*).

Quantitative analysis of P-X_n-G motifs and surrounding Gly residues.

We quantified the abundance of Gly in residue positions surrounding P-X_n-G motifs, namely two residues N-terminal and two residues C-terminal to each motif in a previously reported set of Pro and Gly-rich IDPs (*10*). To understand the significance of the relative abundance of specific P-X_n-G motifs or particular biases in the distribution of Gly residues in the vicinity of such motifs, we evaluated the statistical significance of these biases in relation to the overall amino acid

composition of each protein. This is relevant considering that some of these proteins are highly enriched in specific amino acids and this alone might bias the occurrence of particular amino acid arrangements. To do so, we generated 1000 randomized versions of each protein of interest by assigning the amino acid from a randomly chosen residue position in the parent protein (using a uniform distribution) to each residue position in the randomized equivalent, and then repeating this process to create 1000 such "clones". Note that this process is analogous to scrambling the amino acid sequence of the parent protein, which ensures that all "clones" of the protein have very similar amino acid compositions. The analyses on the abundance of P-X_n-G motifs and Gly distribution around these motifs was carried out on both the parent protein and its 1000 randomized equivalents. By doing this, we were able to generate a distribution of statistically possible values for a quantifiable feature—say the fraction of Pro-Gly dipeptides— in a given protein, which we then used to determine the probability of finding an apparent sequence determinant (i.e., a recurrent amino acid pattern that could be relevant for the properties of the parent protein) by biased amino acid content alone. We considered that a given amino acid pattern was a true sequence determinant of the parent protein when the probability of finding its characteristic value or any other more extreme value in the distribution (i.e. the so-called p-value) was lower than 0.001; in other words, when the characteristic value of a given sequence determinant was not represented in the distribution of statistically possible values generated from 1000 randomized proteins.

Script 1 (included at the end of this section) was used to generate randomized variants of each protein of interest and to quantify the fraction of P-X_n-G motifs for each *n* value with respect to the total number of motifs (i.e., n=0-4), as well as the probability of finding Gly in the vicinity of these motifs.

The data for the fraction of P-X_n-G motifs was used as generated by Script 1. To study the distribution of Gly in the vicinity of these motifs, however, we analyzed a different variable, which we call "Fold change from random Gly" (FCRG), calculated according to Eq. S1

$$FCRG = \frac{(GlyatPmX - Gc)}{Gc} \qquad \text{(Eq. S1)}$$

Where GlyatPmX is the percentage (or probability) of Gly at Pm2, Pm1, Gp1 or Gp2 (equivalent to positions *-2*, *-1*, *+1* and *+2* in the main manuscript) calculated as in Script 1, and Gc is the percentage of Gly (i.e., overall Gly content) in the parent protein also from Script 1. Positive values of FCRG indicate an enrichment of Gly, whereas negative values suggest a depletion of Gly with respect to the probability of occurrence of Gly expected from the overall Gly content of each protein.

Nanoparticle characterization

Temperature-dependent nano-assembly of diblock IDPPs was studied at a fixed concentration of 50 μM in PBS, using UV-visible spectrophotometry (Cary 300) and dynamic light scattering measurements on a Wyatt DynaPro temperature-controlled microsampler. Heating and cooling was performed at ~1 °C/min. Prior to dynamic light scattering measurements, we routinely filtered IDPP solutions (at 25 °C) through 20 nm filters. Cryo-TEM experiments were performed at Duke University's Shared Materials Instrumentation Facility (Durham, NC). Lacey holey carbon grids (Ted Pella, Redding, CA) were glow discharged in a PELCO EasiGlow Cleaning System (Ted Pella, Redding, CA). Nanoparticle assembly was first triggered by incubating samples (50 μM in PBS) at 50 °C for 15 min. Then, inside the vitrification chamber, a 3 μl drop of each sample was deposited onto the grid, blotted for 3 s, and vitrified in liquid ethane using the Vitrobot Mark IV (FEI, Eindhoven, Netherlands). Prior to vitrification, the sample chamber was maintained at either 50 °C (for nanoparticles in Fig. 6D) or 30 °C (for nanoparticles in Fig. 6E) and at 100% relative humidity to prevent sample evaporation. After vitrification, grids were transferred to a Gatan 626 cryoholder (Gatan, Pleasanton, CA) and imaged on a FEI Tecnai G2 Twin TEM (FEI, Eindhoven, Netherlands), operating under low-voltage conditions at 80 keV.

MD simulations

Simulation details are as described in our previous research paper (*30*), but the necessary details are mentioned briefly in the following. Fully atomistic molecular dynamics simulations were performed using Amber 12.0 (*46*) and the ff99SB force field for proteins with explicit solvent using the TIP3P water model (*47*). All simulations were performed using an NPT ensemble with periodic boundary conditions applied. For single-peptide simulation systems, the size of the periodic box is set to be big enough so that no peptide-peptide interactions through periodic

boundary conditions can occur. The initial equilibration process includes several minimization, thermalization and NPT-MD run cycles which result in a suitable initial unbiased system configuration at the desired temperatures —290 K and 350 K for $(VPGVG)_{18}$, $(VGPVG)_{18}$, $(VPAGLG)_{18}$, and $(LGAPVG)_{18}$ and 310 K and 370 K for $(VPAGVG)_{18}$ and $(VGAPVG)_{18}$. MD simulations were performed using a 2 fs time step and produced 60 ns or 70 ns (60 ns for hexamer motifs and 70 ns for pentamer motifs) of molecular dynamics trajectories, following the equilibration process. The statistical and clustering analysis was carried out for the last 30 ns or 40 ns (30 ns for hexamers and 40 ns for pentamers) of MD trajectories. Secondary structure propensities were calculated with the DSSP algorithm. The propensity of unstructured motifs (coil) was calculated as 1-f(turn)-f(helix)-f(sheet).

For the two-chain simulations, the simulation system was constructed in a box by solvating two representative structures of single IDPPs which were placed side-by-side with a certain distance between the centers of mass of the IDPP chains. The representative structures at different temperatures were chosen through the application to the MD simulation trajectories of single IDPP chains described above and using a hierarchical RMSD-based cluster algorithm (*48*) followed by the energetic analysis. The system was allowed to equilibrate at the corresponding temperatures —290 K and 350 K for $(VPGVG)_{18}$, $(VGPVG)_{18}$, $(VPAGLG)_{18}$, and $(LGAPVG)_{18}$, and 310 K and 370 K for $(VPAGVG)_{18}$ and $(VGAPVG)_{18}$. Readers may refer to our previous publication for the information regarding the simulation set-up and equilibration protocols of the two-chain systems. The production simulations were carried out for more than 15 ns for pentamer pairs and 20 ns for hexamer pairs, respectively, with a 2 fs time step. The final system configurations from the two-chain IDPP simulations were taken to be the initial configurations in the “quenching”, cooling simulations. Simulations were carried out for 25 ns in an NPT ensemble at 290 K with a 2 fs time step.

MD trajectories were processed using in-house scripts along with the standard tool suite accompanying Amber12.0. The interaction energy was calculated using the molecular mechanics energy function in NAMD 2.7 (*49*). The hydrophobic contacts between two polypeptides were counted as the number of residues with a distance cutoff of 5 Å between the hydrophobic groups of these residues and residues belonging to the other polypeptide. The hydrogen bond analysis was performed using a distance cutoff of 3.5 Å and an angle cut-off of 30°.

MATLB Script 1

```
%****************** Start of MATLAB Script 1 *********************************
%Definition of main variables
%rndlocation : matrix with random residue positions
%rndProt: matrix with parent protein in the first row followed by its 1000 random variants
%pstarts: stores the residue position of each Pro participating in a P-Xn-G motif
%gends: stores the residue position of each Gly participating in a P-Xn-G motif
%Gpm2: counter for the occurrence of Gly two residues before Pro in a P-Xn-G motif
%Gpm1: counter for the occurrence of Gly one residue before Pro in a P-Xn-G motif
%Ggp1: counter for the occurrence of Gly one residue after Gly in a P-Xn-G motif
%Ggp2: counter for the occurrence of Gly two residues after Gly in a P-Xn-G motif
clear all

fid=fopen('protein_sequence.txt');
x=fscanf(fid,'%s');

% Start of code A to generate 1000 randomized versions of a given protein
N=1000;
rndlocation=unidrnd(length(x), N,length(x));
rndProt(1,:)=x(:);
rndProt(2:N+1,:)=x(rndlocation);
% End of code A (note that the first row of 'rndlocation' is the vector for the parent protein)

% Start of code B to identify the location of Pro and Gly residues forming P-Xn-G motifs
for i=1:N+1
t=1;n=1;
for j=1:length(x)
   if t==1
   if strcmp('P',rndProt(i,j))
      t=2;
      pstarts(n)=j;
   end
   else
      if strcmp('P',rndProt(i,j))
      pstarts(n)=j;
      end
      if strcmp('G',rndProt(i,j))
      t=1;
      gends(n)=j;
      n=n+1;
      end
   end
end
% End of code B

Gpm2=0;Gpm1=0;Ggp2=0;Ggp1=0;t=0; p=0; pg=0;pxg=0;pxxg=0;pxxxg=0;pxxxxg=0;
for j=1:length(gends)
   %Start of code C to calculate the fraction of P-Xn-G motifs
   if pstarts(j)-gends(j)==-1
   pg=pg+1;
   end
   if pstarts(j)-gends(j)==-2
   pxg=pxg+1;
   end
   if pstarts(j)-gends(j)==-3
```

```
   pxxg=pxxg+1;
   end
   if pstarts(j)-gends(j)==-4
   pxxxg=pxxxg+1;
   end
   if pstarts(j)-gends(j)==-5
   pxxxxg=pxxxxg+1;
   end
   % End of code C (requires the loop)

   %Start of code D to calculate the number of Gly residues around P-Xn-G motifs
   if pstarts(j)-2>0
      t=t+1;
      if strcmp('G',rndProt(i,pstarts(j)-2))
         Gpm2=Gpm2+1;
      end
      if strcmp('G',rndProt(i,pstarts(j)-1))
         Gpm1=Gpm1+1;
      end
   end
   if gends(j)+2<length(x)
      p=p+1;
       if strcmp('G',rndProt(i,gends(j)+2))
         Ggp2=Ggp2+1;
       end
       if strcmp('G',rndProt(i,gends(j)+1))
         Ggp1=Ggp1+1;
      end
   end
%End of code D (requires the loop)
end

%Start of code E to calculate the percentage (probability) of Gly at -2, -1, +1 and +2 positions.
ProbGpm2=Gpm2*100/t;
ProbGpm1=Gpm1*100/t;
ProbGgp2=Ggp2*100/p;
ProbGgp1=Ggp1*100/p;
%End of code E
% Start of code F to calculate Pro and Gly content
P=1;G=1;
for j=1:length(x)
   if strcmp('P',rndProt(i,j))
      P=P+1;
   end
   if strcmp('G',rndProt(i,j))
      G=G+1;
   end
end
Pc=100*P/length(x); %Pro content
Gc=100*G/length(x); %Gly content
%End of code F
% Matrices that summarize all the relevant data
GlyatPmX(i,:)=[ProbGpm2 ProbGpm1 ProbGgp1 ProbGgp2 Gc];
FractPXG(i,:)=[pg pxg pxxg pxxxg pxxxxg]/length(pstarts);
End
%****************** End of MATLAB Script 1**********************************
```

Table S1. DNA sequence information for synthesis of genes encoding IDPPs by Pre-RDL. These minimal sequences encode 5 repeats of each motif or a His-tag (added at the C-terminus of these IDDPs). Oligonucleotides were synthesized and 5'-phosphorylated by Integrated DNA Technologies (IDT).

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
LGAPVG (reported in ref (*10*))	GCTGGGTGCTCCAGTTGGTCTGGG CGCCCCGGTGGGCCTGGGCGCGCC TGTGGGCCTGGGCGCGCCGGTCGG CCTGGGTGCACCAGTAGG	TACTGGTGCACCCAGGCCGACCG GCGCGCCCAGGCCCACAGGCGC GCCCAGGCCCACCGGGGCGCCCA GACCAACTGGAGCACCCAGCCC
VPAGLG	GGTACCAGCTGGTCTGGGTGTGCC GGCCGGCCTGGGCGTGCCTGCGGG CCTGGGCGTCCCGGCGGGCCTGGG CGTTCCAGCAGGTCTGGG	CAGACCTGCTGGAACGCCCAGGC CCGCCGGGACGCCCAGGCCCGCA GGCACGCCCAGGCCGGCCGGCA CACCCAGACCAGCTGGTACCCC
VAPVG	GGTTGCTCCAGTTGGTGTGGCCCC GGTGGGCGTGGCGCCTGTGGGCGT CGCGCCGGTCGGCGTAGCACCAGT AGG	TACTGGTGCTACGCCGACCGGCG CGACGCCCACAGGCGCCACGCCC ACCGGGGCCACACCAACTGGAG CAACCCC
TPVAVG	GACCCCAGTTGCTGTTGGTACACC GGTGGCCGTGGGCACTCCTGTGGC GGTGGGCACACCGGTCGCGGTCGG CACGCCAGTAGCAGTAGG	TACTGCTACTGGCGTGCCGACCG CGACCGGTGTGCCCACCGCCACA GGAGTGCCCACGGCCACCGGTGT ACCAACAGCAACTGGGGTCCC
VGPVG	CGTTGGCCCGGTAGGTGTCGGTCC AGTGGGCGTAGGCCCGGTTGGTGT TGGTCCTGTCGGCGTGGGTCCGGT GGG	CACCGGACCCACGCCGACAGGA CCAACACCAACCGGGCCTACGCC CACTGGACCGACACCTACCGGGC CAACGCC
His-tag	CCACCATCATCACCATCACGG	GTGATGGTGATGATGGTGGCC

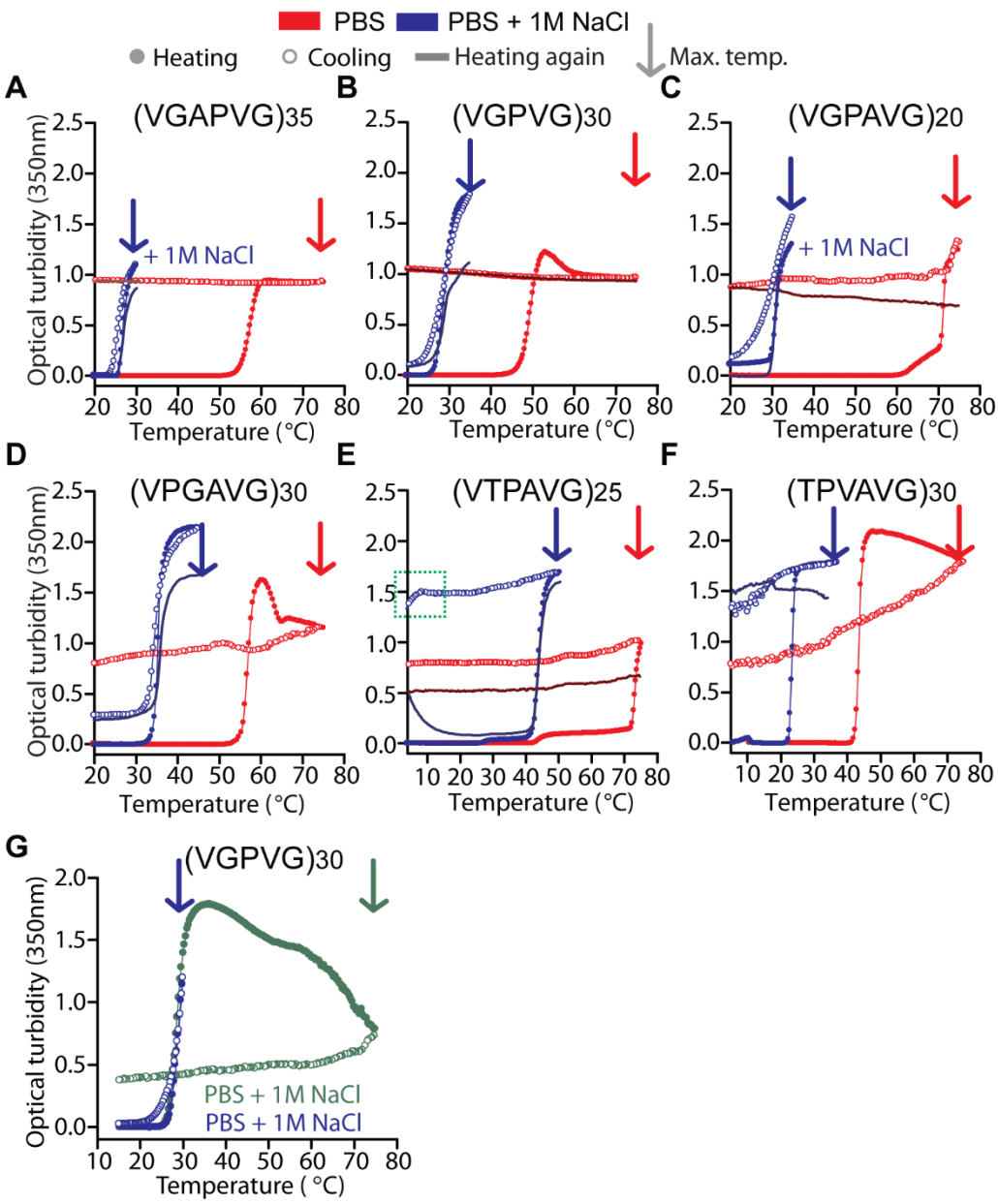

Fig. S1. LCST IDPPs display large, environmentally sensitive hysteresis. (**A-F**) Temperature-dependent optical turbidity for six different IDPPs revealed a form of temperature-sensitive hysteresis. These IDPPs display irreversible phase separation when heated to 75 °C, whereas they exhibit reversible phase transition behavior if heated below a given threshold temperature (indicated by blue arrows). IDPPs with repeat unit VTPAVG (E) exhibit a complex type of phase transition behavior with temperature-sensitive large hysteresis above the threshold temperature (see red curve), and moderate hysteresis —we typically observed zero hysteresis for other sequences— below this temperature (see blue curve). Note that the finite hysteresis of this

IDPP is only partially captured by the cooling trace (open blue symbols) as the absorbance starts decreasing around 10 °C (marked by the green square), whereas the second heating cycle (solid blue line) evidences the ability of this IDPP to undergo an identical phase transition event as in the original heating cycle. Although we were unable to characterize the threshold temperature at which polymers based on repeats of TPVAVG (F) display reversible phase transition behavior in PBS —we do so in urea as shown in Fig. 3D—, we note that this peptide polymer and all others reported in this manuscript were purified exploiting the reversibility of its phase behavior in response to changes in buffer ionic strength (at room temperature). The phase behavior of all polymers was characterized in PBS (red traces) or PBS supplemented with 1 M NaCl (blue traces) at a concentration of 50 μM. (**G**) Heating above the critical threshold temperature results in a switch from negligible hysteresis to seemingly irreversible phase separation regardless of the addition of NaCl. These data are complementary to (B). Both samples were characterized in PBS supplemented with 1 M NaCl at a concentration of 50 μM, with the only difference being the maximum temperature during the heating cycle (indicated by arrows).

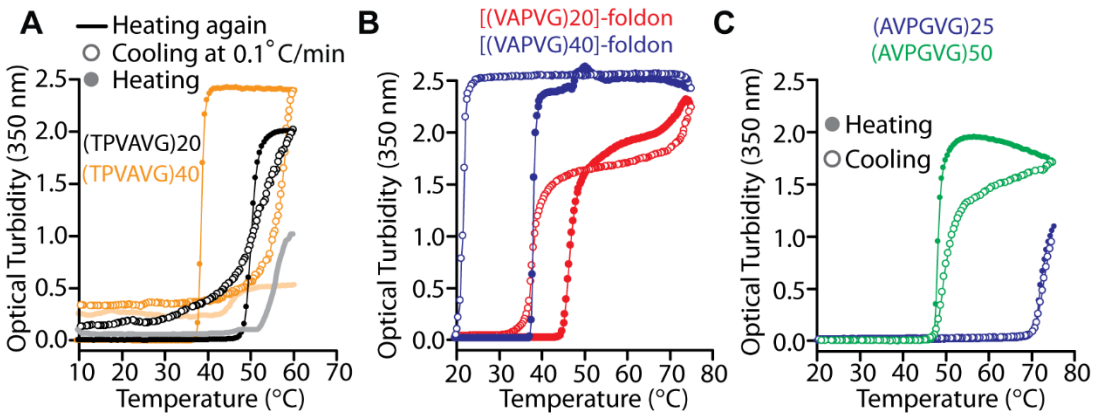

Fig. S2. Repeat number influences the hysteretic phase behavior of LCST IDPPs. (**A**) IDPPs composed of 20 and 40 repeats of the hexapeptide TPVAVG, a motif that upon polymerization exhibits large hysteresis (fig. S1F), show the expected molecular weight dependence of the T_{cp}, but increases in length also augmented their hysteretic phase behavior. Although precipitation when cooling from 60 °C to 5°C —without agitation— at the slow cooling rate of 0.1 °C/min complicates the analysis, a second heating cycle reveals that polymers with a small number of TPVAVG repeats (e.g. 20) display a smaller degree of hysteresis than polymers with twice the number of repeats, as only the latter failed to exhibit a second rapid increase in turbidity upon heating above the expected cloud point temperature. (**B**) By fusing a well-established, short trimerization peptide (Foldon) to the C-terminus of IDPPs, the minimal number of repeats required to exhibit pronounced hysteretic phase behavior can be further decreased, as shown here for the VAPVG motif that would normally require IDPPs with close to 40 repeats to exhibit a sharp transition on cooling (Fig. 1D-E). Note that the degree of thermal hysteresis is still controlled by the total number of repeats. The foldon sequence is GYIPEAPRDGQAYVRKDGEWVLLSTFL. (**C**) IDPPs composed of AVPGVG repeats, a motif in our library that upon polymerization exhibits no hysteresis, display lower cloud point temperatures as the length of the polymer is increased, but no hysteresis emerges as repeat number increases.

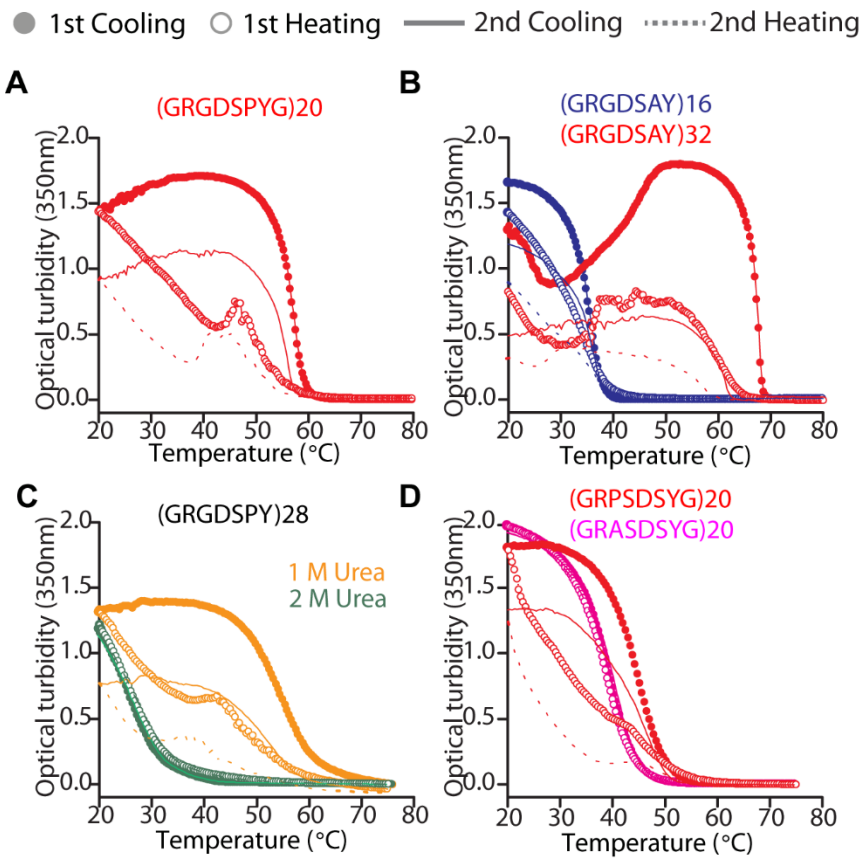

Fig. S3. Forms of irreversible phase behavior in UCST IDPPs. (**A**) UCST IDPPs in our library(*10*) that do not exhibit fully reversible phase behavior through multiple cycles of cooling and heating, typically show a progressive loss of reversibility that is seen as substantial reductions in maximum absorbance at the peak of their temperature-dependent UV-visible absorbance profile. (**B**) This progressive loss of reversibility is influenced by the number of repeats in a given UCST IDPP. (**C**) UCST IDPPs that display progressive loss of reversibility in PBS can be shifted to a form of phase behavior that is fully reversible by adding increasing amounts of urea, which likely counters peptide-peptide hydrogen-bond formation in the aggregated state. (**D**) The progressive aggregation of an UCST IDPP is eliminated by a simple Pro to Ala mutation in the repeat unit. Because Pro rigidifies the backbone of IDPs and intrinsically disordered regions (*50*), we surmise that mutations that reduce chain rigidity (e.g. Pro to Ala, since PolyP is more rigid than polyA(*51*)) are likely to counter the propensity of UCST IDPPs to undergo progressive aging upon multiple cycles of phase separation. We note that while UCST IDPPs do not need Pro to maintain their intrinsic disorder, we have not been

successful in synthesizing LCST-type, Pro-devoid IDPPs, which are enriched in aliphatic residues.

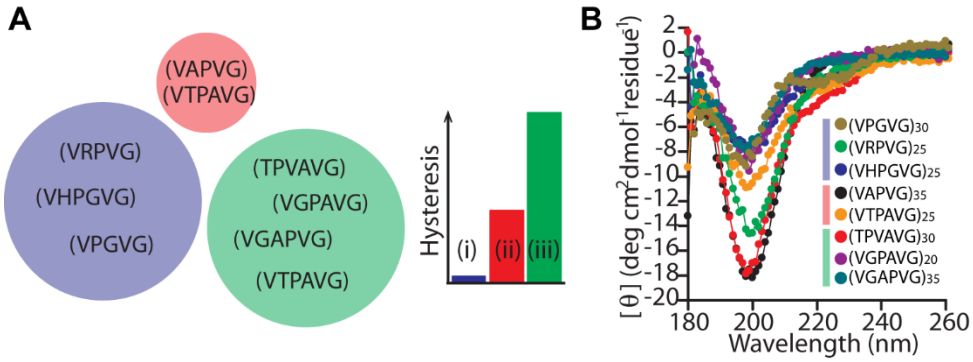

Fig. S4. LCST IDPPs exhibit CD spectra characteristic of intrinsic disorder and regardless of their hysteretic nature. (**A**) Classification of motifs according to the type of hysteretic phase behavior that is encoded in their sequence. (**B**) Circular dichroism data at 25 °C for each motif in (A). Data in (B) was adapted from our previous publication (Fig. 3 in ref (*10*)). Mean Residue Ellipticity (Θ)values are shown as $\Theta*10^{-3}$ for simplicity.

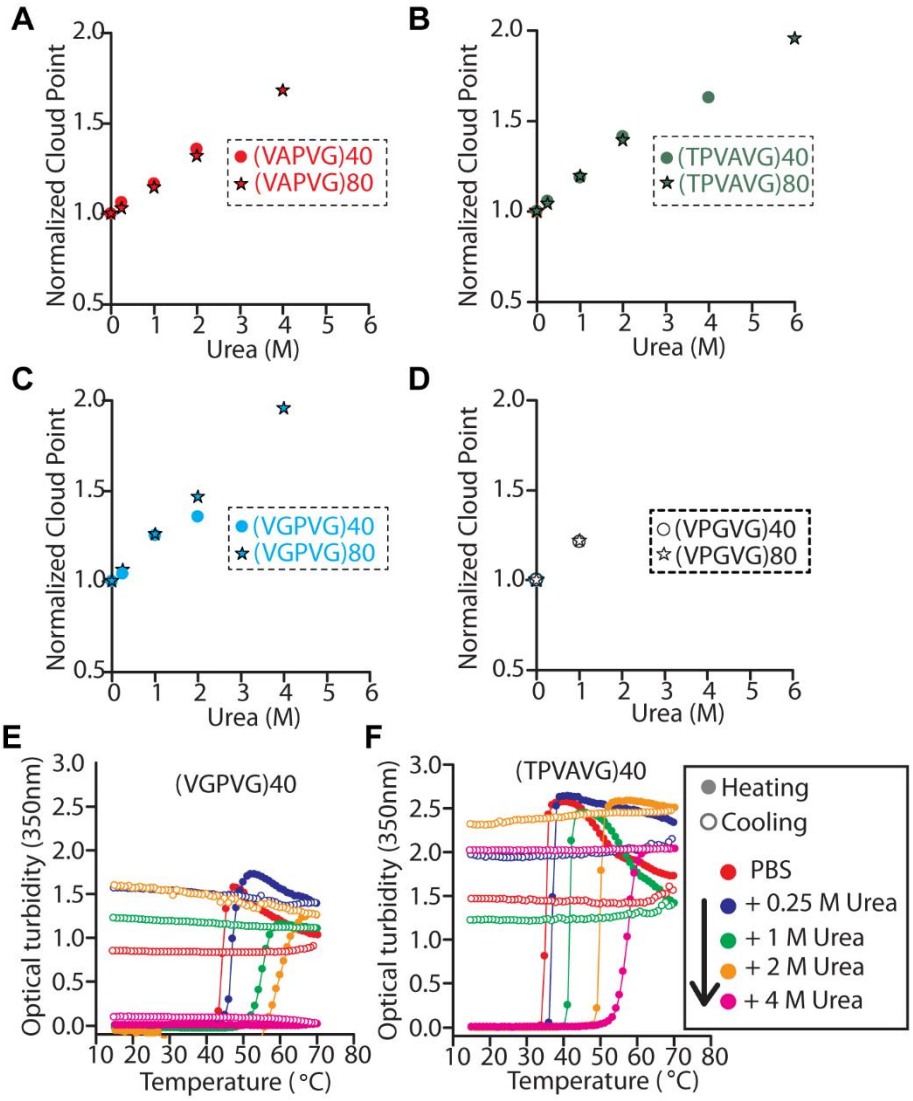

Fig. S5. Effect of urea on the hysteretic phase behavior of IDPPs. (**A-C**) Same data as in Fig. 3A but split data into multiple panels to facilitate data visualization (**D**) Effect of urea on (non-hysteretic) polymers of (VPGVG), which is the canonical ELP motif. (**E-F**) Raw phase behavior (turbidimetry) profiles as a function of urea for $(VGPVG)_{40}$ and $(TPVAVG)_{40}$. Note in (E) that because urea increases the T_{cp} on heating, at 4M urea we are unable to trigger a phase transition for $(VGPVG)_{40}$.

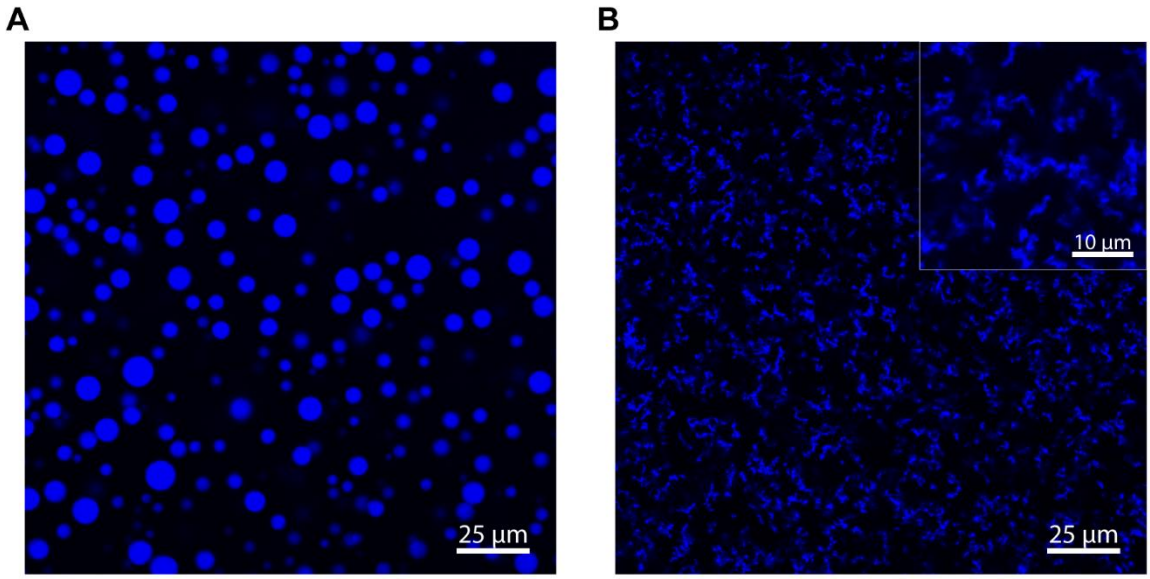

Fig. S6. Imaging of nonhysteretic and hysteretic IDPPs upon phase separation. Single plane confocal images of $(VPGVG)_{80}$ and $(VPAVG)_{45}$ (both at 400 µM, PBS) taken after 5 min of equilibration at 35 ºC (a temperature above their T_{cp}). (**A**) Non-hysteretic $(VPGVG)_{80}$ coacervates into liquid-like droplets commonly observed for IDPs. (**B**) Rather than coalesce into increasingly larger liquid-droplets, hysteretic $(VPAVG)_{45}$ forms a mechanically weak (easily disrupted by shear) arrested network. Similar arrested networks were recently reported by Glassman and Olsen (*27*)

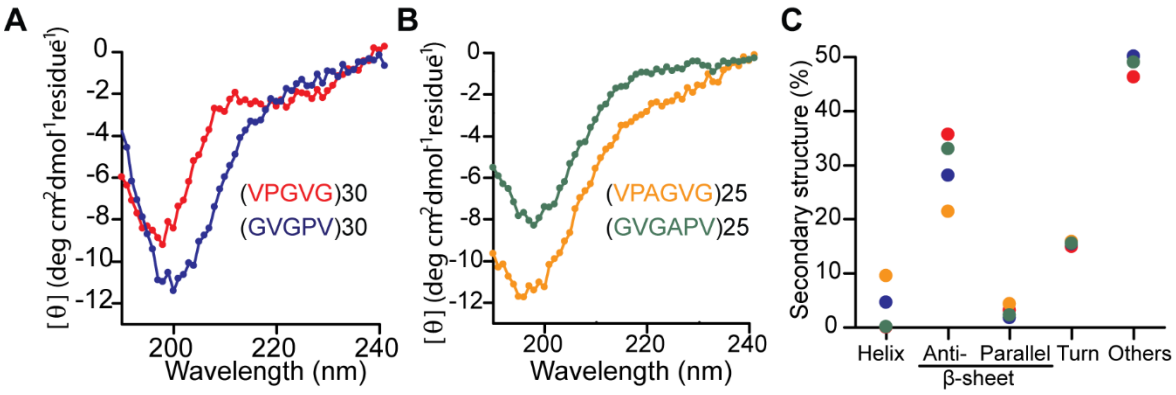

Fig. S7. Secondary structure of IDPPs related by sequence reversal at the repeat level. (**A-B**) Circular dichroism spectra for two pairs of motifs that are related by sequence reversal (corresponding to Fig. 4C-D in the main manuscript). Mean Residue Ellipticity (Θ)values are shown as $\Theta*10^{-3}$ for simplicity. (**C**) Quantification of secondary structure motifs for the CD spectra in (A) and (B) using the BestSel algorithm (*52*). Despite sequence reversal, IDPPs remain predominantly unstructured (~50% classified as "others").

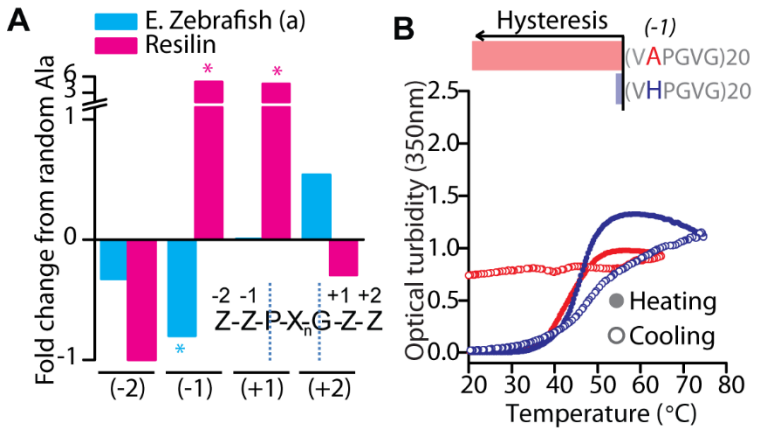

Fig. S8. Steric hindrance at the residue position preceding a P-Xn-G motif influences hysteresis. The occurrence of Gly — the only residue without a side chain— at this position was shown to be a likely modulator of the assembly behavior of IDPs (Fig. 4) as it discriminates between IDP's that go down the coacervation (no Gly) or the fibrillar (with Gly) pathway. Here, we show that this is likely due to the little steric hindrance offered by Gly, since the residue with the next smallest side chain, Ala, can also modulate hysteresis. Ala is relatively abundant in mammalian tropoelastin (~20%), but it is mostly found as poly(Ala) in the crosslinking domains, so the analysis of its distribution around the P-X_n-G motif would not be informative. However, tropoelastin from Zebrafish lacks such Ala enrichment in the crosslinking domains —it also has very small crosslinking domains— and has an overall Ala content of ~7.4%. (**A**) We quantified the enrichment or depletion of Ala —fold change from random occurrence of Ala— in residue positions surrounding P-X_n-G motifs using the same approach previously described for Gly. The residue N-terminal to P-X_n-G was the only surrounding position where we found a significant (p-value<0.001) deviation from random, with Ala being highly avoided. Resilin, in contrast, with an overall Ala content of 4.9% was highly biased to favor the occurrence of Ala at this position. Unlike tropoelastin, however, the LCST behavior of resilin displays hysteresis (*53*). (**B**) Temperature-dependent turbidimetry of IDPPs composed of motifs wherein Ala occurs one residue N-terminal to P-X_n-G and corresponding mutant polymers wherein Ala was substituted by a bulkier amino acid. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 µM, except for and VAPGVG (+ 1 M NaCl).

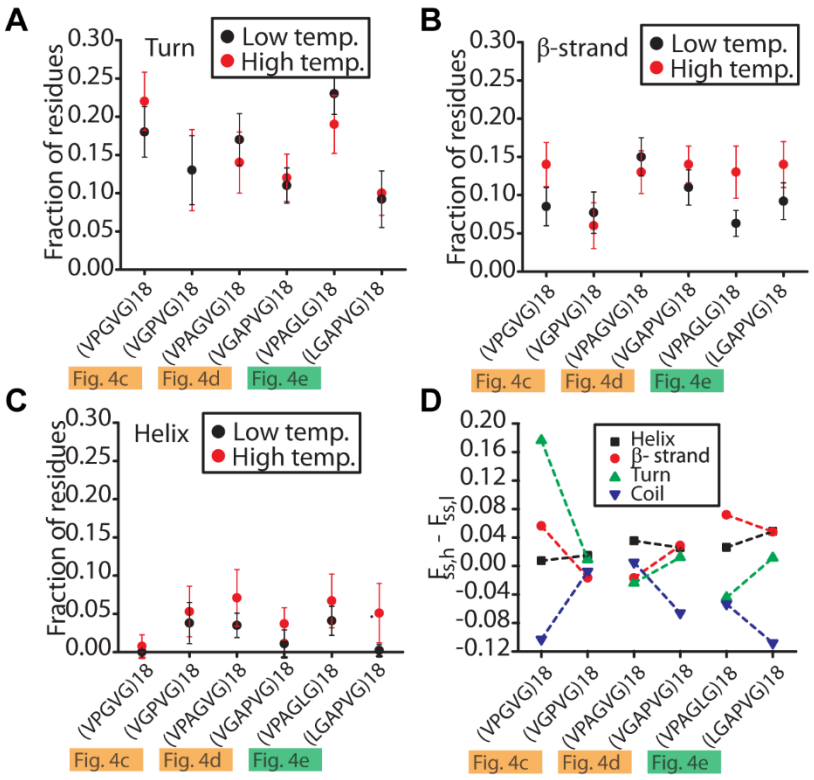

Fig. S9. Secondary structure preferences calculated from single-chain IDPP simulations at low and high temperatures. (**A**)-(**C**) Fraction of residues in the indicated IDPPs that are part of a turn (A), a β-strand (B) or a helix (C), at high and low temperatures. Specific temperatures are described in the methods section. Figure 5A shows data for unstructured motifs that account for the most fraction of residues in these IDPPs. (**D**) Differential between the assigned fractions (F_{ss}, fraction of secondary structure) at high and low temperatures for each structural motif, calculated as $F_{ss,\ high\ temp} - F_{ss,\ low\ temp}$. Note at the single-chain level we only see small temperature-dependent changes in secondary structure propensities that do not distinguish between non-hysteretic and hysteretic IDPPs.