

Syntax of Phase Transition Peptide Polymers with LCST and UCST Behavior

by

Felipe Garcia Quiroz

Department of Biomedical Engineering
Duke University

Date: _____

Approved:

Ashutosh Chilkoti, Supervisor

Gabriel Lopez

Kam Leong

Mark Dewhirst

Xuanhe Zhao

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biomedical Engineering in the Graduate School
of Duke University

2013

ABSTRACT

Syntax of Phase Transition Peptide Polymers with LCST and UCST Behavior

by

Felipe Garcia Quiroz

Department of Biomedical Engineering
Duke University

Date: _____

Approved:

Ashutosh Chilkoti, Supervisor

Gabriel Lopez

Kam Leong

Mark Dewhirst

Xuanhe Zhao

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biomedical Engineering in the Graduate School
of Duke University

2013

Copyright by
Felipe Garcia Quiroz
2013

Abstract

“Smart” polymers that sense stimuli in aqueous environments and that respond with a pronounced change in their solvation are of great utility in biotechnology and medicine. Currently, however, only few peptide polymers are known to display this behavior. Here, we uncover the syntax — defined as the arrangement of amino acids (letters) into repeat units (words) that have a functional behavior of interest — of a novel and extensive family of genetically encoded “smart” peptide polymers, termed syntactomers, that dictates their ability to undergo a soluble to insoluble phase transition at temperatures above a lower critical solution temperature (LCST) or below an upper critical solution temperature (UCST). We show that this syntax ranges from phase transition polymers composed of simple repeats of a few amino acids to polymers whose syntax resembles the complex sequence of peptide drugs and protein domains that exhibit dual levels of function, as seen by their stimulus responsiveness and biological activity. This seamless fusion of materials and protein design embodied by syntactomers promises, we hope, a new generation of designer polymers with multiple levels of embedded functionality that should lead to new functional materials of broad interest.

Dedication

A mi madre, mi hermano y mis tías.

Contents

Abstract	iv
List of Tables	x
List of Figures	xii
Acknowledgements	xxvi
1. Introduction	1
2. Chapter 1: Protein engineering for regenerative medicine and cancer therapy.....	5
I. Introduction	5
II. Engineering of proteins and protein-based materials	6
2.1. Bioinspired design.....	6
2.2. Protein engineering tools.....	7
III. Regenerative medicine and cancer therapy	8
IV. Proteins and protein-based materials for regenerative medicine and cancer therapy	9
V. Perspective.....	12
3. Chapter 2: Overview of LCST and UCST polymers.....	14
3.1. Elastin-inspired, LCST peptide polymers	15
3.1.1. The structure and function of elastin: elasticity and tissue remodeling.....	16
The assembly of elastic fibers: tropoelastin coacervation and crosslinking	20
Mechanical properties of native elastin	23
Physiological role of elastin as a regulator of ECM remodeling and inflammation	26

3.1.2. Synthesis, production and expression systems of elastin-inspired, LCST peptide polymers.....	27
Design and engineering of elastin biopolymers	27
Synthesis of genes encoding elastin biopolymers	31
Expression and purification of recombinant elastin biopolymers	33
3.1.3. The LCST phase behavior of elastin sequences.....	43
Tropoelastin coacervation and self-assembly	44
Self-assembly of Elastin-Derived Polypeptides.....	46
Inverse phase separation and self-assembly of ELPs.....	49
3.2. Resilin-derived UCST peptide polymers	53
4. Chapter 3: Syntax of LCST peptide polymers	56
4.1. Syntax of “smart” peptide polymers governs their function	56
4.2. Supplemental information	76
4.2.1. Supplementary discussion.....	76
4.2.2. Supplemental figures.....	80
4.3. Materials and Methods	93
4.3.1. Materials	93
4.3.2. Bioinformatics studies	94
Digitalization of protein sequences	96
Quantitative analysis of P-X _n -G motifs and surrounding Gly residues.....	97
Distance between P-X _n -G motifs and hydrophathy distribution	105
Identification of Pro-rich (only) proteins.....	106

4.3.3.	Genetically encoded synthesis of peptide polymers.....	107
	Overlap-extension rolling circle amplification (OERCA)	107
	Recursive Directional Ligation by Plasmid reconstruction (PRe-RDL)	112
	Generation and synthesis of a protein-sized, randomized peptide polymer.....	114
4.3.4.	Expression and purification of peptide polymers	126
4.3.5.	Characterization of the phase transition behavior and secondary structure of peptide polymers.....	127
5.	Chapter 4: Syntax of UCST peptide polymers	129
5.1.	Key residue interactions govern the phase behavior of peptide polymers.....	130
5.2.	Supplemental Figures	148
5.3.	Materials and methods	149
5.3.1.	Materials	149
5.3.2.	Genetically encoded synthesis of peptide polymers.....	149
	OERCA and PRe-RDL	149
5.3.3.	Expression and purification of peptide polymers	155
5.3.4.	Characterization of the phase transition behavior and secondary structure of peptide polymers.....	156
6.	Chapter 5: LCST and UCST peptide polymers with intrinsic biological function..	158
6.1.	LCST peptide polymers with intrinsic biological function.....	159
6.2.	UCST peptide polymers with intrinsic biological function.....	166
6.3.	Perspective.....	171
6.4.	Materials and methods	174
6.4.1.	Materials	174

6.4.2. Genetically encoded synthesis of peptide polymers	174
OERCA and PRe-RDL	174
6.4.3. Expression and purification of peptide polymers	178
6.4.4. Evaluation of the anti-tumor activity of SM1 and SM2	180
6.4.5. Evaluation of the cell adhesive activity of RGD-based, UCST peptide polymers	181
Conclusion	182
References	184
Biography	207

List of Tables

Table 1. Expression systems for the production of several types of elastin-based, LCST peptide polymers. Leader amino acid sequences used for expression or purification purposes are indicated when applicable. Typical yields are reported.....	38
Table 2. P-values for the FCRG observed at residue positions surrounding P-X _n -G motifs in Pro- and Gly-rich IDPs. FCRG values are shown in Figure 5B.....	77
Table 3. Full names and accession numbers of all proteins analyzed in this manuscript.....	94
Table 4. DNA sequence information for the synthesis of genes encoding for peptide polymers spanning P-X _n -G motifs from n=0 to n=4. Both oligomers and primers were 5'-phosphorylated. DNA oligomers with 4-10 copies of the starting oligomer sequence (as shown) were constructed by OERCA, with the exception of polymers of the peptide LGAPVG (see Table 5).....	108
Table 5. DNA sequence information for the synthesis of genes encoding peptide polymers via Pre-RDL. Oligomers were 5'-phosphorylated.	113
Table 6. Codon-optimized gene (5'-3') encoding for the protein-sized polymer generated by Script 2. The 5' and 3' untranslated regions that were incorporated for subcloning are not shown.....	125
Table 7. DNA sequence information for the synthesis of genes encoding for peptide polymers spanning P-X _n -G motifs from n=0 to n=4 and designed to incorporate a high density of residues capable of hydrogen bonding or that experience electrostatic interactions. Both oligomers and primers were 5'-phosphorylated. DNA oligomers with 4-10 copies of the starting oligomer sequence (as shown) were constructed by OERCA	150
Table 8. DNA sequence information for the synthesis of genes encoding phase transition peptide polymers via Pre-RDL. Oligomers were 5'-phosphorylated.....	153
Table 9. Design of YIGSR-based peptide polymers with various types of phase behavior. Non-native residues in the repeat unit are highlighted in red. The hydrophathy was averaged for all residues using Urry's scale of hydrophathy [207]. As a reference, the hydrophathy of the "canonical" UCST motif GRGDSPY is 45.3 °C, whereas the canonical LCST motif VPGAGVPGGG averages 36.1 °C.	171

Table 10. Synthesis of genes encoding for drug-like syntactomers by OERCA. Both oligomers and primers were 5'-phosphorylated. 174

Table 11. DNA sequence information for the synthesis of genes encoding drug-like syntactomers via Pre-RDL. Oligomers were 5'-phosphorylated. 177

List of Figures

Figure 1. Current pictures of the subspaces of amino acid sequences that polymerize into peptide polymers that display LCST and UCST phase behavior. The amazingly complex space of proteinogenic sequences comprises the amino acid sequences of both folded peptides and proteins as well as the sequences of unstructured peptides and proteins (that is, IDPs). Within this humungous space, we know that there is a small subspace populated by tropoelastin-derived sequences, namely derivatives of the canonical VPGXG motif, which best describes the current picture of the sequence space of LCST peptide motifs —here depicted in red. An even narrower subspace (here depicted in blue) populated by amino acid sequences derived from UCST-exhibiting proteins, namely collagen and resilin, is known to exist at a distance from the LCST subspace.2

Figure 2. The sequence and domain structure of tropoelastin. A) The sequence of human tropoelastin presents Lys-containing domains (yellow) for crosslinking and is typically divided into Gly- and Pro-rich hydrophobic domains, Lys-Ala (KA) and Lys-Pro (KP) crosslinking domains. Domains (exons) 34-35 are absent in the human gene. KP domains are unlikely to participate in tetrafunctional crosslinks due to their low helical content [64, 81, 82]. Diagram of the domain structure of tropoelastin is from [81]. B) Map of the distribution of the pentapeptide motif VPGXG, descriptive of the elastin biopolymers known as elastin-like polypeptides, along the sequence of various elastin proteins. This motif was generalized by substituting a guest residue X (i.e., any amino acid; except Proline) for the fourth residue (Valine) of the pentapeptide VPGVG identified in the seminal work by Urry.19

Figure 3. Schematic depicting the design of several elastin biopolymers. The majority of these biopolymers are now produced by recombinant synthesis in *E. coli*. ELP: elastin-like polypeptides. SHEL: synthetic human elastin. EDPs: elastin-derived polypeptides.28

Figure 4. The amino acid sequence of Pro- and Gly-rich and Pro-rich (only) IDPs is characterized by discrete arrangements of Pro and Gly residues. (A) Abundance of individual P-X_n-G motifs among Pro- and Gly-rich IDPs, where X is any amino acid different from Pro and Gly and *n* varies from 0 to 4, relative to the total number of P-X_n-G motifs per protein. Asterisks indicate statistical significance (*p* < 0.001) for over-represented motifs compared with the expected distribution in a set of randomized proteins with identical amino acid content (Figure 13 found in *Supplemental Figures*). The inset shows the characteristic recurrence of P-X_n-G motifs in elastin and resilin, as evidenced through a digitalization of their protein sequences in which residues —other than Gly— participating in a P-X_n-G motif are assigned a value of 1, while all other

residues are set to 0. (B) Enrichment or depletion of Gly in residue positions surrounding P-X_n-G motifs among Pro- and Gly-rich IDPs, expressed as a fold change from the random occurrence of Gly based on total Gly content. Asterisks indicate significant (p < 0.001) divergence from a random distribution (see *Supplementary discussion*). (C) Ratio of the median fraction of Pro-Gly motifs to P-X₄-G motifs across 1000 randomized versions (see *Supplementary discussion*) of each Pro and Gly-rich (median % G of 32.8) and Pro-rich (only) (median % G of 6.7) IDP. (D) Analysis of P-X_n-G motifs, as in (A), among Pro-rich IDPs that are not rich in Gly. “E. Homo” is elastin from *Homo sapiens*. Full protein names and accession numbers are reported in Table 3 (found in *Materials and Methods*). 59

Figure 5. Diverse repeat motifs that contain Pro and Gly lead to intrinsically disordered peptide polymers that display LCST behavior. (A-E) Temperature-dependent turbidimetry revealed that peptide polymers with periodic Pro and Gly residues arranged as P-X_n-G units, where n=0 (A), 1 (B), 2 (C), 3 and 4 (D), and having pentapeptide, hexapeptide and nonapeptide repeat units display thermally triggered LCST phase transition behavior. (E) These peptide polymers lack ordered secondary structures as shown by their CD spectra (at 25 °C) that is characteristic of IDPs. (F) Three representative peptide polymers show the three types of phase transition behavior exhibited by all polymers in A-C, corresponding to the degree of hysteresis in the reversibility of their thermally triggered phase transition: (i) zero, (ii) finite, and (iii) heat-sensitive, infinite hysteresis. (G) Polymers that behave as in (iii) display irreversible aggregation at high temperatures but undergo phase transition with zero hysteresis below a critical threshold temperature — arrows indicate the maximum temperature in the turbidity versus temperature profile. Figure 7 shows other peptide polymers with this behavior. (H) CD spectra as a function of temperature for polymer in H-(ii). The LCST was determined using temperature-dependent turbidimetry (Figure 17A found in *Supplemental Figures*). (I) Hysteresis in the reversibility of the phase transition behavior of three peptide polymers with a V-1PVG motif and with Gly, Ala or Arg at the -1 position. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μM, except for VRPVG (+ 1M NaCl), VAPGVG (+ 0.5 M NaCl), TVPGAG (+2 M NaCl), APGVG (+2 M NaCl) and VPSALYGVG (+ 8 M urea). CD studies were conducted in water at a polypeptide concentration of 5 μM. 62

Figure 6. (A) We implemented a method to automatically generate a protein-sized polymer with a sequence that conforms to the amino acid patterns of non-fibrillar IDPs and wherein unconstrained Z residues are randomly selected to yield an average hydropathy that corresponds to an LCST of 37 °C. (B) Hydropathy distribution for all Z residue positions (B) in the resulting protein-sized polymer — 240 amino acids in length.

(C) Temperature-dependent turbidimetry in PBS at a polypeptide concentration of 50 μM indicated an LCST of $\sim 40^\circ\text{C}$. (D) The CD spectrum of this protein-sized polymer is characteristic of intrinsically disordered proteins. The CD data was collected in water at 5 μM and at 25°C . (E) Full amino acid sequence. The sequence includes the N terminal peptide SKGP and the C-terminal tripeptide GWP that were appended at the gene level before recombinant expression; residues shown in black correspond to the randomly assigned amino acids..... 63

Figure 7. Peptide polymers that exhibit LCST behavior may display temperature-dependent hysteresis in their phase transition behavior. (A-F) Temperature-dependent turbidimetry for six different peptide polymers revealed a form of heat-sensitive hysteresis. These polymers 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). Peptide polymers with repeat unit VTPAVG (E) exhibit a complex type of phase transition behavior with heat-sensitive infinite hysteresis above the threshold temperature (see red curve), and finite hysteresis – we typically observed zero hysteresis for other sequences – below this temperature (see blue curve). Note that the finite hysteresis of this peptide polymer 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 the peptide polymer 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, we note that this peptide polymer – like all others reported in this manuscript – was purified exploiting the reversibility of its phase behavior in response to changes in buffer ionic strength. The phase behavior was characterized in PBS at a polypeptide concentration of 50 μM 65

Figure 8. Molecular weight influences the hysteretic phase behavior of LCST peptide polymers. (A) Polymers composed of AVPGVG repeats, a motif that upon polymerization exhibits no hysteresis, display lower LCST values as the length of the polymer is increased, but even the longest polymer shows no hysteresis in its phase behavior. (B-C) Polymers composed of 20, 30 and 40 repeats of the pentapeptide VAPVG, a motif that upon polymerization exhibits finite hysteresis, also display progressively lower LCST values and with evidence of hysteretic phase behavior (B), but more than 20 copies of the motif are needed to obtain polymers with marked hysteresis, as evidenced by a sharp transition on cooling. The degree of thermal hysteresis is also highly reproducible over multiple cycles of heating and cooling (Figure 18 found in *Supplemental Figures*). Moreover, peptide polymers with 40 and more (data not shown)

copies of the repeat are highly kinetically-stable (C) as their LCST on cooling is almost unaffected by a 10 fold decrease in the cooling rate from 1 °C/min to 0.1 °C/min. The LCST's on heating (measured at a heating rate of 1 °C/min) for each temperature scan are depicted in (C) as dotted lines, as the curves are invariable. (D) Polymers composed of 20 and 40 repeats of the hexapeptide TPVAVG, a motif that upon polymerization exhibits infinite hysteresis, show the expected molecular weight dependence of the LCST but this increase in length also influenced 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 fail to exhibit a second phase transition..... 66

Figure 9. LCST peptide polymers exhibit properties intermediate between synthetic polymers and proteins. (A) Backbone-reversed peptide polymers present identical patterns of amino acid side chains, which we illustrate with the structure of a pentapeptide motif and its retro-motif as observed in the crystal structures of two different proteins (PDB id 3MKR_B and 1OZP, respectively). The images were rendered using PyMOL (<http://pymol.org/>). (B-D) Temperature-dependent turbidimetry showed profound changes in the phase behavior of peptide polymers on backbone-reversal, as seen by the change in the hysteresis of the phase transition, despite the intrinsically disordered nature of the parent and retro polymers below their LCST according to CD spectroscopy (Figure 5E). All turbidity measurements were conducted in PBS (unless otherwise indicated) at a polypeptide concentration of 50 µM. CD studies were conducted in water at a polypeptide concentration of 5 µM and at 25 °C. 71

Figure 10. New forms of hysteretic LCST phase behavior are useful in controlling the thermally-triggered nanoassembly of peptide polymers. (A-B) By forming diblock copolymers composed of a hydrophilic ELP block (where X alternates between A and G) and a hydrophobic block that repeats a hysteretic LCST motif with either heat-sensitive, infinite hysteresis (A) or finite hysteresis (B), we synthesized self-assembling nanoparticles that remain assembled despite significant cooling below their critical assembly temperature. The assembly on heating is evidenced by changes in absorbance (filled circles) and hydrodynamic radius (diamonds), and the stability of the particles on cooling is shown by stable absorbance measurements (open circles) at temperatures below the LCST of the ELP corona (~58 °C) and the apparent LCST of the more hydrophobic block. (C) A diblock copolymer with the same composition and block ratio as in (A), but having half the number of repeats in the core –that is, a constant ratio between the size of the core and the corona: 20/40 vs. 40/80 diblocks– fails to remain

assembled through the bulk aggregation event experienced above the LCST of the ELP in the corona. 74

Figure 11. Distribution of the canonical VPGXG motif and its analogs IPGXG and LPGXG along the primary amino acid sequence of tropoelastin from different species. The localization of these motifs is digitized such that all residues forming part of a motif are assigned a value of 1 and any other residues are set to 0. These three motifs altogether account for ~10-20% of the amino acid residues in these proteins. Of these, VPGXG is the most abundant motif. 81

Figure 12. (A) Pro and Gly content of all Pro- and Gly-rich IDPs shown in the figure legend. (B) Average distance between P-X_n-G and P-X_n-A motifs (for n=0-1) across these proteins. Note that some proteins lack P-X_n-A motifs (indicated by asterisks in B). 82

Figure 13. Analysis of randomized Pro- and Gly-rich IDPs to assess the statistical frequency of motifs with Pro and Gly residues in close proximity. First, we generated 1000 randomized versions of each IDP listed in the figure caption (see supplementary methods for details) using the amino acid frequencies characteristic of each protein. We then assessed the abundance of individual P-X_n-G motifs (for n=0-4) relative to the total number of P-X_n-G motifs in both the real (parent) protein and all randomized versions. (A) Fraction of P-X_n-G motifs in the set of randomized and real proteins. For the randomized proteins we plotted the median value of the entire data set. Asterisks indicate statistical significance (p-value<0.001) for over-represented motifs compared with randomized proteins. We then repeated this same analysis for an additional set of Pro-rich (only) IDPs that are not rich in Gly. (B) Randomized versions of these proteins are almost unbiased in the frequency with which different P-X_n-G motifs occur along the protein amino acid sequence. We deemed a result as statistically significant if the p-value was smaller than 0.001. 83

Figure 14. We identified Pro-rich but Gly-poor proteins that display P-X_n-G motifs with n=2, 3 and 4. (A) Quantification of the Pro and Gly content of these proteins (resilin and dragline silk were also included for reference). (B) Abundance of individual P-X_n-G motifs with n=0-4 relative to the total number of motifs per protein. Unlike our previous data set of Pro- and Gly-rich IDPs, these Pro-rich proteins are highly biased toward P-X_n-G motifs with n=2, 3 and 4. The absolute abundance of these motifs in the overall protein sequence is comparable to the abundance of P-X₄-G motifs in resilin (data not shown). (C) The digitalization of the protein sequence for two of these proteins, in which residues —other than Gly—participating in a P-X_n-G motif were assigned a value of 1, while all other residues were set to 0, reveals that these motifs recur at high frequency in

defined regions of the proteins. To facilitate the visualization of the most abundant P-X_n-G units, the squared signals corresponding to the predominant type of motif (n=2 or 3 in these two examples) were colored to differentiate them from the blue, background P-X_n-G motifs. 84

Figure 15. Pro-rich proteins use repeats of P-X₄-G motifs for the presentation of bioactive peptides. Here, we present the digital maps of P-X_n-G motifs across the sequence of these proteins — generated as in Figure 14. P-X₄-G motifs were highlighted in red to facilitate their identification. The right panel for each protein shows the hydrophathy distribution (depicted by the Kyte-Doolittle scale of hydrophobicity) for the residues that form these repeats (i.e., X₄ residues and two Z residues N-terminal and C-terminal to the P-X₄-G motifs), which suggests that different residues in these P-X₄-G motifs can span a broad range of hydrophathies. The two neuroactive proteins FARXamide-related neuropeptides and PRQFVamide display repeats of the bioactive peptides PFLRF and PRQFV, respectively, embedded within P-X₄-G motifs. We observed a similar localized region of P-X₄-G units with highly conserved X₄ residues (right panel) in a transcription factor from yeast (SPT5). 86

Figure 16. Frogs have unusual tropoelastins that are not biased towards a Pro-Gly dipeptide. The expression of two different tropoelastin genes with markedly different amino acid composition and length was recently reported in the amphibian *Xenopus Tropicalis* (XT) [190]. (A) Pro and Gly content for the two tropoelastin proteins from XT and other tropoelastins from bovine and *Xenopus Laevis* (XL). Gly content is a major difference, since Tropoelastin2 has a relatively low Gly content compared with its companion Tropoelastin1 and with all known tropoelastins from other species. A similar tropoelastin has been identified based on bioinformatics analysis in the tetraploid XL (encoded by two different genes, a and b) [190]. (B) Abundance of individual P-X_n-G motifs with n=0-4 relative to the total number of motifs per protein. As in the case of Pro-rich IDP's, tropoelastins with relatively low Gly content are not enriched in Pro-Gly dipeptides. These proteins show a random-like distribution of Pro-X_n-Gly motifs for n values between 0 and 4. Asterisks indicate statistical significance (p-value<0.001) for over-represented motifs compared with randomized proteins (see *Materials and Methods* for details). 88

Figure 17. The heat-sensitive, infinite hysteresis of LCST peptide polymers arises from the emergence of ordered secondary structures that stabilize the insoluble phase. (A-D) The same sample for each polymer was analyzed by temperature-dependent turbidimetry (left panel) and CD spectroscopy as a function of temperature (right panel). Filled arrows in the left panels — turbidity versus solution temperature plots— indicate

the temperature at which CD spectra were recorded upon step-wise heating the sample, while open arrows indicate the temperature at which CD spectra was acquired upon cooling. Whereas a peptide polymer that displays finite hysteresis (A, left) rapidly recovers its conformational disorder on lowering the temperature below the phase transition temperature in accordance with its degree of hysteresis (A, right), we show that peptide polymers with heat-sensitive infinite hysteresis (B-D) undergo conformational changes that persist on cooling. We note that a polymer herein shown to display a very large degree of hysteresis — such that we were unable to observe its reversibility below any given temperature threshold accessible with our instruments (Figure 7F) — also displays the more dramatic ordering on aggregation as seen by the emergence of negative peaks at ~208 nm and ~220 nm at the plateau region of the turbidity profile that are indicative of a newly formed α -helical structure (B). Turbidity and CD data were acquired in water at a polypeptide concentration of 5 μ M..... 90

Figure 18. The degree of thermal hysteresis is highly reproducible over multiple heating and cooling cycles. The peptide polymer was characterized at a concentration of 50 μ M in PBS, at a heating and cooling rate of 1 $^{\circ}$ C/min..... 90

Figure 19. Elastin-like peptide polymers undergo aggregation at the LCST with minor conformational changes — seen by a small decrease in the random coil peak at ~197 nm. (A-B) CD spectroscopy of two elastin-like mutants (at 5 μ M in water) as a function of temperature and accompanying turbidity data (insets) for the same sample. Symbols and arrows indicate the temperatures at which CD spectroscopy was carried out. The most significant conformational changes occur after the polymers are fully aggregated, indicating that major reorganization occurs subsequent to dehydration and collapse of individual polymers chains that trigger the observable LCST, and are mediated by inter-chain interactions as seen by the large decrease in the intensity of the negative peak at ~197 nm. Data were adapted from Ref. [208]..... 91

Figure 20. Gly residues preceding a P-X_n-G unit perturb the behavior of Pro-rich “smart” peptide polymers. We mutated two different LCST peptide polymers to incorporate a Gly residue N-terminal to the repeating P-X_n-G unit, since the occurrence of Gly at this position is a key differentiating factor between fibrillar and non-fibrillar, Pro- and Gly-rich IDPs (Figure 4B). (A-B) Temperature-dependent turbidity data for the parent and mutant peptide polymers showed that these mutations affected the reversibility of the phase transition (A) and the propensity for aggregation (B). Note that (VGPAVG)-20 in (B) displays a lower LCST than a longer polymer of VTPAVG (where N=25), despite the similarity in hydrophobicity between Gly and Thr by both Kyte-Doolittle’s and Urry’s scale of hydrophobicity. In addition to these changes in phase behavior, we observed

interesting changes in the ability of *E. coli* to produce these peptide polymers solubly, as if the Gly-mutants displayed the characteristic expression issues (that is, insolubility in the crowded intracellular space) of fibrillar IDP's. (VGPVG)-30, for instance, which contains the abundant GPXG motif found in collagen, expresses as an insoluble peptide polymer but once purified (i.e., after denaturation and removal of denaturing agents), it displays high solubility in PBS reminiscent of canonical ELPs. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μ M. 91

Figure 21. Steric hindrance at the residue position preceding a P-X_n-G motif modulates 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 IDP's (Figure 4B) 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 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 was not interesting to us. However, tropoelastin from Zebrafish lacks such Ala enrichment in the crosslinking domains —it actually 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 in for Gly (see *Materials and Methods*). The residue N-terminal to P-X_n-G was the only surrounding position where we found a significant (p-value<0.01) 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 large hysteresis [184]. (B-C) Temperature-dependent turbidimetry of “smart” peptide polymers 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. We observed that whereas the parent (Ala-containing) polymers display large hysteresis in their phase transition behavior, the mutant polymers display little to no hysteresis. In (B) we also note a larger degree of hysteresis for peptide polymers with Gly at this position, as compared with Ala, which is in accordance with the reduced bulkiness of the side chain. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μ M, except for VRPVG (+ 1M NaCl) and VAPGVG (+ 1 M NaCl). 92

Figure 22. Hydrophathy distribution of amino acids that surround the repeating Pro-Gly dipeptides in Pro- and Gly-rich IDP's that are known to display LCST. We noted that residues in the vicinity of Pro-Gly dipeptides, across the broader picture of evolutionary

distant tropoelastin proteins and resilin, cover a broader range of hydrophathies than what is evident by studying archetypical LCST tropoelastins (e.g., those from homo sapiens and bovine). Solid circles correspond to the index of individual residues, and the open squares indicate the mean hydrophathy index for all residues at each Z position. Each box delimits the 25 and 75 percentile. The hydrophathy index was plotted using the Kyte-Doolittle scale of amino acid hydrophobicity [209] wherein hydrophobic amino acids have a positive index. The indices corresponding to Ile, Val, Gly and Arg are indicated by dotted lines..... 93

Figure 23. Resilin is built from amino acid repeats that may be tuned to exhibit UCST or LCST phase behavior. (Ai) We postulate that different residue interactions are at play for the design of UCST (charge-charge interactions, hydrogen bonding and salt bridges) and LCST (hydrophobic interactions) peptide polymers. (Aii) The primary structure of resilin (exon 1), which displays both UCST and LCST behavior, is composed of amino acid repeats. Arginine, aspartic acid and serine residues are highly conserved across the repeating units. This sequence alignment was adapted from [184].(B) A polymer composed of a fragment of the consensus repeat unit incorporates residue interactions that lead to UCST phase behavior upon polymerization (Bi), whereas a very similar repeat unit that excludes some of these interactions and dopes in additional hydrophobic residues readily overturns the UCST behavior of the original peptide polymer to yield a peptide polymer with LCST phase behavior (Bii). Turbidimetry data was measured in PBS at 50 μ M..... 131

Figure 24. Syntactomers with repeating P-X_n-G motifs (n=0-4) can be designed to exhibit UCST phase behavior under physiologically relevant conditions. (A-C) Turbidimetry data reveals the UCST behavior of this novel family of genetically encoded peptide polymers. (D) The unique amino acid composition of these polymers tolerates the absence of Pro. Data in (A-D) was acquired in PBS (pH 7.4) at 50 μ M, except for polymers of VPHSRNGG (+2 M NaCl), RGDSPYG (+1 M Urea) and VRPDG (PBS pH 2.0 supplemented with 1.5 M NaCl). (E) Representative example of the overexpression of these polymers (here, GRGDSPYG-20) upon IPTG induction in E. coli and the subsequent purification from the insoluble fraction by UCST transition cycling. Lane A is the cell lysate; lane B is the insoluble fraction and lane C is the result of a first cycle of purification. (F) CD spectroscopy in water at 5 μ M for a subset of these polymers revealed their unstructured character, as their spectra is reminiscent of the spectrum of the canonical VPGXG motif from the LCST-exhibiting tropoelastin..... 133

Figure 25. The UCST behavior of syntactomers is readily tuned by molecular weight, hydrophobicity and concentration. (A) These polymers exhibit fully reversible phase

transition behavior. (B) Molecular weight is a potent modulator of the UCST for relatively short polymers, but the effects become less pronounced as the molecular weight increases. (C) Mutations to the GRGDSPYG motif demonstrate that substitutions for more hydrophilic residues lead to lower UCST values. (D) The UCST is a function of the concentration of the polymer in solution. Turbidimetry data in (A-C) was acquired in PBS at 50 μ M. 135

Figure 26. The zwitterionic character of UCST syntactomers plays an important role in the physiological relevance of their behavior. (A) The response of zwitterionic syntactomers in PBS to molar amounts of additional NaCl is sequence-dependent. (B) Under acidic conditions, protonation of aspartic acid residues in the syntactomer result in positively charged polymers that fail to exhibit phase behavior in PBS, but increasing the buffer ionic strength by adding 0.5 M NaCl enables UCST phase behavior. (C) By mutating the aspartic acid residue in GRGDSPYG into an asparagine, the propensity to display UCST is greatly reduced and UCST phase behavior in PBS (without additional salt) is only observed if the molecular weight of the mutant polymer is significantly increased. (D-E) Syntax controls the UCST but the effects are modest (\sim 4-5 $^{\circ}$ C) compared with LCST syntactomers (Figure 9), at least in the two examples explored herein: (D) a motif and its corresponding sequence-reversed (retro) motif, and (E) peptide polymers with a conservative change in the spacing of the oppositely charged residues. (F) Syntax, however, may offer the possibility to control unusual behaviors in these peptide polymers as observed here for a syntactomer that exhibits both LCST and UCST wherein the LCST is lower than the UCST. Turbidimetry data was acquired at 50 μ M in PBS pH 7.4 unless otherwise indicated, except for data in (D) that was measured at 25 μ M and in (F) were PBS was supplemented with 3 M urea. 141

Figure 27. Reprogramming of a LCST syntactomer into a UCST syntactomer. Two residue positions in the LCST-exhibiting motif VAPVG (A) were selected to introduce parallel (Bi and Bii) and serial (C) mutations that might lead to UCST phase behavior. (A) Polymers of VAPVG exhibit LCST behavior in PBS pH 7.4. (Bi) Substituting R by A in that motif results in cationic polymers that exhibit LCST behavior at high ionic strength. (Bii) Substituting D by V in the parent motif results in anionic polymers that exhibit LCST behavior upon protonation and at high ionic strength. (C) Mutating both residues simultaneously leads to UCST phase behavior upon protonation and at high ionic strength. 144

Figure 28. Design of peptide nanoparticles self-assembled through the phase behavior of UCST syntactomers. (A-C) Dynamic light scattering data demonstrates that diblock copolymers composed of a UCST core-forming block and a LCST block (corona) self-

assemble into nanoparticles with hydrodynamic radius (R_h) of ~ 17-20 nm that then disassemble upon heating and in a reversible manner. (D-E) Turbidimetry (left axis) and dynamic light scattering (right axis) data for three additional diblock copolymers. (D) A diblock copolymer as in (A) but using a LCST block that is half the size, which leads to a core: corona block ratio of ~2:1, exhibits unusual transient bulk aggregation on disassembly. Turbidimetry data was measured at a heating rate of 0.1 °C/min — this event is almost imperceptible when heating at 1 °C/min (Figure 30). (E) Changing the composition of the LCST block for a more hydrophobic peptide polymer composed of VPGVG-30 abolishes self-assembly at low temperatures and results in bulk aggregation upon heating. Open circles correspond to cooling and the solid line to a second heating cycle used to assess reversibility. DLS data was acquired at 25 μ M (F) A diblock copolymer composed of two UCST blocks where the core-forming block displays a UCST around 40°C and the corona-forming block does not show an observable UCST (at least down to 4 °C) fails to self-assemble into nanoparticles at low temperatures and instead undergoes bulk aggregation. Both DLS and turbidimetry data were acquired at 50 μ M in PBS unless otherwise noted. 147

Figure 29. A fine balance between hydrophobic and polar interactions determines whether a syntactomer exhibits LCST or UCST phase behavior. Whereas polymers of VPHSRNGG exhibit UCST phase behavior in PBS at 50 μ M —supplemented with 2 M NaCl—, a slight modification to this motif that favors hydrophobic interactions, by substituting the C-terminal Gly for Leu, as in VPHSRNGL, results in polymers that exhibit LCST phase behavior in PBS at 50 μ M. The latter also exhibits LCST with increasing amounts of NaCl (data not shown). 148

Figure 30. Transient aggregation on disassembly of diblock copolymers composed of UCST-exhibiting hydrophobic cores and short LCST-exhibiting coronas shows a large kinetic dependence. This aggregation event is barely perceptible upon heating at a rate of 1 °C/min (solid circles), but aggregates are significantly stabilized with longer incubation times afforded by a 10 fold decrease in the heating rate (solid stars). In Figure 28D we also show that this event is highly concentration dependent. Turbidimetry data was obtained for 50 μ M samples in PBS. 149

Figure 31. A new picture of the sequence space of “smart” motifs. The research presented in chapters 3 and 4 draws a new picture of the sequence space of amino acid motifs that exhibit “smart” phase behavior upon polymerization. The original pictures (depicted by small, red and blue circles located at a distance in the proteinogenic space) are now replaced by a unifying picture that suggests the existence of a vast sequence space that encompasses amino acid motifs with opposing phase behaviors. This vast

space must also encompass peptide sequences with relevant biological functions, as will be demonstrated in this chapter..... 159

Figure 32. Syntactomers can be designed with a syntax that approaches the complexity of proteins. (A) Amino acid sequence of “smart” matrikine 1 (SM1) and SM2, where only SM1 consists of repeats of a bioactive motif GXXPG found in various extracellular matrix proteins, and SM2 is a scrambled control. (B) SM1 and SM2 displayed identical thermally triggered phase transition behavior in PBS (pH 7.4). (C) Individual tumor volumes measured 17 days after inoculation of 0.5×10^5 HT-1080 tumor cells into the leg of nude mice ($n=10$ per group) and using bioactive SM1 (at $350 \mu\text{M}$ in PBS), non-bioactive SM2 (at $350 \mu\text{M}$ in PBS) and PBS as vehicles for tumor cell inoculation. Asterisk (*) indicates statistical significance with a 95% confidence using a Bonferroni test to compare the mean tumor volumes with respect to the PBS cohort. The anti-tumor activity of SM1-24 and the inertness of SM2-24 were also confirmed in an independent experiment with HT-1080 tumor cells inoculated into the back of nude mice (Fig. S14). (D) Peptide sequences forming the bioactive domains of murine (PDB file: 1DY0) and human endostatin (PDB file: 1BNL), shown here in blue and red, respectively. (E-F) These novel syntactomers display CD spectra characteristic of IDPs (E) and exhibit LCST transition behavior (F) that is accompanied by the characteristic decrease in the intensity of the random-coil peak at ~ 197 nm, as shown by CD spectroscopy as a function of temperature in PBS at pH 6.4 (Fig. S16). CD spectra were measured in water at a polypeptide concentration of $5 \mu\text{M}$ and at 25°C . Images of the 3D structures of endostatin were rendered using PyMOL (<http://pymol.org/>). 163

Figure 33. A syntactomer inspired in the matrikine motif GXXPG, SM1-24 ($250 \mu\text{M}$ in PBS), prevented the grafting of 1×10^6 HT1080 tumor cells inoculated into the back of nude mice. Tumor volumes measured 19 days after tumor inoculation. A control syntactomer, SM2-24 ($250 \mu\text{M}$ in PBS), with a disrupted motif but identical phase transition behavior (Fig. 4) had no effect on tumor engraftment compared with a control group where cell culture medium was used as vehicle..... 164

Figure 34. A syntactomer based on the bioactive domain of murine Endostatin displays an inverse phase transition temperature reminiscent of other syntactomers with simpler syntax. (A) Temperature-dependent turbidimetry of mEndo1-6 at $5 \mu\text{M}$ in PBS (pH 6.4) and (B) corresponding —for the same sample— CD spectroscopy as a function of temperature. The phase transition of this syntactomer was accompanied by a decrease in the disorder of the polypeptide conformation, as evidenced by a greatly diminished random coil negative peak at ~ 197 nm. 164

Figure 35. The ability to design “smart” peptide polymers with monomer units that have local secondary structure propensities, as in human and murine endostatin, may enable the development of a broader set of “smart”, drug-like protein-polymers derived from the growing list of characterized polypeptide hormones. Here we show the primary sequence and predicted secondary structures of polypeptides composed of two repeat units of each hormone. The amino acid sequences of human and murine preptin, as well as the human polypeptide hormones obestatin, glicentin-related polypeptide (GRP), osteostatin, and ghrelin were modified with the C-terminal tripeptide PGG. Residues that are not part of the polypeptide hormone, typically added at the N or C-terminus of the repeat unit, are shown in orange. The UniprotKB/Swiss-Prot code for the pro-hormone corresponding to each peptide is shown under the name of each peptide. Secondary structures were predicted using the Jnet algorithm [233]. H: alpha helix, E: beta-sheet, and ‘-’: random coil. 165

Figure 36. Polymers of the cell adhesion motif GRGDSP display UCST phase behavior and bioactivity. (A) Two peptide polymers composed of 12 and 20 repeats of the octapeptide GRGDSPYG were studied over a range of concentrations. (B) Temperature-dependent turbidimetry. (C-D) Effect of these polymers on the cell adhesion of 0.5×10^5 PC-3 cells to regular 24-well tissue culture plates. The cells were allowed to adhere in the presence of varying concentrations of the peptide polymers for either 3 hours (C) or 18 hours (D), and the cultures were washed with PBS before quantifying cell metabolic activity (indicated by a metabolic product that absorbs strongly at ~ 550 nm). Control cultures ($0 \mu\text{M}$) were prepared by adding PBS supplemented with the maximum amount of residual urea (~ 0.1 M) in these experiments. 167

Figure 37. A syntactomer built from repeats of the GRGDSP cell adhesion motif exhibits UCST phase behavior within the window of mild hyperthermia ($37\text{-}42^\circ\text{C}$). Turbidimetry data was obtained at $25 \mu\text{M}$ in PBS or DMEM (Sigma) supplemented with 10% fetal bovine serum (Gibco). 168

Figure 38. Design of UCST peptide polymers composed of fibronectin and laminin peptides. (A) Fibronectin (PDB file: 1DY0) displays two short peptides (in red), the widely known GRGDSP motif and the synergy motif PHSRN, that act synergistically to control cell adhesion and engage a variety of integrins. (B) Turbidimetry data for a fibronectin-like syntactomer that mimics the distance between these two motifs reveals its reversible UCST phase behavior. VPHSRN is the native sequence found in fibronectin. (C) The synergy motif alone only exhibits UCST phase behavior under high ionic strength. (D) A syntactomer built from repeats of the laminin peptide DPGYIGSR also exhibits UCST phase behavior. Turbidimetry data was acquired in PBS at $50 \mu\text{M}$

unless otherwise indicated. Images of the 3D structures of fibronectin were rendered using PyMOL (<http://pymol.org/>) 169

Acknowledgements

To my advisor, Ashutosh Chilkoti, for his unconditional support, respect and guidance. His example nurtures my passion for Science, teaching and mentoring. I'm also grateful for the unique and motivating research environment that I found in the "Chilkoti Lab". Multiple graduate students in the group, in particular Wafa Hassouneh, Jon McDaniel, Sarah MacEwan, Miriam Santo and Andy Simnick, as well as postdoctoral students and visiting professors, namely Ryan Hill, Weiping Gao, Takeshi Mori, and Daisuke Asai, made of the Chilkoti Lab a perfect place to laugh, learn, reflect and work. This environment was most perfectly complemented by the love, trust and support from Vicky Montenegro — who skillfully managed the lab for a number of years. Vicky and her beautiful Colombian family made me feel at home when I first moved to Durham.

Many friends made my time at Duke very gratifying, especially Michael Sinclair, Jesse Winans, Je Hi An, Lucinda Camras, Pablo Gainza, Wendy González, Eduardo Cuervo, Gabriel Bonilla, Luis Tobón, German Forero, Andrea Velásquez, Sandra Ley, Melina Altamirano and Marco Fernández. I am particularly thankful to Pablo for his unconditional support and friendship.

I am also thankful to the Duke Catholic Center and the many great individuals who I had the chance to meet there. They transformed my time at Duke from a mere — yet amazing— professional and scientific experience into a nurturing, well-rounded

experience built around a spiritual growth that has strengthened my faith in God and my love for the Catholic Church and that will certainly have a permanent impact on all aspects of my life.

Finally, I thank Adriana Villa Moreno, who despite the distance offered me her company, friendship and support throughout my PhD.

1. Introduction

“Smart” polymers that are highly sensitive to temperature and other stimuli and that respond with a rapid change in solubility are driving innovations in nanotechnology [1, 2], drug delivery [3], tissue engineering [4, 5], regenerative medicine [6, 7] and biotechnology [8, 9]. Although a number of synthetic polymers exhibit lower critical solution temperature (LCST) phase transition behavior in water [10], this behavior has only been reported in peptide-based polymers with a narrow range of composition, which has long propagated the idea that stringent rules —met by only a few amino acid motifs— govern their behavior. This idea has prevailed because existing peptide polymers that display LCST¹ are largely derived from peptide motifs found in tropoelastin, a Pro- and Gly-rich, intrinsically disordered protein (IDP) that exhibits LCST behavior [11]. In contrast to the abundance of synthetic polymers that exhibit LCST phase behavior, few polymers phase separate from aqueous solution in response to a decrease —as opposed to an increase— in temperature, below a so-called upper critical solution temperature (UCST) [12, 13]. Moreover, the only existing peptide-based polymer that exhibits UCST phase behavior [14] cannot be genetically encoded as it is synthesized from non-proteinogenic amino acids.

The design of proteinogenic peptide polymers that exhibit LCST and UCST phase behavior in aqueous solution is particularly interesting because of two reasons: 1)

¹ Throughout this dissertation, we will use an operational definition of LCST and UCST. Essentially, they are defined by the onset temperature of the phase transition under specific experimental conditions (e.g., polymer concentration, buffer ionic strength, etc.). This clarification is important because the true (unique) values of the LCST and UCST of a polymer can only be determined from a phase diagram.

the tantalizing complexity of the space of proteinogenic sequences, and (2) the ability to access and harness that complexity through genetic engineering methods. Although the proteinogenic sequence space is humungous, the current pictures of the subspaces of amino acid motifs that polymerize into “smart” peptide polymers (Figure 1) suggest the existence of two separate subspaces both of a miniscule size: a small subspace of LCST motifs and an even smaller subspace of UCST motifs.

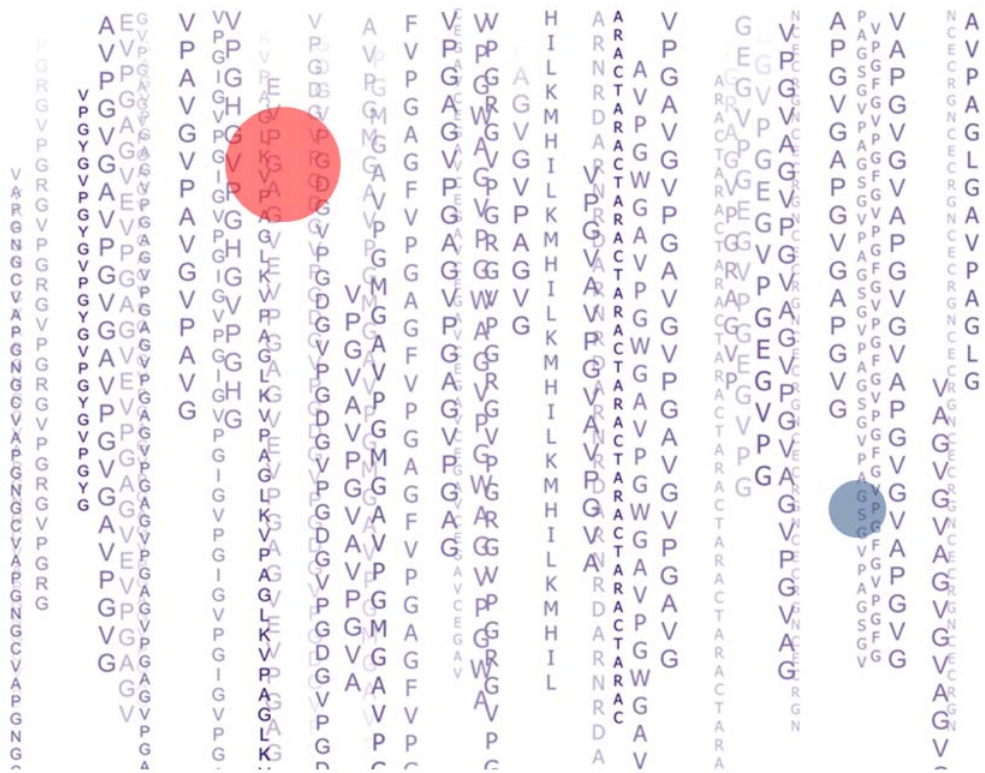


Figure 1. Current pictures of the subspaces of amino acid sequences that polymerize into peptide polymers that display LCST and UCST phase behavior. The amazingly complex space of proteinogenic sequences comprises the amino acid sequences of both folded peptides and proteins as well as the sequences of unstructured peptides and proteins (that is, IDPs). Within this humungous space, we know that there is a small subspace populated by tropoelastin-derived sequences, namely derivatives of the canonical VPGXG motif, which best describes the current picture of the sequence space of LCST peptide motifs —here depicted in red. An even narrower subspace (here depicted in blue) populated by amino acid sequences derived from UCST-exhibiting proteins, namely collagen and resilin, is known to exist at a distance from the LCST subspace.

This dissertation examines the relationship between amino acid sequence space and stimulus responsive phase behavior in proteinogenic peptide polymers, with the underlying hypothesis that the sequence determinants of the intrinsic disorder of Pro and Gly-rich IDP's are overrepresented in the sequence space of amino acid motifs that exhibit "smart" phase behavior upon polymerization. In other words, that the syntax — defined as the arrangement of amino acids into motifs that have a functional behavior— of Pro- and Gly-rich IDP's can shed light into the syntax of LCST and UCST peptide polymers to expand the existing pictures of the subspaces of amino acid motifs that exhibit those behaviors upon polymerization. The rationale for this hypothesis is twofold. First, the few proteins known to display LCST or UCST phase behavior are IDP's, while globular proteins do not exhibit phase behavior unless unfolded [15, 16]. Second, whereas proteins have a high propensity to fold [17], existing protein-based polymers that display stimuli-responsive phase behavior do not readily fold and thus resemble IDP's and synthetic polymers in a good solvent. In addition, the abundance — which has become obvious only recently [18]— and sequence diversity of IDP's suggests the possibility of designing "smart" polymers that exhibit at least two levels of sequence-embedded functionality, namely stimuli-responsiveness and bioactivity.

The significance of this dissertation is most evident in three related areas: first, in the demonstration that the syntax of LCST peptide polymers spans a vast sequence space of amino acid motifs and that this LCST syntax is a potent modulator of the degree of thermal hysteresis in their phase transition behavior. Second, in the development of the first family of genetically encoded peptide polymers with tunable UCST phase

transition behavior, and third, in the development of “smart”, LCST and UCST peptide polymers built from amino acid sequences that exhibit multilevel functionality, namely stimuli-responsiveness and bioactivity. Another area of significance is the exploration of these novel LCST and UCST peptide polymers for the bottom-up synthesis of morphologically-diverse, nano and meso-scale supramolecular structures with tunable stability.

2. Chapter 1: Protein engineering for regenerative medicine and cancer therapy

Overview: Protein engineering enables the design and synthesis of proteins and protein-based materials with advanced properties like self-assembly, self-gelation, stimuli-responsive behavior, super elasticity and a wide spectrum of biological functions. This chapter briefly surveys the use of protein engineering for the design of proteins and protein-based materials for applications in regenerative medicine and cancer therapy. Addressing these two seemingly divergent problems, which may indeed use analogous strategies to either stimulate or inhibit the proliferation of a given cell population, has motivated much of the basic and applied research in this doctoral dissertation. In particular, the desire to address these problems motivated the creation of novel phase transition peptide polymers with intrinsic biological function and new self-assembling behaviors. Protein engineering offers an ample set of tools to capitalize on these and other opportunities in medicine, and the novel LCST and UCST peptide polymers reported in Chapters 3-5 are an important addition to the protein engineering toolbox for the design of highly advanced protein-based polymers.

1. Introduction

Proteins are the molecular machines that propel almost every reaction occurring in the cell, they are the scaffolding material of choice for our tissues, and they ferry messages across distant cells. The extracellular matrix (ECM), for instance, secreted by cells as they construct their perfect niche, acts as a highly advanced proteinaceous scaffold that provides structural support and the on demand release of a large variety of

bioactive proteins and protein fragments that help maintain tissue homeostasis [19]. The ECM hence plays a crucial role in the process of tissue remodeling after injury [20] and in tumor growth and cancer metastasis [19].

Because of this, the engineering of proteins and protein-based materials that mimic the properties of native proteins (that is, antibodies, transcription factors, binding domains, etc.) and natural materials (that is, muscle, tendons, bone, and so forth) or that exhibit behaviors absent in nature, is an exciting field that is fueling innovations in biotechnology, nanotechnology and medicine [1, 6, 21-23].

This short chapter examines the capabilities of protein engineering for the design and development of novel solutions in medicine, and particularly, in the fields of regenerative medicine and cancer therapy.

II. Engineering of proteins and protein-based materials

The rapid progress of the field of protein engineering stems from advances in recombinant technology (that is, restriction enzymes, cell strains, expression systems), DNA sequencing (that is, data from genome-wide sequencing projects, and improved technologies for rapid sequencing), and the development of high-throughput strategies to rapidly evolve proteins with a desired function [24]. These technologies altogether simplify the process of engineering a protein and of constructing materials and materials systems composed of proteins.

2.1. Bioinspired design

Nature provides a plethora of protein-based systems from which to draw inspiration. These systems provide simple recipes to solve myriad engineering

challenges: the design of super tough and elastic materials (as in a spider web) or highly resilient materials (as in resilin and elastin) [25, 26], the programmed assembly of stable nanostructures [27, 28], the controlled release of drugs (as in the ECM) [20], the synthesis of wet adhesives [29], the control of mineral crystallization (as in bone) [28], and the design of general scaffolds to bind protein targets (as in antibodies, fibronectin domains) [30] and to bind and remodel DNA (as in zinc fingers) [31] among many other examples.

2.2. Protein engineering tools

At the heart of protein engineering—defined as the synthesis and manipulation of proteins by the methods of contemporary molecular biology—lies the ability to design and genetically engineer a protein of interest with any desired composition. Although it is possible to construct simple peptide fragments and protein-based materials using chemical synthesis, such as a protein-based polymer with a simple arrangement of monomer units [32] and synthetic polymers that incorporate short bioactive peptides [33], the superior and unique control of composition offered by genetic engineering methods almost invariably favors the genetically encoded synthesis of these materials. Although chemical ligation methods are now available to stitch together proteins from shorter, chemically synthesized peptide fragments, the synthesis of large proteins with complex primary structures, for instance, is still only feasible using genetic engineering tools.

Protein engineering thus involves two major components: i) protein design, and ii) protein synthesis. The latter is a rather standard skill that involves the synthesis of a gene that encodes for the protein of interest, the introduction of the gene into a host (e.g.,

bacteria, fungi, plants or mammalian cells) selected for high yield production, and the purification of the protein from the host cell. The final product, provided that the gene has the correct sequence and that post-translational modifications —if any— can be introduced by the host, is a perfect implementation of the design because the proof-reading machinery of the cell that operates at the replication, transcription and translation levels ensures error-free transmission of the design from the gene level to the protein level. The first component —protein design—, in contrast, may be as simple as copying the protein sequence of a protein of interest or minor modifications therefrom (as in the case of biopolymers inspired in elastin, silk, collagen and others) [34], may involve advanced computational tools to predict protein secondary and tertiary structure (as in the design of protein/protein and protein/drug interactions) [35], or may require highly advanced technical skills to evolve the peptide or protein sequence through extensive rounds of experimental selection (as in the directed evolution of antibodies) [24].

III. Regenerative medicine and cancer therapy

The focus on these two seemingly different problems may be explained by looking at their molecular basis. Essentially, they either involve promoting or inhibiting cell proliferation. On one hand, regenerative medicine aims at controlling tissue regeneration by promoting the proliferation and migration of stem or progenitor cells into the injury site [36, 37]. Cancer therapy, on the other hand, revolves around the uncontrolled proliferation and migration (that lead to tumor metastasis) of cancer cells and aims at exterminating cells with such aberrant proliferation phenotypes [38].

Another link between cancer and tissue regeneration relates to angiogenesis, the process by which new blood vessels are formed from existing vessels [39]. Tissue development and tumor growth are highly dependent on vascularization —blood vessel formation [39]. Without proper blood supply, neither the implanted tissue (cells) nor the tumor would survive. Hence, whereas one may want to prevent angiogenesis for cancer therapy [40], tissue engineers need to devise strategies to promote the rapid vascularization of newly implanted tissues or the generation of a pro-angiogenic environment in regenerating tissues [37].

The role of the immune system and inflammation in particular, in tissue regeneration and cancer is yet another example of the fine line between these opposing processes. The modulation of the immune system is a promising route to suppress the pro-tumorigenic environment created by some immune cells that often infiltrate solid tumors [41], and to drive tissue regeneration as inflammation influences the recruitment of key multipotent cells, as was recently demonstrated for engineered vascular grafts [42].

IV. Proteins and protein-based materials for regenerative medicine and cancer therapy

In order to promote or inhibit cell proliferation, cell migration (e.g., chemotaxis), angiogenesis, apoptosis and any other cellular process involved in tissue growth (whether pathological, as in cancer, or physiological as in tissue regeneration), protein engineering brings to our disposal the entire spectrum of native proteins and protein fragments that cells regularly use for the control of those processes. In theory, one

should be able to produce any protein of interest with any given bioactivity, that is, any biodrug. In practice, however, recombinant synthesis is at times limited by folding issues, poor protein solubility and the inability to reproduce certain posttranslational modifications. Beyond the bountiful and powerful arsenal of proteins provided by nature, the ability to engineer antibodies, antibody fragments and peptides [24, 30] that bind to a target of interest has the capability to provide new biomolecules capable of influencing a plethora of biological processes, as their binding may inactivate a key ligand or receptor [43], act as an agonist [44] or mediate the activation of the immune system [45]. Moreover, recent work has demonstrated the possibility to evolve proteins that bind to specific DNA sequences, like zinc fingers and transcription activator-like (TAL) proteins, that are then fused to DNA processing enzymes in order to design novel biologic drugs capable of fixing genomic defects associated with cancer and tissue degeneration (e.g., muscular dystrophy) [46]. This genome editing strategy, however, is not as appealing for cancer therapy because of the many defects associated with most cancers.

Besides the possibility to design and synthesize biodrugs, protein engineering offers the possibility to construct materials at the nano-, micro- and macro-scales, here referred as protein-based materials. These complex structures result from the interactions between proteins over different length scales: protein receptors, for instance, are often multimeric structures that result from the precise self-assembly of multiple proteins, thus forming highly monodisperse nanostructures [28, 47]. The assembly of collagen and elastin fibers exemplifies designs at both the micro and macro-scale as the

individual microfibers associate to form long fibers that provide support to blood vessels, muscles, skin and other tissues [48]. The ability to exploit well-known protein-protein interactions and assembly behaviors — as in the case of the soluble to insoluble transition displayed by tropoelastin [49]— enables the design of complex assemblies that may be useful for the storage, presentation and release of any biodrug of interest in the contexts of regenerative medicine and cancer therapy.

The self-assembly behavior of some proteins may be used to drive the multivalent presentation of any protein and peptide of interest, which may increase their apparent affinity [50] and prevent rapid clearance from the tissue (due to limited diffusion) to increase the bioavailability of the drug [50, 51]. Koria et al. (2010) recently showed the wound healing benefits of designing a keratinocyte growth factor that forms microparticles on injection into the wound [6]. A similar strategy may be used for other growth factors, antibody fragments and any protein that can affect tissue regeneration or tumor growth. The effect of assembly on drug distribution and pharmacodynamics is particularly evident in the case of small molecules that are lead chemotherapeutic agents in cancer therapy. For instance, transforming a small molecule into a MDa-sized complex may prevent accumulation of the drug in healthy tissues and promotes its accumulation in the extravascular space of the tumor — due to the leaky vasculature in the tumor [39]. Mackay et al (2009) recently demonstrated the potential of protein engineering to develop self-assembled peptide polymer nanoparticles for the delivery of chemotherapeutics [1].

Protein engineering offers a variety of routes for the design of injectable materials since unlike synthetic polymers, protein-based materials may undergo changes in their secondary structure (like a change from a disordered structure to a β -sheet or an α -helix) upon exposure to physiological conditions—they respond to stimuli like buffer ionic strength and temperature—that lead to a rapid change in the viscoelasticity of the materials. Examples of these include the temperature-driven coacervation of elastin-like polypeptides (ELPs) and the salt-driven assembly of β -sheet fibrils [52]. Cells may be considered as very complex biological devices that can produce and release biodrugs with potent effects on tissue regeneration or tumor growth. As such, there is a need to devise ways of delivering cells and ensuring that they are retained at the site of interest. The injection of stem cells into ischemic heart tissue underscores the challenges of successful cell grafting, as only an extremely small percentage of the cells remain at the injection site, which may explain the little (if any) therapeutic value of injecting progenitor cells into the heart [53]. Injectable, self-gelling materials offer an interesting strategy to solve this engraftment issue as they gel *in vivo* on injection and the increase in viscosity that accompanies gelation immobilizes cells at the injection site [52, 54]. *In-situ* gelation is also useful in the context of cancer therapy for the local delivery of bioactive proteins [55] or other therapeutic molecules (e.g., radionuclides) within a tumor that may be toxic to healthy tissues [56, 57].

V. Perspective

The ability to copy, produce, design and reinvent nature's biomolecular workhorse —proteins— provides a unique opportunity for the *de novo* and bioinspired

design of biologic drugs and protein-based materials with promising applications in the regeneration of tissues and the treatment of cancer. This chapter simply scratched the surface of these applications and reviewed the underlying unique features of both proteins and existing protein engineering tools that converge to drive innovation in biomedicine, and that help set the stage to understand the biomedical utility of the novel LCST and UCST peptide polymers described in chapters 3-5.

3. Chapter 2: Overview of LCST and UCST polymers

This chapter provides an overview of existing LCST and UCST polymers with a major focus on peptide-based polymers. Synthetic polymers that display LCST have been extensively reviewed and the reader is referred to reviews by de las Heras et al. [58] and Gil and Hudson [12] for further reading, although a great body of the work revolves around poly(N-isopropylacrylamide) and polymers of other N-alkylacrylamide monomers [12, 58, 59]. Interestingly, synthetic polymers that display tunable UCST phase behavior under physiologically relevant conditions or even useful temperature regimes are rather uncommon in the literature, but recent evidence suggests that a variety of hydrogen bonding polymers chemically-synthesized without ionic initiators –to avoid traces of charged groups– may exhibit useful UCST phase behavior in water and electrolyte solutions that is effectively tunable [60, 61]. The reader is referred to an excellent review by Seuring and Agarwal on the few existing examples of synthetic UCST polymers [13].

The engineering of peptide-based polymers that exhibit LCST and UCST phase behavior has been largely inspired by the study of elastomeric proteins that exhibit these behaviors in aqueous solutions. Elastomeric proteins constitute a group of advanced materials evolved in multiple organisms to provide tissues with mechanisms for the development of elasticity –highly efficient energy storage– that are particularly relevant in the contexts of locomotion, tissue recoil (as in blood vessels subjected to high blood pressures) and energy transduction. Examples of these proteins include elastin, resilin, abductin, lamprin, and flageliform silk among others, which span several phyla

[62]. As in the case of synthetic polymers, the LCST phase behavior of these proteins, particularly elastin, has been the subject of extensive studies, but little work has been conducted on UCST-exhibiting proteins, and only as recently as 2011, Elvin and collaborators characterized the UCST phase behavior of resilin.

Noteworthy, although gelatin —a byproduct of collagen hydrolysis—is long-known to exhibit gelation on cooling, its UCST phase behavior has remained poorly uncharacterized, and despite more recent work that examines this behavior in detail [63], the inability to effectively synthesize collagen in *E. coli* makes it a less than ideal system to study.

3.1. *Elastin-inspired, LCST peptide polymers*

Elastin is a Pro- and Gly-rich, intrinsically disordered protein that abounds in the extracellular matrix of the lungs, arteries, skin, vocal folds, bladder, and elastic cartilage of mammals and most vertebrates [64]. The remarkable properties of elastin have stimulated a growing number of studies centered on the design of elastin-inspired biopolymers. In addition to recapitulating the properties and behaviors of elastin, these biopolymers have been engineered to exploit the architecture and structure-function of elastin in innovative systems that extend well beyond the scenarios and roles wherein the native protein exerts its biological function [65].

This section reviews the seminal work on the design and synthesis of elastin biopolymers that exhibit LCST phase behavior. Elastin-based polymers have found several applications in medicine and biotechnology and the reader is referred to excellent reviews by Simnick *et al.* 2007 and Chow *et al.* 2008 and MacEwan *et al.* 2010

[65-67]. This chapter focuses on the mainstream of elastin-inspired peptide polymers from two different sources, namely recombinant tropoelastin and tropoelastin fragments, and elastin-like polypeptides (ELPs) and other genetically engineered polypeptides containing sequences derived from or inspired by elastin.

3.1.1. The structure and function of elastin: elasticity and tissue remodeling.

Elastin displays a variety of inspiring properties for biomaterials design: it is a durable biomaterial, with half-lives on the order of the lifetime of the parent organisms [62, 68], it displays remarkable resilience, extensibility and self-assembling capacity, and it coacervates in aqueous solution upon a variety of stimuli. Furthermore, elastin plays a physiological role in regulating the remodeling of the extracellular matrix and in controlling inflammation [69, 70].

Studies of the structure-function and the sequence-structure relationship in elastin have been instrumental to the development of elastin biopolymers. Examples of these are the early work by Partridge on the crosslinking of elastin [71], by Bressan and Ronchetti on the coacervation and assembly of tropoelastin [72], by Sandberg on the repetitive sequence of elastin [73], and the work by Urry [74, 75] and later by Tamburro [76] on the biophysical understanding of elastin elasticity. Although this chapter does not intend to provide a detailed picture of the molecular mechanism of elastin elasticity, which is still a subject of discussion [77], this section introduces basic concepts of the molecular architecture and overall properties of elastin that are important for understanding the design and compositions of existing elastin-inspired LCST

biopolymers and that serve as a reference to assess the performance of the polymers developed in chapters 3-5.

Building on the work by Sandberg that resulted in the realization of the repetitive nature of elastin (studies conducted on elastin from swine) and the identification of repeats of the tetrapeptide VPGG, the pentapeptide VPGVG, and the hexapeptide VAPGVG in tryptic digests of porcine elastin, Urry pioneered the development of biopolymers inspired in elastin by chemically synthesizing polypeptides of the pentapeptide VPGVG [78]. It was not until 1987, when Indik *et al.* published the sequence of the human tropoelastin gene, that the order of the peptides previously sequenced by Sandberg for porcine elastin was elucidated [79]. Since then, the primary sequences of elastins from multiple mammalian species, chicken elastin and more recently from frog and zebrafish have been published [80]. The availability of these sequences has revealed several important design parameters that have also been useful for the development of elastin biopolymers, which include: i) elastin sequences present a domain structure with alternating hydrophobic and hydrophilic domains, wherein the hydrophilic domains contain lysine residues for crosslinking and the hydrophobic domains are rich in Proline and Glycine residues; ii) this modular structure allows for the production of various tropoelastin isoforms via alternative splicing of tropoelastin mRNA; and iii) the sequence of the hydrophobic domains is not highly conserved among species. Some of these features are illustrated in Figure 2. Noteworthy, Chapter 3 exploits this evolutionary diversity across elastin proteins from different species, as well

as the even richer composition of a broader set of elastomeric proteins, to gain new insights into the syntax of LCST peptide polymers.

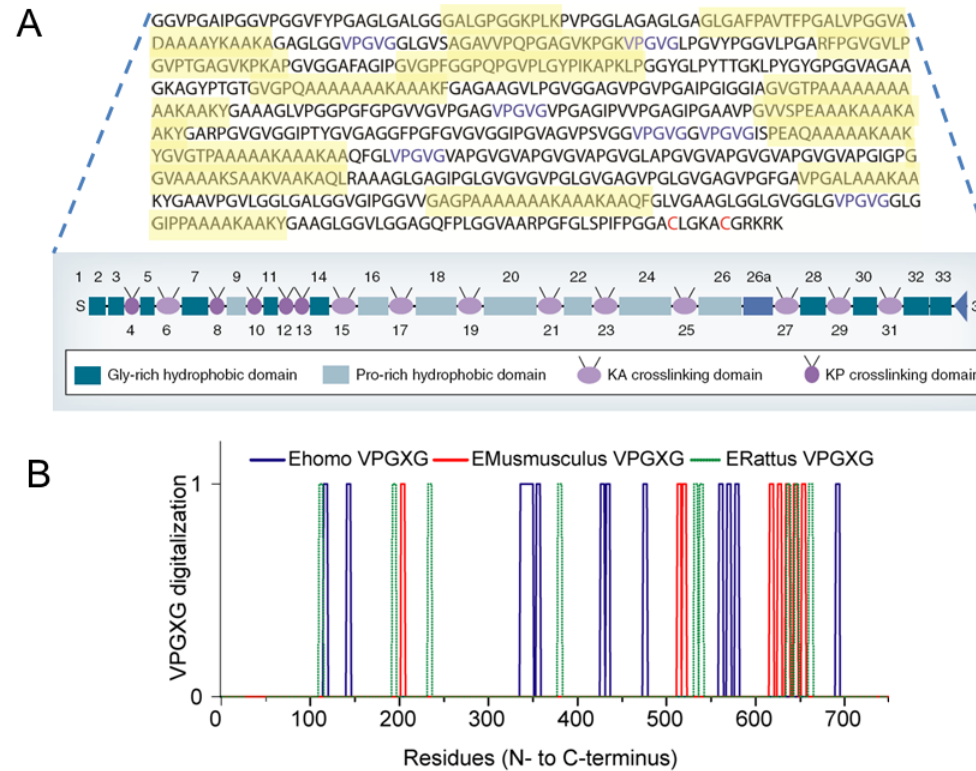


Figure 2. The sequence and domain structure of tropoelastin. A) The sequence of human tropoelastin presents Lys-containing domains (yellow) for crosslinking and is typically divided into Gly- and Pro-rich hydrophobic domains, Lys-Ala (KA) and Lys-Pro (KP) crosslinking domains. Domains (exons) 34-35 are absent in the human gene. KP domains are unlikely to participate in tetrafunctional crosslinks due to their low helical content [64, 81, 82]. Diagram of the domain structure of tropoelastin is from [81]. B) Map of the distribution of the pentapeptide motif VPGXG, descriptive of the elastin biopolymers known as elastin-like polypeptides, along the sequence of various elastin proteins. This motif was generalized by substituting a guest residue X (any amino acid; except Proline) for the fourth residue (valine) in the pentapeptide VPGVG identified in the seminal work by Urry.

The assembly of elastic fibers: tropoelastin coacervation and crosslinking

Tropoelastin, the 60-72 KDa monomer of elastin, displays the unique property of coacervation, an ability to undergo a reversible inverse phase separation from aqueous solution. The process is said to be inverse since, contrary to most proteins, elastin increases its order as it aggregates in response to changes in the aqueous environment, namely the concentration of ionic species, GAG content, pH and temperature [83]. The mechanism behind the coacervation of elastin has puzzled researchers since the initial studies on the coacervation of α -elastin and tropoelastin monomers [75]. This peculiar behavior was suspected to be related to the mechanism of elastin elasticity. The observation that tropoelastin coacervation occurs at an inverse phase transition temperature (or LCST, lower critical solution temperature, in the polymer literature) typically below 37 °C suggested a potential role for tropoelastin coacervation on the self-assembly of elastin during elastic fiber formation [75, 83]. Urry and collaborators demonstrated the self-assembly of α -elastin and tropoelastin into 5 nm wide fibers upon sustained incubation above the LCST, which further supported this idea. Noteworthy, the strong concentration dependence of this coacervation process was largely ignored until more recent studies by Weiss' group —using recombinant tropoelastin— that reported phase transition in a physiologically relevant temperature window only for concentrations of ~10 mg/ml [83]. Despite speculations about whether those concentrations were relevant in the extracellular matrix, or whether high local concentrations in the vicinity of the elastic fiber could drive coacervation, more recent studies have shown the role of glycosaminoglycans in the extracellular matrix,

particularly heparin, in decreasing the critical concentration required for coacervation to concentrations as low as 0.12 mg/ml [84, 85]. These interactions were demonstrated to be mediated by lysine residues, which once blocked with sulfosuccinimidyl acetate, completely inhibited the effect of heparin on the LCST. Hence, tropoelastin environmental sensitivity is exploited to control tropoelastin deposition onto growing elastic fibers in the extracellular matrix. These mechanisms may constitute interesting strategies that can be exploited in the design of injectable gelling materials.

Elastin comprises about 90% of the elastic fiber. Located at its core, crosslinked elastin interacts with surrounding ~10 nm microfibrils of fibrillin-1, and other microfibril associated proteins such as fibulin 5, MAGP-1, and MAGP-2. Interestingly, the arrangement and architecture of elastic fibers varies among tissues. In the artery, for instance, the elastic fibers form concentric rings of elastic lamella, whereas in the ligaments and tendons, the fibers are aligned in parallel to the organization of the tissue and have a larger diameter than in the aorta [86, 87]. Fibrillin-1 microfibrils are essential for proper assembly of tropoelastin into the core of elastic fibers by serving as a template for its deposition and crosslinking. Fibulin-5 was recently proposed to aid in this process by limiting the coalescence of tropoelastin aggregates and by activating LOXL-1, a member of the family of lysyl oxidases responsible for tropoelastin crosslinking [88, 89]. Mecham's group, in a very elegant study, provided a clear picture of the overall dynamics of tropoelastin transport into the extracellular compartment by fusing a timer tag (derived from DsRed) to tropoelastin [90]. This study demonstrated the accumulation of micron-sized tropoelastin globules tethered to the cell surface which

further aggregated before being released and assembled in a microfibril directed manner. In addition, the elastin binding protein (EBP), a 67 KDa spliced variant of β -galactosidase, is known to play an important role in this process [91, 92]. There is evidence that the EBP serves as a chaperon to prevent tropoelastin coacervation intracellularly and assists in its transport to the cell surface; it also prevents its degradation and readily releases tropoelastin in the presence of competing galactosugars (e.g., chondroitin sulfate) [93].

The durability of elastin is not a property of its amino acid composition per se, but rather the highly crosslinked network of elastin chains forming the core of elastic fibers is a key determinant of elastin's superb resistance to degradation. Indeed, uncrosslinked soluble tropoelastin can be readily degraded by a variety of proteases including: cathepsin G, MMP-2, MMP-3, MMP-9, MMP-12, pepsin, neutrophil elastase among others [64, 69, 94]. As a corollary, the design of durable elastin mimics requires recapitulating these highly chemically crosslinked networks. Noteworthy, novel designs of elastin-based polymers that challenge this paradigm and that rely on physical crosslinking are emerging [95]. Tropoelastin crosslinking is mediated by copper-dependent lysyl oxidase (LOX) [96] and other LOX-like proteins (1-4) [97], which catalyze the oxidation of lysine into α -amino adipic- δ -semialdehyde allysine capable of undergoing spontaneous condensation thus forming zero-length crosslinks, namely tetrafunctional desmosines and isodesmosines (for structures of these and other relevant crosslinks see [64, 86]) formed by one lysine and three allysine residues from two different tropoelastin molecules. The prototypical LOX and LOX-like proteins play

different roles in elastin biosynthesis, as demonstrated in studies of LOXL1-null mice, where LOXL1 was shown to be required for elastin deposition during tissue remodeling in adult animals by localizing to the sites of elastogenesis in a fibulin-5 dependent manner. LOX, in contrast, is widely distributed in the tissues and has been suggested to also crosslink fibrillar collagen [97]. Remarkably, it has been shown that tropoelastin coacervation allows for the alignment of specific crosslinking domains to promote the condensation reaction between preferred, juxtaposed lysine residues (even within a domain containing multiple lysine residues) [98], and extensive intermolecular crosslinking has been observed in domains 19-25 of human tropoelastin [99].

Mechanical properties of native elastin

Elastin, considered one of the most elastic biomaterials known, can deform at least 60% in the elastic region (i.e., linear portion of the stress-strain curve), has a Young's modulus (for bovine *ligamentum nuchae*) in the range of 600 KPa to 1200 KPa, tensile strength of ~2000 KPa, and resilience (i.e., spring efficiency) of ~90% [62, 100, 101]. A recent study by Tamburro's group using nanoindentation on crosslinked bovine elastin also suggested a Young's modulus of 1100 KPa [102]. Although these preparations may have small amounts of collagen, the purification usually involves autoclaving or heating at 100 °C for 1 hour to denature remnant collagen [100, 101]. The extensibility (i.e., strain at failure) of elastin has been reported to range between 100 and 200 % [101], and Gosline reported an extensibility of 150 % in his most recent review on elastic proteins [62]. There is agreement that the mechanism of elastin elasticity has an entropic origin, but the debate is ongoing as to what the main source is for such changes

in entropy. Briefly, Urry's model, the librational entropy mechanism, considers highly dynamic non-random conformations in elastin (so called β -spirals comprised of a series of β -turns) which experience high frequency oscillations (high entropy) that become damped upon stretching [103]. Flory's model uses classic theory of rubber elasticity and thus considers a high entropy random coil conformation which upon deformation becomes restricted to a non-random distribution of end-to-end chain lengths [104]. Finally, the exposure of hydrophobic residues during stretching of elastin with the concomitant formation of clathrates of water around these moieties may represent another unfavorable decrease in entropy that drives elastic recoil [77, 105]. Whereas classic rubber theory fails to explain the effects of protein hydration on the elasticity of elastin, Urry's model failed to gain widespread acceptance due to insufficient evidence for the idealized β -spiral structure (even for the simpler VPGXG model peptides), and the minor role it assigned to hydrophobic hydration [103], especially in the light of more recent molecular dynamics studies that demonstrate the instability of the β -spiral [105]. These simulations as well as experimental studies [77] now point to the role of hydrophobic hydration of highly dynamic/disordered polypeptides as a main source of entropy. Noteworthy, as Urry himself acknowledges, the librational model does not necessarily require a β -spiral per se [106], so this mechanism may still help explain some features of the overall elasticity displayed by elastin, particularly under large deformations (i.e., non-physiological) [105]. Further studies are needed to nail down the role of specific dynamic structures (i.e., β -turns), if any, and of hydrophobic hydration on the mechanism of elasticity.

The durability of elastin is evidenced in the billions of stretch/relaxation cycles experienced by elastic fibers in the aortic arch without showing fatigue [107]. This durability is characteristic of entropic elastomers, of which elastin is a formidable example, where the force for elastic recoil arises from the decrease in entropy upon stretching (i.e., limited number of configurations), rather than by direct strain of bonds (i.e., internal energy) in the material that would be prone to breakage [103]. Most of the data available on the fatigue lifetime and other mechanical properties of native elastin have been obtained from pig and bovine but not humans, probably due to limited tissue availability, although at least a few studies on the durability of elastin in humans have reached similar conclusions [64, 108]. Elastin functions as a rather efficient energy storage protein, a spring material which Gosline compares with a steel spring, as he estimated an elastic energy storage capacity for elastin of 95 J Kg^{-1} — for comparison, a steel spring and collagen have energies of 115 J Kg^{-1} and 1000 J Kg^{-1} , respectively [62]. This remarkable elasticity, however, is only displayed under the specific conditions found in tissues where elastin is prevalent, namely conditions of full hydration and of low loading frequency (e.g., cardiac frequencies of 1-3 Hz). For instance, efficient elastic behavior only occurs at frequencies below 0.1 Hz if the relative humidity is reduced to 97% [62]. Also, the poor elastic behavior at high frequencies [62] explains the absence of elastin in the flight system of birds, which can have wing-beat frequencies of 40-70 Hz. Interestingly, resilin, an elastomeric protein with remarkable similarity to elastin in its overall mechanical performance, is found in the wing system of insects with relatively

low wing-beat frequencies (≤ 25 Hz); a frequency range in which resilin still displays over 70% resilience [62].

Physiological role of elastin as a regulator of ECM remodeling and inflammation

The realization in the late 1970's that important components of the extracellular matrix, namely collagen and fibronectin, and peptides derived from these proteins had chemotactic activity for monocytes and fibroblasts, promoted studies on the chemotactic activity of elastin-derived peptides. Mecham and collaborators pioneered this work, demonstrating similar bioactivity for elastin-released fragments upon digestion with neutrophil elastase and suggesting a role for elastin in inflammation [109]. His group then demonstrated similar activity in tropoelastin (i.e., its independence from crosslinking) and later, in collaboration with Urry's group, identified the hexapeptide VGVAPG, which repeats six times in exon 24 of human elastin, as the source of elastin chemotactic activity toward fibroblast and monocytes [110]. Mecham and collaborators (1988) then identified the 67 KDa protein that is now referred as the Elastin Binding Protein (EBP, and also known as S-Gal) [92]. Significant progress has been made since then, namely the identification of all three components of the elastin-binding complex [111], the realization that peptides with the general motif GXXPG present type VIII β -turns and sufficient conformational stability to be specifically recognized by this receptor, and the observation that matrix metalloproteinases (MMPs) are regulated in response to these peptides [112]. The degradation of long-lived elastin protein in tissues is now recognized as a potent signal of extracellular matrix damage capable of triggering its remodeling and of controlling inflammation [69, 113]. These elastin peptides

constitute matrikines, that is, potent cytokine-like molecules derived from the degradation of proteins in the extracellular matrix, and the reader is referred to an excellent review by Antonicelli *et al.*, 2005 on their wide range of biological activities and their signaling mechanisms downstream of the EBP [69]. Little progress has been made in the design of elastin biopolymers that exploit the bioactivity of these matrikines, although Antonicelli and collaborators have already identified opportunities in the repair and regeneration of skin [114]. Notably, the work described in Chapter 5 will demonstrate, for the first time, the utility of matrikine motifs for the design of novel “smart” peptide polymers.

3.1.2. Synthesis, production and expression systems of elastin-inspired, LCST peptide polymers.

Design and engineering of elastin biopolymers

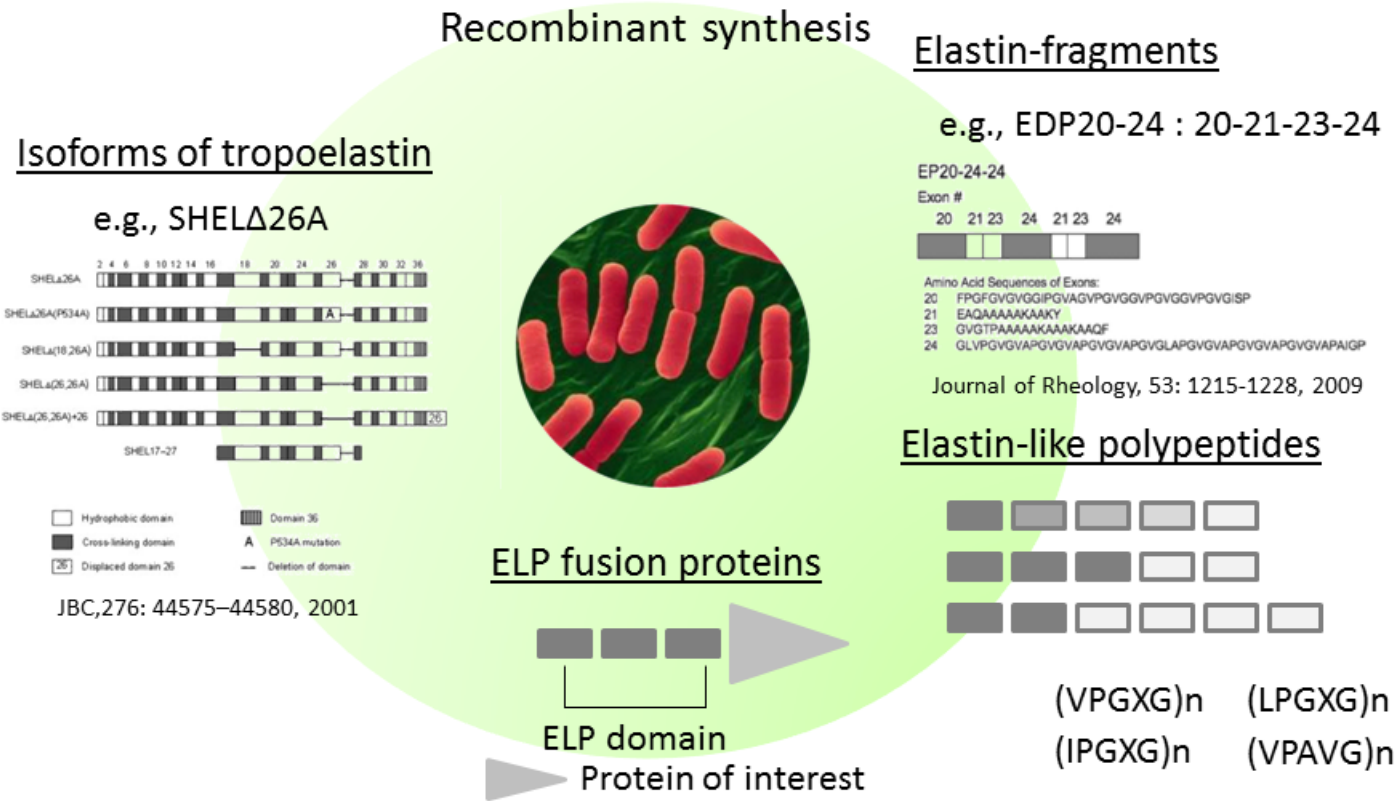


Figure 3. Schematic depicting the design of several elastin biopolymers. The majority of these biopolymers are now produced by recombinant synthesis in *E. coli*. ELP: elastin-like polypeptides. SHEL: synthetic human elastin. EDPs: elastin-derived polypeptides.

First, in order to simplify the reading and comprehension of this and subsequent sections, and to avoid the indistinctive use of the acronym ELPs, here is an explanation of what the main abbreviations mean throughout the text. Elastin biopolymers will not be abbreviated, and will refer to the whole of designs and protein-polymers described in this chapter. ELPs, which stands for Elastin-like Polypeptides, will only make reference to protein-polymers comprised (at least in part) of short amino acid motifs derived or related to elastin (e.g., polymers of VPGXG, VPAVG, and additional functional domains that may be added). Finally, EDPs or Elastin-Derived Polypeptides will refer to those protein-polymers comprised of hydrophobic and crosslinking domains (i.e., entire exons or modifications thereof) found in elastin and combinations thereof. The use of different abbreviations is simply an acknowledgement that despite similarities, different elastin biopolymers may have different biological and biophysical properties, and this chapter and other literature in the field may benefit from such distinction.

The two most obvious elastin biopolymers herein considered are tropoelastin and α -elastin, as they are closely related to the native elastin protein. The latter is a rather uncharacterized soluble byproduct of the acid hydrolysis of crosslinked elastin and thus offers limited design flexibility; yet, it continues to be explored for its simplicity and the possibility to device processing technologies to easily tailor its properties [115]. Tropoelastin, on the other hand, is typically produced as a recombinant protein and there is the possibility to synthesize naturally-occurring and novel isoforms for specific applications, thus enabling the rational design of biomaterials [116]. However, it should be noted that most of the innovation on the uses of recombinant tropoelastin relates to

the processing technologies rather than the protein engineering aspects of the design, since a single isoform that lacks exons 22 and 26A (as they are spliced out in humans) is predominantly used by leaders in the field [85, 117]. A more flexible platform for the design of elastin biopolymers stemmed from the pioneering work by Keeley and collaborators on the design of mini-elastins or elastin-fragments, novel biopolymers comprised of alternating crosslinking and hydrophobic domains (i.e., exons) of human elastin, and repeats of these domains, here referred as EDPs [118]. Keeley's group has studied EDPs composed of different hydrophobic exons, although, as will become evident later, most applications have so far exploited exon 24 as the hydrophobic component. Interestingly, despite the possibility to select any crosslinking domain, EDPs have only been designed incorporating a unique large crosslinking region found in tropoelastin and formed by exons 21 and 23, since the hydrophobic exon 22 does not seem to be expressed in humans [119]. EDPs, however, were introduced rather recently, and the field of elastin biopolymers has been dominated thus far by the applications and designs of ELPs, which constitute without doubt, the most reductionist approach to the design of biomaterials inspired in elastin. ELPs were pioneered by Urry and collaborators as they identified elastin-like properties in protein-polymers exclusively composed of repeats of the pentapeptide VPGVG, which was known to recur in elastin (Figure 2B). His group also demonstrated the possibility to generalize this pentapeptide motif into the pentapeptide VPGXG, as it recurs more frequently in elastin proteins (Figure 2B), and furthermore, the possibility to control their properties by modifying the hydrophobicity of the guest residue X [120]. Urry and others have reported other motifs

that have also been extensively used, namely minor variations of the VPGXG motif such as LPGXG, IPGXG and VPAVG [121, 122]. Tamburro's group has studied somewhat less abundant Gly-rich short amino acid sequences found in elastin and devoid of Pro residues ascribed to a XGGZG motif (where X, Z are L or V), but polypeptides comprised of these motifs do not constitute good mimics of elastin, even though their study has been useful for understanding the mechanisms of elastin elasticity and the potentially clinically relevant amyloidogenesis of elastin [81, 123, 124]. Chilkoti and Meyer pioneered the design of another interesting type of elastin biopolymers, herein referred as ELP fusion proteins. This concept involves the fusion of any type of elastin biopolymer with a protein of interest through protein engineering to impart environmental sensitivity to a variety of interesting proteins with biomedical applications, such as growth factors, soluble receptors, fluorescent tags, and small peptides among others [6, 125]. These and other designs are illustrated in Figure 3.

Synthesis of genes encoding elastin biopolymers

The use of recombinant DNA technology for the production of elastin biopolymers requires the use of specific biomolecular engineering methods for the construction of the genes encoding these protein-polymers. The repetitive nature of elastin biopolymers translates into very repetitive DNA sequences, particularly in the case of ELPs, that tend to be GC-rich and are therefore difficult to handle with conventional cloning methods (e.g., PCR, concatemerization, etc) [126]. However, the preparation of codon optimized recombinant tropoelastin for expression in *E. coli*, first reported by Weiss and collaborators in 1995 [127], or the construction of genes encoding

for EDPs can be achieved using PCR-based strategies and other conventional approaches for DNA cloning. This is due to the lower degree of repetitiveness in both tropoelastin and EDP designs explored thus far, although more complex designs of EDPs or the construction of EDPs with higher molecular weight will certainly benefit from other tools developed for the synthesis of ELPs. Similarly, ELPs of low to moderate molecular weight and relatively high sequence diversity (i.e., different ELP motifs, non-ELP domains, etc.) can be produced by PCR amplification followed by DNA concatemerization, although without any control on the final molecular weight of the construct [128]. Concatemerization of large DNA cassettes (e.g., ~ 150 bp), however, is largely inefficient compared with that of smaller cassettes since large oligomers tend to circularize and thus become unavailable for ligation into the plasmid of interest. The use of chain-terminating capping sequences has been proposed as a strategy to ameliorate this problem [129].

A number of methods have been developed for the synthesis of genes encoding ELPs, and those that allow for the precise control of the molecular weight of the construct are the most widely used, namely recursive directional ligation (RDL) [130]. McDaniel *et al.*, recently reported a new recursive method for the synthesis of elastin biopolymers and any other protein-polymers, that is more versatile and efficient than RDL, which uses a plasmid reconstruction (PRe) strategy to reduce non-specific recombinant products from the self-closing of the vector, as it may occur in RDL, and is thus known as PRe-RDL [126]. Furthermore, unlike RDL, the PRe-RDL cloning steps are performed using an expression vector of the pET series, which facilitates expression of

synthesized constructs and prevents any difficulties in transferring large genes (e.g., > 1.5 Kb) encoding for the elastin biopolymers at the end of the synthesis process. This method also eases the construction of block copolymers and the incorporation of additional N- and C-terminal domains. Hence, PRe-RDL is likely to become the preferred biomolecular engineering tool for the synthesis of genes encoding elastin biopolymers.

We recently developed a new method, Overlap Extension Rolling Circle Amplification (OERCA), for the parallel, “one pot” synthesis of genes that encode repetitive peptide polymers [49]. This work was motivated by our need to develop a high-throughput method to synthesize the genes for the exploration of sequence space for new peptide motifs that display “smart” phase behavior upon polymerization, as will be shown in chapters 3-4. OERCA achieves that in a single PCR-type reaction for the rolling circle amplification (RCA) of a circular DNA template and the simultaneous overlap elongation by polymerase chain reaction (OEPCR), using a high-fidelity thermostable polymerase [49]. We have demonstrated that OERCA is uniquely suited for the rapid synthesis of peptide polymers over a wide and tunable range of molecular weights and that no other existing gene synthesis method for repetitive peptide polymers can achieve such multiplexed gene assembly.

Expression and purification of recombinant elastin biopolymers

One of the critical aspects for the eventual transfer of technologies based on elastin biopolymers into the clinic relates to the ability to produce significant amounts of these materials at low cost. Therefore, it is relevant to review the current expression

systems used for the recombinant production of elastin biopolymers after two decades of innovation since the pioneering work by McPherson et al., 1992 [129], as recombinant methods constitutes the predominant route for their synthesis today. The production of ELPs by recombinant synthesis in *E. coli*, for instance, allows for the production of grams of material from a single fermentation or even from a few shaking cultures in 1 L format, whereas chemical synthesis of these amounts at the laboratory scale would require several months [131]. These recombinant technologies also provide exquisite control on the molecular architecture (i.e., precise control on primary sequence) while ensuring monodispersity. Table 1 summarizes the yields of different biopolymers produced in various expression systems, including *E. coli* (the predominant system as of now), tobacco plants and yeast, with reported yields typically in the range of 100-1000 mg/L of culture. Interestingly, early attempts to produce recombinant elastin biopolymers in multiple hosts, such as *E. coli* [132], fungus [133] and plants [134], suffered from poor expression levels due to non-optimized codon selection of the respective genes. However, in most cases, remarkably high yields have been obtained with very little optimization of the expression conditions, although optimized fermentor systems have been developed as in the case of the gram scale production of tropoelastin (Table 1). Hence, elastin biopolymers can be produced with high yields in a variety of expression systems, with yields typically exceeding those of other repetitive proteins and self-assembling biomaterials (see ref. [135, 136] for comparison). We² have noted

² Throughout this chapter the use of “we” namely refers to the collective constituted by the members of the Chilkoti research group.

that despite the sequence dependence of the overall yields, perhaps partially associated with distinct degradation rates, N-terminal (leader) sequences are capable of boosting the expression levels of these biopolymers. ELPs exclusively composed of the canonical motif VPGVG are produced at very high yields without any leader sequence, but most other VPGXG variants do benefit to a great extent from the incorporation of leader sequences. This effect has also been observed for ELP fusion proteins expressed in plants [137]. Several leader sequences used by different leading groups in the field are also summarized in Table 1, and our research group, in particular, often uses the leader sequence (M)SKGP. Unlike tropoelastin and EDPs (at least those incorporating Y-containing exons like exon 21), ELPs may not contain any aromatic residues with adequate UV absorption properties; therefore, ELP quantification can be greatly simplified by incorporating a C-terminal Y or W residue [138].

In addition to high expression yields, elastin biopolymers are also advantageous as their cost of purification may be significantly lower compared with typical recombinant proteins, and scaling-up their production process is rather straightforward. This is due to simple purification strategies that exploit the environmental sensitivity of elastin biopolymers for their purification, particularly their LCST phase behavior in response to heat and ionic strength. These strategies are known as inverse transition cycling (ITC) [125, 139]. Moreover, the expression of these biopolymers in plants may decrease production costs by 10-50 times compared with bacterial synthesis, and facilitate production of recombinant therapeutic molecules in areas with limited resources and infrastructure, although it remains to be demonstrated whether this

remarkable expression levels can also be achieved in the field [140-142]. Expression of ELP fusion proteins in tobacco seeds seems attractive for long term storage of these proteins without refrigeration [143]. The recent demonstration of ELP overexpression and secretion in fed-batch cultures of *P. pastoris* is also promising for the scale up of these production systems [144]. However, not all elastin biopolymers are purified by ITC, as shown in Table 1, and tropoelastin and EDPs continue to be purified by chromatography-based methods. Although tropoelastin and EDPs may display reversible phase transition behavior and as a result are amenable to purification by ITC, their higher propensity for fibril formation and irreversible aggregation compared with ELPs has favored the use of alternative purification strategies, such as the fusion of EDPs with a glutathione-S-transferase (GST) tag for affinity purification or the purification of tropoelastin by its high solubility in short-chain alcohols followed by high performance liquid chromatography [85, 127]. Nevertheless, most elastin biopolymers, including ELP fusion proteins [125, 142], can be purified by ITC.

The recombinant production of proteins and peptide polymers in bacterial systems often entails the induction of expression using a lactose analogue, IPTG, which removes the repressor from the promoter controlling the expression of the protein, followed by 3-6 hours of expression before cell harvesting and protein purification. However, we and others have observed that IPTG induction does not necessarily result in higher yields of ELPs when compared with non-induced cultures grown overnight and which rely on the leakiness of the T7 promoter used in the pET system [145, 146]. Thus, we typically express these biopolymers in the absence of IPTG, but this may vary

depending on the sequence and molecular weight of the construct, and optimization of these conditions can significantly improve expression yields. However, ELP fusion proteins and relatively low molecular weight ELPs typically benefit from IPTG induction.

The order of the fusion protein may also affect expression and bioactivity, and C-terminal fusion of ELPs has been observed to result in higher purified yields in both *E. coli* and plants. The activity also seems to be better conserved in C-terminal fusion proteins, but the overall effects of the ELP tag may vary depending on the localization of the bioactive site in a given protein [137, 147].

Table 1. Expression systems for the production of several types of elastin-based, LCST peptide polymers. Leader amino acid sequences used for expression or purification purposes are indicated when applicable. Typical yields are reported.

Sequence	Leader sequence	Synthesis and expression	Reported yield	Ref
(1) HHHHHHHHDDDDK (LDGEEIQGHIPREDVYHLYPG((VPGIG) ₂ VPGKG(VPGIG) ₂) ₄)VP) ₃ LKE	MASMTGGQQMG	pET28 expression plasmid (kan). BL21(DE3)pLysS. IPTG induction. 10 L Bioflow 3000 fermentor	(1) 300mg/L	[148]
(2) RKTMG[LD- GEEIQGHIPREDVDYHLYP- G(VPGIG) ₂₅ VP] ₅ -LEKAAKLE			(2) 600mg/L	
ELP[KV6] ₅₆₋₂₂₄ , ELP[QV6] ₁₁₂ , and ELP[KV16] ₅₁₋₂₀₄	SKGPVP	pET25b exp plasmid (Amp). BLR(DE3). 1L cultures with no induction at 37 °C for 24 h	200-400mg/L depending on MW	[138, 149]
ELP[5V3A2G] ₉₀	RFPSIFTAVLFAASSALAAPVNTTTE DETAQIPAEAVIGYSDLEGDFDVAV	Genomic integration of a pPIC9. P. pastoris fed-batch 2.5 L Bioflo	255mg/L	[144]

Sequence	Leader sequence	Synthesis and expression	Reported yield	Ref
	LPFSNSTNNGLLFINTTASIAAKEEG VSLEKREAEA (secretion signal that is cleaved)	3000 fermentors. 48 h induction and pH 6.0 (secreted expression)		
EDPs:				
(1) EDP20-24	Glutathione S-transferase (GST)	pGEX-2T vector (Amp).	(1) 10-15 mg/L	[118, 150]****
(2) EDP20-24-24		BL21(DE3). 1 L cultures with IPTG induction	(2) 20-30 mg/L (3) 50-80 mg/L	
(3) EDP20-24-24-24-24				
ELP[5V2L3G] _n n=20, 40, 90*	<i>GSSHHHHHHSSGLVPRGSHMLEKRE</i> AEAkkGP	pET15b exp. Plasmid (Amp + Tetracycline). BLR(DE3). 0.5 L cultures at 25 °C overnight after IPTG induction	7-10mg/L	[151]
[Y]-[X]-[Y] [Y]={VPAVG[(IPAVG) ₄ (VPAVG)] ₁₆ IP	Not reported	pET24b exp. Plasmid (Kan). BL21-Gold (DE3). Large-scale	(1). 614mg/L (2). 781mg/L	[152]

Sequence	Leader sequence	Synthesis and expression	Reported yield	Ref
AVG} (1). [X] = VPGVG[(VPGVG) ₂ VPGE(VPGVG) ₂] ₃₀ VPGVG (2). [X] = VPGVGVPGVG		fermentation at 37 °C in TB media		
(VPAVG) ₂₂₀ (96KDa)	No leader	pET25 exp. Vector (Amp). BL21 (DE3). Fermentation at 37 °C for 16 h	6mg/L**	[153]
{[(VPGVG) ₂ -(VPGE)- (VPGVG) ₂] ₁₀ - [VGIPG] ₆₀] ₂ V	ESLLP	E. Coli. (presumably a Fermentor system)	520mg/L	[154]
GFP-ELP[V5A2G3] ₉₀	Green fluorescent protein (GFP)	pET25b exp. Vector (Amp). BLR(DE3). 1L shaker flasks at 37 °C for 24 h in optimized media	1620mg/L fusion protein	[145]

Sequence	Leader sequence	Synthesis and expression	Reported yield	Ref
		(no induction)		
SO1- ELP[V5A2G3] ₁₀₀ (Silk-ELP)	LeB4-ER signal peptide*****	pCB301 (Kan). Tobacco		
SO1: recombinant Spindroin 1 (N. clavipes MaSp1) (51.2KDa)	Note: the design includes a C terminal ER retention signal	(Nicotiana tabacum cv. SNN). Growth time was not reported	80mg/Kg	[141]
V _H (TNF)-[SO1-ELP]	LeB4-ER signal peptide*****	pCB301 (Kan). Tobacco	20 mg/Kg fusion protein	
V _H (TNF): variable heavy domain against TNF	Note: the design includes a C terminal ER retention signal	(Nicotiana tabacum cv. SNN). Growth time was not reported		[140]
Tropoelastin (SHELΔ26A) (MW=60KDa)	No leader	pET3d (Amp). E. coli. BioFlo III fermentor.	~1 g/L***	[85, 127]***

* Teeuwen *et al.* 2009 expressed ELP[5V2L3G]₉₀ in the Methionine auxotroph B834(DE3)pLysS and achieved incorporation of the Met analogs azidohomoalanine and homopropargylglycine with similar yields [155].

**We produce polypeptides of VPAVG, which incorporate the short MSKGP leader sequence, with purified yields >100mg/L.

*** Approximate raw expression levels (Weiss, A.S., 2010, personal communication).

**** Yields of purified polypeptides. The yields are expected to be higher under optimized expression conditions (Keeley, F.W., 2010, personal communication).

***** LeB4 signal peptide: MASKPFLSLLSLSLLLFTSTCLA [156].

CNBr: Cyanogen bromide. Kan: Kanamycin. Amp: Ampicillin. SEC: Size exclusion chromatography. TNF: Tumor necrosis factor.

3.1.3. The LCST phase behavior of elastin sequences.

Since the initial reports on the purification of soluble forms of elastin, it was noticed that elastin-derived fragments displayed an interesting ability to reversibly undergo phase separation or reversibly aggregate upon changes in temperature or ionic strength [157]. This ability to sense changes in the environment that trigger pronounced changes in the physical properties of the material constitutes the basis for the “smart” behavior of elastin biopolymers. This process of reversible phase separation is typically studied by quantification of aggregate formation as a function of temperature using spectrophotometric means (i.e., turbidity) or by calorimetric methods to study the thermodynamics of the process. A comprehensive mechanism that explains all aspects of their LCST behavior is still missing, but a large body of literature on biophysical studies of tropoelastin [158, 159], individual tropoelastin exons [82, 160], ELPs [161, 162] and EDPs [163, 164] all point to the importance of an ensemble of disordered or highly dynamic structures that best describe the conformations of these biopolymers below their LCST, namely type II β -turns, sliding β -turns, Polyproline II (PPII), distorted β -sheets or even a complete lack of structure (e.g., random coil) [77], as an important determinant of their self-assembly behavior. The high Pro and Gly content in these polymers contributes to their backbone flexibility and lack of extended conformations (i.e., long-range hydrogen-bonded conformations) that would otherwise promote irreversible aggregation, whereas their hydrophobic nature drives self-association of

apolar residues upon disruption of the clathrate of water surrounding hydrophobic moieties (i.e., waters of hydrophobic hydration that prevent self-aggregation) by stimuli such as increases in temperature and ionic strength. The overall process remains largely entropic, since the decrease in entropy resulting from the self-assembly of elastin biopolymers is smaller than the increase in entropy arising from the disruption of the ordered clathrate of water [103, 163]. Understanding the mechanism responsible for the elasticity and LCST phase behavior of elastin biopolymers may be useful for the engineering of more complex materials systems, such as those that undergo self-assembly in response to stretched-induced pKa shifts [106, 165].

Tropoelastin coacervation and self-assembly

The ability of tropoelastin to undergo reversible phase separation or coacervation in response to a variety of stimuli or to assemble into flexible fibrillar structures [75] is also of interest from a biomaterials and biophysical perspective. As discussed previously, Weiss' group conducted the seminal work describing the dependence of tropoelastin's LCST on protein concentration and buffer ionic strength. Remarkably, the LCST of tropoelastin varied from ~50 °C at 100 mM NaCl to ~30 °C at 200 mM NaCl in a phosphate buffer at 20 mg/ml, which demonstrated the importance of physiological salt concentrations for tropoelastin self-assembly [83]. More recently, they reported the self-assembly of low concentrations of tropoelastin into 200-300 nm spherical particles which coalesce in the presence of negatively charged oligosaccharides that serve as nucleation

sites. This assembly mechanism may help explain the micron-sized particles observed at the cell surface prior to elastin fiber formation [166]. Although pH has relatively small effects on tropoelastin's LCST in the pH range of 5-9 [83], tropoelastin can irreversibly self-assemble under alkaline conditions (pH 10-11), high concentrations (e.g., 100 mg/ml) and prolonged incubation times (~16 h) above its LCST (e.g., 37 °C) into elastic and flexible hydrogels of great mechanical stability with Young's modulus of ~1.7 MPa [117], which further highlights the possibility to harness the self-assembly of tropoelastin for the design of novel biomaterials even in the absence of chemical crosslinking.

The repetitive nature of tropoelastin prompted Tamburro and collaborators to dissect it, exon by exon, in order to study the ability of each domain to drive or promote its overall coacervation behavior. First, they chemically synthesized and studied 24 different exons of human elastin, comprising all of the hydrophobic domains, all KP crosslinking domains and one representative KA crosslinking domain (exon 19) by circular dichroism and nuclear magnetic resonance spectroscopies. Besides exon 19, which secondary structure is predominantly α -helical, all exons were suggested to adopt unordered conformations or PPII at temperatures up to 70 °C in water [82]. Although the LCST of most of these low molecular weight peptides was unlikely to lie in an accessible temperature window, Tamburro and collaborators studied the supramolecular assembly and coacervation properties of the larger exons 18, 20, 24, 26 and 30. Exons 18, 20, 24 and 26 showed coacervation behavior despite their low

molecular weight (< 5 KDa), while exon 30 precipitated irreversibly and formed amyloid-like fibers [81, 160]. Moreover, exons 20, 24 and 26 self-assemble into flexible fibers reminiscent of tropoelastin self-assembled fibers, whereas scrambled analogues of exons 24 and 26 failed to self-assemble [167]. This work demonstrated the intrinsic ability of some tropoelastin exons to drive coacervation and elastin assembly, and evidenced the amyloidogenic propensities of Gly-rich exons.

The processing of tropoelastin at the mRNA level is then critical for the fine tuning of its biophysical properties and self-assembly behavior. For instance, a construct of exons 2-7 displays coacervation at 37 °C, similar to the native protein, whereas a polypeptide including the signal peptide encoded by exon 1 (26 amino acids long) precipitates irreversibly. Therefore, signal peptide removal in the endoplasmic reticulum is important for proper elastin fiber assembly and for minimizing amyloid-like propensities [168].

Self-assembly of Elastin-Derived Polypeptides

Bellingham et al. 2001 were the first to report the design of elastin biopolymers comprised of alternating hydrophobic and hydrophilic domains of human elastin [150]. They characterized the concentration, molecular weight and ionic strength dependence of the coacervation behavior displayed by these polypeptides, which is analogous to that displayed by tropoelastin and ELPs. They compared the post-coacervation assembly of these polypeptides into elastin-like fibers with tropoelastin, and demonstrated

similarities in their self-assembly capacity. In particular, a larger number of alternating hydrophobic and crosslinking domains resulted in fibers more reminiscent of tropoelastin fibers. EDPs such as EP20-24, EP20-24-30 and polypeptides thereof display coacervation behavior in the traditional sense (i.e., they undergo a fully reversible inverse phase transition) despite the high content of VAPGVG motif in their hydrophobic domains or the introduction of Gly-rich exons, which have been reported to precipitate irreversibly after phase separation when expressed as isolated polypeptides [150, 169]. Interestingly, these findings point to the role of crosslinking domains with high helical content, in addition to that of Pro-rich hydrophobic domains, on the reversible phase transition behavior displayed by tropoelastin, perhaps by disrupting potentially extended β -sheet structures that would otherwise arise in these polypeptides and that would eventually result in irreversible aggregation [163]. However, much like tropoelastin, EDPs have a propensity to form extended self-assembled networks and phase separate irreversibly upon sustained heating at moderate temperatures above body temperature (e.g., 40-50 °C) and for relatively short incubation times (5-10 min), thus suggesting a propensity for irreversible assembly [72, 118, 170].

As expected, changes in the hydropathy of the crosslinking domains also affect the LCST of the polypeptides [163]. As previously described for ELPs, more hydrophobic domains also resulted in lower coacervation temperatures, although these

effects were only evident when using the hydrophobicity scale proposed by Urry instead of the widely used Kyte-Doolittle scale. Hence, Urry's hydrophobicity scale does allow for a better prediction of the coacervation temperature of elastin biopolymers. As a case in point, tryptophan is slightly less hydrophobic than Glycine according to Kyte Doolittle's scale, with hydrophobicity indices of -0.9 and -0.8, respectively, whereas tryptophan is considered the most hydrophobic of all amino acids in Urry's scale and the design of ELPs incorporating these residues indeed results in very hydrophobic biopolymers with very low LCST's [171].

The crosslinking domains of EDPs can also be engineered to tailor the phase transition behavior of these polypeptides. Keeley and coworkers demonstrated that by increasing the α -helical content of these domains, the propensity for coacervation is greatly enhanced [164]. These design principles can be incorporated into other elastin biopolymers designs in order to enable polypeptide crosslinking while tailoring their LCST. In fact, the incorporation of K residues in ELPs tends to increase the LCST significantly, which may be undesirable in some applications, and which could be prevented by engineering more elastin-like crosslinking domains into these polypeptides.

The importance of the role of Pro residues in the coacervation and self-assembly properties of EDPs was demonstrated in experiments by Miao *et al.*, 2003 where they showed that P to G mutations in exon 24 in a variety of elastin fragments resulted in the

abrogation of coacervation behavior, formation of fibrillar precipitates and even complete insolubility in water [163]. These findings are equally relevant for other elastin-derived sequences, and bioinformatics studies have shown that Pro and Gly content are key determinants of the self-assembly behavior of repetitive proteins into either elastomeric or amyloid-like fibrils [172].

Temperature alone can also trigger the gelation of EDPs. Keeley's group recently reported that EP20-24 can self-assemble into quite strong gels with storage modulus in the order of 10 KPa at a temperature above the LCST of the polypeptide, in the absence of covalent crosslinking and at relatively low concentrations (~0.34 %W/V) [170].

Inverse phase separation and self-assembly of ELPs

The ability of ELPs to undergo LCST phase behavior or inverse phase separation into a coacervate —ELP-rich— phase has been extensively studied. Urry and collaborators first demonstrated that the phase transition temperature of ELPs could be controlled by the hydrophobicity of the guest residue X in the VPGXG pentapeptide motif, and harnessed this remarkable sensitivity to the hydrophobicity of the guest residue for the construction of a new hydrophobicity scale for amino acids [120]. They have refined this scale in more recent studies and have estimated the Gibbs free energy of hydrophobic association for each amino acid [173]. The utility of this scale relies on its direct relationship with the hydrophobicity of a given amino acid in the context of an elastin biopolymer. However, this scale does not consider two important aspects, the

influence of the polypeptide molecular weight and its concentration, which were clearly elucidated in subsequent studies by Meyer and Chilkoti [171]. These studies suggest that it may be possible to easily predict the length of an ELP so that the LCST lies within a temperature window of interest (e.g., near body temperature), thus simplifying the design and construction of polypeptides useful for biomedical applications. Interestingly, these two interrelated parameters do not influence the LCST equally, and ELPs of different molecular weight coacervate at different temperatures even if the measurement corrects for the overall mass difference [171]. The LCST is a linear function of the logarithm of the biopolymer concentration and a linear function of the reciprocal of the biopolymer molecular weight [83, 163, 171]. Changes in molecular weight have a major effect on the transition temperature for biopolymers of low molecular weight, but have minor effects on polymers with large molecular weights (e.g., >100 KDa). The influence of these two parameters also depends on the motif sequence, with more hydrophilic sequences showing greater dependence and sensibility at high molecular weights and concentrations.

ELP aggregates of both elastic (VPGVG) and non-elastic (VAPGVG) motifs result in fiber formation as shown by electron-microscopy studies [78, 174]. A similar behavior has been shown for the polytetrapeptide VPGG [74]. ELPs based on the VPGXG motif are rather heat stable and display reversible phase transition behavior over a wide range of conditions. The ability of different peptide motifs to form nearly perfect elastomeric

materials —that is, polymers that do not dissipate any stored energy during stretching and thus display nearly zero hysteresis upon relaxation— may be associated with the thermal hysteresis of their phase transition behavior upon heating and cooling. Peptide polymers composed of the VPGXG motif form elastic materials upon crosslinking and in solution display very small if any thermal hysteresis, whereas polymers of VAPGVG do not phase separate reversibly and form inelastic materials upon crosslinking. Interestingly, polymers of VPAVG that display significant thermal hysteresis upon cooling —the solution must be cooled well below the apparent LCST on heating in order to make the polymer soluble [175]— form plastic biopolymers upon crosslinking. In addition, VPAVG has been shown to assemble into well-defined micro- and nanoparticles above the LCST. The size of these particles seems to depend on the polydispersity and molecular weight of the polypeptides, since early reports by Rodriguez-Cabello's group reported ~1-3 μm particles for polydisperse, chemically synthesized $(\text{VPAVG})_n$ [176], whereas recent data using monodisperse recombinantly synthesized polypeptides resulted in particles on the order of ~200 nm [177]. The mechanism behind this assembly behavior and the large hysteresis on cooling compared with other ELP motifs has been suggested to arise from the disruption of the PG dipeptide that is critical for β -turn formation, and from the bulkiness of the Ala side chain, although other β -turn structures involving other residues may still occur [175]. This largely unexplored and poorly understood hysteretic phase behavior will be

discussed in chapter 3 under the light of new findings that help explain its molecular underpinnings and that pave the way for tuning and exploiting these and other forms of hysteretic phase behavior.

The phase transition behavior of ELPs has been shown to remain unaltered when mixing biopolymers of different molecular weights or different amino acid composition [151, 178], but these data are still limited to a few compositions based on the VPGXG motif and further experiments will be needed to extend these results to other elastin biopolymers. Miao *et al.*, 2003 have also shown that EDPs comprised of the same exons but differing in the number of domains coacervate independently. Interestingly, this is in accordance with the effect of displacing hydrophobic domains of tropoelastin on its coacervation propensity, which underscores the role of coacervation in orchestrating the juxtaposition of specific oxidized lysine residues for their condensation by favoring the interaction of specific domains [179].

The LCST behavior of ELPs and other elastin biopolymers constitutes an interesting energy transduction mechanism. Urry has demonstrated that this transition into a “folded” state contracts crosslinked elastin biopolymers, such that tailoring the LCST of the polypeptide can be used to control the development of elastomeric force in the biomaterial [103, 180]. The temperature dependence of this transition does not take away from its biological relevance, despite the tight temperature control exerted in mammals, since there are potent means to tune the LCST by additional environmental

factors or direct chemical modifications, such as the degree of ionization of guest residues (i.e., by pH changes) or more interestingly, through the phosphorylation/dephosphorylation of hydroxyl groups in guest residues along the biopolymer [120].

3.2. Resilin-derived UCST peptide polymers

Resilin is a Pro- and Gly-rich elastomeric protein found in arthropods that, unlike tropoelastin, has a particularly hydrophilic sequence dominated by polar charged and uncharged residues. Despite this dramatic change in composition, resilin and elastin exhibit remarkably similar elastomeric properties [62]. Very recently —indeed concurrently with the experimental work that will be presented in chapter 4—, there has been a growing awareness of the intrinsic UCST phase behavior exhibited by resilin (particularly, exon 1). The seminal work, for instance, was conducted as recently as 2005 when Elvin and collaborators first synthesized this resilin fragment (known as rec1-resilin) recombinantly [25]. Since then, resilin across various insect orders, including jumping and flying insects, have been shown to be intrinsically disordered in solution and to exhibit the unusual propensity to undergo aggregation at low temperatures [181]. Subsequent reports noted that recombinantly synthesized polymers of repetitive sequences found in resilin were also unstructured [182] and that concentrated solutions of these resilin-like polypeptides —obtained by precipitation with 20% ammonium sulfate— had the propensity to phase separate when fortuitously exposed to cold

temperatures during the purification process [135, 183]. This apparent UCST behavior remained uncharacterized for several years until 2011 when Elvin's group succeeded in demonstrating that rec1-resilin displayed UCST phase behavior at very low temperatures and in a pH dependent manner [184]. They also showed that at high temperatures (~65 °C) and independently of the solution pH, rec1-resilin exhibited a form of LCST phase behavior that is largely irreversible. The pH dependence of the UCST behavior was tentatively attributed to concurrent changes in the ability of different prevalent residues to form hydrogen bonds [184].

Resilin-derived peptide polymers are usually composed of consensus repeat sequences found in resilin from various species, such as GGRPSDSYGAPGGGN in rec1-resilin from *Drosophila melanogaster* [25] or AQTSSQYGAP from *Anopheles gambiae* [135], but very little work has been conducted towards understanding the sequence determinants of this behavior [185]. To this date, there are no reports that demonstrate the possibility to tune the UCST phase behavior of these resilin-derived polypeptides, which limits the ability to exploit their intrinsic UCST behavior. For example, recent reports that attempt to use these resilin-like polypeptides as fusion tags for protein purification rely on ammonium sulfate precipitation and not on the intrinsic UCST behavior of the resilin-like tags, as their phase behavior is only appreciable at low temperatures (<4 °C) and at high concentrations [186].

As reported in Table 1 for elastin-based LCST peptide polymers, resilin-derived polymers have also been recombinantly synthesized at high yields in a variety of *E. coli* based expression systems [135, 183, 185]. The production of rec1-resilin has been optimized for expression in a 100 L lactose-induced fermentation system with purified yields of 300 mg per liter of culture. Using a similar expression system, a polymer composed of 16 repeats of the consensus motif AQTPSSQYGAP has been produced with yields of 450 mg/L [135]. Interestingly, polymers of the consensus sequence found in rec1-resilin have been produced only at rather low levels with reported yields of 20 mg/L [183].

4. Chapter 3: Syntax of LCST peptide polymers

Overview: “Smart” polymers that exhibit a physical change in response to stimuli in their aqueous environments are of great utility in biotechnology and medicine. Currently, however, only few peptide polymers are known to display this behavior. Here, we¹ explore a large sequence space to uncover the relationship between the syntax of peptide polymers — defined as the arrangement of amino acids (letters) into repeat units (words) that have a functional behavior of interest — and their lower critical solution temperature (LCST) phase behavior in aqueous solution. We show that the syntax of LCST peptide polymers ranges from polymers composed of repeats of a few amino acids to those with the complex non-repetitive syntax reminiscent of protein domains, and that the concept of syntax can be deployed to tune the thermal hysteresis in LCST phase behavior. By using CD spectroscopy as a function of temperature, we also show that hysteretic LCST phase behavior in peptide polymers results from aggregation-induced folding.

4.1. Syntax of “smart” peptide polymers governs their function

Lower critical solution temperature (LCST) transition polymers, an important class of “smart” polymers that are highly sensitive to temperature and other environmental stimuli and that respond with a sharp change in their solubility, are driving innovations in nanotechnology [1, 2], drug delivery [3], tissue engineering [4, 5],

¹ Throughout this and the remaining chapters, “we” refers to my PhD advisor, Ashutosh Chilkoti, and me.

regenerative medicine [6, 7] and biotechnology [8, 9]. Although a number of synthetic polymers exhibit LCST behavior in water [10], this behavior has only been reported in peptide-based polymers with a narrow range of composition, which has long propagated the dogma that stringent rules — met by only a few amino acid motifs — govern LCST behavior in peptide polymers. This idea has prevailed because existing peptide polymers that display LCST are largely derived from peptide motifs found in tropoelastin, a Pro- and Gly-rich, intrinsically disordered protein (IDP) that exhibits LCST behavior [11].

However, we recently mutated peptide polymers composed of the canonical, LCST-exhibiting Val-Pro-Gly-Xaa-Gly motif — where Xaa is any amino acid except Pro — found in tropoelastin (Figure 11; found in *Supplemental Figures*), by alanine insertion and substitution mutagenesis, and 8 of the 9 variants exhibited LCST behavior [49]. The relative promiscuity of this limited set of elastin-inspired peptide polymers is intriguing as it hints at the existence of a large and diverse set of peptide motifs that should display similar “smart” behavior in a polymeric form. Consistent with this view, resilin, another Pro- and Gly-rich IDP with an overall composition that significantly deviates from tropoelastin, was also recently shown to exhibit a form of LCST phase behavior in water, though only at high temperatures (~65 °C) [184]. Motivated by these observations, herein we examine the relationship between amino acid sequence space and stimulus responsive behavior.

Because tropoelastin and resilin are Pro- and Gly-rich IDPs that exhibit LCST behavior [187, 188], we began our search from this known point in the sequence–function landscape by focusing on IDPs with a high Pro and Gly content (Figure 12A; found in *Supplemental Figures*). We first looked for patterns in the distribution of these residues by mapping the localization and abundance of 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), as well as the distribution of Gly residues that surround these motifs [189]. Although a P-G dipeptide (n=0) is the predominant motif in these proteins, as in the canonical Val-**Pro-Gly**-Xaa-Gly motif that forms the basis of most known LCST peptide polymers [10, 49], we identified an unusually large fraction of recurring repeats of P-X₄-G motifs in resilin (Figure 4A). Compositional profiling of Gly surrounding P-X_n-G motifs [189] further revealed a highly biased localization of Gly residues that differed between IDPs that usually form amorphous aggregates such as elastin, resilin, and gluten and IDPs that tend to form fibrillar structures such as collagen and silk. We found that Gly is non-randomly and consistently (p<0.001, *Supplemental information*) absent one residue N-terminal to Pro (termed the -1 position) in the coacervating group of proteins, whereas it is enriched at that position in the fibril forming group (Figure 4B). This suggests a potential role for Gly at this position in controlling whether Pro- and Gly-rich IDPs go down the coacervation or fibril formation pathway.

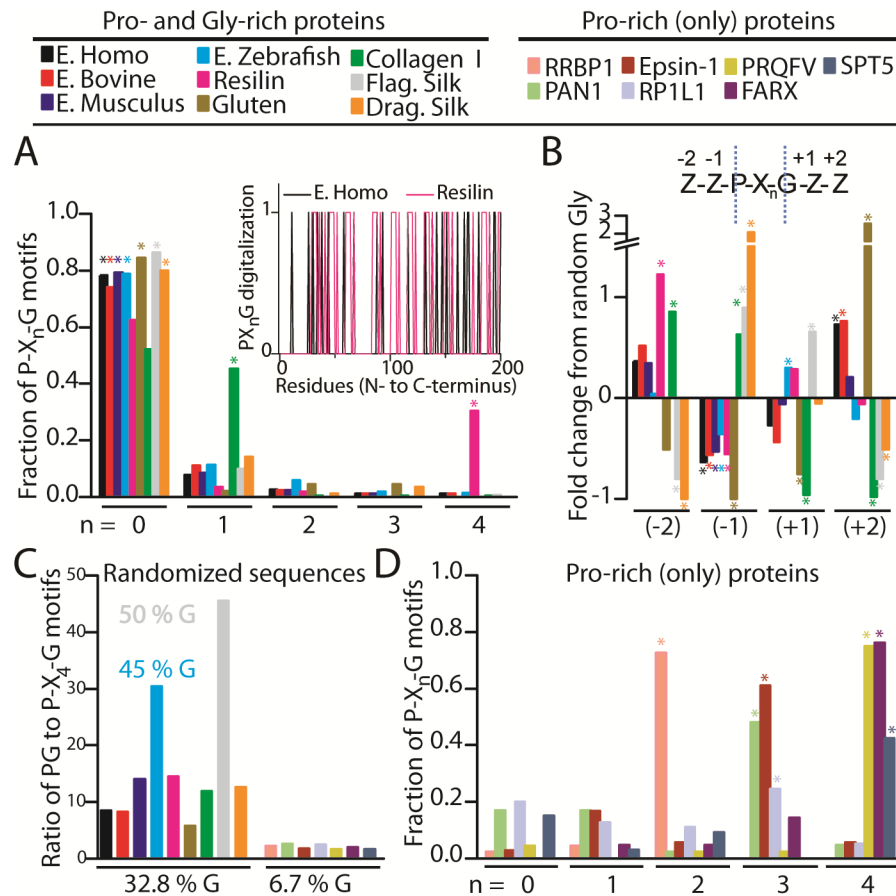


Figure 4. The amino acid sequence of Pro- and Gly-rich and Pro-rich (only) IDPs is characterized by discrete arrangements of Pro and Gly residues. (A) Abundance of individual P-X_n-G motifs among Pro- and Gly-rich IDPs, where X is any amino acid different from Pro and Gly and *n* varies from 0 to 4, relative to the total number of P-X_n-G motifs per protein. Asterisks indicate statistical significance (*p* < 0.001) for over-represented motifs compared with the expected distribution in a set of randomized proteins with identical amino acid content (Figure 13 found in *Supplemental Figures*). The inset shows the characteristic recurrence of P-X_n-G motifs in elastin and resilin, as evidenced through a digitalization of their protein sequences in which residues –other than Gly– participating in a P-X_n-G motif are assigned a value of 1, while all other residues are set to 0. (B) Enrichment or depletion of Gly in residue positions surrounding P-X_n-G motifs among Pro- and Gly-rich IDPs, expressed as a fold change from the random occurrence of Gly based on total Gly content. Asterisks indicate significant (*p* < 0.001) divergence from a random distribution (see *Supplementary discussion*). (C) Ratio of the median fraction of Pro-Gly motifs to P-X₄-G motifs across 1000 randomized versions (see *Supplementary discussion*) of each Pro and Gly-rich (median % G of 32.8) and Pro-rich (only) (median % G of 6.7) IDP. (D) Analysis of P-X_n-G motifs, as in (A), among Pro-rich IDPs that are not rich in Gly. “E. Homo” is elastin from Homo

sapiens. Full protein names and accession numbers are reported in Table 3 (found in *Materials and Methods*).

The overabundance of Pro-Gly dipeptides, however, is statistically expected for any Pro-rich IDP that also has high Gly content, since randomizing the amino acid order in each of these proteins revealed a persistent bias towards the Pro-Gly dipeptide over any other P-X_n-G motif (Figure 4C; and Figure 13A in *Supplemental Figures*). Hence, we next examined IDPs that are rich in Pro but not Gly (Figure 14A; found in *Supplemental Figures*), as they lack such a statistical bias (Figure 4C; and Figure 13B in *Supplemental Figures*). Interestingly, these Pro-rich IDPs show a large fraction of recurring P-X_n-G motifs with characteristic values of n between 2 and 4 (Figure 4D; Figure 14 and Figure 15 in *Supplemental Figures*). In line with this observation, a recently discovered amphibian tropoelastin that has an unusually low Gly content [190] is composed of four types of P-X_n-G motifs (where $n=0-3$) that recur with similar frequency (Figure 16 found in *Supplemental Figures*).

Building upon these insights: (1) the recurrence of P-X_n-G motifs ($n=0-4$) and (2) the absence of Gly at the -1 position in IDPs that show LCST phase behavior, we recombinantly synthesized polymer libraries of 24 unique amino acid motifs that span the entire sequence space of Pro-X_n-Gly motifs (where n ranged from 0 to 4) to experimentally determine whether they exhibit LCST phase behavior [189]. Surprisingly, we found that all these peptide polymers exhibit LCST phase behavior, analogous to that of tropoelastin and elastin-like polypeptides (ELPs) (Figure 5A-D). This finding

significantly expands the vocabulary of LCST peptide polymers and suggests that a variety of amino acid motifs could be incorporated within the P-X_n-G unit, and potentially between them as well, as resilin and tropoelastin have an average distance of ~10 other residues between P-X_n-G units (Figure 12B found in *Supplemental Figures*), yet still exhibit LCST behavior. This notion that sequence diversity can be tolerated in LCST peptide polymers is further supported by our finding that a protein-sized polymer composed of 40 repeats of a Z⁻²-Z⁻¹-P-Z⁰-G-Z⁺¹ hexapeptide motif (Figure 6 and *Supplementary discussion*), where each Zⁱ residue position accommodates a diverse and randomized set of amino acids that span a broad range of hydrophathies (Figure 6B), exhibits a form of LCST phase behavior under physiologically relevant conditions (Figure 6C) wherein the LCST value is controlled by the average hydrophathy of the polymer.

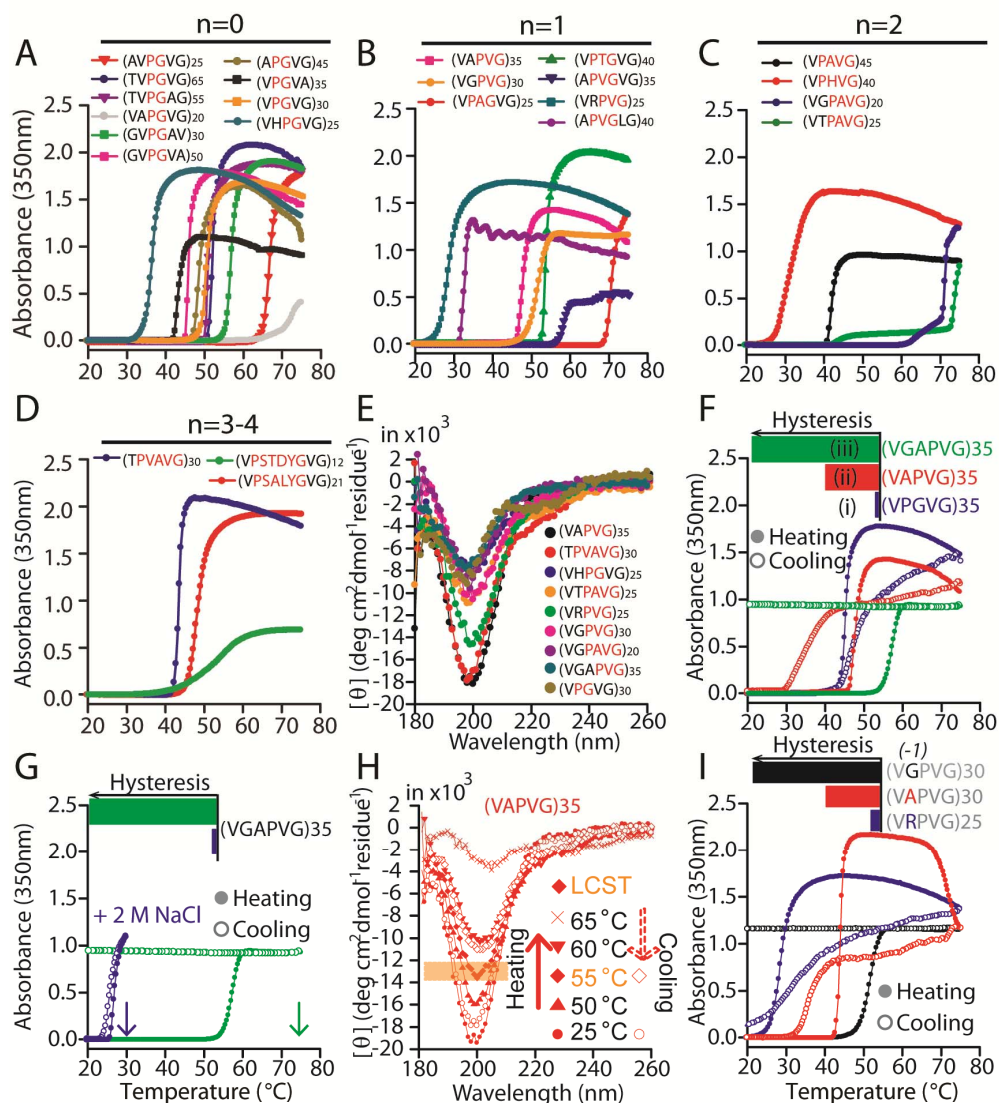


Figure 5. Diverse repeat motifs that contain Pro and Gly lead to intrinsically disordered peptide polymers that display LCST behavior. (A-E) Temperature-dependent turbidimetry revealed that peptide polymers with periodic Pro and Gly residues arranged as P-X_n-G units, where n=0 (A), 1 (B), 2 (C), 3 and 4 (D), and having pentapeptide, hexapeptide and nonapeptide repeat units display thermally triggered LCST phase transition behavior. (E) These peptide polymers lack ordered secondary structures as shown by their CD spectra (at 25 °C) that is characteristic of IDPs. (F) Three representative peptide polymers show the three types of phase transition behavior exhibited by all polymers in A-C, corresponding to the degree of hysteresis in the reversibility of their thermally triggered phase transition: (i) zero, (ii) finite, and (iii) heat-sensitive, infinite hysteresis. (G) Polymers that behave as in (iii) display irreversible aggregation at high temperatures but undergo phase transition with zero hysteresis below a critical threshold

temperature — arrows indicate the maximum temperature in the turbidity versus temperature profile. Figure 7 shows other peptide polymers with this behavior. (H) CD spectra as a function of temperature for polymer in H-(ii). The LCST was determined using temperature-dependent turbidimetry (Figure 17A found in *Supplemental Figures*). (I) Hysteresis in the reversibility of the phase transition behavior of three peptide polymers with a V-IPVG motif and with Gly, Ala or Arg at the -1 position. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μ M, except for VRPVG (+ 1M NaCl), VAPGVG (+ 0.5 M NaCl), TVPGAG (+2 M NaCl), APGVG (+2 M NaCl) and VPSALYGVG (+ 8 M urea). CD studies were conducted in water at a polypeptide concentration of 5 μ M.

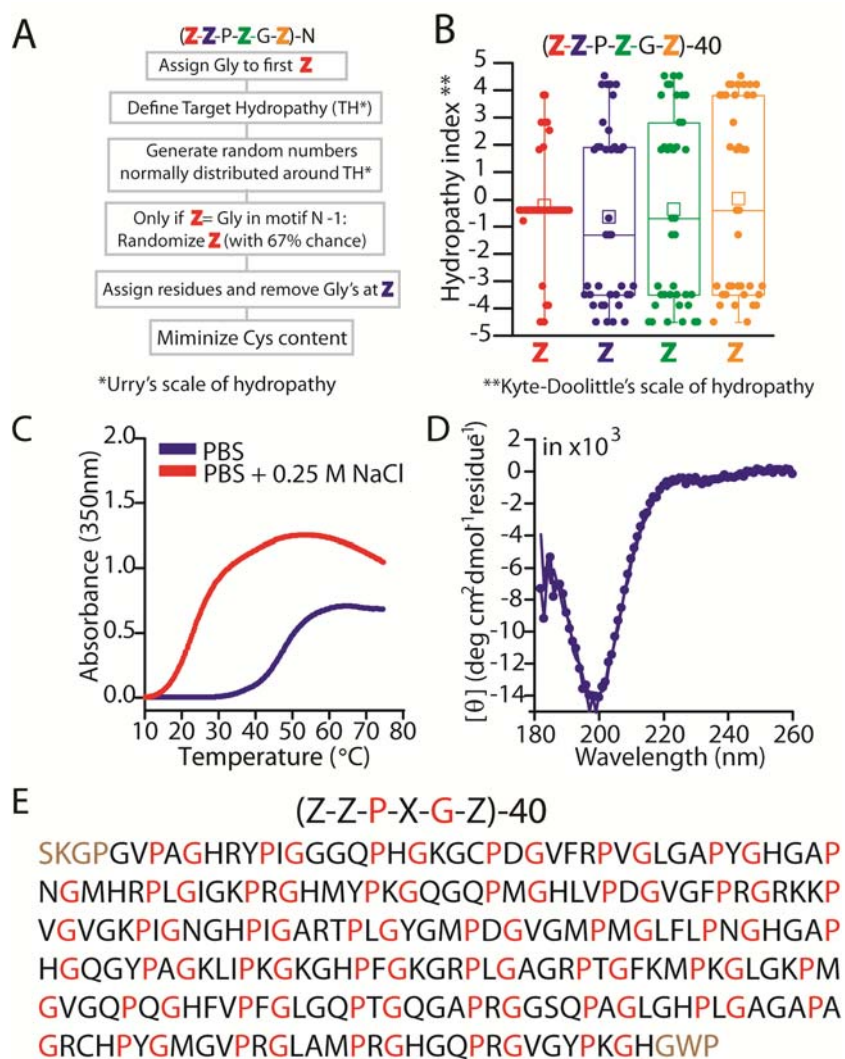


Figure 6. (A) We implemented a method to automatically generate a protein-sized polymer with a sequence that conforms to the amino acid patterns of non-fibrillar IDPs and wherein unconstrained Z residues are randomly selected to yield an average hydropathy that corresponds to an LCST of 37 °C. (B) Hydropathy distribution for all Z residue positions (B) in

the resulting protein-sized polymer — 240 amino acids in length. (C) Temperature-dependent turbidimetry in PBS at a polypeptide concentration of 50 μM indicated an LCST of $\sim 40^\circ\text{C}$. (D) The CD spectrum of this protein-sized polymer is characteristic of intrinsically disordered proteins. The CD data was collected in water at 5 μM and at 25°C . (E) Full amino acid sequence. The sequence includes the N terminal peptide SKGP and the C-terminal tripeptide GWP that were appended at the gene level before recombinant expression; residues shown in black correspond to the randomly assigned amino acids.

We next investigated in detail the LCST phase behavior of these peptide polymers. Circular dichroism (CD) spectroscopy at 25°C , a temperature below the operational definition of the LCST of the characterized polymers — defined by the onset temperature of the phase transition —, showed that these peptide polymers have a secondary structure that is largely disordered as seen by a major negative peak at $\sim 197\text{ nm}$ that is characteristic of random coils, and which is similar to that of ELPs and IDPs below their LCST (Figure 5E) [187]. Temperature-dependent turbidimetry identified peptide polymers with three types of LCST phase behavior that are defined by the degree of thermal hysteresis in their phase transition behavior: i) zero, ii) finite, and iii) heat-sensitive infinite hysteresis (Figure 5F). In the first group, we observed no hysteresis between the measured LCST on heating and cooling, as is observed for tropoelastin and many ELPs, whereas the second group displayed a consistent shift in their LCST between the heating and cooling cycles. The “heat-sensitivity” of the third group refers to the finding that most of these peptide polymers display zero hysteresis below a critical threshold temperature (typically around 40°C for the polymers synthesized herein), but aggregate irreversibly when heated above this threshold as shown by the complete lack of reversibility in their turbidity versus temperature profile

upon cooling (Figure 5G and Figure 7). To investigate the role of molecular weight on this behavior, we characterized polymers with step-wise increases in the number of repeats over a broad range of cooling rates and found that hysteretic LCST phase behavior is sensitive to changes in molecular weight and that a minimum number of repeats is required to exhibit a pronounced hysteretic behavior that is kinetically-stable (Figure 8).

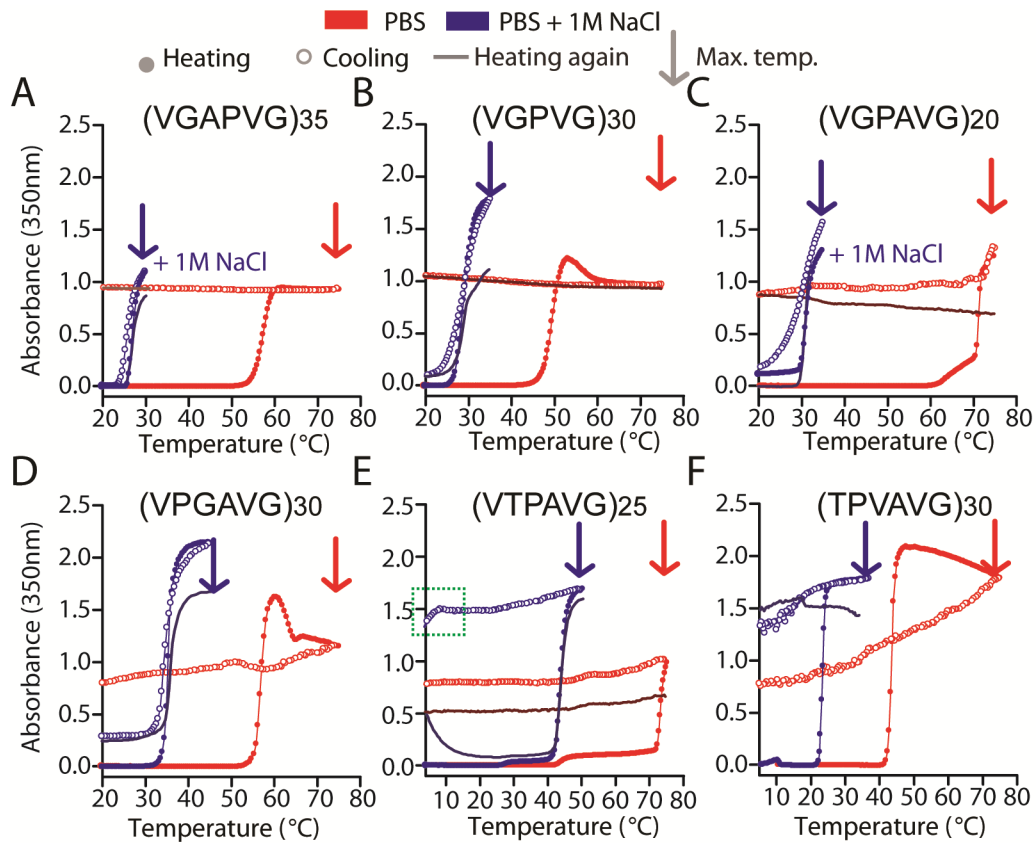


Figure 7. Peptide polymers that exhibit LCST behavior may display temperature-dependent hysteresis in their phase transition behavior. (A-F) Temperature-dependent turbidimetry for six different peptide polymers revealed a form of heat-sensitive hysteresis. These polymers display irreversible phase separation when heated to 75 °C, whereas they exhibit reversible phase transition behavior if heated below a given threshold temperature (indicated by blue arrows). Peptide polymers with repeat unit VTPAVG (E) exhibit a complex type of phase

transition behavior with heat-sensitive infinite hysteresis above the threshold temperature (see red curve), and finite hysteresis — we typically observed zero hysteresis for other sequences — below this temperature (see blue curve). Note that the finite hysteresis of this peptide polymer 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 the peptide polymer 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, we note that this peptide polymer —like all others reported in this manuscript— was purified exploiting the reversibility of its phase behavior in response to changes in buffer ionic strength. The phase behavior was characterized in PBS at a polypeptide concentration of 50 μ M.

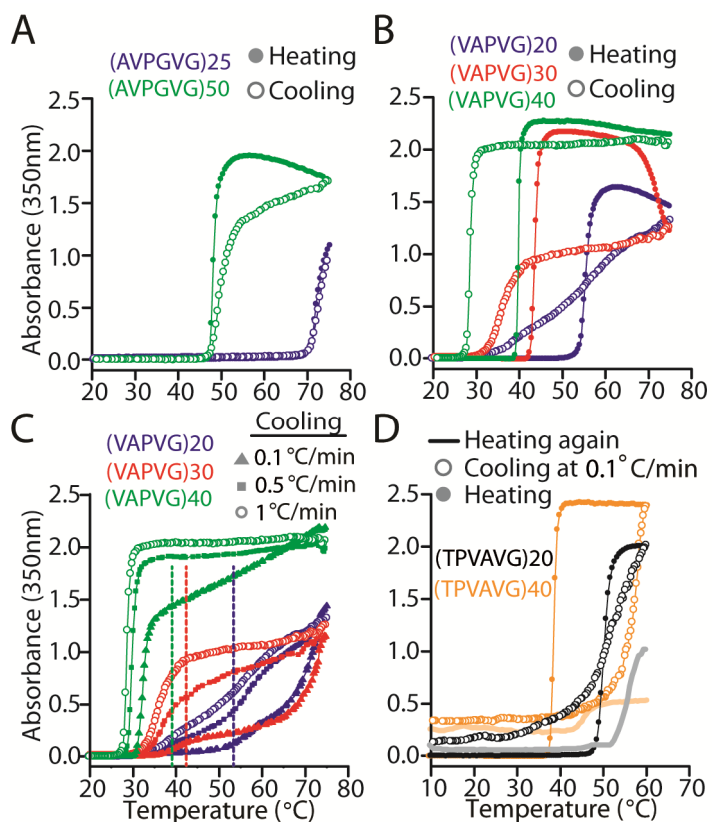


Figure 8. Molecular weight influences the hysteretic phase behavior of LCST peptide polymers. (A) Polymers composed of AVPGVG repeats, a motif that upon polymerization exhibits no hysteresis, display lower LCST values as the length of the polymer is increased, but even the longest polymer shows no hysteresis in its phase behavior. (B-C) Polymers composed of 20, 30 and 40 repeats of the pentapeptide VAPVG, a motif that upon polymerization exhibits finite hysteresis, also display progressively lower LCST values and with evidence of hysteretic phase behavior (B), but more than 20 copies of the motif are needed to obtain polymers with

marked hysteresis, as evidenced by a sharp transition on cooling. The degree of thermal hysteresis is also highly reproducible over multiple cycles of heating and cooling (Figure 18 found in *Supplemental Figures*). Moreover, peptide polymers with 40 and more (data not shown) copies of the repeat are highly kinetically-stable (C) as their LCST on cooling is almost unaffected by a 10 fold decrease in the cooling rate from 1 °C/min to 0.1 °C/min. The LCST's on heating (measured at a heating rate of 1 °C/min) for each temperature scan are depicted in (C) as dotted lines, as the curves are invariable. (D) Polymers composed of 20 and 40 repeats of the hexapeptide TPVAVG, a motif that upon polymerization exhibits infinite hysteresis, show the expected molecular weight dependence of the LCST but this increase in length also influenced 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 fail to exhibit a second phase transition.

We also performed CD spectroscopy on these polymers as a function of solution temperature to understand the structural origins of this hysteresis (Figure 5H; Figure 17 and Figure 19 in *Supplemental Figures*), as the current view that LCST behavior in polymers is entropically-driven as (ordered) waters of hydrophobic hydration become bulk water, falls short to explain such complex hysteretic behavior [32]. First, analogously to the folding of IDPs on binding to a target [18], we show, by measuring the polymer conformation before, during and after the onset of aggregation, that peptide polymers aggregate at the LCST with minor changes in conformation — mainly a small decrease in the intensity of the negative random coil peak at ~197 nm —, but undergo a significant conformational change after macroscopic aggregation, as typically evidenced by a large decrease in the intensity of the negative peak at ~197 nm (Figure 5H) that coincides with the plateau region of the turbidity versus temperature curve (Figure 17A-B and Figure 19 found in *Supplemental Figures*). This suggests that inter-chain contacts drive the emergence of increased order in the aggregated phase. Interestingly, for

polymers with a low to intermediate degree of hysteresis in their LCST behavior, upon cooling significantly below the LCST, the secondary structure reverts back to the disordered state seen prior to heating the sample (Figure 5H). In contrast, for polymers that exhibit a larger degree of thermal hysteresis, the intensity of the random coil peak is not recovered even upon cooling well below the LCST (Figure 17B-D found in *Supplemental Figures*). Moreover, for polymers of the hexapeptide TPVAVG, which exhibited the largest degree of hysteresis among all peptide polymers synthesized herein (Figure 7F), the initial decrease in the intensity of the random coil peak at the onset of the LCST is followed by the emergence of negative peaks at ~208 nm and ~220 nm (Figure 17B found in *Supplemental Figures*) at the plateau region of the turbidity profile that are indicative of a forming α -helical structure. The persistence of these rigid secondary structural elements (Figure 17B-D found in *Supplemental Figures*), once formed above the LCST, can stabilize the aggregated phase by providing stable inter-chain contacts that are not easily disrupted by cooling, thereby providing an interesting molecular insight into the origin of the near-infinite thermal hysteresis in the LCST behavior of this subset of peptide polymers. This is in accordance with the small degree of thermal hysteresis exhibited by polymers of N-isopropylacrylamide, the slight (~3 °C) degree of hysteresis of N-isopropylmethacrylamide — attributed to increased polymer-polymer hydrogen bonding and reduced polymer flexibility [191, 192]—, and the rather large thermal hysteresis (~20 °C) seen in the LCST behavior of polysaccharides such as

methylcellulose [193], as only the latter assume rather rigid conformations and engage in significant hydrogen bonding [194].

We next mutated the residues at the -1 position of three amino acid motifs (V-1PVG, V-1PGVG, and V-1PAVG) that span P-X_n-G motifs with *n* values between 0 and 2 to explore the prediction that Gly at the -1 position promotes aggregation down a fibrillar pathway, which is in turn expected to lead to greater thermal hysteresis or irreversible thermal aggregation. Our results confirm the prediction that the occurrence of Gly one residue N-terminal to repeating P-X_n-G motifs increases the propensity for aggregation (Figure 20 found in *Supplemental Figures*) and stabilizes the aggregate state to increase hysteresis on cooling, whereas bulkier residues at the -1 position (e.g., Arg, His and less effectively Ala) reduce or eliminate hysteresis (Figure 5I; Figure 21 found in *Supplemental Figures*). This may explain the evolutionary pressure against Gly at this position in non-fibrillar, Pro- and Gly-rich IDPs.

The synthesis of this diverse set of LCST peptide polymers also revealed other unexpected features that seem to distinguish peptide polymers from synthetic polymers and proteins. Analogous to the role of directionality in all natural languages (e.g., pan≠nap and loot≠ tool), we found that these peptide polymers exhibit differences in their “meaning” — defined here by their phase behavior — upon reversal of their amino acid sequences. Backbone reversal is interesting because the resulting “retro” polymer has an identical sequence as the parent polymer if read from the C- to the N-terminus,

but the overall composition and other physicochemical properties remain unaltered (Figure 9A). We synthesized six different peptide polymers, wherein each polymer in a pair corresponds to the forward or reverse sequence, and found that backbone reversal led to dramatic changes in the LCST behavior of the peptide polymers as seen by the dramatic difference in the hysteresis of the LCST behavior between members of each pair (Figure 9B-D), despite the fact that all of these peptide polymers are intrinsically disordered below their LCST (Figure 9E). This is an interesting and somewhat unexpected finding because synthetic polymers do not exhibit directionality dependent properties and indeed the concept of directionality is not commonly discussed in the synthetic polymer literature². On the other hand, proteins are inherently directional, in that reversal of the primary amino acid sequence of a protein or well-folded peptide usually alters or abolishes their folding and hence its function [195, 196], suggesting that these peptide polymers occupy a functional space that is intermediate between synthetic polymers and proteins. This finding also calls for an examination of the residue-level determinants of the various behaviors exhibited by Pro and Gly-rich IDP's (e.g., elastomeric vs. amyloidogenic), as the behavior of these retro motifs clearly shows that the overall Pro and Gly content is only one factor that controls their behavior [197].

² For example, synthetic diblock polymers are termed AB block copolymers by convention, as BA block copolymers are considered redundant

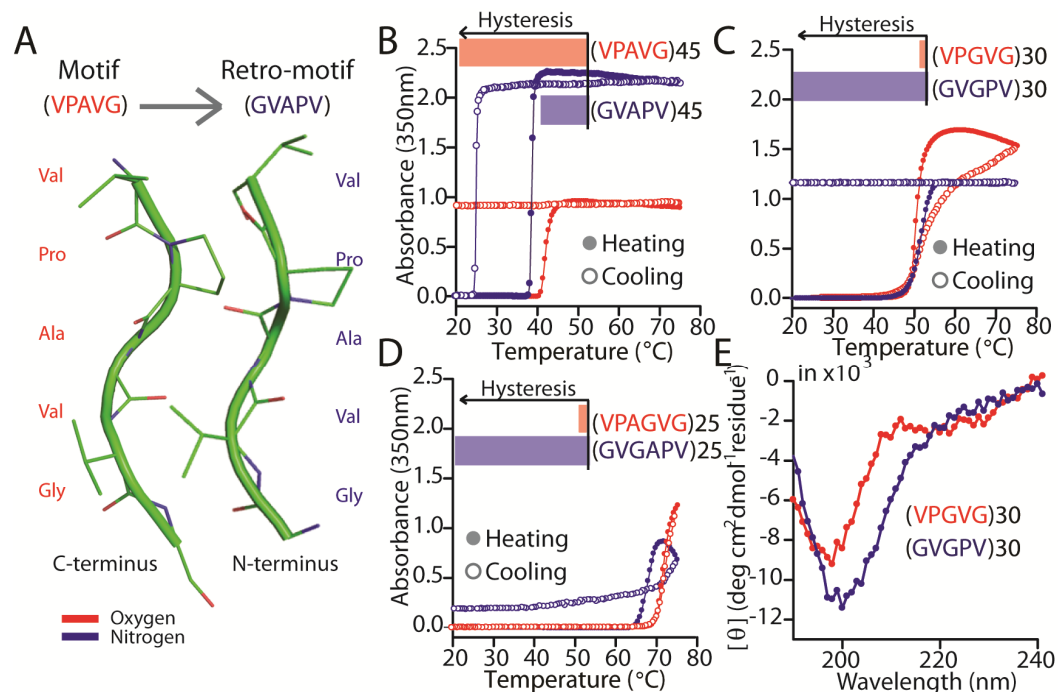


Figure 9. LCST peptide polymers exhibit properties intermediate between synthetic polymers and proteins. (A) Backbone-reversed peptide polymers present identical patterns of amino acid side chains, which we illustrate with the structure of a pentapeptide motif and its retro-motif as observed in the crystal structures of two different proteins (PDB id 3MKR_B and 1OZP, respectively). The images were rendered using PyMOL (<http://pymol.org/>). (B-D) Temperature-dependent turbidimetry showed profound changes in the phase behavior of peptide polymers on backbone-reversal, as seen by the change in the hysteresis of the phase transition, despite the intrinsically disordered nature of the parent and retro polymers below their LCST according to CD spectroscopy (Figure 5E). All turbidity measurements were conducted in PBS (unless otherwise indicated) at a polypeptide concentration of 50 μM . CD studies were conducted in water at a polypeptide concentration of 5 μM and at 25 $^{\circ}\text{C}$.

The tunable thermal hysteresis in the phase behavior of LCST peptide polymers suggests the possibility to control polymeric nanoassembly beyond amphiphilicity as the sole driver of self-assembly, by controlling both the amphiphilicity imparted to the polymer by phase separation and the generation of secondary structure upon aggregation to dictate self-assembly. To explore this possibility, we studied two diblock

copolymers composed of the same corona-forming block — an ELP with an LCST around 58 °C— and different hydrophobic blocks with distinct hysteretic phase behaviors but LCSTs that only differ by ~1°C. We show that they form two types of nanoparticles (Figure 10A-B) of unusually large hydrodynamic radius compared with typical ELP micelles that self-assemble from ELP copolymers of similar length [198], suggesting that these copolymers are not assembling into the structures that we have observed —largely spherical micelles— with existing LCST peptide block copolymers. As expected, the degree of hysteresis directly controls nanoparticle stability on cooling below the critical assembly temperature. Figure 10A shows that the heat-sensitive infinite hysteresis of the hydrophobic core in this copolymer prevented the disassembly of the nanoparticles on cooling well below the critical assembly temperature as evidenced by consistently high absorbance values at all temperatures. In comparison, the finite hysteresis of the hydrophobic core in a different copolymer (Figure 10B) stabilized the nanoparticles below the critical assembly temperature but only for a temperature range that is consistent with the degree of hysteresis of the isolated block (~10 °C). The role of molecular weight in determining the hysteretic phase behavior of these polymers and hence the stability of the nanoparticles was also evidenced by the synthesis of two diblock copolymers with identical block ratios. Interestingly, copolymers with 20 repeats of the motif TPVAVG —unlike the longer polymer studied in Figure 10A— do not remain in a nanoparticle form after undergoing bulk aggregation

above the LCST of the ELP corona, as seen by the large absorbance values upon cooling (Figure 10C). Although we believe that ELP micelles disassemble upon bulk aggregation once the “hydrophilic” block in the corona undergoes phase transition—as the amphiphilicity that drives their assembly vanishes away—the distinct assembly behaviors seen on triggering the bulk aggregation of these two closely related copolymers, and the realization that molecular weight is a strong modulator of their hysteretic phase behavior (Figure 8), suggests that the stability of these nanostructures may be tuned to tolerate their assembly into higher order structures upon bulk aggregation. We thus show first evidence that the synthesis of multi-block copolymers that are self-assembled through the LCST of peptide polymers with different hysteretic phase behaviors leads to interesting changes in the morphology, size and stability of the resulting nanoparticles. This morphological diversity is an exciting outcome, as nanoparticle morphology is a well-known modulator of tissue distribution and pharmacokinetics [51]. The idea that the molecular basis of this hysteretic phase behavior overrides the assembly route dictated by peptide amphiphilicity is an intriguing new concept for the bottom-up assembly of “smart” polymers and is likely to yield diverse nanoparticle morphologies with tunable stability and numerous applications in medicine.

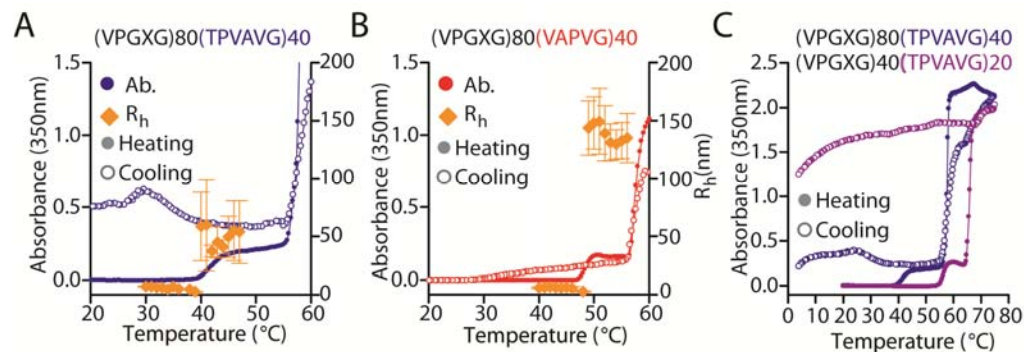


Figure 10. New forms of hysteretic LCST phase behavior are useful in controlling the thermally-triggered nanoassembly of peptide polymers. (A-B) By forming diblock copolymers composed of a hydrophilic ELP block (where X alternates between A and G) and a hydrophobic block that repeats a hysteretic LCST motif with either heat-sensitive, infinite hysteresis (A) or finite hysteresis (B), we synthesized self-assembling nanoparticles that remain assembled despite significant cooling below their critical assembly temperature. The assembly on heating is evidenced by changes in absorbance (filled circles) and hydrodynamic radius (diamonds), and the stability of the particles on cooling is shown by stable absorbance measurements (open circles) at temperatures below the LCST of the ELP corona ($\sim 58^{\circ}\text{C}$) and the apparent LCST of the more hydrophobic block. (C) A diblock copolymer with the same composition and block ratio as in (A), but having half the number of repeats in the core — that is, a constant ratio between the size of the core and the corona: 20/40 vs. 40/80 diblocks— fails to remain assembled through the bulk aggregation event experienced above the LCST of the ELP in the corona.

The unique features exhibited by this library of peptide polymers leads us to suggest that “smart” peptide polymers can be best described as linear macromolecules that are composed of amino acid “letters” that are organized as “words”, with higher order organization of one or more words that repeat or recur to create a “phrase” (the macromolecule) and postulate that the syntax of this class of polymers, understood as the arrangement of amino acids into repeat units that have a functional behavior, controls their function. Hence, by analogy to syntax in natural language — defined as the arrangement of words in a phrase that controls its meaning [199]—, we coin a new term,

syntactomers³, to describe polymers whose properties are controlled by their organization as a collection of letters into words and their higher order organization into functional phrases. We note that the more general idea of language in proteins remains incipient compared with the more established concept of nucleic-acids linguistics [199], but that the idea may be readily applicable to peptide polymers wherein the syntactic units are somewhat obvious [200].

In conclusion, syntactomers are a class of linear macromolecules with properties that are intermediate between synthetic polymers and globular proteins. In their directionality-dependent properties, syntactomers resemble globular proteins, while in their intrinsically disordered structure they most resemble IDPs and synthetic polymers in a good solvent. Unlike synthetic polymers, however, syntactomers can be designed to span a linguistic complexity that ranges from the relatively simple such as a polymer composed of a single repeating word of a few letters to polymers that are a non-repetitive collection of letters with a complex syntax, similar to proteins. The sequence diversity of peptide syntactomers may be exploited to create macromolecules that exhibit dual and orthogonal functions, namely LCST behavior and bioactivity, as will be shown in chapter 5, and may be exploited to study the intriguing phase behavior of other tandem repeats in proteins [201], such as low complexity domains in RNA and DNA-binding proteins [202] and the FXFG repeats that form the permeability barrier in

³ This term was suggested by Prof. Gabriel Lopez to my PhD advisor, Ashutosh Chilkoti.

nuclear pore complexes [203]. The seamless fusion of materials and protein design embodied by syntactomers promises, we hope, a new generation of designer polymers with multiple levels of embedded functionality that should lead to new functional materials.

4.2. Supplemental information

4.2.1. Supplementary discussion

The results of the compositional profiling of Gly residues around P-X_n-G motifs across Pro- and Gly-rich IDPs from Table 3 (Figure 4B) suggested that most residue positions in the vicinity of these motifs have an abnormal Gly content (i.e., either higher or lower than the overall Gly content). However, our analysis of randomized clones for each IDP revealed that many of those biases in local Gly content are not rare for proteins with similar amino acid composition, since the distribution of FCRG (Fold Change from Random Gly) values for the set of randomized clones is populated by similar FCRG values (and more extreme values) with significant probabilities. The summary of these p-values calculated for each residue position are shown in Table 2. Interestingly, this analysis shows that the only residue position with statistically significant biases (p-value<0.001) in local Gly content for all Pro- and Gly-rich proteins is position -1. Moreover, FCRG values at -1 show a consistent trend that differentiates between fibrillar IDPs and non-fibrillar IDPs. The former show positive FCRG values indicative of Gly enrichment, whereas the latter –non-fibrillar– group of proteins are depleted of Gly at

this position. This is interesting since no other residue position showed a consistent trend that separated fibrillar and non-fibrillar IDPs. Noteworthy, our experimental data (Figure 5I; Figure 20 and Figure 21 found in *Supplemental Figures*) also confirmed the relevance of this residue position for the design of peptide polymers composed of P-X_n-G motifs, since the occurrence of Gly at position -1 resulted in polymers that recapitulated the irreversible thermal aggregation expected for fibrillar IDPs.

Table 2. P-values for the FCRG observed at residue positions surrounding P-X_n-G motifs in Pro- and Gly-rich IDPs. FCRG values are shown in Figure 5B.

Protein	p-values for FCRG			
	-2	1	+1	+2
Elastin Homo	0.014	0	0.052	0
Elastin Bovine	0.005	0	0.002	0
Elastin Mus Musculus	0.003	0	0.341	0.061
Elastin Zebrafish	0.31	0	0	0.004
Resilin	0	0	0.048	0.331
Gluten	0.038	0	0	0
Collagen I	0	0	0	0
Flag. Silk	0	0	0	0
Drag. Silk	0	0	0.354	0

The identity of the amino acid at position -1 might affect the reversibility of the phase transition through changes in the propensity for cis-trans isomerization of the Xaa-Pro peptide bond. Tight packing of collagen fibrils, for instance, requires an all-trans conformation for all Xaa-Pro bonds [204]. Residues like Gly and Ala in the Xaa position – for short peptides of the form Ala-Xaa-Pro-Ala-Lys and Gly-Xaa-Pro-Gly – favor isomerization into the trans-Pro form at higher rates than bulkier hydrophobic residues (e.g. Ile, Val, Arg, His, etc.) [204, 205]. However, we note that while Ala and Gly have been reported to have similar cis-trans isomerization rates, in our sequences we have typically observed a greater degree of hysteresis for Gly than for Ala. This discrepancy may favor a simpler explanation where the bulkiness of the side chain modulates reversibility through a steric effect that facilitates or limits chain packing on aggregation (Figure 21 found in *Supplemental Figures*). The challenge in establishing the importance of this cis-trans isomerization for the observed hysteresis is accentuated by the known role of polypeptide conformation on isomerization rates, since the studied peptide polymers might undergo different conformational changes upon phase transition. We believe that the peptide compositions reported herein may be a good starting point for more thorough studies on the role of primary and secondary structure on cis-trans isomerization in protein-like systems, and to study the effect of this isomerization on the stabilization of protein tertiary and quaternary structure, particularly that of fibrillar and amyloidogenic proteins involved in various neurodegenerative diseases. These studies

should benefit from our ability to use force spectroscopy to “image” the process of cis-trans isomerization in Pro-rich polypeptides [206].

To further explore the sequence-function space that lies between typical synthetic polymers and proteins, we asked whether we could rationally synthesize a peptide polymer that would not directly recapitulate the sequence of Pro- and Gly-rich IDPs that exhibit LCST, as they appear to be the closest protein analog of LCST peptide polymers. We hence conceived a peptide polymer composed of widely diverse Z^2 - Z^1 -P-Z-G- Z^{+1} hexapeptides and that conforms to the basic syntax of non-fibrillar Pro- and Gly-rich IDPs, namely the repetition of a P- X_n -G motif, the lack of Gly at Z^1 and the enrichment of Gly at Z^2 (this position is equivalent to positions -2 and +2 in Figure 4B) [189]. We chose a P- X_1 -G repeat (i.e., $n=1$, $X=Z$) to deliberately diverge from the syntax of IDPs, as this motif occurs with low frequency in all proteins that we studied except collagen (Figure 4A and Figure 4D). To control the global polymer hydrophobicity, which will certainly influence the LCST value but not necessarily the type of LCST phase behavior (e.g., degree of hysteresis), we took advantage of Urry’s scale of hydrophobicity — that assigns an experimental LCST to all natural amino acids [207] — by randomizing unconstrained Z residues according to a normal distribution with a target LCST of 37 °C, and a large standard deviation of 50 °C to maximize sequence diversity. We then implemented an algorithm (Figure 6A and *Materials and Methods*) to automatically generate a protein-sized polymer composed of 40 different hexapeptides (i.e., $N=40$)

with this syntax [189]. The resulting peptide polymer (see Figure 6E for sequence) has a remarkable sequence diversity that is evidenced by the broad range of hydrophathies spanned at each Z residue position (Figure 6B). This diversity is indeed strikingly wider compared with that of tropoelastins all the way from humans to zebrafish (compare Figure 6B and Figure 22; the latter is found in *Supplemental Figures*). This protein-sized polymer is intrinsically disordered (Figure 6D) and displays infinite hysteresis in its phase behavior with a LCST of ~40 °C (at 50 μM in PBS, Figure 6C). Global/mean hydrophobicity hence determines the LCST value for an unstructured peptide polymer with a high density of structure breaking Pro and Gly residues as the one studied here, but these primitive syntactic elements are insufficient to design polymers with finite and zero hysteresis in their phase behavior.

4.2.2. Supplemental figures

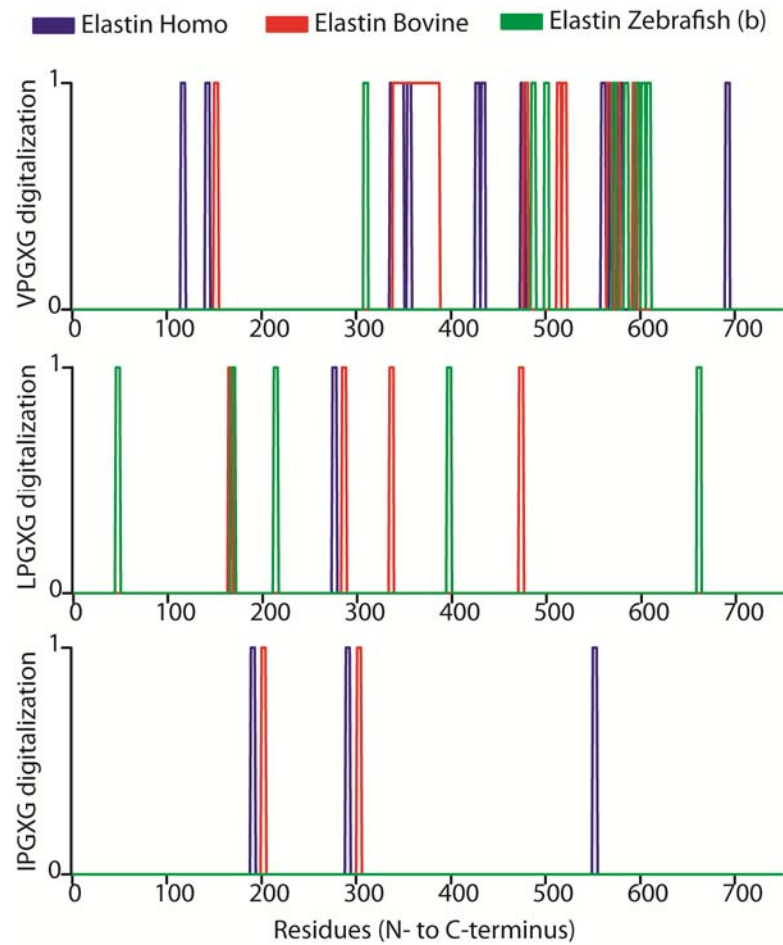


Figure 11. Distribution of the canonical VPGXG motif and its analogs IPGXG and LPGXG along the primary amino acid sequence of tropoelastin from different species. The localization of these motifs is digitized such that all residues forming part of a motif are assigned a value of 1 and any other residues are set to 0. These three motifs altogether account for ~10-20% of the amino acid residues in these proteins. Of these, VPGXG is the most abundant motif.

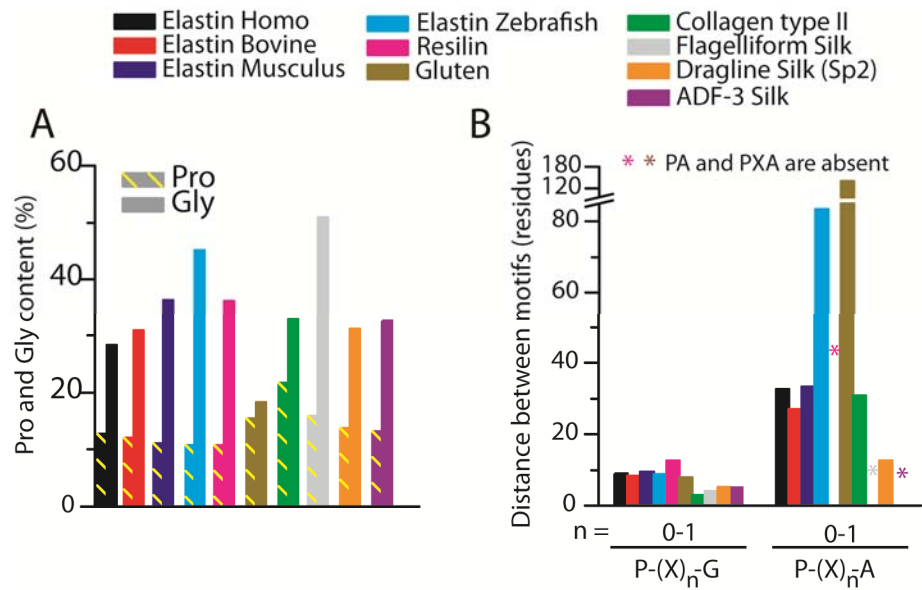


Figure 12. (A) Pro and Gly content of all Pro- and Gly-rich IDPs shown in the figure legend. (B) Average distance between P-X_n-G and P-X_n-A motifs (for n=0-1) across these proteins. Note that some proteins lack P-X_n-A motifs (indicated by asterisks in B).

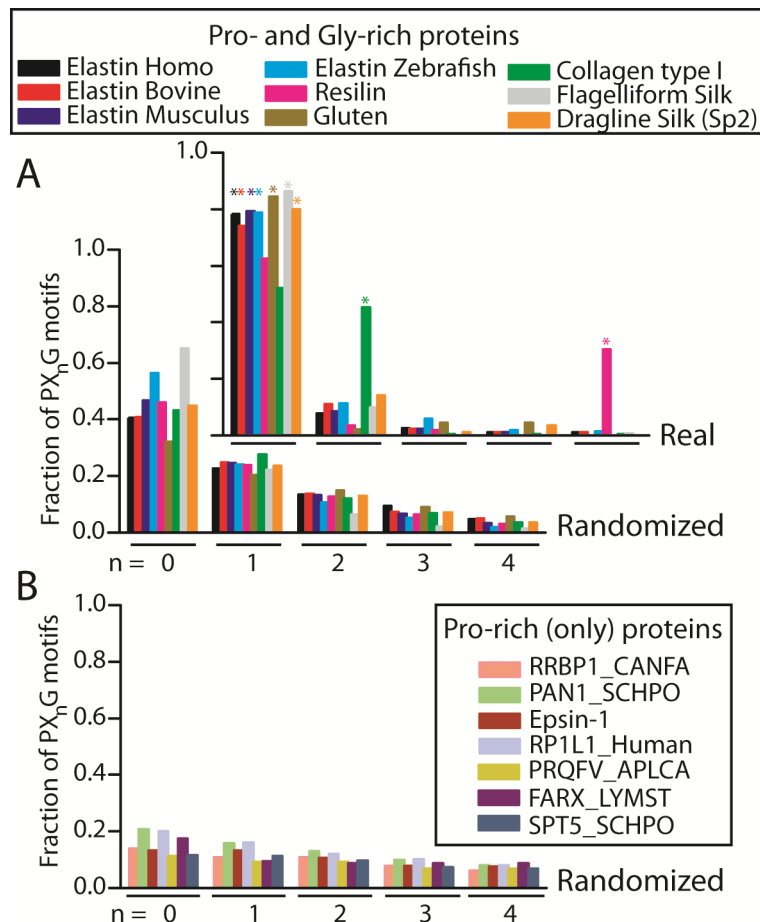


Figure 13. Analysis of randomized Pro- and Gly-rich IDPs to assess the statistical frequency of motifs with Pro and Gly residues in close proximity. First, we generated 1000 randomized versions of each IDP listed in the figure caption (see supplementary methods for details) using the amino acid frequencies characteristic of each protein. We then assessed the abundance of individual P-X_n-G motifs (for n=0-4) relative to the total number of P-X_n-G motifs in both the real (parent) protein and all randomized versions. (A) Fraction of P-X_n-G motifs in the set of randomized and real proteins. For the randomized proteins we plotted the median value of the entire data set. Asterisks indicate statistical significance (p-value<0.001) for over-represented motifs compared with randomized proteins. We then repeated this same analysis for an additional set of Pro-rich (only) IDPs that are not rich in Gly. (B) Randomized versions of these proteins are almost unbiased in the frequency with which different P-X_n-G motifs occur along the protein amino acid sequence. We deemed a result as statistically significant if the p-value was smaller than 0.001.

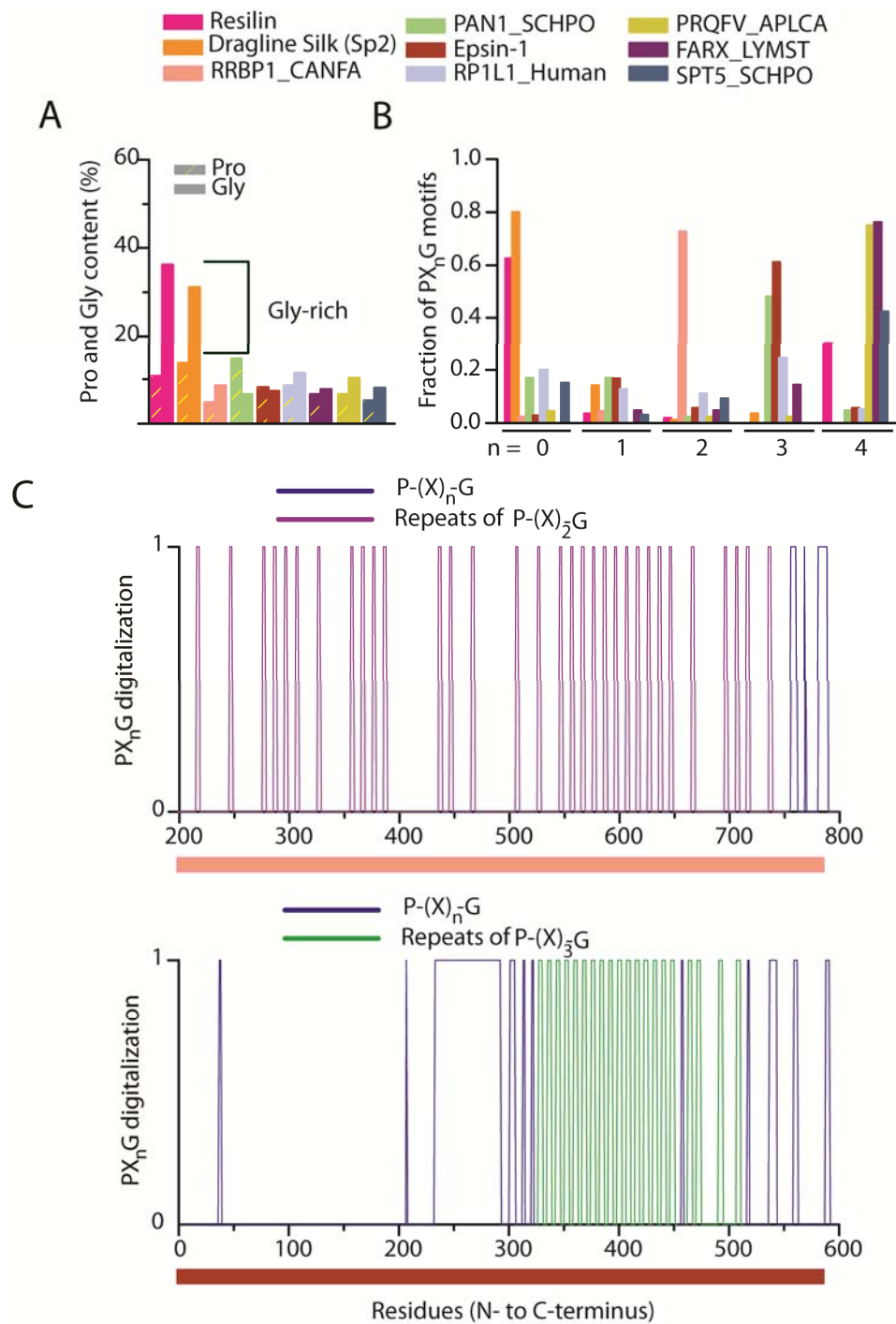


Figure 14. We identified Pro-rich but Gly-poor proteins that display P-X_n-G motifs with n=2, 3 and 4. (A) Quantification of the Pro and Gly content of these proteins (resilin and dragline silk were also included for reference). (B) Abundance of individual P-X_n-G motifs with n=0-4

relative to the total number of motifs per protein. Unlike our previous data set of Pro- and Gly-rich IDPs, these Pro-rich proteins are highly biased toward P-X_n-G motifs with n=2, 3 and 4. The absolute abundance of these motifs in the overall protein sequence is comparable to the abundance of P-X₄-G motifs in resilin (data not shown). (C) The digitalization of the protein sequence for two of these proteins, in which residues —other than Gly— participating in a P-X_n-G motif were assigned a value of 1, while all other residues were set to 0, reveals that these motifs recur at high frequency in defined regions of the proteins. To facilitate the visualization of the most abundant P-X_n-G units, the squared signals corresponding to the predominant type of motif (n=2 or 3 in these two examples) were colored to differentiate them from the blue, background P-X_n-G motifs.

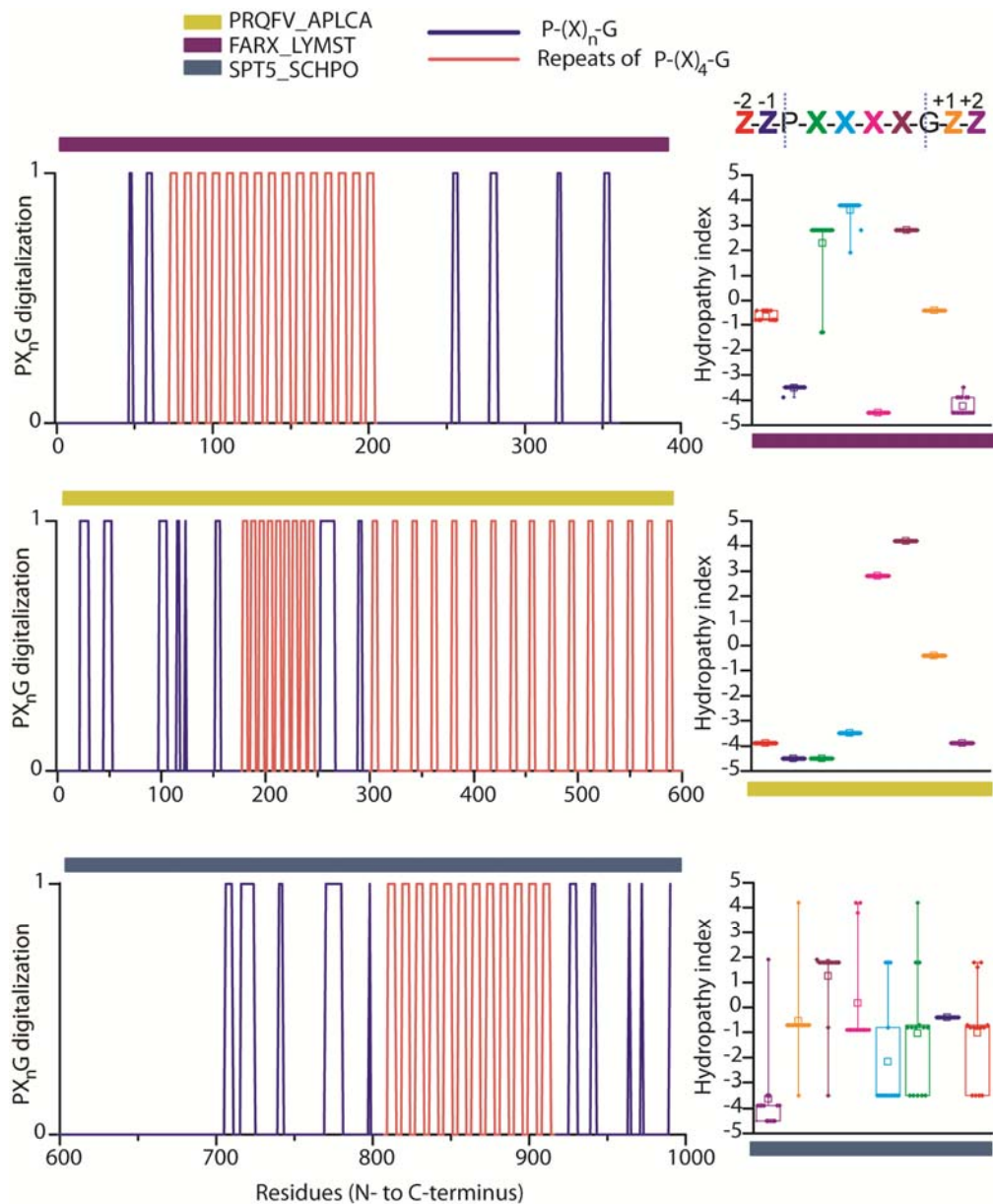


Figure 15. Pro-rich proteins use repeats of P-X₄-G motifs for the presentation of bioactive peptides. Here, we present the digital maps of P-X_n-G motifs across the sequence of these proteins – generated as in Figure 14. P-X₄-G motifs were highlighted in red to facilitate their identification. The right panel for each protein shows the hydropathy distribution (depicted by the Kyte-Doolittle scale of hydrophobicity) for the residues that form these repeats (i.e., X₄ residues and two Z residues N-terminal and C-terminal to the P-X₄-G motifs), which suggests that different residues in these P-X₄-G motifs can span a broad range of hydropathies. The two neuroactive proteins FARXamide-related neuropeptides and PRQFVamide display repeats of the bioactive peptides PFLRF and PRQFV, respectively, embedded within P-X₄-G motifs. We

observed a similar localized region of P-X₄-G units with highly conserved X₄ residues (right panel) in a transcription factor from yeast (SPT5).

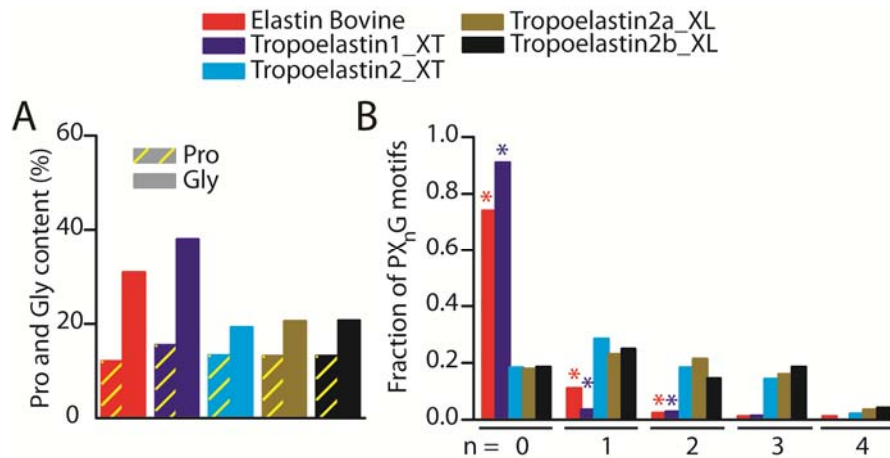


Figure 16. Frogs have unusual tropoelastins that are not biased towards a Pro-Gly dipeptide. The expression of two different tropoelastin genes with markedly different amino acid composition and length was recently reported in the amphibian *Xenopus Tropicalis* (XT) [190]. (A) Pro and Gly content for the two tropoelastin proteins from XT and other tropoelastins from bovine and *Xenopus Laevis* (XL). Gly content is a major difference, since Tropoelastin2 has a relatively low Gly content compared with its companion Tropoelastin1 and with all known tropoelastins from other species. A similar tropoelastin has been identified based on bioinformatics analysis in the tetraploid XL (encoded by two different genes, a and b) [190]. (B) Abundance of individual P-X_n-G motifs with n=0-4 relative to the total number of motifs per protein. As in the case of Pro-rich IDP's, tropoelastins with relatively low Gly content are not enriched in Pro-Gly dipeptides. These proteins show a random-like distribution of Pro-X_n-Gly motifs for n values between 0 and 4. Asterisks indicate statistical significance (p-value<0.001) for over-represented motifs compared with randomized proteins (see *Materials and Methods* for details).

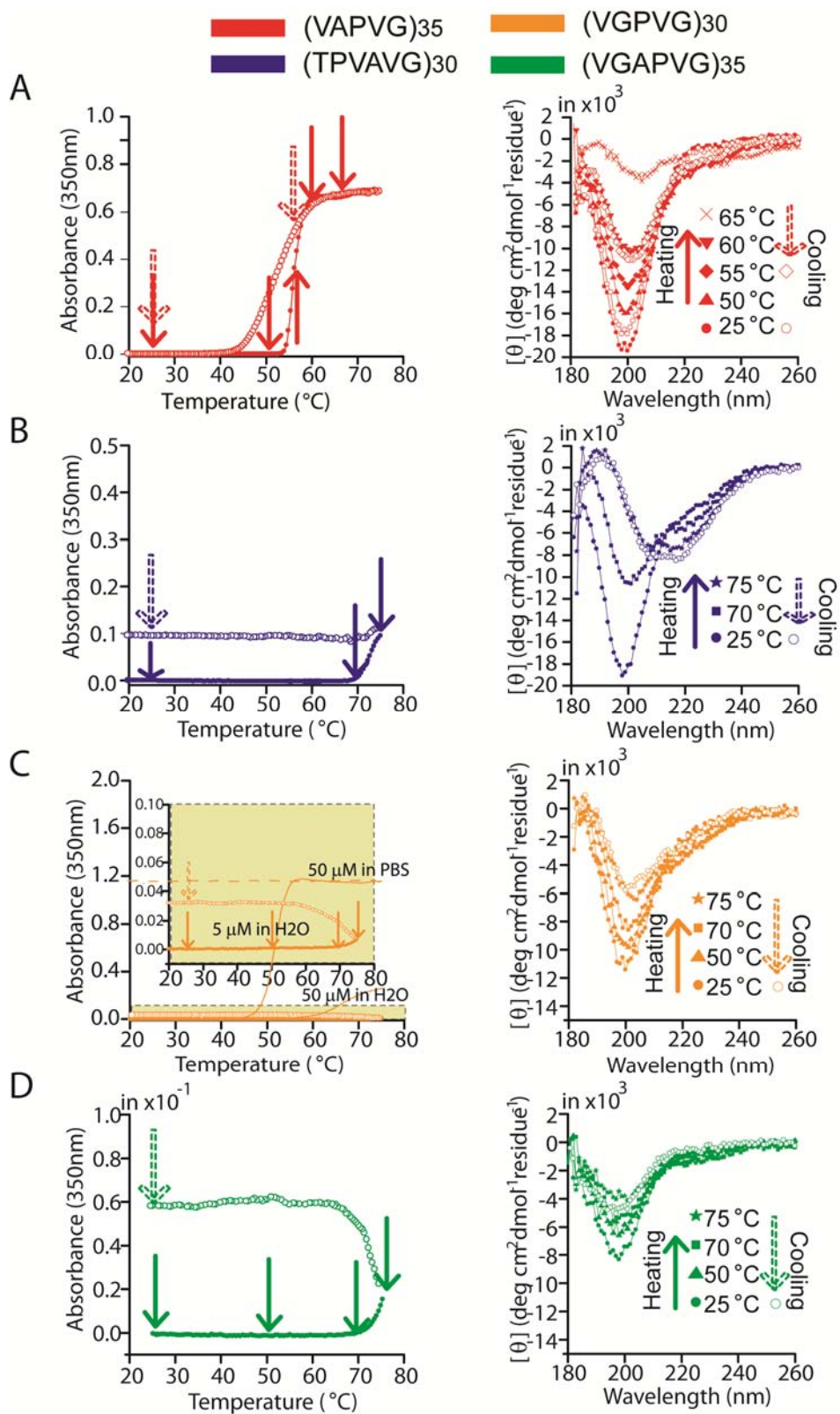


Figure 17. The heat-sensitive, infinite hysteresis of LCST peptide polymers arises from the emergence of ordered secondary structures that stabilize the insoluble phase. (A-D) The same sample for each polymer was analyzed by temperature-dependent turbidimetry (left panel) and CD spectroscopy as a function of temperature (right panel). Filled arrows in the left panels — turbidity versus solution temperature plots — indicate the temperature at which CD spectra were recorded upon step-wise heating the sample, while open arrows indicate the temperature at which CD spectra was acquired upon cooling. Whereas a peptide polymer that displays finite hysteresis (A, left) rapidly recovers its conformational disorder on lowering the temperature below the phase transition temperature in accordance with its degree of hysteresis (A, right), we show that peptide polymers with heat-sensitive infinite hysteresis (B-D) undergo conformational changes that persist on cooling. We note that a polymer herein shown to display a very large degree of hysteresis — such that we were unable to observe its reversibility below any given temperature threshold accessible with our instruments (Figure 7F) — also displays the more dramatic ordering on aggregation as seen by the emergence of negative peaks at ~208 nm and ~220 nm at the plateau region of the turbidity profile that are indicative of a newly formed α -helical structure (B). Turbidity and CD data were acquired in water at a polypeptide concentration of 5 μ M.

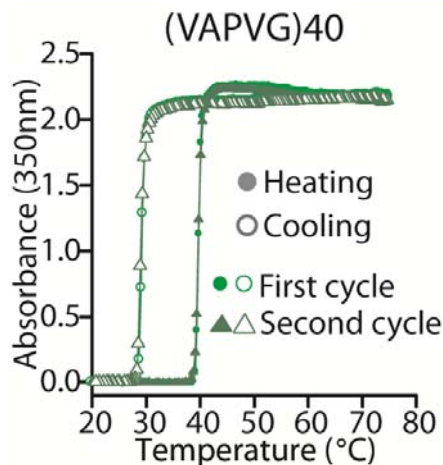


Figure 18. The degree of thermal hysteresis is highly reproducible over multiple heating and cooling cycles. The peptide polymer was characterized at a concentration of 50 μ M in PBS, at a heating and cooling rate of 1 $^{\circ}$ C/min.

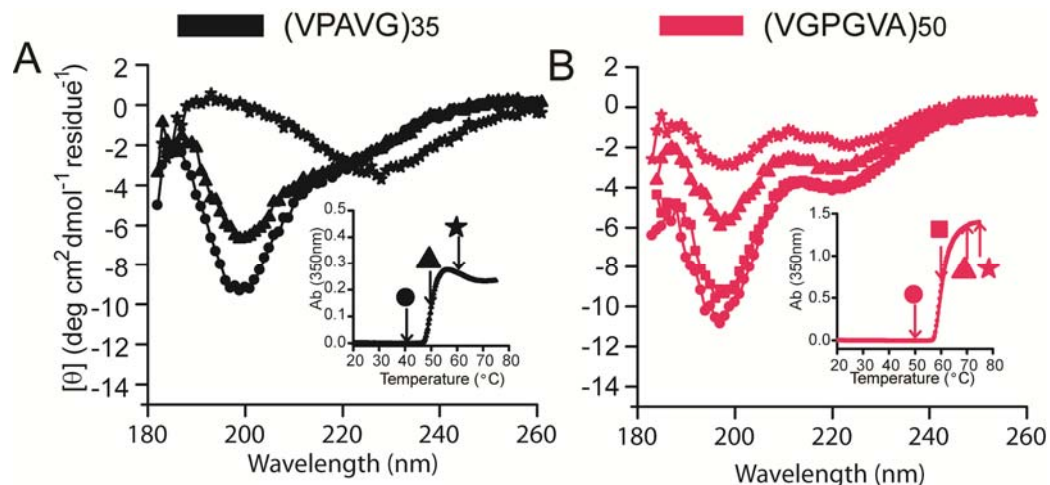


Figure 19. Elastin-like peptide polymers undergo aggregation at the LCST with minor conformational changes —seen by a small decrease in the random coil peak at ~197 nm. (A-B) CD spectroscopy of two elastin-like mutants (at 5 μ M in water) as a function of temperature and accompanying turbidity data (insets) for the same sample. Symbols and arrows indicate the temperatures at which CD spectroscopy was carried out. The most significant conformational changes occur after the polymers are fully aggregated, indicating that major reorganization occurs subsequent to dehydration and collapse of individual polymers chains that trigger the observable LCST, and are mediated by inter-chain interactions as seen by the large decrease in the intensity of the negative peak at ~197 nm. Data were adapted from Ref. [208].

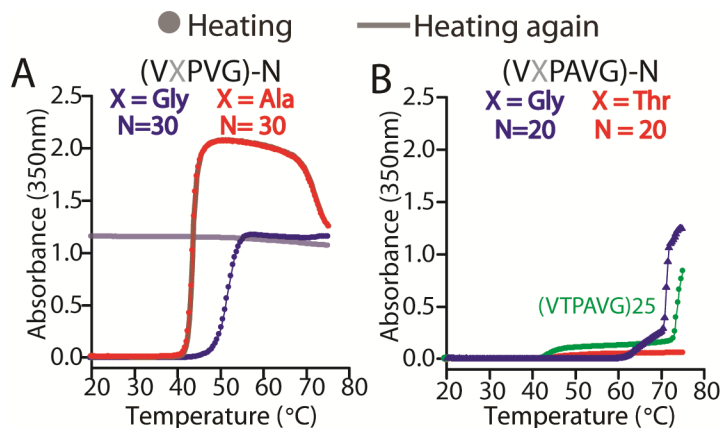


Figure 20. Gly residues preceding a P-X_n-G unit perturb the behavior of Pro-rich “smart” peptide polymers. We mutated two different LCST peptide polymers to incorporate a Gly residue N-terminal to the repeating P-X_n-G unit, since the occurrence of Gly at this position is a key differentiating factor between fibrillar and non-fibrillar, Pro- and Gly-rich IDPs (Figure 4B). (A-B) Temperature-dependent turbidity data for the parent and mutant peptide polymers showed that these mutations affected the reversibility of the phase transition (A) and the propensity for

aggregation (B). Note that (VGPAVG)-20 in (B) displays a lower LCST than a longer polymer of VTPAVG (where N=25), despite the similarity in hydrophobicity between Gly and Thr by both Kyte-Doolittle's and Urry's scale of hydrophobicity. In addition to these changes in phase behavior, we observed interesting changes in the ability of *E. coli* to produce these peptide polymers solubly, as if the Gly-mutants displayed the characteristic expression issues (that is, insolubility in the crowded intracellular space) of fibrillar IDP's. (VGPAVG)-30, for instance, which contains the abundant GPXG motif found in collagen, expresses as an insoluble peptide polymer but once purified (i.e., after denaturation and removal of denaturing agents), it displays high solubility in PBS reminiscent of canonical ELPs. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μ M.

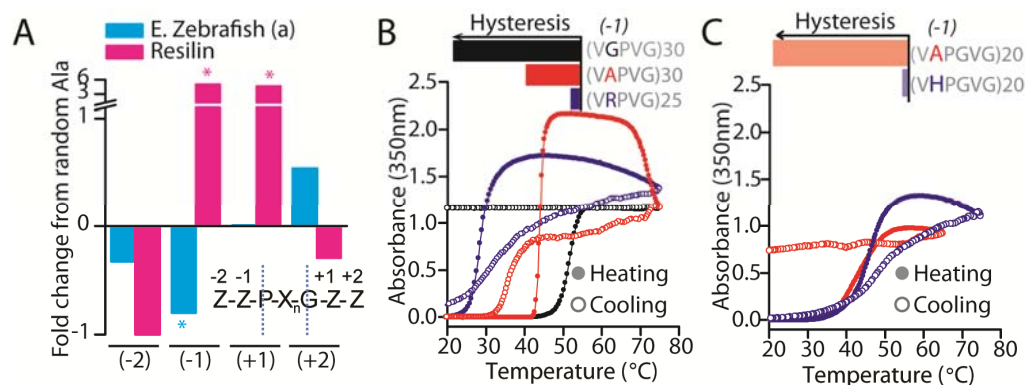


Figure 21. Steric hindrance at the residue position preceding a P-X_n-G motif modulates hysteresis. The occurrence of Gly — the only residue without a side chain— at this position was shown to be a likely modulator of the assembly behavior of IDP's (Figure 4B) as it discriminates between IDP's that go 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 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 was not interesting to us. However, tropoelastin from Zebrafish lacks such Ala enrichment in the crosslinking domains —it actually 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 in for Gly (see *Materials and Methods*). The residue N-terminal to P-X_n-G was the only surrounding position where we found a significant (p-value<0.01) 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 large hysteresis [184]. (B-C) Temperature-dependent turbidimetry of “smart” peptide polymers 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. We observed that whereas the parent (Ala-containing) polymers display large hysteresis in their phase transition behavior, the mutant polymers display little to no hysteresis. In (B) we also note

a larger degree of hysteresis for peptide polymers with Gly at this position, as compared with Ala, which is in accordance with the reduced bulkiness of the side chain. All turbidity measurements were performed in PBS at a polypeptide concentration of 50 μ M, except for VRPVG (+ 1M NaCl) and VAPGVG (+ 1 M NaCl).

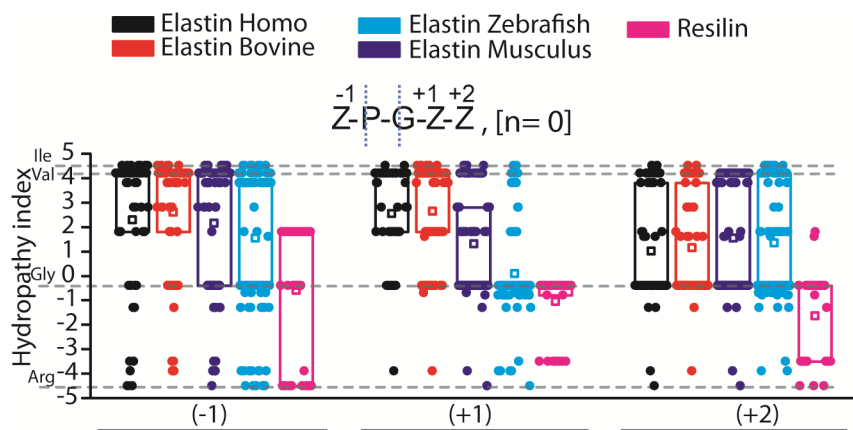


Figure 22. Hydropathy distribution of amino acids that surround the repeating Pro-Gly dipeptides in Pro- and Gly-rich IDP's that are known to display LCST. We noted that residues in the vicinity of Pro-Gly dipeptides, across the broader picture of evolutionary distant tropoelastin proteins and resilin, cover a broader range of hydropathies than what is evident by studying archetypical LCST tropoelastins (e.g., those from homo sapiens and bovine). Solid circles correspond to the index of individual residues, and the open squares indicate the mean hydropathy index for all residues at each Z position. Each box delimits the 25 and 75 percentile. The hydropathy index was plotted using the Kyte-Doolittle scale of amino acid hydrophobicity [209] wherein hydrophobic amino acids have a positive index. The indices corresponding to Ile, Val, Gly and Arg are indicated by dotted lines.

4.3. Materials and Methods

4.3.1. Materials

Restriction endonucleases, calf intestinal alkaline phosphatase (CIP) and T4 DNA ligase (400 U/ μ l) were purchased from New England Biolabs. PfuUltra™ II Fusion HS DNA polymerase (Pfu) was purchased from Stratagene; Go Taq Green Master Mix™ was purchased from Promega; Circligase™ ssDNA ligase was purchased from Epicentre Biotechnologies, and T4 DNA ligase (1 Weiss U/ μ l) was purchased from Fermentas.

Plasmid DNA was purified using Qiaprep™ spin miniprep kits and PCR products were purified using the Qiaquick™ PCR purification kit from QIAGEN. Double stranded DNA was visualized by electrophoresis on agarose gels (EMD chemicals) prestained with SYBR® Safe (Invitrogen). BL21(DE3) and EB5α E. coli strains were purchased from EdgeBio. All cultures were grown in Terrific Broth (TB) from BioExpress, and expression was induced with isopropyl β-D-thiogalactopyranoside (IPTG) from GoldBio. Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride was from Pierce Biotechnology, Inc. (Rockford, IL). Custom oligonucleotides were synthesized by Integrated DNA Technologies, Inc. Proteins were visualized with Precast SDS-PAGE gels (BioRad). Detoxi-Gel Endotoxin Removal Gel and endotoxin-free columns were purchased from Thermo Scientific (Rockford, IL). PYROGENT™ Gel Clot LAL Single Test Vials (0.06 EU/ml sensitivity) was from Lonza.

4.3.2. Bioinformatics studies

We implemented custom-made scripts for analysis of protein sequences using MATLAB R2008a. The full names and sequence identification numbers (gi) for the NCBI (National Center for Biotechnology Information) record of each of the proteins studied herein are reported in **Table 3**. The amino acid sequence of these proteins was retrieved from the NCBI and saved as .TXT files for analysis.

Table 3. Full names and accession numbers of all proteins analyzed in this manuscript.

Protein	Species	Accession Number (gi)
---------	---------	-----------------------

Protein	Species	Accession Number (gi)
Elastin	Homo sapiens	182021
Elastin	Bos taurus	28461173
Elastin	Mus musculus	31542606
Elastin	Rattus norvegicus	55715827
Elastin	Macaca mulatta	13182892
Elastin a	Danio rerio (Zebrafish)	121583675
Alpha-1 Collagen Type I	Homo sapiens	553615
Dragline silk fibroin (Spidroin 2)	Nephila clavipes	159714
Flagelliform silk protein	Nephila clavipes	2833649
Resilin Isoform B	Drosophila melanogaster	45552671
Resilin Isoform A	Drosophila melanogaster	7302880
High molecular weight gluten subunit	Elymus alashanicus	84181091
Fibroin-3 (ADF-3)	Araneus Diadematus	1263287
Protein PRQFV-amide	Aplysia californica	74842069
FMRFamide-related neuropeptides (FARX)	Lymnaea stagnalis	1169643
Transcription elongation factor SPT5	Schizosaccharomyces pombe	74581925

Protein	Species	Accession Number (gi)
Ribosome-binding protein 1 (RRBP1)	Canis lupus familiaris	23822071
Actin cytoskeleton-regulatory complex protein PAN1	Schizosaccharomyces pombe 972h	1723244
Retinitis pigmentosa 1-like 1 protein (RP1L1)	Homo sapiens	317373543
Epsin-1	Schizosaccharomyces pombe	62900127
Tropoelastin 1	Xenopus Tropicalis (XT)	118403463
Tropoelastin 2	Xenopus Tropicalis (XT)	218664459
Tropoelastin 2a	Xenopus Laevis (XL)	296040424
Tropoelastin 2b	Xenopus Laevis (XL)	296040422

Digitalization of protein sequences

We implemented a simple strategy to visualize the repetitive character of motifs across the primary amino acid sequence of proteins. In the case of Elastin, for instance, we were interested in studying the distribution of particular motifs that had been previously identified, such as the canonical VPGXG motif (where X is any residue but Pro) and its variants LPGXG and IPGXG. To do this the algorithm searches for the target motif along the amino acid sequence of the protein and creates a digital vector with either 1 or 0 at every residue position, wherein 1 corresponds to each residue belonging to the target motif and 0 corresponds to any other residue. The resulting vector, when

plotted as a function of residue number (i.e., position in the vector), resembles a square signal where the 'on' state (i.e., Y-axis is equal to 1) depicts the location of the target motif as it recurs along the protein.

We used a similar digitalization strategy to visualize the localization and variety of P-X_n-G motifs across proteins in **Table 3**. In this case, residues occurring in a P-X_n-G motif —other than Gly— are assigned a value of 1 in the digital vector, while all other residues are set to 0. Because this is a general motif with a variable number of X residues (where X is any amino acid but Pro and Gly), the length (measured in number of residues) of the 'on' state provides visual feedback on the specific type of P-X_n-G motif (i.e., the n value). For instance, if n=0 the 'on' state is a sharp peak, whereas for n>0, the signal is increasingly squared.

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

We quantified the abundance of P-X_n-G motifs (where *n* varied from 0 to 4, and X is any residue but Pro and Gly) along the amino acid sequence of all proteins in Table S1. For proteins with high Gly-content (> 20%), we also quantified the abundance of Gly in residue positions surrounding each P-X_n-G motif, namely two residues N-terminal and two residues C-terminal to each motif. 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.

The code below (MATLAB Script 1) 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.

```
%***** 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

        end

    end

end

```

```

        end

        if strcmp('G',rndProt(i,j))

            t=1;

            gends(n)=j;

            n=n+1;

            end

        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
%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*****

```

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. 1.

$$FCRG = \frac{(GlyatPmX - Gc)}{Gc} \quad (\text{Eq. 1})$$

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.

Distance between P-X_n-G motifs and hydropathy distribution

Additional studies involved the quantification of the average distance between P-X_n-G motifs and the hydropathy distribution of residues surrounding Pro-Gly dipeptides in the set of Pro- and Gly-rich proteins. To calculate the average distance

between successive P-X_n-G motifs we only considered those with n=0 and n=1, since these proteins are almost devoid of motifs where n>1. Using a similar strategy to the one shown in Script 1 to store the residue positions of all Pro (vector 'pstarts') and Gly (vector 'gends') residues participating in P-X_n-G motifs, we calculated the average distance between the Gly residue in motif *i-1* and the Pro residue in motif *i*, subjected to the condition that *n* be either 0 or 1 for the two motifs (i.e., *i-1* and *i*) under consideration. A simple variant of this algorithm was used to search for P-X_n-A motifs and to calculate the average distance between the corresponding motifs where n=0-1. Additionally, using the residue positions stored in vectors 'pstrats' and 'gends', we retrieved the identity of the residues N-terminal and C-terminal to each Pro-Gly dipeptide, which were then converted to a numerical vector with the corresponding hydropathy value for each amino acid based on the Kyte-Doolittle scale of hydrophobicity [209].

Identification of Pro-rich (only) proteins

The methods described above were applied to both Pro- and Gly-rich proteins and Pro-rich (only) proteins, although the distribution of Gly residues in the vicinity of P-X_n-G motifs was not studied for the latter group of proteins due to their low Gly content. However, our initial studies only considered Pro- and Gly-rich proteins as they were the most closely related to tropoelastin — the prototypical model for an LCST protein. Once it became evident that Pro- and Gly-rich proteins were statistically biased

presence of the corresponding forward and reverse primers as listed in Table 4. The degree of extension was assessed by DNA electrophoresis and when necessary (i.e., when visible products were ≤ 400 bp in length) constructs were further extended with 10 cycles of OE-PCR in the absence of primers [49]. These OERCA products were blunt ligated (using T4 DNA ligase from Fermentas) into a modified pET25 vector that incorporates an N-terminal MSKGP sequence and a C-terminal HHHHHHYG tag to all peptide polymers synthesized by OERCA. The ligation mixture was then transformed into BL21 cells. The resulting colonies were screened by directional colony PCR (using T7R and the forward primer specific to each polymer) and positive clones were grown overnight, and their plasmids were purified using a plasmid miniprep kit (Qiagen), which were then further screened by restriction endonuclease digestion and DNA sequencing (Eton Bioscience Inc., NC, USA). This resulted in a library of genes encoding for peptide polymers of varying length for each repeat unit in Table 4. Only clones with fully confirmed DNA sequences were retained for further study.

Table 4 only reports the length of the forward and reverse primers for brevity. All primers were designed to flank the two ends of the DNA oligomer. Hence, the DNA sequence of these primers can be easily extracted from the 5'-end (forward primer) and 3'-end (reverse primer) of each oligomer. Note that the DNA sequence of the reverse primer is the reverse complement of the respective sequence in the oligomer.

Table 4. DNA sequence information for the synthesis of genes encoding for peptide polymers spanning P-X_n-G motifs from n=0 to n=4. Both oligomers and primers were 5'-

phosphorylated. DNA oligomers with 4-10 copies of the starting oligomer sequence (as shown) were constructed by OERCA, with the exception of polymers of the peptide LGAPVG (see Table 5).

Repeat unit	Oligomer sequence (5'-3')	F. primer (bp)	R. primer (bp)
	GTACCTGGGGTAGGTGTGCCGGGCGTCCGTGTGCC		
VPGVG	GGGCGTCCGTGTTCCGGGCGTGGGTGTCCCAGGCG TAGGG	17	16
AVPGVG	GCTGTACCAGGTGTTGGTGCCGTGCCGGGCGTGGG CGCGGTGCCTGGCGTGGGCGCGGTCCCAGGCGTCCG GCGCAGTTCAGGTGTAGGT	16	16
VAPGVG	GTAGCTCCAGGTGTTGGTGTGGCCCCGGGCGTGGG CGTGGCGCCTGGCGTGGGCGTCCGCGCCGGGCGTCCG GCGTTGCACCAGGTGTAGGT	16	17
VPGAVG	GTACCAGGTGCTGTTGGTGTGCCGGGCGCCGTGGG CGTGCCTGGCGCGGTGGGCGTCCCAGGCGCGGTCCG GCGTTCCAGGTGCAGTAGGT	16	17
VPGVAG	GTACCAGGTGTTGCTGGTGTGCCGGGCGTGGCCGG CGTGCCTGGCGTGGCGGGCGTCCCAGGCGTCCGCGG GCGTTCCAGGTGTAGCAGGT	15	17
APGVG	GCTCCAGGTGTTGGTGCCCCGGGCGTGGGCGCGCC TGGCGTGGGCGCGCCGGGCGTCCGCGCACCAGGT	16	16

Repeat unit	Oligomer sequence (5'-3')	F. primer (bp)	R. primer (bp)
	GTAGGT		
	GTTCCAGGTGTTGCTGTGCCGGGCGTGGCCGTGCC		
VPGVA	TGGCGTGGCGGTCCCGGGCGTCGCGGTACCAGGTG	17	17
	TAGCA		
	ACCGTACCAGGTGTTGGTACAGTGCCGGGCGTGGG		
TVPGVG	CACTGTGCCTGGCGTGGGCACAGTCCCGGGCGTCG	16	18
	GCACGGTTCAGGTGTAGGT		
	ACCGTACCAGGTGCTGGTACAGTGCCGGGCGCCGG		
TVPGAG	CACTGTGCCTGGCGCGGGCACAGTCCCGGGCGCG	15	15
	GGCACGGTTCAGGTGCAGGT		
	GTACACCCTGGTGTGGAGTGCATCCGGGCGTGGG		
VHPGVG	CGTGCACCCGGGCGTGGGCGTCCATCCGGGCGTCG	19	19
	GCGTTCATCCAGGTGTAGGT		
	GTACCAGCTGGTGTGGTGTGCCGGCCGGCGTGGG		
VPAGVG	CGTGCCTGCGGGCGTGGGCGTCCCGGCGGGCGTCG	16	17
	GCGTTCAGCAGGTGTAGGT		
	GTAGGGCCTGTAGGTGTGGGCCCGGTCCGGTGTGGG		
VGPVG	CCCGGTCGGTGTGGCCCCGGTGGGTGTCGGCCCCAG	15	16
	TAGGG		

Repeat unit	Oligomer sequence (5'-3')	F. primer (bp)	R. primer (bp)
	GTTGCTCCAGTTGGTGTGGCCCCGGTGGGCGTGGC		
VAPVG	GCCTGTGGGCGTCGCGCCGGTCGGCGTAGCACCAG TAGGT	16	16
	GTACGTCCTGTTGGAGTTCGCCCCGGTAGGTGTCCG		
VRPVG	CCCGGTGGGCGTCCGCCCGGTGGGTGTTCGTCCAG TAGGT	18	18
LGAPVG	See Table S3.	N.A.	N.A.
	GTAGGTGCTCCAGTTGGTGTGGGCGCCCCGGTGGG		
VGAPVG	CGTGGGCGCGCCTGTGGGCGTCGGCGCGCCGGTCG GCGTTGGTGCACCAGTAGGT	16	17
	GTACCAACCGGTGTTGGTGTGCCGACAGGCGTGGG		
VPTGVG	CGTGCCTACTGGCGTGGGCGTCCCGACAGGCGTCG GCGTTCCAACGGGTGTAGGT	19	17
	GTTCCAGCTGTTGGTGTGCCGGCCGTGGGCGTGCC		
VPAVG	TGCGGTGGGCGTCCCGGCGGTCCGGCGTACCAGCAG TAGGT	16	16
	GTTCCACACGTTGGTGTGCCGCATGTGGGCGTGCC		
VPHVG	TCACGTGGGCGTCCCGCATGTCGGCGTACCACATG TAGGC	19	18

Repeat unit	Oligomer sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
	GTTGGCCCAGCTGTTGGTGTGGGCCCGCCGTGGG		
VGPAVG	CGTGGGCCCTGCGGTGGGCGTCGGCCCGGCGGTTCG GCGTAGGTCCAGCAGTAGGG	18	18
	GTTACCCCAGCTGTAGGTGTGACACCGGCCGTGGG		
VTPAVG	CGTGACTCCTGCGGTGGGCGTCACACCGGCGGTTCG GCGTAAACGCCAGCAGTTGGG	19	18
	ACCCAGTTGCTGTTGGTACACCGGTGGCCGTGGG		
TPVAVG	CACTCCTGTGGCGGTGGGCACACCGGTTCGCGGTTCG GCACGCCAGTAGCAGTAGGT	16	18
	GTACCTTCGACCGACTATGGGGTAGGCGTGCCGAG		
VPSTDY	CACAGACTACGGCGTCGGTGTGCCGAGCACAGAC	17	17
GVG	TACGGCGTCGGTGTACCATCTACGGATTACGGTGT TGGG		
	GTACCTTCTGCGCTGTATGGGGTAGGTGTGCCGTCC		
VPSALY	GCCCTGTACGGCGTCGGTGTCCCATCGGCACTGTA	18	16
GVG	CGGCGTAGGG		

Recursive Directional Ligation by Plasmid reconstruction (PRe-RDL)

For the synthesis of the genes encoding for polymers of TPVAVG and VAPVG for the study presented in Figure 8 and for the synthesis of the diblock copolymers

studied in Figure 10, as well as polymers of LGAPVG, we used the plasmid reconstruction variant (Pre-RDL) of the recursive directional ligation (RDL) method as it offers a remarkably high cloning efficiency at each step of gene assembly [210]. The oligomers used for the synthesis of these genes are reported in **Table 5**. The overall methodology was as originally described by McDaniel and collaborators [210]. Briefly, the oligomers were annealed to form double stranded DNA cassettes with CC and GG overhangs that enable the concatemerization of the cassettes and their eventual ligation (using T4 DNA Ligase from NEB) into a modified pET24 (JMD3) vector that incorporates a short trailer sequence encoding for the GWP peptide. This led to the construction of a small library of plasmids with 2-4 copies of the original cassettes, which were then assembled through multiple cycles of Pre-RDL to the desired length. The final genes were subjected to an additional cycle of Pre-RDL to incorporate a short leader sequence encoding for the MSKGP peptide. The multiple cycles of gene assembly were carried out using EB5 α cells, whereas the final, full-length gene in the corresponding plasmid was transformed into BL21 cells for expression.

Table 5. DNA sequence information for the synthesis of genes encoding peptide polymers via Pre-RDL. Oligomers were 5'-phosphorylated.

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
LGAPVG	GCTGGGTGCTCCAGTTGGTCTGGG	TACTGGTGCACCCAGGCCGACC
	CGCCCCGGTGGGCCTGGGCGCGC	GGCGCGCCAGGCCACAGGCCG

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
	CTGTGGGCCTGGGCGCGCCGGTCG	CGCCCAGGCCACCGGGGCGCC
	GCCTGGGTGCACCAGTAGG	CAGACCAACTGGAGCACCCAGC CC
	GACCCAGTTGCTGTTGGTACACC	TACTGCTACTGGCGTGCCGACCG
	GGTGGCCGTGGGCACTCCTGTGGC	CGACCGGTGTGCCACCGCCAC
TPVAVG	GGTGGGCACACCGGTCGCGGTCG	AGGAGTGCCACGGCCACCGGT
	GCACGCCAGTAGCAGTAGG	GTACCAACAGCAACTGGGGTCC C
	GGTGCTCCAGTTGGTGTGGCCCC	TACTGGTGCTACGCCGACCGGC
	GGTGGGCGTGGCGCCTGTGGGCGT	GCGACGCCACAGGCGCCACGC
VAPVG	CGCGCCGGTCGGCGTAGCACCAG	CCACCGGGGCCACACCAACTGG
	TAGG	AGCAACCCC

Generation and synthesis of a protein-sized, randomized peptide polymer

We implemented a method (see Script 2 below) to automatically generate a protein-sized polymer with a sequence that conforms to the amino acid patterns of non-fibrillar IDPs. The syntax of the polymer is defined by a repeating $Z^2-Z^1-P-Z^0-G-Z^1$ hexapeptide, wherein unconstrained Z residues are randomly selected to yield an average hydrophathy — according to Urry’s scale of hydrophathy [211]— that corresponds to an LCST of 37 °C. The algorithm considers the following constraints: i) Z^2 is Gly if the

$N-1$ motif lacks Gly at this position (to favor Gly at this position as observed in non-fibrillar IDPs), ii) Z^n is any residue but Pro (to minimize Pro content), iii) Gly is not allowed at Z^{-1} (to conform to the syntax of non-fibrillar IDPs) and Z^0 (to minimize Gly content), and iv) Cys content is $\leq 1\%$ (to avoid undesired disulfide bond formation). To generate this randomized polymer, the algorithm first creates a vector of the desired length (defined by the number of repeats of the hexapeptide) and assigns random temperature values to all Z residues according to a normal distribution around the target hydrophathy (equivalent to a target LCST), while assigning the LCST values associated with either Pro or Gly at the two non-random positions in the hexapeptide. The assignment of random LCST values at Z^{-2} relies upon the presence of Gly at this position in the preceding (Z_{N-1}^{-2}) motif. The algorithm has a 67% chance of assigning a random temperature to Z_{N-2} if Gly is present, but otherwise (with and without Gly at Z_{N-1}^{-2}) assigns the LCST value associated with Gly. This vector of randomized LCST values needs to be approximated to meaningful LCST values that are represented among the natural set of amino acids — according to Urry’s scale of hydrophathy. The algorithm calculates the difference — termed “deltas” — between the temperature at each residue position and the theoretical LCST of each amino acid (note that some amino acids share the same theoretical LCST). These vectors of deltas are assembled into a matrix where each row corresponds to a unique or a few (up to 3) amino acids. By finding the position (i.e., row number) of the minimum delta per residue position (i.e., per column), the

algorithm identifies the amino acid that best approximates the randomly-generated LCST value. At this stage the method applies constraints (ii)-(iv) and removes undesired residues by assigning a large, arbitrary delta at the respective position and by recalculating the new minimum to find the best second choice to approximate the random LCST value. The final step is a direct translation of the identified amino acid LCST values to the corresponding amino acid code.

```
%***** Start of MATLAB Script 2*****
```

```
%Definition of main variables
```

```
%rndEIP: vector with a randomly generated sequence of LCST values at each Z position in (ZZPZGZ)-N
```

```
%Delta: Auxiliary matrix that indicates the amino acid with the LCST that is closest to the random %temperature value in each position of rndEIP.
```

```
%SeqrndEIP: vector with rndEIP sequence translated to amino acid code.
```

```
clear all
```

```
n=1;
```

```
N=40; %Number of repeats of the hexapeptide ZZPZGZ
```

```
HiT=37; % Target hydropathy in °C
```

```

SD=50; % Standard deviation in °C

rndEIP(n)=55;

rndZs=normrnd(HiT,SD,N,4); %Creates a matrix of random temperature values

% Start of code A to assign temperature values to residues in the repeating ZZPZGZ
motif.

for i=1:N

    % Start of code A1 to bias the occurrence of Gly at Pm2/Gp2

    if i~=1

        if rndEIP(n-6)==55

            rndG=unidrnd(3); % 1 and 2 are residues different than G; 3 is G

            if rndG==3

                rndEIP(n)=55;

            else

                rndEIP(n)=rndZs(i,1);

            end

        else

            rndEIP(n)=55;

        end

    end

end

```

```

end

%End of code A1

rndEIP(n+1)=rndZs(i,2); % Potential Gly residues in this position are removed
afterwards

rndEIP(n+2)=-8; % Assigns Pro residue that are part of the P-Xn-G motif
rndEIP(n+3)=rndZs(i,3);

rndEIP(n+4)=55; % Assigns Gly residues that are part of the P-Xn-G motif
rndEIP(n+5)=rndZs(i,4);

n=n+6;

end

%End of code A

%Start of code B to transform a random temperature value to its closest
%integer that matches the temperature (LCST) of an amino acid according to
%Urry's scale of hydrophobicity

DeltaG=abs(rndEIP-(55));

DeltaA=abs(rndEIP-(45));

DeltaV=abs(rndEIP-(24));

DeltaL=abs(rndEIP-(5));

```

```

DeltaI=abs(rndEIP-(10));

DeltaM=abs(rndEIP-(20));

DeltaF=abs(rndEIP-(-30));

DeltaY=abs(rndEIP-(-55));

DeltaW=abs(rndEIP-(-90));

DeltaP=abs(rndEIP-(-8));

DeltaTNS=abs(rndEIP-(50)); % Threonine, Asparragine and Serine

DeltaC=abs(rndEIP-(30));

DeltaRQ=abs(rndEIP-(60)); %Arginine and Glutamine

DeltaK=abs(rndEIP-(120));

DeltaH=abs(rndEIP-(-10));

DeltaD=abs(rndEIP-(150));

DeltaE=abs(rndEIP-(250));

Delta=[DeltaG;DeltaA;DeltaV;DeltaL;DeltaI;DeltaM;DeltaF;DeltaY;DeltaW;Delta
P;DeltaTNS;DeltaC;DeltaRQ;DeltaK;DeltaH;DeltaD;DeltaE];

[deltamin,position]=min(Delta);

%Start of code B1 to remove excess of Cys residues

cysteine=find(position==12);

```

```
if length(cysteine)>length(rndEIP)*0.01 % if more than a 1% cysteines.
```

```
    dCys=length(cysteine)-uint8(length(rndEIP)*0.01);
```

```
    rndC=unidrnd(length(cysteine),1,2*dCys);
```

```
    DeltaC(cysteine(rndC))=100;
```

```
end
```

```
Delta(12,:)=DeltaC;
```

```
%End of code B1
```

```
%Start of code B2 to remove Gly residues at Pm1
```

```
n=1;
```

```
[deltamin,position]=min(Delta);
```

```
for i=1:N
```

```
    if position(n+1)==1
```

```
        DeltaG(n+1)=100;
```

```
    end
```

```
    n=n+6;
```

```
end
```

```
Delta(1,:)=DeltaG;
```

```
%End of code B2
```

%Start of code B3 to avoid ZZPGGZ motifs (to maximize sequence diversity)

n=1;

[deltamin,position]=min(Delta);

for i=1:N

 if position(n+3)==1

 Delta(1,n+3)=100;

 end

n=n+6;

end

%End of code B3

*%Start of code B4 to remove Pro in Z residues while making sure not to introduce new,
undesired Gly's*

n=1;

[deltamin,position]=min(Delta);

for i=1:N

 proline=find([position(n),position(n+1),1,position(n+3),1,position(n+5)]==10);

 if ~isempty(proline)

 Delta(10,n+proline-1)=100;

 badP=n+proline-1;

```

[newdelta,newposition]=min(Delta(:,badP));

newG=find(newposition==1);

if ~isempty(newG)

    Delta(1,badP(newposition))=100;

end

end

n=n+6;

end

%End of code B4


%Start of code B5 to translate the sequence of LCST values to amino acid code

[deltamin,position]=min(Delta);

for i=1:length(rndEIP)

switch position(i)

    case 1

        SeqrndEIP(i)='G';

    case 2

        SeqrndEIP(i)='A';

    case 3

        SeqrndEIP(i)='V';

```

```
case 4

    SeqrndEIP(i)='L';

case 5

    SeqrndEIP(i)='I';

case 6

    SeqrndEIP(i)='M';

case 7

    SeqrndEIP(i)='F';

case 8

    SeqrndEIP(i)='Y';

case 9

    SeqrndEIP(i)='W';

case 10

    SeqrndEIP(i)='P';

case 11

    rndTNS=unidrnd(3); % 1 is T, 2 is N, 3 is S.

    if rndTNS==1 SeqrndEIP(i)='T';end

    if rndTNS==2 SeqrndEIP(i)='N';end

    if rndTNS==3 SeqrndEIP(i)='S';end

case 12
```

```

    SeqrndEIP(i)='C';

case 13

    rndRQ=unidrnd(2); % 1 is R, 2 is Q

    if rndRQ==1 SeqrndEIP(i)='R';end

    if rndRQ==2 SeqrndEIP(i)='Q';end

case 14

    SeqrndEIP(i)='K';

case 15

    SeqrndEIP(i)='H';

case 16

    SeqrndEIP(i)='D';

case 17

    SeqrndEIP(i)='E';

end

end

%End of code B5

%End of code B

%***** End of MATLAB Script 2*****

```

The amino acid sequence of the protein-sized polymer generated using Script 2 was back-translated to its DNA equivalent using SeqBuilder (DNASTAR, Inc). Codon usage was optimized for expression in E. Coli using GeneOptimizer® technology developed by Mr. Gene (now part of Life Technologies Corporation). The DNA sequence (Table S4) was modified at the 5'-end to incorporate a BseRI restriction site (5'-GAGGAGTACATATG) and at the 3'-end to incorporate an AcuI restriction site (5'-CTGAAGATCATTATCAGTAGCC). This gene was purchased from Mr. Gene (GeneArt®, Life Science Technologies). The BseRI and AcuI restriction sites were used for subcloning into the modified pET24 vector (JMD3) described in section 2.2 for pre-RDL. This vector was further restricted with NdeI and BseRI to incorporate a leader sequence (as described in section 2.2) encoding for the peptide MSKGP. The final vector was transformed into BL21 cells for expression.

Table 6. Codon-optimized gene (5'-3') encoding for the protein-sized polymer generated by Script 2. The 5' and 3' untranslated regions that were incorporated for subcloning are not shown.

```

GGTGTTCCGGCAGGTCATCGTTATCCAATCGGGGGAGGCCAACCTCATGGTAAA
GGATGCCCAGACGGCGTTTTTCGTCCTGTGGCCTGGGAGCACCTTATGGACATGGAGCC
CCTAATGGTATGCACCGCCCGCTGGGTATTGGCAAACCTCGTGGCCACATGTATCCGAAA
GGGCAAGGACAGCCGATGGGTCACCTGGTTCAGATGGCGTAGGATTCCCTCGTGGTCGT
AAAAAACCGGTTGGAGTGGGAAAACCAATTGGTAATGGTCACCCTATTGGTGCCCGTAC
ACCACTGGGATATGGAATGCCTGATGGAGTCGGAATGCCGATGGGTCTGTTTCTGCCTAA
TGGACACGGCGCTCCACATGGACAAGGTTATCCGGCGGGTAAACTGATTCCCTAAAGGCA

```

AAGGACATCCGTTCCGGTAAAGGTCGTCCGCTGGGAGCTGGTCGCCCTACTGGGTTTAAAA
TGCCTAAAGGCCTGGGAAAACCGATGGGCGTTGGTCAACCTCAAGGCCATTTTGTGCCGT
TTGACTGGGTCAGCCTACTGGACAGGGAGCACCTCGTGGGGGTTCTCAGCCAGCAGGA
CTGGGCCACCCGCTGGGCGCTGGAGCACCAGCAGGTCGTTGTCATCCGTATGGAATGGGT
GTGCCTCGTGGACTGGCAATGCCTCGTGGTCATGGTCAACCTCGTGGCGTGGGTTATCCG
AAAGGTCAT

4.3.4. Expression and purification of peptide polymers

Starter cultures (5 mL starter culture per 1 L of *E. coli* cells) of Terrific broth (TB) media (BioExpress) supplemented with 100 µg/mL ampicillin (for all peptide polymers synthesized by OERCA) or 45 µg/mL Kanamycin (for peptide polymers prepared by PRe-RDL) 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 to a final concentration of 1 mM. Cells were harvested 24 h after inoculation, and purified by inverse transition cycling (ITC) – with minor modifications – as described elsewhere [212]. Polymers of VPSALYGVG, VGPVG, VGPAVG, and VRPVG 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 GnCl₂. These solutions were centrifuged at 14000 rpm for 30 min at 4 °C and the supernatants were extensively dialyzed against water at 4 °C. These peptide polymers were then purified through regular ITC in PBS, with the exception of polymers of VPSALYGVG, which were purified in PBS with 8 M Urea.

4.3.5. Characterization of the phase transition behavior and secondary structure of peptide polymers

To characterize the phase behavior of the synthesized peptide polymers, the optical density of peptide polymer 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, on a Cary 300 UV-visible spectrophotometer equipped with a multicell thermoelectric temperature controller (Varian Instruments, Walnut Creek, CA).

The secondary structure displayed by these peptide polymers 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. Purified polymers were 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 [213].

The self-assembly behavior of diblock copolymers was studied in PBS at 25 μM by turbidimetry and dynamic light scattering (DynaPro, Wyatt Technologies, CA) measurements as a function of temperature at a heating and cooling rate of 1 $^{\circ}\text{C}/\text{min}$.

5. Chapter 4: Syntax of UCST peptide polymers

Overview: “Smart” polymers that phase separate from solution in response to a decrease — as opposed to an increase — in temperature, below a so-called upper critical solution temperature (UCST), are needed to integrate novel thermally responsive material systems for different applications. The few synthetic polymers that display this behavior, however, are far from ideal, and the only existing peptidic analogue cannot be genetically encoded as it is synthesized from non-proteinogenic amino acids [14]. Here, we demonstrate that the sequence diversity of peptide syntactomers — as described in chapter 3 — can be exploited to uncover the syntax of UCST peptide polymers and to reprogram peptide polymers that exhibit LCST behavior into UCST-exhibiting polymers. We show that the syntax of UCST peptide polymers is more permissive than the syntax of LCST peptide polymers and that it ranges from hydrogen bonding polymers composed of repeats of zwitterionic peptides to those with repeats of charged peptides. The resulting UCST syntactomers constitute the first family of genetically-encoded peptide polymers that exhibits tunable UCST under physiologically relevant conditions. Finally, we show the ability of these materials to control the self-assembly of diblock copolymers into nanostructures that undergo hyperthermia-mediated disassembly. This work is a major step forward in the rational design of highly advanced “smart” peptide polymers for medicine and biotechnology.

5.1. Key residue interactions govern the phase behavior of peptide polymers

First we show that the introduction of key residue interactions into a syntactomer can determine whether phase behavior ensues on heating or cooling (Figure 23). Because of the intriguing LCST and UCST behavior displayed by resilin, we reasoned that the syntax of this protein could inform us about key residue interactions that lead to UCST behavior in peptide polymers. We then synthesized a syntactomer by repeating a fragment of the consensus repeat sequence that most differed from the composition of LCST syntactomers (Figure 5; chapter 3), and in which charge-charge interactions, hydrogen bonding, salt bridges and hydrophobic interactions are all possible. This resilin-derived syntactomer displays a sharp UCST behavior under physiologically relevant conditions (Figure 23B-i). Interestingly, the polymerization of a similar fragment from resilin, but with a reduced ability to engage in hydrogen bonding and that incorporates a few additional hydrophobic residues (here, two additional valine residues), leads to the opposite behavior, showing a LCST phase transition (Figure 23B-ii). This suggests that a fine balance between the polar and the nonpolar interactions experienced by residues in the repeating motif determines whether a syntactomer goes down the LCST or UCST path.

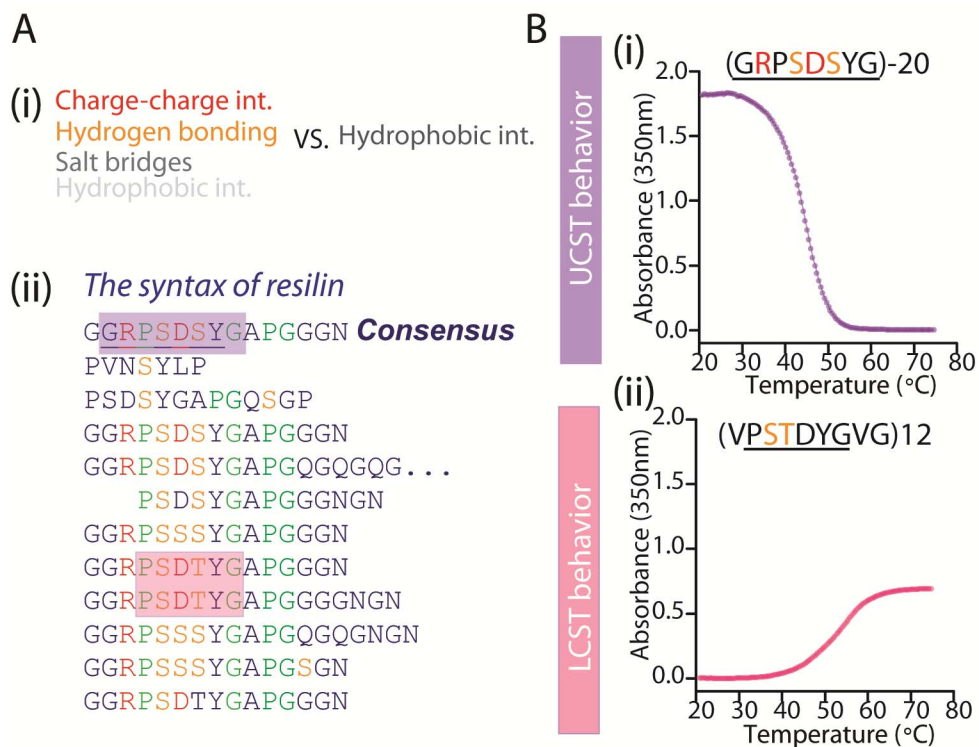


Figure 23. Resilin is built from amino acid repeats that may be tuned to exhibit UCST or LCST phase behavior. (Ai) We postulate that different residue interactions are at play for the design of UCST (charge-charge interactions, hydrogen bonding, salt bridges and some hydrophobic interactions) and LCST (mostly hydrophobic interactions) peptide polymers. (Aii) The primary structure of resilin (exon 1), which displays both UCST and LCST behavior, is composed of amino acid repeats. Arginine, aspartic acid and serine residues are highly conserved across the repeating units. This sequence alignment was adapted from [184].(B) A polymer composed of a fragment of the consensus repeat unit incorporates residue interactions that lead to UCST phase behavior upon polymerization (Bi), whereas a very similar repeat unit that excludes some of these interactions and dopes in additional hydrophobic residues readily overturns the UCST behavior of the original peptide polymer to yield a peptide polymer with LCST phase behavior (Bii). Turbidimetry data was measured in PBS at 50 μ M.

Hence, we hypothesized that the abundance of positively and negatively charged residues in this protein, predominantly arginine and aspartic acid, as well as residues with hydrogen bonding ability (e.g., serine, threonine and arginine), distributed along a scaffold rich in P-X_n-G motifs, is responsible for the UCST phase behavior of resilin. To

test this, we synthesized libraries of syntactomers spanning all P-X_n-G units previously identified in chapter 3 and that incorporate these types of residue interactions into pentapeptide, hexapeptide, heptapeptide and octapeptide motifs (Table 7). Temperature-dependent turbidimetry demonstrated that these peptide polymers display a sharp UCST phase transition in PBS (Figure 24A-C). Interestingly, the syntax of UCST syntactomers tolerates the absence of structure-breaking Pro-X_n-Gly motifs (Figure 24D). Not only do these Pro-devoid syntactomers exhibit UCST behavior, but they are also highly overexpressed in *E. coli* as are all other UCST syntactomers that do contain Pro (Figure 24E) —we typically achieve purified yields between 200-300 mg per liter of culture without any optimization of the expression conditions. This is in stark contrast to LCST syntactomers wherein Pro appears to be a sine qua non element. This independence is likely a result of the contributions of highly polar and charged residues in these motifs to the unstructured character of the polymer, as low complexity domains enriched in such residues populate the space of intrinsically disordered proteins [214, 215]. By performing CD spectroscopy on a subset of tyrosine-devoid polymers, we confirmed the unstructured character of UCST syntactomers by showing that their spectra are reminiscent of the spectrum observed for tropoelastin (Figure 24F). The expected unstructured character of UCST syntactomers that contain aromatic residues (e.g., F and Y), however, was not revealed by CD spectroscopy (data not shown) because

of large contributions from these residues that distort the spectra in the wavelength range of interest [216].

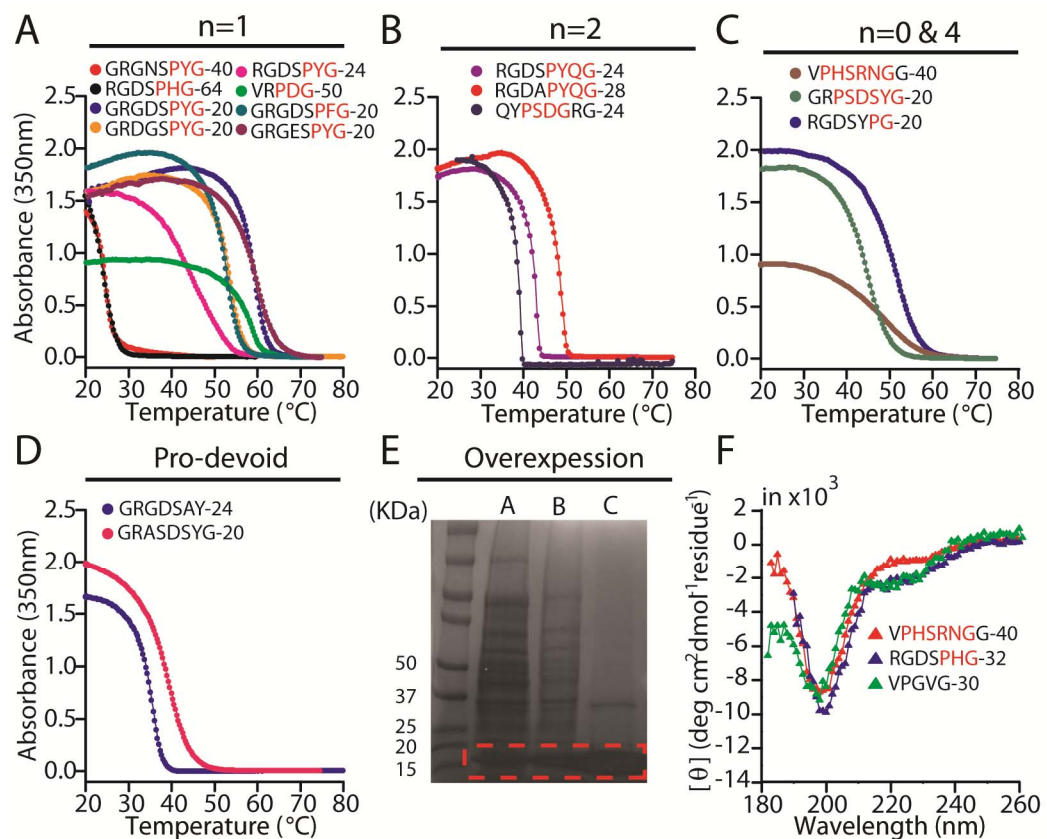


Figure 24. Syntactomers with repeating P-X_n-G motifs (n=0-4) can be designed to exhibit UCST phase behavior under physiologically relevant conditions. (A-C) Turbidimetry data reveals the UCST behavior of this novel family of genetically encoded peptide polymers. (D) The unique amino acid composition of these polymers tolerates the absence of Pro. Data in (A-D) was acquired in PBS (pH 7.4) at 50 μM, except for polymers of VPHSRNNGG (+2 M NaCl), RGDSPYQ (+1 M Urea) and VRPDG (PBS pH 2.0 supplemented with 1.5 M NaCl). (E) Representative example of the overexpression of these polymers (here, GRGDSPYQ-20) upon IPTG induction in *E. coli* and the subsequent purification from the insoluble fraction by UCST transition cycling. Lane A is the cell lysate; lane B is the insoluble fraction and lane C is the result of a first cycle of purification. (F) CD spectroscopy in water at 5 μM for a subset of these polymers revealed their unstructured character, as their spectra is reminiscent of the spectrum of the canonical VPGXG motif from the LCST-exhibiting tropoelastin.

The next relevant question is whether the UCST behavior exhibited by these syntactomers is reversible and tunable by parameters that are readily controlled in our genetically-encoded polymers. First, repeating cycles of cooling and heating demonstrate the reversible UCST phase behavior exhibited by these syntactomers over a range of compositions and molecular weights (Figure 25A). Second, we provide evidence that the UCST of these polymers is readily tuned by the number of repeat units (i.e., molecular weight) (Figure 25B) and by modifying the hydrophobicity of the overall repeat unit (Figure 25C). For instance, replacing one Gly residue in GRGDSPYG by a more hydrophilic residue, as in GRGDSPYQ, allowed for the tuning of the UCST from ~60 °C to a physiologically relevant temperature of ~42 °C (at 50 μM in PBS). The effect of molecular weight is particularly dramatic for low molecular weight polymers but as the number of repeat units increases the UCST continues to increase only marginally (Figure 25B). This is reminiscent —yet opposite in that the phase transition temperature increases— of the effect of increasing molecular weight on the LCST of syntactomers and ELPs, and even the concentration dependence of this UCST phase behavior shares similarities with LCST syntactomers [49, 130]. The level of control on the UCST of these genetically-encoded polymers is in stark contrast with the difficulties encountered when tuning the UCST of synthetic polymers that lack a defined syntax, and in which chain length and copolymerization with hydrophobic monomers have also been shown to modulate the UCST [13, 61].

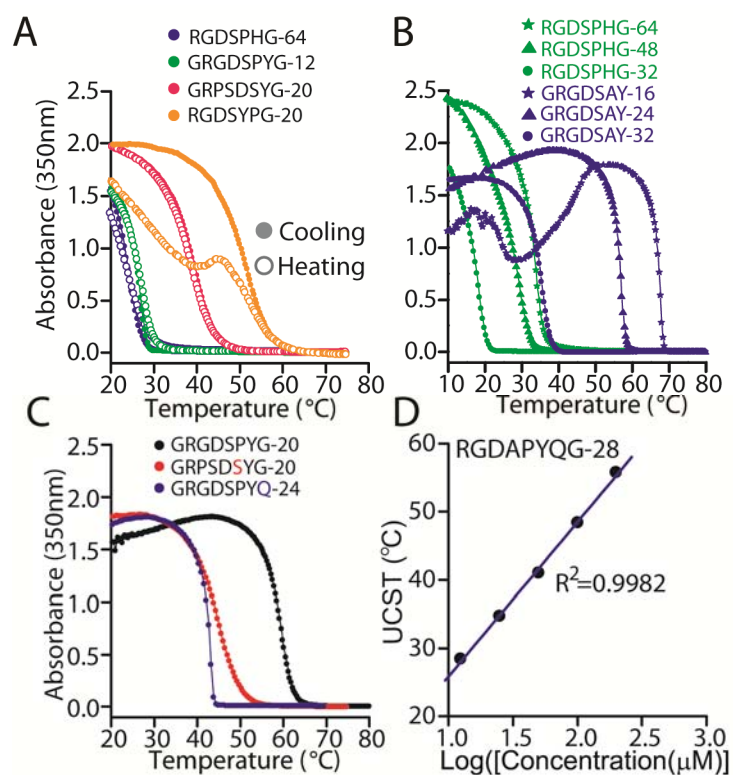


Figure 25. The UCST behavior of syntactomers is readily tuned by molecular weight, hydrophobicity and concentration. (A) These polymers exhibit fully reversible phase transition behavior. (B) Molecular weight is a potent modulator of the UCST for relatively short polymers, but the effects become less pronounced as the molecular weight increases. (C) Mutations to the GRGDSPYG motif demonstrate that substitutions for more hydrophilic residues lead to lower UCST values. (D) The UCST is a function of the concentration of the polymer in solution. Turbidimetry data in (A-C) was acquired in PBS at 50 μ M.

A large number of the peptide motifs explored in this work (Table 7 and Figure 24) are zwitterionic as they contain a positively charged residue — arginine or lysine— and a negatively charged residue — aspartic acid or glutamic acid— and are thus essentially neutral. Zwitterionic polymers (e.g., polysulfobetaines) as well as interpenetrating networks of polyelectrolytes were among the first synthetic polymeric

systems that were shown to exhibit UCST phase behavior in water [61, 217]. Because the phase behavior of these polymers relies on charge-charge interactions that are effectively screened by counter ions in physiologically-relevant solutions, their UCST behavior has not received much attention for biomedical applications [13, 217], although their ability to resist non-specific protein absorption is of great biomedical interest [218, 219]. As a result, more recent work on UCST polymers diverged from this approach by focusing on hydrogen bonding polymers that are not affected by the high ionic strength in physiological solutions [13]. The presence of unexpected residual charges in the polymer, as a result of trace contaminants from the chemical synthesis, was recently suggested as a major contributor to the non-responsive behavior (solubility) of several amide-based polymers that otherwise exhibit UCST phase behavior in electrolyte solutions [60, 61]. In line with this, residual amino groups in ureido-modified poly(allylamine) —the first example of a non-vinyl based UCST polymer that displays phase behavior in electrolyte solution—, which are charged at the physiological pH of 7.4, also decrease the UCST of these polymers [14]. In this context, the UCST phase behavior exhibited by our genetically-encoded zwitterionic peptide polymers under physiological relevant conditions is intriguing.

We hence investigated the importance of this zwitterionic character and the effect of uncompensated charges in these syntactomers on the modulation of their UCST phase behavior in electrolyte solution —here, PBS (10 mM phosphate buffer, 140 mM

NaCl and 3 mM KCl at pH 7.4). Zwitterionic syntactomers displayed an unusual sensitivity to molar amounts of NaCl (Figure 26A) that has not been possible to study with existing zwitterionic polymers, since for the relatively large concentration of NaCl already present in PBS these polymers are essentially soluble over the entire range of accessible temperatures [217, 218]. Here, we observed a consistent decrease in the UCST when adding 1 M NaCl, which in some instances also led to a decrease in the effective fraction of aggregated polymer (i.e., lower maximum absorbance), whereas an even higher amount of NaCl (e.g., 2 M NaCl) usually –but not always– led to an increase in the UCST as would be expected if salt simply changed the quality of the solvent (that is, the enthalpic cost of polymer/solvent interactions). The magnitude of this effect, which is in most cases small compared with the consistent and pronounced decrease in the LCST of neutral and positively charged syntactomers (Figure 17), varied significantly for different syntactomers with slightly different amino acid sequence (Figure 26A). These results evidence the complex solution behavior of zwitterionic syntactomers, which is likely the result of subtle changes in the binding affinities of the positively and negatively charged moieties in these peptide motifs toward the corresponding cationic and anionic counter ions [217]. Further investigation of the effect of different salts in pure water may help elucidate these intriguing effects.

Under acidic conditions (PBS pH 2.0), however, zwitterionic syntactomers lose their virtually uncharged character and are positively charged as aspartic or glutamic

acid residues become protonated. The resulting charge-charge repulsion reduces the UCST below the observable range, but the polymers now respond sharply to the addition of 0.5 M NaCl and exhibit UCST values comparable to those measured in PBS at pH 7.4 (Figure 26B). We then asked whether we could tune the UCST of such positively charged syntactomers. To do this, we mutated the GRGDSPYG motif by substituting aspartic acid (D) for its protonated, uncharged equivalent asparagine (N) to eliminate the zwitterionic character of the polymer without having to study the polymers under acidic conditions. By doubling the molecular weight of this cationic polymer with respect to its zwitterionic equivalent, we tuned the UCST of this syntactomer back to a useful UCST at physiologically relevant conditions (Figure 26C). This demonstrates that even charged peptide motifs can lead to peptide polymers with a sharp UCST phase transition without recurring to non-physiological concentrations of salt. However, we note that this mutant polymer also has a tendency to display LCST, which is likely the result of the increase in hydrophobic content, but this will require further investigation.

This evidence suggests that even if electrostatic interactions are weakened in electrolyte solutions and thus unlikely to play a central role in the observed UCST phase behavior, the virtual charge neutrality in these zwitterionic polymers — perhaps mediated by salt bridges — is important to enable this behavior under physiologically relevant conditions. In chapter 3 we demonstrated that syntax governs the LCST phase

behavior of syntactomers by modulating the secondary structure of peptide polymers, particularly upon aggregation. Although the noted Pro-independence of UCST syntactomers already suggests that these polymers may be somewhat recalcitrant to secondary structure formation, syntax may also influence phase behavior by controlling their virtual neutrality. We studied a UCST syntactomer and its sequence-reversed, retro equivalent —by reading the sequence from C- to the N-terminus—, which share the same amino acid composition, and consistently observed a significant shift in the UCST (~ 5 °C) but without changes in the degree of hysteresis or reversibility of this behavior (Figure 26D). Similarly, two syntactomers with R and D either separated by one residue or forming a dipeptide, also showed a similar shift in the UCST (Figure 26E). Whereas syntax played only a modest role in these two test-cases, we observed that there are UCST “words” that can lead to new types of phase behavior. Here we show that a peptide polymer composed of repeats of the heptapeptide RPLGYDS displays both LCST and UCST phase behavior in a very unusual manner, as its LCST is lower than its UCST (Figure 26F). The design of peptide polymers that display both LCST and UCST wherein $LCST > UCST$ is trivial because fusing a LCST and a UCST motif into a single motif for polymerization is expected to exhibit a solubility region at intermediate temperatures between the UCST and LCST of the individual motifs. This has been demonstrated in synthetic polymers [218] and is now clearly seen for resilin —note that the consensus sequence in resilin (Figure 23B) is essentially composed of a UCST motif

(highlighted in purple) and a LCST motif (roughly APGGGN). The opposite, however, that is $LCST < UCST$, would be expected to result in polymers that are insoluble at all temperatures. We believe that this unusual behavior is likely the result of the precise syntax of this polymer that facilitates marked changes in the ionization state of aspartic acid residues upon aggregation, but this intriguing behavior certainly deserves and needs further investigation. Interestingly, the large sequence space that can be explored suggests that syntactomers may become useful to experimentally address ongoing controversies about the behavior and conformation of polyampholytes in solution [217, 220], as their genetically-encoded sequence and intrinsic phase behavior provide a unique system to study the role of hydrophobicity and charge distribution (symmetric, asymmetric, random) along the polymer.

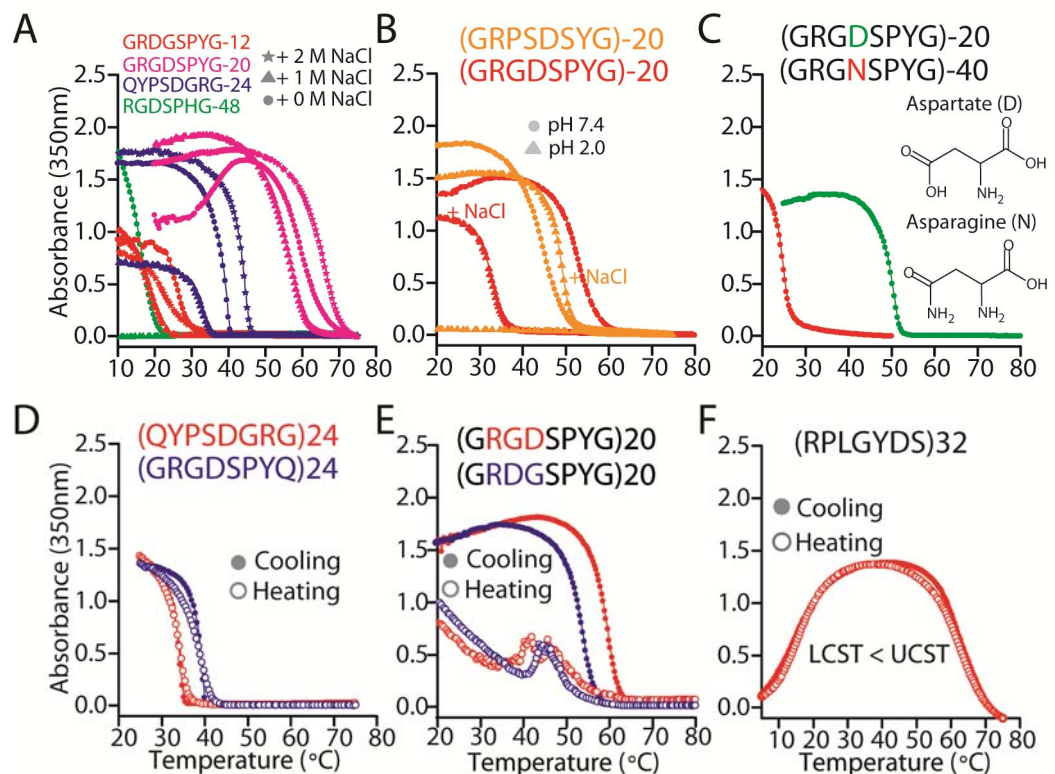


Figure 26. The zwitterionic character of UCST syntactomers plays an important role in the physiological relevance of their behavior. (A) The response of zwitterionic syntactomers in PBS to molar amounts of additional NaCl is sequence-dependent. (B) Under acidic conditions, protonation of aspartic acid residues in the syntactomer results in positively charged polymers that fail to exhibit phase behavior in PBS, but increasing the buffer ionic strength by adding 0.5 M NaCl enables UCST phase behavior. (C) By mutating the aspartic acid residue in GRGDSPLYG into an asparagine, the propensity to display UCST is greatly reduced and UCST phase behavior in PBS (without additional salt) is only observed if the molecular weight of the mutant polymer is significantly increased. (D-E) Syntax controls the UCST but the effects are modest (~4-5 °C) compared with LCST syntactomers (Figure 9), at least in the two examples explored herein: (D) a motif and its corresponding sequence-reversed (retro) motif, and (E) peptide polymers with a conservative change in the spacing of the oppositely charged residues. (F) Syntax, however, may offer the possibility to control unusual behaviors in these peptide polymers as observed here for a syntactomer that exhibits both LCST and UCST, wherein the LCST is lower than the UCST. Turbidimetry data was acquired at 50 μ M in PBS pH 7.4 unless otherwise indicated, except for data in (D) that was measured at 25 μ M and in (F) were PBS was supplemented with 3 M urea.

Besides electrostatic interactions and hydrogen bonding, polypeptide hydrophobicity plays an intriguing role in the phase transition behavior of UCST

syntactomers. The incorporation of highly hydrophobic residues into these otherwise hydrophilic motifs is an important modulator of the UCST. The substitution of tyrosine (Y) in these motifs for histidine (H), for instance, dramatically reduces the UCST (Figure 25B), so hydrophobic interactions strongly modulate the UCST. Because both of these residues can participate in π - π stacking, the role of ring-stacking interactions in facilitating the sharp phase transition event at the UCST cannot be ruled out. However, the UCST phase behavior displayed by polymers of VRPDG (Figure 24A) demonstrates that stacking interactions are not required to mediate UCST behavior in peptide syntactomers.

Hence, a fine balance between the polar residues — whether charged or uncharged — that engage in hydrogen bonding interactions and the nonpolar residues that provide hydrophobic interactions is at play in determining the phase behavior of syntactomers. This is nicely seen by contrasting the UCST phase behavior displayed by polymers of VPHSRNGG (Figure 24C) with the LCST phase behavior displayed by a simple variant of this motif (VPHSRNGL) where the terminal glycine is substituted for leucine (Figure 29; found in *Supplemental Figures*). To further test the hypothesis that the incorporation of key residue interactions into a peptide syntactomer determines their ability to exhibit LCST or UCST phase behavior, we decided to rationally reprogram a LCST syntactomer into a UCST syntactomer (Figure 27). The rationale is that by substituting residues that are responsible for the predominance of hydrophobic

interactions in the original LCST motif by charged residues capable of hydrogen bonding, the “phase balance” in these polymers would be reversed from LCST to UCST. The prototypical VPGXG motif tolerates any amino acid at the X position, suggesting that a single residue capable of hydrogen bonding per pentapeptide repeat is insufficient to achieve UCST behavior. Here, we show that two intermediates of a pentapeptide motif (VAPVG) that exhibits LCST behavior on polymerization (Figure 27B), each with a single mutation, also exhibit LCST behavior, whereas a syntactomer composed of a motif that incorporates the two mutations (Figure 27C) exhibits UCST phase behavior. Interestingly, charge-charge interactions alone do not lead to UCST behavior (see data at pH 7.4), but rather hydrogen bonding mediates the UCST behavior of this syntactomer, although this behavior is only observed at high ionic strength as a result of the cationic character of the polymer under acidic conditions. These results highlight the unique role of aromatic residues (e.g., Y, W, F and H) in modulating the phase behavior of zwitterionic polymers at physiological pH and suggest that syntactomers composed of a sufficiently high density of polar, uncharged residues capable of hydrogen bonding (e.g., N, S, T, Q) may lead to UCST behavior in PBS without the need for highly hydrophobic residues (e.g., Y, F, W, H).

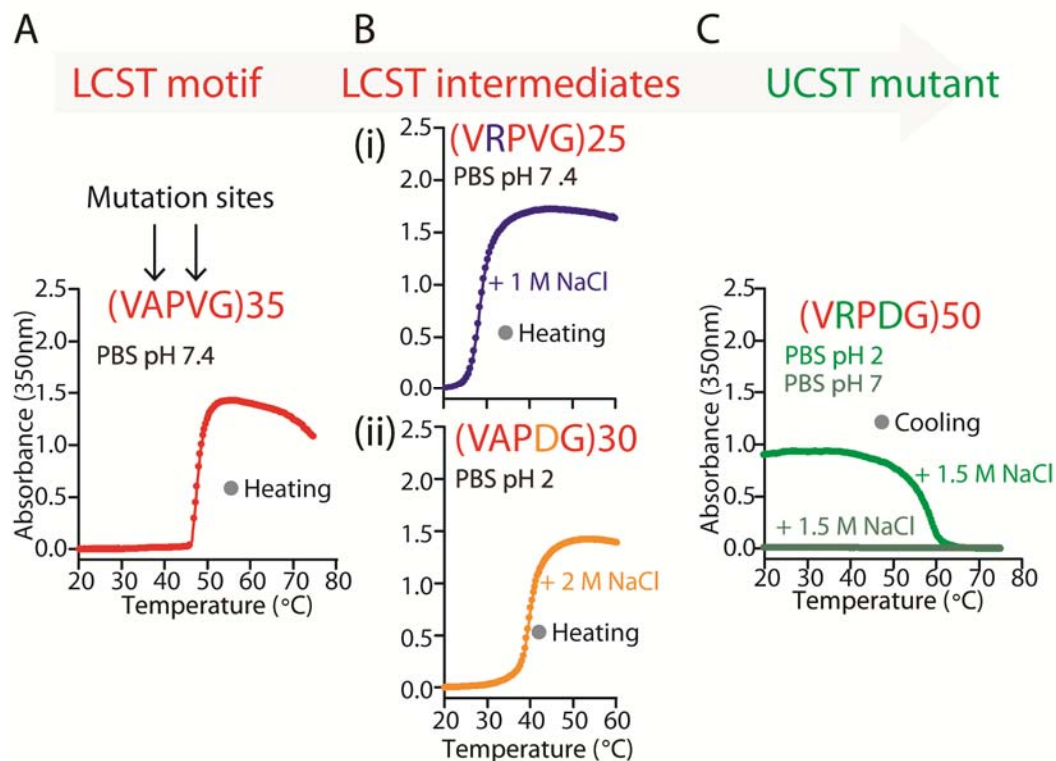


Figure 27. Reprogramming of a LCST syntactomer into a UCST syntactomer. Two residue positions in the LCST-exhibiting motif VAPVG (A) were selected to introduce parallel (Bi and Bii) and serial (C) mutations that might lead to UCST phase behavior. (A) Polymers of VAPVG exhibit LCST behavior in PBS pH 7.4. (Bi) Substituting R by A in that motif results in cationic polymers that exhibit LCST behavior at high ionic strength. (Bii) Substituting D by V in the parent motif results in anionic polymers that exhibit LCST behavior upon protonation and at high ionic strength. (C) Mutating both residues simultaneously leads to UCST phase behavior upon protonation and at high ionic strength.

Nanotechnology has greatly impacted and continues to impact drug delivery to solid tumors [221, 222]. A major determinant of this success is the size-dependent tissue distribution of nanoparticles on intravenous injection, which favors nanoparticle extravasation and accumulation at the leaky and abnormally porous vasculature of the tumor tissue [223]. Size, however, also limits the intratumoral distribution and penetration of nanoparticles, as they primarily accumulate in the tumor perivascular

space while failing to penetrate the highly dense extracellular matrix of the tumor. Moreover, strategies that improve the intratumoral penetration of these drug carriers have been shown to significantly impact their anti-cancer properties [224].

Thermally responsive polymers have been widely explored for the self-assembly of nanoparticles for drug delivery and cancer therapy [1, 3, 225]. Although the usefulness of UCST polymers for cancer drug delivery has been recognized [226], to date, temperature-responsive polymeric drug delivery systems are almost exclusively based on LCST polymers [225]. This is likely the result of the very few polymers that display UCST phase transition behavior under physiologically relevant conditions. Hence, we sought to capitalize on the design of UCST peptide polymers to program the disassembly of peptide nanoparticles using a temperature trigger. Noteworthy, tumor-localized hyperthermia (40-44 °C) is achievable by a number of methods (e.g., highly focused ultrasound, focused radiofrequency and microwaves), is clinically accepted and is well-tolerated by patients [227, 228].

Here, we report for the first time the design of peptide nanoparticles that undergo disassembly on heating and demonstrate that the UCST behavior of a peptide polymer block when fused to a LCST block yields self-assembling nanoparticles with tunable stability (Figure 28A-C). We also briefly explored the effect of changing the length and hydrophobicity of the LCST corona. Reducing the length of the corona from a core: corona block ratio of ~1:1 to ~2:1 still allowed for nanoparticle self-assembly at low

temperatures, but we observed undesired and unusual transient aggregation on disassembly that is not observed in a diblock with a longer LCST corona (Figure 28C) and that is particularly evident at high concentrations (Figure 28D) and low heating rates (Figure 30; found in *Supplemental Figures*). This is suggestive of a transient LCST exhibited by the UCST block at the very high local concentrations of the dissolving nanoaggregates. We also note that the overall amphiphilicity of these UCST-LCST diblock copolymers is still at play in controlling self-assembly. For instance, by replacing the hydrophilic LCST corona (composed of the VPGXG motif with alternating A and G residues in the X position) in the copolymers studied in Figure 28A and Figure 28C with a more hydrophobic LCST corona composed of 30 repeats of the canonical VPGVG motif, we did not observe self-assembly at low temperatures (12-18 °C) and heating resulted in bulk aggregation at ~30 °C (at 25 μ M in PBS), which is surprising considering that the same LCST block alone phase transitions at ~50 °C. This behavior suggests that the ELP corona perceives the UCST block as being hydrophobic —the hydrophobicity of this UCST block is inadequately estimated at ~52 °C according to Urry’s scale of hydrophobicity [207], which is ~16 °C higher (more hydrophilic) than the ELP corona—, which drives its LCST down and eliminates the block amphiphilicity that is key to self-assembly in LCST diblock copolymers. Interestingly, the nature of the residue-level interactions that drive this UCST-mediated assembly is not compatible with the design of diblock copolymers wherein the two blocks are composed of UCST-exhibiting motifs even if their UCST

temperatures are drastically different (Figure 28D). This is a limitation imposed by non-specific hydrogen bonding between the two blocks regardless of their apparent UCSTs.

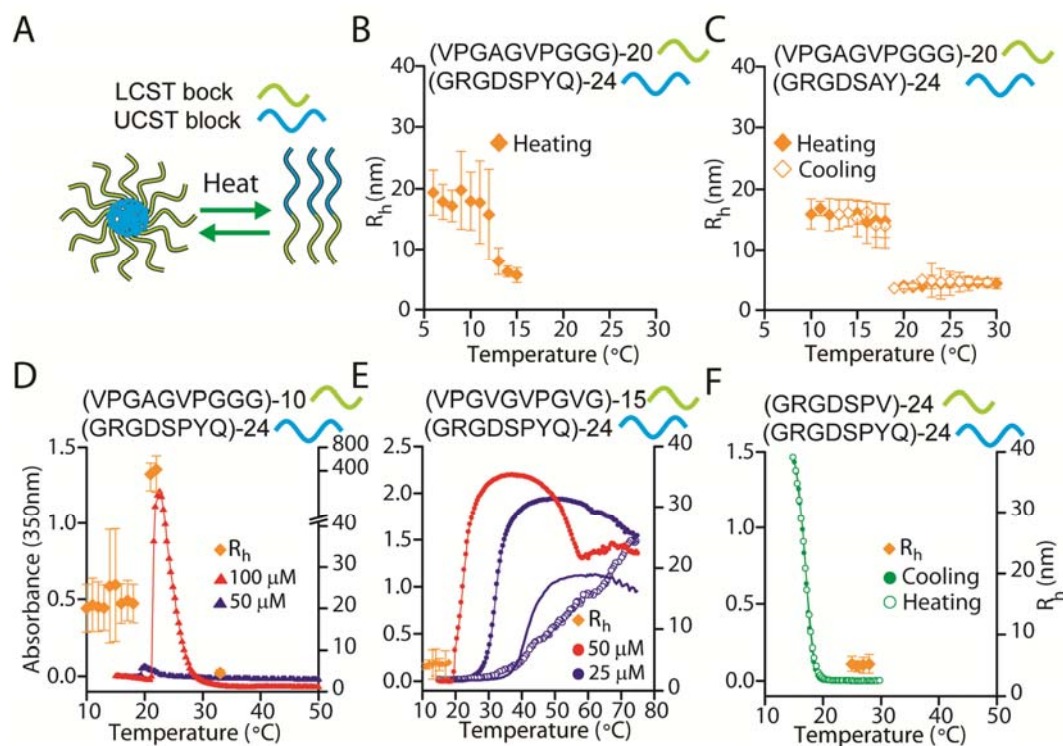


Figure 28. Design of peptide nanoparticles self-assembled through the phase behavior of UCST syntactomers. (A-C) Dynamic light scattering data demonstrates that diblock copolymers composed of a UCST core-forming block and a LCST block (corona) self-assemble into nanoparticles with hydrodynamic radius (R_h) of ~ 17 - 20 nm that then disassemble upon heating and in a reversible manner. (D-E) Turbidimetry (left axis) and dynamic light scattering (right axis) data for three additional diblock copolymers. (D) A diblock copolymer as in (A) but using a LCST block that is half the size, which leads to a core: corona block ratio of $\sim 2:1$, exhibits unusual transient bulk aggregation on disassembly. Turbidimetry data was measured at a heating rate of 0.1 $^{\circ}\text{C}/\text{min}$ – this event is almost imperceptible when heating at 1 $^{\circ}\text{C}/\text{min}$ (Figure 30). (E) Changing the composition of the LCST block for a more hydrophobic peptide polymer composed of VPGVG-30 abolishes self-assembly at low temperatures and results in bulk aggregation upon heating. Open circles correspond to cooling and the solid line to a second heating cycle used to assess reversibility. DLS data was acquired at 25 μM (F) A diblock copolymer composed of two UCST blocks where the core-forming block displays a UCST around 40°C and the corona-forming block does not show an observable UCST (at least down to 4 $^{\circ}\text{C}$) fails to self-assemble into nanoparticles at low temperatures and instead undergoes bulk aggregation. Both DLS and turbidimetry data were acquired at 50 μM in PBS unless otherwise noted.

The development of a large number of genetically-encoded syntactomers that exhibit tunable UCST phase behavior under physiologically relevant conditions and more importantly, the presentation of design principles for the rational design of UCST in peptide polymers is a major contribution toward the design of novel “smart” material systems for medicine and biotechnology.

5.2. Supplemental Figures

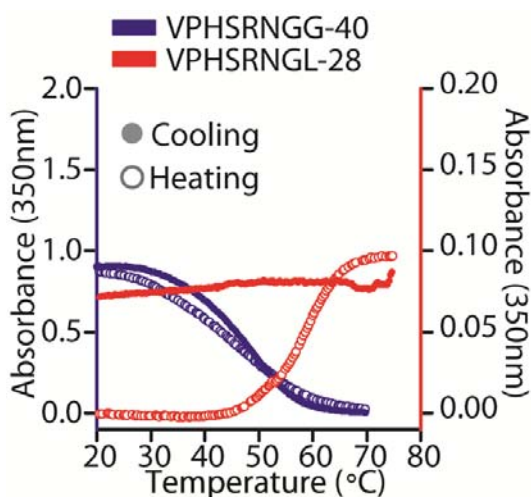


Figure 29. A fine balance between hydrophobic and polar interactions determines whether a syntactomer exhibits LCST or UCST phase behavior. Whereas polymers of VPHSRNGG exhibit UCST phase behavior in PBS at 50 μ M —supplemented with 2 M NaCl—, a slight modification to this motif that favors hydrophobic interactions, by substituting the C-terminal Gly for Leu, as in VPHSRNGL, results in polymers that exhibit LCST phase behavior in PBS at 50 μ M. The latter also exhibits LCST with increasing amounts of NaCl (data not shown).

(GRGDSPYQ)₂₄(VPGAGVPGGG)₁₀

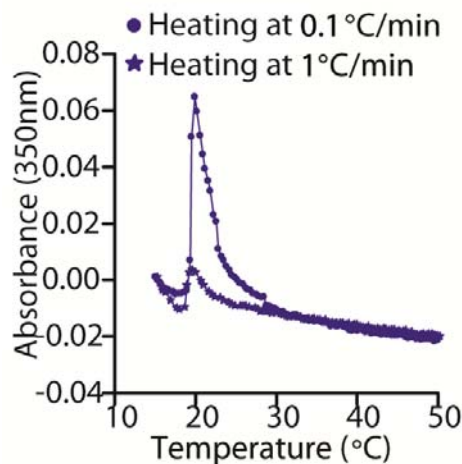


Figure 30. Transient aggregation on disassembly of diblock copolymers composed of UCST-exhibiting hydrophobic cores and short LCST-exhibiting coronas shows a large kinetic dependence. This aggregation event is barely perceptible upon heating at a rate of 1 °C/min (solid circles), but aggregates are significantly stabilized with longer incubation times afforded by a 10 fold decrease in the heating rate (solid stars). In Figure 28D we also show that this event is highly concentration dependent. Turbidimetry data was obtained for 50 μ M samples in PBS.

5.3. Materials and methods

5.3.1. Materials

The reader is referred to the Materials section in Chapter 3 (section 4.3.1) for all materials related to the biomolecular engineering work.

5.3.2. Genetically encoded synthesis of peptide polymers

OERCA and PRe-RDL

Constructs were synthesized by OERCA as described in section 4.3.3 and using the oligonucleotides reported in Table 7. Polymers of GRGDSPYQ, GRGNSPYG, RPAGYDS, QYPSDGRG, GRGDSPV, GRGDSPH, GRGDSPY, GRGDSAY and diblock

copolymers therefrom were synthesized by Pre-RDL as described in section 4.3.3 and using the oligonucleotides in Table 8.

Table 7. DNA sequence information for the synthesis of genes encoding for peptide polymers spanning P-X_n-G motifs from n=0 to n=4 and designed to incorporate a high density of residues capable of hydrogen bonding or that experience electrostatic interactions. Both oligomers and primers were 5'-phosphorylated. DNA oligomers with 4-10 copies of the starting oligomer sequence (as shown) were constructed by OERCA

Repeat unit	Oligomer sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
	GGGCGCGGTGACTCTCCATATGGTCGTGGCGACA		
GRGDSPY	CCCCGTATGGTCGTGGCGACAGTCCGTATGGTCG TGGTGATTCGCCTTAC	18	18
	GGGCGCGGTGACTCTCCACATGGTCGTGGCGAC		
GRGDSPH	AGCCCGCATGGTCGTGGCGACAGTCCGCATGGTC GTGGTGATTCGCCTCAC	18	16
	GGGCGCGGTGACTCTCCAGTGGGTCGTGGCGAC		
GRGDSPV	AGCCCGGTTGGTCGTGGCGACAGTCCGGTGGGTC GTGGTGATTCGCCTGTC	18	18
	GGGCGCGGTGACTCTCCATACGGCGGTCGTGGC		
GRGDSPYG	GACAGCCCGTATGGCGGTCGTGGCGACAGTCCG TATGGCGGTCGTGGTGATTCGCCTTACGGT	18	19
	CGCCATTAGGTTATGACTCTCGTCCGCTGGGCT		
RPLGYDS	ATGACAGCCGTCGCTGGGCTATGACAGTCGTCC	20	20

Repeat unit	Oligomer sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
	TTTGGGTTACGATTCG		
	GGGCGCGGTGACTCTTATCCAGGTCGTGGCGACA		
GRGDSYP	GCTATCCGGGTCGTGGCGACAGTTATCCGGGTCG TGGTGATTCGTACCCT	21	19
	GTTCCACACAGCCGTAACGGTGGTGTGCCGCACT		
VPHSRNG	CTCGTAATGGTGGTGTTCGCACTCCCGTAATGG TGGTGTGCCGATTCTCGCAATGGCGGT	19	19
G			
	CGCCCAGCTGGTTATGACTCTCGTCCGGCGGGCT		
RPAGYDS	ATGACAGCCGTCCGGCGGGCTATGACAGTCGTCC TGCAGGTTACGATTCG	19	19
	GGTCGTGGCGATGCTCCATAACCAGGGCCGCGGT		
GRGDAPY	GACGCCCCGTATCAGGGCCGCGGTGACGCGCCG TATCAGGGCCGTGGTGTGACACCTTACCAA	20	21
Q			
	GGGCGTCCATCTGACTCCTACGGCGGTGCCCCGA		
GRPSDSYG	GCGATAGCTATGGCGGTGCCCCGAGCGACAGTT ACGGCGGTGTCCTTCCGATTCGTATGGT	19	20
	GTTGCTCCAGATGGTGTGGCGCCGGATGGCGTGG		
VAPDG	CCCCTGACGGCGTCGCGCCGGATGGCGTAGCAC CAGACGGT	20	18

Repeat unit	Oligomer sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
	GGGCGTGCTTCTGACTCCTACGGCGGTTCGCGCAA		
GRASDSYG	GCGATAGCTATGGTGGCCGCGCCAGCGACAGTT ACGGCGGTTCGTGCATCCGATTCGTATGGT	19	20
	GGGCGCGACGGTTCTCCATACGGCGGTTCGTGAC		
GRDGSPYG	GGCAGCCCGTATGGCGGTTCGTGACGGCAGTCCG TATGGCGGTTCGTGATGGTTCGCCTTACGGT	19	20
	GTACGTCCTGATGGAGTTCGCCCGGATGGTGTCC		
VRPDG	GCCCGGACGGCGTCCGCCCGGATGGTGTTCGTCC AGACGGT	20	20
	GGGAAGGGTGACTCTCCATACGGCGGTAAAGGC		
GKGDSPY G	GACAGCCCGTATGGCGGTAAAGGCGACAGTCCG TATGGCGGTAAAGGTGATTCGCCTTACGGT	18	18
	GGGCGCGGTGACTCTCCATTCGGCGGTTCGTGGCG		
GRGDSYFG	ACAGCCCGTTTGGCGGTTCGTGGCGACAGTCCGTT TGGCGGTTCGTGGTATTTCGCCTTTTCGGT	20	19
	GGGCGCGGTGAATCTCCATACGGCGGTTCGTGGC		
GRGESPYG	GAAAGCCCGTATGGCGGTTCGTGGCGAAAGTCCG TATGGCGGTTCGTGGTATTTCGCCTTACGGT	19	19
	GTACCTTCTGACGACTATGGGCAGGGTGTGCCGA		
VPDDYD		20	19

Repeat unit	Oligomer sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
QG	GCGATGACTACGGCCAGGGTGTGCCGAGCGATG		
	ACTACGGCCAGGGTGTACCATCGGATGATTACG		
	GCCAAGGG		
	GTACCTTCTGACGACTATGGGGTAGGTGTGCCGA		
VPSDDYG	GCGATGACTACGGCGTCGGTGTGCCGAGCGATG	20	19
VG	ACTACGGCGTCGGTGTACCATCGGATGATTACGG		
	CGTTGGG		

Table 8. DNA sequence information for the synthesis of genes encoding phase transition peptide polymers via Pre-RDL. Oligomers were 5'-phosphorylated.

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
GRGDSPYQ	GCGCGGTGACTCTCCATACCAGG	TGGTAAGGCGAATCACCACGA
	GTCGTGGCGACAGCCCGTATCAG	CCCTGATACGGACTGTCGCCAC
	GGTCGTGGCGACAGTCCGTATCA	GACCCTGATACGGGCTGTCCGC
	GGGTCGTGGTGATTGCCTTACCA	ACGACCCTGGTATGGAGAGTCA
	AGG	CCGCGCCC
GRGNSPYG	GCGCGGTAACTCTCCATACGGCG	ACCGTAAGGCGAATTACCACGA
	GTCGTGGCAACAGCCCGTATGGtG	CCGCCATACGGACTGTTGCCAC
	GTCGTGGCAACAGTCCGTATGGC	GACCACCATACGGGCTGTTGCC

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
	GGTCGTGGTAATTCGCCTTACGGT	ACGACCGCCGTATGGAGAGTTA
	GG	CCGCGCCC
	GCAGTACCCATCTGACGGTCGCG	ACGACCATCCGAAGGGTATTGA
	GTCAGTATCCGAGCGACGGCCGT	CCACGGCCGTCACTCGGATACT
QYPSDGRG	GGTCAGTATCCGAGTGACGGCCG	GACCACGGCCGTCGCTCGGATA
	TGGTCAATACCCTTCGGATGGTCG	CTGACCGCGACCGTCAGATGGG
	TGG	TACTGCCC
	CCGCGGTGACTCTCCACATGGTCG	GTGAGGCGAATCACCACGACCA
	TGGCGACAGCCCGCATGGTCGTG	TGCGGACTGTCGCCACGACCAT
GRGDSPH	GCGACAGTCCGCATGGTCGTGGTG	GCGGGCTGTCGCCACGACCATG
	ATTCGCCTCACGG	TGGAGAGTCACCGCGGCC
	CCGCGGTGACTCTCCATACGGTCG	GTAAGGCGAATCACCACGGCCA
	TGGCGACAGCCCGTATGGTCGTGG	TACGGACTGTCGCCACGACCAT
GRGDSPY	CGACAGTCCGTATGGCCGTGGTG	ACGGGCTGTCGCCACGACCGTA
	ATTCGCCTTACGG	TGGAGAGTCACCGCGGCC
	CCGCGGTGACTCTGCTTACGGTCG	GTATGCCGAATCACCACGGCCA
	TGGCGACAGCGCCTATGGTCGTGG	TACGCACTGTCGCCACGACCAT
GRGDSAY	CGACAGTGCATATGGCCGTGGTG	AGGCGCTGTCGCCACGACCGTA
	ATTCGGCATAACGG	AGCAGAGTCACCGCGGCC

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
	TCGTGTAGGTCCaCATGGCGATGT	GTGCGGACCAACGTCGCCATGC
VGPHGDVGP	GGGTCCGCACGGCCGCGTGGGTC	GGACCCACGCGGCCGTGCGGAC
HGR	CGCATGGCGACGTTGGTCCGCAC	CCACATCGCCATGtGGACCTACA
	GG	CGACC
	TCCaCGCGGCGATCATGGTCCGCG	GTGGTCACCACGCGGACCATGA
	TGGCGACCACGGTCCGCGCGGCG	TCGCCGCGCGGACCGTGGTCGC
GPRGDH	ATCATGGTCCGCGTGGTGACCACG	CACGCGGACCATGATCGCCGCGt
	G	GGACC

5.3.3. Expression and purification of peptide polymers

Expression of plasmid-borne genes encoding all peptide polymers was carried out as previously described in section 4.3.4. The purification of these polymers was carried out with a new purification strategy that exploited the UCST behavior of the peptide polymers. Because of the high yield expression of these UCST peptide polymers in E.coli, they form an insoluble phase that can be readily collected 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 clear supernatant and the viscous phase were discarded and the pellet was resuspended in 3-5 ml of PBS or PBS supplemented with 2-4 M Urea. These solutions were heated to 65 °C for 10 min and centrifuged at 15000 rpm for 15 min at room temperature. The supernatants were further purified by

one or two rounds of UCST transition cycling. The samples supplemented with Urea were diluted 0.5 to 1 times their volume in PBS and heated to 65 °C for 10 min, followed by centrifugation at room temperature for 15 min at 15000 rpm. The resulting supernatant was incubated on ice for a 5 min, typically followed by 2 min at -20 °C, to trigger a pronounced phase transition. The coacervate was then harvested by centrifugation at 14000 rpm for 3 min at -4 °C and the pellet was resuspended in PBS supplemented with 2-4 M Urea. This solution was then heated to 65 °C for 5 min and centrifuged at 15000 rpm for 5 min at room temperature to remove any contaminants. Although most peptide polymers were pure at this stage, we often conducted an additional cycle of UCST transition cycling.

5.3.4. Characterization of the phase transition behavior and secondary structure of peptide polymers

UCST peptide polymers were characterized by turbidimetry and CD spectroscopy as a function of temperature as previously reported for LCST peptide polymers in section 4.3.5. However, because of the UCST behavior exhibited by these constructs, the turbidimetry studies were run from an elevated temperature (typically 75 °C) down to temperatures between 5-15 °C at a cooling rate of 1 °C/min, followed by additional cycles of heating and cooling to assess hysteresis. These experiments were conducted in pure PBS but small amounts of residual urea were usually present in the range of 0.01-0.05 M —because of the dilution of concentrated samples stored in PBS supplemented with 2-4 M urea. To study the concentration dependence of the UCST

behavior without varying amounts of residual urea, the final concentration of urea was fixed at 0.2 M for all polymer concentrations. For experiments conducted in PBS pH 2.0, particular care was taken to minimize heating at high temperatures in order to avoid degradation of the polypeptides. CD spectroscopy was only measured in constructs devoid of tyrosine residues, as aromatic residues interfered with the CD signal in the wavelength range of interest [216].

The self-assembly behavior of diblock copolymers was studied in PBS at 50 μ M by turbidimetry and dynamic light scattering (DynaPro, Wyatt Technologies, CA) measurements as a function of temperature at a heating and cooling rate of ~ 1 $^{\circ}$ C/min.

6. Chapter 5: LCST and UCST peptide polymers with intrinsic biological function

Overview: “Smart” polymers that sense stimuli in aqueous environments and that respond with pronounced physical changes are of great utility in biotechnology and medicine. These polymers are usually composed of synthetic building blocks or very simple amino acid motifs that fail to embed the multilevel functionality that is characteristic of materials composed of proteins, wherein amino acid sequence dictates function at different levels —mechanical behavior, heat-sensitivity, ligand binding and bioactivity are some examples. The work presented in chapter 3 and chapter 4, however, demonstrates that peptide polymers with unprecedented diversity in their amino acid composition may be designed to exhibit lower or upper critical solution temperature (LCST or UCST, respectively) phase transition behavior (Figure 31). Here, we sought to demonstrate the utility of this new paradigm for the design of novel “smart” peptide polymers that exhibit multilevel functionality, namely stimuli-responsiveness and bioactivity.



Figure 31. A new picture of the sequence space of “smart” motifs. The research presented in chapters 3 and 4 draws a new picture of the sequence space of amino acid motifs that exhibit “smart” phase behavior upon polymerization. The original pictures (depicted by small, red and blue circles that are far apart in the proteinogenic space) are now replaced by a unifying picture that suggests the existence of a vast sequence space that encompasses amino acid motifs with opposing phase behaviors. Hence, we hypothesized that this vast sequence space must also encompass peptide sequences with relevant biological functions, as will be demonstrated in this chapter.

6.1. LCST peptide polymers with intrinsic biological function

In an attempt to explore the sequence diversity of LCST peptide polymers, we asked whether existing peptides that exhibit biological activity could be reprogrammed to exhibit LCST behavior. First, we chose the Gly-X-X-Pro-Gly motif as it closely resembles the syntax of many LCST syntactomers, and because specific peptide motifs that belong to this class are matrikines that bind the elastin receptor to exert a variety of

biological functions [229]. We synthesized syntactomers with a repeat unit composed of a recurring matrikine motif (SM1) or a scrambled motif that completely disrupts the Gly-X-X-Pro-Gly pattern (SM2) (Figure 32A), and wherein each repeat unit included a single Cys residue to promote *in vivo* retention by formation of intermolecular disulfide crosslinks [55, 189]. We then examined the bioactivity of two syntactomers with 24 copies of these repeat units (i.e., N=24) in a tumor engraftment model. Nude mice inoculated *s.c.* with 0.5×10^5 HT-1080 tumor cells mixed with the SM1-24 peptide polymer showed a strong reduction in tumor growth by 17 days after inoculation compared with PBS controls ($p < 0.05$) in two independent experiments (Figure 32C and Figure 33). In contrast, the syntactomer with a scrambled motif (SM2-24) that only differed in its sequence from SM1-24 by the transposition of Ala from the second to the fourth position in the repeating hexapeptide, showed similar LCST phase behavior (Figure 32B) but no activity (Figure 32C and Figure 33), as the tumor size was not statistically different from the PBS cohort. The mechanism behind the observed anti-tumor activity, however, is currently unknown and will have to be investigated in future studies.

To explore our ability to reprogram more complex bioactive peptides into LCST syntactomers, we identified a number of candidate peptide drugs that are predicted to be predominantly disordered, and chose the 25-27 amino acid long bioactive peptide domains of endostatin from humans and mice —hEndo and mEndo, respectively (Figure 32D)— for polymerization, as they have been shown to retain the anti-

angiogenic activity of endostatin [230]. As there is only one P-X₂-G motif (P-L-S-G) in these peptides, we incorporated a new PG motif at the C-terminus of mEndo and at the N-terminus of hEndo prior to polymerization to reduce the average distance between P-X_n-G motifs in the peptide polymer. The resulting polymers are intrinsically disordered as seen by their CD spectra (Figure 32E) and exhibit LCST behavior (Figure 32F) that is accompanied by a decrease in the intensity of the random coil peak at ~197 nm (Figure 34) as observed for most syntactomers. The synthesis of LCST polymers with a predominantly disordered repeat unit and yet with stretches of amino acids with a high propensity to fold, which seems a common feature in many bioactive peptides of interest (Figure 35), suggests that even sequences that have some secondary structure and useful biological function can yield syntactomers that display dual functions and that span materials science and biology. Accessing the bioactivity of these endostatin-like and other drug-like “smart” polymers, however, may require the release of monomers from the polymer. As an example, some peptides require an exposed N- or C-terminus to exert their bioactivity [231, 232]. At the organ level, for instance, some peptide drugs that must cross the blood brain barrier to exert a neuromodulatory function may suffer from transport issues across the barrier when circulating in a polymeric form. In such cases, however, one can design a “smart” peptide polymer as an inactive pro-drug that is programmed to release the bioactive peptide at the site of interest in a protease-responsive manner. This can be achieved by the incorporation of protease-cleavable

peptide motifs between repeat units to release bioactive, intact peptide drugs at the site of interest, as Chilkoti *et al.* recently demonstrated with an inert polymer of a glucagon-like peptide-1 analog fused to an ELP to effectively treat type II diabetes in mice [232].

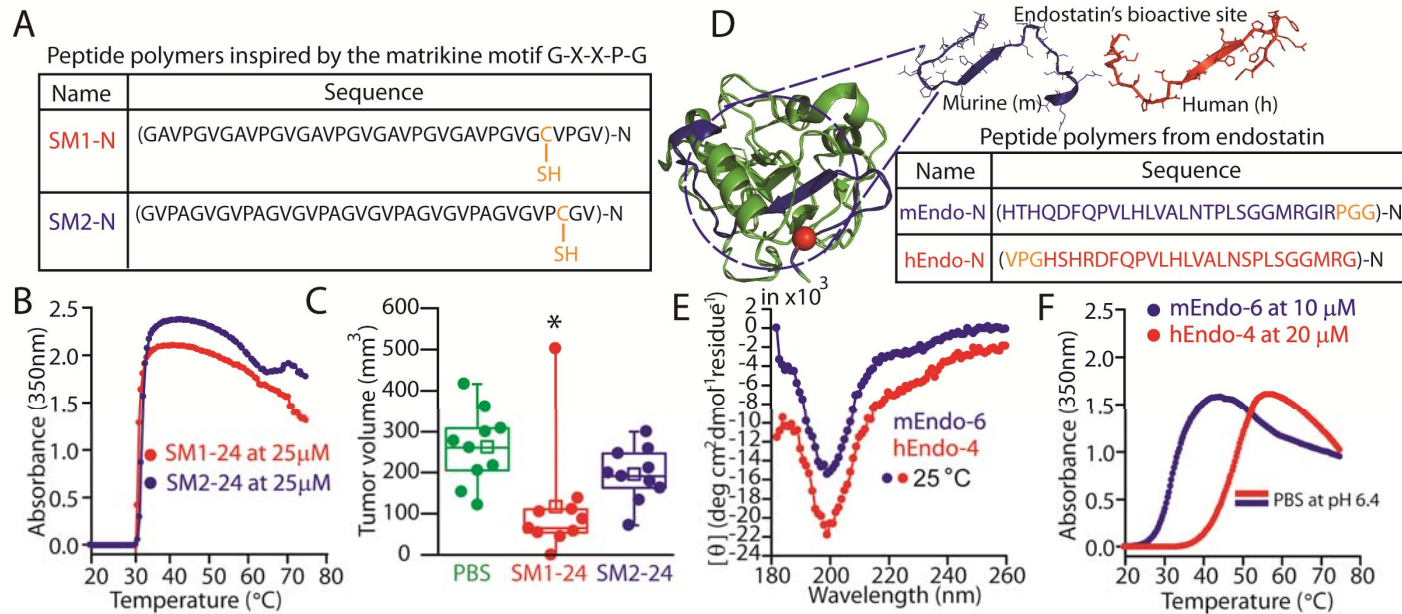


Figure 32. Syntactomers can be designed with a syntax that approaches the complexity of proteins. (A) Amino acid sequence of “smart” matrikine 1 (SM1) and SM2, where only SM1 consists of repeats of a bioactive motif GXXPG found in various extracellular matrix proteins, and SM2 is a scrambled control. (B) SM1 and SM2 displayed identical thermally triggered phase transition behavior in PBS (pH 7.4). (C) Individual tumor volumes measured 17 days after inoculation of 0.5×10^5 HT-1080 tumor cells into the leg of nude mice ($n=10$ per group) and using bioactive SM1 (at $350 \mu\text{M}$ in PBS), non-bioactive SM2 (at $350 \mu\text{M}$ in PBS) and PBS as vehicles for tumor cell inoculation. Asterisk (*) indicates statistical significance with a 95% confidence using a Bonferroni test to compare the mean tumor volumes with respect to the PBS cohort. The anti-tumor activity of SM1-24 and the inertness of SM2-24 were also confirmed in an independent experiment with HT-1080 tumor cells inoculated into the back of nude mice (Fig. S14). (D) Peptide sequences forming the bioactive domains of murine (PDB file: 1DY0) and human endostatin (PDB file: 1BNL), shown here in blue and red, respectively. (E-F) These novel syntactomers display CD spectra characteristic of IDPs (E) and exhibit LCST transition behavior (F) that is accompanied by the characteristic decrease in the intensity of the random-coil peak at $\sim 197 \text{ nm}$, as shown by CD spectroscopy as a function of temperature in PBS at pH 6.4 (Fig. S16). CD spectra were measured in water at a polypeptide concentration of $5 \mu\text{M}$ and at $25 \text{ }^\circ\text{C}$. Images of the 3D structures of endostatin were rendered using PyMOL (<http://pymol.org/>).

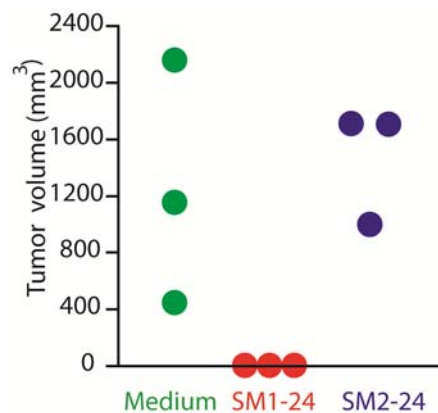


Figure 33. A syntactomer inspired in the matrikine motif GXXPG, SM1-24 (250 μM in PBS), prevented the grafting of 1×10^6 HT1080 tumor cells inoculated into the back of nude mice. Tumor volumes measured 19 days after tumor inoculation. A control syntactomer, SM2-24 (250 μM in PBS), with a disrupted motif but identical phase transition behavior (Fig. 4) had no effect on tumor engraftment compared with a control group where cell culture medium was used as vehicle.

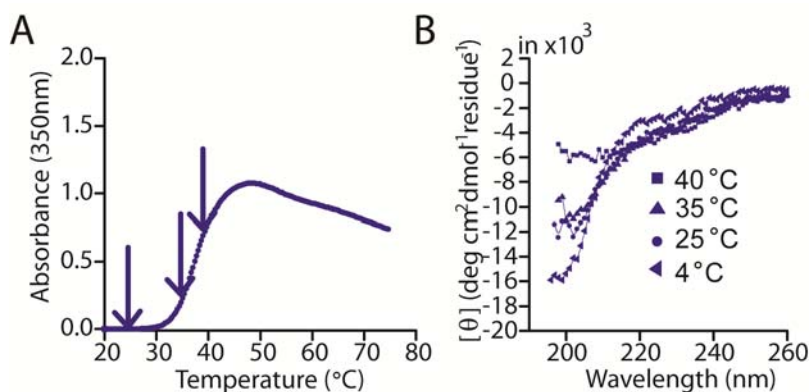


Figure 34. A syntactomer based on the bioactive domain of murine Endostatin displays an inverse phase transition temperature reminiscent of other syntactomers with simpler syntax. (A) Temperature-dependent turbidimetry of mEndo1-6 at 5 μM in PBS (pH 6.4) and (B) corresponding —for the same sample— CD spectroscopy as a function of temperature. The phase transition of this syntactomer was accompanied by a decrease in the disorder of the polypeptide conformation, as evidenced by a greatly diminished random coil negative peak at ~ 197 nm.

6.2. UCST peptide polymers with intrinsic biological function

We have also designed a variety of “smart” peptide polymers built from the cell adhesion motif GRGDSP. Their UCST phase behavior was already discussed in chapter 4 (Figure 24). GRGDSP is a fibronectin-derived sequence and a prototypical cell adhesion motif with highly promiscuous integrin binding [234]. Here, we show that the UCST behavior exhibited by these polymers may constitute a novel bioactivity switch, as we observed that polymers with a UCST below 37 °C, which are thus soluble under regular cell culture conditions (Figure 36B), reduce cell adhesion in a dose-dependent manner (Figure 36C), whereas polymers with a UCST above 37 °C failed to affect cell adhesion in a similar manner —despite having more repeats of the RGD motif, which has been reported to lead to stronger adhesion properties [235]. The anti-adhesion property of this RGD-containing polymer was no longer observed when the cells were allowed to adhere for 18 hours, which demonstrates that the observed response after 3 hours specifically reports a delayed adhesion process and not a cytotoxic effect (Figure 36B). The future utility of these RGD-containing, “smart” peptide polymers is further motivated by our ability to effectively decouple the bioactivity of the peptide and the phase behavior of the polymer, as seen for syntactomers composed of scrambled GRGDSP motifs, such as QYPSDGRG (Figure 26D) and GRDGSPYG (Figure 26E). Moreover, these peptide polymers already exhibit UCSTs in a temperature window that is relevant for mild hyperthermia, as they are insoluble at 37 °C and soluble at 41-42 °C

(Figure 37) in cell culture media supplemented with serum and at 25 μM ; a concentration at which we already observe a marked response for the short peptide polymer studied in Figure 36C.

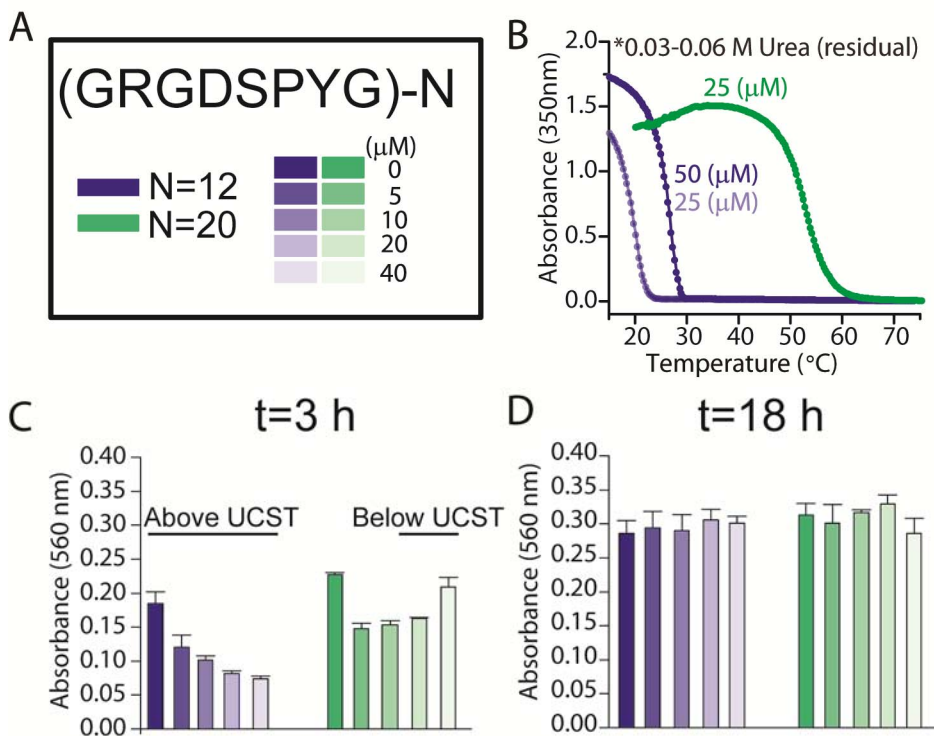


Figure 36. Polymers of the cell adhesion motif GRGDSP display UCST phase behavior and bioactivity. (A) Two peptide polymers composed of 12 and 20 repeats of the octapeptide GRGDSPYG were studied over a range of concentrations. (B) Temperature-dependent turbidimetry. (C-D) Effect of these polymers on the cell adhesion of 0.5×10^5 PC-3 cells to regular 24-well tissue culture plates. The cells were allowed to adhere in the presence of varying concentrations of the peptide polymers for either 3 hours (C) or 18 hours (D), and the cultures were washed with PBS before quantifying cell metabolic activity (indicated by a metabolic product that absorbs strongly at ~ 560 nm). Control cultures ($0 \mu\text{M}$) were prepared by adding PBS supplemented with the maximum amount of residual urea (~ 0.1 M) in these experiments.

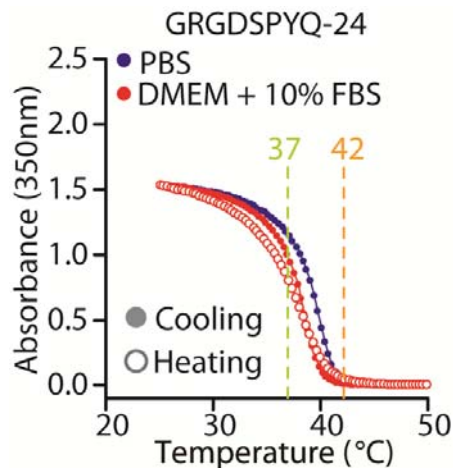


Figure 37. A syntactomer built from repeats of the GRGDSP cell adhesion motif exhibits UCST phase behavior within the window of mild hyperthermia (37-42 °C). Turbidimetry data was obtained at 25 μ M in PBS or DMEM (Sigma) supplemented with 10% fetal bovine serum (Gibco).

Because fibronectin uses a second adhesion motif, PHSRN —the so-called synergy motif—, to modulate binding to a variety of integrins (Figure 38A), we sought to create novel “smart” polymers that better mimic fibronectin. Building on previous work by Hubbell’s group, which demonstrated that the 3D spacing of these motifs in fibronectin can be mimicked by a simple linker composed of six consecutive Gly residues [236], we designed a 19-mer peptide that uses a GGNSGG peptide to link these two motifs. Polymers built from repeats of this novel fibronectin-mimic exhibit UCST phase behavior that is fully reversible under physiologically relevant conditions (Figure 38B), unlike polymers of the synergy motif alone that only exhibit UCST behavior under high ionic strength (Figure 38C) due to the highly charged character of the peptide polymer —yet both motifs have a net charge of +1. We also note that these fibronectin-

like syntactomers already possess a UCST within the relevant temperature range for mild hyperthermia (Figure 38B data at 25 μ M). The intrinsic cell adhesion activity of fibronectin-like and RGD-containing syntactomers represents an exciting direction for the construction of novel “smart” scaffolds with unique capabilities to promote cell adhesion and modulate cell migration. We also envision the design of novel thermally-responsive surfaces for cell sheet engineering that achieve rapid cell detachment by disengaging polymer-integrin interactions upon incubation at temperatures below the apparent UCST of surface-grafted UCST syntactomers.

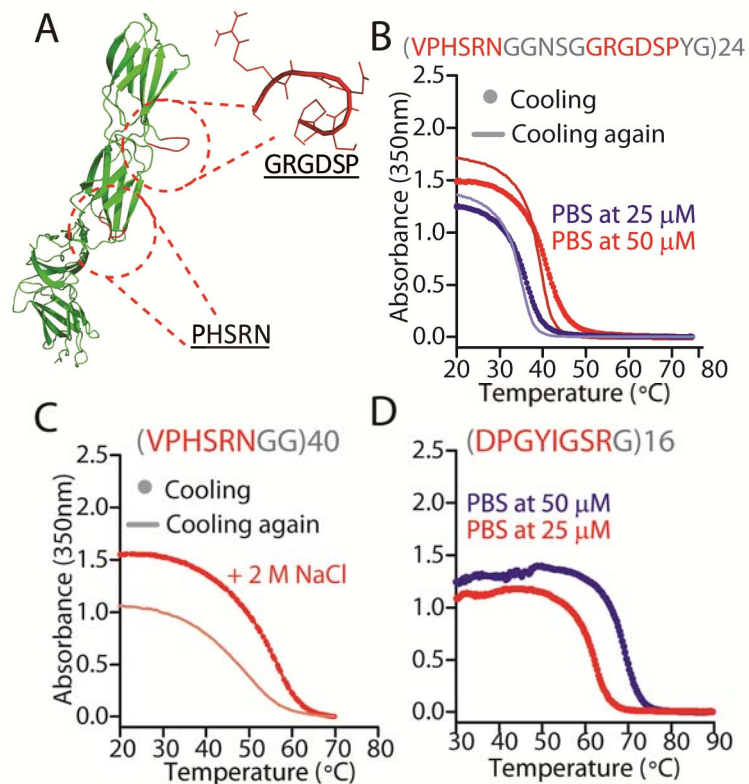


Figure 38. Design of UCST peptide polymers composed of fibronectin and laminin peptides. (A) Fibronectin (PDB file: 1DY0) displays two short peptides (in red), the widely known GRGDSP motif and the synergy motif PHSRN, that act synergistically to control cell adhesion

and engage a variety of integrins. (B) Turbidimetry data for a fibronectin-like syntactomer that mimics the distance between these two motifs reveals its reversible UCST phase behavior. VPHSRN is the native sequence found in fibronectin. (C) The synergy motif alone only exhibits UCST phase behavior under high ionic strength. (D) A syntactomer built from repeats of the laminin peptide DPGYIGSR also exhibits UCST phase behavior. Turbidimetry data was acquired in PBS at 50 μ M unless otherwise indicated. Images of the 3D structures of fibronectin were rendered using PyMOL (<http://pymol.org/>)

The peptide-drug YIGSR was originally derived from laminin (B1 chain) and has been studied for over two decades for its anti-cancer activity, namely anti-angiogenic activity and anti-metastatic potential [237-239]. Interestingly, the activity of this pentapeptide motif is equivalent to that of the parent nonapeptide found in laminin, CDPGYIGSR, and both bind to the laminin receptor [237]. Moreover, linear polymers of YIGSR appear to be more potent than the parent peptide drug (on a weight basis) [240]. Hence, this pentapeptide and the parent octapeptide DPGYIGSR offer an interesting opportunity to design two types of “smart” polymers, with minimal incorporation of non-native residues into the bioactive repeat, that exhibit opposing phase behaviors (LCST and UCST) despite being built from essentially the same peptide drug.

With this idea in mind, we recombinantly synthesized the peptide polymers in Table 9 –by OERCA– to rapidly screen a first set of peptide motifs designed to exhibit LCST or UCST phase transition behavior. Constructs predicted to exhibit LCST were very sensitive to ionic strength and although they appear to exhibit LCST phase behavior, the characterization of these constructs is still ongoing. Polymers of DPGYIGSRG were characterized in detail as it was the only peptide motif in Table 9 that was predicted to yield UCST polymers. The turbidimetry data in Figure 38D

demonstrates that these laminin-derived syntactomers display UCST in a concentration dependent manner as all other UCST syntactomers with similar molecular weight (Figure 25D). Based on this initial screen, however, and in order to assess their bioactivity *in vitro*, we will have to design more hydrophilic variants of this motif to identify laminin-derived UCST polymers that are fully soluble at 37 °C (up to 100 μM).

Table 9. Design of YIGSR-based peptide polymers with various types of phase behavior. Non-native residues in the repeat unit are highlighted in red. The hydrophathy was averaged for all residues using Urry’s scale of hydrophathy [207]. As a reference, the hydrophathy of the UCST motif GRGDSPY is 45.3 °C, whereas the canonical LCST motif VPGAGVPGGG averages 36.1 °C.

Repeat unit	Hydrophathy	Expected phase behavior
DPGYIGSRG	44.4 °C	UCST
KPGYIGSR	32.6 °C	LCST
KPGYIGSRG	35.1 °C	LCST
YIGSRGKP	32.6 °C	LCST
YIGSRPK	29.4 °C	LCST

6.3. Perspective

The ability to confer “smart” behavior to peptidic drugs may overcome major roadblocks in the development of peptides as useful therapeutic molecules, namely their rapid clearance and degradation *in vivo*. These problems are usually addressed by the synthesis of chemically-modified peptide analogs [241, 242], by fusing a peptide drug to a peptide polymer [243, 244] or by conjugation to a synthetic hydrophilic polymer such as PEG [245]. Gene level fusion or covalent conjugation to a large soluble macromolecule

increases the apparent MW of the peptide, thereby slowing its renal clearance, and in many instances can also shield the peptide from proteolytic degradation [245, 246], whereas chemically-modified peptides are designed to resist enzymatic degradation [241]. However, this comes at the cost of including an extraneous polymer “carrier” or the use of unnatural chemical components that can create new issues [247]. First, fusing or covalently attaching large soluble macromolecules typically decreases the affinity of the peptide by at least one order of magnitude or more [248]. Second, synthetic “stealth” polymers like PEG are not biodegradable and have been shown to accumulate *in vivo* with undesired side effects such as renal failure due to the accumulation of polymer-containing vacuoles in the kidney [249]. Third, the drug loading is typically low, as only one (N or C-terminal) to two (fusion to both termini) copies of the drug are carried by a single fusion molecule and chemical conjugation of multiple copies of a peptide to a polymer is undesirable as it yields a heterogeneous product [250].

Reprogramming peptide drugs into a peptide polymer composed of repeats of a bioactive peptide may solve these problems by a residue-level fusion of the peptide and polymer carrier: as (1) each “drug-like” polymer is composed of dozens or hundreds of repeats of the original peptide as an unstructured polymer with tunable solubility; (2) the large size of these polymers can prevent rapid clearance by renal filtration; and (3) unmodified copies of individual, monomer peptide drugs can be released by incorporation of protease cleavage sites between peptide repeats [232]. Furthermore, by

converting a peptide drug into a “smart” polymer (as opposed to an inert hydrophilic polymer) we can add additional functionality that inert polymers cannot impart, such as the ability to convert the polymeric peptide drug into its own injectable depot for sustained delivery. This builds upon the concept of PODs —Protease Operated Depots—, as recently demonstrated by Chilkoti *et al.* [232], except that the peptide drug and carrier are now fused at the residue level into a single motif that is then polymerized, in contrast to PODs that contain two separate domains: a depot forming ELP segment and a drug segment that consists of multiple repeats of the peptide drug. Such two-component systems where the “smart” polymer carrier typically accounts for a large percentage (80-90%) of the injected mass, greatly limits the maximum injected dose of the peptide drug and the active life-time of the depot.

Besides the design of a peptide drug into a “smart” peptide polymer that undergoes a phase transition into an insoluble viscous depot upon subcutaneous injection, these drug-like, phase transition polymers can also be designed as di- or triblock copolymers that self-assemble into micelles or vesicles [198, 251], thereby enabling the transformation of inert nanoparticles that only serve a structural role (as scaffolds) in existing polymeric drug delivery systems into highly bioactive nanoparticles with new therapeutic value. Noteworthy, this dream of truly bioactive nanovehicles is far from being accomplished with the existing approaches to nanoparticle functionalization [252].

6.4. Materials and methods

6.4.1. Materials

The reader is referred to the Materials section in Chapter 3 (section 4.3.1) for all materials related to the biomolecular engineering work.

6.4.2. Genetically encoded synthesis of peptide polymers

OERCA and PRe-RDL

Endostatin-based, RGD-based and laminin-derived constructs were synthesized by OERCA as described in section 4.3.3 and using the oligonucleotides reported in Table 10. Fibronectin-like polypeptides and constructs inspired in the matrikine motif GXXPG were synthesized by Pre-RDL as described in section 4.3.3 and using the oligonucleotides in Table 11.

Table 10. Synthesis of genes encoding for drug-like syntactomers by OERCA. Both oligomers and primers were 5'-phosphorylated.

Repeat unit	Oligomer Sequence (5'-3')	F.	R.
		primer (bp)	primer (bp)
hEndo:	GTGCCGGGTCATAGCCATCGTGACT	20	19
VPGHSHRDFQPVLHLVALNSPL	TCCAGCCTGTTCTGCACCTGGTTGC		
SGGMRG	TCTGAACAGCCCGCTGTCTGGCGGT ATGCGTGGC		
mEndo:	CATACCCACCAGGACTTCCAGCCTG	20	19

HTHQDFQPVHLVALNTPLSG	TTCTGCATCTGGTTGCTCTGAACAC		
GMRGIRPGG	CCCGCTGTCTGGCGGTATGCGTGCC ATTCGTCCGGGTGGT		
KPGYIGSR	AAGCCAGGCTACATCGGTTCTCGCA	21	21
	AACCGGGCTATATCGGCAGCCGTA AACCGGGCTATATCGGCTCTCGTAA ACCTGGTTACATTGGGAGTCGT		
KPGYIGSRG	AAGCCAGGCTACATCGGTTCTCGCG	21	21
	GTAAACCGGGCTATATCGGCAGCC GTGGCAAACCTGGCTATATCGGCTC TCGCGGCAAACCGGTTACATTGG GAGTCGTGGT		
YIGSRGKP	TACATCGGTTCTCGCGGCAAACCAT	20	20
	ATATCGGCTCTCGTGGCAAGCCGTA TATCGGCAGCCGTGGCAAACCTTAC ATTGGGAGTCGTGGTAAACCG		
YIGSRPK	TACATCGGTTCTCGCCAAAATATA	19	20
	TCGGCTCTCGTCCGAAGTATATCGG CAGTCGTCCGAAGTACATTGGGAG CCGTCCTAAA		
DPGYIGSRG	GACCCAGGCTACATCGGTTCTCGCG	19	21
	GTGACCCGGGCTATATCGGCAGCC		

	GTGGGGATCCGGGCTATATCGGCTC		
	TCGTGGTGATCCGGGTTATATTGGC		
	AGTCGTGGC		
VPHSRNGG	GTTCCACACAGCCGTAACGGTGGT	19	19
	GTGCCGCACTCTCGTAATGGTGGTG		
	TTCCGCACTCCCGTAATGGTGGTGT		
	GCCGCATTCTCGCAATGGCGGT		
VPHSCNNG	GTTCCACACAGCTGTAACGGTGGTG	20	19
	TGCCGCACTCTTGTAATGGTGGTGT		
	TCCGCACTCCTGTAATGGTGGTGTG		
	CCGCATTCTTGCAATGGCGGT		
GRGDSPYG	GGGCGCGGTGACTCTCCATACGGC	18	19
	GGTCGTGGCGACAGCCCGTATGGC		
	GGTCGTGGCGACAGTCCGTATGGC		
	GGTCGTGGTGATTTCGCCTTACGGT		
GRDGSPYG	GGGCGCGACGGTTCTCCATACGGC	19	20
	GGTCGTGACGGCAGCCCGTATGGC		
	GGTCGTGACGGCAGTCCGTATGGC		
	GGTCGTGATGGTTCGCCTTACGGT		

Table 11. DNA sequence information for the synthesis of genes encoding drug-like syntactomers via Pre-RDL. Oligomers were 5'-phosphorylated.

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
SM1:	GGCTGTACCAGGTGTTGGTGCCGT	TACACCTGGAACGCAGCCCACG
AVPGVGAVPG	GCCGGGCGTGGGCGCGGTGCCTG	CCCGGCACGGCGCCGACGCCCCG
VGAVPGVGA	GCGTGGGCGCGGTCCCGGGCGTC	GGACCGCGCCCACGCCAGGCAC
VPGVGAVPGV	GGCGCCGTGCCGGGCGTGGGCTG	CGCGCCCACGCCCGGCACGGCA
GCVPGVG	CGTTCAGGTGTAGG	CCAACACCTGGTACAGCCCC
SM2:	GGTACCAGCTGGTGTGGTGTGCC	TACACCGCATGGAACGCCACG
VPAGVGPAG	GGCCGGCGTGGGCGTGCCTGCGG	CCGGCCGGCACGCCGACGCCCCG
VGVPAGVGP	GCGTGGGCGTCCCGGCGGGCGTC	CCGGGACGCCACGCCCGCAGG
AGVGPAGV	GGCGTGCCGGCCGGCGTGGGCGT	CACGCCACGCCGGCCGGCACA
GVPCGVG	TCCATGCGGTGTAGG	CCAACACCAGCTGGTACCCC
	CGTTCACACAGCCGTAACGGTGT	ATTGCGAGAATGCGGCACACCA
VPHSRNG	GCCGCACTCTCGTAATGGCGTTCC	TTACGGGAGTGCGGAACGCCAT
	GCACTCCCGTAATGGTGTGCCGCA	TACGAGAGTGCGGCACACCGTT
	TTCTCGCAATGG	ACGGCTGTGTGGAACGCC
	CGTTCGCACAGCCGTAACGGTG	GTACGGAGAATCGCCGCGACCG
VPHSRNGGNS	GCAACTCTGGTGGCCGCGGTGAC	CCGCTGTTACCGCCATTACGGGA
GGRGDSPYG	AGCCCATATGGTGTGCCGATTCC	ATGCGGCACACCATATGGGCTG
	CGTAATGGCGGTAACAGCGGCGG	TCACCGCGGCCACCAGAGTTGC
	TCGCGGCGATTCTCCGTACGG	CACCGTTACGGCTGTGCGGAAC

Repeat unit	Forward Oligomer (5'-3')	Reverse Oligomer (5'-3')
		GCC
mEndo1:	GCATACCCACCAGGACTTCCAGC	ACCCGGACGAATGCCACGCATA
HTHQDFQPVL	CTGTTCTGCATCTGGTTGCTCTGAA	CCGCCAGACAGCGGGGTGTTCA
HLVALNTPLS	CACCCCGCTGTCTGGCGGTATGCG	GAGCAACCAGATGCAGAACAGG
GGMRGIRPGG	TGGCATTCTGTCGGGTGG	CTGGAAGTCCTGGTGGGTATGCC
		C
	GCGCGGTGACTCTCCATACCAGG	TGGTAAGGCGAATCACCACGA
	GTCGTGGCGACAGCCCGTATCAG	CCCTGATACGGACTGTCGCCAC
GRGDSPTYQ	GGTCGTGGCGACAGTCCGTATCA	GACCCTGATACGGGCTGTCCGC
	GGGTCGTGGTGATTGCCTTACCA	ACGACCCTGGTATGGAGAGTCA
	AGG	CCGCGCCC
	GCAGTACCCATCTGACGGTCGCG	ACGACCATCCGAAGGGTATTGA
	GTCAGTATCCGAGCGACGGCCGT	CCACGGCCGTCACTCGGATACT
QYPSDGRG	GGTCAGTATCCGAGTGACGGCCG	GACCACGGCCGTCGCTCGGATA
	TGGTCAATACCCTTCGGATGGTCG	CTGACCGCGACCGTCAGATGGG
	TGG	TACTGCC

6.4.3. Expression and purification of peptide polymers

LCST and UCST constructs were expressed as described in section 4.3.4.

Polypeptides that display UCST were purified according to the protocol described in section 5.1.3. For the LCST peptide polymers with repeating Cys residues (SM1 and

SM2), purification was carried out in 20 mM TCEP (pH 7.0) to prevent undesired disulfide bond formation throughout the purification process [55]. These polymers were extensively dialyzed against water to remove the TCEP and were lyophilized for storage at -20 °C. Polymers of murine and human endostatin 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 G_nCl₂. These solutions were centrifuged at 14000 rpm for 30 min at 4 °C and the supernatants were extensively dialyzed against water at 4 °C. These peptide polymers were then purified through regular ITC in PBS.

For the *in vivo* studies involving SM1 and SM2, the purified polymer solutions were subjected to an endotoxin-removing protocol to procure that the remaining endotoxin levels were below the FDA required limit of 5 EU/dose. Briefly, ~800 μM solutions of SM1 and SM2 (in endotoxin-free DPBS from Gibco with C²⁺ and Mg²⁺) were run through endotoxin-free columns containing a 1ml Detoxi-Gel Endotoxin Removal Gel, following the instructions of the manufacturer. The samples were eluted in endotoxin-free DPBS and sterile filtered through a 0.2 μm Cellulose Acetate Membrane (VWR, USA), which resulted in a final concentration of ~400 μM for both peptide polymers. We tested the endotoxin levels of these samples using 25 μM solutions and the PYROGENT™ Gel Clot LAL Single Test Vials (0.06 EU/ml sensitivity), following the instructions of the manufacturer.

6.4.4. Evaluation of the anti-tumor activity of SM1 and SM2

Female nude mice (Balb/C nu/nu) with an average body weight of about 20 g were provided by Duke Cancer Facility Isolated Facility (Durham, North Carolina). These experiments were performed in accordance with the regulations from the Duke University Institutional Animal Care and Use Committee. We purchased HT-1080 cells (passage 23) from the Duke Cell Culture Facility (provided by ATCC), which were provided mycoplasma-free as a confluent monolayer in a T75 cm² flask. The cells were grown in DMEM (Sigma, D6046) supplemented with 10% FBS (Heat Inactivated, Gibco) and 1% Streptomycin/Penicillin (Gibco). The cells were subcultured using 0.05% Trypsin-EDTA (Gibco) and seeded onto T175 cm² cell culture-treated flasks (Corning, USA). The cultures were harvested as before and resuspended in DPBS alone or in sterile solutions of SM1 and SM2 in DPBS (for a final concentration of 350 μM) to yield a cell density of 0.5x10⁶ HT-1080 cells per 30 μl. These concentrated cell suspensions were kept on ice before inoculation. A total of 30 mice (n=10 per group) were inoculated with 30 μl of the corresponding cell suspensions into the subcutaneous tissue of the right leg. Mice were monitored for general well-being, weight, and tumor volume. The latter was determined by the equation: volume (mm³)= (width in mm)² × (length in mm) × π/6. A second tumor grafting study (reported in Fig. S14) involved the inoculation of 1x10⁶ HT-1080 cells (30 μl injections) into the subcutaneous tissue of the back of the mice (n=3 per group). The same three vehicles (DPBS, SM1 and SM2) were tested but using 250 μM

solutions of SM1 and SM2 (from a different batch from the polymers used for inoculation into the leg).

6.4.5. Evaluation of the cell adhesive activity of RGD-based, UCST peptide polymers

5×10^5 PC3-luc-C6 cells per well were seeded (24-well plate) and immediately after (while at room temperature) the indicated amounts of the filter-sterilized UCST syntactomer (or PBS supplemented with Urea) were added to the media. Because there is a residual concentration of Urea of up to 0.1 M for the cultures with the highest polymer concentration, control cultures with matching concentrations of urea were also prepared. The cultures were incubated for either 3 or 18 hours at 37 °C and 5% CO₂. The media was then removed and the number of adhered cells was measured using a conventional MTT assay. Briefly, media was replaced and MTT (5mg/ml in PBS) was added to a final concentration of 0.5 mg/ml before incubating the cultures for 1 h at 37 °C/5% CO₂ while protected from light. Cultures were then washed with PBS. The formazan crystals resulting from the metabolization of the MTT salt were solubilized by adding 200 µl DMSO per well. A 50 µl aliquot per well was transferred to a 96-well plate to measure the absorbance at 560 nm using a Microplate reader.

Conclusion

Overall, this dissertation presents a radically new picture of the sequence space of amino acid motifs that exhibit “smart” phase behavior upon polymerization. Two major attributes of this new picture underscore its significance and novelty: 1) it captures the unexpected vastness of this sequence space, and 2) it reveals the space as a continuum that encompasses amino acid motifs that exhibit opposing phase behaviors upon polymerization, namely LCST and UCST. This new picture enabled the invention of novel stimuli responsive peptide polymers with intrinsic biological function, the synthesis of LCST peptide polymers with various degrees of thermal hysteresis and the creation of the first family of genetically-encoded peptide polymers with tunable UCST phase behavior in physiologically-relevant solutions. The biomedical utility of this new viewpoint is particularly evident in the conception and implementation of multifunctional peptide polymers built from bioactive peptide repeat units, but also in the design and synthesis of self-assembling nanovehicles that exhibit shape memory or hyperthermia-mediated disassembly. At a fundamental level, this work demonstrated that the balance of key polar and hydrophobic residue interactions incorporated into an unstructured peptide polymer — a polymeric blank slate— determines whether their soluble to insoluble phase transition occurs above or below a critical solution temperature, and revealed that the syntax of “smart” peptide polymers governs important details of their phase behavior by influencing the emergence of secondary

structure upon aggregation. The seamless fusion of materials and protein design embodied by these novel phase transition peptide polymers may be exploited to study other biologically-relevant forms of stimuli-responsive phase behavior, and promises a new generation of designer polymers with multiple levels of embedded functionality that should lead to new functional materials of broad interest.

References

- [1] MacKay JA, Chen M, McDaniel JR, Liu W, Simnick AJ, Chilkoti A. Self-assembling chimeric polypeptide–doxorubicin conjugate nanoparticles that abolish tumours after a single injection. *Nature Materials*. 2009;8:993-9.
- [2] Stuart MAC, Huck WTS, Genzer J, Müller M, Ober C, Stamm M, et al. Emerging applications of stimuli-responsive polymer materials. *Nature Materials*. 2010;9:101-13.
- [3] McDaniel JR, Callahan DJ, Chilkoti A. Drug delivery to solid tumors by elastin-like polypeptides. *Advanced drug delivery reviews*. 2010.
- [4] Caves JM, Kumar VA, Martinez AW, Kim J, Ripberger CM, Haller CA, et al. The use of microfiber composites of elastin-like protein matrix reinforced with synthetic collagen in the design of vascular grafts. *Biomaterials*. 2010;31:7175-82.
- [5] Nishida K, Yamato M, Hayashida Y, Watanabe K, Yamamoto K, Adachi E, et al. Corneal reconstruction with tissue-engineered cell sheets composed of autologous oral mucosal epithelium. *New England Journal of Medicine*. 2004;351:1187-96.
- [6] Koria P, Yagi H, Kitagawa Y, Megeed Z, Nahmias Y, Sheridan R, 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.
- [7] Caves JM, Cui W, Wen J, Kumar VA, Haller CA, Chaikof EL. Elastin-like protein matrix reinforced with collagen microfibers for soft tissue repair. *Biomaterials*. 2011;32:5371-9.
- [8] Meyer DE, Chilkoti A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nature biotechnology*. 1999;17:1112-5.
- [9] Stayton PS, Shimoboji T, Long C, Chilkoti A, Ghen G, Harris JM, et al. Control of protein–ligand recognition using a stimuli-responsive polymer. *Nature*. 1995;378:472 - 4
- [10] De Las Heras Alarcon C, Pennadam S, Alexander C. Stimuli responsive polymers for biomedical applications. *Chem Soc Rev*. 2005;34:276-85.
- [11] Yeo GC, Keeley FW, Weiss AS. Coacervation of Tropoelastin. *Advances in Colloid and Interface Science*. 2011;167:94–103.
- [12] Gil ES, Hudson SM. Stimuli-reponsive polymers and their bioconjugates. *Progress in Polymer Science*. 2004;29:1173-222.

- [13] Seuring J, Agarwal S. Polymers with Upper Critical Solution Temperature in Aqueous Solution. *Macromolecular Rapid Communications*. 2012.
- [14] Shimada N, Ino H, Maie K, Nakayama M, Kano A, Maruyama A. Ureido-derivatized polymers based on both poly (allylurea) and poly (L-citrulline) exhibit UCST-type phase transition behavior under physiologically relevant conditions. *Biomacromolecules*. 2011.
- [15] San Biagio P, Bulone D, Emanuele A, Palma M. Self-assembly of biopolymeric structures below the threshold of random cross-link percolation. *Biophysical Journal*. 1996;70:494-9.
- [16] Tobitani A, Ross-Murphy SB. Heat-induced gelation of globular proteins. 1. Model for the effects of time and temperature on the gelation time of BSA gels. *Macromolecules*. 1997;30:4845-54.
- [17] Schaefer C, Schlessinger A, Rost B. Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics*. 2010;26:625-31.
- [18] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*. 2005;6:197-208.
- [19] Hynes RO. The extracellular matrix: not just pretty fibrils. *Science*. 2009;326:1216.
- [20] Barker TH. The role of ECM proteins and protein fragments in guiding cell behavior in regenerative medicine. *Biomaterials*. 2011;32:4211-4.
- [21] Holst J, Watson S, Lord MS, Eamegdool SS, Bax DV, Nivison-Smith LB, et al. Substrate elasticity provides mechanical signals for the expansion of hemopoietic stem and progenitor cells. *Nature biotechnology*. 2010;28:1123-8.
- [22] Lv S, Dudek DM, Cao Y, Balamurali M, Gosline J, Li H. Designed biomaterials to mimic the mechanical properties of muscles. *Nature*. 2010;465:69-73.
- [23] Ostrov N, Gazit E. Genetic Engineering of Biomolecular Scaffolds for the Fabrication of Organic and Metallic Nanowires. *Angewandte Chemie International Edition*. 2010;49:3018-21.
- [24] Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature*. 2011;472:499-503.

- [25] Elvin CM, Carr AG, Huson MG, Maxwell JM, Pearson RD, Vuocolo T, et al. Synthesis and properties of crosslinked recombinant pro-resilin. *Nature*. 2005;437:999-1002.
- [26] Lazaris A, Arcidiacono S, Huang Y, Zhou JF, Duguay F, Chretien N, et al. Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science*. 2002;295:472.
- [27] Whitesides GM, Grzybowski B. Self-assembly at all scales. *Science*. 2002;295:2418.
- [28] Fang PA, Conway JF, Margolis HC, Simmer JP, Beniash E. Hierarchical self-assembly of amelogenin and the regulation of biomineralization at the nanoscale. *Proceedings of the National Academy of Sciences*. 2011;108:14097-102.
- [29] Lee H, Lee BP, Messersmith PB. A reversible wet/dry adhesive inspired by mussels and geckos. *NATURE-LONDON*-. 2007;448:338.
- [30] Holliger P, Hudson PJ. Engineered antibody fragments and the rise of single domains. *Nature biotechnology*. 2005;23:1126-36.
- [31] Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKolver RC, et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature biotechnology*. 2009;27:851-7.
- [32] Chen Y, Guan Z. Bioinspired Modular Synthesis of Elastin-Mimic Polymers To Probe the Mechanism of Elastin Elasticity. *Journal of the American Chemical Society*. 2010;132:4577-9.
- [33] Bott K, Upton Z, Schrobback K, Ehrbar M, Hubbell JA, Lutolf MP, et al. The effect of matrix characteristics on fibroblast proliferation in 3D gels. *Biomaterials*. 2010.
- [34] Charati MB, Kiick KL. Elastomeric polypeptide-based biomaterials. *Polymer Chemistry*. 2010;1:1160-70.
- [35] Karanicolas J, Kuhlman B. Computational design of affinity and specificity at protein-protein interfaces. *Current opinion in structural biology*. 2009;19:458-63.
- [36] Gurtner GC, Werner S, Barrandon Y, Longaker MT. Wound repair and regeneration. *Nature*. 2008;453:314-21.

- [37] Borselli C, Storrie H, Benesch-Lee F, Shvartsman D, Cezar C, Lichtman JW, et al. Functional muscle regeneration with combined delivery of angiogenesis and myogenesis factors. *Proceedings of the National Academy of Sciences*. 2010;107:3287.
- [38] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-74.
- [39] Jain RK. Molecular regulation of vessel maturation. *Nature Medicine*. 2003;9:685-93.
- [40] Kerbel R, Folkman J. Clinical translation of angiogenesis inhibitors. *Nature Reviews Cancer*. 2002;2:727-39.
- [41] De Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. *Nature Reviews Cancer*. 2006;6:24-37.
- [42] Roh JD, Sawh-Martinez R, Brennan MP, Jay SM, Devine L, Rao DA, et al. Tissue-engineered vascular grafts transform into mature blood vessels via an inflammation-mediated process of vascular remodeling. *Proceedings of the National Academy of Sciences*. 2010;107:4669.
- [43] Labrijn AF, Aalberse RC, Schuurman J. When binding is enough: nonactivating antibody formats. *Current opinion in immunology*. 2008;20:479-85.
- [44] Jones DS, Tsai PC, Cochran JR. Engineering hepatocyte growth factor fragments with high stability and activity as Met receptor agonists and antagonists. *Proceedings of the National Academy of Sciences*. 2011;108:13035-40.
- [45] Reichert JM, Valge-Archer VE. Development trends for monoclonal antibody cancer therapeutics. *Nature Reviews Drug Discovery*. 2007;6:349-56.
- [46] Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology*. 2011.
- [47] Rasmussen SGF, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, et al. Crystal structure of the [bgr] 2 adrenergic receptor-Gs protein complex. *Nature*. 2011.
- [48] O'Leary LER, Fallas JA, Bakota EL, Kang MK, Hartgerink JD. Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nature Chemistry*. 2011.

- [49] 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;10:141-8.
- [50] Webber MJ, Tongers J, Newcomb CJ, Marquardt KT, Bauersachs J, Losordo DW, et al. Supramolecular nanostructures that mimic VEGF as a strategy for ischemic tissue repair. *Proceedings of the National Academy of Sciences*. 2011;108:13438-43.
- [51] Geng Y, Dalhaimer P, Cai S, Tsai R, Tewari M, Minko T, et al. Shape effects of filaments versus spherical particles in flow and drug delivery. *Nature nanotechnology*. 2007;2:249-55.
- [52] Haines-Butterick L, Rajagopal K, Branco M, Salick D, Rughani R, Pilarz M, et al. Controlling hydrogelation kinetics by peptide design for three-dimensional encapsulation and injectable delivery of cells. *Proceedings of the National Academy of Sciences*. 2007;104:7791.
- [53] Rosenzweig A. Cardiac cell therapy—mixed results from mixed cells. *New England Journal of Medicine*. 2006;355:1274-7.
- [54] Suuronen EJ, Veinot JP, Wong S, Kapila V, Price J, Griffith M, et al. Tissue-engineered injectable collagen-based matrices for improved cell delivery and vascularization of ischemic tissue using CD133+ progenitors expanded from the peripheral blood. *Circulation*. 2006;114:I-138.
- [55] Asai D, Xu D, Liu W, Garcia Quiroz F, Callahan DJ, Zalutsky MR, et al. Protein polymer hydrogels by *in situ*, rapid and reversible self-gelation. *Biomaterials*. 2012.
- [56] Liu W, MacKay JA, Dreher MR, Chen M, McDaniel JR, Simnick AJ, et al. Injectable intratumoral depot of thermally responsive polypeptide-radionuclide conjugates delays tumor progression in a mouse model. *Journal of Controlled Release*. 2010;144:2-9.
- [57] Liu W, McDaniel J, Li X, Asai D, Quiroz FG, Schaal J, et al. Brachytherapy Using Injectable Seeds That Are Self-Assembled from Genetically Encoded Polypeptides In Situ. *Cancer research*. 2012;72:5956-65.
- [58] de las Heras Alarcón C, Pennadam S, Alexander C. Stimuli responsive polymers for biomedical applications. *Chem Soc Rev*. 2005;34:276-85.
- [59] Braun D, Cherdron H, Rehahn M, Ritter H, Voit B. *Functional Polymers. Polymer Synthesis: Theory and Practice*: Springer; 2013. p. 375-95.

- [60] Seuring J, Bayer FM, Huber K, Agarwal S. Upper Critical Solution Temperature of Poly (N-acryloyl glycinamide) in Water: A Concealed Property. *Macromolecules*. 2011;45:374-84.
- [61] 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;45:3910-8.
- [62] Gosline J, Lillie M, Carrington E, Guerette P, Ortlepp C, Savage K. Elastic proteins: biological roles and mechanical properties. *Philosophical Transactions B*. 2002;357:121.
- [63] Gupta A, Mohanty B, Bohidar H. Flory temperature and upper critical solution temperature of gelatin solutions. *Biomacromolecules*. 2005;6:1623-7.
- [64] Vrhovski B, Weiss A. Biochemistry of tropoelastin. *European Journal of Biochemistry*. 1998;258:1-18.
- [65] MacEwan S, Chilkoti A. Elastin-like polypeptides: Biomedical applications of tunable biopolymers. *Peptide Science*. 2010;94:60-77.
- [66] Simnick A, Lim D, Chow D, Chilkoti A. Biomedical and biotechnological applications of elastin-like polypeptides. *Polymer Reviews*. 2007;47:121.
- [67] Chow D, Nunalee M, Lim D, Simnick A, Chilkoti A. Peptide-based biopolymers in biomedicine and biotechnology. *Materials Science & Engineering R*. 2008.
- [68] Lillie M, Gosline J. The viscoelastic basis for the tensile strength of elastin. *International Journal of Biological Macromolecules*. 2002;30:119-27.
- [69] Antonicelli F, Bellon G, Debelle L, Hornebeck W. Elastin-elastases and inflamm-aging. *Current topics in developmental biology*. 2007;79:99.
- [70] Li D, Brooke B, Davis E, Mecham R, Sorensen L, Boak B, et al. Elastin is an essential determinant of arterial morphogenesis. *Nature*. 1998;393:276-80.
- [71] Partridge S, Elsdon D, Thomas J, Dorfman A, Telser A, Ho P. Biosynthesis of the desmosine and isodesmosine cross-bridges in elastin. *The Biochemical journal*. 1964;93:30C.
- [72] Bressan G, Pasquali-Ronchetti I, Fornieri C, Mattioli F, Castellani I, Volpin D. Relevance of aggregation properties of tropoelastin to the assembly and structure of elastic fibers. *Journal of ultrastructure and molecular structure research*. 1986;94:209-16.

- [73] Foster J, Bruenger E, Gray W, Sandberg L. Isolation and amino acid sequences of tropoelastin peptides. *Journal of Biological Chemistry*. 1973;248:2876.
- [74] Long MM, Rapaka RS, Volpin D, Pasquali-Ronchetti, DW U. Spectroscopic and electron micrographic studies on the repeat tetrapeptide of tropoelastin: (Val-Pro-Gly-Gly)_n *Archives of Biochemistry and Biophysics*. 1980;201:445-52
- [75] Cox BA, Starcher BC, DW U. Coacervation of Tropoelastin Results in Fiber Formation. *Journal of Biological Chemistry*. 1974;249:997-8.
- [76] Tamburro A, Guantieri V, Daga-Gordini D, Abatangelo G. Conformational transitions of [alpha]-elastin. *Biochimica et Biophysica Acta (BBA)-Protein Structure*. 1977;492:370-6.
- [77] Chen Y, Guan Z. Bioinspired Modular Synthesis of Elastin-Mimic Polymers To Probe the Mechanism of Elastin Elasticity. *Journal of the American Chemical Society*.561-73.
- [78] Urry D, Long M, Cox B, Ohnishi T, Mitchell L, Jacobs M. The synthetic polypentapeptide of elastin coacervates and forms filamentous aggregates. *Biochimica et Biophysica Acta (BBA)-Protein Structure*. 1974;371:597-602.
- [79] Indik Z, Yeh H, Ornstein-Goldstein N, Sheppard P, Anderson N, Rosenbloom J, et al. Alternative splicing of human elastin mRNA indicated by sequence analysis of cloned genomic and complementary DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 1987;84:5680.
- [80] Chung M, Miao M, Stahl R, Chan E, Parkinson J, Keeley F. Sequences and domain structures of mammalian, avian, amphibian and teleost tropoelastins: clues to the evolutionary history of elastins. *Matrix Biology*. 2006;25:492-504.
- [81] Tamburro A. A never-ending love story with elastin: a scientific autobiography. *Nanomedicine*. 2009;4:469-87.
- [82] Tamburro A, Bochicchio B, Pepe A. Dissection of Human Tropoelastin: Exon-By-Exon Chemical Synthesis and Related Conformational Studies. *Biochemistry*. 2003;42:13347-62.
- [83] Vrhovski B, Jensen S, Weiss A. Coacervation characteristics of recombinant human tropoelastin. *European Journal of Biochemistry*. 1997;250:92-8.

- [84] Tu Y, Weiss A. Glycosaminoglycan-mediated coacervation of tropoelastin abolishes the critical concentration, accelerates coacervate formation, and facilitates spherule fusion: implications for tropoelastin microassembly. *Biomacromolecules*. 2008;9:1739-44.
- [85] Wu W, Vrhovski B, Weiss A. Glycosaminoglycans mediate the coacervation of human tropoelastin through dominant charge interactions involving lysine side chains. *Journal of Biological Chemistry*. 1999;274:21719.
- [86] Daamen W, Veerkamp J, van Hest J, van Kuppevelt T. Elastin as a biomaterial for tissue engineering. *Biomaterials*. 2007;28:4378-98.
- [87] O'Connell M, Murthy S, Phan S, Xu C, Buchanan J, Spilker R, et al. The three-dimensional micro- and nanostructure of the aortic medial lamellar unit measured using 3D confocal and electron microscopy imaging. *Matrix Biology*. 2008;27:171-81.
- [88] Choi J, Bergdahl A, Zheng Q, Starcher B, Yanagisawa H, Davis E. Analysis of dermal elastic fibers in the absence of fibulin-5 reveals potential roles for fibulin-5 in elastic fiber assembly. *Matrix Biology*. 2009.
- [89] Cirulis J, Bellingham C, Davis E, Hubmacher D, Reinhardt D, Mecham R, et al. Fibrillins, fibulins, and matrix-associated glycoprotein modulate the kinetics and morphology of in vitro self-assembly of a recombinant elastin-like polypeptide. *Biochemistry*. 2008;47:12601.
- [90] Kozel B, Rongish B, Czirok A, Zach J, Little C, Davis E, et al. Elastic fiber formation: a dynamic view of extracellular matrix assembly using timer reporters. *Journal of cellular physiology*. 2006;207:87-96.
- [91] Wise S, Weiss A. Tropoelastin. *International Journal of Biochemistry and Cell Biology*. 2009;41:494-7.
- [92] Hinek A, Wrenn D, Mecham R, Barondes S. The elastin receptor: a galactoside-binding protein. *Science*. 1988;239:1539.
- [93] Hinek A, Rabinovitch M. 67-kD elastin-binding protein is a protective "companion" of extracellular insoluble elastin and intracellular tropoelastin. *Journal of Cell Biology*. 1994;126:563.
- [94] Taddese S, Weiss A, Jahreis G, Neubert R, Schmelzer C. In vitro degradation of human tropoelastin by MMP-12 and the generation of matrikines from domain 24. *Matrix Biology*. 2009;28:84-91.

- [95] Sallach R, Cui W, Balderrama F, Martinez A, Wen J, Haller C, et al. Long-term biostability of self-assembling protein polymers in the absence of covalent crosslinking. *Biomaterials*. 2010;31:779-91.
- [96] Bedell-Hogan D, Trackman P, Abrams W, Rosenbloom J, Kagan H. Oxidation, cross-linking, and insolubilization of recombinant tropoelastin by purified lysyl oxidase. *Journal of Biological Chemistry*. 1993;268:10345.
- [97] Liu X, Zhao Y, Gao J, Pawlyk B, Starcher B, Spencer J, et al. Elastic fiber homeostasis requires lysyl oxidase-like 1 protein. *Nature genetics*. 2004;36:178-82.
- [98] Wise S, Mithieux S, Raftery M, Weiss A. Specificity in the coacervation of tropoelastin: solvent exposed lysines. *Journal of Structural Biology*. 2005;149:273-81.
- [99] Mithieux S, Wise S, Raftery M, Starcher B, Weiss A. A model two-component system for studying the architecture of elastin assembly in vitro. *Journal of Structural Biology*. 2005;149:282-9.
- [100] Fung Y. *Biomechanics: mechanical properties of living tissues*. 2 ed. New York: Springer-Verlag; 1993.
- [101] Aaron B, Gosline J. Elastin as a random-network elastomer: a mechanical and optical analysis of single elastin fibers. *Biopolymers*. 1981;20:1247-60.
- [102] Loretta L, Maruccio G, Pompa P, Bochicchio B, Tamburro A, Cingolani R, et al. Amyloid-like fibrils in elastin-related polypeptides: Structural characterization and elastic properties. *Biomacromolecules*. 2008;9:796-803.
- [103] Urry D. Entropic elastic processes in protein mechanisms. I. Elastic structure due to an inverse temperature transition and elasticity due to internal chain dynamics. *Journal of protein chemistry*. 1988;7:1-34.
- [104] Hovee C, Flory P. The elastic properties of elastin. *Biopolymers*. 1974;13:677-86.
- [105] Li B, Alonso D, Bennion B, Daggett V. Hydrophobic hydration is an important source of elasticity in elastin-based biopolymers. *J Am Chem Soc*. 2001;123:11991-8.
- [106] Urry D. Physical Chemistry of Biological Free Energy Transduction As Demonstrated by Elastic Protein-Based Polymer†. *J Phys Chem B*. 1997;101:11007-28.
- [107] Rosenbloom J, Abrams W, Mecham R. Extracellular matrix 4: the elastic fiber. *The FASEB Journal*. 1993;7:1208-18.

- [108] Shapiro S, Endicott S, Province M, Pierce J, Campbell E. Marked longevity of human lung parenchymal elastic fibers deduced from prevalence of D-aspartate and nuclear weapons-related radiocarbon. *Journal of Clinical Investigation*. 1991;87:1828.
- [109] Senior R, Griffin G, Mecham R. Chemotactic activity of elastin-derived peptides. *Journal of Clinical Investigation*. 1980;66:859.
- [110] Senior R, Griffin G, Mecham R, Wrenn D, Prasad K, Urry D. Val-Gly-Val-Ala-Pro-Gly, a repeating peptide in elastin, is chemotactic for fibroblasts and monocytes. *Journal of Cell Biology*. 1984;99:870.
- [111] Hinek A, Pshezhetsky A, Von Itzstein M, Starcher B. Lysosomal sialidase (neuraminidase-1) is targeted to the cell surface in a multiprotein complex that facilitates elastic fiber assembly. *Journal of Biological Chemistry*. 2006;281:3698.
- [112] Brassart B, Fuchs P, Huet E, Alix A, Wallach J, Tamburro A, et al. Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts. *Journal of Biological Chemistry*. 2001;276:5222.
- [113] Robinet A, Fahem A, Cauchard J, Huet E, Vincent L, Lorimier S, et al. Elastin-derived peptides enhance angiogenesis by promoting endothelial cell migration and tubulogenesis through upregulation of MT1-MMP. *Journal of cell science*. 2005;118:343.
- [114] Antonicelli F, Bellon G, Lorimier S, Hornebeck W. Role of the elastin receptor complex (S-Gal/Cath-A/Neu-1) in skin repair and regeneration. *Wound Repair and Regeneration*. 2009;17:631-8.
- [115] Leach J, Wolinsky J, Stone P, Wong J. Crosslinked alpha-elastin biomaterials: towards a processable elastin mimetic scaffold. *Acta Biomaterialia*. 2005;1:155-64.
- [116] Mithieux S, Rasko J, Weiss A. Synthetic elastin hydrogels derived from massive elastic assemblies of self-organized human protein monomers. *Biomaterials*. 2004;25:4921-7.
- [117] Mithieux S, Tu Y, Korkmaz E, Braet F, Weiss A. In situ polymerization of tropoelastin in the absence of chemical cross-linking. *Biomaterials*. 2009;30:431-5.
- [118] Keeley F, Bellingham C, Woodhouse K. Elastin as a self-organizing biomaterial: use of recombinantly expressed human elastin polypeptides as a model for investigations of structure and self-assembly of elastin. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*. 2002;357:185-9.

- [119] Dyksterhuis L, Baldock C, Lammie D, Wess T, Weiss A. Domains 17-27 of tropoelastin contain key regions of contact for coacervation and contain an unusual turn-containing crosslinking domain. *Matrix Biology*. 2007;26:125-35.
- [120] Urry D, Gowda D, Parker T, Luan C, Reid M, Harris C, et al. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers*. 1992;32:1243-50.
- [121] Rincon A, Molina-Martinez I, de Las Heras B, Alonso M, Bailez C, Rodriguez-Cabello J, et al. Biocompatibility of elastin-like polymer poly (VPAVG) microparticles: in vitro and in vivo studies. *Journal of Biomedical Materials Research-Part A*. 2006;78:343-51.
- [122] Lee J, Macosko C, Urry D. Elastomeric polypentapeptides cross-linked into matrixes and fibers. *Biomacromolecules*. 2001;2:170-9.
- [123] Flamia R, Zhdan P, Martino M, Castle J, Tamburro A. AFM study of the elastin-like biopolymer poly (ValGlyGlyValGly). *Biomacromolecules*. 2004;5:1511-8.
- [124] Martino M, Coviello A, Tamburro A. Synthesis and structural characterization of poly (LGGVG), an elastin-like polypeptide. *International Journal of Biological Macromolecules*. 2000;27:59-64.
- [125] Meyer D, Chilkoti A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nature biotechnology*. 1999;17:1112-5.
- [126] McDaniel J, MacKay J, Quiroz F, Chilkoti A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules*. 2010;11:944-52.
- [127] Martin S, Vrhovski B, Weiss A. Total synthesis and expression in *Escherichia coli* of a gene encoding human tropoelastin. *Gene*. 1995;154:159-66.
- [128] Girotti A, Reguera J, Rodríguez-Cabello J, Arias F, Alonso M, Testera A. Design and bioproduction of a recombinant multi (bio) functional elastin-like protein polymer containing cell adhesion sequences for tissue engineering purposes. *Journal of Materials Science: Materials in Medicine*. 2004;15:479-84.
- [129] Mcpherson D, Morrow C, Minehan D, Wu J, Hunter E, Urry D. Production and purification of a recombinant elastomeric polypeptide, G-(VPGVG) 19-VPGV, from *Escherichia coli*. *Biotechnology progress*. 1992;8:347-52.

- [130] Meyer D, 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;3:357-67.
- [131] Urry D, Hugel T, Seitz M, Gaub H, Sheiba L, Dea J, et al. Elastin: a representative ideal protein elastomer. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*. 2002;357:169.
- [132] Indik Z, Abrams W, Kucich U, Gibson C, Mecham R, Rosenbloom J. Production of recombinant human tropoelastin: Characterization and demonstration of immunologic and chemotactic activity* 1. *Archives of Biochemistry and Biophysics*. 1990;280:80-6.
- [133] Herzog R, Singh N, Urry D, Daniell H. Expression of a synthetic protein-based polymer (elastomer) gene in *Aspergillus nidulans*. *Applied microbiology and biotechnology*. 1997;47:368-72.
- [134] Zhang X, Urry D, Daniell H. Expression of an environmentally friendly synthetic protein-based polymer gene in transgenic tobacco plants. *Plant Cell Reports*. 1996;16:174-9.
- [135] Lyons R, Nairn K, Huson M, Kim M, Dumsday G, Elvin C. Comparisons of Recombinant Resilin-like Proteins: Repetitive Domains Are Sufficient to Confer Resilin-like Properties. *Biomacromolecules*. 2009;10:3009--14.
- [136] Kyle S, Aggeli A, Ingham E, McPherson M. Production of self-assembling biomaterials for tissue engineering. *Trends in Biotechnology*. 2009.
- [137] Conley A, Joensuu J, Jevnikar A, Menassa R, Brandle J. Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants. *Biotechnology and bioengineering*. 2009;103:562-73.
- [138] McHale M, Setton L, Chilkoti A. Synthesis and in vitro evaluation of enzymatically cross-linked elastin-like polypeptide gels for cartilaginous tissue repair. *Tissue Engineering*. 2005;11:1768-79.
- [139] McPherson D, Xu J, Urry D. Product Purification by Reversible Phase Transition Following *Escherichia coli* Expression of Genes Encoding up to 251 Repeats of the Elastomeric Pentapeptide GVGVP. *Protein Expression and Purification*. 1996;7:51-7.
- [140] Conrad U, Plagmann I, Malchow S, Sack M, Floss D, Kruglov A, et al. ELPylated anti-human TNF therapeutic single-domain antibodies for prevention of lethal septic shock. *Plant Biotechnology Journal*. Epub ahead of print.

- [141] Scheller J, Henggeler D, Viviani A, Conrad U. Purification of spider silk-elastin from transgenic plants and application for human chondrocyte proliferation. *Transgenic research*. 2004;13:51-7.
- [142] Floss D, Schallau K, Rose-John S, Conrad U, Scheller J. Elastin-like polypeptides revolutionize recombinant protein expression and their biomedical application. *Trends in Biotechnology*. 2009.
- [143] Scheller J, Leps M, Conrad U. Forcing single-chain variable fragment production in tobacco seeds by fusion to elastin-like polypeptides. *Plant Biotechnology Journal*. 2006;4:243-9.
- [144] Schipperus R, Teeuwen R, Werten M, Eggink G, de Wolf F. Secreted production of an elastin-like polypeptide by *Pichia pastoris*. *Applied microbiology and biotechnology*. 2009;85:293-301.
- [145] Chow D, Dreher M, Trabbic-Carlson K, Chilkoti A. Ultra-high expression of a thermally responsive recombinant fusion protein in *E. coli*. *Biotechnology progress*. 2006;22:638.
- [146] Guda C, Zhang X, McPherson D, Xu J, Cherry J, Urry D, et al. Hyper expression of an environmentally friendly synthetic polymer gene. *Biotechnology Letters*. 1995;17:745-50.
- [147] Christensen T, Amiram M, Dagher S, Trabbic-Carlson K, Shamji M, Setton L, et al. Fusion order controls expression level and activity of elastin-like polypeptide fusion proteins. *Protein Science*. 2009;18:1377-87.
- [148] Heilshorn S, DiZio K, Welsh E, Tirrell D. Endothelial cell adhesion to the fibronectin CS5 domain in artificial extracellular matrix proteins. *Biomaterials*. 2003;24:4245-52.
- [149] Trabbic-Carlson K, Setton L, Chilkoti A. Swelling and mechanical behaviors of chemically cross-linked hydrogels of elastin-like polypeptides. *Biomacromolecules*. 2003;4:572-80.
- [150] Bellingham C, Woodhouse K, Robson P, Rothstein S, Keeley F. Self-aggregation characteristics of recombinantly expressed human elastin polypeptides. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular Enzymology*. 2001;1550:6-19.

- [151] Teeuwen R, de Wolf F, Zuilhof H, van Hest J. Elastin-like polypeptides of different molecular weights show independent transition temperatures when mixed. *Soft Matter*. 2009;4305-10.
- [152] Wright E, McMillan R, Cooper A, Apkarian R, Conticello V. Thermoplastic elastomer hydrogels via self-assembly of an elastin-mimetic triblock polypeptide. *Advanced Functional Materials*. 2002;12:149.
- [153] Machado R, Ribeiro A, Padrão J, Silva D, Nobre A, Teixeira J, et al. Exploiting the sequence of naturally occurring elastin: construction, production and characterization of a recombinant thermoplastic protein-based polymer. *Journal of Nano Research*. 2009;6:133-45.
- [154] Martín L, Arias F, Alonso M, García-Arévalo C, Rodríguez-Cabello J. Rapid micropatterning by temperature-triggered reversible gelation of a recombinant smart elastin-like tetrablock-copolymer. *Soft Matter*. 2010;6:1121-4.
- [155] Teeuwen R, Berkel S, Dulmen T, Schoffelen S, Meeuwissen S, Zuilhof H, et al. "Clickable" elastins: elastin-like polypeptides functionalized with azide or alkyne groups. *Chemical Communications*. 2009;2009:4022-4.
- [156] Scheller J, Guhrs K, Grosse F, Conrad U. Production of spider silk proteins in tobacco and potato. *Nature biotechnology*. 2001;19:573-7.
- [157] Adair G, Davis H, Partridge S. A soluble protein derived from elastin. *Nature*. 1951;167.
- [158] Muiznieks L, Weiss A. Flexibility in the Solution Structure of Human Tropoelastin. *Biochemistry*. 2007;46:8196-205.
- [159] Muiznieks L, Weiss A, Keeley F. Structural disorder and dynamics of elastin. *Biochem Cell Biol*. 88:239-50.
- [160] Pepe A, Guerra D, Bochicchio B, Quaglino D, Gheduzzi D, Pasquali Ronchetti I, et al. Dissection of human tropoelastin: supramolecular organization of polypeptide sequences coded by particular exons. *Matrix Biology*. 2005;24:96-109.
- [161] Li B, Alonso D, Daggett V. The molecular basis for the inverse temperature transition of elastin. *Journal of molecular biology*. 2001;305:581-92.

- [162] Cook W, Einspahr H, Trapane T, Urry D, Bugg C. Crystal structure and conformation of the cyclic trimer of a repeat pentapeptide of elastin, cyclo-(L-valyl-L-prolylglycyl-L-valylglycyl) 3. *Journal of the American Chemical Society*. 1980;102:5502-5.
- [163] Miao M, Bellingham C, Stahl R, Sitarz E, Lane C, Keeley F. Sequence and structure determinants for the self-aggregation of recombinant polypeptides modeled after human elastin. *Journal of Biological Chemistry*. 2003;278:48553.
- [164] Miao M, Cirulis J, Lee S, Keeley F. Structural Determinants of Cross-linking and Hydrophobic Domains for Self-Assembly of Elastin-like Polypeptides. *Biochemistry*. 2005;44:14367-75.
- [165] Rodríguez-Cabello J, Reguera J, Girotti A, Alonso M, Testera A. Developing functionality in elastin-like polymers by increasing their molecular complexity: the power of the genetic engineering approach. *Progress in Polymer Science*. 2005;30:1119-45.
- [166] Tu Y, Weiss A. Transient tropoelastin nanoparticles are early-stage intermediates in the coacervation of human tropoelastin whose aggregation is facilitated by heparan sulfate and heparin decasaccharides. *Matrix Biology*. 2009.
- [167] Pepe A, Flamia R, Guerra D, Quaglino D, Bochicchio B, Pasquali Ronchetti I, et al. Exon 26-coded polypeptide: An isolated hydrophobic domain of human tropoelastin able to self-assemble in vitro. *Matrix Biology*. 2008;27:441-50.
- [168] Ostuni A, Bochicchio B, Armentano M, Bisaccia F, Tamburro A. Molecular and supramolecular structural studies on human tropoelastin sequences. *Biophysical Journal*. 2007;93:3640-51.
- [169] Bellingham C, Lillie M, Gosline J, Wright G, Starcher B, Bailey A, et al. Recombinant human elastin polypeptides self-assemble into biomaterials with elastin-like properties. *Biopolymers*. 2003;70:445-55.
- [170] Cirulis J, Keeley F, James D. Viscoelastic properties and gelation of an elastin-like polypeptide. *Journal of Rheology*. 2009;53:1215.
- [171] Meyer D, Chilkoti A. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules*. 2004;5:846-51.
- [172] Rauscher S, Baud S, Miao M, Keeley F, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*. 2006;14:1667-76.

- [173] Urry D. The change in Gibbs free energy for hydrophobic association:: Derivation and evaluation by means of inverse temperature transitions. *Chemical Physics Letters*. 2004;399:177-83.
- [174] Volpin D, Urry D, Pasquali-Ronchetti I, Gotte L. Studies by electron microscopy on the structure of coacervates of synthetic polypeptides of tropoelastin. *Micron* (1969). 1976;7:193-8.
- [175] Reguera J, Lagaron J, Alonso M, Reboto V, Calvo B, Rodriguez-Cabello J. Thermal behavior and kinetic analysis of the chain unfolding and refolding and of the concomitant nonpolar solvation and desolvation of two elastin-like polymers. *Macromolecules*. 2003;36:8470-6.
- [176] Herrero-Vanrell R, Rincon A, Alonso M, Reboto V, Molina-Martinez I, Rodriguez-Cabello J. Self-assembled particles of an elastin-like polymer as vehicles for controlled drug release. *Journal of Controlled Release*. 2005;102:113-22.
- [177] Bessa P, Machado R, Nürnberger S, Dopler D, Banerjee A, Cunha A, et al. Thermoresponsive self-assembled elastin-based nanoparticles for delivery of BMPs. *Journal of Controlled Release*. 2009.
- [178] Meyer D, Kong G, Dewhirst M, Zalutsky M, Chilkoti A. Targeting a genetically engineered elastin-like polypeptide to solid tumors by local hyperthermia. *Cancer Research*. 2001;61:1548.
- [179] Toonkool P, Jensen S, Maxwell A, Weiss A. Hydrophobic domains of human tropoelastin interact in a context-dependent manner. *Journal of Biological Chemistry*. 2001;276:44575.
- [180] Urry D, Urry K, Szaflarski W, Nowicki M. Elastic-contractile model proteins: Physical chemistry, protein function and drug design and delivery. *Advanced drug delivery reviews*. 2010.
- [181] Lyons RE, Wong DC, Kim M, Lekiuffre N, Huson MG, Vuocolo T, et al. Molecular and functional characterisation of resilin across three insect orders. *Insect biochemistry and molecular biology*. 2011;41:881-90.
- [182] Nairn KM, Lyons RE, Mulder RJ, Mudie ST, Cookson DJ, Lesieur E, et al. A synthetic resilin is largely unstructured. *Biophysical Journal*. 2008;95:3358-65.

- [183] Lyons RE, Lesieur E, Kim M, Wong DC, Huson MG, Nairn KM, et al. Design and facile production of recombinant resilin-like polypeptides: gene construction and a rapid protein purification method. *Protein Engineering Design and Selection*. 2007;20:25-32.
- [184] Dutta NK, Truong MY, Mayavan S, Roy Choudhury N, Elvin CM, Kim M, et al. A Genetically Engineered Protein Responsive to Multiple Stimuli. *Angewandte Chemie International Edition*. 2011;50:4428-31.
- [185] Charati MB, Ifkovits JL, Burdick JA, Linhardt JG, Kiick KL. Hydrophilic elastomeric biomaterials based on resilin-like polypeptides. *Soft Matter*. 2009;5:3412-6.
- [186] Lyons RE, Elvin CM, Taylor K, Lekieffre N, Ramshaw JA. Purification of recombinant protein by cold-coacervation of fusion constructs incorporating resilin-inspired polypeptides. *Biotechnology and bioengineering*. 2012.
- [187] Muiznieks LD, Keeley FW. Proline periodicity modulates the self-assembly properties of elastin-like polypeptides. *Journal of Biological Chemistry*. 2010;285:39779.
- [188] Miao M, Bellingham CM, Stahl RJ, Sitarz EE, Lane CJ, Keeley FW. Sequence and structure determinants for the self-aggregation of recombinant polypeptides modeled after human elastin. *Journal of Biological Chemistry*. 2003;278:48553-62.
- [189] Online MaaaaasmoS.
- [190] Miao M, Stahl RJ, Petersen LF, Reintsch WE, Davis EC, Keeley FW. Characterization of an unusual tropoelastin with truncated C-terminus in the frog. *Matrix Biology*. 2009;28:432-41.
- [191] Maeda Y, Nakamura T, Ikeda I. Changes in the hydration states of poly (N-n-propylmethacrylamide) and poly (N-isopropylmethacrylamide) during their phase transitions in water observed by FTIR spectroscopy. *Macromolecules*. 2001;34:8246-51.
- [192] Konák C, Oupický D, Chytrý V, Ulbrich K, Helmstedt M. Thermally controlled association in aqueous solutions of diblock copolymers of poly [N-(2-hydroxypropyl) methacrylamide] and poly (N-isopropylacrylamide). *Macromolecules*. 2000;33:5318-20.
- [193] Fujishige S, Kubota K, Ando I. Phase transition of aqueous solutions of poly (N-isopropylacrylamide) and poly (N-isopropylmethacrylamide). *The Journal of Physical Chemistry*. 1989;93:3311-3.
- [194] Jeong B, Kim SW, Bae YH. Thermosensitive sol–gel reversible hydrogels. *Advanced drug delivery reviews*. 2012;64:154-62.

- [195] Sabaté R, Espargarç A, de Groot NS, Valle-Delgado JJ, Fernández-Busquets X, Ventura S. The role of protein sequence and amino acid composition in amyloid formation: scrambling and backward reading of IAPP amyloid fibrils. *Journal of molecular biology*. 2010;404:337-52.
- [196] Dai M, Haghpanah J, Singh N, Roth EW, Liang A, Tu RS, et al. Artificial Protein Block Polymer Libraries Bearing Two SADs: Effects of Elastin Domain Repeats. *Biomacromolecules*. 2011;12:4240-6.
- [197] Rauscher S, Baud S, Miao M, Keeley FW, Pomès R. Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure*. 2006;14:1667-76.
- [198] Dreher MR, Simnick AJ, Fischer K, Smith RJ, Patel A, Schmidt M, et al. Temperature triggered self-assembly of polypeptides into multivalent spherical micelles. *Journal of the American Chemical Society*. 2008;130:687-94.
- [199] Gimona M. Protein linguistics—a grammar for modular protein assembly? *Nature Reviews Molecular Cell Biology*. 2006;7:68-73.
- [200] Loose C, Jensen K, Rigoutsos I, Stephanopoulos G. A linguistic model for the rational design of antimicrobial peptides. *Nature*. 2006;443:867-9.
- [201] Matsushima N, Yoshida H, Kumaki Y, Kamiya M, Tanaka T, Izumi Y, et al. Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. *Current Protein and Peptide Science*. 2008;9:591-610.
- [202] Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell*. 2012;149:753-67.
- [203] Frey S, Richter RP, Görlich D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*. 2006;314:815-7.
- [204] Dugave C, Demange L. Cis-trans isomerization of organic molecules and biomolecules: implications and applications. *Chemical reviews*. 2003;103:2475-532.
- [205] Reimer U, Scherer G, Drewello M, Kruber S, Schutkowski M, Fischer G. Side-chain effects on peptidyl-prolyl cis/trans isomerisation¹. *Journal of molecular biology*. 1998;279:449-60.

- [206] Valiaev A, Lim DW, Oas TG, Chilkoti A, Zauscher S. Force-induced prolyl cis-trans isomerization in elastin-like polypeptides. *Journal of the American Chemical Society*. 2007;129:6491-7.
- [207] Urry DW. The change in Gibbs free energy for hydrophobic association: Derivation and evaluation by means of inverse temperature transitions. *Chemical physics letters*. 2004;399:177-83.
- [208] Quiroz FG. "Smart" Behavior of Non-Canonical Elastin-Like Polypeptides: Duke University; 2011.
- [209] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*. 1982;157:105-32.
- [210] McDaniel JR, MacKay JA, Quiroz FG, Chilkoti A. Recursive Directional Ligation by Plasmid Reconstruction Allows Rapid and Seamless Cloning of Oligomeric Genes. *Biomacromolecules*. 2010;11:944-52.
- [211] Urry DW, Gowda D, Parker TM, Luan CH, Reid MC, Harris CM, et al. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers*. 1992;32:1243-50.
- [212] Hassouneh W, MacEwan SR, Chilkoti A. Fusions of Elastin-Like Polypeptides to Pharmaceutical Proteins. *Methods in Enzymology*. 2012;502:215.
- [213] Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*. 2007;1:2876-90.
- [214] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics*. 2000;41:415-27.
- [215] Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins: Structure, Function, and Bioinformatics*. 2001;42:38-48.
- [216] Manning MC, Woody RW. Theoretical study of the contribution of aromatic side chains to the circular dichroism of basic bovine pancreatic trypsin inhibitor. *Biochemistry*. 1989;28:8609-13.

- [217] Mary P, Bendejacq DD, Labeau M-P, Dupuis P. Reconciling low-and high-salt solution behavior of sulfobetaine polyzwitterions. *The Journal of Physical Chemistry B*. 2007;111:7767-77.
- [218] Chang Y, Chen W-Y, Yandi W, Shih Y-J, Chu W-L, Liu Y-L, et al. Dual-Thermoresponsive Phase Behavior of Blood Compatible Zwitterionic Copolymers Containing Nonionic Poly (N-isopropyl acrylamide). *Biomacromolecules*. 2009;10:2092-100.
- [219] Azzaroni O, Brown AA, Huck WT. UCST Wetting Transitions of Polyzwitterionic Brushes Driven by Self-Association. *Angewandte Chemie*. 2006;118:1802-6.
- [220] Yamakov V, Milchev A, Jörg Limbach H, Dünweg B, Everaers R. Conformations of random polyampholytes. *Physical review letters*. 2000;85:4305-8.
- [221] Ferrari M. Cancer nanotechnology: opportunities and challenges. *Nature Reviews Cancer*. 2005;5:161-71.
- [222] Hrkach J, Von Hoff D, Ali MM, Andrianova E, Auer J, Campbell T, et al. Preclinical Development and Clinical Translation of a PSMA-Targeted Docetaxel Nanoparticle with a Differentiated Pharmacological Profile. *Science Translational Medicine*. 2012;4:128ra39-ra39.
- [223] Torchilin V. Tumor delivery of macromolecular drugs based on the EPR effect. *Advanced drug delivery reviews*. 2011;63:131-5.
- [224] Sugahara KN, Teesalu T, Karmali PP, Kotamraju VR, Agemy L, Greenwald DR, et al. Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. *Science*. 2010;328:1031-5.
- [225] Meng F, Zhong Z, Feijen J. Stimuli-responsive polymersomes for programmed drug delivery. *Biomacromolecules*. 2009;10:197-209.
- [226] Andrew Mackay J, Chilkoti A. Temperature sensitive peptides: Engineering hyperthermia-directed therapeutics. *International Journal of Hyperthermia*. 2008;24:483-95.
- [227] Kong G, Braun RD, Dewhirst MW. Hyperthermia enables tumor-specific nanoparticle delivery: effect of particle size. *Cancer research*. 2000;60:4440.
- [228] Falk M, Issels R. Hyperthermia in oncology. *International Journal of Hyperthermia*. 2001;17:1-18.

- [229] Brassart B, Fuchs P, Huet E, Alix AJP, Wallach J, Tamburro AM, et al. Conformational dependence of collagenase (matrix metalloproteinase-1) up-regulation by elastin peptides in cultured fibroblasts. *Journal of Biological Chemistry*. 2001;276:5222.
- [230] Tjin Tham Sjin RM, Satchi-Fainaro R, Birsner AE, Ramanujam V, Folkman J, Javaherian K. A 27-amino-acid synthetic peptide corresponding to the NH₂-terminal zinc-binding domain of endostatin is responsible for its antitumor activity. *Cancer research*. 2005;65:3656.
- [231] Teesalu T, Sugahara KN, Kotamraju VR, Ruoslahti E. C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *Proceedings of the National Academy of Sciences*. 2009;106:16157-62.
- [232] Amiram M, Luginbuhl KM, Li X, Feinglos MN, Chilkoti A. Injectable protease-operated depots of glucagon-like peptide-1 provide extended and tunable glucose control. *Proceedings of the National Academy of Sciences*. 2013;110:2792-7.
- [233] Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. *Nucleic acids research*. 2008;36:W197-W201.
- [234] Pierschbacher MD, Ruoslahti E. Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule. *Nature*. 1984;309:30.
- [235] Lee BW, Schubert R, Cheung YK, Zannier F, Wei Q, Sacchi D, et al. Strongly Binding Cell-Adhesive Polypeptides of Programmable Valencies. *Angewandte Chemie International Edition*. 2010;49:1971-5.
- [236] 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.
- [237] Iwamoto Y, Robey FA, Graf J, Sasaki M, Kleinman HK, Yamada Y, et al. YIGSR, a synthetic laminin pentapeptide, inhibits experimental metastasis formation. *Science*. 1987;238:1132-4.
- [238] Nomizu M, Yamamura K, Kleinman HK, Yamada Y. Multimeric forms of Tyr-Ile-Gly-Ser-Arg (YIGSR) peptide enhance the inhibition of tumor growth and metastasis. *Cancer research*. 1993;53:3459.

- [239] Sarfati G, Dvir T, Elkabets M, Apte RN, Cohen S. Targeting of polymeric nanoparticles to lung metastases by surface-attachment of YIGSR peptide from laminin. *Biomaterials*. 2011;32:152-61.
- [240] Saiki I, Murata J, Iida J, Nishi N, Sugimura K, Azuma I. The inhibition of murine lung metastasis by synthetic polypeptides [poly (arg-gly-asp) and poly (tyr-ile-gly-ser-arg)] with a core sequence of cell adhesion molecules. *British journal of cancer*. 1989;59:194.
- [241] Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews Drug Discovery*. 2007;6:404-14.
- [242] Bernal F, Wade M, Godes M, Davis TN, Whitehead DG, Kung AL, et al. A stapled p53 helix overcomes HDMX-mediated suppression of p53. *Cancer cell*. 2010;18:411-22.
- [243] Schellenberger V, Wang C-w, Geething NC, Spink BJ, Campbell A, To W, et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nature biotechnology*. 2009;27:1186-90.
- [244] Simnick AJ, Valencia CA, Liu R, Chilkoti A. Morphing low-affinity ligands into high-avidity nanoparticles by thermally triggered self-assembly of a genetically encoded polymer. *ACS nano*. 2010;4:2217-27.
- [245] Harris JM, Chess RB. Effect of pegylation on pharmaceuticals. *Nature Reviews Drug Discovery*. 2003;2:214-21.
- [246] Hubbell JA. Drug development: Longer-lived proteins. *Nature*. 2010;467:1051-2.
- [247] Lien S, Lowman HB. Therapeutic peptides. *Trends in Biotechnology*. 2003;21:556-62.
- [248] Miranda LP, Winters KA, Gegg CV, Patel A, Aral J, Long J, et al. Design and synthesis of conformationally constrained glucagon-like peptide-1 derivatives with increased plasma stability and prolonged in vivo activity. *Journal of medicinal chemistry*. 2008;51:2758-65.
- [249] Webster R, Elliott V, Park BK, Walker D, Hankin M, Taupin P. PEG and PEG conjugates toxicity: towards an understanding of the toxicity of PEG and its relevance to PEGylated biologicals. *PEGylated protein drugs: Basic science and clinical applications*. 2009:127-46.

[250] Gao W, Liu W, Mackay JA, Zalutsky MR, Toone EJ, Chilkoti A. In situ growth of a stoichiometric PEG-like conjugate at a protein's N-terminus with significantly improved pharmacokinetics. *Proceedings of the National Academy of Sciences*. 2009;106:15231-6.

[251] Martín L, Castro E, Ribeiro A, Alonso M, Rodríguez-Cabello JC. Temperature-Triggered Self-Assembly of Elastin-Like Block Co-Recombinamers: The Controlled Formation of Micelles and Vesicles in an Aqueous Medium. *Biomacromolecules*. 2012;13:293-8.

[252] Garanger E, Lecommandoux S. Towards Bioactive Nanovehicles Based on Protein Polymers. *Angewandte Chemie International Edition*. 2012;51:3060-2.

Biography

I was born in Medellin, Colombia on 21 April 1986. I received a B.S in Biomedical Engineering from Escuela de Ingeniería de Antioquia (EIA) and Universidad CES in 2008. This 5 year program — as most undergraduate programs in Colombia — was an incredible experience, both professionally and personally. During those years I became a passionate engineer, a scientist and a religious person, and I had the fortune to meet friends who I love and admire tremendously. There I also met Adriana Villa, who I dated for several years, many of those during my PhD. Her testimony has had a tremendous impact in my life.

Besides the excellent training in medicine, biology and engineering that I received at EIA-CES, I was fortunate to be involved in bone tissue engineering and stem cell research under the mentorship of Prof. Luis Ernesto López and Dr. Daniel Gallego-Pérez. Several scientific publications resulted from that work. I also had the chance to complete a 5 month undergraduate internship with Prof. Michael Mayer in the Biomedical Engineering Department of the University of Michigan.

In 2008, I joined the PhD program in Biomedical Engineering of Duke University to work in the research group of Prof. Ashutosh Chilkoti. I received a M.S degree in Biomedical Engineering from Duke in 2011 for my work on the “smart” behavior of non-canonical elastin-like polypeptides. That work served as the basis for this doctoral dissertation.