Supporting Information for

Engineering the dynamic properties of protein networks through sequence variation

*Lawrence J. Dooling and David A. Tirrell**

Division of Chemistry and Chemical Engineering
California Institute of Technology
Pasadena, CA, USA

Additional Experimental Details

Cloning The EPE variants were generated by site-directed mutagenesis of the sequence encoding the P domain on a pUC19 plasmid (pUC19 P). The primer sequences are listed in Table S1. After sequencing to confirm the correct mutant, the resulting pUC19 P-*mutant* plasmids (pUC19 P T40A, pUC19 P Q54A, pUC19 P I58A, pUC19 P L37A, pUC19 P L37V, pUC19 P L37I) were digested with SacI and SpeI (New England BioLabs, Ipswich, MA) to isolate the fragments encoding P-*mutants*. Note that the residue numbering convention is based on the position of the residues in rat cartilage oligomeric matrix protein from which the P domain is derived.[1] As in previous artificial proteins containing P, Cys68 and Cys71 were mutated to serine to prevent covalent cross-linking in the coiled-coil domain.[2] The fragments encoding P-*mutants* were then ligated into pQE-80L EPE (ref. [3]) digested with SacI and SpeI. This replaces the original P domain with the mutated variant. Site-directed mutagenesis was not carried out directly on pQE-80L EPE due to difficulties created by the highly repetitive, GC-rich elastin-like domains. Chemically competent BL21 *E. coli* (New England BioLabs) were transformed with the pQE-80L EPE-*mutant* plasmids (pQE-80L EPE T40A, pQE-80L EPE Q54A, pQE-80L EPE I58A, pQE-80L EPE L37A, pQE-80L EPE L37V, pQE-80L EPE L37I) for protein expression. The amino acid sequences of the proteins are listed in Table S2.

The pQE-80L EAE plasmid was constructed by subcloning the sequence encoding the A domain from pQE-9 PC10A (ref. [2]) into the pQE-80L EPE plasmid in place of the sequence encoding the P domain. The forward and reverse primers used to amplify A are listed in Table S1. The forward primer contained a SacI overhang while the reverse primer bound the plasmid downstream of an in-frame SpeI site flanking the A domain coding sequence in pQE-9 PC10A.

The amplicon was digested with these enzymes and the restriction fragment was ligated into the pQE-80L EPE plasmid in place of P. The BL21 strain was also used for expression of EAE.

Protein characterization Fractions were saved from each step in the temperature cycling purification. Proteins were extracted from the pelleted fractions with a volume of 8 M urea equal to the volume of the supernatant from that step. These solutions were then diluted with water to adjust the final urea concentration to 4 M. Samples of the supernatant were diluted with 8 M urea to obtain a final urea concentration of 4 M. All of the fractions were then mixed with 2x SDS loading buffer containing 5% (v/v) βME and boiled for 10 min before loading 2 μL per well in a 15 well, 4-12% NuPage Bis-Tris SDS-PAGE gel (Thermo Fisher Scientific, Waltham, MA). The gel was run in MES-SDS running buffer (Boston BioProducts, Ashland, MA) for 45 min at 180 V, fixed, and stained with colloidal blue protein staining solution (Life Technologies, Carlsbad, CA). Gels were imaged on a Typhoon Trio (GE Healthcare, Marlborough, MA).

Protein solutions (0.2 mg mL^{-1} in 0.1% (v/v) formic acid) were analyzed by LC-MS using a Waters (Milford, MA) UPLC/LCT Premier XE TOF mass spectrometer by electrospray ionization in the positive ion mode with a MassPREP Micro desalting column. The mass spectrometer was calibrated with NaI using standard procedures and calibration was subsequently verified by running a standard solution of myoglobin. The mass spectrometer settings were: capillary voltage = 2.8 kV, cone voltage = 40, source temperature = 120 °C, and desolvation temperature = 350 °C, desolvation gas = 750 L hr^{-1}, acquisition range = 500 to 2000 in V mode, ion guide = 5. The mobile phase consisted of a gradient of water and acetonitrile with 0.1% formic acid. Electrospray mass spectra were deconvoluted using MaxEnt1 software (Waters).

The free thiol content of each protein was measured using Ellman's assay.[4] Lyophilized proteins were dissolved in reaction buffer (100 mM sodium phosphate, 1 mM EDTA, pH 8) at a

concentration of 5 mg mL^{-1}. Upon mixing 250 μL of the protein solutions and 50 μL of Ellman's reagent (Sigma) stock solution (5 mg mL^{-1} in reaction buffer) in 2.5 mL of reaction buffer, a yellow product formed. After 15 min incubation, the absorbance at 412 nm was measured on a Cary50 UV/Vis spectrophotometer (Agilent, Carlsbad, CA). The concentration of thiol groups was calculated from the absorbance value and the extinction coefficient 14,150 M^{-1} cm^{-1}.[5]

The free thiol content was also estimated by nonreducing SDS-PAGE. Samples prepared for thiol quantitation by Ellman's assay were diluted 1:10 in SDS loading buffer, and 2 μL of each solution was loaded in a 15-well, 4-12% NuPage Bis-Tris SDS-PAGE gel. Control lanes contained samples that were reduced by boiling in the presence of 5% (v/v) βME for 5 min. The gel was run in MES-SDS running buffer at 180 V for 45 min. Proteins were visualized with InstantBlue protein stain (Expedion, San Diego, CA). The intensity of each band on the protein gel was quantified using ImageQuant software (GE Healthcare).

Hydrogel swelling measurements For swelling experiments, hydrogels were transferred to 6-well plates containing 3 mL of PBS plus 6 M guanidinium chloride per well. The gels were swollen to equilibrium and the swollen mass was measured after 48 hr. The guanidinium chloride concentration was gradually decreased from 6 M to 0 M in PBS as described in the Materials and Methods section of the main text. Swollen masses were recorded for gels in PBS after 48 hr of swelling. After measuring the swollen mass in PBS, the gels were washed at least five times with ddH_2O over the course of 72 hr to remove salts. The dry mass was obtained by lyophilizing the gels to remove water. The mass swelling ratio Q_m was calculated for gels in PBS plus 6 M guanidinium chloride and for gels in PBS by dividing the swollen mass for each condition by the dry mass of the network.

Table S1. Oligonucleotides for mutagenesis of pUC19 P and subcloning of A. For oligonucleotides used in site-directed mutagenesis, (+) represents the sense strand and (-) represents the antisense strain. Codons for the mutated residue are in bold, underlined text.

P variant (strand)	**Primer Sequence** (5'to 3')
T40A (+)	GAA CTG CAG GAA **GCC** AAT GCC GCG C
T40A (-)	G CGC GGC ATT **GGC** TTC CTG CAG TTC
Q54A (+)	GAA TTG CTT CGT CAA **GCG** GTC AAG GAG ATA AC
Q54A (-)	GT TAT CTC CTT GAC **CGC** TTG ACG AAG CAA TTC
I58A (+)	CAG GTC AAG GAG **GCA** ACG TTC TTG AAG
I58A (-)	CTT CAA GAA CGT **TGC** CTC CTT GAC CTG
L37A (+)	CAA ATG CTG CGT GAA **GCG** CAG GAA ACC AAT GCC
L37A (-)	GGC ATT GGT TTC CTG **CGC** TTC ACG CAG CAT TTG
L37V (+)	CAA ATG CTG CGT GAA **GTG** CAG GAA ACC AAT GCC
L37V (-)	GGC ATT GGT TTC CTG **CAC** TTC ACG CAG CAT TTG
L37I (+)	CAA ATG CTG CGT GAA **ATT** CAG GAA ACC AAT GCC
L37I (-)	GGC ATT GGT TTC CTG **AAT** TTC ACG CAG CAT TTG
A (forward)	ATT GCA *GAG CTC* ATG CCG ACT AGC GGT GAC CTG (SacI)
A (reverse)	CTT GGC TGC AGG TCG ACG G

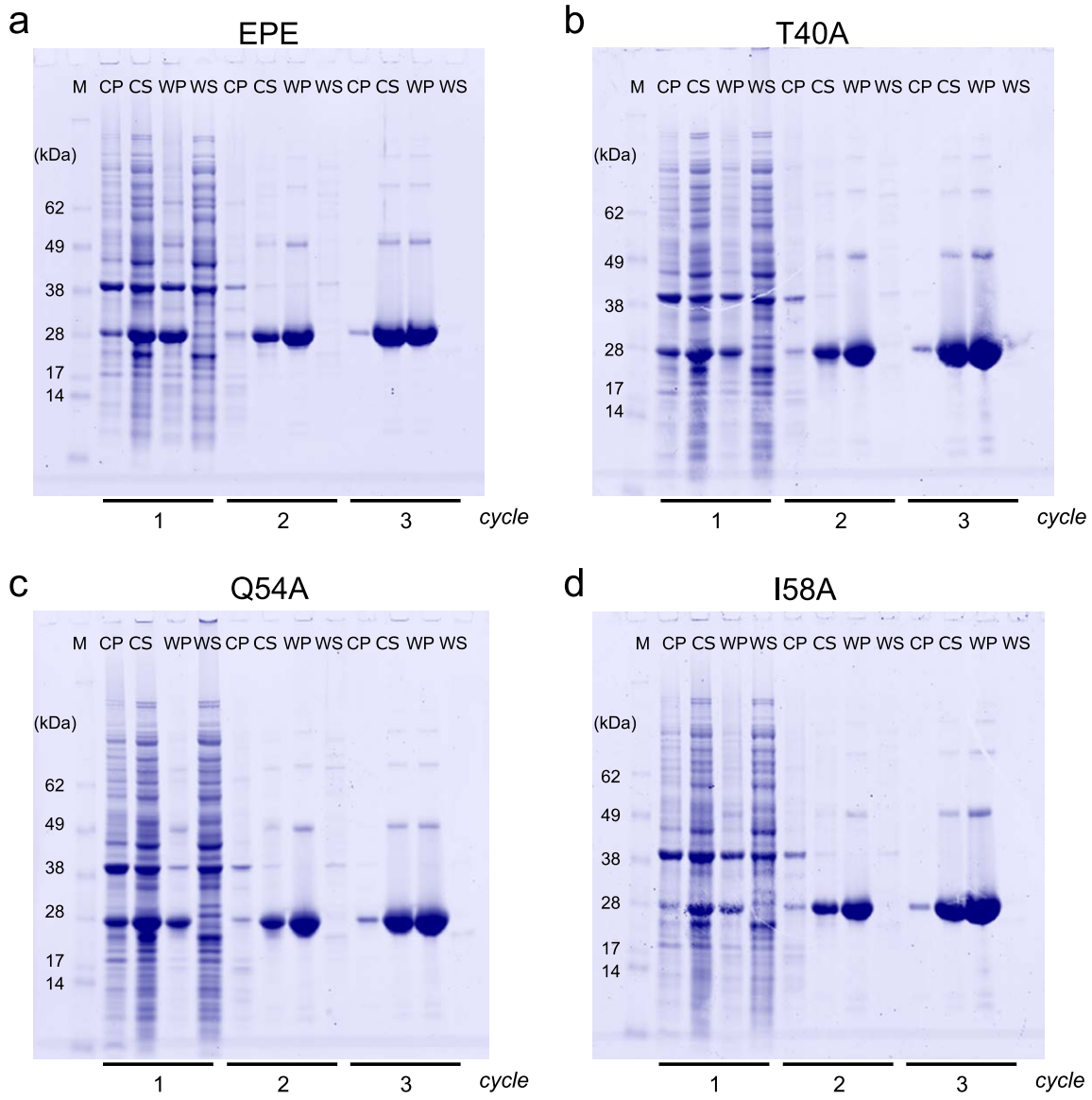
Tables S2. Amino acid sequences of EPE, EPE variants, and EAE.

EPE variant	**Protein Sequence** (mutations bolded and underlined)
EPE	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRELQETNAALQDVRELLRQQVKEITFLKNTVMESDA SKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVP GVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
T40A	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRELQE**A**NAALQDVRELLRQQVKEITFLKNTVMESD ASKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEG VPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
Q54A	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRELQETNAALQDVRELLRQ**A**VKEITFLKNTVMESDA SKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVP GVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
I58A	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRELQETNAALQDVRELLRQQVKE**A**TFLKNTVMESD ASKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEG VPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
L37A	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRE**A**QETNAAAQDVRELLRQQVKEITFLKNTVMESD ASKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEG VPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
L37V	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRE**V**QETNAAAQDVRELLRQQVKEITFLKNTVMESD ASKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEG VPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM

L37I	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELGSGLGSAPQMLRE**I**QETNAAAQDVRELLRQQVKEITFLKNTVMESDA SKLNTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVP GVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM
EAE	MRCSSHHHHHHVDGHGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPG VGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVG ELMPTSGDLENEVAQLEREVRSLEDEAAELEQKVSRLKNEIEDLKAE IGDHVAPRDTSVPGVGVPGVGVPGEGVPGVGVPGVGVPGVGVPGVGVP GEGVPGVGVPGVGVPGVGVPGVGVPGEGVPGVGVPGVGVPGGLLECM

Figure S1. SDS-PAGE analysis of inverse temperature cycling. Samples of each fraction were saved throughout the inverse temperature cycling purification. (**a-h)** EPE, T40A, Q54A, I58A, L37A, L37V, L37I, and EAE. Elastin-like proteins are soluble in the cold step of each cycle (4 °C, low ionic strength) and can be sedimented in the warm step of each cycle (37 °C, 2 M NaCl). Pure proteins were obtained after three cycles. The gels also reveal that the target proteins were not completely solubilized in the first cold step, but the yields were high enough that the insoluble fraction could be discarded.

Key:
CP: cold pellet WP: warm pellet
CS: cold supernatant WS: warm supernatant
M: SeeBlue protein molecular weight marker (Life Technologies)



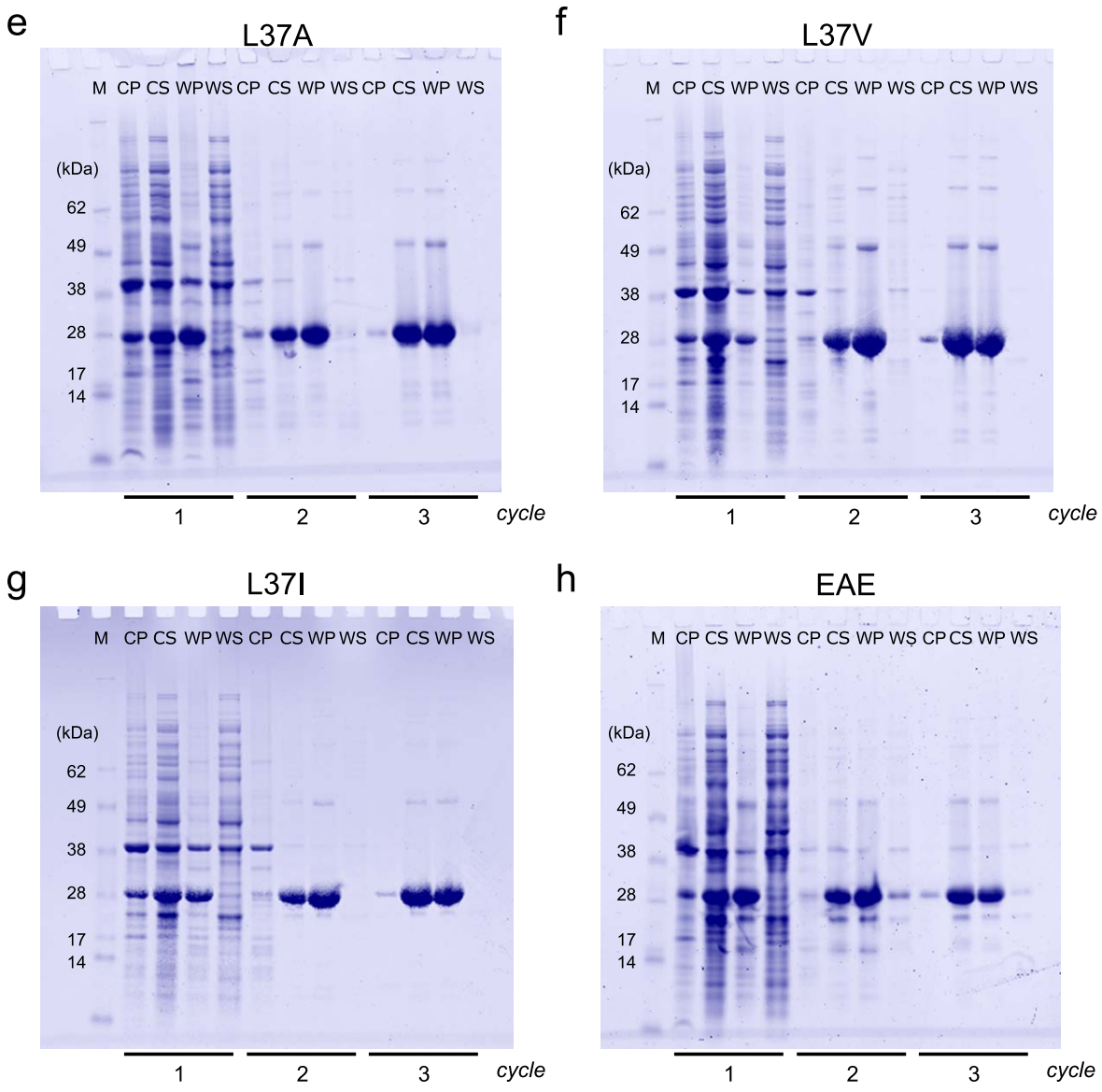

e
L37A
M CP CS WP WS CP CS WP WS CP CS WP WS
(kDa)
62
49
38
28
17
14
1
2
3
cycle
f
L37V
M CP CS WP WS CP CS WP WS CP CS WP WS
(kDa)
62
49
38
28
17
14
1
2
3
cycle
g
L37I
M CP CS WP WS CP CS WP WS CP CS WP WS
(kDa)
62
49
38
28
17
14
1
2
3
cycle
h
EAE
M CP CS WP WS CP CS WP WS CP CS WP WS
(kDa)
62
49
38
28
17
14
1
2
3
cycle

Figure S2. Ellman's assay quantitation of the free thiol content in artificial proteins. The fraction of free thiol groups was calculated by dividing the concentration of thiols measured by Ellman's assay by the expected concentration of thiols assuming that each protein chain contains two cysteine residues. The values all fall between 83% and 91%. Analysis of the proteins by nonreducing SDS-PAGE (Figure S4 and Table S3, below) suggests that deviations from the expected concentration of free thiols arise from the formation of intermolecular disulfides (dimers, trimers, etc.) and intramolecular disulfides (cyclized monomers and higher order species).

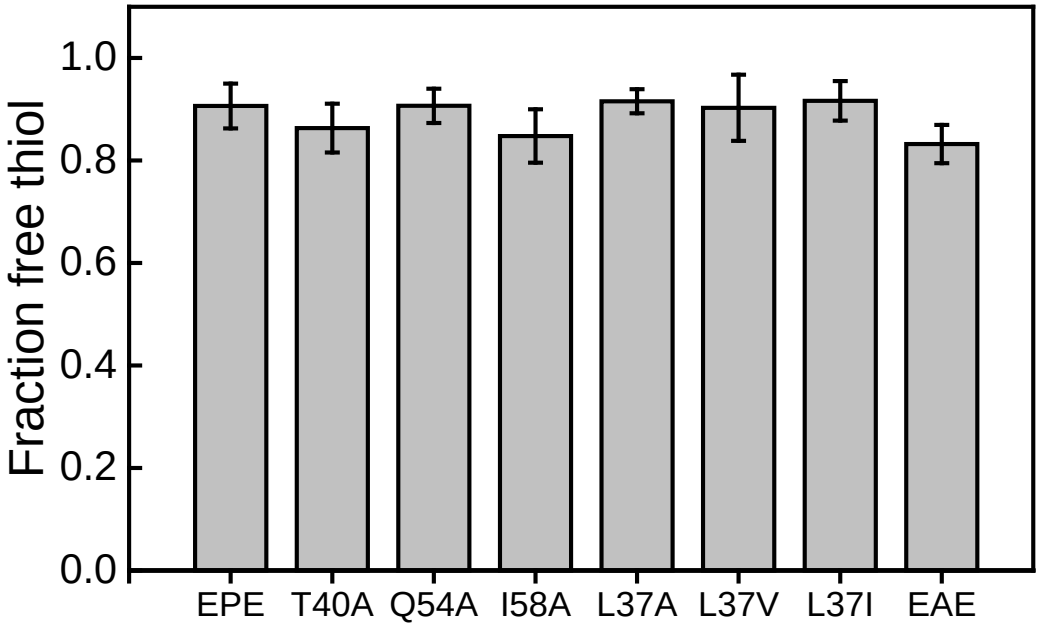


Figure S3. Analysis of protein oligomerization state by nonreducing SDS-PAGE for EPE (**a**), EPE variants (**b-g**), and EAE (**h**). Each gel image contains three lanes. Lane 1: SeeBlue protein molecular weight marker with the molecular weight of select bands labeled. Lane 2: artificial protein prepared under nonreducing conditions, denoted as "–" βME. Lane 3: artificial protein prepared under reducing conditions by boiling the sample in 5% (v/v) βME, denoted as "+" βME. In (a) and (h), the bands assigned as protein trimers (3°), dimers (2°), linear monomers (1°-ℓ), and cyclized monomers (1°-c) are labeled on the right-hand side of the gel. All artificial proteins are predominantly monomeric and linear, although in general more higher-order oligomers and cyclized monomers are present under nonreducing conditions (−) than in the reduced control lanes (+).

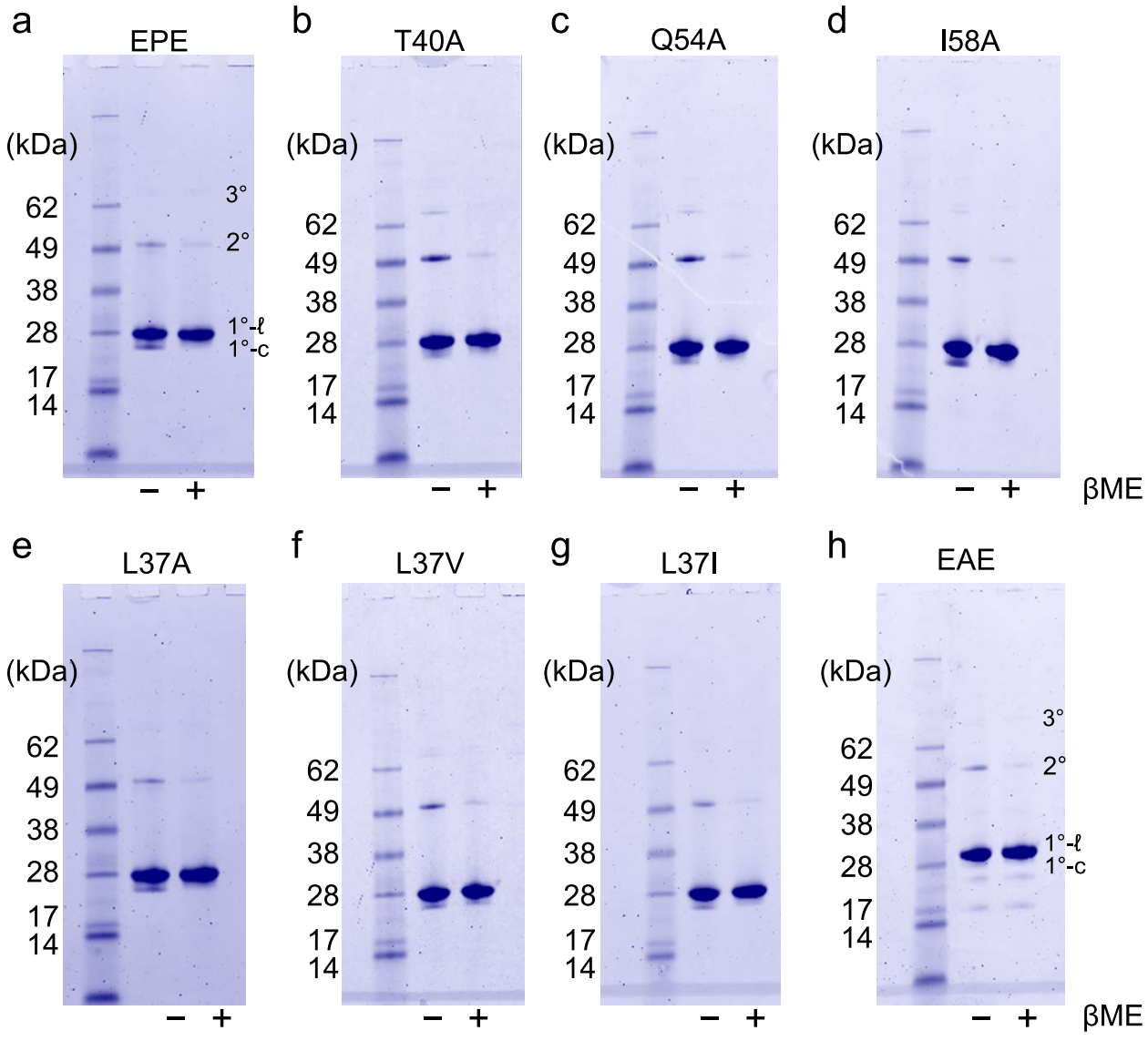


Table S3. Gel densitometry of protein bands in nonreducing SDS-PAGE. Lane profiles were created for the nonreducing sample lanes of the gel images in Figure S4. The intensities of bands assigned as linear monomers, cyclic monomers, dimers, and trimers in Figure S4 were quantified by integrating the peak corresponding to each species. The data are reported as the percentage of the total area of all peaks detected in the lane.

Protein	**EPE**	**T40A**	**Q54A**	**I58A**	**L37A**	**L37V**	**L37I**	**EAE[a]**
Trimer	< 1	2	2	1	< 1	1	< 1	< 1
Dimer	6	15	14	11	5	9	7	7
Monomer (linear)	83	81	80	75	87	85	86	84
Monomer (cyclic)	9	2	5	12	8	5	7	4

[a] The EAE protein preparation has two small impurities that are likely degradation products. The integrated intensities of these bands on the SDS-PAGE gel were 3% and 2% of the total.

Table S4. Percent free thiol by Ellman's assay and by nonreducing SDS-PAGE. The percent free thiol measured by Ellman's assay is compared to the percent free thiol calculated from the relative amount of each band in the nonreducing SDS-PAGE gels (Figure S4 and Table S3). The data in Table S3 are multiplied by the expected number of free thiols per protein chain for each species (2 for linear monomers, 0 for cyclized monomers, 1 for dimers, and 2/3 for trimers). This value is then divided by 2, the expected number of thiols per chain if all proteins were in the reduced form.

Protein	**EPE**	**T40A**	**Q54A**	**I58A**	**L37A**	**L37V**	**L37I**	**EAE**
Ellman's assay	91	86	91	85	92	90	92	83
SDS-PAGE	86	89	87	81	90	90	90	88

Table S5. Protein mass determination by ESI-MS. Proteins were analyzed by LC-MS with electrospray ionization. The deconvoluted masses were all within 0.02% of the masses calculated from the protein sequence.

EPE variant	Calculated mass	Observed mass
EPE	21 464	21 462
T40A	21 434	21 434
Q54A	21 407	21 406
I58A	21 422	21 421
L37A	21 422	21 423
L37V	21 450	21 449
L37I	21 464	21 462
EAE	21 908	21 909

Figure S4. Mass swelling ratio. Hydrogels swollen for 48 hours in PBS buffer (n = 6, avg. ± s.d.).

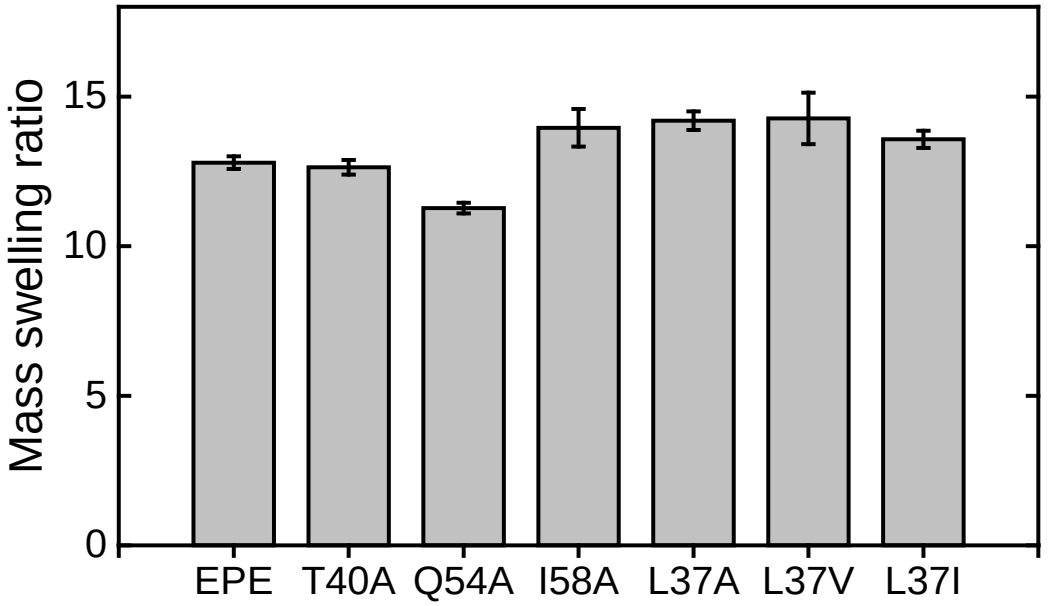


Figure S5. Rheology of L37V gels at different temperatures. Dynamic oscillatory frequency sweeps showing *G'* and *G"* at 2% strain amplitude and 5, 15, 25, and 35 °C. The behavior is qualitatively similar, though relaxation occurs faster with increasing temperature.

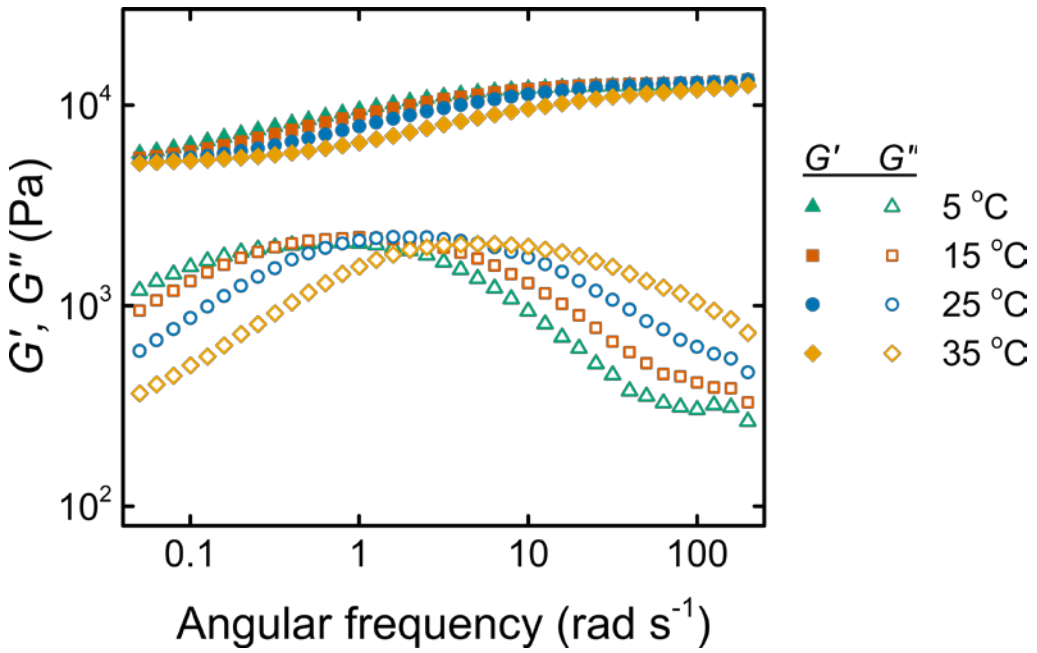


Figure S6. Rheology and swelling of EPE and EPE variants under denaturing conditions. (**a**) Hydrogels were swollen in PBS, pH 7.4 with 6 M guanidinium chloride. In frequency sweeps at 2% strain amplitude, 25 °C, the storage moduli *G'* are nearly independent of the oscillation frequency and the loss moduli *G"* do not exhibit different local maxima. The rise in *G"* at low frequency (0.001-0.01 rad s^{-1}) may be due to slip. (**b**) Hydrogels swollen in PBS containing 6 M guanidinium chloride exhibit similar mass swelling ratios, with the exception of Q54A, which is slightly less swollen than the other gels (n = 6, avg. ± s.d.).

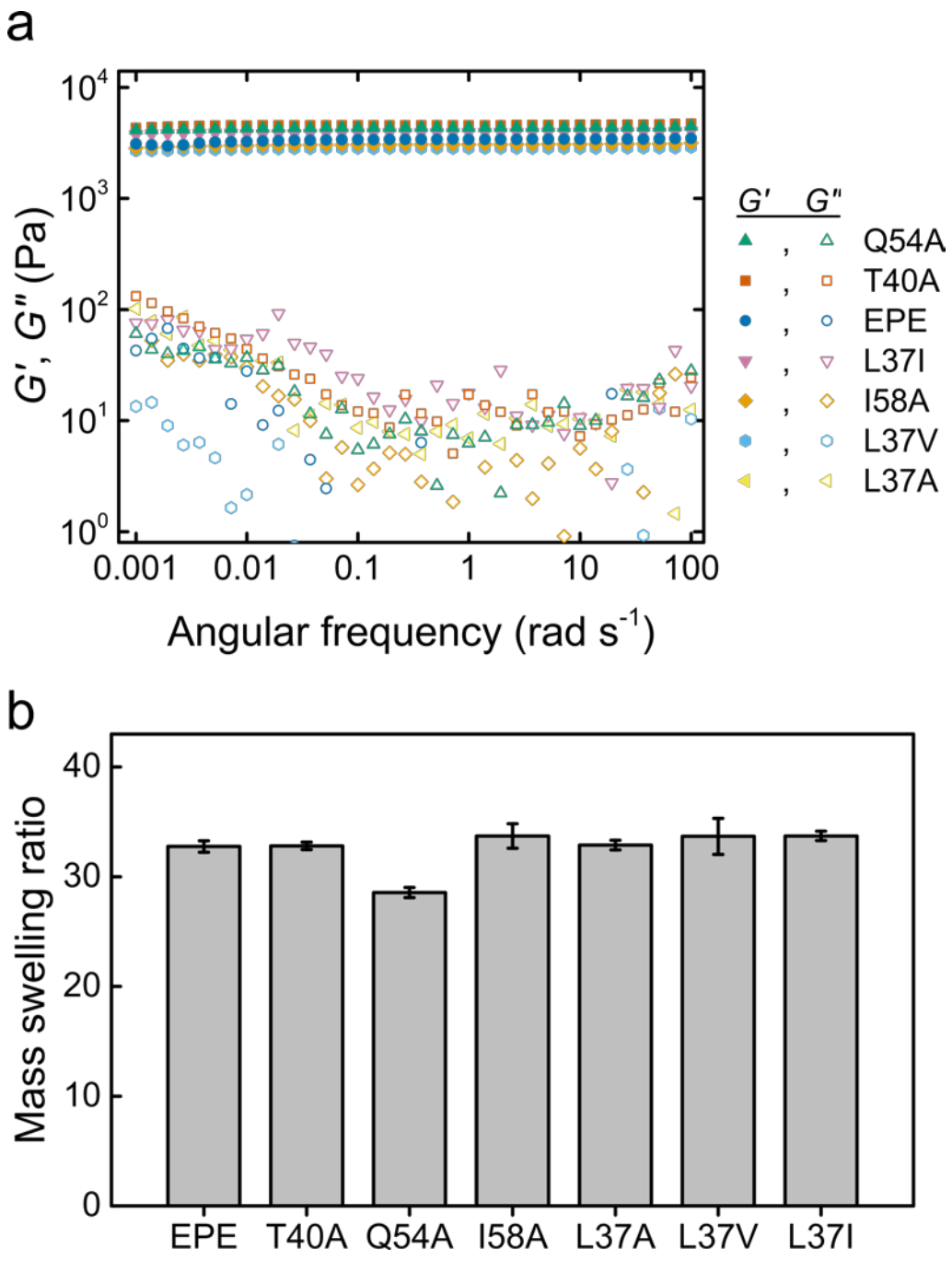


Figure S7. Stress relaxation of L37V:EAE and L37V:Q54A mixed composition gels. (**a**) The relaxation function $G(t)$ is plotted for single protein networks L37V and EAE and the mixed protein network L37V:EAE. (**b**) $G(t)$ was also plotted for the Q54A single protein network and the L37V:Q54A mixed protein network. The solid lines are fits of the stretched exponential model (eq. 1) for single protein networks and a double stretched exponential model (eq. 3) for mixed protein networks. $G(t)$ is well fit by the double stretched exponential model for the L37V:EAE network, which contains an orthogonal pair of physical cross-linking domains, but not for the L37V:Q54A.

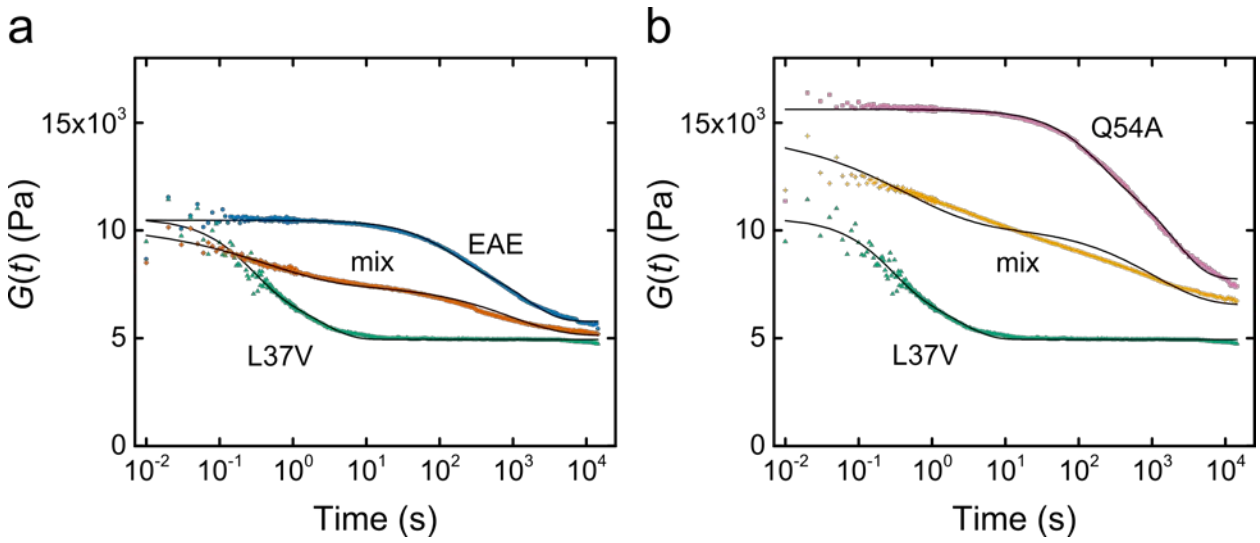
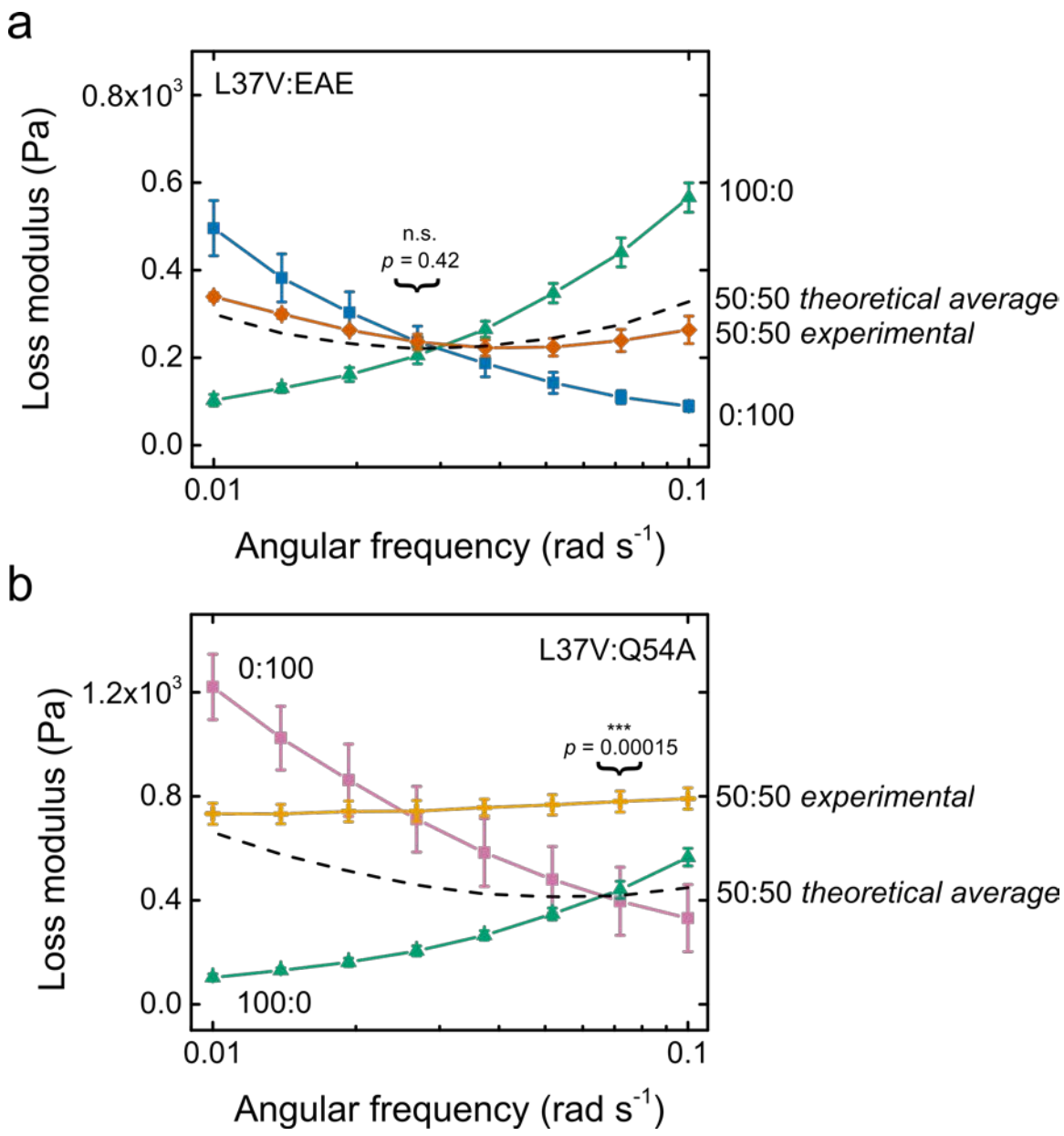


Figure S8. Loss moduli intersection analysis. (**a**) L37V, EAE, and L37V:EAE gels. (**b**) L37V, Q54A, L37V:Q54A gels (n = 3, avg. ± s.d.). The theoretical average of the curves for the single protein networks is plotted as the dashed black line. Significance testing (*vide infra*) was performed to compare the theoretical average of *G"* from the single protein networks and the experimental value of the loss modulus at the point nearest the predicted local minimum (i.e. the intersection of the *G"* curves from the single protein networks; ω = 0.027 rad s^{-1} for L37V and EAE and ω = 0.072 rad s^{-1} for L37V and Q54A).



Statistical test comparing loss moduli in mixed protein networks

The loss modulus of the L37V:EAE network closely matches the theoretical average of the loss moduli of the L37V and EAE single protein networks, and the intersection of the L37V and EAE *G"* curves coincides with a local minimum in *G"* of the mixed protein network. This is not the case for the non-orthogonal protein pair, L37V and Q54A; the loss modulus does not agree well with the theoretical average of the single protein networks. In particular, the local minimum predicted by the theoretical average to occur at the intersection of the L37V and Q54A loss moduli is not present in the experimental data. Instead, *G"* is greater than the theoretical average indicating that more energy is being dissipated at this angular frequency than would be expected if the L37V and Q54A cross-links acted independently. To determine whether the difference in the measured and calculated values of *G"* at the predicted local minimum is statistically significant, we performed hypothesis testing on a linear combination of the means[6] of *G"* at this point.

The mean values of *G"* for the L37V:EAE, L37V, and EAE networks are μ_m, μ_V, and μ_A, respectively. It is assumed that the experimental values of *G"* were sampled from populations that are normally distributed around these means with equal variances. Let γ be the linear combination of these means,

$$\gamma = m_m\mu_m + m_V\mu_V + m_A\mu_A$$

with constants $m_m = 1$, $m_V = -0.5$, and $m_A = -0.5$. In other words,

$$\gamma = \mu_m - 0.5(\mu_V + \mu_A),$$

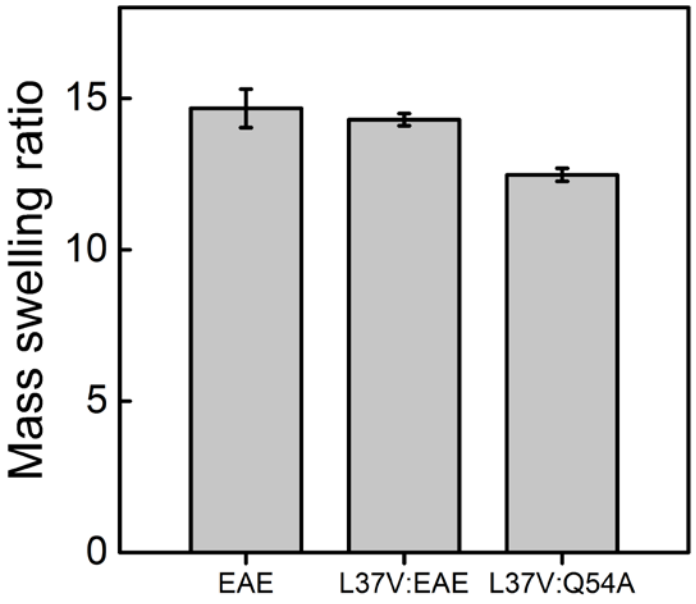
the difference between the mean of the mixed protein network and the average value of the means of the single protein networks. The null hypothesis, H_0, is $\gamma = 0$. The alternative hypothesis, H_1, is $\gamma \neq 0$.

The experimental data were used to compute the average values $\langle G_i'' \rangle$ of the loss modulus at the data point nearest the predicted minimum and the standard deviation (s.d.) for each set i = 1, 2, 3. The total number of sets is $I = 3$. Each set i has three measurements $n_i = 3$ for a total of n = 9 measurements.

Let $c = \sum_{i=1}^{3} m_i \langle G_i'' \rangle$ and $s_c = s_p \sqrt{\sum_{i=1}^{3} \left(\frac{m_i}{n_i} \right)^2}$ where s_p is the pooled standard deviation. The t-test statistic is $t = \frac{c - \gamma}{s_c}$ and there are n - I = 6 degrees of freedom.

For the L37V:EAE network, t = 0.87 and p = 0.418 from the two-tailed distribution. Therefore, the null hypothesis was not rejected. For the same analysis of the L37V:Q54A network, t = 8.48 and p = 0.00015 from the two-tailed distribution. Therefore, the null hypothesis was rejected.

Figure S9. Mass swelling ratio of EAE, L37V:EAE, and L37V:Q54A hydrogels. Hydrogels were swollen for 48 hours in PBS buffer (n = 6, avg. ± s.d.).



References

1. Efimov, V. P.; Lustig, A.; Engel, J., The thrombospondin-like chains of cartilage oligomeric matrix protein are assembled by a five-stranded α-helical bundle between residues 20 and 83. *FEBS Lett.* **1994,** *341*, 54-58.

2. Shen, W.; Zhang, K.; Kornfield, J. A.; Tirrell, D. A., Tuning the erosion rate of artificial protein hydrogels through control of network topology. *Nat. Mater.* **2006,** *5*, 153-158.

3. Dooling, L. J.; Buck, M. E.; Zhang, W.-B.; Tirrell, D. A., Programming molecular association and viscoelastic behavior in protein networks. *Adv. Mater.* **2016,** *28*, 4651-4657.

4. Ellman, G. L., Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959,** *82*, 70-77.

5. Eyer, P.; Worek, F.; Kiderlen, D.; Sinko, G.; Stuglin, A.; Simeon-Rudolf, V.; Reiner, E., Molar absorption coefficients for the reduced Ellman reagent: reassessment. *Anal. Biochem.* **2003,** *312*, 224-227.

6. Wine, R. L., *Statistics for Scientists and Engineers*. Prentice-Hall: Englewood Cliffs, N.J., 1964.