

Supporting Information:

Reversible amyloids encode biomolecular memory through hysteresis

Alexander J. Dear^{†,1,2} Dorota M. Pfizenmaier^{‡,1,2} Sophie Rüdiger^{1,2} Anastasiia Kovalenko^{1,2}
Sung Sik Lee^{1,2} Paolo Arosio^{3,2} Matthias Peter^{1,2} and Thomas C. T. Michaels^{*1,2}

¹*Department of Biology, Institute of Biochemistry,
ETH Zurich, Otto Stern Weg 3, 8093 Zurich, Switzerland*

²*Bringing Materials to Life Initiative, ETH Zurich, Switzerland*

³*Department of Chemistry and Applied Biosciences,
Institute for Chemical and Bioengineering, ETH Zurich, Zurich, Switzerland*
(Dated: September 29, 2025)

I. EXPERIMENTAL METHODS

A. Fibril production

Peptide fibril formation was induced using a stock solution of human PKM2 peptide, obtained in lyophilized form (sequence: EAEAAIYHLQLFEEL from peptides&elephants GmbH). The peptide was dissolved in dimethyl sulfoxide (DMSO; Sigma-Aldrich, D5879) to a concentration of 10 mg/mL and stored at -20°C until use. For fibril formation, the peptide stock was diluted to a final concentration of 0.5 mg/mL in citrate phosphate buffer (CPS) containing 2 mM citric acid, 50 mM Na₂HPO₄, and 150 mM NaCl. The pH of the buffer was adjusted to a final value between 4.5 and 8.5 using HCl or NaOH. The buffer was filtered through a 0.22 µm Steriflip Millipore Express PLUS Membrane before use. Fibril formation was induced by incubating the peptide solution at pH 5.0 in low-binding Eppendorf Protein LoBind Tubes (Eppendorf, 0030 108 442) at 30°C for 24 hours.

B. Seed Preparation

Following incubation, fibrils were fragmented to produce seeds using a Sonoplus HD 4050 homogenizer (Bandelin) equipped with a Sonoplus TS 102 titansonotrode (Bandelin). Fragmentation was performed by sonicating the fibrils on ice at 4°C. The sonication parameters were set to a frequency of 19,925 Hz and 20% amplitude (corresponding to 10 W power output), with a pulse cycle of 1 second on and 1 second off. The sonication process was conducted for 30 seconds, repeated 10 times, with a 1-minute rest interval between each sonication cycle.

C. Aggregation yield experiments

The peptide was diluted in CPS buffers at the indicated pH values (see figure legends) to a final working concentration of 0.5 mg/mL in 50 µL samples. For the seeded experiments, 5 µL, 10 µL, and 15 µL seed stock were added to the samples, resulting in seed concentrations of 10%, 16%, and 24%, respectively, while maintaining the overall protein concentration at 0.5 mg/mL. Given the molecular weight of the PKM2 construct is 1775.2 g/mol, this corresponds to 282 µM. The final pH of the mixtures was validated using a PH 9500 meter (Apera Instruments) to ensure accuracy, as the peptide may influence the pH. The samples were subsequently incubated in low-binding tubes (Eppendorf, Protein LoBind Tubes, 0030 108 442) at 30°C for incubation periods of 1 day, 3 days, and 1 week. 18 µL aliquots of each sample were transferred into individual wells of a 384-well, non-binding surface assay plate (Corning 3766, Black/Clear Flat Bottom Polystyrene NBS Microplate). Thioflavin T (ThT; arcoss organics, 211760250) was dissolved in water to a final concentration of 2.5 mM and filtered (Millipore, 0.2 µm filter). The samples were then supplemented with 0.25 µM of the ThT solution. The ThT fluorescence signal was recorded using a CLARIOstar Plus plate reader (BMG Labtech) (450 nm excitation, 490 nm emission), at 30°C. To prevent the samples from clumping, the samples were subjected to double orbital shaking for 5 seconds at 500 rpm before each cycle. Each measurement cycle lasted 120 seconds, with 20 flashes per well and cycle.

D. Aggregation kinetic experiments with and without seeds

Fibril seeds were generated essentially as described above, with peptide concentrations of 0.0791 mg/mL, 0.25 mg/mL, 0.791 mg/mL and 2.08 mg/mL followed by five cycles of sonication. Samples of 110 µL were prepared at the specified pHs (see figure legends) by first mixing CPS buffer and ThT (final concentration 0.25 mg/mL) in PCR tubes (Sarstedt AG & Co. KG, 72.985.002). The mixtures were precooled on ice for at least 30 min-

¹ *To whom correspondence should be addressed. E-mail: thomas.michaels@bc.biol.ethz.ch

[‡]These authors contributed equally to this work.

utes using a metal rack. After precooling, peptide stock was added to a final concentration of 0.25 mg/ml or 0.5 mg/ml to the mixture, along with 3.3 μ l or 6.6 μ l of the sonicated seed stock, respectively, resulting in final seed percentages of 0.95%, 3%, 9.5% and 25% relative to the peptide concentration. Aliquots of 30 μ L from each sample were placed into individual wells of a 384-well, non-binding surface assay plate (Corning 3766, Black/Clear Flat Bottom Polystyrene NBS Microplate). Aggregation was monitored by measuring ThT fluorescence through the bottom of the plate (450 nm excitation, 490 nm emission), at 30°C. To ensure uniform sample mixing, orbital shaking was performed between readings at 200 rpm with alternating on/off times of 30 seconds and 10 seconds. Additionally, double orbital shaking at 500 rpm for 3 seconds was performed prior to each cycle. The signal was measured every 2 minutes for at least 8 hours with 20 flashes per well and cycle.

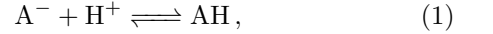
E. Disaggregation experiments

Fibrils were produced as described above. After 24 hours of incubation, the fibrils were pelleted by centrifugation at 14 000 rpm with an Eppendorf 5427 R centrifuge for 10 minutes. The supernatant was carefully removed with a pipette, and the fibrils were resuspended in an equal volume of the desired buffer with the same pHs as used in the aggregation experiments. The tube was tapped until the fibrils were fully resuspended. The final pH was rechecked as explained before (1.2). The samples were then incubated again for 24 hours at 30°C. ThT fluorescence measurement was performed as described above (1.2).

II. ANALYSIS OF EQUILIBRIUM FIBRIL YIELD

A. Extent of protonation

PKM2 monomers carry a substantial net negative charge, that must be partially neutralized to enable aggregate formation. This partial neutralization can be achieved at physiological pH ranges by protonation of a certain amino acid's side group (Histidine in WT). To describe this protonation event as a function of pH, we consider the following protonation-deprotonation reaction scheme



where AH stands for the protonated residue and A^- is the deprotonated residue. The equilibrium constant for this reaction is:

$$K_A = \frac{[A^-][H^+]}{[AH]}. \quad (2)$$

Since by definition $pH = -\log_{10}([H^+])$ and $pK_{a_A} = -\log_{10}(K_A)$, we can rewrite Eq. (2) as

$$\frac{K_A}{[H^+]} = \frac{[A^-]}{[AH]} = 10^{pH - pK_{a_A}}. \quad (3)$$

From this and using the conservation relation $[A]_{\text{tot}} = [A^-] + [AH]$, we can calculate the fraction x_A of monomers with the correct charge state for aggregate formation as

$$x_A = \frac{[AH]}{[A]_{\text{tot}}} = \frac{1}{1 + 10^{pH - pK_{a_A}}}. \quad (4)$$

We will write $m(t)$ as the total concentration of protonated and deprotonated monomeric protein, i.e. $m(t) = [A]_{\text{tot}}$.

Protonation is typically very rapid in aqueous solution, and expected to be far faster than the aggregation timescale. Therefore, the pre- or partial-equilibrium approximation is warranted. The concentration of protonated monomers can thus be modelled within the rate laws for protein aggregation at all times by $m_{AH} = x_A \cdot m(t)$.

TABLE I.

$t = t$	$t = \infty$	$t = 0$
$M(t)$	$M(\infty)$	$M(0)$
$m_{AH} = x_A \cdot m(t)$	m_{CAC}	$x_A \cdot m(0)$
$m_{A^-} = (1 - x_A) \cdot m(t)$	$m_{A^-, \infty}$	$(1 - x_A) \cdot m(0)$
$m(t) = m_{A^-} + m_{AH}$	$m_{CAC, eff} = m_{CAC} / x_A$	$m(0)$

B. Fibril yield and critical aggregation concentration

We write $M(t)$ for the concentration of protein in fibrillar form. In a closed system, the total protein concen-

tration m_{tot} is constant, so $m_{\text{tot}} = M(t) + m(t)$ (and in

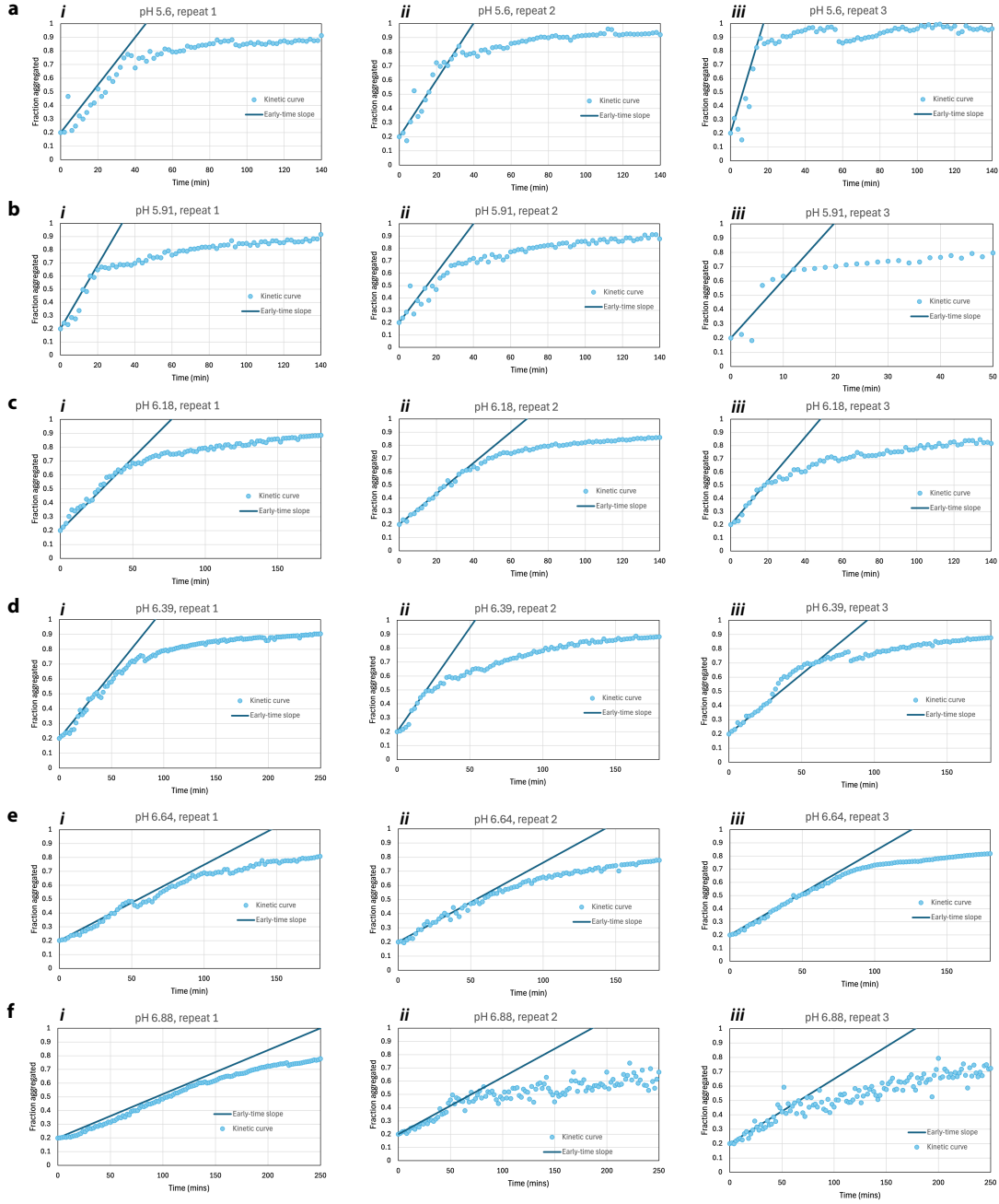


FIG. 1. **Initial slopes of kinetic curves for highly-seeded aggregation reactions.** 3 repeats per condition. **a:** pH 5.6. **b:** pH 5.91. **c:** pH 6.18. **d:** pH 6.39. **e:** pH 6.64. **f:** pH 6.88. Fitted values for initial slopes are listed in Table II.

the absence of seed, $m(0) = m_{\text{tot}}$).

As large aggregate species, fibrils co-exist at equilibrium with a constant concentration of monomers, the “critical aggregation concentration”, no matter what m_{tot} is, for the same reason that micelles possess a constant critical micelle concentration. Historically the charge state of the monomers has not been accounted for explicitly and they have consequently been treated as a single species, with the critical aggregation concentration referring implicitly to the total monomer concentration at equilibrium. Since our experiments are carried

out across multiple pH values this approximation is invalid, since from a thermodynamic perspective only the concentration of monomers with the correct charge state to form fibrils remains constant at equilibrium even as the pH changes. Therefore, it is more convenient and more terminologically accurate to use “critical aggregation concentration” specifically for the concentration of monomers with the correct charge to form fibrils when the fibril formation reaction has attained equilibrium. We henceforth denote this as m_{CAC} .

By contrast, we continue use the conventional defi-

TABLE II. Absolute and relative initial slopes determined by fitting in Fig. 1

pH	Initial slope	Baselined to pH 6.88
6.88	$(40 \pm 7) \times 10^{-4}$	1 ± 0.23
6.64	$(58 \pm 5) \times 10^{-4}$	1.45 ± 0.34
6.39	$(11 \pm 3) \times 10^{-3}$	2.74 ± 0.64
6.18	$(13 \pm 3) \times 10^{-3}$	3.34 ± 0.78
5.91	$(21 \pm 5) \times 10^{-3}$	5.16 ± 1.20
5.6	$(28 \pm 11) \times 10^{-3}$	6.97 ± 1.62

nition of solubility s_A , i.e. the total concentration of monomeric protein at equilibrium, which is therefore $s_A = m_{CAC}/x_A$. The total equilibrium yield of fibrils is then $M_\infty = m_{tot} - m_{CAC}/x_A$. We can finally relate m_{CAC} to rate constants as follows. At equilibrium the rates of elongation and disaggregation become equal:

$$2k_+m_{CAC}P = 2k_{off}P, \quad (5)$$

where $P(t)$ is the total concentration of amyloid fibrils, k_+ the elongation rate constant and k_{off} the disaggregation rate constant. Therefore:

$$m_{CAC} = \frac{k_{off}}{k_+} = \frac{c_0}{K_{el}}, \quad (6)$$

where c_0 is a standard concentration and K_{el} the equilibrium constant for elongation. These relations also permit k_{off} to be rewritten as k_+m_{CAC} .

C. Determining pKa

For an aggregation reaction with significant and non-zero preformed fibril seed concentration $P(0)$, the initial slope of elongation of fibrils can be modelled as:

$$\frac{dM}{dt} = 2k_+(x_A m(0) - m_{CAC})P(0). \quad (7)$$

The increase in yield of fibrils over the reaction time-course at any given pH is $M_\infty - M(0) = m(0) - m_{CAC}/x_A$. So, when the data are normalized, the initial slope becomes:

$$\text{initial slope} = 2k_+x_AP(0). \quad (8)$$

This only depends on pK_{aA} , not m_{CAC} . Now, we can measure this initial slope at many pH values (see Fig. 1) and then normalize them again to one of these initial slopes, evaluated at given pH^* . This relative initial slope is then:

$$\text{relative initial slope} = x_A(pH)/x_A(pH^*). \quad (9)$$

We performed aggregation experiments at monomer concentration of 0.5 mg/mL, with preformed seeds added at 25% of this value, across a series of pH values. Fitting Eq. (9) (see main text Fig. 2) to this gave a pK_{aA} of 5.9,

close to the value of 6 guessed in ref. [1] (being the pK_{aA} of the histidines in the core region). Once the pK_{aA} is known, so is $x_A(pH^*)$, allowing finally the fraction of protonated PKM2(284-398), x_A , to be plotted directly (see main text Fig. 2). This also tells us that the charge state required for fibril formation is -2, as achieved when the H residue is protonated. At higher pH values almost all PKM2(284-398) carries instead a net charge of -3.

D. Determining the critical aggregation concentration

The fraction of protein in aggregated state at a given pH was measured using disaggregation reactions. This is just equivalent to the yield, i.e.:

$$M_\infty = m_{tot} - \frac{m_{CAC}}{x_A}. \quad (10)$$

Since x_A depends only on pK_{aA} and pH, we can use the above-calculated value of pK_{aA} in this expression and then fit it to the disaggregation reaction yield at different pH values to obtain the CAC, which we find to be 0.019 mg/mL (see main text Fig. 2).

E. Effective solubility of fibrils

It is convenient to consider the pH-dependent effective solubility of fibrils s_A , i.e. the amount of monomer required at different pH values for fibrils to become thermodynamically stable. This can be expressed as:

$$0 \leq M_\infty = m_{tot} - \frac{m_{CAC}}{x_A} \quad (11)$$

$$\therefore m_{tot} \geq s_A \equiv m_{CAC}(1 + 10^{pH-pK_{aA}}). \quad (12)$$

Assuming m_{tot} has been chosen to be $\gg m_{CAC}$, such that elongation is rapid at lower pH values, the expression for s_A simplifies to one depending on a single combined parameter σ_A :

$$s_A \approx 10^{pH-\sigma_A}, \quad (13)$$

$$\sigma_A = pK_{aA} - \log_{10} m_{CAC}. \quad (14)$$

The pH required for fibril thermodynamic stability is therefore:

$$pH \leq \sigma_A + \log_{10}[m_{tot}]. \quad (15)$$

Note, although we have separately determined pK_{aA} and m_{CAC} , this is not strictly required to characterize the upper arm of the hysteresis loop, which is uniquely determined (apart from very minor adjustments to its shape as it approaches unity) by the combined parameter σ_A via the simplified yield formula:

$$M_\infty = \begin{cases} m_{tot} - 10^{pH-\sigma_A} & m_{tot} > 10^{pH-\sigma_A} \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

III. MODELLING PH-DEPENDENT AGGREGATION WITHOUT NUCLEATION INTERMEDIATES

In the case that there are no intermediates of fibril nucleation, the only species to consider are monomers and fibrils.

A. Deriving a kinetic model

For primary nucleation and its inverse, the rate law reads:

$$\frac{dP}{dt} = k_n(x_A m(t))^{n_c} - k_{-n} f_{n_c}(t), \quad (17)$$

where f_{n_c} is the concentration of new fibril nuclei before they are further elongated. Obviously we do not want this term in the model. However, when we are well above m_{CAC} this is governed by a steady-state relation, between elongation of the nuclei and their formation, i.e.:

$$0 = k_n(x_A m)^{n_c} - 2k_+ x_A m f_{n_c} \\ \therefore f_{n_c} = \frac{k_n}{2k_+} (x_A m)^{n_c-1}. \quad (18)$$

So, we can overall write the rate law for primary nucleation and its inverse as:

$$\frac{dP}{dt} = k_n(x_A m(t))^{n_c-1} (x_A m(t) - k_{-n}/2k_+). \quad (19)$$

Since nucleation is much slower than elongation, generally the second term can be neglected. However, as we approach m_{CAC} the above steady-state relation no longer applies. n_c does not diverge since the fibrils are linear, and beyond the forming of the first two fibril planes further monomer addition therefore involves proportional increases of surface area and volume. Consequently, provided $m > m_{CAC}$ further monomer addition is always favoured after the assembly of the first two fibril planes.

Only once m_{CAC} is actually reached from above does anything change. Now the free energy increases ad infinitum with aggregate size, such that there is no free energy maximum occurring at a finite aggregate size. Therefore, nucleation according to classical nucleation theory ceases to occur. (Of course this is an approximation as the Boltzmann distribution still applies, but it remains true that no significant fractional aggregation occurs.)

To satisfy these two limits, we can therefore use the following approximate rate law:

$$\frac{dP}{dt} = k_n(x_A m(t))^{n_c-1} (x_A m(t) - m_{CAC}). \quad (20)$$

The exact same logic applies to secondary nucleation;

therefore, the overall kinetic equations are:

$$\frac{dP}{dt} = k_n(x_A m(t))^{n_c-1} (x_A m(t) - m_{CAC}) \\ + k_2(x_A m(t))^{n_2-1} (x_A m(t) - m_{CAC}) M(t) \quad (21a)$$

$$\frac{dM}{dt} = 2k_+ q(t) P(t) \quad (21b)$$

$$q(t) = x_A m(t) - m_{CAC}, \quad m(t) + M(t) = m_{tot}. \quad (21c)$$

B. Determining rate constants by fitting kinetic data

At pH 5.75, $x_A m \gg m_{CAC}$ and so we can write $x_A m(t) = q(t)$ without appreciable loss of accuracy. The kinetic equations for the no-intermediate model at pH 5.75 then simplify to:

$$\frac{dP}{dt} = \tilde{k}_n m(t)^{n_c} + \tilde{k}_2 m(t)^{n_2} M(t) \quad (22a)$$

$$\frac{dM}{dt} = 2\tilde{k}_+ m(t) P(t) \quad (22b)$$

$$m(t) + M(t) = m_{tot}, \quad (22c)$$

where:

$$\tilde{k}_n = x_A (\text{pH} = 5.75)^{n_c} k_n \quad (23a)$$

$$\tilde{k}_2 = x_A (\text{pH} = 5.75)^{n_2} k_2 \quad (23b)$$

$$\tilde{k}_+ = x_A (\text{pH} = 5.75) k_+. \quad (23c)$$

These are identical to the standard rate equations typically used for protein aggregation with secondary processes, but with modified rate constants. They therefore have the usual analytical solutions [2, 3], albeit with these modified rate constants. Note that the usual analytical solutions do not depend on these rate constants independently, but instead as products $k_+ k_n$ and $k_+ k_2$. The products $k_+ k_2 x_A^{n_2+1}$ and $k_+ k_n x_A^{n_c+1}$ of the modified rate constants can therefore be determined by fitting the solutions of the standard protein aggregation rate equations to the data using Amylofit (Fig. 2). To extract the pH-independent rate constants, we must divide our fitted values by the appropriate powers of x_A . We do precisely this after global fitting across a range of concentrations from 0.2 to 0.45 mg/mL of PKM2 at pH 5.75. We find $n_c = 2$, $n_2 = 1.93$, and the corrected $k_+ k_n = 6.1 \times 10^{-5}$ and $k_+ k_2 = 0.41$.

C. Misfitting the hysteresis loop

With these values and the previously determined CAC and pKa_A , we can now plot the theoretical prediction of this model for the lower arm of the hysteresis loop. For this we must use the unapproximated no-intermediate

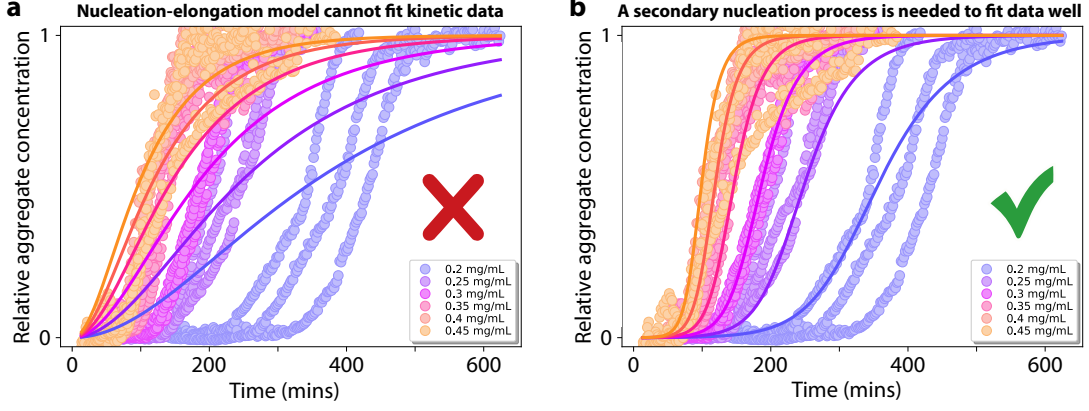


FIG. 2. **Kinetic model fitting to concentration series data.** 3 repeats per concentration. **a:** Misfit to model featuring only elongation and primary nucleation. **b:** Fit to model featuring elongation, primary nucleation and secondary nucleation. Unmodified rate constants back-calculated using Eqs. (23) are $n_c = 2$, $n_2 = 1.93$, $k_+k_n = 6.1 \times 10^{-5}$ and $k_+k_2 = 0.41$.

model, i.e. Eqs. (21):

$$\frac{dP}{dt} = k_n(q(t) + m_{CAC})^{n_c-1}q(t) + k_2(q(t) + m_{CAC})^{n_2-1}q(t)M(t) \quad (24a)$$

$$\frac{dM}{dt} = 2k_+q(t)P(t) \quad (24b)$$

$$q(t) = x_A m(t) - m_{CAC}, \quad m(t) + M(t) = m_{tot} \quad (24c)$$

We find that this can explain only part of the hysteresis seen in the experiments. This implies that this model is incorrect. (As explained above, n_c is not expected to diverge as m_{CAC} is approached from above due to the linearity of the fibrils.)

IV. MODELLING PH-DEPENDENT AGGREGATION WITH NUCLEATION INTERMEDIATES

The nucleation of new amyloid fibrils requires conformational conversion of usually-oligomeric intermediates. These intermediates are generally expected to be much smaller than mature fibrils but large enough to possess a critical monomer concentration required for their formation, similar to micelles. We label this “critical intermediate concentration” m_{CIC} . If the concentration of monomeric PKM2 with appropriate protonation state is below this concentration, the concentration of intermediates is always negligible, and their rate of conversion to new fibrils is therefore also negligible. This aspect of the physical chemistry is unchanged if the intermediates are filamentous rather than oligomeric. All that matters for hysteresis to occur is that the intermediate is significantly less stable than the product. However, the protonation state required for intermediate formation may not be the same as that required for fibril formation. Therefore, the effective pKa for protonation to this required level may be different; we denote it pK_{A_I} . So, the fraction of

monomeric PKM2 with appropriate charge for intermediate formation, x_I , can be modelled as:

$$x_I = \frac{[AH]}{[A]_{tot}} = \frac{1}{1 + 10^{pH - pK_{A_I}}}, \quad (25)$$

where A now represents the residue that must be protonated for intermediate formation to proceed. The hypothesis that different pH-sensitivity of nucleation intermediates may drive hysteresis behaviour is supported by the fact that the pH-dependence of elongation (Fig. 3a) is different to that for secondary nucleation (Fig. 3b) and primary nucleation (Fig. 3c). Moreover, primary and secondary nucleation both generate nucleation intermediates and both appear to have similar pH-dependence to one another.

A. Deriving a kinetic model

Crudely, a kinetic model of nucleation with intermediates can be modelled as a nucleation reaction without nucleation intermediates but with available monomer concentration $r(t) = x_I m(t) - m_{CIC}$ and multiplied by $H[r(t)]$, where $H[\dots]$ is the Heaviside step function, such that no nucleation occurs below the critical intermediate concentration. Since we do not have (or really need) measurements of intermediate concentrations, this crude approach should suffice for our purposes. The kinetic equations for the with-intermediate model then read:

$$\frac{dP}{dt} = H[r(t)] (k_n r(t)^{n_c} + k_2 r(t)^{n_2} M(t)) \quad (26a)$$

$$\frac{dM}{dt} = 2k_+ q(t) P(t) \quad (26b)$$

$$m(t) + M(t) = m_{tot} \quad (26c)$$

$$r(t) = x_I m(t) - m_{CIC} \quad (26d)$$

$$q(t) = x_A m(t) - m_{CAC}. \quad (26e)$$

We may condense these equations by eliminating q and r :

$$\frac{dP}{dt} = H \left[m(t) - \frac{m_{\text{CIC}}}{x_I} \right] \cdot \left(\hat{k}_n \left(m - \frac{m_{\text{CIC}}}{x_I} \right)^{n_c} + \hat{k}_2 \left(m - \frac{m_{\text{CIC}}}{x_I} \right)^{n_2} M \right) \quad (27a)$$

$$\frac{dM}{dt} = 2\hat{k}_+ (m - s_A) P \quad (27b)$$

$$m(t) + M(t) = m_{\text{tot}}, \quad (27c)$$

where:

$$\hat{k}_n = k_n x_I (\text{pH})^{n_c} \quad (28a)$$

$$\hat{k}_2 = k_2 x_I (\text{pH})^{n_2} \quad (28b)$$

$$\hat{k}_+ = k_+ x_A (\text{pH}), \quad (28c)$$

and s_A is the pH-dependent effective solubility of fibrils, given (as shown above) by $s_A = m_{\text{CAC}}/x_A$.

B. Effective solubility of intermediates

As it stands, the kinetic model appears to depend on both m_{CIC} and pKa_I via the pH-dependent effective solubility of intermediates s_I :

$$s_I \equiv \frac{m_{\text{CIC}}}{x_I} = m_{\text{CIC}} (1 + 10^{\text{pH} - \text{pKa}_I}). \quad (29)$$

This is analogous to s_A , the above-defined effective solubility of fibrils. Nucleation proceeds only when $m(t) > s_I$.

To determine m_{CIC} and pKa_I separately requires detailed model fitting of extensive seeded and unseeded kinetic experiments. However, in reality this is not necessary so long as $m(0)$ has been chosen to be $\gg m_{\text{CIC}}$, such that aggregation can complete rapidly at lower pH values. This is because to a very good approximation the kinetic model then depends on these parameters only via a single combined parameter. This follows because the most important effect of s_I is via the argument of the Heaviside function, which dictates the pH at which aggregation ceases. Multiplying out the bracket in x_I^{-1} , this can now be expressed as:

$$m(t) - s_I \equiv m(t) - m_{\text{CIC}} - m_{\text{CIC}} \cdot 10^{\text{pH} - \text{pKa}_I}. \quad (30)$$

If $m(0) \gg m_{\text{CIC}}$, then the second term can always be neglected in front of the first at early times. Moreover, nucleation is no longer important to the kinetics at late times, whereupon the precise form of the terms entering Eq. (27a). Therefore, without loss of accuracy the expression for s_I can be simplified at all times to one depending on a single combined parameter σ_I :

$$s_I \approx 10^{\text{pH} - \sigma_I}, \quad (31)$$

$$\sigma_I = \text{pKa}_I - \log_{10} m_{\text{CIC}}. \quad (32)$$

Consequently, using Eqs. (13), (29) and (31) we can further simplify the rate equations (Eqs. (27) to:

$$\frac{dP}{dt} = H[m - 10^{\text{pH} - \sigma_I}] \cdot \left(\hat{k}_n (m - 10^{\text{pH} - \sigma_I})^{n_c} + \hat{k}_2 (m - 10^{\text{pH} - \sigma_I})^{n_2} M \right) \quad (33a)$$

$$\frac{dM}{dt} = 2\hat{k}_+ (m - 10^{\text{pH} - \sigma_A}) P \quad (33b)$$

$$m(t) + M(t) = m_{\text{tot}}. \quad (33c)$$

C. Position and width of hysteresis loops

Since $m(t)$ is monotonic decreasing, the switching condition described by the first Heaviside function at $t = 0$ translates to a switching condition for whether or not there is ever any amyloid formation. Again assuming $m_{\text{tot}} \gg m_{\text{CIC}}$ (as will typically be the case), this switching condition can alternatively be expressed as:

$$\text{pH} \leq \sigma_I + \log_{10}[m(0)]. \quad (34)$$

The parameter σ_I therefore controls the position (and most of the shape) of the lower arm of the hysteresis loop, analogously to σ_A for the upper arm (as explained above). Using Eq. (15) and Eq. (34), the pH values at which the lower and upper arms of the hysteresis loop reach zero, pH_L and pH_U , are therefore given by:

$$\text{pH}_L = \sigma_I + \log_{10}[m_{\text{tot}}] = \text{pKa}_I + \log_{10} \left[\frac{m_{\text{tot}}}{m_{\text{CIC}}} \right] \quad (35a)$$

$$\text{pH}_U = \sigma_A + \log_{10}[m_{\text{tot}}] = \text{pKa}_A + \log_{10} \left[\frac{m_{\text{tot}}}{m_{\text{CAC}}} \right], \quad (35b)$$

where we have used the fact that $m(0) = m_{\text{tot}}$ in the lower arm. We can then calculate the width of hysteresis loops, $\Delta\text{pH} = \text{pH}_U - \text{pH}_L$, using Eqs. (35). These equations imply that this width is just $\sigma_A - \sigma_I$, giving:

$$\Delta\text{pH} = \sigma_A - \sigma_I = \Delta\text{pKa} - \log_{10} \left[\frac{m_{\text{CAC}}}{m_{\text{CIC}}} \right], \quad (36)$$

where $\Delta\text{pKa} = \text{pKa}_A - \text{pKa}_I$.

If $m_{\text{tot}} \gg m_{\text{CIC}}$, the switching condition described by the first Heaviside function, $m(0) - s_I > 0$, is instead (from Eq. (29), and using again the fact that $m(0) = m_{\text{tot}}$ in the lower arm):

$$\frac{m(0)}{m_{\text{CIC}}} > 1 + 10^{\text{pH} - \text{pKa}_I} \quad (37)$$

$$\Rightarrow \text{pH} < \text{pKa}_I + \log_{10} \left[\frac{m_{\text{tot}}}{m_{\text{CIC}}} - 1 \right]. \quad (38)$$

The width of the hysteresis loop is then instead the difference between Eq. (12) and Eq. (38), i.e.:

$$\Delta\text{pH} = \Delta\text{pKa} - \log_{10} \left[\frac{\frac{m_{\text{tot}}}{m_{\text{CIC}}} - 1}{\frac{m_{\text{tot}}}{m_{\text{CAC}}} - 1} \right]. \quad (39)$$

Either way, we see that only modest differences in pK_{aA} are required to cause significant width in hysteresis loops. Differences in solubility discounting protonation or deprotonation must be much larger to cause significant hysteresis.

V. QUANTIFYING HYSTERESIS INDUCED BY KINETIC TRAPPING

This kinetic model implies that monomers are kinetically trapped when the pH is low enough for fibrils to be thermodynamically stable, but not for intermediates to be thermodynamically stable. To quantify hysteresis we must therefore first quantify the lower arm of the hysteresis loop. Assuming the time at which hysteresis is measured has been chosen such that aggregation has time to complete when $x_I m(0) \gg m_{\text{CIC}}$, then the lower arm deviates from the upper arm only as $x_I m(0) \rightarrow m_{\text{CIC}}$, or equivalently $r(0) \rightarrow 0$. Then only the Heaviside functions in the kinetic model equations (Eqs. (27)) are important for simulating the lower arm, which is zero when $r(0) \leq 0$ and coincides with the upper arm when $r(0) > \delta m(0)$, and δ is a small positive number. For the narrow pH range where $\delta m(0) > r(0) > 0$, the lower arm rapidly transitions between these limiting values, with the rapidity of this transition governed by the time at which hysteresis is measured and the rate constants for aggregation. $r(0)$ effectively acts like a switch, with close to 0% aggregation happening when $r(0) = 0$ and the equilibrium reaction yield almost fully achieved when $r > 0$.

A. Determining rate constants at pH 5.75

To convert the naively fitted rate constants to the correct rate constants, we first eliminate the effective solubility of aggregates by noting that, as before, at pH 5.75 $m(t) \gg s_A$ over most of the reaction timecourse. The first simplification of the rate equations therefore yields:

$$\frac{dP}{dt} = H[r(t)] (k_n r(t)^{n_c} + k_2 r(t)^{n_2} M(t)) \quad (40a)$$

$$\frac{dM}{dt} = 2\tilde{k}_+ m(t) P(t) \quad (40b)$$

$$m(t) + M(t) = m_{\text{tot}} \quad (40c)$$

$$r(t) = x_I (m(t) - s_I), \quad (40d)$$

with $\tilde{k}_+ = \hat{k}_+(\text{pH} = 5.75) = k_+ x_A(\text{pH} = 5.75)$ as before. We next note that the nucleation reaction steps are important predominantly only early in the reaction timecourse, when $m(t) \simeq m(0)$ and $r(t) > 0$ at pH 5.75. Our next simplification for the rate equations at this pH

therefore yields:

$$\frac{dP}{dt} = \tilde{k}_n m(t)^{n_c} + \tilde{k}_2 m(t)^{n_2} M(t) \quad (41a)$$

$$\frac{dM}{dt} = 2\tilde{k}_+ m(t) P(t) \quad (41b)$$

$$m(t) + M(t) = m_{\text{tot}}, \quad (41c)$$

where:

$$\tilde{k}_n = k_n x_I (\text{pH} = 5.75)^{n_c} \left(1 - \frac{s_I(\text{pH} = 5.75)}{m(0)}\right)^{n_c} \quad (42a)$$

$$\tilde{k}_2 = k_2 x_I (\text{pH} = 5.75)^{n_2} \left(1 - \frac{s_I(\text{pH} = 5.75)}{m(0)}\right)^{n_2} \quad (42b)$$

$$\tilde{k}_+ = k_+ x_A(\text{pH} = 5.75). \quad (42c)$$

These are just standard kinetic equations for aggregation with secondary processes, but with modified rate constants. Fitting using standard kinetic models via Amylofit will therefore determine these modified rate constants.

B. Fitting WT hysteresis loop

For use in modelling kinetic curves at other pH values using Eqs. (27), \tilde{k}_n , \tilde{k}_2 and \tilde{k}_+ must in principle be converted back to pH-independent rate constants k_n , k_2 and k_+ . However, we cannot calculate pK_{aI} for any of the peptides considered in the main text, which would require detailed kinetic experiments and modelling to obtain. Fortunately, as shown above these rate constants only affect the shape of the hysteresis lower arm in a very narrow range, and do not significantly affect its position.

For simulating hysteresis loops, we can therefore instead simply approximate \hat{k}_n , \hat{k}_2 and \hat{k}_+ as follows:

$$\hat{k}_n \approx \tilde{k}_n \left(1 - \frac{s_I(\text{pH} = 5.75)}{m(0)}\right)^{-n_c} \quad (43a)$$

$$\hat{k}_2 \approx \tilde{k}_2 \left(1 - \frac{s_I(\text{pH} = 5.75)}{m(0)}\right)^{-n_2} \quad (43b)$$

$$\hat{k}_+ \approx \tilde{k}_+. \quad (43c)$$

This will underestimate the rate of reaction at pH values below 5.75, but this does not matter as the rate at pH 5.75 is already fast enough for the hysteresis lower arm to reach the (equilibrium) upper arm. This underestimate will therefore have no effect on the hysteresis loop at lower pH. The only effect will be to overestimate the rate constants slightly between pH 5.75 and pH_L (≈ 6), but since at pH 5.75 the reaction completes well within the time allotted for hysteresis measurements, this overestimate should not noticeably change the lower arm between these pH values.

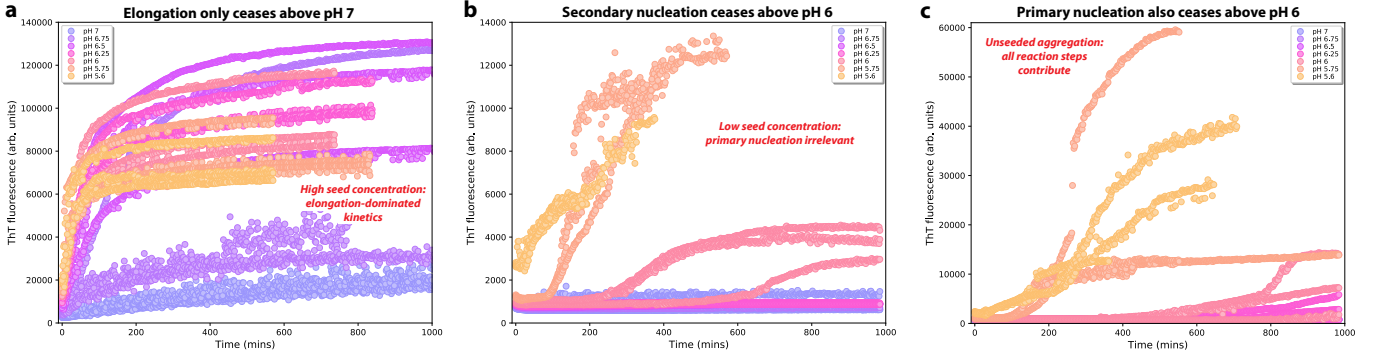


FIG. 3. **Kinetic experiments performed with varying seed levels verify that kinetic trapping occurs with secondary as well as primary nucleation.** Initial monomer concentration $m(0) = 0.5$ mg/ml in all panels. **a:** Highly seeded reaction kinetics depend predominantly on elongation. They confirm that elongation is not suppressed until pH increases above 7. Seed concentration 25% of $m(0)$. Nonetheless the fibril yield is low at pH 6.75 and pH 7 because the solubility s_A is close to $m(0)$ at these pH values. **b:** Reaction kinetics at low seeding levels depend on both elongation and secondary nucleation but not primary nucleation [4]. They demonstrate that secondary nucleation ceases once pH increases above 6. Seed concentration 0.95% of $m(0)$. **c:** Unseeded kinetics depend also on primary nucleation and show that the latter also ceases above pH 6, or at least is exceedingly slow. Note, ThT fluorescence varies between experiments not just due to differing fibril concentrations but also due to different settings on plate readers. ThT fluorescence plateau height variance between repeats can reflect variations in yield but also different morphologies with different fluorescence efficiencies, slightly different temperatures in different wells, differing levels of heterogeneity in the distribution of fibrils within the well, etc.

We can estimate the pH at which $r(0) = 0$ separately by visually examining the low seeding data, and seeing at what pH positive curvature is lost. The loss of curvature happens between pH 6 and pH 6.25 (SI Fig. 3b). Therefore, using Eq. (15) gives:

$$6.25 = \sigma_I + \log_{10}[0.5]. \quad (44)$$

This gives an estimate for σ_I of 6.55. Fitting the rate equations Eqs. (27) to the hysteresis lower arm data using modified rate constants Eqs. (43) gives us instead a value of $\sigma_I = 6.64$ - but this is close enough to be consistent with the low seeding data and implies a degree of robustness in our calculation.

Were $\text{pK}_{aI} = \text{pK}_{aA}$, this would imply that $m_{\text{CIC}} \approx 0.19$ mg/ml. However, this is inconsistent with the observation of spontaneous aggregation at pH 5.75 when $m(0) = 0.2$ mg/ml (see main text Fig. 3), which instead implies $m_{\text{CIC}} \ll 0.2$ mg/ml. So, it must be true that $\text{pK}_{aI} < \text{pK}_{aA}$. In other words, to enable intermediate formation it must be necessary to protonate an additional residue with a lower pK_a than the one whose protonation enables fibril formation. So, the charge state required for intermediate formation is -1 as opposed to the -2 charge state required for fibril formation.

C. Fitting hysteresis loops for mutants

We produced a series of single-point mutants of PKM2 and measured hysteresis loops for them. We determined σ_I and σ_A by fitting Eqs. (27) to the lower arm data and Eq. (16) to the upper arm data. We used the same estimates (Eqs. (43)) for the modified rate constants as for

WT since, as discussed above, the hysteresis loops are fairly insensitive to the exact values of the rate constants provided a late enough measurement time has been provided that aggregation completes at low pH values. It was therefore not necessary to perform kinetic experiments on the mutants and directly determine their aggregation rate constants. We present as the result of the fitting the values of σ_I and σ_A in Table III. Note the combined parameters for upper and lower arms can be interpreted as the pH required for the respective arm to increase above zero, minus $\log_{10} m(0)$.

Although it is therefore not necessary for fitting purposes, we nonetheless estimate in the main text how the pK_{aA} is changed for the relevant protonated species by calculating how the single-point mutation should change the charge state.

TABLE III. Combined parameters $\sigma_A = \text{pK}_{aA} - \log_{10} m_{\text{CAC}}$ and $\sigma_I = \text{pK}_{aI} - \log_{10} m_{\text{CIC}}$ determined from fitting mutant hysteresis loops.

Mutant	σ_A	σ_I
WT	7.65	5.94
E3Q	11.03	6.41
E13Q	11.03	7.41
Q10E	6.41	6.06
H8K	6.93	6.08
H8R	7.53	6.53
E3QH8R	8.13	6.48
H8E	6.53	6.04

VI. FREE ENERGY DIAGRAMS

For a chemical reaction not at equilibrium, the molar free energy change is:

$$\Delta_r G = \Delta_r G^\circ + RT \ln Q_r = RT \ln Q_r / K_{eq}, \quad (45)$$

where K_{eq} is the reaction's equilibrium constant and Q_r the (time-dependent) reaction quotient. For the reaction $\sum_i a_i A_i \rightleftharpoons \sum_i b_i B_i$, where A_i and B_i are the reactant and product species and a_i and b_i their stoichiometries, Q_r is defined as:

$$Q_r = \frac{\prod_i [B_i]^{b_i}}{\prod_i [A_i]^{a_i}}. \quad (46)$$

Since this represents the difference in free energy between reactants and products, we can construct energy level diagrams using $\Delta_r G$ at $t = 0$ (and therefore $m(t) = m(0)$) to inform on the spontaneity of elongation and nucleation.

Monomer-fibril stability can be computed using the relation $\Delta G^\circ = -RT \ln K_{el} = RT \ln m_{CAC}$ where we have chosen a standard concentration of 1 M in the second step. Since we can represent the general elongation step as $x_A m + P \rightleftharpoons P$, the reaction quotient at $t = 0$ is just $(x_A m(0))^{-1}$, and we are left with:

$$\Delta_r G = \Delta_r G^\circ + RT \ln Q_r = RT \ln \left[\frac{m_{CAC}}{x_A m(0)} \right]. \quad (47)$$

The formation of a nucleation intermediate S can be represented as $n_p m \rightleftharpoons S$. The equilibrium constant is $K_{nuc} = S_{eq} / m_{eq}^{n_p}$. Therefore:

$$S_{eq} = K_n^{-1} (K_n m_{eq})^{n_p}, \quad (48)$$

where we have defined $K_{nuc} = K_n^{n_p-1}$. Defining also $m_{s0} = x_I m(0) + n_p S(0)$, the initial concentration of reactants and intermediates, we can write:

$$K_n m_{s0} = K_n m_{eq} + n_p (K_n m_{eq})^{n_p}. \quad (49)$$

For significant n_p we have that $m_{eq} \approx m_{s0}$ if $K_n m_{eq} < 1$, since the first term on the RHS dominates. If otherwise $K_n m_{eq} > 1$, the second term dominates and $m_{eq} \approx K_n^{-1} (K_n m_{s0} / n_p)^{1/n_p}$. For large n_p this reduces to $m_{eq} \approx K_n^{-1}$. We therefore identify $K_n^{-1} = m_{CIC}$, or $K_{nuc} = m_{CIC}^{1-n_p}$.

The reaction quotient at $t = 0$ for intermediate formation is $Q_r = S(0) / (x_I m(0))^{n_p}$. So the Gibbs free energy change is:

$$\begin{aligned} \Delta_r G &= RT \ln \frac{m_{CIC}^{n_p-1} S(0)}{(x_I m(0))^{n_p}} \\ &= RT \ln \frac{S(0)}{x_I m(0)} + (n_p - 1) RT \ln \frac{m_{CIC}}{x_I m(0)}. \end{aligned} \quad (50)$$

For the purposes of constructing free energy diagrams, if n_p is large and we choose $S(0)$ not to be many orders of

magnitude different to m_{CIC} , this can be well approximated as:

$$\Delta_r G \approx n_p RT \ln \frac{m_{CIC}}{x_I m(0)}. \quad (51)$$

We then construct the free energy diagrams in Fig. 4 of the main text using the per-monomer free energy value, i.e. dividing through by n_p . Since n_p is expected to be large, the actual free energy barrier to nucleation of new fibrils is therefore many times higher than that implied by Fig. 4 when intermediate formation is unfavourable thermodynamically. Since m_{CIC} and pK_{aI} affect the hysteresis loop mainly as the combined parameter $m_{CIC} \cdot 10^{-pK_{aI}}$ via m_{CIC} / x_I , we can use the pK_{aI} value determined during fitting, alongside the value of m_{CIC} arbitrarily chosen for the fitting.

Note the difference in free energy for a monomer in an intermediate or in a fibril is:

$$\Delta_r G \approx RT \ln \left[\frac{m_{CAC}}{x_A m(0)} \right] - RT \ln \frac{m_{CIC}}{x_I m(0)}. \quad (52)$$

The arguments of the logarithms are set from fitting the hysteresis loops. Therefore, if $pK_{aA} \neq pK_{aI}$, these arguments do not change.

VII. ESTIMATING PEPTIDE NET CHARGE VS PH FOR WT AND MUTANTS

To estimate the net charge carried by PKM2(384-398) and its single point mutants at various pH values, we made the rough approximation that residue side chains and the C and N terminal groups are protonated and deprotonated independently and do not affect one another's pK_a values. We then used a simple equilibrium model for the protonation of each functional group X (total concentration $[X]$ and pK_a value denoted pK_{aX}):

$$\frac{[XH]}{[X]} = \frac{1}{1 + 10^{pH - pK_{aX}}}, \quad (53)$$

where $[XH]$ is the concentration of the functional group. For each side group that can be at least partially charged at the pH range we use in our experiments (4-8.5) we constructed such an equation. We constructed one also for the N-terminal end group, which can carry a positive charge at most of these pH values. (We did not do so for the C-terminal group which in our case is an uncharged -ONH₂ group.) We then modelled the overall charge C as being the sum of all these charges, or the sum of all these equations, i.e.:

$$C(pH) = \sum_X \frac{[XH]}{[X]} = \sum_X \frac{1}{1 + 10^{pH - pK_{aX}}}. \quad (54)$$

This allowed us to graph the charge vs pH for WT. This was extended to the various mutants by adding and removing protonation terms for new and old residues respectively.

VIII. FITTING HYSTERESIS LOOPS FOR ADDITIONAL MUTANTS

We produced and measured hysteresis loops for 3 more mutants beyond those illustrated in the main text: H8K, H8R and E3QH8R. Although the H8K mutation increases charge by +1 similarly to EXQ, it appears to shift the hysteresis loop to the left rather than the right, and reduce hysteresis width (Fig. 4a). This implies that the charge state for both intermediate and fibril formation must be higher, thus still requiring E protonation in both cases. This could be explained by the protonated form of H playing a key structural role in aggregation that is disrupted if the side group is changed at this position, weakening the bonding strength in these aggregates such that it is less able to overcome the charge repulsion in net negatively charged PKM2(284-298). Alternatively, it could be the case that the much higher solubility in water of K vs H stabilizes the monomeric vs aggregated form.

Similarly to H8K, H8R replaces the histidine with one that is always positively charged at the pH's considered here. This appears to stabilize intermediates and have little effect on fibril stability (Fig. 4b), unlike both EXQ (another mutant that increases the charge by +1 at higher pH) and H8K. This again supports the idea that not just the charge of the side chain but also its structure at this position in the peptide can be important for ag-

gregation. However, the ThT signal is actually non-zero at all pH values measured with this mutant, even very high pH's. So, this apparent reduction in yield at high pH could potentially be illusory, and instead be due to a lower ThT fluorescence for fibrils composed of this mutant at high pH. If so, this mutant would have a similar hysteresis loop to E3Q, as would be expected on purely charge grounds.

Finally, E3QH8R has a charge increased by +1 below around pH 6, and +2 above this pH. As discussed in the main text, the differing behaviours of the lower arms of the hysteresis loops of E3Q and E13Q suggest it may be necessary to protonate an E residue in the C-terminal end of the peptide to enable intermediate formation. This would lead us to predict a more positive charge state is required for intermediate formation of E3QH8R vs WT, and that therefore the lower arm of its hysteresis loop should be similar to E3Q. This is indeed observed (Fig. 4c). But on charge grounds alone, we would expect fibrils of this mutant to be stable at all pH values measured, with the upper arm therefore never reaching zero, similar to EXQ. This is also indeed observed, but the upper arm nonetheless appears to approach zero at the highest pH values. This could either be due to non-charge effects, such as differing solubilities of the monomeric forms of WT and E3QH8R, or it could simply be due to pH decreasing ThT fluorescence of fibrils of this mutant.

-
- [1] G. Cereghetti, *et al.*, An evolutionarily conserved mechanism controls reversible amyloids of pyruvate kinase via pH-sensing regions. *Developmental Cell* (2024).
 - [2] S. I. A. Cohen, *et al.*, Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9758–9763 (2013), doi:10.1073/pnas.1218402110, <https://doi.org/10.1073/pnas.1218402110>.
 - [3] G. Meisl, *et al.*, Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* **11** (2), 252–272 (2016), doi:10.1038/nprot.2016.010, <http://dx.doi.org/10.1038/nprot.2016.010>.
 - [4] A. J. Dear, G. Meisl, J. Hu, T. P. J. Knowles, S. Linse, Kinetics of seeded protein aggregation: theory and application (2025), <https://arxiv.org/abs/2503.20941>.

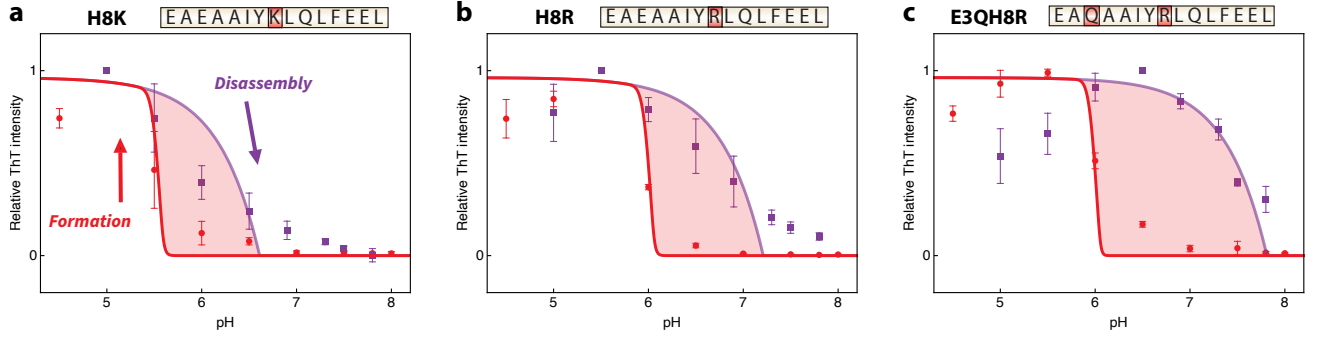


FIG. 4. **Hysteresis loops displayed by additional mutants.** Initial monomer concentration $m(0) = 0.5$ mg/ml in all panels. **a:** The mutant H8K has less stable fibrils and consequently less hysteresis than WT (see main text Fig. 5d). **b:** The mutant H8R has more stable intermediates and consequently less hysteresis. **c:** Both fibrils and nucleation intermediates of the mutant E3QH8R are increased similarly in stability; consequently, the extent of hysteresis is similar to WT.