Supplemental Information for

αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics

Andrew J. Baldwin[1,*], Hadi Lioe[2,*,†], Carol V. Robinson[2],

Lewis E. Kay[1], Justin L.P. Benesch[2]

[1] Departments of Molecular Genetics, Biochemistry and Chemistry, The University of Toronto, 1 Kings College Circle, Toronto, Ontario, M5S 1A8, Canada

[2] Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3QZ, U.K.

[†] Current address: Bio21 Institute, School of Chemistry, The University of Melbourne, Victoria, 3010, Australia

Correspondence to: AJB (ajb204@pound.med.utoronto.ca), LEK (kay@pound.med.utoronto.ca) and JLPB (justin.benesch@chem.ox.ac.uk)

* Both authors contributed equally.

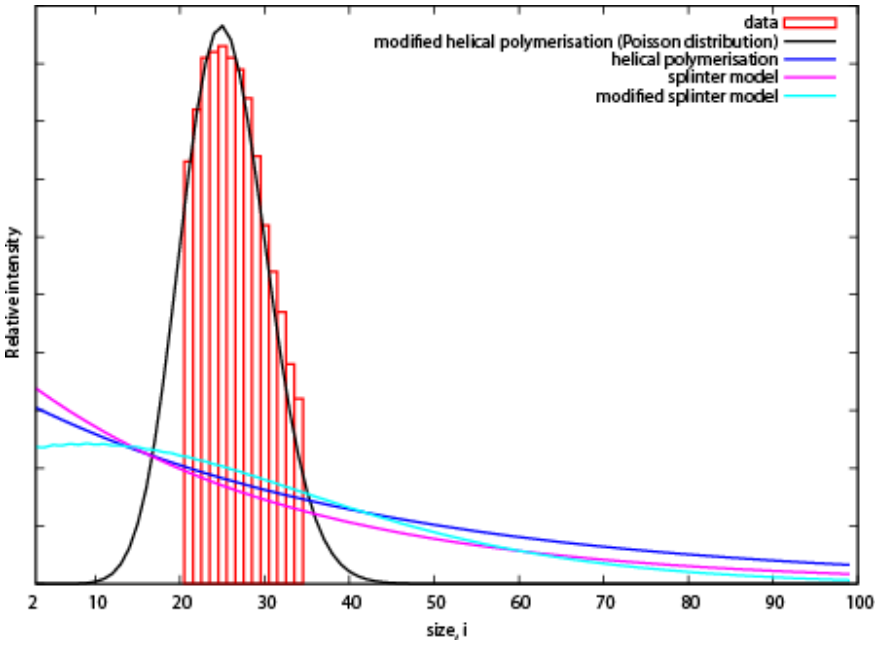

Figure S1. A comparison between the experimentally determined equilibrium oligomer distribution of αB-crystallin at pH 5, 37^{o}C (red) with the distributions predicted from four simple aggregation models, each characterised by a single free parameter. Two of the aggregation models assume that subunit exchange is due solely to monomer hopping (helical polymerisation, dark blue and the model considered in the text and referred to here as a modified helical polymerisation model, black). An additional two models, termed here 'splinter' models, assume that subunit exchange is mediated by the spontaneous dissociation and recombination of oligomers of arbitrary size. The distributions shown were obtained by numerically evaluating the kinetic equations shown below to determine the equilibrium oligomer size profiles for a given ratio of the association and dissociation rates $M_A = \frac{k^+[P_1]}{k^-}$ in the case of helical and modified helical polymerisation models and $\frac{k^+}{k^-}$ in the case of the splinter models. These ratios were

systematically varied to find the distribution that best matches the experimental data. Only the model described in the text (black) gives rise to a peaked distribution function that resembles the experimental data.

Helical Polymerisation Model:

Subunit exchange involving only monomers is considered, $P_{i-1} + P_1 \leftrightarrow P_i$. The rate of change of an oligomer of size *i*, is given by

$$\frac{d[P_i]}{dt} = k_i^+[P_{i-1}][P_1] - k_i^-[P_i] - k_{i+1}^+[P_i][P_1] + k_{i+1}^-[P_{i+1}]$$

for 2≤i<N where k_i^+ and k_i^- are oligomer dependent association and dissociation rates and N is the largest oligomer size in the distribution. Setting $k_i^+ = k^+$ and $k_i^- = k^-$ reduces the number of required rate constants from 2(N-1) to 2, and results in the equilibrium distributions observed in both the linear and helical polymerisation aggregation models [1]. The distribution at equilibrium depends only on the product of the association rate and the free monomer concentration, and on dissociation rate, $M_A = k^+[P_1]/k^-$. The only stable equilibrium distribution arises in the limit where $k^+[P_1] < k^-$ and follows an exponential decay, $[P_i] \propto \exp(i \ln M_A)$. Such a distribution is in poor agreement with the experimentally observed distribution (dark blue). Setting $k_i^+ = k^+$ and $k_i^- = ik^-$, as in the model described in the main text, results in a Poisson distribution at equilibrium that is in excellent accord with the experimentally observed profile (black).

Splinter Model:

An additional exchange scheme can be conceived, termed here the 'Splinter model', where formation and destruction of an oligomer of size *i* proceeds by the collision and

dissociation of combinations of fragments of different sizes (for example, $P_k + P_j \leftrightarrow P_{i=j+k}$). In this scheme the time evolution of an oligomer of size *i* is given by

$$\frac{d[P_i]}{dt} = \sum_{j=1}^{\beta}\left(k_i^+[P_{i-j}][P_j] - k_i^-[P_i]\right) + \sum_{j=i+1}^{\infty}\left(-k_j^+[P_i][P_{j-i}] + k_j^-[P_j]\right) + \left(-k_{2i}^+[P_i]^2 + k_{2i}^-[P_{2i}]\right)$$

where $\beta = i/2$ and $(i\text{-}1)/2$ for even and odd *i*, respectively and the forward and backward rates are k_i^+ and k_i^-. The terms grouped together in the first sum reflect the fact that an oligomer of size *i* can be formed by the collision of any two smaller oligomers whose combined size is equal to *i* (first term), or destroyed by breaking into two oligomers of smaller sizes (second term). The terms in the second sum correspond to removal of an oligomer of size *i* from solution through a binding event involving another oligomer to form a new, larger oligomer of size *j* (first term) or formation of an oligomer of size *i* by decomposition of a larger aggregate of size *j* (second term). The final sum 'corrects' for the fact that when a pair of equivalent oligomers (P_i) combine 2 molecules of P_i are lost (first term) and when P_{2i} dissociates into P_i, two molecules of P_i are produced (second term).

Using similar simplifying assumptions for the rate equations as described for the monomer hopping model, we set $k_i^+ = k^+$ and $k_i^- = k^-$ and numerically evolve the system from an arbitrary starting distribution to its natural equilibrium distribution, that depends only on the ratio k^+/k^-. This ratio was optimised to give the distribution in best agreement with the experimental data (pink). Finally, setting $k_i^+ = k^+$ and $k_i^- = ik^-$ and again optimising the ratio of k^+/k^- results in the distribution shown in light blue. The best fitting distributions in both cases are similar in form to those arising from the helical

polymerisation model and do not reproduce the experimental data. Of all the models considered here, (with similar levels of complexity) only the scheme considered in detail in the text predicts equilibrium distributions that are consistent with the experimentally observed oligomer size distribution of αB-crystallin at pH5.

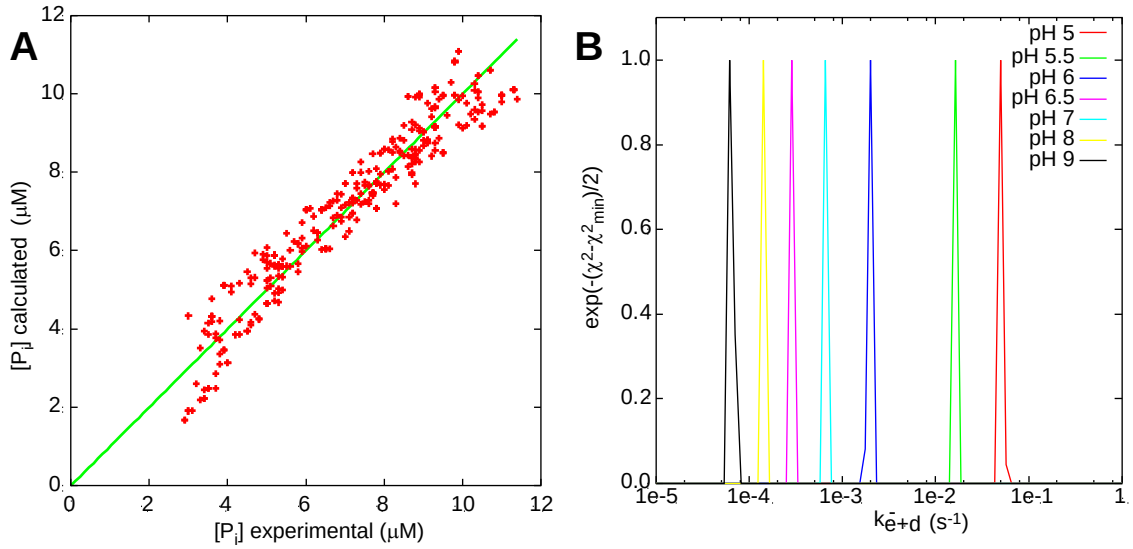

Figure S2. Quality of fits of the monomer exchange model (see text) to the MS data. **A** - Oligomers in the range 10-40mers were detected in the MS experiment. The 14 oligomers in the range 21-34 were present at concentrations where they could be reliably quantified at each solution condition. 25 pH and temperature combinations were examined in this study. The experimentally observed oligomer distributions at each pH and temperature condition was fitted to the exchange model described in the text so that two independent free energies, ΔG_{e+d} and ΔG_d, were obtained. The correlation between the experimentally determined oligomer concentrations and those calculated from the fitted parameters for all experimentally determined distributions are shown (350 individual oligomer concentrations), requiring 50 parameters, two for each pH/temperature condition. The Pearson's R^2 correlation coefficient for the entire dataset of 350 measurements is 0.92, and the global RMSD to the y=x line is 0.65 μM, indicating that the average uncertainty in each concentration measurement is approximately 6% of the concentration of the most populated oligomer. **B** – The rate k_{e+d}^{-} was determined by back calculating time dependent m/z vs intensity profiles as described in the text. The quantity $\exp(-(\chi^2-\chi^2_{min})/$

2) is proportional to the probability of a given k_{e+d}^{-} rate being the best-fitting parameter[2]. The width of the resulting peaks on the probability surface is directly related to the uncertainty in the given fitting parameter. Here, the fitting error for the rates extracted from the 25 datasets was found to be on the order of 5%, a value in line with the variations obtained when the measurement was repeated. Shown are the results for the data obtained at 37^{o}C rates as a function of pH.

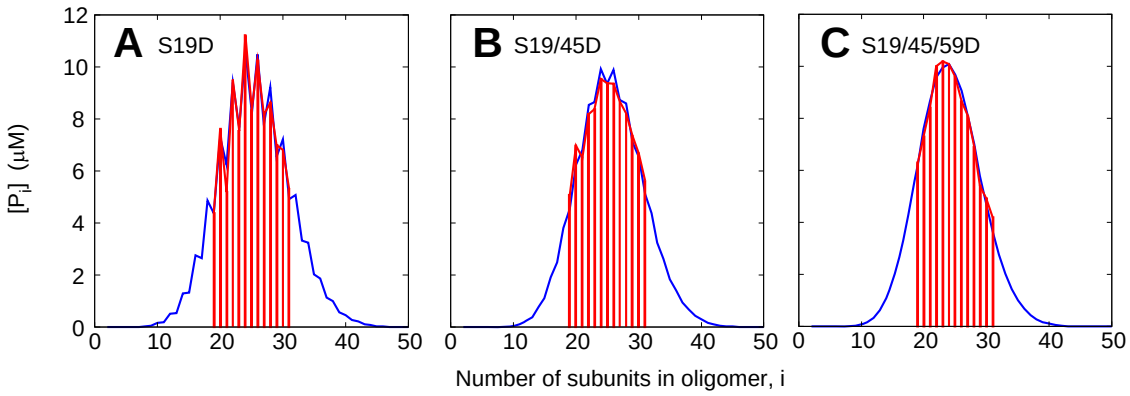

Figure S3. Distributions and fits of the phosphomimics S19D, S19/45D and S19/45/59D at 37^{o}C and pH 7. The distribution of S19D is comparable to the wild type under the same conditions, the distribution of S19/45D is at pH7 is comparable to the wild type distribution at pH6 and the distribution of S19/45/59D at pH7 is comparable to the wild type distribution at pH5.

Table S1 - Activation parameters from the temperature dependencies of $k^+[P_1]$, k^-_{e+d} and k^-_e obtained from fitting mass spectrometry kinetic data (Fig. 5) to the Eyring equation Eq. S1[a].

		ΔH^* (kJmol^{-1})	ΔS^* (Jmol^{-1}K^{-1})
pH 5	$k^+[P_1]$	158 ± 12	259 ± 39
	k^-_{e+d}	156 ± 11	277 ± 38
	k^-_e	172 ± 14	289 ± 47
pH 9	$k^+[P_1]$	298 ± 10	663 ± 32
	k^-_{e+d}	298 ± 10	633 ± 32
	k^-_e	298 ± 10	661 ± 32

[a]Values of $k^+[P_1]$, k^-_{e+d} and k^-_e, obtained as a function of temperature, in the range 20^{o}C - 50^{o}C, and at pH 5 and pH 9 were found to show Arrhenius behaviour, allowing activation parameters ΔH^* and ΔS^* to be calculated. These were obtained, following transition state theory

$$k_{rate} = \nu\kappa \exp\left(-\frac{\Delta G^*}{RT}\right) = \nu\kappa \exp\left(-\frac{\Delta H^* - T\Delta S^*}{RT}\right) \quad \text{(S1)}$$

where κ is the transmission coefficient, $\nu = \frac{k_B T}{h}$, T is the temperature, k_B is Boltzmann's constant and h is Planck's constant. The value of $\nu\kappa$ was taken to be 3000 s^{-1} [3].

Supplemental References

1. Oosawa, F. & Kasai, M. (1962). A theory of linear and helical aggregations of macromolecules. *J Mol Biol* **4**, 10-21.
2. (2007). *Numerical Recipes*, Cambridge University Press.
3. Fersht, A. (1998). *Structure and Mechanism in Protein Science*, W.H. Freeman, New York.