Microfluidic Diffusion Analysis of the Sizes and Interactions of Proteins Under Native Solution Conditions

Paolo Arosio[1†], Thomas Müller[1,2†], Luke Rajah[1†], Emma V. Yates[1], Francesco A. Aprile[1], Yingbo Zhang[1], Samuel I. A. Cohen[1], Duncan A. White[1], Therese W. Herling[1], Erwin J. De Genst[1], Sara Linse[3], Michele Vendruscolo[1], Christopher M. Dobson[1], and Tuomas P. J. Knowles[1*]

[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK*

[2]*Fluidic Analytics Ltd, Cambridge, UK*

[3]*Department of Biochemistry and Structural Biology, Lund University, Box 124, SE221 00 Lund, Sweden P. O. Box 124, SE221 00 Lund, Sweden*

[†] these authors contributed equally

[*]to whom correspondence should be addressed: tpjk2@cam.ac.uk

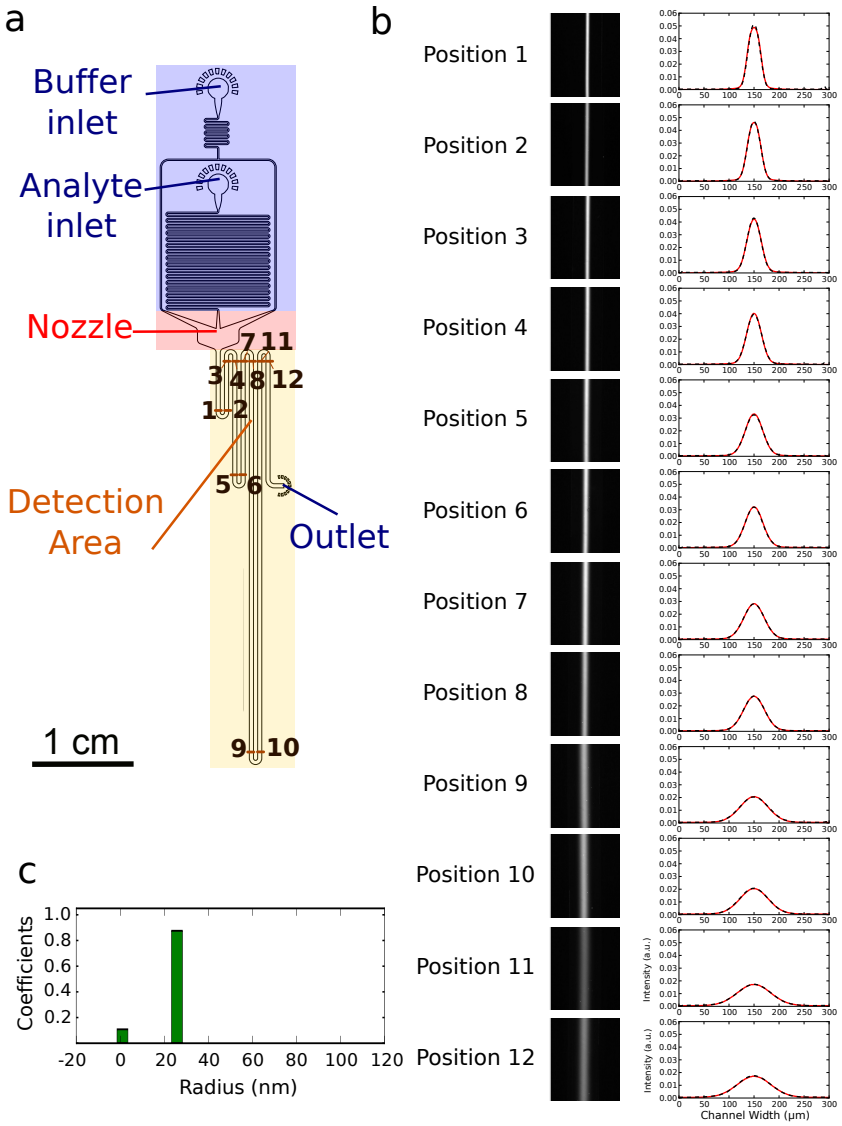

Figure S1: Global Fitting Procedure. a) Schematic of the diffusion microfluidic device. b) Images of diffused analyte (standard colloids with radius of 23.5 nm flowing at 40 μL/h) at 12 different positions along the channel as indicated in a), and comparison between the corresponding measured (black dot lines) and simulated (continuous red lines) diffusion profiles. The constraints of the fitting in time and space allow the evaluation of the diffusion coefficient with high accuracy. c) Size distribution of the analyte evaluated from the fit shown in b).

	Microfluidic device R_H (nm)	Dynamic Light Scattering R_H (nm)
Glucagon	1.18 ±0.28	1.29 ± 0.30
Insulin	1.76 ± 0.27	2.14 ± 0.05
β-Lactoglobulin	2.1 ± 0.28	2.37 ± 0.14
Bovine Serum Albumin	3.9 ± 0.64	4.4 ± 0.04
Nanoparticles 23.5 nm	22.2 ± 1.85	26.0 ± 0.11
Nanoparticles 100 nm	100	101.65 ± 1.69

Table S1: Comparison between the sizes of several proteins and standard nanoparticles as measured by the microfluidic device and by conventional dynamic light scattering. The values correspond to the data reported in Fig. 2a in the main text.

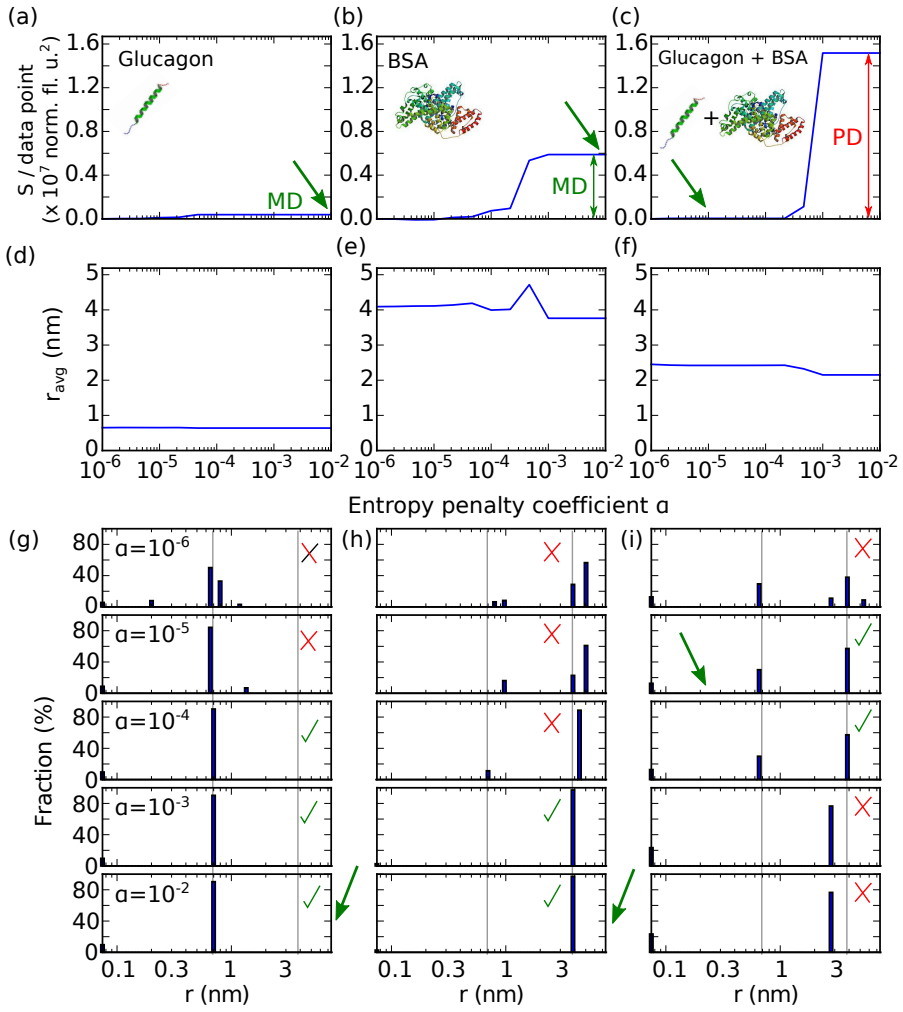

Figure S2: (a)-(c) Change in the sum of squared residuals, S, per data point (in normalized fluorescence units squared) upon increase of the entropy penalty coefficient α for pure glucagon, pure BSA, and a 1:1 mixture, respectively. The flow rate in experiments and simulated basis functions was 160 μl/h, and all experiments were performed in 20 mM phosphate buffer at pH 8.0 containing 20 $\%$ DMSO to prevent interactions between proteins in the mixture. These data were fitted to a linear combination of simulated basis functions using penalised least squares minimisation with a basin hopping algorithm implemented in python (Wales, D J, and Doye J P K, Journal of Physical Chemistry A, 1997, 101, 5111) performing 250 random displacements. According to Fig. 3a-b in the main text, these data show that we have monodisperse solutions (MD) in (a) and (b), and a polydisperse mixture (PD) in (c). Consequenlty, we should choose the coefficient α to be maximal in the analysis of (a) and (b), and consider the smallest coefficient α that provides a bimodal distribution in (c), which in this case leads to $\alpha = 10^{-5}$ (marked with the green arrow). (d)-(f) Average radius extracted by our analysis as a function of α. It can be seen that the value of the average radius is robust against changes in the entropy penalty, and that the average radius in the mixture lies between the values of the individual species. (g)-(i) Distribution of components with radius r in the samples as extracted from the entropy-penalised least squares fitting. At low entropy penalties, many species seem to be present; by increasing α, however, we can reduce the number of apparent species to only those that are relevant to describe the experimental data. In the extreme case of $\alpha = 10^{-2}$, the distributions are forced to resume a single peak corresponding to the average radius of the distribution. A small artefact peak at the minimal simulated radius is present at all entropy values.

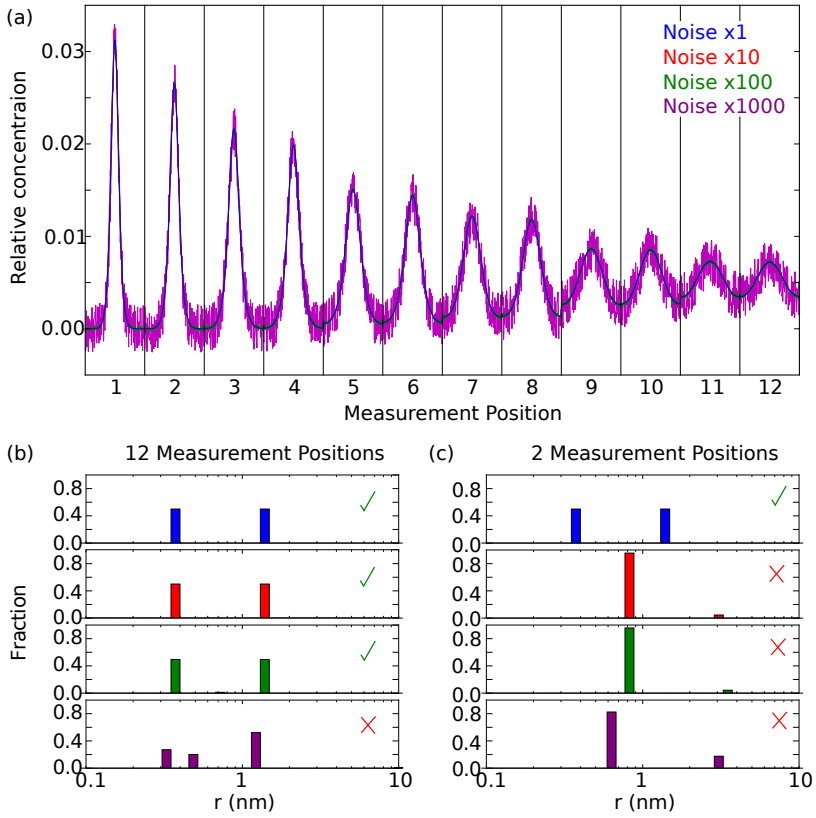

Figure S3: (a) Simulated profiles for a mixture between particles with radii 3.7 Å and 14.0 Å with random noise added at every data point. The base noise amplitude was uniformly distributed between $\pm 2.6 \times 10^{-6}$ in normalized fluorescence units (blue curve) and was increased by factors of ten each for the red, green and magenta lines, respectively. (b) Distribution of radii obtained by fitting the above profiles at 12 measurement positions with our standard fitting algorithm using 1000 basin hops. Only when the noise starts to have similar amplitude of the signal does the assignment of the radii become faulty. (c) Same as (b) but considering only the last 2 measurement positions. In this case, the analysis is much more error-prone.

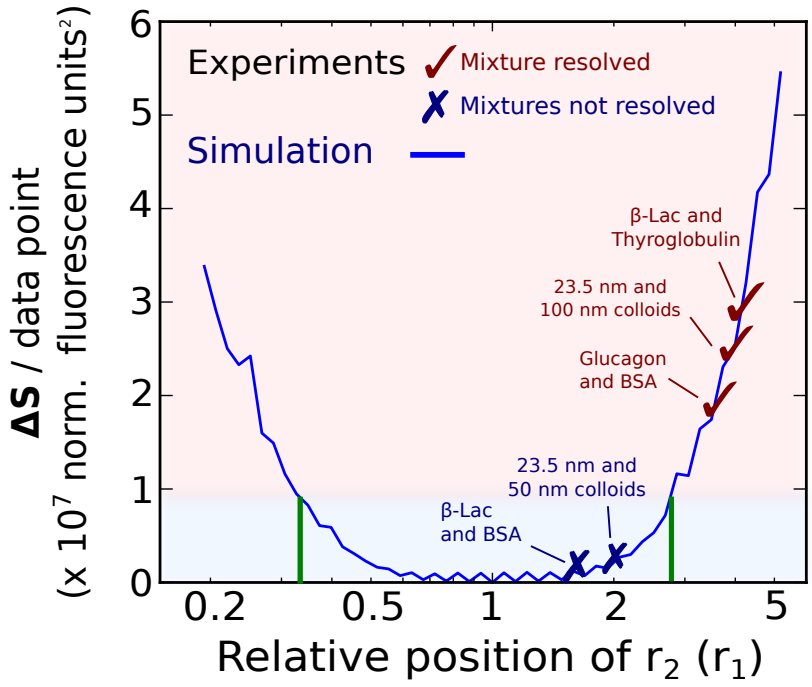

Figure S4: Simulated change in the sum of squared residuals per data point (in normalized fluorescence units squared) when forcing binary mixtures of particles with radius $r_1 = 3.7$ nm and variable r_2 to be represented by a single species. The oscillations for near-unity ratios r_2/r_1 are due to the fact that only a discrete number of radii are available, and hence the average radius can only be matched properly for even steps of r_2 away from r_1. Comparison with Fig. 3b in the main text indicates that, with the current experimental deviations, the sizes of two individual species must differ of a factor of roughly 3 in order to be resolved. This estimation is confirmed by experiments performed with binary mixtures of species exhibiting different ratios between the larger and the smaller component: in the situations where the ratio of sizes between the larger and the smaller species exceeds a factor of 3, the technique is able to resolve the sizes of the individual components within the binary mixtures.

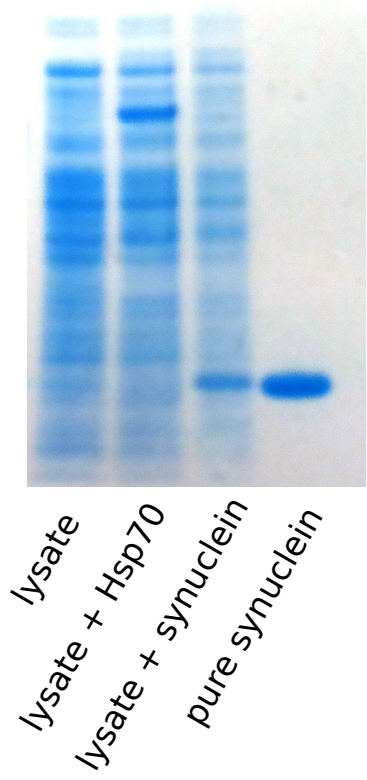

Figure S5: SDS-PAGE analysis of the samples considered in the immuno-binding assay between the antibody fragment and α-synuclein.

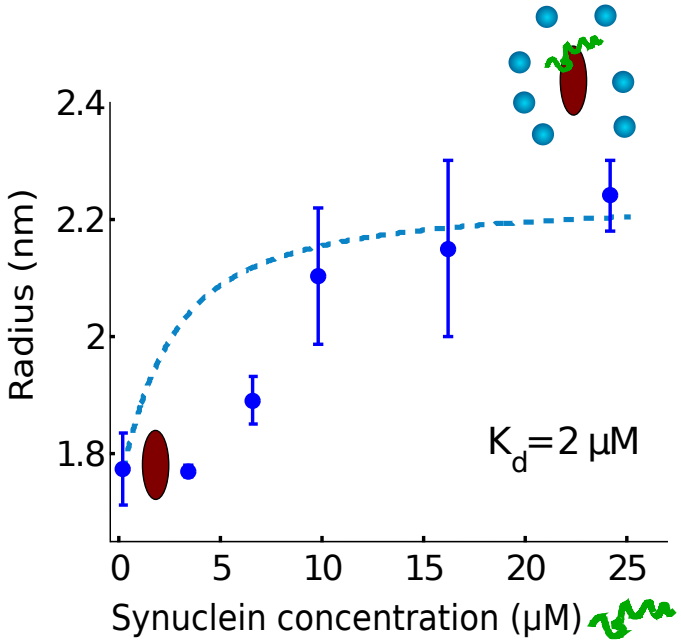

Figure S6: Titration curve of the binding between the nanobody and α-synuclein at increasing concentrations of α-synuclein and at the reference nanobody concentration of 1 μM. The increase of the average size indicates the formation of the complex when the concentration of the substrate exceeds the K_{d}. For details of the experiment see Methods section.

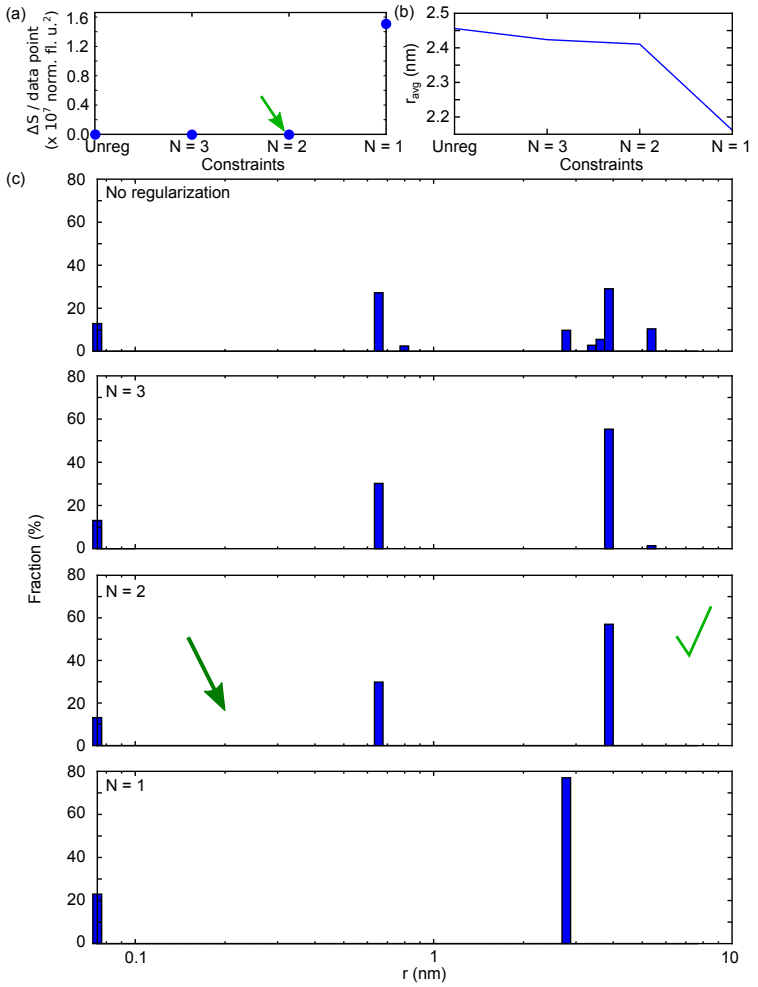

Figure S7: The approximately 1:1 mixture of glucagon and BSA discussed in Fig. S2 was re-analyzed by an alternative, equivalent method which costrains directly the number of components present within the mixture. (a) Change in the sum of squared residuals, S, per data point (in normalized fluorescence units squared) upon decreasing the number of components considered in the fit: from left to right, the data points represent the results for a completely unconstrained fit and for the constraint of 3, 2, or 1 species allowed in the fit. In addition to the indicated number of components, an artefact peak at the minimal simulated radius is allowed for all contraints. As in Fig. S2, the data were fitted considering a linear combination of numerically simulated basis functions using least squares minimisation with a basin hopping algorithm, performing 100 random displacements. The increase in residuals upon constraining the number of species to three or two with respect to the unconstrained fit is negligible. By contrast, forcing the fit to consider only a single species (plus the artefact peak) leads to a significant jump in the value of the residuals, in agreement with the behavior expected for a polydisperse solution discussed in Fig. S2. (b) Average radius extracted by our analysis upon constraining the fits. (c) Distribution of components with radius r in the analyte as extracted from the unconstrained and the constrained least squares fitting.