# Arginine to lysine mutations increase the aggregation stability of a single chain variable fragment through unfolded state interactions

**DOI:**
10.1021/acs.biochem.9b00367

**Document Version**
Accepted author manuscript

Biochemistry



Article

# Arginine to lysine mutations increase the aggregation stability of a single chain variable fragment through unfolded state interactions

James Austerberry, Angela Thistlethwaite, Karl Fisher, Alexander P. Golovanov, Alain Pluen, Reza Esfandiary, Christopher F. van der Walle, Jim Warwicker, Jeremy P. Derrick, and Robin Curtis

*Biochemistry*, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.9b00367 • Publication Date (Web): 17 Jul 2019

# Arginine to lysine mutations increase the aggregation stability of a single chain variable fragment through unfolded state interactions

James I. Austerberry [ab], Angela Thistlethwaite [a], Karl Fisher [b], Alexander P. Golovanov [bc], Alain Pluen [d], Reza Esfandiary [e], Christopher F. van der Walle [f], Jim Warwicker [bc], Jeremy P. Derrick [a], Robin Curtis [bg*]

a Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, United Kingdom
b Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, United Kingdom
c School of Chemistry, University of Manchester, Manchester, M1 7DN, United Kingdom
d Manchester Pharmacy School, University of Manchester, M13 9PL, United Kingdom
e Dosage Form Design & Development, AstraZeneca, Gaithersburg, MD20878,
f USA Dosage Form Design & Development, AstraZeneca, Granta Park, Cambridge, CB21 6GH, United Kingdom
g School of Chemical Engineering and Analytical Science, University of Manchester, M1 7DN, United Kingdom

*denotes corresponding author. Email address: R.curtis@manchester.ac.uk

**Abstract**

Increased protein solubility is known to correlate with an increase in the proportion of lysine over arginine residues. Previous work has shown that the aggregation propensity of a single-chain variable fragment (scFv) does not correlate with its conformational stability or native-state protein-protein interactions. Here we test the hypothesis that aggregation is driven by the colloidal stability of partially unfolded states, studying the behaviour of scFv mutants harbouring single or multiple site-specific arginine/lysine mutations in denaturing buffers. In 6 M guanidine hydrochloride (GdmCl) or 8 M urea, repulsive protein-protein interactions were measured for the wild-type and lysine enriched (4RK) scFvs on account of weakened short-ranged attractions and increased excluded volume, in contrast to the arginine enriched (7KR) scFv which demonstrated strong reversible association. In 3 M GdmCl, the minimum concentration at which the scFvs were unfolded, the hydrodynamic radius of 4RK remained constant but increased for the wild-type and especially for 7KR. Individually swapping lysine back to arginine in 4KR indicated that the observed aggregation propensity of arginine in denaturing conditions was non-specific. Interestingly, one such swap generated a scFv with especially low aggregation rates under low/high ionic strength and denaturing buffers; molecular modelling identified hydrogen-bonding between the arginine side chain and main chain peptide groups, stabilising the structure. The arginine/lysine ratio is not routinely considered in biopharmaceutical scaffold design, or current amyloid prediction methods. This work therefore suggests a simple method to increase the stability of a biopharmaceutical protein against aggregation.

## Introduction

Design and optimisation of bioprocessing steps for the production and manufacture of therapeutic monoclonal antibodies (mAbs) requires control over their physical and chemical instabilities such as unfolding, adsorption, deamidation, oxidation and aggregation.[1] Control over physical instabilities such as aggregation and the generation of particulates becomes more challenging as mAb concentration increases to meet, for example, high concentration (> 100 mg/ml) liquid formulations for subcutaneous injection.[2,3] Understanding and prediction of such instabilities involves multi-disciplinary effort between bioprocess and design teams in which the solution behaviour of mAbs can be reviewed and optimised.[4]

Protein aggregation can occur through different mechanisms, which can be generally described by models including an aggregate formation step through a nucleation event followed by aggregate growth. The formation of the smallest stable aggregate can involve reversible association of partially folded states to form an oligomer. Aggregate growth can occur through monomer addition, commonly referred to as chain polymerisation, or by aggregate condensation and precipitation.[5–7] The rates of the individual steps are determined by multiple characteristics including the colloidal stability of the aggregate precursors, the native-state conformational stability, and the propensity to undergo conformational rearrangements. The complex nature of protein aggregation makes prediction problematic, with no single measureable parameter able to describe the protein's propensity to aggregate.[8,9]

A number of aggregate prediction algorithms focus on the propensity of the protein to form amyloid-like structural motifs, using as inputs amino acid sequence and secondary structure to identify regions prone to rearrange into β-sheet structures.[10–13] Unfortunately, these aggregation predictors do not take into account colloidal or conformational stability,[14,15] although surface hydrophobic patches on native, transient or dynamic structures are indirect measures.[14,16,17] Applying these algorithms to mAbs will require improved prediction of aggregation prone regions on intermediate structures.[18,19] Identifying aggregation prone regions via protein mutation and analysis provides complementary data to *in silico* methods. Aggregation prone regions for mAb variable, heavy ($V_H$) domain fragments have been generally determined in the vicinity of the complementarity determining region (CDR).[20,21] Partial unfolding also exposes aggregation prone hydrophobic residues buried in the native core[8,22] Aggregation pathways driven through buried residues in transient structural intermediates are difficult to predict *in silico* and experimentally verify.[22–24]

Contributions of the protein surface electrostatic and short range attractive interactions to aggregation propensity can be elucidated through the study of native state protein-protein interactions. A large repulsive electrostatic surface charge in combination with short-range attraction will generally result in aggregate growth via chain polymerization since aggregate-aggregate coalescence must overcome the long range electrostatic barrier. If this electrostatic barrier is removed by changing pH to minimise surface net charge, aggregate coalescence occurs and leads to larger, denser aggregates that appear more amorphous.[7,25–29] Thus, measurement of reversible self assembly of protein monomers may not predict aggregation driven by transient, partially unfolded states, which presents a challenge to a full understanding of protein aggregation propensity.

Temperature is often used to promote protein unfolding in order to accelerate and study protein aggregation.[30–32] Properties such as the melting temperature can be used as an indicator of aggregation propensity in early mAb formulation screening. However, because high temperatures can alter the relative populations of fully unfolded to partially unfolded states and the relative rates of aggregate growth and nucleation, extrapolating temperature-induced behaviour to long-term room or storage temperature behaviour does not always follow Arrhenius modelling for the calculation of the mAb product shelf-life.[19,29,33–35] Non-native protein aggregation can also be investigated using chemical denaturants such as urea and guanidine hydrochloride (GdmCl). GdmCl induced aggregation differs significantly with respect to thermal-induced aggregation since chemical denaturants weaken intermolecular association leading to reduced aggregate growth.[36–38] At denaturant concentrations above the unfolding transition of a protein, aggregation no longer occurs and the colloidal stability of the fully unfolded protein can be directly quantified in terms of the osmotic second virial coefficient ($B_{22}$) or the protein-protein interaction parameter ($k_D$) obtained from diffusion coefficient measurements.[39] A key question is whether or not such measurements for fully unfolded proteins reflects the behaviour of transient intermediates involved in aggregation under native conditions. $B_{22}$ measurements of eight different proteins in solutions containing 6 or 8 M GdmCl were found to correlate with the fraction of the protein refolded from inclusion bodies in solutions containing 1 M GdmCl. This provides evidence that protein-protein interactions can be scaled with denaturant concentration.[39]

As described above, control of mAb instability is a multi-disciplinary effort and before excipients are explored in the context of stabilising a candidate mAb drug substance, the mAb sequence may be modified during lead selection. Approaches include the introduction of amino acid mutations which: i) supercharge the mAb protein by introducing multiple charge mutations of either acidic [8,40] or basic [41] residues; ii) target charged residues near aggregation 'hotspots' in the CDR [21,42], or residues in amyloidogenic sequences;[43] iii) change the arginine to lysine ratio, wherein a higher lysine content is correlated with increased solubility.[44] Our previous work has shown that mutating arginine into lysine reduces the aggregation propensity of a human recombinant single chain fragment variable (scFv).[29] However, the mechanism underpinning our observed increase in solubility was not fully determined and here we study the effect of changing the arginine to lysine ratio of the scFv in more detail. Protein-protein interactions and aggregation behaviour were determined under both native and denaturing conditions and the results provide insight into why the arginine-enriched mutants are more aggregation prone.

## Methods

### Mutant Design

Individual mutants of an anti-c-Met scFv[45] were created based on the four locations of a lysine rich mutant termed 4RK, described previously.[29] These mutants were termed R98K, R156K, R214K and R252K, using the single letter amino acid code and numbering according to sequence position, and generated by ThermoFisher GeneArt, UK, with optimised codon usage for *E.coli*. A further mutant, termed 7KR, had 7 arginine residues substituted for 7 lysine residues, with the sites located more extensively over the protein surface.[29] Extinction coefficients were determined from sequence using ProtParam.[46]

### Protein Purification

Protein purification was carried out as previously stated using Protein A affinity and size exclusion chromatography, the latter equilibrated in 25 mM sodium acetate, pH 5.[29] Following purification, protein samples were concentrated to 10 mg/ml in 25 mM sodium acetate, pH 5, and stored at -80 °C.

**Sample preparation**

250 µl of a protein sample was placed in 3.5 kDa MWCO GeBAflex-Midi Dialysis tubes (Generon, UK) and dialysed against 1 L of a solution containing 25 mM sodium acetate, pH 5, for 1 hour and then a fresh dialysis was carried out overnight. The same procedure was used to also dialyse protein samples into the same buffer with addition of 150 mM NaCl. After dialysis, protein samples were concentrated to 40 mg/ml by ultracentrifugation using Vivaspin 500 10,000 MWCO, and then filtered through a 0.02 µm polypropylene syringe filter (Whatman, UK). Final protein concentration was measured by absorbance at 280 nm. Stock denaturant solutions of 10 M urea and 8 M GdmCl by mass were prepared with gentle (~50 °C) solubilisation in 25 mM sodium acetate, with final pH adjusted to pH 5, before 0.02 µm filtration (Whatman, UK). Stock solutions were rapidly diluted with 40 mg/ml protein solution in 25 mM sodium acetate, pH 5, to give final concentrations of 8 and 10 mg/ml in 8 M urea and 6 M GdmCl respectively. These solutions were then used for determining the protein-protein interaction parameter $k_{\mathrm{D}}$ or for isothermal aggregation studies.

**Static light scattering (SLS) measurements of aggregation**

Thermal-induced aggregation behaviour was monitored by SLS, undertaken on an Uncle instrument (Unchained Labs, UK), which contains a laser with wavelength of 473 nm. The sample environment was equilibrated at 36.5 °C for 10 mins before measurements. Increases in light scattering intensity reflect concomitant increases in weight average molecular weight of the protein.[47] 9 µL of protein sample at 1 mg/mL in 25 mM sodium acetate, pH 5, with/without 150 mM NaCl, was loaded into each microcuvette. SLS experiments were carried out under isothermal conditions at 36.5 °C and using a temperature ramp with 0.5 °C increments between 20 and 80 °C with a 30 s delay after each increment to allow for equilibration. Each sample scan took ~2 s. All measurements were taken in triplicate.

**Dynamic light scattering (DLS) measurements of aggregation and determination of interaction parameter $k_D$.**

The Wyatt Dynapro Nanostar system was used for DLS measurements (Wyatt, UK). DLS was used for quantifying isothermal aggregation at 25 °C in terms of the time evolution of an apparent hydrodynamic radius, which is calculated from the measured protein diffusion coefficient using the Stokes-Einstein relation. In addition, DLS measurements were carried out under non-aggregating conditions as a function of protein concentration to determine the protein-protein interaction parameter $k_{\mathrm{D}}$ The Nanostar uses laser light scattering at 658 nm and a scattering angle of 90 °C. 9 µL of protein in 25 mM sodium acetate, pH 5, with/without 150 mM NaCl, was loaded for each measurement in a low volume quartz cuvette.

Diffusion coefficients were obtained from cumulant fits to the correlation function over the time delays in the range of 10 to 6 × $10^4$ µs using the DYNAMICS software. The minimum cut-off in the time delay was chosen to avoid artefacts in the analysis arising from the effect of high

concentrations of either urea or GdmCl on the autocorrelation function. For protein free solutions, there is an observable decay in the correlation function between 0.5 and 4 µs (Figure S1). This effect was omitted by only using the interval over delay times in the range of 10 and 100000 µs when fitting the correlation function.

***Sample preparation and collection for $k_D$ determination:*** Protein samples in denaturant (see sample preparation) were diluted with 6 M GdmCl or 8 M urea in 25 mM sodium acetate, pH 5, at room temperature to a concentration range between 2 to 10 mg/ml for $k_D$ determination. The DLS acquisition time was set to 5 s and 15 collections were taken per sample and each sample was run in triplicate

$k_{\mathrm{D}}$ was determined from measurements of diffusion coefficients *D* (under non-aggregating conditions) as a function of protein concentration over the range of 1 and 10 mg/mL at 25 °C using the relation

$$D = D_0[1 + k_{\mathrm{D}}c] \quad (1)$$

where $D_0$ is the infinite dilution diffusion coefficient. An infinite-dilution hydrodynamic radius $R_{\mathrm{H0}}$ is calculated from the extrapolated value for $D_0$ using the Stokes-Einstein equation:

$$R_{\mathrm{H0}} = \frac{k_{\mathrm{B}}T}{6\pi\mu D_0} \quad (2)$$

where $k_{\mathrm{B}}$ is the Boltzmann constant, *T* is temperature and solvent viscosity $\mu$ values were taken from literature to account for the high concentration of chemical denaturant.[48,49]

***Sample preparation for isothermal aggregation:*** Samples at either 4 mg/ml and 2 mg/ml protein concentration were initially prepared in solutions with 6 M GdmCl, in 25 mM sodium acetate, pH 5, buffer. The sample was then diluted to 3 M GdmCl concentration by 50:50 dilution with a buffer-only solution to a total volume of 40 µl. The DLS acquisition time was set to 5 s and 3 collections were taken per time point with measurements taken immediately after dilution every 15 s up to 2500 s.

**Modelling**

To investigate whether the structural environment of R214 may aid understanding of the importance of this site upon mutation to lysine, the scFv sequence was searched against the structural database[50] using BLAST.[51] From the aligned sequences that maintain an arginine in the equivalent position, an Fv fragment (Protein Data Bank, PDB, ID 2OTU) was selected. This scaffold was used to assess the change of electrostatic interactions for an arginine to mutation at the site equivalent to R214 of scFv. Side chain interactions for the arginine in 2OTU are largely with the main chain component of two loops, each associated with a pair of β-strands. Additionally, the arginine side chain is relatively buried, so that the effect of mutation to lysine is dominated by whether the lysine side chain can reproduce the hydrogen-bonding interactions between arginine side chain and neighbouring loop main chains. Electrostatic interactions of arginine and modelled lysine side chains were assessed using pKa calculations.[52] A lysine side chain was modelled with Swiss PDB Viewer [53], maintaining a conformer that is, so far as possible, isosteric with the native arginine.

**Fluorescence based protein unfolding**

Protein samples at 100 µg/ml in 25 mM sodium acetate, pH 5, were prepared over a range of GdmCl concentrations from 0 to 3 M in 0.2 M increments, in the same buffer. 40 µl samples of each were pipetted into a Corning Cyclic Olefin Copolymer 384 well plate and measured using a Varioscan Lux fluorescence plate reader (Thermo Fisher Scientific, UK). Excitation was performed using a laser at 295 nm, where the emission was measured at 325 and 350 nm to determine the ratio between the intensities at these wavelengths.

**Results and Discussion**

*Arginine to lysine mutations reduce aggregation propensity*

Static light scattering profiles as a function of temperature for the WT, 4RK or 7KR scFv mutants in 25 mM sodium acetate, pH 5, with/without 150 mM sodium chloride are shown in Figure 1. The temperature onset of aggregation $T_{agg}$ refers to the temperature when the light scattering signal is greater than the mean baseline intensity by 10% of its value. $T_{agg}$ decreased in order of: 4RK > WT > 7KR (~50, 47 and 40 °C, respectively); with higher $T_{agg}$ values indicating lower aggregation propensity. Upon addition of 150 mM NaCl to the buffer solution, this trend was still observed although the aggregation temperatures were lower. In solutions with 150 mM NaCl, the $T_{agg}$ for 4RK, WT and 7KR was 43, 40 and 27 °C, respectively. This reduction is due to screening of the repulsive double layer forces by sodium chloride leading to a lower colloidal stability which increases the aggregation growth rates.[29,43,54–56] The mutant-specific effects observed at high and low ionic strength indicates the changes in aggregation propensities cannot be only rationalized in terms of differences in electrostatic interactions. In addition, the aggregation propensities are not controlled by the conformational stabilities, as the 4RK mutant has been shown to be less conformationally stable than the WT, while the 7KR native structure is the most stable.[29] The similar rate of SLS intensity increase observed for each of the mutants may indicate that the mechanisms of aggregate growth of the three proteins are similar to each other.[57]

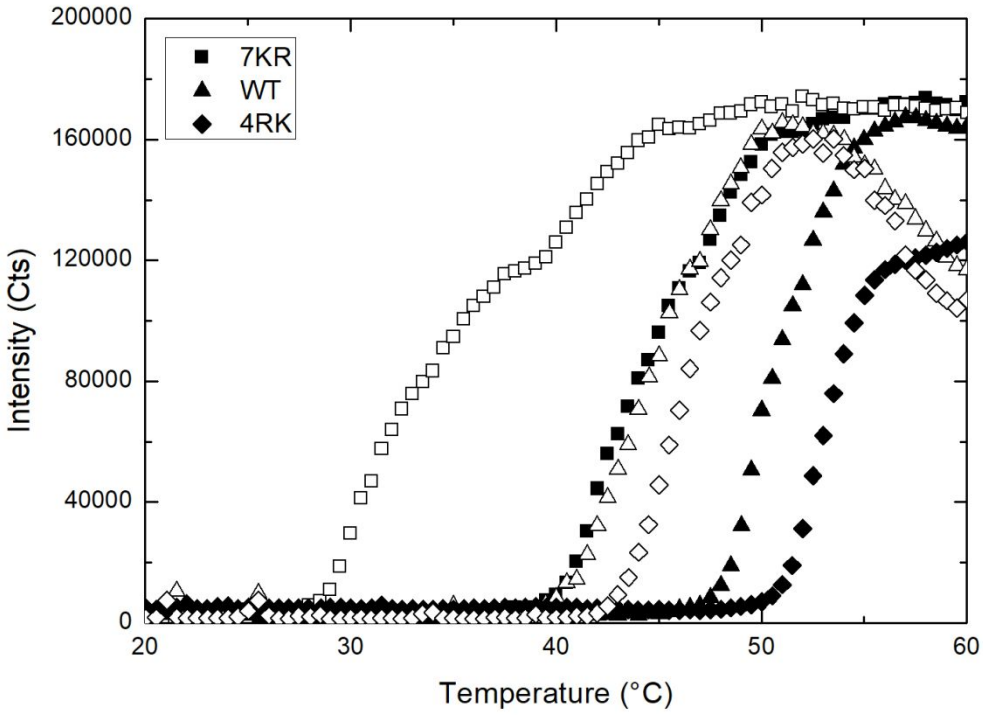


**Figure 1:** SLS intensity measured for WT, 7KR and 4RK scFvs during heating from 20 to 60 °C at 2 mg/ml in solutions containing 25 mM sodium acetate, pH 5 (closed symbols) or with addition of 150 mM sodium chloride (open symbols).

In our previous work [29], we found that the aggregation propensity of the mutants did not correlate with the conformational stability or the native-state protein-protein interactions, which, led us to hypothesize that the main impact of the mutations was to alter the interactions or the colloidal stability of the unfolded or partially folded states. The other possibility that the mutation alters the β-sheet forming propensity was disregarded because predictors of aggregation prone regions, which primarily reflect the propensity to form β-structure, do not distinguish between lysine and arginine. However, there is no direct evidence for the effect of the mutation on the colloidal stability of the unfolded states. As such, we have carried out protein-protein interaction measurements in solutions with 6 M GdmCl, conditions where the scFv proteins only undergo reversible interactions in the unfolded states. In this case, there are no conformation rearrangements to β-structure which would lead to irreversible association or aggregation, so the measurement only reflects the colloidal stability.

*7KR mutant exhibits much stronger self association under highly denaturing conditions*

The protein-protein interaction parameter, $k_D$, is often used as a surrogate parameter for the osmotic second virial coefficient, $B_{22}$, which is related to the interaction free energy between a pair of proteins averaged over their relative orientations and separations. There exists a monotonic correlation between $k_{\mathrm{D}}$ and $B_{22}$ for solutions of lysozyme or monoclonal antibodies under native conditions[58–60] and for lysozyme under chemically denaturing conditions.[61] Determining the net effect of protein-protein interactions first requires knowledge of the excluded volume contribution

to $k_D$, which exists for all proteins. Increasingly positive values reflect increasing protein-protein repulsion, while attractive protein-protein interactions are reflected by decreasing values of $k_D$. The change in attractive versus repulsive interactions can be observed clearly in Figure 2 for the three mutants in buffer with/without denaturant. Under native conditions, the excluded volume contribution can be approximated by using the result for hard spheres given by $k_D = 1.55 V_p N_A / M$ where $N_A$ and $M$ are Avogadro's number and protein molecular weight respectively, and $V_p$ is the protein molecular volume assuming spherical geometry, $V_p = 4\pi R_H^3/3$. For example, from Table 1, measured values of $R_H$ equal to 2.7 nm, give an excluded volume contribution equal to 3 mL/g.

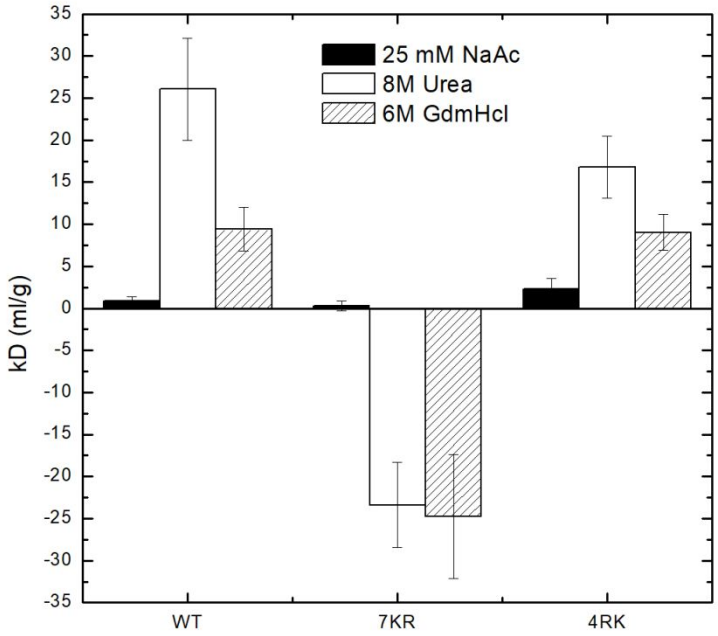


**Figure 2 :** Protein-protein interaction ($k_D$) values for WT, 7KR and 4RK scFvs in 25 mM sodium acetate, pH 5, alone or containing 6 M GdmCl, or 8 M urea denaturant, at 25 °C.

**Table 1 :** $R_{H0}$ values of WT, 4RK and 7KR scFvs in 25 mM sodium acetate, pH 5, alone or with addition of 8 M urea, or 6 M GdmCl

| scFv | $R_{H0}$ (nm) | | |
|---|---|---|---|
| | Native | 8 M Urea | 6 M GdmHCl |
| WT | 2.8 ± 0.1 | 4.3 ± 0.1 | 3.6 ± 0.1 |
| 4RK | 2.7 ± 0.1 | 4.5 ± 0.1 | 4.1 ± 0.1 |
| 7KR | 2.7 ± 0.1 | 4 ± 0.1 | 3.4 ± 0.1 |

The effects of urea and GdmCl on the $k_D$ of the WT and 4RK mutants can be rationalized in terms of the denaturants ability to weaken inter and intramolecular interactions. Under native conditions, the $k_D$ values are near to the excluded volume value, which indicates that the net contribution from other forces cancel each other out. It is likely that there exists an electrostatic repulsion arising from the net charge on the proteins due to the low ionic strength conditions, which is balanced by a short ranged attraction.[27] Increasing urea concentration is expected to weaken any short-ranged attraction and increase the excluded volume contribution due to the increase in protein size upon unfolding.[39,62,63] It is the combination of these effects that causes the increase in protein-protein repulsion for 4RK and for WT. The $k_D$ values are lower for solutions of GdmCl versus urea, indicating the protein-protein interactions for WT and 4RK are less repulsive. The likely explanation is that GdmCl has screened the electrostatic repulsion, although, the sizes are also smaller in GdmCl so that the excluded volume contribution is also expected to be smaller.

While there is no direct evidence for self association of WT or 4RK at high denaturant concentrations, the negative values of $k_D$ measured for 7KR provide a clear indication of a strong reversible association in solutions of either denaturant. Replacing the lysines with arginines in the 7KR mutant causes association of the unfolded state that exists even at a denaturant concentration of 6 M GdmCl. While this finding might seem surprising, it is consistent with measured patterns of protein-protein interactions in solutions containing the amino-acid arginine. Ho and Middelberg (2004) showed that protein-protein attractions, characterized in terms of $B_{22}$ measurements, covering a set of eight proteins when dissolved in solutions containing 6 M GdmCl are weakened when adding 500 mM arginine chloride. The attenuation of attractive interactions indicates preferential binding of arginine to the protein is not screened by the high concentration of guanidinium ion. It is not surprising then that arginine, when part of the protein surface, can form preferential binding interactions in solutions at high denaturant concentration. The stronger interactions of arginine likely occur with aromatic groups on the protein surface, which has been deduced from the stronger effectiveness of arginine over guanidnium at salting in aromatic amino acids[64] and other small molecules containing aromatic ring groups.[65,66] The strong preference for aromatic groups has also been observed from molecular simulation studies,[67,68] the identification of arginine binding sites near to aromatic groups in protein crystal structures,[69] and a survey of the macromolecular structure database indicating approximately one-half of all protein complexes contain at least one arginine-aromatic stabilizing interaction.[70] As such, it is very likely that the sticky interactions observed with 7KR over the WT arise from the additional arginines interacting with aromatic groups exposed in the unfolded states.

The $R_{H0}$ values (Table 1) follow a similar pattern to the $k_D$ measurements indicating the mutations have similar effects on intra and intermolecular interactions. The arginine-enriched mutant is more compact in the presence of denaturant reflecting more attractive intramolecular interations. In addition each of the mutants has a smaller size in GdmCl versus urea possibly due to screening of intramolecular electrostatic repulsion by the salt.

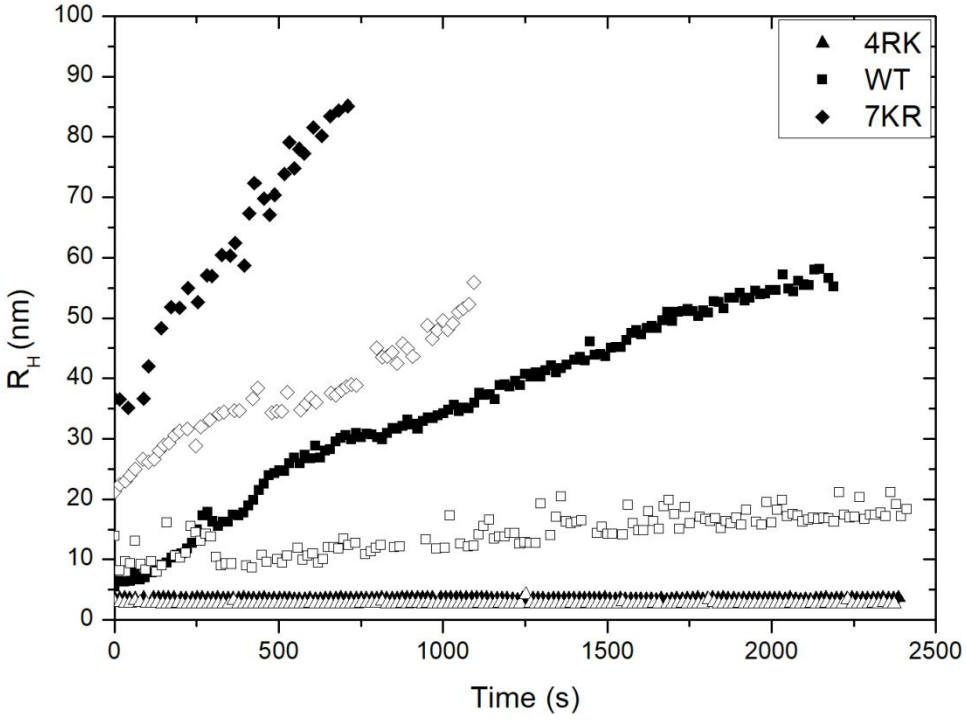


**Figure 3**: $R_h$ values for WT, 4RK and 7KR scFvs at 2 mg/ml (solid) and 1 mg/ml (open) in 25 mM sodium acetate, pH 5, containing 3 M GdmCl, at 25 °C.

The $k_D$ measurements provide strong evidence that the 7KR mutant has much stronger self-association in the unfolded state (Figure 2). To further quantify the unfolded-state interactions lower denaturant concentrations were used so that effects of protein-protein association should become more apparent. As such, protein aggregation behaviour was measured in solutions with 3 M GdmCl (Figure 3), which corresponds to the minimum denaturant concentration where each protein is in the unfolded state.[29] An increase in $R_{\mathrm{H}}$ was observed for WT and 7KR mutants over time, reflecting an aggregation process, whilst $R_{\mathrm{H}}$ for 4RK remained relatively constant indicating very little, if any, aggregate growth. Under these conditions, the WT is significantly more aggregation prone than the 4RK mutant, which further correlates with the trend that arginine promotes association of the unfolded state.

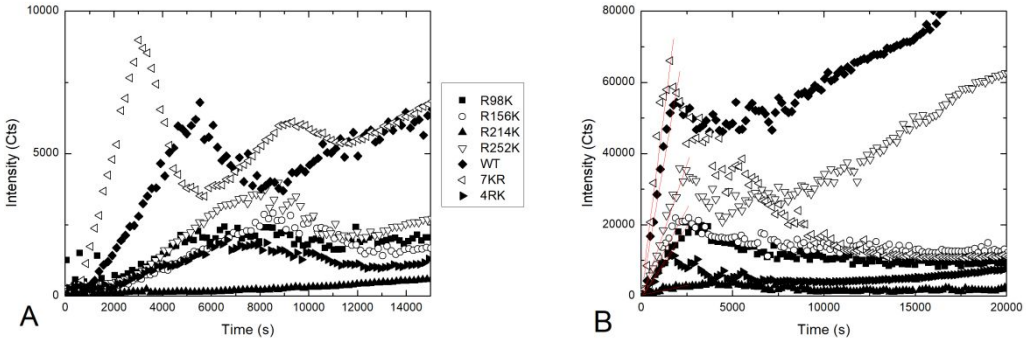

**Figure 4**: SLS intensity over time for R98K, R156K, R214K, R252K, WT, 7KR and 4RK scFvs at 2 mg/mL in 25 mM sodium acetate, pH 5, (A) and in the same buffer with 150 mM NaCl (B), at 36.5 °C.

**Table 2**: Aggregation rates for mutants at 2 mg/ml in 25 mM sodium acetate, pH 5, with/without 150 mM NaCl (high/low ionic strength buffer), at 36.5 °C.

| scFv | Aggregation rate low ionic strength (Cts/s ± S.E.) | Aggregation rate high ionic strength (Cts/s ± S.E.) | Standard Error |
|---|---|---|---|
| R98K | 0.29 ±0.04 | 8.0 | 0.44 |
| R156K | 0.32 ±0.01 | 9.5 | 0.70 |
| R214K | 0.02 ± 0.001 | 1.1 | 0.12 |
| R252K | 0.49 ± 0.02 | 15.0 | 0.86 |
| WT | 1.32 ± 0.04 | 29.4 | 2.32 |
| 7KR | 3.26 ± 0.12 | 39.9 | 4.77 |
| 4RK | 0.25 ± 0.01 | 7.5 | 0.59 |

Four single-point mutants were produced to ascertain how each mutation in 4RK contributed to its aggregation propensity relative to the WT. Each mutant contained a single arginine to lysine mutation corresponding to the 4RK protein at positions 98 (R98K), 165 (R165K), 214 (R214K) and 252 (R252K). In order to understand the contributions of each residue to the stability of the 4RK mutant, the SLS of each mutant was monitored during an isothermal hold at 36.5 °C (Figure 4). For each mutant there is an initial linear rate of scattering intensity increase, which was used to estimate an aggregation rate from the corresponding slope of the plot which is given in Table 2. Under both low and high ionic strength conditions, the aggregation rates for the scFvs increased in order of R214K < 4RK ≈ R98K ≈ R156K < R252K <WT < 7KR.

The aggregation propensity of the 4RK mutant is similar to two of the individual R to K mutants (R98K and R156K) and indicates that the effects of R to K mutations on aggregation propensity may not be additive. In order to check whether this discrepancy could be rationalized in terms of changes to protein conformational stability or protein structure, the denaturant unfolding curve for R214K was determined and compared against 4RK and the WT (Figure S2). There is a slight decrease in the unfolding curve to lower denaturant concentration for the R214K mutant, which is expected as arginine groups are known to contribute favourably to protein stability. The decrease however is less than observed for 4RK. Although the denaturant unfolding curves have not been measured for the other single mutations, it is likely that each of these either does not contribute or contributes unfavourably to the stability and that the larger reduction in stability observed for 4RK is due to the net additive effect from all the mutations. It has been shown that the introduction of arginine in place of lysine increases the conformational stability of mutants of GFP[71,72] through increased opportunities for stabilising hydrogen bond and salt bridge interactions. This is also consistent with the finding that 7KR has a much increased conformational stability which is likely due to the stabilizing effects of multiple KR mutations, rather than one mutation. Because each of the RK

mutations likely lowers the conformational stability, we hypothesize that the non-specific reduction in aggregation brought about by each mutation may be due to changes in how the protein-protein interactions occur between (partially) unfolded states, while in the WT, the native arginine residue located at position 214 may be located in a region responsible for self-association between monomers.

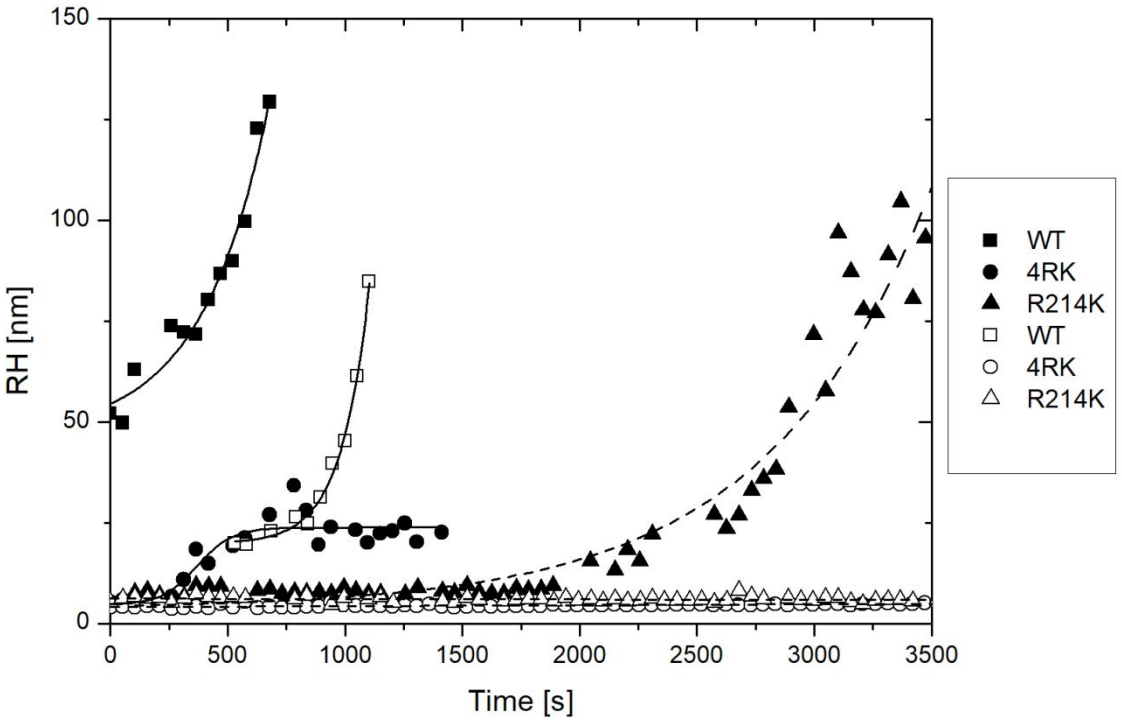


**Figure 5**: Change in $R_H$ over time as a measure of aggregation of 4RK, R214K and WT scFvs, initiated by dilution from 6 to 2.6 M (solid) or 2.8 M (open) GdmCl, final concentration 2 mg/ml, at 25 °C. Lines provided are purely illustrative.

Isothermal aggregation studies were carried out using DLS for the proteins in 2.6 or 2.8 M GdmCl solutions, chosen to differentiate between the WT scFv and 4RK/R214K mutants. The small increase in GdmCl concentration to 2.8 M attenuated the aggregation only of the mutants (Fig 5), i.e. substitution of K for R reduced aggregation likely through reduced association of the unfolded states, which were prevalent at these denaturant concentrations. The fact that the R214K mutant aggregates later than the 4RK mutant in 2.6 M GdmCl is of interest. If the whole population of each protein is entirely unfolded, it might be expected that the mutant containing the 4 lysine residues would be more stable than the mutant only containing one. However, aggregation pathways of unfolded proteins are complicated involving multiple steps including nucleation and aggregation growth. As an example there is no simplistic mechanism behind aggregate nucleation of the intrinsically disordered protein α-synuclein.[73] It is also interesting to note that the 4RK mutant achieves a stable aggregate size of ~20 nm which is an aggregation behaviour that differs from the WT and R214K mutant which exhibit an apparent exponential growth rate. This may be indicative of

different rate controlling steps in the aggregation pathway of 4RK versus either WT or the R214K mutant.

Interrogation of the structural database, through sequence alignment, yielded an Fv fragment (in PDB ID 2OTU) that has arginine at an equivalent position to R214 of scFv. From this comparison, the arginine is located between two structural loops of the scFv, each of which connects antiparallel β-strands. The arginine side chain is relatively buried and makes hydrogen-bond interactions with the main chain regions of the two loops. Lysine was also modelled at this position, to assess the difference between lysine side chain amino and arginine side chain guanidine groups, using pKa calculations that estimate charge interactions, including hydrogen-bonding. From calculated pKa changes relative to the side chains in free solution, the effect of interactions in the protein environment is predicted. For arginine in the scFv, a stabilisation of 5 kJ/mole is calculated, in comparison with a destabilisation of 7 kJ/mole for a lysine side chain at this location. It is concluded that that arginine side chain is involved in a hydrogen-bonding network with main chain peptide groups, which stabilises the structure, whilst the lysine side chain is not complementary to the surrounding loops, leading to destabilisation. These calculations refer to the folded scFv. Interestingly, one of the loops interacting with R214 (by analogy to 2OTU), neighbours R214 in sequence, so that that part of the arginine side chain to main chain interaction could be replicated in an unfolded form. Modelling this putative unfolded state local interaction gives a small stabilisation of the arginine side chain from electrostatic interactions with the main chain (1 kJ/mole). It would be possible, in principle, to reproduce the elements of the folded interaction between arginine side chain and two loops, from this local interaction in the unfolded state, with the juxtaposition of β-turn from a second protein molecule, and formation of a hydrogen-bonding network similar to that of the folded form. In this model, the R214K mutation would not only impact on stability of the folded state, but also on aggregation potential of the unfolded state.

A critical step in aggregation pathways often corresponds to forming intermolecular $\beta$-sheet structures that irreversibly trap the protein in an oligomeric state. A key question to address, therefore, is whether or not swapping between arginine and lysine can alter the $\beta$-sheet propensity, which can be assessed using amyloid predictors. We have tested a number of these predictors on our arginine-lysine mutants and found a negligible difference in their aggregation propensity scores and no change to the aggregation-prone regions of the protein (Table S1),[10,14,74] which is expected because the predictors have similar scoring functions between lysine and arginine. There is such a low propensity for lysine or arginine to be buried in $\beta$-sheets that the groups are classified as gatekeeper residues. Indeed, in some of the predictors, arginine is a more effective $\beta$-sheet breaker than lysine.[75] The effect of the arginine to lysine mutation is non-specific, as mutations are located both close to, and away from predicted aggregation-prone regions and β-sheet rich regions (including R252K located on the C terminus), yet each reduces the aggregation propensity of the protein (Figure 4). This provides further evidence that the mutations are not changing conformational rearrangement steps- there would otherwise be positional dependent changes based on their proximal location to the aggregation prone regions. If the R/K mutations do not impact conformational changes, it is likely that they promote protein 'stickiness' through favourable interactions with aromatics exposed in the unfolded states. This stickiness will decrease the colloidal stability of the protein, by increasing the percentage of successful collisions between proteins that form an aggregate, which is often quantified in terms of the Fuchs ratio.[55]

## Conclusion

We show that solvent-exposed arginine residues make ideal charge swap mutation candidates for improvement of colloidal stability of scFvs, where rational mutagenesis can be undertaken in a non-specific manner. This approach avoids the need to first identify aggregation prone regions.[76] Since arginine residues promote interactions with the unfolded state, we propose that their mutation to lysine will reduce the likelihood that protein-protein collision will result in aggregation, due to a reduction in surface stickiness. This method of charge swap mutation is also a successful approach for the identification of aggregation hotspot regions, as the difference in aggregation behaviour from a single residue substitution is large, without significant changes to the native protein structure. For the scFv and the arginine-lysine swap mutants, the protein-protein interactions measured under chemically denaturing conditions, rather than under native conditions, correlate with their aggregation propensity. This suggests the measurements could be a complementary tool for developing protein therapeutics, either in the discovery phase when choosing from a panel of therapeutics based, or when screening solution conditions to find the most stable formulation.

## Acknowledgements

Funding was provided by the Biotechnology and Biological Sciences Research Council (BBSRC grants BB/L006391/1 and BB/M006913/1) and AstraZeneca. Proteins were produced at the Manchester Institute of Biotechnology Protein Expression Facility.

## Supporting information files

Supporting information: A word document containing supplementary experimental data referenced in the main text.

**References**


1 Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M. and Gajiwala, K. (2014) Contribution of hydrogen bonds to protein stability. *Protein Sci.* 23, 652–661.

2 Parenky, A., Myler, H., Amaravadi, L., Bechtold-Peters, K., Rosenberg, A., Kirshner, S. and Quarmby, V. (2014) New FDA Draft Guidance on Immunogenicity. *AAPS J.* 16, 499–503.

3 Alford, J. R., Kendrick, B. S., Carpenter, J. F. and Randolph, T. W. (2008) High Concentration Formulations of Recombinant Human Interleukin-1 Receptor Antagonist: II. Aggregation Kinetics. *J. Pharm. Sci.* 97, 3005–3021.

4 Meisl, G., Kirkegaard, J. B., Arosio, P., Michaels, T. C. T., Vendruscolo, M., Dobson, C. M., Linse, S. and Knowles, T. P. J. (2016) Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* 11, 252.

5 Andrews, J. M. and Roberts, C. J. (2007) A Lumry−Eyring Nucleated Polymerization Model of Protein Aggregation Kinetics: 1. Aggregation with Pre-Equilibrated Unfolding. *J. Phys. Chem. B* 111, 7897–7913.

6 Gülseren, İ., Güzey, D., Bruce, B. D. and Weiss, J. (2007) Structural and functional changes in ultrasonicated bovine serum albumin solutions. *Ultrason. Sonochem.* 14, 173–183.

7 Brummitt, R. K., Nesta, D. P., Chang, L., Chase, S. F., Laue, T. M. and Roberts, C. J. (2011) Nonnative aggregation of an IgG1 antibody in acidic conditions: Part 1. Unfolding, colloidal interactions, and formation of high-molecular-weight aggregates. *J. Pharm. Sci.* 100, 2087–2103.

8 Carpenter, J. F., Ph, D., Barnard, J., Liu, L., Torres, R., Caplan, L., Winter, G. G., Freitag, A., Lewis, N., Dahl, K., et al. (2014) Stability engineering of the human antibody repertoire. *J. Phys. Chem. B* (Sanchez-Ruiz, J. M., ed.) 105, 1086–1096.

9 Andrews, J. M., Weiss IV, W. F. and Roberts, C. J. (2008) Nucleation, growth, and activation energies for seeded and unseeded aggregation of α-chymotrypsinogen A. *Biochemistry* 47, 2397–2403.

10 Tartaglia, G. G. and Vendruscolo, M. (2008) The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* 37, 1395–1401.

11 Garbuzynskiy, S. O., Lobanov, M. Y. and Galzitskaya, O. V. (2009) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* 26, 326–332.

12 Zibaee, S., Makin, O. S., Goedert, M. and Serpell, L. C. (2007) A simple algorithm locates β-strands in the amyloid fibril core of α-synuclein, Aβ, and tau using the amino acid sequence alone. *Protein Sci.* 16, 906–918.

13 Trovato, A., Seno, F. and Tosatto, S. C. E. (2007) The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* 20, 521–523.

14 Weiss, W. F., Young, T. M. and Roberts, C. J. (2009) Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J. Pharm. Sci.* 98, 1246–1277.

15 Meric, G., Robinson, A. S. and Roberts, C. J. (2017) Driving forces for nonnative protein aggregation and approaches to predict aggregation-prone regions. *Annu. Rev. Chem. Biomol. Eng.* 8, 139–159.

16 Chennamsetty, N., Voynov, V., Kayser, V., Helk, B. and Trout, B. L. (2009) Design of therapeutic proteins with enhanced stability. *Proc. Natl. Acad. Sci.* 106, 11937–11942.

17 Morriss-Andrews, A. and Shea, J.-E. (2015) Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* 66, 643–666.

18 Vagenende, V., Yap, M. G. S. and Trout, B. L. (2009) Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. *Biochemistry* 48, 11084–11096.

19 Chakroun, N., Hilton, D., Ahmad, S. S., Platt, G. W. and Dalby, P. A. (2016) Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Mol. Pharm.* 13, 307–319.

20 Perchiacca, J. M., Bhattacharya, M. and Tessier, P. M. (2011) Mutational analysis of domain antibodies reveals aggregation hotspots within and near the complementarity determining

regions. *Proteins Struct. Funct. Bioinforma.* 79, 2637–2647.
21 Perchiacca, J. M., Ladiwala, A. R. A., Bhattacharya, M. and Tessier, P. M. (2012) Aggregation-resistant domain antibodies engineered with charged mutations near the edges of the complementarity-determining regions. *Protein Eng. Des. Sel.* 25, 591–602.
22 Rayner, L. E., Hui, G. K., Gor, J., Heenan, R. K., Dalby, P. A. and Perkins, S. J. (2015) The solution structures of two human IgG1 antibodies show conformational stability and accommodate their C1q and FcγR ligands. *J. Biol. Chem.* jbc-M114.
23 Polverino de Laureto, P., Taddei, N., Frare, E., Capanni, C., Costantini, S., Zurdo, J., Chiti, F., Dobson, C. M. and Fontana, A. (2003) Protein Aggregation and Amyloid Fibril Formation by an SH3 Domain Probed by Limited Proteolysis. *J. Mol. Biol.* 334, 129–141.
24 Chiti, F., Stefani, M., Taddei, N., Ramponi, G. and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808.
25 Kim, N., Remmele, R. L., Liu, D., Razinkov, V. I., Fernandez, E. J. and Roberts, C. J. (2013) Aggregation of anti-streptavidin immunoglobulin gamma-1 involves Fab unfolding and competing growth pathways mediated by pH and salt concentration. *Biophys. Chem.* 172, 26–36.
26 Olsen, S. N., Andersen, K. B., Randolph, T. W., Carpenter, J. F. and Westh, P. (2009) Role of electrostatic repulsion on colloidal stability of Bacillus halmapalus alpha-amylase. *Biochim. Biophys. Acta (BBA)-Proteins Proteomics* 1794, 1058–1065.
27 Roberts, C. J. (2014) Therapeutic protein aggregation: mechanisms, design, and control. *Trends Biotechnol.* 32, 372–380.
28 Barnett, G. V., Razinkov, V. I., Kerwin, B. A., Laue, T. M., Woodka, A. H., Butler, P. D., Perevozchikova, T. and Roberts, C. J. (2015) Specific-ion effects on the aggregation mechanisms and protein-protein interactions for anti-streptavidin immunoglobulin gamma-1. *J. Phys. Chem. B* 119, 5793–5804.
29 Austerberry, J. I., Dajani, R., Panova, S., Roberts, D., Golovanov, A. P., Pluen, A., van der Walle, C. F., Uddin, S., Warwicker, J., Derrick, J. P., et al. (2017) The effect of charge mutations on the stability and aggregation of a human single chain Fv fragment. *Eur. J. Pharm. Biopharm.* 115, 18–30.
30 Rosa, M., Roberts, C. J. and Rodrigues, M. A. (2017) Connecting high-temperature and low-temperature protein stability and aggregation. *PLoS One* (Sanchez-Ruiz, J. M., ed.) 12, 1–12.
31 Singla, A., Bansal, R., Joshi, V. and Rathore, A. S. (2016) Aggregation Kinetics for IgG1-Based Monoclonal Antibody Therapeutics. *AAPS J.* 18, 689–702.
32 Barnett, G. V., Drenski, M., Razinkov, V., Reed, W. F. and Roberts, C. J. (2016) Identifying protein aggregation mechanisms and quantifying aggregation rates from combined monomer depletion and continuous scattering. *Anal. Biochem.* 511, 80–91.
33 Weiss, W., Hodgdon, T., Kaler, E., Lenhoff, A. and Roberts, C. (2007) Nonnative protein polymers: structure, morphology, and relation to nucleation and growth. *Biophys. J.* 93, 4392–4403.
34 Wang, W. and Roberts, C. J. (2013) Non-Arrhenius Protein Aggregation. *AAPS J.* 15, 840–851.
35 Austerberry, J. I. and Belton, D. J. (2018) Dynamic Measurements of Bovine Serum Albumin Aggregation Using Small Angle Neutron Scattering.
36 Hevehan, D. L. and De Bernardez Clark, E. (1997) Oxidative renaturation of lysozyme at high concentrations. *Biotechnol. Bioeng.* 54, 221–230.
37 De Bernardez Clark, E., Hevehan, D., Szela, S. and Maachupalli-Reddy, J. (1998) Oxidative renaturation of hen egg-white lysozyme. Folding vs aggregation. *Biotechnol. Prog.* 14, 47–54.
38 Cleland, J. L. and Wang, D. I. C. (1990) Refolding and aggregation of bovine carbonic anhydrase B: quasi-elastic light scattering analysis. *Biochemistry* 29, 11072–11078.
39 Ho, J. G. S. and Middelberg, A. P. J. (2004) Estimating the potential refolding yield of recombinant proteins expressed as inclusion bodies. *Biotechnol. Bioeng.* 87, 584–592.
40 Lawrence, M. S., Phillips, K. J. and Liu, D. R. (2007) Supercharging proteins can impart unusual

resilience. *J. Am. Chem. Soc.* 129, 10110–10112.
41 Miklos, A. E., Kluwe, C., Der, B. S., Pai, S., Sircar, A., Hughes, R. A., Berrondo, M., Xu, J., Codrea, V., Buckley, P. E., et al. (2012) Structure-Based Design of Supercharged, Highly Thermoresistant Antibodies. *Chem. Biol.* 19, 449–455.
42 Dudgeon, K., Rouet, R., Kokmeijer, I., Schofield, P., Stolp, J., Langley, D., Stock, D. and Christ, D. (2012) General strategy for the generation of human antibody variable domains with increased aggregation resistance. *Proc. Natl. Acad. Sci.* 109, 10879–10884.
43 Lee, C. C., Julian, M. C., Tiller, K. E., Meng, F., DuConge, S. E., Akter, R., Raleigh, D. P. and Tessier, P. M. (2016) Design and optimization of anti-amyloid domain antibodies specific for β-amyloid and islet amyloid polypeptide. *J. Biol. Chem.* 291, 2858–2873.
44 Warwicker, J., Charonis, S. and Curtis, R. A. (2013) Lysine and arginine content of proteins: computational analysis suggests a new tool for solubility design. *Mol. Pharm.* 11, 294–303.
45 Edwardraja, S., Neelamegam, R., Ramadoss, V., Venkatesan, S. and Lee, S.-G. (2010) Redesigning of anti-c-Met single chain Fv antibody for the cytoplasmic folding and its structural analysis. *Biotechnol. Bioeng.* 106, 367–375.
46 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook, pp 571–607.
47 Doty, P. and Edsall, J. T. (1951) Light scattering in protein solutions. In Advances in protein chemistry, pp 35–121.
48 Singh, M. and Ram, B. (1998) Effect of Urea, its Concentration and Temperature on Water Structure. *Asian J. Chem.*
49 Kawahara, K. and Tanford, C. (1966) Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J. biol. Chem* 241, 3228–3232.
50 Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. and Bourne, P. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
51 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
52 Sazanavets, I. and Warwicker, J. (2015) Computational Tools for Interpreting Ion Channel pH-Dependence. *PLoS One* 10, e0125293.
53 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723.
54 Bickel, F., Herold, E. M., Signes, A., Romeijn, S., Jiskoot, W. and Kiefer, H. (2016) Reversible NaCl-induced aggregation of a monoclonal antibody at low pH: Characterization of aggregates and factors affecting aggregation. *Eur. J. Pharm. Biopharm.* 107, 310–320.
55 Nicoud, L., Arosio, P., Sozo, M., Yates, A., Norrant, E. and Morbidelli, M. (2014) Kinetic Analysis of the Multistep Aggregation Mechanism of Monoclonal Antibodies. *J. Phys. Chem. B* 118, 10595–10606.
56 Arosio, P., Rima, S. and Morbidelli, M. (2013) Aggregation mechanism of an IgG2 and two IgG1 monoclonal antibodies at low pH: from oligomers to larger aggregates. *Pharm. Res.* 30, 641–654.
57 Minton, A. P. (2016) Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Anal. Biochem.* 501, 4–22.
58 Muschol, M. and Rosenberger, F. (1995) Interactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 103, 10424–10432.
59 Connolly, B. D., Petry, C., Yadav, S., Demeule, B., Ciaccio, N., Moore, J. M. R., Shire, S. J. and Gokarn, Y. R. (2012) Weak interactions govern the viscosity of concentrated antibody solutions: high-throughput analysis using the diffusion interaction parameter. *Biophys. J.* 103, 69–78.
60 Lehermayr, C., Mahler, H.-C., Mäder, K. and Fischer, S. (2011) Assessment of net charge and

protein–protein interactions of different monoclonal antibodies. *J. Pharm. Sci.* 100, 2551–2562.

61 Liu, W., Cellmer, T., Keerl, D., Prausnitz, J. M. and Blanch, H. W. (2005) Interactions of lysozyme in guanidinium chloride solutions from static and dynamic light-scattering measurements. *Biotechnol. Bioeng.* 90, 482–490.

62 Ghosh, R., Calero-Rubio, C., Saluja, A., Roberts, C. J., Bickel, F., Herold, E. M., Signes, A., Romeijn, S., Jiskoot, W., Kiefer, H., et al. (2014) Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *J. Pharm. Sci.* 105, 1086–1096.

63 Dong, X.-Y., Liu, J.-H., Liu, F.-F. and Sun, Y. (2009) Self-interaction of native and denatured lysozyme in the presence of osmolytes, l-arginine and guanidine hydrochloride. *Biochem. Eng. J.* 43, 321–326.

64 Arakawa, T., Tsumoto, K., Nagase, K. and Ejima, D. (2007) The effects of arginine on protein binding and elution in hydrophobic interaction and ion-exchange chromatography. *Protein Expr. Purif.* 54, 110–116.

65 Hirano, A., Arakawa, T. and Shiraki, K. (2008) Arginine increases the solubility of coumarin: comparison with salting-in and salting-out additives. *J. Biochem.* 144, 363–369.

66 Ariki, R., Hirano, A., Arakawa, T. and Shiraki, K. (2011) Arginine increases the solubility of alkyl gallates through interaction with the aromatic ring. *J. Biochem.* 149, 389–394.

67 Li, J., Garg, M., Shah, D. and Rajagopalan, R. (2010) Solubilization of aromatic and hydrophobic moieties by arginine in aqueous solutions. *J. Chem. Phys.* 133, 1–10.

68 Shah, D., Shaikh, A. R., Peng, X. and Rajagopalan, R. (2011) Effects of arginine on heat-induced aggregation of concentrated protein solutions. *Biotechnol. Prog.* 27, 513–520.

69 Ito, L., Shiraki, K., Matsuura, T., Okumura, M., Hasegawa, K., Baba, S., Yamaguchi, H. and Kumasaka, T. (2011) High-resolution X-ray analysis reveals binding of arginine to aromatic residues of lysozyme surface: implication of suppression of protein aggregation by arginine. *Protein Eng. Des. Sel.* 24, 269–274.

70 Crowley, P. B. and Golovin, A. (2005) Cation–π interactions in protein–protein interfaces. *Proteins Struct. Funct. Bioinforma.* 59, 231–239.

71 Sokalingam, S., Madan, B., Raghunathan, G. and Lee, S.-G. (2013) In silico study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability. *Biotechnol. Bioprocess Eng.* 18, 18–26.

72 Sokalingam, S., Raghunathan, G., Soundrarajan, N. and Lee, S.-G. (2012) A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. *PLoS One* (Permyakov, E. A., ed.) 7, 1–13.

73 Buell, A. K., Galvagnion, C., Gaspar, R., Sparr, E., Vendruscolo, M., Knowles, T. P. J., Linse, S. and Dobson, C. M. (2014) Solution conditions determine the relative importance of nucleation and growth processes in α-synuclein aggregation. *Proc. Natl. Acad. Sci.* 111, 7671–7676.

74 Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306.

75 Lopez de la Paz, M. and Serrano, L. (2004) Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci.* 101, 87–92.

76 Lee, C. C., Perchiacca, J. M. and Tessier, P. M. (2013) Toward aggregation-resistant antibodies by design. *Trends Biotechnol.* 31, 612–620.

For Table of Contents Use Only