An Expanded Conformation of an Antibody Fab Region by X-Ray Scattering, Molecular Dynamics and smFRET Identifies an Aggregation Mechanism

Nuria Codina, David Hilton, Cheng Zhang, Nesrine Chakroun, Shahina S. Ahmad, Stephen J. Perkins, Paul A. Dalby

PII: S0022-2836(19)30086-5
DOI: https://doi.org/10.1016/j.jmb.2019.02.009
Reference: YJMBI 66013

To appear in: *Journal of Molecular Biology*

Received date: 22 November 2018
Revised date: 6 February 2019
Accepted date: 6 February 2019

Please cite this article as: N. Codina, D. Hilton, C. Zhang, et al., An Expanded Conformation of an Antibody Fab Region by X-Ray Scattering, Molecular Dynamics and smFRET Identifies an Aggregation Mechanism, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.009

An expanded conformation of an antibody Fab region by X-ray scattering, molecular dynamics and smFRET identifies an aggregation mechanism

Nuria Codina[a], David Hilton[a], Cheng Zhang[a], Nesrine Chakroun[a], Shahina S. Ahmad[a], Stephen J. Perkins[b], and Paul A. Dalby[a]

[a]Department of Biochemical Engineering, University College London, Gordon Street, London, WC1E 7JE, UK

[b]Department of Structural and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

Corresponding Authors:

Prof. Stephen J. Perkins, Department of Structural and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK. Tel: +44-20-7679-7048. Email: s.perkins@ucl.ac.uk.

Prof. Paul A. Dalby, Department of Biochemical Engineering, University College London, Gordon Street, London, WC1E 7JE, UK. Tel: +44-20-7679-9566. Email: p.dalby@ucl.ac.uk.

Abstract

Protein aggregation is the underlying cause of many diseases, and also limits the usefulness of many natural and engineered proteins in biotechnology. Better mechanistic understanding and characterization of aggregation-prone states, is needed to guide protein engineering, formulation, and drug-targeting strategies that prevent aggregation. While several final aggregated states - notably amyloids – have been characterized structurally, very little is known about the native structural conformers that initiate aggregation. We used a novel combination of small-angle X-ray scattering (SAXS), atomistic molecular dynamics (MD) simulations, single-molecule FRET (smFRET), and aggregation-prone region (APR) predictions, to characterize structural changes in a native humanized Fab A33 antibody fragment, that correlated with the experimental aggregation kinetics. SAXS revealed increases in the native state radius of gyration, R_g, of 2.2% to 4.1%, at pH 5.5 and below, concomitant with accelerated aggregation. In a cutting-edge approach, we fitted the SAXS data to full molecular dynamics simulations from the same conditions, and located the conformational changes in the native state to the constant domain of the light chain (C_L). This C_L displacement was independently confirmed using smFRET measurements with two dual-labelled Fabs. These conformational changes were also found to increase the solvent exposure of a predicted aggregation-prone region (APR), suggesting a likely mechanism through which they promote aggregation. Our findings provide a means by which aggregation-prone conformational states can be readily determined experimentally, and thus potentially used to guide protein engineering, or ligand binding strategies, with the aim of stabilizing the protein against aggregation.

Key Words

Antibody fragment; protein aggregation; X-ray scattering; molecular dynamics; single-molecule FRET

Abbreviations

Fab, fragment antigen-binding; SAXS, Small-angle X-ray scattering; MD, molecular dynamics; smFRET, Single molecule fluorescence resonance energy transfer; APR, aggregation-prone regions; R_g, radius of gyration; SASA, solvent accessible surface area.

Introduction

Elucidating how proteins misfold or partially unfold before they aggregate remains a key challenge with significant impact in medicine and biotechnology [1]. In the last 30 years, antibody-based products have become the main drug class for new approvals in the pharmaceutical industry [2], used to treat human diseases, mainly in oncology, autoimmunity and chronic inflammation [3]. However, many natural or newly engineered proteins suffer from aggregation during development, which is highly undesirable as they may elicit adverse immune responses in patients, and so must be avoided [4,5].

Protein stability is only marginal, and stresses to the system (such as changes in pH, salt concentration or temperature), can perturb the native structure of the protein and trigger aggregation [6]. Aggregates are also observed at all stages of drug product development, including expression, purification, shipping and storage [7,8]. Protein engineering and formulation are two potential avenues to minimize protein aggregation. However, in order to improve success, the molecular mechanisms that lead to aggregation need to be understood first, so that informed rational decisions can be made. Thus, understanding the mechanisms by which soluble protein molecules misfold and aggregate is of fundamental and biomedical importance.

The most widely accepted aggregation mechanism involves two steps: (i) a conformational change to the protein native state and (ii) the assembly of protein molecules into aggregates [9]. Initially, the native state experiences a conformational change to form an aggregation-prone state. This intermediate or native-like state is believed to expose aggregation-prone regions, which are normally shielded in the native protein [10]. Then, assembly into aggregates is driven by the hydrophobic effect, or the propensity of exposed sequences to form cross-β sheets [11]. The first step is controlled by the conformational stability of the native protein relative to aggregation-prone states, though this is often probed indirectly

by the free energy of unfolding ΔG_{unf} of the native protein relative to the fully unfolded state. The second step, the assembly into aggregates, is controlled by the persistence time, or relative population of aggregation-prone states, their ability to form specific intermolecular interactions, and their colloidal stability in terms of intermolecular attractive and repulsive forces. Many proteins aggregate with first-order kinetics, implying a unimolecular rate-limiting step linked to conformational changes or partial unfolding, rather than a rate-limiting bimolecular association of two protein molecules [9]. It is therefore important to characterize the nature of any conformational changes in the native state that can promote aggregation.

For many years, experiments have tried to characterize the states that precede aggregation [12]. Initial studies suggested that aggregation takes place from the fully-unfolded state of the protein, drawn from early observations of proteins at elevated temperatures. However, increasing evidence suggests that, at storage temperatures below the melting temperature (T_{m}) of the protein, aggregation takes place from near-native states, where only partial or transient local-unfolding of the protein occurs [13]. This hypothesis is supported by recent work on Fab A33, where it was found that the melting temperatures of the protein under different conditions were only correlated with aggregation kinetics that were determined at temperatures elevated to just below the T_{m} of the protein, where aggregation from the unfolded state therefore predominated [14]. By contrast, T_{m} did not correlate with the aggregation kinetics determined at lower storage temperatures, indicating that global unfolding was no longer the cause of aggregation.

Many studies have reported the presence of near-native states of proteins that are aggregation-prone. For example, a combined analysis of kinetics and solution thermodynamics of recombinant human interferon-γ, found that only a 9% expansion of the native-state surface area was necessary to form the intermediate state that preceded aggregation [15]. Similar results were found for human granulocyte colony stimulating factor, in which the expanded

intermediate state preceding aggregation represented only 15% of the change in surface area observed for the completely unfolded conformation [16]. More recently, only transient local unfolding was found necessary to show faster aggregation for variants of human lysozyme, using hydrogen/deuterium exchange experiments [17]. Studies on hyperthermophilic acylphosphatase, superoxide dismutase 1, transthyretin, β_2-microglobulin and Fyn SH3 also showed that global unfolding was not necessary, and that aggregation could be initiated from locally unfolded states [18,19]. NMR was able to resolve a structural folding intermediate of the 6.4kDa Fyn SH3 domain that was more aggregation-prone than the native state [19]. However, this relied upon mutations that stabilized the folding intermediate, and so the use of NMR to characterize directly pre-aggregational states in unmutated native-ensembles remains very challenging, particularly for larger proteins such as the 48 kDa humanized antibody fragment Fab A33.

As an alternative new approach, we combined small angle X-ray scattering (SAXS) with molecular dynamics (MD) simulations, and verified the results using single-molecule Förster Resonance Energy Transfer (smFRET), to characterize a pH-dependent shift in the native ensemble conformation of the humanized antibody fragment Fab A33 (Fig. 1; Fig. S1). Such antibody fragments bring therapeutic advantages, including deeper tissue penetration [20], and are also simpler and less costly to manufacture in prokaryotes, by removing the F_c domain and the accompanying need for glycosylation [21]. The conformational shift was correlated with the accelerated aggregation kinetics measured under the same native conditions. SAXS showed that Fab A33 adopted an expanded conformation at pH below 7.0, with a 2.2-4.1% increase in the radius of gyration. This expanded conformation was more aggregation-prone, and small amounts of aggregates were also detected by SAXS. To understand the molecular structure of this expanded conformation, the SAXS data were combined with structures generated using MD simulations. We found that the constant domain of the light chain (C_L) became more open and dynamic as the pH decreased. This domain displacement was confirmed using smFRET, in

which donor and acceptor fluorophores were attached at specific locations in Fab A33, and the distance between them was monitored. Our results support increasing evidence that aggregation at temperatures well below the T_m of the protein takes place from near-native states, and does not require global unfolding. The conformational states preceding aggregation are very similar to the native structure at pH 7.0, but they are more expanded due to local regions with increased flexibility. The latter then expose sequence regions that favor aggregation.

Results

Small-angle scattering identified an expanded aggregation-prone conformation

SAXS is a diffraction technique that characterizes the structure of a protein in solution [22]. By studying the protein in different solution conditions, it is possible to elucidate changes in conformation. Here, we investigated the effect of pH and salt concentration upon the Fab A33 structure (Fig. 2). X-ray scattering curves were acquired for Fab A33 at 1 mg/ml in 20 different conditions that combined five pH values (3.5, 4.5, 5,5. 7.0 and 9.0) and four ionic strengths (IS) (20, 50, 150 and 250 mM). These yielded the radius of gyration R_g and the intensity at zero Q I(0), the latter being proportional to the molecular weight. The R_g and I(0) values were obtained using both Guinier analyses of the low Q region, and pair density distribution P(r) analyses from the full scattering curve. Both analyses gave consistent results (Fig. 2; Fig. S2). The Guinier plots revealed trace amounts of aggregates at low Q, especially for the samples below pH 7.0 (red; Fig. 2a). The aggregates were seen when the I(Q) intensities curved upward. When the intensities were fitted to larger Q values, linear non-aggregated Guinier plots with satisfactory $Q.R_g$ ranges were identified that were distinct from the Guinier fits for aggregated Fab A33. This analysis was confirmed from the I(0)/c values from the fits at larger Q (Fig. 2b), where these were seen to be similar at around 40 (39 ± 3) for almost all 20 solution conditions. This outcome indicated that the R_g values of monomeric Fab A33 could be determined independently from its

aggregation. Thus, for all samples, two Q ranges were fitted, namely the larger Q range of 0.37-0.5 nm^{-1} that monitored the structure of the monomer, and the shorter Q range of 0.14-0.3 nm^{-1} to monitor the degree of sample aggregation. As previously used [23], aggregation levels were monitored using $\Delta I(0)/c$ [= $(I(0)_{Q:\ 0.14\text{-}0.3\ nm}^{-1} - I(0)_{Q:\ 0.37\text{-}0.5\ nm}^{-1})/c$], which used the monomeric I(0) values as baseline to monitor the aggregation that had taken place (Fig. 2d).

At acidic pH (5.5, 4.5 and 3.5), SAXS detected the presence of a more expanded or partially unfolded conformation of Fab A33 when compared to neutral pH (7.0 and 9.0) (Fig. 2c). The R_g values increased from 2.62 nm at pH 7.0 and 2.64 nm at pH 9.0 to 2.70 nm at pH 3.5, 2.73 nm at pH 4.5, and 2.71 nm at pH 5.5 (mean ± SEM of 0.01 nm). These correspond to R_g increases of between 2.2% and 4.1% from neutral pH (7.0 and 9.0) to acidic pH (5.5, 4.5 and 3.5). We also observed that pH had a bigger effect on the conformation of Fab A33 than salt, whereby an increase in the IS of the solution from 20 to 250 mM had little effect on R_g (e.g., R_g increased at pH 7.0 from 2.60 nm to 2.64 nm, and at pH 4.5 from 2.72 nm to 2.73 nm (mean ± SEM of 0.01 nm)). These results are consistent with previous reports that showed that pH had a bigger influence on protein conformations [24].

Protein aggregation was monitored using $\Delta I(0)/c$. Small amounts of aggregates in the samples at acidic pH (5.5, 4.5 and 3.5) were detected, in contrast to the samples at neutral pH (7.0 and 9.0) where no aggregates were found (Fig. 2d). Thus, the same conditions that resulted in an expanded conformation of Fab A33 led to its accelerated aggregation, to indicate that the expanded conformation of Fab A33 is aggregation-prone. Interestingly, no aggregates were detected in the pH 3.5 samples at a low IS of 20 mM, unlike for the pH 4.5 and 5.5 samples. This is probably due to colloidal stabilization at pH 3.5 and low IS, where Fab A33 is highly protonated, and so repulsive forces between positively charged proteins prevents their aggregation. When salt concentration is increased (Fig. 2d, pH 3.5 and IS of 50, 150 and 250 mM), the long-range repulsions between charged Fab molecules become shielded, which favors

aggregation. These results are consistent with the Fab A33 aggregation kinetics observed previously [14] for which the pH 3.5 samples at low IS also aggregated much more slowly than at higher pH or IS. Our findings also suggest that the addition of salt mainly contributes to charge shielding, and does not destabilize the native conformation sufficiently to induce global unfolding.

The R_g results were correlated with our previously-reported aggregation rates (v) [14], obtained by monitoring monomer loss by SEC-HPLC in the same 20 experimental conditions at 23 °C. The initial rates of monomer loss as a function of pH and ionic strength were increased to 0.027 ± 0.003 % day^{-1} at pH 3.5-5.5, compared to rates of 0.009 ± 0.0018 % day^{-1} at pH 7-9 (Fig. S3). Thus, the experimental conditions that caused an increase in R_g (pH 3.5, 4.5, and 5.5) also resulted in faster aggregation rates than at neutral pH (7.0 and 9.0) (Fig. 3). These results confirm that the expanded conformation of Fab is more aggregation-prone.

MD simulations captured pH-induced unfolding and identified flexible regions

A homology model of Fab A33 was generated using Rosetta, given that no crystal structure of Fab A33 was available. PropKa software was used with this model to determine the protonation state of the ionizable residues [25]. This gave the following total charges: +35 (pH 3.5), +18 (pH 4.5), +12 (pH 5.5), +9 (pH 7.0) and +5 (pH 9.0). This structure was used as the starting point for MD simulations using Gromacs [26–28]. These were carried out for 50 ns at pH 3.5, 4.5, 5.5, 7.0 and 9.0 for an IS of 50 mM. Simulations were performed at three different temperatures: 300 K, 340 K and 380 K. For each of the 15 conditions, three independent simulations were performed.

The MD simulations at 300 K were used to calculate the R_g value of Fab A33 and its solvent accessible surface area (SASA) as a function of simulation time for each pH (Figs. 4a and 4b). For each pH, the three simulation repeats were averaged at each time point (10 ps) to show their variability, then smoothed by window averaging consecutive data for clarity. The

simulations were able to capture the increase in R_g and SASA as the pH decreased, in agreement with the SAXS results (Fig. 2c). Note that the R_g values from the simulations were smaller than those obtained from SAXS because no account was taken of the hydration shell visible by SAXS, nonetheless the trends were clear. The MD simulations at pH 7.0 and 9.0 gave an R_g of 2.52 nm and 2.51 nm (mean ± SEM of 0.01 nm), respectively, whereas at pH 3.5 this increased to 2.58 nm ± 0.02 nm. At pH 4.5 and 5.5, the R_g increases were not yet as large as those measured experimentally with SAXS. MD simulations do not update the protonation state of molecules continuously as the protein structure unfolds, and this would limit the rate of structural change during simulation, most critically at pH 4.5-5.5, which overlaps the pKa range of acidic residues. In accord with the R_g changes, the SASA increased from 210 nm^2 at pH 7.0 and 9.0, to 220 nm^2 at pH 3.5.

The MD simulations also provided information about protein dynamics potentially down to the level of individual residues. The flexible regions of Fab A33 were assessed using the root mean square fluctuation (RMSF), this being the average distance that a residue moves during the simulation. The RMSF for the last 30 ns of each simulation (20-50 ns) were averaged for each residue, and visualized by color in Fig. 4c and 4d for pH 7.0 and pH 3.5 respectively. Less flexible residues are shown in blue and more flexible residues in red. For pH 7.0, the most flexible regions were found to be the loop regions, followed by the α-helical regions, then the β-strand regions. The highest flexibility was seen in the CDR loops, the C-terminus of the heavy chain, and several loops and α-helices. The β-strands of the C_L and C_H1 domains were more flexible than the V_L and V_H domains. For pH 3.5, Fab A33 was seen to be more flexible than at pH 7.0. In addition to the regions seen to be flexible at pH 7.0, which were also flexible at pH 3.5, both the C_L and C_H1 domains showed increased flexibility at low pH. Together with the R_g and SASA changes, the MD simulations showed that at low pH, Fab A33 adopted both an

expanded conformation and specific regions with increased flexibility. These are characteristics expected of aggregation-prone conformers.

Atomistic modelling of SAXS data to characterize the expanded conformation

The atomistic modelling of full SAXS curve to large Q values enables additional information about the structure of Fab A33 to be obtained beyond the low resolution R_g analyses. Theoretical X-ray scattering curves can be calculated from each atomistic model for comparison to the experimental SAXS curves to identify best fit structures that are representative of the average solution structure of Fab A33. This has been applied to a range of protein structures [29–31].

Each MD simulation above recorded 5,000 structural snapshots of the 50 ns simulations at every 10 ps, i.e. 45,000 structural models for each pH value. A theoretical scattering curve was calculated for each model for comparison against its corresponding experimental SAXS curve at the same pH and NaCl concentration. The R-factor goodness-of-fits were calculated by comparing the theoretical and experimental curves in the Q range of 0.37-1.6 nm^{-1}. Each theoretical R_g value was also calculated from the theoretical scattering curves from Guinier analysis in the same Q range (0.37-0.5 nm^{-1}) as that used above. The sets of R-factor vs. R_g plots for each pH show the extent to which the 45,000 models converge with the experimental data. All five plots showed minima of less than 2% at theoretical R_g values close to the experimental R_g values (vertical lines; Fig. 5a), indicating very good fits had been obtained. The presence of these minima indicated that enough Fab A33 conformations had been sampled, although simulations at higher temperatures were required at pH 4.5 and 5.5 to verify that minima had been reached in these cases. This outcome reflected the observation above that the simulations at pH 5.5 and 300 K did not fully capture the conformational change seen by SAXS. The ten best fit models for each pH are highlighted in yellow, all being within one standard deviation of the experimental R_g values. Notably, the best fit minima were reached at

different R_g values for each pH, implying that the above conformational differences had been followed in the fits. The best-fit modelled SAXS and experimental I(Q) curves were overlaid, and likewise the experimental and best-fit modelled P(r) curves were overlaid (Fig. 5b). Visual inspection showed very good fits. The best-fit structures showed that Fab A33 had a more compact structure at pH 7.0 and 9.0, and was partially unfolded to a more expanded conformation at pH 5.5, 4.5 and 3.5.

Alignments of the sets of ten best-fit structures were performed to gain insights into the expanded conformation of Fab A33 at low pH. The best-fit structures from pH 7.0 were first considered (Fig. 6). These structures did not align perfectly, indicating that Fab A33 is flexible at pH 7.0. The RMSD of each Fab domain was calculated relative to the structure that best fitted the SAXS experimental data at pH 7.0, as reference. As expected from their antigen-binding role, the CDR loops showed high flexibility, with a median RMSD of 0.18 nm and an interquartile range of 0.13-0.21 nm. Interestingly, the C_H1 domain also showed high flexibility with an RMSD of 0.17 nm and range of 0.10-0.19 nm. In particular, the C_H1 C-terminal β-strand connected to the hinge peptide showed wide conformational variability. In the full-length antibody, the hinge is attached to the F_C region, which may provide additional stability. The V_H and C_L domains showed RMSDs and ranges of 0.15 nm (0.14-0.16 nm) and 0.10 nm (0.09-0.12 nm) respectively. The V_L domain showed the least variability with a RMSD and range of 0.08 nm (0.08-0.09 nm) and good alignment of all the β-strands.

Next, the alignment of the ten best-fit structures at pH 7.0 to those at pH 3.5 provided structural information about the pH-induced conformational change of Fab A33 (Fig. 7). The RMSDs at pH 3.5 were also calculated relative to the reference best fit structure at pH 7.0. Notably, the C_L domain was the only domain to show a significant increase in RMSD as the pH was decreased. The RMSD and range of the C_L domain increased to 0.16 nm (0.13-0.17 nm) at pH 3.5, compared to 0.10 nm (0.09-0.12 nm) at pH 7.0. The structure alignments revealed a

displacement of this domain at low pH (magenta; Fig. 7a), being more open to solvent at pH 3.5. This pH-dependent domain displacement was clearly visualized in two loops (light chain residues 150-159 and 198-205; arrowed in Fig. 7a) which connect C_L β-strands. The α-helix at residues 183-188 of the C_L domain was displaced. The β-sheet structure was lost in 8 out of 10 of the C_L best structures at pH 3.5 in residues 144-147, which suggests an increased flexibility in this segment. Additional views of the best-fit structures at pH 7.0, 5.5 and 3.5 in Fig. S4 provide further visual support for the conformational shift at low pH in the C_L domain.

No corresponding systematic displacements were found at pH 3.5 for the other three domains, even though these showed comparable RMSDs and ranges of 0.10 nm (0.09-0.10 nm) for the V_L domain, 0.17 nm (0.16-0.17 nm) for the V_H domain and 0.18 nm (0.16-0.20 nm) for the C_H1 domain, each relative to the best fit Fab A33 structure at pH 7.0. As seen at pH 7.0, the C_H1 domain was relatively flexible in sampling a wide range of conformations, particularly in the C-terminal β-strand connected to the C-terminal hinge (Fig. 7b). The hinge itself was highly extended at pH 3.5, and adopted a range of conformations. The five sets of ten best-fit MD structures at each pH value are downloadable in Supplementary Data Online.

In order to obtain a better picture of the displacements experienced by the whole Fab A33 molecule at low pH, distances between the four domains were also measured, using one cysteine in each domain. Six distances were monitored in total (V_L-V_H, C_L-C_H1, V_L- C_L, V_H- C_H1, V_L-C_H1 and V_H- C_L), using the four cysteines located in outer ß-strands (C23, C194, C236, C414), (Fig. S5). The six distances were calculated for the ten best SAXS fit structures at pH 7.0 and the ten best SAXS fit structures at pH 3.5, and their averages and SEM are reported (Table 1). Results confirmed that the displacement at low pH occurred in the C_L domain, given that the only inter-domain distances that increased between pH 7.0 and 3.5, were the distances where the C_L domain was involved. Distances increased between the outer ß-strand cysteines

of V_L-C_L (0.29 ± 0.07 nm), V_H-C_L (0.27 ± 0.03 nm) and C_L-C_H1 (0.05 ± 0.02 nm). By contrast, the distances between the other domains did not change significantly.

Table 1. Inter-domain distance differences between the best SAXS fit structures at pH 7.0 and 3.5, using one cysteine in each domain (V_L, V_H, C_L and C_H1). Six distances were monitored between the four Fab domains (V_L-V_H, C_L-C_H1, V_L- C_L, V_H- C_H1, V_L-C_H1 and V_H- C_L) using the four cysteines (C23, C194, C236, C414) located in the outer ß-strands. These are shown in the Fab A33 structure in Fig. S5.

	pH 7.0		pH 3.5		ΔpH (3.5 - 7.0)	
	Dist (nm)	SEM	Dist (nm)	SEM	ΔDist (nm)	SEM
C23 (V_L) - C236 (V_H)	3.00	0.02	2.95	0.01	-0.05	0.03
C194 (C_L) - C414 (C_H1)	2.47	0.02	2.52	0.01	0.05	0.02
C23 (V_L) - C194 (C_L)	4.18	0.02	4.47	0.06	0.29	0.07
C236 (V_H) - C414 (C_H1)	4.11	0.07	4.09	0.06	-0.02	0.09
C23 (V_L) - C414 (C_H1)	4.68	0.06	4.64	0.02	-0.05	0.06
C236 (V_H) - C194 (C_L)	4.99	0.02	5.26	0.02	0.27	0.03

smFRET to confirm the C_L domain displacement at low pH

To confirm the MD and SAXS modelling of displacements within the C_L domain at low pH, we used smFRET. FRET is the radiation-less transfer of energy from a donor fluorophore to an acceptor in a range of 2-10 nm distances [32–35]. smFRET is thus sensitive to such distance changes, and can study conformational changes in proteins [36–39]. Here, the apparent FRET transfer efficiency (E_{app}) [32,40,41] that measures the fraction of photons absorbed by the donor

that have been transferred to the acceptor was used to report the separation between the donor and acceptor.

Two donor-acceptor constructs were generated to probe an intra-C_L separation and a separation between the C_L domain and the heavy-chain linker (Fig. 8a). Specifically, these were (Dist 1) LC-K126pAzF + LC-S156C, and (Dist 2) HC-S117pAzF + LC-S156C. Each construct contained one unnatural amino acid [42], p-azido-l-phenylalanine (pAzF), and one solvent-exposed cysteine [43,44] to attach the fluorophores Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor). The labelling was confirmed using ESI mass spectrometry and UV-vis absorption spectra (Figs. S6 and S7). The confocal detection of freely diffusing molecules was used to obtain the apparent transfer efficiency histograms (E_{app}) of Fab A33 at pH 7.0 and 3.5 in an ionic strength of 50 mM (Fig. 8a). The peaks at a FRET efficiency of zero correspond to molecules with no active acceptor fluorophore, and these peak backgrounds are in grey. The peaks at high FRET efficiencies correspond to Fab A33 molecules with one donor and one acceptor. To determine their mean transfer efficiencies, these were fitted to Gaussian peak functions (black lines). Raw data were shown in the form of distributions of inter-photon delays (Fig. S8), together with controls of Fab A33 unfolding using guanidinium chloride (GdmCl) (Figs. S9 and S10).

smFRET showed that the intra-C_L distance (Dist 1) did not change with pH, as the same FRET efficiency value (E_{app} = 0.97) was found at pH 7.0 and pH 3.5 (Fig. 8a). In contrast, the distance between C_L and the heavy chain linker (Dist 2) increased at pH 3.5, with a decrease in FRET efficiency from E_{app} = 0.87 at pH 7.0 to E_{app} = 0.78 at pH 3.5 (Fig. 8a). These results indicated the displacement of the C_L domain and the partial unfolding of Fab A33 at low pH. On comparing this outcome to the atomistic modelling of the SAXS data, the ten best fit structures at pH 7.0 and pH 3.5 were examined (Fig. 8b). Dist 1 was unchanged with pH, being 2.5 ± 0.1 nm at pH 7.0 and pH 3.5. However, Dist 2 increased from 2.8 ± 0.4 nm at pH 7.0 to 3.5 ± 0.1 nm at pH 3.5, corresponding to an increased separation between C_L and the heavy chain linker.

Dist 1 and Dist 2 were also monitored during the MD simulations at pH 7.0 and 3.5 (Fig. 8c). Dist 1 was unchanged during the simulation, while Dist 2 increased from 2.9 ± 0.3 nm at pH 7.0 to 3.3 ± 0.2 nm at pH 3.5. Both the atomistic SAXS modelling and the MD simulations confirmed the experimentally observed displacement at low pH of the C_L domain by smFRET.

Identification of Aggregation-Prone Regions (APR) suggests aggregation mechanism

Computational biology tools were used to predict the regions in proteins most likely to form and stabilize the cross-β structure characteristic of aggregates. These aggregation-prone regions (APRs) are mostly hydrophobic, possess a low net charge, and have a high propensity to form β-sheets. Several methods have been developed to predict the presence of APRs in a protein. The first methods only used the protein sequence as input, this being equivalent to the fully unfolded state. Predictions were based on either the intrinsic properties of amino acids, or their compatibility with protein structural features in known amyloid fibril structures [10]. Examples include TANGO [45], AGGRESCAN [46], PASTA [47], MetAmyl [48], FoldAmyloid [49], FishAmyloid [50] and Waltz [51]. As these predictions do not always agree, Amylpred2 generates a consensus from up to eleven existing algorithms [52]. However, it is known that APRs are frequently buried inside the hydrophobic core of globular proteins, and so their ability to trigger aggregation would depend upon solvent accessibility, i.e. the potential of the APR to become solvent exposed through structural dynamics or partial unfolding. Thus, more recent methods include aspects of the protein structure to predict APRs, including AGGRESCAN 3D [53], AggScore [54], SAP [55] and Solubis [56].

Here, we have already identified the solution conformations of Fab A33 in different solution conditions via SAXS atomistic modelling. Thus, we determined APRs using only sequence-based predictors, and combined them with the best experimentally identified structures at pH 7.0 and 3.5 to identify differences in their solvent exposure. Four sequence-

based APR predictors were used in total, PASTA 2.0, TANGO, AGGRESCAN and MetAmyl, to predict the APRs in Fab A33. APRs, where three out of the four predictors identified an aggregation-prone sequence and were selected (Fig. S11a). Seven segments showed the highest aggregation propensity values, namely residues 31-36, 47-51, 114-118 and 129-139 in the light chain and residues 261-165, 325-329 and 387-402 in the heavy chain. Additionally, these APRs were confirmed using Amylpred2, which identified the same APRs in addition to others (Fig. S11a).

To display the aggregation-prone regions on the Fab A33 homology model as shown in Fig. 9a, each aggregation propensity was normalized between 0 and 1, and weighted equally (Fig. S11b). Red represented high aggregation propensities and blue low aggregation propensities. The seven APRs were co-located as three regions of largely buried β-strands within the folded structure, and all were protected from the solvent. Next, the difference in solvent accessibility of the APRs in the SAXS best-fit structures at pH 7.0 and pH 3.5 were analysed. The solvent accessibility of one APR visibly increased at pH 3.5 due to the displacement of the C_L domain (circled; Figs. 9b and 9c). Quantitatively, the SASA of the seven APRs were calculated for the ten best-fit structures at pH 7.0 and pH 3.5, and summed (Table S1). While most APR showed small decreases in solvent accessibility, the APR at residues 387-402 increased by 83 $Å^2$ from 536 ± 43 $Å^2$ at pH 7.0 to 619 ± 39 $Å^2$ at pH 3.5 (3% increase), due to the displacement of the C_L domain at low pH. These data illustrate the potential of combining biophysical methods that determine conformational changes, with sequence-based APR prediction tools, for determining aggregation hotspots. For Fab A33, the aggregation prediction tools suggested a possible molecular explanation for the observed increase in aggregation at low pH as the result of structural instabilities.

Discussion

In order to prevent protein aggregation, the aggregation-prone conformations that lead to aggregation need to be elucidated first, so that rational strategies can be taken to prevent it. However, such aggregation-prone conformations for near-native solution conditions have proven most challenging to characterize over the years, and have remained elusive within unmutated native-protein ensembles. In this study, we combined orthogonal methods (SAXS, MD simulations, SAXS atomistic modelling, smFRET and APR predictions) to characterize the structural perturbations that take place within the native ensemble of the humanized antibody Fab A33 over a range of different pH and ionic strengths. Formation of an expanded conformation of Fab A33 correlated with its increased aggregation propensity under native conditions. The conformational change was localized on the C_L domain of the Fab, revealing that aggregation at near-native conditions proceeds through a state that is only slightly perturbed in structure relative to the native state. In the case of Fab A33, the radius of gyration increased between 2.2% and 4.1% for the aggregation-competent species relative to the native state.

To explain the increased aggregation propensity of the expanded conformations of Fab A33, we used online tools to predict the aggregation-prone regions (APR) that are more likely to form cross-β structures found in aggregates. All the APRs predicted for Fab A33 were buried in the protein interior and shielded from solvent. However, the displacement of the C_L domain at low pH increased the solvent accessibility of one APR, presenting a likely route to aggregation. Future work to confirm this proposed aggregation mechanism, could include mutagenesis of the ionizable residues inferred to drive the pH-induced change, or to reduce the aggregation propensity of the exposed APR. Our findings are in accordance with an aggregation mechanism that proceeds through partially unfolded and transiently-populated conformations within the native ensemble [18,57]. The initial oligomers formed would thus retain high structure similarity to the native state. We hypothesize that in later stages of the aggregation process, a structural

re-arrangement takes place to form the typical cross-β structure of amyloids, as indicated in previous studies [16,58,59].

Collectively, this work provides compelling evidence of how local unfolding can lead to transiently-formed structural conformers within the native ensemble that promote aggregation. It also highlights the promise of SAXS combined with molecular dynamics simulations to resolve aggregation-prone conformers within native ensembles, particularly for large proteins that are less accessible by NMR. This also provides a new route to gaining molecular level knowledge of potential target sites for the rational engineering of more stable proteins, either via protein engineering or formulation, or for the design of drugs that bind to and stabilize proteins against aggregation *in vivo*.

Materials and Methods

Cloning, site-directed mutagenesis, expression and purification of Fab A33

Wild-Type C226S Fab A33

The gene coding for Fab A33 (Fig. S1) was kindly provided by UCB (Slough, UK) in the plasmid pTTOD in *E. coli* W3110, and the C226S heavy-chain variant was used to avoid the formation of linked Fab dimers. This variant was termed the wild-type Fab A33, which was expressed and purified as described previously [14].

Fab A33 mutants for smFRET

Fab A33 mutants with one non-natural amino acid, p-azido-l-phenylalanine (pAzF), and one engineered solvent-exposed cysteine were generated to allow attachment of donor and

acceptor fluorophores. Two different constructs were generated: (i) LC-K126pAzF + LC-S156C and (ii) HC-S117pAzF + LC-S156C. To incorporate pAzF, the plasmid encoding for Fab A33 was co-transformed into *E. coli*, with plasmid pEVOL-pAzF (Plasmid ID: 31186) (Addgene, Cambridge, USA), which encodes an engineered tyrosyl-tRNA synthetase/amber suppressor tRNA derived from *Methanococcus jannaschii*, to incorporate pAzF at the amber stop codon [60,61]. As pTTOD and pEVOL-pAzF had the same origin of replication p15A, the tac promoter and Fab A33 gene from pTTOD was sub-cloned into pET-29a(+) which has a ColE1 origin of replication using circular polymerase extension cloning (CPEC) [62]. CPEC primers (Eurofins, Wolverhampton, UK), were:

(Insert.REV) GGCTTTGTTAGCAGCGATATGACGACAGGAAGAGTTTGTAGAAACG

(Vector.REV) TTCCTGTCGTCATATCGCTGCTAACAAAGCCCGAAAGG

(Insert.FOR) TGATGTCGGCGATACCATCGGAAGCTGTGGTATGG

(Vector.FOR) CAGCTTCCGATGGTATCGCCGACATCACCGATGGG.

The insert and vector were amplified by PCR using the CPEC primers, purified using QiaQuick gel purification kit (Qiagen, Hilden, Germany), assembled using a 30:1 (insert:vector) ratio with 100 ng of vector and 10 cycles, then directly transformed into NEB *10β* competent cells (New England Biolabs, Ipswich, US). Final assembly was confirmed by sequencing (pET29Fab_for: AGGAATGGTGCATGCAAGG, pET29Fab_mid: AGTGGAAGGTGGATAACGC, T7 term: CTAGTTATTGCTCAGCGG), using Source Bioscience (UK).

Site-specific mutations were introduced using QuickChange Lightning Site-Directed Mutagenesis (Agilent Technologies, Santa Clara, USA) to form the double mutants: (i) LC-K126pAzF + LC-S156C and (ii) HC-S117pAzF + LC-S156C. In order to incorporate pAzF, we mutated the native codon to the amber stop codon (TAG). The pEVOL-pAzF and variant pET-29a plasmids were co-transformed into the engineered "amberless" *E. coli* (C321.ΔA.exp) (ID: 49018) (Addgene, Cambridge, USA) [63], then grown in a 180 ml DASbox Mini Bioreactor.

Expression was equivalent to wild-type Fab A33, with the addition of 1 mM pAzF (Chem-Impex International, Wood Dale, US), 34 µg/ml chloramphenicol, 50 µg/ml kanamycin, 0.2% L-(+)-arabinose (as an inducer for pEVOL-pAzF), 0.2 mM IPTG (as an inducer for variant pET-29a containing Fab A33) and 2.5 µg/ml D-biotin (because the C321. ΔA.exp strain is auxotrophic for D-biotin) (all final concentrations).

Site-Specific Labeling of Fab A33

Fab A33 was buffer-exchanged into PBS using 10 kDa cut-off centrifugal filters (Merck, Kenilworth, UK) and adjusted to 0.5 mg/ml. The donor fluorophore dibenzocyclooctyne Alexa Fluor 488 (Thermo Fisher Scientific, Waltham, USA) was reacted using click chemistry at a 5:1 molar ratio (fluorophore:protein) for 24 h at room temperature with gentle shaking in the dark. To attach the acceptor to the Fab solvent-exposed cysteine using maleimide-thiol chemistry, TCEP was added to 0.5 mM (50-fold molar excess of TCEP to Fab), incubated for 1.5 h at room temperature [64], then removed by buffer exchange into PBS. Incubation for 24 h allowed reconstitution of the correct disulfide-bridges. Maleimide-activated Alexa Fluor 594 was added in a 5:1 molar ratio of fluorophore:protein, and incubated for 16-18 h at room temperature. 10 kDa centrifugal filters were used to remove the unreacted dye. The correct labelling of constructs i and ii was confirmed using ESI mass spectrometry and UV-vis absorption (Fig. S7).

X-ray scattering data for Fab A33

Small-angle X-ray scattering measurements were carried out on beamline BM29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Scattering data *I(Q)* were collected using a 2D Pilatus detector located 2.867 m from the sample, yielding a *Q*-range of 0.025-5 nm^{-1}, where $Q = 4\pi \sin\theta / \lambda$; 2θ is the scattering angle and λ is the wavelength (0.09919 nm). Protein solutions were loaded using a quartz flow-through capillary (diameter 1.833 mm,

wall thickness 0.02 mm) and measurements were performed at 20 °C. Data were collected in 10 successive 1 second frames, to minimize the effects of radiation damage, with pre- and post-sample buffer measurements for subsequent background subtraction. The 2D data were normalized to an absolute scale calibrated using the scattering from water and azimuthally averaged to obtain 1D intensity profiles. Profiles with observable radiation damage were discarded prior to averaging and buffer subtraction. Scattering data was collected for Fab A33 at 1.0 mg/ml and 20 °C, at all combinations of pH 3.5, 4.5, 5.5, 7.0 and 9.0 and ionic strengths 20, 50, 150 and 250 mM. Buffers at pH 3.5, 4.5 and 5.5 were 20 mM sodium acetate, at pH 7.0 in 20 mM sodium phosphate, and at pH 9.0 in 20 mM Tris.HCl buffer. Ionic strength was set via the addition of NaCl.

Two types of SAXS analyses were performed to determine the radius of gyration R_g of the protein and its molecular weight from the forward scattered intensity I(0). Guinier analyses at low Q values up to a $Q.R_g$ of 1.5 were based on linear fits using ln I(Q) vs Q^2 plots:

$$\ln I(Q) = \ln I(0) - R_g^2 Q^2/3$$

Fourier transformation of the full scattering curve I(Q) in reciprocal space into real space gave the distance distribution function P(r) that represented all the distances between the atoms in the protein. The maximum in P(r) corresponded to the most commonly occurring distance and the cutoff at large r represented the protein length. Guinier fits and P(r) transformations utilized the programs SCT and GNOM [29,65].

Fab A33 homology model using Rosetta

We generated a homology model of Fab A33 using the Rosetta method “minirosetta” and the crystal structure of human germline antibody 5-51/O12 (PDB ID 4KMT) as the template [66,67]. After residue replacement, 6,811 out of 20,000 structure models retained the five

disulfide bonds intact. From these, 1000 structures with the lowest Rosetta Energy Units were selected, and clustered based on their similarities. The largest category in the clustering step contained 573 structures, and the structure with the lowest score in this category was selected as the homology model of Fab A33 [68].

Molecular dynamic simulations

Molecular dynamic (MD) simulations on the Fab A33 homology model were carried out using Gromacs v5.0 [26]. MD simulations of 50 ns were carried at pH 3.5, 4.5, 5.5, 7.0 and 9.0, all at an ionic strength of 50 mM, using the OPLS-AA/L all-atom force field [69]. Simulations were generated at 300 K, 340 K and 380 K to increase the range of energy to the system and generate more variable structures. Three independent simulations were carried out for all conditions. The topology file retained the five (four intra and one inter) disulfide bonds of Fab. The protonation state of each residue was entered manually, and these were determined at each pH using the PDB2PQR server, which performed the pKa calculations by PropKa [25]. The Fab A33 structure was placed in a cubic box with a layer of water up to at least 10.0 Å from the protein surface. The box was solvated with SPC/E water molecules, Cl^- added to neutralize the net charges, and NaCl added to 50 mM. The system was energy minimized (relaxed), using the steepest descent algorithm (2000 steps) followed by the conjugate gradient method (5000 steps), then the solvent and ions around the protein were equilibrated for 100 ps under NVT ensemble to stabilize at 300 K, and then at 100 ps under NPT ensemble to stabilize at atmospheric pressure (position-restricted simulations). MD simulations were carried out on the UCL Legion High Performance Computing Facility. The time step of the simulations was set to 2 fs, trajectories were saved every 10 ps, and analyses were performed using standard Gromacs tools.

SCT software for SAXS curve calculations

The theoretical X-ray scattering curve was calculated using SCT software for each of the MD models (45,000 models per pH) for comparison with the experimental SAXS data [29]. Coarse-grained models were created from the atomistic structures by placing the latter in a cubic grid of boxes and replacing boxes with spheres if sufficient atoms were present. We used a standard box side of 0.54 nm and cutoff of 4 atoms (confirmed using one of the structures at the end of a pH 3.5 simulation), and adding a hydration shell of 0.3 g water per gram protein because SAXS visualizes the layer of water in contact with the protein. The theoretical scattering curves were calculated using Debye's Law adapted to spheres. The experimental and theoretical curves were compared using the R-factor, with low R-factors representing the better fits, which were computed in the Q range: 0.37-1.6 nm^{-1}.

$$R = \frac{\sum \left\| \|I_{Expt}(Q)\| - \eta \|I_{Theor}(Q)\| \right\|}{\sum \|I_{Expt}(Q)\|} x\ 100$$

Single-Molecule FRET

Confocal fluorescence spectroscopy

Single-molecule fluorescence measurements were carried out on a MicroTime 200 confocal microscope (PicoQuant, Germany). For excitation, a diode laser at the donor excitation wavelength was used (LDH-D-C-485, PicoQuant, Germany), at 20 MHz (laser pulse every 50 ns) and a laser power of 100 µW at the back aperture of the objective. The laser was focused into the sample solution with an UPlanApo 60x/1.20W objective (Olympus). Measurements were performed by placing the confocal volume 50 µm into the solution relative to the cover slide surface. The fluorescence signal was collected by the same objective and filtered with a 485/595 dual-band dichroic mirror (Chroma Technology). Afterwards, the photons passed

through a 100 µm pinhole. Donor and acceptor photons were separated by a second dichroic mirror, 585 DCXR, and further filtered by band-pass filters, ET525/50M for donor, and ET645/75M for acceptor (all Chroma Technology). Finally, photons were detected using two single-photon avalanche photodiodes (τ-SPAD) (PicoQuant). The arrival time of every detected photon was recorded with a HydraHarp 400 counting module (PicoQuant).

Single-molecule measurements were acquired at a protein concentration of <100 pM. The measurements were performed in 20 mM sodium phosphate buffer pH 7.0 and 20 mM sodium citrate buffer pH 3.5, both 50 mM final ionic strength adjusted with NaCl. Despite the low pH, the fluorescence quantum yields of the dyes remained the same [70]. Each sample was measured for 30 min at room temperature.

Single-Molecule Data Analysis

First, the raw data was converted to the Photon-HDF5 file format (.h5) using the open source software Photon-HDF5 [71]. Next, single-molecule FRET data was analyzed using the open source software FRETBursts [72]. We followed the steps for background estimation, burst search, burst selection and computation of FRET efficiency histograms. Background rates were calculated first, by plotting a histogram of inter-photon delay times in windows of 30 s. Signal from single-molecules can be differentiated from background because single molecules show short delay times whereas background signal follows a Poisson process that is exponentially distributed. By fitting the long delay times to an exponential, the background rates were calculated. After background calculation, bursts corresponding to single-molecules traversing the excitation volume were identified. We identified a burst if the rate of photons was 6 times faster than the local background rate, and we used 10 consecutive photons to compute the local count rate. For this calculation, all photons were taken into account (donor and acceptor). After burst identification, corrections were applied. Bursts were corrected for background and donor leakage into the acceptor channel, the later was calculated to be 8%. No acceptor direct

excitation and γ-factor correction were applied, thus the conversions of FRET efficiencies to distances were not possible. In this study, we refer to the calculated FRET efficiencies as "apparent FRET efficiencies" (E_{app}) [32,40,41], which allowed the relative comparison between Fab A33 constructs and solution conditions. We applied a size filter to the previous bursts found, where only bursts with more than 30 photons were kept. Lastly, apparent transfer efficiency histograms were calculated for each burst using the expression $E_{app} = n_A/(n_A+n_D)$; where n_D and n_A are the numbers of donor and acceptor photons in the burst, respectively, and apparent FRET efficiencies were fitted to Gaussian functions.

Aggregation Prediction Regions Software

Sequence-based aggregation prone regions (APR) of Fab A33 were predicted using PASTA 2.0 [47], TANGO [45] AGGRESCAN [46] and MetAmyl [48]. The regions in which three out of the four software identified an aggregation-prone region were selected, resulting in seven APRs. Amylpred2 consensus tool was used to confirm the presence of these APRs. Additionally, a consensus was created between the four sequence-based software, (Normalized TANGO * 1/4 + Normalized PASTA 2.0 * 1/4 + Normalized AGGRESCAN * 1/4 + Normalized MetAmyl * 1/4).

ACKNOWLEDGMENTS

We thank the Engineering and Physical Sciences Research Council (EPSRC), the Centre for Doctoral Training in Emergent Macromolecular Therapies (EP/L015218/1) (N.C.C.), the EPSRC Future Targeted Healthcare Manufacturing Hub (EP/P006485/1, EP/I033270/1) (N.C. and D.H.),

EPSRC EP/N025105/1 (C.Z.), the Biotechnology and Biological Sciences Research Council (BBSRC) Bioprocess Research Industry Club (BB/I017119/1) and BBSRC BB/H015795/1 (S.S.A.), and the CCP-SAS project through a joint EPSRC (EP/K039121/1) and National Science Foundation (CHE-1265821) grant (S.J.P.)

REFERENCES

[1] F. Chiti, C.M. Dobson, Protein Misfolding, Functional Amyloid, and Human Disease, Annu. Rev. Biochem. 75 (2006) 333–366. doi:10.1146/annurev.biochem.75.101304.123901.

[2] D.M. Ecker, S.D. Jones, H.L. Levine, The therapeutic monoclonal antibody market., MAbs. 7 (2015) 9–14. doi:10.4161/19420862.2015.989042.

[3] P.J. Carter, G.A. Lazar, Next generation antibody drugs: Pursuit of the “high-hanging fruit,” Nat. Rev. Drug Discov. 17 (2018) 197–223. doi:10.1038/nrd.2017.227.

[4] M.C. Manning, D.K. Chou, B.M. Murphy, R.W. Payne, D.S. Katayama, Stability of protein pharmaceuticals: An update, Pharm. Res. 27 (2010) 544–575. doi:10.1007/s11095-009-0045-6.

[5] W. Wang, Protein aggregation and its inhibition in biopharmaceutics, Int. J. Pharm. 289 (2005) 1–30. doi:10.1016/j.ijpharm.2004.11.014.

[6] F. Chiti, C.M. Dobson, Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade, Annu. Rev. Biochem. 86 (2017) 1–42. doi:10.1146/annurev-biochem-061516-045115.

[7] W. Wang, S. Nema, D. Teagarden, Protein aggregation-Pathways and influencing

factors, Int. J. Pharm. 390 (2010) 89–99. doi:10.1016/j.ijpharm.2010.02.025.

[8] S. Frokjaer, D.E. Otzen, Protein drug stability: a formulation challenge., Nat. Rev. Drug Discov. 4 (2005) 298–306. doi:10.1038/nrd1695.

[9] E.Y. Chi, S. Krishnan, T.W. Randolph, J.F. Carpenter, Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation, Pharm. Res. 20 (2003) 1325–1336. doi:10.1023/A:1025771421906.

[10] G. De Baets, J. Schymkowitz, Predicting aggregation-prone sequences in proteins, Essays Biochem. 56 (2014) 41–52. doi:10.1042/BSE0560041.

[11] C.J. Roberts, Therapeutic Protein Aggregation: Mechanisms, Design, and Control, Trends Biotechnol. 32 (2014) 372–380. doi:10.1016/j.tibtech.2014.05.005.

[12] T.R. Jahn, S.E. Radford, Folding versus aggregation: Polypeptide conformations on competing pathways, Arch. Biochem. Biophys. 469 (2008) 100–117. doi:10.1016/j.abb.2007.05.015.

[13] M.J. Robinson, P. Matejtschuk, A.F. Bristow, P.A. Dalby, Tm-Values and Unfolded Fraction Can Predict Aggregation Rates for Granulocyte Colony Stimulating Factor Variant Formulations but Not under Predominantly Native Conditions, Mol. Pharm. 15 (2018) 256–267. doi:10.1021/acs.molpharmaceut.7b00876.

[14] N. Chakroun, D. Hilton, S.S. Ahmad, G.W. Platt, P.A. Dalby, Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment, Mol. Pharm. 13 (2016) 307−319. doi:10.1021/acs.molpharmaceut.5b00387.

[15] B.S. Kendrick, J.F. Carpenter, J.L. Cleland, T.W. Randolph, A transient expansion of the native state precedes aggregation of recombinant human interferon-gamma., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 14142–6. doi:10.1073/pnas.95.24.14142.

[16] S. Krishnan, E.Y. Chi, J.N. Webb, B.S. Chang, D. Shan, M. Goldenberg, M.C. Manning, T.W. Randolph, J.F. Carpenter, Aggregation of granulocyte colony stimulating factor under physiological conditions: Characterization and thermodynamic inhibition,

Biochemistry. 41 (2002) 6422–6431. doi:10.1021/bi012006m.

[17] D. Canet, A.M. Last, P. Tito, M. Sunde, A. Spencer, D.B. Archer, C. Redfield, C. V. Robinson, C.M. Dobson, Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme, Nat. Struct. Biol. 9 (2002) 308–315. doi:10.1038/nsb768.

[18] F. Chiti, C.M. Dobson, Amyloid formation by globular proteins under native conditions, Nat. Chem. Biol. 5 (2009) 15–22. doi:10.1038/nchembio.131.

[19] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundström, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, L.E. Kay, Structure of an intermediate state in protein folding and aggregation, Science. 336 (2012) 362–366. doi:10.1126/science.1214203.

[20] A.L. Nelson, Antibody fragments: Hope and hype, MAbs. 2 (2010) 77–83. doi:10.4161/mabs.2.1.10786.

[21] C. Enever, T. Batuwangala, C. Plummer, A. Sepp, Next generation immunotherapeutics-honing the magic bullet, Curr. Opin. Biotechnol. 20 (2009) 405–411. doi:10.1016/j.copbio.2009.07.002.

[22] S.J. Perkins, A.I. Okemefuna, R. Nan, K. Li, A. Bonner, Constrained solution scattering modelling of human antibodies and complement proteins reveals novel biological insights, J. R. Soc. Interface. 6 (2009) S679–S696. doi:10.1098/rsif.2009.0164.focus.

[23] R. Nan, S. Tetchner, E. Rodriguez, P.J. Pao, J. Gor, I. Lengyel, S.J. Perkins, Zinc-induced self-association of complement C3b and factor H: Implications for inflammation and age-related macular degeneration, J. Biol. Chem. 288 (2013) 19197–19210. doi:10.1074/jbc.M113.476143.

[24] E. Sahin, A.O. Grillo, M.D. Perkins, C.J. Roberts, Comparative effects of pH and ionic strength on protein-protein interactions, unfolding, and aggregation for IgG1 antibodies, J. Pharm. Sci. 99 (2010) 4830–4848. doi:10.1002/jps.22198.

[25] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pK a values, Proteins Struct. Funct. Bioinforma. 61 (2005) 704–721.

doi:10.1002/prot.20660.

[26] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindah, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.

[27] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. Van Der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics. 29 (2013) 845–854. doi:10.1093/bioinformatics/btt055.

[28] R.D. Toofanny, V. Daggett, Understanding protein unfolding from molecular simulations, Wiley Interdiscip. Rev. Comput. Mol. Sci. 2 (2012) 405–423. doi:10.1002/wcms.1088.

[29] D.W. Wright, S.J. Perkins, SCT: A suite of programs for comparing atomistic models with small-angle scattering data, J. Appl. Crystallogr. 48 (2015) 953–961. doi:10.1107/S1600576715007062.

[30] K.T. Walker, R. Nan, D.W. Wright, J. Gor, A.C. Bishop, G.I. Makhatadze, B. Brodsky, S.J. Perkins, Non-linearity of the collagen triple helix in solution and implications for collagen function, Biochem. J. 474 (2017) 2203–2217. doi:10.1042/BCJ20170217.

[31] L.E. Rayner, G.K. Hui, J. Gor, R.K. Heenan, P.A. Dalby, S.J. Perkins, The solution structures of two human IgG1 antibodies show conformational stability and accommodate their C1q and FcγR ligands, J. Biol. Chem. 290 (2015) 8420–8438. doi:10.1074/jbc.M114.631002.

[32] R. Roy, Sungchul, T. Ha, A Practical Guide To single-molecule FRET, Nat. Methods. 5 (2008) 507–516. doi:10.1038/NMETH.1208.

[33] B. Schuler, Application of Single Molecule Förster Resonance Energy Transfer to Protein Folding, Protein Fold. Protoc. 350 (2006) 115–138. doi:10.1038/NMETH.1208

[34] E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S.Y. Chung, X. Michalet, S. Weiss, Toward dynamic structural biology: Two decades of single-molecule förster resonance

energy transfer, Science. 359 (2018). doi:10.1126/science.aan1133.

[35] B. Schuler, W.A. Eaton, Protein folding studied by single-molecule FRET, Curr. Opin. Struct. Biol. 18 (2008) 16–26. doi:10.1016/j.sbi.2007.12.003.

[36] H. Hofmann, F. Hillger, S.H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. a Lipman, B. Schuler, Single-molecule spectroscopy of protein folding in a chaperonin cage., Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 11793–11798. doi:10.1073/pnas.1002356107.

[37] K. a Merchant, R.B. Best, J.M. Louis, I. V Gopich, W. a Eaton, Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations., Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1528–1533. doi:10.1073/pnas.0607097104.

[38] M.B. Borgia, A. Borgia, R.B. Best, A. Steward, D. Nettels, B. Wunderlich, B. Schuler, J. Clarke, Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins, Nature. 474 (2011) 662–5. doi:10.1038/nature10099.

[39] F. Hillger, D. Nettels, S. Dorsch, B. Schuler, Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy, J. Fluoresc. 17 (2007) 759–765. doi:10.1007/s10895-007-0187-z.

[40] D.S. Majumdar, I. Smirnova, V. Kasho, E. Nir, X. Kong, S. Weiss, H.R. Kaback, Single-molecule FRET reveals sugar-induced conformational dynamics in LacY, Proc. Natl. Acad. Sci. 104 (2007) 12640–12645. doi:10.1073/pnas.0700969104.

[41] R. Roy, A.G. Kozlov, T.M. Lohman, T. Ha, SSB protein diffusion on single-stranded DNA stimulates RecA filament formation, Nature. 461 (2009) 1092–1097. doi:10.1038/nature08442.

[42] C.C. Liu, P.G. Schultz, Adding New Chemistries to the Genetic Code, Annu. Rev. Biochem. 79 (2010) 413–444. doi:10.1146/annurev.biochem.052308.105824.

[43] Y. Shiraishi, T. Muramoto, K. Nagatomo, D. Shinmi, E. Honma, K. Masuda, M. Yamasaki,

Identification of Highly Reactive Cysteine Residues at Less Exposed Positions in the Fab Constant Region for Site-Specific Conjugation, Bioconjug. Chem. 26 (2015) 1032–1040. doi:10.1021/acs.bioconjchem.5b00080.

[44] R.B. Pepinsky, L. Walus, Z. Shao, B. Ji, S. Gu, Y. Sun, D. Wen, X. Lee, Q. Wang, E. Garber, S. Mi, Production of a PEGylated Fab′ of the anti-LINGO-1 Li33 antibody and assessment of its biochemical and functional properties in vitro and in a rat model of remyelination, Bioconjug. Chem. 22 (2011) 200–210. doi:10.1021/bc1002746.

[45] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins., Nat. Biotechnol. 22 (2004) 1302–6. doi:10.1038/nbt1012.

[46] O. Conchillo-Solé, N.S. de Groot, F.X. Avilés, J. Vendrell, X. Daura, S. Ventura, AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides., BMC Bioinformatics. 8 (2007) 65. doi:10.1186/1471-2105-8-65.

[47] I. Walsh, F. Seno, S.C.E. Tosatto, A. Trovato, PASTA 2.0: An improved server for protein aggregation prediction, Nucleic Acids Res. 42 (2014) 301–307. doi:10.1093/nar/gku399.

[48] M. Emily, A. Talvas, C. Delamarche, MetAmyl: A METa-Predictor for AMYLoid Proteins, PLoS One. 8 (2013). doi:10.1371/journal.pone.0079722.

[49] S.O. Garbuzynskiy, M.Y. Lobanov, O. V Galzitskaya, FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence, Bioinformatics. 26 (2010) 326–332. doi:10.1093/bioinformatics/btp691.

[50] P. Gasior, M. Kotulska, FISH Amyloid – a new method for finding amyloidogenic segments in proteins based on site specific co-occurence of aminoacids, BMC Bioinformatics. 15 (2014) 1–8. doi:10.1186/1471-2105-15-54.

[51] S. Maurer-stroh, M. Debulpaep, N. Kuemmerer, M. Lopez, D. Paz, I.C. Martins, J. Reumers, K.L. Morris, A. Copland, L. Serpell, L. Serrano, J.W.H. Schymkowitz, F. Rousseau, Exploring the sequence determinants of amyloid structure using position-

specific scoring matrices, Nat. Methods. 7 (2010) 237–242. doi:10.1038/nmeth.1432.

[52] A.C. Tsolis, N.C. Papandreou, V.A. Iconomidou, S.J. Hamodrakas, A Consensus Method for the Prediction of Aggregation-Prone' Peptides in Globular Proteins, PLoS One. 8 (2013) 1–6. doi:10.1371/journal.pone.0054175.

[53] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik, S. Ventura, AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures, Nucleic Acids Res. 43 (2015) 306–313. doi:10.1093/nar/gkv359.

[54] K. Sankar, S.R. Krystek, J. Stephen, M.C. Tyler, J.K.X. Maier, AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches, Proteins Struct. Funct. Bioinforma. (2018) 1–10. doi:10.1002/prot.25594.

[55] N. Chennamsetty, V. Voynov, V. Kayser, B. Helk, B.L. Trout, Design of therapeutic proteins with enhanced stability, Proc. Natl. Acad. Sci. 106 (2009) 11937–11942. doi:10.1073/pnas.0904191106.

[56] J. Van Durme, G. De Baets, R. Van Der Kant, M. Ramakers, A. Ganesan, H. Wilkinson, R. Gallardo, F. Rousseau, J. Schymkowitz, Solubis: a webserver to reduce protein aggregation through mutation, Protein Eng. Des. Sel. 29 (2016) 285–289. doi:10.1093/protein/gzw019.

[57] F. Bemporad, F. Chiti, "Native-like aggregation" of the acylphosphatase from Sulfolobus solfataricus and its biological implications, FEBS Lett. 583 (2009) 2630–2638. doi:10.1016/j.febslet.2009.07.013.

[58] A. Orte, N.R. Birkett, R.W. Clarke, G.L. Devlin, C.M. Dobson, D. Klenerman, Direct characterization of amyloidogenic oligomers by single-molecule fluorescence., Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 14424–14429. doi:10.1073/pnas.0803086105.

[59] M. Iljina, G.A. Garcia, M.H. Horrocks, L. Tosatto, M.L. Choi, K.A. Ganzinger, A.Y. Abramov, S. Gandhi, N.W. Wood, N. Cremades, C.M. Dobson, T.P.J. Knowles, D. Klenerman, Kinetic model of the aggregation of alpha-synuclein provides insights into

prion-like spreading, Proc. Natl. Acad. Sci. 113 (2016) E1206–E1215. doi:10.1073/pnas.1524128113.

[60] T.S. Young, I. Ahmad, J.A. Yin, P.G. Schultz, An Enhanced System for Unnatural Amino Acid Mutagenesis in E . coli, J. Mol. Biol. 395 (2010) 361–374. doi:10.1016/j.jmb.2009.10.030.

[61] S.I. Lim, Y.S. Hahn, I. Kwon, Site-specific albumination of a therapeutic protein with multi-subunit to prolong activity in vivo, J. Control. Release. 207 (2015) 93–100. doi:10.1016/j.jconrel.2015.04.004.

[62] J. Quan, J. Tian, Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries, Nat. Protoc. 6 (2011) 242–251. doi:10.1007/978-1-62703-764-8_8.

[63] M.J. Lajoie, A.J. Rovner, D.B. Goodman, H.R. Aerni, A.D. Haimovich, G. Kuznetsov, J.A. Mercer, H.H. Wang, P.A. Carr, J.A. Mosberg, N. Rohland, P.G. Schultz, J.M. Jacobson, J. Rinehart, G.M. Church, F.J. Isaacs, Genomically recoded organisms expand biological functions, Science. 342 (2013) 357–360. doi:10.1126/science.1241459.

[64] S. Jevševar, M. Kusterle, M. Kenig, PEGylation of Antibody Fragments for Half-Life Extension, Antib. Methods Protoc. 901 (2012) 233–246. doi:10.1007/978-1-61779-931-0.

[65] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P. V. Konarev, D.I. Svergun, New developments in the ATSAS program package for small-angle scattering data analysis, J. Appl. Crystallogr. 45 (2012) 342–350. doi:10.1107/S0021889812007662.

[66] D. Chivian, D. Baker, Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection, Nucleic Acids Res. 34 (2006) 1–18. doi:10.1093/nar/gkl480.

[67] S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. Dimaio, O. Lange, L. Kinch, W. Sheffler, B.H. Kim, R. Das, N. V. Grishin, D. Baker,

Structure prediction for CASP8 with all-atom refinement using Rosetta, Proteins Struct. Funct. Bioinforma. 77 (2009) 89–99. doi:10.1002/prot.22540.

[68] C. Zhang, M. Samad, H. Yu, N. Chakroun, D. Hilton, P.A. Dalby, Computational Design To Reduce Conformational Flexibility and Aggregation Rates of an Antibody Fab Fragment, Mol. Pharm. 15 (2018) 3079–3092. doi:10.1021/acs.molpharmaceut.8b00186.

[69] G.A. Kaminski, R.A. Friesner, J. Tirado-rives, W.L. Jorgensen, Evaluation and Reparametritzation of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B. 2 (2001) 6474–6487. doi:10.1021/jp003919d.

[70] H. Hofmann, D. Nettels, B. Schuler, Single-molecule spectroscopy of the unexpected collapse of an unfolded protein at low pH, J. Chem. Phys. 139 (2013). doi:10.1063/1.4820490.

[71] A. Ingargiola, T. Laurence, R. Boutelle, S. Weiss, X. Michalet, Photon-HDF5: An Open File Format for Timestamp-Based Single-Molecule Fluorescence Experiments, Biophys. J. 110 (2016) 26–33. doi:10.1016/j.bpj.2015.11.013.

[72] A. Ingargiola, E. Lerner, S.Y. Chung, S. Weiss, X. Michalet, FRETBursts: An open source toolkit for analysis of freely-diffusing Single-molecule FRET, PLoS One. 11 (2016) 1–27. doi:10.1371/journal.pone.0160716.

FIGURE LEGENDS

Fig 1. Structure of native Fab A33. Fab is composed of light (magenta) and heavy (yellow) chains. Each chain contains variable (V_L and V_H) and constant (C_L and C_H1) domains. The antigen-binding region at the complementary determining regions (CDRs; blue), are located in the variable domains. There are five disulphide bonds (gray highlights), four of them being intra-domain in V_L, V_H, C_L and C_H1, and the fifth is at the C-terminus between the light and heavy chains. The Fab A33 sequence is reported in Fig. S1.

Fig 2. SAXS Guinier analyses. Twenty experimental conditions were studied for Fab A33 using five pH (3.5, 4.5, 5.5, 7.0, 9.0) and four ionic strengths (20, 50, 150, 250 mM). (a) Guinier plots of ln *I(Q)* vs. Q^2 gave the R_g and I(0) values. Five representative fits are shown for each of pH 3.5, 4.5, 5.5, 7.0 and 9.0 in an ionic strength of 50 mM. The fits for native Fab A33 were determined using the *Q* range of 0.37-0.5 nm^{-1} (green) and those for aggregated Fab A33 was determined from the *Q* range of 0.14-0.3 nm^{-1} (red). (b) I(0) values for native Fab A33, where I(0)/c is proportional to the molecular weight, and error bars are the SEM of three measurements. (c) R_g values for native Fab A33 for each of the 20 experimental conditions studied, with error bars are the SEM of three measurements. (d) The amount of aggregate

present was determined from ΔI(0)/c, defined as $(I(0)_{Q:\ 0.14\text{-}0.3\ nm^{-1}} - I(0)_{Q:\ 0.37\text{-}0.5\ nm^{-1}})/c$.

Fig 3. Correlation between the R_g values and aggregation rates v. For each pH (see inset), the averaged R_g values for native Fab A33 and the aggregation rates v are shown for the four ionic strengths. Error bars are the SEM.

Fig 4. MD simulations of native Fab A33 at 300 K. (a, b) The R_g values and solvent accessible surface area (SASA) of Fab A33 are shown as a function of simulation time for five pH values as labelled, using an ionic strength of 50 mM for each. For each pH, three simulation repeats were averaged at every time frame, from which a window average is shown in a darker colour. (c, d) The root mean square fluctuation (RMSF) of the simulations at pH 7.0 and pH 3.5 respectively are shown in blue (low values) and red (high values) to highlight the dynamic regions in the structure. Front and back views are shown to follow the view of Figure 1. The RMSF values were added as notional B-factors to the PDB file for the Fab A33 homology model.

Fig 5. Comparison of the SAXS data with the MD simulations. (a) Comparison of the experimental X-ray scattering curve for native Fab A33 with the structures generated from the MD simulations for the five pH values as shown. MD simulations were carried out using an ionic strength of 50 mM at three temperatures of 300 K (green), 340 K (purple) and 380 K (grey). In total, each experimental SAXS curve was compared against 45,000 simulated structures per pH value. The goodness of fit was monitored using R-factors (Methods). The vertical lines represent the experimental R_g values with their experimental errors (SEM). The R_g value of each model was calculated from the theoretical scattering curve using the same Q-range used experimentally. The 10 best fit models with the lowest R-factors are highlighted in yellow. (b) Comparison of each experimental SAXS scattering curve (black) with its best-fit modelled curve

(red). The inset shows the comparison between the experimental and best fit modelled P(r) curves.

Fig 6. Alignment of the best-fit Fab A33 structures at pH 7.0. The ten best-fit simulated structures determined for pH 7.0 and an ionic strength of 50 mM are aligned in a cartoon representation. (a) Alignment of the light chain; (b) Alignment of the heavy chain.

Fig 7. Alignment of the best fit Fab A33 structures at pH 7.0 and 3.5. The ten best-fit simulated structures determined for pH 7.0 are shown in cyan and those for pH 3.5 are shown in magenta; both at an ionic strength of 50 mM. (a) Alignment of the light chain in which the C_L domain is highlighted to show its loop and helix displacements at low pH in three views. (b) Alignment of the heavy chain, in which a side view of the C_H1 domain is shown.

Fig 8. Distance monitoring using smFRET, SAXS and MD simulations. Two separations termed Dist 1 (residues LC-K126 and LC-S156) and Dist 2 (residues HC-S117 and LC-S156) were monitored at pH 7.0 and pH 3.5 in an ionic strength of 50 mM each. (a) Apparent single-molecule FRET efficiency (E_{app}) histograms of Dist 1 (green) and Dist 2 (gray) at pH 7.0 (dark colour) and pH 3.5 (light colour). At a FRET efficiency of 0.0, the population of molecules without an active acceptor fluorophore is shaded in gray. At higher FRET efficiencies, there is a population that corresponds to Fab A33 containing both fluorophores. This population was fitted with a Gaussian function and the peak is shown with a vertical line. (b) The averaged Dist 1 and 2 separations and their SD were measured from the ten best-fit SAXS structures at each of pH 7.0 (cyan) and pH 3.5 (magenta). (c) The Dist 1 (green) and Dist 2 (gray) separations as a function of simulation time for pH 7.0 (dark colour) and 3.5 (light colour) are shown from the MD simulations. Three simulation repeats were averaged at every time frame, from which a window average is shown in a darker colour.

Fig 9. Aggregation prone regions in Fab A33. (a) The consensus aggregation propensity of residues in Fab A33 was determined using PASTA 2.0, TANGO, AGGRESCAN and MetAmyl software. Using the native Fab A33 homology model, regions with greater aggregation propensities are shown in red and reduced propensities in blue. (b, c) Aggregation propensities in the SAXS best-fit structure for pH 7.0 and 3.5, respectively, are shown using a CPK spheres representation. The circled residues highlight the increase in SASA of APR 387-402 at pH 3.5 compared to pH 7.0. For reference, the residues used in smFRET labelling are highlighted in yellow.

Highlights

1. Elucidation of local changes in native conformation that promote protein aggregation.
2. Unprecedented resolution of native-like conformers by fitting MD simulations to SAXS data.
3. smFRET of dual-labelled Fab A33 confirms conformational changes.
4. Local unfolding of Fab A33 exposes a predicted aggregation-prone region.

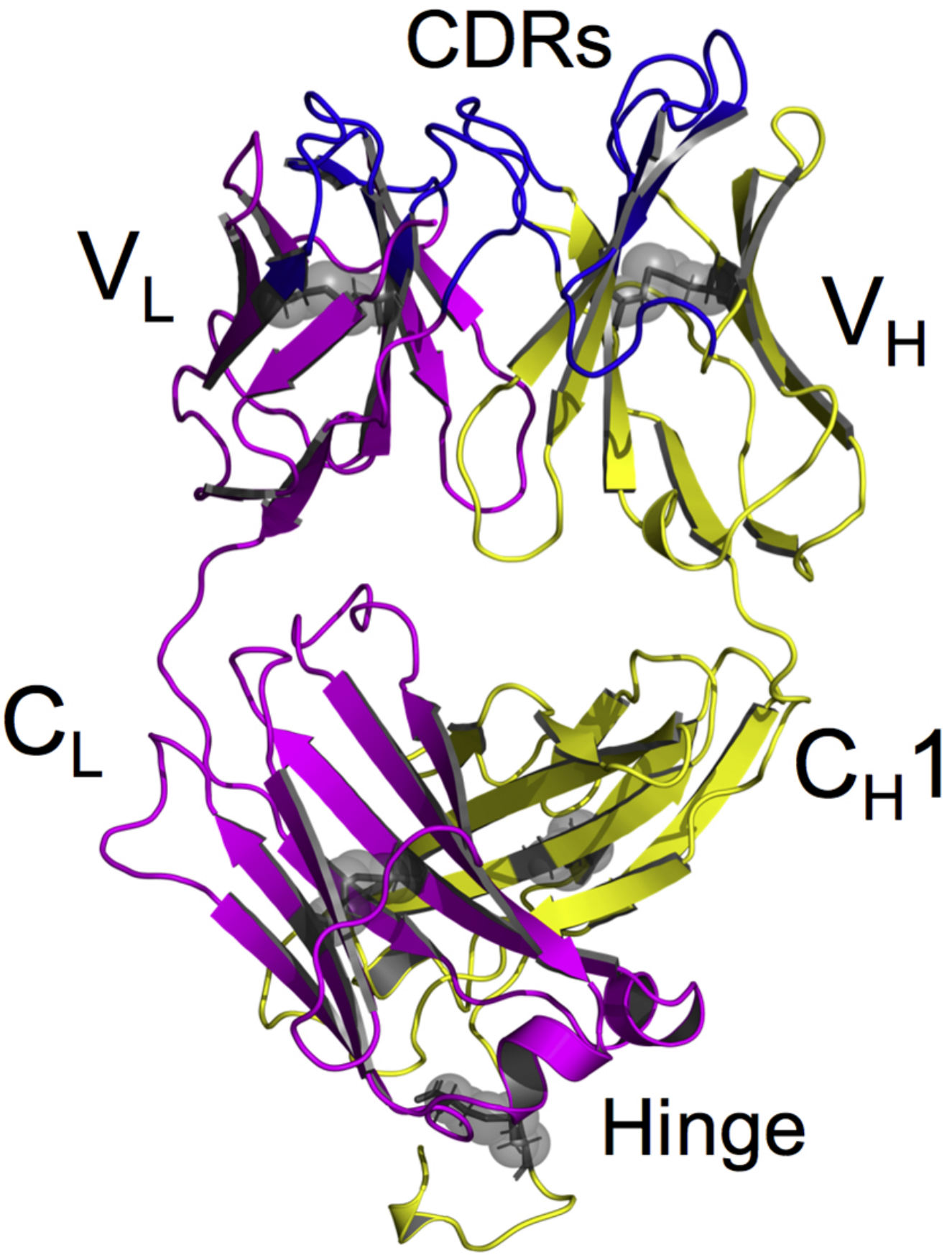

Figure 1

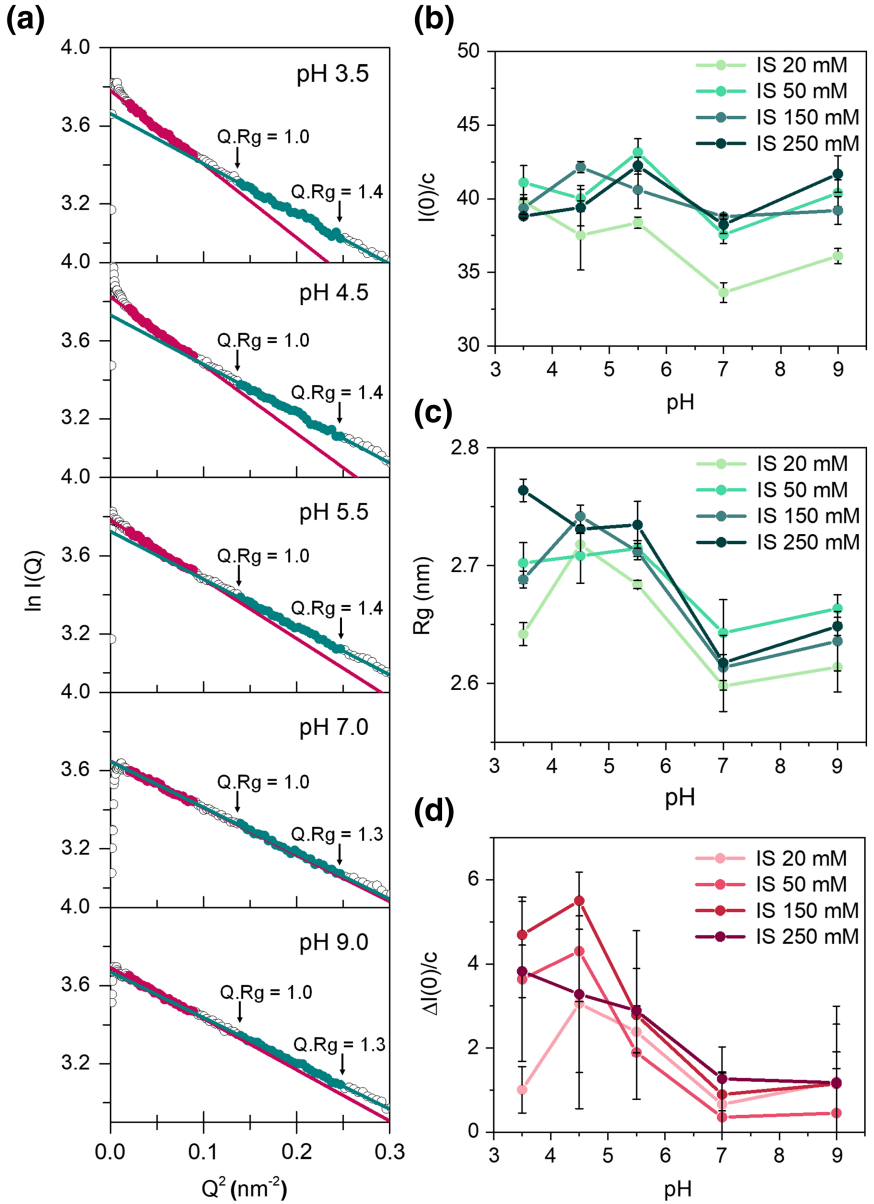
(a)
pH 3.5
Q.Rg = 1.0
Q.Rg = 1.4
pH 4.5
Q.Rg = 1.0
Q.Rg = 1.4
pH 5.5
Q.Rg = 1.0
Q.Rg = 1.4
pH 7.0
Q.Rg = 1.0
Q.Rg = 1.3
pH 9.0
Q.Rg = 1.0
Q.Rg = 1.3
ln I(Q)
Q² (nm⁻²)
(b)
IS 20 mM
IS 50 mM
IS 150 mM
IS 250 mM
I(0)/c
pH
(c)
IS 20 mM
IS 50 mM
IS 150 mM
IS 250 mM
Rg (nm)
pH
(d)
IS 20 mM
IS 50 mM
IS 150 mM
IS 250 mM
ΔI(0)/c
pH

Figure 2

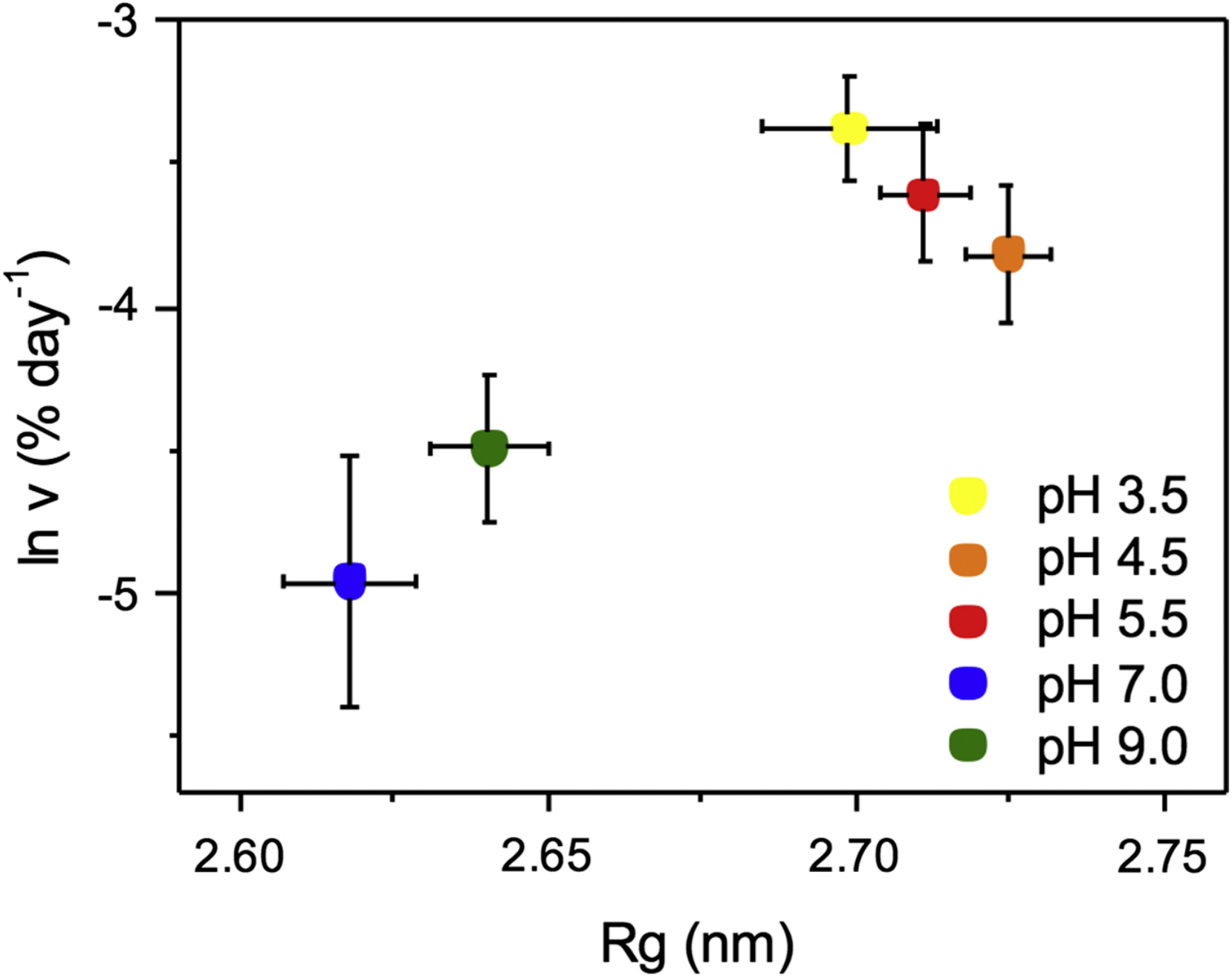

Figure 3

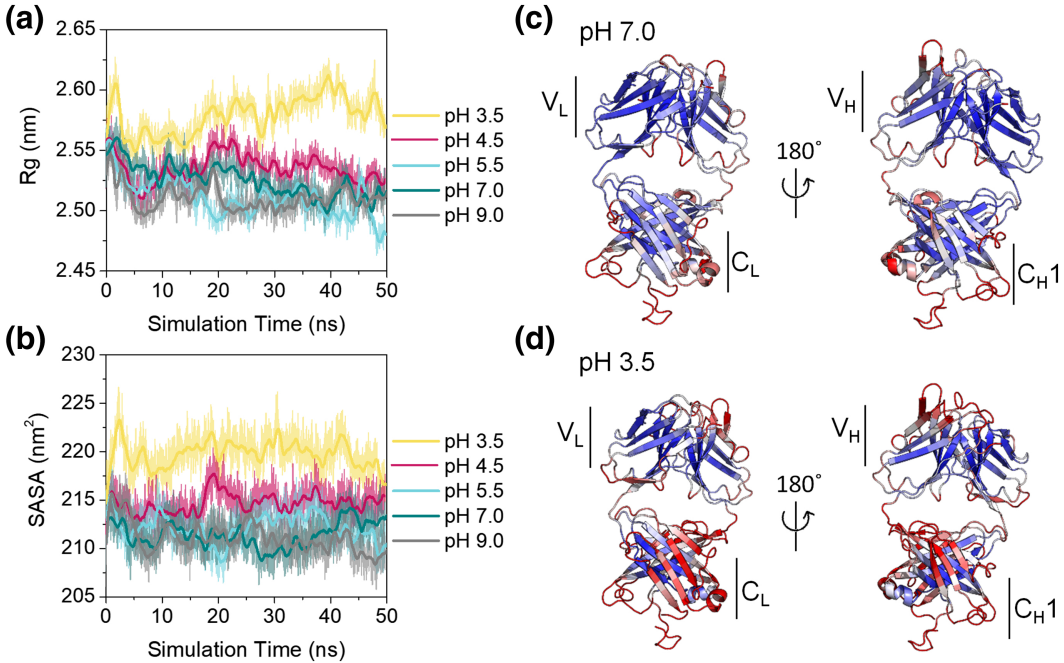
(a)
Rg (nm)
2.65
2.60
2.55
2.50
2.45
0
10
20
30
40
50
Simulation Time (ns)
pH 3.5
pH 4.5
pH 5.5
pH 7.0
pH 9.0
(b)
SASA (nm^2)
230
225
220
215
210
205
0
10
20
30
40
50
Simulation Time (ns)
pH 3.5
pH 4.5
pH 5.5
pH 7.0
pH 9.0
(c)
pH 7.0
V_L
C_L
180°
V_H
C_H1
(d)
pH 3.5
V_L
C_L
180°
V_H
C_H1

Figure 4

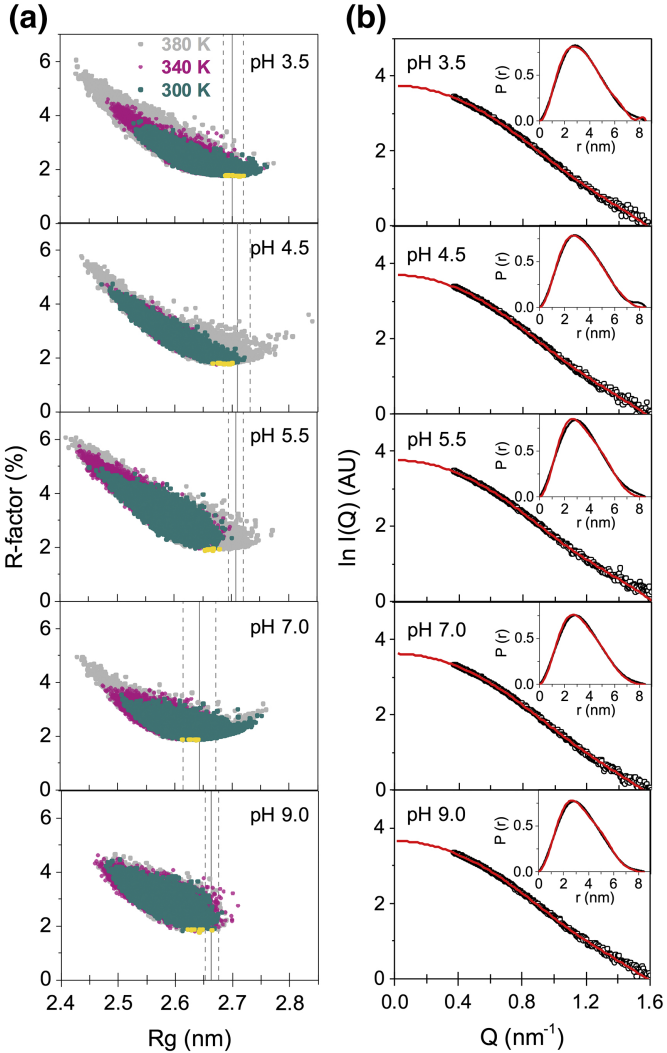
(a)
380 K
340 K
300 K
pH 3.5
pH 4.5
pH 5.5
pH 7.0
pH 9.0
R-factor (%)
Rg (nm)
(b)
ln I(Q) (AU)
Q (nm-1)
P (r)
r (nm)

Figure 5

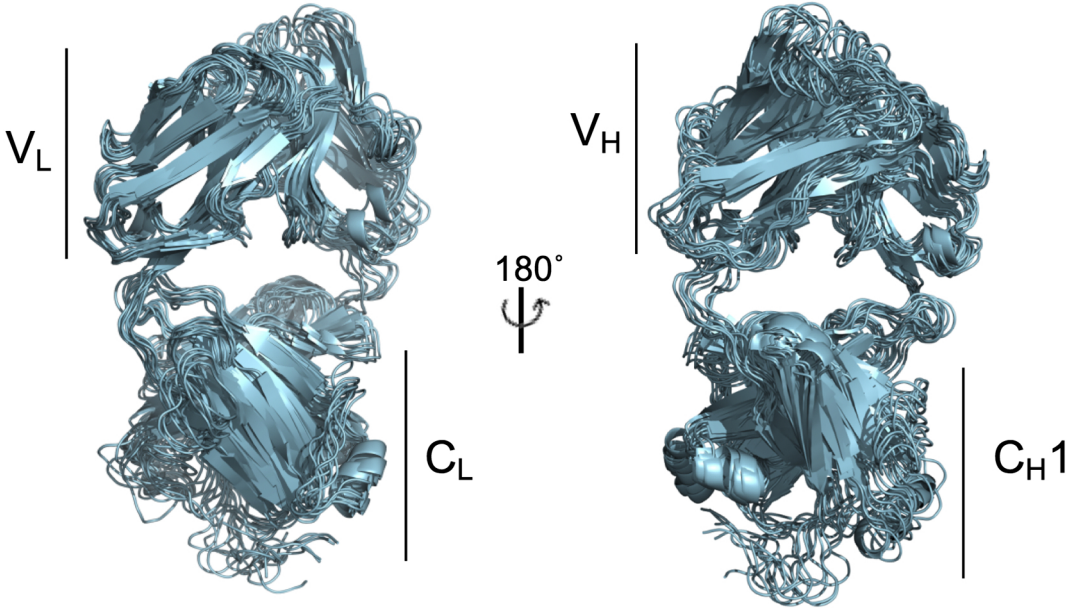

Figure 6

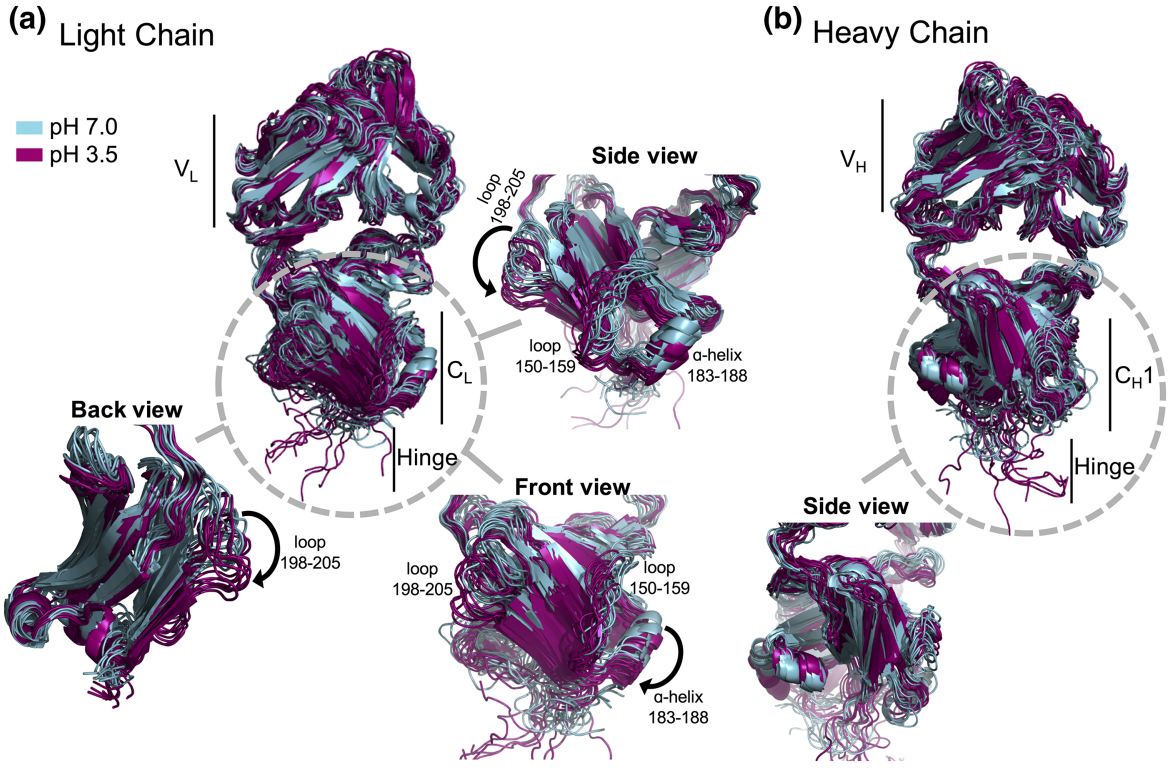
(a) Light Chain
pH 7.0
pH 3.5
V_L
C_L
Hinge
Side view
loop 198-205
loop 150-159
α-helix 183-188
Back view
loop 198-205
Front view
loop 198-205
loop 150-159
α-helix 183-188
(b) Heavy Chain
V_H
C_H1
Hinge
Side view

Figure 7

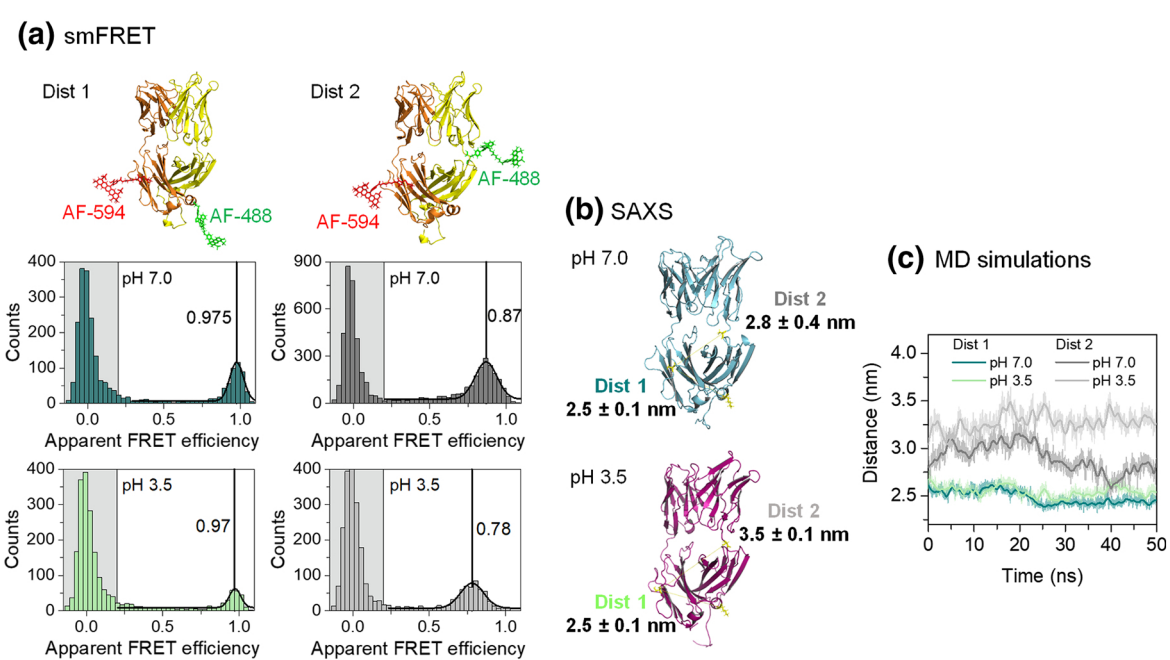
(a) smFRET
Dist 1
AF-594
AF-488
Dist 2
AF-488
AF-594
pH 7.0
0.975
Counts
Apparent FRET efficiency
pH 7.0
0.87
Counts
Apparent FRET efficiency
pH 3.5
0.97
Counts
Apparent FRET efficiency
pH 3.5
0.78
Counts
Apparent FRET efficiency
(b) SAXS
pH 7.0
Dist 2
2.8 ± 0.4 nm
Dist 1
2.5 ± 0.1 nm
pH 3.5
Dist 2
3.5 ± 0.1 nm
Dist 1
2.5 ± 0.1 nm
(c) MD simulations
Dist 1
pH 7.0
pH 3.5
Dist 2
pH 7.0
pH 3.5
Distance (nm)
Time (ns)

Figure 8

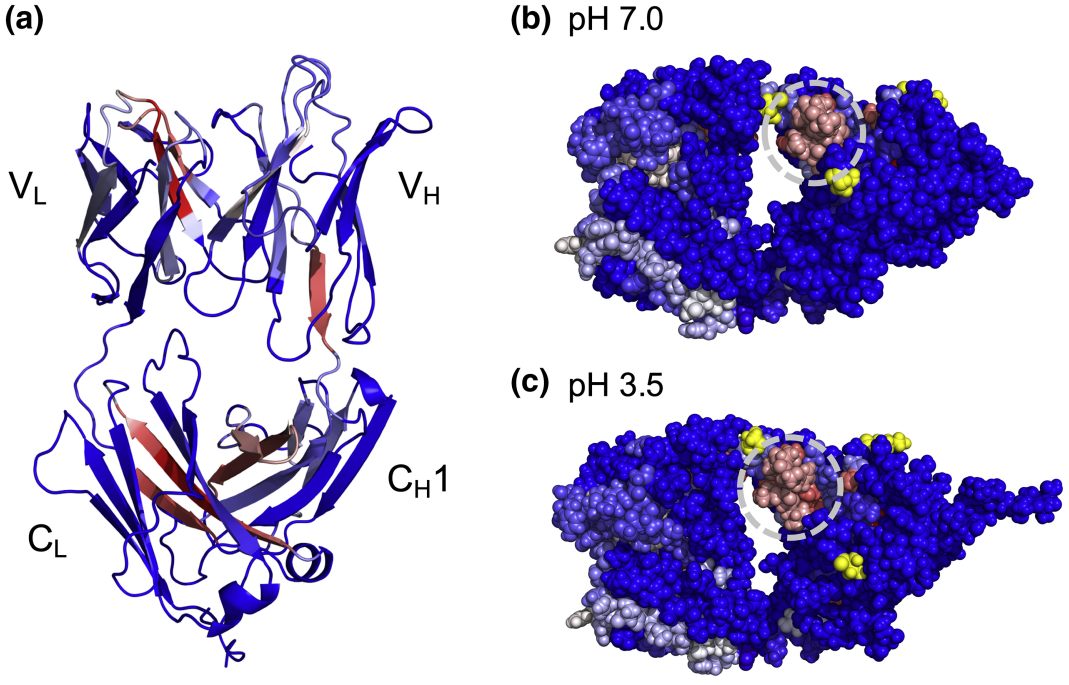

Figure 9