Hierarchical molecular dynamics of bovine serum albumin in concentrated aqueous solution below and above thermal denaturation

Electronic Supplementary Information

Marco Grimaldo,[a,b] **Felix Roosen-Runge,**[a] **Marcus Hennig,**[a,b] **Fabio Zanini,**[b,e] **Fajun Zhang,**[b] **Niina Jalarvo,**[c,d] **Michaela Zamponi,**[c] **Frank Schreiber,**[b] **Tilo Seydel**[*a]

1 Effect of an increase of the hydrodynamic radius on the theoretical and experimentally determined translational diffusion coefficients

The translational diffusion coefficient d_t becomes systematically lower than the value expected from the theory of colloids. The effect can be probably attributed to thermal expansion, which is however difficult to quantify. For example, as shown in Figure 1 by the gray dashed lines a linear increase of the hydrodynamic radius R_H by 3% from 295 to 330 K lowers the theoretical translational diffusion coefficient, since the protein volume fraction is increased (even though there is a competing decrease of water density with T) and the dilute limit translational diffusion coefficient $d_0(T) = k_\mathrm{B}T/(6\pi\eta(T)R_H)$ is decreased (k_B Boltzmann constant, $\eta(T)$ solvent viscosity) . The decrease of the rotational contribution to the experimental apparent d yields slightly larger d_t (gray symbols).

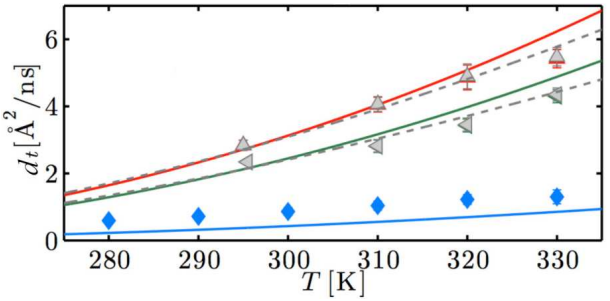


Figure 1 Translational diffusion coefficients of BSA solutions as a function of temperature at three protein concentrations (from top to bottom, $c_p = 150, 200$ and 500 mg/ml). The solid lines are the values expected for a colloidal suspension at fixed R_H, while the gray symbols and dashed lines are the experimental and theoretical coefficients assuming a thermal expansion of R_H by 3%.

[a] *Institut Max von Laue - Paul Langevin (ILL), CS 20156, F-38042 Grenoble, France*
[b] *Institut für Angewandte Physik, Universität Tübingen, D-72076 Tübingen, Germany*
[c] *Jülich Centre for Neutron Science (JCNS), Forschungszentrum Jülich, D-52425 Jülich, Germany*
[d] *Chemical and Engineering Materials Division, Neutron Sciences Directorate, and JCNS Outstation at the Spallation Neutron Source (SNS), Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[e] *Max-Planck Institute for Developmental Biology, Spemannstraße 35, D-72076 Tübingen, Germany*
[*] *e-mail: seydel@ill.eu*

2 Alternative models for the analysis of the scattering function

In the following, we present the analysis of the data with a series of models for different physical pictures. In Section 2.1 we show that the use of two components, one for the global and one for the internal dynamics, results in an unusual non-monotonicity of Γ as a function of q^2. In the following sections we show that the models tested here yield unphysical or inconsistent results.

2.1 Single Lorentzian for the global protein motions and single Lorentzian with HWHM Γ for the internal motions

The quasielastic scattering function can be modeled by[1,2]

$$S(q,\omega) = \mathscr{R} \otimes \{\beta(q)\left[A_0(q)\mathscr{L}_\gamma(\omega) + (1-A_0(q))\mathscr{L}_{\gamma+\Gamma}(\omega)\right] \ldots \\ \ldots + \beta_{D_2O}\mathscr{L}_{\gamma_{D_2O}}(\omega)\} \quad (1)$$

where $\mathscr{R}$ denotes the instrumental resolution function, modeled by a combination of gaussian functions, $\beta(q)$ is a scalar, and $A_0(q)$ represents the elastic incoherent structure factor (EISF). The two Lorentzians $\mathscr{L}_\gamma(\omega)$ and $\mathscr{L}_{\gamma+\Gamma}(\omega)$ account for two processes occurring at distinct time scales. The HWHM Γ accounts for fast internal protein motions, while γ describes the apparent diffusion of the entire protein in solution. Finally, the fixed term $\beta_{D_2O}\gamma_{D_2O}$ models the solvent contribution as explained in Ref.[3]. Γ is plotted in Figure 2 for three temperatures T (symbols), as indicated in the legend. At rather low temperatures, Γ can be well described by a jump-diffusion model as that by Singwi and Sjölander[4] (dotted blue line). However, at higher T, Γ has a peak at $q^2 \sim 0.4\,$Å, goes through a minimum and increases again at higher q^2. This non-monotonicity is inconsistent with a single processs, since it would imply that a particle explores a larger area faster than a smaller. The jump-diffusion model mentioned above cannot describe such a behavior (cf. Figure 2, green dot-dashed and red dashed lines), suggesting the presence of a multi-state diffusion.

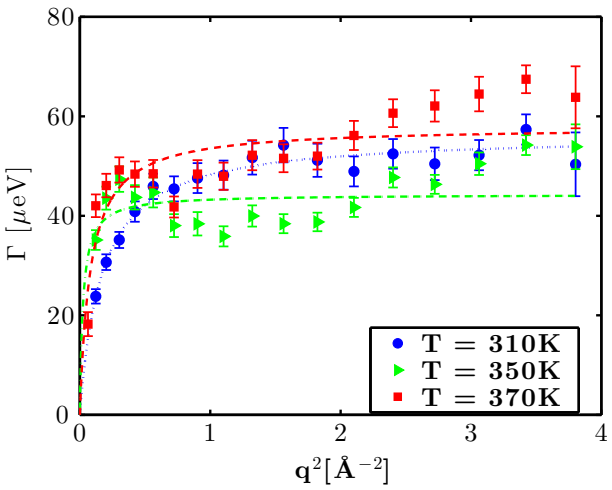


Figure 2 HWHM Γ of the Lorentzian function describing the internal motions as a function of q^2 for the model in equation (1). Using only one Lorentzian function for the fit of global diffusion and one for the internal motions results in a non-monotonic Γ at higher temperatures, suggesting the presence of at least one additional diffusive process.

2.2 Static distribution of two populations (clusters and monomers) - Homogeneous internal dynamics

Physical picture. There are two species in solution: clusters and monomers. Their volume fractions are ϕ_c and ϕ_m, respectively, with $\phi_m = 1 - \phi_c$. Proteins are thus either in a monomeric form or in a cluster, and do not exchange in the time-scale of the experiment. The overall internal dynamics can be described by a single effective Lorentzian function and is the same for the monomers and for the proteins in a cluster. The corresponding scattering function is:

$$S(q,\omega) = \beta(q)\{\left[\phi_c \mathscr{L}_{\gamma_c}(\omega) + (1-\phi_c)\mathscr{L}_{\gamma_m}(\omega)\right] \otimes [A_0\,\delta(\omega) + (1-A_0)\mathscr{L}_\Gamma(\omega)]\} \tag{2}$$

With the assumption that at such short time-scales the global dynamics can be described by a Fickian law, also at high temperatures, $\gamma_j = d_j q^2$ (d is the diffusion coefficient, $j = m,c$) and a global fit over all the spectra at different q can be performed.
Parameters. Cluster fraction ϕ_c and cluster apparent diffusion coefficient d_c and monomer apparent diffusion coefficient d_m are global parameters, $\beta(q)$, A_0 and Γ are q-dependent.
Expectations. ϕ_c should be ~ 0 at room temperature and up to the denaturation temperature; then, increase with T and become the main component. d_c should be considerably lower than d_m. On the other hand, d_m should not be higher than $d_0 q^2$, and, as long as the cluster fraction is low, be consistent with colloid theory.

The obtained parameters are plotted as a function of T in Figure 3.

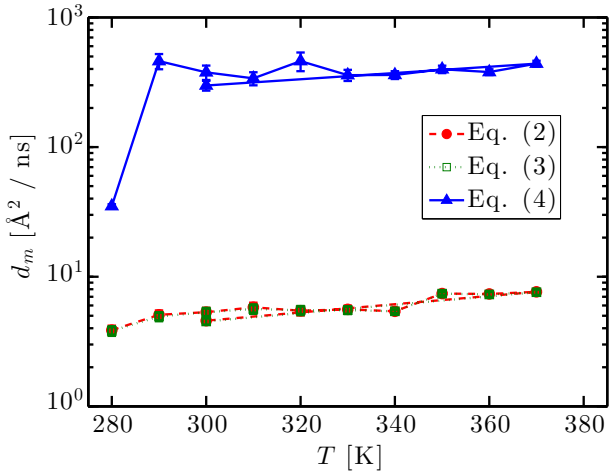
- ϕ_c indicates that, even at temperatures well below denaturation, the main population is composed by clusters (Figure 3(c)), and is thus inconsistent with previous findings[1].
- The faster diffusion coefficient d_m associated with monomers, is much higher than $d(c_p = 500mg/ml)$, while the slow component d_c which should account for clusters is closer to the value expected for monomers (cf. Figures 3(a) and (b)).
- Having introduced the component $d_c q^2$ removes the "unusual" non-monotonicity of Γ discussed in the previous section: Γ seems now rather consistent with a Singwi-Sjölander jump-diffusion model, as shown in Figure 4, *left*.

2.3 Distribution of dynamical clusters and monomers: switching model for the global diffusion (2 alternating diffusive states) - Homogeneous internal dynamics

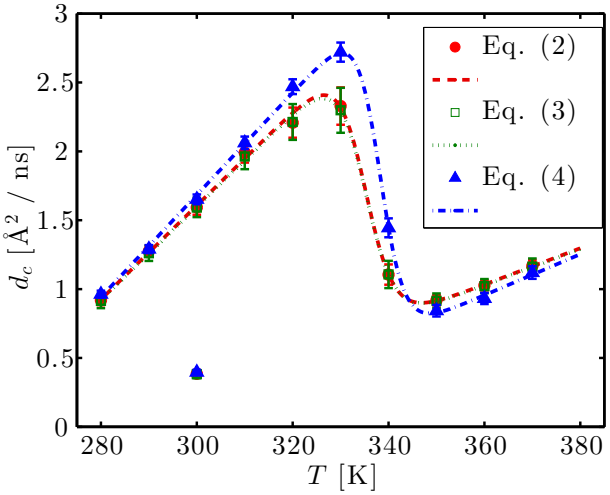
Physical picture. Proteins are alternating between a state diffusing with a cluster and a state diffusing as monomers. Such exchange is visible on the time-scale of the experiment. The overall internal dynamics can be described by a single effective Lorentzian function and is the same for the monomers and for the proteins in a cluster. The corresponding scattering function is:

$$S(q,\omega) = \beta\{S_{\mathrm{sw}}(q,\omega)) \otimes [A_0\,\delta(\omega) + (1-A_0)\mathscr{L}_\Gamma(\omega)]\}\ , \tag{3}$$

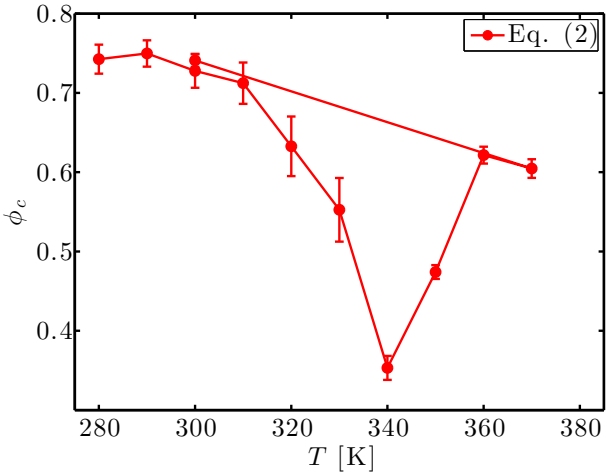
where $S_{\mathrm{sw}}(q,\omega)$ is the scattering function of the switching model (Equation (2) in the article). Assuming that at such short time-scales the global diffusion of monomers and clusters follow a Fickian law, $\gamma_j = d_j q^2$ (d is the diffusion coefficient, $j = m,c$) and a global fit over all the spectra at different q can be done.
Parameters. The monomer d_m and cluster d_c apparent diffusion coefficients, and the residence times τ_m and τ_c in the two states are global parameters, β, A_0 and Γ are q-dependent.



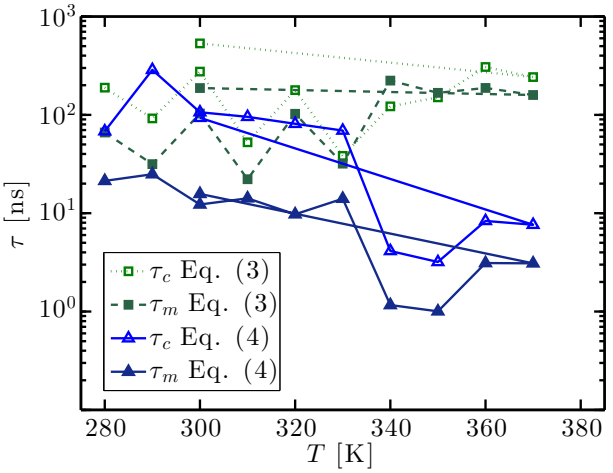
(a) Apparent diffusion coefficient d_m as a function of temperature T resulting from the models in equations (1)-(3), as reported in the legend, for $c_p = 500$ mg/ml. d_m is associated with the diffusion of monomers, since it is the faster global diffusive process in the respective models. In all cases shown here, and especially for the last model, d_m is significantly higher than the theoretical diffusion coefficient at $c_p = 500$ mg/ml, being thus inconsistent with the assumptions in the models.



(b) Apparent diffusion coefficient d_c as a function of temperature T resulting from the models in equations (1)-(3), as reported in the legend, for $c_p = 500$ mg/ml (symbols). The lines are guides to the eye. d_c is associated with the diffusion of clusters, since it is the slower global diffusive process in the respective models. In all cases shown here, the extraction of the translational diffusion coefficient from the apparent coefficient d_c would be close to the theoretical translational diffusion coefficient of monomers at $c_p = 500$ mg/ml. Together with the values of d_m in Figure 3(a), this suggests that our data are not consistent with the presence of monomers *and* clusters in solution.



(c) Fraction ϕ_c of proteins in clusters calculated from model (2) as a function of T. This result would suggest a high amount of clusters in solution, even far away from denaturation, which seems highly unlikely, also combined with the results in Figures 3(a) and (b).



(d) Residence times τ_m and τ_c as a function of T for $c_p = 500$ mg/ml, calculated from the fits of the scattering data by equations (3) and (4), as indicated in the legend. In general, their values are much higher than the time-range accessible by the instrument, and are thus inconsistent with the assumption made in the model that monomers bind and unbind from clusters on the time-scale sampled by the instrument. Furthermore, τ_c is always greater than τ_m, meaning that clusters are the main component, which is again inconsistent with expectations.

Figure 3 Fit parameters d_m, d_c, ϕ_c, τ_m and τ_c from various models as discussed in the captions below the subfigures. These parameters are essentially unphysical or inconsistent with the assumptions made in the respective models.

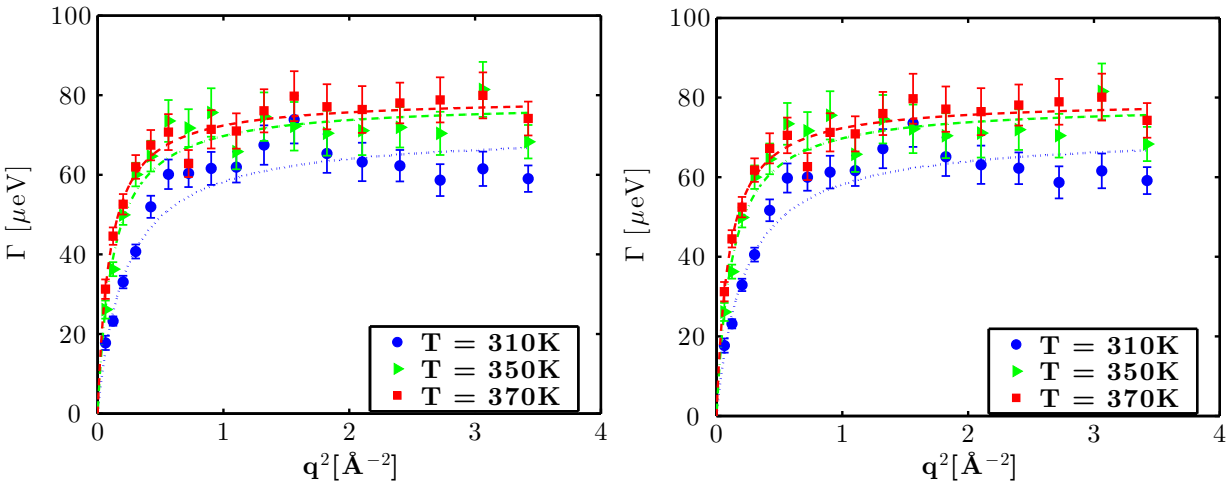


Figure 4 HWHM Γ of the Lorentzian function describing the internal motions as a function of q^2 for the model in equation (2) (*left*) and that in equation (3) (*right*). Adding a component for the protein global diffusion essentially removes the non-monotonicity of Γ, although the diffusion coefficients obtained for the global diffusion are not consistent with the physical picture (see text).

Expectations. d_m and d_c should be consistent with the theoretical diffusion coefficients of monomers and clusters, respectively. Moreover, it is reasonable to expect that at low temperatures τ_c is ~ 0, meaning that clusters are almost completely absent.

The obtained parameters are plotted as a function of T in Figure 3 and are almost exactly the same as for the previous model.

We observe that:

- d_m seems too big, at least at low T, and d_m and d_c are not consistent with a population of monomers and clusters.
- Both τ_c and τ_m are of the order of several hundreds of nanoseconds, which is not consistent with the assumption that the proteins are switching from one to the other diffusive state on the time-scale accessible by the instrument. Such result would rather point back to a static picture (if it wasn't for the first observation). Note that $\tau(T)$ is plotted without the error bars, since these would be very large.
- As with the previous model, the addition of the Lorentzian with HWHM $d_c q^2$ removes the non-monotonicity of Γ, which is here consistent with jump-diffusion (see Figure 4 *right*).

2.4 Distribution of dynamical clusters and monomers: switching model for the global diffusion (2 alternating diffusive states) - Heterogeneous internal dynamics modeled by 2 alternating diffusive states

Physical picture. Proteins are alternating between a state diffusing with a cluster and a state diffusing as monomers. Such exchange happens within the time-scale of the experiment. Also the internal dynamics is modeled by two alternating diffusive states. The corresponding scattering function is:

$$S(q,\omega) = \beta\{S^g_{\mathrm{sw}}(q,\omega)) \otimes \left[A_0\,\delta(\omega) + (1-A_0)\,S^i_{\mathrm{sw}}(q,\omega))\right]\}\ , \tag{4}$$

where $S^g_{\rm sw}(q,\omega)$ and $S^i_{\rm sw}(q,\omega)$ are the scattering functions of the switching models for the global and internal motions respectively. Assuming that at such short time-scales the global dynamics can be described by a Fickian law, $\gamma_j = d_j q^2$ (d is the diffusion coefficient, $j = m,c$) and a global fit over all the spectra at different q can be done.
Parameters. d_m,d_c, τ_m, τ_c, D_1,D_2, $\tau_1^{(\rm int)}$, $\tau_2^{(\rm int)}$ are global parameters, β and A_0 are q-dependent.
Expectations. d_c and d_m should be consistent with the diffusion of a cluster and a monomer. Moreover, it is reasonable to expect that at low temperatures $\tau_c \ll \tau_m$, meaning that clusters are almost completely absent.

The obtained parameters are plotted as a function of T in Figure 3.

- d_m is several hundreds Å^2 / ns . Not consistent with diffusion of monomers.
- Both τ_c and τ_m are of the order of some nanoseconds, which is not consistent with the assumption that the proteins are switching from one to the other diffusive state on the time-scale accessible by the instrument.

3 Effect of the H/D-exchange between the native proteins and the deuterated solvent water

Labile H-atoms of the native proteins in solution may exchange with the deuterium from the solvent water. This exchange increases the amplitude of the solvent water in the contribution to the total scattering signal. In the following we estimate the maximum possible H/D-exchange and with the increased solvent amplitude repeat the fits reported in the article to which the present document is the Supplementary Information. We display the fit results assuming the H/D-exchange in the figures 5,6, and 7. Due to the remaining uncertainty on the actual magnitude of the H/D-exchange and the water structure factor, we do not display these figures in the article itself.

We obtain the number 776 of labile, i.e. exchangeable H-atoms per BSA protein from the protein data base (PDB) file[5], using VMD to determine the surface amino acids and a MATLAB script to count the number of exchangeable H-atoms according to Ref.[6]. Subsequently, we assume that all labile H-atoms actually exchange. The number density of exchangeable H-atoms then reads $n_H = 776 n_{BSA}$ where $n_{BSA} = c_{BSA}/M_{BSA}/(c_{BSA}\vartheta_{BSA} + 1\text{ml})$ is the number density of protein at the nominal concentration c_{BSA} (mg/ml). $\vartheta_{BSA} = 0.74\,$ml/g is the specific volume and $M_{BSA} = 66500\,$g/mol is the molecular weight of BSA. The number density of D-atoms from the water solvent is given by $n_D = 2\cdot 55.5M[1 - c_{BSA}\vartheta_{BSA}/(c_{BSA}\vartheta_{BSA} + 1\text{ml})]$. The contamination fraction of H-atoms α in the solvent is then simply calculated via $r_H = n_H/(n_H + n_D)$. Using this estimation, we obtain $\alpha \approx 1\%$, 2%, and 5% H-contamination of the solvent water by atom number density, respectivly, for the protein concentrations 100 mg/ml, 200 mg/ml, and 500 mg/ml, respectively.

Assuming a similar dynamics of all atomic components in the water, the total scattering cross section $\sigma_{tot}^{(X)}$ of the H/D-mixture $X = \alpha H + (1-\alpha)D$ can then simply be calculated from the tabulated values for the total scattering cross sections of H, D, and O:

$$\sigma_{tot}^{(X)} = 2\alpha\sigma_{tot}^{(H)} + 2(1-\alpha)\sigma_{tot}^{(D)} + \sigma_{tot}^{(O)} \tag{5}$$

With this expression, we obtain the following factors for the increase of the total scattering of the H-contaminated D_2O-solvent with respect to pure D_2O: 1.08 for 1% contamination corresponding to 100 mg/ml BSA, 1.16 for 2% contamination (200 mg/ml), 1.38 for 5% contamination (500 mg/ml).

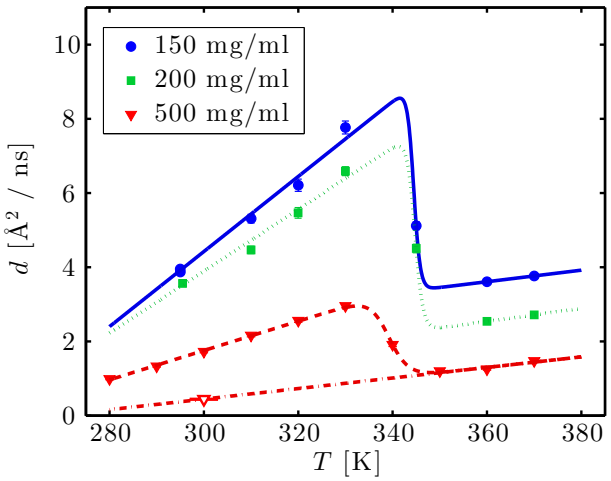


Figure 5 Apparent self-diffusion coefficient d as a function of T for $c_p = 150, 200$ and $500\,$mg/ml (symbols) and fits with the two-state switching diffusion model (see article, lines), assuming an H/D-exchange between the proteins and the solvent water as outlined in section 3.

We have performed the analysis with these assumptions for the H/D-exchange, and the results are depicted in the Figures 5, 6 and 7. We observe that the fit results are overall very similar to those obtained when not taking into account the H/D-exchange.

References

[1] F. Roosen-Runge, M. Hennig, F. Zhang, R. M. J. Jacobs, M. Sztucki, H. Schober, T. Seydel and F. Schreiber, *PNAS*, 2011, **108**, 11815–11820.

[2] M. Grimaldo, F. Roosen-Runge, F. Zhang, T. Seydel and F. Schreiber, *J. Phys. Chem. B*, 2014, **118**, 7203–7209.

[3] M. Grimaldo, F. Roosen-Runge, N. Jalarvo, M. Zamponi, F. Zanini, M. Hennig, F. Zhang, F. Schreiber and T. Seydel, *EPJ Web of conferences, in press*, 2015.

[4] K. S. Singwi and A. Sjölander, *Phys. Rev.*, 1960, **119**, 863–871.

[5] A. Bujacz, *Acta Crystallographica Section D*, 2012, **68**, 1278–1289.

[6] B. Jacrot, *Reports on Progress in Physics*, 1976, **39**, 911.

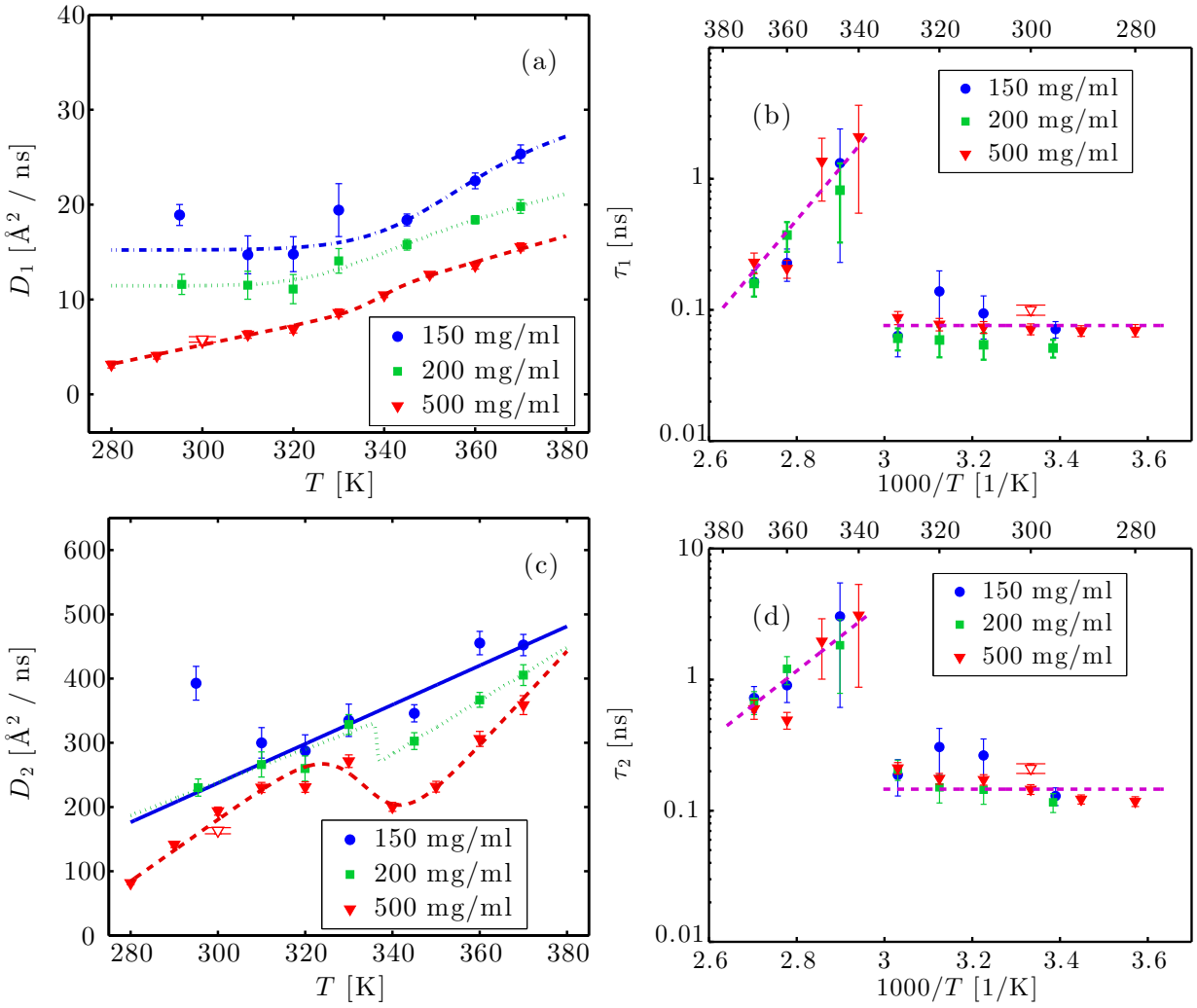


Figure 6 Fit results for the two-state switching difusion model assuming an H/D-exchange between the proteins and the solvent water as outlined in section 3: (a) D_1 as a function of T (symbols). The lines are guides to the eye. (b) Arrhenius plot of the residence time τ_1 between two jumps of the side-chains versus T for three c_p (symbols). The data above denaturation were fitted with an Arrhenius equation (line), while at low temperatures the line is a guide to the eye. (c) D_2 as a function of T for the concentrations reported in the legend (symbols). The lines are guides to the eye. The illustrations depict solvent-inaccessible side-chains in the folded protein (left) becoming solvent-exposed in the unfolded protein (right). (f) Arrhenius plot of the residence time τ_2 as a function of T (symbols). The data above denaturation were fitted with an Arrhenius equation (line). The red open triangles in (a), (c)-(f) refer to the sample at 500 mg/ml cooled back to room temperature after denaturation.

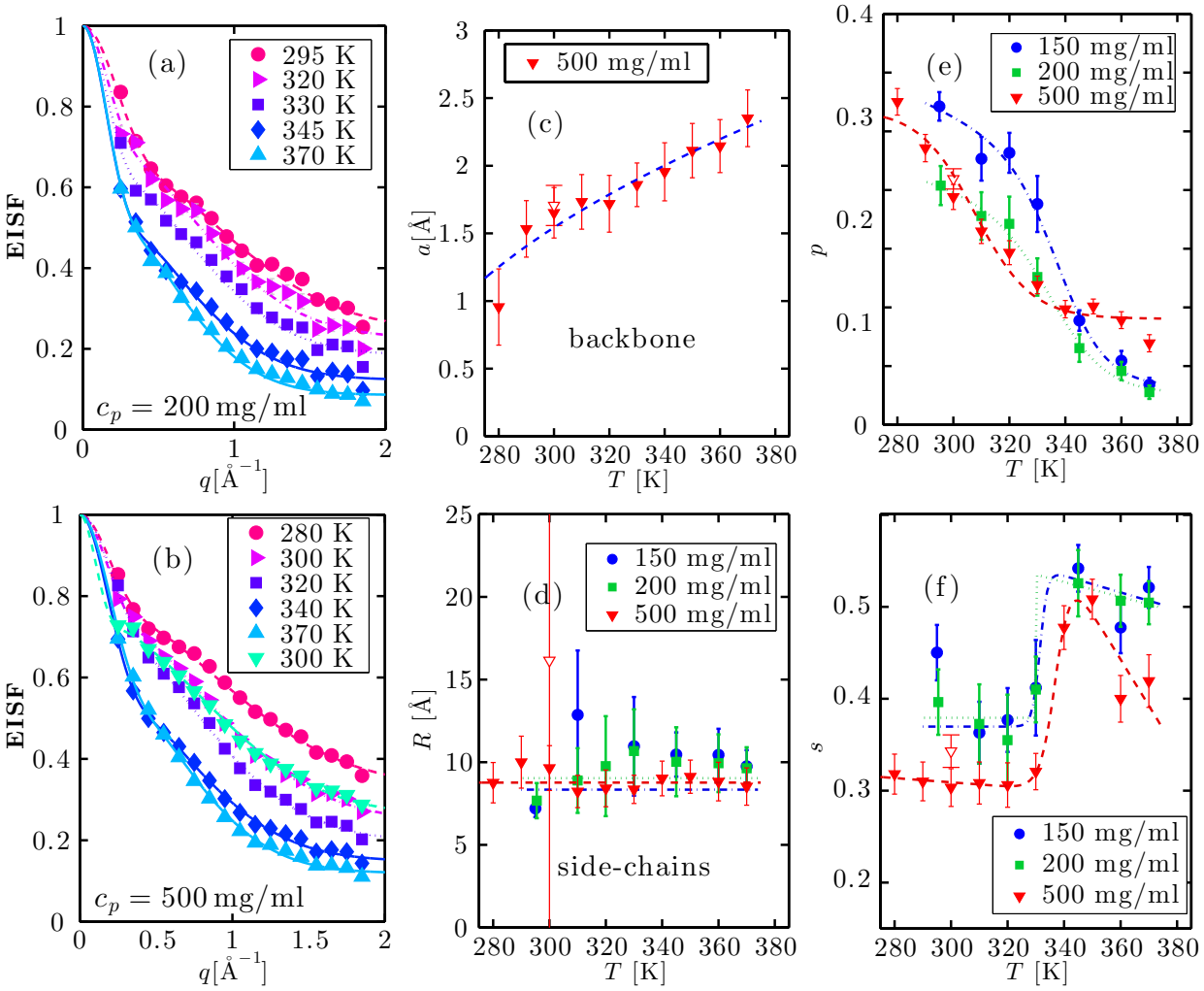


Figure 7 Fit results for the EISF associated with the two-state switching difusion model assuming an H/D-exchange between the proteins and the solvent water as outlined in section 3: (a) and (b): EISF as a function of q at the temperatures given in the legend for $c_p = 200$ and 500 mg/ml, respectively, (symbols) and fits as described in the article (solid lines). (c) radius a as a function of T for $c_p = 500$ mg/ml (symbols) and fit (blue solid line). a is associated with the effective sphere accessible by backbone atoms. (d) Radius R as a function of T (symbols). The lines are guides to the eye. This radius defines the sphere accessible by side-chain motions. (e) Fraction of immobile atoms p as a function of T (symbols). The lines are guides to the eye. (f) s as a function of T (symbols) defining the ratio of side-chains describable with a diffusion in a sphere model to the total amount of mobile side-chains. The line is a guide to the eye. The parameters in Figures 3(c)-(f) are obtained from the fit of the EISF, and the open symbols in Figures 3(b)-(f) refer to the sample at $c_p = 500$ mg/ml cooled down to room temperature after irreversible denaturation.