Anisotropic protein-protein interactions in dilute and concentrated solutions

Pasquier, Coralie; Midtgaard, Soren Roi; Polimeni, Marco; Jorgensen, Christian Isak; Arleth, Lise; Callisen, Thomas H.; Lund, Mikael

Published in:
Journal of Colloid and Interface Science

DOI:
10.1016/j.jcis.2022.08.054

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Pasquier, C., Midtgaard, S. R., Polimeni, M., Jorgensen, C. I., Arleth, L., Callisen, T. H., & Lund, M. (2023). Anisotropic protein-protein interactions in dilute and concentrated solutions. Journal of Colloid and Interface Science, 629, 794-804. https://doi.org/10.1016/j.jcis.2022.08.054
Anisotropic protein-protein interactions in dilute and concentrated solutions

Coralie Pasquier a,b,* Søren Roi Midtgaard c, Marco Polimeni b, Christian Isak Jørgensen d, Lise Arleth c,* Thomas H. Callisen b, Mikael Lund b,*

a ICSM, CNRS, CEA, Univ Montpellier, ENSCM, Marcoule, France
b Division of Theoretical Chemistry, Lund University, POB 124, SE-221 00 Lund, Sweden
c X-ray and Neutron Science, Niels Bohr Institute, University of Copenhagen, Denmark
d Novozymes A/S, Denmark
e LINXS – Lund Institute of Advanced Neutron and X-ray Science, Scheelevägen 19, SE-223 70 Lund, Sweden

ABSTRACT

Interactions between biomolecules are ubiquitous in nature and crucial to many applications including vaccine development; environmentally friendly textile detergents; and food formulation. Using small angle X-ray scattering and structure-based molecular simulations, we explore protein–protein interactions in dilute to semi-concentrated protein solutions. We address the pertinent question, whether interaction models developed at infinite dilution can be extrapolated to concentrated regimes? Our analysis is based on measured and simulated osmotic second virial coefficients and solution structure factors at varying protein concentration and for different variants of the protein Thermomyces Lanuginosus Lipase (TLL). We show that in order to span the dilute and semi-concentrated regime, any model must carefully capture the balance between spatial and orientational correlations as the protein concentration is elevated. This requires consideration of the protein surface morphology, including possible patch interactions. Experimental data for TLL is most accurately described when assuming a patchy interaction, leading to dimer formation. Our analysis supports that the dimeric proteins predominantly exist in their open conformation where the active site is exposed, thereby maximising hydrophobic attractions that promote inter-protein alignment.

© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Introduction

Understanding and controlling protein colloidal stability associated with aggregation phenomena are besides biological interest, of significant importance in biotechnology where it affects expression, fermentation, purification, formulation, and eventually efficacy of produced protein molecules [1]. It is also central to formulation of protein-based pharmaceuticals such as insulin, where the controlled colloidal properties of the proteins directly couple to their release profile when used as a drug [2]. The factors influencing stability are already intricate in controlled and pure systems, but in native systems there is an added level of complexity due to e.g. glycosylation and unprocessed (signal) pro-peptide that biotech-produced protein samples typically have, depending on the choice of host organism and production conditions [3–5].

In this work we investigate protein–protein interactions of a 30 kDa lipase (TLL) from the fungal organism Thermomyces lanuginosus (formerly Humicola lanuginosa), The TLL enzyme protein is a triacylglycerol lipase (EC 3.1.1.3) that catalyses the hydrolysis of ester-bonds in substrates such as triglycerides. This type of lipase has widespread use from stereo-specific synthesis of drugs to laundry detergents where it enables removal of fatty and oily soil and allows for washing under more sustainable conditions, e.g. at lower temperatures and with reduced use of chemicals [6,7]. TLL exhibits interfacial activation, i.e. the activity of TLL is strongly enhanced upon recruitment to an interface where the substrate is enriched. In the course of this process, TLL undergoes a conformational change upon shifting the local environment: from the aqueous environment of high dielectric constant to a more hydrophobic (substrate-rich) environment of lower dielectric constant. The lid domain – made of residues 82–98 and covering the hydrophobic active-site region – rolls over, thereby exposing the active site to the interface [8–10].

To study intermolecular interactions between different variants of TLL in aqueous solutions we performed small-angle X-ray Scattering (SAXS) combined with molecular modelling. Both SAXS and the related technique, small-angle neutron scattering (SANS), are commonly used to study non-ideal macro-molecular solutions [11–14], including the effect of protein glycosylation [15,16]. We here complement experimental scattering studies with computer simulations aiming to shed light on the origin and detail of the observed aggregation properties. Molecular Dynamics (MD) is used to establish models for equilibrium structures of the different protein molecules and are subsequently used in coarse grained Monte Carlo simulations (MC) to model one-body, two-body, and many-body interactions. With this analysis we address the pertinent question: can parameters valid at dilute conditions be reliably transferred to more concentrated regimes?

2. Methods and Theory

2.1. Protein samples

Purified TLL lipase samples were prepared for SAXS experiments with and without N-linked deglycosylation, and with and without the presence of a SPIRR (Ser-Pro-Ile-Arg-Arg) propeptide at the N-terminus. For the deglycosylation, Endoglycosidase H from Streptomyces plicatus was purchased from Sigma-Aldrich, (Saint-Louis, USA) and was used by incubation in a pH 5.5, 50 mM NH₄Ac + 200 mM NaCl buffer at 30°C for 1 h. For SAXS experiments, samples were buffer exchanged by size exclusion and run over a Superdex200 column at an Äkta purifier using a pH 8, 10 mM Tris + 100 mM NaCl buffer. Based on stock concentrations obtained in the range of 6–8 mg/ml, we prepared dilution series of the samples that went down to 0.3–0.6 mg/ml (sample dependent). The final samples were characterised by intact molecular weight analyses by mass spectrometry (see following paragraph).

2.1.1. Protein concentrations

The protein samples were dried, then hydrolysed in 18.5% HCl + 0.1% phenol at 110°C for 16 h. Amino acid analyses were performed by precolumn derivatisation using the Waters AccQ-Tag Ultra Method. In short, amino acids were derivatised by the AccQ-Tag Ultra Reagent and separated with reversed-phase UPLC (UPLC, Waters Corp., Milford, MA), and the derivatives quantitated using UV absorbance with a molar extinction coefficient of 3592 M⁻¹ cm⁻¹ at 280 nm [17,18].

2.2. Mass Spectrometry

The intact molecular weight analyses were performed using a MAXIS II electrospray mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). The samples were first diluted to 0.1 mg/ml in MilliQ water. The diluted samples were applied to an AdvanceBio Desalting-RP column (Agilent Technologies) followed by washing and elution from the column running an acetonitrile linear gradient and introduced to the electrospray source with a flow of 400 ml/min by an Ultimate 3000 LC system (Dionex). Data analysis was performed with DataAnalysis version 4.3 (Bruker Daltonik GmbH, Bremen, DE). The molecular weight was calculated by deconvolution of the raw data in the range 20 to 60 kDa.

The lipase sequence contains one N-glycosylation site (AsnXaa-Thr, sequence motif) that is glycosylated with a high mannose type glycan structure. The N-glycosylation is efficiently removed by EndoH (Endo–N-acetylglucosaminidase H). EndoH cleaves the bond in the diacetyltibiose core of the oligosaccharide between two N-acetylglucosamine (GlcNAc) subunits directly proximal to the asparagine residue, generating a truncated sugar molecule with one N-acetylglucosamine residue remaining on the asparagine. Thus by measuring the intact protein mass after EndoH treatment it is possible to precisely measure the percentage of glycosylated molecules in the protein sample by simply comparing the intensity of the unmodified mass peak with the mass peak + 1xGlcNAc. A spreadsheet showing the calculations is included in the supporting information.

2.3. Small-Angle X-ray Scattering

SAXS experiments were performed on the BioXolver L (Xenocs) placed at Department of Drug Design and Pharmacology, at the University of Copenhagen. During the measurements, the instrument was equipped with a 250 W liquid gallium alloy X-ray source (MetalJet) with a wavelength of λ = 1.34 Å. The scattering data were collected at 298 K in 20 x 30 s exposures with a sample-detector distance of d = 654 mm. This provided a q-range of 0.011 – 0.50 Å⁻¹, with q defined as q = 4π/λ sin(θ), where 2θ is the scattering angle. Data was averaged using RAW [19] and absolute scale calibrated using water as a reference. Pair distance distribution functions, p(r), were obtained by indirect Fourier transformation performed using BayesApp [20].

2.4. All Atom Molecular Dynamics

As the structure of Thermomyces lanuginosus lipase (TLL) with the SPIRR fragment attached to the N-terminus (TLL+) is not already available, all atom molecular dynamics (MD) simulations were performed to generate those structures, for the closed and open forms of TLL. The PDB structures of the closed (1DU4) and open (1EIN) form of TLL [21] were used as starting points for the
SPIRR-containing propeptides, see Fig. 1. The SPIRR fragment was attached to the N-terminus of the two forms of TLL using PyMol [22] and protonation states of pH 8 were obtained using PDB2PQR [23]. To relax the generated structure, MD simulations were performed using Amber 16.12 [24], with the ff14SB forcefield [25], SPC/E water [26] and Li/Merz parameters for monovalent ions for the ff14SB forcefield [25]. Harmonic $C_2$ constraints of 10 kcal/mol/Å$^2$ were applied during the minimisation. Constant pressure (1 bar) was maintained using a Monte Carlo (MC) barostat [28], with volume change attempts every 100 steps. Langevin dynamics with a collision frequency of 1.0 ps$^{-1}$ were used to maintain a constant temperature of 298 K [29]. A real-space cutoff of 8 Å was applied to non-bonded interactions and the remaining electrostatic interactions were accounted for using the particle-mesh Ewald procedure with a Fourier spacing of 72 Å. The system was equilibrated for 500 ps, followed by 500 ns of sampling. A principal component analysis was applied to the trajectory of atoms in order to identify discrete conformational states and the structure corresponding to the lowest free energy region was selected as the final structure.

2.5. Metropolis Monte Carlo Simulations

For studying protein–protein interactions at multiple concentrations, we constructed a coarse grained (CG) interaction model for the closed and open forms of TLL starting from the PDB structures 1DU4, 1EIN, and as well as MD generated variants with SPIRR attached. Each residue was represented by a spherical bead, which could be charged or neutral, depending on residue type and solution pH. In all simulations pH = 8, giving a wildtype TLL net charge of −30 elementary unit charges, and protonation states with their molecular weight, and $q$ is the density defined in [30].

For hydrophobic residues (ALA, ILE, LEU, MET, PHE, PRO, TRP, VAL) residues with a solvent accessible surface area (SASA) larger than 30 Å$^2$ [31], a specific hydrophobic-hydrophobic LJ strength, $\varepsilon_{m-n}$, was used to match experimental virial coefficients at low protein concentrations, see Results and Discussion. SASA was measured using a spherical probe of radius 1.5 Å.

De-glycosylation of TLL leaves a mannose unit on ASN 33, and we thus added an uncharged bead of radius equivalent to that of a residue (2.9 Å) along the surface of ASN 33 and with $\varepsilon_{m} = 0.005$, assuming negligible attraction to protein residues. For the cases where TLL is glycosylated (TLLm), a neutral bead of radius 20 Å was attached to the previously mentioned mannose unit.

2.5.1. Two-body Interactions

Two proteins were placed inside a cylindrical cell ($L = 500$ Å, $R = 80$ Å) with hard sides and periodic ends, see Fig. 2, middle. Proteins were allowed to rotate and translate along the $z$-axis of the cylinder only. For TLLm+ and TLL+, we considered that the proportion of proteins having the SPIRR propeptide attached was of 63%, as determined by mass spectrometry. Thus, the simulation box contained one protein with SPIRR and one protein without SPIRR. The angularly averaged two-body protein–protein distribution function, $g(r)$, was sampled via the histogram method as a function of mass-center separation, $r$. The potential of mean force (PMF) is obtained by Boltzmann inversion, $\beta w(r) = -\ln g(r) + c$ where the free energy of the reference state, $c$, is chosen such that $w(r) \rightarrow 0$ for large $r$. The second virial coefficient was estimated from the PMF via $B_2 = 2 \pi \int_0^\infty \left( e^{w(r)} - 1 \right) r^4 dr$ and was normalised by the same second virial coefficient of hard spheres ($B_{2HS}$) as used for the experimental data.

During the simulations, the contacts between the residues of the two proteins were recorded. A contact between a residue of protein 1 and a residue of protein 2 was considered effective if the distance between the surfaces of the two residues was less than 3 Å. The number of contacts for each residue was then summed at the end of the simulation and normalised so that the total number of contacts for one protein is 100%.

2.5.2. Many-body simulations

325 rotating and translating proteins were placed in a cubic cell with periodic boundaries and volumes matching the experimental protein concentrations. A spherical cut-off of 5 times the Debye length was used to truncate the pair-potentials to reduce the computational time. Due to oligomer formation, a cluster move with a

**Fig. 1.** Two different views of the superimposed PDB structures of the closed TLL (PDB 1DU4, cyan/orange) and open TLL (PDB 1EIN, purple/red). The lid is coloured in orange for the closed form, and red for the open form. Left and right show side and top view, respectively.
threshold of 6 Å was used to rotate and translate aggregated proteins. Protein mass center positions were used to compute the concentration dependent structure factor, \( S(q) \), by explicitly averaging Eq. 2 over the 3+6+4 directions obtained by permuting the crystallographic index \( \frac{100}{C_138} \), \( \frac{110}{C_138} \), \( \frac{111}{C_138} \) to define the scattering vector \( q = \frac{2\pi}{L} \) where \( p = 1,2,\ldots,p_{\text{max}} \). The sampled \( q \)-interval is thus \( 2\pi/L, 2\pi p_{\text{max}} \sqrt{3}/L \), \( L \) being the cube side length and \( p_{\text{max}} = 15 \).

\[
S(q) = \frac{1}{N_p} \left( \sum_{i=1}^{N_p} \sin(qr_i) \right)^2 + \left( \sum_{i=1}^{N_p} \cos(qr_i) \right)^2
\]

Equation 2

\( N_p \) is the number of proteins; \( r_i \) protein mass center positions. For comparison, we also evaluate \( S(q) \) using the Debye formula \([33,34]\),

\[
S_d(q) = 1 + \frac{2}{N_p} \sum_{j=1}^{N_p} \sum_{j=1}^{N_p} \sin(qr_{ij})/qr_{ij}
\]

Equation 3

Although Eq. 3 is strictly not valid for periodic cells, the interaction range in our system is short compared to the box size and it is interesting to compare with the more slowly converging Eq. 2.

2.5.3. One-body simulations

In both two-body and many-body simulations, salt implicitly accounted for using a Debye length, \( 1/\kappa \), c.f. Eq. 1. To validate this approximation, we performed a series of one-body, Grand canonical Monte Carlo (GCMC) simulations of a single protein in explicit salt solution. The 1:1 NaCl salt particles and protein counter ions consisting of charged beads (\( \sigma_{\text{Na-Cl}}=4.6 \) Å, \( \sigma_{\text{Na-Cl}}=0.01 \) \( kT \)) with a constant chemical potential, are inserted in a spherical cell with radius \( R = (3/4\pi N_s)^{1/3} \). The mobile configurational space was sampled by a combination of single ion translational moves and grand canonical insertion and deletion of electroneutral salt pairs (Fig. 2, left). The interaction energy is still Eq. 1 albeit with a Debye length of infinity as salt is now handled explicitly, plus a term accounting for the energy fluctuation due to the residue protonation state, \( kT \sum_{i} (\beta_{\text{pKa}-i} - \phi_{\text{pH}})(\ln(10)) \).

3. Results and Discussion

We initially study four variations of TLL, see Fig. 3, and in the first part we discuss the experimental SAXS observations, followed by a dissection using theory and molecular modelling.

---

Table 1

| Residue type      | \( \sigma_{ii}/\text{Å} \) | \( \mu_{ii} \) |
|-------------------|-----------------|--------------|
| hydrophobic       | 7.0             | variable, denoted “eH” |
| non-hydrophobic   | (6Mw/4\pi\rho)^{1/3} | 0.05 or 0.005, denoted “eLJ” |

Fig. 2. Top: Simulation models for one, two, and many-body proteins used to predict \( \gamma_s \), \( B_2 \), and \( S(q) \), respectively (see text for details). Amino acid residues interact via Eq. 1 and attractive interactions between hydrophobic residues (black) are, solely using \( \epsilon_{HH} \), calibrated to match experimental virial coefficients for wildtype TLL at low protein concentrations. Bottom: Lennard-Jones interaction matrix between bead types (left) and an illustration of the Lennard-Jones interaction using two different \( \epsilon_{LJ} \) parameters for non-hydrophobic beads (right).
and a structure factor, introduction of the second virial coefficient

\[ p \text{ concentration (units of g/m}^3; \]

monomers, such that for a set of SAXS data at different protein concentrations, \( c_p \), the structure factor at \( q = 0 \), \( S(c_p, 0) \), is related to a virial expansion [38],

\[ \frac{1}{S(c_p, 0)} = \frac{P(0)}{I(c_p, 0)/A(c_p)} = 1 + 2M_aA_2c_p + \ldots \]  

where \( A_2 \) is the second virial coefficient in mol·m\(^3\)·g\(^{-2} \), \( c_p \) the protein concentration in g/m\(^3\), and \( M_a \) the protein molar mass in g/mol. \( P(0) = 1 \) as per definition. As mentioned above, the experimental \( I(c_p, 0) \) values, were approximated through IFT analysis yielding the \( p(r) \) functions. For each series, we then plotted \( 1/S(c_p, 0) \) as a function of the protein concentration and performed a linear regression on the data points. The slope of this curve equals \( 2M_aA_2 \). We converted \( A_2 \) to m\(^3\)·mol\(^{-2} \), \( B_{278} = 2\pi(2R_p)^3/3 \), where \( R_p \) is the radius of the protein, for which we take the hydrodynamic radius estimated by Zhu et al. [39], 27.8 Å.

**Table 2** shows the obtained \( B_2/B_{278} \) and we note that for the same state of glycosylation, the addition of the SPIRR propedte decreases the value of \( B_2/B_{278} \). This attraction is reasonable as the addition of the SPIRR propedte reduces the protein net charge from \(-7e\) to \(-5e\) and adds two hydrophobic residues. Similarly, glycosylation of the protein leads to a \( B_2 \) increase as expected from the effect of increased steric repulsion between the proteins.

| 3.2. Metropolis Monte Carlo Simulations |

In this section we develop a structural model to understand and predict the scattering intensities and osmotic second virial coefficients. The strategy is to first calibrate a *two-body* model – valid at low protein concentrations – with the experimental virial coefficient for wildtype TLL ([Table 2](#)). Next, we test if the model is able to predict virial coefficients for the variants as well as *many-body* effects and SAXS spectra occurring at higher protein concentrations.

| 3.2.1. Two-body Interactions |

The osmotic second virial coefficient reports exactly on the pair-interaction between two proteins in the limit of infinite dilution. TLL exists in a closed and open form, with the latter having \( \sim 20\% \) more hydrophobic surface area exposed to the solvent. To determine how to model short-range attractive interactions, we explore two schemes for non-electrostatic interactions between the proteins (see also [Fig. 2](#)):

1. Non-hydrophobic residues have a *low* LJ-parameter, \( \beta_{c_i} = 0.005 \), effectively meaning that the LJ interaction is chiefly repulsive and therefore contributes to excluded volume only. This model has previously been used to calculate structure factors for protein solutions [30].

2. Non-hydrophobic residues have a *moderate* LJ-parameter, \( \beta_{c_i} = 0.05 \) which mimics a generic short-range attraction distributed uniformly over all residues. This model has been used in a number of studies for protein–protein interactions [40,41] and would in the limit of perfectly spherical proteins correspond to a colloidal Hamaker constant [42].

For both models, a single LJ interaction parameter between hydrophobic residues, \( \beta_{c_{H-H}} \), is fitted to match the experimental virial coefficient for wildtype TLL.

Two proteins are placed in a cylindrical geometry ([Fig. 2](#), middle) to calculate \( B_2/B_{278} \) for the open and closed forms as a function of the interaction strength between surface hydrophobic residues, see [Fig. 5](#). All curves possess the same features: a region at low \( g_{c,2}\) where the protein–protein interaction is predominantly repulsive \( (B_2/B_{278} > 0) \) and independent of the value of the hydrophobic attraction, therefore likely dominated by the electrostatic compo-
...then a transition to a region where the value of $B_2 = B_{2\text{HS}}$ is extremely sensitive to $e_{\text{H}/C_0\text{H}}$. Note that in all simulated cases, the experimental value of $B_2 = B_{2\text{HS}}$ is located in this latter region, showing the importance of taking the attractive interactions into account in order to describe the behaviour of the studied lipase solutions.

The trends clearly show an effect of the protein conformation on the interactions, as the open conformation greatly reduces (0.255 or 0.34 $kT$ depending on the value of $e_{\text{LJ}}$) the value of $e_{\text{H}/C_0\text{H}}$ needed to match the experiments. This is due to a higher exposure of hydrophobic residues in the case of the open conformation: the total SASA of the hydrophobic residues of the open form is indeed 1.2 times the one of the closed form, primarily since four residues become solvent-accessible due to the opening of the lid. The contact between the two proteins is indeed strongly predominant in the lid region, as can be seen on Fig. 6 for wild-type TLL.

The values of $e_{\text{H}/C_0\text{H}}$ best fitting experiment are shown in Table 3 and are now used to predict the virial coefficients for the TLL variants. Fig. 7 shows that the parameterisation, based purely on wild-type data, well reproduces the TLLm variant for all interaction models, i.e. open/closed, $e_{\text{LJ}}=0.05$ or 0.005. However, the agreement between simulation and experiment becomes worse in the case of the SPIRR-containing forms of the lipase (TLL+ and TLLm+). Further, simulating using the closed form of TLL gives larger deviations than using the open form, i.e. the protein–protein interaction is too attractive when attaching SPIRR. This result fits the expectation that when the proteins are in close proximity, they do indeed take the open form which is stabilised by short-range interactions.

It is important to note that the results presented here are valid at low protein concentrations and the question is if the developed...
models can be transferred to higher concentrations? This is the topic of the next section.

### 3.2.2. Many-body Interactions

The values of $e_{i,j}$ matching the experimental data determined in the dilute two-body simulations (Table 3) have been kept as a reference in the coming many-body simulations at higher protein concentrations. We now exclusively focus on wildtype TLL, and investigate the effect of the open/closed conformations, and values of $e_{LJ}$. Fig. 8, left shows the simulated structure factors obtained for the four different interaction models and it is clear that large differences arise as the protein concentration increases. Remember that by design, all four models give the same virial coefficient in the dilute regime. The right-hand side of Fig. 8 shows the concentration normalised scattering intensities, $I(q) = I_{exp}$ from the simulations in comparison with experiment. Clearly the interaction scheme using a closed structure of TLL and a moderate LJ parameter ($e_{LJ} = 0.05$) differs significantly from experiment due to strong attraction as the proteins move closer. The remaining three models all give similar experimental agreement, although using the open structure and a moderate LJ interaction gives the best overall agreement at all concentrations.

Fig. 9 shows the values of $1/S(0)$ for each condition at different concentrations, each obtained by fitting the low $q$ part of the corresponding $S(q)$. The root-mean-square deviation of the simulated values of $1/S(0)$ to the experimental ones is presented in Table 4.

This confirms our previous notion that the open TLL structure and an intermediate LJ parameter of 0.05 $kT$ gives the best agreement with experiment. It also highlights the danger of transferring interaction models from infinite dilution to high concentration where multiple proteins may simultaneously interact. A particularity of the present system is that the protein–protein interaction is driven largely by a hydrophobic patch [14], i.e. the active site. This calls for a particularly well-balanced interaction between the generic LJ interaction (working between all residues) and the specific hydrophobic attraction. We can balance this at infinite dilution to match e.g. osmotic second virial coefficients, but further introducing concentration series is a much more stringent, and here crucial, test.

### 3.2.3. Effect of Structural Flexibility

In the previous section we observed that the open and closed forms of TLL gave rise to significantly different protein–protein interactions. When free in solution, we expect the closed form to dominate as it is unfavorable to expose hydrophobic groups to the solvent. Upon association, the lid is expected to open, thereby stabilising the dimer through hydrophobic interactions as previ-

---

**Table 3**

| Conformation | $e_{LJ}$ | $e_{H}/C_0 H$ |
|--------------|---------|--------------|
| closed       | 0.05    | 0.61         |
| closed       | 0.005   | 0.825        |
| open         | 0.05    | 0.355        |
| open         | 0.005   | 0.485        |

**Fig. 6.** Simulated contact map for wild-type TLL, for the closed and open forms, with a general Lennard-Jones $e_{ij}$ of 0.05 and 0.005. The residues composing the catalytic triad (Ser146, Asp201 and His258) are shown in white. Top: view of the lid/active site region; bottom: 90 degree rotation of the protein.

**Fig. 7.** Absolute deviation from the experimental values of the normalised simulated second virial coefficients ($B_2/B_{2HS}$) for TLL, TLLm, TLL+, and TLLm+ obtained with the parameters described in Table 3. The blue background colours reflect the cell values.
ously shown for *Pseudomonas fluorescens* lipase [36] and suggested for several other lipases, including TLL [35]. We now mimic this situation in the coarse grained simulations by introducing a MC swap move [43–45]. Here there is a dynamic exchange between the open and closed form. A caveat is that the intrinsic probability or free energy barrier to exchange between the two forms is currently unknown. We therefore arbitrarily set this barrier to zero, corresponding to the artificial situation where the two forms are equally likely. While this is certainly a harsh simplification, the missing term is somewhat captured in the heuristic parameterisation of the hydrophobic residue interaction, $\epsilon_{LJ}/C_0$. For the two LJ interaction models, i.e. “weak” ($\beta L J = 0.005$) and “moderate” ($\beta L J = 0.05$) the hydrophobic interaction, $\epsilon_{LJ}$, is 0.537 and 0.404, respectively. Using this parameterisation from the low protein concentration regime we now investigate the effect of increasing the concentration.

**Fig. 8.** Left: Simulated structure factors, $S(q)$, for wildtype TLL at different protein concentrations, $c_p$, and for four different interaction schemes: open/closed protein; weak/strong Lennard-Jones attraction. At infinite dilution, all four interaction schemes reproduce the experimental virial coefficient. Lines represent averages from three simulations, while shaded zones represent the span of values from those three simulations. Right: Simulated and zoomed concentration-normalised intensities, $I(q)/c_p$, at different protein concentrations and for different interaction schemes. The lines represent the averages from three simulations, while the dotted lines represent the normalised experimental intensities.

**Fig. 9.** Experimental and simulated $1/S(0)$ values as a function of protein concentration, $c_p$, obtained by extrapolating measured and calculated scattering intensities, $I(q)$. Simulated data is shown for both the open and closed structures of TLL as well as for two different interaction models (red and blue) as discussed in the text.

**Table 4**

| Conformation | $\beta L J$ | RMSD |
|--------------|------------|------|
| closed       | 0.005      | 0.766|
| closed       | 0.005      | 0.012|
| open         | 0.05       | 0.001|
| open         | 0.005      | 0.091|

**Fig. 10.** Simulated fraction of proteins in the open form as a function of protein concentration, $c_p$, and two different interaction models, both matching the experimental virial coefficient at low concentrations, see text. Data points represent averages over three independent simulation runs.
tions. As expected, in the dilute limit we recover the 1:1 ratio between the open and closed form, bearing in mind that this is an artificial choice.

Similar to the discussion in the Many-body Interaction section we have different models based on either a “moderate” or a “weak” general Lennard-Jones interaction scheme that both reproduce the experimental second virial coefficient in the limit of infinite protein dilution. To more stringently test these models, we perform simulations at finite protein concentration and compare scattering intensities with experimental SAXS data. As shown in Fig. 11, the two models give increasingly different results as the protein concentration is elevated, highlighting the importance of correctly balancing the strength and range of site-site interactions. Incorporating dynamic exchange between the open and closed form, the best overall agreement is seen for a “weak” to negligible LJ attraction as also used in previous studies [30]. It is important to reiterate that in our current implementation of the MC swap move, we artificially assume a 1:1 distribution of the open and closed form, bearing in mind that this is an artificial choice.

3.2.4 Effect of Protein on the Salt Excess Chemical Potential

For protein solutions prepared by dialysis against a buffer solution, the salt chemical potential in the sample, by definition, equals that of the buffer. That is, for different protein concentrations the salt activity is constant, yet the salt concentration, $c_p$, may vary due to the chemical environment in the protein solution. In models for protein–protein interactions it is however often assumed that the salt concentration (not activity) in the buffer solution, $c_s$, equals $c_p$, or $\gamma_s = c_s/c_p \approx 1$. In this section we investigate if this approximation holds as it affects the ionic strength used to calculate the Debye screening length. The approach is to use Grand Canonical MC moves for salt particle insertion [46] and explore different salt activities and protein concentrations. As a reference point, we performed salt only simulations (bulk solution, $c_s$), then added a coarse grained TLL in the center of a spherical simulation cell. In this so-called cell model description [47–49], the radius of the simulation box now determines the protein concentration.

For different salt activities, $a_s$. Fig. 12 shows the ratio $\gamma_s = c_s/c_p$, where $c_s$ is the salt concentration in the bulk solution (reference point). For all activities, $\gamma_s$ increases with the protein concentration, indicating a gradual deviation from the bulk salt solution. In the SAXS experiments and in the previous MC simulations, the ionic strength is 0.105 M, while $c_p$ ranges from 0.01 mM to 0.3 mM. This region is marked in Fig. 12 (red square) and $\gamma_s$ is here close to unity, suggesting that $c_s \approx c_p$ is a good approximation. However, large deviations occur at higher protein concentration and care must be taken when estimating the salt concentration. The departure from ideality is driven mainly by (i) the protein excluded volume, and (ii) the non-uniform electric potential surrounding the charged proteins. Within the cell model, (i) can be estimated by assuming spherical proteins that exclude a fraction of the total volume:

$$\gamma_s = \frac{c_s}{c_p} = \frac{1}{1 - (R_p/R_{cell})} = e^{l_s^* k_BT}.$$  

Here $R_{cell}$ is the radius of the simulation cell; $R_p = 22.7 \, \text{Å}$ [50] is the radius of the protein; and $l^*$ corresponds to the excluded volume excess chemical potential of transferring salt between the buffer and protein solution. While Eq. 7 is strictly valid for neutral spheres only, Fig. 12 shows good agreement with MC simulations of TLL where all charges have been artificially neutralised. The remaining increase in $\gamma_s$ is due to electrostatics which depends strongly on the protein charge distribution and the valency of salt and counter-ions. While the used cell model captures these effects, it is limited to homogeneous protein samples. Mixtures of e.g. oppositely charged proteins or cellular environments may further influence $\gamma_s$ in unknown ways, and would require a more sophisticated model with at least two proteins.
4. Conclusion

Aiming to understand protein–protein interactions in dilute to semi-concentrated protein solutions, we have studied variants of the enzyme Thermomyces lanuginosus Lipase (TLL) dissolved in aqueous salt solutions. Using a combination of small-angle X-ray scattering and molecular simulations we have simultaneously measured and calculated osmotic second virial coefficients, $B_2$, and solution structure factors, $S(q)$. Via experimental data from dilute protein solutions, we created and tested several coarse grained protein interaction models where short range attraction is (i) accumulated solely on the hydrophobic, active site of TLL or (ii) also smeared over the entire protein surface. We also explored the effect of experimentally known open and closed form of TLL. Our results clearly show that while a given model may well reproduce experimental data ($B_2$) at infinite dilution, it is no guarantee to predict the solution structure at even mildly elevated concentrations. For the present system we attribute this to the fact that the TLL system interacts through a single-patch interaction that promotes dimer formation. Therefore, the generic attraction (between all surface residues, $i_{	ext{II}-m}$) and specific patch attraction (between hydrophobic residues, $i_{	ext{H}-m}$) must be balanced to correctly capture both spatial and angular correlations between the proteins as the concentration increases. This highlights the importance of careful consideration of protein structure when performing a priori predictions of solution stability.

We further explored a two-state conformation model where Monte Carlo swap moves dynamically account for the known open and closed forms of TLL. While this added degree of freedom in principle elevates realism, the intrinsic free energy difference between the two forms is currently unknown and requires further experimental and/or theoretical investigation. For all interaction models, we do observe a transition from closed to open form as the concentration increases. This agrees well with existing experimental knowledge of TLL. Using an open form – with the hydrophobic patch exposed – consistently gives best agreement with measured scattering intensities as a function of protein concentration. Overall the above findings all support the idea of a single patch interaction.

In summary, we investigated variants of TLL as typically found in industrial protein production pipelines. Fixing model parameters to capture the experimental $B_2$ of the wildtype, we explored if variant behaviour could be predicted at dilute conditions. When using a fixed, open form of TLL, the coarse grained interaction model reasonably well reproduces the effect of glycosylation (TLLm) and to some extent also the effect of added a pro-peptide (TLL+ and TLLm+). In all cases, using a closed TLL structure reduces agreement with experiment, again suggesting that dimer formation occurs through interactions between the active site patches, whereby a restricted angular space is a fundamental model requirement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We acknowledge financial support from the Swedish Research Foundation; the European Union through the Innovative Training Network, PIPPI. The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the center for scientific and technical computing at Lund University.

C. Pasquier, Saren Roi Midtgaard, M. Polimeni et al. Journal of Colloid and Interface Science 629 (2023) 794–804

Appendix A

An electronic notebook for running simulations, analysing, and plotting the presented data is available at Zenodo: https://doi.org/10.5281/zenodo.7053143.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jcis.2022.08.054.

References

[1] C.J. Roberts, Non-native protein aggregation kinetics, Biotechnol. Bioeng. 98 (2007) 927–938.
[2] N. Willems, M. Lelimousin, J. Skjold-Jørgensen, A. Svendsen, M.S. Sansom, The H. Nothaft, C.M. Szymanski, Protein glycosylation in bacteria: sweeter than ever, Nat. Rev. Microbiol. 8 (2010) 765–778.
[3] R. Fernandez-Lafuente, Lipase from Thermomyces lanuginosus: Uses and prospects as an industrial biocatalyst, J. Mol. Catal. B: Enzym. 62 (2010) 197–212.
[4] A.W. Sonesson, T.H. Calisien, U.M. Eloffsson, H. Brismar, Imaging the Detergency of Single Cotton Fibers with Confocal Microscopy: the Effect of Surfactants and Lipoysis, J. Surfactants Deterg. 10 (2007) 211–218.
[5] Y. Cai, J. Prat, J.D. Bolds, M.A. Alinsa, A. Svendsen, Spectroscopic studies of the interfacial binding of Humicola lanuginosa lipase, The Analyst 123 (1998) 2229–2233.
[6] M. Martinelle, M. Holminquist, K. Hult, On the interfacial activation of Candida antarctica lipase A and B as compared with Humicola lanuginosa lipase. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 1258 (1995) 272–276.
[7] N. Willems, M. Lelimousin, J. Skjold-Jørgensen, A. Svendsen, M.S. Sansom, The effect of mutations in the lid region of Thermomyces lanuginosus lipase on interactions with triglyceride surfaces: A multi-scale simulation study, Chem. Phys. Lipids 211 (2018) 4–15.
[8] D. Swergun, M. Koch, Small-angle scattering studies of biological macromolecules in solution, Rep. Prog. Phys. 66 (2003) 1735–1782.
[9] A. Stradner, H. Sedgwick, F. Cardinaux, W. Poon, S. Egelhaaf, P. Schureren, Equilibrium cluster formation in concentrated protein solutions and colloids, NATURE 432 (2004) 492–495.
[10] Hura, G.L.; Menon, A.L.; Hammel, M.; Rambo, R.P.; Poole, F.L., II; Tsutakawa, S. E.; Jenney, F.E., Jr.; Classen, S.; Frankel, K.A.; Hopkins, R.C.; Yang, S.-j.; Scott, J. W.; Dillard, B.D.; Adams, M.W.W.; Tainer, J.A. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature Methods 2009, 6, 606–U83.
[11] W. Li, B.A. Persson, M. Lund, J. Bergenholz, M.Z. Osolkova-Konzelmann, Lipase-Induced Association in a Protein System Caused by a Highly Directional Patch Interaction, J. Phys. Chem. B 120 (2016) 8953–8959.
[12] R. Hoiberg-Nielsen, P. Westh, L.K. Skov, L. Arleth, Interrelationship of Steric Stabilization and Self-Crowding of a Glycosylated Protein, Biophys. J. 97 (2009) 1445–1453.
[13] R. Hoiberg-Nielsen, P. Westh, L.K. Skov, L. Arleth, The Effect of Glycosylation on Interparticle Interactions and Dimensions of Native and Denatured Phytase, Biophys. J. 96 (2009) 153–161.
[14] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 192 (1989) 319–326.
[15] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423.
[16] J.B. Hopkins, R.E. Gillilan, S. Skou, BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis, J. Appl. Crystallogr. 50 (2017) 1545–1553.
[17] S. Hansen, BayesApp: a web site for indirect transformation of small-angle scattering data. J. Appl. Crystallogr. 45 (2012) 566–567.
[18] A.M. Brzozowski, H. Savage, C.S. Verma, J.P. Turkenburg, D.M. Lawson, A. Svendsen, S. Patkar, Structural Origins of the Interface Activation in Thermomyces (Humicola) lanuginosa Lipase, Biochemistry 39 (2000) 15071–15082.
[19] Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8.
[20] T.J. Doliński, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667.
C. Pasquier, Søren Roi Midtgaard, M. Polimeni et al.
Journal of Colloid and Interface Science 629 (2023) 794–804