# Estimation of the mass density of biological matter from refractive index measurements

Conrad Möckel,[1, 2, 3] Timon Beck,[1, 2] Sara Kaliman,[1, 2] Shada Abuhattum,[1, 2] Kyoohyun Kim,[1, 2] Julia Kolb,[1, 2, 3] Daniel Wehner,[1, 2] Vasily Zaburdaev,[2, 3] and Jochen Guck[1, 2, 3, *]

[1]Max Planck Institute for the Science of Light, 91058 Erlangen, Germany
[2]Max-Planck-Zentrum für Physik und Medizin, 91058 Erlangen, Germany
[3]Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany
(Dated: December 4, 2023)

The quantification of physical properties of biological matter gives rise to novel ways of understanding functional mechanisms by utilizing models that explicitly depend on physical observables. One of the basic biophysical properties is the mass density (MD), which determines the degree of crowdedness. It impacts the dynamics in sub-cellular compartments and further plays a major role in defining the opto-acoustical properties of cells and tissues. As such, the MD can be connected to the refractive index (RI) via the well known Lorentz-Lorenz relation, which takes into account the polarizability of matter. However, computing the MD based on RI measurements poses a challenge as it requires detailed knowledge of the biochemical composition of the sample. Here we propose a methodology on how to account for *a priori* and *a posteriori* assumptions about the biochemical composition of the sample as well as respective RI measurements. To that aim, we employ the Biot mixing rule of RIs alongside the assumption of volume additivity to find an approximate relation of MD and RI. We use Monte-Carlo simulations as well as Gaussian propagation of uncertainty to obtain approximate analytical solutions for the respective uncertainties of MD and RI. We validate this approach by applying it to a set of well characterized complex mixtures given by *bovine* milk and intralipid emulsion. Further, we employ it to estimate the mass density of trunk tissue of living zebrafish (*Danio rerio*) larvae. Our results enable quantifying changes of mass density estimates based on variations in the *a priori* assumptions. This illustrates the importance of implementing this methodology not only for MD estimations but for many other related biophysical problems, such as mechanical measurements using Brillouin microscopy and transient optical coherence elastography.

Keywords: Mass density, Refractive index, Optical diffraction tomography, Brillouin microscopy, Optical coherence elastography

## I. INTRODUCTION

Quantifying the physical properties of biological matter has become increasingly important over recent decades. By now, the notion that biological function of cells and tissues is affected by their physical phenotype and vice versa has been validated in many experimental studies (see e.g., [1–3]). A fundamental property of matter, including living matter, is the mass density (MD) [4], which is not only associated with buoyancy, crowdedness [5], biomolecular condensation [6] and inherent dynamical processes of the sample of interest [7, 8], but also plays a major role in elastography, particularly Brillouin microscopy [9–12] and transient optical coherence elastography (see e.g., [13, 14]). However, measuring the *in vivo* MD distribution in a direct manner poses a challenge which has not been resolved so far. One way of inferring the *in vivo* MD of a sample is to measure the refractive index (RI) *via* microscopy techniques such as optical diffraction tomography (ODT) [15]. The Lorentz-Lorenz relation then connects the RI with the mass density if the molar refractivity and partial specific volume (PSV) of the dry mass composition as well as the solvent

content are known. This knowledge, however, is not trivially obtainable. A customary assumption regarding biological matter is that the dry mass composition is given by proteins only [15–17] and that the solvent content is then indirectly constrained by the measured RI. While this approximation holds for binary polymer-solutions, it cannot be directly extended to samples with a 'complex dry mass composition'. In the context of cells and tissues, the dry mass composition may be thought of as a mixture of (phase separated) proteins, lipids, sugars, etc. [18]. By employing e.g., mass spectrometry (MS) and/or (stimulated) Raman spectroscopy (SRS), individual components and their respective concentrations in the sample can be identified [19–21]. Additionally, correlative fluorescence information could be employed to segment RI maps acquired by ODT [12, 22]. However, these experimental modalities might not be available or applicable for certain samples, which creates a degree of ignorance about the dry mass composition which should be accounted for in the inference process of obtaining an MD estimate. Another, closely related aspect is the robust estimation of the uncertainty of the MD. Considering the previously mentioned degree of ignorance, these uncertainties are clearly not only of statistical but also of systematic nature. And further, even if universal knowledge about the true distributions of the molar refractivity and PSV was available, in order to estimate

* jochen.guck@mpl.mpg.de

the uncertainty of the MD adequately, the uncertainties of the individual parameters should be propagated.

Here we present a robust methodology for estimating the uncertainties of the MD and the correlative RI by employing Monte-Carlo (MC) simulations. Furthermore, we provide analytical approximations for both the MD and RI distributions in dependence of the dry mass composition and the solvent content, employing Gaussian propagation of uncertainty (GPU).

To this end, we first motivate a simple mixture model to estimate the MD and the RI from two material constants, namely the RI increment $\alpha$ and the PSV $\theta$. We then extend the model towards unimodal distributions of $\alpha$ and $\theta$, for which previously only precise values were assumed.

The distributions of the RI increment and PSV are remarkably narrow, when only considering proteins in the mixture [23], resulting in sharp distributions of RI and MD. However, taking a second type of molecule, such as lipids or sugars, into account adds an additional complexity to the MD estimations since their values of RI increment and PSV differ drastically from those of proteins. Therefore, we derive an effective description of the system based on weighted mixture distributions. This allows for a correlative prediction of the RI and the MD, accounting e.g., for the lipid and water content of the sample as well as for fluctuations in both quantities. We then apply this approach to 1) *bovine* milk, a well characterized mixture of water, proteins and lipids and 2) 20% intralipid emulsion, which mainly consists of water and soybean oil. Comparing the measured values of the MD and RI acquired by pycnometry and Abbe refractometry, respectively, with our theoretical estimates, we find both to be in good agreement with each other.

After demonstrating the applicability of our method on *bovine* milk and intralipid emulsion, we explore the MD distribution of larval zebrafish trunk, compromising major tissue including muscle, spinal cord etc., employing the recent RI and MS measurements of [19]. This purely optical and computational approach shows how the MD can be estimated also in complex *in vivo* specimens, enabling a more profound interpretation of mechanical measurements.

## II. A BINARY MIXTURE MODEL FOR MASS DENSITY ESTIMATIONS

Considering a mixture of some molecule (i.e. solute content) in a solvent, the total mass $m$ and volume $v$ of the solution follow the form

$$
\begin{aligned}
m &= m_1 + m_2, \\
v &= v_1 + m_2 \tilde{\theta},
\end{aligned}
\tag{1}
$$

where index 1 denotes the solvent, index 2 denotes the solute and $\tilde{\theta}$ denotes the so called apparent specific volume of the solute (ASV), which describes the volume per gram of the solute *in solution*. As such, the ASV may be dependent on the mass of the solute $m_2$, since it accounts for solute-solute interactions under constant temperature $T$, pressure $p$ and solvent mass $m_1$. The change of the total volume of the solution $v$ with respect to a change in the mass of the solute is then characterized by the partial specific volume (PSV) *via*

$$
\theta \equiv \left( \frac{\partial v}{\partial m_2} \right)_{T,p,m_1} = \tilde{\theta} + m_2 \left( \frac{\partial \tilde{\theta}}{\partial m_2} \right)_{T,p,m_1}, \tag{2}
$$

as motivated in [24].

For the sake of simplicity, for all the following considerations, we employ the concept of volume additivity, for which it is straightforward to show that $\theta = \tilde{\theta} = 1/\rho_2$, where $\rho_2$ is the solute density (see SI; Eqs. (S3) [25]). By denoting the solute concentration as $c_2 \equiv m_2/v$ and expressing the mass of the solvent as

$$
m_1 = \rho_1 v \left( 1 - c_2 \theta \right), \tag{3}
$$

we obtain an expression for the MD of the mixture as

$$
\rho = \frac{m}{v} = c_2 + \rho_1 (1 - c_2 \theta). \tag{4}
$$

In the next step, we connect the RI of the solution to the solute concentration $c_2$ *via* the phenomenological mixing rule

$$
n = \varphi_1 n_1 + \varphi_2 n_2, \tag{5}
$$

with $\varphi_i = c_i/\rho_i = v_i/v$ being the volume fraction of component $i$ [26]. In the following, we refer to Eq. (5) as Biot equation or mixing rule. Note that in a typical ODT experiment we determine the RI contrast $\delta n \equiv n - n_1$ which has to fulfill a non-negativity constraint (i.e., $\delta n \geq 0$). Evaluating Eq. (5) while assuming volume additivity $[(1 - \varphi_2) = \varphi_1]$, we arrive at

$$
\frac{\delta n}{c_2} = \frac{n_2 - n_1}{\rho_2} = \alpha, \tag{6}
$$

where we identify the customarily designated RI increment $\alpha$. Finally, by replacing $c_2$ in Eq. (4) with the expression given in Eq. (6), we obtain an estimate of the solution MD in dependence of the RI as

$$
\rho = \frac{\delta n}{\alpha} + \rho_1 \left( 1 - \theta \frac{\delta n}{\alpha} \right), \tag{7}
$$

which has been employed in [17], or stated equivalently in [27, 28]. While relations similar to Eq. (7) can be found for different RI mixing rules (see e.g., [29]), the assumption of volume additivity is central for our considerations. Eq. (7) is the basis of all following considerations and will be employed frequently throughout this study.

The first term of Eqs. (7) and (4), respectively, computes the solute concentration and the second term accounts for the MD of the solvent as well as the volume

uptake of the solute content. In other words, the second term can be interpreted as the MD contribution of the solvent content of the sample. In this interpretation, the PSV, $\theta$, scales the solvent content.

Experimentally, the PSV of a solute is determined by measuring the density of the solution $\rho$ in dependence of the solute concentration $c_2$ [24], which can be motivated by solving Eq. (4) for $\theta$ as

$$\theta = \frac{1}{\rho_1}\left(1 - \frac{\rho - \rho_1}{c_2}\right). \tag{8}$$

Hence, any experimentally observed deviation from a constant slope of $\rho(c_2)$ can be directly interpreted as a violation of the volume additivity assumption.

A straightforward and common approach to obtain the MD of a liquid is given by pycnometry, where the mass of a precisely fixed volume is measured. The MD is then simply given by the mass to volume ratio. Further methods to measure MD are reviewed in [24] and references therein. Similarly, the RI increment is customarily determined by measuring the RI of the solution in dependence of the solute concentration (see Eq. (6)), e.g., by employing an Abbe refractometer.

In the context of biological matter, one may think of the 'solute' as a complex composition of many constituents, which makes the experimental determination of both, PSV and RI increment for all components and their combinations practically impossible; the human proteome alone consists of $\sim 10^4$ proteins [18, 30]. In order to resolve this problem, at least partially, not accounting for volume inadditivities, a method of determining the correlative distributions $\mathcal{P}(\alpha, \theta)$ of the proteomes of different organisms was introduced and experimentally validated for two amino acid sequences in [23]. The authors computed the weight averages of the residue refractivity per gram and according PSV $\theta$ of proteins based on their respective amino acid sequence. The refractivity per gram $R_i$ of a molecule $i$ is proportional to its polarizability $\widehat{\alpha}$ and molar mass $M$ as $R_i \sim \widehat{\alpha}_i/M_i$. Further, the refractivity per gram can be connected to the RI and the MD of the solute via the Lorentz-Lorenz (or Clausius–Mossotti) relation (see e.g., [31]) as

$$R_i = \frac{1}{\rho_i}\frac{n_i^2 - 1}{n_i^2 + 2}, \tag{9}$$

which may be solved for the RI of the molecule $n_i$, to obtain

$$n_i = \sqrt{\frac{2R_i + \theta_i}{\theta_i - R_i}}. \tag{10}$$

Considering that the PSV of a binary solution is given by the mass average of the PSVs of the solvent and solute as

$$\frac{1}{\rho} = \frac{m_1}{m}\frac{1}{\rho_1} + \frac{m_2}{m}\frac{1}{\rho_2}, \tag{11}$$

consequently, one could assume that the refraction per gram of the solution is given by the mass average of the constituents as well

$$R = \frac{m_1}{m}R_1 + \frac{m_2}{m}R_2, \tag{12}$$

as it carries the same unit [ml/g] [32]. By inserting Eq. (9) into Eq. (12), we obtain the customarily denoted Lorentz-Lorenz mixing rule of RIs

$$\frac{n^2 - 1}{n^2 + 2} = \varphi_1\frac{n_1^2 - 1}{n_2^2 + 2} + \varphi_2\frac{n_2^2 - 1}{n_2^2 + 2}. \tag{13}$$

We note that Eq. (13) could be employed to derive an expression of the MD in dependence of the RI, similar to Eq. (7), using the Biot Eq. (5). Furthermore, we want to stress that neither the Lorentz-Lorenz mixing rule of RIs (Eq. (13)), nor the Biot mixing rule of RIs (Eq. (5)) do explicitly require volume additivity (see [26, 33] for further details).

However, assuming dilute solutions ($n \to n_1$) as well as volume additivity, the authors of [23] computed the RI increment for each protein, denoted by the index p, as derived in [26], as

$$\alpha_{\mathrm{p}}^{\mathrm{W}} \simeq \frac{3}{2}\theta_{\mathrm{p}}n_1\frac{n_{\mathrm{p}}^2 - n_1^2}{n_{\mathrm{p}}^2 + 2n_1^2}, \tag{14}$$

by employing the Wiener mixing rule of RIs

$$\frac{n^2 - n_1^2}{n^2 + 2n_1^2} = \varphi_2\frac{n_{\mathrm{p}}^2 - n_1^2}{n_{\mathrm{p}}^2 + 2n_1^2}, \tag{15}$$

and Eq. (10) to compute the RI of each protein $n_{\mathrm{p}}$ from the mass averaged refractions per gram and PSVs of the respective amino acid sequences, as described earlier. Repeating this procedure for all proteins presumed to be abundant in the different organisms under study, they obtained the bivariate distribution of RI increment and PSV. Zhao et al. [23] then fitted normal distributions to the univariate histograms to obtain the means and standard deviations for different organisms, of which we depicted two in Tab. I. We repeated the computations presented in [23] (see SI [25]) for the updated human and zebrafish proteomes obtained from [30]. Besides computing the RI increment from the Wiener relation for dilute solutions (Eq. (14)), we also employed the Biot Eq. (5) in combination with the volume additivity assumption to obtain an expression of the RI increment as given in Eq. (6)

$$\alpha_{\mathrm{p}}^{\mathrm{B}} = \theta_{\mathrm{p}}\left(n_{\mathrm{p}} - n_1\right). \tag{16}$$

Furthermore, different to [23], we employed the consensus averages for the amino acid residue molecular volumes of [34]. The full list of parameters employed here is given in Tab. SII [25]. The resulting bivariate distributions $\mathcal{P}(\alpha, \theta)$ are given in Fig. 1. The corresponding mean values and standard deviations of the fits of the univariate histograms with a normal distribution are given in

4

Tab. I. Evidently, the PSV and RI increment values obtained here coincide very well, with the values from [23]. However, since our values were computed using the consensus values of the molar residue volumes of the amino acids of [34] instead of the ones of Cohn and Edsall [35], they are systematically lower. Furthermore, the RI increments obtained using Eq. (16) are systematically higher than the ones obtained from the Wiener equation for dilute solutions, given in Eq. (14).

While it is not clear why the calculations proposed in Zhao *et al.* [23] result in distributions $\mathcal{P}(\alpha, \theta)$ that resemble uncorrelated bivariate normal distributions for the proteomes of different species, they facilitate the idea of taking the whole distribution $\mathcal{P}(\alpha, \theta)$ into account when estimating the MD *via* Eq. (7). Consistently, in order to obtain a reliable estimate of the MD distribution, a precise characterization of the solute composition is needed.

To our knowledge, the justification for or against the assumption of volume additivity in complex biological matter is yet to be given (experimentally). However, we may interpret Eq. (2) in first order approximation as

$$\theta \approx \tilde{\theta} \pm \vartheta, \qquad (17)$$

where any potential deviation from the volume additivity $\vartheta$ could be treated as an additional factor that will also contribute to the uncertainty of the RI and MD estimates.

In the following, we construct and employ a theoretical framework in which the information of the solute composition is incorporated into the prediction of the MD. For that purpose, we employ the Biot Eq. (5) for the majority of our further considerations and denote $\alpha = \alpha^{\mathrm{B}}$.

## III. EXTENSION OF THE BINARY MIXTURE MODEL

As motivated earlier, when dealing with biological matter, the complexity of the solute composition should be taken into account in order to obtain reliable estimates of the MD. To that aim, we first extend the expression of the MD of the binary mixture (Eq. (7)) to the case of different solute constituents, e.g., proteins of the human proteome, lipids and sugars, being dissolved in a solvent with corresponding RI $n_1$ and MD $\rho_1$.

We describe this problem by discretizing the total sample volume into $N_{\mathrm{v}}$ voxels with volumes $v_{\mathrm{v}}$

$$v = N_{\mathrm{v}} v_{\mathrm{v}}. \qquad (18)$$

Furthermore, we discretize the voxels into $N_0$ 'voxelinos' with volumes $v_0$ as

$$v_{\mathrm{v}} = N_0 v_0. \qquad (19)$$

The motivation to divide a voxel into $N_0$ voxelinos is to obtain small, yet macroscopic, volume fractions with a constant volume $v_0$ that contain one and only one solution constituent. Hence, each voxelino can be either a
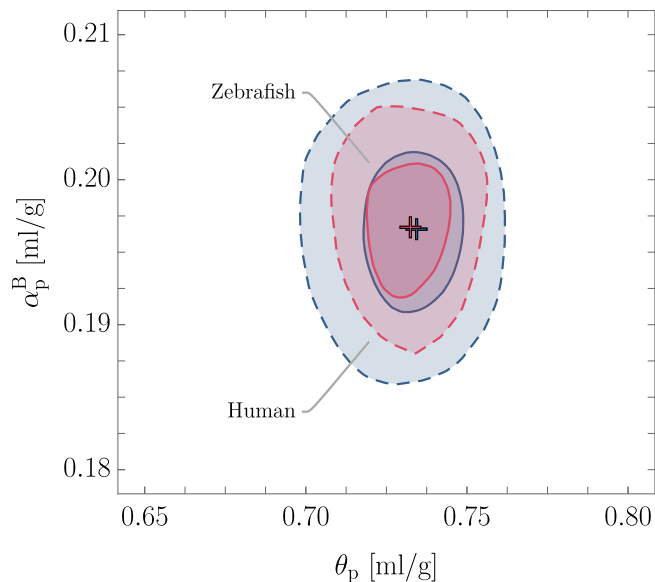


FIG. 1: $1\sigma$ (solid) and $2\sigma$ (dashed) likelihood contours of the correlated distributions of RI increment $\alpha_{\mathrm{p}}^{\mathrm{B}}$ in ml/g and PSV $\theta_{\mathrm{p}}$ in ml/g, for the human proteome; 82127 proteins (blue), and the zebrafish proteome; 46517 proteins (red).

solvent voxelino, or a solute voxelino and is inherently characterized by its respective partial specific volume $\theta_i$ and refraction per gram $R_i$, i.e., its MD and RI. The number of solvent voxelinos in the voxel, $N_1$, is given by $N_1 = N_0 - N_{\mathrm{s}}$, where $N_{\mathrm{s}}$ is the number of solute voxelinos. Accordingly, the solvent volume fraction of a voxel is given by

$$\varphi_1 \equiv 1 - \frac{N_{\mathrm{s}}}{N_0}. \qquad (20)$$

Accounting for multiple types of solute molecules, e.g., proteins, lipids, sugars, etc., being present in the solution, we choose the PSVs and refractions per gram of the solute voxelinos $\theta_{i+1}$ and $R_{i+1}$ to be random values, drawn from a weighted mixture distribution

$$\mathcal{P}_{\mathrm{mix}}(R, \theta) = \sum_j x_j \mathcal{P}_j(R, \theta), \qquad (21)$$

where $j$ denotes the solute constituents. Further, the $\mathcal{P}_j(R, \theta)$ represent the bivariate probability distributions of the refraction per gram and PSV of the respective constituents and the associated weights $x_j$ are given by the relative volume fractions as

$$x_j = \frac{v_j}{v_{\mathrm{s}}} = \frac{N_j}{N_{\mathrm{s}}}, \qquad (22)$$

where $v_{\mathrm{s}}$ denotes the total solute volume. We note that Eq. (21) should be seen as a way of denoting that $N_j$ out of $N_{\mathrm{s}}$ solute voxelinos of constituent $j$ are present

TABLE I: Mean and standard deviations of the RI increments $\alpha_{\mathrm{p}}^{i}$ and partial specific volumes $\theta_{\mathrm{p}}$ distributions based on amino acid sequences of the proteome of the human and the zebrafish and the trunk tissue of the larval zebrafish at 96 hpf. The calculation of $\alpha_{\mathrm{p}}^{\mathrm{B}}$ employed the Biot equation given in Eq. (5), while $\alpha_{\mathrm{p}}^{\mathrm{W}}$ was derived from the dilute limit of the Wiener mixing rule of refractive indices (Eq. (14)).

| | $\overline{\alpha}_{\mathrm{p}}^{\mathrm{W}}$ in ml/g | $\Delta\alpha_{\mathrm{p}}^{\mathrm{W}}$ in ml/g | $\overline{\alpha}_{\mathrm{p}}^{\mathrm{B}}$ in ml/g | $\Delta\alpha_{\mathrm{p}}^{\mathrm{B}}$ in ml/g | $\overline{\theta}_{\mathrm{p}}$ in ml/g | $\Delta\theta_{\mathrm{p}}$ in ml/g |
|---|---|---|---|---|---|---|
| Human | | | | | | |
| Proteome[a] | 0.1899 | 0.0030 | N/A | N/A | 0.735 | 0.010 |
| Proteome[b] | 0.188 | 0.004 | 0.196 | 0.004 | 0.734 | 0.012 |
| Zebrafish | | | | | | |
| Proteome[a] | 0.1904 | 0.0030 | N/A | N/A | 0.735 | 0.010 |
| Proteome[b] | 0.1887 | 0.0031 | 0.1968 | 0.0034 | 0.732 | 0.010 |
| trunk tissue[b] | 0.1883 | 0.0029 | 0.1963 | 0.0032 | 0.734 | 0.009 |

[a] [23]
[b] this work

in a voxel. Consequently, $N_j$ solute voxelinos have the RI $n_j$ and MD $\rho_j$ and we have that $N_{\mathrm{s}} = \sum_j N_j$, i.e., $\sum_j x_j = 1$.

Employing Eqs. (4) and (5), we readily obtain the relation between the MD $\rho$ and the RI $n$ of one voxel

$$
\begin{aligned}
\rho &= \sum_{i=1}^{N_{\mathrm{s}}} c_{i+1} + \rho_1 \left( 1 - \sum_{i=1}^{N_{\mathrm{s}}} \theta_{i+1} c_{i+1} \right) \\
&\equiv \frac{\delta n}{\alpha_{\mathrm{eff}}} + \rho_1 \left( 1 - \theta_{\mathrm{eff}} \frac{\delta n}{\alpha_{\mathrm{eff}}} \right),
\end{aligned}
\tag{23}
$$

where we defined the effective RI increment $\alpha_{\mathrm{eff}}$ and PSV $\theta_{\mathrm{eff}}$. The RI of the solution $n$ is then given by the Biot Eq. (5), where the RIs of the individual solute voxelinos $n_{i+1}$ could be known directly from RI measurements, or may be computed by employing the Lorentz-Lorenz relation, given in Eq. (10).

It can be shown (see SI; Eqs. (S17) and (S18) [25]) that the effective parameters may be expressed by the mass averages of the respective parameters of the solute voxelinos as

$$
\begin{aligned}
\theta_{\mathrm{eff}} &= \sum_{i=1}^{N_{\mathrm{s}}} \left( \frac{x_{i+1}}{\theta_{i+1}} \right)^{-1} = \sum_{i=1}^{N_{\mathrm{s}}} y_{i+1} \theta_{i+1}, \text{ and} \\
\alpha_{\mathrm{eff}} &= \theta_{\mathrm{eff}} (n_{\mathrm{s}} - n_1) = \sum_{i=1}^{N_{\mathrm{s}}} y_{i+1} \alpha_{i+1},
\end{aligned}
\tag{24}
$$

where we denoted the solute concentration and RI by $c_{\mathrm{s}}$ and $n_{\mathrm{s}}$, respectively, and the relative mass fraction of a solute voxelino by

$$
y_i = \frac{m_i}{m_{\mathrm{s}}},
\tag{25}
$$

with $m_{\mathrm{s}}$ being the total solute mass. Hence, the effective parameters absorb the 'complexity' of the mixture, while the functional relationship of Eq. (4) is obeyed. With this at our disposal, we are able to compute the MD $\rho$ and

the corresponding RI $n$ of a voxel, given a distribution $\mathcal{P}_{\mathrm{mix}}(R, \theta)$ and a solvent volume fraction of the voxels $\varphi_1$, i.e., the ratio $N_{\mathrm{s}}/N_0$. Repeating this procedure for $N_{\mathrm{v}}$ voxels provides a map of RI values in resemblance of an ODT measurement, alongside the corresponding MDs.

As the framework introduced above is strongly dependent on *a priori* assumptions of the individual model parameters ($\varphi_1$, $x_j$, $\alpha_j$, $\theta_j$) for different complex mixtures, in the next step, we shall discuss the impact of these assumptions on the uncertainties of the RI and MD predictions.

## IV. UNCERTAINTY QUANTIFICATION OF THE EXTENDED MODEL

Given access to experimental data, i.e., ODT tomograms of a sample of interest, we may not only compare the mean values of the measured and predicted RI distributions, but also their widths. This in turn enables a more reliable estimate of the MD distribution. Hence, in the next step we study the dependence of the uncertainty of the MD estimate $\rho$, defined in Eq. (23), and the RI $n$ on the sample properties, namely, the solvent volume fraction $\varphi_1$, the effective RI increment $\alpha_{\mathrm{eff}}$, the effective partial specific volume $\theta_{\mathrm{eff}}$ as well as their respective associated uncertainties.

To that aim, we investigate the impact of the presence of a second type of macro molecule in the mixture of proteins and water. Since lipids make up for about 13% of the solute mass fraction in mammalian cells [18], and typically exhibit an MD lower than water, they merit detailed scrutiny. In the following, we assume that the lipids are present in the form of lipid droplets and form an emulsion in the water+protein phase. For the sake of simplicity, we further assume that the lipid droplets consist only of the neutral lipid triolein (TO), neglecting sterol esters, triacylglycerols and phospholipids [36–38]. The corresponding values of the RI, refraction per

gram, PSV and RI increment of the two types of solute molecules under investigation are given in Tab. SI [25].

Throughout this study we assume that both RI and MD of the solvent are precisely known. We further point out that we implement the values given in Tab. SI [25] in our calculations as follows: if a quantity is stated as mean ± standard deviation, we account for it as normally distributed with respective mean and standard deviation. For the cases where we could not estimate an uncertainty, we assume the quantity to be delta distributed.

### A.  Effective RI increment and PSV

Examining the definitions of $\alpha_{\mathrm{eff}}$ and $\theta_{\mathrm{eff}}$, given in Eqs. (24), we observe a dependence of both quantities on the number of solute voxelinos per voxel $N_{\mathrm{s}}$, given as the upper limit of the sum. To obtain an intuition about the implications on the respective uncertainties, we first consider the case of proteins dissolved in water, which may be approximated by uncorrelated normal distributions $\mathcal{N}(\mu, \sigma)$ of the RI increment and PSV, as shown in Fig. 1

$$
\begin{aligned}
\alpha^{\mathrm{N}} &\sim \mathcal{N}\left(\overline{\alpha}, \Delta\alpha\right), \\
\theta^{\mathrm{N}} &\sim \mathcal{N}\left(\overline{\theta}, \Delta\theta\right),
\end{aligned}
\tag{26}
$$

with respective mean values and standard deviations (see e.g., Tab. I). Applying GPU to Eqs. (24) we have

$$
\begin{aligned}
\Delta\theta_{\mathrm{eff}} &= \frac{\Delta\theta}{\sqrt{N_{\mathrm{s}}}}, \text{ and} \\
\Delta\alpha_{\mathrm{eff}} &= \frac{\Delta\alpha}{\sqrt{N_{\mathrm{s}}}},
\end{aligned}
\tag{27}
$$

where we used that the relative mass fraction of each solute voxelino is given by $y_{i+1} = 1/N_{\mathrm{s}}$. The result of Eq. (27) is in concordance with the central limit theorem (CLT). In other words, the standard deviation of the effective distributions corresponds to the 'standard error of the mean' of the initial distributions, given that the voxel contains $N_{\mathrm{s}}$ protein voxelinos. Thus, considering that in biological matter the number of voxelinos per voxel is typically larger than $\sim 10^{8}$ for experimentally accessible voxel sizes in the order of $1\,\mu\mathrm{m}^{3}$, the deviations of the effective RI increment and PSV, for the case of proteins dissolved in water, are negligibly small.

Next, we study the impact of lipids in the protein+water mixture. In this scenario, Eq. (21) takes the form

$$
\begin{aligned}
\mathcal{P}_{\mathrm{p+lip}}(R, \theta) &= (1 - x_{\mathrm{lip}})\mathcal{P}_{\mathrm{p}}(R, \theta) \\
&\quad + x_{\mathrm{lip}}\mathcal{P}_{\mathrm{lip}}(R, \theta).
\end{aligned}
\tag{28}
$$

Assuming that $\mathcal{P}_{\mathrm{p}}$ and $\mathcal{P}_{\mathrm{lip}}$ follow uncorrelated bivariate normal distributions, using Eqs. (27), we find that the distributions of the effective PSV and RI increment follow normal distributions as

$$
\begin{aligned}
\theta_{\mathrm{eff}} &\sim \mathcal{N}\left(\overline{\theta}_{\mathrm{eff}}, \Delta\theta_{\mathrm{eff}}^{0}/\sqrt{N_{\mathrm{s}}}\right), \text{ and} \\
\alpha_{\mathrm{eff}} &\sim \mathcal{N}\left(\overline{\alpha}_{\mathrm{eff}}, \Delta\alpha_{\mathrm{eff}}^{0}/\sqrt{N_{\mathrm{s}}}\right),
\end{aligned}
\tag{29}
$$

where the respective mean values of effective PSV and RI increment are given in Eqs. (24) and the standard deviations $\Delta\theta_{\mathrm{eff}}^{0}$, $\Delta\alpha_{\mathrm{eff}}^{0}$ follow the standard deviation of a mixture distribution, provided in the supplemental information (Eq. (S14) [25]).

With this, we further examine the implications of deviations of the relative lipid volume fraction $\Delta x_{\mathrm{lip}}$ from voxel to voxel. Such deviations may be interpreted as a form of inhomogeneity of the system, which have been experimentally quantified in cells and tissues by SRS measurements [20]. Employing GPU we find the following analytical expression of the deviation of the mean effective PSV and RI increment with respect to $\Delta x_{\mathrm{lip}}$ as

$$
\begin{aligned}
\Delta\theta_{\mathrm{eff}} &= \sqrt{\left(\frac{\Delta\theta_{\mathrm{eff}}^{0}}{\sqrt{N_{\mathrm{s}}}}\right)^{2} + \left(\frac{\partial\overline{\theta}_{\mathrm{eff}}}{\partial\overline{x}_{\mathrm{lip}}}\Delta x_{\mathrm{lip}}\right)^{2}}, \text{ and} \\
\Delta\alpha_{\mathrm{eff}} &= \sqrt{\left(\frac{\Delta\alpha_{\mathrm{eff}}^{0}}{\sqrt{N_{\mathrm{s}}}}\right)^{2} + \left(\frac{\partial\overline{\alpha}_{\mathrm{eff}}}{\partial\overline{x}_{\mathrm{lip}}}\Delta x_{\mathrm{lip}}\right)^{2}}.
\end{aligned}
\tag{30}
$$

The partial derivatives in Eqs. (30) are given in the supplemental information (Eqs. (S19) and (S20) [25]). A visual depiction of Eqs. (30) in dependence of the number of voxelinos per voxel $N_{0} = N_{\mathrm{s}}/(1 - \varphi_{1})$ and the corresponding results of MC simulations for certain parameter configurations is shown in Fig. 2(a).

As becomes apparent, the analytical solution, employing GPU, is in concordance with the simulated values. However, we note that this is due to the assumption of normal distributions for the individual effective RI increments and PSVs. For non-normal distributions, GPU might not be applicable.

By assuming the mixture distribution, given by Eq. (28), consequently, the deviations $\Delta\alpha_{\mathrm{eff}}$ and $\Delta\theta_{\mathrm{eff}}$ are maximized for $\overline{x}_{\mathrm{lip}} = 0.5$ (see SI [25]). Secondly, while for non-fluctuating $x_{\mathrm{lip}}$, i.e., $\Delta x_{\mathrm{lip}} = 0$, the respective deviations of the effective RI increment and the PSV exhibit the $1/\sqrt{N_{0}}$ proportionality, for $\Delta x_{\mathrm{lip}} > 0$, we obtain non-vanishing deviations for a large number of voxelinos per voxel $N_{0}$.

Throughout this study we assume that all volume fractions, in particular $x_{\mathrm{lip}}$, follow a normal distribution, truncated in the domain $[0, 1]$, since values outside of this interval are non-physical under the assumption of volume additivity. For a normal distribution with mean $\mu$ and standard deviation $\sigma$, the corresponding truncated distribution is denoted by $\mathcal{T}(\mu, \sigma)$. The probability density function (PDF) of $\mathcal{T}$ is defined as

$$
t(y) \equiv \begin{cases} \frac{f(y)}{F(1) - F(0)} & 0 \le y \le 1 \\ 0 & \text{else} \end{cases},
\tag{31}
$$

where $f(y)$ and $F(y)$ denote the PDF and cumulative distribution function (CDF), respectively, of said normal distribution $\mathcal{N}(\mu, \sigma)$. We note that another choice of distribution could be given by the beta distribution $\mathcal{B}(a, b)$ with shape parameters $a$ and $b$, which is inherently bounded between 0 and 1. However, the interpretation of the shape parameters in this context is not as straightforward as the truncated normal distribution. Alternatively, a uniform distribution with a certain domain $\left[x_{\text{lip}}^{\min}, x_{\text{lip}}^{\max}\right]$ could be employed. For more than two types of macro molecules, the Dirichlet distribution could be of use. However, we expect a similar qualitative behaviour for all mentioned distributions.

### B. Solvent volume fraction

Next, we study the uncertainty associated to the solvent volume fraction $\varphi_1$, defined in Eq. (20), in dependence of the number of solute voxelinos $N_s$ per voxel. By employing GPU we find

$$\Delta\varphi_1 = (1 - \overline{\varphi}_1)\frac{\Delta N_s}{\overline{N}_s}, \qquad (32)$$

where we used the presumption that the number of voxelinos per voxel $N_0$ is constant. From Eq. (32) we obtain that a change in the solvent content from voxel to voxel is directly proportional to a relative change in the number of solute voxelinos per voxel $\Delta N_s/\overline{N}_s$. However, it is not clear whether $N_s$ follows a statistical distribution, and if so, how this distribution would be governed by active/passive processes in biological matter. An ingenuous guess is given by the equilibrium assumption that the number of solute voxelinos per voxel $N_s$ is binomially distributed as

$$N_s \sim \text{Bin}\left(N_0, (1 - \overline{\varphi}_1)\right), \qquad (33)$$

which may be interpreted as finding $N_s$ out of $N_0$ voxelinos in a voxel with a probability of $1 - \overline{\varphi}_1$. The according standard deviation and mean value of $N_s$ is then given by

$$\begin{aligned} \Delta N_s^{\text{Bin}} &= \sqrt{\overline{N}_s^{\text{Bin}}\overline{\varphi}_1}, \text{ and} \\ \overline{N}_s^{\text{Bin}} &= N_0(1 - \overline{\varphi}_1), \end{aligned} \qquad (34)$$

respectively, which is consistent with Eq. (20). Besides a statistical argument, we may also compute a change in the number of solute voxelinos from voxel to voxel by considering

$$N_s = \frac{v_s}{v_0} = \frac{m_s\theta_{\text{eff}}}{v_0}, \qquad (35)$$

from which we have

$$\frac{\Delta N_s}{\overline{N}_s} = \sqrt{\left(\frac{\Delta m_s}{\overline{m}_s}\right)^2 + \left(\frac{\Delta\theta_{\text{eff}}}{\overline{\theta}_{\text{eff}}}\right)^2 + \left(\frac{\Delta N_s^{\text{Bin}}}{\overline{N}_s^{\text{Bin}}}\right)^2}, \qquad (36)$$

using the presumption of a constant voxelino volume $\Delta v_0 = 0$. Herewith, Eq. (32) may be written as

$$\Delta\varphi_1 = (1 - \overline{\varphi}_1)\sqrt{\left(\Delta\varphi_1^0\right)^2 + \left(\Delta\varphi_1^\infty\right)^2}, \qquad (37)$$

with

$$\begin{aligned} \Delta\varphi_1^0 &= \sqrt{\frac{1}{N_0(1 - \overline{\varphi}_1)}\left(\overline{\varphi}_1 + \left(\frac{\Delta\theta_{\text{eff}}^0}{\overline{\theta}_{\text{eff}}}\right)^2\right)}, \\ \Delta\varphi_1^\infty &= \sqrt{\left(\frac{\Delta m_s}{\overline{m}_s}\right)^2 + \left(\frac{\partial\overline{\theta}_{\text{eff}}}{\partial\overline{x}_{\text{lip}}}\frac{\Delta x_{\text{lip}}}{\overline{\theta}_{\text{eff}}}\right)^2}. \end{aligned} \qquad (38)$$

This indicates that, similarly to the effective RI increment and PSV, given by Eqs. (27), the deviation of the solvent volume fraction $\Delta\varphi_1$ has two components; for once $\Delta\varphi_1^0$, which shows the $1/\sqrt{N_0}$ dependence, following the CLT. Secondly, $\Delta\varphi_1^\infty$, which connects fluctuations in the solvent volume fraction to fluctuations in the solute mass and/or fluctuations in the solute composition. Hence $\Delta\varphi_1^\infty$ may be interpreted as quantification of the degree of inhomogeneity of the sample in the solute and its composition. Consequently, for a large number of voxelinos per voxel $N_0$, these inhomogeneities become the dominant contribution to the deviation of the solvent volume fraction.

We note that the presumption of a constant voxelino volume $\Delta v_0 = 0$ and a constant number of voxelinos per voxel $\Delta N_0 = 0$ from voxel to voxel is necessitated by the experimental boundary condition, that all voxels have the same volume, i.e., $\Delta v_v = 0$.

### C. Refractive index

Considering the previous derivations of the uncertainties of the effective PSV and RI increment, as well as the water volume fraction, we now examine the refractive index of the solution. For that purpose, we rewrite Eq. (5) as

$$n = (1 - \varphi_1)\frac{\alpha_{\text{eff}}}{\theta_{\text{eff}}} + n_1. \qquad (39)$$

With this at hand, we readily obtain an estimate of the uncertainty of $n$ by employing GPU (neglecting potential correlations) as

$$\Delta n = \sqrt{\sum_i\left(\frac{\partial n}{\partial\beta_i}\Delta\beta_i\right)^2}, \qquad (40)$$

where the sum is taken over all $\beta = \{\alpha_{\text{eff}}, \theta_{\text{eff}}, \varphi_1\}$ and the respective deviations $\Delta\beta_i$ are given in Eqs. (30) and (37). The graphical representation of Eq. (40) in dependence of $N_0$ is given in Fig. 2(a).

As a consequence we have that for a vanishing relative deviation of the solute $\Delta m_s/\overline{m}_s$, the $1/\sqrt{N_0}$ scaling

behaviour of the effective RI increment and PSV determines the scaling behaviour of the deviations in the RI $\Delta n$. However for $\Delta m_\mathrm{s}/\overline{m}_\mathrm{s} > 0$ we obtain a constant deviation of the RI for large $N_0$, consistent with the broad RI distributions obtained by ODT measurements of different cells and tissues [12, 15, 19, 39], and fluctuations in the water volume fraction measured by SRS [20]. Furthermore, while the dependence of $\Delta n$ on the mean relative lipid volume fraction $\overline{x}_\mathrm{lip}$ is determined by the deviations of the effective RI increment and PSV, we observe a vanishing impact of the deviation of the relative lipid volume fraction $\Delta x_\mathrm{lip}$. This fact is due to the numerically small differences of the mean refractions per gram of the particular choice of proteins and the lipid $\left| \overline{R}_\mathrm{lip} - \overline{R}_\mathrm{p} \right| \approx 0.05$ in combination with larger differences of the mean PSVs $\left| \overline{\theta}_\mathrm{lip} - \overline{\theta}_\mathrm{p} \right| \approx 0.35$ (see SI; Fig. S4 [25]).

### D. Mass density

Finally, we investigate the uncertainty associated with the MD. For convenience, we may express Eq. (23) as

$$\rho = \frac{(1 - \varphi_1)}{\theta_\mathrm{eff}} + \varphi_1 \rho_1, \tag{41}$$

to obtain the corresponding deviation, employing GPU, as

$$\Delta\rho = \sqrt{\left( \frac{\overline{\theta}_\mathrm{eff}\rho_1 - 1}{\overline{\theta}_\mathrm{eff}} \Delta\varphi_1 \right)^2 + \left( \frac{\overline{\varphi}_1 - 1}{\overline{\theta}_\mathrm{eff}^2} \Delta\theta_\mathrm{eff} \right)^2}, \tag{42}$$

displayed in Fig. 2(a). As discussed for $\Delta n$, the magnitude of $\Delta\rho$ is impacted by the mean relative lipid fraction $\overline{x}_\mathrm{lip}$ as a consequence of the mixture distribution, given in Eq. (21). Furthermore, we have a non-vanishing deviation of the MD for $\Delta m_\mathrm{s}/\overline{m}_\mathrm{s} > 0$ and large $N_0$. To no surprise, we observe a remarkable impact of deviation of the relative lipid fraction $\Delta x_\mathrm{lip}$ on $\Delta\rho$ due to the strong scaling with the deviation of the effective PSV $\Delta\theta_\mathrm{eff}$.

### E. Correlation of refractive index and mass density

Having studied the deviations associated with RI and MD, we next sought to illuminate the correlation between the distributions of the RI contrast $\delta n$, and the MD $\rho$, denoted by $\mathcal{P}(\rho, \delta n)$, in dependence of the model parameters introduced earlier. To that aim, we performed MC simulations of Eq. (23) for the case of human proteins and triolein in water for a range of different mean relative lipid and water volume fractions, as shown in Fig. 2(b).

Besides the intuitive behaviour of $\rho(\delta n)$ for the cases of $\overline{x}_\mathrm{lip} = 0$ (MD increases with increasing refractive index) and $\overline{x}_\mathrm{lip} = 1$ (MD decreases with increasing refractive index), for $\overline{x}_\mathrm{lip} = 0.8$ the MD is roughly constant for all refractive index values. This feature is quite remarkable since it demonstrates that for certain solute compositions, the MD is decoupled from the refractive index

for all possible water volume fractions. Furthermore, as motivated earlier, and shown in Fig. 2(c), measuring a refractive index distribution, e.g., *via* ODT, may correspond to a range of different water and relative lipid volume fractions, resulting in drastically different estimates on the distribution of the MD from case to case. This in turn strongly motivates the necessity for detailed knowledge about not only the solute composition, but also the solvent content of the sample.

In the light of the theoretical implications delineated above, in the next step we want to examine the predictive capabilities of the model for a set of physiological complex mixtures that are well characterized in terms of their solute composition.

## V. EXPERIMENTAL VALIDATION AND APPLICATION

In the following, we scrutinize the applicability of previous findings to a set of well characterized, physiological substances, i.e., *bovine* skim-milk powder (SM, Millipore 70166) and 20% intralipid emulsion (IL, Sigma-Aldrich I141). To that aim, we measured the solute concentration-dependent RI and MD with an Abbe refractometer and a pycnometer, respectively. Both samples are particularly intriguing since they should exhibit different $\rho(\delta n)$ dependencies; SM mainly consists of lactose and milk proteins, while IL is a stabilized emulsion of soybean oil (see Fig. 2(b) for a reference).

According to the chemical certificate of analysis, provided by the manufacturers, the SM exhibits $y_1^\mathrm{SM} = 4\%$ w/w and the IL exhibits $y_1^\mathrm{IL} = 76\%$ w/w of water. Hence, we computed the respective solute concentrations as

$$c_\mathrm{s}^\mathrm{k} = \frac{m_\mathrm{k}(1 - y_1^\mathrm{k})}{\left( m_\mathrm{w} + y_1^\mathrm{k}m_\mathrm{k} \right)/\rho_1}, \tag{43}$$

where $m_\mathrm{k}$ denotes the mass of the sample (SM or IL) and $m_\mathrm{w}$ is the mass of water added to the sample. The results of the measurements are shown in Figs. 3(a) and 3(b), where each point represents $N = 5$ technical repetitions.

Using Eqs. (4) and (6) we fitted the data *via* a $\chi^2$-minimization approach to obtain experimental values of $\theta_\mathrm{eff}$ and $\alpha_\mathrm{eff}$ with according uncertainties, respectively, for both, SM, and IL. Examining the fitting residuals, evidently, the RI contrast and MD scale linearly with the solute concentration, justifying the assumption of volume additivity.

We then employed the information about the chemical composition of the respective substances, provided by the manufacturers (see Tab. SI [25]), to compute the correlative RI contrast and MD according to Eq. (23) and the Biot RI mixing rule (Eq. (5)), employing MC sampling. A graphical representation of the respective RI increments and PSVs of all substances considered, as well as the experimentally determined and predicted values of SM and IL is given in Fig. 3(c). The numerical
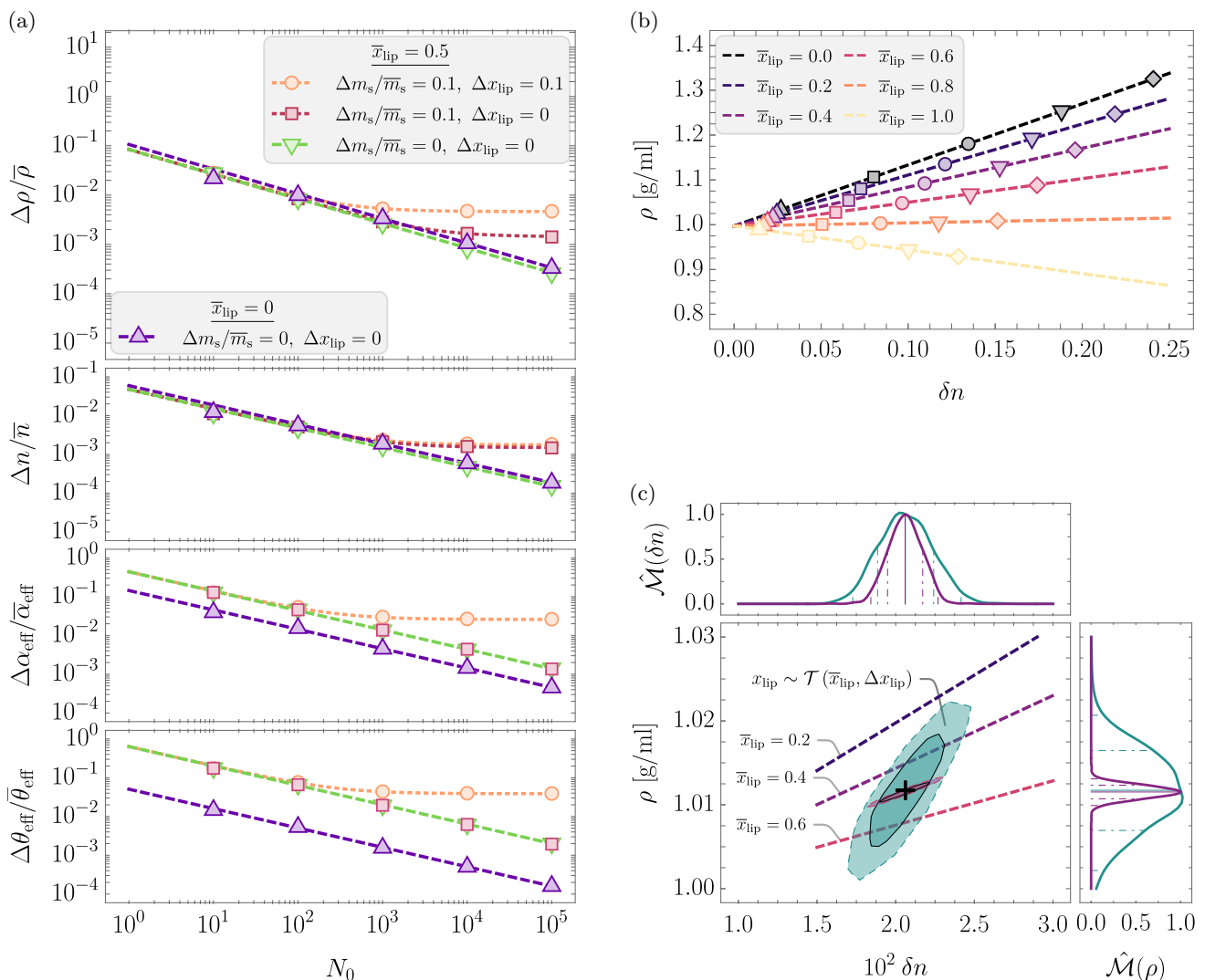
FIG. 2: Results of MC simulations and according analytical solutions for the mixture of human proteins and the neutral lipid triolein in water; (a): Relative deviations of the MD $\rho$, the RI $n$, the effective RI increment $\alpha_{\mathrm{eff}}$ and PSV $\theta_{\mathrm{eff}}$ in dependence of the number of molecules per voxel $N_0$ obtained from MC simulations (symbols) and analytical solutions (dashed lines) for different mean relative lipid volume fractions $\overline{x}_{\mathrm{lip}}$, associated deviations $\Delta x_{\mathrm{lip}}$ and relative deviations of the number of the solute mass $\Delta m_{\mathrm{s}}/\overline{m}_{\mathrm{s}}$. The MC simulations were performed for a mean water volume fraction of $\overline{\varphi}_1 = 0.9$ and $N_{\mathrm{v}} = 10^3$. (b): MD $\rho$ in g/ml in dependence of the RI contrast $\delta n$ for different mean relative lipid volume fractions $\overline{x}_{\mathrm{lip}}$ and mean water volume fractions ($\diamond \overline{\varphi}_1 = 0.1$, $\triangledown \overline{\varphi}_1 = 0.3$, $\bigcirc$ $\overline{\varphi}_1 = 0.5$, $\square \overline{\varphi}_1 = 0.7$, $\triangle \overline{\varphi}_1 = 0.9$) for $\Delta x_{\mathrm{lip}} = 0$, $N_0 = 10^3$ and $N_{\mathrm{v}} = 10^2$. The dashed lines indicate the analytical solutions of Eq. (23). (c): Correlative distribution of the MD $\rho$ in g/ml and the RI contrast $\delta n$ (the solid and dashed lines indicate the 68.3% and 95.5% confidence contours) with the corresponding normalized marginal probability density distributions $\hat{\mathcal{M}}(\rho)$ and $\hat{\mathcal{M}}(\delta n)$, respectively (the solid line represents the median, the dash-dotted and dashed lines indicate the 68.3% and 95.5% confidence intervals), for the exact relative lipid volume fraction $\overline{x}_{\mathrm{lip}} = 0.5$ (purple) and the relative lipid volume fraction following a truncated normal distribution $x_{\mathrm{lip}} \sim \mathcal{T}\left(\overline{x}_{\mathrm{lip}} = 0.5, \Delta x_{\mathrm{lip}} = 0.1\right)$ (cyan, see main text). The MC simulations were performed for a water volume fraction following a truncated normal distribution $\varphi_1 \sim \mathcal{T}\left(\overline{\varphi}_1 = 0.9, \Delta\varphi_1 = 0.1\right)$, $N_0 = 10^5$ and $N_{\mathrm{v}} = 10^3$.

values and respective references are stated in Tab. SI [25]. As becomes apparent, the measured and predicted values of the SM and IL are in good agreement, while potential uncertainties regarding the exact chemical composition

might be underappreciated here as we have no means of estimating them.

Furthermore, matching the experimental concentrations, we obtained a prediction of the MD in dependence

of the RI, which was found to coincide well with the measurements for both SM, and IL, as shown in Fig. 3(d).

### A. Larval zebrafish trunk tissue

Having delineated the good agreement between the assumptions used to derive an expression of the MD in dependence of the RI for two dilute, physiological samples, we now want to examine the capability of the proposed model in an *in vivo* scenario. To that aim we chose the larval zebrafish model system at 96 hours post-fertilization (hpf), for which mass spectrometry (MS) and RI data, employing ODT, of the trunk tissue were recently obtained [19]. The RI data is given in Tab. SIII [25]. As the MS data provides the protein content, we are able to estimate the RI increment and PSV distribution of the proteins present in the tissue, as demonstrated earlier, following Zhao *et al.* [23] (see Tab. I).

In the following, we assume that the larval zebrafish trunk tissue primarily consists of mentioned proteins, lipids and water, based on the findings of [40, 41]. The lipid composition of zebrafish larvae was determined in [41], from which we adapted the four phospholipid fatty acids (PLFA) with the highest abundance. The respective relative volume fractions of the PLFAs for zebrafish larvae at 96 hpf were roughly digitally obtained from [41] and are provided in Tab. SI [25]. These four PLFAs make up for about 72% of the total PLFA amount in the zebrafish larva. Additionally, based on [41], we assume that the overall lipid composition of the tissue is only given by triolein (TO) and said PLFAs.

Next, we inferred the distributions of the relative lipid volume fraction $x_{\mathrm{lip}}$ and the water volume fraction $\varphi_1$ based on the measurements of [40] (see SI [25]). From these initial estimations, employing Eq. (23), we obtained the correlative RI and MD distribution, displayed in Fig. 4(b), by MC sampling.

Evidently, both measured and predicted RI distributions do not coincide within one standard deviation. This fact is not unexpected since the measurements of [40] were done on whole animals, including the yolk sac which is rich in lipids. Hence, the assumed relative volume fraction of the lipids and the water volume fraction are likely to be different from the trunk tissue. However, given the deviations on the distributions of the relative lipid volume fraction and the water volume fraction, the resulting deviation of the predicted RI distribution matches remarkably well with the deviation of the measured RI distribution.

Hence, in an attempt to find the mean water volume fraction $\overline{\varphi}_1$ and the mean relative lipid volume fraction $\overline{x}_{\mathrm{lip}}$ that resembles the experimental condition, we evaluated the quantile comparison effect size (QCES, see SI; Eq. (S11) [25]), as proposed in [42], denoted by $\Xi$, between predicted and measured RI distributions for a range of different $\overline{\varphi}_1$ and $\overline{x}_{\mathrm{lip}}$, shown in Fig. 4(a). The

QCES provides an elegant way of comparing any two distributions by means of the vertical quantile comparison divergence. It is bounded on the domain $[0, 1]$; for $\Xi = 0$ the two distributions are identical and for $\Xi = 1$, they are maximally distinct. To give further intuition on the QCES, we may consider two normal distributions with mean values $\mu_1$ and $\mu_2$ and the standard deviations $\sigma_1 = \sigma_2 = \sigma$. For the case of $\mu_2 - \mu_1 = 1\sigma$, which corresponds to a Cohen's $d = 1$, the QCES $\Xi \approx 0.52$. Hence, we consider any two distributions to be significantly different when $\Xi > 0.52$.

With this at our disposal, we obtain a region in the $\overline{\varphi}_1 - \overline{x}_{\mathrm{lip}}$ plane that establishes concordance between the predicted and measured RI distributions, as indicated by the white dashed lines in Fig. 4(a). Evidently, the MD is not well constrained by this set of parameters, since combinations of $\overline{\varphi}_1$ and $\overline{x}_{\mathrm{lip}}$ for a wide range of values result in the same RI, but drastically different MD distributions (see Fig. 4(b)). Consequently, minimizing the QCES under the assumption that either the initial guess of the water volume fraction or the initial guess of the relative lipid volume fraction is adequate, one obtains incompatible MD estimates, as indicated by the cases '$\overline{x}_{\mathrm{lip}}$ adjusted' and '$\overline{\varphi}_1$ adjusted' in Fig. 4(b). This fact convincingly shows the ambiguity of estimations of the MD given a measured RI distribution without correlative constraints on the (local) chemical composition of the sample. However, reconsidering that the measurements of [40] were performed on whole animals, including the yolk sac, the real relative lipid volume fraction of the trunk tissue should be equal or lower than the 'initial guess'. This in turn makes the case of '$\overline{\varphi}_1$ adjusted' our best guess, yielding a MD estimate of $\rho^{\mathrm{B}} = (1.032 \pm 0.004)$ g/ml.

We note that, as indicated in Fig. 4(a), when employing the Lorentz-Lorenz mixing rule of RIs (Eq. (13)) instead of the Biot Eq. (5), the RI distribution obtained by the 'initial guess' is in good agreement with the measured RI distribution and we obtain an MD estimate for the case of '$\overline{\varphi}_1$ adjusted' of $\rho^{\mathrm{LL}} = (1.034 \pm 0.004)$ g/ml, which is, considering the uncertainties, well compatible with the estimate we obtained employing the Biot Eq. (5). However, it is worth pointing out that the Lorentz-Lorenz mixing rule did not agree well with the RI measurements of the validation samples presented earlier.

In essence, at present, it is not clear which RI mixing rule should be employed in this *in vivo* scenario, without a more comprehensive understanding regarding the biochemical composition of the sample. However, once this insight becomes available, e.g., by measuring (S)RS, different mixing rules could be compared to each other, as presented earlier. This in turn would also allow to study whether the 'best fitting' RI mixing rule is conserved across different specimen.

Finally, we want to point out that if we use the customary simplifying assumption that the solute composition of the trunk tissue is only given by proteins, employing Eq. (7), the RI increment and PSV given in Tab. I as well as the measured RI data (Tab. SIII [25]), we obtain
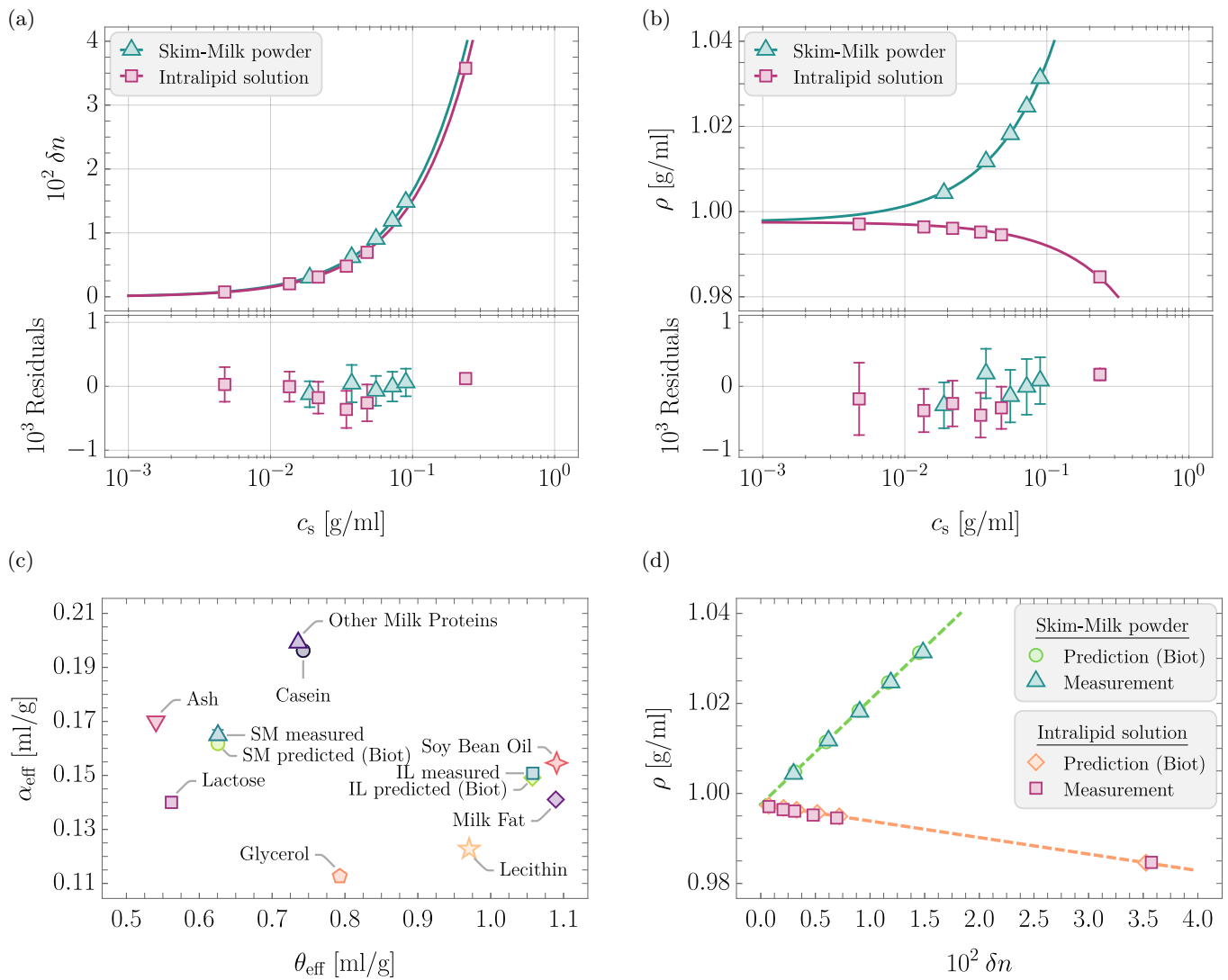
FIG. 3: Results of the concentration dependent measurements of the MD and RI of *bovine* skim-milk powder (SM) and 20% intralipid emulsion (IL) in water as well as the theoretical predictions based on the chemical composition using Eq. (23); (a): RI contrast $\delta n$ in dependence of the solute concentration $c_s$ in g/ml of SM ($\triangle$, $N = 5$ technical repetitions) and IL ($\square$, $N = 5$ technical repetitions) with the respective fits of Eq. (16) (solid lines) and fit residuals. (b): MD $\rho$ in g/ml in dependence of the solute concentration $c_s$ in g/ml of SM ($\triangle$, $N = 5$ technical repetitions) and IL ($\square$, $N = 5$ technical repetitions) with the respective fits of Eq. (4) (solid lines) and fit residuals. (c): Effective RI increment $\alpha_{\text{eff}}$ in ml/g and PSV $\theta_{\text{eff}}$ in ml/g of various substances that compose SM and IL, as well as the measured and predicted values for SM and IL. (d): MD $\rho$ in g/ml in dependence of the RI contrast $\delta n$ for different concentrations of SM in water and IL. The symbols represent measured values ($N = 5$ technical repetitions) and the predicted values using the Biot mixing rule (Eq. (5)) for $N_0 = 10^4$ and $N_v = 10^3$. The dashed lines indicate Eq. (23) for the predicted values of the effective RI increment and PSV.

$\rho_{\text{p}} = (1.040 \pm 0.004)$ g/ml, using GPU. Apparently, this value does not coincide well with the values determined above, $\rho^B$ and $\rho^{LL}$. In fact, as outlined earlier, not taking lipids into account results in a systematic overestimation of the mass density of biological matter.

## VI. DISCUSSION AND OUTLOOK

Considering the complexity of the chemical composition of biological matter in order to infer the 'optical' MD, based on a RI measurement is not well established in contemporary literature. Here we present a theoretical macroscopic model that is capable of describing the problem, employing a minimal set of assumptions, namely the
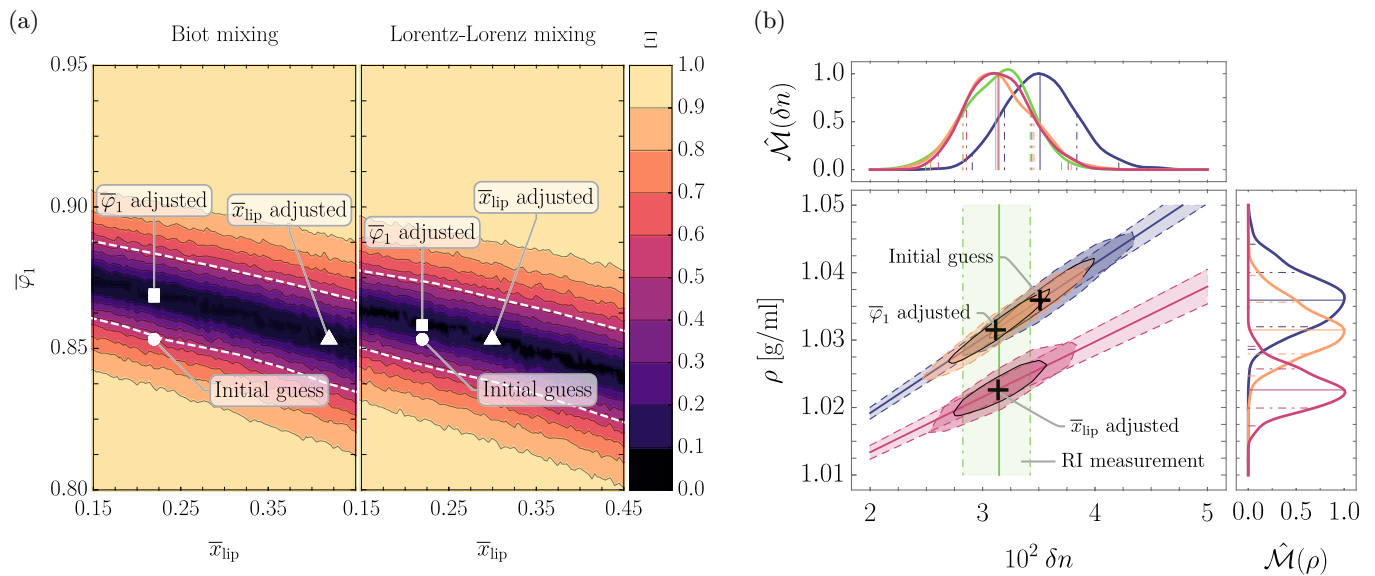
FIG. 4: Results of MC simulations and according analytical solutions for the trunk tissue of larval zebrafish; (a): Quantile comparison effect size (QCES) $\Xi$ between the RI distribution of larval zebrafish trunk tissue measured in [19] and the predictive RI distributions obtained from MC simulations of Eq. (23) using the Biot mixing rule (left, Eq. (5)) and the Lorentz-Lorenz mixing rule (right, Eq. (13)) for different mean water volume fractions $\overline{\varphi}_1$ and mean relative lipid volume fraction $\overline{x}_{lip}$. The symbols indicate $\bigcirc$: the initial guess, based on the measurements of [40], $\square$: the local minimum of $\Xi$ for the relative lipid volume fraction of the initial guess and $\triangle$: the local minimum of $\Xi$ for the water volume fraction of the initial guess. The white dashed lines indicate a $1\,\sigma$ deviation of the predicted RI distribution to the measured RI distribution. All MC simulations were performed for $N_0 = 10^3$ and $N_v = 10^3$. (b): Predicted correlative distributions of the MD $\rho$ in g/ml and the RI contrast $\delta n$ of larval zebrafish trunk tissue with the corresponding normalized marginal probability density distributions $\hat{\mathcal{M}}(\rho)$ and $\hat{\mathcal{M}}(\delta n)$, respectively (the solid line represents the median, the dash-dotted and dashed lines indicate the 68.3% and 95.5% confidence intervals). The ellipsoids indicate the different cases shown in Fig. (a): the initial guess (blue), $\overline{\varphi}_1$ adjusted (orange) and $\overline{x}_{lip}$ adjusted (red) with the corresponding analytical predictions for the different relative lipid volume fractions and varying water volume fraction. The green band indicates the RI measurement of [19]. The MC simulations were performed for $N_0 = 10^5$ and $N_v = 10^3$.

Biot mixing rule of RIs and the assumption of volume additivity.

We evaluated the possible sources of uncertainties associated with the model and showed that based on the chemical composition of the sample and the associated degree of inhomogeneity, the resulting RI and MD distributions might drastically differ from the customary assumption of biological matter consisting of proteins and water only. For that purpose, we provided analytical solutions as well as consistent simulation results for the case of a binary solute, composed of proteins and lipids.

While it is shown that for singular proteins in solution the assumption of volume additivity might not be justified [43] (see SI [25]), we provided experimental evidence that, for the set of validation samples under investigation, i.e., *bovine* skim-milk powder and 20% intralipid emulsion, it holds well within the measurement uncertainties.

Further, we provided evidence that the predictions of the correlative MD and RI based on our model agree with the experimentally obtained values of the validation samples, thus establishing confidence in the application of the

theoretical considerations presented here to estimate MD in dependence of the RI, given the chemical composition of the sample under study.

When applying the model to an *in vivo* specimen, i.e., the trunk tissue of the larval zebrafish, we observed that the mean value of RI measurements of [19] does not coincide within one standard deviation of our predictions, based on the estimations of the chemical composition of the tissue, employing the measurements of [40], where the authors determined the masses of water, proteins and lipids of whole animals. However, our initial predictions are remarkably close to the measurements, given the multitude of assumptions and simplifications employed.

We then explored the possibility of adjusting the hypothetical protein to lipid ratio and the water content of larval zebrafish trunk tissue in order to establish agreement between the measured and predicted RI distributions. To that aim we visited the topic of distribution comparisons. Employing the quantile comparison effect size, proposed in [42], we found a range of possible combinations of relative lipid volume fractions and water volume fractions

for which the predicted RI distribution coincides with the measured one. However, for these inferred ranges, the MD varies significantly, resulting in large deviations. In order to resolve this problem, complementary (S)RS measurements would be necessary. With the information about the local concentrations of proteins and lipids in the trunk tissue, obtained by such (S)RS measurements, the predictive capability of our model would be significantly increased [20, 21].

Going forward, the estimation of the MD of biological matter from RI measurements, as outlined in this study, will have interesting implications for inferring the mechanical properties from opto-acoustical measurements, e.g., *via* Brillouin microscopy. The problem of a varying solute composition within a sample is well appreciated (see e.g., [12]), but not resolved in a cohesive manner. Evidently, combining Brillouin microscopy not only with ODT, but also incorporating the local biochemical composition (e.g., *via* (S)RS measurements) allows for a better estimation of the longitudinal (elastic) modulus. We note that the aforementioned problem could also be resolved by performing stimulated Brillouin microscopy in combination with ODT. Here, the MD can be obtained directly from measurements of the Brillouin resonance gain factor, which in turn is connected to the Brillouin gain and the pump laser power, as well as the RI, Brillouin frequency shift and Brillouin line width [44]. Accurately determining the local (*in vivo*) MD, will conceivably enable a more profound interpretation of functional mechanisms at play in biological matter.

The possibility of linking the MD estimates of a sample with the inherent dynamics of the system, i.e., the statistical processes, poses an intriguing endeavor. Moreover, explorations of the predictive RI and MD distributions of cells, in particular *Xenopus* egg extract, i.e., cytoplasm, for which the chemical composition could be accurately determined, will be insightful for inferences regarding the applicability and predictive capability of our model.

Finally, describing the problem of MD estimations, given uncertain *a priori* assumptions, in a Bayesian framework should be investigated.

In order to make the application of the findings of this study more accessible, we delineate strategies on how to estimate MD, given certain experimental paradigms, in the supplemental information [25].

## VII. MATERIALS AND METHODS

### A. Sample preparation

The skim-milk powder was dissolved in distilled water while carefully stirring the solution to avoid foaming. The solution was then left on a tilt/roller mixer for approximately 30 minutes. For the intralipid emulsion, according amounts of water were added to the emulsion

and the solution was left on a tilt/roller mixer for approximately 30 minutes, as well. All samples were freshly prepared before the measurements were performed.

### B. Abbe refractometry and pycnometry

For measuring the refractive index of a liquid sample, 100 μl of the sample were loaded into an Abbe refractometer (KERN ORT 1RS) and a commercially available flashlight LED was employed as illumination source.

To determine the density of a liquid sample, a pycnometer (Blaubrand 43305) was employed. First, the volume of the pycnometer was determined by employing distilled water as a calibration sample ($N = 10$ technical repetitions) as

$$v_{\text{pyc}} = \frac{m_{\text{w}} - m_{\text{pyc}}}{\rho_{\text{w,lit.}}} = (4.9455 \pm 0.0017) \text{ ml}, \quad (44)$$

where $m_{\text{w}}$ is the mass of the pycnometer filled with water, $m_{\text{pyc}}$ is the mass of the empty pycnometer and $\rho_{\text{w,lit.}} = 0.997$g/ml is the literature value of the density of water at $23°$C. The respective masses were measured using a high precision lab scale (Ohaus Pioneer PX124). The density of the liquid sample under study was then computed by

$$\rho = \frac{m - m_{\text{pyc}}}{v_{\text{pyc}}}. \quad (45)$$

### C. Measurement uncertainties

The systematic uncertainties of the respective measurement devices under use were taken from the manuals and considered in all calculations together with the statistical uncertainties by employing Gaussian propagation of uncertainty as

$$\Delta z = \sqrt{\Delta z_{\text{sys}}^2 + \Delta z_{\text{stat}}^2}, \quad (46)$$

where $z$ is an arbitrary observable.

### D. Data analysis

All data analysis, plotting and simulations were performed using custom scripts in `Wolfram Research, Inc., Mathematica, Version 12.2` [45].

### ACKNOWLEDGMENTS

[1] J. Guck, S. Schinkinger, B. Lincoln, F. Wottawah, S. Ebert, M. Romeyke, D. Lenz, H. M. Erickson, R. Ananthakrishnan, D. Mitchell, J. Käs, S. Ulvick, and C. Bilby, Optical deformability as an inherent cell marker for testing malignant transformation and metastatic competence, Biophysical Journal **88**, 3689 (2005).

[2] A. J. Engler, S. Sen, H. L. Sweeney, and D. E. Discher, Matrix elasticity directs stem cell lineage specification, Cell **126**, 677 (2006).

[3] P. Egan, R. Sinko, P. R. LeDuc, and S. Keten, The role of mechanics in biological and bio-inspired systems, Nature Communications **6**, 7418 (2015).

[4] A. Biswas, O. Munoz, K. Kim, C. Hoege, B. M. Lorton, D. Shechter, J. Guck, V. Zaburdaev, and S. Reber, Conserved nucleocytoplasmic density homeostasis drives cellular organization across eukaryotes, bioRxiv 10.1101/2023.09.05.556409 (2023).

[5] S. Iida, S. Ide, S. Tamura, T. Tani, T. Goto, M. Shribak, and K. Maeshima, Orientation-independent-dic imaging reveals that a transient rise in macromolecular crowding contributes to mitotic chromosome condensation, bioRxiv 10.1101/2023.11.11.566679 (2023).

[6] J. L. Watson, E. Seinkmane, C. T. Styles, A. Mihut, L. K. Krüger, K. E. McNally, V. J. Planelles-Herrero, M. Dudek, P. M. McCall, S. Barbiero, M. Vanden Oever, S. Y. Peak-Chew, B. T. Porebski, A. Zeng, N. M. Rzechorzek, D. C. S. Wong, A. D. Beale, A. Stangherlin, M. Riggi, J. Iwasa, J. Morf, C. Miliotis, A. Guna, A. J. Inglis, J. Brugués, R. M. Voorhees, J. E. Chambers, Q.-J. Meng, J. S. O'Neill, R. S. Edgar, and E. Derivery, Macromolecular condensation buffers intracellular water potential, Nature **623**, 842 (2023).

[7] M. Schürmann, J. Scholze, P. Müller, J. Guck, and C. J. Chan, Cell nuclei have lower refractive index and mass density than cytoplasm, Journal of Biophotonics **9**, 1068 (2016).

[8] A. Biswas, K. Kim, G. Cojoc, J. Guck, and S. Reber, The xenopus spindle is as dense as the surrounding cytoplasm, Developmental Cell **56**, 967 (2021).

[9] G. Antonacci and S. Braakman, Biomechanics of subcellular structures by non-invasive Brillouin microscopy, Scientific Reports **6**, 1 (2016).

[10] G. Yan, S. Monnier, M. Mouelhi, and T. Dehoux, Probing molecular crowding in compressed tissues with brillouin light scattering, Proceedings of the National Academy of Sciences **119**, e2113614119 (2022), https://www.pnas.org/doi/pdf/10.1073/pnas.2113614119.

[11] G. Scarcelli, W. J. Polacheck, H. T. Nia, K. Patel, A. J. Grodzinsky, R. D. Kamm, and S. H. Yun, Noncontact three-dimensional mapping of intracellular hydromechanical properties by Brillouin microscopy, Nature Methods **12**, 1132 (2015).

[12] R. Schlüßler, K. Kim, M. Nötzel, A. Taubenberger, S. Abuhattum, T. Beck, P. Müller, S. Maharana, G. Cojoc, S. Girardo, A. Hermann, S. Alberti, and J. Guck, Correlative all-optical quantification of mass density and mechanics of subcellular compartments with fluorescence specificity, eLife **11**, e68490 (2022).

[13] B. F. Kennedy and J. C. Bamber, Introduction to Optical Coherence Elastography, in *Optical Coherence Elastography: Imaging Tissue Mechanics on the Micro-Scale* (AIP Publishing LLC).

[14] K. V. Larin and D. D. Sampson, Optical coherence elastography - oct at work in tissue biomechanics [invited], Biomed. Opt. Express **8**, 1172 (2017).

[15] K. Kim and J. Guck, The relative densities of cytoplasm and nuclear compartments are robust against strong perturbation, Biophysical Journal **119**, 1946 (2020).

[16] B. Bhaduri, H. Pham, M. Mir, and G. Popescu, Diffraction phase microscopy with white light, Opt. Lett. **37**, 1094 (2012).

[17] R. Schlüßler, S. Möllmert, S. Abuhattum, G. Cojoc, P. Müller, K. Kim, C. Möckel, C. Zimmermann, J. Czarske, and J. Guck, Mechanical mapping of spinal cord growth and repair in living zebrafish larvae by brillouin imaging, Biophysical Journal **115**, 911 (2018).

[18] R. Rollin, J.-F. Joanny, and P. Sens, Physical basis of the cell size scaling laws, eLife **12**, e82490 (2023).

[19] J. Kolb, V. Tsata, N. John, K. Kim, C. Möckel, G. Rosso, V. Kurbel, A. Parmar, G. Sharma, K. Karandasheva, S. Abuhattum, O. Lyraki, T. Beck, P. Müller, R. Schlüßler, R. Frischknecht, A. Wehner, N. Krombholz, B. Steigenberger, D. Beis, A. Takeoka, I. Blümcke, S. Möllmert, K. Singh, J. Guck, K. Kobow, and D. Wehner, Small leucine-rich proteoglycans inhibit cns regeneration by modifying the structural and mechanical properties of the lesion environment, Nature Communications **14**, 6814 (2023).

[20] S. Oh, C. Lee, W. Yang, A. Li, A. Mukherjee, M. Basan, C. Ran, W. Yin, C. J. Tabin, D. Fu, X. S. Xie, and M. W. Kirschner, Protein and lipid mass concentration measurement in tissues by stimulated raman scattering microscopy, Proceedings of the National Academy of Sciences **119**, 10.1073/pnas.2117938119 (2022).

[21] M. Alunni Cardinali, A. Di Michele, M. Mattarelli, S. Caponi, M. Govoni, D. Dallari, S. Brogini, F. Masia, P. Borri, W. Langbein, F. Palombo, A. Morresi, and D. Fioretto, Brillouin–raman microspectroscopy for the morpho-mechanical imaging of human lamellar bone, Journal of The Royal Society Interface **19**, 20210642 (2022).

[22] S. Abuhattum, P. Kotzbeck, R. Schlüßler, A. Harger, A. Ariza de Schellenberger, K. Kim, J.-C. Escolano, T. Müller, J. Braun, M. Wabitsch, M. Tschöp, I. Sack, M. Brankatschk, J. Guck, K. Stemmer, and A. V. Taubenberger, Adipose cells and tissues soften with lipid accumulation while in diabetes adipose tissue stiffens, Scientific Reports **12**, 10325 (2022).

[23] H. Zhao, P. H. Brown, and P. Schuck, On the distribution of protein refractive index increments, Biophysical Journal **100**, 2309 (2011).

[24] H. Durchschlag, Specific volumes of biological macromolecules and some other molecules of biological interest (Springer Berlin Heidelberg, Berlin, Heidelberg, 1986) pp. 45–128.

[25] See supplemental material at http://link-to-SI.com for further information, equations and figures.

[26] W. Heller, Remarks on refractive index mixture rules, *The Journal of Physical Chemistry*, The Journal of Physical Chemistry **69**, 1123 (1965).

[27] P. Walstra, J. T. M. Wouters, and T. J. Geurts, *Dairy Science and Technology* (CRC Press, 2005).

[28] B. F. Hanley, A practical method for estimating specific refractive index increments for flexible non-electrolyte polymers and copolymers in pure and mixed solvents using the gladstone-dale and lorentz-lorenz equations in conjunction with molar refraction structural constants, and solvent physical property databases, Materials Today Communications **25**, 101644 (2020).

[29] F. Pretorius, W. W. Focke, R. Androsch, and E. du Toit, Estimating binary liquid composition from density and refractive index measurements: A comprehensive review of mixing rules, Journal of Molecular Liquids **332**, 115893 (2021).

[30] T. U. Consortium, UniProt: the Universal Protein Knowledgebase in 2023, Nucleic Acids Research **51**, D523 (2022).

[31] T. G. Mayerhöfer and J. Popp, Beyond beer's law: Revisiting the lorentz-lorenz equation, ChemPhysChem **21**, 1218 (2020).

[32] P. Brocos, Piñeiro, R. Bravo, and A. Amigo, Refractive indices, molar volumes and molar refractions of binary liquid mixtures: concepts and correlations, Phys. Chem. Chem. Phys. **5**, 550 (2003).

[33] Y. Liu and P. H. Daum, Relationship of refractive index to mass density and self-consistency of mixing rules for multicomponent mixtures like ambient aerosols, Journal of Aerosol Science **39**, 974 (2008).

[34] S. J. PERKINS, Protein volumes and hydration effects, European Journal of Biochemistry **157**, 169 (1986).

[35] C. E. J., Density and apparent specific volume of protein, Proteins, Amino Acids and Peptides as Ions and Dipolar Ions , 370 (1943).

[36] A. Chorlay and A. R. Thiam, Neutral lipids regulate amphipathic helix affinity for model lipid droplets, Journal of Cell Biology **219**, 10.1083/jcb.201907099 (2020), e201907099.

[37] G. Onal, O. Kutlu, D. Gozuacik, and S. Dokmeci Emre, Lipid droplets in health and disease, Lipids in Health and Disease **16**, 128 (2017).

[38] A. R. Thiam and E. Ikonen, Lipid droplet nucleation, Trends in Cell Biology **31**, 108 (2021).

[39] K. Kim, V. R. Gade, T. V. Kurzchalia, and J. Guck, Quantitative imaging of caenorhabditis elegans dauer larvae during cryptobiotic transition, Biophysical Journal **121**, 1219 (2022).

[40] Y. Long, L. Li, Q. Li, X. He, and Z. Cui, Transcriptomic characterization of temperature stress responses in larval zebrafish, PLOS ONE **7**, 1 (2012).

[41] N. Hachicho, S. Reithel, A. Miltner, H. J. Heipieper, E. Küster, and T. Luckenbach, Body mass parameters, lipid profiles and protein contents of zebrafish embryos and effects of 2,4-dinitrophenol exposure, PLOS ONE **10**, 1 (2015).

[42] D. A. Wooff and A. Jamalzadeh, Robust and scale-free effect sizes for non-normal two-sample comparisons, with applications in e-commerce, *Journal of Applied Statistics*, Journal of Applied Statistics **40**, 2495 (2013).

[43] A. Zamyatnin, Protein volume in solution, Progress in Biophysics and Molecular Biology **24**, 107 (1972).

[44] R. W. Boyd, Chapter 9 - stimulated brillouin and stimulated rayleigh scattering, in *Nonlinear Optics (Third Edition)*, edited by R. W. Boyd (Academic Press, Burlington, 2008) third edition ed., pp. 429–471.

[45] W. R. Inc., Mathematica, Version 12.2, champaign, IL, 2020.

[46] ChemSpider (Accessed Feb 2, 2023)
- Triolein, CSID:4593733, http://www.chemspider.com/Chemical-Structure.4593733.html
- Palmitic acid , CSID:960, http://www.chemspider.com/Chemical-Structure.960.html
- Oleic acid, CSID:393217, http://www.chemspider.com/Chemical-Structure.393217.html
- Docosahexaenoic acid, CSID:393183, http://www.chemspider.com/Chemical-Structure.393183.html
- Stearic acid, CSID:5091, http://www.chemspider.com/Chemical-Structure.5091.html

.

[47] Merck (Accessed Aug 31, 2023)
- Soybean oil, 47122, https://www.sigmaaldrich.com/DE/de/product/supelco/47122
- Glycerol, 104057, https://www.merckmillipore.com/DE/de/product/Glycerol,MDA_CHEM-104057?ReferrerURL=https%3A%2F%2Fwww.google.com%2F

.

[48] N. Bansal, T. Truong, and B. Bhandari, Feasibility study of lecithin nanovesicles as spacers to improve the solubility of milk protein concentrate powder during storage, Dairy Science & Technology **96**, 861 (2017).

[49] Oxford Lab Fine Chem LLP (Accessed Aug 31, 2023)
- Soybean lecithin, https://www.oxfordlabchem.com/msds/(S-08441)SOYA%20LECITHIN%2030%20Extra%20Pure.pdf

.

[50] D. Khago, J. C. Bierma, K. W. Roskamp, N. Kozlyuk, and R. W. Martin, Protein refractive index increment is determined by conformation as well as composition, Journal of Physics: Condensed Matter **30**, 435101 (2018).

[51] Wolfram Research, Inc., Wolfram|Alpha Knowledgebase, Champaign, IL (2022).

[52] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, PubChem 2023 update, Nucleic Acids Research **51**, D1373 (2022).

[53] T. L. McMeekin, M. Wilensky, and M. L. Groves, Refractive indices of proteins in relation to amino acid composition and specific volume, Biochemical and Biophysical Research Communications **7**, 151 (1962).