Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2025, **21**, 8669

Predicting the structure and swelling of microgels with different crosslinker concentrations by combining machine learning with numerical simulations

Susana Marín-Aguilar *[a] and Emanuela Zaccarelli [ab]

Received 22nd August 2025,
Accepted 15th October 2025

DOI: 10.1039/d5sm00852b

rsc.li/soft-matter-journal

Microgels made of poly(*N*-isopropylacrylamide) are the prototype of soft, thermoresponsive particles widely used to study fundamental problems in condensed matter physics. However, their internal structure is far from homogeneous, and existing mean-field approaches, such as Flory–Rehner theory, provide only qualitative descriptions of their thermoresponsive behavior. Here, we combine machine learning and numerical simulations to accurately predict the concentration and spatial distribution of crosslinkers, the latter hitherto unknown experimentally, as well as the full swelling behavior of microgels, using only polymer density profiles. Our approach provides unprecedented insight into the structural and thermodynamic properties of any standard microgel.

1 Introduction

Microgels are colloidal particles produced by the simple yet versatile process of precipitation polymerization, and are nowadays used as a favourite model system of soft, deformable particles.[1,2] Since they are entirely made of polymers, arranged in a disordered network *via* chemical crosslinking, they are able to respond to external stimuli. For example, they can easily change their volume in response to a temperature variation when the constituent polymer is thermoresponsive, such as poly(*N*-isopropylacrylamide) (pNIPAM).[3–5] Therefore, at a characteristic temperature, the polymer network undergoes a so-called volume phase transition (VPT), from a swollen to a collapsed state. This can be exploited to study fundamental physics problems *via* an *in situ* tuning of the suspension packing fraction.[6,7] One important parameter that can be easily varied experimentally is the molar concentration of crosslinking molecules, hereafter abbreviated as crosslinker concentration c, which controls the overall softness of a microgel.[2] It is well-known that, in the standard synthesis protocol—precipitation polymerization—crosslinkers react faster than monomers, thereby accumulating within a central region of the microgel, the so-called core, which can be distinguished from the outer corona, where crosslinkers are almost absent. Despite several attempts made using chemical kinetic modeling, to our knowledge, there is no available work capable of predicting the crosslinker distribution within pNIPAM microgels. Instead, to describe the VPT, there exists the celebrated theory of Flory and Rehner,[8] which combines mean-field-like theoretical assumptions with more phenomenological approaches, to describe the available experimental swelling curves. However, it is well-known that the theory suffers some important drawbacks, such as the presence of many fit parameters, whose physical meaning is often obscure.[9] Therefore, it would be important to have an alternative way to predict the swelling behavior of a microgel.

In this article, we take advantage of a recently put forward monomer-resolved microgel model, which has been shown to accurately describe the internal structure of the experimental systems, both in bulk suspensions[10] and at liquid–liquid interfaces,[11] to fill this gap. To this aim, we perform extensive computer simulations of individual microgels with different crosslinker concentrations to create a comprehensive numerical database of microgel structures at many different temperatures across the VPT. Using a subset of these structures as the training set, we apply machine learning (ML) techniques to predict crosslinker properties. Building on the concept that density profiles can encode thermal observables[12,13] and all relevant structural information, we demonstrate that the total polymer density profiles of microgels in the swollen state are enough to accurately predict the crosslinker concentration, c, *via* unsupervised ML. This approach is subsequently validated against simulations of microgels with different sizes and c outside the training set, as well as for available experimental data. Following a similar approach, we then train a supervised neural network

[a] *Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro 2, 00185 Roma, Italy. E-mail: susana.marinaguilar@uniroma1.it, emanuela.zaccarelli@cnr.it*
[b] *CNR Institute of Complex Systems, Uos Sapienza, Piazzale Aldo Moro 2, Roma, 00185, Italy*

(NN) to predict the distribution of the crosslinkers, which is found to follow the fuzzy sphere model,[14,15] in analogy to the overall microgel. Our approach thus enables, for the first time, the prediction of the crosslinker distribution in experimental systems. Finally, we use our database to build a phenomenological framework to predict the complete swelling behavior of the microgels for a given c, similar in spirit to the Flory–Rehner theory. Therefore, from the sole knowledge of the microgel total density profile at low temperature, which can be extracted from the fit of the form factor, routinely measured nowadays by small-angle scattering techniques[14] or directly obtained by super-resolution microscopy,[15] our ML approach is able to reliably predict the concentration of crosslinkers and their radial distribution, as well as the full swelling behavior of any standard pNIPAM microgel.

2 Models and methods

2.1 Molecular simulations

The *in silico* synthesis of microgels follows the approach established in ref. 4, 10 and 16, which was shown to accurately describe the structure of experimental systems.[10,17] The protocol involves two steps. First, we perform NVT molecular dynamics (MD) simulations of a binary mixture of $N = N_p + N_c$ patchy particles confined within a sphere of radius Z. The mixture contains N_p particles with two patches representing monomers and N_c particles with four patches acting as crosslinkers. These simulations are performed using the OxDNA package.[18] To reproduce the core–corona nature of microgels, a radial force is exerted on the crosslinking particles, as described in ref. 10. In the second step, once the microgel is assembled, we replace the patchy interactions with bead-spring bonds, thereby fixing the structure.

We simulate the assembled polymer networks with different crosslinker molar fractions $c \in [0.5, 1.25, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 10.0, 12.5, 15.0]$, covering the full range of experimentally synthesized pNIPAM microgels. The total number of particles in the microgel is set to either $N \sim 42\,000$ or $N \sim 336\,000$. The number of crosslinkers is given by $N_c = cN/100$, with the rest being monomers. All particles interact with the Weeks–Chandler–Andersen (WCA) potential,

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \varepsilon & r \le 2^{1/6}\sigma \\ 0 & r > 2^{1/6}\sigma, \end{cases} \quad (1)$$

where r is the distance between two beads, σ is the diameter of each bead, setting the unit of length, and ε controls the energy scale and sets the unit of energy. For bonded beads, there is an additional interaction modeled by the finite-extensible-nonlinear-elastic (FENE) potential:[19]

$$U_{\mathrm{FENE}}(r) = -\varepsilon k_F R_0^2 \log\left[1 - \left(\frac{r}{R_0\sigma}\right)^2\right] \quad \text{if } r < R_0\sigma, \quad (2)$$

with $R_0 = 1.5$ and $k_F = 15$ being the maximum bond extension and stiffness, respectively. To simulate the effect of temperature, we include a solvophobic attraction,[16,20] which implicitly captures the change in the affinity of the monomers with respect to the solvent. This potential reads as

$$U_\alpha(r) = \begin{cases} -\varepsilon\alpha & r \le 2^{1/6}\sigma \\ \frac{1}{2}\alpha\varepsilon\left[\cos\left(\gamma\left(\frac{r}{\sigma}\right)^2 + \beta\right) - 1\right] & 2^{1/6}\sigma < r \le R_0, \\ 0 & r > R_0 \end{cases} \quad (3)$$

where $\gamma = (\pi(2.25 - 2^{1/3}))^{-1}$, $\beta = 2\pi - 2.25\gamma$, and α is the strength of the attractive interaction, thus amounting to an effective temperature. The network is in good solvent conditions for $\alpha = 0$, undergoes the volume phase transition (VPT) for $\alpha \sim 0.63$ and reaches the collapsed state observed in experiments for $\alpha \sim 0.8$–1.0. Previous studies have shown that α varies approximately linearly with temperature across the VPT range[10,17] (see the SI, Section S4).

We perform NVT MD simulations of individual microgels using the LAMMPS package[21] with a time step of $\delta t = 0.002\tau$, where $\tau = \sqrt{m\sigma^2/\varepsilon}$ corresponds to the time unit. The temperature is fixed to $k_BT/\varepsilon = 1.0$, where k_B is the Boltzmann constant, while we vary the solvophobic parameter α to mimic the effect of the real temperature, as described above. The center of mass of the microgel is fixed to the center of the simulation box. We equilibrate the system for at least 3×10^6 timesteps, followed by a production simulation of 1×10^7 timesteps. To account for the role of disorder in the assembled polymer networks, the results are averaged over at least three independent topologies.

2.2 Structural analysis

The main structural information used as input for the neural networks is obtained from the radial density profile of the monomers $\rho(r)$, with respect to the microgel center of mass. This is averaged over different time intervals and it is defined as

$$\rho(r) = \left\langle \sum_{i=1}^{N} \delta(|\mathbf{r}_i - \mathbf{r}_{\mathrm{cm}}| - r) \right\rangle_{\Delta t}, \quad (4)$$

where $\mathbf{r}_{\mathrm{cm}}$ corresponds to the position of the center of mass of the microgel, $\mathbf{r}_i$ that of particle i, and $\langle\,\rangle$ represents a time average over the window of time Δt. In order to generate a large pool of density profiles, we generate such profiles every 1×10^4 timesteps, averaging every $\Delta t = 1 \times 10^5$ timesteps inside the previous time window. This choice ensures that each sampled configuration represents an independent snapshot of the system's structure. In addition, we calculate the radial density profiles of the crosslinker monomers only, $\rho_c(r)$, by substituting N in eqn (4) with N_c.

We fit $\rho(r)$ using the well-established fuzzy sphere model,[14] which combines the profile of a solid sphere of radius R_c with a Gaussian function representing the corona of the particle.[14] In an approximated version[22] that holds for standard core–corona microgels, this reduces to

$$\rho(r) \sim A\ \mathrm{erfc}\left(\frac{r - R_c}{\sqrt{2}\sigma_s}\right), \quad (5)$$

where A is a fit parameter, R_c denotes the radius of the core, and σ_s is the half-width of the corona shell.

We also compute the hydrodynamic radius, R_H, of the microgel, following the method established in ref. 4. First, we construct the convex-hull that encloses all of the microgel monomers. This convex hull is tessellated into triangular facets by construction, and their vertex coordinates are used to build the gyration tensor. The R_H is finally computed as

$$R_H = 2\left[\int_0^\infty \frac{1}{\sqrt{(a_1^2+\theta)(a_2^2+\theta)(a_3^2+\theta)}}\mathrm{d}\theta\right]^{-1}, \quad (6)$$

with a_1, a_2, and a_3 being the principal semiaxes of the gyration tensor, obtained from its eigenvalues, λ_i, as $a_i = \sqrt{3\lambda_i}$. This definition approximates the microgel to an effective ellipsoid with the same gyration tensor, providing a meaningful measure of its overall size in solution. Further details and validations of this approximation can be found in ref. 4 and 23. The swelling behavior is characterized by the α-dependent swelling ratio S_α, defined as

$$S_\alpha = R_H(\alpha)/R_H(\alpha^\star = 0.86), \quad (7)$$

where $R_H(\alpha^\star = 0.86)$ roughly corresponds to the hydrodynamic radius of the collapsed microgel at the effective temperature of $\alpha^\star = 0.86$. The latter is estimated from the experimental data of ref. 17.

2.3 Machine learning methods

To analyze and predict different structural features of the microgels, we employ both unsupervised and supervised machine learning (ML) approaches. We use autoencoders (AE)[24,25] to determine the crosslinker concentration and neural networks (NN)[26] to predict crosslinker distributions. Both models are implemented and trained using PyTorch.[27]

2.4 Autoencoders

An autoencoder is a nonlinear dimensionality reduction method based on a neural network (NN) trained to reproduce its input at the output. It consists of an encoder that maps high-dimensional input data to a low-dimensional latent space and a decoder that reconstructs the input from this latent representation. The AE is trained by minimizing a loss function that quantifies the difference between the input and the reconstructed output. Once trained, we discard the decoder and retain only the encoder using the latent space representation to predict the crosslinker concentration of a microgel.

As input, we use a collection of discretized radial density profiles from MD simulations of microgels with different c values, represented by a set of vectors $\mathbf{P}(i) \in \mathbb{R}^d$, where $\mathbf{P}(i) = \rho_i$ denotes the discretized density profile of sample i and d denotes the dimensionality of the vector. The choice of d allows us to control the resolution of the density profile. Here, we choose d = 115. In addition, the entries $P_j(i)$ of the input vector correspond to the value of the density profile evaluated at r_j, and thus, we ensure that for all considered data sets, regardless of the crosslinker concentration, the entry j always corresponds to the same distance r_j. In order to guarantee this condition and to be able to compare microgels of different sizes, we preprocess the data obtained from the radial density profiles as follows. First, we use a scaled radial coordinate $r^\star$ that preserves the characteristic decay of each state point. This is done using the value of the integral of the density profile, which in simulations corresponds to the total number of particles in the microgel $N = \int \rho \mathrm{d}V$, and by rescaling $r^\star = r/N^{1/3}$. The latter normalizes the integral of $\rho(r^\star)$ to one. Second, to ensure that entry j corresponds to the same $r^\star$ across the dataset, we interpolate $\rho(r^\star)$ using splines and construct $\mathbf{P}(i)$ by evaluating the splines in intervals of $r^\star$ = 0.025 starting from $r^\star$ = 0.125. In the output layer, the density profiles, represented by a vector $\mathbf{P}'(i) \in \mathbb{R}^d$, are recovered.

Inside the encoder, we perform a nonlinear projection onto a low-dimensional space $\mathbf{L}(i) \in \mathbb{R}^{d'}$, where $d' < d$. We choose d' in such a way that more than 99% of the variance of the input data set is recovered, in this case, corresponding to d' = 1, as shown later. We use a hyperbolic tangent activation function,[28] varying the number of neurons in the hidden layer from 32 up to 160 and adjusting the batch size to optimize the parameters that better minimize the error function. In particular, we fix the number of neurons in the hidden layer to 80, as we do not find relevant changes in the results obtained with different architectures.

To train the AE, we minimize a loss function using the mini-batch stochastic gradient descent with momentum.[29–31] In particular, we use the mean squared error function with the addition of a weight decay regularization term[29] defined as

$$E = \frac{1}{M}\sum_{i=1}^{M}|\mathbf{P}(i) - \mathbf{P}'(i)|^2 + \lambda\sum_{j=1}^{W} w_j^2, \quad (8)$$

where W is the total number of weights w, M is the total number of samples in the training data set, and $\lambda = 10^{-5}$ is a regularization parameter that penalizes large weight values. In addition, we use a learning rate of 0.01. We measure the performance of the AE by calculating the fraction of variance explained (FVE) by the NN, defined as

$$\mathrm{FVE} = 1 - \frac{\mathrm{MSE}}{\mathrm{var}(P)} = 1 - \frac{\sum_{i=1}^{M}|\mathbf{P}(i) - \mathbf{P}'(i)|^2}{\sum_{i=1}^{M}|\mathbf{P}(i) - \langle\mathbf{P}\rangle|^2} \quad (9)$$

where the MSE corresponds to the mean-squared-error and var denotes the variance of $\mathbf{P}$.

2.5 Neural networks

To predict the distribution of the crosslinkers within the microgel, we use a simple NN with a single hidden layer and a rectified linear unit (ReLU) activation function.[28] The same normalized discretized radial density profiles $\mathbf{P}(i) \in \mathbb{R}^d$ with d = 115 from the AE are used as input to the NN. In this case, the feature to learn corresponds to the crosslinker distribution ρ_c represented by the vector $\mathbf{P}'$. As for the AE, the radial coordinate is normalized with the total number of particles $1/N^{1/3}$. Due to the small values of ρ_c, the y-axis is also normalized by the corresponding value of c. The number of neurons in the hidden

layer is varied from 32 to 736. We minimize the error function defined in eqn (8) between the predicted crosslinker density profiles $\mathbf{P}'$ and the original ones from the training set, using mini-batch stochastic gradient descent with momentum[29–31] and a learning rate of 0.1. The NN is trained until the number of epochs reaches 5×10^6 or the MSE difference between the last 500 epochs is less than 1×10^{-7}. We finally fix the number of neurons to 96 reaching a RMSE of 1.3×10^{-6}.

3 Results and discussion

3.1 Effect of the crosslinker concentration on the microgel structure

In order to generate a database of microgel particles, we perform molecular dynamics (MD) simulations of monomer-resolved microgels interacting with the bead-spring potential,[19,20] as described in the Models and methods section. Each microgel contains a total of either $N \sim 42\,000$ or $N \sim 336\,000$ beads, with the molar crosslinker concentration c ranging from $c = 0.5\%$ to $c = 15\%$. This parameter plays a key role in determining the microgel topology, resulting in significant structural and elastic differences.[2,17] We start by reporting the dependence of the microgel structure on the crosslinker concentration in the swollen state, *i.e.* $\alpha = 0.0$. Fig. 1(a) shows the radial density profiles $\rho(r)$ of the microgels with $N \sim 336\,000$ beads. The change in c has a direct effect on the shape of the microgel, as reflected in the profiles, which become progressively sharper with increasing c. In particular, the behavior of $\rho(r)$ displays an almost constant regime, representing the core, followed by a decay at large r in the corona region. The latter becomes progressively more extended and less distinct from the core with decreasing c. Microgels with $N \sim 42\,000$ beads exhibit similar behavior, as shown in the SI (Fig. S1). Typical snapshots of the microgels are shown in Fig. 1(b), changing from a rather compact structure for $c = 10\%$ to a much more heterogeneous one for $c = 0.5\%$, denoting the presence of so-called dangling ends in the exterior of the microgel. All density profiles are well-described by the fuzzy sphere model,[14] as shown in Fig. 1(a), with the corresponding fit parameters reported in Table S1 of the SI.

In Fig. 1(c), we show the corresponding crosslinker radial density profiles, ρ_c, of the microgels with $N \sim 336\,000$ beads. Similar to the total density profiles, the crosslinkers alone exhibit a similar profile. In the core region, they are roughly constant, although with larger statistical noise due to the small amount of crosslinkers. However, they decay to zero at much shorter distances in r with respect to $\rho(r)$, signaling the accumulation of the crosslinkers within the core. In addition, this decay distance does not strongly depend on c in contrast to what is observed for $\rho(r)$. It is worth noting that crosslinker profiles are also found to be quite well-described by the fuzzy sphere model. This description smoothens out local fluctuations, particularly at small r, where structural inhomogeneities are more pronounced. In this sense, the fuzzy-sphere model provides a useful representation that captures the overall shape of the distribution, while averaging out microscopic details. The corresponding fuzzy-sphere fits are also shown in Fig. 1(c), with continuous lines with the fit parameters reported in Table S1 of the SI.

Finally, variations in c also affect the microgel size. For instance, the hydrodynamic radius R_H displays a power-law dependence $R_H \sim c^{-0.21\pm0.01}$, in good agreement with the predictions of the Flory–Rehner theory for polymer networks.[8,32] This behavior, also shared by the calculated gyration radius, was also previously reported in experimental works.[17,33] This is shown in Fig. 1(d), where the results from simulations of microgels with both $N \sim 42\,000$ and $N \sim 336\,000$ beads as a function of c are shown. Instead, at the local scale, for individual polymer chains belonging to the network, the average end-to-end distance varies as $R_{ee} \sim c^{-0.54\pm0.01}$, as shown in Fig. 1(e). This behavior is relatively close to the Flory scaling for polymer chains in good solvent,[17] with a minor deviation

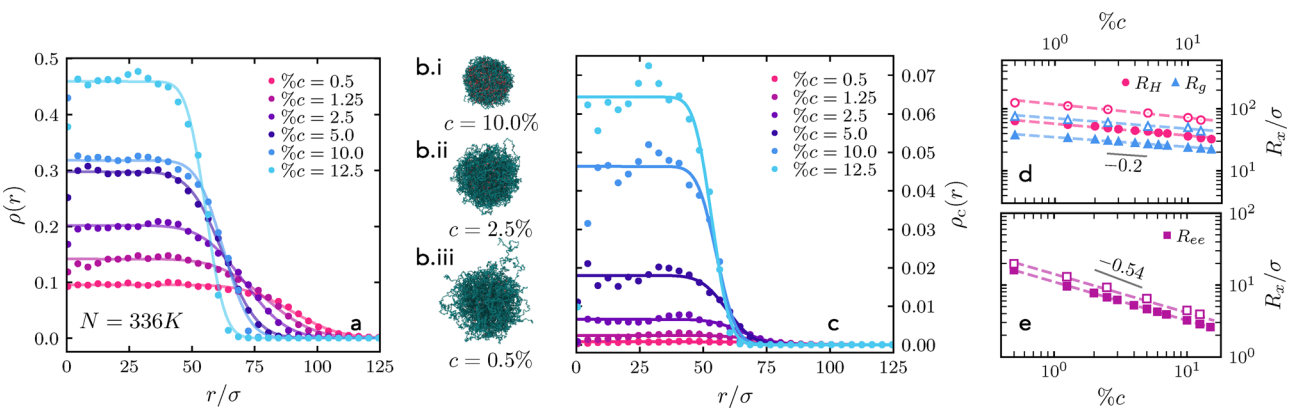

Fig. 1 Crosslinker concentration and structural effects. (a) Averaged radial density profiles of microgels with $N \sim 336\,000$ beads for various crosslinker molar fractions c, obtained from simulations (symbols) and fuzzy sphere fits (lines); (b) snapshots of high (i), medium (ii) and low (iii) crosslinked microgels; (c) corresponding crosslinker radial density profiles, $\rho_c(r)$, from simulations (symbols) for different c values. Lines correspond to fuzzy sphere fits (Table S1); (d) crosslinker concentration c dependence of characteristic microgel lengths (R_x): hydrodynamic radius R_H (circles), gyration radius R_g (triangles) and (e) end-to-end distance R_{ee} (squares) for microgels with $N \sim 42\,000$ (fill symbols) and $N \sim 336\,000$ (open symbols) beads. Dashed lines are the corresponding power-law fits.

likely attributed to the effect of the crosslinkers. Both scaling behaviors and related exponents are essentially independent of the microgel size, as expected.

3.2 Crosslinker concentration prediction

The previous results suggest that the knowledge of the crosslinker concentration is encoded in some of the microgel features, in particular, in the total density profile of the microgels. Although c can be estimated by fitting the fuzzy-sphere parameters as a function of c, as described in detail in the SI (see Section S2), the resulting fit parameters do not follow the expected scaling behavior for microgels, namely, the Flory or Flory–Rehner relations. Moreover, this approach is limited to systems that can be adequately described by the fuzzy-sphere model, making the use of this approach specific to the model itself. To overcome this problem and to establish a more general and transferable method, we thus propose to use unsupervised ML by means of autoencoders (AE) to directly extract c from $\rho(r)$. This nonlinear dimensionality-reduction method consists of a neural network that encodes a high-dimensional input and projects it into a low-dimensional latent space, from which it then reconstructs the original input. The key information that distinguishes the initial input data is presumably conserved in the latent space, and hence, we expect this to correspond to the crosslinker concentration.

To train the AE, we employ discretized density profiles obtained from simulations of microgels with varying c values. The input of the AE is a vector $\mathbf{P}$ of dimension $d = 115$, where each entry corresponds to the density profile value at a given radial distance. To make the method scalable and transferable to microgels of different sizes, we preprocess the input data by rescaling the radial coordinate as described in the Model and methods section. This allows the method to be applied to any microgel, whose density profile is known, including experimental ones. Furthermore, to provide sufficient data and variability to the AE, we employ averaged radial density profiles in small windows of time. We use $c = 1.25$, 2.5, 5.0, and 12.5% as a training set, while the simulations for additional c values are later used as test data sets. The encoder consists of 80 neurons, followed by the bottleneck that projects the data onto a d'-dimensional space.

We first determine the minimum number of dimensions to be used in the bottleneck d'. To this aim, we train the AE for at least 3×10^3 epochs, with varying d' and then we compute the final mean squared error (MSE) and the fraction of variance explained (FVE) as defined in the Methods section. The resulting FVE is reported in Fig. 2(a) as a function of d', showing that, even for $d' = 1$, the AE is already able to capture more than 99% of the variance of the input data by projecting it onto a single number. At higher dimensions, $d' \geq 2$, this improves, reaching a plateau of FVE $\approx$ 0.993. However, since the crosslinker concentration c is a scalar quantity, we choose to work with a bottleneck dimension of $d' = 1$ for its prediction. This choice simplifies the model and already yields very accurate results.

After training the entire AE, we discard the decoder and use the encoder to project the data into the latent space. We report in Fig. 2(b) the probability distributions $P(L)$ of the latent space

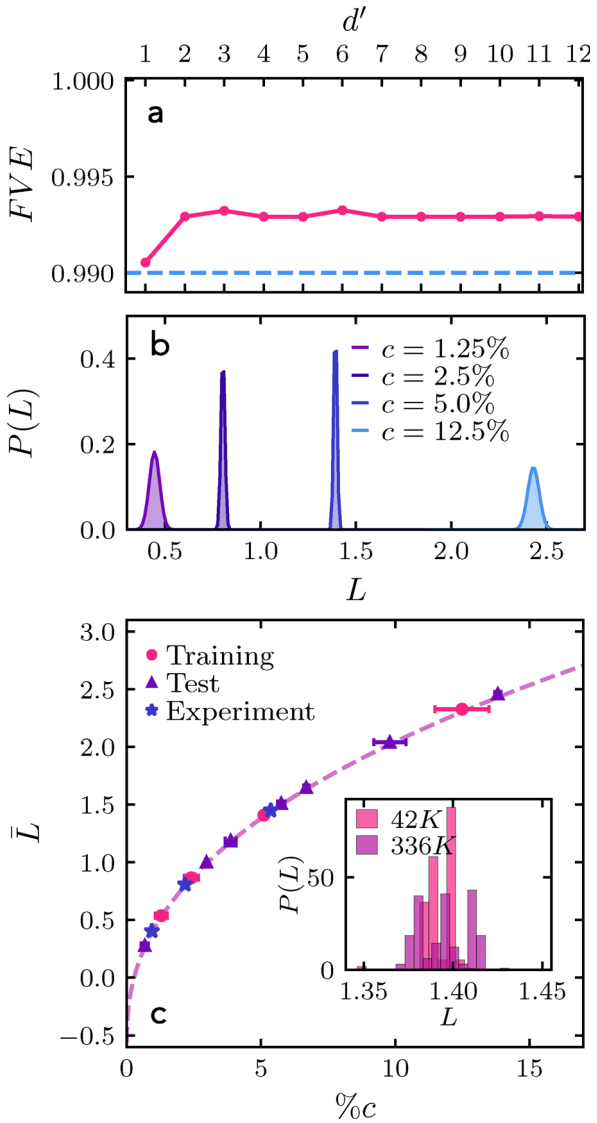

Fig. 2 Crosslinker concentration prediction. (a) Fraction of variance explained (FVE) for the autoencoder with varying dimensions d' of the latent space L. The horizontal line indicates FVE = 99%; (b) probability distributions of L for microgels with $N \sim 336\,000$ beads with different c values for the training data; (c) $\bar{L}$ as a function of c for all simulated microgels, including training (circles) and test (triangles) data sets and experiments from ref. 17, corresponding to microgels with c = 1.25, 2.5, and 5.0% (stars). The dashed line is a power-law fit. Inset: $P(L)$ for two different microgel sizes with c = 10.0%.

values L for different c values of the training data set with microgels consisting of $N \sim 336\,000$ beads. These results confirm that $d' = 1$ is already able to distinguish between microgels with different c values, since the distributions in the latent space are well separated, each following a normal distribution with a maximum standard deviation of 0.01. We also observe that the mean values of the distributions, $\bar{L}$, shown in Fig. 2(c), as a function of c, are unevenly distributed in the latent space, following a power law relation with c, given by $\bar{L}(c) \sim A_0 c^{\nu}$, where $A_0 \sim 0.514$ and $\nu \sim 0.59$ are fit parameters. The latter is remarkably close to the Flory exponent for polymer

chains in good solvent. A similar value of ν is recovered when varying the number of neurons in the AE hidden layer. Although $\bar{L}$ does not correspond to any physical observable, this scaling behavior suggests that the encoder does not simply differentiate the samples, but rather captures the key structural changes associated with variations in c, consistent with theoretical expectations for polymer networks.

Overall, these results demonstrate that the AE effectively captures the underlying physics of the system, with information taken solely from the density profile. Finally, the crosslinker concentration c can be easily predicted for any microgel by inverting $\bar{L}(c) \sim A_0 c^{\nu}$ and using the projection on the latent space of a given density profile.

We report in Fig. 2(c) the latent values as a function of the predicted c of all simulated microgels, including the ones outside the training. The AE predictions show remarkable agreement with the expected c-values of training and test data sets, with a confidence value ranging from $\pm 1\%$ to $\pm 10\%$. We further validate the method by applying it to microgels of different sizes. Apart from a small shift, the latent space values for the two microgel sizes are found to be very similar, as shown in the inset of Fig. 2(c). In addition, the AE can also be directly applied to experimental data to estimate c. We show in Fig. 2(c) the predicted c values for the experimental data from ref. 17. Despite using a single density profile obtained from the fit of the form factor fuzzy sphere,[17] the predictions fall within an error margin of ± 0.3. Finally, when applying the AE trained with $\alpha = 0$ to larger values of α (see the SI, Fig. S4), the latent-space values follow trends similar to those of the swelling behavior, discussed in following sections. All these results further demonstrate the robustness and transferability of the approach.

3.3 Crosslinker distribution prediction

Once c is obtained, we focus on the prediction of the crosslinker density profiles $\rho_c(r)$, calculated directly from simulations and shown in Fig. 1(c). Although the overall shapes of $\rho_c(r)$ and the total monomer density profile $\rho(r)$ are similar, their relationship is non-linear, particularly in the decaying region corresponding to the corona, as shown in Fig. S2. Therefore, to perform this task, we employ a NN to learn $\rho_c(r)$ from the overall density profile. The same training and test data sets of the AE are used. As before, our input data set is obtained from the discretized total density profile with the r-axis normalized. The output now corresponds to the ρ_c. However, since this quantity is subject to large statistical noise due to the small number of involved crosslinkers, we use the fuzzy sphere fits of ρ_c, averaged over different microgel configurations, for each c. We normalize the radial axis and the y-axis by the same factor of the total density profile and by the crosslinker concentration c, respectively. The chosen NN has 1 hidden layer with 96 neurons. To train the NN, we minimize the error function between the target ρ_c and the NN output defined in eqn (8). The input dataset includes data from microgels assembled with crosslinker concentrations c = 1.25, 2.5, 5.0, and 10.0%, with the additional simulated values of c used as test data sets. In the same spirit of the previous section, we add variability to the input dataset by using short-time-average density profiles as input.

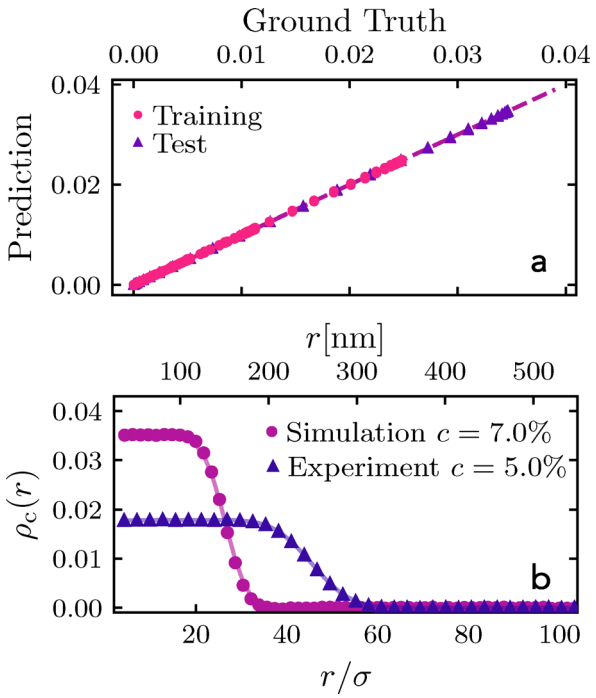

Fig. 3 Crosslinker distribution prediction: (a) parity plot of target and predicted $\rho_c(r)$ values for the training set corresponding to c = 1.25, 2.5, 5.0, and 10.0% (circles) and the test set (triangles); (b) predicted $\rho_c(r)$ values outside the training set for simulations with c = 7.0% (circles) and experimental data with c = 5.0% (triangles) from ref. 17; lines are fuzzy sphere fits.

The parity plot between the ground truth and the predicted $\rho_c(r)$ is shown in Fig. 3(a), where we observe remarkably good predictions for the training and also test data sets. This is confirmed by directly comparing the ML-predicted $\rho_c(r)$ for a c = 7.0% microgel outside the training set with the one obtained from the simulations in Fig. 3(b).

We further test our method by applying it to the c = 5.0% experimental data taken from ref. 17, whose ML-predicted $\rho_c(r)$ is also reported in Fig. 3(b). In both cases, the predicted profiles are in excellent agreement with those obtained by simulations, confirming the predictive capability of our approach.

3.4 Swelling behavior prediction

Our final goal is to predict the swelling behavior of a microgel, given the ML-predicted crosslinker concentration. To do so, we first simulate the various microgels with different c values at many different temperatures, by employing the solvophobic potential U_α in eqn (3)[16] (see the Model and methods section). We then calculate the swelling ratio S from eqn (7) and report it in Fig. 4(a) as a function of c for different values of α for microgels with $N \sim 42\,000$ beads. We then fit the dependence of S on c with a generalized power law as

$$S(\alpha, c) = \frac{R_H}{R_{H0}} = h(\alpha) c^{f(\alpha)}, \tag{10}$$

where R_{H0} corresponds to the hydrodynamic radius of the

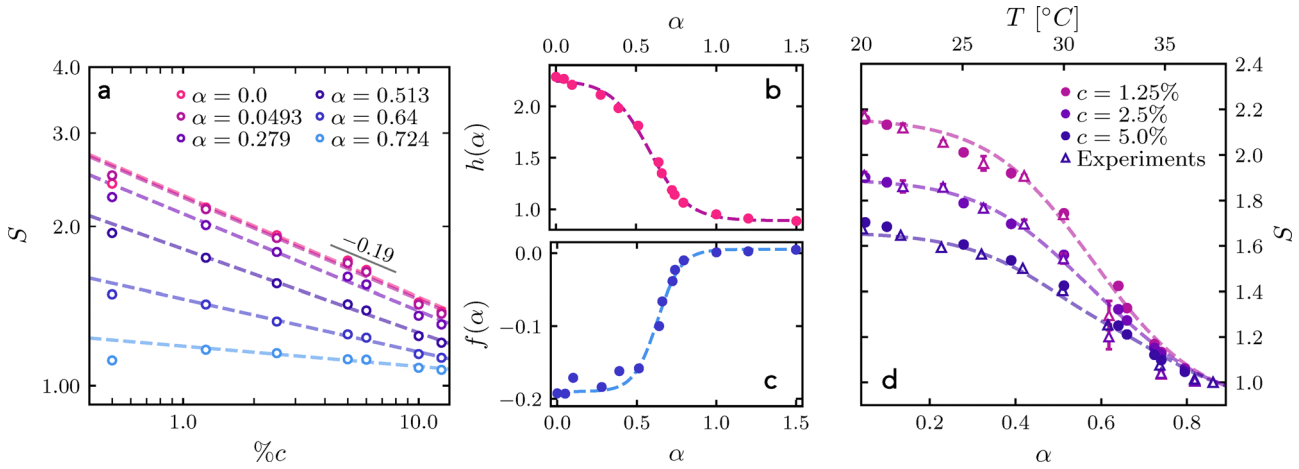

Fig. 4 Swelling behavior: (a) swelling ratio S as a function of crosslinker molar concentration c for various α values, ranging from $\alpha = 0$, the swollen state, to $\alpha = 0.724$. Dashed lines are power-law fits with eqn (10), determining $h(\alpha)$ and $f(\alpha)$, reported in (b) and (c) as a function of α. (d) S as a function for α for various values of c from simulations (circles), experimental data from ref. 17 (triangles) and predictions from eqn (10) and (11) (dashed lines).

collapsed microgel. The functions $h(\alpha)$ and $f(\alpha)$ solely depend on α and are reported in Fig. 4(b) and (c), respectively. The α-dependence of both functions can be written as

$$F(\alpha) \sim A_0^X + A_1^X \zeta((\alpha + A_2^X)/A_3^X), \qquad (11)$$

where $\zeta(x) = 1/(1 + e^x)$ is a logistic function, X refers to either f or h and A_i^X are fit parameters associated with each function, reported in Table S3 of the SI. We find that the physics of the VPT is directly encoded in these parameters for $f(\alpha)$,with $A_1^f \sim 0.19$ being the expected Flory–Rehner exponent at $\alpha = 0.0$ and $A_2^f \sim -0.63$ coinciding with the α value at which the VPT occurs. A similar value is also obtained for A_2^h.

Once we know the dependence of h and f on temperature and c, we can finally reconstruct the full swelling behavior of any standard pNIPAM microgel. To test this hypothesis, we show the predictions of eqn (10) and (11) in Fig. 4(d) for different c values in comparison to simulation results and to the experimental data of ref. 17. In particular, to enable a proper comparison, the relationship between temperature and α reported in ref. 10 and 17 is used, as shown in Section S4 of the SI. Overall, we find very good agreement at all temperatures, with minor deviations close to the VPT, where experimental error bars are largest. Interestingly, once $S(\alpha)$ is known, for example, from dynamic light scattering (DLS) measurements, it can be used backwards to estimate c using eqn (10) and then obtain an approximate density profile, if not available experimentally, by using our database, following the mapping described in the SI (Section S6).

4 Conclusions

In this work, we established a novel ML approach that, through the sole knowledge of the total density profile of any standard microgel, is able to predict its crosslinker concentration and crosslinker distribution and the full swelling curve.

While the monomer density profile of a microgel can be directly or indirectly obtained from experiments, crosslinker profiles remain experimentally challenging. Techniques such as contrast-variation SANS with deuterated monomers[34,35] could, in principle, allow their determination. However, to our knowledge, such measurements have not yet been reported, and careful experimental design is required to be able to detect the weak crosslinker signal arising from their low concentration. Alternative super-resolution microscopy measurements, where crosslinkers are attached to fluorophores, could be employed, though they would present comparable experimental challenges. The present study exploits neural networks to predict them for the first time also for experimental systems. While we have validated the approach against the corresponding numerical profiles, it remains a challenge to detect them also experimentally to verify whether they also satisfy a fuzzy sphere decay, as shown in the present simulations.

Our work further highlights the power of machine learning techniques, trained on carefully selected data obtained from simulations, to uncover hidden relations. Specifically, by using autoencoders (AEs), we extract the information regarding the crosslinker concentration encoded in the density profiles. The associated latent space is found to follow a power-law dependence on c with an exponent of $\sim$0.59, remarkably close to the Flory prediction for polymers in a good solvent, thereby demonstrating the ability of the AE to correctly recover the underlying physics of the system. Since the overall structure of the microgels was shown to be conserved across different length scales,[10] the proposed methodology is robust against size variations. This structural invariance, combined with the use of radially normalized profiles, ensures that the predicted features remain robust and transferable, allowing the method to be applied to experimental microgels of different sizes, including smaller ones with a size of $\lesssim$100 nm.

In addition, our approach can be readily extended to different microgels, either of different topologies, *e.g.*, hollow ones,[36] or obtained through a different synthetic protocol, amounting to networks with a different internal structure. We will pursue such generalization of the method in the future.

Finally, we can also foresee to use the method to calculate other microgel properties, such as elasticity or charge

distribution, providing sufficient training data. Similarly, it could be applied to assess microgel deformation at interfaces as a function of crosslinker concentration.[11]

Overall, our results show that machine learning models trained on monomer-resolved simulations can successfully predict key structural features of microgels across a wide range of synthesis conditions. The quality of the approach will be established by its employment against different sets of experimental systems, in order to establish the accuracy of the predictions. We hope that experimental groups will take advantage of our model and assess its predictive capability. In parallel, future work will be devoted to leveraging similar ML approaches for the inverse design of microgel architectures with tailored swelling ratios and density profiles, optimized for specific applications.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article are available at Zenodo at **https://doi.org/10.5281/zenodo.17376766**.

Supplementary information is available. See DOI: **https://doi.org/10.1039/d5sm00852b**.

Acknowledgements

We thank L. Tavagnacco for useful discussions. This project was funded by the European Union HORIZON-MSCA-2022-Postdoctoral Fellowships under grant agreement no. 101106848, MGELS. EZ also acknowledges funding from ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU – PNRR, Missione 4 Componente 2 Investimento 1.4. We gratefully acknowledge the CINECA award under the ISCRA initiative for providing high-performance computing resources and support.

Notes and references

1 M. Karg, A. Pich, T. Hellweg, T. Hoare, L. A. Lyon, J. Crassous, D. Suzuki, R. A. Gumerov, S. Schneider and I. I. Potemkin, *et al.*, *Langmuir*, 2019, **35**, 6231–6255.
2 A. Scotti, M. F. Schulte, C. G. Lopez, J. J. Crassous, S. Bochenek and W. Richtering, *Chem. Rev.*, 2022, **122**, 11675–11700.
3 M. Otten, M. Hildebrandt, B. Pfeffing, V. C. Voigt, F. Scheffold, T. Hellweg and M. Karg, *Langmuir*, 2024, **40**, 14515–14526.
4 G. Del Monte, D. Truzzolillo, F. Camerin, A. Ninarello, E. Chauveau, L. Tavagnacco, N. Gnan, L. Rovigatti, S. Sennato and E. Zaccarelli, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2109560118.
5 G. M. Conley, S. Nöjd, M. Braibanti, P. Schurtenberger and F. Scheffold, *Colloids Surf., A*, 2016, **499**, 18–23.
6 M. Schelling and J.-M. Meijer, *Phys. Rev. E*, 2024, **109**, L062601.
7 J. Appel, B. Fölker and J. Sprakel, *Soft Matter*, 2016, **12**, 2515–2522.
8 P. J. Flory, *Principles of polymer chemistry*, Cornell University Press, 1953.
9 C. G. Lopez and W. Richtering, *Soft Matter*, 2017, **13**, 8271–8280.
10 A. Ninarello, J. J. Crassous, D. Paloli, F. Camerin, N. Gnan, L. Rovigatti, P. Schurtenberger and E. Zaccarelli, *Macromolecules*, 2019, **52**, 7584–7592.
11 F. Camerin, M. A. Fernández-Rodrguez, L. Rovigatti, M.-N. Antonopoulou, N. Gnan, A. Ninarello, L. Isa and E. Zaccarelli, *ACS Nano*, 2019, **13**, 4548–4559.
12 F. Sammüller and M. Schmidt, *J. Phys.: Condens. Matter*, 2024, **37**, 083001.
13 F. Sammüller, S. Robitschko, S. Hermann and M. Schmidt, *Phys. Rev. Lett.*, 2024, **133**, 098201.
14 M. Stieger, W. Richtering, J. S. Pedersen and P. Lindner, *J. Chem. Phys.*, 2004, **120**, 6197–6206.
15 S. Bergmann, O. Wrede, T. Huser and T. Hellweg, *Phys. Chem. Chem. Phys.*, 2018, **20**, 5074–5083.
16 N. Gnan, L. Rovigatti, M. Bergman and E. Zaccarelli, *Macromolecules*, 2017, **50**, 8777–8786.
17 N. Hazra, A. Ninarello, A. Scotti, J. E. Houston, P. Mota-Santiago, E. Zaccarelli and J. J. Crassous, *Macromolecules*, 2024, **57**, 339–355.
18 E. Poppleton, M. Matthies, D. Mandal, F. Romano, P. Šulc and L. Rovigatti, *J. Open Source Software*, 2023, **8**, 4693.
19 K. Kremer and G. S. Grest, *J. Chem. Phys.*, 1990, **92**, 5057–5086.
20 T. Soddemann, B. Dünweg and K. Kremer, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2001, **6**, 409–419.
21 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. In't Veld, A. Kohlmeyer, S. G. Moore and T. D. Nguyen, *et al.*, *Comput. Phys. Commun.*, 2022, **271**, 108171.
22 F. Scheffold, *Soft Matter*, 2024, **20**, 8181–8184.
23 J. B. Hubbard and J. F. Douglas, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1993, **47**, R2983.
24 M. Tschannen, O. Bachem and M. Lucic, *arXiv*, 2018, preprint, arXiv:1812.05069, DOI: **10.48550/arXiv.1812.05069**.
25 E. Boattini, S. Marn-Aguilar, S. Mitra, G. Foffi, F. Smallenburg and L. Filion, *Nat. Commun.*, 2020, **11**, 5479.
26 K. Gurney, *An introduction to neural networks*, CRC Press, 2018.
27 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *Advances in Neural Information Processing Systems*, 2019.
28 S. Sharma, S. Sharma and A. Athaiya, *Towards Data Sci*, 2017, **6**, 310–316.

29 C. Bishop, *Neural Networks for Pattern Recognition*, New York, USA, 1995.
30 D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, 1986, **323**, 533–536.
31 I. Sutskever, J. Martens, G. Dahl and G. Hinton, *International conference on machine learning*, 2013, pp. 1139–1147.
32 J. A. Paulin, J. E. Lopez-Aguilar, B. Fouconnier, R. O. Vargas and F. Lopez-Serrano, *Polym. Bull.*, 2022, **79**, 6709–6732.
33 A. Fernández-Barbero, I. J. Suárez, B. Sierra-Martn, A. Fernández-Nieves, F. J. de Las Nieves, M. Marquez, J. Rubio-Retama and E. López-Cabarcos, *Adv. Colloid Interface Sci.*, 2009, **147**, 88–108.
34 M. Cors, L. Wiehemeier, O. Wrede, A. Feoktystov, F. Cousin, T. Hellweg and J. Oberdisse, *Soft Matter*, 2020, **16**, 1922–1930.
35 M. Brugnoni, A. C. Nickel, L. C. Kröger, A. Scotti, A. Pich, K. Leonhard and W. Richtering, *Polym. Chem.*, 2019, **10**, 2397–2405.
36 L. Rank and E. Zaccarelli, *Soft Matter*, 2025, **21**, 3979–3990.