

Automated matching of two-time X-ray photon correlation maps from phase-separating proteins with Cahn–Hilliard-type simulations using auto-encoder networks

Sonja Timmermann,^{a*} Vladimir Starostin,^b Anita Girelli,^b Anastasia Ragulskaya,^b Hendrik Rahmann,^a Mario Reiser,^c Nafisa Begam,^b Lisa Randolph,^a Michael Sprung,^d Fabian Westermeier,^d Fajun Zhang,^b Frank Schreiber^b and Christian Gutt^a

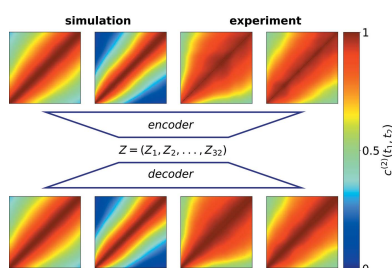
^aDepartment Physik, Universität Siegen, Walter-Flex-Strasse 3, 57072 Siegen, Germany, ^bInstitut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany, ^cDepartment of Physics, AlbaNova University Center, Stockholm University, SE-106 91 Stockholm, Sweden, and ^dDeutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany. *Correspondence e-mail: sonja.timmermann@uni-siegen.de

Machine learning methods are used for an automated classification of experimental two-time X-ray photon correlation maps from an arrested liquid–liquid phase separation of a protein solution. The correlation maps are matched with correlation maps generated with Cahn–Hilliard-type simulations of liquid–liquid phase separations according to two simulation parameters and in the last step interpreted in the framework of the simulation. The matching routine employs an auto-encoder network and a differential evolution based algorithm. The method presented here is a first step towards handling large amounts of dynamic data measured at high-brilliance synchrotron and X-ray free-electron laser sources, facilitating fast comparison with phase field models of phase separation.

1. Introduction

X-ray photon correlation spectroscopy (XPCS) is an experimental method capable of accessing the dynamics of protein systems over length scales that range from the atomic to micrometre scale, and timescales from microseconds to hours (X. Lu *et al.*, 2008; Zhang *et al.*, 2011, 2017; Madsen *et al.*, 2016; Möller *et al.*, 2016; Zinn *et al.*, 2018; Perakis & Gutt, 2020; Ruta *et al.*, 2012; Begam *et al.*, 2021). In XPCS, time series of coherent X-ray speckle patterns are measured, giving access to dynamical properties via time-resolved correlation maps – the two-time correlation function (TTC) (Brown *et al.*, 1997; Bikondoa, 2017). With fast megapixel 2D X-ray detectors, XPCS experiments can produce large quantities of data in a short time interval with up to thousands of TTCs per hour of beamtime. Given the typical duration of a synchrotron experiment of a few days, the resulting quantities of TTCs are difficult to handle. Therefore, there is a generic need for methods facilitating fast and reliable analysis including classification of the data. This is especially important with regards to steering, selecting and controlling the experimental parameters during the beamtime. However, in the aftermath of the experiment, quick classification methods are important for benchmarking models which, for example, simulate a gel transition or the solidification process of a protein solution upon liquid–liquid phase separation (LLPS).

Phase separation is a ubiquitous process in nature, with applications and consequences for a wide range of scientific



disciplines such as solid-state physics, material sciences and biology. Phase separation in biological systems attracted considerable attention with the discovery that LLPS in protein solutions (Ishimoto & Tanaka, 1977; Schurtenberger *et al.*, 1989; Broide *et al.*, 1991; Berland *et al.*, 1992; Muschol & Rosenberger, 1997) constitutes a possible pathway for organizing membraneless structures in living cells (Brangwynne *et al.*, 2009; Shin & Brangwynne, 2017; Berry *et al.*, 2018). Detailed investigations were carried out with respect to the biological functions of these protein condensates, encompassing biochemical reaction rates, buffering protein concentrations, and sensing or signaling (Shin & Brangwynne, 2017). A variety of diseases which result from a loss or change of function of these condensates are accompanied by phase transitions (Malinowska *et al.*, 2013; Weber & Brangwynne, 2012).

Phase separation is initiated by quenching a system into the metastable region of the phase diagram, launching a process of out-of-equilibrium self-organization (Dong & Granick, 2021). The state of the condensates depends on the dynamic and kinetic processes during their formation, with dynamical asymmetries between the two phases on a hierarchy of length and timescales and invoking viscoelastic properties of the resulting network structures (Berry *et al.*, 2018; Zaccarelli, 2007; Tanaka, 2000). In the biological context, an understanding of the component timescales is important as they are expected to meet the intrinsic timescales of biochemical processes in the condensates.

Upon phase separation, the dynamics slow down on molecular length scales due to local concentration changes. This microscopic deceleration can eventually lead to the arrest of phase separation on larger length scales complemented by the formation of bicontinuous gel network structures – as observed in colloidal and protein systems (Manley *et al.*, 2005; Conrad *et al.*, 2010; P. J. Lu *et al.*, 2008). Employing time-resolved scattering experiments, the kinetics during arrested phase separations have been observed frequently as a slow-down of the growth of the static structure factor $S(Q)$ in both Q position and peak height for low-temperature quenches (Gibaud & Schurtenberger, 2009; Gibaud *et al.*, 2011; Cardinaux *et al.*, 2007; Da Vela *et al.*, 2016, 2017, 2020; Bucciarelli *et al.*, 2015). In contrast, the dynamics of protein solutions evolving into an arrested phase transition are still poorly understood as studies require the concurrent monitoring of an extraordinarily broad range of length and timescales. Likewise, experimental validation of models of the dynamics of critical phenomena during LLPS like the Cahn–Hilliard equation (CHE) (Cahn, 1965) and other correlated models is still needed, particularly with respect to glass–gel transitions displaying considerable dynamical asymmetries between the dilute and concentrated phase (Berry *et al.*, 2018; Girelli *et al.*, 2021; Ragulskaya *et al.*, 2021).

Machine learning (ML) has already proven to be useful in the analysis of various structural X-ray data from small-angle X-ray scattering (Archibald *et al.*, 2020; Chen & Pollack, 2020; Franke *et al.*, 2018; Wang *et al.*, 2017) and wide-angle X-ray scattering (Chen & Pollack, 2020; Wang *et al.*, 2017), diffrac-

tion (Berntson *et al.*, 2003; Oviedo *et al.*, 2019; Vecsei *et al.*, 2019), and reflectometry (Greco *et al.*, 2019) experiments. The algorithms provide real-time feedback (Wang *et al.*, 2017; Ke *et al.*, 2018; Greco *et al.*, 2019) during the experiment or identify and remove bad images to reduce the stored data volume (Wang *et al.*, 2017; Ke *et al.*, 2018). In many cases the ML algorithms have been trained with synthetic structural data generated from existing databases on proteins (Franke *et al.*, 2018; Archibald *et al.*, 2020), RNA (Chen & Pollack, 2020) or crystal structures (Oviedo *et al.*, 2019; Vecsei *et al.*, 2019), which enables them to classify the results obtained in experiments without an expert spending time on labeling large quantities of data for training (Berntson *et al.*, 2003). Recently, ML algorithms have also been trained to reduce the noise of experimental TTCs (Konstantinova *et al.*, 2021). The CHE has also been used in different ML tasks (Wight & Zhao, 2021; Pokuri *et al.*, 2019; Zhang & Garikipati, 2020; Farimani *et al.*, 2018).

The focus of X-ray scattering applications of ML has been mainly on static properties up to now. We note, however, that many processes in nature, such as the LLPS of protein solutions, display dynamic phenomena which require the development of schemes and methods capable of classifying the corresponding dynamic X-ray signatures of the processes. ML-based classification of process data evolving in time is computationally very expensive due to the large quantity of training data required. Employing a minimum 2D model which allows a network with finite computer resources to be trained, we report on a first step towards this goal and present a neural network method for analysis and classification of dynamical X-ray information during LLPS.

The CHE in a simplified 2D version still captures the essence of the physics involved (Sciortino *et al.*, 1993; Sappelt & Jäckle, 1997; Lamorgese & Mauri, 2009), while allowing us to generate a large number of simulations on the timescale of hours, and couple it with a concentration-induced gelation as proposed by Sciortino *et al.* (1993) and Sappelt & Jäckle (1997). We aim for an automated assignment of the measured TTCs in the work of Girelli *et al.* (2021) with TTCs from simulations that differ in the parameters ϵ , *i.e.* proportional to reduced quench depth, and Ψ_{gel} , *i.e.* the order parameter at which the mobility drops by a factor of 1/2. We train an auto-encoder neural network with TTC simulation data and employ a differential evolution based algorithm for matching encoded experimental TTCs with simulations. Our work constitutes a first step to making use of the static, kinetic and dynamical information contained in XPCS data for ML-based analysis of processes in phase separation.

2. Analysis and results

2.1. Simulation of the Cahn–Hilliard equation

We model the dynamic processes during the spinodal decomposition with the help of the CHE (Sciortino *et al.*, 1993; Gunton *et al.*, 1983),

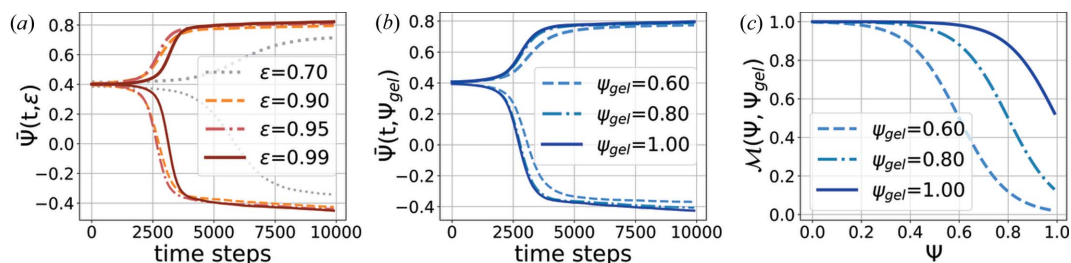


Figure 1

(a), (b) Order parameter averaged over all lattice sites having an order parameter higher/lower than the initial order parameter ($\Psi = 0.4$) as a function of time. In (a), Ψ_{gel} was fixed to $\Psi_{\text{gel}} = 1$ and ϵ was varied. For $\epsilon = 0.7$, the phase separation is much slower and the concentration of the dense phase is lower at the end of the chosen time window. This makes it hard to see the effects of high Ψ_{gel} , which is why ϵ is restricted to $[0.88, 0.99]$ for the generation of the simulated training data. In (b), ϵ was fixed to $\epsilon = 0.9$ and Ψ_{gel} was varied. The phase separation starts later for larger values of ϵ due to the smaller initial fluctuations. For constant ϵ , the phase separations show similar behavior at early times but the droplet formation is faster for larger values of Ψ_{gel} . (c) Mobility as a function of the order parameter and the gelation point Ψ_{gel} marking the order parameter at which the mobility drops to 0.5.

$$\frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = \nabla \cdot M \nabla [\Psi^3(\mathbf{r}, t) - \epsilon \Psi(\mathbf{r}, t) - \nabla^2 \Psi(\mathbf{r}, t)]. \quad (1)$$

Here $\Psi(\mathbf{r}, t)$ is the order parameter at a given time t and position \mathbf{r} which is related to the local protein concentration Φ by $\Phi = (\Psi + 1)/2$ (Sciortino *et al.*, 1993). The symmetry breaking is introduced by the parameter ϵ which can be identified as being proportional to the reduced quench depth $\epsilon \propto (T_c - T)/T_c$, where T_c is the upper critical temperature of the system. If ϵ fulfills the condition $\epsilon > 3\Psi^2$ (e.g. Gunton *et al.*, 1983), the system is in an unstable state below the spinodal curve and separates into regions with higher and lower local protein concentrations [Figs. 1(a) and 1(b)].

The phase separation by spinodal decomposition is coupled to a gelation process by introducing a concentration-dependent mobility parameter M in equation (1) (Sappelt & Jäckle, 1997) which reduces the mobility of the highly concentrated phase via

$$M(\Psi) = \frac{1}{1 + \exp[\alpha(\Psi - \Psi_{\text{gel}})]}. \quad (2)$$

In this model the mobility decreases from values $M = 1$ to $M = \frac{1}{2}$ when the order parameter Ψ exceeds the value of Ψ_{gel} [Fig. 1(c)]. The parameter α is fixed to $\alpha = 10$, providing a steep decrease of the mobility as required for an arrested phase separation (Sappelt & Jäckle, 1997).

We perform the simulation on a 2D grid of size 256×256 . The grid uses periodic boundary conditions and is initialized with a mean order parameter of $\Psi_0 = 0.4$ such that we observe droplets with a low concentration instead of an interconnected network, which would be the case for $\Psi_0 = 0$. The fluctuations necessary for LLPS are added as random noise with a maximum amplitude of $0.5(1 - \epsilon)$. Ten thousand time steps are calculated for each simulation with a fixed time step of $\Delta t = 0.03$. The temporal evolution of the order parameter for different parameter configurations is depicted in Figs. 1(a) and 1(b). Apart from the initial noise configuration, the simulations differ only in the values of the parameters ϵ and Ψ_{gel} . Both were systematically varied within the intervals

$$\epsilon \in [0.88, 0.99] \quad \text{and} \quad \Psi_{\text{gel}} \in [0.55, 1]. \quad (3)$$

These intervals were chosen with $\Psi_{\text{gel}} > \Psi_0$, thus avoiding an arrest of the phase separation before the formation of droplets. We restrict our analysis to large values of ϵ as for small values of ϵ the concentration of the dense phase is too low [gray line in Fig 1(a)] to introduce visible changes in the simulations for different values of Ψ_{gel} .

A single simulation run yields a time series of the protein density's real-space configuration $\Psi(\mathbf{r}, t)$ [Fig. 2(a)] which is converted to an X-ray speckle pattern by means of Fourier transform, yielding a 256×256 image. The temporal evolution of the square of the azimuthal integration of these Fourier images $|\Psi(R_Q, t)|^2$ is shown in Fig. 2(b) and is equivalent to the X-ray intensity. We introduce R_Q as the distance from the center of the Fourier image in units of pixels. In Fig. 2(b) the occurrence of a peak indicates the phase separation.

The dynamics on a certain length scale in Fourier space can be traced by correlating pixelwise the corresponding intensities with a TTC,

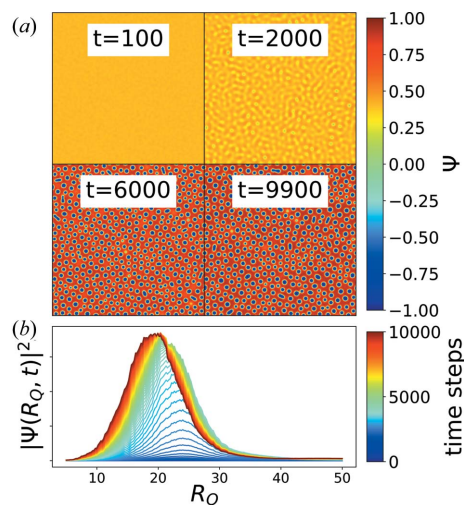


Figure 2

(a) Evolution of the real-space configuration of the sample on the 256×256 grid with simulation parameters $\Psi_{\text{gel}} = 0.6$ and $\epsilon = 0.955$, showing the creation and coarsening of droplets with a low concentration. (b) Corresponding temporal evolution of the azimuthally integrated scattering intensity as a function of the absolute value of the distance R_Q to the origin of the Fourier map. The peak in the intensity profile is shifting towards smaller R_Q with time, indicating the coarsening of the droplets.

$$c^{(2)}(t_1, t_2) = \frac{\langle [I(t_1) - \langle I(t_1) \rangle][I(t_2) - \langle I(t_2) \rangle] \rangle}{\{[\langle I^2(t_1) \rangle - \langle I(t_1) \rangle^2][\langle I^2(t_2) \rangle - \langle I(t_2) \rangle^2]\}^{1/2}}, \quad (4)$$

with $\langle \cdot \rangle$ being the average over pixels within a distance $[R_Q - dR_Q, R_Q + dR_Q]$ to the center.

The azimuthally integrated intensity profile [Fig. 2(b)] shows the growth dynamics of the droplets by a shift of the peak position towards smaller values of R_Q . These dynamics can be observed in the TTC by choosing a value of R_Q such that it is located at the falling slopes of the peaks in Fig. 2(b), which is $R_Q \in [26, 45]$. For our training data, we chose R_Q to be an integer in this range, while dR_Q increases linearly from 0.7 at $R_Q = 26$ to 3 at $R_Q = 45$, as done in experimental analyses as well to compensate for lower photon counts at higher Q values.

2.2. Experimental data

The XPCS experiments (Fig. 3) were conducted at the Coherence Applications beamline P10 at PETRA III, DESY, employing an X-ray beam of photon energy 8.54 keV, size $100 \times 100 \mu\text{m}$ and maximum photon density $10^7 \text{ photons s}^{-1} \mu\text{m}^{-2}$ [for further experimental details, see Girelli *et al.* (2021)]. We measured time series of coherent diffraction patterns that were collected with an EIGER 4 Mpixel detector covering a Q range from 3 to $50 \mu\text{m}^{-1}$. The samples consisted of immunoglobulin G (IgG) with polyethylene glycol (PEG) and NaCl, and were quenched from 37°C to six different quench temperatures below the binodal line. Further details about the sample preparation can be found in the work of Da Vela *et al.* (2017).

In the experiment, 4000 frames were recorded for every quench temperature. The sample was illuminated for 0.02 s followed by 0.1 s where the shutter was closed such that the measurement covers a time range of $4000 \times 0.12 \text{ s} = 480 \text{ s}$. The experimental TTCs were calculated similarly to the TTCs from the simulation using equation (4). Regarding the temporal shift of the integrated intensity profiles as a function of Q , the falling edge can be restricted to $Q \in [7.5, 10] \mu\text{m}^{-1}$. The contrast of the experimental TTCs is lower than that of

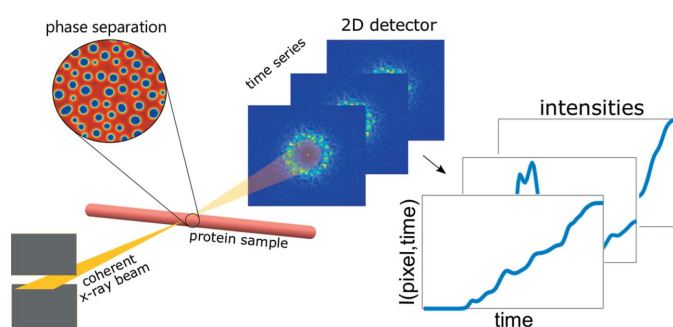


Figure 3 Schematic of the experimental setup. The coherent X-ray beam is scattered from the protein solution during the LLPS yielding a time series of X-ray speckle patterns. A constant Q ring is selected in the speckle patterns and the intensities are correlated pixel by pixel for different times.

the simulated ones where we assume completely coherent scattering and detection. This issue is addressed by normalizing the experimental TTCs to the speckle visibility which is determined by the values of the pixels adjacent to the diagonal of the experimental TTC, as the values directly on the diagonal are distorted by shot noise (Inoue *et al.*, 2012).

2.3. Data preparation and auto-encoding

For the training data, 12 000 pairs of the parameters ϵ and Ψ_{gel} were sampled from a uniform distribution within the intervals. To increase the simulation robustness against the random initial conditions, TTCs from five simulations with different initial noise were averaged. Thus, $5 \times 12\,000 = 60\,000$ simulations were performed in total, which took 7 h on two NVIDIA 2080 Ti graphic cards. TTCs were calculated for 20 R_Q ranges, so that the training data consisted of $20 \times 12\,000 = 240\,000$ TTCs.

All simulated TTCs display a broad, slow part in the beginning (Fig. 4) which is absent in the experimental data. This part is associated with the still homogeneous phase at early times before the onset of density differences [Fig. 2(a), $t = 2000$]. Approaching smaller values of R_Q this slow part becomes even more pronounced.

These broad parts are eliminated by cropping the TTCs. The maximum of the intensity in the associated Q rings is well defined for both simulated and experimental TTCs and is chosen to be the time where the TTCs are cropped. By doing this, the early part of the phase separation is removed from the TTC, such that only the coarsening stages of simulated and experimental TTCs are compared. The cutting procedure is illustrated in Fig. 4 for three simulated TTCs. For different models, this preprocessing via cropping needs to be adapted to remove, for example, well known deficiencies from simulated TTCs or make sure that the timescales of simulation and experiment are comparable. We derive the latter from the facts that the intensity curves show similar behavior within the coarsening stage for both experimental and simulated data and the TTCs show the same overall features in the whole

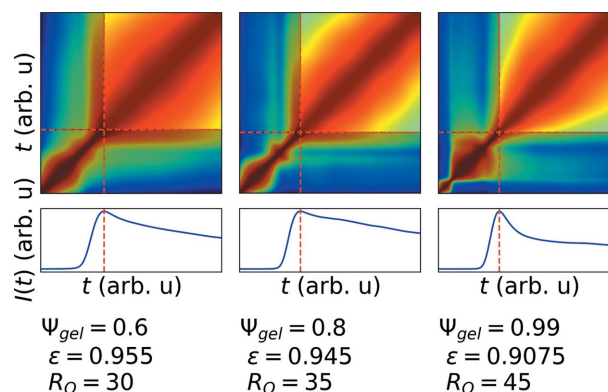


Figure 4 The simulated TTCs show a broad feature at early times (lower-left corner of the TTCs). In the second row, the mean intensity of the respective Q range is shown as a function of time. The maximum of this intensity curve defines the point where the TTCs are cropped.

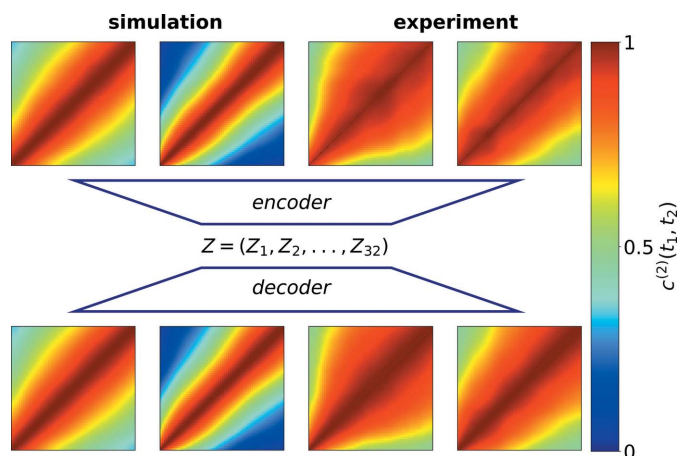


Figure 5
Performance of the auto-encoder on simulated and experimental TTCs. The encoder network encodes the TTCs into a 32-dimensional vector from which the decoder network restores the TTC. On the experimental data, this procedure reduces the experimental TTC to the main dynamic components, enabling a comparison with simulated TTCs.

measurement time (Girelli *et al.*, 2021). The cropped TTCs are finally scaled down to a 64×64 resolution.

Instead of matching raw TTCs between the experiment and the simulations, we compare their encoded representations via an auto-encoder (Rumelhart *et al.*, 1986). The auto-encoder network (Starostin, 2022) was trained with the simulated TTCs, encoding them in 32-dimensional vectors \mathbf{Z} . For the encoder, the first three layers of the pretrained ResNet-18 model (He *et al.*, 2016) were used as a feature extractor, followed by adaptive average pooling with an output size of 3×3 and a fully connected network. The decoder architecture follows the work of Hou *et al.* (2017).

The performance of the auto-encoder network on simulated and experimental data is shown in Fig. 5, from which one can see that the network is capable of reducing the experimental TTCs to the main dynamic components. Thus, the encoded representations allow comparison of TTCs in a more reliable and faster way. If a series of TTCs from the same simulation but at different values of R_Q is created, the entries of the \mathbf{Z} vector change continuously (Fig. 6) which enabled us to

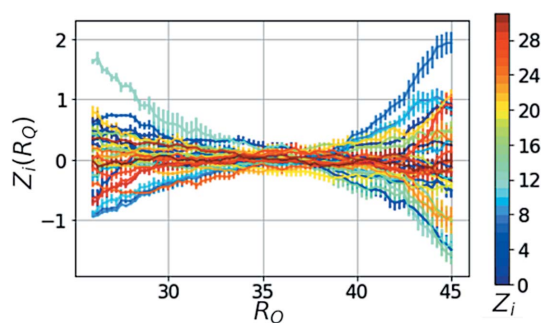


Figure 6
Example of encoded vector components' Z_i dependence on R_Q for a simulation with $\epsilon = 0.88$ and $\Psi_{\text{gel}} = 0.7$. The error bars result from averaging over ten simulations with different initial noise.

interpolate TTCs for arbitrary values of R_Q in between the selected values.

2.4. Results

Our goal is to develop tools for a fast classification of experimental TTCs in the context of different models. We approach this task via matching encoded TTCs from measurements at different quench temperatures with encoded TTCs from simulations of our simplified model. During this matching process, we find the quench temperature dependence of the two simulation parameters and the values of the three additional calibration parameters $t_{\text{low}}, t_{\text{up}}, k_r$, where t_{low} and t_{up} are the times at which the experimental TTCs are cropped such that they match the timescale of the simulation. k_r denotes the factor for converting the Q range in units of pixels, which were used in the simulation, to units of inverse micrometre as they were used in the experiment via

$$R_Q(\text{pixels}) \times k_r = Q (\mu\text{m}^{-1}). \quad (5)$$

A differential evolution (Price, 1996) based algorithm was employed for this matching process. For each of the six experimental measurements, TTCs were calculated at five different Q values,

$$Q = 7.02, 7.35, 7.66, 7.96, 8.2 \mu\text{m}^{-1}, \quad (6)$$

such that we obtain a set of 30 experimental TTCs. The differential evolution algorithm optimizes the mean dot product D between the normalized encoded representations of 30 experimental TTCs and 30 simulated TTCs from the training data set with respect to $\epsilon, \Psi_{\text{gel}}, k_r, t_{\text{up}}$ and t_{low} :

$$D = \frac{1}{30} \sum_{q=1}^5 \sum_{T=1}^6 \mathbf{Z}_{\text{sim}}^q[\epsilon(T), \Psi_{\text{gel}}(T), k_r] \cdot \mathbf{Z}_{\text{exp}}^{T,q}(t_{\text{up}}, t_{\text{low}}). \quad (7)$$

The choice of boundaries for the calibration parameters was based on the assumption that the duration of the experiment is longer than that of the simulation, and that the chosen Q range $[7.02; 8.2] \mu\text{m}^{-1}$ lies within the used R_Q range for the simulations. Knowing that ϵ is proportional to $(T_c - T)/T_c$, we apply additional boundaries to consider only those solutions where the average value of ϵ is decreasing as a function of experimental quench temperature.

During the fitting procedure, the experimental TTCs were cropped and encoded for each generated combination of the parameters. A caching technique was used to accelerate these calculations. The encoded vectors of the simulated TTCs were taken from the training data and interpolated on the basis of the generated calibration coefficient k_r . After 1500 iterations the result of the fit converged to $D \simeq 0.72$.

Fig. 7(a) shows the averaged outcomes for the simulation parameters Ψ_{gel} and ϵ . The rise in Ψ_{gel} for shallower quenches indicates that the gelation is occurring at higher concentrations at these temperatures. The $(\Psi_{\text{gel}}, \epsilon)$ pairs were integrated into a phase diagram [Fig. 7(b)] which displays the spinodal and binodal lines as determined from the Landau free energy [equation (1)]. The predictions of the neural network enable us to estimate the corresponding gel line [line in the lower-

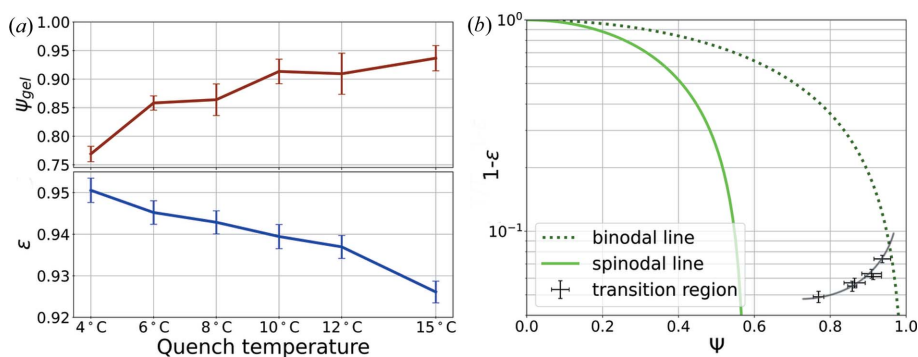


Figure 7 (a) Fitted parameters ϵ and Ψ_{gel} as a function of quench temperature. (b) Phase diagram derived from the free energy density used in the simulation displaying the spinodal and binodal line. The points represent the predictions of the neural networks for the experimental TTCs and mark the region of the gelation (guide to the eye) in the framework of the simulation.

right part of Fig. 7(b) as a guide to the eye]. This line bends towards the spinodal line, resulting in lower concentrations of the dense phase for deeper quenches. Such a behavior of the gel line in the coexistence region has also been observed for the LLPS of BSA-YCl₃ (Da Vela *et al.*, 2020), lysozyme (Cardinaux *et al.*, 2007) and γ -globulin (Da Vela *et al.*, 2017).

Averaging over the 30 outcomes for the calibration parameters yields

$$k_r = (0.185 \pm 0.002) \mu\text{m}^{-1} \text{ pixel}^{-1}, \quad (8)$$

$$t_{low} = (5.02 \pm 1.72) \text{ s}, \quad (9)$$

$$t_{up} = (353.3 \pm 5.7) \text{ s}. \quad (10)$$

A further validation of these results with respect to ϵ and Ψ_{gel} for the experimental TTCs is difficult as neither parameter is directly accessible in the experiment.

3. Discussion and conclusion

In conclusion, we successfully trained an auto-encoder neural network with TTC data originating from Cahn–Hilliard simulations. This auto-encoder was used to extract the main dynamic components out of TTCs originating from XPCS experiments of LLPS in a model protein solution of IgG. We used a differential evolution based algorithm for matching encoded TTCs from the experiment and the simulation and determine thereby the parameters necessary for connecting length and timescales of simulation to experiment. The simulation parameters of the classified experimental TTCs were used to construct a phase diagram with the experimental data indicating the position of a gel line.

The training of neural networks based on dynamical simulation data is challenging. The Cahn–Hilliard model employed here, including the gelation, is rather simplistic and uses two labeling parameters only. However, we emphasize the possibility of extending this methodology to more complicated models capable of making strong predictions and eventually also benchmarking of theoretical models based on very large XPCS data sets, possibly by including the kinetic information

in the model as well. This is a first step towards handling large amounts of XPCS data on short timescales and will be of general interest not only for protein dynamics but for material science aspects of spinodal decomposition.

Acknowledgements

ST and VS contributed equally to this work. Open access funding enabled and organized by Projekt DEAL.

Funding information

ST and CG acknowledge BMBF for financial support under grant Nos. 05K19PS1 and 05K20PSA. FZ and FS acknowledge BMBF for financial support under grant No. 05K20VTA. AR thanks the Studienstiftung des deutschen Volkes for a personal fellowship. NB thanks the Alexander von Humboldt Foundation for a postdoctoral research fellowship.

References

Archibald, R. K., Doucet, M., Johnston, T., Young, S. R., Yang, E. & Heller, W. T. (2020). *J. Appl. Cryst.* **53**, 326–334.
 Begam, N., Ragulskaia, A., Girelli, A., Rahmann, H., Chandran, S., Westermeier, F., Reiser, M., Sprung, M., Zhang, F., Gutt, C. & Schreiber, F. (2021). *Phys. Rev. Lett.* **126**, 098001.
 Berland, C. R., Thurston, G. M., Kondo, M., Broide, M. L., Pande, J., Ogun, O. & Benedek, G. B. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 1214–1218.
 Bertson, A., Stojanoff, V. & Takai, H. (2003). *J. Synchrotron Rad.* **10**, 445–449.
 Berry, J., Brangwynne, C. P. & Haataja, M. (2018). *Rep. Prog. Phys.* **81**, 046601.
 Bikondoa, O. (2017). *J. Appl. Cryst.* **50**, 357–368.
 Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoeg, C., Gharakhani, J., Jülicher, F. & Hyman, A. A. (2009). *Science*, **324**, 1729–1732.
 Broide, M. L., Berland, C. R., Pande, J., Ogun, O. O. & Benedek, G. B. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 5660–5664.
 Brown, G., Rikvold, P. A., Sutton, M. & Grant, M. (1997). *Phys. Rev. E*, **56**, 6601–6612.
 Bucciarelli, S., Casal-Dujat, L., De Michele, C., Sciortino, F., Dhont, J., Bergenholtz, J., Farago, B., Schurtenberger, P. & Stradner, A. (2015). *J. Phys. Chem. Lett.* **6**, 4470–4474.
 Cahn, J. W. (1965). *J. Chem. Phys.* **42**, 93–99.
 Cardinaux, F., Gibaud, T., Stradner, A. & Schurtenberger, P. (2007). *Phys. Rev. Lett.* **99**, 118301.
 Chen, Y.-L. & Pollack, L. (2020). *IUCrJ*, **7**, 870–880.
 Conrad, J. C., Wyss, H. M., Trappe, V., Manley, S., Miyazaki, K., Kaufman, L. J., Schofield, A. B., Reichman, D. R. & Weitz, D. A. (2010). *J. Rheol.* **54**, 421–438.
 Da Vela, S., Begam, N., Dyachok, D., Schäufele, R. S., Matsarskaia, O., Braun, M. K., Girelli, A., Ragulskaia, A., Mariani, A., Zhang, F. & Schreiber, F. (2020). *J. Phys. Chem. Lett.* **11**, 7273–7278.
 Da Vela, S., Braun, M. K., Dörr, A., Greco, A., Möller, J., Fu, Z., Zhang, F. & Schreiber, F. (2016). *Soft Matter*, **12**, 9334–9341.
 Da Vela, S., Exner, C., Schäufele, R. S., Möller, J., Fu, Z., Zhang, F. & Schreiber, F. (2017). *Soft Matter*, **13**, 8756–8765.
 Dong, R.-Y. & Granick, S. (2021). *Nat. Commun.* **12**, 911.

- Farimani, A., Gomes, J., Sharma, R., Lee, F. L. & Pande, V. (2018). arXiv:1803.08993.
- Franke, D., Jeffries, C. M. & Svergun, D. I. (2018). *Biophys. J.* **114**, 2485–2492.
- Gibaud, T., Cardinaux, F., Bergenholtz, J., Stradner, A. & Schurtenberger, P. (2011). *Soft Matter*, **7**, 857–860.
- Gibaud, T. & Schurtenberger, P. (2009). *J. Phys. Condens. Matter*, **21**, 322201.
- Girelli, A., Rahmann, H., Begam, N., Ragulskaya, A., Reiser, M., Chandran, S., Westermeier, F., Sprung, M., Zhang, F., Gutt, C. & Schreiber, F. (2021). *Phys. Rev. Lett.* **126**, 138004.
- Greco, A., Starostin, V., Karapanagiotis, C., Hinderhofer, A., Gerlach, A., Pithan, L., Liehr, S., Schreiber, F. & Kowarik, S. (2019). *J. Appl. Cryst.* **52**, 1342–1347.
- Gunton, J., San Miguel, M. & Sahni, P. (1983). *Phase Transitions and Critical Phenomena*, edited by C. Domb & J. Lebowitz, Vol. 8, 2nd ed., pp. 269–466. London: Academic Press.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778. IEEE.
- Hou, X., Shen, L., Sun, K. & Qiu, G. (2017). *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1133–1141. IEEE.
- Inoue, I., Shinohara, Y., Watanabe, A. & Amemiya, Y. (2012). *Opt. Express*, **20**, 26878–26887.
- Ishimoto, C. & Tanaka, T. (1977). *Phys. Rev. Lett.* **39**, 474–477.
- Ke, T.-W., Brewster, A. S., Yu, S. X., Ushizima, D., Yang, C. & Sauter, N. K. (2018). *J. Synchrotron Rad.* **25**, 655–670.
- Konstantinova, T., Wiegart, L., Rakitin, M., DeGennaro, A. M. & Barbour, A. M. (2021). *Sci. Rep.* **11**, 14756.
- Lamorgese, A. G. & Mauri, R. (2009). *Phys. Fluids*, **21**, 044107.
- Lu, P. J., Zaccarelli, E., Ciulla, F., Schofield, A. B., Sciortino, F. & Weitz, D. A. (2008). *Nature*, **453**, 499–503.
- Lu, X., Mochrie, S. G. J., Narayanan, S., Sandy, A. R. & Sprung, M. (2008). *Phys. Rev. Lett.* **100**, 045701.
- Madsen, A., Fluerasu, A. & Ruta, B. (2016). *Synchrotron Light Sources and Free-electron Lasers*, edited by E. J. Jaeschke, S. Khan, J. R. Schneider & J. B. Hastings, pp. 1617–1641. Cham: Springer International Publishing.
- Malinowska, L., Kroschwald, S. & Alberti, S. (2013). *Biochim. Biophys. Acta*, **1834**, 918–931.
- Manley, S., Wyss, H. M., Miyazaki, K., Conrad, J. C., Trappe, V., Kaufman, L. J., Reichman, D. R. & Weitz, D. A. (2005). *Phys. Rev. Lett.* **95**, 238302.
- Möller, J., Chushkin, Y., Prevost, S. & Narayanan, T. (2016). *J. Synchrotron Rad.* **23**, 929–936.
- Muschol, M. & Rosenberger, F. (1997). *J. Chem. Phys.* **107**, 1953–1962.
- Oviedo, F., Ren, Z., Sun, S., Settens, C., Liu, Z., Hartono, N. T. P., Ramasamy, S., DeCost, B. L., Tian, S. I. P., Romano, G., Gilad Kusne, A. & Buonassisi, T. (2019). *npj Comput. Mater.* **5**, 60.
- Perakis, F. & Gutt, C. (2020). *Phys. Chem. Chem. Phys.* **22**, 19443–19453.
- Pokuri, B. S. S., Ghosal, S., Kokate, A., Sarkar, S. & Ganapathysubramanian, B. (2019). *npj Comput. Mater.* **5**, 95.
- Price, K. (1996). *Proceedings of North American Fuzzy Information Processing*, pp. 524–527. IEEE.
- Ragulskaya, A., Begam, N., Girelli, A., Rahmann, H., Reiser, M., Westermeier, F., Sprung, M., Zhang, F., Gutt, C. & Schreiber, F. (2021). *J. Phys. Chem. Lett.* **12**, 7085–7090.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). *Learning Internal Representations by Error Propagation*, pp. 318–362. Cambridge: MIT Press.
- Ruta, B., Chushkin, Y., Monaco, G., Cipelletti, L., Pineda, E., Bruna, P., Giordano, V. M. & Gonzalez-Silveira, M. (2012). *Phys. Rev. Lett.* **109**, 165701.
- Sappelt, D. & Jäckle, J. (1997). *Physica A*, **240**, 453–479.
- Schurtenberger, P., Chamberlin, R. A., Thurston, G. M., Thomson, J. A. & Benedek, G. B. (1989). *Phys. Rev. Lett.* **63**, 2064–2067.
- Sciortino, F., Bansil, R., Stanley, H. E. & Alstrøm, P. (1993). *Phys. Rev. E*, **47**, 4615–4618.
- Shin, Y. & Brangwynne, C. P. (2017). *Science*, **357**, eaaf4382.
- Starostin, V. (2022). *XPCS Autoencoder*, https://github.com/schreiberlab/xpcs_autoencoder.
- Tanaka, H. (2000). *J. Phys. Condens. Matter*, **12**, R207–R264.
- Vecsei, P. M., Choo, K., Chang, J. & Neupert, T. (2019). *Phys. Rev. B*, **99**, 245120.
- Wang, B., Yager, K., Yu, D. & Hoai, M. (2017). *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 697–704. IEEE.
- Weber, S. C. & Brangwynne, C. P. (2012). *Cell*, **149**, 1188–1191.
- Wight, C. & Zhao, J. (2021). *Commun. Comput. Phys.* **29**, 930–954.
- Zaccarelli, E. (2007). *J. Phys. Condens. Matter*, **19**, 323101.
- Zhang, F., Allen, A. J., Levine, L. E., Ilavsky, J., Long, G. G. & Sandy, A. R. (2011). *J. Appl. Cryst.* **44**, 200–212.
- Zhang, Q., Bahadur, D., Dufresne, E. M., Grybos, P., Kmon, P., Leheny, R. L., Maj, P., Narayanan, S., Szczygiel, R., Ramakrishnan, S. & Sandy, A. (2017). *Phys. Rev. Lett.* **119**, 178006.
- Zhang, X. & Garikipati, K. (2020). *Comput. Methods Appl. Mech. Eng.* **372**, 113362.
- Zinn, T., Homs, A., Sharpnack, L., Tinti, G., Fröjdh, E., Douissard, P.-A., Kocsis, M., Möller, J., Chushkin, Y. & Narayanan, T. (2018). *J. Synchrotron Rad.* **25**, 1753–1759.