Optical spectroscopy of crystal nucleation one nucleus at a time

Oscar Urquidi[1,†], Johanna Brazard[1,†], Natalie LeMessurier[2], Lena Simine[2], Takuji B. M. Adachi[1,*]

Affiliations:

[1] Department of Physical Chemistry, Sciences II, University of Geneva, 30, Quai Ernest Ansermet, 1211 Geneva, Switzerland

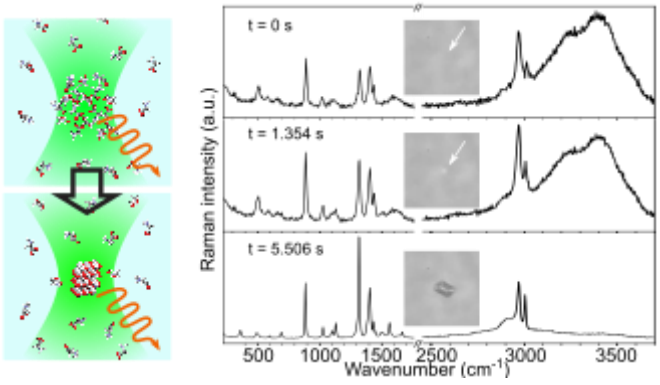
[2] Department of Chemistry, McGill University, 801 Sherbrooke Street W, Montreal, Quebec H3A 0B8, Canada

[*] Correspondence to: Takuji.Adachi@unige.ch

[†] These authors contributed equally to this work

Abstract

While crystallization is a ubiquitous and an important process, the microscopic picture of crystal nucleation is yet to be established. Recent studies agree that the nucleation process is more complex than the view offered by the classical nucleation theory. Here, we implement Single Nucleus Spectroscopy (SNS) by combining Raman micro-spectroscopy and optical trapping induced crystallization (OTIC) to spectroscopically investigate crystallization one nucleus at a time. Raman spectral evolution during a single glycine crystal nucleation from water, measured by SNS and analyzed by a non-supervised spectral decomposition technique, uncovered the Raman spectrum of pre-nucleation aggregates as well as their critical role as an intermediate species in the dynamics. The agreement between the SNS results and our simulation suggested that the structure of pre-nucleation aggregates is possibly loosely-bound hydrogen-bonded linear chains of glycine. The present work provides a strong impetus for accelerating the investigation of crystal nucleation by optical spectroscopy.



Crystallization is an important process in a wide range of disciplines from fundamental science to industrial application.[1–3] Despite the importance of controlling the crystallization and its morphology (*e.g.* polymorphism), the lack of microscopic description of crystal nucleation limits the rational approach to its engineering and control.[2] A major challenge in establishing the molecular level understanding of crystal nucleation is its stochastic and heterogeneous nature at the nanoscale.[1] This has long prevented experimentalists from directly observing the nucleation event in real-time. Statistical approaches, such as counting nucleation rate under various crystallization conditions, have been the major experimental methods to circumvent the difficulty.[1,4] The discrepancy between the nucleation rate obtained from the statistical experiments and the one predicted by classical nucleation theory (CNT) has led to a general agreement that crystal nucleation is a more complex phenomenon than how it is described by CNT.[4–7]

A real-time observation of nucleation dynamics has been long desired as a critical step in establishing its microscopic picture. For example, crystallization of colloidal particles has been used as a model system to mimic the dynamics of crystallization because the assembly process of particles can be easily visualized in real-time by optical microscopy.[8,9] A recent breakthrough in tackling this problem was the development of cryo- and liquid cell transmission electron microscopy (TEM) for visualizing crystallization of atomic/molecular systems.[10–12] The presence of morphologically ”featureless” or “amorphous” clusters was confirmed in these TEM studies,[13,14] and *in-situ* TEM results caught the moment at which clusters became well-defined crystals.[15–17] More and more studies show the involvement of pre-nucleation aggregates in crystal nucleation process and confirm the complexity of crystal nucleation process in contrast to CNT.[5–7,15,13,18] The next critical steps toward establishing a microscopic picture include the detailed understanding of the structure of pre-nucleation aggregates and how the crystalline order emerges from them.

Optical spectroscopy could be a powerful technique to extract the molecular level details of the structure of pre-nucleation aggregates and their structural dynamics toward the phase transition. The application of optical spectroscopy on crystal nucleation problem has been, however, limited because the stochastic and heterogeneous nature of the nucleation process is detrimental to the interpretation of

the spectroscopic signals. If one probes a large volume of a sample to capture a nucleation event that can occur anywhere at any time, the obtained spectroscopic signal is obscured by the average of various species in solution (*e.g.* monomers, aggregates, and crystals). Although not applied in this field yet, there is a well-established powerful concept to deal with a stochastic, complex and heterogenous system: *Single molecule spectroscopy*.[19] The key to bring optical spectroscopy with its full potential is to probe *one* nucleation event at a time, which requires one to predict precisely where a nucleation occurs.

Herein, we report the first demonstration of single nucleus spectroscopy (SNS) which spectroscopically probes crystal nucleation from aqueous solution one nucleus at a time. SNS is based on a novel extension of optical trapping Raman micro-spectroscopy, a well-established tool to study a trapped single object.[20–22] In this work, optical trapping was used to induce a crystal nucleation from a focused laser spot instead of as a tweezer to trap an object. Optical trapping induced crystallization (OTIC) was first demonstrated by Sugiyama *et al.* in 2007 by focusing a near infrared (NIR) laser in supersaturated glycine/D_2O solution.[23] When a laser beam is tightly focused in solution, optical gradient force attracts particles toward the focus spot. Optical gradient force is, however, not large enough to trap a single molecule, and therefore it is generally assumed that the crystallization occurs as a result of local concentration increase by trapping aggregates in solution. Since the discovery, OTIC has been applied to a variety of systems,[24–26] and furthermore high quality single crystals can be prepared while the polymorphs can be chosen by laser polarization.[26,27] Our approach applies OTIC to spatially control a single crystal nucleation event (*i.e.* at a focused laser spot), so that a probe beam can be placed at the same position to track the nucleation process. We achieved measuring Raman spectral evolution of glycine crystal formation from water at the single nucleus level with 46 ms time resolution at room temperature. The fast spectral acquisition allowed us to extract the Raman spectrum of pre-nucleation aggregates and its kinetics during a nucleation. The comparison between the experimental and simulated Raman spectra of glycine aggregates provided new insights into the glycine crystallization dynamics such as the formation of loosely-bound hydrogen-bonded linear chains toward the nucleation.

Results and Discussion

The *in-situ* SNS setup to track *single* crystal nucleation dynamics was custom built based on an inverted microscope (Fig. 1a). While OTIC has been performed using a NIR laser in the past,[23–28] we employed a 532 nm CW laser that served the dual role of inducing a crystal nucleation and Raman excitation (Fig. 1b). This design has two major benefits: i) the setup is as simple as conventional confocal microscope; ii) high laser power (> 1 W at the sample) required to induce crystallization produces high Raman signal (*c.f.* typical Raman spectroscopy is performed using ~10 mW with seconds to minutes accumulation time), which increases the temporal resolution of Raman spectroscopy (46 ms in this report). The other benefits of a 532 nm laser compared with a NIR laser is that Raman scattering is more efficient and the temperature of water at laser focus does not increase (estimated as ~ 2 mK W^{-1}).[29,30] The polarization of the laser beam was randomized by a depolarizer to minimize any laser polarization effect on crystallization process, to mimic a "natural" crystallization occurring in solution. In fact, we have never observed the formation of γ-glycine crystal in the 112 performed experiments with a depolarized beam, while it is reported that γ-glycine crystal can be prepared when linear or circularly polarized beam is used.[27] The use of the depolarized beam also simplified the interpretation of Raman spectroscopy by eliminating Raman spectral variation due to polarization selection rules.

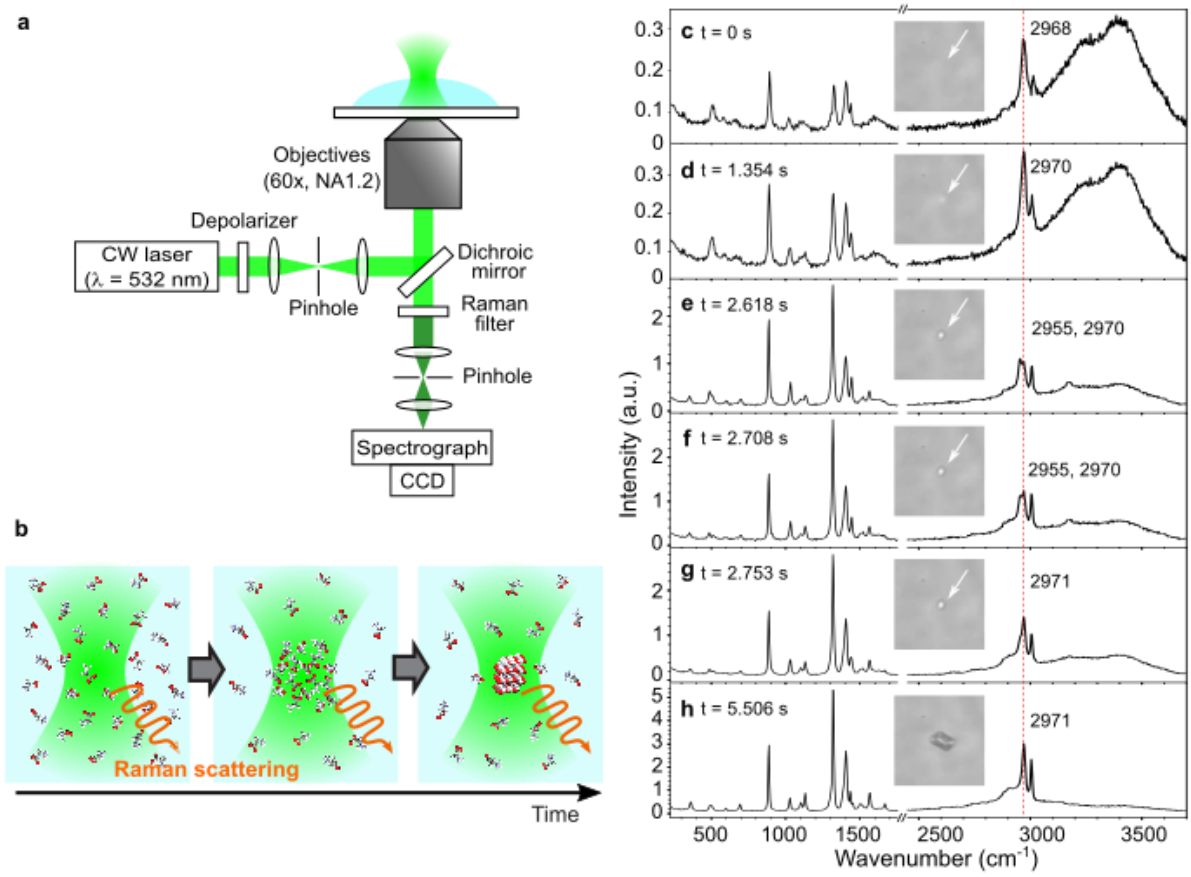


Fig. 1 Single Nucleus Spectroscopy (SNS): I*n-situ* Raman spectroscopy during α-glycine crystallization one nucleus at a time. a, A simple scheme of the home-built SNS setup to achieve *in-situ* Raman measurement during *single* crystal nucleation. **b**, A cartoon representing the dual role of 532 nm CW laser. A tightly focused CW laser induces crystallization at the focus while it serves as Raman excitation laser to track the crystallization dynamics. **c-h**, Snapshots (46 ms time resolution) of Raman spectral evolution during α-glycine crystallization. Insets are the bright field microscopy images (16 x 16 μm) taken at the corresponding time to each acquired spectrum.

Glycine crystallization from water was chosen as a model system in this study. While glycine has been widely studied as it is the simplest amino acid, its crystallization process is complex and has been under active debates.[31–38] The major hypothesis has been that α-glycine forms in water as a result of the assembly of doubly hydrogen-bonded cyclic dimers in a highly concentrated glycine solution.[31] Molecular dynamics simulations on glycine aqueous solutions have shown that glycine molecules do not form the doubly hydrogen-bonded cyclic dimers but instead form clusters mediated by charge-assisted hydrogen bonding.[34,36] Recent studies based on X-ray diffraction, solid-state NMR and cryo-TEM on flash cooled glycine aqueous solution proposed a new hypothesis that α-glycine crystallization occurs via glycine dihydrates and β-glycine.[35,37,39] While the formation of β-glycine from water was

observed at room temperature for example in nanopores,[40,41] the presence of glycine dihydrates has not yet been confirmed at room temperature.

SNS on glycine crystallization was performed by focusing a depolarized beam (1.2 W) in the glycine aqueous solution (the degree of supersaturation: SS = 1.2). Fig. 1c-f shows a series of Raman spectra obtained *in-situ* during a *single* crystal nucleation of glycine, which highlights the dynamic nature of crystal nucleation (see the movie SI1 for the whole experiment). Over the course of a nucleation, broad Raman bands around 3200-3500 cm^{-1} due to O-H vibration of water molecules decreased. This coincides with the appearance of a small blurry object in the bright field image, starting from the frame shown in Fig. 1d. These observations can be interpreted as the local concentration of glycine molecules increased and pushed water molecules out of the confocal detection volume. The formation of a crystal is clearly captured with the increase of Raman intensity and the appearance of new peaks (Fig. 1e), even if the bright field image remains blurry. Among various spectral changes upon nucleation, the appearance of the Raman peak at 2955 cm^{-1} caught our immediate attention because α-glycine (C-H stretching mode: ~2970 cm^{-1}) was expected to form in water. From the comparison of features with the reference spectra of α-, β- and γ-glycine measured on the same setup (Fig. SI1), the species was identified as β-glycine. The Raman spectral feature of β-glycine was short-lived and quickly became that of α-glycine (Fig. 1f-g).

The spectral evolution during a glycine nucleation was analyzed using non-negative matrix factorization (NMF), a type of unsupervised principal component analysis used in machine learning analysis (Fig. 2).[42] NMF has been used to deconvolute spectra in the field of Raman spectroscopy. A powerful aspect of this analysis algorithm is that a set of partial Raman spectrum (PRS), as well as the amplitude of each PRS during the spectral evolution, can be obtained without any assumption concerning the spectral shapes. The number of PRS used for the deconvolution is the only pre-determined parameter. The analysis was performed on the spectral series in which clear features of β-glycine were observed for several seconds before converting to α-glycine, to minimize the complexity of the crystallization dynamics. We first applied NMF analysis with two PRS to the spectral series (Fig. 2a), but it did not reach good fitting results (Fig. SI3) This indicates that the glycine crystal nucleation

does not occur through CNT mechanism, where the whole spectral evolution should be reproduced by a linear combination of the spectrum of solution and crystal. Instead, good quality fit was achieved with three PRS (Fig. 2b-f). One of the PRS (PRS-2) matches well with the Raman spectrum of β-glycine. Interestingly, the spectra before the appearance of β-glycine were deconvoluted to two PRS (PRS-1 and 3). PRS-1 resembled the spectrum of glycine aqueous solution at SS = 1.2, while we could not find a reference spectrum for PRS-3.

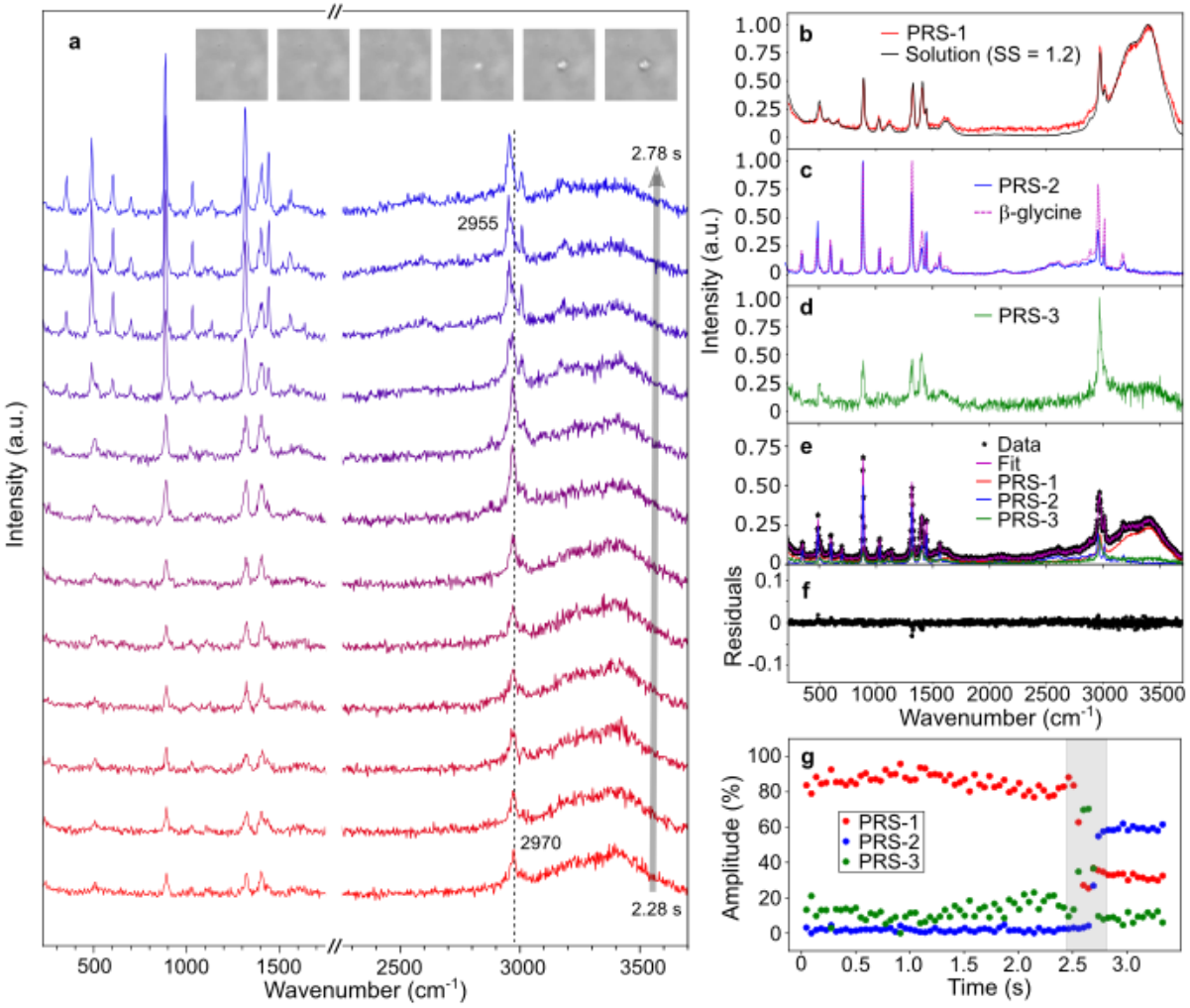


Fig. 2 Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis. a, Snapshots of Raman spectra showing the phase transition from solution (bottom) to crystal (top) with the bright field microscopy images (16x16 μm) corresponding to every other spectrum (at 2.28 s to 2.78 s from the left to right) as insets. **b-d**, Three spectra obtained by non-supervised data decomposition of the series of spectra in the panel **a** (PRS: Partial Raman Spectrum 1-3) with reference spectra. **e**, An example of the fit to the data (the frame at 2.69 s) by three constituents and **f**, the residuals of the fit. **g**, Temporal evolution of each constituent amplitude during the crystal nucleation.

If we assumed that crystal nucleation occurred via non-classical nucleation pathway, PRS-3 can be interpreted as the Raman spectrum of pre-nucleation aggregates. This assumption fits well with the temporal evolution of each PRS amplitude (Fig. 2g). Initially, the spectra were composed of 80~85 % of PRS-1, 0 % of PRS-2 and 15~20 % of PRS-3. Just before the nucleation occurred at ~2.65 s (the sudden increase of the amplitude of PRS-2), the amplitude of PRS-1 decreased while that of PRS-3 increased (highlighted by grey color in Fig. 2g). Once a crystal formed, the amplitude of PRS-3 dropped down to almost zero. PRS-1 does not go down to zero at this time range, because there is still solution remaining in the focal volume, as the size of a crystal is still too small to occupy the whole focal volume. These dynamics were reproduced qualitatively in 33 experiments where three components were necessary and sufficient in the NMF, although there are some variances between the experiments due to the stochastic nature of crystallization (Fig. SI4 and Fig. SI5 for more examples).

It is worth noting that the amplitude of PRS-3 was steadily 15~20 % before the nucleation event. This could mean that the aggregates of glycine molecules are abundant in a supersaturated aqueous solution. To test this hypothesis, we measured Raman spectrum of glycine aqueous solution as a function of concentration (Fig. 3a). To minimize an effect of optical trapping on the local concentration, a low laser power (50 mW) was used for this series of measurements. The intensity of Raman spectrum increased with the concentration, as expected. The intensity, however, did not increase proportionally as a function of the concentration but rather showed a saturation behavior (Fig. 3b). Already at the lower concentration range of SS = 0.05 to 0.2, the constant of proportionality of the Raman peak increase is not one. This suggests that the glycine solution at SS = 0.1 is not purely composed of monomers and the concentration of aggregates increases with higher SS, as seen by the saturation behavior of the peak intensities. NMF analysis with two PRS resulted in an excellent fit to the series of Raman spectra (Fig. SI7). Due to the nature of principle component analysis, PRS-1' resembles the spectrum of the lowest concentration of the data set fed into the algorithm. We applied the NMF analysis for different concentration regions (*e.g.* SS = 0.2 ~1.6, 0.4 ~ 1.6, 0.1 ~ 1.2 *etc.*) and verified that PRS-1' depends on the spectrum with the lowest intensity. PRS-2' was, however, not affected by the choice of concentration range (Fig. SI8, Fig. 3d). PRS-2' represents the spectral variation that occurs nonlinearly

upon the increase of the glycine concentration. As the change is due to the concentration increase, we assign PRS-2' as the spectrum of aggregates. For the comparison with the SNS results described earlier (Fig. 2), the NMF analysis was applied to the concentration range of SS = 1.2 ~ 1.6 to obtain PRS-1' and 2' (Fig. 3c-d).

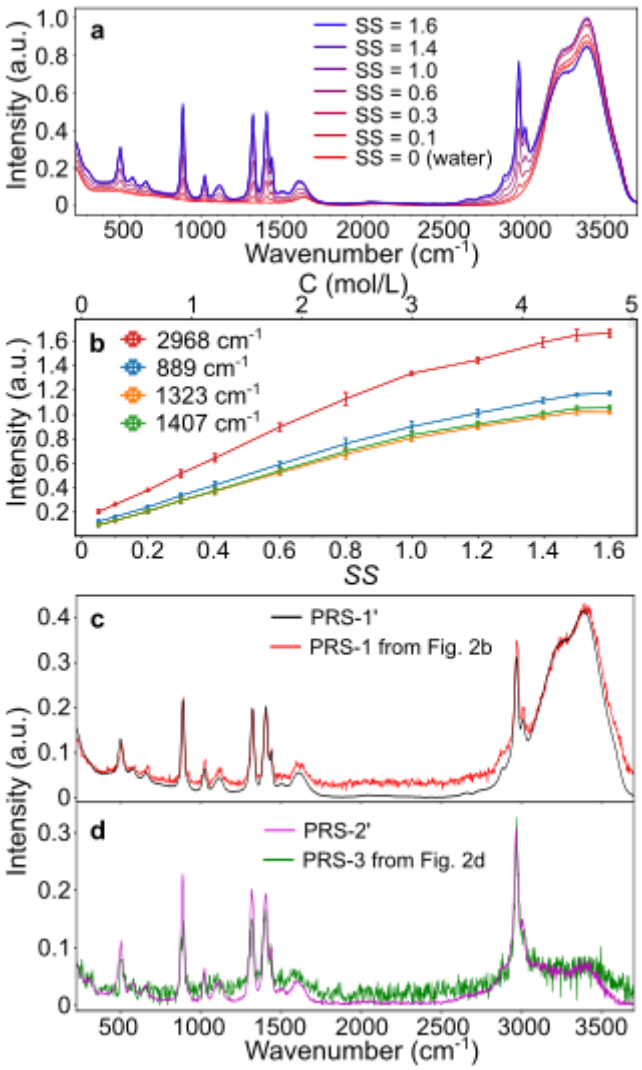


Fig. 3 A series of Raman spectra of glycine solution at different concentration and its NMF analysis revealing the spectrum of glycine monomers and aggregates. a, Raman spectra of glycine solution at different concentration. **b**, Raman peak intensity as a function of concentration at four positions. **c**, The overlay of PRS-1' (obtained from NMF analysis of spectra at SS = 1.2 ~1.6) and PRS-1 obtained from the nucleation dynamics (Fig. 2b), and **d**, PRS-2' and PRS-3 (Fig. 2d).

Notably, the PRS assigned as aggregates (PRS-2') from the concentration series matches the spectral features of PRS-3 obtained from β-crystal nucleation dynamics (Fig. 3d). The PRS-1 from Fig. 2b matches the PRS-1' obtained from the concentration series of SS = 1.2 ~ 1.6 (Fig. 3c). These

comparisons support our interpretation of the observed nucleation dynamics (Fig.2) discussed earlier. Crystal nucleation of glycine occurs in a non-classical pathway, where pre-nucleation aggregates are formed and converted to a crystal. While there are always some aggregates present under the optical trapping laser, the rapid increase in the amplitude of aggregates which leads to crystal nucleation starts rather stochastically. It remains unclear what is the critical difference between the aggregates readily present in solution and the pre-nucleation aggregates that convert to a crystal. As their Raman spectra matched well (Fig. 3d), aggregates that are present in solution and pre-nucleation aggregates may be structurally similar, except that we can only compare the time-averaged spectrum over 46 ms. It will be crucial to track the structural fluctuation at faster time scale to gain further insight. The differences between individual aggregates and "pre-nucleation aggregates" will be certainly an important aspect to be spectroscopically explored in the future.

While recent studies based on molecular dynamics simulation suggested that glycine molecules form large clusters through charge-assisted hydrogen bonding in aqueous solution,[36,43,44] it has been challenging to experimentally test this hypothesis. In this work, we were determined to gain an insight into the conformations of the pre-nucleation aggregates using the Raman spectrum of glycine aggregates as a link between the experiments and simulations. To this end, we first ran molecular dynamics (MD) simulations using a polarizable force-field at two concentrations: 3.3 mol L^{-1} representing the initial concentration in the experiments and 5.2 mol L^{-1} modeling a putative concentration before nucleation. We have analyzed the sampled configurations and performed density functional theory electronic structure calculations to obtain Raman spectra from glycine clusters of varying sizes (see Methods for the description of methods and technical details).

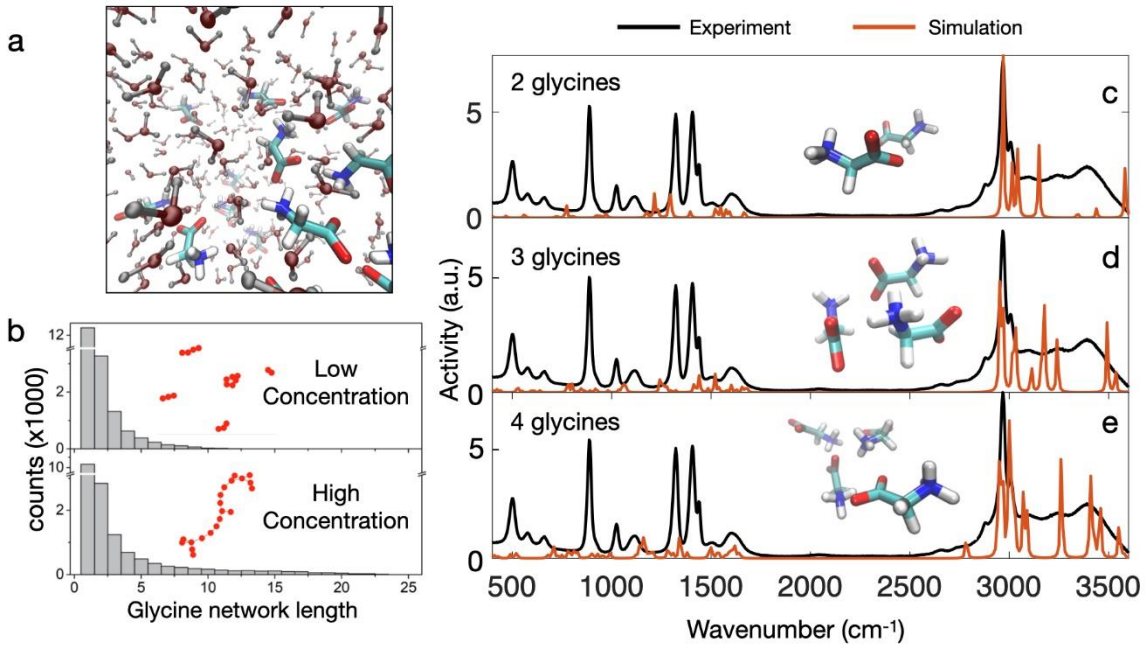


Fig. 4 Molecular dynamics (MD) simulation of glycine solutions and Raman spectrum calculations of glycine clusters. a, A snapshot from a MD trajectory of glycines in water; **b,** the effect of concentration on the formation of glycine-glycine contacts: at low concentration glycine predominantly exist as monomers, and as the concentration increases loosely-bound linear chains form. The insets are showing samples from the network visualizations: red circles are individual glycines, and the lines between them indicate contacts. Disconnected glycines are omitted. **c-e,** (red) simulated Raman spectra from glycine clusters of increasing sizes sampled using MD: **c**) two , **d**) three , and **e**) four glycines; the experimental PRS2'assigned to aggregates is shown in black.

Snapshots obtained from the MD simulation of glycine aqueous solution (Fig.4a, Fig. SI9) showed that glycine forms clusters mediated by charge-assisted hydrogen-bonding within the simulation box, which agrees with the previous reports.[34,36] These clusters dynamically form and dissociate as reported before. The results of the MD simulation were analyzed by defining the clusters using a 2.5Å cut-off distance between any two atoms belonging to two different glycines. This cut-off value is just under the van der Waals distance between O--H and N--H in order to capture the thermal fluctuations in hydrogen bonding. Fig. 4b shows the histograms of the size of glycine clusters at two different concentration. At the low concentration, monomers and open dimers dominate the solution. Doubly hydrogen bonded cyclic dimers[31] were never observed in our simulations. We attribute this to the merits of the polarizable forcefields (AMOEBA was used) that capture the nuanced dynamics of a charged species in water. As the concentration increases, larger clusters emerge as seen in the histogram at the bottom of Fig. 4b. The insets show the results of the network visualization analysis: red circles

are individual glycines and the lines between them indicate contacts. This analysis captures the glycine clusters forming hydrogen-bonded networks in a linear fashion. These chains are rather "loose" in the sense that the networks are dynamic systems that can assemble and disassemble over time and they curve up in the simulation box without collapsing into a homogeneous aggregate.

The formation of the hydrogen-bonded networks affects the vibration frequency of individual glycine molecules. The calculated Raman spectra for the clusters of two, three, four glycines (Fig. 4c-e) show that the frequency of vibration can be shifted as a result of the intermolecular interactions. A broadening of Raman spectrum is resulting from the heterogeneity of the conformation of glycines in the hydrogen bonded clusters (See movie SI2 and SI3). As a result of heterogenous broadening, Raman peaks gradually 'fill in' the broad spectroscopic feature observed in experiments (black), especially in the region of 2500–3500 cm^{-1} as the size of the clusters increases. The agreement between theory and experiment indicates that the loosely bound linear chains observed in the simulations may be the structure of pre-nucleation aggregates that are the precursors to constitute the hydrogen bonded networks observed in the crystal structure of β-glycine.

In summary, we demonstrated a novel *in-situ* optical spectroscopy to study crystallization dynamics of glycine one nucleus at a time, which overcomes the difficultly imposed by stochastic and heterogeneous nature of the nucleation. Spectral dynamics obtained during single glycine nucleation events from aqueous solution at room temperature, along with non-supervised spectral decomposition analysis, provided new evidences to understand the nucleation dynamics: i) glycine crystallization occurs via non-classical nucleation pathway, where pre-nucleation aggregates grow and nucleate; ii) the comparison between the experiments and simulations using the Raman spectrum of aggregates suggests that glycine molecules form loosely-bound linear chains, which may be the precursor to β-glycine crystal. An *in-situ* spectroscopic approach presented in this study, "single nucleus spectroscopy", can be regarded as a basic platform to which more advanced spectroscopy techniques can be coupled. The present work identifies a modality for time-resolved study of crystal nucleation that provides a strong impetus for accelerating such investigations using additional capabilities of optical spectroscopy.

Methods (3000 words max)

Sample preparation

Glycine (Sigma-Aldrich) was used without further purification. Glycine aqueous solution was prepared using ultrapure water (MilliQ). Glycine was dissolved at 80 ℃ using ultrasonication, filtered with a 0.2 µm Target2^{TM} Nylon syringe filter (Thermo Scientific), and then slowly cooled down (~3 hours) to room temperature in a dry bath (Thermo Scientific). The degree of supersaturation (SS) was calculated using the solubility of glycine in water at 20 ℃ (0.225 g mL^{-1}).[45]

For SNS, a sample was assembled by placing a silicone isolator sheet (0.25 mm thick, Grace Bio-Labs 664475) with a 10 mm hole on a piranha cleaned cover glass. The silicone sheet was cleaned in methanol before the use. 10 µL of glycine aqueous solution (SS = 120 %) was dropped in the hole, which resulted in a thin film of glycine solution (100 µm thickness). Another cover glass was placed on top to prevent the evaporation of water. No spontaneous crystallization was observed for ~1 hour.

Conventional crystallization was performed to obtain α-, β-, and γ-glycine, to measure Raman spectrum of each phase (Fig. SI1). α-glycine was prepared by the slow evaporation of glycine aqueous solution. β-glycine was prepared by adding methanol to aqueous solution of glycine. γ-glycine was obtained by the slow evaporation of glycine aqueous solution with potassium nitrate as additive (see more details in Supplementary Information).

Single Nucleus Spectroscopy (SNS) setup

1) General scheme

The optical setup (Fig.1a) was home-built based on an inverted microscope (Olympus IX73). A 532 nm CW laser (Laser Quantum, Opus 532) was used with a dual role for optical trapping and Raman excitation. The diameter of the laser beam was properly adjusted by a telescope to slightly overfill the back aperture of water-immersion objective lens (Olympus UPLSAPO60XW, NA1.2). Liquid crystal polymer depolarizer (Thorlabs, DPP25-A) was used to randomize the polarization of the laser beam. A periodic retardation pattern of the depolarizer generates two focus spots due to diffraction. The beam after the depolarizer was once focused by a lens to spatially filter one of the spots and then was

recollimated (Fig. SI10). A dichroic beamsplitter (AHF analysentechnik AG, Raman beamsplitter RT 532 rdc, F78-535) and a 532 nm RazorEdge® ultrasteep long-pass edge filter (Semrock, LP03-532RE-25) were used to remove the excitation beam from Raman scattering. The signal was spatially filtered at the conjugate plane using a 25 μm pinhole to remove the contribution of extrafocal volume. The size of the pinhole adopted in this setup is smaller than the one in typical setups (~100 μm). Therefore, the contribution of Raman signal from the extrafocal volume is significantly reduced and the spectral sensitivity to the nucleation dynamics occurring at the laser focus is improved.[46] Raman spectrum was recorded by using a spectrograph (Andor, Kymera 193) and a CCD (Andor, iDus420) with a 900 l/mm density diffraction grating, blazed at 550 nm (Quantum Design AG, A-SR2-GRT-0900-0550).

2) Technical details of *in-situ* glycine crystallization

The depolarized laser beam was focused just below (~1 μm) the liquid-air interface of glycine solution. While it has been reported that the height of the liquid-air interface continuously changes during OTIC with a 1064 nm laser,[47] almost no change of the height was observed in the experiments with a 532 nm laser. This could simply be because there is no heating at the focus when 532 nm is used. OTIC with a 532 nm laser not only allows one to use aqueous solution instead of deuterated (D_2O) solution, but also reduces complications related with heat-induced flows around the focus spot. Although the height of the liquid-air interface was stable, we also verified that the spectral shape near the liquid-air interface does not change due to the focus position relative to the interface (Fig. SI11). When the beam is focused at or above the interface, Raman signal is lower because half or more of the focal volume probes air. Besides the intensity, the spectral shape remained unchanged.

Data analysis

The raw data were noise-filtered by Singular Value Decomposition (SVD, see Fig. SI2). The cleaned spectra were analyzed by Non-negative Matrix Factorization (NMF) using the Scikit_Learn library in Python.[48] The python code and the data are available on Yareta. For more details, see the Supplementary Information.

Computational simulations

Molecular dynamics simulations were run on the zwitterionic form of glycine in a 8 nm^3 water box using the AMOEBA 2013 forcefield[49] implemented in the OpenMM toolkit[50] at two concentrations: 3.3 mol L^{-1} and 5.2 mol L^{-1}. The Raman spectra calculations were performed on clusters obtained from the MD trajectories using the Gaussian 16 suite of programs[51] with B3LYP/6-311G(d)[33]. For the complete technical discussion, please see the Supplementary Information Section SI2.

Acknowledgements

T.B.M.A., O.U., and J.B. are thankful for the financial support from the University of Geneva. Computations were made on the supercomputer Béluga, managed by Calcul Québec (https://www.calculquebec.ca/) and Compute Canada (www.computecanada.ca). The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI).

Author Contributions

T.B.M.A. designed the initial experiments. T.B.M.A., O.U. and J.B. built the setup and finalized the design of the experiments. O.U. performed most of the measurements under supervision of T.B.M.A and J.B.. J.B. introduced NMF analysis, and O.U. and J.B. performed most of the NMF analysis. N.L and L.S. performed the computational simulations and analysis. All authors were involved in drafting the manuscript.

Competing Interests Statement

The authors declare no competing interests.

References (30 max, can be 50 in special case)

1. Mullin, J. W. *Crystallization.* (Elsevier Science, 2014).
2. Bernstein, J. *Polymorphism in molecular crystals*. (Oxford University Press, 2002).

3. Tamura, R. & Miyata, M. *Advances in Organic Crystal Chemistry Comprehensive Reviews 2015*. (Springer, 2015).
4. Sosso, G. C. *et al.* Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chemical Reviews* **116**, 7078–7116 (2016).
5. Erdemir, D., Lee, A. Y. & Myerson, A. S. Nucleation of Crystals from Solution: Classical and Two-Step Models. *Accounts of Chemical Research* **42**, 621–629 (2009).
6. Gebauer, D., Kellermeier, M., Gale, J. D., Bergström, L. & Cölfen, H. Pre-nucleation clusters as solute precursors in crystallisation. *Chem. Soc. Rev.* **43**, 2348–2371 (2014).
7. Jehannin, M., Rao, A. & Cölfen, H. New Horizons of Nonclassical Crystallization. *J. Am. Chem. Soc.* **141**, 10120–10136 (2019).
8. Schall, P., Cohen, I., Weitz, D. A. & Spaepen, F. Visualizing dislocation nucleation by indenting colloidal crystals. *Nature* **440**, 319–323 (2006).
9. Dong, F., Liu, M., Grebe, V., Ward, M. D. & Weck, M. Assembly of Shape-Tunable Colloidal Dimers in a Dielectrophoretic Field. *Chem. Mater.* **32**, 6898–6905 (2020).
10. De Yoreo, J. J. & N. A. J. M., S. Investigating materials formation with liquid-phase and cryogenic TEM. *Nat Rev Mater* **1**, 16035 (2016).
11. Patterson, J. P., Xu, Y., Moradi, M.-A., Sommerdijk, N. A. J. M. & Friedrich, H. CryoTEM as an Advanced Analytical Tool for Materials Chemists. *Acc. Chem. Res.* **50**, 1495–1501 (2017).
12. Ross, F. M. Opportunities and challenges in liquid cell electron microscopy. *Science* **350**, aaa9886 (2015).
13. Tsarfati, Y. *et al.* Crystallization of Organic Molecules: Nonclassical Mechanism Revealed by Direct Imaging. *ACS Central Science* **4**, 1031–1036 (2018).
14. Schreiber, R. E. *et al.* Real-time molecular scale observation of crystal formation. *Nature Chemistry* **9**, 369–373 (2016).

15. Yamazaki, T. *et al.* Two types of amorphous protein particles facilitate crystal nucleation. *Proc Natl Acad Sci USA* **114**, 2154–2159 (2017).
16. Nielsen, M. H., Aloni, S. & De Yoreo, J. J. In situ TEM imaging of CaCO3 nucleation reveals coexistence of direct and indirect pathways. *Science* **345**, 1158–1162 (2014).
17. Nakamuro, T., Sakakibara, M., Nada, H., Harano, K. & Nakamura, E. Capturing the Moment of Emergence of Crystal Nucleus from Disorder. *J. Am. Chem. Soc.* **143**, 1763–1767 (2021).
18. Vekilov, P. G. The two-step mechanism of nucleation of crystals in solution. *Nanoscale* **2**, 2346 (2010).
19. Barbara, P. F. Single-Molecule Spectroscopy. *Acc. Chem. Res.* **38**, 503–503 (2005).
20. Buajarern, J., Mitchem, L. & Reid, J. P. Manipulation and Characterization of Aqueous Sodium Dodecyl Sulfate/Sodium Chloride Aerosol Particles. *J. Phys. Chem. A* **111**, 13038–13045 (2007).
21. Chan, J. W. Recent advances in laser tweezers Raman spectroscopy (LTRS) for label-free analysis of single cells. *J. Biophoton.* **6**, 36–48 (2013).
22. Mrad, R., Kruglik, S. G., Ben Brahim, N., Ben Chaâbane, R. & Negrerie, M. Raman Tweezers Microspectroscopy of Functionalized 4.2 nm Diameter CdSe Nanocrystals in Water Reveals Changed Ligand Vibrational Modes by a Metal Cation. *J. Phys. Chem. C* **123**, 24912–24918 (2019).
23. Sugiyama, T., Adachi, T. & Masuhara, H. Crystallization of Glycine by Photon Pressure of a Focused CW Laser Beam. *Chemistry Letters* **36**, 1480–1481 (2007).
24. Yuyama, K., Islam, M. J., Takahashi, K., Nakamura, T. & Biju, V. Crystallization of Methylammonium Lead Halide Perovskites by Optical Trapping. *Angew. Chem. Int. Ed.* **57**, 13424–13428 (2018).

25. Cheng, A.-C., Masuhara, H. & Sugiyama, T. Evolving Crystal Morphology of Potassium Chloride Controlled by Optical Trapping. *J. Phys. Chem. C* **124**, 6913–6921 (2020).
26. Sugiyama, T., Yuyama, K. & Masuhara, H. Laser Trapping Chemistry: From Polymer Assembly to Amino Acid Crystallization. *Accounts of Chemical Research* **45**, 1946–1954 (2012).
27. Yuyama, K., Rungsimanon, T., Sugiyama, T. & Masuhara, H. Selective Fabrication of α- and γ-Polymorphs of Glycine by Intense Polarized Continuous Wave Laser Beams. *Crystal Growth & Design* **12**, 2427–2434 (2012).
28. Yuyama, K., Sugiyama, T. & Masuhara, H. Laser Trapping and Crystallization Dynamics of L -Phenylalanine at Solution Surface. *J. Phys. Chem. Lett.* **4**, 2436–2440 (2013).
29. Ito, S., Sugiyama, T., Toitani, N., Katayama, G. & Miyasaka, H. Application of Fluorescence Correlation Spectroscopy to the Measurement of Local Temperature in Solutions under Optical Trapping Condition. *J. Phys. Chem. B* **111**, 2365–2371 (2007).
30. Hale, G. M. & Querry, M. R. Optical Constants of Water in the 200-nm to 200-μm Wavelength Region. *Appl. Opt.* **12**, 555 (1973).
31. Weissbuch, I., Torbeev, V. Yu., Leiserowitz, L. & Lahav, M. Solvent Effect on Crystal Polymorphism: Why Addition of Methanol or Ethanol to Aqueous Solutions Induces the Precipitation of the Least Stable β Form of Glycine. *Angew. Chem. Int. Ed.* **44**, 3226–3229 (2005).
32. Huang, J., Stringfellow, T. C. & Yu, L. Glycine Exists Mainly as Monomers, Not Dimers, in Supersaturated Aqueous Solutions: Implications for Understanding Its Crystallization and Polymorphism. *J. Am. Chem. Soc.* **130**, 13973–13980 (2008).
33. Friant-Michel, P. & Ruiz-López, M. F. Glycine Dimers: Structure, Stability, and Medium Effects. *Chem. Eur. J. of Chem. Phys.* **11**, 3499–3504 (2010).

34. Yani, Y., Chow, P. S. & Tan, R. B. H. Glycine Open Dimers in Solution: New Insights into α-Glycine Nucleation and Growth. *Crystal Growth & Design* **12**, 4771–4778 (2012).
35. Xu, W., Zhu, Q. & Hu, C. T. The Structure of Glycine Dihydrate: Implications for the Crystallization of Glycine from Solution and Its Structure in Outer Space. *Angew. Chem.* **129**, 2062–2066 (2017).
36. Bushuev, Y. G., Davletbaeva, S. V. & Koifman, O. I. Molecular dynamics simulations of aqueous glycine solutions. *CrystEngComm* **19**, 7197–7206 (2017).
37. Cerreia Vioglio, P. *et al.* Insights into the Crystallization and Structural Evolution of Glycine Dihydrate by In Situ Solid-State NMR Spectroscopy. *Angew. Chem. Int. Ed.* **57**, 6619–6623 (2018).
38. Broadhurst, E. T. *et al.* Polymorph evolution during crystal growth studied by 3D electron diffraction. *IUCrJ* **7**, 5–9 (2020).
39. Surovtsev, N. V. *et al.* Glycine phases formed from frozen aqueous solutions: Revisited. *The Journal of Chemical Physics* **137**, 065103 (2012).
40. Jiang, Q., Hu, C. & Ward, M. D. Stereochemical Control of Polymorph Transitions in Nanoscale Reactors. *J. Am. Chem. Soc.* **135**, 2144–2147 (2013).
41. Juramy, M. *et al.* Monitoring Crystallization Processes in Confined Porous Materials by Dynamic Nuclear Polarization Solid-State Nuclear Magnetic Resonance. *J. Am. Chem. Soc.* (2021) doi:10.1021/jacs.0c12982.
42. Woelffel, W. *et al.* Analysis of soda-lime glasses using non-negative matrix factor deconvolution of Raman spectra. *Journal of Non-Crystalline Solids* **428**, 121–131 (2015).
43. Hamad, S., Hughes, C. E., Catlow, C. R. A. & Harris, K. D. M. Clustering of Glycine Molecules in Aqueous Solution Studied by Molecular Dynamics Simulation. *J. Phys. Chem. B* **112**, 7280–7288 (2008).

44. Di Gioacchino, M., Ricci, M. A., Imberti, S., Holzmann, N. & Bruni, F. Hydration and aggregation of a simple amino acid: The case of glycine. *Journal of Molecular Liquids* **301**, 112407 (2020).
45. Datta, A. & Hossain, A. Solubility Data of Glycine in Water and Justification of Literature Results. *Asian J. Chem.* **32**, 1525–1533 (2020).
46. Wilson, T. Resolution and optical sectioning in the confocal microscope: PROPERTIES OF THE FLUORESCENT CONFOCAL MICROSCOPE. *Journal of Microscopy* **244**, 113–121 (2011).
47. Yuyama, K., Sugiyama, T. & Masuhara, H. Millimeter-Scale Dense Liquid Droplet Formation and Crystallization in Glycine Solution Induced by Photon Pressure. *The Journal of Physical Chemistry Letters* **1**, 1321–1325 (2010).
48. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
49. Esser, A., Belsare, S., Marx, D. & Head-Gordon, T. Mode specific THz spectra of solvated amino acids using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* **19**, 5579–5590 (2017).
50. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).
51. Frisch, M. J. *et al. Gaussian 16 Rev. C.01*. (2016).