Supplementary information

In-situ optical spectroscopy of crystallization: one crystal nucleation at a time

Oscar Urquidi1,†, Johanna Brazard1,†, Natalie LeMessurier2, Lena Simine2, Takuji B. M. Adachi1,*

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

Contents:
Section SI1. Supplementary figures:
  Figure S1. Raman spectra of α-, β-, and γ- glycine crystal.
  Figure S2. Raw data and filtered data by SVD analysis.
  Figure S3. NMF analysis with two PRS on the dataset shown in Fig. 2 of the main text.
  Figure S4. Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis: Example 2.
  Figure S5. Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis: Example 3.
  Figure S6. The difficulty of NMF analysis with four components on data set where β-glycine quickly converts to α-glycine.
  Figure S7. A series of Raman spectra of glycine solution at different concentration (A complete data set of Fig. 3).
  Figure S8: Two components NMF analysis of glycine solution at different concentration and data reconstruction of an intermediate spectrum.
  Figure S9: Two components NMF analysis of glycine solution at different concentration ranges.
  Figure S10: An example of glycine clusters forming hydrogen-bonded networks in a linear fashion.
  Figure S11: The comparison of Partial Raman Spectrum assigned as glycine aggregates (PRS-3) and the Raman spectrum of formamide
Figure S12. Images of the focal spot after and before spatial filtering of the depolarized 532 nm laser beam.

Figure S13. Characterization of the confocal volume of SCNS by Fluorescence correlation spectroscopy (FCS).

Figure S14. Raman spectra of glycine solution at different focus positions near the liquid-air interface.

Section SI2. Computational simulations:
   SI2.A Molecular Dynamics of Glycine in water.
   SI2.B Raman spectra calculations.

Section SI3. Movie files:
   Movie S1. An example of Single Crystal Nucleation Spectroscopy (SCNS) using Raman spectroscopy as a probe during single crystal nucleation of glycine.
   Movie S2. Vibration modes corresponding to the simulated Raman peaks shown in Fig. 4c-e in the main text. This movie contains the six peaks in the range 2900 – 3200 cm$^{-1}$.
   Movie S3. Vibration modes corresponding to the simulated Raman peaks shown in Fig. 4c-e in the main text. This movie contains the four peaks in the range 3200 – 3600 cm$^{-1}$.
Section S11. Supplementary figures.

Raman spectra of α-, β-, and γ- glycine crystal:

The Raman spectra of α-, β-, and γ-glycine crystal were measured using a depolarized 532 nm CW laser beam (100 mW after objective, 15 times five seconds exposure time).

Firstly, α-glycine crystal was grown by slow evaporation of a supersaturated glycine/water (MilliQ) solution. Secondly, β-glycine crystal was prepared by adding 2 ml of methanol to 1 mL of a glycine/water solution (SS = 0.2). After 15 minutes, needle-like crystals were produced.(1) Finally, γ-glycine was prepared by mixing to a 1:1 molar ratio nitrate potassium (KNO₃) and glycine in water (MilliQ) to achieve a SS = 1 glycine solution in water. Then the solution was slowly evaporated.(2)

![Raman spectra of three glycine crystal polymorphs.](image)

**Fig. S1: Raman spectra of three glycine crystal polymorphs.** a, Overlaid Raman spectrum of α-, β- and γ-glycine crystal. b, Raman spectrum of α-glycine crystal. c, Raman spectrum of β-glycine crystal. d, Raman spectrum of γ-glycine crystal.
Raw data and filtered data by SVD analysis:

Raman spectra of a SS = 1.2 glycine/water solution were recorded at 46 ms time resolution (Fig. S2a) and analyzed by Singular Value Decomposition (SVD). After keeping the six main singular values, the raw data was noise-filtered as shown on Fig. S2b. The residuals map does not present any spectral signature (Fig. S2c). Fig SI 2d shows a selected spectrum before and after SVD-filtering.

The filtered data are the ones represented on Fig. 1 and 2 in the main text.

Fig. S2: SVD filtering of Raman spectra of a 120% SS glycine/water solution. a. Temporal evolution of Raman spectra recorded with a 46 ms time resolution. b. Temporal evolution of Raman spectra after SVD filtering. c. Residuals maps after SVD filtering. d. Raman spectrum at 3.330 s before (black) and after (red) SVD filtering. The residuals are shown in black dots.
NMF analysis with two PRS on the dataset shown in Fig. 2 of the main text:

The Raman spectral evolution during the β-glycine crystal nucleation described in the Fig. 2 in the main text was analyzed first by NMF analysis but with two PRS, instead of three. The data (the frame at 2.69 s), the fit, and the residuals of the fit are shown on Fig. S3. Some spectral features around 900, 1400 and 2900 cm\(^{-1}\) remain on the residuals. The NMF analysis with two PRS did not describe properly the dynamics of this system.

Fig. S3: Data reconstruction with two PRS obtained by NMF analysis. a, Fit to the data (the frame at 2.69 s) by two components. b, The residuals of the fit.
Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis: Example 2:

An additional example of β-glycine nucleation is shown on Fig. S4. The time resolution of this Raman spectral evolution is 46 ms, as the one reported in the main text.

Fig. S4: Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis (46 ms time resolution). 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 spectrum (from 6.95 to 7.49 s, 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). e, An example of the fit to the data (the frame at 7.33 s) by three constituents and f, the residuals of the fit. g, Temporal evolution of each constituent amplitude during the crystal nucleation.
Another example of β-glycine nucleation is shown on Fig. S5. The time resolution of this Raman spectral evolution is 164 ms.

**Fig. S5: Raman spectral evolution during a β-glycine crystal nucleation and its non-supervised data decomposition analysis (164 ms time resolution).**

- **a**, Snapshots of Raman spectra showing the phase transition from solution (bottom) to crystal (top) with the bright field microscopy images (20x20 μm) corresponding to every spectrum (from 3.72 to 4.80 s, 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).
- **e**, An example of the fit to the data (the frame at 4.13 s) by three constituents and **f**, the residuals of the fit.
- **g**, Temporal evolution of each constituent amplitude during the crystal nucleation.
The difficulty of NMF analysis with four components on data set where β-glycine quickly converts to α-glycine:

As mentioned in the main text, the NMF analysis was performed on the spectral series in which clear features of β-glycine were observed for several seconds before converting to α-glycine in this work. The NMF analysis on the spectral series when β-glycine quickly converted to α-glycine is yet challenging even at the spectral time resolution of 46 ms. Fig. S6 shows the result of NMF analysis on the spectral evolution that was presented in Fig. 1 of the main text. Because the lifetime of β-glycine is short, its spectrum is observed on top of the spectrum of either intermediate species or α-glycine. This limits the NMF analysis to execute a clean spectral deconvolution of each species, and it results in PRS that are somewhat spectrally mixed (Fig. S6a, c and d). Despite the imperfect deconvolution, it was still possible to assign each PRS to reference spectra (Fig. S6a-c). The temporal evolution of each PRS (Fig. S6g) shows that the crystal nucleation occurs as the amplitude of solution (PRS-2, the red filled circle) decreases while that of “prenucleation aggregates” (PRS-4, the green filled circle) increases (highlighted by grey color). Then β-glycine (PRS-1) and α-glycine (PRS-3) appears. Overall, the nonclassical nucleation was observed where the amplitude of prenucleation aggregates increases before a crystal appeared. Because the quality of the deconvolution was not as good as those shown in the main text (Fig. 2) and Fig. S4 and S5, we focused on discussing the spectral series in which clear features of β-glycine were observed for several seconds before converting to α-glycine in this report.

Fig. S6: Non-supervised data decomposition analysis (46 ms time resolution) on the Raman spectral evolution in which a β-glycine crystal quickly converted to α-glycine. a-d, Four spectra
obtained by non-supervised data decomposition of the series of Raman spectra (Part of them are shown in Fig. 1 in the main text, PRS: Partial Raman Spectrum 1-4). e, An example of the fit to the data (the frame at 0.27 s) by four constituents and f, the residuals of the fit. g, Temporal evolution of each constituent amplitude during the crystal nucleation.
A series of Raman spectra of glycine solution at different concentration (A complete data set of Fig. 3):

Raman spectrum of glycine solution was measured as a function of concentration. Glycine aqueous solutions were prepared using ultrapure water (MilliQ). Glycine was dissolved at 80 °C using ultrasonication and filtered with a 200 nm Nylon membrane. Up to SS = 1.2, solutions were separated in individual vials and then slowly cooled down (~3 hours) to room temperature in a dry bath. For this set of measurements, a 100 µl droplet was placed on a glass coverslip.

For the other set of concentrations (SS = 1.4 and 1.6), the solutions were transferred at 80 °C to a container composed of two glass coverslips separated by a 1.5 mm silicone spacer with a 10 mm hole. In each case, 35 µl of solution was placed inside this container, making sure the droplet touched just the top and bottom glass surfaces. These samples were then slowly cooled down to room temperature on a hot plate (~3 hours). The microscope height was set at 100 µm from the glass-liquid interface and the laser power was kept at 50 mW (after objective) for all measurements. The degree of SS was calculated using the solubility of glycine in water at 20 °C (0.225 g ml$^{-1}$)(3).

![Fig. S7: Raman spectra of glycine/water solution at multiple super saturation (SS) concentrations. a, Raman spectra of glycine/water solution at different SS. b, Raman intensity at different SS for four main peaks: 889 (blue), 1323 (orange), 1407 (green) and 2968 cm$^{-1}$ (red).](image)
Two components NMF analysis of glycine solution at different concentrations:

The series of spectra shown in Fig. S7a was analyzed by NMF and two PRS were extracted (Fig. S8a). A spectrum at an intermediate concentration (SS = 0.6) was reconstructed as a linear combination of the components. The data, and the fit, and the residuals are shown on Fig. S8. No prominent spectral features are identified in the residuals.

Fig. S8: Two components NMF analysis of glycine solution at different concentration and data reconstruction of an intermediate spectrum. a, Two PRS constituting the series of spectra in the panel a of Fig. S7 (PRS-1' and 2'), obtained from the NMF analysis. b, An example of the fit to the data of an intermediate spectrum (SS = 0.6) by two components. c, Residuals of the fit.
Two components NMF analysis of glycine solution at different concentration ranges:

The spectra shown in Fig. S7a were analyzed by NMF for different concentration ranges (Fig. S9). i) The lowest concentration of the series of spectra was varied from SS = 0.1~1.6 to 0.8~1.6. ii) The highest concentration of the series of spectra was varied from SS = 0.1~1.6 to 0.1~0.8. For each series of spectra analyzed by NMF, two PRS were extracted. PRS-1’ (Fig. S9a-d) changes depending on the lowest concentration of the spectral series. Fig. S9a-b shows that PRS-1’ varies as the lowest concentration in the series increases from 0.1 to 0.8, while PRS-1’ in the Fig. S9c-d barely varies as the highest concentration decreases. After the normalization at 1325 cm⁻¹, the main peaks are identical (Fig. S9b and d). On the other hand, PRS-2’ (Fig. S9e-h) is insensitive to the range of concentration chosen for the analysis. PRS-2’ represents the spectra 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. Remarkably, PRS-2’ is universal over different range of concentration variation.

![Fig. S9: Two components NMF analysis of glycine solution at different concentration ranges. a-b, PRS-1’ obtained by NMF analysis of glycine solution for the concentration range from SS = 0.1~1.6 to SS = 0.8~1.6 (b is the normalized spectra of a at the peak intensity at 1325 cm⁻¹) and c-d, from SS = 0.1~1.6 to 0.1~0.8 (d is the normalized spectra of c at the peak intensity at 1325 cm⁻¹). e-f, PRS-2’ obtained by NMF analysis of glycine solution for the concentration range from SS = 0.1~1.6 to SS = 0.8~1.6 (f is the normalized spectra of e at the maximum intensity) and g-h, from SS = 0.1~1.6 to 0.1~0.8 (h is the normalized spectra of g at the maximum intensity).]
An example of larger glycine clusters:

Fig. S10: An example of glycine clusters forming hydrogen-bonded networks in a linear fashion. 

a, A snapshot of molecular dynamics (MD) simulation of glycine aqueous solution with both glycine and water molecules shown and b, with only glycine molecules shown. The opacity of the molecular model is used to represent the 3-D depth of the position of each molecule.
The comparison of Partial Raman Spectrum assigned as glycine aggregates (PRS-3) and the Raman spectrum of formamide:

The Raman spectrum of liquid formamide (Sigma-Aldrich 47670-250ML-F, >99.5% purity) was measured on the same Raman micro-spectroscopy setup. The laser power was 50 mW after the objective lens. The structure of formamide has been studied extensively in literature and it has been known to form hydrogen-bonded linear networks. Due to the networks, a broad band between 3000 to 3500 cm\(^{-1}\) has been observed and assigned as the signature of N-H---O hydrogen bonding networks.

![Figure S11: The comparison of Partial Raman Spectrum assigned as glycine aggregates (PRS-3, PRS-2') and the Raman spectrum of formamide. a, Raman spectrum of glycine aggregates (PRS-2'/PRS-3) and b, formamide (inset is one of the proposed hydrogen-bonded liquid structures in literature(5)).](image)
Images of the focal spot after and before spatial filtering of the depolarized 532 nm laser beam:

The 532 nm laser beam passing through a liquid crystal polymer depolarizer (Thorlabs, DPP25-A) generated a diffraction pattern with two focus spots separated by 2 µm (Fig. S12a). The collimated beam after the depolarizer was focused. At the Fourier space, one of the two spots was spatially filtered. After spatial filtering, a single focus spot was observed after the microscope objective (Fig. S12b), and the laser power was twice reduced.

Fig. S12: Images of the focal spot as the depolarized 532 nm laser beam is focused at the glass interface. a, Before and b, After spatial filtering at the Fourier space.
Characterization of the confocal volume of SCNS by Fluorescence correlation spectroscopy (FCS):

FCS was used to characterize the effective confocal volume of the SCNS setup. The setup was slightly modified to be able to measure fluorescence of Rhodamine 6G dye aqueous solution (12.5 nM) on a single photon counting module (Laser components GmbH, Count® Blue). An additional dichroic filter (AHF analyseotechnik AG, RazorEdge LP Edge, F76-785) was placed in front of the detector to reject Raman signals from water. Rhodamine 6G in water has been used as a standard sample for determining the confocal volume, as the diffusion constant, $D$, is well determined as $367 \, \mu m^2/s$ at 21°C.(6) A time correlated photon counting system (PicoQuant, MultiHarp-150) was used to register the time correlated single photon events. Autocorrelation curve was obtained from the fluorescence time transient using a home-written Python code based on literature,(7) and the curve was fitted to a 3D diffusion model (equation 1) to extract an effective volume approximated as a 3D gaussian (equation 2). (6)

\[
G(t) = A \cdot \left(1 + \frac{4Dt}{w_0^2}\right)^{-1} \cdot \left(1 + 4Dt k^2 w_0^2\right)^{-1/2} \tag{1}
\]

\[
V_{eff} = \frac{\pi^{3/2} w_0^3 k}{2} \tag{2}
\]

$A$ is an amplitude, $w_0$ is the lateral 1/e$^2$-radius of the confocal volume, $k$ is the eccentricity of the confocal volume (the ratio of the axial and lateral 1/e$^2$-radius), and $t$ is the lagtime.

Fig. S13 shows an example of the fluorescence time transient, autocorrelation curve and fit results of single experiment using 25 and 50 μm pinholes. For all the experiments, laser power was set to 0.67 mW after the objective, corresponding to the saturation region (where a significant change in power does not have a big impact on the counts). Measurements were performed five times for each pinhole size to obtain the mean and the standard deviation of each parameter. For the 25 μm pinhole, mean values and standard deviations were obtained as: $A=1.82 \pm 0.03$, $w_0=0.273 \pm 0.002 \, \mu m$, $k=4.6 \pm 0.2$ and $V_{eff} = 0.52 \pm 0.03 \, \mu m^3$. For the 50 μm pinhole, mean values and standard deviations are: $A = 0.912 \pm 0.003$, $w_0=0.351 \pm 0.002 \, \mu m$, $k=4.4 \pm 0.2$, $V_{eff} = 1.05 \pm 0.05 \, \mu m^3$. 

![Example of the fluorescence time transient, autocorrelation curve and fit results](image)
Fig. S13. FCS measurement of a standard dye solution. Representative data of fluorescence time transient, autocorrelation curve and the fit results of Rhodamine 6G aqueous solution with a, 25 μm, and b, 50 μm pinhole. The inset shows the list of parameters obtained by fitting of 3D diffusion model.
Raman spectra of glycine solution at different focus positions near the liquid-air interface:

Raman spectrum of a glycine/water solution (SS = 1.2) was measured at different focal depths with respect to the liquid-air interface. At this interface, the focus position is specified as zero \( (z = 0 \, \mu m) \), where negative values go into solution and positive values go into the air region. As the confocal volume approached the interface from solution, Raman signal intensity decreased as the confocal volume included more of the air region \( (z = \pm 2 \, \mu m, \text{Fig. S14a}) \). All spectra were normalized at 3391 cm\(^{-1}\) (stretching mode of OH in water) to assess any spectral shape variation (Fig. S14b). In addition, the intensity of the main normalized peaks showed no significant variation as the focal depth was modified (Fig. S14c).

**Fig. S14.** Raman spectra of glycine/water solution (SS = 1.2) at different focus positions. a, Overlay of Raman spectra at different focus positions. b, Normalized Raman spectra at different focus position. c, Intensity of four different peaks as a function of focus position, relative to the liquid-air interface.
Section SI2. Computational simulations:

SI2.A Molecular Dynamics of Glycine in water:

Molecular dynamics simulations were run on the zwitterionic form of glycine in a 8 nm³ water box using the OpenMM toolkit(8). Simulations were run with 16 and 25 glycines (3.3 mol L⁻¹ and 5.2 mol L⁻¹) using the AMOEBA 2013 forcefield, water14 parameters, and zwitterionic glycine parameters obtained from(9). The input .pdb files were created using the OpenMM Setup platform to generate a water box with an explicit water model. All simulations were run with a step size of 0.5 fs, a Verlet integrator with a constraint tolerance of 1x10⁻⁶, and an Andersen thermostat with a temperature of 300 K and a collision frequency of 1 ps⁻¹. The nonbonded method was set to the Particle Mesh Ewald method to apply periodic boundary conditions with a cutoff of 1 nm and an error tolerance of 1x10⁻⁵. Additionally, the Lennard Jones interaction cutoff was set to 1 nm and the extrapolated polarization method was used. The simulations of 16 Gly with 211 water molecules (3.3 mol L⁻¹) and 25 Gly with 171 water molecules (5.2 mol L⁻¹) were run using the CUDA platform with mixed CUDA precision. First, an energy minimization was performed followed by 1000 steps for equilibration and 40,000,000 simulation steps (20 ns). Trajectory frames were recorded every 20,000 steps (0.01 ns). The trajectory frames were centered using the “center_coordinates()” method from the MDTraj Python library and analyzed using the MDAnalysis Python toolkit (10). The “select_atoms()” method was used to identify all other Gly-component atoms within 2.5 Å of each Gly residue and to construct instantaneous contact maps. The network analysis on the glycine contact maps reported in Fig. 4b in the main text was performed using the networkX python library. Building on the network analysis, for each snapshot in the high concentration (5.2 mol L⁻¹) ensemble, we have identified glycine residues which were involved in clusters with four or more members. We have counted the number of connections for each, added them together and divided by the total number of glycines involved in all large networks (four or more members) to get the average number of nearest neighbors. We then built a histogram of all averages for nearest neighbors in large networks in the trajectory (Fig. 4c in the main text) and found that on average in large networks glycines had only 2.05 neighbors. This result suggests that large aggregates of glycines in our simulations tend to be elongated chain-like structures.

SI2.B Raman spectra calculations:

The Raman spectra calculations were performed using the Gaussian 16 suite of programs(11) on the Compute Canada supercomputer infrastructure. Three typical clusters with two, three, and four glycine molecules were isolated from snapshots in the molecular dynamics trajectories described in SI2.A. These clusters are shown in Fig. 4c-e in the main text. A quantum chemistry calculation on glycines only (without waters) was set up using the B3LYP functional with 6-311G(d) basis set. Constraints on distances between key atoms were imposed in order to prevent the configuration of the glycine cluster from changing significantly during the geometry optimization step, particular care was taken to prevent new hydrogen bonds from forming during the optimization. The resulting Raman spectra were shifted to align the peaks corresponding to the C-H stretch mode with the C-H stretch mode in the experimental spectra and the intensity was increased by a factor 8000 in order to match the experimental intensities.
References

1. Q. Jiang, A. G. Shtukenberg, M. D. Ward, C. Hu, Non-Topotactic Phase Transformations in Single Crystals of β-Glycine. *Cryst. Growth Des.* **15**, 2568–2573 (2015).

2. M. E. Peter, P. Ramasamy, Growth of gamma glycine crystal and its characterisation. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **75**, 1417–1421 (2010).

3. A. Datta, A. Hossain, Solubility Data of Glycine in Water and Justification of Literature Results. *Asian J. Chem.* **32**, 1525–1533 (2020).

4. B. D. Saksena, Raman spectrum and molecular association in formamide. *Proc. Indian Acad. Sci. - Sect. A* **11**, 53–61 (1940).

5. S. E. M. Colaianni, O. F. Nielsen, Low-frequency Raman spectroscopy. *J. Mol. Struct.* **347**, 267–283 (1995).

6. V. Buschmann, B. Krämer, F. Koberling, R. Macdonald, S. Rüttinger, Quantitative FCS: determination of the confocal volume by FCS and bead scanning with the microtime 200. *Appl. Note PicoQuant GmbH Berl.* 2009.

7. T. A. Laurence, S. Fore, T. Huser, Fast, flexible algorithm for calculating photon correlations. *Opt. Lett.* **31**, 829 (2006).

8. P. Eastman, *et al.*, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **13**, e1005659 (2017).

9. A. Esser, S. Belsare, D. Marx, T. Head-Gordon, Mode specific THz spectra of solvated amino acids using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* **19**, 5579–5590 (2017).

10. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

11. M. J. Frisch, *et al.*, *Gaussian 16 Rev. C.01* (2016).