Biophysical Journal, Volume *119*

Supplemental Information

Unraveling VEALYL Amyloid Formation Using Advanced Vibrational Spectroscopy and Microscopy

Steven J. Roeters, Mathias Sawall, Carl E. Eskildsen, Matthijs R. Panman, Gergely Tordai, Mike Koeman, Klaus Neymeyr, Jeroen Jansen, Age K. Smilde, and Sander Woutersen

Supplementary material for:
Unraveling VEALYL amyloid formation using advanced vibrational spectroscopy and microscopy

Steven J. Roeters[1,7*], Mathias Sawall[2], Carl E. Eskildsen[3], Matthijs Panman[1,8], Gergely Tordai[1], Mike Koeman[4], Klaus Neymeyr[2,5], Jeroen Jansen[4], Age K. Smilde[6], and Sander Woutersen[1]

[1]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands
[2]Institut für Mathematik, Universität Rostock, Ulmenstraße 69, 18057 Rostock, Germany
[3]Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands
[4]Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6500 GL Nijmegen, The Netherlands
[5]Leibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock, Germany
[6]Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands
[7]Current address: Kemisk Institut, Aarhus Universitet, Langelandsgade 140, 8000 C Aarhus, Denmark
[8]Current address: Department of Chemistry and Molecular Biology, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden
[*]Correspondence: s.j.roeters@chem.au.dk

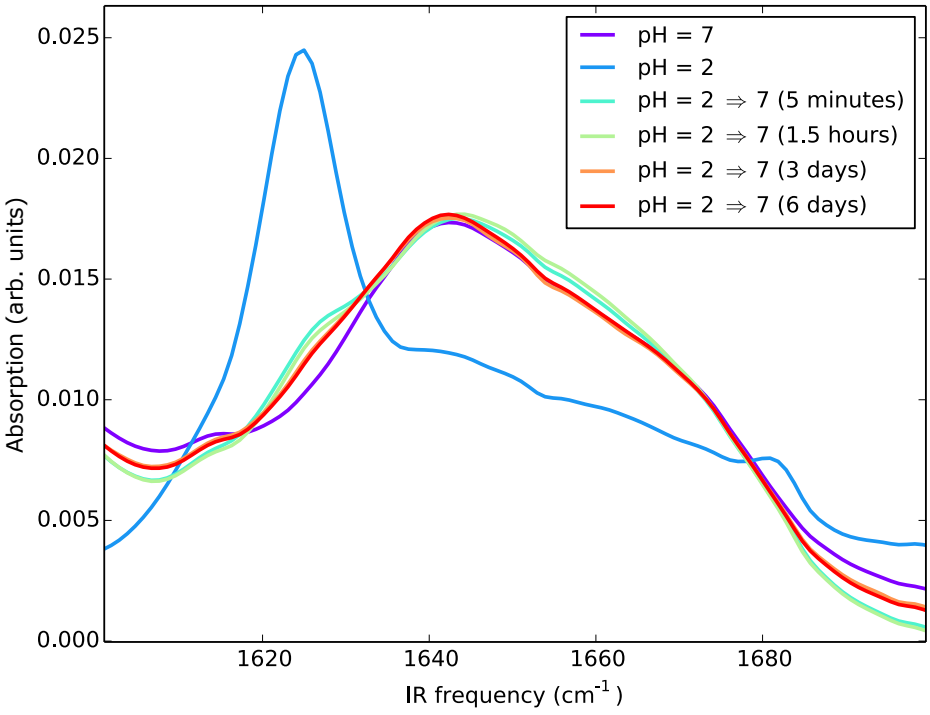

Figure S1: pH Reversibility of VEALYL aggregation at a 25 mM concentration, showing that already 5 minutes after raising the pH from 2 to 7, most of the amyloid β-sheet structure is dissolved. However, even after 6 days the spectrum is not completely similar to the monomeric, random coil spectrum of freshly-solvated VEALYL at pH = 7.

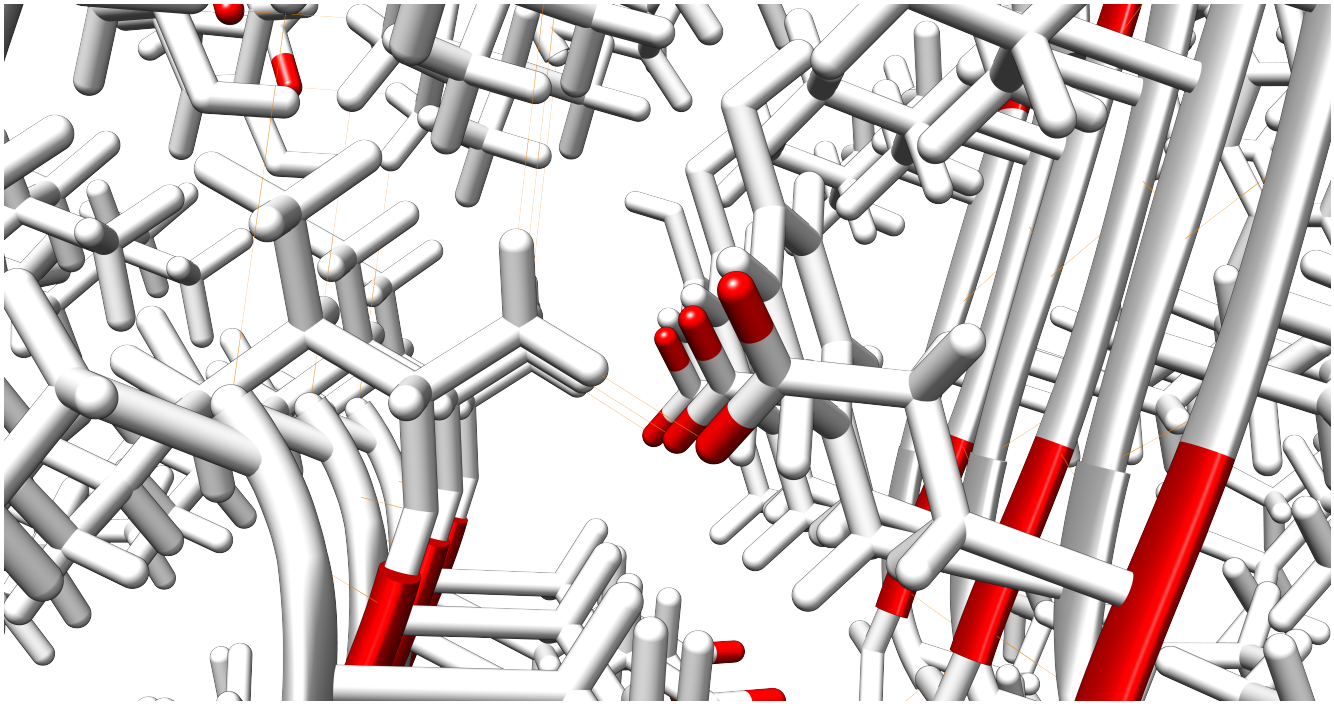

Figure S2: A zoom-in on the environment of the Glu sidechains (with its oxygen and — in ribbon style — backbone atoms marked in red) in the crystal structure of VEALYL (PDB-entry 2OMQ), that shows incomplete hydrogen bonding of the Glu carboxyl/carboxylate group. The thin orange lines indicate hydrogen bonds.

Computation of the feasible spectra and the feasible concentration profiles

For the analysis of the rotational ambiguity of the data sets we use the *FACPACK*-implementation of the polygon inflation method, introduced in ref. 1. For the AFS, only the first row of T is analyzed and a special scaling with $T(1,1) = 1$ is used. The AFS reads:

$$\mathcal{M} = \{x \in \mathbb{R}^s : \exists \boldsymbol{T} \in \mathbb{R}^{s\times s} \text{ with } \mathrm{rank}(\boldsymbol{D}) = s,\ \boldsymbol{T}(1,:) = (1, x^T) \text{ and } \boldsymbol{C},\ \boldsymbol{A} \geq 0\}.$$

If three physical components are assumed to describe the experimental data ($s = 3$), the AFS is a subset of $\mathbb{R}^2$.

Since the data $\boldsymbol{D}$ includes perturbations, we used a control parameter $\varepsilon \geq 0$ (like in equation (6) of ref. 1), to allow a certain degree of negativity. This would otherwise pose a problem to the non-negativity constraint of both the concentration ($\boldsymbol{C}$) and spectral ($\boldsymbol{A}$) profiles. For the AFS computation a certain factorization with $\boldsymbol{C}$ and $\boldsymbol{A}$ is accepted if:

$$\frac{\min(\boldsymbol{A}(i,:))}{\max(\boldsymbol{A}(i,:)} \geq -\varepsilon,\ i = 1,\ldots,3, \qquad \text{as well as} \qquad \frac{\min(\boldsymbol{C}(:,j))}{\max(\boldsymbol{C}(:,j)} \geq -\varepsilon,\ j = 1,\ldots,3. \tag{S1}$$

In other words: the control parameter is a relative boundary for negative entries for each profile (see ref. 1 for more details on ε, which is set to 0.035 in this calculation).

The subsets of the concentration AFS are computed analogue.

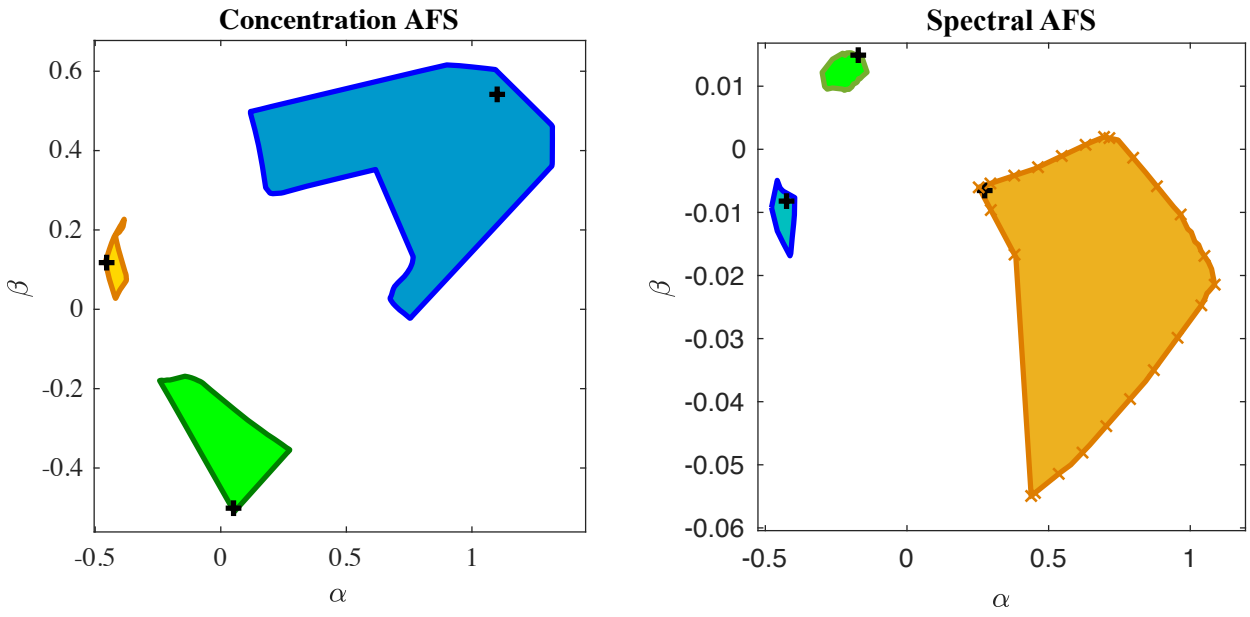

Figure S3: The AFS segments of the time-dependent FTIR spectra of aggregating VEALYL peptides, with the same color coding as in all other graphs: species 1, 2 and 3 are depicted in blue, green and orange. Herein $T \in \mathbb{R}^{3\times3}$ is used for the transformation from the SVD to the original factors as in main text eq. 1.

MCR-ALS calculation and results

In a multiple curve resolution with alternating least square (MCR-ALS) analysis, a $m \times n$ data set $\boldsymbol{D}$ is fitted as follows (2):

$$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^T + \boldsymbol{E} \tag{S2}$$

with $\boldsymbol{C}$ the $m \times s$ 'score matrix' (with s the number of physical components), $\boldsymbol{S}$ the $n \times s$ 'loading matrix' and $\boldsymbol{E}$ the 'residual error' or 'noise matrix', while taking certain constraints relating to the underlying physics into consideration. In an MCR-ALS procedure, $\mathbf{E}$ is minimized with alternating least-squares fitting, meaning that the concentration and spectral profiles are fitted in an alternating fashion, *i*.e. first optimizing the time profiles assuming a certain initial guess for the spectral profiles, then optimizing the spectral profiles assuming the optimal values for the time profiles from the previous step, et cetera — or vice versa, with respect to time and frequency space. In each iteration a least-squares problem is solved, possibly with constraints.

Examples of commonly applied constraints for a time-dependent spectral data set are non-negativity of the spectral and time profiles, the fixation of the first and/or last spectrum to the spectrum of the first and/or last component, and 'concentration closure' (conservation of the total amount of molecules). When such constraints are applied, the spectral and temporal profiles of the 3 significant (as judged by the singular values of the SVD analysis, see main text fig. 2) species can be resolved by the MCR-ALS analysis (see fig. S4).

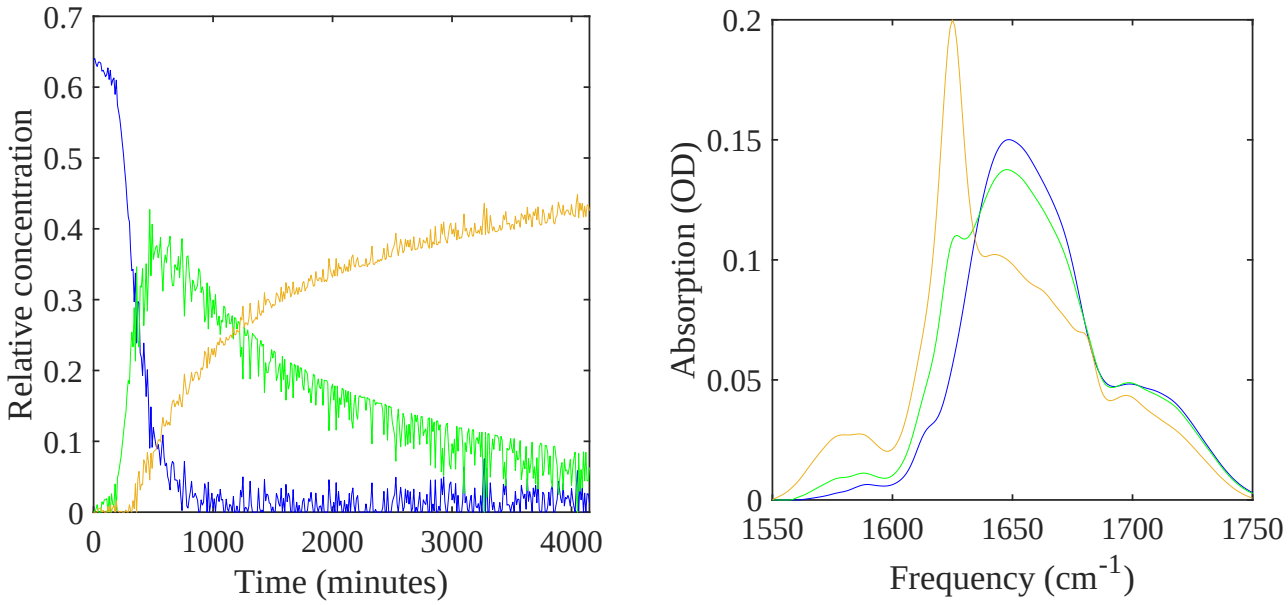

Figure S4: Concentration profiles (left) and spectral profiles (right) resulting from an MCR-ALS analysis with only non-negativity constraints for both the spectra and time profiles, with the spectra at t = 0, 250 and 5250 minutes as the initial guess. The solution has a lack of fit of 0.461%.

The MCR-ALS method allows for an arbitrary number of species to be fitted to the data (as opposed to the FACPACK software that currently maximally handles four species), so another assessment of the number of resolvable species can be made by determining what the lack of fit is as a function of the number of species fitted (see fig. S5).

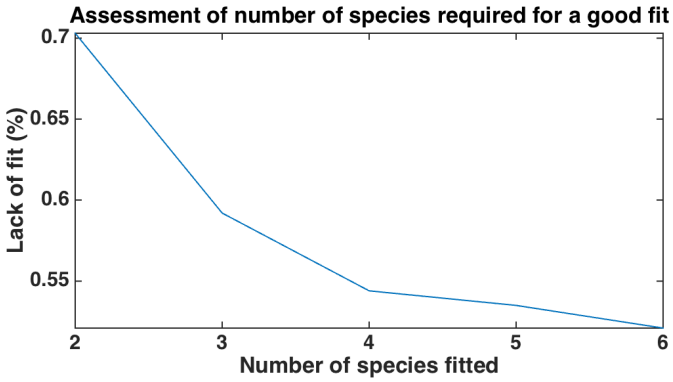

Figure S5: The lack-of-fit versus the number of species used in an MCR-ALS fit assuming concentration closure (*i.e.* conservation of the number oscillators and — in 1D-IR spectroscopy — also conservation of the total absorption). The experimental error was determined to be 0.03%, by measuring 100 spectra of a constant protein sample with a comparable absorbance.

Also with this approach the amount by which the match between the experimental and modeled data decreases when going from one to two intermediates. When more than four species (*i.e.* two intermediates) are assumed, the fit does not improve, indicating that it is impossible to resolve more species. The presence of a second MCR-ALS resolvable intermediate could however still be argued from the improved fit (fig. S5), but the resulting profiles appear to be very sensitive to the initial guess (see fig. S6).

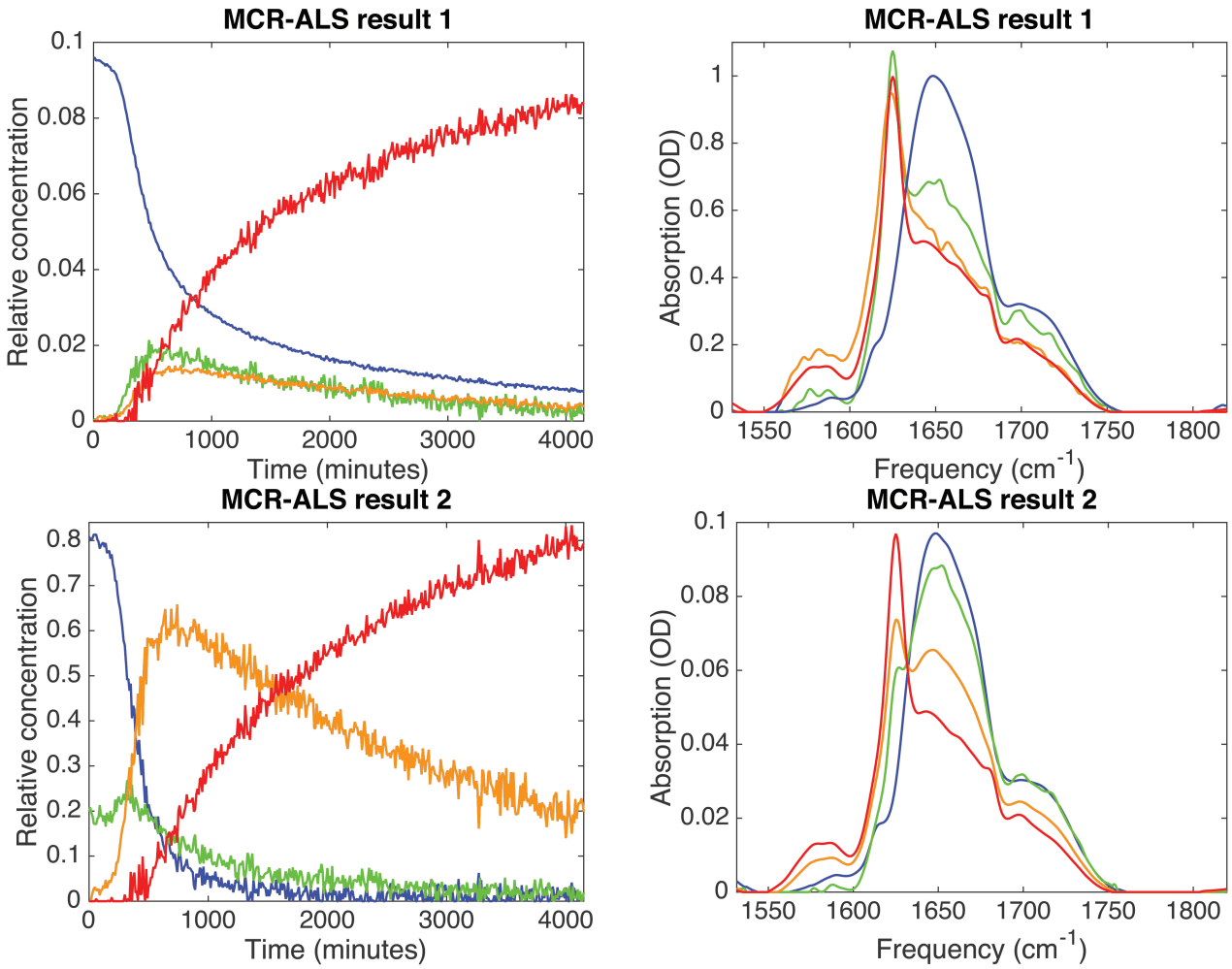

Figure S6: Two different solutions obtained with an MCR-ALS analysis with only non-negativity constraints for both the spectra and time profiles, showing the dependency on the initial guess. When the selected solution from the FACPACK analysis is chosen as an initial guess for the spectra, the upper result is obtained, while for a fit that has the spectra at t = 0, 250, 1000 and 5250 minutes as the initial guess, the lower result is obtained. The solutions have a lack of fit of 0.341% and 0.343%, respectively.

The variations resulting from the different initial guesses however do not lead to differences in the qualitative interpretation, not even with respect to the three-species solution: in all cases an increased β-sheet content and an increased deprotonation occurs during the conversion of the monomers into the intermediates and then into the fibrillar species. The number of species that can be resolved appears to depend both on the signal-to-noise ratio of the data, as well as on the methodology of the analysis, but as discussed in the main text *Conclusion* section, in reality there will probably be a continuum of aggregating species that is increasing in size with time, until the aggregates are converted into fibrils.

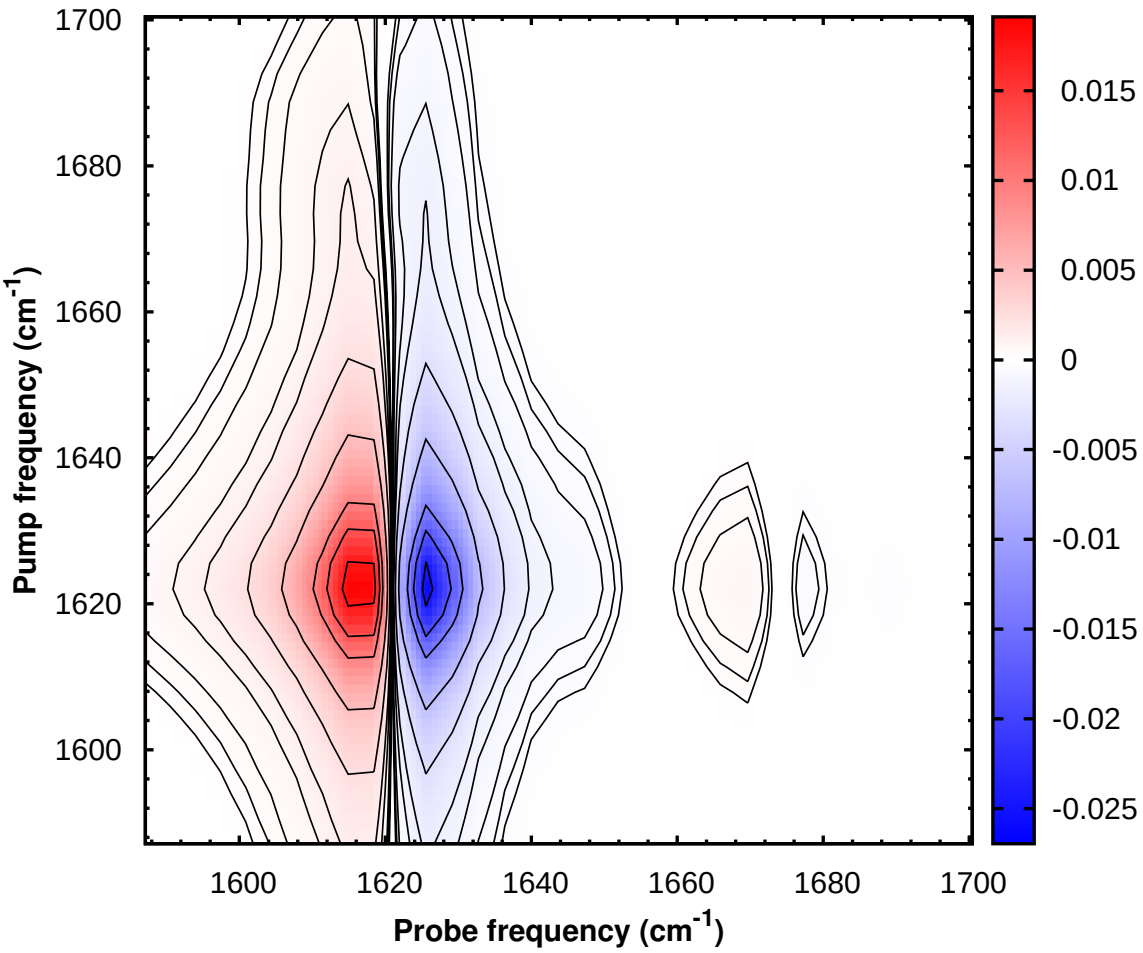

Figure S7: 2D-IR spectrum calculated with the one-exciton Hamiltonian formalism, based on an 50 amino-acid-long amyloid VEALYL fibril, generated from the unit cell of PDB entry 2OMQ in Chimera(3). The couplings between the amide-I modes of the peptide groups of the peptide backbone are estimated with the transition-dipole coupling model(4), resulting in a well-reproduced peak splitting and cross-peak pattern. The lack of intensity around 1640 cm^{-1} in comparison with the experimental spectrum (see main text fig. 4) is probably due to the fact that there is still a small amount of random-coil structure, which is consistent with the result of the MCR analyses.

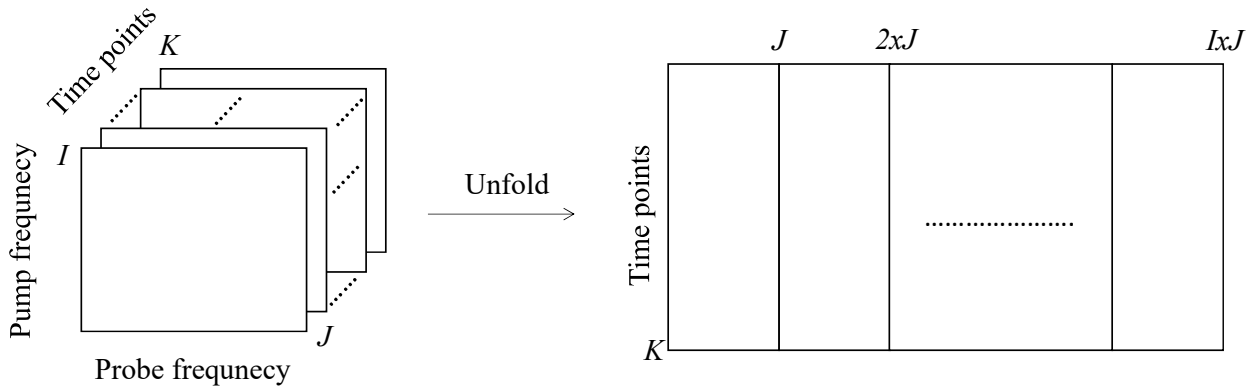

Figure S8: Visual representation of how the time-dependent 32 probe pixels x 32 pump pixels x 122 time points data set is unrolled, or flattened, into a 1024 x 122 data set, so that the SVD analysis can be performed in a straightforward fashion.

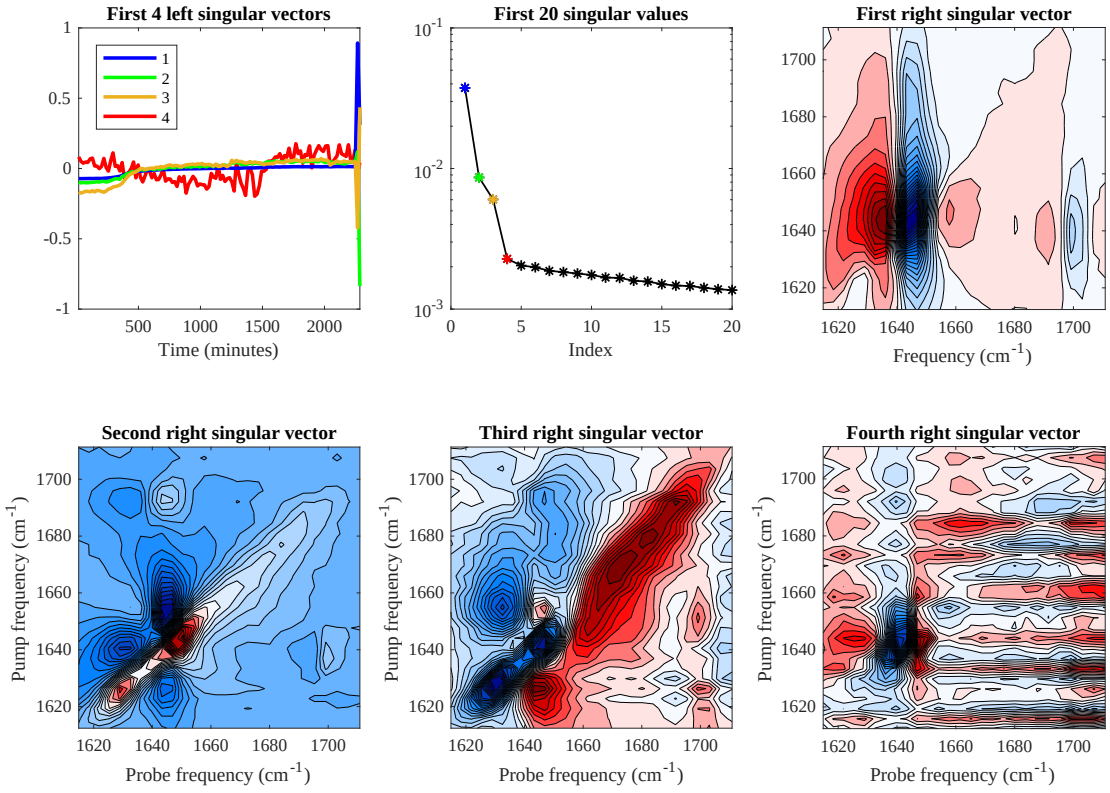

Figure S9: SVD analysis of the unfolded, or flattened, time-dependent 2D-IR data set. The singular values corroborate the conclusion from the 1D-IR SVD, namely that there are three molecular species that can be discerned with IR spectroscopy. The right singular vectors are depicted in their refolded 32 x 32 form. The right singular vector corresponding to the fourth largest singular value shows large noise contributions. To aid the analysis, also the spectrum recorded at 400 μ m away from the original spot and the 5-days incubated spectrum (the last two spectra depicted in main text fig. 4, not shown in this figure) were included at the end of the time-dependent spectral data set, providing more mature fibril structures.

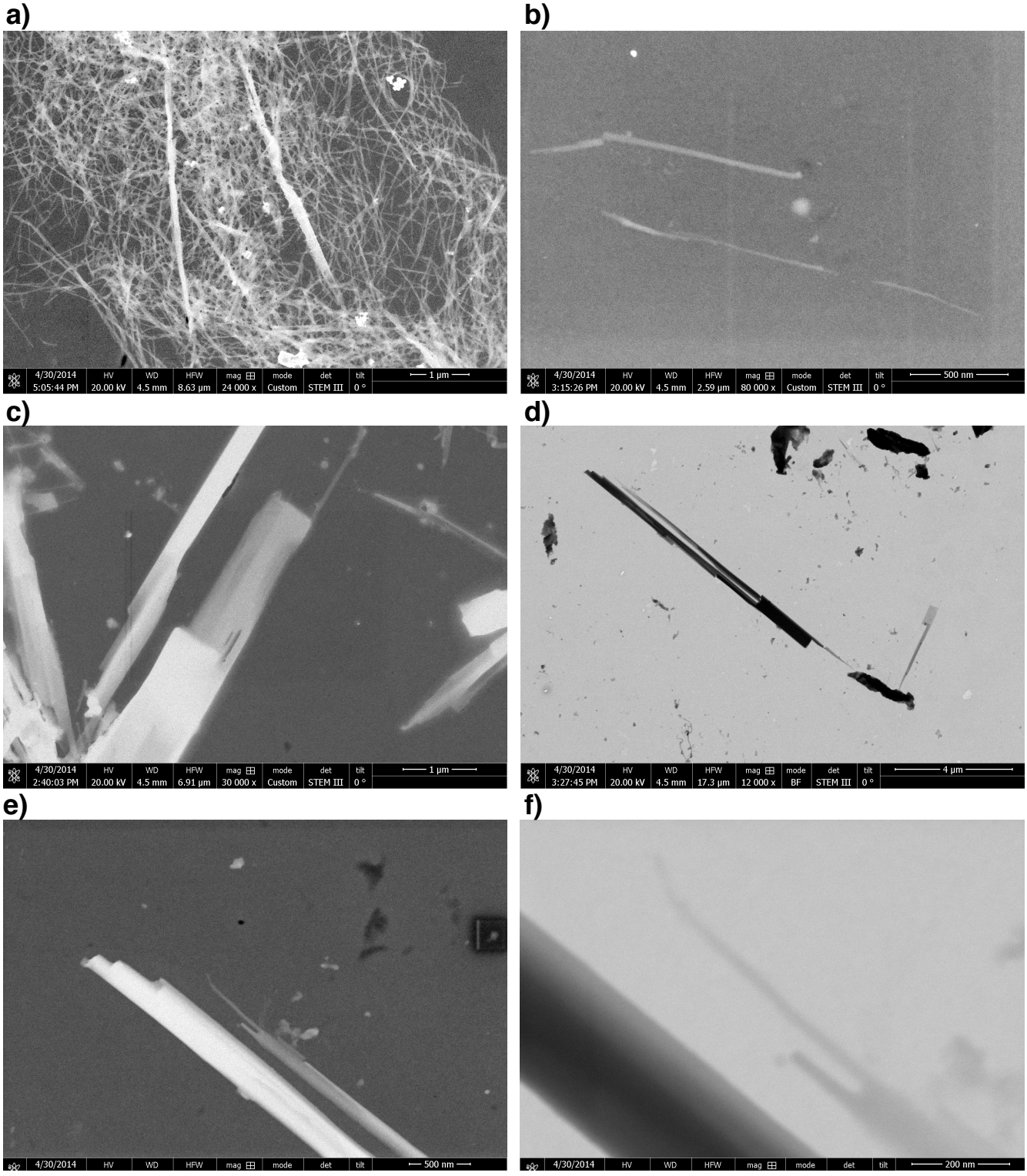

Figure S10: STEM images of a) the sample at 50% fibrillation (similar to the 3rd and 4th image on row 7 of main text fig. 7) and b-f) the end state of VEALYL aggregation (similar to the 5th image on row 7 of main text fig. 7), indicating the lack of any discernible twist, which is in line with the observed 'reverse fibril' VCD spectrum.

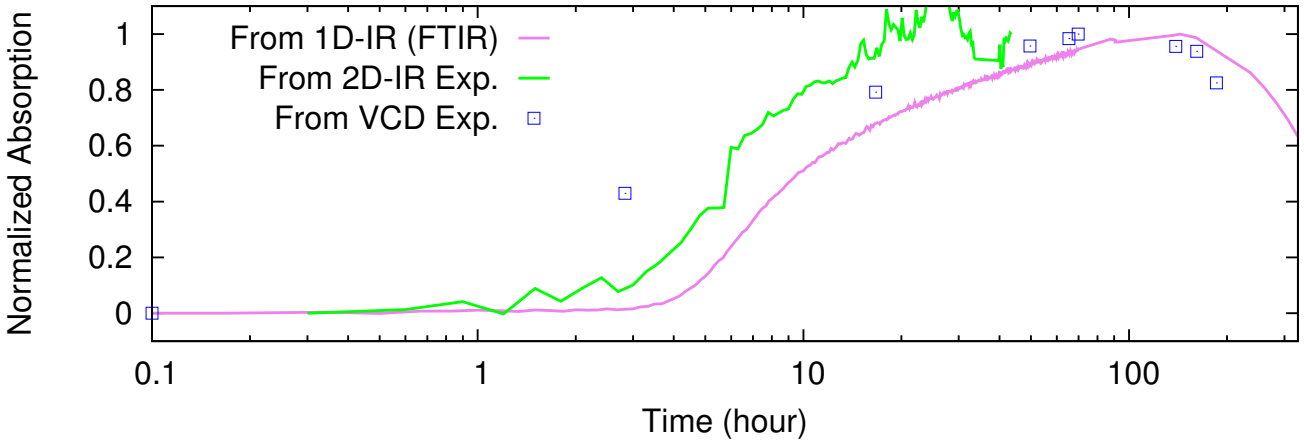

Figure S11: Time traces of the maximum of the 1D-IR absorption peak of the low-frequency amyloid β-sheet peak (~1620 cm^{-1}), measured with-, or derived from the three different vibrational-spectroscopy techniques used in this study. The purple trace comes directly from the data set which is analyzed in the SVD, FACPACK and MCR-ALS analyses and whose time dependence is depicted in the fifth row of main text fig. 7, while the green line is derived from the 2D-IR spectra depicted in main text fig. 4, and the blue squares are derived from the VCD experiments depicted in main text fig. 5, whose corresponding 1D-IR spectra are depicted in main text fig. 1. These traces show the variation in the lag time in the three different experiments, and how the (derived) 1D-IR response can be used as an internal calibration of the progress of the process under study. The internal calibration can be especially useful when spontaneously aggregating amyloid systems are studied, which often exhibit a large variation in the duration of the lag phase (5) that precedes the observation of β-sheets. The short-time scale variation in the 2D-IR signal may be due to small changes in the pointing of the beam that irradiates the heterogeneous sample, as the signal only stems from a beam spot on the order of 250 μm.

SUPPORTING REFERENCES

1. Sawall, M., C. Kubis, D. Selent, A. Börner, and K. Neymeyr, 2013. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems. I: concepts and applications. *J. Chemom.* 27:106–116.

2. Jaumot, J., A. de Juan, and R. Tauler, 2015. Chemometrics and Intelligent Laboratory Systems. *Chemom. Intell. Lab. Syst.* 140:1–12.

3. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, 2004. UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–12.

4. Krimm, S., and J. Bandekar, 1986. Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *In* Advances in protein chemistry, Elsevier, volume 38, 181–364.

5. Arosio, P., T. P. J. Knowles, and S. Linse, 2015. On the lag phase in amyloid fibril formation. *Phys. Chem. Chem. Phys.* 17:7606–7618.