

Quantum Cascade Laser Infrared Microscopy and 2D IR Correlation Spectroscopy: Improves Crystallization Screening of a Protein Complex

Sherly Nieves and Belinda Pastrana*

Protein Dynamic Solutions, Inc., Wakefield, MA

*Corresponding author: Protein Dynamic Solutions, Inc., 9 Audubon Road, Wakefield, MA 01880; Phone: 978-594-6955; E-mail: belinda@pdsbio.com

Abstract

X-ray diffraction of crystallized protein continues to be the preferred means of obtaining high-resolution structures of proteins and their complexes. These structures are crucial for drug design, but the screening and optimization of the conditions that produce well-diffracting crystals represent a bottleneck in the structure determination process. A quantum cascade laser (QCL) infrared microscope, was used to determine protein aggregation, distinct from self-association, which is crucial to the success of any crystallization effort. Hyperspectral images of an aliquot from a vapor diffusion hanging drop crystallization screen were acquired over a small temperature range (30 – 38 °C) at intervals of 2 °C. QCL infrared (IR) spectral data were subjected to 2D IR correlation analysis to describe *Homo sapiens* centrin 2(E32A)-Sfi1p₂₁ complex (1:1.5 molar ratio) and the selective aggregation of the target peptide. To our knowledge this is the first time such a level of molecular understanding has been achieved.

Key words: quantum cascade laser infrared microscopy, hyperspectral image, 2D infrared correlation spectroscopy, protein complex crystallization screening, protein aggregation

Introduction

Protein crystallization is the preferred means of obtaining high-resolution structures of proteins and their complexes. These structures are crucial for drug design, yet methodological barriers to their determination still exist [1-4]. Efforts to make the structure determination process more efficient include data collection methods such as the synchrotron beam lines [5], detection via sensitive focal plane arrays, and improved cryogenic and mounting procedures for crystals. Yet, one key step in successful high-throughput X-ray crystallography continues to be a bottleneck, namely, the screening and optimization of the conditions required to produce well-diffracting crystals. This process includes the optimization of multiple variables: precipitant, pH, temperature, protein sequence and concentration, etc. [4]. New techniques are needed if we hope to increase the rate of success of this step, which is currently ~18%.

There are three stages in the crystallization process: nucleation; crystal growth, which is governed by diffusion; and cessation of crystal growth. Crystal growth requires specific interactions between individual protein molecules that lead to an organized crystal lattice. If the crystallizability of the protein of interest is dependent on the type of protein aggregate generated during the crystallization process, then our proposed method should allow for this assessment. Conceptually, the aggregate should be one of self-association in which chemical associations driven by the weak interactions between proteins and/or the precipitating agent are critical to maintaining the protein's conformational stability. Self-association in principle would account for the generation of a crystal lattice, while excluding solvent molecules from the immediate hydration shell of the protein and increasing the influence of the surface charge of the protein to account for the crystallization event. The evaluation of the crystallization process of proteins and protein complexes using this direct and highly informative approach would allow for further understanding of the process of crystallization and should be explored even further during crystallization optimization as part of the screening process. It may also be possible to evaluate the precipitant's role in changing the dynamics of the protein's interaction with its aqueous environment, causing a synergistic effect that will lead to the successful crystallization of the protein or the protein complex.

Here we have employed quantum cascade laser (QCL) microscopy for both the determination of protein aggregation, distinct from self-association, under various crystallization conditions. The platform includes a QCL infrared (IR) microscope with enhanced signal-to-noise ratio, accessories, cells under thermal control, and software for QCL IR spectral analysis, and assessment of crystallizability. Hyperspectral images (HSIs) within the mid-IR spectral region of 1800 - 1000 cm^{-1} provide data distinguishing an aggregate from a crystal as well as a rough determination of the crystal or aggregate size. The QCL IR spectral data are subjected to 2D IR correlation spectroscopy analysis [6, 7] to determine the extent and molecular mechanism of aggregation and/or crystallization under thermal stress.

A high-resolution structure for a yeast homologue of centrin known as CDC31 in the form of CDC31-Sfi1p complex exists as PDB ID: 2DOQ (Figure 1) [8], and our group has

extensively studied *Hscentrin* isoforms using multiple biophysical techniques including differential scanning calorimetry and circular dichroism, FT-IR, and 2D IR correlation spectroscopies [9]. The current study involved a complex between a centrin variant (*Hscentrin* 2 (E32A)) and its target peptide, Sfi1p₂₁, which is presented as proof-of-concept. 2D IR correlation spectroscopy provided molecular insight into which of the components was aggregated and how the initial interaction of Sfi1p₂₁ with the centrin variant led to the selective aggregation of the peptide. To our knowledge this is the first time that such a level of molecular understanding has been achieved for a crystal screen. The results obtained demonstrate the breakthrough capabilities of the quantum cascade laser infrared microscope along with 2D IR correlation analysis for the evaluation of the crystallizability of a protein or protein complex.

Experimental

Crystal screen setup

We performed a vapor diffusion hanging drop crystal screen containing 2 μ L of the *Hscentrin* 2 (E32A) variant and *HsSfi1p*₂₁ synthetic peptide (1:1.5 molar ratio) for a 5 mM total protein concentration and 2 μ L of the Index screen from Hampton Research (Aliso Viejo, CA) were placed in a vapor diffusion setup tray from the Easy Xtal Tool from Qiagen Sciences (Germantown, MD). The visual evaluation of the screen was done using a Nikon SMZU microscope for selection of potentially diffracting crystals. We identified 1 condition to contain both amorphous and microcrystals: condition #31 of the index screen. This crystallization solution contained 0.1 M Tris at pH 8.5, 0.5% (w/v) polyethylene glycol monomethyl ether 5000, and 0.8 M potassium sodium tartrate tetrahydrate.

A 1 μ L aliquot from a selected hanging drop identified to contain amorphous and/or microcrystals was drawn from the hanging drop, placed in a predefined well of a custom milled CaF₂ slide, and covered with a fully polished crystal to make up the slide cell. The slide cell was then placed in a heated accessory with accurate thermal control, and HSIs were acquired within the temperature range of 30-38 °C with 2 °C intervals and 4 min equilibration periods using the ProteinMentor™ a QCL IR microscope from Protein Dynamic Solutions, Inc. (Wakefield, MA).

Hyperspectral image (HSI) acquisition

The QCL infrared microscope acquires the HSIs, allowing for a linear response microbolometer focal plane array (480 x 480 pixels) detector to be used. A low magnification objective (4x) with a numerical aperture of 0.3 numerical aperture and a 2 x 2 mm² field of view provides a pixel size of 4.25 x 4.25 μ m². The HSIs are composed of 223,000 QCL IR spectra and were collected at 4 cm⁻¹ resolution within the spectral region of 1750-1500 cm⁻¹. To prevent coherence effects due to QCL IR fluctuations, the background was collected at each temperature once thermal equilibrium (4 min) was achieved.

QCL IRM spectral overlay and 2D IR correlation analysis for this sample was performed using the Kinetics program of MATLAB (MathWorks, Natick, MA), which was generously provided by Dr. Erik Goormaghtigh from the Free University of Brussels,

Belgium. However, we have developed fully automated Correlation Dynamics software to analyze an array of samples.

Results and Discussion

Size of the aggregate

HSIs for the *Hscentrin* 2 (E32A)-Sfi1p₂₁ sample at five different temperatures (30, 32, 34, 36, and 38 °C) are shown in Figure 2. Initially, microcrystals were observed in a circular arrangement due to the dispensing of the pipette, which allows for the amorphous or microcrystals to flow from the tip to the well during loading (Figure 2). A micron scale was used to determine the size of the aggregates observed in the HSI the size of the aggregate was determined to be 30 μm x 40 μm at 30 °C. Furthermore, as the temperature was increased, the aggregate continued to grow, reaching 300 μm x 400 μm in size at 38 °C. The corresponding QCL IR spectra (Figure 3a) were indicative of the presence of aggregate, thus confirming the HSIs (Figure 2). Specifically, the full-width at half height of the amide I band and the presence of a shoulder at 1613 cm^{-1} are consistent with the presence of aggregates. While 2D IR correlation analysis yielded the typical synchronous plot observed for a protein aggregation process during thermal perturbation (as indicated by the prominent auto peak 1613 cm^{-1} present in the synchronous plot), the asynchronous plot suggested an unfolding event during the aggregation process, as indicated by the cross peaks located within 1613 cm^{-1} .

Band assignments

2D IR correlation spectroscopy was performed to improve our understanding of the backbone dynamics and side chain interactions involved in the aggregation process within a crystallization condition. Band assignments were made using the amino acid sequence of *Hscentrin* 2 obtained from the National Center for Biotechnology Information database (NCBI accession number: EAW72900). The band assignments are summarized in Table I. The vibrational modes associated with *Hscentrin* 2 (E32A) include the p-aromatic overtones for phenylalanine residues (1702.0 and 1717.5 cm^{-1}) and calcium binding site loops (1682.6 and 1672.0 cm^{-1}). EF-hand proteins also have hinge loops, which we have assigned to 1662.5 cm^{-1} . We also assigned the helical component (1652.5 cm^{-1}) and the short β -sheet segments (1632.0 cm^{-1}). Finally, the aspartate carboxylate stretching side chain vibrational mode $\nu(\text{COO}^-)$, mainly located within the calcium binding sites within centrin, was assigned to 1577.0 cm^{-1} . Shared vibrational modes were the glutamate carboxylate side chain mode $\nu(\text{COO}^-)$ (1550.3 cm^{-1}), found in both centrin and the Sfi1p₂₁ target peptide. Exclusive to Sfi1p₂₁ were two vibrational modes due to aggregation (1613.0 cm^{-1}) and the two sets of contiguous histidine residues with the stretching vibrational mode $\nu(\text{C}=\text{C})$ (1597.2 cm^{-1}). These band assignments are consistent with the secondary structure information available in the high-resolution crystal structure of the complex and with previous work from our laboratory [8, 9]. Also, the side chain modes and their molar extinction coefficients have been determined [10] and reviewed [11, 12].

2D IR correlation analysis

The protein sample is in an aqueous environment and the molar extinction coefficient of pure H₂O is high at 55.5 $\text{M}^{-1} \text{cm}^{-1}$, yet like any protein containing sample it has effectively diluted the contribution of H₂O in the overall absorbance spectrum. Also, important is the

decreased path-length allowing for the management of samples that exhibit high absorptivity. Finally, the QCL IRM transmission absorbance spectra have an enhanced signal-to-noise ratio allowing for the difference spectra approach to be used thus overcoming the common deterrents of using IR spectroscopy for aqueous samples. The QCL IR spectrum for the region of interest at the initial temperature (30 °C) was subtracted from all subsequent spectra, thereby generating the difference spectra. It is to this data set that the 2D IR correlation function is applied to generate the synchronous and asynchronous plots.

Difference spectra used in 2D IR correlation spectroscopy are defined as:

$$\tilde{A}(\nu_j, t_k) = \begin{cases} A(\nu_j, t_k) - \bar{A}(\nu_j) & \text{for } 1 \leq k \leq m \\ 0 & \text{otherwise} \end{cases} \quad (\text{i})$$

where $\bar{A}(\nu_j)$ is the initial spectrum of the data set to generate the covariance spectra. Synchronous 2D correlation intensities of the covariance spectral data are defined by:

$$\Phi(\nu_1, \nu_2) = \tilde{A}(\nu_1, t_j) \cdot \tilde{A}(\nu_2, t_j) \quad (\text{ii})$$

where the resulting correlation intensity $\Phi(\nu_1, \nu_2)$ as a function of two independent wavenumber axes, ν_1 and ν_2 , is the synchronous plot.

Asynchronous 2D correlation intensities of the covariance spectral data are defined by:

$$\Psi(\nu_1, \nu_2) = \tilde{A}(\nu_1, t_j) \cdot N_{ij} \tilde{A}(\nu_2, t_i) \quad (\text{iii})$$

where the term N_{ij} is the element of the so-called Hilbert-Noda transformation matrix, given by:

$$N_{ij} = \begin{cases} 0 & \text{for } i = j \\ \frac{1}{\pi(j-i)} & \text{otherwise} \end{cases} \quad (\text{iv})$$

These 2D IR correlation plots provide enhanced resolution of the peak intensity changes and positions within the spectral region of interest that were due to the temperature increase. Specifically, the *synchronous plot* was used to establish relationships between peaks that change in the same phase with one another (Figure 3b). This plot has peaks on the diagonal known as *auto peaks*, which provide the main peak intensity changes. In addition, peaks off the diagonal provide the relationship between the auto peaks and are known as *cross peaks*. In this case, as shown in Figure 3b, the only intensity change was at 1613 cm⁻¹ and was due to the aggregation process. The synchronous plot did not contain any negative peaks, thus simplifying the analysis even further. The *asynchronous plot* (Figure 3c) establishes the relationship between peaks that are changing out-of-phase from

one another. In general, there are no peaks on the diagonal and only cross peaks are observed. This plot provides enhanced resolution and is used to determine the sequential order of molecular events that describe the aggregation process during the temperature perturbation. The order of molecular events was determined by following Noda's rules [6, 7]. Keeping in mind that asynchronous plots are symmetrical in nature, and by convention we always refer to the top triangle for analysis, we apply the rules as follows:

- I. If asynchronous cross peak v_2 is positive, then v_2 is perturbed prior to v_1 ($v_2 \rightarrow v_1$).
- II. If asynchronous cross peak v_2 is negative, then v_2 is perturbed after to v_1 ($v_2 \leftarrow v_1$).
- III. If the corresponding synchronous cross peak is positive, then the order of the events is established using the asynchronous plot (rules I and II).
- IV. However, if the corresponding synchronous cross peak is negative and the asynchronous cross peak is positive, then the order is reversed.

The order of events can be established for each peak observed in the v_2 axis (Figure 3c). In this specific case, the only two rules that apply are rules I and II. Our group has successfully applied 2D IR correlation spectroscopy to the study of numerous proteins and peptides, including a comparative analysis of related human centrin [9].

Description of the molecular aggregation process

Figure 4 gives a schematic representation of the molecular events that lead to the initial interaction of Sfilp₂₁ with Hscentrin 2 (E32A) and the selective aggregation of the target peptide, based on the features of the asynchronous plot. Initially the glutamate residues (1550.3 cm⁻¹), which could be attributed to either Hscentrin 2 (E32A) or the Sfilp₂₁ target peptide, were perturbed, followed by the aspartates (1577.0 cm⁻¹), which are exclusive to centrin and are located within the calcium binding sites. Next, the short β -sheets (1632.0 cm⁻¹) that are located close to the calcium binding sites were perturbed, presumably due to the initial interaction with the target peptide. At this point, the optimum orientation was not achieved between the Hscentrin 2 (E32A) and Sfilp₂₁, leading to the selective aggregation (1613 cm⁻¹) of the target peptide via its contiguous histidine residues (1597.2 cm⁻¹), located both in the middle of the sequence (H₁₃-H₁₅, $\epsilon = 210$ M⁻¹ cm⁻¹) and near the C-terminal end (H₂₉-H₃₀, $\epsilon = 140$ M⁻¹ cm⁻¹). These histidine vibrational modes are exclusive to Sfilp₂₁, and the histidine pairs are presumably involved in the aggregation process as two separate but simultaneous molecular events during the thermal stress. We theorize that the Hscentrin 2 (E32A) variant may predispose the initial interaction with Sfilp₂₁ to the EF-hand motif located in the N-terminal end instead of the EF-hand motif located within the C-terminal end, preventing the optimum orientation that would normally lead to Sfilp₂₁ adopting a helical conformation. The helical region of Hscentrin (E32A) (1652.5 cm⁻¹) was then perturbed, followed by opening of the EF-hand motif, which perturbs the hinge loop (1662.5 cm⁻¹) and associated calcium binding site loops (1672.0 and 1682.6 cm⁻¹, which represent the apo- and holo- forms of the calcium binding sites, respectively). The opening of the EF-hand exposes the phenylalanine residues (1715 and 1702 cm⁻¹, p-aromatic overtones) to their aqueous environment, leading to an unstable centrin variant-target peptide interaction which resulted in the selective aggregation of Sfilp₂₁.

Conclusion

The QCL IRM platform, combined with 2D IR correlation spectroscopy, has proven useful for obtaining a molecular description of the aggregation distinct from the desired self-association of a protein-peptide complex during crystal screening. A successful crystal screen would include the self-association of proteins or the association observed in protein complexes, which would lead to nucleation and crystal growth. 2D IR correlation spectroscopy would be capable of distinguishing between the two types of processes, while providing an unprecedented level of molecular detail. This approach may provide valuable information leading to increased success rates for the crystallization of protein complexes by identifying the variables that lead to the optimum chemical and physical properties associated with well-diffracting crystals.

Accepted in Spectroscopy

Author contributions

Crystallization screens were performed and HSIs acquired by Sherly Nieves. 2D IR spectral data analysis was performed by Belinda Pastrana. Figures were generated by Sherly Nieves.

Acknowledgments

The authors would like acknowledge Dr. Melissa Stauffer (Scientific Editing Solutions, Walworth, WI) for editing the manuscript.

Funding

The work presented herein was made possible by support from the National Science Foundation (SBIR PII Award 1632420 (BP)).

Disclosures

Belinda Pastrana is the CEO of Protein Dynamic Solutions, Inc. and a Professor in the Department of Chemistry at the University of Puerto Rico-Mayagüez, Mayagüez, PR. Sherly Nieves is a scientist at Protein Dynamic Solutions, Inc.

References

- [1] I.D. Hoffman, *Methods Mol. Biol.* **841**, 67-91 (2012).
- [2] I. Russo Krauss, A. Merlino, A. Vergara, and F. Sica, *Int. J. Mol. Sci.* **14**, 11643-11691 (2013).
- [3] C. Sauter, B. Lorber, A. McPherson, and R. Giege, In *International Tables for Crystallography. Volume F: Crystallography of Biological Macromolecules*, 2nd ed. E. Arnold, D.M. Himmel, and M.G. Rossman, Eds. (Wiley: Chichester, NH, USA, 2012), pp. 99-121.
- [4] Y. Lin, *Expert opinion on Drug Discovery* **13**, 691-695 (2018).
- [5] R.A. Judge, S. Takahashi, K.L. Longenecker, E.H. Fry, C. Abad-Aapatero, M.L. Chiu Cryst. Growth Des. **9**, 3463-3469 (2009).
- [6] I. Noda, *Vibrat. Spectrosc.* **36**, 143-165 (2004).
- [7] I. Noda, *J. Mol. Struct.* **1069**, 23-49 (2014).
- [8] S. Li, A.M. Sandercock, P. Conduit, C.V. Robinson, R.L. Williams, and J.V. Kilmartin, *J. Cell. Biol.* **173**, 867-877 (2006).
- [9] B. Pastrana-Rios, J. De Orbata, V. Meza, M. Reyes, D. Narváez, A.M. Gómez, A. Rodríguez Nassif, R. Almodovar, A. Díaz Casas, J. Robles, A.M. Ortiz, L. Irizarry, M. Campbell, and M. Colón, *Biochemistry*. **52**, 1236-1248 (2013).
- [10] S.Y. Venyaminov and N.N. Kalnin, *Biopolymers*. **30**, 1243-1257 (1990).
- [11] J. Bandekar. *Biochim Biophys Acta*. **1120**, 123-143 (1992).
- [12] A. Barth. *Biochim. Biophys. Acta*. **1767**, 1073-1101 (2007).

Figure Captions

Figure 1: Ribbon model representation for the yeast centrin homologue known as CDC31 CDC31-Sfi1p complex based on high resolution X-ray structure determination PDB ID: 2DOQ (7). The centrin (light gray) wraps around the Sfi1p peptide (dark gray central helix), causing the peptide to adopt its helical conformation. Centrin, a calcium-binding protein belonging to the EF-hand superfamily, is also complexed with calcium, shown as orange spheres.

Figure 2: HSIs of *Hscentrin 2* (E32A)-Sfi1p crystallization screening sample. HSIs as a function of temperature (30-38 °C), where the amorphous aggregate is observed to grow. These Hyperspectral Images are generated from 223,000 QCL IRM transmission spectra.

Figure 3: QCL IRM spectra, and 2D IR correlation spectra of *Hscentrin 2* (E32A)-Sfi1p crystallization screening sample. (a) The corresponding QCL IRM spectral overlay with enhanced S/N ratio within the spectral region of 1750 -1485 cm^{-1} . Spectral features observed are commonly associated with protein aggregation. (b-c) 2D IR correlation plots using the QCL IRM spectral data. (b) The synchronous plot includes an auto peak due to protein aggregation as the main event during the thermal stress, which correlates with the HSIs shown in the top row. The color bar defines the synchronous intensity changes within the plot. (c) The asynchronous plot provides the enhanced resolution needed for the molecular understanding of the mechanism of aggregation for this crystallization condition. Also, shown is the color bar that defines the intensity changes within the asynchronous plot.

Figure 4: Schematic representation of the sequential order of molecular events that led to the selective aggregation of the Sfi1p₂₁ peptide. (a) Sequential order of molecular events. The 2D IR correlation analysis resulted in the elucidation of molecular events that were exclusive to either the centrin variant or the Sfi1p₂₁ peptide, which provided direct evidence of the aggregation process. (b) Helical wheel representation of Sfi1p₂₁ typically adopted when in complex with centrin [8]. (c) Amino acid sequence of Sfi1p₂₁ target peptide; the underlined residues are the hydrophobic triad (L₁₉L₂₃W₂₆) required for binding.

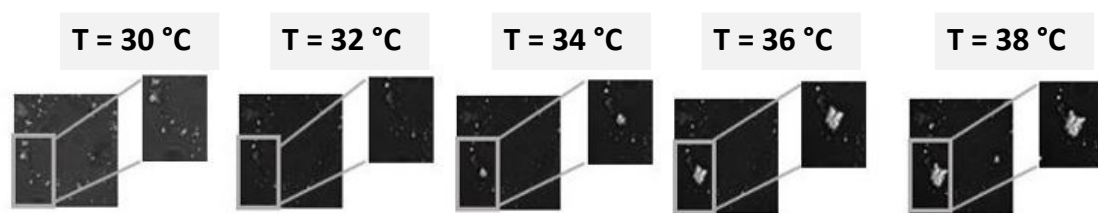
Table I. Summary of QCL IR spectral band assignments

Protein	Wavenumber (cm ⁻¹)	Band Assignment
Protein Backbone		
<i>Hs</i> centrin 2(E32A)	1672.0	Apo-CaBS loop
<i>Hs</i> centrin 2(E32A)	1682.6	Holo-CaBS loop
<i>Hs</i> centrin 2(E32A)	1662.5	hinge loop
<i>Hs</i> centrin 2(E32A)	1652.5	helical
<i>Hs</i> centrin 2(E32A)	1632.0	β-sheet
<i>Hs</i> Sfi1p ₂₁	1613.0	aggregation
Side Chains		
<i>Hs</i> centrin 2(E32A)	1717.5	Phe (overtones (p-aromatic))
<i>Hs</i> centrin 2(E32A)	1702.5	Phe (overtones (p-aromatic))
<i>Hs</i> centrin 2(E32A)	1577.0	Asp ⁻ (ν _{as} (COO ⁻))
<i>Hs</i> centrin 2(E32A) and <i>Hs</i> Sfi1p ₂₁	1550.3	Glu ⁻ (ν _{as} (COO ⁻))
<i>Hs</i> Sfi1 p ₂₁	1597.2	His (C=C)

Figure 1



Figure 2



Accepted in Spectroscopy

Figure 3

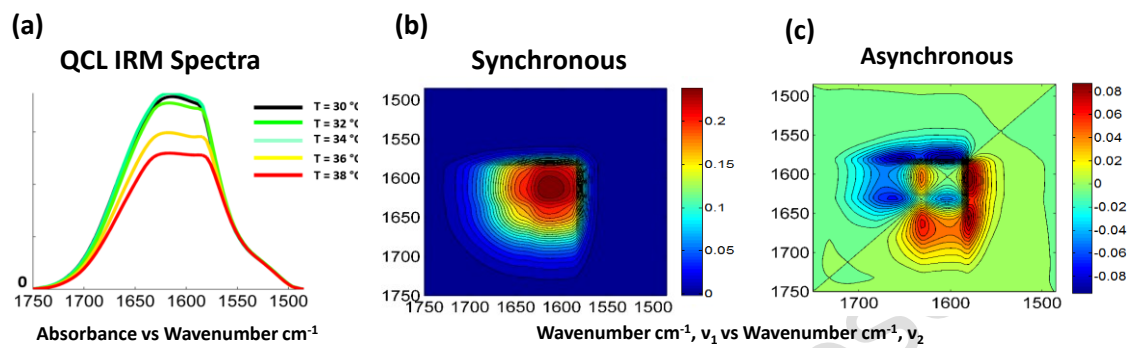


Figure 4

