

Running Title: *Developability of IgG C_H2 domain*

Developability Assessment of an Isolated C_H2 Immunoglobulin Domain

Belinda Pastrana^{1*}, Sherly Nieves¹, Wei Li ^{2,3}, Xianglei Liu³, and Dimitar S. Dimitrov^{2,3}

¹ Protein Dynamic Solutions, 9 Audubon Road, Wakefield, MA 01880-1256

² National Cancer Institute, Frederick, MD 21702-1201

³ University of Pittsburgh, Department of Medicine, Center for Antibody Therapeutics, 3550 Terrace Street, Pittsburgh, PA 15261

This study was funded by NSF SBIR Phase II Award 1632420 (BP) and UPMC (DSD).

*Corresponding author:

Belinda Pastrana, Ph.D.

Protein Dynamic Solutions, Inc.

9 Audubon Road

Wakefield, MA 01880

Phone: (978) 594-6955

E-mail: belinda@pdsbio.com

Abstract

The IgG C_H2 domain continues to hold promise for the development of new therapeutic entities due to its bi-functional role as a biomarker and effector protein. The need for further understanding of molecular stability and aggregation in therapeutic proteins has led to the development of a breakthrough quantum cascade laser microscope to allow for real-time comparability assessment of an array of related proteins in solution upon thermal perturbation. Our objective was to perform a comprehensive developability assessment of three similar monoclonal antibody fragments: C_H2, C_H2s, and m01s. The C_H2 construct consists of residues Pro238 to Lys340 of the IgG1 heavy chain sequence. C_H2s has a 7-residue deletion at the N-terminus and a 16-residue C-terminal extension containing a histidine tag. The m01s construct is identical to C_H2s except for two cysteines introduced at positions 242 and 334. A series of hyperspectral images was acquired during thermal perturbation from 28 to 60 °C for all three proteins in an array. Co-distribution and 2D IR correlation spectroscopies yielded the mechanism of aggregation and stability for these three proteins. The level of detail is unprecedented, identifying the regions within C_H2 and C_H2s that are prone to self-association and establishing the differences in stability. Furthermore, C_H2 helical segments, β -sheets, β -turns, and random coil regions were less stable than in C_H2s and m01s, due to the presence of the N-terminal 3₁₀-helix and β -turn type III. The engineered disulfide bridge in m01s eliminated the self-association process and rendered this mAb fragment the most stable.

Key words: protein array, mAb fragments, developability, stability assessment, quantum cascade laser infrared microscopy, 2D IR correlation and co-distribution correlation spectroscopy, sequential molecular order of events, thermal dependence

Proteins are increasingly being used as pharmacotherapies against cancer, neurological and immunological disorders, infectious diseases, and organ transplant rejection.¹ Because they are large on the molecular scale, therapeutic proteins require special considerations for production, formulation, and administration. The objectives of therapeutic protein formulation are the optimization of stability, the preservation of efficacy, and the minimization of immunogenicity. One challenge is the selection of therapeutic proteins on the basis of their biochemical and biophysical properties, such as structural integrity, charge, stability, aggregation and deamidation propensity, post-translational modifications (PTMs), and binding to their intended target.²⁻⁸ These features contribute to the developability of a protein.

Unstable proteins tend to lose structural integrity, unfold, and precipitate or aggregate in solution. Aggregation is a common occurrence during development of therapeutic proteins due to the different stresses to which they may be subjected. An understanding of the molecular nature of instability is required to re-engineer the protein of interest and decrease its risk for aggregation. Protein aggregation can be assessed by a variety of chromatographic and analytical methods, but none provide more detailed insights than two-dimensional infrared (IR) correlation spectroscopy (2D-COS) and co-distribution correlation spectroscopy (2D-CDS).⁹⁻¹¹ 2D IR correlation detects changes in protein secondary structure and side chain interactions as a function of thermal stress or other perturbations.¹² A typical IR spectrum provides a basis for peak assignments, while the 2D-COS plot reveals which structural events are correlated and in what order they occur. More importantly, it allows for both the spatial and temporal resolution to define the molecular dynamics of the protein, including the disruption of weak interactions that lead to conformational instability.¹³⁻¹⁸ The use of this technique reveals the relative stability within different regions of a protein and provides insight into the structural determinants of aggregation and unfolding, domain stability, and the effect of site-directed mutation on the conformational stability of recombinant proteins and protein-protein interactions.¹²⁻¹⁸ In addition, the 2D co-distribution correlation routine allows for the determination of the common dynamic factors that lead to the behavior of the population of proteins within the sample.

Here we describe the use of these two correlation algorithms (2D-COS and 2D-CDS) to characterize the molecular dynamics of a series of proteins in solution, combined with quantum cascade laser microscopy (QCLM) in the ProteinMentor platform. This platform uses a unique slide array and an innovative microscope with a high-intensity infrared light source. The detector

provides real-time image acquisition, yielding a series of hyperspectral images (HSIs) of the protein solutions during their thermal perturbation, which allows for a side-by-side comparison of the samples in terms of stability and dynamics. To our knowledge, this is the first time this approach has been employed in evaluating the developability of therapeutic proteins.

As a model system, we used an isolated, unglycosylated C_H2 domain of human IgG1, a small monomeric protein construct that has the potential to be utilized as a scaffold for the design of custom antibodies.¹⁹ The crystal structure of human IgG1 C_H2 (residues Gly236 to Gln342, molecular weight (MW) of 14 kDa) showed that it possesses the typical immunoglobulin fold consisting of 7 β -strands organized into 2 antiparallel β -sheets.²⁰ A native disulfide bond at the interface between strands B and F (residues Cys261 and Cys321) stabilizes the domain in solution and also presumably *in vivo*. In previous studies, a shortened construct beginning at Pro238 (named C_H2s) was found to have increased thermal stability, i.e., a higher melting temperature by 5 °C compared to the wild-type C_H2.²¹ Introduction of a second disulfide bond into C_H2s between strands A and G, via site-directed mutation of residues 242 and 334 to cysteine, resulted in a variant called m01s and further stabilized the domain, as was evident from an additional 20 °C increase in the melting temperature.^{22,23} The mechanism underlying the distinct thermal stabilities of these three variants is intriguing.

In this study, we characterized the structural determinants of stability of C_H2, C_H2s, and the m01s variant using QCLM. QCLM analyses allowed for the monitoring of an array of these proteins in solution under thermal stress. 2D-COS and 2D-CDS analysis revealed the molecular basis for destabilization with increasing temperature. The results yield insight into the differential stability of the three protein constructs. Our study also demonstrates the ability of QCLM to identify protein constructs with the best thermal stabilities and good potential for further drug development.

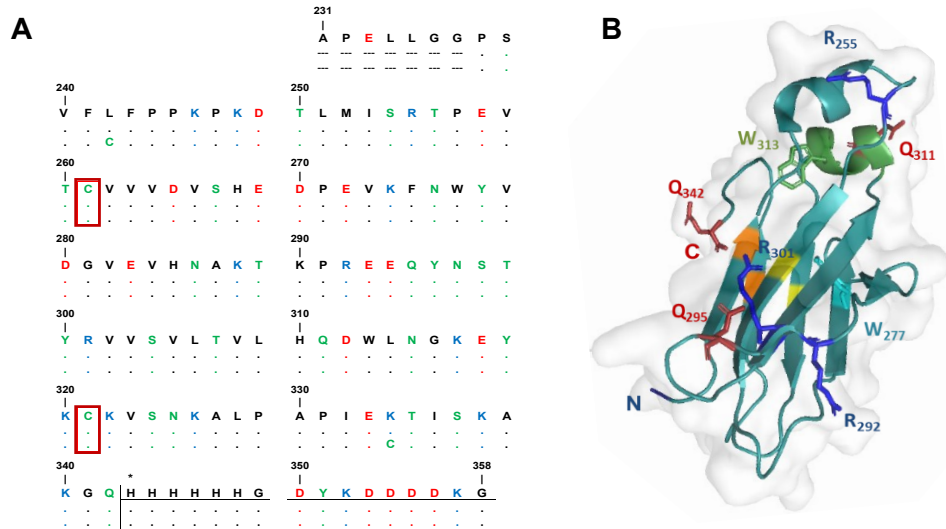
Methods

Protein constructs

The C_H2 construct consists of residues Pro238 to Lys340 of the IgG1 heavy chain sequence shown in Figure 1 (using the numbering from Prabakaran et al. (2008)²⁰). C_H2s has a 7-residue deletion at the N-terminus and a 16-residue C-terminal extension containing a histidine (His) tag (Figure 1). The m01s construct is identical to C_H2s except for the two cysteines introduced at positions 242 and 334 (Figure 1). The design, expression, purification, and characterization of these constructs

have been published.^{21,22} We performed SDS-PAGE under non-reducing conditions and circular dichroism for these three monoclonal antibody (mAb) fragments. The results are shown in the Supporting Information section as Figures S1 and S2.

Figure 1. Sequence and structural features for C_H2, C_H2s and m01s constructs



(A) The sequences of the C_H2 (top), C_H2s (middle), and m01s (bottom) constructs are aligned. The sequence is identical to the crystal structure through Lys340, after which the sequence of the C-terminal His tag is underlined. Colors correspond to the polarity of each residue: red, negatively charged; blue, positively charged; and green, polar. Cysteines involved in the native disulfide bridge (positions 261 and 321) are indicated with red boxes. The bottom sequence shows the two positions (242 and 334) in which cysteine was substituted.

(B) Ribbon diagram of the secondary structural elements of CH2s (PDB ID: 3DJ9).²⁰ The N- and C-termini are indicated, along with the arginines (positions 255, 292 and 301), glutamines (positions 295, 311 and 342), and tryptophans (positions 277 and 313) used as internal probes in the 2D-COS analyses. Yellow coloring indicates the positions of the native disulfide bridge, and orange indicates the disulfide bridge formed after double cysteine substitution in the m01s construct. Green indicates the second helical segment as the α -helix, containing two of the internal probes, Trp313 and Gln311. The ribbon representation with surface rendering was generated using the PyMOL Molecular Graphics System, Version 2.3, Schrödinger, LLC.

QCLM

All protein samples were in the same buffer conditions (phosphate buffered saline: 8 mM Na_2HPO_4 , 137 mM NaCl, 1.94 mM KH_2PO_4 , 2.7 mM KCl, at pH 7.4). Triplicates of each protein sample ($\text{C}_{\text{H}2}$, $\text{C}_{\text{H}2\text{s}}$, and mO1s) were loaded into a pre-defined well array on a single slide cell with optimum fixed path-length from Protein Dynamic Solutions, Inc. (Wakefield, MA). A quality-by-design QCL mid-IR transmission microscope (model ProteinMentor) from Protein Dynamic Solutions, Inc. (Wakefield, MA) equipped with a low magnification (4x) objective with a 0.25 numerical aperture and a $2 \times 2 \text{ mm}^2$ field of view, providing a spatial resolution $4.25 \times 4.25 \mu\text{m}$, with a dedicated interface, stage, heated accessory, and focal plan array (480×480 pixels) detector, was used to assess the stability of the three mAb fragments simultaneously under the same conditions. The dedicated accessory was heated within a temperature range of 28-60 °C with intervals of 4 °C with real-time monitoring of both temperature and humidity. Also, to prevent coherence effects due to QCL fluctuations, a background image (a region on the slide cell to assess instrument response) was collected at each temperature step once thermal equilibrium (4 min) was achieved. The HSIs are composed of 223,000 QCLM spectra that were collected at 4 cm^{-1} spectral resolution within the spectral region of 1780-1485 cm^{-1} .

2D-COS and 2D-CDS

To further understand the molecular dynamics of these proteins in solution, we performed both 2D-COS and 2D-CDS analyses on the QCLM difference spectra within the spectral region of 1780-1485 cm^{-1} . Difference spectra were generated by subtracting the initial spectrum from all subsequent spectra, allowing for the removal of the absorption of the H_2O bending vibration and all protein associated absorptions that did not change due to the thermal perturbation. The mathematical definitions of both algorithms are explained in the Supporting Information. 2D-CDS spectra reflect the distribution of the protein population in solution, and reveal the main structural events that are common to the majority of the proteins in the sample. The spectral correlations observed can be compared directly between protein samples, allowing for determinations of relative stability and aggregation. The 2D-COS analysis provides detailed molecular events that the protein undergoes upon thermal stress, highlighting any aggregation-prone regions of the protein. Such analysis includes two separate plots. The first is the synchronous plot, which includes all of the in-phase peak intensity changes, while the asynchronous plot comprises the out-of-phase peak intensity changes. The asynchronous plot also provides increased resolution of the spectra. The spectral features are specific to certain

residues and occur in a specified order, such that the mechanism of aggregation and stability within the different structural elements of the protein can be ascertained (see Supporting Information for further details).

Water absorption

Protein samples are in an aqueous environment, and the molar extinction coefficient of pure H₂O is high at 55.5 M⁻¹ cm⁻¹. Like any protein-containing sample, there is an effective dilution of the contribution of HOH bending vibration within the spectral region of 1700-1600 cm⁻¹.^{12,24,26} Also important is the optimum path-length used, allowing for the management of samples that exhibit high absorptivity. Finally, the QCLM transmission absorbance spectrum has an enhanced signal-to-noise ratio, allowing for the difference spectral approach to be used, thus overcoming the common deterrents of using IR spectroscopy for aqueous samples.

HSIs and QCLM spectral analysis

The HSIs were analyzed using DataPDS™ software from Protein Dynamic Solutions, Inc. (Wakefield, MA), while the QCLM spectral baseline correction, difference spectral data, overlays, 2D-CDS and 2D-COS plots were generated using Correlation Dynamics™ software, also from Protein Dynamic Solutions, Inc. (Wakefield, MA). The difference spectra thermal dependence plots were generated using Excel in Microsoft Office Professional 2016 (Microsoft; Redmond, WA).

Results

HSIs

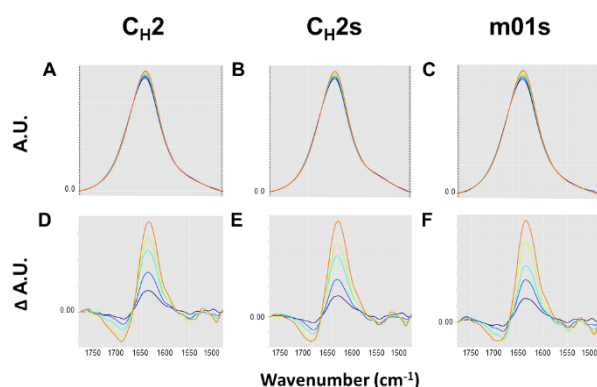
HSIs of C_H2, C_H2s, and m01s at 28 and 60 °C are shown for the entire field of view (Supporting Information Figure S3). The HSIs represented clear and homogeneous solutions with a spatial resolution of 4.25 μm, suggesting no existing aggregation. However, it could not be excluded that aggregates with a size below the spatial resolution may present, which requires further evaluation of the QCLM spectral data.

Quality of the QCLM and difference spectra

The overlay of baseline-corrected QCLM and subsequent difference spectra in the spectral region of 1780-1485 cm⁻¹ for C_H2, C_H2s, and m01s in the temperature range of 28-48 °C are shown in

Figure 2. The QCLM spectral overlays provide evidence of incremental intensity changes as the temperature is increased for the amide I band and overlapping H₂O absorbance within 1700-1600 cm⁻¹ and the amide II band within 1600-1500 cm⁻¹.^{12,26} The enhanced signal-to-noise ratios are provided by using the QCLM (Figure 2A-C). The vertical dashed lines represent the 2-point linear baseline correction performed. Also shown are the difference spectra using the spectral region of 1780-1485 cm⁻¹ generated by subtracting the first spectrum from all subsequent spectra, allowing for the subtraction of the H₂O bending absorption and all spectral contributions that did not change during the thermal perturbation (Figure 2D-F). Both positive and negative peaks are observed within the difference spectra; these occurred due to thermal perturbation.

Figure 2. QCLM spectral and difference overlay for C_H2, C_H2s, and m01s



Baseline-corrected and difference spectral overlays within the spectral region of 1780-1485 cm⁻¹ and temperature range of 28 - 48 °C to address comparability. **(A-C)** baseline corrected QCLM spectra. The vertical dashed lines represent the exact point selected for the linear baseline correction. The color temperature code for each spectral line used was: black, 28 °C; blue, 32 °C; green, 36 °C; yellow, 40 °C; orange, 44 °C; and red, 48 °C. **(D-F)** The difference QCLM spectra were generated by subtracting the first spectrum from all subsequent spectra. For **(A, D)** C_H2, **(B, E)** C_H2s, and **(C, F)** m01s, a high signal-to-noise ratio was observed. The color temperature code for each spectral line used was: dark blue, 32 °C; blue, 36 °C; cyan, 40 °C; yellow, 44 °C; and orange, 48 °C. The resulting line for 28 °C from the subtraction is not shown.

Band assignments

The band assignments for the backbone vibrational and side chain modes are an integral part of the description of the molecular dynamics that occur during thermal perturbation.²⁷⁻³⁰ The band positions used for the comparability assessment represent the average of the 2D IR correlation

peaks for each mAb fragment analyzed. The amide I band within 1700-1600 cm^{-1} is mainly due to C=O stretches, with minor contributions of C-N stretches and to a lesser extent N-H deformation modes; are all sensitive to conformational changes. In general, for all three mAb fragments, the QCLM spectra are composed of: the different types of β -turns²⁷ (1696, 1688, 1682, and 1680 cm^{-1}), π -, α - and 3_{10} -helix (1656, 1653, and 1648 cm^{-1} , respectively), random coil (1643 cm^{-1}), and β -sheets (1636 cm^{-1}). Moreover, the 3_{10} -helix (1648 cm^{-1}) is accompanied by a type III β -turn (1688 cm^{-1}). Also present is the β -turn (1696 cm^{-1}) peak when coupled or correlated with the β -sheet (1636 cm^{-1}) peak, resulting in the structural assignment of antiparallel β -sheets. Protein aggregation was assigned to 1624 cm^{-1} and was observed for C_H2 and C_H2s.

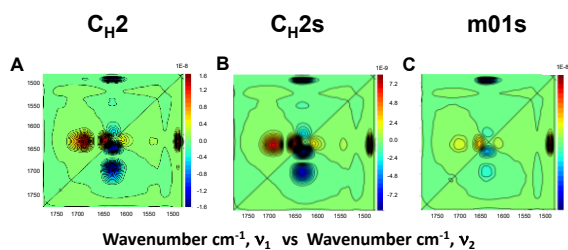
For the side chain modes, there are some vibrational modes that overlap with the amide I band, and others are located within the amide II band (1600-1500 cm^{-1}). Side chain modes located just prior to or overlapping with the amide I band are aspartates or glutamates (1727, 1722, 1714, and 1708 cm^{-1}) that are located within the β -turns (1696, 1682, and 1680 cm^{-1}) and experience hydrogen bonding interactions with their aqueous environment, other side chains, or with the backbone. Side chain modes that fully overlap with the amide I band are attributed to arginine $\nu_{\text{as}}(\text{CN}_3\text{H}_5^+)$ (1673 cm^{-1}), glutamine $\nu(\text{C=O})$ (1670 cm^{-1}), and asparagine bending vibration $\delta(\text{NH}_2)$ (1620 cm^{-1}). Side chain modes located within the amide II band are histidine $\nu(\text{C=C})$ (1610, 1604 and 1600 cm^{-1}), the carboxylate of the C-terminal end $\nu(\text{COO}^-)$ (1596 cm^{-1}), and glutamine $\delta(\text{NH}_2)$ (1588 cm^{-1}). Different aspartates, one of which is presumably involved in a salt-bridge interaction $\nu(\text{COO}^-)$ (1569.0 cm^{-1} and 1567 cm^{-1}); different glutamates, two of which are presumably involved in salt-bridge interactions $\nu(\text{COO}^-)$ (1554, 1548, and 1534 cm^{-1}); and lysine $\delta_s(\text{NH}_3^+)$ (1525 cm^{-1}), tyrosine $\nu(\text{C=C})$ (1511 cm^{-1}), and tryptophan $\nu(\text{C=C})$ (1496 cm^{-1}) also appear in this spectral region. These are the vibrational modes identified based on the peak pattern observed in the 2D-COS asynchronous plot that can be used to probe the molecular dynamics of these mAb fragments in solution during thermal perturbation.^{29,30}

2D-CDS

2D-CDS analysis provides the distribution of the population of the mAb fragments in solution during the thermal perturbation (Figure 3; see Supporting Information for additional details on spectral interpretation). It is a preferred method of analysis to compare stability among similar proteins studied simultaneously in an array. Key cumulative intensity changes for each mAb fragment are shown especially in the cloverleaf like peaks located near the center of the plot,

resulting in the determination of the most stable protein. In this study, m01s was determined to be the most stable mAb fragment (Figure 3C), while both C_H2 and C_H2s have evidence of aggregation (Figure 3A, B). Also observed was the similar extent to which the tryptophan residues (1496 cm⁻¹) were perturbed during the temperature increase. The perturbation of glutamates (1548 cm⁻¹) was also common to C_H2 and C_H2s.

Figure 3. Co-distribution asynchronous plots for C_H2, C_H2s, and m01s



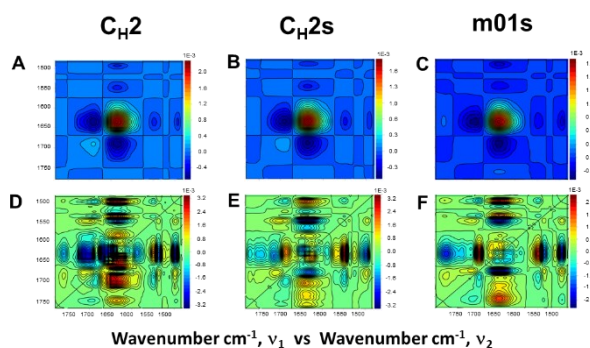
Co-distribution asynchronous correlation plots within the spectral region of 1780-1485 cm⁻¹ and temperature range of 28-48 °C for (A) C_H2, (B) C_H2s, and (C) m01s. These plots represent the distribution of the population of the mAb fragments in solution. Also shown is the color bar that defines the intensity changes within the asynchronous plot. For C_H2 and C_H2s, evidence of aggregation is observed within the set of cloverleaf-shaped peaks near the center of the plot. Furthermore, m01s is the most stable of the three mAb fragments due to the much attenuated intensity changes observed.

2D-COS

A correlation function is applied to the difference spectra, providing enhanced spectral resolution and correlations between cross peaks that allow for their assignment, resulting in a description of the molecular dynamics of proteins. This approach identifies both spatial and temporal molecular events that occur in proteins under thermal stress, yielding an amino acid-level resolution for these mAb fragments. We are focusing our discussion on the temperature range of 28-48 °C to allow for comparability (Figure 4). The results for the full temperature range of 28-60 °C are shown in the Supporting Information (Figure S4; see this file for additional explanations of spectral interpretation). The **synchronous plots** shown in Figure 4A-C show all the changes that occur synchronously with one another. On the diagonal are the *auto peaks*. The lack of the β -turn peak (1696 cm⁻¹) for m01s (Figure 4C) is due to the stability of this secondary structure, while C_H2 and C_H2s have greater flexibility within their β -turns (1696 cm⁻¹). Off-diagonal *cross peaks* are evident for each protein, associating the β -turns (1696 cm⁻¹) with the β -sheets (1636 cm⁻¹), suggesting

that all variants have the antiparallel β -sheet structure. Furthermore, cross peaks associated with side chains such as glutamates (1534 and 1554 cm^{-1}) and tryptophans (1496 cm^{-1}) are observed, demonstrating that these proteins are highly similar to each other. The ***asynchronous plots*** are shown in Figure 4D-F. For C_H2 , the cross peaks associated with conformational stability were observed to have greater intensity changes (Figure 4D) when compared to C_H2s and $m01s$ (Figure 4E, F). The aggregation peak at 1624 cm^{-1} is evident for C_H2 and C_H2s (Figure 4D, E), but not for $m01s$ (Figure 4F). Furthermore, we did not observe glutamine or asparagine deamidation in any of the mAb fragments. The sequential order of molecular events obtained from our analysis of the 2D-COS plots is discussed below.

Figure 4. 2D-COS plots within the temperature range of 28-48 °C for C_H2 , C_H2s and $m01s$



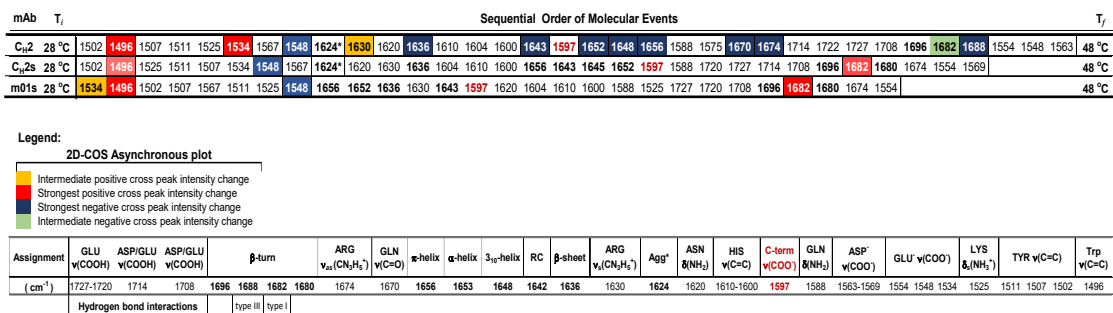
(A-C) Synchronous and (D-F) asynchronous 2D-COS plots within the spectral region of $1780\text{--}1485\text{ cm}^{-1}$ and temperature range of 28-48 °C for (A, D) C_H2 , (B, E) C_H2s , and (C, F) $m01s$. The asynchronous plot provides the enhanced resolution needed for the molecular understanding of the mechanism of aggregation. No glutamine or asparagine deamidation was observed for these mAb fragments. Also shown is the color bar that defines the intensity changes for each contour plot.

Sequential order of molecular events

Key residues can serve as internal probes. For this study they were tryptophans, arginines, and glutamines, which are all located in different secondary structures, allowing us to distinguish them (Supporting Information Figure S5). For example, there are only two tryptophans in the sequence: one is relatively buried and located within a β -sheet motif, while the other is located in one of the helical segments. 2D-COS takes advantage of the cross-talk between the cross peaks (see the Supporting Information for details), allowing for both spatial and temporal correlation to be established.

The 2D-COS analysis resulted in the elucidation of molecular events that led to the understanding of the distinct aggregation process of C_H2 and C_H2s. *Comparability of the molecular dynamics is established and the molecular understanding of the increased stability for m01s is achieved.* For C_H2 (Figures 1B and 5), initially tyrosine (1502, 1507, 1511 cm⁻¹) and tryptophan (1496 cm⁻¹) ring vibrations are perturbed, followed by the disruption of a salt-bridge interaction between a lysine (1525 cm⁻¹) and glutamate (1534 cm⁻¹), then lysine (1525 cm⁻¹) and aspartate (1567 cm⁻¹), and then interactions form presumably with the stretch of poly-aspartates (as part of the His tag) located within the C-terminal end causing the self-association/aggregation (1624 cm⁻¹). Next, the salt-bridge interaction is disrupted between glutamate (1546 cm⁻¹) and arginine (1630 cm⁻¹), both located in β -sheets, followed by the perturbation of the hydrogen bonding interaction between asparagine (1620 cm⁻¹) and glutamate (1546 cm⁻¹), then the β -sheets (1636 cm⁻¹) themselves are perturbed, followed by the poly-histidine stretch (1610, 1604, and 1600 cm⁻¹) located within the random coil (1642 cm⁻¹) C-terminal end (1597 cm⁻¹); presumably all of these are involved in the self-association process. As the temperature increases, the helical segments (1648, 1652, and 1656 cm⁻¹) are perturbed, leading to the loss of 3₁₀-helical secondary structure (1648 cm⁻¹), then the loss of hydrogen bonding between the glutamine (1588 cm⁻¹) and histidine (1600 cm⁻¹) located within the helical segment involved in capping of this helical segment is lost, followed by the loss of hydrogen bonding interactions of glutamates/aspartates (1714, 1722, 1727, and 1708 cm⁻¹) with lysines (1575 cm⁻¹), arginines (1670 cm⁻¹), and asparagines (1674 cm⁻¹) located within the β -turns (1696, 1682, and 1688 cm⁻¹). This loss of hydrogen bonding results in the perturbation of glutamates (1554, 1548 cm⁻¹) and aspartates (1563 cm⁻¹).

Figure 5. Sequential order of molecular events in C_H2, C_H2s, and m01s



Schematic representation of the sequential order of molecular events that lead to aggregation/self-association or unfolding of the mAb fragments during the thermal perturbation of 28-48 °C, focusing on the molecular changes that occurred in the pre-transition period. The background colors shown are directly related to the corresponding asynchronous plot. Also included are the peak assignments as defined in the legend. The 2D-COS analysis resulted in the elucidation of molecular events that led to the understanding of the distinct aggregation process of C_H2 and C_H2s. *Comparability of the molecular dynamics is established and the molecular understanding of the increased stability for m01s is achieved.*

For C_H2s (Figures 1B and 5) with the truncated N-terminal end, tyrosines (1502 cm⁻¹) and tryptophans (1496 cm⁻¹) are perturbed, followed by the disruption of hydrogen bonding between lysine (1525 cm⁻¹) and tyrosine (1511 cm⁻¹), then the disruption of hydrogen bonding between asparagine (1620 cm⁻¹) and glutamates (1534 cm⁻¹), both located within β -sheets. A similar mechanism of self-association/aggregation occurs involving the arginine (1630 cm⁻¹) and the poly-aspartates (1567 cm⁻¹) within the random coil (1642 cm⁻¹), and also perturbed are the poly-histidines (1610, 1604, and 1600 cm⁻¹) near the C-terminal end (1597 cm⁻¹). This causes a perturbation of the helical segments (1656 and 1653 cm⁻¹), followed by the disruption of the hydrogen bonding between glutamine (1588 cm⁻¹) and histidine (1600 cm⁻¹) located within the same helical segment, followed by hydrogen bonding perturbations involving glutamates/aspartates (1720, 1727, 1714, and 1708 cm⁻¹) located within β -turns, followed by the perturbations of the β -turns themselves (1696, 1682 and 1680 cm⁻¹), leading to the disruption of hydrogen bonding interaction between arginine (1674 cm⁻¹) and glutamate (1554 cm⁻¹) and/or aspartate (1569 cm⁻¹).

The m01s variant contains an additional disulfide bridge that is absent in the mAb fragments discussed above, providing greater structural stability and eliminating the propensity for self-association. For m01s (Figures 1B and 5), the initial perturbation involves glutamates (1534 cm⁻¹) followed by tryptophan (1496 cm⁻¹), then tyrosine residues (1502 and 1507 cm⁻¹), followed by the disruption of hydrogen bonding interactions involving three different groups mentioned in the order of their perturbation: (1) aspartate (1567 cm⁻¹) and lysine (1525 cm⁻¹), (2) tyrosine (1511 cm⁻¹) and lysine (1525 cm⁻¹), and (3) lysine (1525 cm⁻¹) and glutamate (1546 cm⁻¹). These hydrogen bonding disruptions are then followed by α -helix (1653 cm⁻¹) and β -sheet (1636 cm⁻¹) perturbations, and by the disruption of hydrogen bonding between lysine (1630 cm⁻¹) and

asparagine (1620 cm^{-1}), then the perturbation of the C-terminal end (1597 cm^{-1}) comprised of a random coil structure (1642 cm^{-1}) along with poly-histidines (1610 and 1604 cm^{-1}). Next, hydrogen bonding interaction of lysines (1525 cm^{-1}) with glutamine (1588 cm^{-1}) is disrupted, followed by disruption of hydrogen bonding involving glutamates/aspartates (1727 , 1720 , and 1708 cm^{-1}) that are located within β -turns (1696 , 1682 and 1680 cm^{-1}), and finally by the disruption of the hydrogen bonding interaction involving arginine (1674 cm^{-1}) and glutamate (1554 cm^{-1}).

In summary, C_H2 helical segments, β -sheets, β -turns (type III and I), and random coil regions were less stable than in C_H2s and $m01s$, due to the presence of a 3_{10} -helix (1648 cm^{-1}) and β -turn type III (1688 cm^{-1}) within the N-terminal end.

Validation of comparability

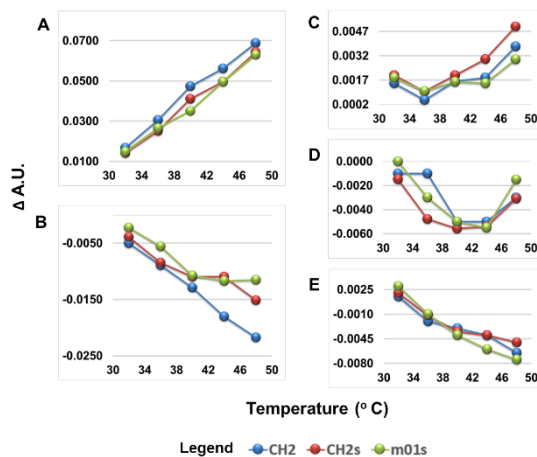
Thermal dependence evaluation using the difference spectral data verified the relative stability of these mAb fragments within the temperature range of $32\text{--}48\text{ }^{\circ}\text{C}$, well within the pre-transition region (Figures 6 and 7). The β -sheets represent the major secondary structure of these proteins, and for all three proteins, a linear incremental change in β -sheet composition is observed with increasing temperature (Figure 6A), presumably within the fourth β -strand due to the neighboring random coil segments that can convert to the β -strand. In $m01s$, the β -turns stabilize above $40\text{ }^{\circ}\text{C}$ compared to both C_H2 and C_H2s (Figure 6B). The evaluation of stability must also include the side chain dynamics involving changes in weak interactions such as salt-bridge and hydrogen bonding interactions. Hence, we included lysine, glutamate, and tryptophan residues, identified by 2D-COS as key in the molecular dynamics, to explore the disruption of these interactions (Figure 6C-E). For lysines, two changes in salt bridge interaction occur at 35 and $44\text{ }^{\circ}\text{C}$, with accompanying changes to the signal (Figure 6C). For glutamates, a curve is observed for all three proteins, suggesting the stabilizing weak interactions involving glutamates are common for all three proteins (Figure 6D). In all three proteins, the two tryptophans are observed to have a progressive destabilization as the temperature increases (Figure 6E).

Similarly, we carried out a thermal dependence evaluation using the peak maxima or minima for the peaks discussed above (Figure 7A-E). The β -sheets within $m01s$ are more stable when compared to C_H2 and C_H2s (Figure 7A). When evaluating the β -turns, the C_H2 mAb is destabilized within the temperature range of $36\text{--}44\text{ }^{\circ}\text{C}$ when compared to C_H2s and $m01s$ (Figure 7B). Our study

also considers the side chain dynamics as a vital component of the stability of these mAb fragments. The lysines within the C_H2 and C_H2s mAbs are more stable compared with m01s (Figure 7C). For glutamates, the loss in weak interactions is suggested by a shift towards higher wavenumbers in all three proteins (Figure 7D). For tryptophans, the minimum observed at 35 °C for all mAb proteins suggests the disruption of hydrogen bonding interactions (Figure 7E).

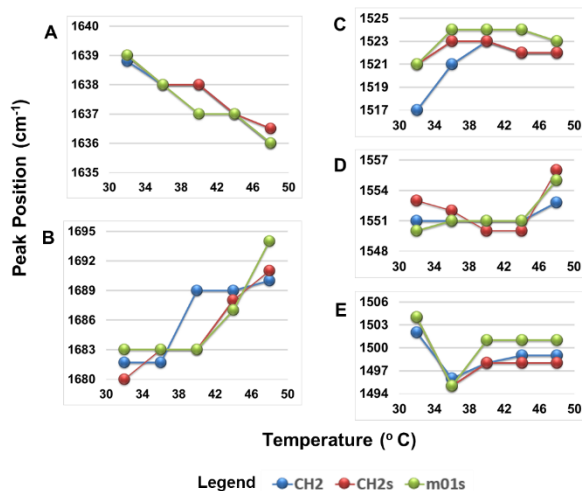
Implications for m01s conformational stability are observed in the combined intensity (Figure 6A and B) and peak maxima and minima position changes (Figure 7A and B). These results lead to m01s having the potential for generation of anti-parallel β -sheets (1639 cm⁻¹ stabilizes at 1637 cm⁻¹ with continued shift to 1636 cm⁻¹) and by its coupling to β -turn (1683 cm⁻¹ shift to 1694 cm⁻¹). This is not observed for C_H2 or C_H2s where the β -sheets (1639 cm⁻¹ stabilizes at 1638 cm⁻¹ and continues to shifts to 1636 cm⁻¹) and couples to different types of β -turns (1680 cm⁻¹ increasing to ~1690 cm⁻¹), suggesting a less stable conformation. Moreover, the C_H2 β -turn intensity change is linear (Figure 6B), suggesting continued instability of this secondary structure when compared to C_H2s and m01s which plateau at 40 °C. Yet, for C_H2s β -turn intensity change continues to decrease above 44 °C following the trend observed for C_H2.

Figure 6. Thermal dependence study monitoring the intensity changes as a function of temperature for C_H2, C_H2s, and m01s.



QCLM difference spectral intensity changes for the temperature range of 32-48 °C for **(A)** β -sheets, **(B)** β -turns, **(C)** lysine, **(D)** glutamate, and **(E)** tryptophan residues in C_H2, C_H2s, and m01s. These data confirm the greater stability observed for m01s when comparing conformational changes of the main structural components of these mAb fragments and disruption of dominant side chain interactions.

Figure 7. Thermal dependence study monitoring the peak position as a function of temperature for C_H2, C_H2s, and m01s.



QCLM difference spectra peak position changes within the temperature range of 32-48 °C for **(A)** β-sheets, **(B)** β-turns, **(C)** lysine, **(D)** glutamate, and **(E)** tryptophan residues in C_H2, C_H2s, and m01s.

Discussion

The architecture and design of therapeutic proteins to date continues to depend on the biochemical and biophysical determinants governing antibody domain stability that are not yet understood in sufficient detail. Understanding stability at the molecular level is crucial to the potential development of therapeutic and diagnostic antibodies and can effectively reduce the risk of immunogenicity.

The C_H2 domain of IgG comprises three short helical segments, several types of β-turns, and three- and four-strand β-sheets pinned by an intra-chain disulfide bond. Herein, we have compared the C_H2 wild-type domain with the truncated N-terminal C_H2s and the double disulfide bonded m01s. We chose to perform a comparability assessment using both 2D-CDS and 2D-COS within the temperature range of 28-48 °C, taking into account the low thermal transition temperature of the wild-type C_H2 (~57 °C).²¹ This temperature range allowed for the investigation of the early elements involved in the destabilization of wild-type C_H2, C_H2s, and m01s prior to their unfolding. Our approach took advantage of certain side chains that served as internal probes for the evaluation of the disruption of weak interactions and their effect on the conformational stability of these mAb fragments.

For C_H2, the helical segments, β -sheets, β -turns (type III and I), and random coil regions were less stable than in C_H2s and m01s, due to the presence of a 3_{10} -helix (1648 cm⁻¹) and β -turn type III (1688 cm⁻¹) within the N-terminal end. Specifically, the N-terminal end disrupts the capping interaction observed for the helices by disrupting hydrogen bonding interactions observed for lysine (Lys248, 1575 cm⁻¹), glutamine (Gln311, 1588 cm⁻¹), and arginine (Arg255, 1674 cm⁻¹) located within the π - and α -helices. These findings are consistent with the NMR ¹³C _{α} chemical shift differences determined for C_H2 and C_H2s as reported by Gong and Weber 2018.²⁵ The perturbation of these weak interactions contributes to the instability of the helical segments (1652 and 1656 cm⁻¹), potentially leading to the lower thermal transition temperature reported by Gong *et al.* 2011 and 2013.^{21,22} However, the tryptophans were equally destabilized in all three mAb fragments, suggesting no significant unfolding event had occurred within the temperature range of 28-48 °C studied. This is in accordance with the fluorescence studies performed by Gong *et al.* 2013,²¹ where the fluorescence intensity ratios were 350/330 within the pre-transition range and stable within the temperature range of 30-50 °C, for both C_H2 and C_H2s. We also observed aggregation in the form of self-association for C_H2 and C_H2s. These results are also validated by the dynamic light scattering results in which two peaks corresponding to higher hydrodynamic radius were observed, suggesting C_H2 had greater propensity for aggregation.²¹

The mechanism of self-association will be discussed in the context of our experimental findings through 2D-COS analysis (i.e., the sequential order of molecular events) and its relation to the available structural information (PDB ID: 3DJ9) reported by Prakabaran 2008²⁰, allowing for a detailed molecular dynamic description that identifies the region prone to aggregation within the C_H2 domain. For C_H2 and C_H2s, the region prone to aggregation was the *C-terminal end involving the arginine-containing β sheets and nearby asparagine*. Specifically, the mechanism of self-association (1624 cm⁻¹) involved the disruption of several key weak interactions between the poly-aspartates (1567 cm⁻¹) located within the His tag and glutamine (Gln342, 1588 cm⁻¹) at the C-terminal end (1598 cm⁻¹) comprised of a random coil (1642 cm⁻¹). This caused the disruption of a salt-bridge interaction involving the arginine (Arg301, 1630 cm⁻¹) located in the fifth β -strand of the β -sheet (1636 cm⁻¹) motif, presumably with a glutamate (Glu293, 1534 cm⁻¹). The C-terminal end also experienced disruption of several local hydrogen bonding interactions within the β sheets (1636 cm⁻¹) involving asparagine (Asn297, 1620 cm⁻¹), arginine (Arg292, 1630 cm⁻¹), and lysines (Lys290 and Lys334, 1525 cm⁻¹) with the β -sheet (1636 cm⁻¹) backbone, as well as

glutamate side chains (Glu294 and Glu333, 1548 cm⁻¹) and their aqueous environment. This mechanism of self-association was observed for both C_H2 and C_H2s.

The most stable mAb fragment, m01s, did not undergo aggregation, due to a second disulfide bond which stabilized the structural motif even further, while preventing self-association from occurring by eliminating the flexibility of the C-terminal end. However, for m01s, the glutamate side chains are significantly destabilized due to the loss of the interactions that were available to them in the absence of the second disulfide bond. In general, the majority of the weak interactions involved hydrogen bonding, but there were two distinct disruption events which involved salt-bridge interactions within the temperature range of 28-48 °C.

Common to all three mAb fragments was the stability of selected β -turns (1695, 1682 cm⁻¹) which were stabilized by hydrogen bonding interactions, consistent with the C_H2 and m01s ¹H-¹⁵N NOE NMR studies reported by Gong *et al.* 2009.²³ Finally, in our analysis of the 2D-COS asynchronous plots, no glutamine or asparagine deamidation was observed for C_H2, C_H2s, or m01s mAb fragments within the temperature range studied (28-60 °C).

To our knowledge, this is the first time that QCLM has been combined with 2D-CDS and 2D-COS approaches for the developability assessment of a series of related proteins. QCLM with enhanced signal-to-noise ratio allows the real-time acquisition of HSIs for an array of mAb fragments in solution under controlled thermal perturbation.¹² This single experiment within the temperature range of 28-60 °C occurred in less than 2.5 hours, requiring 1 μ L of sample per well. The HSIs provided a 4.3 μ m resolution of homogeneous solutions. However, the subsequent analysis of the QCLM difference spectra revealed the aggregation-prone regions and the differential stability of the C_H2 domain-derived mAbs with dual antigen recognition and effector functions. Furthermore, the increased stability of m01s correlated with increased binding of effector proteins.²²

Conclusion

The results reported herein validate previous biophysical findings, while providing exquisite molecular-level detail necessary to define the region and mechanism of aggregation as well as the stability of these mAb fragments in solution. More importantly, this technical approach can be used to inform protein design. The significance of the use of this QCLM platform lies in the predictive results obtained towards developability for an array of proteins, the minimization of the potential for immunogenicity, and its implementation in both the evaluation of stability and

the detection of aggregate species. Thus, this technology has the potential to accelerate the drug development pipeline.

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)) and UPMC (DSD).

Disclosures

The results reported herein are funded by NSF SBIR PII Award 1632420 including Salaries for Belinda Pastrana and Sherly Nieves. Belinda Pastrana has a license without pay from the University of Puerto Rico Mayaguez Campus and is currently full-time at Protein Dynamic Solutions, Inc. Sherly is currently a full-time employee at Protein Dynamic Solutions, Inc. Protein Dynamic Solutions has a patent pending from USPTO. The decision to publish was a joint scientific decision by Dr. Dimitrov and Dr. Belinda Pastrana.

Supporting Information

A single Word document containing 5 supplemental figures and additional details about the 2D-COS and 2D-CDS methods.

References

1. Reichert, J.M. Marketed therapeutic antibodies compendium. *MAbs*. **2012**, *4*, 413-415.
2. Kola, I.; Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **2004**, *3*, 711–15. DOI:10.1038/nrd1470.
3. Yang, X.; Xu, W.; Dukleska, S.; Benchaar, S.; Mengisen, S.; Antochshuk, V.; Cheung, J.; Mann, L.; Babadjanova, Z.; Rowand, J.; Gunawan, R.; McCampbell, A.; Beaumont, M.; Meininger, D.; Richardson, D.; Ambrogelly, A. Developability studies before initiation of process development: Improving manufacturability of monoclonal antibodies. *MAbs*. **2013**, *5*, 787–794. DOI:10.4161/mabs.25269.
4. Jarasch, A.; Koll, H.; Regula, J.T.; Bader, M.; Papadimitriou, A.; Kettenberger, H. Developability assessment during the selection of novel therapeutic antibodies. *J. Pharm. Sci.* **2015**, *104*, 1885–98. DOI:10.1002/jps.24430.
5. Dobson, C.L.; Devine, P.W.; Phillips, J.J.; Higazi, D.R.; Lloyd, C.; Popovic, B.; Arnold, J.; Buchanan, A.; Lewis, A.; Goodman, J.; van der Walle, C.F.; Thornton, P.; Vinall, L.; Lowne, D.; Aagaard, A.; Olsson, L.-L.; Wollberg Riderstad, A.; Welsh, F.; Karamanos, T.K.; Pashley, C.L.; Iadanza, M.G.; Ranson, N.A.; Ashcroft, A.E.; Kippen, A.D.; Vaughan, T.J.; Radford, S.E.; Lowe, D.C. Engineering the surface properties of a human monoclonal antibody prevents self-association and rapid clearance in vivo. *Nat. Sci. Rep.* **2016**, *6*, 38644. DOI:10.1038/srep38644.
6. Fogel, D.B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp. Clin. Trials Commun.* **2018**, *11*, 156–64. DOI:10.1016/j.conctc.2018.08.001.
7. Yingda, X.; Wang, D.; Mason, B.; Rossomando, T.; Li, N.; Liu, D.; Cheung, J.K.; Xu, W.; Raghava, S.; Katiyar, A.; Nowak, C.; Xiang, T.; Dong, D.D.; Sun, J.; Beck, A.; Liu, H. Structure, heterogeneity and developability assessment of therapeutic antibodies. *MAbs*. **2019**, *11*, 39-264. DOI:10.1080/19420862.2018.1553476.
8. Goulet, D.R.; Atkins, W.M. Considerations for the design of antibody-based therapeutics. *J. Pharm. Sci.* **2020**, *109*, 74-103.
9. Noda, I. Techniques of two-dimensional (2D) correlation spectroscopy useful in life science research. *Biomed. Spectrosc. Imaging*. **2015**, *4*, 109-127.

10. Noda, I. Techniques useful in two-dimensional correlation and co-distribution spectroscopy (2DCOS and 2DCDS) analysis. *J. Mol. Struct.* **2016**, *1124*, 29-41. DOI:10.1016/j.molstruc.2016.01.089.
11. Noda, I. Vibrational two-dimensional correlation spectroscopy (2DCOS) study of proteins *Spectrochim. Acta part A: Molecular and Biomolecular Spectroscopy* **2017**, *187*, 119-129. DOI:10.1016/j.saa.2017.06.034.
12. Pastrana-Rios, B.; Rodriguez-Toro, J.J. USPTO Method Patent: Method and System for Spectral Data Analysis Publication number: 16/240,441; Issued Date 7/14/2020.
13. Pastrana-Ríos, B. Understanding the mechanism of peptide unfolding. *Biochemistry* **2001**, *40*, 9074-9081.
14. Sosa, L.D.V.; Alfaro, E.; Santiago, J.; Narváez, D.; Rosado, M.C.; Rodríguez, A.; Gómez, A.M.; Schreiter, E.R.; Pastrana-Ríos, B. The structure, molecular dynamics and energetic of centrin-melittin complex. *Proteins: Struct. Funct. Bioinf.* **2011**, *79*, 3132-3143.
15. Pastrana-Ríos, B.; Reyes, M.; De Orbeta, J.; Meza, V.; Narváez, D.; Gómez, A. M.; Rodríguez Nassif, A.; Almodovar, R.; Díaz Casas, A.; Robles, J.; Ortiz, A. M.; Irizarry, L.; Campbell, M.; Colón, M. Relative stability of centrins and its relationship with calcium binding. *Biochemistry* **2013**, *52*, 1236-1248.
16. Pastrana-Rios, B.; Sosa, L.D.V.; Santiago, J. TFA as excipient destabilizes melittin causing the selective aggregation of melittin within the centrin-melittin-trifluoroacetic acid complex. *Struct. Dyn.* **2015**, *2*, 041711. DOI: 10.1063/1.4921219.
17. Diaz Casas, A.; Chazin, W.J.; Pastrana-Rios, B. Prp40 Homolog A is a novel centrin target. *Biophys. J.* **2017**, *112*, 2529-2539. DOI: 10.1016/j.bpj.2017.03.042.
18. Rodriguez-Nassif, A.; Arada, I.; Arrondo, J.L.; Pastrana-Rios, B. Krr1 KH domain peptide aggregation in the presence and absence of an excipient. *Anal. Chem.* **2017**, *89*, 5765–5775. DOI: 10.1021/acs.analchem.6b04800.
19. Dimitrov, D.S. Engineered CH2 domains (nanoantibodies). *MAbs*. Jan-Feb **2009**, *1*(1), 26-28.,

20. Prabakaran, P.; Vu, B.K.; Gan, J.; Feng, Y.; Dimitrov, D.S.; Ji, X. Structure of an isolated unglycosylated antibody C(H)2 domain. *Acta Crystallogr. D Biol. Crystallogr.* Oct **2008**, *64*(Pt 10), 1062-1067.
21. Gong, R.; Wang, Y.; Ying, T.; et al. N-terminal truncation of an isolated human IgG1 CH2 domain significantly increases its stability and aggregation resistance. *Mol. Pharm.* Jul 01 **2013**, *10*(7), 2642-2652.
22. Gong, R.; Wang, Y.; Feng, Y.; Zhao, Q.; Dimitrov, D.S. Shortened engineered human antibody CH2 domains: increased stability and binding to the human neonatal Fc receptor. *J. Biol. Chem.* **2011**, *286*(31), 27288-27293.
23. Gong, R.; Vu, B.K.; Feng, Y.; Prieto, D.A.; Marzena, A.D.; Walsh, J.D.; Prabakaran, P.; Veenstra, T.D.; Tarasov, S.G.; Ishima, R.; Dimitrov, D.S. Engineered human antibody constant domains with increased stability. *J. Biol. Chem.* **2009**, *284*(21), 14203-14210.
24. Venyaminov, S.Y.; Prendergast, F.G. Water (H₂O and D₂O) molar absorptivity in the 1000-4000 cm⁻¹ Range and Quantitative infrared spectroscopy of aqueous solutions. *Anal. Biochem.* **1997**, *248*, 234-245.
25. Weber, B.; Brandl, M.J.; Pullido Cendales, M.D.; Berner, C.; Pradhan, T.; Feind, G.M.; Zacharias, M.; Reif, B.; Buchner, J. A single residue switch reveals principles of antibody domain integrity. *J. Biol. Chem.* **2018**, *293*, 17107-17118.
26. Nieves, S.; Pastrana, B. Quantum cascade laser infrared microscopy and 2D IR correlation spectroscopy improves crystallization screening of a protein complex. *Spectroscopy* **2019**, *34*, 34-39.
27. Bandekar, J. Amide modes and protein conformation. *Biochim. Biophys. Acta* **1992**, *1120*, 123-143.
28. Arrondo, J.L.; Muga, A.; Castresana, J.; Goñi, F.M. Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **1993**, *59*, 23-56.
29. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta* **2007**, *1767*, 1073-1101.

- 30.** Venyaminov, S.Y.; Kalnin, N.N. Quantitative IR spectrophotometry of peptide compounds in water (H₂O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers* **1990**, *30*, 1243-1257.

For Table of Contents Only

