JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

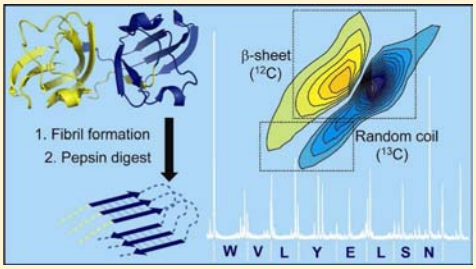
# Structural and Sequence Analysis of the Human γD-Crystallin Amyloid Fibril Core Using 2D IR Spectroscopy, Segmental $^{13}C$ Labeling, and Mass Spectrometry

Sean D. Moran,[†,§] Sean M. Decatur,[§] and Martin T. Zanni*[,†]

[†]Department of Chemistry, University of Wisconsin—Madison, 1101 University Avenue, Madison, Wisconsin 53706, United States
[§]Department of Chemistry and Biochemistry, Oberlin College, 119 Woodland Street, A263, Oberlin, Ohio 44074, United States

Supporting Information

**ABSTRACT:** Identifying the sequence and structural content of residues that compose the core of amyloid fibrils is important because core regions likely control the process of fibril extension and provide potential drug targets. Human γD-crystallin is an eye lens protein that aggregates into amyloid fibrils under acidic conditions. In this manuscript, we use a pepsin enzymatic digest to isolate the core of the amyloid fibrils. The sequence of the core is identified with MALDI MS/MS and its structure is probed with 2D IR spectroscopy and $^{13}C$ isotope labeling. Mass spectrometry of the digest identifies residues 80−163 as the amyloid core, which spans most of the C-terminal domain, the linker, and a small portion of the N-terminal domain. From 2D IR spectroscopy of the digested fibrils, we learn that only the C-terminal domain contributes to the amyloid β-sheets while the N-terminal and linker residues are disordered. A comparison to the native crystal structure reveals that loops and α-helices in the native state must undergo conformational transitions to β-strands upon aggregation. These locations may be good drug binding targets. Besides providing new information about γD -crystallin, this study demonstrates the complementarity of mass spectrometry and 2D IR spectroscopy to obtain both sequence and structure information that neither technique provides individually, which will be especially useful for samples only available in microgram quantities.




## INTRODUCTION

Amyloid fibrils are ordered protein aggregates containing a characteristic cross-β architecture in which intermolecular β-sheets are formed with β-strands running perpendicular to the fibril axis.[1−3] Cross-β structures are very stable and are typically resistant to further chemical or enzymatic modification.[4,5] These aggregates, and intermediates in their formation, have been linked to more than 20 human diseases.[6] High resolution structures of small fibril forming peptides, including the ∼40 residue human peptides hIAPP[7] and Aβ,[8,9] have shown that their conformations in the fibril state are dominated by β-strands.[1,3,10] However, amyloid fibril structures may be complex, and in larger fibril forming proteins, it is often unclear which regions of sequence form the stable cross-β fibril core.[3,10] For example, ∼50% or less of the residues in human γD-crystallin[11] and α-synuclein[12−14] form β-sheets in their fibril states, and the remaining residues exist in other conformations outside the fibril core. Regions of helical or random coil secondary structure may also be interspersed between individual core β-strands.[15,16] Identifying the core residues, and their secondary structures, is an important task because they act as templates for additional protein molecules and may also contain nucleation sites involved in the initial aggregate formation.[17] Both of these properties are relevant to the design of pharmaceuticals that can prevent or reverse aggregation.[6] Biophysical methods such as NMR spectroscopy,[7−10] EPR spin labeling,[18] mass spectrometry,[12−14,19] and infrared spectroscopy[5,11,20] have been applied to identify core structures and sequences in amyloid fibrils and other aggregates and to understand how small molecules and peptides can bind to core residues and inhibit aggregation.[21−24] In this article, we utilize 2D IR spectroscopy in conjunction with isotope labeling, expressed protein ligation, enzymatic digestions, and mass spectrometry to identify the sequence and analyze the structure of the human γD-crystallin (S84C) amyloid fibril core region.

Cataracts, which are caused by the aggregation of lens crystallin proteins into opaque deposits that blur vision, are one of the most common conformational diseases.[25,26] Over 50% of the population over the age of 55 develops age-related cataracts, as a result of damage from UV radiation, heat, oxidants, and a number of other environmental factors.[27] The aggregate structures in cataracts are poorly understood, but various modes of aggregation of the component proteins have been identified *in vitro*, including amorphous precipitation and amyloid fibril formation.[28,29] Recently, we utilized 2D IR spectroscopy and segmental $^{13}C$ labeling to investigate the acid-induced amyloid state of the cataract-related lens protein γD-crystallin.[11] 2D IR spectroscopy is a useful technique for the study of protein aggregates because it is sensitive to secondary

**Received:** August 10, 2012
**Published:** October 19, 2012

   dx.doi.org/10.1021/ja307898g | J. Am. Chem. Soc. 2012, 134, 18410−18416

structure, solvent exposure, structural dynamics, and even higher-order structures through vibrational frequencies, lineshapes, and cross-peaks.[30−36] Using expressed protein ligation, we $^{13}C$ labeled each of γD-crystallin's domains, so that they are resolved in 2D IR spectra from which we were able to independently assess their structural features in the soluble native state as well as the fibril state.[11] We observed amyloid β-sheet secondary structure in the C-terminal domain (CTD), with some disordered structures, and the N-terminal domain (NTD) becomes completely disordered.[11] Thus, we concluded that the C-terminal domain and not the N-terminal domain formed the amyloid core, in contrast to previous reports.

Another way of identifying amyloid cores is by utilizing a nonspecific protease, such as pepsin, to physically isolate the core by digesting unstructured regions. The undigested fragment is then identified using mass spectrometry. Similar approaches have been applied to lysozyme and α-synuclein.[12,13,19] As we report below, when we applied this mass spectrometry to γD-crystallin (S84C), our MALDI MS/MS sequencing analysis reveals that both the N-terminal and C-terminal domains contribute to the core, in apparent contradiction to our previous results. 2D IR spectroscopy of the digested fibrils resolves this apparent contradiction, because the N-terminal and loop residues that contribute to the core are in fact disordered and do not contribute to amyloid β-sheets. Thus, the β-sheet region is smaller than predicted by mass spectrometry. These results provide insights into the γD-crystallin (S84C) fibril structure, sequence information that may be useful for drug studies, and demonstrates that mass spectrometry and 2D IR are complementary techniques, since mass spectrometry reveals sequence information and 2D IR spectroscopy provides secondary structure content. The structural analysis of the enzymatic digest is possible because of the small quantities of material required for 2D IR spectroscopy and the ability to use isotopic labeling for residue specificity.

## MATERIALS AND METHODS

**Materials.** All chemical were obtained from Sigma-Aldrich and used as received, unless otherwise noted.

**Preparation of Human γD-Crystallin Amyloid Fibrils.** Human γD-crystallin (S84C) and its individual domains were expressed in *Escherichia coli* and purified as described previously.[11] Unlabeled and $^{13}C$-labeled domains were linked using expressed protein ligation. Amyloid fibrils of all γD-crystallin samples were prepared by incubating ∼60 μM samples under mildly acidic conditions (pH/pD ∼ 3) at 37 °C for 18 h. Samples for 2D IR analysis were prepared in deuterated buffers (20 mM sodium phosphate, 100 mM NaCl), while those used for MALDI MS/MS analysis were prepared in undeuterated buffers. Transmission electron micrographs of the full-length fibrils were collected at the University of Wisconsin Medical School Electron Microscopy Facility, using a Philips CM 120 microscope. Fibrils were loaded onto copper grids and stained with methylamine tungstate as described previously.[23] Excess stain was blotted off prior to the collection of images.

**Pepsin Digestion of Amyloid Fibrils.** Porcine pepsin from a 1 mg/mL stock was added to samples of γD-crystallin amyloid fibrils in a 1:30 pepsin/γD-crystallin mass ratio. Samples were placed in a shaking incubator at 200 rpm and 37 °C, and aliquots were taken after 30 min, 1 h, 1.5 h, 5 h, 15 h, and 30 h and added to SDS-PAGE sample buffer to denature the fibrils and stop the reaction. Additional aliquots taken at 15 and 30 h were centrifuged at 17 000*g* and resuspended in 20 mM sodium phosphate (pH 3) and 100 mM NaCl a total of three times to remove soluble fragments. The insoluble material obtained in this manner was then redissolvedin SDS-PAGE sample buffer. The protein components of each sample were then separated by SDS-PAGE using a 15% acrylamide/bis-acryalmide resolving gel, and visualized with Coomassie Blue R-250 stain.

**Enzymatic In-Gel Digestion.** In-gel digestion and mass spectrometric analysis was done at the University of Wisconsin—Madison Biotechnology Center Mass Spectrometry Facility. The digestion was performed as outlined on the Web site: http://www.biotech.wisc.edu/facilities/massspec/protocols/ingelprotocol. Coomassie Blue R-250 stained gel pieces were destained twice for 5 min in $CH_3OH/H_2O/NH_4HCO_3$ [50%:50%:100 mM], dehydrated for 5 min in $CH_3CN/H_2O/NH_4HCO_3$ [50%:50%:25 mM] then once more for 1 min in 100% $CH_3CN$. Destained gel pieces were then dried in a Speed-Vac for 1 min, reduced in 25 mM dithiothreitol in 25 mM $NH_4HCO_3$ for 30 min at 56 °C, alkylated with 55 mM iodoacetamide in 25 mM $NH_4HCO_3$ in darkness at room temperature for 30 min, washed once in $H_2O$ for 30 s, and equilibrated in 25 mM $NH_4HCO_3$ for 30 s. Gel pieces were then dehydrated for 5 min in $CH_3CN/H_2O/NH_4HCO_3$ [50%:50%:25 mM], washed for 30 s in 100% $CH_3CN$, dried again, and rehydrated with 20 μL of trypsin solution [10 ng/μL trypsin Gold (PROMEGA) in 25 mM $NH_4HCO_3$/0.01% ProteaseMAX w/v (PROMEGA)]. An additional 30 μL of digestion solution [25 mM $NH_4HCO_3$/0.01% ProteaseMAX w/v (PROMEGA)] was added to facilitate complete rehydration and provide excess overlay needed for peptide extraction. The digestion was conducted for 3 h at 42 °C. Peptides generated from digestion were transferred to a new tube and acidified with 2.5% trifluoroacetic acid to 0.3% final concentration. Degraded ProteaseMAX was removed via centrifugation [max. speed, 10 min] and the peptides were solid phase extracted using ZipTip C18 pipet tips (Millipore, Billerica, MA).

**Mass Spectrometry MALDI TOF/TOF Analysis.** Peptides were eluted off the C18 column with 1 μL of $CH_3CN/H_2O$/TFA (60%:40%:0.1%) into a 0.5 mL Protein LoBind tube (Eppendorf). then, 0.5 μL was deposited onto an Opti-TOF 384-well plate (Applied Biosystems, Foster City, CA), and recrystallized with 0.4 μL of matrix [10 mg/mL α-cyano-4-hydroxycinnamic acid in $CH_3CN/H_2O$/TFA (75%:25%:0.1%)]. Peptide Map Fingerprint result-dependent MS/MS analysis was performed on a 4800 Matrix-Assisted Laser Desorption/Ionization-Time of Flight-Time of Flight (MALDI TOF-TOF) mass spectrometer (AB SCIEX, Foster City, CA). A peptide fingerprint was generated scanning 700−4000 Da mass range using 1000 shots acquired from 20 randomized regions of the sample spot at 3600 intensity of OptiBeam on-axis Nd:YAG laser with 200 Hz firing rate and 3−7 ns pulse width in positive reflectron mode. The 15 most abundant precursors, excluding trypsin autolysis peptides and sodium/potassium adducts, were selected for subsequent tandem MS analysis where 1200 total shots were taken with 4200 laser intensity and 1.5 kV collision induced activation (CID) using air. Postsource decay (PSD) fragments from the precursors of interest were isolated by timed-ion selection and reaccelerated into the reflectron to generate the MS/MS spectrum. Raw data was deconvoluted using GPS Explorer software and submitted for peptide mapping and MS/MS ion search analysis against the *E. coli* protein sequence database (4205 entries) containing the sequence of the recombinant, His-tagged human γD-crystallin (S84C) construct with an in-house licensed Mascot search engine (Matrix Science, London, U.K.). Cys carbamidomethylation, Met/His/Trp oxidation, and Asn/Gln deamidation were included in the ion search as variable modifications. CID spectra and fragment masses for identified peptides are provided in Figures S1−S6 and Tables S1−S6, respectively.

**2D IR Spectroscopy.** 2D IR spectra of pepsin-digested γD-crystallin samples were collected as described previously.[37] Samples were prepared by resuspending the insoluble material remaining after 30 h of pepsin digestion in 20 mM sodium phosphate (pD ∼ 3), 100 mM NaCl. Samples were placed between $CaF_2$ windows with a 56 μM spacer for data collection. 2D IR spectra were normalized to the intensities of their sharp amyloid β-sheet peaks, at $\omega_{pump}$ = 1617 $cm^{-1}$ (unlabeled) or $\omega_{pump}$ = 1575 $cm^{-1}$ ($^{13}C$ labeled). Diagonal slices ($\omega_{pump} = \omega_{probe}$) taken through the resulting 2D IR spectra were used to visualize the attenuation of the disordered region through the diagonal intensity of the amide I fundamental peaks. Total disordered

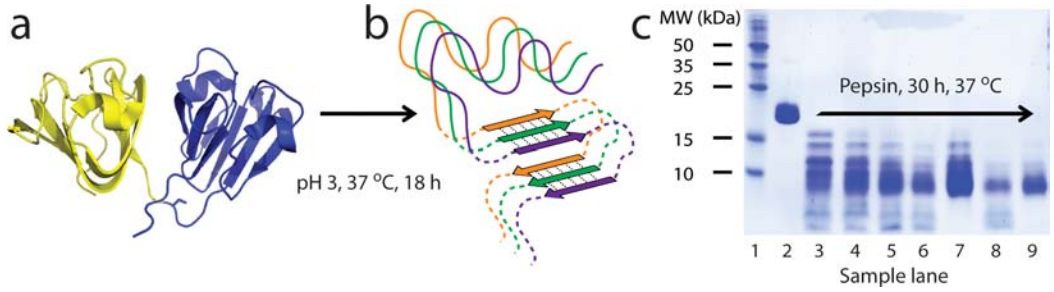



**Figure 1.** Amyloid fibril formation and pepsin digestion of human γD-crystallin (S84C). Native γD-crystallin (a) is converted to the fibril form (b) at 37 °C and pH 3. (c) SDS-PAGE shows that the full-length protein (lane 2) is degraded over 30 h (lanes 3−6, 8) with a major product at ∼8.5 kDa. Insoluble components at 15 h (lane 7) and 30 h (lane 9) contain the stable fibril core.

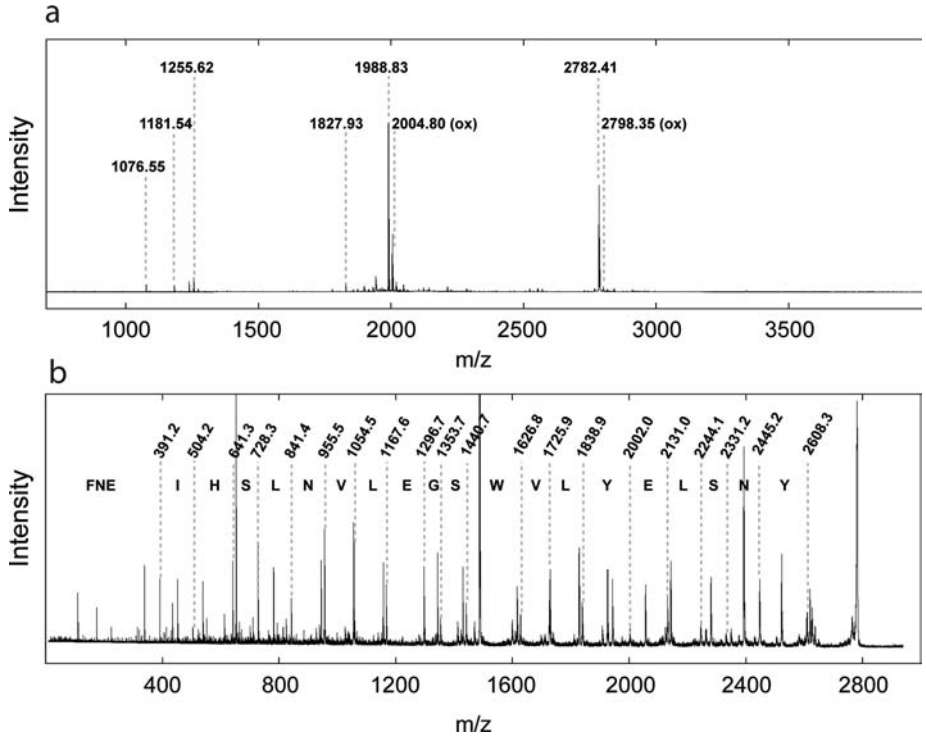



**Figure 2.** (a) MALDI mass spectrum of γD-crystallin (S84C) tryptic digest. Peaks from which MS/MS sequence data were derived are labeled with precursor masses. Peaks resulting from partial oxidation of peptides are labeled (ox). (b) Representative CID spectrum of a γD-crystallin (S84C) tryptic peptide (FNEIHSLNVLEGSWVLYELSNYR + 1, mass = 2782.41), displaying extensive sequence coverage in the ***b*** series. Additional peaks representing ***a*** and ***y*** series ions, as well as internal fragments, are tabulated in the Supporting Information.

peak intensities were compared by volume integration of the 2D IR peaks.

## ■ RESULTS AND DISCUSSION

The pepsin-stable core of γD-crystallin (S84C) amyloid fibrils was isolated as described in Figure 1. Acid-induced aggregation of the native protein (Figure 1a) resulted in the formation of amyloid fibrils with a mixture of β-sheets and disordered regions (Figure 1b). Fibril suspensions prepared in this manner are uniform, and TEM images (Figure S7) show no evidence of nonfibrillar aggregates. Digestion with porcine pepsin at pH 3 and 37 °C resulted in the removal of disordered regions from the fibrils, as indicated by a shift in apparent molecular mass observed using SDS-PAGE (Figure 1c). A strong band near 8.5 kDa persisted even at long times (15−30 h), and was isolated by centrifugation of the reaction mixture. The presence of a large undigested fragment indicates that the fibril core is compact, and does not contain long loops susceptible to cleavage with pepsin. However, the apparent mass alone does not preclude a structure with coil-like elements that are occluded from solvent or protected from digestion by stable interactions with β-sheets.

To identify the residues that make up the ∼8.5 kDa fragment, an in-gel tryptic digest and MALDI MS/MS analysis was performed. The tryptic peptide fingerprint (Figure 2a) is dominated by two very intense peaks at *m*/*z* values of 1988.83 and 2782.41, and their related oxidation products. Numerous other peaks were also observed, and the top 15 most intense peaks were selected for *de novo* sequencing by MS/MS. Among these, five tryptic fragments and one semitryptic fragment were positively identified based on comparison of the CID spectra to the known sequence of the recombinant human γD-crystallin (S84C) protein used in this study. A Mascot search of the MS/MS data of tryptic and semitryptic fragments against the sequence of the recombinant protein and the *E. coli* protein sequence database revealed significant matches to only the


dx.doi.org/10.1021/ja307898g | J. Am. Chem. Soc. 2012, 134, 18410−18416

**Table 1. Tryptic Fragments Positively Identified by MS/MS Sequencing Analysis**

| fragment | expected mass | measured mass | Mascot peptide score | sequence[a] |
|---|---|---|---|---|
| 80−88 | 1075.5418 | 1076.5491 | 38 | LIPHC$^c$GSHR |
| 100−115 | 1987.8268 | 1988.8341 | 108 | GQMIEFTEDC$^c$SC$^c$LQDR |
| 118−140 | 2781.3985 | 2782.4058 | 119 | FNEIHSLNVLEGSWVLYELSNYR |
| 126−140 | 1826.9234 | 1827.9307 | 74 | *VLEGSWVLYELSNYR |
| 143−152 | 1254.6104 | 1255.6177 | 54 | QYLLMPGDYR |
| 154−163 | 1180.5327 | 1181.5400 | 67 | YQDWGATNAR |

[a]C$^c$ = Carbamidomethyl cysteine.*Semitryptic fragment.

recombinant and wild type γD-crystallin sequences. Sequence matches, above the 90% confidence level versus random chance based on Mascot peptide scores, were found for only the six peptides identified by mass in Figure 2a. Since γD-crystallin is the only protein component the sample, this level of confidence is sufficient for positive identification of the peptides. Other peaks did not yield extensive sequence information. A representative CID spectrum, with extensive sequence coverage in the ***b*** series of fragment ions, is shown in Figure 2b. CID spectra for all six identified peptides, and tables of observed fragment masses, are provided in the Supporting Information. Masses, sequence matches, and Mascot peptide scores for all six peptides are summarized in Table 1. The sequences of these peptides contain a number of sites, such as Phe, Tyr, Trp, and Leu, where pepsin is expected to cleave efficiently.[19] Since cleavage is not observed at these residues, we conclude that the structure of the fibril, and not the absence of cleavage sites, dictates the protection of these sequences from digestion. The sequence of the protein analyzed here differs from the wild-type by a single Ser to Cys mutation in the interdomain linker, at residue 84, near the boundary of the pepsin-resistant core. Because of the similar steric and hydrogen-bonding properties of their side chains, the fact that they are both fully protonated at pH 3, and the placement of the substitution in a flexible interdomain linker, this change is not expected to strongly influence the folding properties of γD-crystallin. In our previous work,[11] the S84C mutation was shown to have no measurable effect on the stability of the protein in its native state. However, the current data does not rule out that the cysteine may slightly alter the fibril structure.

The core peptides identified above are mapped to the native structure of wild type human γD-crystallin[38] (PDB ID: 1HK0) in Figure 3. Small gaps (light blue) in arginine-rich regions likely arise from the generation of tryptic precursor peptides with masses outside the 700−4000 Da (+1 charge) range that was analyzed, and not from the removal of these residues by pepsin, based on the presence of a single intense band in the SDS-PAGE gel which indicates a continuous fragment (Figure 1c). Likewise, sequences from ∼71−80 and ∼164−174 cannot be definitively excluded from the core based on the presence of multiple arginine residues in those regions. The total calculated mass of the continuous sequence fragment containing residues 81−163 is 10.1 kDa, consistent with the estimated mass of the pepsin digest product derived from the SDS-PAGE analysis. Thus, according to enzymatic digestion, residues 80−81 of the N-terminal domain, the entire loop, and residues 89−163 of the C-terminal domain contribute to the core of γD-crystallin fibrils. From the sequence overlay in Figure 3, it is apparent that the core residues are derived from sequences that exist in the native monomers as a mixture of β-strands (42%), loops (37%), and α-helices (21%).

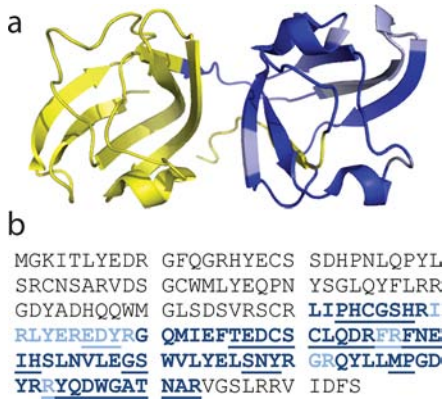



**Figure 3.** Native structural (a) and sequence (b) location of core peptides identified by MS/MS. An N-terminal methionine residue, not present in the wild-type sequence, is present, and numbering of amino acids begins at the first glycine residue. Positively identified peptides (bold, dark blue) lie primarily in the C-terminal domain. Gaps (bold, light blue) occur in arginine-rich regions where precursor tryptic peptide masses fall below the limit of the analyzed range ($m/z$ < 700). Underlined residues in (b) exist in loop or helical conformations in the native state.

To assess the structure of the digested fibrils, we turn to 2D IR spectroscopy. The spectra of undigested fibrils are shown in Figure 4a−c. In the spectrum of the unlabeled fibrils (Figure 4a), two regions of secondary structure are immediately recognizable: a region composed of ordered, extended β-sheets ($\omega_{pump}$ = 1617 cm$^{-1}$) and a region of disordered structures that results in a highly broadened peak pair ($\omega_{pump}$ = 1630−1675 cm$^{-1}$). In the spectra shown in Figure 4b,c, these features appear at different frequencies due to $^{13}$C labeling, even though the structures of the fibrils are unchanged. In Figure 4b, the NTD is $^{13}$C labeled, which shifts the frequencies of amino acids 1−83 down by 39 cm$^{-1}$. Since the sharp β-sheet feature remains at 1617 cm$^{-1}$ with no sharp features near 1578 cm$^{-1}$, only residues from the CTD, and not the NTD, contribute to the β-sheets of the amyloid fibril core. In addition, since there are still features caused by disordered structures at $\omega_{pump}$ = 1630−1675 cm$^{-1}$, some regions of the unlabeled CTD are disordered in the fibrils. In Figure 4c, the CTD is $^{13}$C labeled. This spectrum confirms the conclusions from Figure 4b, and reveals that no region of the NTD has any resolvable secondary structure as indicated by the extremely inhomogeneous features at $\omega_{pump}$ = 1630−1700 cm$^{-1}$. A more complete description of the interpretation of these spectra was presented previously.[11]

In the 2D IR spectra of the isolated cores (Figure 4d−f), there is a loss of intensity in disordered regions compared to the β-sheets. The attenuation of disordered signal is clearly shown in diagonal slices taken through the undigested and digested spectra (Figure 4g−i). However, some disordered

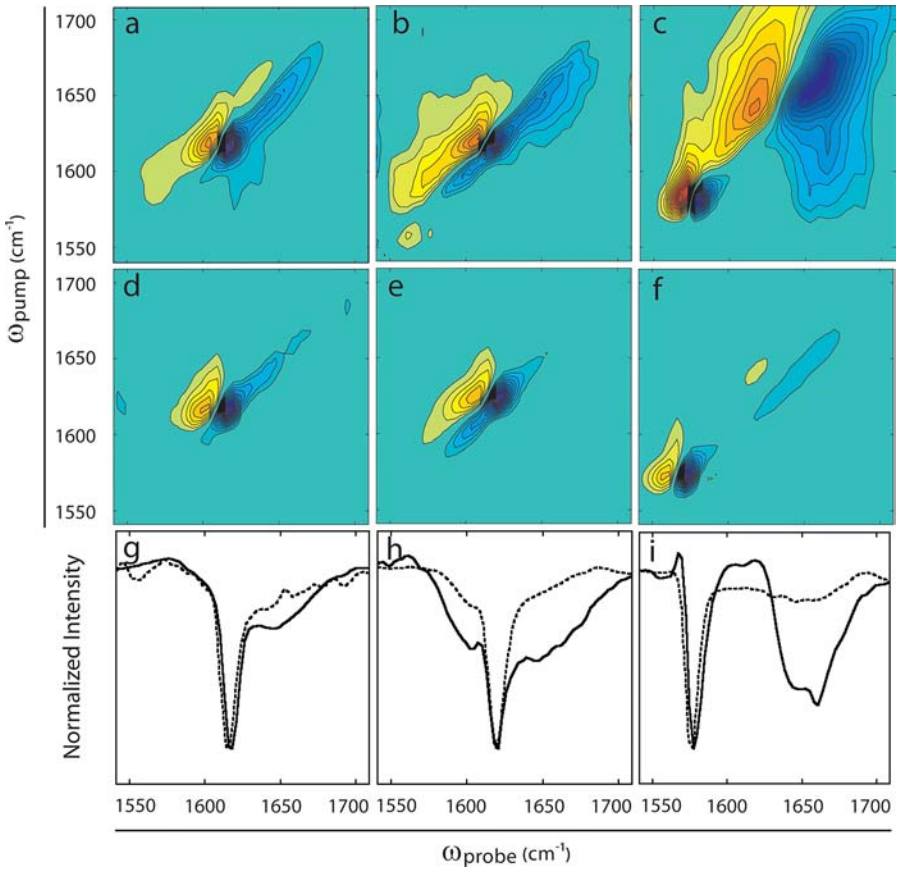



**Figure 4.** 2D IR spectra of γD-crystallin (S84C) amyloid fibrils and isolated pepsin-digested cores. Spectra of full-length unlabeled (a), N-terminally labeled (b), and C-terminally labeled (c) fibrils.[11] After pepsin digestion, signal from disordered structures is strongly attenuated, compared to the narrow signal from amyloid β-sheets (d–f). Diagonal slices (g–i) reveal the attenuation of signal in disordered regions in pepsin-digested samples (dashed lines) compared to undigested samples (solid lines). Disordered signals observed after pepsin digestion correlate to the labeling of the N-terminal domain (e and f). Disordered signals from the C-terminal domain are not resolved.

signal still remains. In the spectrum of pepsin digested, unlabeled γD-crystallin fibrils (Figure 4d), a small residual disordered signal is visible along the diagonal at 1630–1670 $cm^{-1}$, indicating that disordered and/or loop-like regions remain in the fibril core. In the N-terminally labeled sample, the disordered signal is visible at 1580–1610 $cm^{-1}$, whereas in the C-terminally labeled sample, it appears at 1630–1670 $cm^{-1}$. These features indicate that after pepsin digestion, structurally disordered residues from the NTD remain in the insoluble core of γD-crystallin amyloid fibrils. On the basis of the attenuation of disordered signals in the C-terminally labeled samples in Figure 4c versus Figure 4f, about 90% of the NTD is digested by pepsin, in agreement with the core sequence identified via MS/MS. The frequencies of the residual disordered signals in Figure 4e and 3f are largely dependent on the labeling of the NTD, indicating that the majority of residual disordered residues and loops derive from the NTD and not the CTD. However, since disordered regions give rise to broad signals, the 39 $cm^{-1}$ frequency shift due to $^{13}C$ labeling cannot completely resolve disordered regions in the two domains. Thus, some disordered or loop-like regions may exist in the CTD. On the basis of the intensity of the signal in the 1580–1620 $cm^{-1}$ region in Figure 4f, we estimate that less than 10% of the core residues derived from the CTD are disordered. Additionally, since the secondary structure contents of the pepsin-digested fibril cores are similar in both the unligated, full-length unlabeled protein (Figure 4d) and the ligated, segmentally $^{13}C$ labeled proteins (Figure 4e,f) are nearly identical, we conclude that the separate expression and reconstitution of domains has no effect on fibril structure.

A comparison of the our results to the crystal structure of the native protein (Figure 3) reveals which regions undergo a structural change upon fibril formation and highlights promising regions to target with pharmaceuticals aimed at preventing fibril formation. In the native structure, the sequence corresponding to the core exists as a mixture of β-strands (42%), loops (37%), and α-helices (21%).[38] Because the vast majority of core residues form β-sheets in the fibril state, major secondary structure rearrangements must occur during fibril formation. Thus, long loops and helical regions between β-strands in the native Greek key structure must be reorganized into β-strands in the fibril core. Additionally, since few β-strands longer than 10–12 residues have been reported in protein structures,[39] it is unlikely that the residues which do form β-strands in the fibril are continuous. While our results do not quantify the number of β-strands per protein molecule, the stretch of ∼80 residues in the pepsin-stable core suggests that multiple strands are likely. Since our previous 2D IR experiments found that each protein molecule only contributes 1–2 β-strands to each β-sheet,[11] multiple stacked sheets separated by short turns are probable, suggesting that the core forms a compact β-superpleated structure.[3] Finally, the transition of helices and loop-like regions of the native structure to β-strands in the fibril suggests a number of target


dx.doi.org/10.1021/ja307898g | J. Am. Chem. Soc. 2012, 134, 18410–18416

sequences for drug design. Since the transition involves a conformational change from α-helices and loop-like regions to β-strands and not just simple reorganization of existing β-strands, molecules that bind and stabilize helices and loops in the native state are likely to act as aggregation inhibitors.[6,40] For instance, stabilizing the long segment of loop and helical structures spanning residues 106−122 would likely lead to inhibition of aggregation. On the other hand, the N-terminal and interdomain linker residues within the core would probably not be good targets because they do not undergo such a dramatic change in secondary structure.

## CONCLUSION

In conclusion, by combining 2D IR spectroscopy with enzymatic digestion and mass spectrometry, we defined the sequence and secondary structure content of the compact amyloid fibril core of human γD-crystallin. The two sets of data are roughly consistent with each other, validating conclusions drawn from each technique. However, comparison between the two data sets resolves ambiguities in both results, by showing that while the interdomain linker and part of the NTD is found in the pepsin-stable core, the β-sheets in the core lie exclusively in the CTD. The experiments indicate that the core is most likely formed by a β-superpleated structure and point to loop and helical structures in the native state that may be fruitful targets for the design of aggregation inhibitors.[6] Although the data do not define secondary structure at the residue level, this method does provide a guide for more specific labeling schemes that could identify the length and number of β-strands, α-helices, and loops in the fibrils. Such schemes may be implemented by choosing different sites for the ligation of $^{12}C$ and $^{13}C$ protein fragments[11] or individual residue labels.[5,20,37] Short segments or individual residue labels could identify the locations of loops and short helices by taking advantage of the effects of extended vibrational coupling frequency and line shape in 2D IR spectra.[5,20,37,41,42] The techniques described here are powerful because they provide detailed structural information on protein aggregates with minimal sample requirements, because samples on the order of 1 μg of each protein are sufficient for the entire analysis. Thus, even samples that are difficult to prepare can be studied with this combined method under a wide range of experimental conditions. Moreover, similar results would be expected with standard FTIR spectroscopy, just so long as it was coupled with isotope labeling (although 2D IR spectroscopy has the advantage of emphasizing secondary structures over random coils[37,42]). Thus, we expect the combination of infrared spectroscopy, isotope labeling, and mass spectrometry to be broadly applicable to the structural analysis of many ordered protein aggregates involved in human disease, and in the development of drugs that prevent their formation.

## ASSOCIATED CONTENT

### Supporting Information

A TEM image of γD-crystallin amyloid fibrils, and detailed MS/MS data and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

### Corresponding Author

zanni@chem.wisc.edu

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Support for this research was provided by the National Science Foundation (USA) through a Collaborative Research in Chemistry (CRC) grant (0832584). The authors would like to thank G. Sabat of the University of Wisconsin Biotechnology Center Mass Spectrometry Facility for help with data analysis and useful discussions.

## REFERENCES

(1) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773−778.
(2) Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. *J. Mol. Biol.* **2010**, *395*, 717−727.
(3) Eisenberg, D.; Jucker, M. *Cell* **2012**, *148*, 1188−1203.
(4) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2011**, *133*, 14160−14163.
(5) Wang, L.; Middleton, C. T.; Singh, S.; Reddy, A. S.; Woys, A. M.; Strasfeld, D. B.; Marek, P.; Raleigh, D. P.; de Pablo, J. J.; Zanni, M. T.; Skinner, J. L. *J. Am. Chem. Soc.* **2011**, *133*, 16062−16071.
(6) Bartolini, M.; Andrisano, V. *ChemBioChem* **2010**, *11*, 1018−1035.
(7) Luca, S.; Yau, W.-M.; Leapman, R.; Tycko, R. *Biochemistry* **2007**, *46*, 13505−13522.
(8) Petkova, A. T.; Yau, W.-M.; Tycko, R. *Biochemistry* **2005**, *45*, 498−512.
(9) Bertini, I.; Gonnelli, L.; Luchinat, C.; Mao, J.; Nesi, A. *J. Am. Chem. Soc.* **2011**, *133*, 16013−16022.
(10) Tycko, R. *Q. Rev. Biophys.* **2006**, *39*, 1−55.
(11) Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 3329−3334.
(12) Miake, H.; Mizusawa, H.; Iwatsubo, T.; Hasegawa, M. *J. Biol. Chem.* **2002**, *277*, 19213−19219.
(13) Giasson, B. I.; Murray, I. V. J.; Trojanowski, J. Q.; Lee, V. M.-Y. *J. Biol. Chem.* **2001**, *276*, 2380−2386.
(14) Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15477−15482.
(15) Teoh, C. L.; Pham, C. L. L.; Todorova, N.; Hung, A.; Lincoln, C. N.; Lees, E.; Lam, Y. H.; Binger, K. J.; Thomson, N. H.; Radford, S. E.; Smith, T. A.; Müller, S. A.; Engel, A.; Griffin, M. D. W.; Yarovsky, I.; Gooley, P. R.; Howlett, G. J. *J. Mol. Biol.* **2011**, *405*, 1246−1266.
(16) Kurouski, D.; Deckert-Gaudig, T.; Deckert, V.; Lednev, I. K. *J. Am. Chem. Soc.* **2012**, *134*, 13323−13329.
(17) Dupuis, N. F.; Wu, C.; Shea, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2011**, *133*, 7240−7243.
(18) Margittai, M.; Langen, R. *Q. Rev. Biophys.* **2008**, *41*, 265−297.
(19) Frare, E.; Mossuto, M. F.; Polverino de Laureto, P.; Dumoulin, M.; Dobson, C. M.; Fontana, A. *J. Mol. Biol.* **2006**, *361*, 551−561.
(20) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6614−6619.
(21) Lamberto, G. R.; Binolfi, A.; Orcellet, M. L.; Bertoncini, C. W.; Zweckstetter, M.; Griesinger, C.; Fernández, C. O. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21057−21062.
(22) Daval, M.; Bedrood, S.; Gurlo, T.; Chang-Jiang, H.; Costes, S.; Butler, P. C.; Langen, R. *Amyloid* **2010**, *17*, 118−128.
(23) Middleton, C. T.; Marek, P.; Cao, P.; Chiu, C. C.; Singh, S.; Woys, A. M.; de Pablo, J. J.; Raleigh, D. P.; Zanni, M. T. *Nat. Chem.* **2012**, *4*, 355−360.
(24) Zheng, X.; Gessel, M. M.; Wisniewski, M. L.; Viswanathan, K.; Wright, D. L.; Bahr, B. A.; Bowers, M. T. *J. Biol. Chem.* **2012**, *287*, 6084−6088.
(25) Wang, Y.; King, J. A. In *Protein Misfolding Diseases*; John Wiley & Sons, Inc.: Hoboken, NJ, 2010; pp 487−515.
(26) Surgucheva, A.; Surguchov, A. *Brain Res. Bull.* **2010**, *81*, 12−24.

(27) Bloemendal, H.; de Jong, W.; Jaenicke, R.; Lubsen, N. H.; Slingsby, C.; Tardieu, A. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 407−485.
(28) Ecroyd, H.; Carver, J. A. *Cell. Mol. Life Sci.* **2009**, *66*, 62−81.
(29) Papanikolopoulou, K.; Mills-Henry, I.; Tho, S. L.; Wang, Y. T.; Gross, A. A. R.; Kirschner, D. A.; Decatur, S. M.; King, J. *Mol. Vision* **2008**, *14*, 81−89.
(30) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: New York, 2011.
(31) Remorino, A.; Korendovych, I. V.; Wu, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Science* **2011**, *332*, 1206−1209.
(32) Woys, A. M.; Lin, Y. S.; Reddy, A. S.; Xiong, W.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832−2838.
(33) Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 3528−3533.
(34) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17751−17756.
(35) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2010**, *114*, 10913−10924.
(36) Remorino, A.; Hochstrasser, R. M. *Acc. Chem. Res.* **2012**, in press. DOI: 10.1021/ar3000025.
(37) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. *Methods* **2010**, *52*, 12−22.
(38) Basak, A.; Bateman, O.; Slingsby, C.; Pande, A.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. *J. Mol. Biol.* **2003**, *328*, 1137−1147.
(39) Tsutsumi, M.; Otaki, J. M. *J. Chem. Inf. Model.* **2011**, *51*, 1457−1464.
(40) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T. *Nat. Chem.* **2011**, *3*, 172−177.
(41) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. *J. Chem. Phys.* **2005**, *123*, 084905.
(42) Strasfeld, D. B.; Ling, Y. L.; Gupta, R.; Raleigh, D. P.; Zanni, M. T. *J. Phys. Chem. B* **2009**, *113*, 15679−15691.