JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Glutamine side chain 13C=18O as a non-perturbative IR probe of amyloid fibril hydration and assembly

Haifan Wu, Daniel J. Saltzberg, Huong T. Kratochvil, Hyunil Jo, Andrej Sali, and William F. DeGrado

J. Am. Chem. Soc., **Just Accepted Manuscript** • DOI: 10.1021/jacs.9b00577 • Publication Date (Web): 17 Apr 2019

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Glutamine side chain $^{13}C{=}^{18}O$ as a non-perturbative IR probe of amyloid fibril hydration and assembly

Haifan Wu[1], Daniel J. Saltzberg[2], Huong T. Kratochvil[1], Hyunil Jo[1,*], Andrej Sali[2,*], and William F. DeGrado[1,*]

[1]Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California, United States

[2]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco, San Francisco, California, United States

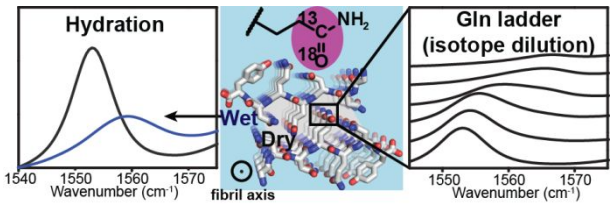
Corresponding authors:

E-mail: hyunil.jo@ucsf.edu, sali@salilab.org, william.degrado@ucsf.edu

Abstract

Infrared (IR) spectroscopy has provided considerable insights into the structures, dynamics, and formation mechanisms of amyloid fibrils. IR probes, such as main chain $^{13}C=^{18}O$, have been widely employed to obtain site-specific structural information, yet only secondary structures and strand-to-strand arrangements can be probed. Very few non-perturbative IR probes are available to report on side chain conformation and environments, which are critical to determining sheet-to-sheet arrangements in steric zippers within amyloids. Polar residues, such as glutamine, contribute significantly to the stability of amyloids and thus are frequently found in core regions of amyloid peptides/proteins. Furthermore, polyglutamine (polyQ) repeats form toxic aggregates in several neurodegenerative diseases. Here we report the synthesis and application of a new non-perturbative IR probe—glutamine side chain $^{13}C=^{18}O$. We use side chain $^{13}C=^{18}O$ labeling and isotope dilution to detect the presence of intermolecularly hydrogen-bonded arrays of glutamine side chains (Gln ladders) in amyloid-forming peptides. Moreover, the linewidth of the $^{13}C=^{18}O$ peak is highly sensitive to its local hydration environment. The IR data from side chain labeling allows us to unambiguously determine the sheet-to-sheet arrangement in a short amyloid-forming peptide, GNNQQNY, providing insight that was otherwise inaccessible through main chain labeling. With several different fibril samples, we further show the versatility of this IR probe in studying the structures and aggregation kinetics of amyloids. Finally, we demonstrate the capability of modeling amyloid structures with IR data using the Integrative Modeling Platform (IMP) and the potential of integrating IR with other biophysical methods for more accurate structural modeling. Together, we believe that side chain $^{13}C=^{18}O$ will complement main chain isotope labeling in the future IR studies of amyloids and integrative modeling using IR data will significantly expand the power of IR spectroscopy for elucidating amyloid assemblies.

TOC



Introduction

Extensive efforts have been dedicated to understanding the structures, dynamics, and aggregation pathways of amyloid fibrils to inform principles underlining amyloid formation, strategies of inhibition, and designs of amyloid-based biomaterials.[1-4] High-resolution techniques, such as X-ray diffraction, micro-electron diffraction (MicroED), solid-state NMR (ssNMR) and cryo-electron microscopy (cryoEM), have provided atomic details of amyloid assemblies,[5-13] yet they give little insights into the pathways of amyloid formation. By contrast, infrared (IR) spectroscopy is one of the few techniques that can probe not only amyloid structures and dynamics but also formation kinetics.[14-18]

A major drawback of IR spectroscopy is the lack of residue-specificity, which can nevertheless be overcome through site-specific IR probes.[14-15] For instance, main chain $^{13}C=^{18}O$,[16, 19-20] which redshifts the amide I band of a particular residue away from the congested main band, has been used to follow conformational changes at specific sites during amyloid aggregation[21-22] and to investigate strand-to-strand arrangements in amyloids[23-24]; however, little information is available to derive the tertiary and quaternary packing of amyloids. On the other hand, unnatural side chain probes, such as thiocyanates, azides, nitriles and esters, have been employed to determine the inter-sheet packing, local electrostatic environment, and hydrogen bonding dynamics within amyloids.[15, 25-29] These unnatural side chain probes are advantageous because of their distinctive IR absorption frequencies. However, the need for introducing unnatural groups[30] has greatly limited their application to amyloids because amyloid conformations are highly sensitive to even subtle changes[31]. Thus, we explored non-perturbative intrinsic side chain IR probes.

High-resolution structures of fibril-like micro-crystals formed by short peptide segments have revealed the steric zipper as the basic unit of amyloid assembly and identified key motifs and interactions that stabilize amyloid structures.[3, 32] One particular type of interactions involves side chains of polar residues, which form hydrogen bonding networks to achieve extraordinary high stability.[5-6, 33] As a critical residue in these polar interactions, glutamine is frequently observed in the core region of amyloid peptides/proteins.[34-37] Moreover, pathogenic mutations to Gln cause more rapid aggregation and are associated with earlier disease onsets.[38-39] Furthermore, polyglutamine (polyQ) peptides form toxic aggregates in several neurodegenerative diseases.[40] Given the important role of glutamine residues in amyloids, the glutamine side chain amide is a promising intrinsic IR probe. However, it has never been explored due to a lack of isotope labeling strategies that can differentiate its absorption band from the main amide band. Thus, there is a great need for a synthetic method for introducing $^{13}C=^{18}O$ in the side chain of Gln regioselectively. Here we developed a synthetic methodology for selective and efficient $^{13}C=^{18}O$ labeling of the glutamine side chain. We found that this probe is highly sensitive to its arrangement and local hydration environment

within amyloid fibrils. We also present the first attempt to use IR data in integrative modeling of amyloid structures using the open source Integrative Modeling Platform (IMP) program[41].

Results and discussion

A robust synthetic method of side chain $^{13}C{=}^{18}O$ labeling. Previously, several strategies have been reported for the main chain $^{13}C{=}^{18}O$ labeling, including an acid catalyzed ^{18}OH exchange[19, 42] and hydrolysis of an activated ester[43] (**Fig. 1a**). Gai and coworkers also reported the ^{13}C labeling of aspartate side chain via an asymmetric alkylation reaction.[44] However, these approaches have potential limitations in selective side chain $^{13}C{=}^{18}O$ labeling of glutamine. For instance, when applied to amides, the acid-catalyzed ^{18}OH exchange approach will cause many undesired side reactions. Hydrolysis of activated ester can only provide acid rather than the required amide side chain. Then it was envisaged that hydration of nitrile by $H_2{}^{18}O$ could provide a solution. To test our strategy, we first performed a substitution reaction with $Na^{13}CN$ to give the nitrile intermediate **2** in 73% yield (**Fig. 1b**). It is also worth mentioning that this approach can provide one of the most affordable routes to the introduction of selective ^{13}C labeling on the side chain of Gln. For the key transformation—hydrolysis of nitrile into amide, we employed the Ghaffar-Parkins catalyst[45] since its functional group selectivity has been demonstrated in total syntheses of a variety of natural products[46-48]. Gratifyingly, hydrolysis with ^{18}O water afforded the $^{13}C{=}^{18}O$ labeled amide **3** in 88% yield, and isotopic enrichment was more than 95% by mass analysis (**Fig. S1**). We then installed trityl (Trt) with Trt-OH catalyzed by concentrated H_2SO_4. The removal of Cbz and methyl ester and fluorenylmethyloxycarbonyl (Fmoc) protection produced the final product **5** in 88% yield over three steps. Our method provided 4 g product ready for solid-phase peptide synthesis, and we foresee no difficulties scaling the procedure further.

To provide further confirmation of the successful installation of side chain $^{13}C{=}^{18}O$, we measured the Fourier-transform infrared (FTIR) spectra of labeled and unlabeled glutamine methyl ester in D_2O. As shown in **Fig. S2**, side chain $^{13}C{=}^{18}O$ displayed a single peak at 1575 cm^{-1} and, compared with unlabeled amide, the peak is redshifted by 64 cm^{-1}. This shift exactly matched the previously reported value for main chain $^{13}C{=}^{18}O$.[20] Thus, for the first time, we are able to isolate the side chain amide vibrational transition from the main chain amide I band.

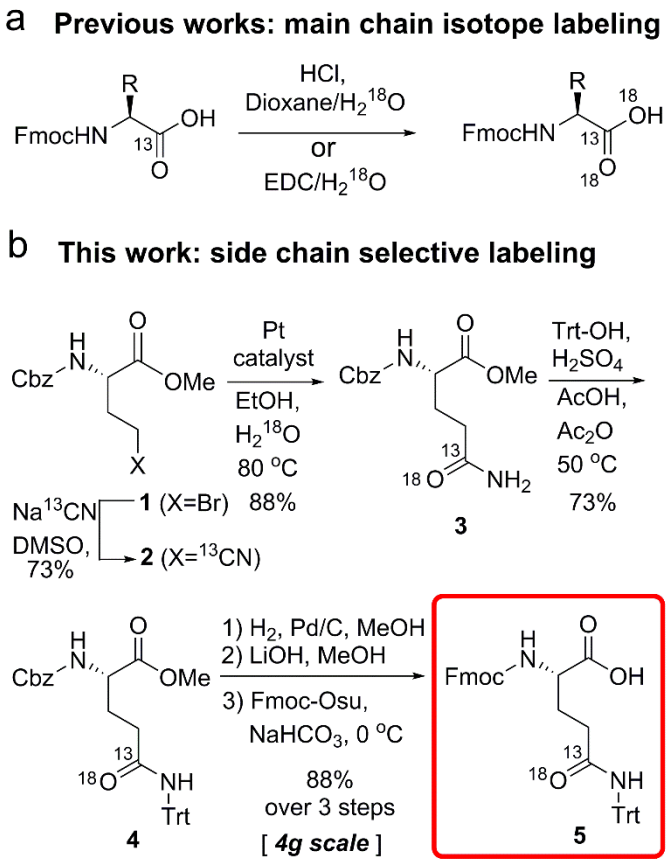


Figure 1. (**a**) Previously reported methods for main chain $^{13}C=^{18}O$ labeling. (**b**) The new synthetic route for preparing side chain $^{13}C=^{18}O$ labeled glutamine.

Probing the Gln ladder by IR and isotope dilution. To test whether side chain $^{13}C=^{18}O$ is sensitive to its local structure and environments, we examined GNNQQNY by FTIR. This yeast prion Sup35-derived peptide forms micro-crystals with a known fibril-like structure, featuring a steric zipper arrangement and arrays of hydrogen-bonded glutamine side chains, known as Gln ladders (**Fig. 2b**). [5, 7] In the crystal structure, Q4 is in the steric zipper interface and forms a Gln ladder. We introduced side chain $^{13}C=^{18}O$ onto Q4 of GNNQQNY (underlined residue is the labeling site). The attenuated total reflection (ATR-FTIR) spectrum of GNNQQNY micro-crystals display a peak at 1625 cm^{-1} and no absorption around 1685 cm^{-1} (**Fig. 2c**), consistent with parallel beta-sheet.[5] The sharp $^{13}C=^{18}O$ peak at 1552 cm^{-1} differs from the broad 1575 cm^{-1} absorption of a fully hydrated $^{13}C=^{18}O$ (**Fig. S2**), indicating a homogenous environment of Q4 in the steric zipper interface. The difference in the absorption frequency can be attributed to the vibrational coupling of aligned $^{13}C=^{18}O$ in the Gln ladder.[49-50]

We then performed an isotope dilution experiment in which we mixed labeled and unlabeled peptides to effectively break the vibrational coupling.[50] As expected (**Fig. 2c**), the $^{13}C=^{18}O$ peaks of micro-

crystals from isotopically diluted samples gradually blueshifted to higher frequencies as we increased the ratios of unlabeled peptides. Additionally, we observed the broadening of the isotope label with increasing ratios of the unlabeled peptide, which recapitulates the results observed in the simulated isotope dilution experiments (**Fig. S3**). We calculated the coupling constant β from $\nu=\nu_0+2\beta$, where ν and ν_0 are the frequencies of coupled and uncoupled transitions, respectively. The obtained coupling constant (-7 cm^{-1}) differs from the reported value (-9 to -11 cm^{-1}) for two hydrogen-bonded carbonyls of adjacent strands in a perfect parallel β-sheet.[16, 51-52] This deviation might reflect the difference in coupling between primary amides (side chain) and secondary amides (main chain). Another possible explanation is the contribution of a positive vibrational coupling between Q4 carbonyls of mating sheets in steric zippers. In other words, two Gln ladders pack against each other. Therefore, isotope dilution is an effective and convenient way to probe Gln ladders, and analyzing the vibrational coupling constant could provide additional insights into the arrangement of ladders.

The peak linewidth is sensitive to the local hydration environment. We next ask whether side chain $^{13}C=^{18}O$ can detect the local hydration status. In GNNQQNY, although both Q4 and Q5 form Gln ladders, they experience completely different hydration environments. Q4 is in the dry steric zipper interface, while Q5 sits in the wet interface (**Fig. 3a**). To investigate the differences, we prepared another peptide GNNQQNY with side chain $^{13}C=^{18}O$ on Q5. The ATR-FTIR spectrum of GNNQQNY micro-crystals showed a much broader $^{13}C=^{18}O$ peak at 1559 cm^{-1} with the peak linewidth (~20 cm^{-1}) almost doubled compared with that of GNNQQNY micro-crystals (~10 cm^{-1}) (**Fig. 3b**). Since isotope dilution confirmed the ladder arrangement of Q5 (**Fig. S4**), the peak broadening is not likely due to conformational heterogeneity. Instead, it should reflect the heterogeneous interaction of Q5 side chains with water molecules. The difference in the absorption frequency might be attributed to the differences in hydrogen bonding strength and/or the local electric field,[53-54] which would require further investigation. Nevertheless, the $^{13}C=^{18}O$ linewidth provides a convenient means to determine whether the probe is in or outside the dry steric zipper interface.

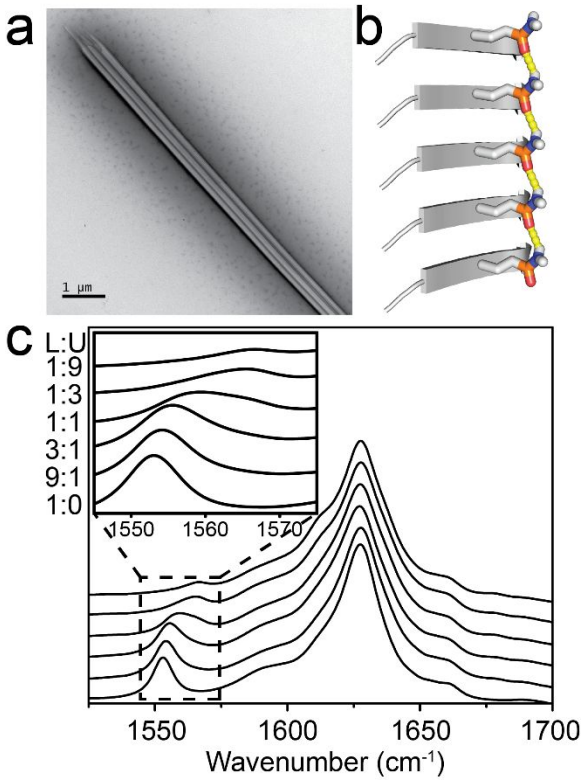


Figure 2. Isotope dilution detected Gln ladders in GNNQQNY micro-crystals. (**a**) The EM image of GNNQQNY micro-crystals. Scale bar: 1 μm. (**b**) The alignment of side chain carbonyls to form a Gln ladder (orange). The graph was created from PDB 5K2G.[7] (**c**) ATR-FTIR spectra of micro-crystal samples with various ratios of GNNQQNY (labeled, L) and GNNQQNY (unlabeled, U). The intensity is normalized according to the 1625 cm^{-1} peak.

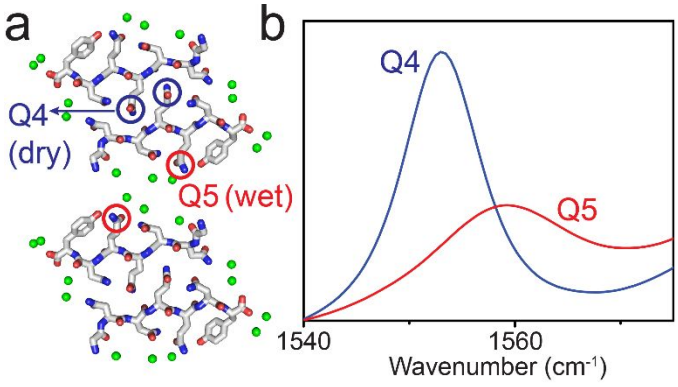


Figure 3. Peak linewidth is sensitive to local hydration status. (**a**) In the crystal structure of GNNQQNY (PDB 5K2G), glutamine side chains are in either dry (Q4, blue) or wet interfaces (Q5, red). Green spheres represent water molecules. (**b**) ATR-FTIR spectra of micro-crystals formed by GNNQQNY (Q4, blue) and GNNQQNY (Q5, red).

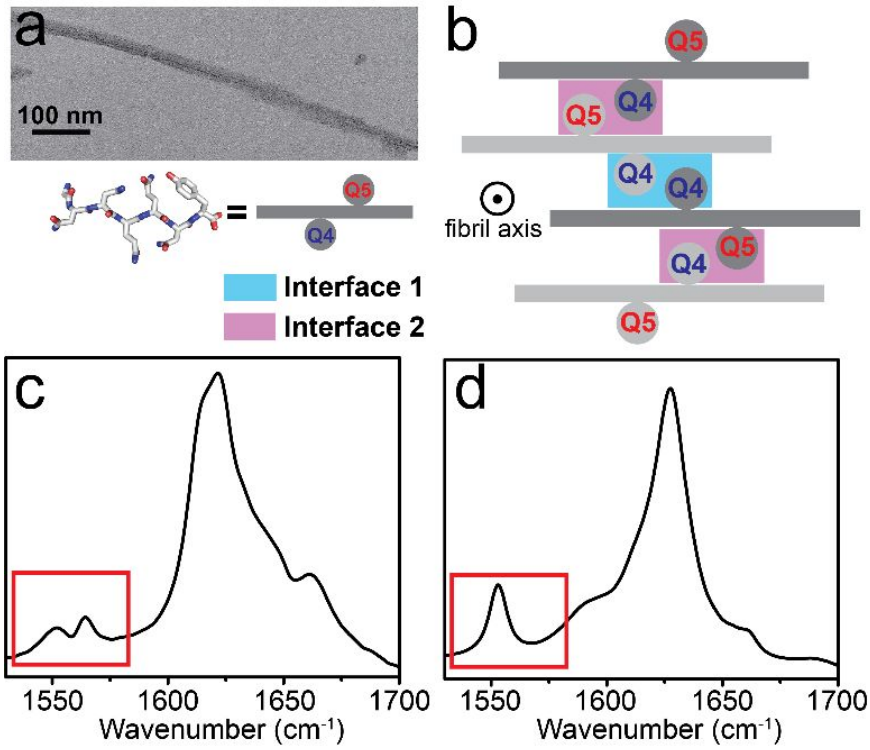


Figure 4. The characterization of GNNQQNY fibrils. (**a**) The EM image of GNNQQNY fibrils. Scale bar: 100 nm. (**b**) The GNNQQNY protofilament model proposed by Griffin and coworkers.[55] For clarity, side chains other than Q4 and Q5 are omitted. In the model, Q4 side chains are in two different interfaces—interface 1 (cyan) and interface 2 (pink), which occur in 1:1 molar ratio. A second model (not shown) proposed by Griffin and coworkers[55] has a similar structure, although the interfaces occur in different molar ratios. (**c**) The FTIR spectrum of fibrils is most consistent with the model shown in **b** and is quite distinct from what was observed in micro-crystals (**d**) formed by the same peptide.

Application of $^{13}C{=}^{18}O$ to structural and kinetical studies of amyloid fibrils. In the above studies, we demonstrated the high sensitivity of side chain $^{13}C{=}^{18}O$ using micro-crystalline GNNQQNY as a model system. To begin to explore side chain $^{13}C{=}^{18}O$ in amyloid fibrils, we further studied three fibril samples formed by GNNQQNY as well as two tau-derived peptide segments, Ac-VQIVYK-NH_2 and $tau_{306\text{-}321}$. These peptides differ in their original proteins and peptide lengths. In particular, $tau_{306\text{-}321}$ is significantly longer and contains an Asp residue, whose side chain carboxylate vibration could potentially overlap with Gln $^{13}C{=}^{18}O$.[56]

*GNNQQNY fibrils***.** In the above section, we investigated microcrystalline GNNQQNY formed at low concentrations in the presence of crystal seeds. The same peptide also forms amyloid fibrils at high concentrations, but the conformation is quite different from that in the microcrystalline form.[57-58] Using

solid-state NMR (ssNMR), Griffin and coworkers have extensively studied GNNQQNY fibril and proposed two closely related models.[55, 58-60] In each of these models, the sheets pack in two distinct zipper interfaces (**Fig. 4b,** light blue and pink). Q4 forms infinite hydrogen-bonded ladders that alternately pack with a second Q4 ladder or with another ladder from Q5. However, no other studies exist to support this packing arrangement with two distinct environments. We thus studied GNNQQNY fibrils by IR spectroscopy. As shown in **Fig. 4c**, the amide I' band of the fibrils displayed significantly higher complexity compared with that of micro-crystals (**Fig. 4d**). In contrast to the single $^{13}C=^{18}O$ peak observed in micro-crystals, fibrils displayed two peaks, suggesting that Q4 side chains are in two different environments. The data are consistent with the proposed models.

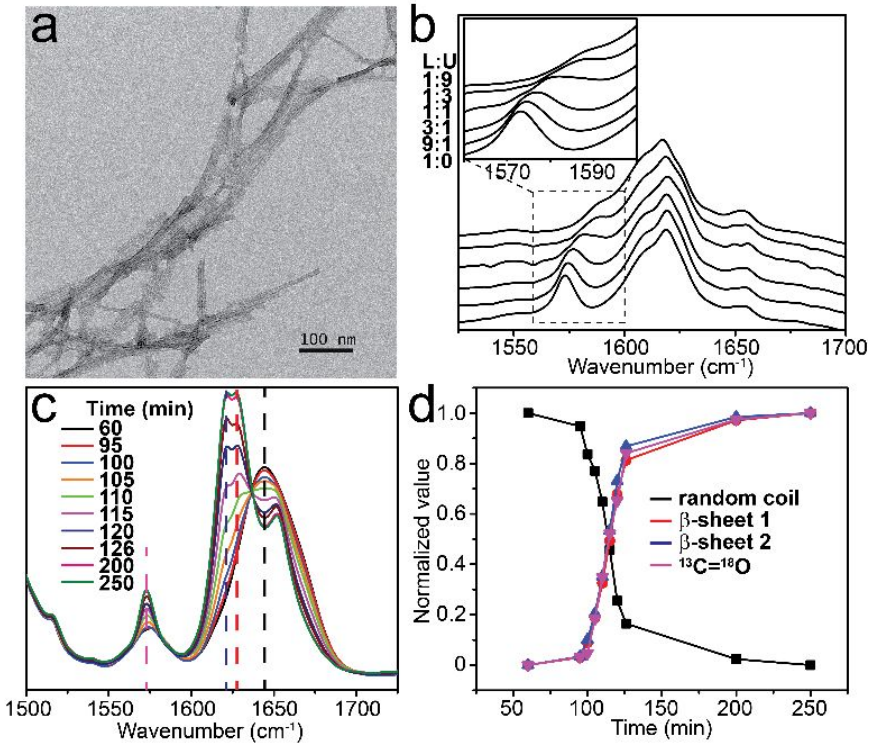


Figure 5. (**a**) The EM image of Ac-VQIVYK-NH_2 fibrils. Scale bar: 100 nm. (**b**) ATR-FTIR spectra of fibril samples with various ratios of Ac-VQIVYK-NH_2 (labeled, L) and Ac-VQIVYK-NH_2 (unlabeled, U). The intensity is normalized according to the 1619 cm^{-1} peak. (**c**) FTIR spectra were taken following the aggregation of $tau_{306-321}$ in 1xPBS (D_2O). Dashed lines denote different structural components. Black: random coil; red: beta-sheet 1; blue: beta-sheet 2; magenta: Gln side chain $^{13}C=^{18}O$. (**d**) The evolution of different structural components over time. The same color code as the dashed line is used.

Ac-VQIVYK-NH_2 fibrils. The hexapeptide ($tau_{306-311}$) is essential for the aggregation of tau that is associated with several neurodegenerative diseases.[61-62] Eisenberg and coworkers also solved the crystal structure of this peptide, showing the class 1 steric zipper.[6] However, few studies directly support the same

inter-sheet packing in the fibril form. We thus examined the fibril sample formed by Ac-VQIVYK-NH_2 with side chain $^{13}C=^{18}O$ on Q2. As shown in **Fig. 5b**, the main amide I' band is centered at 1618 cm^{-1}, and $^{13}C=^{18}O$ display a single peak at 1575 cm^{-1}. Moreover, similar to the study with GNNQQNY micro-crystals, isotope dilution caused a blue shift, which identified Gln ladders. Taken together, these results are consistent with the class 1 steric zipper in Ac-VQIVYK-NH_2 fibrils (also see discussions in the next section).

Kinetics of Gln ladder formation in tau$_{306-321}$. We next evaluate the use of side chain $^{13}C=^{18}O$-labeled Gln in a longer peptide, which also includes a carboxylate-containing side chain, potentially interfering with the side chain $^{13}C=^{18}O$ absorption.[56] To examine these issues, we studied tau$_{306-321}$ (^{306}VQIVYKPVDLSKVTSK^{321}), a peptide fragment derived from the third repeat region of tau and containing an Asp residue.[31] This peptide aggregates within a few hours and thus is well-suited for the kinetics measurement. As shown in **Fig. 5c**, this peptide initially adopted a random coil conformation with a broad amide I' band centered at 1645 cm^{-1} and another broad peak at 1575 cm^{-1}, contributed by the Asp side chain and unstructured $^{13}C=^{18}O$. As the aggregation proceeded, the random coil component started to decrease, and the spectral signature of β-sheet emerged. Concurrently, the 1575 cm^{-1} peak sharpened. Because no change for the 1575 cm^{-1} peak was observed in the unlabeled peptide (**Fig. S5**), this sharpening must be associated with the alignment of side chain $^{13}C=^{18}O$ to form a Gln ladder. By monitoring the changes of these components over time (**Fig. 5d**), sigmoidal curves typical of nucleation growth kinetics were observed and matched the previously reported kinetics trace by thioflavin T (ThT) signal.[31] Interestingly, these structural changes followed similar trends, suggesting a well-concerted process involving both main chain and side chain groups during the aggregation of tau$_{306-321}$. The isosbestic point at 1636 cm^{-1} (**Fig. 5c**) suggests a transition between two structural states (initial and final) with little intermediate species, consistent with the nucleation growth mechanism in which the nuclei grow rapidly once formed.[63] Our study represents the first investigation of Gln ladder formation during amyloid aggregation and should facilitate further investigation into this subject. Importantly, due to the homogeneity and much higher absorptivity of $^{13}C=^{18}O$ in Gln ladders, an Asp residue did not interfere with the analysis. We believe that, for more difficult systems containing multiple Asp/Glu residues, side chain $^{13}C=^{18}O$ study is highly feasible, particularly with the aid of two-dimensional methods.

IR information in structural modeling. In the absence of a high-resolution structure, IR data from side chain $^{13}C=^{18}O$ provides sensitive environmental information that can be used to assess alternative models of amyloid structures, thus producing a more precise final model. To demonstrate this application, we modeled amyloid protofilament structures with the open source *Integrative Modeling Platform* (IMP) program (http://integrativemodeling.org) using microcrystalline GNNQQNY as a test case (**Fig. 6a**).

As classified by Eisenberg and coworkers, the steric zipper can be described by eight different types of packing—classes 1 to 8—depending on strand-to-strand and sheet-to-sheet arrangements.[2-3, 6] Seven of the eight classes (all but class 3) have been observed experimentally. We first determined the zipper type using IR data of microcrystalline GNNQQNY (**Fig. 6a**). As discussed above, the main amide I' band supports parallel β-sheets, so we can rule out classes 5 to 8 that consist of anti-parallel β-sheets. However, information of the main chain alone is not sufficient to further distinguish among classes 1, 2, and 4. Using side chain $^{13}C{=}^{18}O$ and isotope dilution, we identified a hydrogen-bonded array of Q4 side chains, therefore supporting in-register β-sheets. Moreover, since both Q4 and Q5 side chains showed singlet peaks, classes 2 and 4 can be eliminated; otherwise, due to the translational symmetry in classes 2 and 4, we would observe alternating solvent exposed and buried environments for both Q4 and Q5 side chains. As a result, the class 1 steric zipper is the only one of the original eight classes that is consistent with the IR data.

Next, we generated an unbiased ensemble of GNNQQNY models that satisfy basic physical and geometrical restraints of the class 1 steric zipper, using a modified version of the kinematics module of IMP[64] to sample backbone and side chain dihedral angles and fibril geometry parameters (**Fig. 6b**). The resulting ensemble of 3.81 million models (**Fig. S6**, top left) was further filtered to include only those that satisfied the IR data of microcrystalline GNNQQNY, determined by the solvent accessible surface areas of Q4 and Q5 as well as a hydrogen bonding network along Q4 (see Table S1 for a list of the exact geometrical thresholds used). The resulting filtered ensemble (**Fig. S6**) was significantly reduced from 3.81 million to 82,500 models (2.6% of the original ensemble). In addition, the IR filter increased the accuracy of the ensemble, defined as the fibril root mean square deviation (fibril RMSD, detailed in **Fig. 6** and the method section) of fibril models to the crystal structure (PDB 1YJP)[5], from 8.5 Å for the entire ensemble to 5.2 Å for the IR filtered ensemble (**Fig. 6c**, left).

Many biophysical methods, including NMR spectroscopy, can give a broad range of distance constraints. Therefore, we explore how the IR filter performs compared with distance constraints and whether or not they are synergistic with each other. We thus included two simulated distance constraints between two interdigitated strands of mating sheets—6 Å for N2:CB–Q4:CD (Distance 1) and 5 Å for N2:CB–N6:CG (Distance 2). When applied individually, these constraints provided ensemble reductions to 6.9% and 2.9% and accuracy of 6.2 Å and 4.3 Å, respectively (**Fig. 6c**). The results are comparable to those from the IR filter. Importantly, combining the IR filter with Distance 1 (**Fig. 6c**, red dots) provided a further ensemble reduction to 1.4% and an increase in accuracy (1.0 Å). Integrating IR with Distance 2, which is more stringent, provided a more modest gain in accuracy (0.1 Å), yet with further ensemble reduction to 0.5%. In both cases, we benefit from considering all information (integrative modeling).

Interestingly, the IR information is more advantageous in determining the side chain intercalation (**Fig. S6**) and the sheet-sheet geometry in the steric zipper, particularly because it can be used to narrow down the class of packing. These results show that the utility of information from the IR is comparable to that for a single simulated distance constraint and that the two types of constraints can be used in an integrative fashion to produce more precise and accurate structural ensembles than either type of constraint alone.

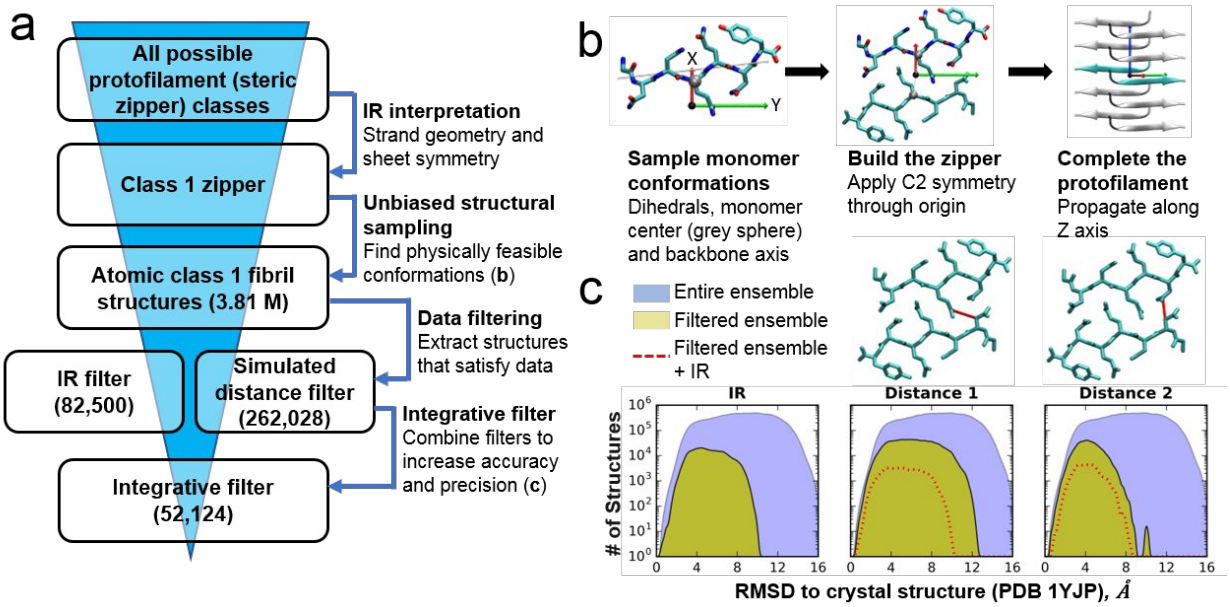


Figure 6. IR information on modeling structural ensembles of GNNQQNY. (**a**) A flowchart depicting how IR data was used to filter the ensemble of possible structures. (**b**) Building a protofilament by sampling only monomer parameters. The position of the monomer center (grey sphere) and yaw/pitch/roll of the fibril axis (grey line) are sampled along with backbone and side chain dihedral angles. The zipper was built by applying the C2 symmetry through the origin, and the protofilament was created by propagating the fibril along the Z axis. A total of 3.81 million models were generated. (**c**) Histograms showing the increase in accuracy and precision from data filtering using IR data and two simulated distance constraints and the integrative data filter. Fibril RMSD is defined as the sum of the all heavy atom RMSD of the model monomer to the crystal monomer plus the distance between the monomer centers.

Conclusions

In summary, we report glutamine side chain $^{13}C{=}^{18}O$ as an intrinsic vibrational probe and its application to the study of side chain environments within amyloid fibrils. The synthesis is robust and has been achieved on a multi-gram scale. This non-perturbative IR probe is highly sensitive to its local environment, including the carbonyl alignment and hydration status, making it well-suited for probing the

sheet-to-sheet arrangements in amyloids. We have successfully applied this probe to the studies of fibril structures and kinetics. Finally, we report a strategy of modeling amyloid structures using IR information and demonstrate the capability of integrating IR and other structural information for more accurate modeling. Thus, the combination of this sensitive IR probe and integrative modeling strategy will significantly advance IR spectroscopy in deciphering the structures and aggregation mechanisms of amyloid fibrils.

Supporting information

Experimental details and additional data, including Table S1 and Figures S1-S6.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

This work is supported by grants from the National Institute of Health: P01AG002132 (W.F.D. and A.S.), and R35GM122603 (W.F.D.). H.W. was supported by a CTSI TL1 Postdoctoral Fellowship (TL1TR001871). H.T.K. was supported by a Ruth L. Kirschstein NRSA Postdoctoral Fellowship (F32GM125217). We thank Dr. Michael Sawaya (UCLA) for suggestions regarding the micro-crystal preparation, as well as Dr. Feng Gai (UPenn) for helpful discussions.

References

1. Chiti, F.; Dobson, C. M., Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **2006,** *75*, 333-66.
2. Eisenberg, D.; Jucker, M., The amyloid state of proteins in human diseases. *Cell* **2012,** *148* (6), 1188-203.
3. Riek, R.; Eisenberg, D. S., The activities of amyloids from a structural perspective. *Nature* **2016,** *539* (7628), 227-235.
4. Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W., Functional amyloid--from bacteria to humans. *Trends Biochem Sci* **2007,** *32* (5), 217-24.
5. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D., Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **2005,** *435* (7043), 773-8.
6. Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T.; Madsen, A. O.; Riekel, C.; Eisenberg, D., Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **2007,** *447* (7143), 453-7.
7. Sawaya, M. R.; Rodriguez, J.; Cascio, D.; Collazo, M. J.; Shi, D.; Reyes, F. E.; Hattne, J.; Gonen, T.; Eisenberg, D. S., Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc Natl Acad Sci U S A* **2016,** *113* (40), 11232-11236.
8. Tycko, R., Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* **2011,** *62*, 279-99.
9. Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W., Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017,** *547* (7662), 185-190.
10. Fitzpatrick, A. W.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Muller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V.; Griffin, R. G.; Dobson, C. M., Atomic structure and hierarchical assembly of a cross-beta amyloid fibril. *Proc Natl Acad Sci U S A* **2013,** *110* (14), 5468-73.
11. Lu, J. X.; Qiang, W.; Yau, W. M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R., Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **2013,** *154* (6), 1257-68.
12. Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M.; George, J. M.; Rienstra, C. M., Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat Struct Mol Biol* **2016,** *23* (5), 409-15.
13. Lee, M.; Wang, T.; Makhlynets, O. V.; Wu, Y.; Polizzi, N. F.; Wu, H.; Gosavi, P. M.; Stohr, J.; Korendovych, I. V.; DeGrado, W. F.; Hong, M., Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proc Natl Acad Sci U S A* **2017,** *114* (24), 6191-6196.
14. Moran, S. D.; Zanni, M. T., How to get insight into amyloid structure and formation from infrared spectroscopy. *J Phys Chem Lett* **2014,** *5* (11), 1984-1993.
15. Ma, J.; Pazos, I. M.; Zhang, W.; Culik, R. M.; Gai, F., Site-specific infrared probes of proteins. *Annu Rev Phys Chem* **2015,** *66*, 357-77.
16. Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M., Two-dimensional infrared spectra of isotopically diluted amyloid fibrils from Abeta40. *Proc Natl Acad Sci U S A* **2008,** *105* (22), 7720-5.
17. Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M., 2D IR provides evidence for mobile water molecules in beta-amyloid fibrils. *Proc Natl Acad Sci U S A* **2009,** *106* (42), 17751-6.
18. 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., Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor. *Nat Chem* **2012,** *4* (5), 355-60.

19. Torres, J.; Adams, P. D.; Arkin, I. T., Use of a new label, (13)==(18)O, in the determination of a structural model of phospholamban in a lipid bilayer. Spatial restraints resolve the ambiguity arising from interpretations of mutagenesis data. *J Mol Biol* **2000,** *300* (4), 677-85.
20. Torres, J.; Kukol, A.; Goodman, J. M.; Arkin, I. T., Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic (13)C=(18)O group as a novel infrared probe. *Biopolymers* **2001,** *59* (6), 396-401.
21. Shim, S. H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T., Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc Natl Acad Sci U S A* **2009,** *106* (16), 6614-9.
22. Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P. N.; Chiu, C. C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T., Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient beta-sheet. *Proc Natl Acad Sci U S A* **2013,** *110* (48), 19285-90.
23. Buchanan, L. E.; Carr, J. K.; Fluitt, A. M.; Hoganson, A. J.; Moran, S. D.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T., Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc Natl Acad Sci U S A* **2014,** *111* (16), 5796-801.
24. Nagy-Smith, K.; Beltramo, P. J.; Moore, E.; Tycko, R.; Furst, E. M.; Schneider, J. P., Molecular, local, and network-level basis for the enhanced stiffness of hydrogel networks formed from coassembled racemic peptides: predictions from Pauling and Corey. *ACS Cent Sci* **2017,** *3* (6), 586-597.
25. Oh, K. I.; Lee, J. H.; Joo, C.; Han, H.; Cho, M., Beta-azidoalanine as an IR probe: application to amyloid Abeta(16-22) aggregation. *J Phys Chem B* **2008,** *112* (33), 10352-7.
26. Pazos, I. M.; Ghosh, A.; Tucker, M. J.; Gai, F., Ester carbonyl vibration as a sensitive probe of protein local electric field. *Angew Chem Int Ed Engl* **2014,** *53* (24), 6080-4.
27. Gao, Y. C.; Zou, Y.; Ma, Y.; Wang, D.; Sun, Y.; Ma, G., Infrared probe technique reveals a millipede-like structure for Abeta(8-28) Amyloid Fibril. *Langmuir* **2016,** *32* (4), 937-946.
28. Jia, B.; Sun, Y.; Yang, L.; Yu, Y.; Fan, H.; Ma, G., A structural model of the hierarchical assembly of an amyloid nanosheet by an infrared probe technique. *Phys Chem Chem Phys* **2018,** *20* (43), 27261-27271.
29. Pazos, I. M.; Ma, J.; Mukherjee, D.; Gai, F., Ultrafast hydrogen-bonding dynamics in amyloid fibrils. *J Phys Chem B* **2018**, 122 (49), 11023–11029.
30. Adhikary, R.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E., IR probes of protein microenvironments: utility and potential for perturbation. *Chemphyschem* **2014,** *15* (5), 849-53.
31. Stohr, J.; Wu, H.; Nick, M.; Wu, Y.; Bhate, M.; Condello, C.; Johnson, N.; Rodgers, J.; Lemmin, T.; Acharya, S.; Becker, J.; Robinson, K.; Kelly, M. J. S.; Gai, F.; Stubbs, G.; Prusiner, S. B.; DeGrado, W. F., A 31-residue peptide induces aggregation of tau's microtubule-binding region in cells. *Nat Chem* **2017,** *9* (9), 874-881.
32. Martial, B.; Lefevre, T.; Auger, M., Understanding amyloid fibril formation using protein fragments: structural investigations via vibrational spectroscopy and solid-state NMR. *Biophys Rev* **2018**, 10(4), 1133-1149.
33. Gallagher-Jones, M.; Glynn, C.; Boyer, D. R.; Martynowycz, M. W.; Hernandez, E.; Miao, J.; Zee, C. T.; Novikova, I. V.; Goldschmidt, L.; McFarlane, H. T.; Helguera, G. F.; Evans, J. E.; Sawaya, M. R.; Cascio, D.; Eisenberg, D. S.; Gonen, T.; Rodriguez, J. A., Sub-angstrom cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat Struct Mol Biol* **2018,** *25* (2), 131-134.
34. Davies, S. W.; Turmaine, M.; Cozens, B. A.; DiFiglia, M.; Sharp, A. H.; Ross, C. A.; Scherzinger, E.; Wanker, E. E.; Mangiarini, L.; Bates, G. P., Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **1997,** *90* (3), 537-48.
35. Michelitsch, M. D.; Weissman, J. S., A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc Natl Acad Sci U S A* **2000,** *97* (22), 11910-5.

36. Perutz, M. F.; Pope, B. J.; Owen, D.; Wanker, E. E.; Scherzinger, E., Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid beta-peptide of amyloid plaques. *Proc Natl Acad Sci U S A* **2002,** *99* (8), 5596-600.
37. Kurt, T. D.; Aguilar-Calvo, P.; Jiang, L.; Rodriguez, J. A.; Alderson, N.; Eisenberg, D. S.; Sigurdson, C. J., Asparagine and glutamine ladders promote cross-species prion conversion. *J Biol Chem* **2017,** *292* (46), 19076-19086.
38. Wattendorff, A. R.; Bots, G. T.; Went, L. N.; Endtz, L. J., Familial cerebral amyloid angiopathy presenting as recurrent cerebral haemorrhage. *J Neurol Sci* **1982,** *55* (2), 121-35.
39. Abrahamson, M.; Jonsdottir, S.; Olafsson, I.; Jensson, O.; Grubb, A., Hereditary cystatin C amyloid angiopathy: identification of the disease-causing mutation and specific diagnosis by polymerase chain reaction based analysis. *Hum Genet* **1992,** *89* (4), 377-80.
40. Adegbuyiro, A.; Sedighi, F.; Pilkington, A. W. t.; Groover, S.; Legleiter, J., Proteins containing expanded polyglutamine tracts and neurodegenerative disease. *Biochemistry* **2017,** *56* (9), 1199-1217.
41. Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A., Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci* **2018,** *27* (1), 245-258.
42. Marecek, J.; Song, B.; Brewer, S.; Belyea, J.; Dyer, R. B.; Raleigh, D. P., A simple and economical method for the production of 13C,18O-labeled Fmoc-amino acids with high levels of enrichment: applications to isotope-edited IR studies of proteins. *Org Lett* **2007,** *9* (24), 4935-7.
43. Seyfried, M. S.; Lauber, B. S.; Luedtke, N. W., Multiple-turnover isotopic labeling of Fmoc- and Boc-protected amino acids with oxygen isotopes. *Org Lett* **2010,** *12* (1), 104-6.
44. Abaskharon, R. M.; Brown, S. P.; Zhang, W.; Chen, J.; Smith, A. B., 3rd; Gai, F., Isotope-labeled aspartate sidechain as a non-perturbing infrared probe: Application to Investigate the Dynamics of a Carboxylate Buried Inside a Protein. *Chem Phys Lett* **2017,** *683*, 193-198.
45. Ghaffar, T.; Parkins, A. W., A new homogeneous platinum-containing catalyst for the hydrolysis of nitriles. *Tetrahedron Lett* **1995,** *36* (47), 8657-8660.
46. Richter, M. J. R.; Schneider, M.; Brandstatter, M.; Krautwald, S.; Carreira, E. M., Total synthesis of (-)-Mitrephorone A. *J Am Chem Soc* **2018,** 140 (48), 16704–16710.
47. Mercado-Marin, E. V.; Garcia-Reynaga, P.; Romminger, S.; Pimenta, E. F.; Romney, D. K.; Lodewyk, M. W.; Williams, D. E.; Andersen, R. J.; Miller, S. J.; Tantillo, D. J.; Berlinck, R. G. S.; Sarpong, R., Total synthesis and isolation of citrinalin and cyclopiamine congeners. *Nature* **2014,** *509* (7500), 318-324.
48. Hirooka, Y.; Ikeuchi, K.; Kawamoto, Y.; Akao, Y.; Furuta, T.; Asakawa, T.; Inai, M.; Wakimoto, T.; Fukuyama, T.; Kan, T., Enantioselective synthesis of SB-203207. *Org Lett* **2014,** *16* (6), 1646-9.
49. Silva, R. A.; Barber-Armstrong, W.; Decatur, S. M., The organization and assembly of a beta-sheet formed by a prion peptide in solution: an isotope-edited FTIR study. *J Am Chem Soc* **2003,** *125* (45), 13674-5.
50. Paul, C.; Wang, J.; Wimley, W. C.; Hochstrasser, R. M.; Axelsen, P. H., Vibrational coupling, isotopic editing, and beta-sheet structure in a membrane-bound polypeptide. *J Am Chem Soc* **2004,** *126* (18), 5843-50.
51. Choi, J. H.; Ham, S. Y.; Cho, M., Local amide I mode frequencies and coupling constants in polypeptides. *Journal of Physical Chemistry B* **2003,** *107* (34), 9132-9138.
52. Dunkelberger, E. B.; Buchanan, L. E.; Marek, P.; Cao, P.; Raleigh, D. P.; Zanni, M. T., Deamidation accelerates amyloid formation and alters amylin fiber structure. *J Am Chem Soc* **2012,** *134* (30), 12658-67.
53. Bakker, H. J.; Skinner, J. L., Vibrational spectroscopy as a probe of structure and dynamics in liquid water. *Chem Rev* **2010,** *110* (3), 1498-517.
54. Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L., Development and validation of transferable amide I vibrational frequency maps for peptides. *J Phys Chem B* **2011,** *115* (13), 3713-24.
55. Lewandowski, J. R.; van der Wel, P. C.; Rigney, M.; Grigorieff, N.; Griffin, R. G., Structural complexity of a composite amyloid fibril. *J Am Chem Soc* **2011,** *133* (37), 14686-98.

56. Barth, A., The infrared absorption of amino acid side chains. *Prog Biophys Mol Biol* **2000,** *74* (3-5), 141-73.
57. Balbirnie, M.; Grothe, R.; Eisenberg, D. S., An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc Natl Acad Sci U S A* **2001,** *98* (5), 2375-80.
58. van der Wel, P. C.; Lewandowski, J. R.; Griffin, R. G., Solid-state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J Am Chem Soc* **2007,** *129* (16), 5117-30.
59. Debelouchina, G. T.; Bayro, M. J.; van der Wel, P. C.; Caporini, M. A.; Barnes, A. B.; Rosay, M.; Maas, W. E.; Griffin, R. G., Dynamic nuclear polarization-enhanced solid-state NMR spectroscopy of GNNQQNY nanocrystals and amyloid fibrils. *Phys Chem Chem Phys* **2010,** *12* (22), 5911-9.
60. van der Wel, P. C.; Lewandowski, J. R.; Griffin, R. G., Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR. *Biochemistry* **2010,** *49* (44), 9457-69.
61. von Bergen, M.; Friedhoff, P.; Biernat, J.; Heberle, J.; Mandelkow, E. M.; Mandelkow, E., Assembly of tau protein into Alzheimer paired helical filaments depends on a local sequence motif ((306)VQIVYK(311)) forming beta structure. *Proc Natl Acad Sci U S A* **2000,** *97* (10), 5129-34.
62. Mandelkow, E. M.; Mandelkow, E., Biochemistry and cell biology of tau protein in neurofibrillary degeneration. *Cold Spring Harb Perspect Med* **2012,** *2* (7), a006247.
63. Jarrett, J. T.; Lansbury, P. T., Jr., Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **1993,** *73* (6), 1055-8.
64. Carter, L.; Kim, S. J.; Schneidman-Duhovny, D.; Stohr, J.; Poncet-Montange, G.; Weiss, T. M.; Tsuruta, H.; Prusiner, S. B.; Sali, A., Prion protein-antibody complexes characterized by chromatography-coupled small-angle x-ray scattering. *Biophys J* **2015,** *109* (4), 793-805.