

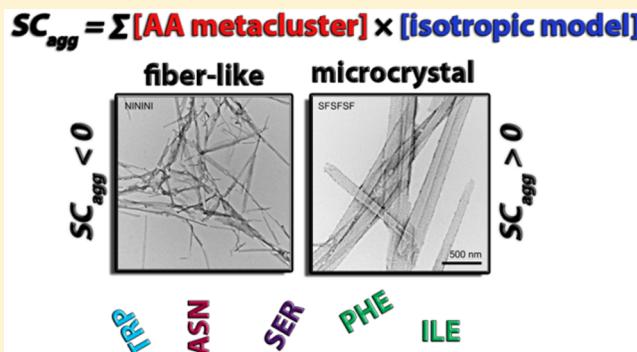
# Amino Acid Metaclusters: Implications of Growth Trends on Peptide Self-Assembly and Structure

Thanh D. Do,<sup>†</sup> Natália E. C. de Almeida,<sup>†</sup> Nichole E. LaPointe,<sup>‡</sup> Ali Chamas,<sup>†</sup> Stuart C. Feinstein,<sup>‡</sup> and Michael T. Bowers\*,<sup>†</sup>

<sup>†</sup>Department of Chemistry and Biochemistry and <sup>‡</sup>Neuroscience Research Institute and Department of Molecular Cellular and Developmental BiologyUniversity of California, Santa Barbara, California 93106, United States

Supporting Information

**ABSTRACT:** Ion-mobility mass spectrometry is utilized to examine the metacluster formation of serine, asparagine, isoleucine, and tryptophan. These amino acids are representative of different classes of noncharged amino acids. We show that they can form relatively large metaclusters in solution that are difficult or impossible to observe by traditional solution techniques. We further demonstrate, as an example, that the formation of Ser metaclusters is not an ESI artifact because large metaclusters can be detected in negative polarity and low concentration with similar cross sections to those measured in positive polarity and higher concentration. The growth trends of tryptophan and isoleucine metaclusters, along with serine, asparagine, and the previously studied phenylalanine, are balanced among various intrinsic properties of individual amino acids (e.g., hydrophobicity, size, and shape). The metacluster cross sections of hydrophilic residues (Ser, Asn, Trp) tend to stay on or fall below the isotropic model trend lines whereas those of hydrophobic amino acids (Ile, Phe) deviate positively from the isotropic trend lines. The growth trends correlate well to the predicted aggregation propensity of individual amino acids. From the metacluster data, we introduce a novel approach to score and predict aggregation propensity of peptides, which can offer a significant improvement over the existing methods in terms of accuracy. Using a set of hexapeptides, we show that the strong negative deviations of Ser metaclusters from the isotropic model leads a prediction of microcrystalline formation for the SFSFSF peptide, whereas the strong positive deviation of Ile leads to prediction or fibril formation for the NININI peptide. Both predictions are confirmed experimentally using ion mobility and TEM measurements. The peptide SISI is predicted to only weakly aggregate, a prediction confirmed by TEM.

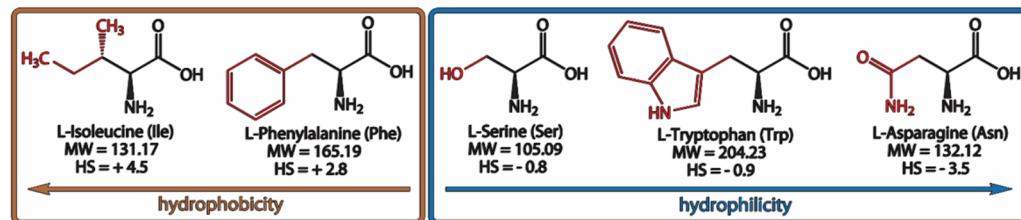


Protein aggregation is an intriguing phenomenon that has been linked to pathological agents in devastating diseases (e.g., amyloid  $\beta$  ( $A\beta$ )<sup>1–4</sup> and tau proteins in Alzheimer's,<sup>5–8</sup>  $\alpha$ -synuclein in Parkinson's,<sup>9,10</sup> polyglutamine in Huntington's,<sup>11,12</sup> islet amyloid polypeptide (IAPP) in Diabetes type II,<sup>13,14</sup> etc.) and at the same time has been found to have promising applications in the development of new functional materials.<sup>15,16</sup> Aggregation prone regions (e.g.,  $A\beta$ (16–22) [KLVFFAE] in  $A\beta$ ,<sup>17,18</sup> PHF6 [VQIVYK] in Tau,<sup>19,20</sup> GNNQQNY in Prion,<sup>21</sup> hIAPP(23–27) [FGAIL] in IAPP,<sup>22</sup> etc.) often dictate the self-assembly kinetics and aggregate morphologies of the parent protein. Of note, many of these peptides aggregate into amyloid-like fibrils (typical width of 5–20 nm), whereas a few of them form microcrystals in addition to fibrils.<sup>21</sup> Although previous work has shown that the individual peptide conformation inside a fibril or microcrystalline structure may be very similar,<sup>21</sup> the reason the system forms one type of morphology and not another remains elusive. The microcrystalline structures often have a diameter that is 10× wider than a traditional fibril.

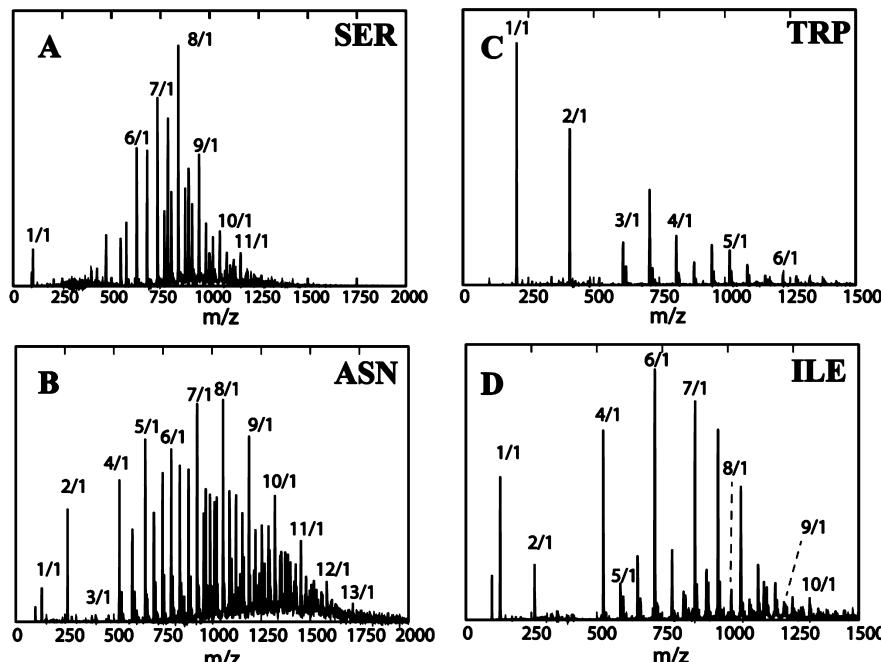
Recent research in the past decade has focused on identifying aggregation-prone regions (often referred to as “hot-spots”) within a large protein and using these regions to predict their aggregation propensities (for review, see ref 23). A majority of methods utilize either statistical or parametrized approaches in prediction,<sup>24–29</sup> largely because intermolecular interactions at the amino acid level are unknown for most peptides. Major factors such as hydrophobicity, side chain packing, secondary structure propensity, and residue pattern are often parametrized independently and considered as additive factors when calculating aggregation propensity,<sup>27,28</sup> although in nature they all correlate and affect each other. There is a need for experimental data that quantify these various factors as a single but unique value for each amino acid.

Proteins are composed of amino acids. Diverse, complex, and highly functional systems arise as the result of interactions between a finite number of residue types. A majority of the

**Received:** September 10, 2015  
**Accepted:** December 3, 2015  
**Published:** December 3, 2015



**Figure 1.** Structural formulas of the five amino acids whose metacluster growth trends are investigated. Each amino acid is shown together with its molecular weight (MW, g/mol) and hydrophathy score (HS).



**Figure 2.** Representative nano-ESI-q mass spectra of 6 mM amino acid in water for (A) L-serine, (B) L-asparagine, (C) L-tryptophan, and (D) L-isoleucine. A number of mass spectral peaks are annotated with the ratio  $n/z$ , where  $n$  is the metacluster number and  $z$  the charge. The peaks that must have  $z > 1$  are not annotated, but all mass spectral peaks are utilized to determine the metacluster cross sections.

interactions occurring within and across systems, especially in self-assembly and aggregation, are pairwise between residues with similar properties, such as hydrophobicity or hydrophilicity, which are often regulated by the intrinsic properties of individual amino acids (e.g., size, shape, charge, and chirality). Amino acids can come together via these interactions to form metaclusters (i.e., clusters that are metastable) with the most notable examples being serine,<sup>30–39</sup> phenylalanine,<sup>40–43</sup> arginine,<sup>44</sup> tryptophan,<sup>45</sup> proline,<sup>46</sup> and tyrosine.<sup>47</sup> These studies have provided invaluable insights into the pairwise interactions of amino acids, which although noncovalent and transiently weak, are extremely important to nucleation and self-assembly of biological structures.<sup>42,47</sup>

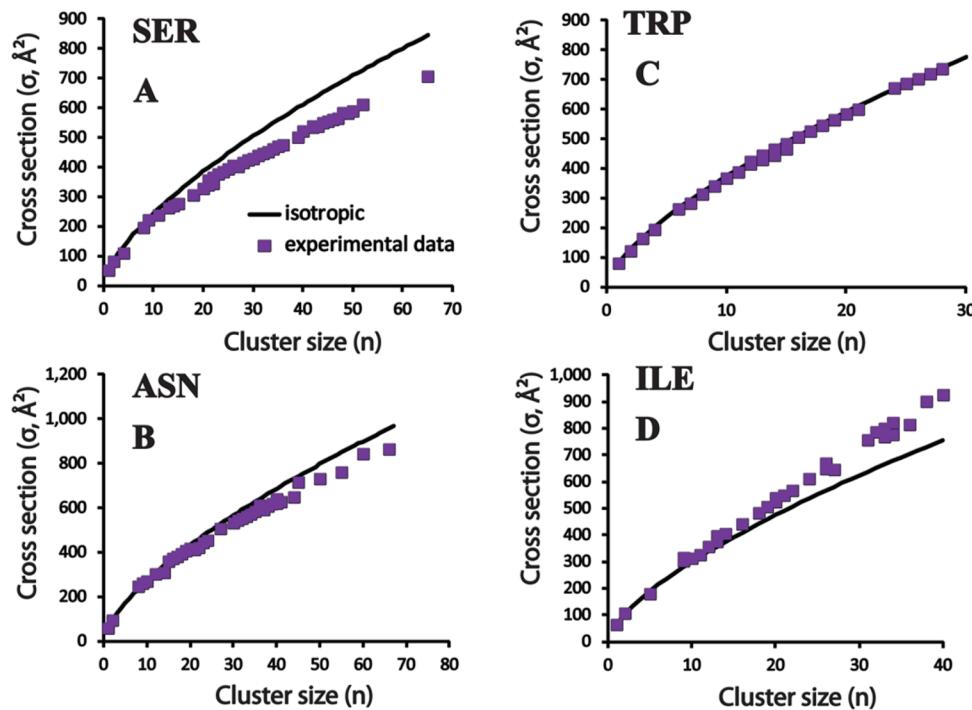
Recent advances in the field of mass spectrometry allow a detailed determination of molecular clusters formed by amino acids.<sup>35,37–39,44,48</sup> Here, we use ion-mobility mass spectrometry (IM-MS) to examine the ability of several amino acids to form molecular metaclusters and the properties that contribute to the relationship between cluster size and shape. Accurate measurements of collisional cross sections achieved from high resolution ion mobility experiments in combination with electrospray ionization (ESI) allow unambiguous characterization of stable amino acid metaclusters. We study the growth trends of molecular clusters of some representative hydro-

phobic and hydrophilic/polar amino acids. The growth trends of amino acid metaclusters provide unique information about the intermolecular interactions at the residue level. These data provide a novel approach toward the development of a new empirical method for predicting aggregation propensities of short peptides and distinguishing amyloid from nonamyloid formation.

## RESULTS AND DISCUSSION

**Amino Acids Form Large Metaclusters with Distinctive Growth Trends.** During the self-assembly process, mesoscopic structures are formed by hydrophobic and hydrophilic interactions while at the same time excluding water.<sup>49,50</sup> Here, we use IM-MS to investigate the metacluster growth trends of representative hydrophilic and hydrophobic amino acids specifically. These amino acids include serine, asparagine, tryptophan, and isoleucine. The side chains of these amino acids carry different functional groups that can facilitate various types of interactions: hydroxyl, carboxamide, alkyl, and aromatic side chains as seen in Figure 1.

**Serine and Asparagine.** Because of its implications for biomolecule homochirality,<sup>32,51</sup> serine (Ser) cluster formation has been investigated by theoretical calculations<sup>38,39</sup> and by multiple mass spectrometry based techniques including H/D



**Figure 3.** Plots of metacluster size ( $n$ ) versus experimental cross section ( $\sigma, \text{Å}^2$ ). The experimental values are compared to the isotropic model  $\sigma_n = \sigma_1 \times n^{2/3}$ , where  $\sigma_1$  is the monomer cross section for L-serine ( $\sigma_1 = 52 \text{ Å}^2$ ), L-asparagine ( $\sigma_1 = 59 \text{ Å}^2$ ), L-tryptophan ( $\sigma_1 = 80 \text{ Å}^2$ ), and L-isoleucine ( $\sigma_1 = 65 \text{ Å}^2$ ).

exchange,<sup>35,37</sup> ion-mobility spectrometry,<sup>38,39</sup> and tandem MS/MS.<sup>31,32,34</sup> Ser octamer is a stable magic-number cluster originally reported by Cooks and co-workers.<sup>34</sup> The octamer can be generated through means of aerosol ionization techniques (e.g., electrospray, cold spray, sonic spray, and electrosonic spray), evaporation, and sublimation.<sup>32</sup> When putting Ser in a peptide and protein context, previous studies have found conflicting effects of Ser on aggregation propensity. For example, Wetzel and co-workers have shown that Ser plays an important role in full-length human glutamine-rich Huntington pathogenesis and that Ser phosphorylation or mutation abrogates aggregation and disease symptoms.<sup>52</sup> On the other hand, Ser phosphorylation or mutation in  $\alpha$ -synuclein, a protein linked genetically and pathologically to Parkinson's diseases,<sup>10,11</sup> can either enhance or protect against aggregation and toxicity.<sup>53,54</sup> The ESI-mass spectrum of 6 mM L-serine in water is shown in Figure 2A. We note that the concentration used here is comparable or lower than the concentration used in previous work (~10 mM).<sup>32,38,39</sup> The mass spectrum populates many intense mass spectral peaks, including both singly and multiply charged state peaks. An arrival time distribution (ATD) is obtained for each peak (see Figure S1), and the features in the ATDs are assigned and cross sections obtained for each as described in the Experimental Methods. The metacluster cross sections are compared to the values predicted by the isotropic model. This model provides a good approximation for systems that grow isotropically to form spherical-like oligomers.<sup>55</sup> We note that the presence of the multiply charged state mass spectral peaks corresponding to even-sized clusters with odd charge states or odd-sized clusters with even charge states are very useful for ATD assignments. The cross sections of odd  $n$  serine metaclusters are notably smaller than the values predicted by the isotropic model (see Figure 3A) and consistent with our assignments for even  $n$

mass spectral peaks (see Tables S1). The growth trend of Ser indicates that the amino acid monomers are tightly packed to form the metaclusters, favoring neither globular shape nor isotropic growth, and that individual monomers within clusters become more compact as cluster size increases. This is the first system we have observed in which the growth trend falls significantly below the isotropic model prediction. We note that the Ser monomer cross section ( $\sigma_1 = 52.24 \text{ Å}^2$ ) is in good agreement with the theoretical cross section of the Ser crystal structure obtained by Benedetti et al.<sup>56</sup> and by Kistenmacher et al.<sup>57,58</sup> ( $\sigma_{\text{theory}} = 50 \text{ Å}^2$ ). We also note that the negative deviation of the Ser growth trend from the isotropic model is observed using negative ESI polarity and at lower concentration (250  $\mu\text{M}$ ) (see Figures S8 and S9 for the ATDs and growth curve), indicating that the structure of Ser metaclusters reflect solution assembly and are not driven by ESI polarity or solution concentration.

Moggach et al. solved six different X-ray crystal structures of L-serine at room temperature and at pressures between 0.3 and 5.4 GPa.<sup>58</sup> We used these X-ray structures to construct the theoretical models of serine 24-mer and computed their collision cross sections using the trajectory method (TJ)<sup>59,60</sup> and the projected superposition approximation (PSA) method.<sup>61,62</sup> The 24-mer ( $n/z = 24/3$ ) is the dominant species observed in the ATDs of the "magic" cluster peak at 841  $m/z$ .<sup>38,39</sup> The theoretical cross sections of these models (see Table S6 and Figure S10) agree well with our experimental data ( $\sigma_{\text{exp}} = 385 \text{ Å}^2$  in positive mode and  $373 \text{ Å}^2$  in negative mode ESI, respectively) and are significantly smaller than the isotropic cross section of  $435 \text{ Å}^2$ . From these data, we suggest that the negative deviation of the Ser growth trend from the isotropic curve may correlate to the microcrystalline (nonamyloid) aggregation of Ser.<sup>39</sup>

There is strong evidence that asparagine (Asn)-rich regions promote amyloid fibril formation through the creation of hydrogen-bonded spines ("asparagine ladders") between neighboring repetitive residues.<sup>63</sup> Lindquist and co-workers have shown that replacing glutamine (Glu) with Asn increases amyloid formation by glutamine-rich proteins.<sup>64</sup> In a previous study, we showed that Asn is essential to maintain the amyloid formation of Sup35 NNQQNY mutants. Mutating one or all Gln to hydrophobic residues enhances amyloid formation (e.g., NVVVYY, NNVVNY, NNVQIY, NVQVYY), whereas replacing the Asn residues (VIQVYY) drives the aggregation into nonamyloid pathways.<sup>65</sup> Similar to Ser, the mass spectrum of 6 mM L-asparagine (Figure 2B) populates mass spectral peaks corresponding to singly and multiply charged metaclusters. The Asn metaclusters follow the isotropic curve up to  $n = 10$  and then begin to negatively deviate from the curve (see Figure 3B); however, the difference between the model and experimental data is smaller as compared to the case of Ser. This can be partially attributed to the physical bulkiness of the carboxamide side chain of Asn, which is larger than the hydroxyl side chain of Ser.

**Isoleucine and Tryptophan.** In the above analysis, the physical bulkiness of amino acid side chains is proposed to contribute to the deviation of metacluster growth trends from the isotropic predictions. The question is, how much deviation should be attributed to this factor? We address this question by examining the growth trends of hydrophilic tryptophan (Trp, hydropathy index = -0.9) and hydrophobic isoleucine (Ile, hydropathy index = 4.5). A recent study on L-phenylalanine (Phe) aggregation has shown that the metacluster growth of Phe deviates positively from the isotropic model.<sup>42</sup> Further analysis on the growth trends suggests that Phe metaclusters assemble in tubular structures.<sup>42,43</sup> Of note, Phe is a quite hydrophobic amino acid (hydropathy score = 2.8) with strong  $\pi$ -stacking among residues.<sup>42</sup> If side chain bulkiness is the sole factor, the indole side chain of Trp should promote a more pronounced positive deviation from the isotropic model than that of Phe, which possesses a less bulky benzyl side chain, whereas the even smaller hydrocarbon side chain of Ile should make the growth trend of Ile metaclusters deviate much less from or even fall below the isotropic curve.

The mass spectra of 6 mM L-tryptophan and L-isoleucine in water are shown in Figure 2C and D, respectively. There are fewer multiply charged state peaks as compared to the mass spectra of Ser and Asn. The growth trend of Trp metaclusters follows the isotropic model with no significant deviation, and that of Ile grows almost linearly after  $n = 10$  (Figure 3C). These observations eliminate the possibility that side chain bulkiness is a major contributor to the characteristics of amino acid growth trends. With respect to aggregation propensity, it is interesting that the growth trend of Ile resembles those of amyloid peptides.<sup>55,65–67</sup> Hydrophobic amino acids, such as Ile, Leu, and Phe, often score higher in aggregation propensity than hydrophilic ones.<sup>68</sup> Prediction algorithms score these two residues very high in terms of aggregation propensity.<sup>27,28,69</sup> We note that "strong/weak aggregation propensity" in an amino acid context should be understood as a strong/weak contribution to aggregation of a peptide in which the amino acid is a residue. Dipeptides made of Ile and Phe (e.g., Ile-Ile, Phe-Phe, Ile-Phe) are some of the smallest aggregating systems.<sup>41,43,70–72</sup> However, we have previously found only modest agreement between predictions and experiments.<sup>65,67</sup> Furthermore, it is unclear whether nonaromatic hydrophobic

amino acids (e.g., isoleucine, leucine, and valine) are better aggregation promoters than aromatic amino acids (e.g., phenylalanine, tryptophan, and tyrosine). Although there is not yet a systematic study on aggregation propensity of nonaromatic and aromatic-containing peptides, we have shown that YVIFL, a [leu-5] Enkephalin mutant containing isoleucine, aggregates much faster than YVVFL or YVVFV mutants, which contain phenylalanine but no isoleucine.<sup>67</sup> From our data here, the growth trend of Ile metaclusters deviates further from the isotropic model than Phe or Trp, indicating that Ile is the better aggregation promotor than Phe or Trp.

Trp forms smaller metaclusters than either Ile and Phe ( $n < 30$ ) under our experimental conditions, and their cross sections follow the isotropic growth curve. Trp is not often involved in interactions leading to self-assembly or aggregation. In fact, it appears to be an aggregation blocker in a few systems.<sup>73</sup> This amino acid is expected to play a neutral role in controlling aggregation propensities of peptides and proteins.

### An Empirical Approach to Predicting Peptide

**Aggregation Propensity.** The growth trends of the amino acids suggest that hydrophobicity is a very important factor in the amyloid self-assembly process. The growth trend of Ile metaclusters (hydropathy score = 4.5) deviates more from the isotropic curve than that of Phe (2.8), and Phe deviates more than Trp (-0.9).<sup>74</sup> Although the hydropathic properties of amino acids correlate well with the metacluster growth trends, independent of the physical bulkiness of the side chains, it may not be the only factor that significantly affects the assemblies. Additional factors may include the secondary structure propensity of each amino acid and hydrophilic interactions such as hydrogen bonding and electrostatics. Recent studies have focused on identifying aggregation-prone fragments and predicting amyloid fibril propensity based on parametrizing each of these factors separately.<sup>23</sup> Several sophisticated algorithms for predicting  $\beta$ -sheet aggregation propensity have been developed based on this strategy;<sup>24–29,75,76</sup> however, these still need improvement.<sup>65</sup> One important drawback of approaches taken in the literature is the neglect of the strong correlation among different factors affecting aggregation. This is where amino acid metacluster data may be able to offer major improvement, because each amino acid's growth trend should naturally reflect how various factors interact and affect assembly.

An additional benefit of our approach is that these growth trends may discriminate well between amyloid and nonamyloid assemblies. Sawaya et al. described several cross- $\beta$  structures capable of forming amyloid fibrils and microcrystals.<sup>21</sup> In many instances, their analysis does not discriminate between these two tendencies, although they result in very different aggregates. An amyloid fibril is 5–20 nm in width, whereas microcrystals often have larger dimensions and appear in well-defined two-dimensional shapes. In contrast, our analysis of the amino acid metacluster data takes into account both positive and negative deviation from the isotropic model. When the deviation is positive, the system is expected to be extended as in  $\beta$ -sheet formation, suggesting eventual fibril formation. A negative deviation, on the other hand, implies that the structures rearrange into compact structures, implying eventual crystal formation. Including both types of deviation results in a meaningful indication of the nature of the aggregation.

Here, we used the present limited data set to develop an empirical approach to score peptide aggregation propensity. Because the growth trend of Trp shows the smallest deviation

from the isotropic model out of the amino acids presented in this work,<sup>42</sup> we choose to consider Trp as a reference in our scoring system. The aggregation propensity score  $SC_{agg}$ , expressed as a function of the Trp growth curve, is defined as

$$SC_{agg} = - \left[ \frac{1}{l} \sum_i^{\text{all amino acids}} n_i \left( \frac{\sigma_{1,\text{Trp}}}{\sigma_{1,i}} \right) (\Delta i - 1) \times 100 \right] \quad (\text{eq } 1)$$

where  $l$  is the number of residues in the peptide (e.g., hexapeptide  $l = 6$ ),  $n_i$  is the number of different amino acids  $i$  in the peptide (e.g., in NININI,  $n_{\text{Ile}} = 3$ ),  $\sigma_{1,i}$  is the experimental cross section of amino acid  $i$  monomer, and  $\Delta i$  is the ratio of the value of the experimental growth trend of amino acid  $i$  to that of the isotropic model, as determined by IM-MS experiments. The quantity  $(\sigma_{1,\text{Trp}}/\sigma_{1,i})$  normalizes the cross section of the monomer of the amino acid  $i$  to the cross section of the Trp monomer. The value for  $\Delta i$  can be approximated by taking the ratios of very large metaclusters ( $n > 30$ ) of amino acid  $i$  to the corresponding values predicted by the isotropic model.<sup>55,65</sup> Lastly, the negative sign makes our scoring consistent with the convention established by most existing methods in which a negative score indicates high amyloid aggregation propensity. Zyggregator is an exception in which a more positive score indicates stronger aggregation propensity. As mentioned above, in addition to being an empirical method, the second advantage of our scoring system is that it is bidirectional. That is, a negative score predicts a strong amyloid fibril formation propensity and a positive score implies three-dimensional microcrystal formation. When the score is close to zero, the system is predicted to be weakly or nonaggregating. In other methods, the scoring scheme is unidirectional (e.g., in PASTA,<sup>24–26</sup> a high negative score predicts a high amyloid aggregation propensity, but a positive score is not predictive).

As an empirical test of our method, we constructed a set of hexapeptides from the four amino acids, Phe, Ile, Asn, and Ser, namely NININI, NFNFNF, SISISI, and SFSFSF. Here, we will evaluate the aggregation propensities of these peptides experimentally and compare the results to the predictions from our peptide scoring method. Table 1 lists the aggregation propensities predicted by our method as well as some widely known algorithms, specifically PASTA (Prediction of Amyloid Structural Aggregation; <http://protein.bio.unipd.it/pasta2/>),<sup>24–26</sup> 3D RosettaProfile (<http://services.mbi.ucla.edu/zipperdb/>),<sup>76</sup> and Zyggregator (<http://www-mvsoftware.ch.cam.ac.uk/index.php/zyggregator>)<sup>27,28</sup> algorithms. For all of

the methods, a more negative score in Table 1 predicts a higher  $\beta$ -sheet aggregation propensity because the Zyggregator scores were multiplied by  $(-1)$  to make them consistent in sign with the other scoring methods.

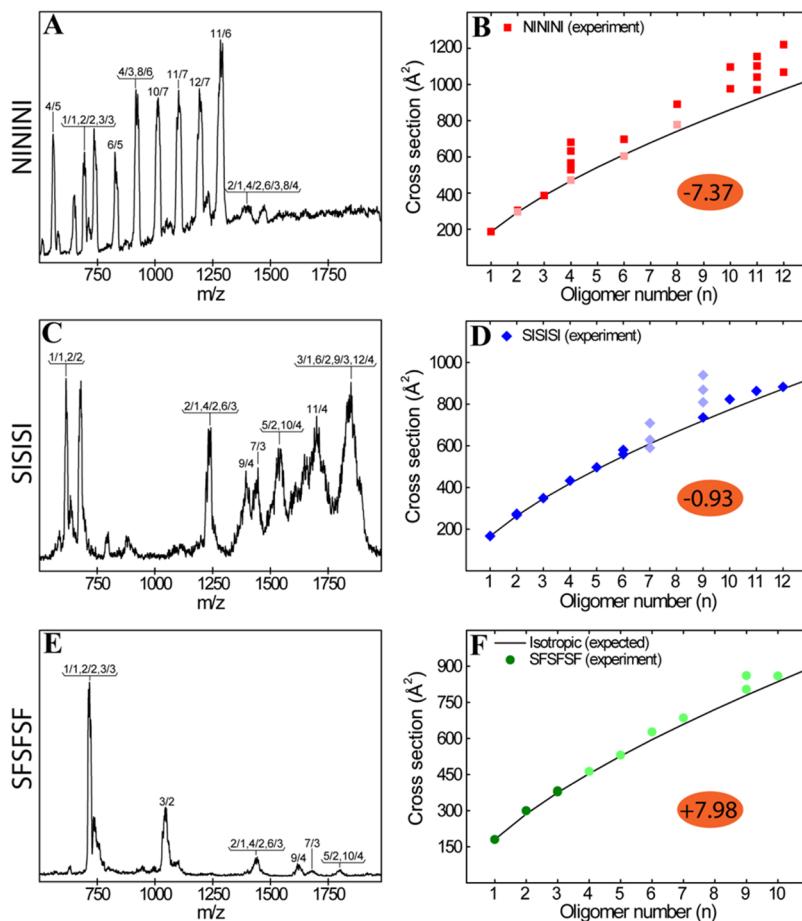
Our method ranks the order of amyloid aggregation propensities identically to that of the PASTA algorithm. However, our method predicts only NININI to have strong amyloid fibril formation propensity, whereas PASTA predicts that both NININI and SISISI should form  $\beta$ -sheet fibrils (i.e., a PASTA score lower than  $-4$  indicates very high amyloid propensity<sup>24–26</sup>). Our scoring method makes predictions for these peptides opposite to that of Zyggregator, which favors the phenylalanine-containing peptides over isoleucine-containing peptides. Also, according to Zyggregator, NININI and SISISI are indistinguishable, as are NFNFNF and SFSFSF. The 3D Profile method has some issues with scoring these peptides. The C-Score of NININI obtained from 3D Profile is zero, although it has an uncharacteristically high Rosetta Energy ( $-27.6$  kcal/mol) and shape complementary (0.8640), both predictive of amyloid aggregation. Furthermore, 3D Profile scores serine-containing peptides higher than asparagine-containing peptides, whereas this pattern is reversed in scores generated by our method. Of note, our scoring scheme is the only method attempting to distinguish fibril formation from microcrystalline formation.

To examine the experimental aggregation propensities of these model peptides, we turn to IM-MS experiments to monitor the early oligomer formation and TEM to image the final morphologies of the aggregates. NFNFNF was not soluble under the same experimental conditions employed for the other three peptides, and thus, we were unable to obtain experimental data on that peptide. The IM-MS data were obtained using a mass spectrometer having a short drift cell (5.0 cm) but capable of detecting low charged oligomers.<sup>77</sup> As the mass resolution of this mass spectrometer is low and that the fwhm of the mass spectral peaks can be broad, it could be a problem for assigning  $m/z$  values for large peptide or protein samples. However, here, we worked with small peptides; the  $m/z$  assignment is straightforward, and there is no ambiguity in the assignments. Figure 4 shows the nano-ESI-q mass spectra (left panels) of the three peptides (NININI, SISISI, and SFSFSF) and the plots of experimental oligomer cross section ( $\sigma$ , Å<sup>2</sup>) as a function of oligomer number ( $n$ ) (right panels). Similar to the cases of individual amino acids, each peptide forms multiple oligomers with different sizes and the same  $m/z$  ratio. The ATDs of each  $m/z$  peak were collected, and the cross section of each feature was determined (see Figures S5–S7). These values were compared with the isotropic cross sections to examine the structures of early oligomers (a positive deviation from the isotropic model is associated with  $\beta$ -rich structures<sup>55,65</sup>) and to predict whether the peptides would aggregate. The degree of deviation from isotropic growth is large for NININI and small for SFSFSF and SISISI. In NININI, the transition from isotropic to  $\beta$ -rich structures occurs at  $n = 4$ , which is consistent with the high amyloid aggregation propensity predicted by our method (Table 1). SISISI has isotropic oligomers up to  $n = 12$ . Deviations above the isotropic line are observed at  $n = 7$  and  $n = 9$ . However, the ATDs (see Figure S6) indicate that the large majority of SISISI clusters have cross sections very near the isotropic line and only a very minor percentage of structures deviate from the line as shown by the color coding of the data points in Figure 4D. For SFSFSF, the clusters grow in an isotropic manner up to the largest oligomer detected under our

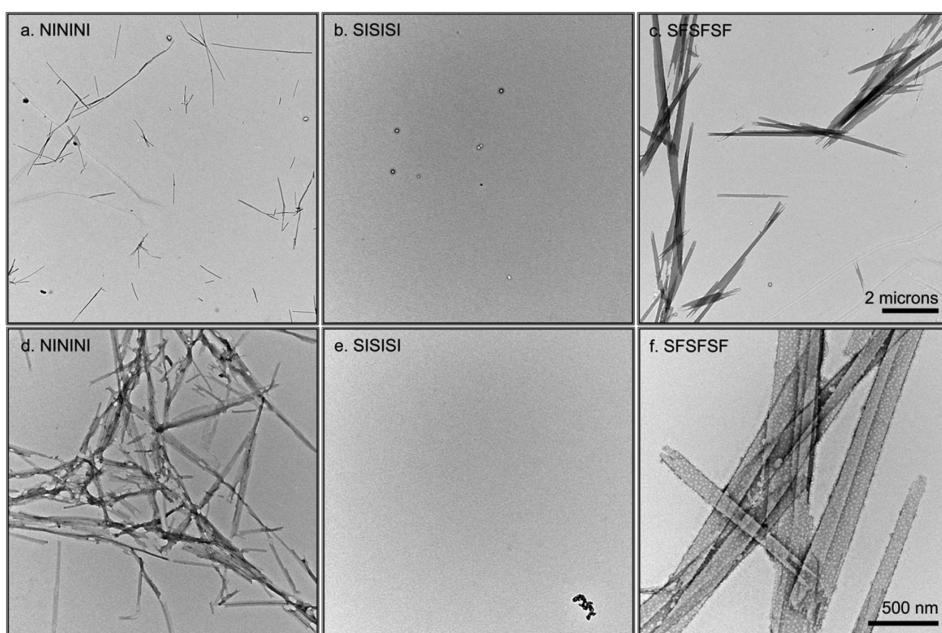
**Table 1. Amyloid Aggregation Predictions for Model Peptides According to the PASTA, 3D Profile, and Zyggregator Algorithms and Our Method<sup>a</sup>**

peptides	our method <sup>c</sup>	PASTA	3D Profile <sup>b</sup>	Zyggregator <sup>d</sup>
NININI	-7.37 (1)	-5.87 (1)	+0.00 (4)	+0.940 (3)
SISISI	-0.93 (2)	-4.58 (2)	-48.3 (1)	+0.940 (3)
NFNFNF	+1.54 (3)	-2.62 (3)	-37.9 (3)	+0.309 (1)
SFSFSF	+7.98 (4)	-2.35 (4)	-40.9 (2)	+0.309 (1)

<sup>a</sup>Number inside the parentheses is the rank for each peptide based on predicted aggregation propensity from the highest to the lowest. <sup>b</sup>For the 3D Profile method, the composite score (C-Score) of each peptide is reported. <sup>c</sup>See Supporting Information section 2.5 for how the scores were computed. <sup>d</sup>Zyggregator values have been multiplied by  $(-1)$  to give scores consistent in sign with the other methods in the table.



**Figure 4.** (left panels) Nano-ESI-q-mass spectra of the three hexapeptides. Each mass spectral peak is annotated by  $n/z$ , where  $n$  is the oligomer number and  $z$  the charge. (right panels) Plots of collision cross section ( $\sigma$ ) as a function of oligomer size ( $n$ ). The cross section data are colored based on intensities of the corresponding peaks with a lighter color indicating a less intense feature and a darker color indicating a more intense feature. The predicted isotropic growth curves are shown as black lines. The predicted scores from our method are also shown (circled values).



**Figure 5.** Fibrillar and microcrystalline aggregates. NININI at 1 week in 8% DMSO/water shows a mixture of fibers and microcrystals (panels a and d). SISI SI did not aggregate appreciably in 30% DMSO/water over the time frame of the experiment (panels b and e). Microcrystalline aggregates of SFSFSF after 24 h in 30% DMSO/water (panels c and f). The difference in DMSO concentrations was due to the solubilities of the peptides.

experimental condition at  $n = 10$ . Of note, none of the oligomer growth trends fall below the isotropic curves. The amino acid connectivity through peptide bonds limited the degree of freedom in interactions, leading to isotropic cross sections approximating the most compact structures that the oligomers could attain.

TEM experiments were performed to examine aggregate morphologies of the three hexapeptides (see Figure 5). NININI, which our method ranked highest, and hence most likely to form  $\beta$ -sheet fibrils, forms a mixture of mostly fibrillar aggregates and fewer microcrystalline aggregates (Figure 5a,d). No differences were observed between 24 h and 1 week incubations. The fibrils were fairly uniform in width ( $16.0 \pm 6.6$  nm), characteristic of  $\beta$ -sheet structure, but showed greater variation in length. SISI did not aggregate appreciably at incubation times up to 1 week (the longest time point examined), forming only an occasional amorphous aggregate (Figure 5b,e). In contrast, SFSFSF ranked lowest on our list in terms of amyloid tendency and had a strong positive predictive capacity for crystal formation. TEM results indicate that SFSFSF forms microcrystalline aggregates with forked, tapered ends (Figure 5c,f). SFSFSF microcrystals were roughly 10-fold wider ( $121.2 \pm 32.3$  nm) than NININI aggregates and twice as long. We observed no major differences between 24 h and 1 week incubations, although at the longer time point more microcrystal ends appeared blunted, possibly as a result of breakage (data not shown).

## SUMMARY AND CONCLUSIONS

In this work, we study the metacluster formation of several hydrophobic and hydrophilic amino acids (i.e., serine, asparagine, tryptophan, and isoleucine) and include prior work on phenylalanine.<sup>42</sup> The cross sections of Ser metaclusters are more compact than predicted by the isotropic growth curve. Although Asn is bulkier than serine, its growth trend also falls below that predicted by the isotropic model. For the hydrophobic amino acids, isoleucine is smaller in size than phenylalanine, but its trend deviates above the isotropic growth curve more significantly than Phe and is nearly linear at large cluster sizes. Trp is larger in size than Phe, but its growth trend follows the isotropic line. Hence, it is clear that side chain bulkiness has only a minor effect on metacluster growth, but hydropathy index has a significant effect on growth. Recognizing this dependence, we developed an empirical method for scoring peptide aggregation propensity based on amino acid metacluster growth trends. This method is bidirectional and has the potential to distinguish amyloid from crystalline formation. Our predictions on a set of hexapeptides are in good agreement with the experimental results and represent a significant improvement over current algorithms. The current data set is small, however, and future work will further assess the model by examining additional amino acid metaclusters and further testing the assembly tendencies of more peptides. In particular, the impact of primary sequence on assembly tendency needs to be addressed.

## EXPERIMENTAL METHODS

L-Serine, L-tryptophan, L-isoleucine, and L-asparagine were purchased from Sigma-Aldrich (St. Louis, MO). The IM-MS experiments on individual amino acids were performed on a home-built ion-mobility mass spectrometer with a nano-ESI source, an entrance funnel, a 2 m long drift cell, an exit funnel,

and a quadrupole detector.<sup>78</sup> The hexapeptides NININI, SISI, NFNFN and SFSFSF were purchased from GenScript Corporation (Piscataway, NJ). Ion-mobility experiments for the short peptides were performed on another home-built mass spectrometer with a shorter drift cell (~5 cm) that can detect low charge state mass spectral peaks.<sup>77</sup> Water was purchased from J. T. Baker. The concentration of amino acids in the spray sample is 6 mM, which is comparable to previous experiments (~10 mM).<sup>31,34,35,38,39</sup> The concentration of hexapeptides in the spray sample is 100  $\mu$ M in water at pH 8 (ammonium hydroxide was added to increase the pH to allow the peptide powders to be dissolved). In the IM-MS experiments, ions were generated through the means of nano-ESI and pulled through a helium-filled drift cell under the influence of a weak electrical field balanced by a drag force created from collisions with buffer gas. Metaclusters were mass-selected and their arrival time distributions measured at different pressure-to-drift voltage ratios, allowing the determination of reduced mobility  $K_0$  and experimental collision cross section  $\sigma$ .<sup>79,80</sup> Details are given in the Supporting Information in section S2. The model structures of the serine 24-mer were constructed based on the X-ray structures reported by Moggach et al.<sup>58</sup> using the Super Cell Builder feature available in the Avogadro software.<sup>81</sup> For TEM experiments, aliquots of the hexapeptide solutions in H<sub>2</sub>O/DMSO were adsorbed onto mesh Formvar/carbon copper grids and imaged on a JEOL 123 microscope with an ORCA camera and AMT Image Capture Software v. 5.24. The use of a small percentage of DMSO (8–30%) is required for peptide solubility and compatible with TEM imaging conditions. However, the presence of DMSO appeared to suppress the TIC signal in the ESI process, and hence, pure water was used as a solvent for IM-MS. In limited experiments, we did not find any differences in IM-MS results with or without DMSO added, only a reduction in signal when DMSO was added.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.analchem.5b03454](https://doi.org/10.1021/acs.analchem.5b03454).

Tables of cross sections, representative ATDs, and detailed protocols for obtaining theoretical scores of the hexapeptide models (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [bowers@chem.ucsb.edu](mailto:bowers@chem.ucsb.edu). Tel: +1-805-893-2673.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from National Science Foundation Grant CHE-1301032 (M.T.B.), instrumental support from Air Force Office of Scientific Research Grant FA9550-11-0113 (M.T.B.), and a grant from Santa Barbara Cottage Hospital and the University of California, Santa Barbara (N.E.L.). N.E.C.A. acknowledges Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for a postdoctoral fellowship (204613/2014-0). We also acknowledge the use of the NRI Microscopy Facility at UCSB.

## ■ REFERENCES

- (1) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. *Nat. Chem.* **2009**, *1*, 326–331.
- (2) Wyss-Coray, T. *Nat. Med.* **2006**, *12*, 1005–1015.
- (3) Karra, E.; Mercken, M.; De Strooper, B. *Nat. Rev. Drug Discovery* **2011**, *10*, 698–712.
- (4) Hardy, J.; Selkoe, D. J. *Science* **2002**, *297*, 353–356.
- (5) Buee, L.; Bussiere, T.; Buee-Scherrer, V.; Delacourte, A.; Hof, P. R. *Brain Res. Rev.* **2000**, *33*, 95–130.
- (6) Wolfe, M. S. *J. Biol. Chem.* **2009**, *284*, 6021–6025.
- (7) Lewis, J.; Dickson, D. W.; Lin, W. L.; Chisholm, L.; Corral, A.; Jones, G.; Yen, S. H.; Sahara, N.; Skipper, L.; Yager, D.; Eckman, C.; Hardy, J.; Hutton, M.; McGowan, E. *Science* **2001**, *293*, 1487–1491.
- (8) Goedert, M.; Spillantini, M. G. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2000**, *1502*, 110–121.
- (9) Maries, E.; Dass, B.; Collier, T. J.; Kordower, J. H.; Steele-Collier, K. *Nat. Rev. Neurosci.* **2003**, *4*, 727–738.
- (10) Recchia, A.; Debetto, P.; Negro, A.; Guidolin, D.; Skaper, S. D.; Giusti, P. *FASEB J.* **2004**, *18*, 617–626.
- (11) Scherzinger, E.; Lurz, R.; Turmaine, M.; Mangiarini, L.; Hollenbach, B.; Hasenbank, R.; Bates, G. P.; Davies, S. W.; Lehrach, H.; Wanker, E. E. *Cell* **1997**, *90*, 549–558.
- (12) Chen, S. M.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11884–11889.
- (13) Westermark, P.; Andersson, A.; Westermark, G. T. *Physiol. Rev.* **2011**, *91*, 795–826.
- (14) Mandrup-Poulsen, T. *Nat. Immunol.* **2010**, *11*, 881–883.
- (15) Tanaka, M.; Komi, Y. *Nat. Chem. Biol.* **2015**, *11*, 373–377.
- (16) Knowles, T. P.; Buehler, M. J. *Nat. Nanotechnol.* **2011**, *6*, 469–479.
- (17) Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. *Biochemistry* **2000**, *39*, 13748–13759.
- (18) Krone, M. G.; Hua, L.; Soto, P.; Zhou, R. H.; Berne, B. J.; Shea, J. E. *J. Am. Chem. Soc.* **2008**, *130*, 11066–11072.
- (19) Ganguly, P.; Do, T. D.; Larini, L.; LaPointe, N. E.; Sercel, A. J.; Shade, M. F.; Feinstein, S. C.; Bowers, M. T.; Shea, J. E. *J. Phys. Chem. B* **2015**, *119*, 4582–4593.
- (20) von Bergen, M.; Barghorn, S.; Li, L.; Marx, A.; Biernat, J.; Mandelkow, E. M.; Mandelkow, E. *J. Biol. Chem.* **2001**, *276*, 48165–48174.
- (21) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. O.; Riek, C.; Eisenberg, D. *Nature* **2007**, *447*, 453–457.
- (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. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 19285–19290.
- (23) Caflisch, A. *Curr. Opin. Chem. Biol.* **2006**, *10*, 437–444.
- (24) Walsh, I.; Seno, F.; Tosatto, S. C.; Trovato, A. *Nucleic Acids Res.* **2014**, *42*, W301–307.
- (25) Trovato, A.; Chiti, F.; Maritan, A.; Seno, F. *PLoS Comput. Biol.* **2006**, *2*, 1608–1618.
- (26) Trovato, A.; Seno, F.; Tosatto, S. C. E. *Protein Eng., Des. Sel.* **2007**, *20*, 521–523.
- (27) Pawar, A. P.; Dubay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J. Mol. Biol.* **2005**, *350*, 379–392.
- (28) Tartaglia, G. G.; Pawar, A. P.; Campioni, S.; Dobson, C. M.; Chiti, F.; Vendruscolo, M. *J. Mol. Biol.* **2008**, *380*, 425–436.
- (29) Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat. Biotechnol.* **2004**, *22*, 1302–1306.
- (30) Wyttenbach, T.; Bowers, M. T. *Top. Curr. Chem.* **2003**, *225*, 207–232.
- (31) Koch, K. J.; Gozzo, F. C.; Zhang, D. X.; Eberlin, M. N.; Cooks, R. G. *Chem. Commun.* **2001**, 1854–1855.
- (32) Nanita, S. C.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 554–569.
- (33) Yang, P. X.; Xu, R. F.; Nanita, S. C.; Cooks, R. G. *J. Am. Chem. Soc.* **2006**, *128*, 17074–17086.
- (34) Cooks, R. G.; Zhang, D. X.; Koch, K. J.; Gozzo, F. C.; Eberlin, M. N. *Anal. Chem.* **2001**, *73*, 3646–3655.
- (35) Takats, Z.; Nanita, S. C.; Schlosser, G.; Vekey, K.; Cooks, R. G. *Anal. Chem.* **2003**, *75*, 6147–6154.
- (36) Spencer, E. A. C.; Ly, T.; Julian, R. R. *Int. J. Mass Spectrom.* **2008**, *270*, 166–172.
- (37) Mazurek, U.; Geller, O.; Lifshitz, C.; McFarland, M. A.; Marshall, A. G.; Reuben, B. G. *J. Phys. Chem. A* **2005**, *109*, 2107–2112.
- (38) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2001**, *105*, 8092–8096.
- (39) Julian, R. R.; Hodyss, R.; Kinnear, B.; Jarrold, M. F.; Beauchamp, J. L. *J. Phys. Chem. B* **2002**, *106*, 1219–1228.
- (40) Adler-Abramovich, L.; Vaks, L.; Carny, O.; Trudler, D.; Magno, A.; Caflisch, A.; Frenkel, D.; Gazit, E. *Nat. Chem. Biol.* **2012**, *8*, 701–706.
- (41) Reches, M.; Gazit, E. *Science* **2003**, *300*, 625–627.
- (42) Do, T. D.; Kincannon, W. M.; Bowers, M. T. *J. Am. Chem. Soc.* **2015**, *137*, 10080–10083.
- (43) German, H. W.; Uyaver, S.; Hansmann, A. U. *J. Phys. Chem. A* **2015**, *119*, 1609–1615.
- (44) Julian, R. R.; Hodyss, R.; Beauchamp, J. L. *J. Am. Chem. Soc.* **2001**, *123*, 3577–3583.
- (45) Feketeova, L.; Khairallah, G. N.; Brunet, C.; Lemoine, J.; Antoine, R.; Dugourd, P.; O'Hair, R. A. J. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3255–3260.
- (46) Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **2004**, *108*, 4885–4898.
- (47) Shaham-Niv, S.; Adler-Abramovich, L.; Schnaider, L.; Gazit, E. *Sci. Adv.* **2015**, *1*, e1500137.
- (48) Hodyss, R.; Julian, R. R.; Beauchamp, J. L. *Chirality* **2001**, *13*, 703–706.
- (49) Thirumalai, D.; Reddy, G.; Straub, J. E. *Acc. Chem. Res.* **2012**, *45*, 83–92.
- (50) Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riek, C. *Acc. Chem. Res.* **2006**, *39*, 568–575.
- (51) Julian, R. R.; Myung, S.; Clemmer, D. E. *J. Phys. Chem. B* **2005**, *109*, 440–444.
- (52) Mishra, R.; Hoop, C. L.; Kodali, R.; Sahoo, B.; van der Wel, P. C. A.; Wetzel, R. *J. Mol. Biol.* **2012**, *424*, 1–14.
- (53) McFarland, N. R.; Fan, Z.; Xu, K.; Schwarzschild, M. A.; Feany, M. B.; Hyman, B. T.; McLean, P. J. *J. Neuropathol. Exp. Neurol.* **2009**, *68*, 515–524.
- (54) Gorbatyuk, O. S.; Li, S.; Sullivan, L. F.; Chen, W.; Kondrikova, G.; Manfredsson, F. P.; Mandel, R. J.; Muzyczka, N. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 763–768.
- (55) Bleiholder, C.; Dupuis, N. F.; Wyttenbach, T.; Bowers, M. T. *Nat. Chem.* **2011**, *3*, 172–177.
- (56) Benedetti, E.; Pedone, C.; Sirigu, A. *Gazzetta Chim. Ital.* **1973**, *103*, 555–561.
- (57) Kistenmacher, T. J.; Rand, G. A.; Marsh, R. E. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1974**, *30*, 2573–2578.
- (58) Moggach, S. A.; Allan, D. R.; Morrison, C. A.; Parsons, S.; Sawyer, L. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 58–68.
- (59) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082–16086.
- (60) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86–91.
- (61) Bleiholder, C.; Contreras, S.; Do, T. D.; Bowers, M. T. *Int. J. Mass Spectrom.* **2013**, *345*, 89–96.
- (62) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass Spectrom.* **2011**, *308*, 1–10.
- (63) Jenkins, J.; Pickersgill, R. *Prog. Biophys. Mol. Biol.* **2001**, *77*, 111–175.
- (64) Halfmann, R.; Alberti, S.; Krishnan, R.; Lyle, N.; O'Donnell, C. W.; King, O. D.; Berger, B.; Pappu, R. V.; Lindquist, S. *Mol. Cell* **2011**, *43*, 72–84.

- (65) Do, T. D.; Economou, N. J.; LaPointe, N. E.; Kincannon, W. M.; Bleiholder, C.; Feinstein, S. C.; Teplow, D. B.; Buratto, S. K.; Bowers, M. T. *J. Phys. Chem. B* **2013**, *117*, 8436–8446.
- (66) Do, T. D.; Lapointe, N. E.; Economou, N. J.; Buratto, S. K.; Feinstein, S. C.; Shea, J. E.; Bowers, M. T. *J. Phys. Chem. B* **2013**, *117*, 10759–10768.
- (67) Do, T. D.; LaPointe, N. E.; Sangwan, S.; Teplow, D. B.; Feinstein, S. C.; Sawaya, M. R.; Eisenberg, D. S.; Bowers, M. T. *J. Phys. Chem. B* **2014**, *118*, 7247–7256.
- (68) Marshall, K. E.; Morris, K. L.; Charlton, D.; O'Reilly, N.; Lewis, L.; Walden, H.; Serpell, L. C. *Biochemistry* **2011**, *50*, 2061–2071.
- (69) Azriel, R.; Gazit, E. *J. Biol. Chem.* **2001**, *276*, 34156–34161.
- (70) de Groot, N. S.; Parella, T.; Aviles, F. X.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732–1741.
- (71) Gorbitz, C. H. *Chem. - Eur. J.* **2001**, *7*, 5153–5159.
- (72) Do, T. D.; Bowers, M. T. *Anal. Chem.* **2015**, *87*, 4245–4252.
- (73) Kumar, C. V.; Swetha, R. G.; Ramaiah, S.; Anbarasu, A. J. *Biomol. Struct. Dyn.* **2015**, *33*, 1695–1709.
- (74) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105–132.
- (75) Thompson, M. J.; Sievers, S. A.; Karanikolas, J.; Ivanova, M. I.; Baker, D.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4074–4078.
- (76) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 3487–3492.
- (77) Wytttenbach, T.; Kemper, P. R.; Bowers, M. T. *Int. J. Mass Spectrom.* **2001**, *212*, 13–23.
- (78) Kemper, P. R.; Dupuis, N. F.; Bowers, M. T. *Int. J. Mass Spectrom.* **2009**, *287*, 46–57.
- (79) Mason, E. A. *Transport Properties of Ions in Gases*; 99th ed.; John Wiley & Sons, 1988.
- (80) Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 15132–15140.
- (81) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminf.* **2012**, *4*, 17.