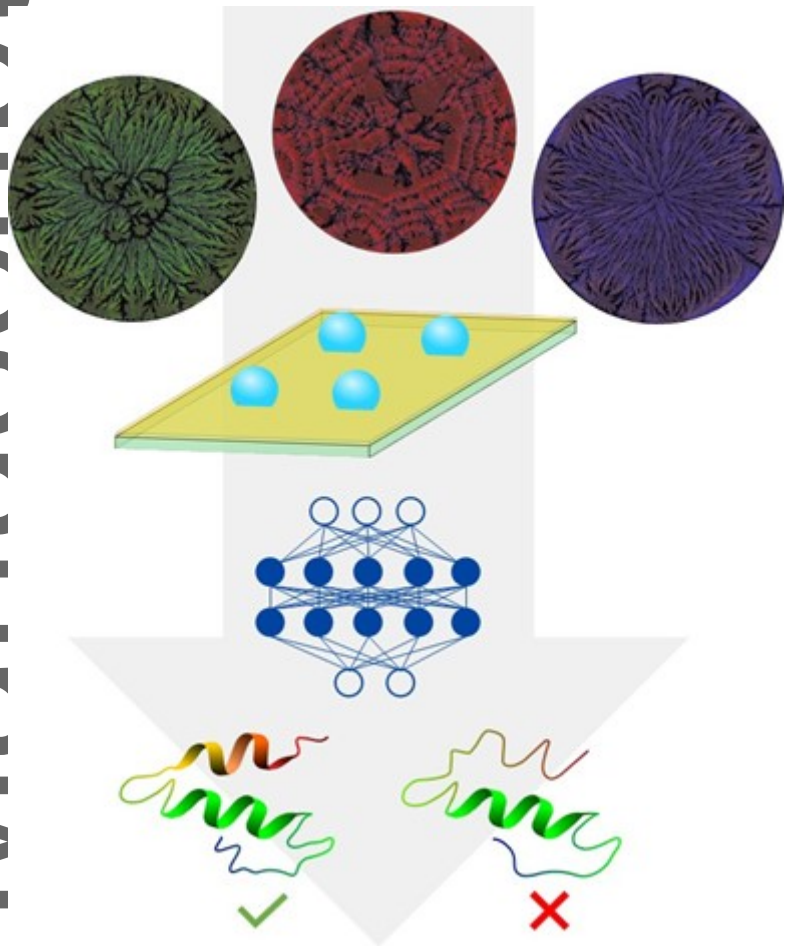
Author Manuscript



Deep learning assisted stratification of amyloid beta mutants using drying droplet patterns. A process as simple and scalable as allowing a peptide droplet to dry onto a surface can disclose critical structural mismatches of amyloid beta peptides. When analyzed with pretrained neuronal networks, the peptide deposits predict single amino acid mismatches and conformational misfolding with 99% accuracy.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/adma.202110404.

This article is protected by copyright. All rights reserved.

Author Manuscript

Deep learning assisted stratification of amyloid beta mutants using drying droplet patterns

Azam Jeihanipour[1], Jörg Lahann[1,2*]

[1] Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

[2] Biointerfaces Institute, Departments of Chemical Engineering, Materials Science and Engineering, and Biomedical Engineering, and the Macromolecular Science and Engineering Program, University of Michigan, Ann Arbor, MI 48109, USA

* *Corresponding author: Prof. J. Lahann; Email: lahann@umich.edu*

Abstract

This article is protected by copyright. All rights reserved.

The development of simple and accurate methods to predict mutations and misfolding in proteins remains one of the unsolved challenges in modern biochemistry. We have discovered that critical information about primary and secondary peptide structures can be inferred from the stains left behind by their drying droplets. To analyze the complex stain patterns, we challenged deep learning (DL) neuronal networks with polarized light micrograph (PLM) images derived from the drying droplet deposits of a range of amyloid beta (1-42) ($A\beta_{42}$) peptides. These peptides differed in a single amino acid residue and represented hereditary mutants of Alzheimer's Disease. Stain patterns were not only highly reproducible, but resulted in comprehensive stratification of eight amyloid beta (Aβ) variants with predictive accuracies above 99%. Similarly, peptide stains of a range of distinct $A\beta_{42}$ peptide conformations were identified with accuracies above 99%. Our results suggest that a method as simple as drying a droplet of a peptide solution onto a solid surface may serve as an indicator of minute, yet structurally meaningful differences in peptides' primary and secondary structures. Scalable and accurate detection schemes for stratification of conformational and structural protein alterations are critically needed to unravel pathological signatures in many human diseases such as Alzheimer's and Parkinson's disease.

Keywords:

amyloid beta; protein misfolding; self-assembly; deep learning; coffee ring

Author Manuscript

This article is protected by copyright. All rights reserved.

Introduction

Proteins draw biological functions from their hierarchical biochemical structure; abundant evidence exists that even minute structural or conformational alterations can promote pathogenesis [1]. For example, numerous neurodegenerative diseases have been traced to misfolding of peptides and proteins [2]. Among the best studied examples is the 42-residue amyloid beta ($A\beta_{42}$) peptide, which is implicated with Alzheimer's disease (AD) and hereditary cerebral hemorrhage with amyloidosis (HCHWA) [2-3]. In the case of hereditary AD, single amino acid mutations in the amyloid beta (Aβ) peptide sequence, such as A21G (Flemish) [4], E22K (Italian) [5], E22G (Arctic) [6], and E22Q (Dutch) [7], result in the formation of aggregation-prone and pathogenic variants [8]. Peptide aggregation is a precursor for the growth of insoluble fibrils and plaques that can cause synapse failure and memory weakening [9]. The stratification of minute alterations in a peptide's structure, such as single amino acid substitutions or more transient conformational changes, remains an active area of scientific research [10]. Amyloid beta misfolding has been studied by Raman spectroscopy [11], fluorescence-based assays [12], [13], electron microscopy (EM) [14],[15], nuclear magnetic resonance (NMR) [16], and circular dichroism (CD) spectroscopy [17], [18]. However, these methods require multimodal analytics (e.g., NMR) [19], the use of complex amplification schemes (e.g., fluorescence assays) [20], [21], or are predicated on the late-stage formation of fibrils (e.g., cryo-EM) [15, 22], an event that is typically associated with later stages of disease progression [23] (see table **SI.1**).

Due to their experimental simplicity, studies into the staining patterns of drying droplets have been pursued in various fields, such as material science [24], food technology [25], bioscience [26], and medical diagnostics [27]. Numerous investigations focusing on the physicochemical dynamics in droplet wetting and evaporation exist and underscore the importance of solute factors [28],

Author Manuscript

This article is protected by copyright. All rights reserved.

atmospheric conditions [28a, 29], and the substrate [28b, 30]. Even the most elementary case, the deposition patterns of simple aqueous droplets on a material's surface, the so-called "coffee-ring" stains, are governed by a range of physical processes including fluid convection, surface forces at both liquid-air and liquid-solid interfaces, pinning of contact lines [31], and Marangoni flows [32]. Additional solute components, such as the presence of a peptide or protein, characteristically alter the stains due to the multifaceted interplay between transport and dewetting effects and nucleation and growth processes [28d, 33]. Published stain patterns range from uniform films [34] to dendritic and soccer-ball patterns [33], as well as more complex arrangements [28d, 35]. In this paper, we leverage the power of pre-trained deep learning (DL) algorithms to dissect structural features of pathogenic Aβ peptides based on stains left behind by their respective drying droplets.

Results

To investigate the nature of the stains formed during the drying of peptide-containing droplets, we analyzed their deposition patterns using polymerized light microscopy (PLM). **Fig. 1**A shows a PLM image of a characteristic drying pattern of the $A\beta_{42}$ peptide. To prepare homogenous, hydrophobic substrates, a prerequisite for regular droplet deposition over large areas, chemical vapor deposition (CVD) polymerization [36] was employed to obtain poly(*p*-xylylene)-coated glass surfaces with an average water contact angle of 80 ± 1°. After 2 µl circular droplets were deposited and allowed to dry under humidity- and temperature-controlled conditions for 40 min, consistent and reproducible droplets stains with an average diameter of 2.0 ± 0.1 mm were observed. Throughout this study, we used a solution of 0.1 mg/ml peptide in an aqueous bicarbonate buffer (100 mM). Bicarbonate buffer was selected because it is a kosmotropic salt according to the Hofmeister series [37] and promotes salting out effects that generally favor protein-protein interactions [38].

Author Manuscript

This article is protected by copyright. All rights reserved.

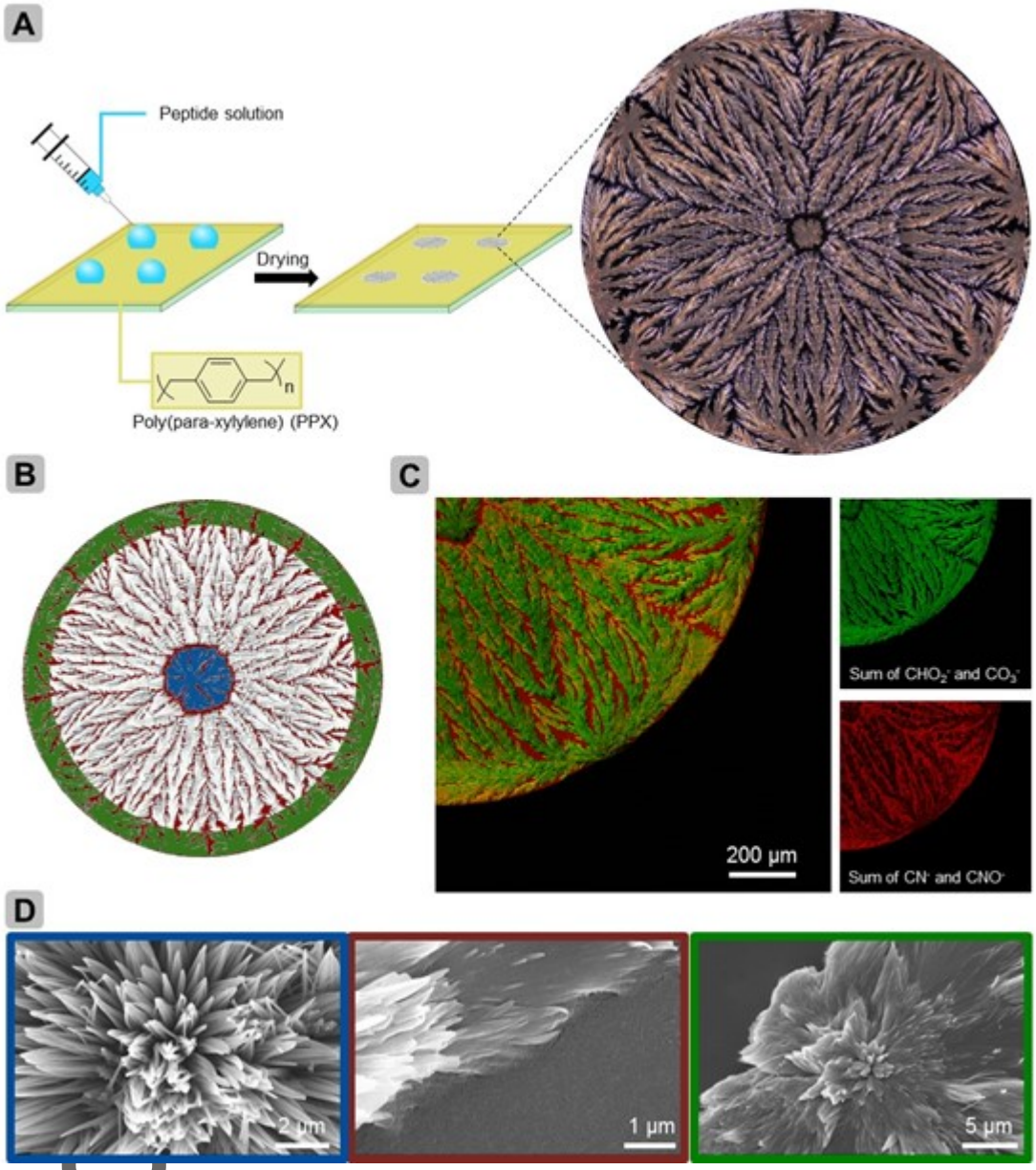


Fig. 1: Deposition stains of peptide-containing droplets are reproducible and reveal complex information about a peptide's structure. A) Peptide stains are obtained by depositing 2 µL droplets of an aqueous carbonate buffer solution onto hydrophobic poly(para-xylylene) (PPX) coated glass wafers. Drying of the droplets is performed under controlled humidity and temperature conditions. A representative polarized light micrograph (PLM) of the dried stain obtained from Aβ_{42} reveals complex deposition patterns. B) Microscopic analysis based on time-of-flight secondary ion mass spectrometry (ToF-SIMS) and (scanning electron microscopy) SEM identifies three distinct regions of interested marked by blue (center), red (a transition region), and green (rim). C) Chemical analysis of a dried Aβ_{42} peptide stain using ToF-SIMS imaging: salt crystals, marked in green, are identified by CHO_2^-, and CO_3^- fragment ions and the Aβ_{42} peptide, marked in red, is identified by CN^-, and CNO^- fragments. D) Analysis of microscopic features using SEM

This article is protected by copyright. All rights reserved.

supports the presence of three regions as proposed in B. Scale bars are 1, 2, and 5 µm.

Consequently, drying droplet stains, such as the one shown in **Fig. 1**A, constitute compositionally simple, yet structurally rich and complex supramolecular systems, governed by locally and temporally coupled multiscale processes [32c]. The drying patterns are the result of a complex interplay between several physico-chemical factors, such as interfacial properties, heat and mass transfer effects, fluid instabilities and heterogenous nucleation[32]. Therefore, controlling parameters, such as the substrate upon which the droplet is deposited or the atmosphere surrounding the droplet, directly influences the nature of deposited stain patterns. During the drying of the peptide solution, the salt can reach its saturation limit, and promotes peptide association. Once the peptide reaches saturation, it precipitates in a process that appears to occur concomitantly with the crystallization of the bicarbonate salt (Supplemental Video 1). These observations suggest that deposition patterns, obtained under controlled conditions, are not only complex and characteristic, but also highly reproducible.

To further elucidate chemical differences within the drying patterns, we employed time-of-flight secondary ion mass spectrometry (ToF-SIMS) (**Fig. 1**C). We attributed the sum of CHO_2^- and CO_3^- ions to bicarbonate crystals, whereas the sum of CN^- and CNO^- ions were used to identify the presence of peptide. The ToF-SIMS analysis of the top 5 nm of the deposited droplets revealed three characteristic regions: The core region was predominantly composed of salt, the rim region that coincided with the contact line of the original droplet and displayed convoluted signals of both, salt and peptide, and a transition region characterized by alternating areas of salt and peptide. The existence of the three zones was further confirmed by scanning electron microscopy (SEM) (Fig. 1B&D). The center is mainly characterized by crystalline structures corroborating the dominance of

Author Manuscript

This article is protected by copyright. All rights reserved.

salt, the rim region has a film-like morphology, whereas the transition region is characterized by dendritic growth structures indicating the coexistence of separate salt and peptide domains. Consequently, the deposition patterns may thus be best characterized as heterogenous co-deposits of a biological component ($A\beta_{42}$ peptide) and salt ($NaHCO_3$), where the protein appeared to template the salt. Under defined experimental conditions, the deposition patterns were highly conserved and reproducible (Fig. SI.2). These complex stain patterns displayed a high level of information content as indicated by an average Shannon's entropy [39] of 5.14 ± 0.13, which approaches the range of 5.55 to 7.17 obtained as the average Shannon's entropy of the most famous paintings of mankind [40] (Table SI.4). However, discerning systematic differences between PLM images with the naked eye has been very challenging due to their similar appearance (Fig. 2A). We thus hypothesized that the stain patterns could be exploited for categorizing structural elements of peptides (Fig. 2B). To test this hypothesis, we deposited a library of drying droplet stains derived from eight structurally similar variants of the $A\beta_{42}$ peptide and recorded approximately 400 PLM images per peptide. For pattern analysis, we selected a deep learning (DL) approach that builds upon previous machine learning studies [25b, 41], but minimizes extensive image processing and cumbersome feature extraction.

Author Manuscript

This article is protected by copyright. All rights reserved.

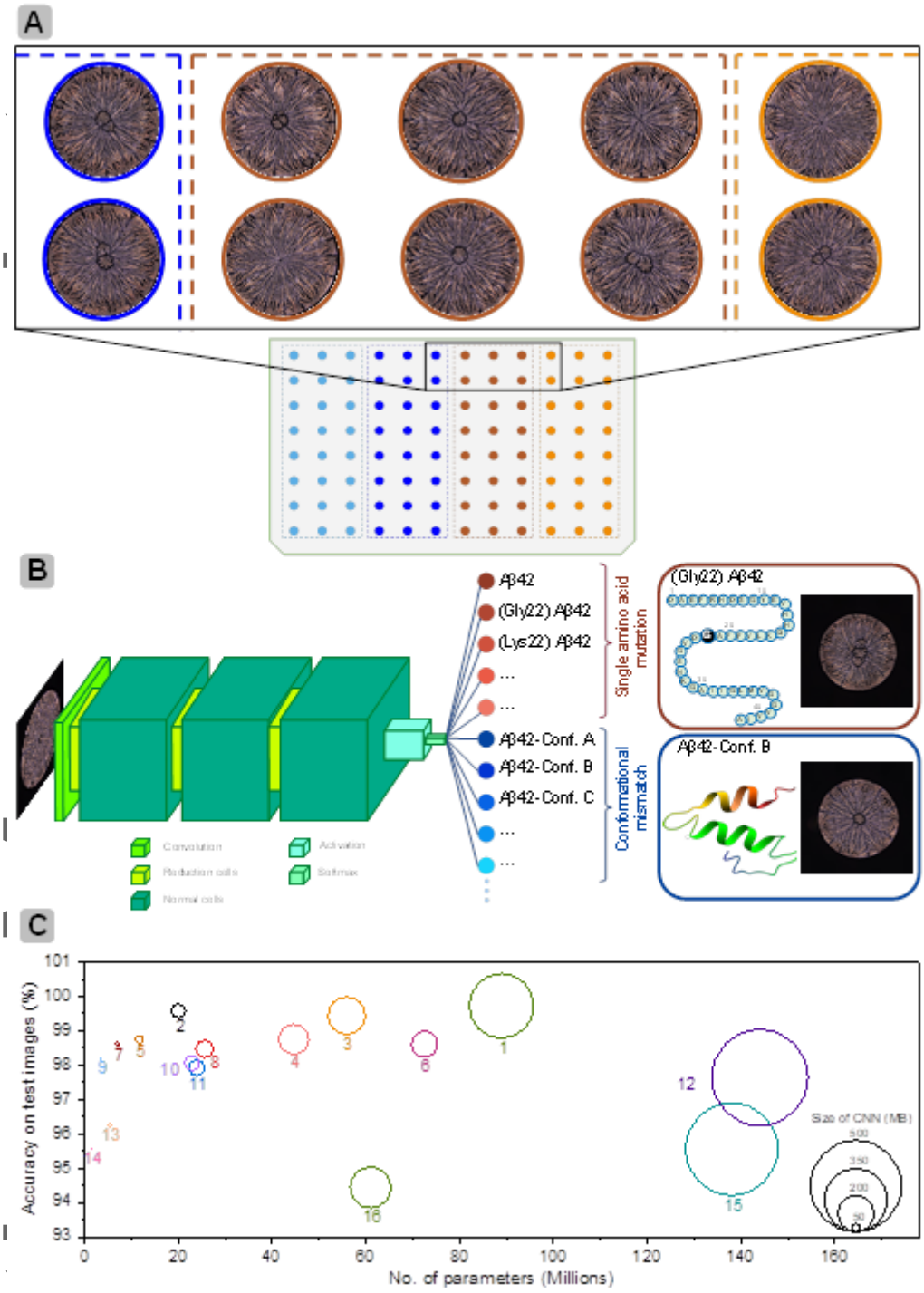


Fig. 2: Stratification of amyloid beta mutants based on image analysis of their deposition stains using a deep learning (DL) approach. A) Examples of 10 representative PLM images of peptide stains from three different peptides that were deposited onto a CVD-coated glass wafer to demonstrate the pattern variability within (6 stains in the center: A21G) and between groups of peptide mutants (outside

This article is protected by copyright. All rights reserved.

left: E22Q; outside right: wtA42beta). B) Using an established transfer learning approach, a pretrained deep convolutional neuron networks (CNN) is first trained with a medium sized number of PLM images of peptide stains with distinct primary and secondary structures. Because of the simplicity of the deposition process, about 200 images can be obtained within 60 min. In a second step, the CNN is challenged with never-seen images in order to assess its ability to categorize amyloid beta variants with single protein mutations and conformational misfolds. C) Performance of various deep learning networks with respect to network accuracy, speed, and size, when trained and challenged with PLM images of 8 different $A\beta_{42}$ with different mutations. Networks included in this study are (1) NASNet-Large, (2) DenseNet-201, (3) Inception-ResNet-v2, (4) ResNet-101, (5) ResNet-18, (6) DarkNet-53, (7) GoogleNet, (8) ResNet-50, (9) MobileNet-v2, (10) Xception, (11) Inception-v3, (12) Vgg-19, (13) NASNet-Mobile, (14) ShuffleNet, (15) VGG-16, and (16) AlexNet. The NasNet-Large, a medium-to-small sized DL network had the highest accuracy and is hence used in this study.

Transfer learning approaches using commercially available DL neural networks [42] [43] can reduce the burden of creating overly large training datasets and should thus be well-suited for the analysis of the stain images obtained in this study [44]. To identify a manageable DL network that offered high accuracy combined with a sufficiently small network size (**Fig. 2**C), we compared 16 different algorithms [42e, 45], some of which had been previously employed in medical image analysis [42a, 46] and biological research [47]. As shown in Fig. 2B, NasNet-Large, a medium-to-small sized DL network has the highest accuracy and was thus used in all subsequent experiments.

In **Fig. 3**, the NasNet-Large was employed for stratifying eight $A\beta_{42}$ peptides that differed by a single amino acid residue. It is well known that subtle changes in the primary structure of a peptide can affect a peptide's secondary structure and alter its physical properties [8a, 48]. The positions in which the amino acid residues were exchanged were thus distributed throughout the primary peptide structure with one being a part of the disordered N-terminal region (Ala^{2}), two in the central region (Ala^{21} and Glu^{22}) that has been critically implicated with self-aggregation and neuropathogenesis of $A\beta_{42}$ peptides [49], and one in a β-sheet region closer to the C-terminus (Gly^{37}, **Fig. 3**A). Four of the eight peptides were mutants of Ala^{21} or Glu^{22} because they are located in a critical transition region between a β-sheet (Gln^{15}-Ala^{21}) and a short turn (Glu^{22}-Asp^{23}) that continues into a second β-sheet

This article is protected by copyright. All rights reserved.

Author Manuscript

[48]. Specifically, the Ala^{21} residue was replaced with a glycine residue (A21G), while the Glu^{22} residue of the E22K, E22Q, and E22G peptide was replaced with amino acid residues characterized by a higher (Lys^{22}), identical (Gln^{22}) and lower (Gly^{22}) hydropathy index [50]. In addition, the Ala^{2} residue was exchanged with either the more hydrophilic tyrosine (hydropathy index: -3.5, A2T), or the more hydrophobic valine residue (hydropathy index: 4.2, A2V), and the Gly^{37} residue was exchanged with asparagine (Asp^{37}, G37D). Except for Val^{2}, the exchanged amino acid residues had a higher tendency to be buried inside of the peptide core compared to the wild type peptide. Accordingly, the A2V variant was characterized by a lower number of buried side chains, whereas the buried side chains increased in the A2T variant. In the case of E22K, this effect is further magnified by Lys's ability to engage in intramolecular hydrogen bonding. Steric effects can also impact the physical properties of the peptides [51]. While the molar volumes of the exchanged amino acid residues were generally increased relative to the wild type peptide, the Ala^{21} and Glu^{22} residues had significantly larger molar volumes. In general, the hydropathy indices of the exchanged amino acid residues were comparable. An exception was the exchange of Glu^{22} with Gly^{22}, which lowered the overall hydropathy of the E22G variant relative to the wild type peptide. Several of these peptide structures corresponded to familial mutations of the Aβ_{42} peptide including the Flemish (A21G) [4], Arctic (E22G) [6], Dutch (E22Q) [7], and Italian (E22K) [5] variants. In addition, an Aβ_{42} peptide structure was included where the Gly^{37} residue was replaced with asparagine (G37D). In **Fig. 3**A characteristic PLM images of deposition patterns obtained for the eight peptide variants are presented. Initial attempts to categorize the PLM images by the naked eye resulted in low accuracies, which can be attributed to similar physicochemical properties of the Aβ variants including molecular weight, isoelectric point and average hydrophilicity (**Table SI.2**). In contrast, the DL network NasNet-Large was able to stratify the PLM images with an accuracy of 99% across the eight Aβ variants as summarized in the confusion

Author Manuscript

This article is protected by copyright. All rights reserved.

matrix shown in **Fig. 3**B. This analysis was based on a training set of about 3200 images (i.e., about 400 images per peptide variant) and a test set of 720 randomized images of the eight different peptides; all images were new to the DL network, i.e., not included in the original training cohort.

To identify features of discriminatory regions of the PLM images, we used gradient-weighted class activation mapping (Grad-CAM) [52] to generate activation maps of the final convolutional layer of the DL network [52-53]. Heat map layers of the PLM images are presented in Fig. SI3 and highlight a particular focus of the DL network on both, the center and the transition regions within the stain patterns **(Fig. 1** B, C, and D). Further insights about the clustering of the peptide stains can be derived by applying a t-distributed stochastic neighbor embedding (t-SNE) [54] algorithm. After non-linear dimensionality reduction of NasNet-Large's "depth concatenation" layer, pronounced clustering of the test set of PLM images corresponding to eight different amyloid-beta variants was observed providing further support that the stain patterns are highly reproducible and characteristic of the various amyloid-beta peptides.

Author Manuscript

This article is protected by copyright. All rights reserved.

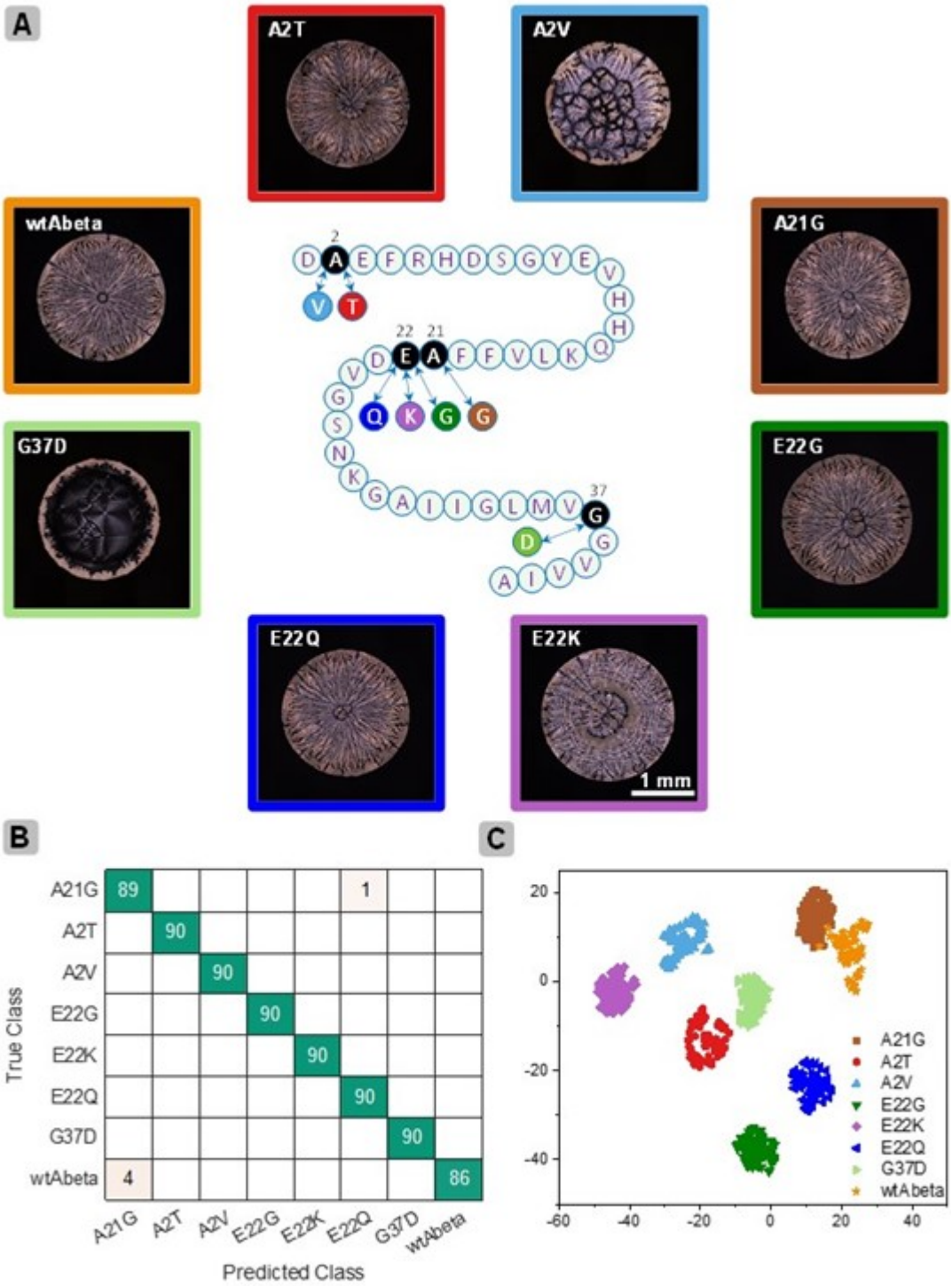


Fig. 3: Peptide stains accurately predict single amino acid mismatches in the $A\beta_{42}$ structure. A) Characteristic PLM images of deposition patterns of wild type $A\beta_{42}$ peptide and seven distinct variants with chemical structures that varied by a single point mutation. B) Peptide stains were analyzed using NasNet-

This article is protected by copyright. All rights reserved.

Large, which uses the ImageNet [55] dataset (1.28 million images over 1,000 generic object classes) for pretraining. The training dataset contained 3200 PLM images derived from the eight peptides shown in A. The test dataset contained 720 never-seen images. The trained NasNet-large network classified all peptides according to the multiclass confusion matrix with a total top-1 accuracy above 99%. C) The t-SNE plot of the "depth concatenation" layer of the trained CNN indicates excellent clustering of all peptides in spite of the fact that their structures varied only by a single amino acid mutation.

In addition to single amino acid mutations, conformational alterations are known to play a critical role in Aβ aggregation and subsequent fibril formation, but are particularly difficult to predict [18]. In its native form, the $A\beta_{42}$ peptide is comprised of two helical regions encompassing residues 8-25 and 28-38 that are connected by a regular type I beta-turn [56]. However, misfolded Aβ peptide typically has a cross-β structure in which the individual β-strands are oriented perpendicular to the fibril axis. To model controlled alterations in the secondary structure of Aβ peptides [3a, 57], $A\beta_{42}$ peptide was incubated with different concentrations of an aqueous hexafluoroisopropanol (HFIP) solution (**Fig. 4**). The resulting conformational alterations were quantitatively assessed by circular dichroism (CD) spectroscopy (**Fig. 4**A, B). These data confirmed predictive changes in the ratio of alpha-helix and beta-sheet content with increasing HFIP concentration (**Fig. 4**A) that were sufficiently long-lasting to be preserved even after removal of the HFIP and transfer into the bicarbonate buffer prior to CD analysis and droplet deposition. As shown in **Fig. 4**B, exposure of the $A\beta_{42}$ peptide to 0.1% v/v HFIP lowered the helix content by 2%. Further increase of the HFIP concentration corresponded to a continuous reduction in helical content until a plateau of about 3% in helix content was reached at HFIP concentrations above 40% v/v. **Fig. 4**C compares the PLM images of the droplet stains from eight $A\beta_{42}$ peptide conformations sampling the conformational range defined in **Fig. 4**B. Conformations with similar secondary structures, such as Hc6, Hc3a, and Hc3b, gave raise to stain patterns that appeared undistinguishable to the human eye. In contrast, Hc9, a conformer with 9% helical content, gave rise to a stain pattern that was very distinct from that of Hc6, which had a 3%

Author Manuscript

This article is protected by copyright. All rights reserved.

lower helical content. Using the NasNet-Large neural network with a training set of about 400 images per class, conformational alterations relative to the native Aβ conformer were stratified with a 99% accuracy (**Fig. 4**D). Within a test cohort of 720 randomized PLM images that had never been presented to the DL network before, NasNet-Large reported six misclassifications. It is helpful to understand these misclassifications in further details: Four misclassifications came from the alpha helix content below 3%, i.e., Hc6, Hc3a, and Hc3b. The remaining two misclassifications were Hc27 and Hc10, both misclassified as Hc7. For samples with more significant differences in their alpha helix content, stratification of peptide conformers was 100% accurate. This is further underscored by the t-SNE plot shown in **Figure 4E** that unambiguously confirms the utility of drying droplet stains for stratification of even minute alterations in peptides' secondary structures. While the data associated with the different peptide conformations generally clustered tightly together, different degrees of spreading was observed for the different peptides, potentially indicating different levels of polymorphism.

Author Manuscript

This article is protected by copyright. All rights reserved.

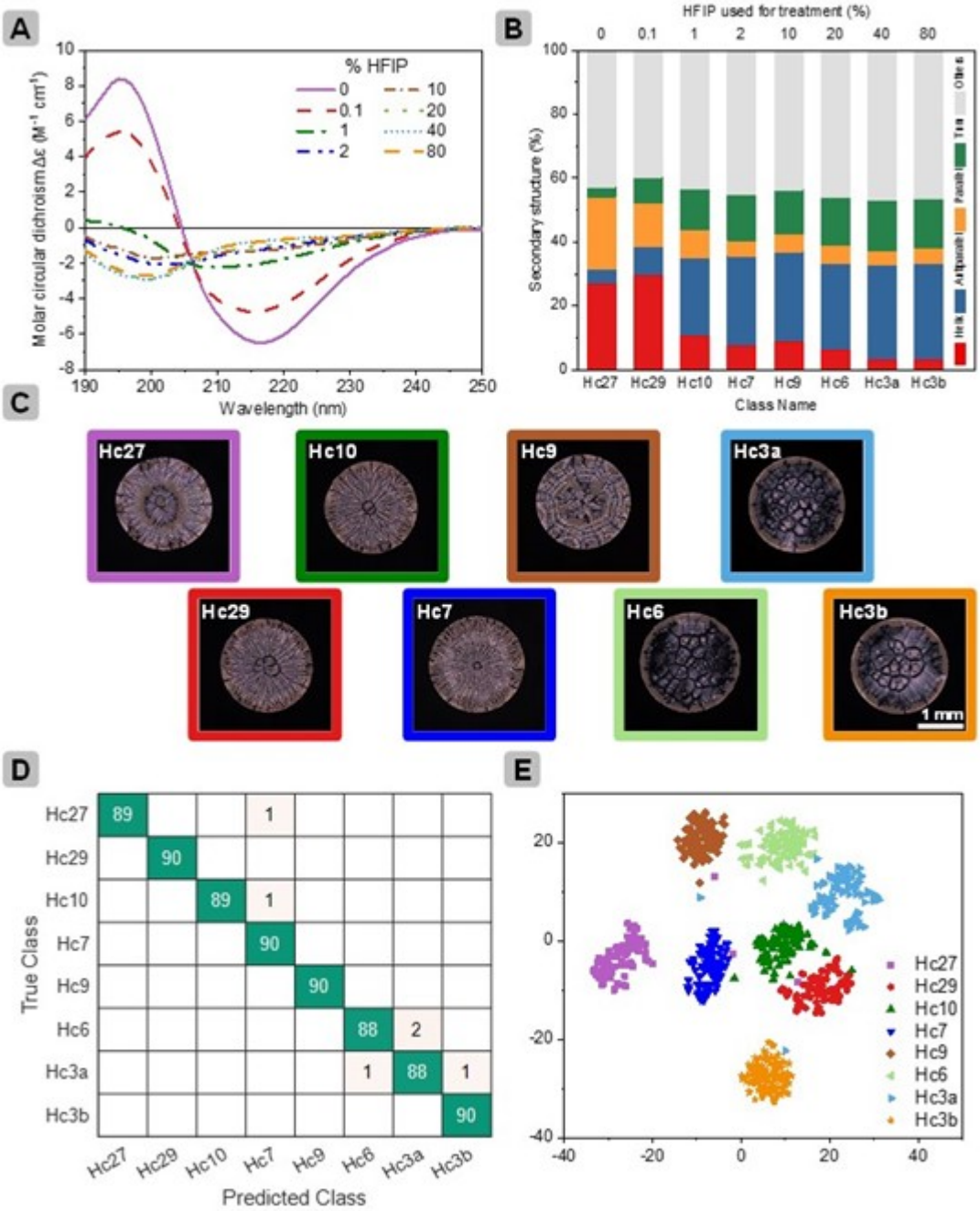


Fig. 4: Peptide stains can accurately predict Aβ$_{42}$ peptide misfolding. A) Far-UV CD spectra reveal reproduceable conformational changes of the Aβ$_{42}$ peptide. All circular dichroism (CD) measurement were conducted in a carbonate buffer under identical conditions used for the droplet deposition. To induce conformational changes in the Aβ$_{42}$ peptide, the peptide solutions were exposed to 0-80 vol-%

This article is protected by copyright. All rights reserved.

hexafluoroisopropanol (HFIP)[57]. After 15 min, the conformational variants were lyophilized to remove the HFIP and reconstituted in a carbonate buffer. **B)** Analysis of the content of helix, β-strand (including parallel and antiparallel), turn, and irregular (other) structures for all conformational variants of the $A\beta_{42}$ peptide based on their respective CD spectra that were analyzed using the BeStSel algorithm [58]. **C)** PLM images of the corresponding peptide stains labeled based on the degree of their helical content. For example, the structure of Hc27 has a helicity of 27%. **D)** Confusion matrix of the test set of 8 different secondary structures of $A\beta_{42}$ obtained from 720 PLM images as classified by the trained NasNet-Large (2880 training images). The total accuracy is above 99%. **E)** t-SNE graph of one of the last hidden layers of NasNet-Large, namely "depth concatenation layer", when classifying the PLM images test sets (720 images) of 8 different $A\beta_{42}$ conformation indicating accurate clustering of all misfolded peptide variants.

Conclusions

Simple protein stains deposited on a material's surface can be complex, characteristic, and reproducible. In this paper, we demonstrate that stain patterns of amyloid beta peptides serve as accurate fingerprints revealing a peptide's structural and conformational identity. In our hands, well-established, off-the-shelf DL neuronal networks trained on a limited number of stain images stratified structural mutations within $A\beta_{42}$ variants with exceptional accuracies (> 99%) within minutes. Due to their unprecedented simplicity, deposition patterns hold ample potential as indicators of minute differences in primary and secondary proteins structures, and will likely be broadly applicable to questions of protein aggregation and interactions, far beyond amyloid beta peptides. Our findings thus establish a major scientific advance with respect to both, fundamental research into the role of protein misfolding as well as the quest for simpler and potentially earlier diagnostic tools for neurodegenerative diseases.

Author Manuscript

This article is protected by copyright. All rights reserved.

Methods

Amyloid beta peptide solutions

Wild type $A\beta_{42}$ peptide and all peptide variants were purchased from Bachem (Bubendorf, Switzerland, Table S1). To prepare the Aβ peptide solutions, the lyophilized peptide was dissolved in 100 mM carbonate buffer to a concentration of 0.1 mg/ml. The buffer had a pH of 9.2 and was prepared using ultrapure water from a Milli-Q Plus system (Millipore, Schwalbach, Germany) and contained 91 mM $NaHCO_3$ (Merck Chemicals GmbH) and 9 mM Na_2CO_3 (Merck Chemicals GmbH). The peptide solution was mixed using an SB3 tube rotator (Stuart, Stone, UK) at 40 rpm for 15 min at room temperature, and then stored in aliquots at -20 °C. The $A\beta_{42}$ peptide stock solution (Bachem, Bubendorf, Switzerland) was prepared by dissolving 2 mg/ml lyophilized peptide in 100 mM carbonate buffer. After mixing for 15 min at 40 rpm at room temperature, the peptide solution was diluted with hexafluoroisopropanol (HFIP) (Sigma Aldrich) and MQ water to the desired HFIP concentrations with a final carbonate buffer concentration of 20 mM and peptide concentration of 0.2 mg/ml. The selected concentrations of HFIP were 0, 0.1, 1, 2, 5, 10, 20, 40, 60, and 80% (v/v).

Surface preparation via chemical vapor deposition polymerization

Glass wafers with the specification of extra white float, clear, and uncoated were custom made with a dimension of 120 mm × 80 mm and thickness of 1.0 ± 0.05 mm (Optrovision, München, Germany). Before the coating process, the glass wafers were washed first with Piranha solution (4 min floating in Piranha bath) and then several times with extra MQ water until a pH of 6.0 was reached. The clean glass plates were coated with poly(p-xylylene) (PPX) via CVD polymerization following a previously reported procedure [36]. The starting material [2.2]paracyclophane (Curtiss-Wright Surface Technologies (Galway, Ireland)) was sublimed under vacuum and converted by pyrolysis into the

Author Manuscript

This article is protected by copyright. All rights reserved.

quinomethane, which spontaneously polymerized upon condensation to the glass surface. A constant argon flow of 20 sccm was used as the carrier gas. The sublimation temperature was 100-110 °C followed by pyrolysis at 650 °C. The coating pressure was 0.5 mbar.

Droplet Deposition

All peptide solutions had a peptide concentration of 0.1 mg/ml in 100 mM carbonate buffer. An automated 96-well microplate pipetting system (epMotion 5070, Eppendorf AG, Hamburg, Germany), equipped with a 1-channel-dispenser (TS10, Eppendorf AG, Hamburg, Germany) was used for controlled deposition of defined arrays of droplets onto the glass slide. To control the environmental conditions, the pipetting system was placed inside a climate chamber (ICH 750, Memmert GmbH + Co. KG, Schwabach, Germany) and the environmental conditions were controlled at a temperature of 23°C± 0.5°C and humidity of 40% ± 3%. Each droplet was deposited with a dispensing speed of 3 mm/s and had a volume of 2 µl. The pipetting system was programmed to dispense 96 droplets per glass plate, in the form of 12 columns by 8 rows. In each experiment, 8 different solutions were dispensed in random arrangement, where 16 droplets of each solution were placed on the same slide (Figure SI.1). After the droplets were allowed to dry for 40 min ± 5 min, images of the deposition patterns were captured using a polarized light microscope (Olympus Polarizing Optical Microscope BX-53F, Tokyo, Japan) equipped with an automated stage. PLM images were selected over light micrographs, because the machine learning algorithms categorized them with higher accuracy. All the images were acquired at a consistent light intensity, using a 10x objective and stitched together using the Multi Image Alignment (MIA) algorithm included in

Author Manuscript

This article is protected by copyright. All rights reserved.

CellSens software (Olympus, Tokyo, Japan), with a 15% overlap. The acquired image of each dried droplet had a size of 2344×1878 pixels in the format of JPG.

Training and testing of the CNN

All raw images acquired from polarized light microscopy were imported into the MATLAB (Release 2020a, Math Works Inc.) for further processing and training the CNN. The images were first cropped to a size of 1878×1878 pixels by cutting 233 pixels from the left and 233 pixels from the right side of the image and then resized to the proper size compatible with the input layer of the CNN. For deep learning, several pre-trained CNN networks were used for training and testing purposes to evaluate the performance of different CNNs and selecting the most accurate one for our purpose in this paper. The one that we selected for further examination was NasNet-Large developed by google [59]. The NasNet-Large network structure contains two main modules, i.e. normal cells and reduction cells, in total 1244 layers, being part of automated machine learning. The NasNet-Large has been pre-trained to 96.2% top-5 accuracy on the 1,000 object classes (1.28 million images) of the ImageNet [55, 59]. Therefore, it has learned rich feature representations for a wide range of images. The network has an image input size of 331×331 pixels.

Following a transfer learning approach, [44] a network that was pre-trained with a large set of image features, was fine-tuned with a relative small set of new images. During transfer learning, the final classification layer was removed from the network and retrained with the dataset. Fine-tuning of the parameters occurred across all layers using the same global learning rate of 0.001, a minimum batch size of 32 images, and maximum epochs of 20. To prevent the network from overfitting and memorizing the exact details of the training set, the images were augmented using a random reflection function, in which each image is reflected horizontally and vertically with a 50%

Author Manuscript

This article is protected by copyright. All rights reserved.

probability. About 360 images per class were used for training, about 40 images per class for validation during training, and 90 images per class for testing the network after the training. There was no overlap between training, validation, and test sets.

The t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm [54], a method for visualizing high-dimensional data, was applied to one of the last hidden layers of the trained CNN ,namely the "depth concatenation layer", to show how well the network clusters the peptides. We used a MATLAB machine learning package to perform the t-SNE with the perplexity of 30 and the learning rate of 500 [54].

The visualization algorithm gradient-weighted class activation mapping (Grad-CAM) [52] was used to understand which regions of the image have the maximum influence on the classification decisions of the DL network.

Mass Spectrometry

Time-of-flight secondary ion mass spectrometry (ToF-SIMS) was performed using a ToF-SIMS instrument (ION-TOF GmbH, Münster, Germany) equipped with a Bi cluster liquid metal primary ion source and a non-linear time-of-flight analyzer. For spectrometry, short primary ion pulses (<1 ns) of the Bi source was operated in the "bunched" mode providing Bi_1^+ or Bi_3^+ ion pulses at 25 keV energy and a lateral resolution of approximately 4 µm. Since the droplets were larger than the maximum deflection range of the primary ion gun of 500×500 μm^2, the images were obtained using the manipulator stage scan mode. Negative polarity spectra were calibrated on the C^-, CH^-, CH^{2-} peaks. Spectrometry was performed in static SIMS mode by limiting the primary ion dose to $<10^{11}$ ions/cm^2. High lateral resolution images were acquired in a primary ion source mode providing a lateral resolution of about 200 nm with nominal mass resolution in "burst alignment" mode. Charge

This article is protected by copyright. All rights reserved.

Author Manuscript

compensation was necessary because of the glass substrate so that an electron flood gun providing electrons of 21 eV was applied and the secondary ion reflectron tuned accordingly.

Scanning Electron Microscopy

The morphology of the deposited peptide and salt in dried droplets were analyzed using scanning electron microscopy (SEM) (LEO 1530 Gemini, Zeiss, Germany). A thin layer of gold was sputtered onto the samples prior to SEM imaging to minimize surface charging. All SEM images were measured at an electron accelerating voltage of 10 kV and a working distance of 2.4 mm.

Circular Dichroism (CD) spectroscopy

The far-UV CD spectra of the peptide solutions were recorded using a J-810 spectropolarimeter (JASCO, Groß-Umstadt, Germany) at 20°C, in quartz glass cuvettes of 500 µm optical path length (Suprasil, Hellma Optik GmbH, Jena, Germany) between 260 and 180 nm, at 0.5 nm intervals. Three repeat scans at a scan rate of 10 nm/min, 8 s response time, and 1 nm bandwidth were averaged for each sample and its respective baseline of the protein-free sample. The peptide concentration was 0.15 mg/ml in 20 mM carbonate buffer (pH 9.2). To predict the content of secondary structures of the peptides and calculate the percentage of each specific conformation, the CD spectra were analyzed using the BeStSel webserver [58].

Acknowledgments

This research was supported by Virtual Materials Design (VirtMat) funding at Karlsruhe Institute of Technology. We thank Dr. Alexander Welle (Institute of Functional Interfaces (IFG), Karlsruhe

Author Manuscript

This article is protected by copyright. All rights reserved.

Institute of Technology (KIT)) for assistance with TOF-SIMS analyses, Dr. Divya Varadharajan (Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology (KIT)) for assistance with SEM imaging, and Dr. Joachim Bürck (Chair of Biochemistry (IOC) and Department of Molecular Biophysics (IBG-2), Karlsruhe Institute of Technology (KIT)) for assistance with CD analyses. We thank Alexander Bruckmann, Gilberth Andrés González Castellón, and Daniel Barón for assistance with data acquisition. Prof. Ronald Larson, University of Michigan, and Prof. Matthias Franzreb, Karlsruhe Institute of Technology, have had valuable input in form of scientific discussions and feedback throughout the studies.

References

[1] J. H. C. Smetana, G. Misra, in *Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics*, (Ed: G. Misra), Springer Singapore, Singapore 2017.

[2] F. Chiti, C. M. Dobson, *Annual Review of Biochemistry* **2017**, 86, 27.

[3] a)M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. I. Elliott, W. E. Van Nostrand, S. O. Smith, *Nature Structural & Molecular Biology* **2010**, 17, 561; b)R. Riek, D. S. Eisenberg, *Nature* **2016**, 539, 227.

[4] L. Hendriks, C. M. van Duijn, P. Cras, M. Cruts, W. Van Hul, F. van Harskamp, A. Warren, M. G. McInnis, S. E. Antonarakis, J.-J. Martin, A. Hofman, C. Van Broeckhoven, *Nature Genetics* **1992**, 1, 218.

[5] G. Rossi, G. Macchi, M. Porro, G. Giaccone, M. Bugiani, E. Scarpini, G. Scarlato, G. E. Molini, F. Sasanelli, O. Bugiani, F. Tagliavini, *Genetic, neuropathologic, and biochemical study of a patient from a new Italian kindred* **1998**, 50, 688.

[6] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, J. Näslund, L. Lannfelt, *Nature Neuroscience* **2001**, 4, 887.

[7] C. Van Broeckhoven, J. Haan, E. Bakker, J. Hardy, W. Van Hul, A. Wehnert, M. Vegter-Van der Vlis, R. Roos, *Science* **1990**, 248, 1120.

Author Manuscript

This article is protected by copyright. All rights reserved.

[8] a)L. Xu, S. S. Shan, X. C. Wang, *J. Phys. Chem. B* **2013**, 117, 6206; b)M. G. Krone, A. Baumketner, S. L. Bernstein, T. Wyttenbach, N. D. Lazo, D. B. Teplow, M. T. Bowers, J.-E. Shea, *Journal of Molecular Biology* **2008**, 381, 221.

[9] a)D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, D. J. Selkoe, *Nature* **2002**, 416, 535; b)I. Benilova, E. Karran, B. De Strooper, *Nature Neuroscience* **2012**, 15, 349.

[10] F. A. Aprile, P. Sormanni, M. Podpolny, S. Chhangur, L.-M. Needham, F. S. Ruggeri, M. Perni, R. Limbocker, G. T. Heller, T. Sneideris, T. Scheidt, B. Mannini, J. Habchi, S. F. Lee, P. C. Salinas, T. P. J. Knowles, C. M. Dobson, M. Vendruscolo, *Proceedings of the National Academy of Sciences* **2020**, 117, 13509.

[11] P. Hoang, N. M. Khashab, *Small Methods* **2019**, 3, 1900611.

[12] L. Chen, F. Rashid, A. Shah, H. M. Awan, M. Wu, A. Liu, J. Wang, T. Zhu, Z. Luo, G. Shan, *Proceedings of the National Academy of Sciences* **2015**, 112, 10002.

[13] X. Fang, A. Sen, M. Vicens, W. Tan, *ChemBioChem* **2003**, 4, 829.

[14] H. K. Narang, *Journal of Neuropathology & Experimental Neurology* **1980**, 39, 621.

[15] L. Gremer, D. Schölzel, C. Schenk, E. Reinartz, J. Labahn, R. B. G. Ravelli, M. Tusche, C. Lopez-Iglesias, W. Hoyer, H. Heise, D. Willbold, G. F. Schröder, *Science* **2017**, 358, 116.

[16] L. Hou, H. Shao, Y. Zhang, H. Li, N. K. Menon, E. B. Neuhaus, J. M. Brewer, I.-J. L. Byeon, D. G. Ray, M. P. Vitek, T. Iwashita, R. A. Makula, A. B. Przybyla, M. G. Zagorski, *Journal of the American Chemical Society* **2004**, 126, 1992.

[17] J. C. Stroud, C. Liu, P. K. Teng, D. Eisenberg, *Proceedings of the National Academy of Sciences* **2012**, 109, 7717.

[18] G.-f. Chen, T.-h. Xu, Y. Yan, Y.-r. Zhou, Y. Jiang, K. Melcher, H. E. Xu, *Acta Pharmacologica Sinica* **2017**, 38, 1205.

[19] K. H. Lim, H. H. Collver, Y. T. H. Le, P. Nagchowdhuri, J. M. Kenney, *Biochemical and Biophysical Research Communications* **2007**, 353, 443.

[20] G. P. Saborio, B. Permanne, C. Soto, *Nature* **2001**, 411, 810.

[21] R. Atarashi, K. Satoh, K. Sano, T. Fuse, N. Yamaguchi, D. Ishibashi, T. Matsubara, T. Nakagaki, H. Yamanaka, S. Shirabe, M. Yamada, H. Mizusawa, T. Kitamoto, G. Klug, A. McGlade, S. J. Collins, N. Nishida, *Nature Medicine* **2011**, 17, 175.

Author Manuscript

This article is protected by copyright. All rights reserved.

[22] M. Kollmer, W. Close, L. Funk, J. Rasmussen, A. Bsoul, A. Schierhorn, M. Schmidt, C. J. Sigurdson, M. Jucker, M. Fändrich, *Nature Communications* **2019**, 10, 4760.

[23] B. Olsson, R. Lautner, U. Andreasson, A. Öhrfelt, E. Portelius, M. Bjerke, M. Hölttä, C. Rosén, C. Olsson, G. Strobel, E. Wu, K. Dakin, M. Petzold, K. Blennow, H. Zetterberg, *The Lancet Neurology* **2016**, 15, 673.

[24] a)M. Layani, M. Gruchko, O. Milo, I. Balberg, D. Azulay, S. Magdassi, *ACS Nano* **2009**, 3, 3537; b)A. Shimoni, S. Azoubel, S. Magdassi, *Nanoscale* **2014**, 6, 11084; c)T.-S. Wong, T.-H. Chen, X. Shen, C.-M. Ho, *Analytical Chemistry* **2011**, 83, 1871.

[25] a)J. González-Gutiérrez, R. Pérez-Isidoro, J. C. Ruiz-Suárez, *Review of Scientific Instruments* **2017**, 88, 074101; b)N. Kim, Z. Li, C. Hurth, F. Zenhausern, S.-F. Chang, D. Attinger, *Analytical Methods* **2012**, 4, 50; c)J. O. Andersen, C. B. Henriksen, J. Laursen, A. A. Nielsen, *Computers and Electronics in Agriculture* **1999**, 22, 51; d) T. Yakhno, A. Pakhomov, A. Sanin, V. Kazakov, R. Ginoian, V. Yakhno, *arXiv preprint arXiv:2006.15565* **2020**.

[26] E. Alberts, C. Warner, E. Barnes, K. Pilkiewicz, E. Perkins, A. Poda, *Colloids and Surfaces B: Biointerfaces* **2018**, 161, 210.

[27] a)K. Sefiane, *Advances in Colloid and Interface Science* **2014**, 206, 372; b)R. Chen, L. Zhang, D. Zang, W. Shen, *Advances in Colloid and Interface Science* **2016**, 231, 1; c)J. M. Cameron, H. J. Butler, D. S. Palmer, M. J. Baker, *Journal of Biophotonics* **2018**, 11, e201700299; d)C. Hurth, T. Contente-Cuomo, M. Murtaza, F. Zenhausern, *J. Nanomed.* **2018**, 2, 1010; e)T. A. Yakhno, A. A. Sanin, R. G. Ilyazov, G. V. Vildanova, R. A. Khamzin, N. P. Astascheva, M. G. Markovsky, V. D. Bashirov, V. G. Yakhno, *Journal of Biomedical Science and Engineering* **2015**, 8, 1.

[28] a) D. Brutin, *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2013**, 429, 112; b)H. H. Lee, S. C. Fu, C. Y. Tso, C. Y. H. Chao, *International Journal of Heat and Mass Transfer* **2017**, 105, 230; c)X. Zhong, H. Xie, F. Duan, *Applied Thermal Engineering* **2017**, 111, 1565; d)B. Pathak, J. Christy, K. Sefiane, D. Gozuacik, *Langmuir* **2020**, 36, 9728; e)I. I. Smalyukh, O. V. Zribi, J. C. Butler, O. D. Lavrentovich, G. C. L. Wong, *Physical Review Letters* **2006**, 96, 177801.

[29] a)V. H. Chhasatia, A. S. Joshi, Y. Sun, *Applied Physics Letters* **2010**, 97, 231909; b)A. Askounis, K. Sefiane, V. Koutsos, M. E. R. Shanahan, *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2014**, 441, 855.

[30] a)Y.-F. Li, Y.-J. Sheng, H.-K. Tsao, *Langmuir* **2013**, 29, 7802; b)D. Orejon, K. Sefiane, M. E. R. Shanahan, *Langmuir* **2011**, 27, 12834.

[31] R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, T. A. Witten, *Nature* **1997**, 389, 827.

Author Manuscript

This article is protected by copyright. All rights reserved.

[32] a)H. Hu, R. G. Larson, *The Journal of Physical Chemistry B* **2006**, 110, 7090; b)R. G. Larson, *Angewandte Chemie International Edition* **2012**, 51, 2546; c)R. G. Larson, *AIChE Journal* **2014**, 60, 1538, d) D. Bruitin, V. Starov, *Chem. Soc. Rev.* **2018,** 47, 558..

[33] R. G. G. Larson, M. A. López, D. W. Lim, J. Lahann, *MRS Proceedings* **2010**, 1273, 1273, 1273-mm03-01.

[34] R. Bhardwaj, X. Fang, P. Somasundaran, D. Attinger, *Langmuir* **2010**, 26, 7833.

[35] G. Chen, G. J. Mohamed, *The European Physical Journal E* **2010**, 33, 19.

[36] a)X. Deng, C. Friedmann, J. Lahann, *Angewandte Chemie International Edition* **2011**, 50, 6522; b)H.-Y. Chen, Y. Elkasabi, J. Lahann, *Journal of the American Chemical Society* **2006**, 128, 374.

[37] Y. Zhang, P. S. Cremer, *Current Opinion in Chemical Biology* **2006**, 10, 658.

[38] J. Zhang, *Protein-protein interactions-computational and experimental tools* **2012**, 1.

[39] a)J. C. Mello Román, J. L. Vázquez Noguera, H. Legal-Ayala, D. P. Pinto-Roa, S. Gomez-Guerrero, M. García Torres, *Entropy* **2019**, 21, 244; b) C. E. Shannon, *The Bell System Technical Journal* **1948**, 27, 379.

[40] S. Saxena, 2020 50 Most Famous Paintings of All Time in the Art History (Ranked). Available from: https://bookmypainting.com/blog/famous-paintings/.

[41] L. Hamadeh, S. Imran, M. Bencsik, G. R. Sharpe, M. A. Johnson, D. J. Fairhurst, *Scientific Reports* **2020**, 10, 3313.

[42] a)D. Zhou, F. Tian, X. Tian, L. Sun, X. Huang, F. Zhao, N. Zhou, Z. Chen, Q. Zhang, M. Yang, Y. Yang, X. Guo, Z. Li, J. Liu, J. Wang, J. Wang, B. Wang, G. Zhang, B. Sun, W. Zhang, D. Kong, K. Chen, X. Li, *Nature Communications* **2020**, 11, 2961; b)E. J. Topol, *Nature* **2019**, 572, 36; c)E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, D. Van Valen, *Nature Methods* **2019**, 16, 1233; d)S. Webb, *Nature* **2018**, 554, 555; e)Y. LeCun, Y. Bengio, G. Hinton, *Nature* **2015**, 521, 436.

[43] a)H. C. Shin, H. R. Roth, M. C. Gao, L. Lu, Z. Y. Xu, I. Nogues, J. H. Yao, D. Mollura, R. M. Summers, *Ieee Transactions on Medical Imaging* **2016**, 35, 1285; b)N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. M. Liang, *Ieee Transactions on Medical Imaging* **2016**, 35, 1299.

[44] S. J. Pan, Q. Yang, *IEEE Transactions on knowledge and data engineering* **2010**, 22, 1345.

[45] A. Krizhevsky, I. Sutskever, G. E. Hinton, presented at *Advances in neural information processing systems*, **2012**.

Author Manuscript

This article is protected by copyright. All rights reserved.

[46] a)A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, *Nature* **2017**, 542, 115; b)N. Tomašev, X. Glorot, J. W. Rae, M. Zielinski, H. Askham, A. Saraiva, A. Mottram, C. Meyer, S. Ravuri, I. Protsyuk, A. Connell, C. O. Hughes, A. Karthikesalingam, J. Cornebise, H. Montgomery, G. Rees, C. Laing, C. R. Baker, K. Peterson, R. Reeves, D. Hassabis, D. King, M. Suleyman, T. Back, C. Nielson, J. R. Ledsam, S. Mohamed, *Nature* **2019**, 572, 116; c)P. Wang, X. Xiao, J. R. Glissen Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, X. Yang, L. Li, J. He, X. Yi, J. Liu, X. Liu, *Nature Biomedical Engineering* **2018**, 2, 741.

[47] a)T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, *Nature methods* **2019**, 16, 67; b)C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle, *Molecular systems biology* **2016**, 12, 878.

[48] A. Morimoto, K. Irie, K. Murakami, Y. Masuda, H. Ohigashi, M. Nagao, H. Fukuda, T. Shimizu, T. Shirasawa, *Journal of Biological Chemistry* **2004**, 279, 52781.

[49] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, *Nature neuroscience* **2001**, 4, 887.

[50] J. Kyte, R. F. Doolittle, *Journal of Molecular Biology* **1982**, 157, 105.

[51] G. Fischer, *Chemical Society Reviews* **2000**, 29, 119.

[52] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, presented at *Proceedings of the IEEE International Conference on Computer Vision*, **2017**.

[53] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, *arXiv preprint arXiv:1611.07450* **2016**.

[54] L. v. d. Maaten, G. Hinton, *Journal of machine learning research* **2008**, 9, 2579.

[55] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, presented at *2009 IEEE conference on computer vision and pattern recognition*, **2009**.

[56] O. Crescenzi, S. Tomaselli, R. Guerrini, S. Salvadori, A. M. D'Ursi, P. A. Temussi, D. Picone, *European Journal of Biochemistry* **2002**, 269, 5642.

[57] Y. Lin, B. R. Sahoo, D. Ozawa, M. Kinoshita, J. Kang, M. H. Lim, M. Okumura, Y. H. Huh, E. Moon, J. H. Jang, H.-J. Lee, K.-Y. Ryu, S. Ham, H.-S. Won, K.-S. Ryu, T. Sugiki, J. K. Bang, H.-S. Hoe, T. Fujiwara, A. Ramamoorthy, Y.-H. Lee, *ACS Nano* **2019**, 13, 8766.

[58] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, *Nucleic Acids Research* **2018**, 46, W315.

[59] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, presented at *Proceedings of the IEEE conference on computer vision and pattern recognition*, **2018**.

Author Manuscript

This article is protected by copyright. All rights reserved.