ADVANCED MATERIALS

Supporting Information

for *Adv. Mater.,* DOI: 10.1002/adma.202110404

Deep-Learning-Assisted Stratification of Amyloid Beta Mutants Using Drying Droplet Patterns

*Azam Jeihanipour and Jörg Lahann**

Supplementary Information

Table SI.1: Comparison of the DL-resolved staining method (presented in this table) vs. some of competing methods in terms of cost and applicability

	Method			
	Deep learning resolved staining (presented in this paper)	Mass spectrometry (MS)-based proteomics[1]	Atomic force microscopy (AFM)[2]	Cryo electron microscopy (cryo-EM)[3]
Application	Stratification of structural and conformational mismatches of peptides	Measuring amino acid substitutions directly in proteins down to a level of 0.001%[1]	Obtaining visual information about the aggregates (e.g. size, shape and aggregation extent)	Detecting late-stage formation of fibrils[3b]
Clinical application	Promising to diagnose and stratify, e.g., AD patients	Promising to be used in oncology[4]	Not currently used	Not currently used
Portability	Portable	Non-portable	Non-portable	Non-portable
Miniaturization	Minaturizible	Large-scale	Large-scale	Large-scale
Point-of-care usage	compatible	Incompatible	Incompatible	Incompatible
Cost	Low	Up to 100 k USD	Up to 500 K USD	Up to 1000K USD
Summary	Simple sample preparation, high accuracy detection potential for early stage detection of diseases	Hindered by the low abundance of substitutions[5] and complicated sample preparation	Complicated sample preparation, detection only at later stages of disease progression	Detection only at later stages of disease progression

References:
1) Z. Zhang, B. Shah, P. V. Bondarenko, *Biochemistry* **2013**, 52, 8165.
2) J. A. J. Housmans, G. Wu, J. Schymkowitz, F. Rousseau, *The FEBS Journal*, n/a.
3) 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; b)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.
4) A. Macklin, S. Khan, T. Kislinger, *Clinical Proteomics* **2020**, 17, 17.
5) E. Mordret, O. Dahan, O. Asraf, R. Rak, A. Yehonadav, G. D. Barnabas, J. Cox, T. Geiger, A. B. Lindner, Y. Pilpel, *Molecular cell* **2019**, 75, 427.

A)

Wild type $A\beta_{1-42}$ sequences: Asp-Ala[2]-Glu-Phe-Arg-His-Asp-Ser-Gly-Tyr-Glu-Val-His-His-Gln-Lys-Leu-Val-Phe-Phe-Ala[21]-Glu[22]-Asp-Val-Gly-Ser-Asn-Lys-Gly-Ala-Ile-Ile-Gly-Leu-Met-Val-Gly[37]-Gly-Val-Val-Ile-Ala

B)

Physico-chemical properties	**wtAβ**	**Val2**	**Thr2**	**Gly21**	**Gly22**	**Gln22**	**Lys22**	**Asp37**
Short abbreviation	-	A2V	A2T	A21G	E22G	E22Q	E22K	G37D
Familial mutation name	wtAβ	A2V	-	Flemish	Arctic	Dutch	Italian	-
Molecular weight (g/mol)	4514.1	4542.2	4544.1	4500.1	4442.0	4513.1	4513.2	4572.1
Isoelectric point	5.2	5.2	5.2	5.2	5.8	5.8	6.3	4.8
Net charge at pH 7.0	-2.7	-2.7	-2.7	-2.7	-1.7	-1.7	-0.7	-3.7
Average hydrophilicity	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.1	0.0
Ratio of hydrophilic residues /total number of residues	31%	31%	31%	31%	29%	31%	31%	33%

Table SI.2: Physio-chemical properties of Aβ peptides with amino acid mismatches
A) Sequences of wild type $A\beta_{42}$. B) The physcio-chemical properties of wild-type $A\beta_{42}$ and the ones with single point mutations.

References:

1) Lehninger, A. L., et al. (1982). "Principles of Biochemistry Worth Publishers." Inc., New York: 615-643.
2) Hopp, T. P. and K. R. Woods (1981). "Prediction of protein antigenic determinants from amino acid sequences." Proceedings of the National Academy of Sciences **78**(6): 3824-3828.
3) https://www.bachem.com/knowledge-center/peptide-calculator/

Test Images	Validation Accuracy (%)	Test-Images Accuracy (%)
Single amino acid mismatches		
AI	100.0	96.4
BH	98.8	97.3
CG	98.8	98.6
DI	98.9	93.3
EF	98.4	98.9
Average	99.0	96.9
STDEVA	0.6	2.3
Conformational mismatches		
AI	97.5	96.2
BH	96.4	96.8
CG	97.0	90.4
DI	95.3	97.1
EF	97.0	96.8
Average	96.7	95.5
STDEVA	0.8	2.9

Table SI.3: Reproducibility testing of the method

For training of the network with the images, two random sets of data, including 8 plates, e.g., A and I in Figure SI. 1, were selected as test images, called test images AI, and the rest of images were selected as training images. Then about 400 random images of the training set and 90 random images of the test set (different images, not included in training set) were used for training and testing the DL network. Additionally, during the training, 10% of the images were used for validation of the training process. The accuracy of the trained DL network on the validation set is reported as "validation accuracy". The validation set has been obtained from the same plates than the training set. The "test image accuracy" is the percentage of corrected classification of the test images by the trained DL network which were never seen by the network and obtained on different runs than the training images. This process was repeated several times for different selected test and training sets, to examine the reproducibility of the method. In each trial, a new pair of the runs were selected as test images and the rest as training images. The results are reported in the table for both single amino acid mismatches and conformational mismatches. The network used for this part of study was VGG16.

Images	Nr. of Images	Min. Entropy	Max. Entropy	Average Entropy
World famous paintings	48	5.55	7.92	7.17 ± 0.51
Coffee ring images from literature	60	3.71	7.46	5.84 ± 1.30
PLM images of peptide stains from current study	100	4.34	5.90	5.14 ± 0.31

Table SI.4: Shannon's entropy of different image cohorts

Shannon's entropy of three sets of images was calculated using Matlab's entropy function. The PLM images of peptide stains from the current study were randomly selected from all studied groups. Images of the world-famous paintings were obtained from https://bookmypainting.com/blog/famous-paintings/. The coffee ring images used in the control cohort (Coffee ring images from literature) have been obtained from the following references.

1. Cameron JM, Butler HJ, Palmer DS, Baker MJ. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *Journal of Biophotonics* 2018, **11**(4)**:** e201700299.
2. Devineau S, Anyfantakis M, Marichal L, Kiger L, Morel M, Rudiuk S, *et al.* Protein Adsorption and Reorganization on Nanoparticles Probed by the Coffee-Ring Effect: Application to Single Point Mutation Detection. *Journal of the American Chemical Society* 2016, **138**(36)**:** 11623-11632.
3. Hu H, Larson RG. Marangoni Effect Reverses Coffee-Ring Depositions. *The Journal of Physical Chemistry B* 2006, **110**(14)**:** 7090-7094.
4. Li Y, Zhao Z, Lam ML, Liu W, Yeung PP, Chieng C-C, *et al.* Hybridization-induced suppression of coffee ring effect for nucleic acid detection. *Sensors and Actuators B: Chemical* 2015, **206:** 56-64.
5. Li X, Sanderson AR, Allen SS, Lahr RH. Tap water fingerprinting using a convolutional neural network built from images of the coffee-ring effect. *Analyst* 2020, **145**(4)**:** 1511-1523.
6. Lin S-Y, Yang K-C, Chen L-J. Effect of Surface Hydrophobicity on Critical Pinning Concentration of Nanoparticles To Trigger the Coffee Ring Formation during the Evaporation Process of Sessile Drops of Nanofluids. *The Journal of Physical Chemistry C* 2015, **119**(6)**:** 3050-3059.
7. Nguyen TAH, Hampton MA, Nguyen AV. Evaporation of Nanoparticle Droplets on Smooth Hydrophobic Surfaces: The Inner Coffee Ring Deposits. *The Journal of Physical Chemistry C* 2013, **117**(9)**:** 4707-4716.

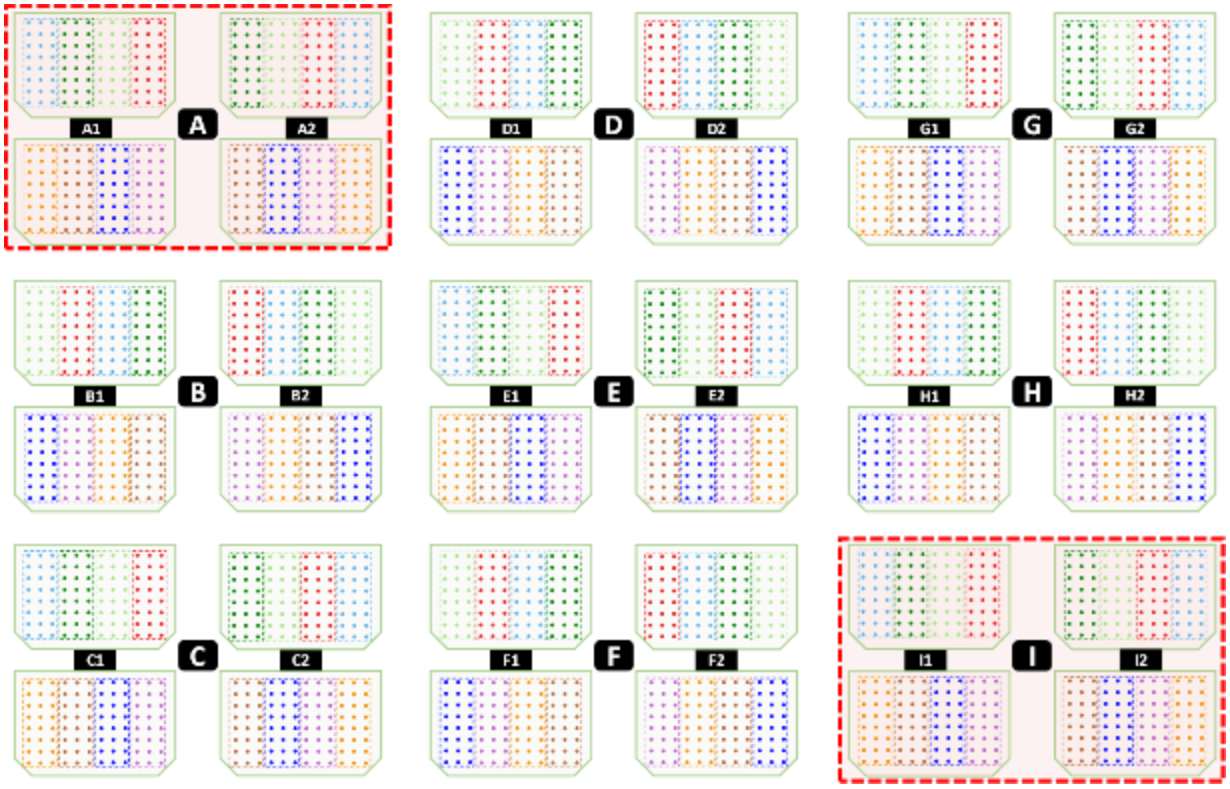

Figure SI.1: Experimental set up and approach used for depositing peptide droplets
In order to increase the robustness of the stratification of Aβ variants using DL network, nullifying the effect of possible surface non-homogeneity on the method performance, and avoid artifacts due to the position of the droplets on the plate, the following approach was taken. The pipetting system was programmed to put 96 droplets on each coated glass plate. In each run of pipetting, 192 droplets were placed on 2 plates, in the format of the standard 96-well plates, i.e., 12 columns by 8 rows on each plate. In each run, 8 different variants were placed simultaneously following a random arrangement on both plates. An example of this random arrangement is shown in this figure. In every run the location of each peptide variant on each plate was different than the previous run. Each run was named with an alphabet and a number of either 1 or 2. The pair plates A1, for example, are prepared in one run of pipetting system and subsequently the paired plate A2 was obtained. For each set of peptides, more than 20 runs were performed. For illustration, 18 runs result in a total of 36 plates decorated with 3456 droplets.

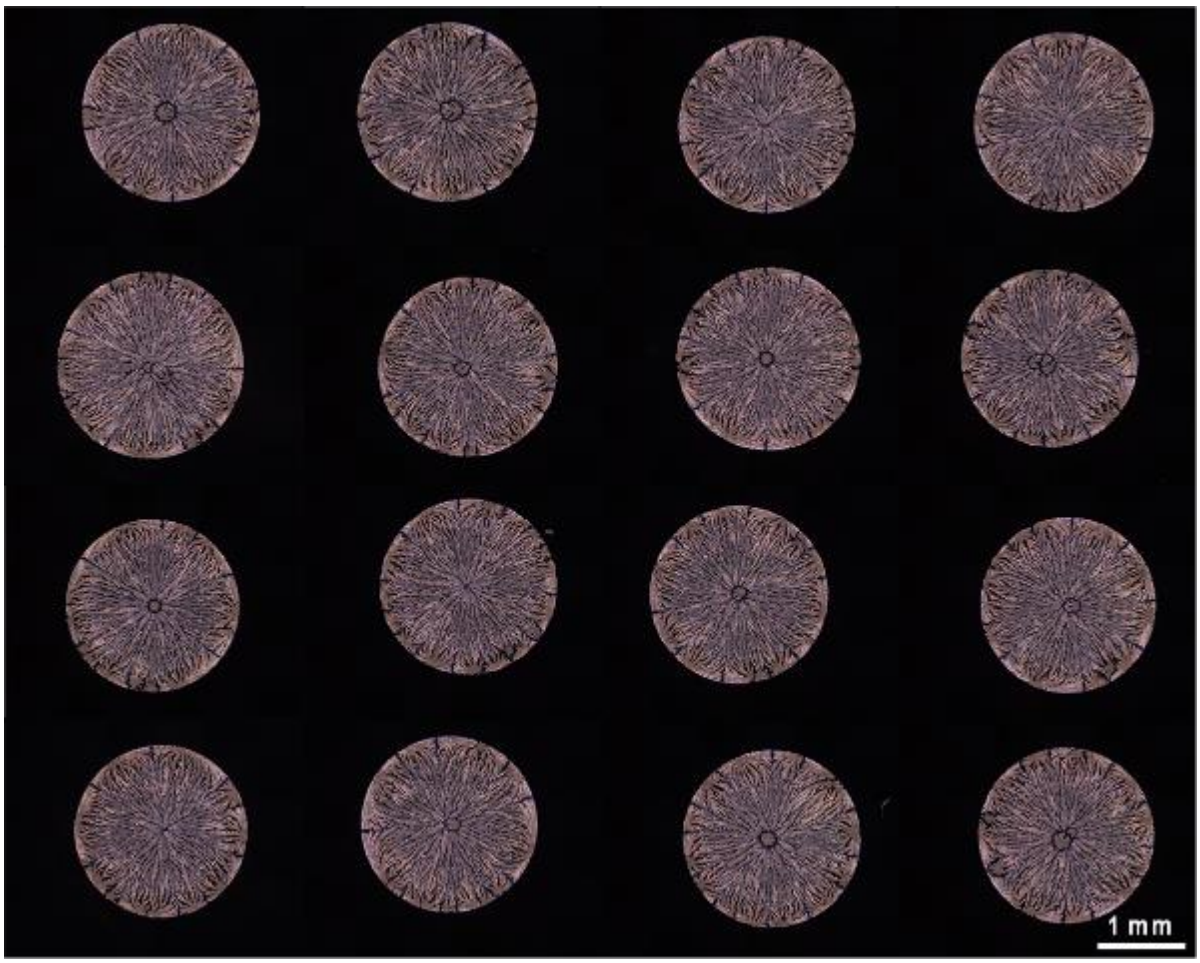

Figure SI.2: Reproducibility of dried droplet deposition stains
A few randomly selected PLM image of deposition pattern of wild type $A\beta_{42}$ peptide. By drying peptide solution droplets under defined experimental conditions, the deposition patterns are highly conserved and reproducible.

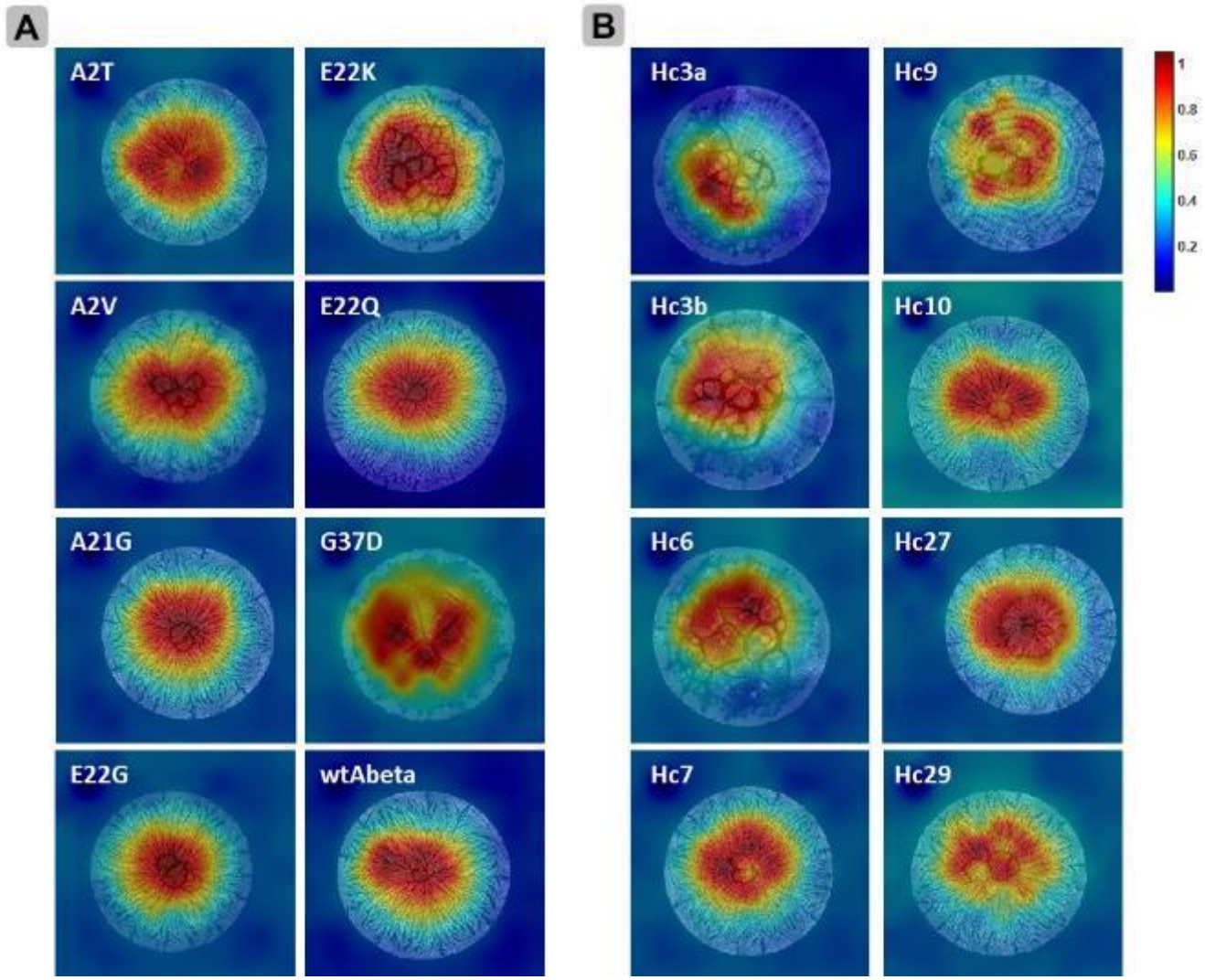

Figure SI.3: Grad-CAM activation maps
Gradient-weighted class activation mapping (Grad-CAM) was used to generate activation maps of the final convolutional layer of the DL network to identify features of discriminatory regions of the PLM images. Heat map layers of the PLM images of deposition patter of peptides with A) single amino acid mismatches and B) protein misfolding. The DL network is NasNetLarge and all images are correctly classified by the trained network.

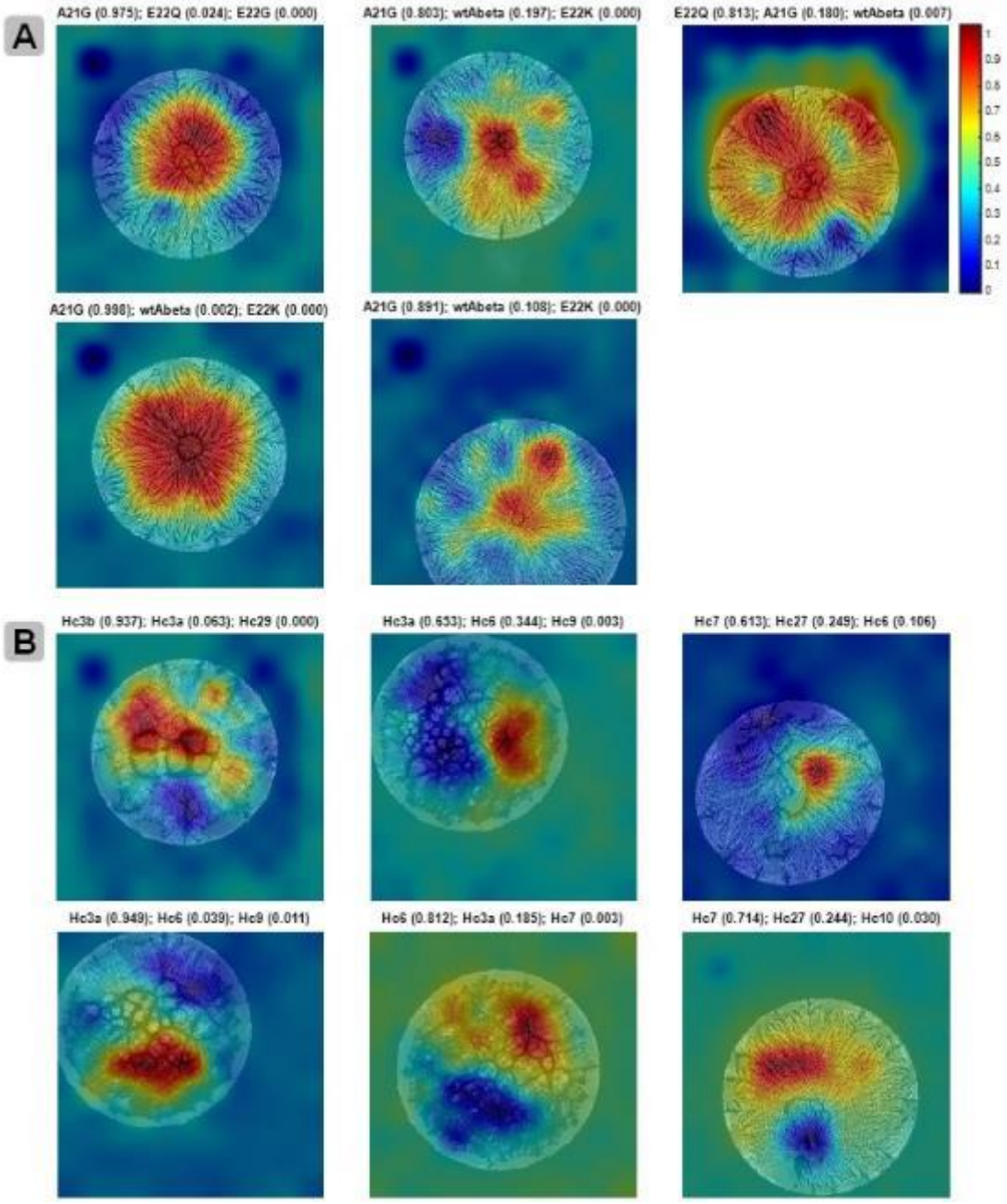

Figure SI. 4: Grad-CAM activation maps of misclassified PLM images
The misclassified images by trained NasNet-Large are analyzed by Grad-CAM. The heat map of the final convolutional layer of the DL network and the top three scored classes are shown. **A)** Out of 720 never-seen PLM images of deposition pattern of peptides with single amino acid mismatches, 5 images were classified wrong. **B)** There were 6 misclassifications out of 720 never-seen PLM images of deposition pattern of misfolded peptides.

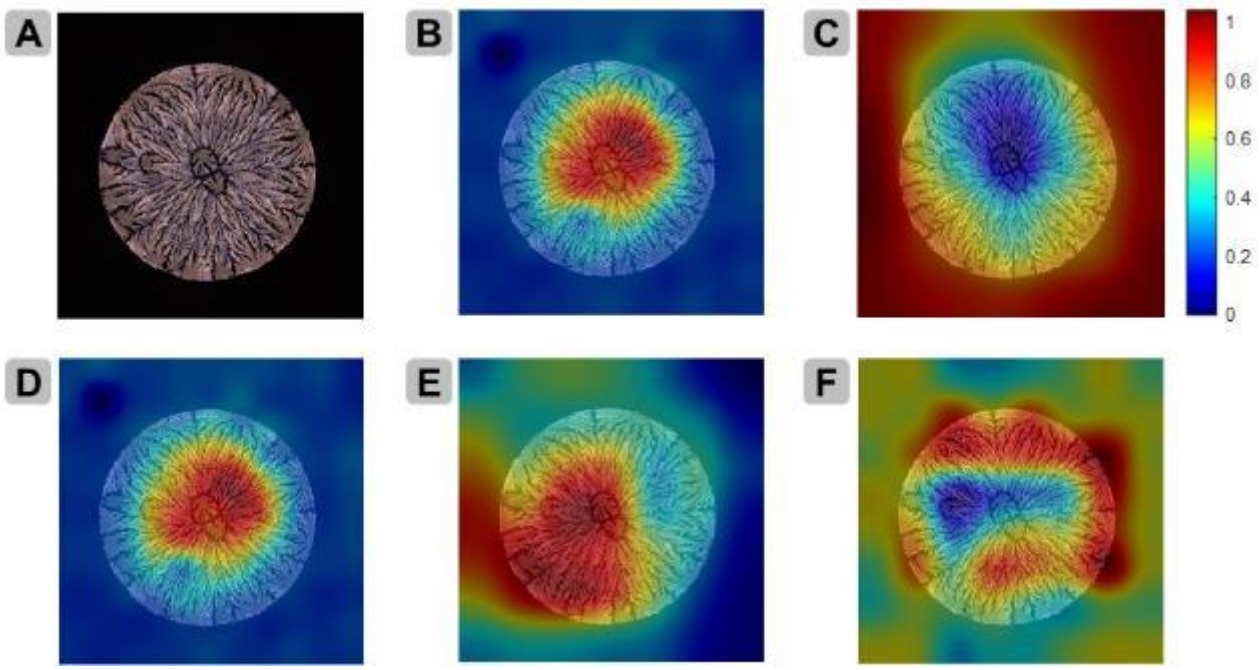

Figure SI. 5: Comparison of Grad-CAM activation maps for a misclassified image from different DL networks

One of the PLM images misclassified by trained NasNet-Large was also misclassified by other trained DL networks. (A) a deposition pattern PLM image of wild type $A\beta_{42}$ was classified as A21G by (B) NasNet-Large, (C) DenseNet-201 (D) InceptionResNet V2, (E) GoogleNet, and (G) DarkNet-53. The Grad-CAM heat map shows that for some of the DL networks, like DenseNet-201, the background of this image is affecting the final scores, given for the classification.

Video SI.1: Drying process of a wild type $A\beta_{42}$ peptide droplet
Drying process of a sessile droplet of wild type $A\beta_{42}$ peptide solution on poly(p-xylylene) (PPX) coated surface under polarized light. The volume of droplet is 2 µL and the drying and recording is performed under normal laboratory conditions. The video is accelerated 8 times.