Modelling amyloidogenic scaffolds of biofilms

Michał Burdukiewicz
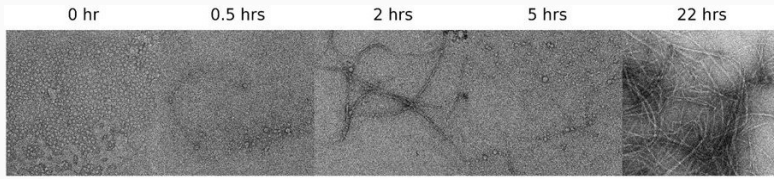
University of Wrocław, Department of Genomics

# The role of amyloids in biofilm formation

Curli fibers:

- involved in adhesion to surfaces, cell aggregation, and biofilm formation.
- formed by highly aggregative structural subunits, e.g, curlin (CsgA).

CsgA fibre formation through self-assembly (amyloidogenic aggregation).

Scale bar = 200nm. Taylor et al. (2016).

A similar aggregation mechanism is demonstrated by proteins associated with various neurodegenerative disorders (e.g., Alzheimer's, Parkinson's, Creutzfeldta-Jakob's diseases).
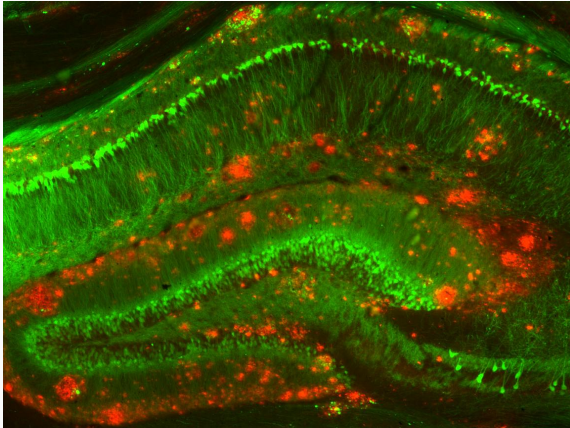

β-amyloid aggregates (red) around neurons (green). Strittmatter Laboratory, Yale University.

The aggregation of amyloids is initiated by 6- to 15-residue segments called hot spots, diverse subsequences that form unique zipper-like β-structures.
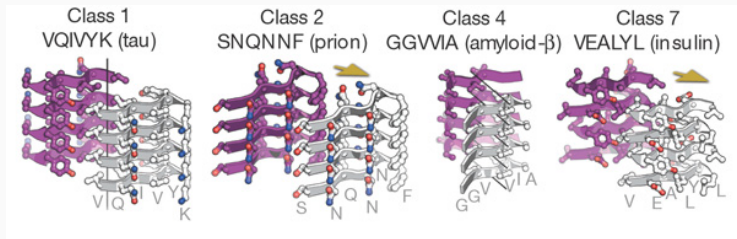
Sawaya et al. (2007)

CsgC was reported to effectively (molar ratios as low as 1:500) inhibit CsgA aggregation. *In vivo*, it prevents CsgA from aggregating within the periplasm (Evans et al., 2015).
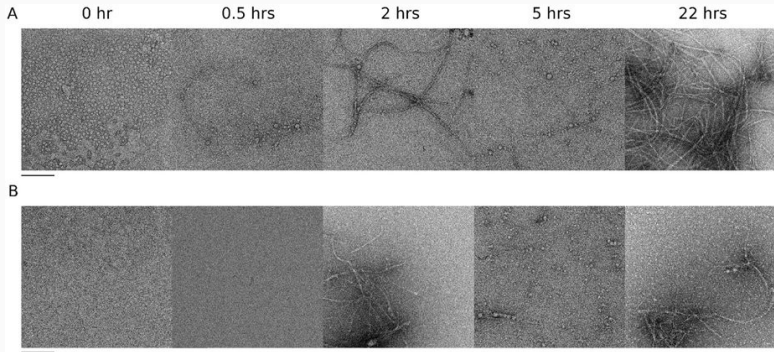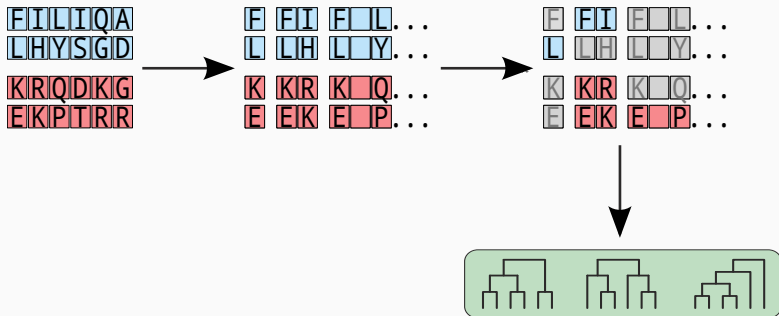
A. CsgA fibre formation. B. CsgA+CsgC 200:1 (molar ratio) fibre formation.

Scale bar = 200nm. Taylor et al. (2016).

Amyloidogenic motifs

Which motifs (countinous or gapped subsequences of amino acids) are associated with amyloidogenicity?

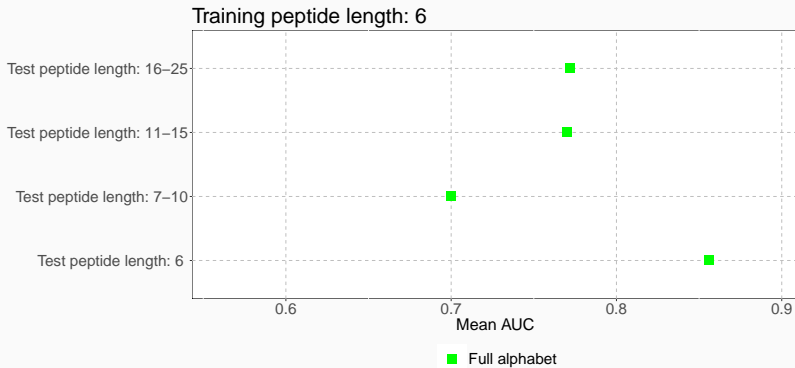Can we build a predictor of amyloidogenicity using these motifs?

QuiPT

Quick **P**ermutation **T**est is a fast alternative to permutation tests for motifs. It also allows precise estimation of p-value.

QuiPT is avaible as part of the **biogram** R package.

Cross-validation

Full alphabet: 20 standard amino acids.

Area Under the Curve (AUC): a performance measure between 0 (completely wrong predictions) and 1 (completely right predictions).
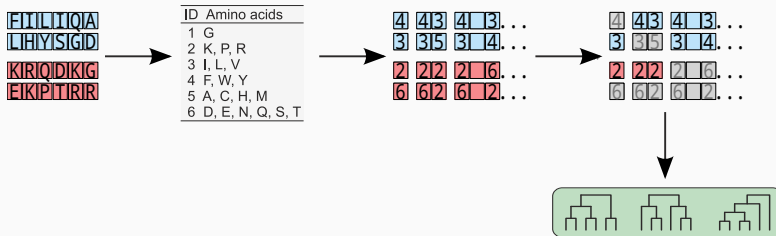
Reduced amino acid alphabets

Does amyloidogenicity depend on the exact sequence of amino acids?

Standard reduced amino acid alphabets

To date, several reduced amino acid alphabets have been proposed, which have been applied to (among others) protein folding and protein structure prediction (Kosiol et al., 2004; Melo and Marti-Renom, 2006).

Standard reduced amino acid alphabets

Cross-validation


Training peptide length: 6
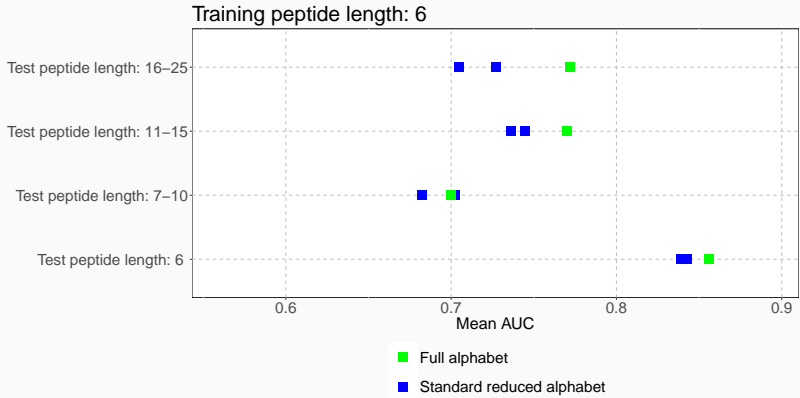
Standard reduced amino acid alphabets do not enhance discrimination between amyloidogenic and non-amyloidogenic proteins.
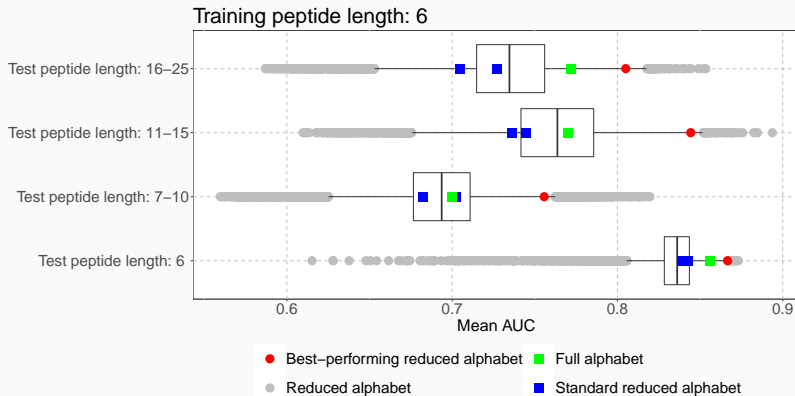
Novel reduced amino acid alphabets

- 17 measures handpicked from AAIndex database:
  - size of residues,
  - hydrophobicity,
  - solvent surface area,
  - frequency in β-sheets,
  - contactivity.
- 524 284 amino acid reduced alphabets with different level of amino acid alphabet reduction (three to six amino acid groups).

Cross-validation

Training peptide length: 6
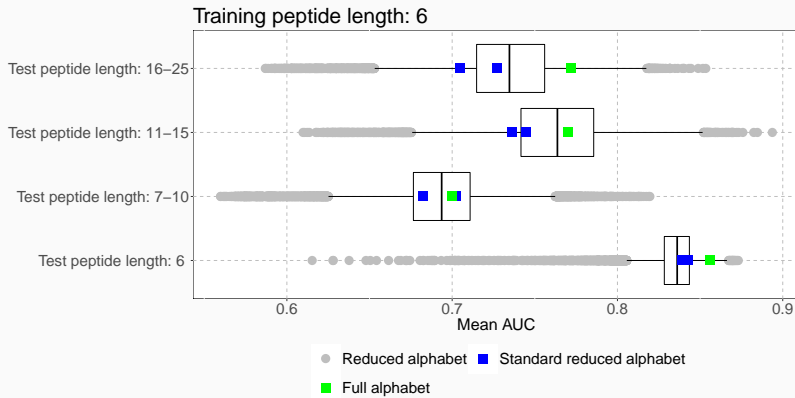
Hinges of boxes correspond to the 0.25 and 0.75 quartiles. The bar inside the box represents the median. The gray circles correspond to the reduced alphabets with the AUC outside the 0.95 confidence interval.

Training peptide length: 6

The best-perfoming reduced alphabet

Subgroup ID	Amino acids
1	G
2	K, P, R
3	I, L, V
4	F, W, Y
5	A, C, H, M
6	D, E, N, Q, S, T

The best-perfoming reduced alphabet

Subgroup ID	Amino acids
1	G
2	K, P, R
3	I, L, V
4	F, W, Y
5	A, C, H, M
6	D, E, N, Q, S, T

Group 3 and 4 - hydrophobic amino acids.

The best-perfoming reduced alphabet

Subgroup ID	Amino acids
1	G
2	K, P, R
3	I, L, V
4	F, W, Y
5	A, C, H, M
6	D, E, N, Q, S, T

Group 2 - charged breakers of β-structures.

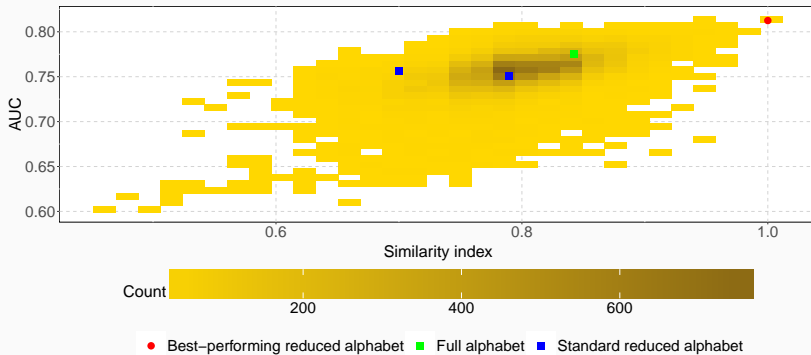Is the best-performing reduced amino alphabet associated with amyloidogenicity?

Similarity index
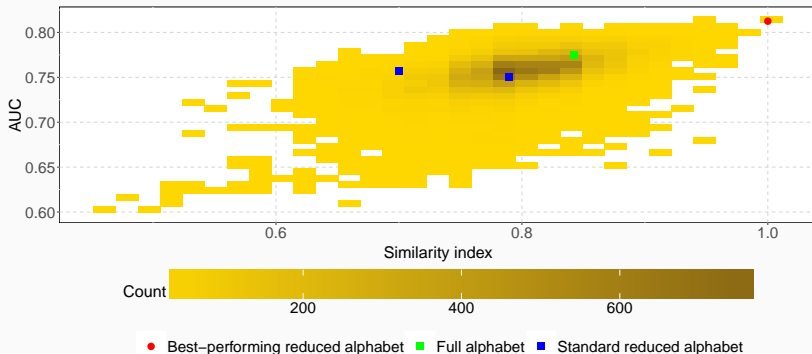
Similarity index (Stephenson and Freeland, 2013) measures the similarity between two reduced alphabets (1 - identical, 0, totally dissimilar).

The color of a square is proportional to the number of reduced alphabets in its area.
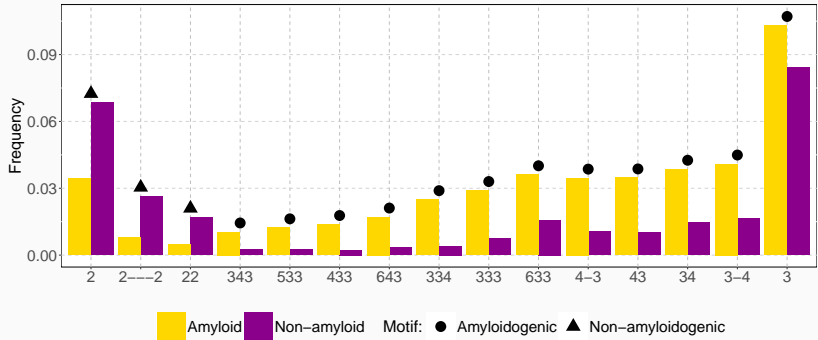
23

Similarity index

The correlation between mean AUC an similarity index is significant ($\text{p-value} \leq 2.2^{-16}$; $\rho = 0.51$).

Knowledge-discovery

Are informative n-grams found by QuiPT associated with amyloidogenicity?

Informative n-grams

Out of 65 the most informative n-grams, 15 (23%) were also found in the motifs validated experimentally (López De La Paz et al., 2002).

Benchmark and summary

Is performance of the AmyloGram, the classifier based on the best-performing reduced amino acid alphabet, also adequate on the independent dataset?

Benchmark results

Classifier	AUC	MCC
AmyloGram	**0.8972**	**0.6307**
PASTA 2.0 (Walsh et al., 2014)	0.8550	0.4291
FoldAmyloid (Garbuzynskiy et al., 2010)	0.7351	0.4526
APPNN (Família et al., 2015)	0.8343	0.5823

The predictor based on the best-performing alphabet, called AmyloGram, was benchmarked against the most popular tools for the detection of amyloid peptides using an external data set *pep424*.
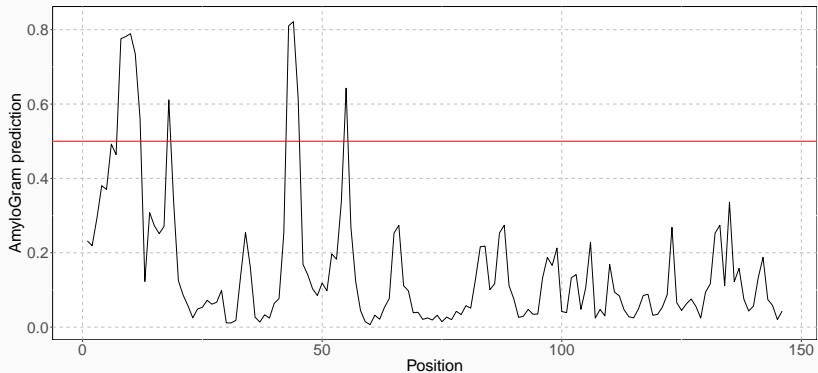
We verified experimentally with Fourier transform infrared spectroscopy (FTIR) **8** non-amyloidogenic peptides from the AmyLoad database that AmyloGram assessed with a high probability of amyloidogenicity.
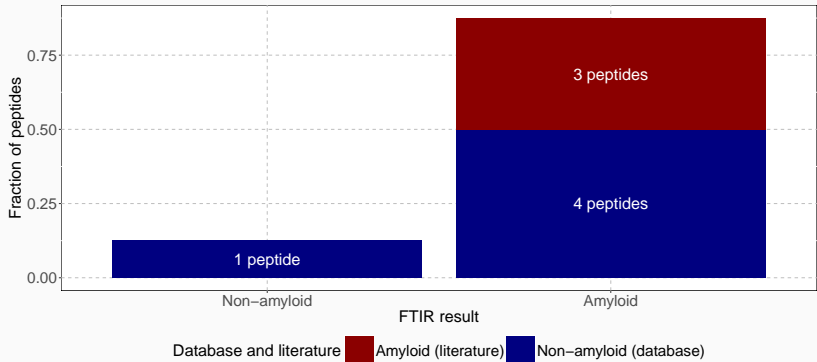
CsgA protein (E. coli K12)

Summary

We identified a group of reduced amino acid alphabets which capture properties of amyloids.

Our algorithm was also capable of extracting n-gram associated with amyloidogenicity, partially confirming experimental results.

Our software is available as a web-server: smorfland.uni.wroc.pl/amylogram.

n-gram analysis workflow is implemented in the R package **biogram**: https://cran.r-project.org/package=biogram.

31

Acknowledgements and funding

References

Evans, M. L., Chorell, E., Taylor, J. D., Åden, J., Götheson, A., Li, F., Koch, M., Sefer, L., Matthews, S. J., Wittung-Stafshede, P., Almqvist, F., and Chapman, M. R. (2015). The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation. *Molecular Cell*, 57(3):445–455.

Família, C., Dennison, S. R., Quintas, A., and Phoenix, D. A. (2015). Prediction of Peptide and Protein Propensity for Amyloid Formation. *PLOS ONE*, 10(8):e0134679.

References II

Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics (Oxford, England)*, 26(3):326–332.

Kosiol, C., Goldman, N., and Buttimore, N. H. (2004). A new criterion and method for amino acid classification. *Journal of Theoretical Biology*, 228(1):97–106.

López De La Paz, M., Goldie, K., Zurdo, J., Lacroix, E., Dobson, C. M., Hoenger, A., and Serrano, L. (2002). De novo designed peptide-based amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25):16052–16057.

Melo, F. and Marti-Renom, M. A. (2006). Accuracy of sequence alignment and fold assessment using reduced amino acid alphabets. *Proteins*, 63(4):986–995.

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. ., Riekel, C., and Eisenberg, D. (2007). Atomic structures of amyloid cross-spines reveal varied steric zippers. *Nature*, 447(7143):453–457.

Stephenson, J. D. and Freeland, S. J. (2013). Unearthing the root of amino acid similarity. *Journal of Molecular Evolution*, 77(4):159–169.

References IV

Taylor, J. D., Hawthorne, W. J., Lo, J., Dear, A., Jain, N., Meisl, G., Andreasen, M., Fletcher, C., Koch, M., Darvill, N., Scull, N., Escalera-Maurer, A., Sefer, L., Wenman, R., Lambert, S., Jean, J., Xu, Y., Turner, B., Kazarian, S. G., Chapman, M. R., Bubeck, D., de Simone, A., Knowles, T. P. J., and Matthews, S. J. (2016). Electrostatically-guided inhibition of Curli amyloid nucleation by the CsgC-like family of chaperones. *Scientific Reports*, 6:24656.

Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Research*, 42(W1):W301–W307.