Co-evolution of curli components CsgA and CsgB

Jarosław Chilimoniuk[1]*, Paweł Mackiewicz[1] and Michał Burdukiewicz[2]

*jaroslaw.chilimoniuk@gmail.com

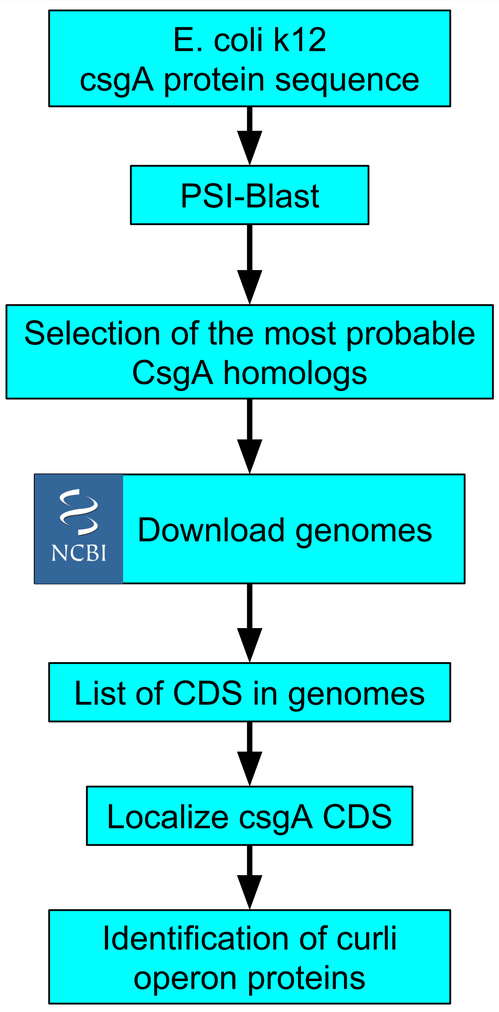
[1]University of Wrocław, Department of Genomics, Wrocław, POLAND [2]Warsaw University of Technology, Warsaw, POLAND

Introduction

CsgA, also known as a major curli component, is a secreted protein ubiquitous in biofilms of gram-negative bacteria. Thanks to its ability to create durable fibers, CsgA is a dominant proteinaceous scaffold of biofilms. In fact, CsgA belongs to amyloids, proteins that form fibers during a spontaneous aggregation.

The presence of pre-formed amyloid fibers can accelerate aggregation of other amyloids. This process is known as cross-seeding. It is extremely sequence specific and can be restricted by a difference in a single amino acid. CsgA can be in vivo cross-seeded by its nucleator protein, CsgB, but also other CsgA fibrils.

Methods



We used CsgA from *E. coli* K12 as a starting point of our search for CsgA homologs.

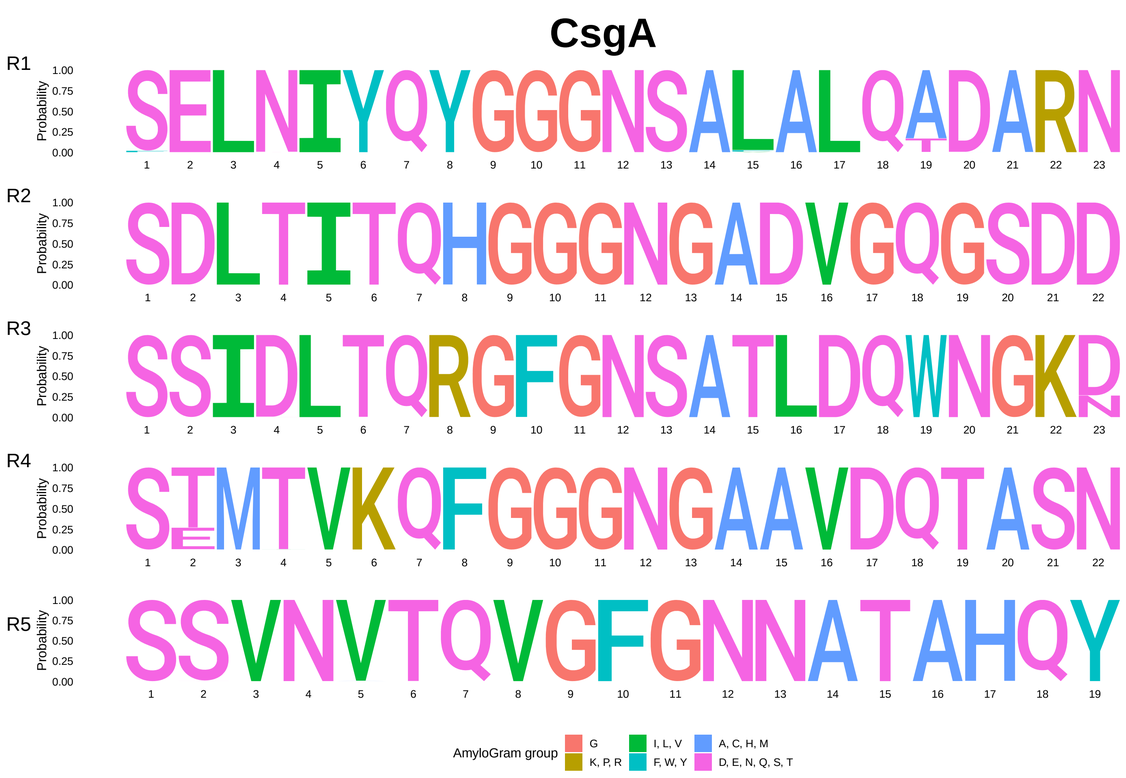
After five iterations of PSI-Blast, we found 5007 sequences producing significant alignments with E-value lower than the threshold.

We evaluated sequences using a simple heuristic approach to find the most probable candidates for the CsgA proteins.

Proteins were used to find appropriate genomes in the Nucleotide database.

Using genomic information, we reconstructed CsgBAC operons.

Results



AmyloGram (Burdukiewicz et al., 2017) was used to create reduced aa alphabet. The software assigned amino acids into 6 groups, by using eleven combinations of physicochemical properties:

I) Lowest propensity to form β-sheets (Glycine)

II) The most hydrophilic, includes two strongly basic amino acids, highly flexible (Lysine, Proline, Arginine)

III) Strongly hydrophobic, highest propensity to form β-sheets (Isoleucyne, Leucine, Valine)

IV) Aromatic properties, the most hydrophilic, the least flexible, highest propensity to form β-sheets (Phenylalanine, Tryptophan, Tyrosine)

V) The least flexible (Alanine, Cysteine, Histidine, Methionine)

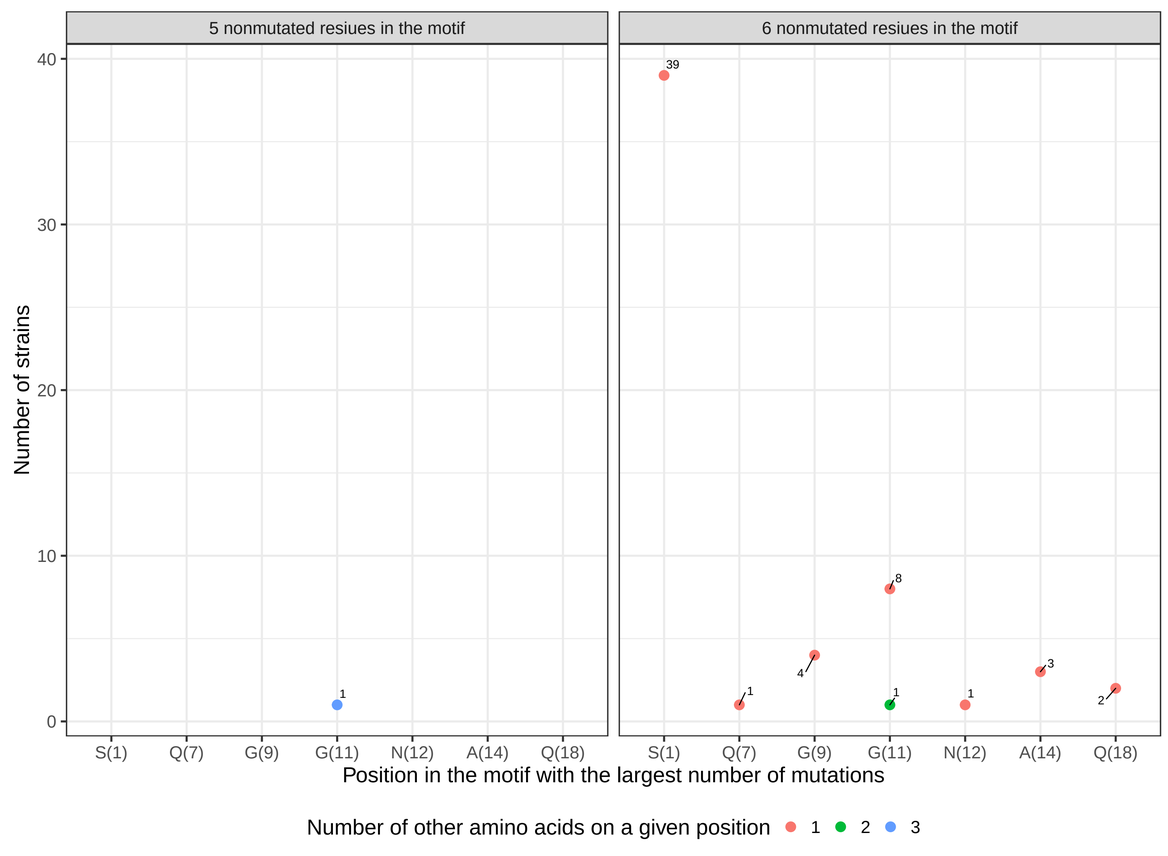
VI) Strongly hydrophilic and highly flexible (Aspartic Acid, Glutamic Acid, Asparagine, Glutamine, Serine, Threonine)

Results

Both CsgA and CsgB are characterized by a regional structure of five repeated motifs. We found out that the general motif (S-X5-Q-X-G-X2-N-X-A-X3-Q) (Evans and Chapman, 2014) (the serin in absent in the case of CsgB) is faithfully preserved among different variants of CsgA and CsgB. The residual variability in motifs of both proteins does not affect the sequence of other protein.

Aligned repeat motifs from CsgA and CsgB show that the variability of these sequences is insignificant. We can also see invariable places, which probably are responsible for protein amyloidogenic properties.

Stability of CsgA curli motif



Among 1877 investigated strains, only 60 (3.2%) had mutations in any repeat of the CsgA motif. The majority (59) of strains had only a single mutation, mostly in S (1). An insignificant fraction had more than 2 mutations. 1 strain (0.05%) had two mutations S (1).

Conclusions

The interplay of CsgA and CsgB suggests that if a mutation occurs in the region responsible for protein interaction, it should be compensated by mutations in other protein. We have not identified any simultaneous mutations between CsgA and CsgB. This may be due that single mutation in one region is not enough to change the protein function and to cause mutations in another protein. Probably, the compensation of single mutations by the regional structure of mentioned proteins is sufficient.

Bibliography

Burdukiewicz, M., Sobczyk, P., Rödiger, S., Duda-Madej, A., Mackiewicz, P., and Kotulska, M. (2017). Amyloidogenic motifs revealed by n-gram analysis. *Scientific Reports*, 7.

Evans, M. L. and Chapman, M. R. (2014). Curli biogenesis: Order out of disorder. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8):1551–1558.