

# CsgA, CsgB and CsgC - evolutionary interplay in curli biogenesis

Jarosław Chilimoniuk<sup>1</sup>, Paweł Mackiewicz<sup>1</sup> and Michał Burdukiewicz<sup>2\*</sup>

\*jaroslaw.chilimoniuk@gmail.com

<sup>1</sup>University of Wrocław, Department of Genomics, Wrocław, POLAND <sup>2</sup>Warsaw University of Technology, Warsaw, POLAND

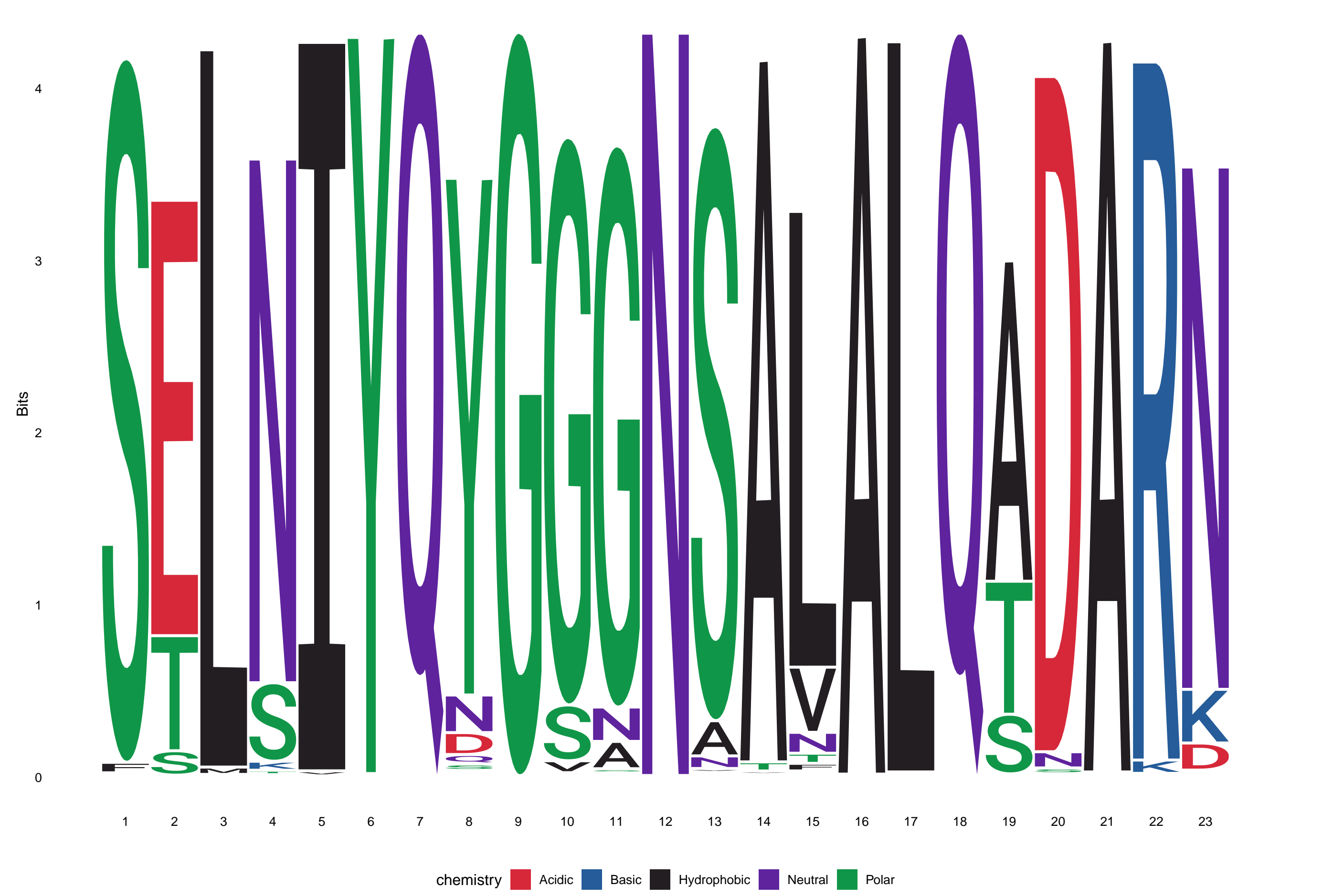
## Introduction

CsgA, also known as a major curli component, is a secreted protein ubiquitous in biofilms of gram-negative bacteria (Zhou et al., 2012a). 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. CsgC, a sequence-specific inhibitor, is able to arrest aggregation of CsgA in periplasmic space (Evans et al., 2015). 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 (Zhou et al., 2012b). CsgA can be in vivo cross-seeded by its nucleator protein, CsgB, but also other CsgA fibrils.

## Methods

We recently published AmyloGram (Burdukiewicz et al., 2017). It is an efficient tool for prediction of amyloid proteins as CsgA and CsgB. Among others, our algorithm produces a simplified amino acid alphabet for amyloid proteins. It groups amino acid together into six groups based on their physicochemical features. We combined sequences of around 500 triplets CsgA-CsgB-CsgC from closely related gram-negative bacteria and used EVmutation software (Hopf et al., 2017) to find out which residues are especially susceptible to simultaneous mutations.

## 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) (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. The opposite situation happens in CsgC, where mutations are not limited to specific motifs. On the other hand, CsgA and CsgC keep a similar level of the sequence divergence aside from species of Enterobacter and Salmonella genera, where CsgC accumulates more mutations.

## Conclusions

The interplay of CsgA, CsgB and CsgC suggests that if a mutation occurs in the region responsible for protein interaction, it should be compensated by mutations in other two proteins. For example, the presence of mutations that decrease the aggregation rate of CsgA, allows CsgC to become a less effective inhibitor without affecting the wellbeing of bacteria. Our preliminary study partially confirms this hypothesis by finding out that genera with less efficient CsgA, as Salmonella, can accumulate more mutations in the sequence of CsgC. 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.