In silico and *in vivo* combinatorial design of Octarellin VI, an artificial protein modeled on the (beta/alpha)8 fold

Maximiliano Figueroa[1], Sébastian Taralla[1]; Marco Buscetta[1]; Wilmer Sanjuan-Pacheco[1]; André Matagne[2]; Annabelle Lejeune[1]; Joseph A. Martial[1]; Cécile Van de Weerdt[1]

[1] GIGA-Research, Molecular Biology and Genetic Engineering Unit, University of Liège. Avenue de l'hopital 1, 4000 Liège, Belgium
[2] Centre of Protein Engineering, University of Liège. Institut de Chimie B6, Sart-Tilman, 4000 Liège, Belgium.

One way to gain insight into the sequence-structure-function relationship in proteins is to perform de novo design of artificial proteins. The applications of such a study are varied. For example, in medicine and industry, it would give us the ability to precisely engineer proteins to perform a specific function under a wider range of conditions. Despite impressive successes in the de novo protein design, designing a folded protein of more than 100 amino acids remains a challenge. In our lab, four generations of Octarellins, de novo polypeptides of more than two hundred amino acids modelled on the (beta/alpha)8 barrel fold, have been built and structurally characterized using biophysical and spectroscopic methods. The last generation of Octarellins was designed following a hierarchical method combining the specificity of rational design and the power of computational design. The resulting artificial protein, named Octarellin VI, was expressed in E. coli and purified from inclusion bodies. The biophysical characterization showed a monomeric protein, with a secondary structure level similar to the computationally designed model and thermostability. However, the poor solubility in bacteria and low stability of the protein at long term make impossible determine its structure to criticize the model. To improve these negative features, we performed a directed evolution process over the Octarellin, following the improvement at solubility level in the bacteria, thanks to the fusion of Octarellin to the fluorescent folding reporter GFP. After 8 cycles of directed evolution by Error Prone PCR technique, we obtained a most soluble protein, with a 92% of sequence identity with the original protein. This soluble variant is under study to characterize its structural features. The combination between in silico design and directed evolution process emerges as a powerful tool for protein engineering, showing be complementaries techniques and the information obtained by the whole process of design and posterior comparison between 3D structure of Octarellin with the computational model will allow to improve the algorithms for protein design.