

## **A pre-competitive consortium to share data and build predictive models for recombinant protein expression**

◊ Lovisa Holmberg Schiavone, ‡ Aurelie Bornot

◊ *Discovery Biology, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden*

‡ *Data Sciences & Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK*

### **Abstract**

Producing recombinant proteins is the first and rate-limiting step in most drug discovery projects. In AstraZeneca we design thousands of protein constructs every year using a part-rational/part-empirical process. We use experimental expression data at small scale to identify successful constructs to scale-up. The data capture affinity purified protein. This is quantified as mg/L expression culture and SDS-PAGE assessment of protein quality. Half of the targets requires two or more eight-week design/screening iterations. Using AI to increase the success rate of the first experimental iteration would have a huge impact on timelines and cost for protein delivery. Machine learning compatible AZ data is limited. Therefore, we are exploring sharing protein expression screen data in a pre-competitive consortium. We have initiated discussions via the EBI industry forum and will be part of an EBI-hosted workshop in October 2021. The ongoing MELLODDY project involving 10 large pharmaceutical companies (including AZ) joining forces to train AI for accelerating drug discovery has been a paradigm shift in the industry. Our proposal is building on this success and aim at federating large pharmaceutical companies and large academic institutions in a consortium. We suggest to build a platform for sharing pre-competitive experimental data and for harnessing the collective knowledge using ML algorithms.