MACHINE LEARNING TO ENGINEER ANTIBODY FRAMEWORKS FOR DEVELOPABILITY

Claes Gustafsson, ATUM
cgustafsson@atum.bio
Sridhar Govindarajan, ATUM
Mark Welch, ATUM
James Love, ATUM

Key Words: antibody engineering, developability, aggregation, thermostability, machine learning.

Monoclonal antibodies (mAbs) have revolutionized medicine in the last 20 years and today represents ~$70B/yr in total pharmaceutical sales, most notably in the areas of oncology and autoimmune disorders. The function and developability of mAbs depend on the expression, folding and integrity of their structure. Protein pharmaceuticals must be tolerant to factors such as heat, interfacial stress, aggregation, pH and more in order to reach the market. We here apply systematic variance, inductive machine learning, synthetic genes and our high-throughput transient mammalian protein expression platform to engineer a humanized IgG1 scaffold for high developability independent of the hypervariable region present in the mAb.

All amino acid substitutions present in the framework of human IgG1 antibodies were derived from human sequences in the public domain databases and assembled in a set of 96 partial factorial IgG1 variants (aka 'infologs') using Design of Experiment (DoE) variable distribution. Total explored sequence diversity was ~2x1019. Hypervariable regions were derived from two commercial antibodies for a total of 2x96 genetic constructs. Synthesis of the 2x96 antibodies was done by transient transfection in HEK293 cells and purified in high throughput. Several independent machine learning algorithms were compared for cross validation and model accuracy and used to build iterative sequence-function correlation models to identify and quantify independent and/or synergistic variables affecting one or more of the developability functionalities [1]. The study resulted in markedly improved mAbs frameworks as well as a deeper understanding on how different machine learning algorithms are dependent on different types of data sets.

[1] Issued US patents 8635029, 8412461, 8005620 and related pending applications.

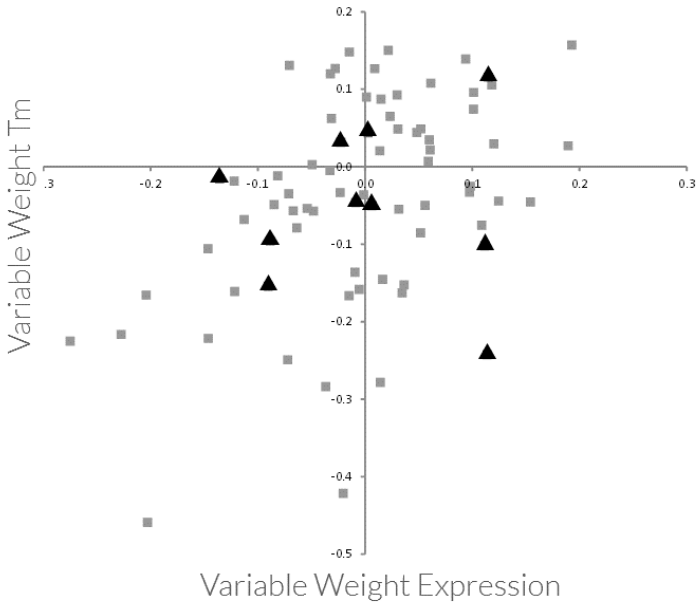

Figure 1 – Variables in mAb frameworks distributed by their quantitative weight in 1) Expression yield (X dimension), 2) Thermostability (Y dimension), and 3) Affinity (square vs triangle).