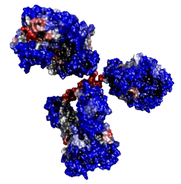
Application Of Machine Learning In Drug Design

Leo Bleicher[†], Ton van Daelen[†], Moises Hassan[†], Dana Honeycutt[†], Kevin Merlo[*], Matthew Sage[†], Adrian Stevens[‡]

[†]Dassault Systemes BIOVIA Corp., 5005 Wateridge Vista Drive, San Diego, CA 92121, United States, [‡]Dassault Systèmes BIOVIA Ltd, 334/335 Cambridge Science Park Milton Road, CB4 0WN Cambridge, United Kingdom, [*]Dassault Systèmes SE | Rue Evariste Galois, Immeuble Emerald Square D, CS60354, 06906 Sophia Antipolis Biot, France



It can take four years, and up to *ca.* 4,000 antibodies to find a lead candidate. What if we could reduce this timeline and express only a fraction of these proteins?

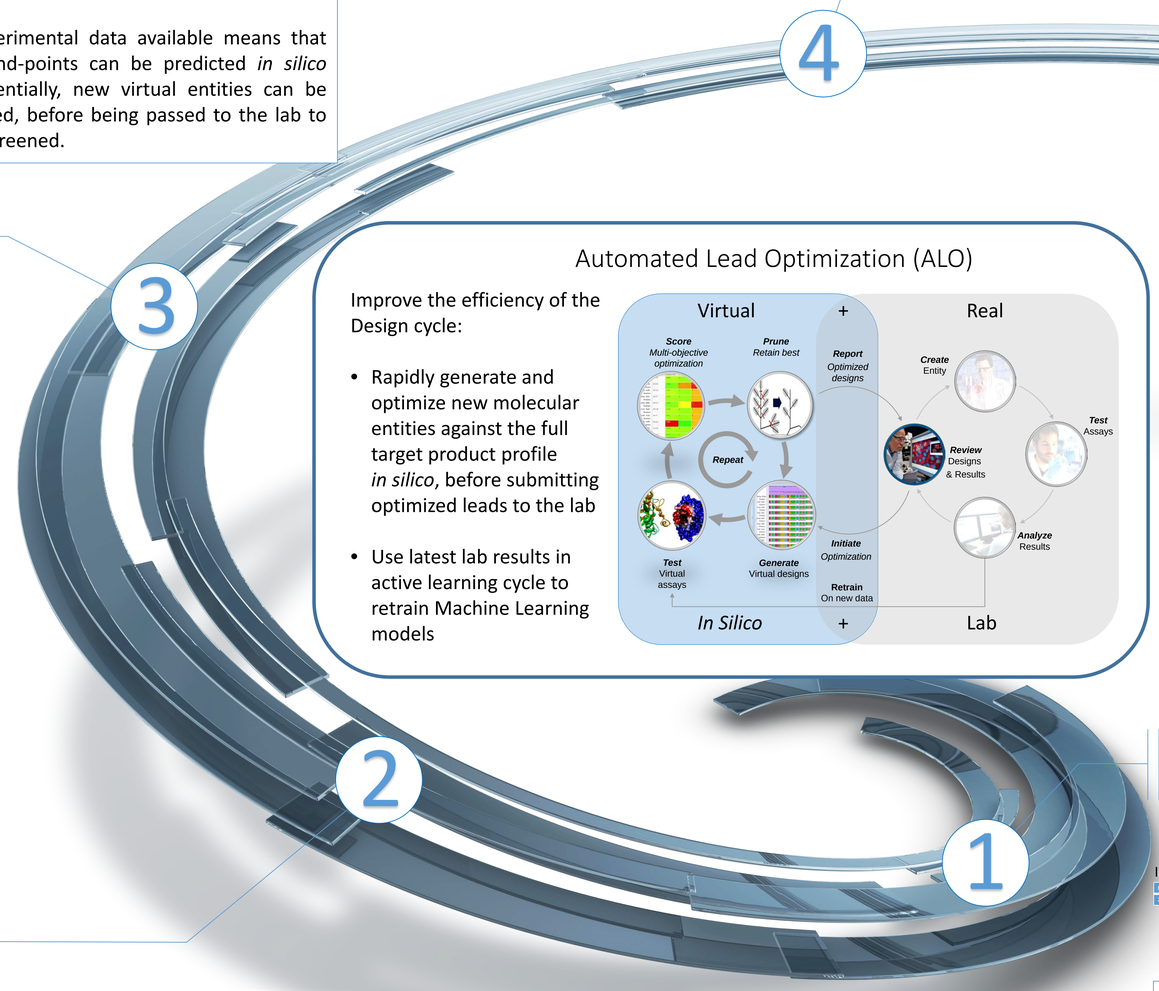
The Innovation Cycle

Design | Make | Test | Analyze

Machine Learning (ML) and Artificial Intelligence (AI) are transforming scientific innovation

Scientific innovation process is an experimentally-driven Design-Make-Test-Analyze cycle. Its success rate is driven by making the best informed design decisions on '*what to make next?*'

The wealth of historical experimental data available means that many biologically relevant end-points can be predicted *in silico* using machine learning. Potentially, new virtual entities can be designed and rapidly optimized, before being passed to the lab to be physically expressed and screened.



1 Generative Lead Mutations

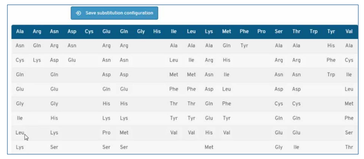
Each cycle begins with either experimental hits and leads, known drugs or the top members of a previous ALO cycle. From these sequences, novel virtual analogs are generated and can be either relatively similar or highly novel, depending on the level of desired diversity.

Input Sequence(s)

Name	Identity	Similarity	Start	End	Sequence
2FJF_L_VL_KAPPA_VDenv_1_95	100	100	1	95	DIQMTQSPSSLSASVGDRVTITCRASQDVSTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSYTTP

- Input sequences are compared to collection of known Human Abs
- Regions for modification identified, avoiding CDRs, with potential to enhance Ab properties
- Amino Acid Substitution Matrix is applied at suggested locales, generating Output Sequences

Substitution Matrix



Output Sequences

Molecule Name	Transformation
2FJF_LH_FAB_LH_FV:210ASP	H:LEU111>ASP
2FJF_LH_FAB_LH_FV:208GLU	H:PHE109>GLU
2FJF_LH_FAB_LH_FV:53GLU	L:PHE66>GLU
2FJF_LH_FAB_LH_FV:53LYS	L:PHE66>LYS
2FJF_LH_FAB_LH_FV:207ARG	H:VAL108>ARG
2FJF_LH_FAB_LH_FV:50ARG	L:SER56>ARG
2FJF_LH_FAB_LH_FV:159ARG	H:THR57>ARG
2FJF_LH_FAB_LH_FV:210ARG	H:LEU111>ARG
2FJF_LH_FAB_LH_FV:208ASP	H:PHE109>ASP
2FJF_LH_FAB_LH_FV:209ASP	H:PHE110>ASP

2 Predicting Properties Associated with the Target Product Profile

Using available experimental data, Machine Learning (ML) algorithms and molecular simulations are applied to generate predictive models.

- Use a set of models to assess each virtual library member relative to the target product profile
- Rapidly assess the suitability of each member of a virtual generation before physically making and testing
- Dynamic: Models are not static. As new experimental data is generated, models are retrained and updated

Sequence-based
- Molecular weight
- Isoelectric point
- Charge
- hydrophilicity & hydrophobicity

3D Structure-based
- Aggregation
- Developability Index
- Binding Affinity
- Mutation Energy

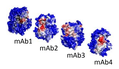
Molecule Name	Transformation
2FJF_LH_FAB_LH_FV:210ASP	H:LEU111>ASP
2FJF_LH_FAB_LH_FV:208GLU	H:PHE109>GLU
2FJF_LH_FAB_LH_FV:53GLU	L:PHE66>GLU
2FJF_LH_FAB_LH_FV:53LYS	L:PHE66>LYS

Calculate sequence-based Properties

mAb	Mutation	E (stability)	Effect	Solubility	Aggregation	Half-Life
1	L:Phe66>Glu	-0.06	Neutral	423	-522	Same
2	Phe66>Lys	0.32	Neutral	420	-523	Same
3	L:Wild	0	Neutral	407	-480	-
4	L:Ser56>Arg	0.56	Neutral	422	-555	Same

Generate Homology Models

Calculate Structure-based Properties



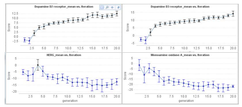
3 Optimization: Scoring and Ranking

Each member of a virtual generation is scored according to how closely the predicted properties meet the target product profile requirements.

Score is assessed either by:
- Multi-Objective Optimization (MOO) methods such as Pareto-based optimization
- or by transforming the MOO into a Single Objective Optimization (SOO) problem using approaches such as Weighted Sum or Derringer Desirability

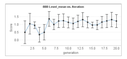
Automated Molecular Optimization: Multi-Objective Progress

Target activities Increasing, Balanced

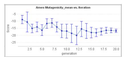


Anti-target activities Decreasing

ADME property Improving



Toxicity Decreasing



Automated Lead Optimization (ALO)

Improve the efficiency of the Design cycle:

- Rapidly generate and optimize new molecular entities against the full target product profile *in silico*, before submitting optimized leads to the lab
- Use latest lab results in active learning cycle to retrain Machine Learning models

4 Pruning and Evolving new Generations

At the end of each enumeration cycle, the virtual members are ranked and the most optimal entities of the population retained.

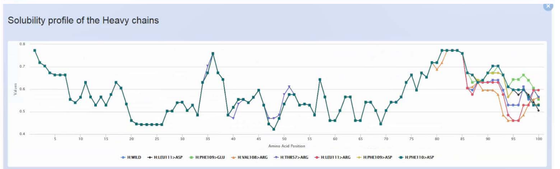
The optimization cycle ends if:
- The target product profile has been met
- The maximum number of iterations has been completed or
- No further improvement can be achieved

Otherwise, the top entities are used as starting points for a next virtual generation of new entities and optimization

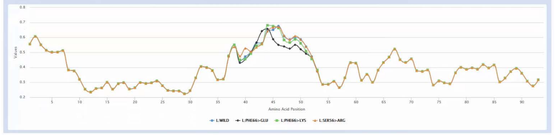
Best candidates

Molecule Name	Transformation	Aggregation	Mutation Energy Stability	EFFECT	VDW	ELEC	ENTROPY	Objective
2FJF_LH_FAB_LH_FV:210ASP	H:LEU111>ASP	-575.639954	0.76	NEUTRAL	1.16	0.46	-0.06	-13.386
2FJF_LH_FAB_LH_FV:208GLU	H:PHE109>GLU	-577.852722	2.03	DESTABILIZING	3.54	0.91	-0.25	-13.315
2FJF_LH_FAB_LH_FV:53GLU	L:PHE66>GLU	-568.759521	0.68	NEUTRAL	-1.15	1.86	0.41	-13.233
2FJF_LH_FAB_LH_FV:53LYS	L:PHE66>LYS	-553.424255	-0.16	NEUTRAL	-1.77	0.38	0.67	-12.956
2FJF_LH_FAB_LH_FV:207ARG	H:VAL108>ARG	-532.901123	-0.17	NEUTRAL	-3.75	0.74	1.67	-12.477
2FJF_LH_FAB_LH_FV:50ARG	L:SER56>ARG	-522.823792	-1.19	STABILIZING	-5.09	0.82	1.18	-12.341
2FJF_LH_FAB_LH_FV:159ARG	H:THR57>ARG	-523.671753	-0.88	STABILIZING	-5.27	1.8	1.07	-12.33
2FJF_LH_FAB_LH_FV:210ARG	H:LEU111>ARG	-579.764709	0.63	NEUTRAL	0.3	0.62	0.21	-12.097
2FJF_LH_FAB_LH_FV:208ASP	H:PHE109>ASP	-582.604126	1.83	NEUTRAL	2.8	1.36	-0.31	-12.037
2FJF_LH_FAB_LH_FV:209ASP	H:PHE110>ASP	-570.168152	-0.05	NEUTRAL	-0.76	0.65	0.01	-11.963

Solubility profile of the Heavy chains



Immunogenicity Profile of the Light chains



DASSAULT SYSTEMES | The 3DEXPERIENCE® Company