

Screenshots der Präsentation vom 16.07.2024

# YOUR GMP "READY" PROTEIN

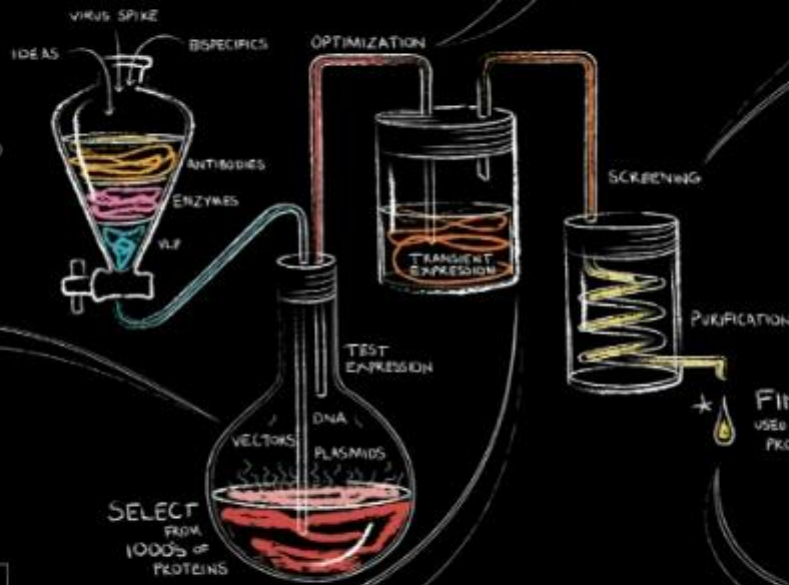
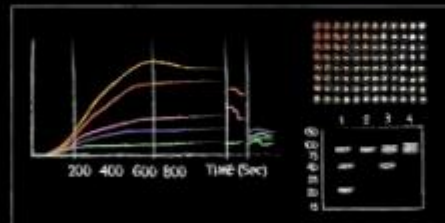
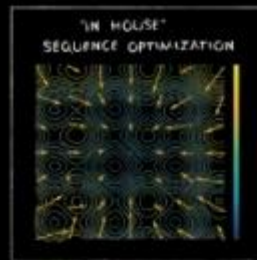
COLLABORATIVE DISCOVERY  
WHAT IS THE PROTEIN?  
WHAT INFO IS AVAILABLE?  
WHAT IS THE USE?  
WHAT QUANTITY IS REQUIRED?

DGE EXPRESSIONS TO IDENTIFY  
OPTIMUM HOST SYSTEM AND  
UPSTREAM CRITICAL PARAMETERS

DOWNSTREAM  
PROCESS  
OPTIMIZATION

CLARIFICATION  
ULTRA/DIA  
FILTRATION  
VIRAL INACTIVATION?  
BIND/ELUTE  
VIRAL INACTIVATION?  
POLISH  
VIRAL FILTRATION?  
ULTRA/DIA  
FILTRATE  
FORMULATE

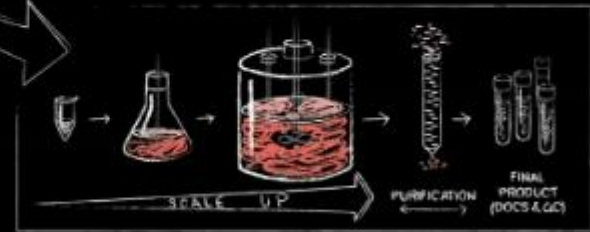
DOWNSTREAM  
PROCESS DEVELOPMENT  
IS AN ITERATIVE PROCESS



FINAL PROTEIN

USED IN:  
PROCESS DISCOVERY ASSAY DEVELOPMENT  
PHASE 1 PHASE 2 PHASE 3  
CLINICAL DEVELOPMENT FOR FUTURE

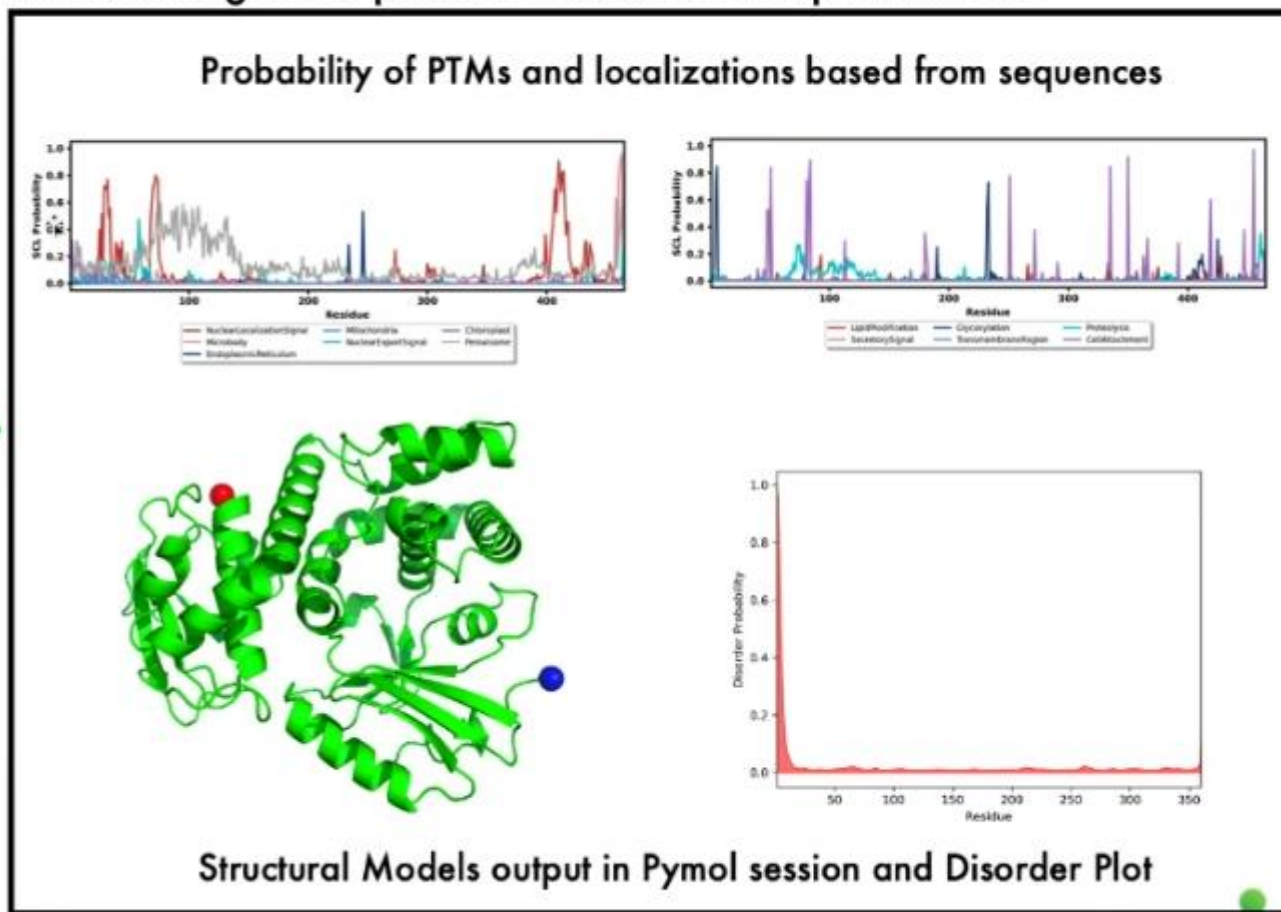
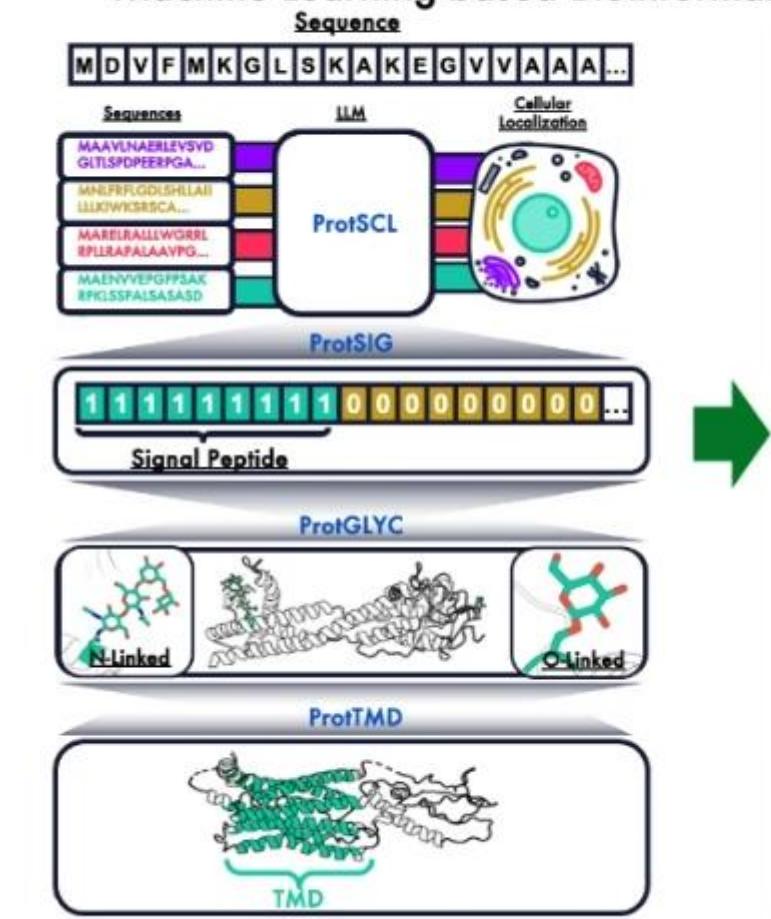
SUPPORT DOSSIER FROM  
KEMP PROTEINS



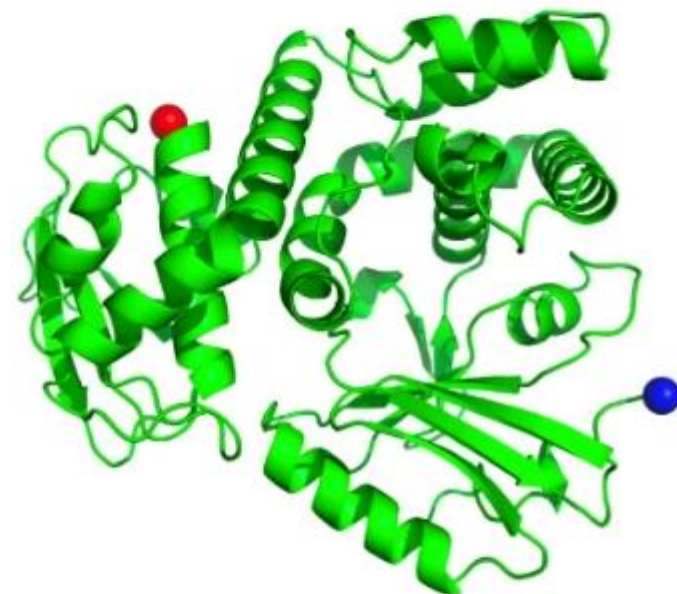
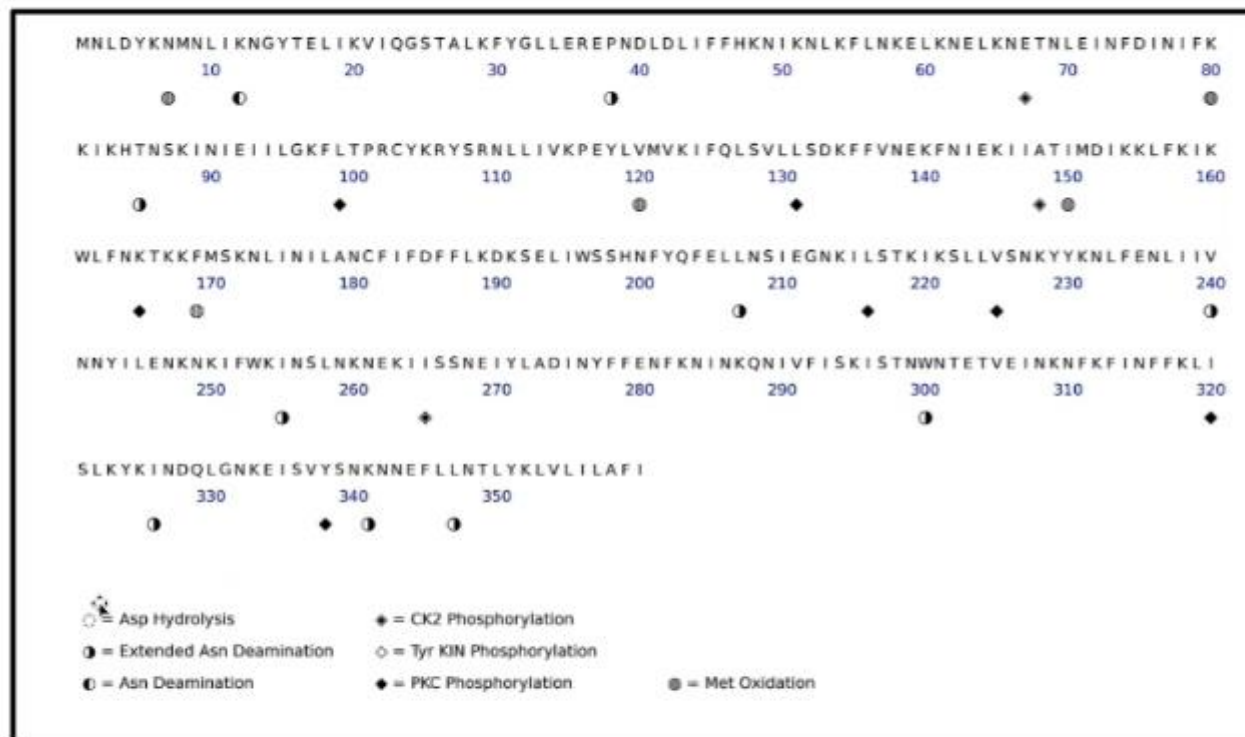
**KEMP**  
PROTEINS  
QUALITY EXPRESSED

# PROTiQ<sup>SM</sup>: Novel Generalizable ML Tools

Machine Learning based Bioinformatic Tools designed to predict features from sequence alone



# PROTiQ<sup>SM</sup>: Sequence Liabilities Prediction



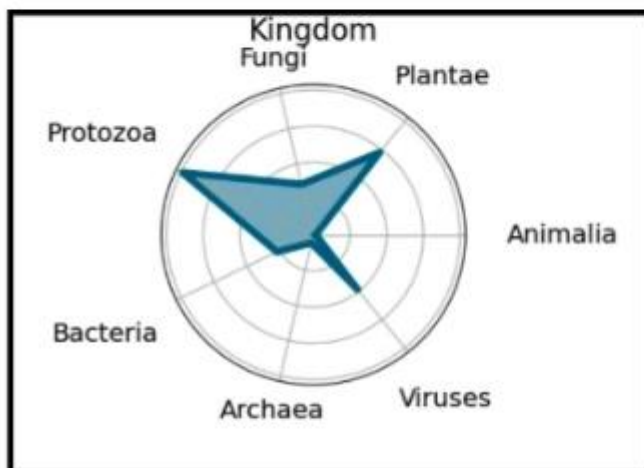
RoseTTAFold Model + Gromacs  
Energy Minimization

*The sequence is processed through the generalizable ML tools to identify and sequences liabilities as a function of recombinant expression host. Not shown here are proteolytic liabilities as a function of recombinant host*

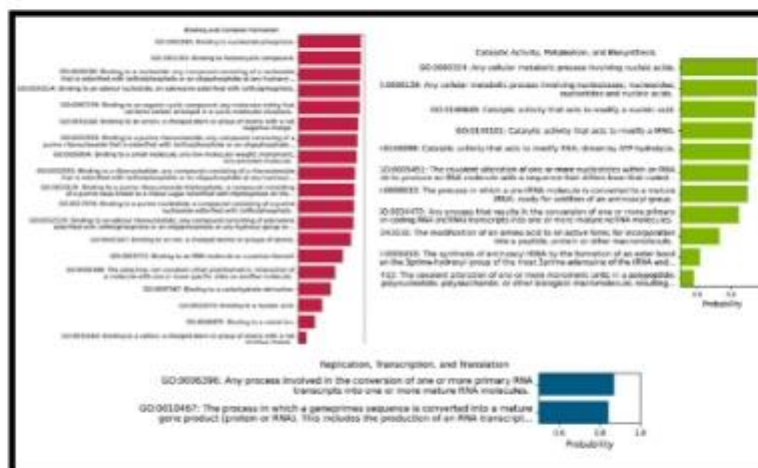


# PROTiQ<sup>SM</sup>: FAUNA

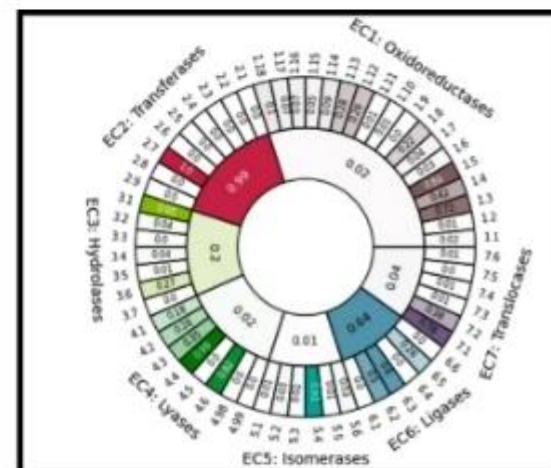
## Kingdom Prediction & GO / Enzyme Classifiers



ProtFAUNA Kingdom - assigns probabilities that a given sequence belongs to a Kingdom





ProtFAUNA GO - assigns probabilities that a given sequence satisfies GO terms



ProtFAUNA EC - assigns probabilities that a given sequence belongs to an Enzyme Commission

# PROTiQ<sup>SM</sup>: FAUNA Report

	
	
Report Date: 22 March 2024	
Client: FNL – Alpha-Twist and Beta-Turns	
Client Contact: Carter A. Mitchell	
<p><b>About PROTiQ<sup>SM</sup></b> Decades of research enabled the creation of extensive databases that provide a about the structure and action of various proteins. We have localized this information learning process to combine this information with knowledge from our own experts, provide our clients with a cutting edge system for designing proteins that may challenges during initial protein development. This system PROTiQ<sup>SM</sup> provides a design document between Client and Kemp Proteins regarding a Protein Development</p> <p>Following initial discussions and review of the data provided by PROTiQ<sup>SM</sup> our clients Services Agreement as part of the project initiation and expression of the protein sequence</p> <p>Kemp Proteins <a href="mailto:ProtiQ@kempproteins.com">ProtiQ@kempproteins.com</a> 240-829-8804</p> <p>PROTiQ<sup>SM</sup> is owned and trademarked by Kemp Proteins. It was built in strategic Licensing and AI consulting opportunities exist for PROTiQ<sup>SM</sup>, contact us if you want an arrangement.</p> <p>The information in this report is for research purposes only and it is understood by the recipient experimental in nature and as such Kemp Proteins makes no representation or warranty, or the reliability of the protein sequences generated, nor does Kemp Proteins guarantee that report may contain links to third-party websites or content and Kemp Proteins does not party websites, products, or services. All content on this report is the property of Kemp Proteins of Kemp Proteins until such time that a client enters into commercial arrangements for right of the information contained in this report agree to indemnify and hold harmless Kemp Proteins from use of the information contained in this report.</p>	
1	Table of Contents
Contents	
1	Table of Contents
2	Client Sequence
3	Executive Summary
4	ProtFauna Analysis
5	Potential Liabilities
5.1	Asn Deamination Site
5.2	Extended Asn Deamination Site
5.3	DP hydrolysis Site
5.4	PKC Phosphorylation Site
5.5	Tyrosine Kinase Phosphorylation Site
5.6	CK2 Phosphorylation Site
5.7	Met Oxidation Site
6	Subcellular Localization Prediction
7	Structural Analysis Blast
8	Structural Analysis Folded Models
9	Disordered Prediction
10	Protease Prediction
10.1	Proteolysis Liability in Homo sapiens
10.2	Sequence Fragments in Homo sapiens
10.3	Proteolysis Liability in Cricetulus griseus
10.4	Sequence Fragments in Cricetulus griseus
10.5	Proteolysis Liability in Trichoplusia ni
10.6	Sequence Fragments in Trichoplusia ni
10.7	Proteolysis Liability in Spodoptera frugiperda
10.8	Sequence Fragments in Spodoptera frugiperda
10.9	Proteolysis Liability in Bacillus subtilis
10.10	Sequence Fragments in Bacillus subtilis
10.11	Proteolysis Liability in Saccharomyces cerevisiae
10.12	Sequence Fragments in Saccharomyces cerevisiae
10.13	Proteolysis Liability in Pichia pastoris
10.14	Sequence Fragments in Pichia pastoris
10.15	Proteolysis Liability in Escherichia coli
10.16	Sequence Fragments in Escherichia coli
10.17	Proteolysis Liability in Laboratory
10.18	Sequence Fragments in Laboratory

## Data input into a LLM Summarizer:

### FAUNA EC Final Analysis:

" ... the evidence isn't strong enough to confidently classify any section of this sequence as part of an enzymatic domain based on the traditional threshold."

### FAUNA GO Final Analysis:

" ... the protein seems to be crucial for RNA metabolism and protein synthesis, likely localizing to the cytoplasm or specific organelles like the mitochondria where these processes occur."

### FAUNA Host Suggestion:

#### Choice of Organism

- *E. coli*, *Saccharomyces cerevisiae*, and *Pichia pastoris* might be preferred due to simple PTM patterns and potentially lower maintenance costs. –

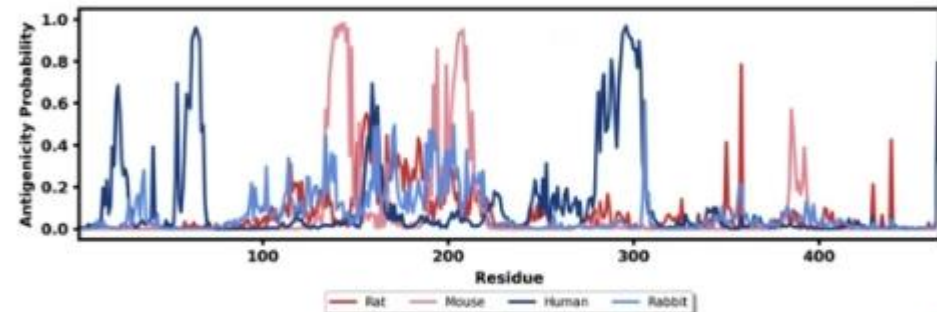
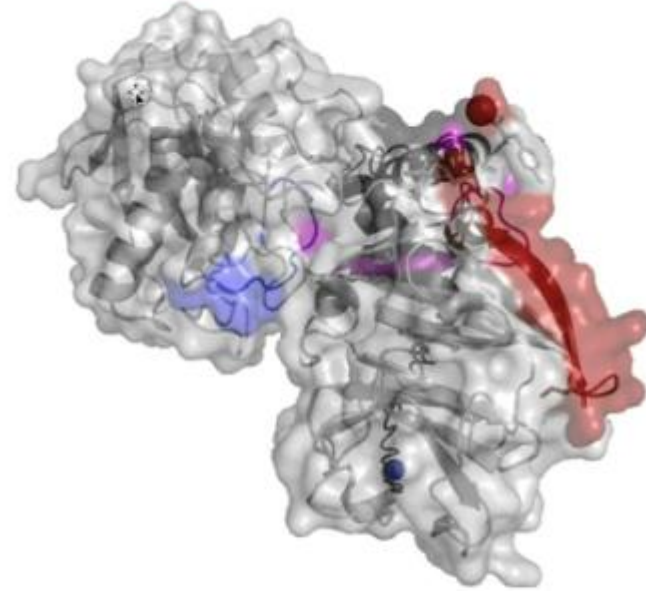


# PROTiQ<sup>SM</sup>: Antigenicity Scoring Function

*The Antigenicity Score is generated from analysis of known interactions taken from the literature and curated databases that include structural information when available.*

**Current rules include:**

- >10 residues < 40 residues
- Sequence Identity  $\geq 55\%$ ; Similarity  $\geq 65\%$
- Solvent accessible, noninteraction regions
- Identification of preferred immunization animal as a function of sequence homology including
  - 1) Mouse
  - 2) Rat
  - 3) Human
  - 4) Rabbit
  - 5) Camelid





# Antibody Expression at Kemp in Various Platforms



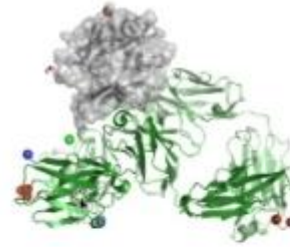
- **Full-Length mAB**
  - Grafted
  - Isotype Switching
  - Variables provided



- **IgM**
  - Pentameric
  - Hexameric



- **VHH**
  - Bacterial
  - Mammalian



- **BiS**
  - KiH
  - Reductive Pairing
  - Common Light
  - DVD-IgG, Charge Repulsion
  - Cross mAB, etc



- **scFv**
  - Bacterial
  - Mammalian
  - Refolded if needed



- **Fc-Fusion**
  - Common engineering mode



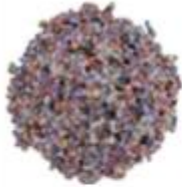
# Proteins Expressed at Kemp in Various Platforms

- Antibody Class of Molecule



- Full-Length - AB Fragments
- Multi-specific Antibodies
- Nanobodies – VHH & scFv
- IgM

- Virus Related Proteins and Particles



- Viral Glycoproteins
- Virus-Like Particles
- Nanoparticles

- Multi-protein Complexes

- E3 ligases with targets

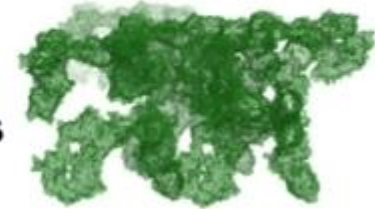


- Toxins

- Ricin, Botox, CtXB

- Membrane Proteins

- GPCRs
- Extracellular Domains
- Soluble TM mutants



- Immunologic Proteins / Hormones

- Hormones
- Interleukins
  - IL27, IL23, IL13, IL15



- Enzymes

- DNA pol
- Methyltransferases



- Cytoskeletal Elements

- Dystrophin SILAC / unlabeled<sup>1</sup>



# Rapid Small Scale Expression & Purification

DNA Synthesis

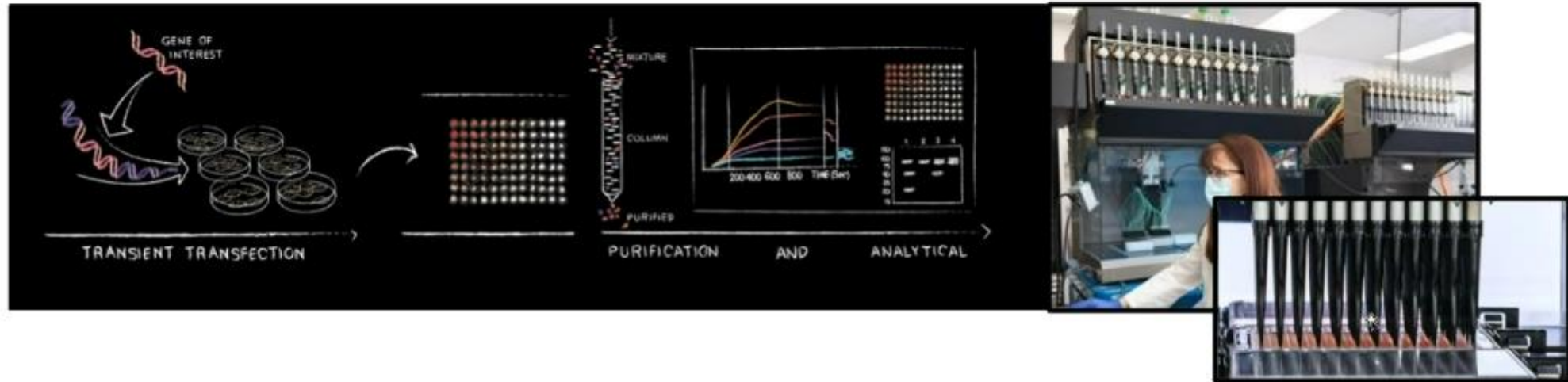
Plasmid

Expression

Purification

QC

QA



## ML Derived Sequence Assessment

- 24-96 purifications from 0.5 mL - >1L
- Flexible expression & purification modalities
- Quick TAT ~15-20 days to get 100's protein variants and associated data (plasmids provided)

## Final Deliverables:

- Go / No-Go on Constructs
- Customizable QC packages to feed into ML models
- ng to >100 mg of purified protein buffer exchanged