DOE Bioenergy Technologies Office (BETO) 2021 Project Peer Review

Integration of ABF Informatic Modules with TeselaGen's BioCAD/CAM Platform
and
Evaluation of Emerging TeselaGen Functionality Support for ABF Workflows

Wednesday March 10, 2021

Agile BioFoundry Consortium

Michael Fero

TeselaGen Biotechnology, Inc.

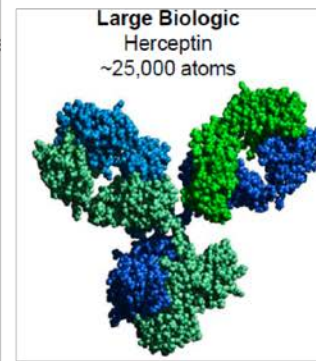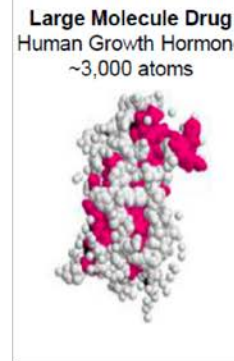# Project Overview

problem...

Biotechnology is one of the most important technologies for building a healthy sustainable future

FUTURE FABRICS   FUTURE FOOD   FUTURE PHARMA   FUTURE FUELS

Large Molecule Drug
Human Growth Hormone
~3,000 atoms

Large Biologic
Herceptin
~25,000 atoms

But biotech product development is fragmented, heterogeneous and bespoke

NOT COLLABORATIVE   LOW PRODUCTIVITY   ERROR PRONE

Agile BioFoundry
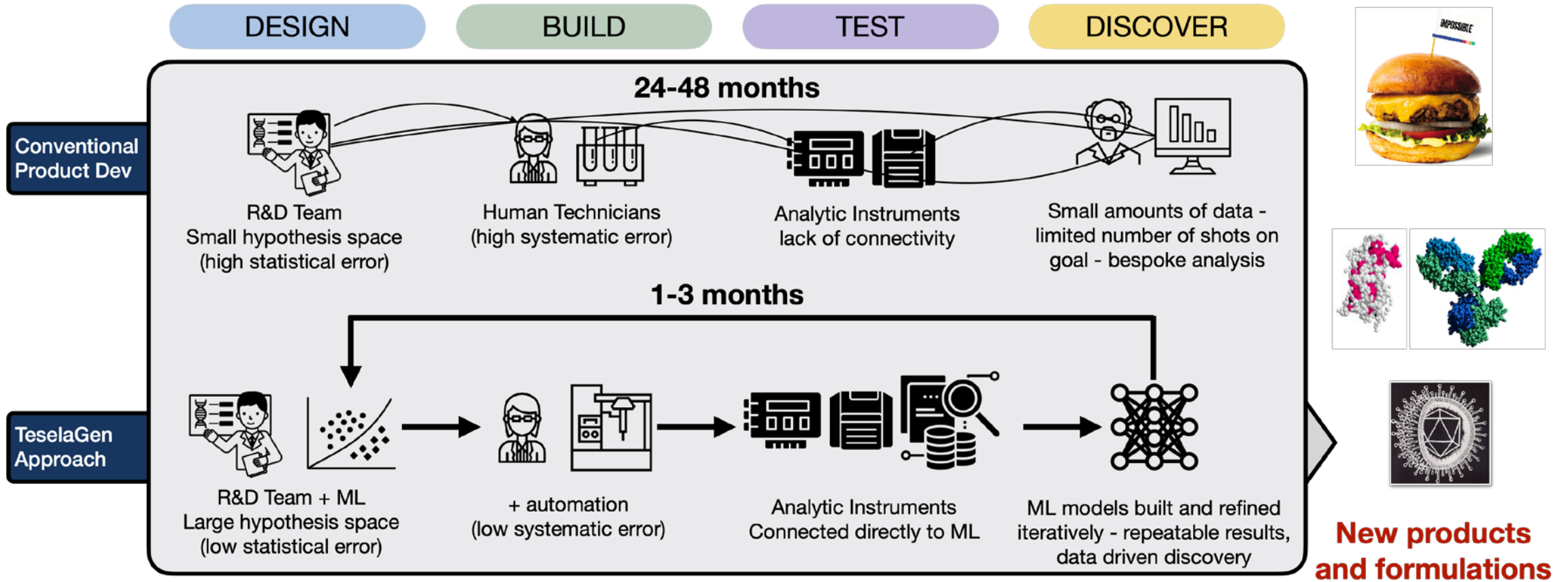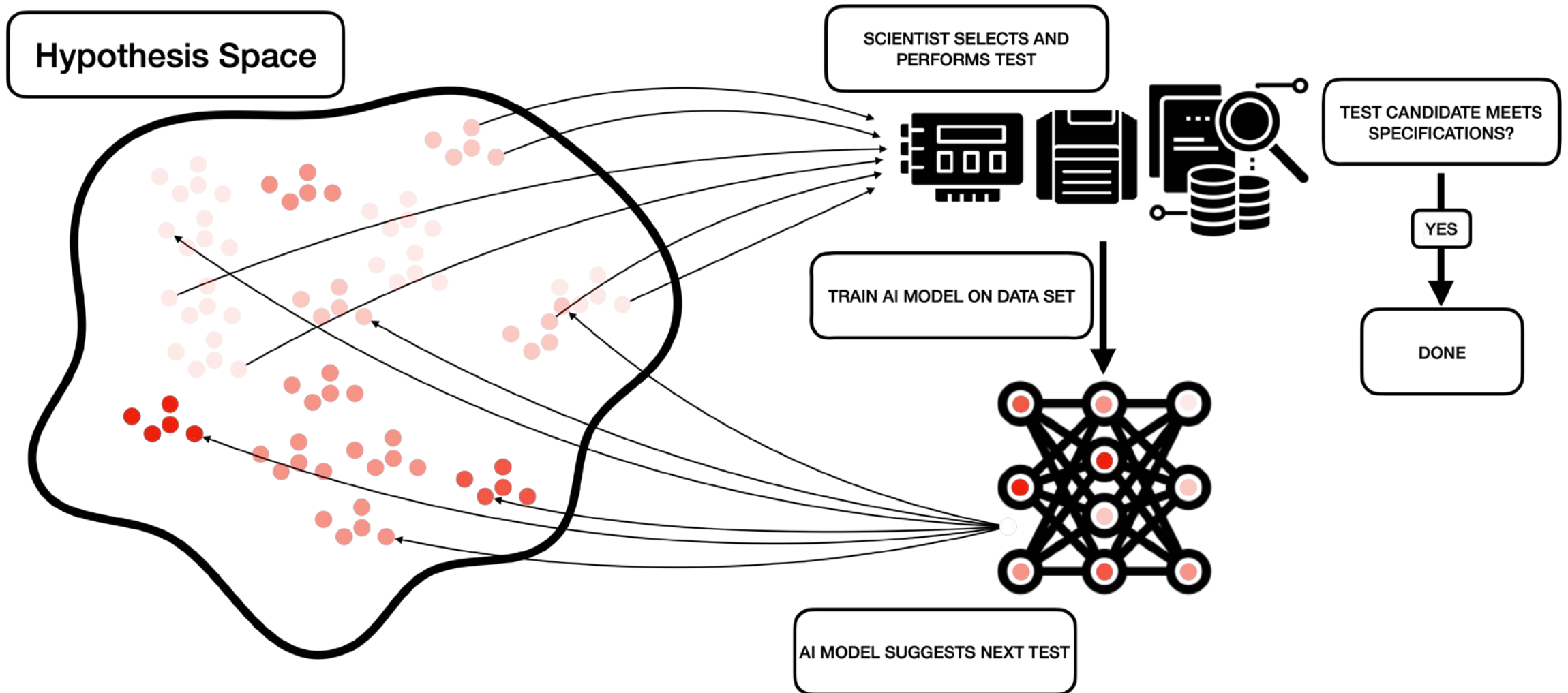
Rebuilding the product development and formulation framework as an AI platform for biotechnology
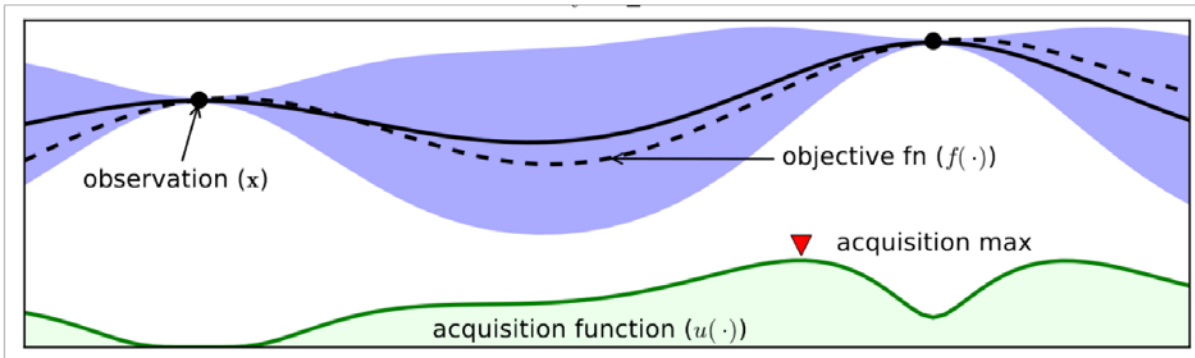

A unified data driven discovery platform, using machine learning to uncover and optimize bioproducts

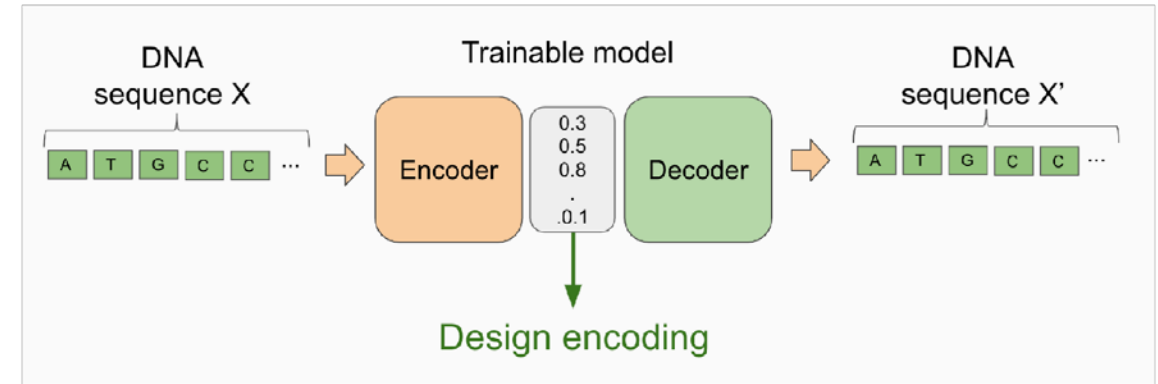# Synthetic Evolution® iterative model building...

…using AI techniques

Proprietary framework to evolve biological systems through an iterative and cost-effective manner (patent pending)

observation (x)
objective fn ($f(\cdot)$)
acquisition max
acquisition function ($u(\cdot)$)

We have developed a mathematical framework based on deep learning and bayesian optimization to predict biological outcomes and generate experimental recommendations to evolve biological systems.

Proprietary encoding of the search space using transfer learning techniques, to effectively train predictive models (patent pending)

DNA sequence X
A T G C C ...
Trainable model
Encoder
0.3
0.5
0.8
.
.0.1
Decoder
DNA sequence X'
A T G C C ...
Design encoding

We encode biochemical compounds (small molecules, DNA, proteins) using proprietary embeddings trained with large datasets., empowering us to search and optimize these compounds within their corresponding the design space.

2 patents granted, 4 patent applications in the pipeline

Agile BioFoundry

secure scalable solutions… built for automation

THIS COOPERATIVE RESEARCH AGREEMENT

THE COOLABORATORS

THE TOOLS

BOOST — Build Optimization Software Tools (optimizes sequence building through codon juggling and polishing)

BLiSS — Black List Sequence Screening (screens sequences for biosecurity issues.

EDD — Experiment Data Depot (stores data for use with modeling, machine learning and data mining)

Synthetic Evolution® — TeselaGen's Synthetic Evolution Platform (AI-enabled operating system for biotechnology)

THIS COOPERATIVE RESEARCH AGREEMENT

teselagen
BIOTECHNOLOGY

THE COOLABORATORS	THE TOOLS	THE GOALS

THE COOLABORATORS

BERKELEY LAB

teselagen
BIOTECHNOLOGY

Pacific Northwest
NATIONAL LABORATORY

Sandia
National
Laboratories

THE TOOLS

BOOST

BLiSS

EDD

Synthetic Evolution®

THE GOALS

1. Enhance selected ABF design tools (BOOST and BLISS) and integrate them with TeselaGen's platform FY19 Q2

2. Integrate Experiment Data Depot and analytical workflows into TeselaGen's platform FY19 Q4

3. Demonstrate ABF functionality integrated into TeselaGen's platform with industrial partners FY20 Q2

4. Evaluation of Emerging TeselaGen Functionality Support for ABF Workflows FY20 Q1

5. Final Report FY20 Q2

1 – Management
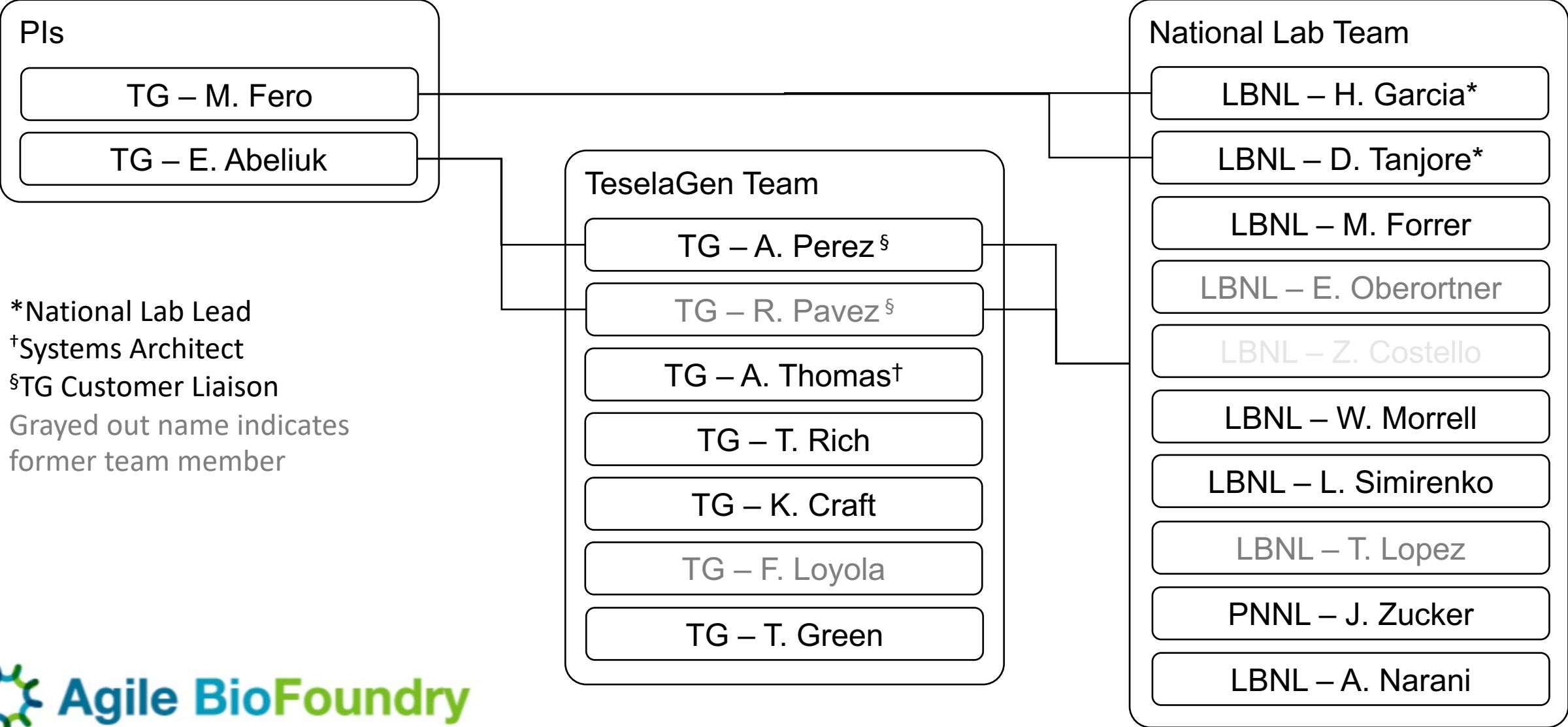
management... tasks & goals

No.	Task	Goal
1	Enhance selected ABF design tools and integrate with TeselaGen	Enhance BOOST modularity and API to allow ease of integration with commercial platforms.
		Enhance user ability to modulate between functionally driven segmentation of a DNA construct and synthesis optimal segmentation. This may include incorporation of TeselaGen HDE (Hierarchical Design Editor).
		Improve API to allow direct integration of BLISS with ABF workflows and design platforms.
2	Integrate experiment data depot and analytical workflows into TeselaGen's platform	Advance the EDD API to allow integration with TeselaGen platform.
		Develop analytical tools, statistical algorithms and visualization tools that will facilitate search, and provide exploratory data analysis of data in the EDD.
		Provide assistance in developing output formats and queries that facilitate the generation of data sets suitable for both mechanistic modelling and machine learning.
		Work to integrate and support PNNL analytical workflows/data with EDD (e.g. metabolomics/proteomics/flux analysis)
3	Demonstrate ABF functionality integrated into TeselaGen's platform with industrial partners	Develop commercial user stories and requirements relevant to proposed enhancements.
		Deploy at least one enhancement at one major industrial partner. Evaluate enhancement for improved productivity.
4	Evaluation of Emerging TeselaGen Functionality Support for ABF Workflows	Evaluate the effectiveness of HDE for facilitating and speeding up ABF workflows.
		Evaluate the effectiveness of TeselaGen LIMS to facilitating and speeding up ABF workflows.
E	Extension to test data acquisition and analysis capabilities against ABPDU fermentation data	Evaluate the effectiveness of TeselaGen BUILD/TEST modules for collection and analysis of fermentation data.

management... tasks & team

No.	Task	Responsible Parties
1	Enhance selected ABF design tools and integrate with TeselaGen	TeselaGen, LBNL
2	Integrate experiment data depot and analytical workflows into TeselaGen's platform	TeselaGen, NTESS, PNLL
3	Demonstrate ABF functionality integrated into TeselaGen's platform with industrial partners	TeselaGen, LBNL, NTESS, PNLL
4	Evaluation of Emerging TeselaGen Functionality Support for ABF Workflows	TeselaGen, LBNL, NTESS, PNLL
E	Extension to test data acquisition and analysis capabilities against ABPDU fermentation data	TeselaGen, ABPDU

Progress tracked through weekly or bi-weekly meetings

Well organized meetings due to Mark Forrer (LBNL) diligence

12

management... organization and accountability

teselagen
BIOTECHNOLOGY

PIs

TG – M. Fero
TG – E. Abeliuk

*National Lab Lead
†Systems Architect
§TG Customer Liaison
Grayed out name indicates
former team member

National Lab Team

LBNL – H. Garcia*
LBNL – D. Tanjore*
LBNL – M. Forrer
LBNL – E. Oberortner
LBNL – Z. Costello
LBNL – W. Morrell
LBNL – L. Simirenko
LBNL – T. Lopez
PNNL – J. Zucker
LBNL – A. Narani

TeselaGen Team

TG – A. Perez §
TG – R. Pavez §
TG – A. Thomas†
TG – T. Rich
TG – K. Craft
TG – F. Loyola
TG – T. Green

Agile BioFoundry

2 – Approach

technical approach, background...

Experiment Data Depot

TeselaGen Design/Build/Test modules

TeselaGen's TEST module and parts of BUILD and DISCOVER map to certain features in EDD
- Each has elements missing from the other
- Might it not make sense to standardize the feature set, and more importantly the interfaces?

technical approach, background…

With respect to EDD's analog TEST in TeselaGen:
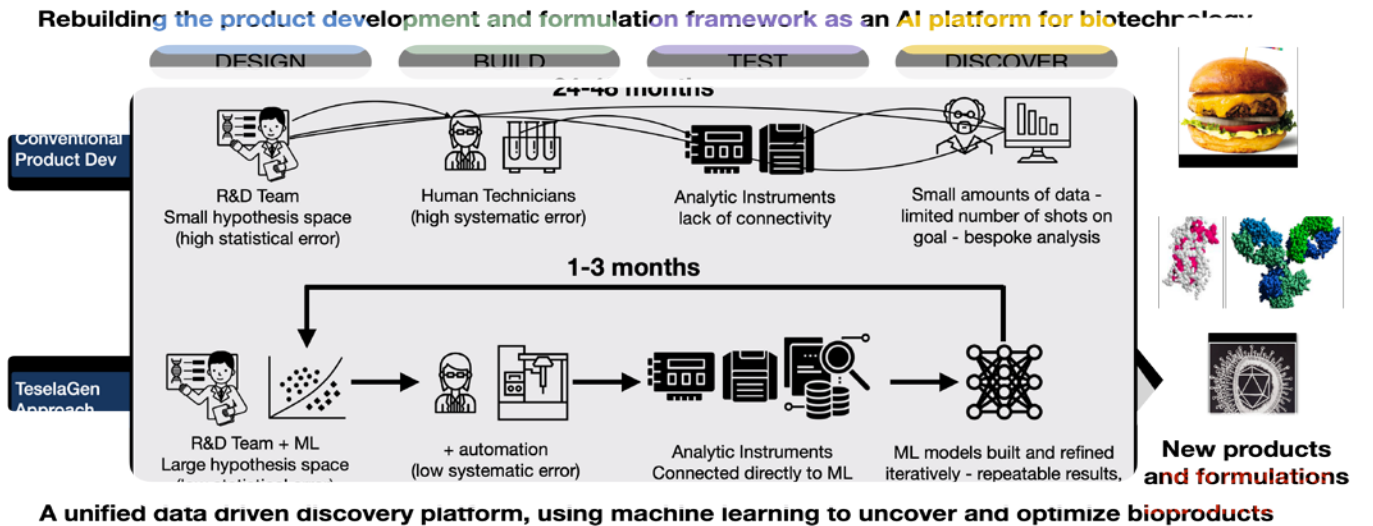
- The EDD/TEST module plays a vital role as that "clean well-lighted place for data" that is so important to successful data analytics, whether machine learning based or otherwise.

- The EDD/TEST module performs less as a stand-alone product, and more as a vital piece of infrastructure that ties our laboratory data acquisition tooling in the TeselaGen BUILD module, to the machine learning models developed and deployed in the DISCOVER module.

top challenges …

TeselaGen is a commercial product
- Only partially open source
- Stringent security requirements
- Rigorous development process and testing required
- Industry standard advanced software frameworks
- Advanced interfaces required, APIs, CLIs, Microservices, Integration frameworks like Node-Red
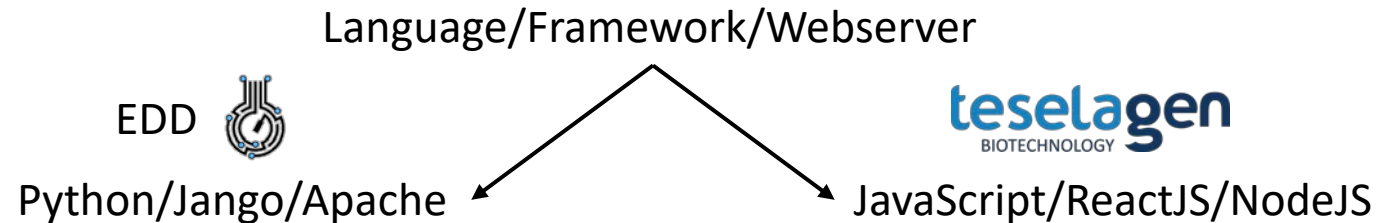- Often not published

EDD, BLiSS, BOOST etc., are academic projects
- Some open source, some not, potential licensing fees and negotiations for commercial use
- Often written in less-commercial languages like Python or Perl
- Small teams with ad-hoc testing regimes
- Missing advanced interfaces
- Supported until publication, may lie dormant and unsupported thereafter

decision points…

It was taking longer to make TeselaGen's TEST module compatible with EDD that to simply re-create EDD functionality de novo. TEST is written in pure JavaScript from front to back, using NodeJS as a backend. EDD is written in Python.

→ With the guidance of the EDD team, it was decided to simply rebuild EDD functionality in TeselaGen's TEST module in pure JavaScript for speed and tight integration with advanced web technologies.

Language/Framework/Webserver

EDD

Python/Jango/Apache   JavaScript/ReactJS/NodeJS

metrics used to measure progress...

As a commercial company, TeselaGen's metrics are bottom line based... we measure progress against future potential commercial viability of the platform.

Some new relevant commercial data points:

Biopharma	Agriculture	Industrial Bio	Academia

3 – Impact

impact on the state of industry...

Designing, building, and testing large molecules is both ambitious and complex!

It puts considerable demand on available information technologies

Improved efficiency and time to market is often cited as motive

Exploiting emerging methods can make the difference between having a viable product and having nothing at all to show for years of research.
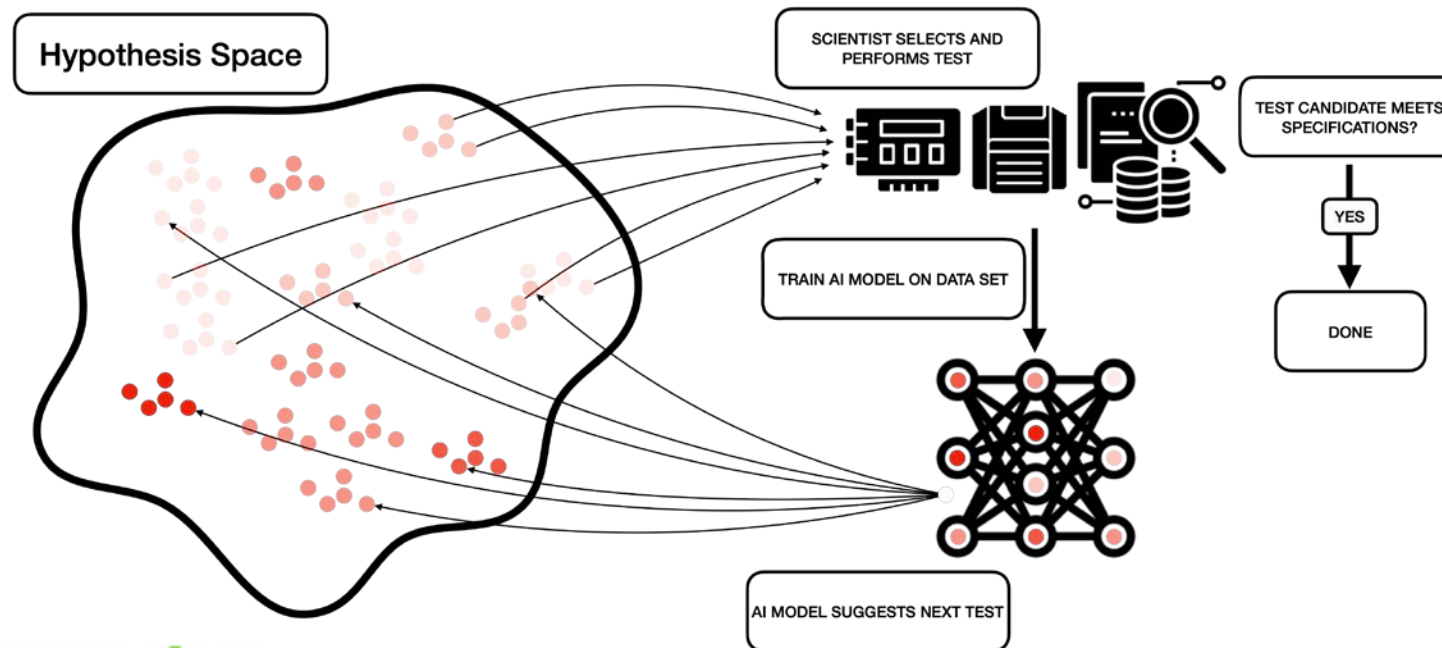
impact on the state of industry…

- A biological design exercise differs fundamentally from traditional physics-driven engineering by the nature of our understanding. (Regardless of how much synthetic biologists might wish otherwise)

- Physics driven engineering disciplines like electrical engineering have exploited simulation and monte carlo methods to good effect since the 1970's.

- Now, with the help of AI, biological engineering problems like protein folding are finding similar success.

- Problems in biology naturally lend themselves to a statistical learning approach, with genotype and phenotype taking the place of independent and dependent variables in a machine learning exercise (for example).

- Up until recently however, the combinatorial complexity of even a simple enzyme would appear hopelessly intractable. Furthermore, the amount of experimental data needed to build decent supervised or unsupervised machine learning models seems quite daunting to most investigators and beyond the reach of most biologists.

impact on the state of industry…

Our impact is that we are changing the way companies think about data flow and data driven discovery. There is not physical or chemical reason the discovery process cannot proceed at a much faster pace.

Covid-19 has taught us that with a strong enough push things happen fast!

The challenge for the future will to be to replace a massively capital and human resource dependent acceleration with one enabled by automation, artificial intelligence and well-organized data flows.

impact of results...dissemination

teselagen
BIOTECHNOLOGY

For TeselaGen, dissemination is largely in the form of getting our platform's advanced capabilities in the hands of our customers!

LanzaTech and TeselaGen Biotechnology Sign New Multi-Year Deal to Advance Carbon Remediation via Biological Processes

LanzaTech and TeselaGen ink new 5-year deal to reduce the human carbon footprint using synthetic biology, extending their relationship through 2025

SAN FRANCISCO AND CHICAGO (PRWEB) DECEMBER 21, 2020

Today, LanzaTech announced the signing of a new multi-year contract with TeselaGen Biotechnology, extending their relationship through 2025. The two companies have been collaborating since 2016 on the informatics behind high-throughput synthetic biology workflows. Over that time, LanzaTech has continued to experience rapid growth, showing the feasibility to synthesize more than 100 different molecules using its carbon-eating bacteria, and demonstrating the need for a significant scaleup of its R&D operations.

"Designing and optimizing biology is not easy, and we are in a race to recycle more carbon before it is too late. This collaboration with TeselaGen will extend our capabilities and help us achieve our goals", said Dr. Sean Simpson Chief Scientific Officer and Co-founder at LanzaTech. "TeselaGen has developed one of the most advanced cloud-based solutions for

" The ability to economically recycle poisonous greenhouse gasses like carbon oxides into valuable products via a biological process is an amazing achievement. "

nature > nature communications > articles > article

Article | Open Access | Published: 25 September 2020

Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism

DTU

Jie Zhang, Søren D. Petersen, Tijana Radivojevic, Andrés Ramirez, Andrés Pérez-Manríquez, Eduardo Abeliuk, Benjamín J. Sánchez, Zak Costello, Yu Chen, Michael J. Fero, Hector Garcia Martin, Jens Nielsen, Jay D. Keasling & Michael K. Jensen

Nature Communications **11**, Article number: 4880 (2020) | Cite this article

This very productive collaboration has also led to a relationship that has yielded a publication in a high impact journal... and a new customer!

Agile BioFoundry

24

4 – Progress and Outcomes

progress and outcomes... milestones & status

No.	Milestones	Task No.	Status
1	BOOST enhanced	1	Complete
2	BOOST integrated with TeselaGen's platform; evaluation of TeselaGen's HDE functionality applied to ABF workflows begins	1, 4	Complete
3	Test BOOST enhancements with 10 DNA constructs and compare with results without BOOST	1	Complete
4	BLiSS integrated with TeselaGen's platform; evaluation of TeselaGen's LIMS functionality applied to ABF workflows begins. (Tasks 1, 4)	1, 4	Complete
5	Enhanced analytics/visualization capabilities developed for EDD.	2	Complete
6	EDD integrated with TeselaGen's platform; ABF evaluations of TeselaGen's HDE and LIMS functionality complete. Test new capabilities with 2 studies present in EDD.	2, 4	Complete
7	Deployment of new capabilities (BOOST, BliSS, EDD) with industry partners.	3	Complete
8	Evaluation of deployment with industry partners complete.	3	Complete
E	Extension to test data acquisition and analysis capabilities against ABPDU fermentation data.	E	In Progress

progress and outcomes… notes

Other algorithms developed by LBNL:

- BOOST* has useful code but is protected by a license agreement. We will be happy to implement when we have pull from customer. In the mealtime the developer started a company to commercialize BOOST which may make integrations easier.

- BLiSS** is important from a public safety point of view but is also difficult to integrate into a commercial product and lacks support and interfaces. The DOE should do more to support BLiSS and make it freely available to trusted parties through a web service.

*Build OptimizatiOn Software Tools
**Biosecurity List Sequence Screening

progress and outcomes... extension

We finished our project with money to spare, so we took up the challenge of trying to extend our reach downstream into fermentation processes.

Collaboration with D. Tanjore at the Advanced Bioprocess Development Unit (ABPDU) and H. Garcia at the ABF to see if we can interpret and utilized fermentation data in a way that helps accelerate strain development and scale-up.

summary

- EDD serves a great model for the type and form of data needed by data scientists in industrial and synthetic biology. *EDD and TeselaGen users have the same needs.* Implementation might differ between commercial software product and academic software project.

- BOOST is being commercialized by the creator which should make it more amenable to integration via well thought out interfaces.

- BLiSS is a very important tool. The DOE should do more to support BLiSS and make it freely available to trusted parties through a web service.

- The collaboration has opened commercial opportunities as well as seeding academic collaborations.

- The extension to working with data from the ABPDU will allow us to be responsive to users in the strain development world which have similar needs and pain points as basic researchers.

Quad Chart Overview

Timeline

- Project start date: June 29, 2018
- Project end date: June 29, 2020
- Extension start date: Jan 27, 2021

	FY20 Costed	Total Award
DOE Funding	$400,000	$800,000
Project Cost Share	$171,419	$342,858

Project Partners

- NTESS (DOE)
- PNNL (DOE)
- LBNL (DOE)
- TeselaGen Biotechnology, Inc.

Project Goal

The objective of this project is to enhance software tools that facilitate biomanufacturing design. These tools allow scientists to work from fundamental DNA level designs for a biobased product, all the way through to manufacturing those products.

Milestones

1) Enhance selected ABF design tools (BOOST and BLISS) and integrate them with Teselagen's platform (15 %, FY19 Q2).

2) Integrate Experiment Data Depot and analytical workflows into Teselagen's platform (50 %, FY19 Q4).

3) Demonstrate ABF functionality integrated into Teselagen's platform with industrial partners (0%, FY20 Q2).

4) Evaluation of Emerging Teselagen Functionality Support for ABF Workflows (0%, FY20 Q1).

5) Final Report (0%, FY20 Q2).

Funding Mechanism

AgileBioFoundry MODULAR CRADA
STEVENSON-WYDLER (15 U.S.C. 3710a)
COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT