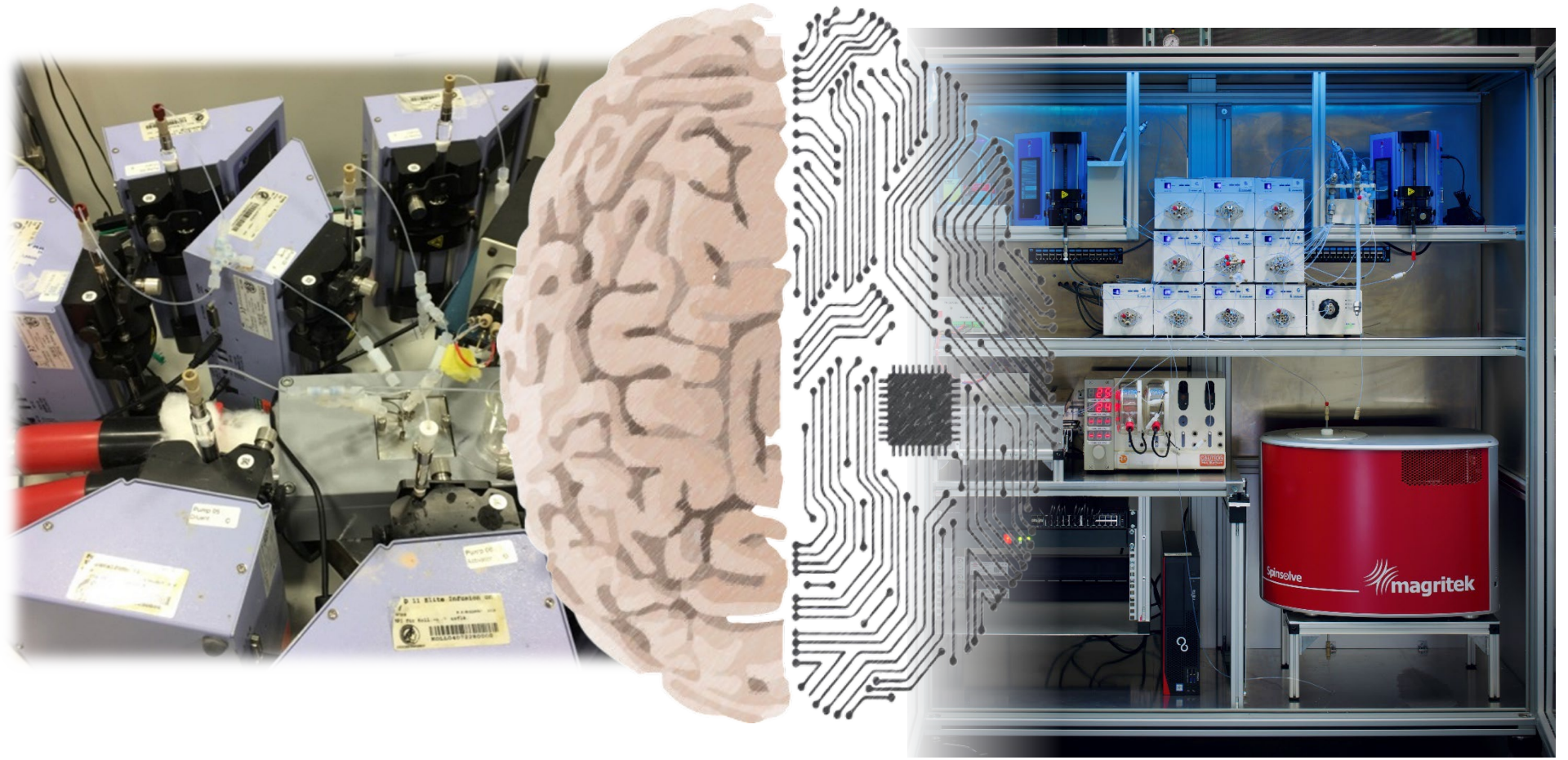
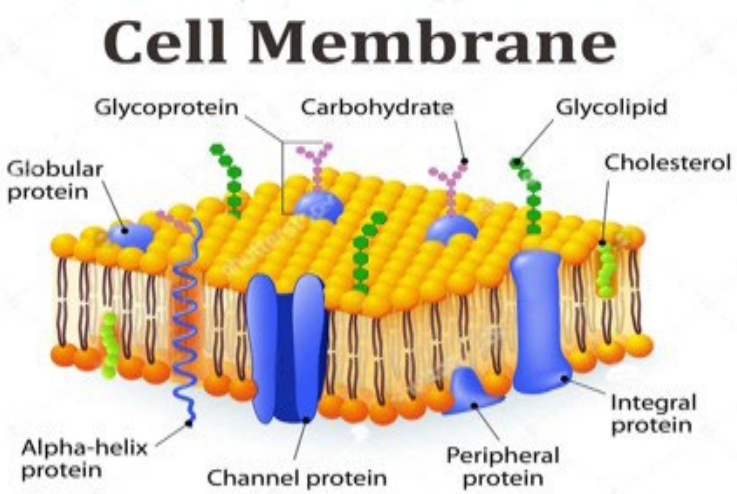
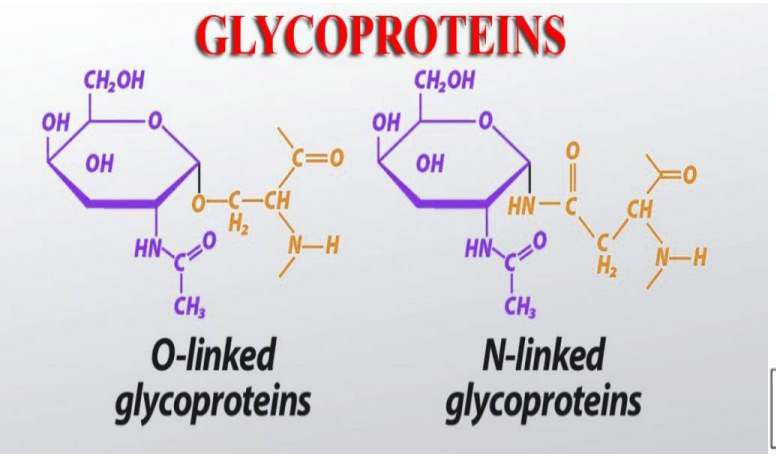
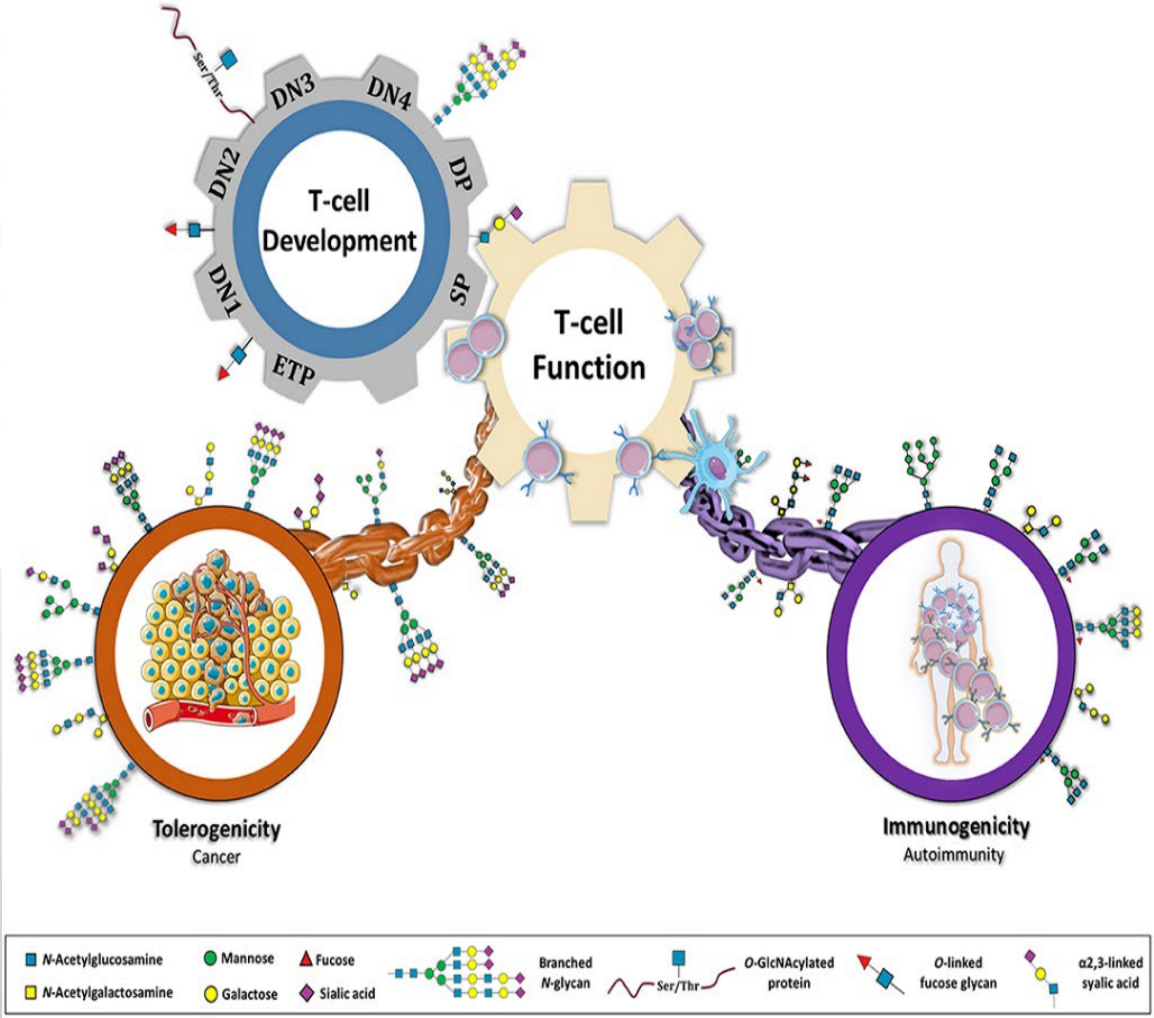
Application of AI guided Automated Flow Platforms for Understanding Complex Chemical Systems

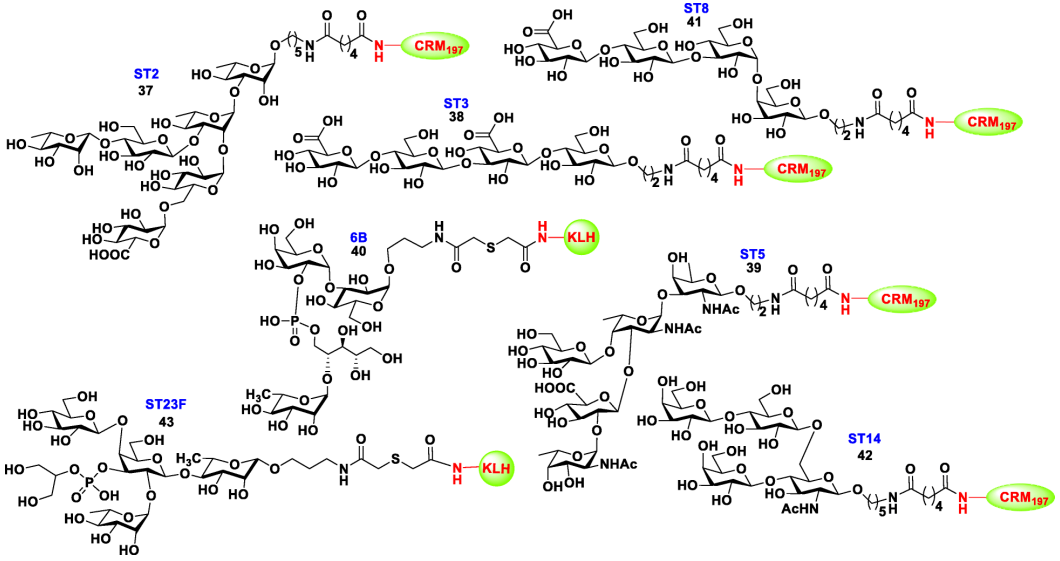


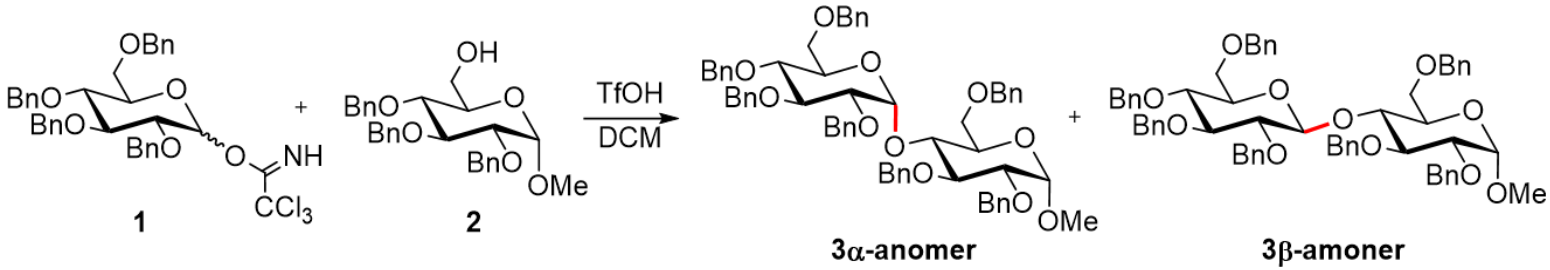

Cell Membrane
Glycoprotein
Carbohydrate
Glycolipid
Cholesterol
Globular protein
Integral protein
Alpha-helix protein
Channel protein
Peripheral protein


GLYCOPROTEINS
O-linked glycoproteins
N-linked glycoproteins


DN3
DN4
DN2
DP
T-cell Development
DN1
SP
ETP
Ser/Thr
T-cell Function
Tolerogenicity
Cancer
Immunogenicity
Autoimmunity
N-Acetylglucosamine
N-Acetylgalactosamine
Mannose
Galactose
Fucose
Sialic acid
Branched N-glycan
Ser/Thr
O-GlcNAcylated protein
O-linked fucose glycan
α2,3-linked syalic acid

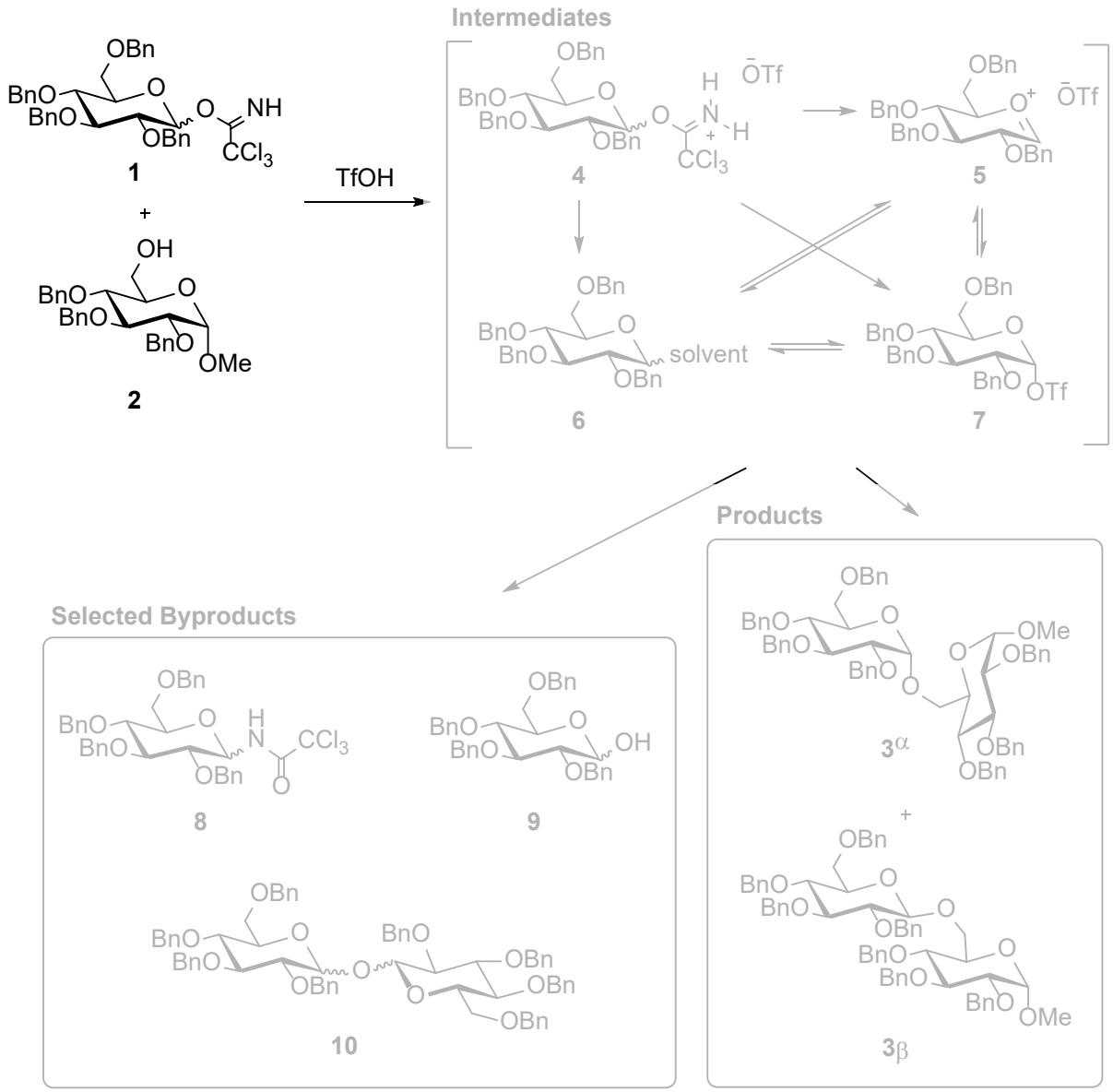
Pneumococcal polysaccharide vaccine (Sugar based vaccines)



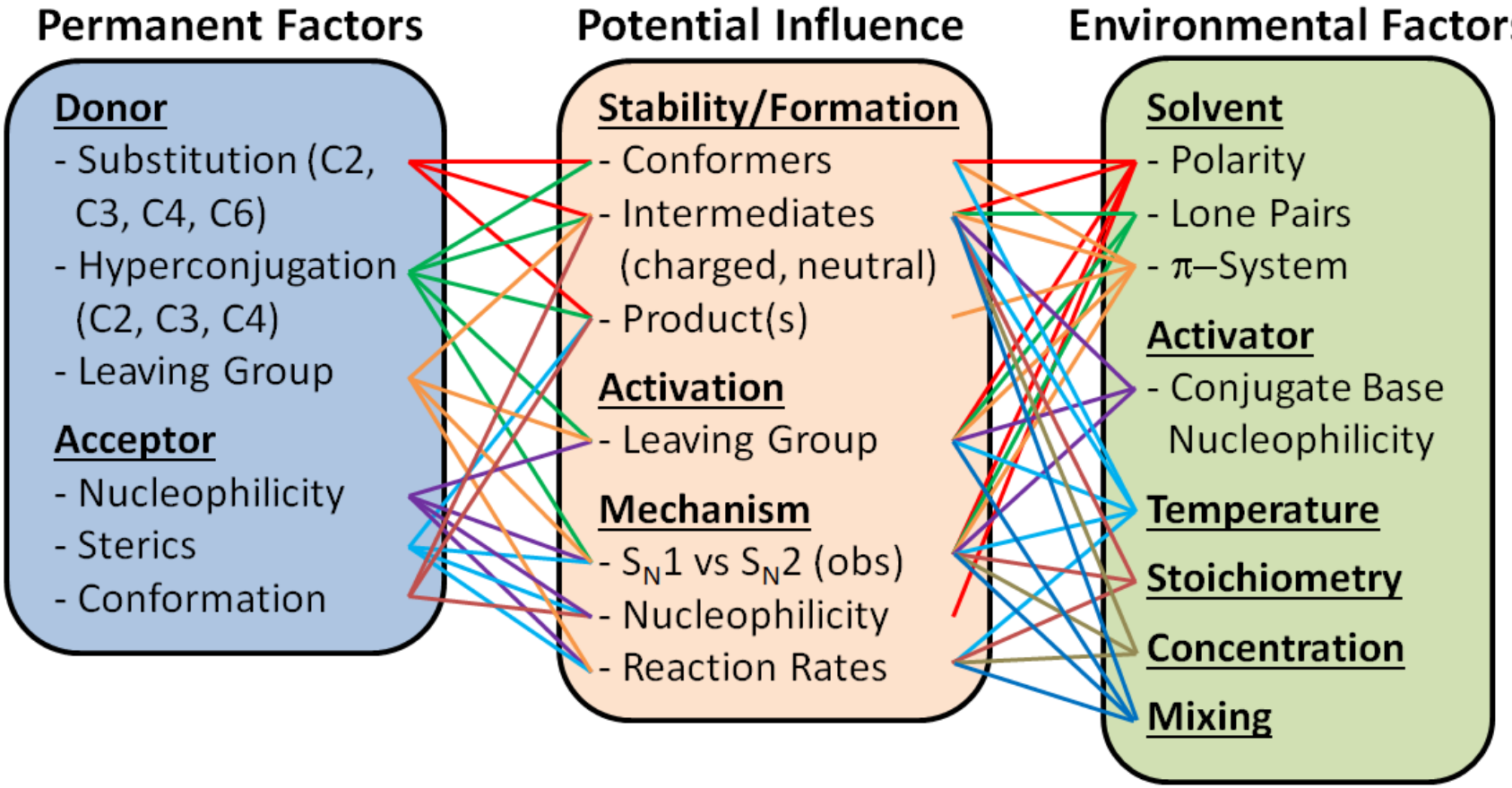


- One of the key reaction in Organic and Biochemistry
- Understanding will help to make better Carbohydrate based Vaccines

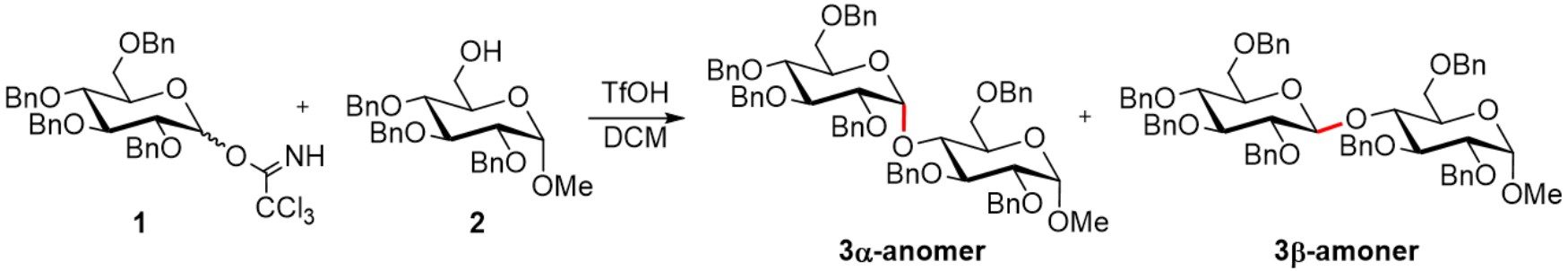
Chemical Glycosylation – A Challenge



Challenges Of Glycosylation



Chemical Glycosylation



- Studied for more than 100 years!
- Mechanistically a black box
- Factors determining selectivity are inconclusive

A Simple Reaction?

Reaction

Reactor

Flowrate

Residence Time

Solvent

Temperature

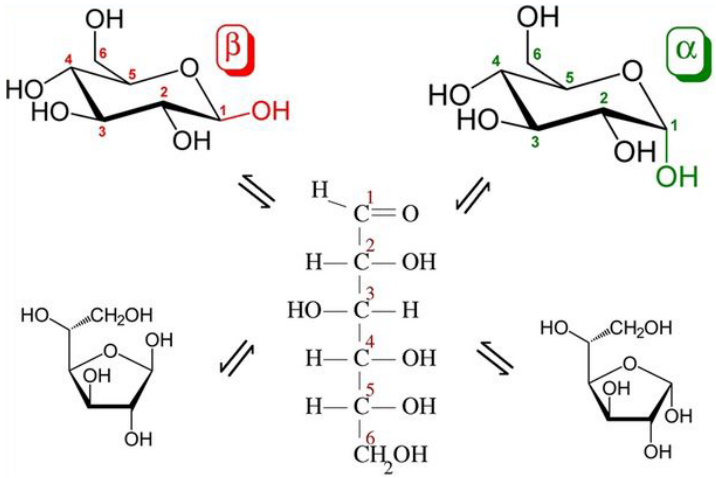
Concentration

Reaction Mechanism

Kinetics

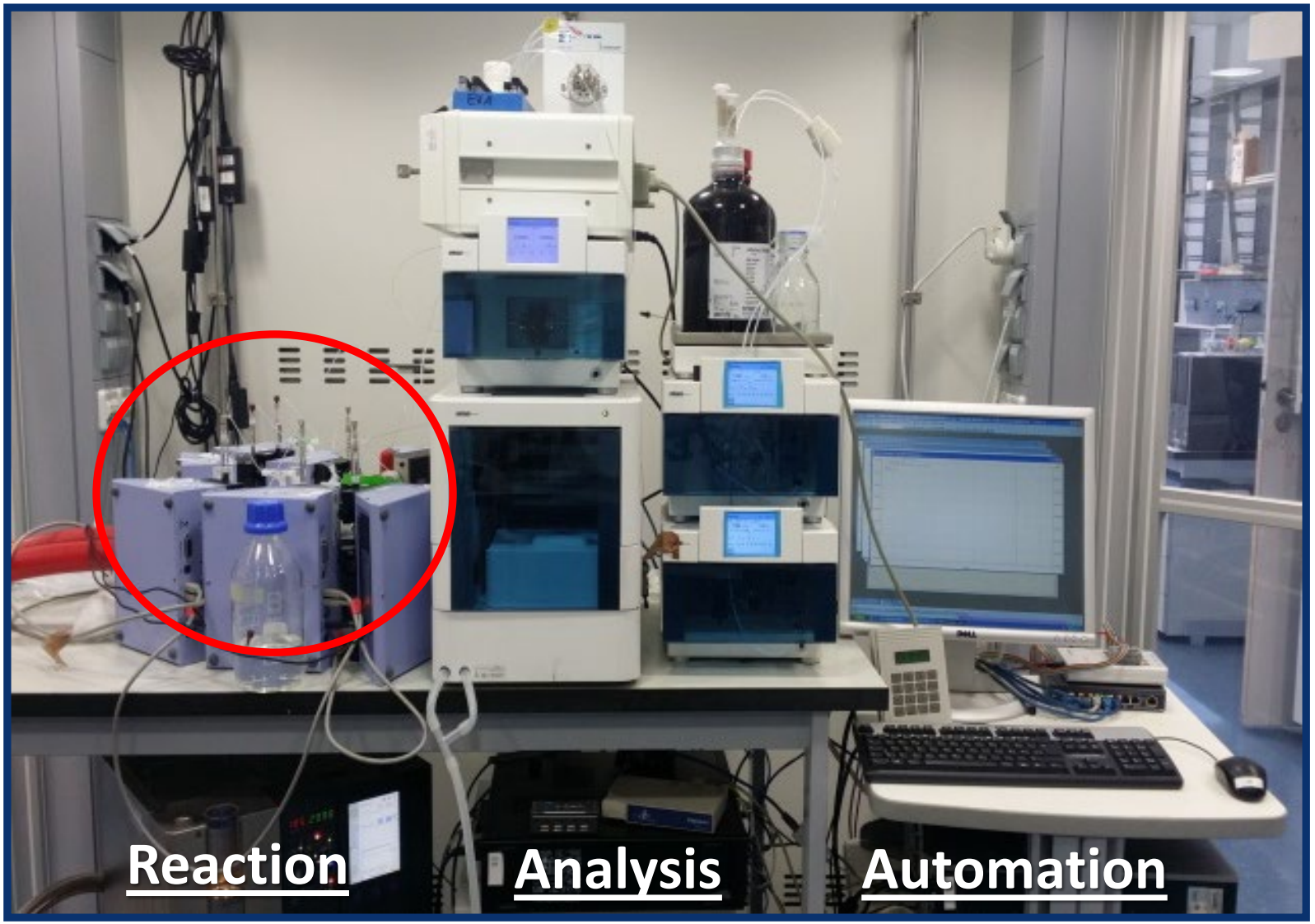
Selectivity , Yield

Diastereomers

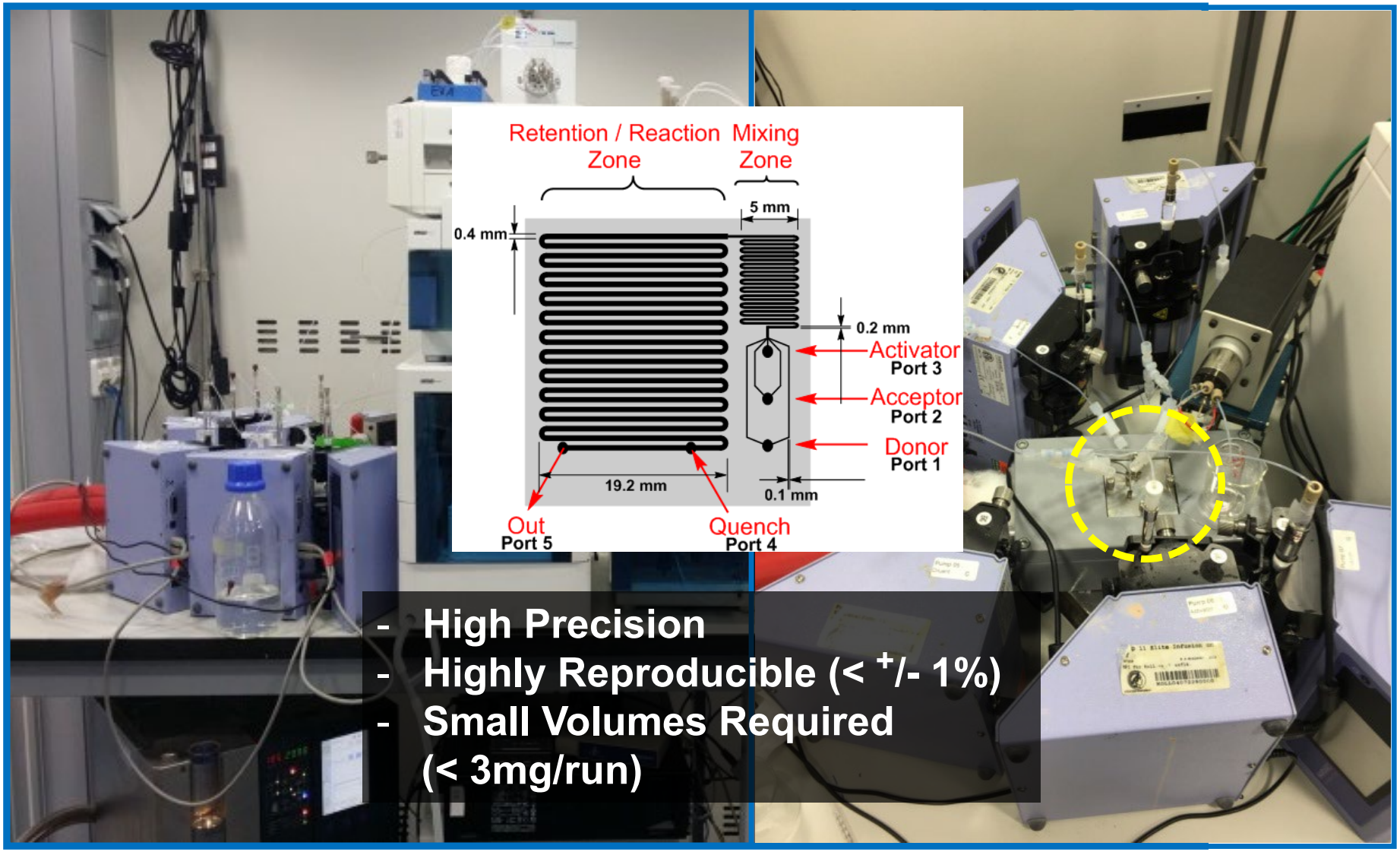


By-Products

Precise Control Over Reaction Conditions Via An Automated Flow Platform

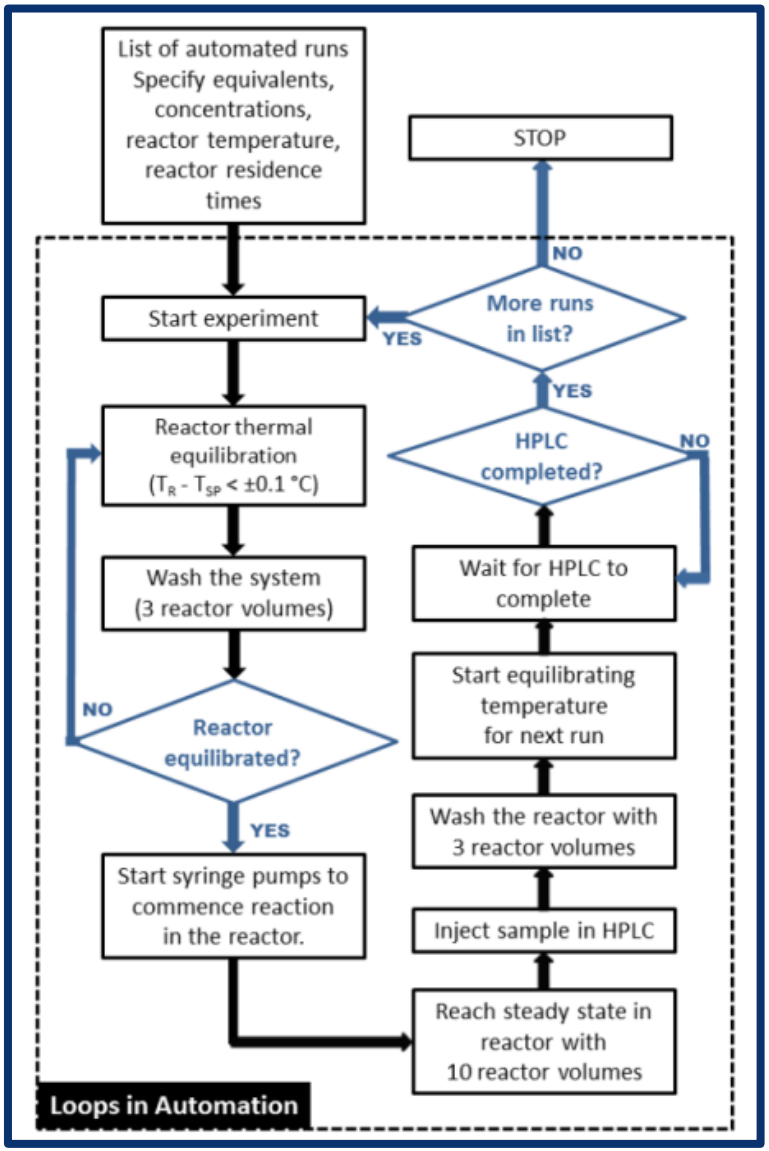


Overview Of The Automated Platform



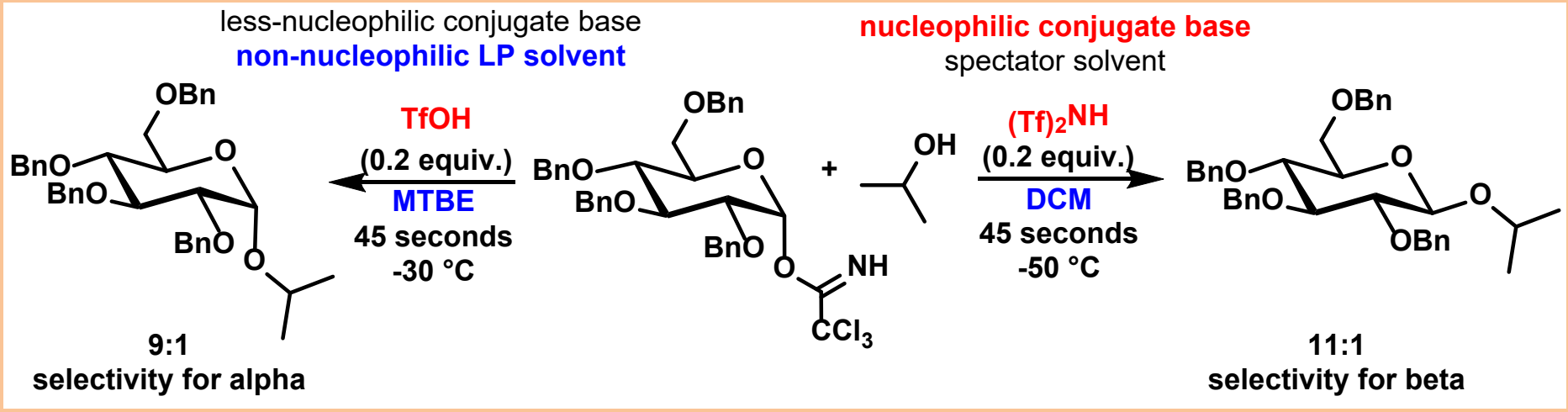
- High Precision
- Highly Reproducible (< +/- 1%)
- Small Volumes Required (< 3mg/run)

Control Algorithm of Flow Chemical Platform



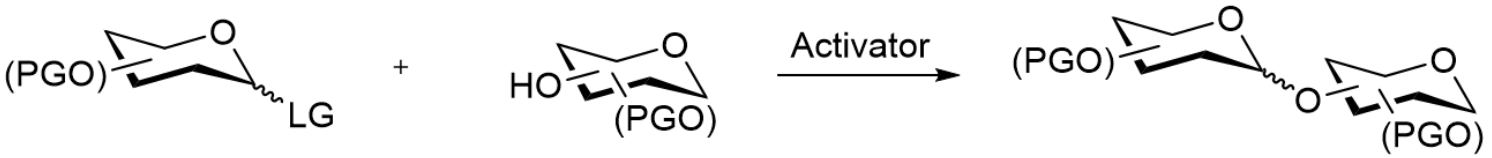
Chatterjee, S. *et al. JACS,* **2018**, *140,* 11942

Use Automated Flow Platform To Gain A Greater Control on Stereoselectivity

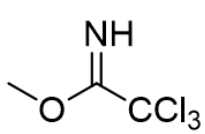
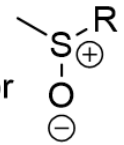


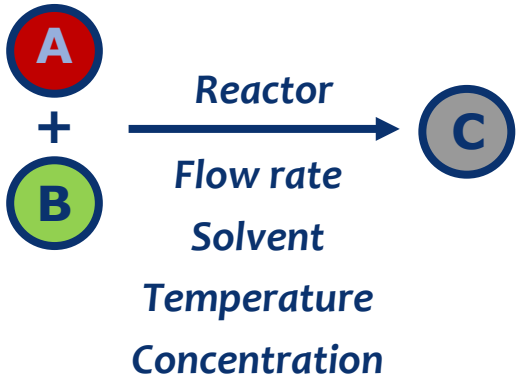
Chatterjee, S. *et al. JACS,* **2018**, *140,* 11942

Single Step Programmed Synthesis

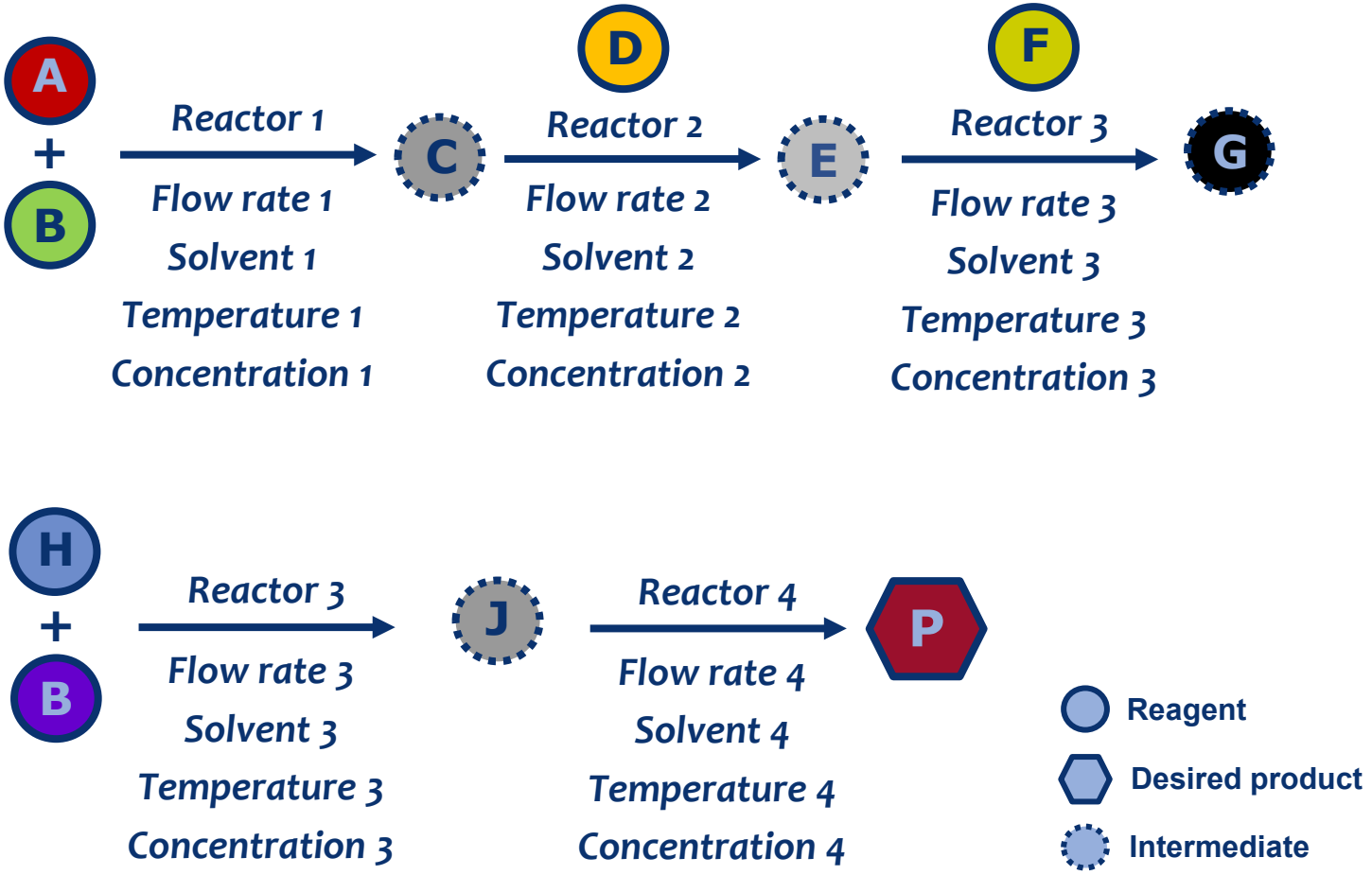


glycosyl donor **glycosyl acceptor** **disaccharide**

LG (leaving group) = Br, F, SR,  ,or 



Research Overview: Programmed Multistep Synthesis



A Conceptually New Approach To Organic Synthesis

Linear Approach

Synthesis of Small Molecules

Stepwise

Continuous Multistep

- Equipment Redundancy
- Reconfigurations Required

What Do We NEED?

- Deliver Reagents
- Controlled Conditions
- Variable Conditions
- Link Multiple Steps
- Analysis

- **Multiple Modules**
- **Diverse Chemistry**
- **Route Screening**
- **No Reconfiguration**

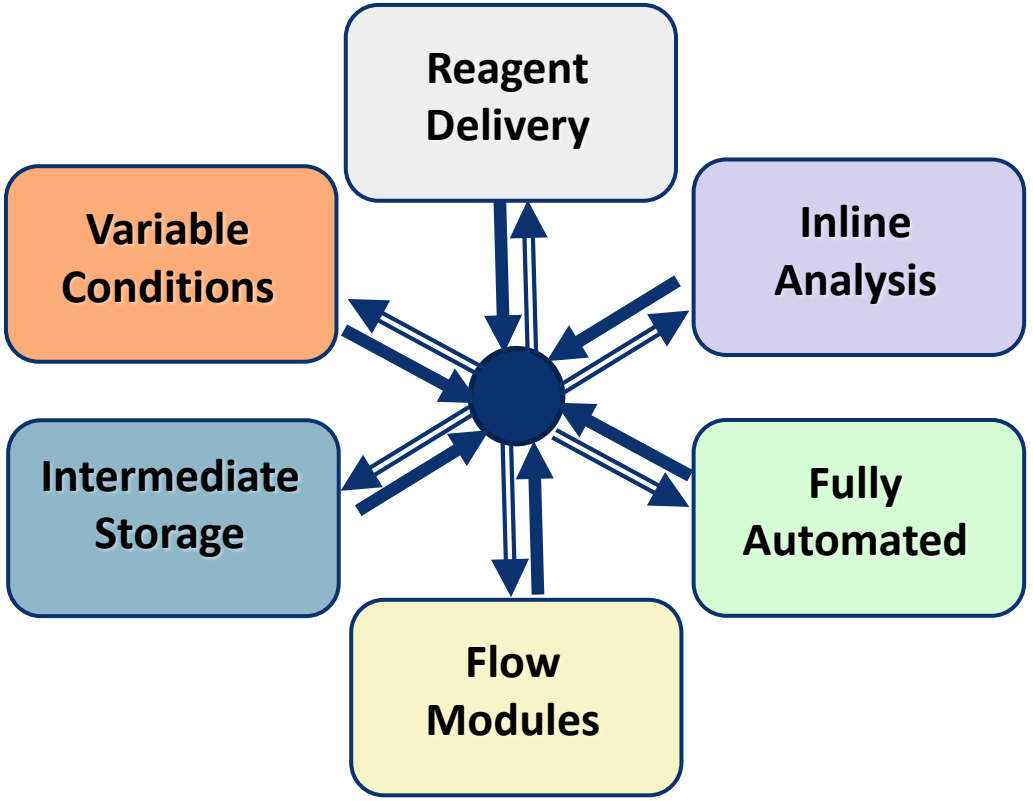
Cyclic Approach

Synthesis of Biopolymers

- Single Module
- One Type Chemical Bond
- Pre-Functionalized Substrates

Radial Synthesizer: Key Concept

Central Switching Station (CSS)

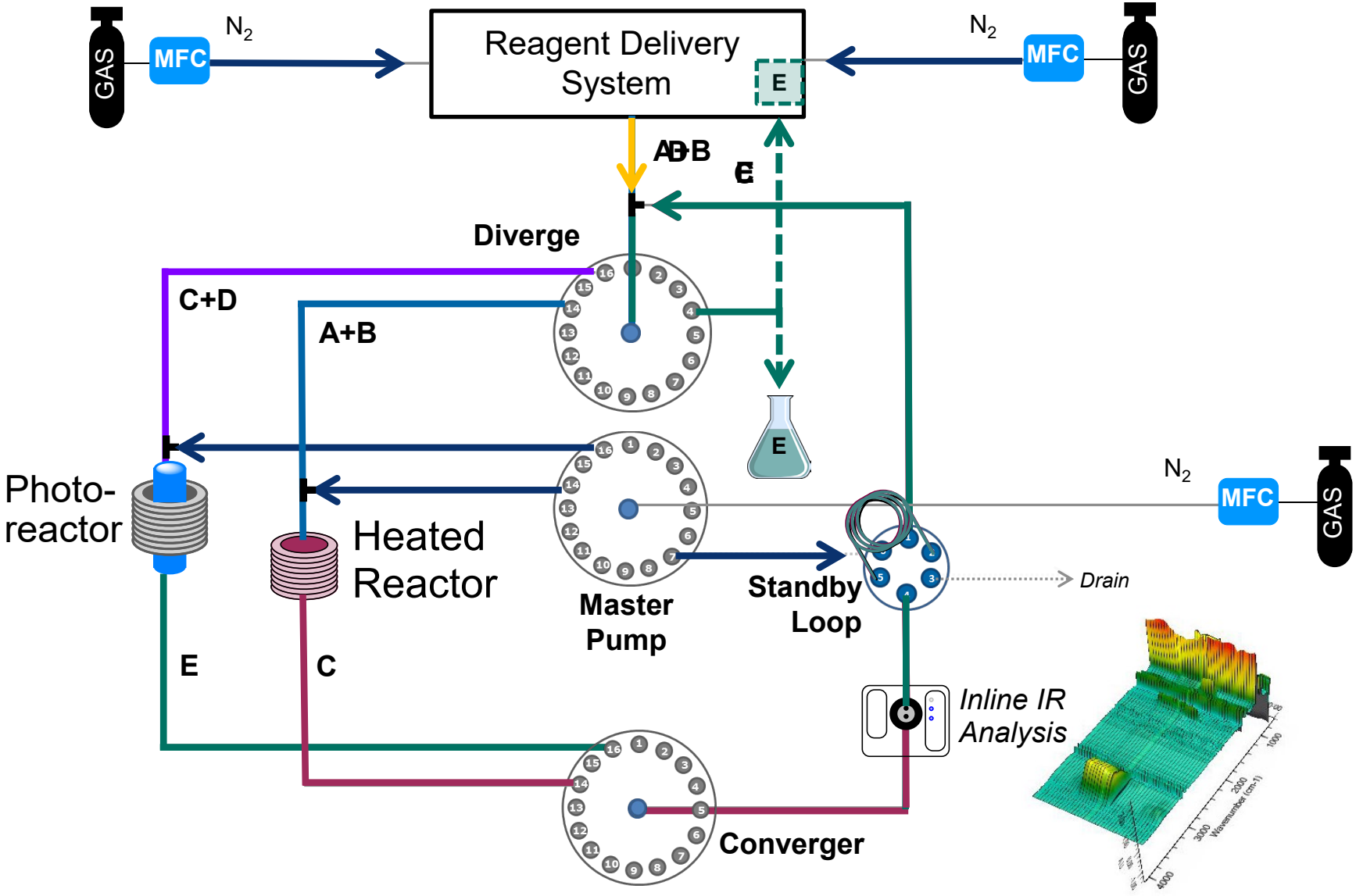


Radial Synthesizer

- Modular assembly
- 13 reagents and 5 solvents
- Any conditions accessible
- Single- or multistep
- ***No Reconfiguration Required***

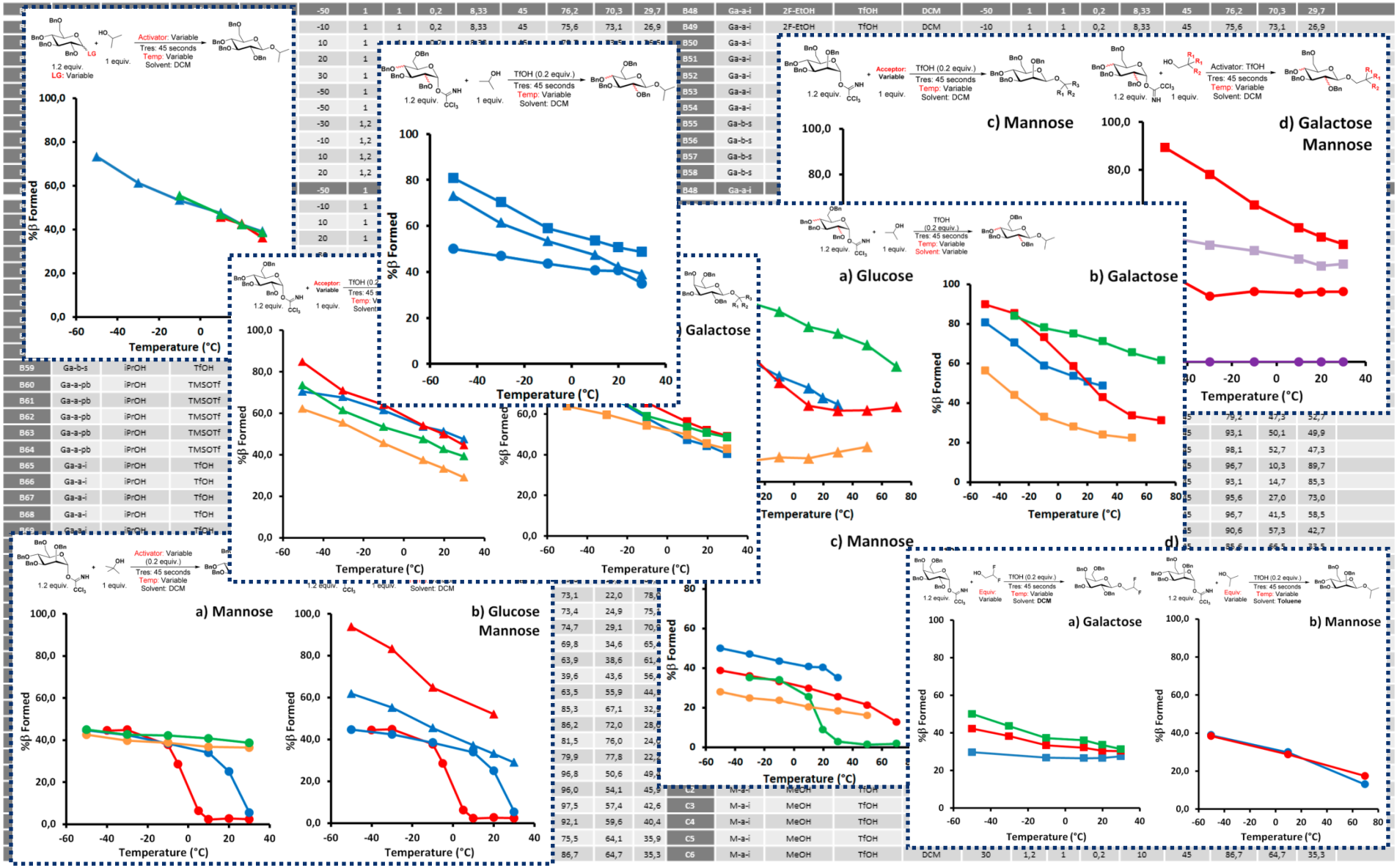
Chatterjee, S. *et al. Nature* **2020**, *579*, 379
"Automated radial synthesis of organic molecules"

Multistep Radial Synthesis

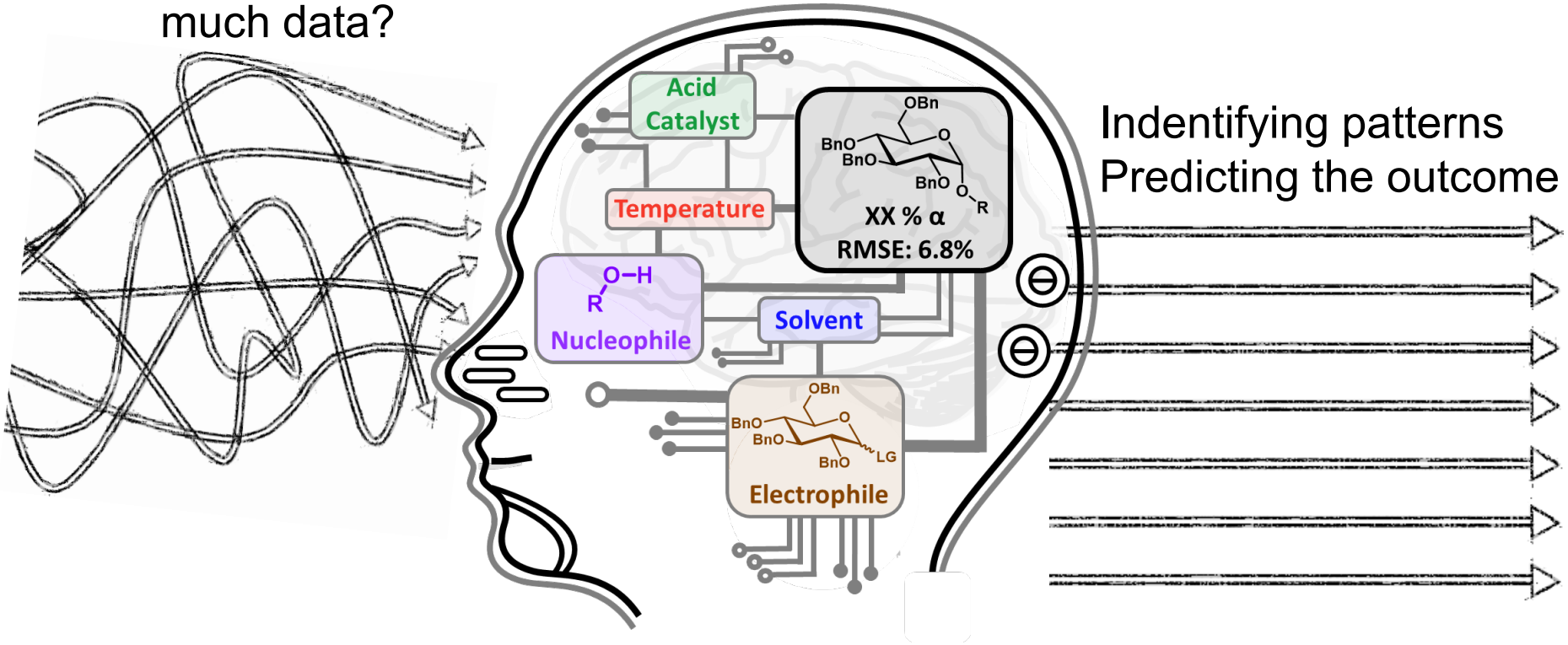


***Chatterjee, S.** et al. Nature* **2020**, *579*, 379
"Automated radial synthesis of organic molecules"

Over 400 Reproducible and Curated Data!



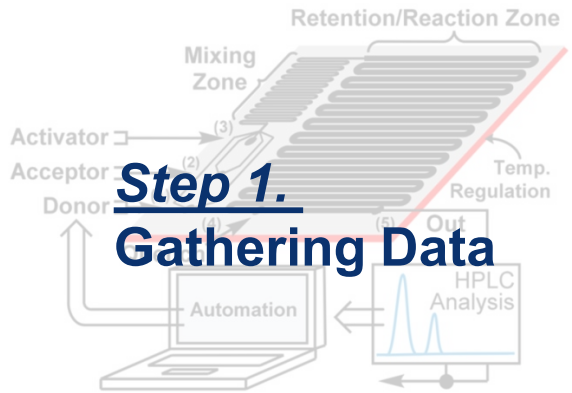
Automated flow platform: what can we do with so much data?



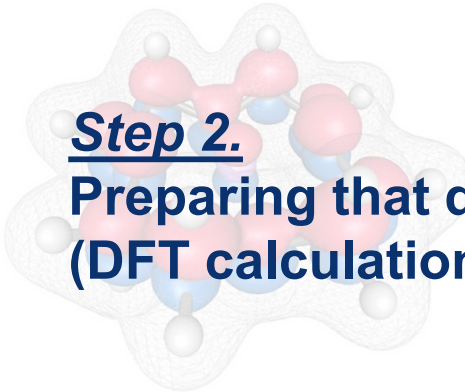
Chatterjee, S. *et al. Chem. Sci.* **2021**,*12,* 2931
"Predicting glycosylation stereoselectivity using machine learning"

Machine Learning Workflow

Input Data → Learning Algorithms (Decision Trees) → Predictions



Step 1.
Gathering Data



Step 2.
Preparing that data (DFT calculations)

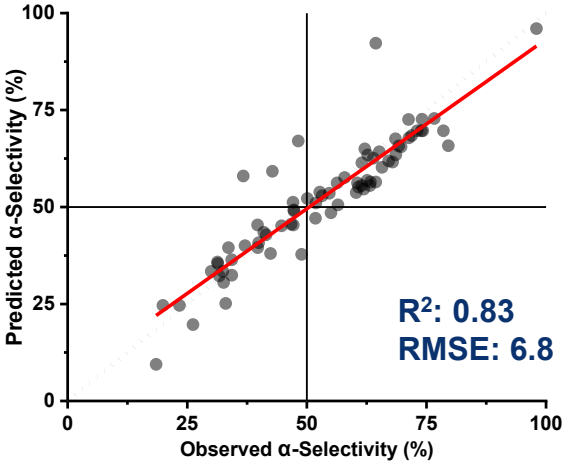
Random Forest

- Machine learning software based on Random Forest in MATLAB
- Trained using five-fold cross validation strategy
- Generates models, known as decision trees
- These trees are combined with averaged weighted sum to create the final ensemble (collection) of trees

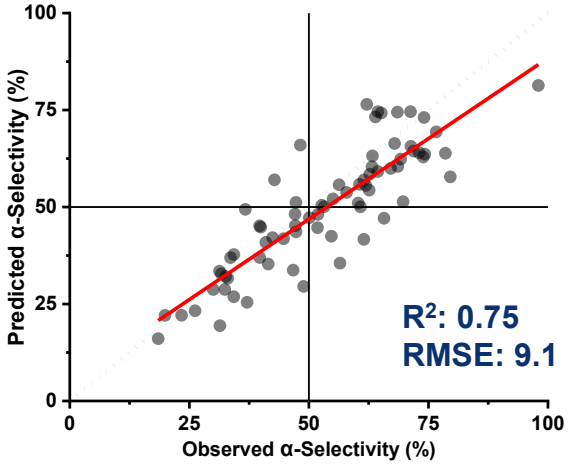
Step 6.

Comparing Other ML Algorithms

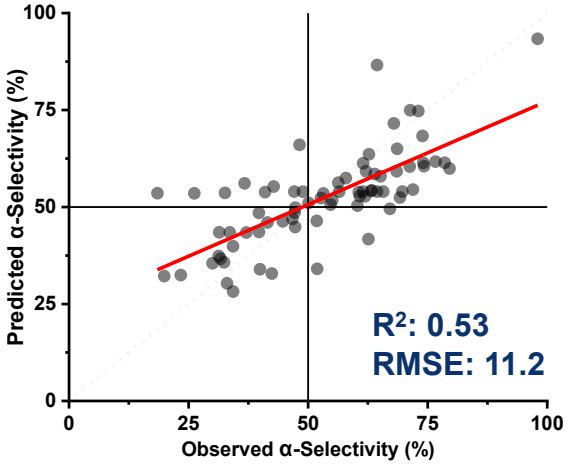
Random Forest



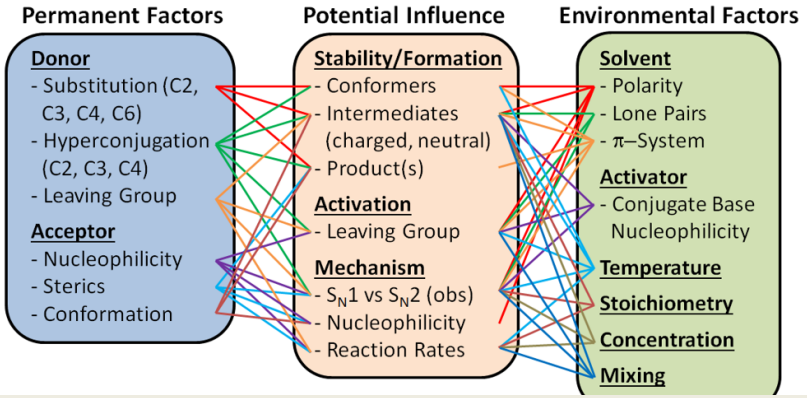
Gaussian Process Regression

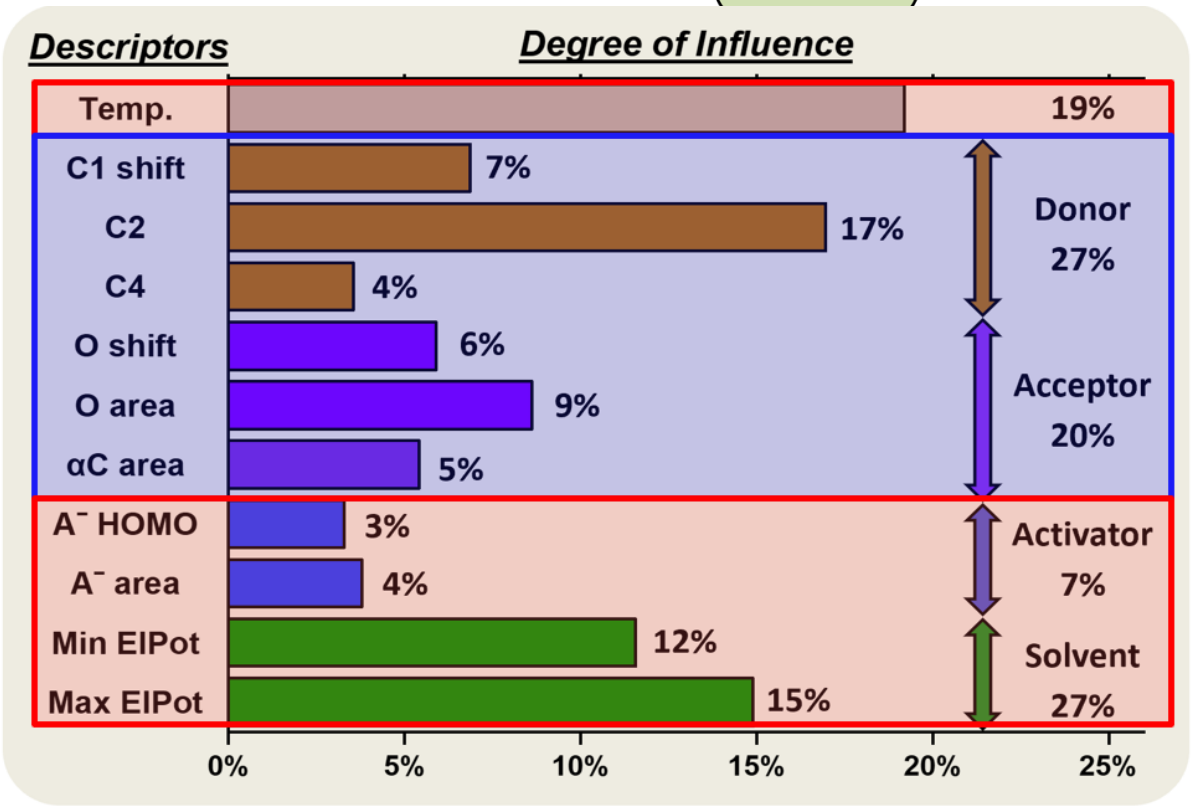


Support Vector Machine



Overall Degree of Influence of Descriptors





Acknowledgements and Collaborations

Prof. Peter Seeberger, Dr. Kerry Gilmore, Dr. Sooyeon Moon

Max Planck Institute of Colloids and Interfaces

Dr. Paco Laveille, Prof. Christophe Copéret

ETH zürich

DARPA

Dr. Alex Cresswall, Prof. John Chew, Dr. Uli Hintermair

UNIVERSITY OF BATH

Prof. Matthew Davidson

Institute for Sustainability

Science and Technology Driven Sustainability