Optimization of glycosylation with glycosyl fluoride by using machine learning

（機械学習を用いたグリコシルフルオリドによるグリコシル反応の最適化）

Doctoral Thesis

戴 長浩

Department of chemistry

Graduate School of Science

Osaka University

Japan

2023

Content

List of Abbreviations

Ac	acetyl
Allyl	2-propenyl
Ab	antibody
APC	antigen presenting cell
Boc	*t*-butoxycarbonyl
DCM	dichloromethane
DMF	*N,N*-dimethylformamide
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
Et	ethyl
Gal	galactose
Glc	glucose
GlcA	glucuronic acid
GlcN	glucosamine
GlcNAc	*N*-acetylglucosamine
HOBt	hydroxybenzotriazole
HPLC	high performance liquid chromatography
Me	methyl
MS	mass spectrometry
MS	molecular sieves
NIS	*N*-iodosuccinimide
NMR	nuclear magnetic resonance
o	ortho position
p	para position
Ph	phenyl
rt	room temperature
TAA	tumor associated antigen
t-Bu	*tert*-butyl
TBSCl	*tert*-butyldimethylsilyl chloride
TFA	trifluoroacetic acid
TfOH	trifluoromethanesulfonic acid
THF	tetrahydrofuran
TMS	trimethylsilyl
TMSOTf	trimethylsilyl trifluoromethanesulfonate
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
Troc	2,2,2-trichloroethoxycarbonyl

Abstract

Efficient glycosylation methods have been the focus of research in glycochemistry as it is a fundamental reaction for the preparation of glycosides and the synthesis of sugar chains. Research on glycosylation reactions has focused on high yields and high stereoselectivity. Glycosylations are complicated processes as they are influenced by many factors, such as the choice of donors and acceptors, type of catalyst, choice of solvent, reaction time, and temperature, all of which have an impact on yield and stereoselectivity. In terms of donor selection, glycosyl fluorides offer significant advantages over other donors. Due to the stability of the carbon-fluorine bond, glycosyl fluorides are thermally and chemically stable and can therefore be purified by column chromatography, distillation, and other operations, and can be stored for long periods. In contrast, glycosyl fluorides can be readily activated by hard Lewis acids for use in glycosylation reactions. In recent years, with the development of artificial intelligence, machine learning can be a powerful tool for the efficient optimization of reaction conditions, thereby reducing time and saving the efforts of researchers.

In this study, the author applied machine learning to glycochemistry; machine learning was used to analyze data to improve the yield and stereoselectivity of glycosylation reactions. Firstly, the author conducted α-glucosylation as a model study to investigate the integration of machine learning with glycosylation reaction. The author then chose to investigate the glycosylation reaction for the first step of the α-gal synthesis. After performing conditional exploration for collecting data sets for machine learning analysis, machine learning improved the reaction yield by 13% compared to the dataset.

The author then investigated the α-selective glycosylation of xylose. Machine learning improved the α-selectivity from 1.3/1 to 1.5/1 in the batch analysis. The author next describes flow chemistry techniques, which are reproducible, easy to scale up, and effective in collecting data. By applying flow chemistry, machine learning optimization further improved the α-selectivity to 3.2.

Chapter 1. Introduction

1.1. Background on Glycosylation Reactions

Glycosylation reactions play a vital role in the synthesis of glycosides and oligosaccharides, which are essential components of numerous biological processes, including cell signaling, molecular recognition, and immune response.[1] A key challenge in glycosylation reactions is to achieve high yields and stereoselectivity to control the formation of the desired product. The complexity of glycosylation reactions arises from the need to consider multiple factors, such as the donor and acceptor selection, the choice of solvent, the type of catalyst, reaction time, and temperature.[2, 3]

Among many glycosylation donors, including thioglycosides, hemiacetals, glycosyl halides, and activated sugar derivatives,[4] glycosyl fluorides, have gained significant attention due to their enhanced stability and versatility in glycosylation reactions. The unique properties of glycosyl fluorides, such as the strength and stability of the carbon-fluorine bond, offer several advantages, including high thermal and chemical stability, ease of purification, and long-term storage.[4]

Despite the considerable progress made in the field of glycochemistry, achieving high yields and selectivity in glycosylation reactions remains challenging. Recent advances in artificial intelligence (AI) and machine learning (ML) offer an opportunity to address these challenges by leveraging computational methods to optimize reaction conditions and analyze complex datasets.[5] Integrating machine learning into glycochemistry can significantly improve research efficiency, reduce experimental time, and save researchers' efforts.

1.2. Glycosyl Fluorides as Glycosylation Donors

Glycosyl fluorides have emerged as promising glycosylation donors due to their unique properties and advantages over other donor types. The carbon-fluorine bond in glycosyl fluorides is characterized by its high stability, resistance to hydrolysis, and low

reactivity towards nucleophiles under mild conditions.[6] These features confer significant benefits to glycosyl fluorides in glycosylation reactions; glycosyl fluorides have emerged as valuable glycosylation donors due to their stability, stereoselectivity, versatility, and compatibility with mild reaction conditions. The integration of machine learning techniques into glycosylation reactions involving glycosyl fluorides has the potential to further enhance their utility by optimizing yields and selectivity.

1.3. Machine Learning in Chemical Synthesis

Machine learning (ML), a subset of artificial intelligence (AI), has recently gained considerable attention for its potential applications in various scientific disciplines, including chemistry. The use of ML algorithms in chemical synthesis has demonstrated remarkable potential for optimizing reaction conditions, predicting reaction outcomes, and assisting in the design of new molecules and materials.[10] This section will explore the impact of machine learning on chemical synthesis and its potential for advancing the field of glycochemistry.

1.3.1. Reaction Optimization: One of the key applications of machine learning in chemical synthesis is the optimization of reaction conditions. ML algorithms can process vast amounts of experimental data to identify patterns and correlations between variables, such as temperature, catalyst, solvent, and concentration, and the resulting yields or selectivities.[11-14] This information can be utilized to predict the most favorable conditions for a given reaction, thereby improving efficiency, and reducing experimental time.

1.3.2. Predictive Modeling: Machine learning algorithms can be used to develop predictive models that estimate the outcome of a reaction based on specific input variables.[10] These models can be particularly useful in guiding the design of new molecules or identifying promising synthetic routes, ultimately saving researchers' time and resources by reducing the need for trial-and-error experimentation.

1.3.3. Catalyst and Ligand Design: ML techniques have also been employed in the

design of new catalysts and ligands for chemical reactions.[17] By analyzing existing data on catalyst performance and structural features, ML algorithms can generate predictions for the performance of new catalysts or ligands, guiding their synthesis and evaluation in experimental settings.

1.3.4. Application to Glycochemistry: The integration of machine learning techniques into glycochemistry, such as in the optimization of glycosylation reactions involving glycosyl fluorides, has the potential to significantly improve yields and selectivities. By leveraging ML algorithms to analyze and optimize reaction conditions, researchers can enhance the efficiency of glycosylation reactions, reduce the time required for experimentation, and accelerate the discovery and synthesis of novel glycoconjugates with potential biological applications.[18]

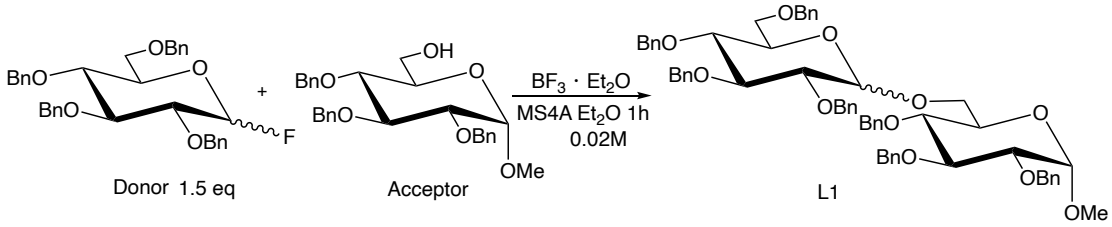
In conclusion, the application of machine learning in chemical synthesis offers numerous opportunities for improving reaction efficiency, predicting outcomes, and designing new catalysts and ligands. As the field of glycochemistry continues to evolve, the integration of ML techniques into the study of glycosylation reactions has the potential to significantly advance our understanding and control of these complex processes.

Chapter 2. Model Study: Integrating Machine Learning with Glycosylation Reactions

2.1. Model Selection and Rationale

In the application of machine learning to improve the yield of chemical reactions, key reaction conditions can be experimentally varied to obtain yield data under various conditions. By creating a dataset containing all the reaction conditions and their corresponding yields, machine learning can be applied to analyze the data and subsequently provide optimized reaction conditions.[13, 14, 19] Several factors can influence the yield and stereoselectivity of glycosylation reactions, such as temperature, solvent, catalyst, concentration, and the types of protecting groups on the monosaccharides.[20]

In this study, glycosylation shown in Scheme 2.1 was chosen as a model glycosylation reaction. $BF_3 \cdot Et_2O$ was employed as the catalyst. This study primarily investigated the influence of temperature and catalyst equivalents on the yield of the α-configuration. Upon completing the dataset collection, machine learning was used to optimize reaction conditions for enhancing the yield and selectivity of the glycosylation reaction.



Scheme 2.1. Glycosylation reaction for model study

2.2. Data Collection and Analysis

To obtain a high-quality dataset for machine learning analysis, the reaction conditions in the dataset should cover a range of temperatures and catalyst equivalents, from low to high values. Under an argon atmosphere, the donor (1.5 eq) and acceptor were dissolved in diethyl ether. The solution was stirred for 10 minutes at temperature X °C,

followed by the addition of Y equivalents of $BF_3 \cdot Et_2O$. After one hour, the reaction was quenched with a saturated sodium bicarbonate solution. The organic phase was extracted with dichloromethane, and dried over anhydrous sodium sulfate, and the residual solvent was removed under reduced pressure by rotary evaporation. The yield and stereoisomeric ratios of the products were determined using proton nuclear magnetic resonance (H^1NMR) spectroscopy, with tetrachloroethane as the internal standard. After an initial exploration of reaction conditions, the dataset's temperature range was controlled from -20°C to 34.6 °C, and the catalyst equivalents ranged from 8 eq to 15 eq (Table 2.2). Nine sets of data were obtained through experimentation. A preliminary analysis revealed that the total yield of the glycosylation reaction increased from -20°C to 30°C and slightly decreased from 30°C to 34.6°C. The decline in yield might be due to the degradation of the donor under high-temperature conditions. The stereoselectivity of the α-configuration increased with temperature, from -20°C to 34.6°C.

Regarding catalyst equivalents, an increase in catalyst equivalents improved the reaction yield under both low and high-temperature conditions. Under low-temperature conditions at -20°C and 0°C, increasing the catalyst equivalents led to a decrease in the stereoselectivity of the α-configuration. Conversely, under high-temperature conditions

Table 2.2. Data set for model study

Entry	Temp.(°C)	Eq(BF_3Et_2O)	L1(%)			
			α	β	α/β	$\alpha+\beta$
1	-20	10	7	9	0.78	16
2	-20	15	8	12	0.67	20
3	0	10	21	26	0.81	47
4	0	15	23	33	0.70	56
5	24	14	43	51	0.84	94
6	30	8	46	49	0.94	95
7	30	12	48	46	1.04	94
8	34.6	10	39	35	1.11	74
9	34.6	14	42	38	1.11	80

* Estimated by H^1NMR using tetrachloroethane as an internal standard

at 30°C, increasing the catalyst equivalents enhanced the stereoselectivity of the α-configuration. This indicates that the influence of temperature and catalyst equivalents on the yield and stereoselectivity of glycosylation reactions is complex and variable.

2.3. Machine Learning Optimization and Results

2.3.1. Gaussian Process Machine Learning Method

Gaussian process (GP) is a powerful and flexible machine learning method that has gained popularity in various fields, including chemistry, for its ability to model complex, nonlinear relationships between input features and output values.[19] Gaussian process regression (GPR), also known as kriging, is a Bayesian non-parametric method that provides not only predictions but also uncertainty estimates for those predictions, making it particularly useful for optimization problems (Figure 2.3.1).[22]

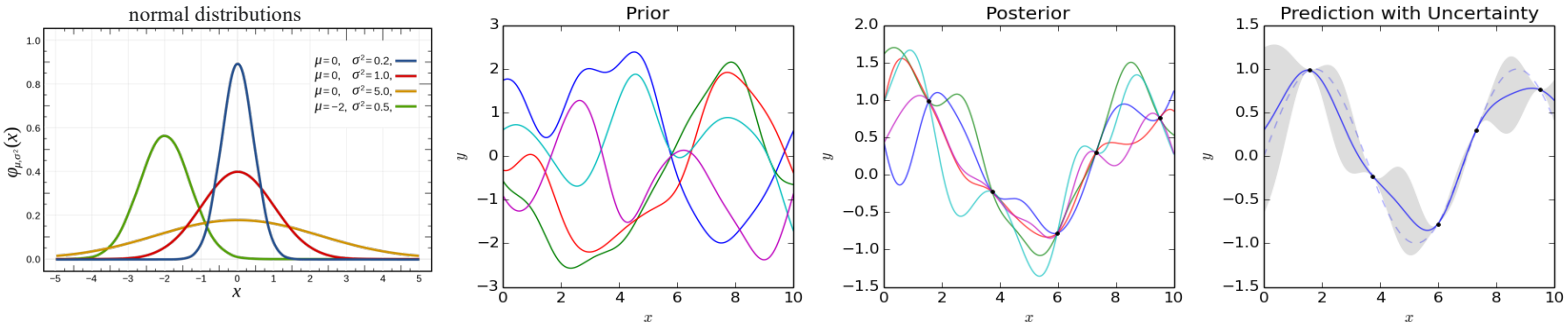


Figure 2.3.1. Gaussian process regression

In a Gaussian process, the prior distribution over functions is defined by a mean function and a covariance function, often referred to as the kernel.[23] The kernel determines the similarity between data points, and various kernel functions can be used depending on the problem at hand. In essence, a Gaussian process models the function as an infinite-dimensional Gaussian distribution, where the function values are correlated according to the kernel function.[24]

One of the key advantages of Gaussian process regression is its ability to handle uncertainty in both input data and model predictions.[25] This is particularly useful when modeling chemical reactions, as experimental data may have noise or uncertainties. The Gaussian process can incorporate these uncertainties during learning and provide predictions along with their confidence intervals.[26]

This study employed Gaussian process regression to model the relationship between reaction conditions (temperature and catalyst equivalents) and the glycosylation reaction yield and stereoselectivity. The Gaussian process model provided optimized reaction conditions that led to improved yields and selectivities by optimizing the kernel parameters and incorporating the uncertainties in the experimental data.

2.3.2 Machine learning optimization

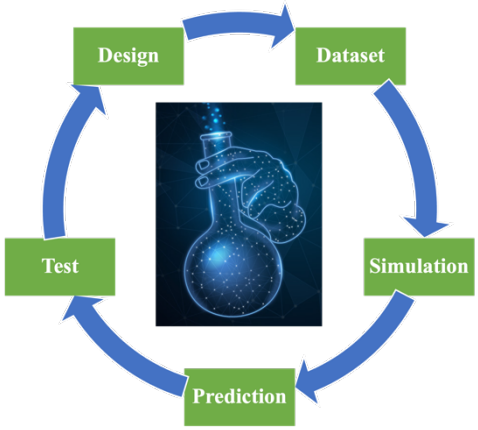


Figure 2.3.2.1 Machine learning optimization procedure

The dataset was processed by taking temperature and catalyst equivalents as independent variables, and the yield of the α-product as the dependent variable. This information was read in array form by a Python program, which utilized Gaussian process regression to fit the data and generate a model. High-confidence points within the model were chosen as optimized conditions for experimental verification.[12] Dr. Takizawa and Prof. Sasai's group have applied machine learning to the reaction optimization of several organic syntheses.[12-14] The program code is based on the code framework from Kondo, M. *et al.*[12,13] and the GPy authors.[15, 16] It has been rewritten, and now includes an automatic peak location function for the model, which facilitates optimization. Additionally, a 3D model display has been added, allowing for a more intuitive observation of the reaction model. The obtained results were added as new data to the dataset, and a new round of optimization was conducted until the highest yield was achieved (Figure 2.3.2.1).

The first round of machine learning simulation results is shown in Figure 2.3.2.2. The blue solid line represents the fitted curve, and the light blue region represents the

confidence interval. High points within the confidence interval were chosen as

Table 2.3.2. Optimization process of glycosylation model study

Entry	Temp.(℃)	Eq(BF_3Et_2O)	L1(%)			
			α	β	α/β	$\alpha+\beta$
10	25	6.2	37	31	1.20	68
11	25	12.7	45	43	1.05	88
12	33	7.4	47	42	1.12	89
13	33	12.5	47	40	1.18	87
14	30	7.4	45	37	1.22	82
15	30	12.1	50	45	1.11	95

* Estimated by H^1 NMR using tetrachloroethane as an internal standard

optimized conditions, with catalyst equivalents of 6.2 and 12.7, and a temperature of 25°C (Table 2.3.2 Entry 10, 11).

After obtaining the experimental results, the data was added to the dataset for the next round of machine learning analysis and optimization, followed by experimental verification. After three rounds of optimization, the optimal glycosylation reaction conditions were obtained (Table 2.3.2 Entry 15): a temperature of 30°C and 12.1 equivalents of catalyst. Under these conditions, the yield of the α-product was 50%. Although this yield improvement was relatively small compared to the initial dataset, the model study confirmed the potential applicability of machine learning in optimizing glycosylation reactions.

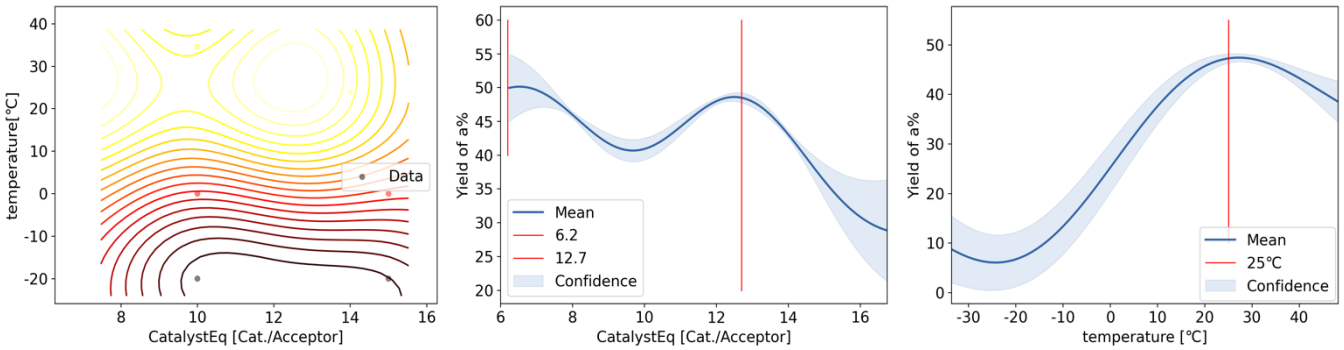


Figure 2.3.2.2. Simulation result of the first run

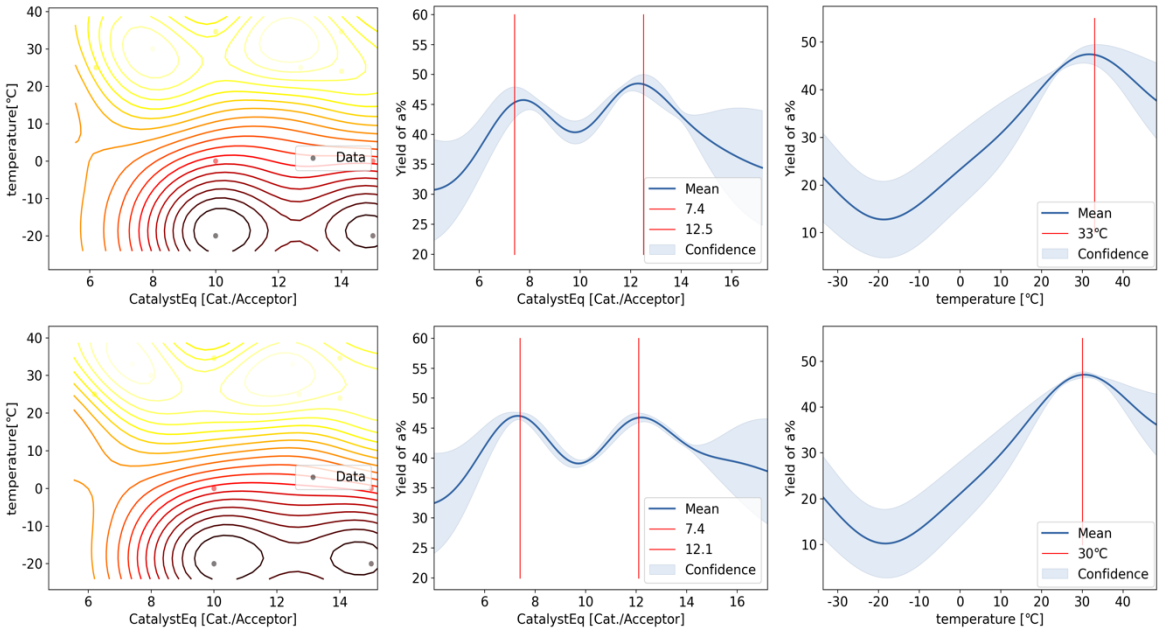


Figure 2.3.2.3 Simulation result of 2nd and 3rd run

Summary

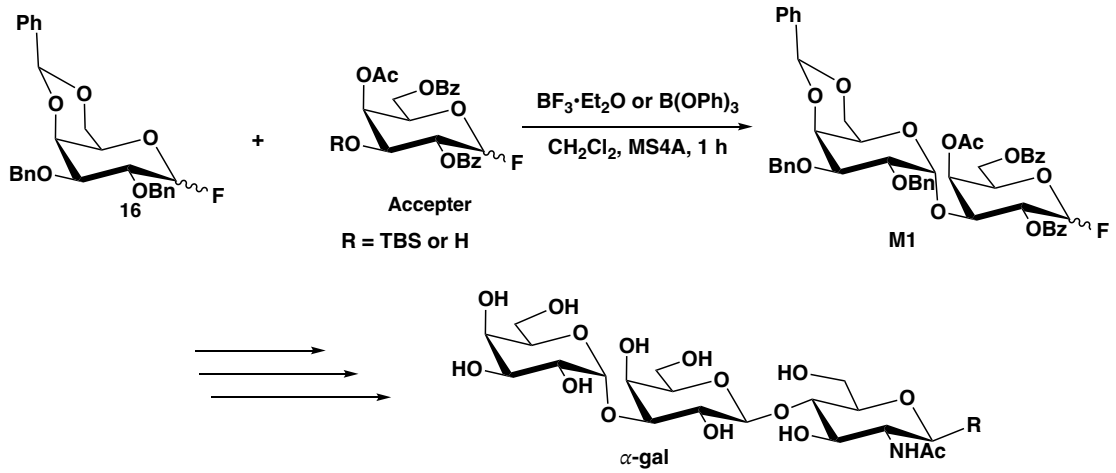
In Chapter 2, the authors conducted a model study to establish a general approach for using machine learning to optimize glycosylation reactions. The temperature and catalyst equivalents, these two parameters of the glycosylation, were optimized. Compared to traditional optimization methods that involve controlling a single variable, the Gaussian process regression machine learning method allows for the simultaneous optimization of two reaction conditions. This approach is more efficient than traditional methods, saving researchers time and effort. The findings of this study confirmed the potential applicability of machine learning in optimizing glycosylation reactions, providing a basis for further research in this area.

Chapter 3. Application to the Chemo-selective glycosylation reaction

In Chapter 3, the author applied the machine learning optimization method developed in Chapter 2 to chemo-selective glycosylation reactions. Chemo-selective glycosylation is a crucial synthetic strategy in glycochemistry for constructing complex glycoconjugates.[27] Achieving high chemo-selectivity in glycosylation reactions is critical for the development of novel pharmaceuticals, vaccines, and other biologically active compounds.[26] For example, the author focused on synthesizing the α-gal epitope, which has significant implications in various biomedical applications.[29, 30]

The α-gal epitope (Galα1-3Galβ1-4GlcNAc-R) is a carbohydrate structure found in various glycoconjugates, such as glycoproteins and glycolipids, and is present in non-primate mammals, prosimians, and New World monkeys.[31] The α-gal epitope has been widely studied due to its immunogenic properties and its role in the immune response to xenotransplantation, tick-borne diseases, and cancer.[32] Humans and Old World monkeys lack the α-gal epitope due to the inactivation of the α1,3-galactosyltransferase (α1,3GT) gene, which leads to the production of anti-α-gal antibodies.[33] These antibodies can cause hyperacute rejection in xenotransplantation procedures and can be exploited for the development of cancer immunotherapies and novel vaccine strategies.[34]

The selected reaction for this study is the first step in the synthesis of the α-gal epitope, which involves the formation of a disaccharide through a glycosylation reaction. (**Scheme 3**) This reaction exhibits α-stereoselectivity due to the presence of a donor with a 2-equatorial phenyl group in a trans-decalin ring system, which hinders the attack of the acceptor from the β-face.[35]



Scheme 3. Synthesis of α-gal epitope

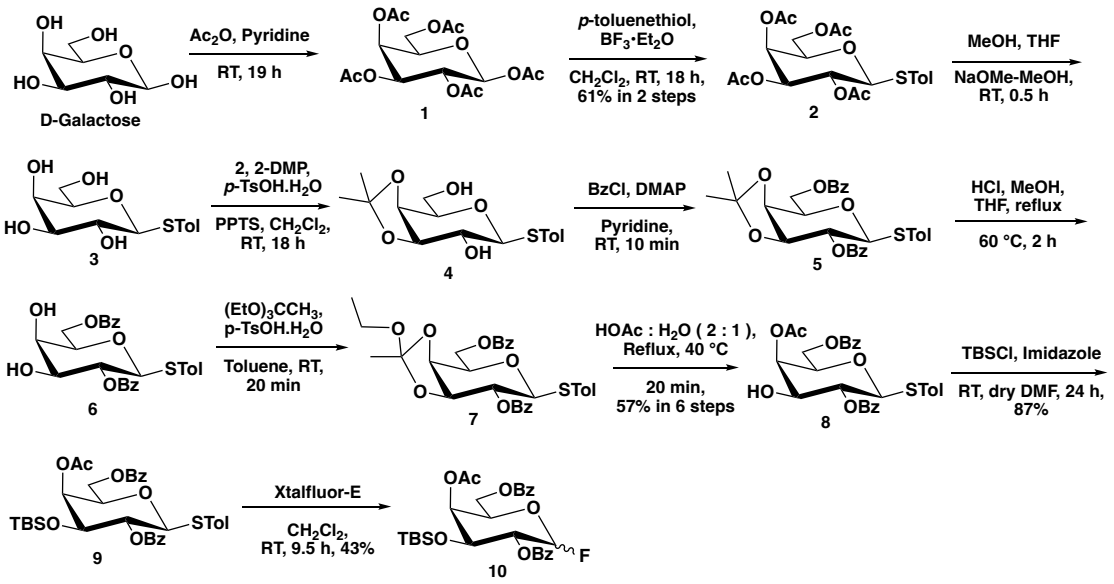
The main object of this chapter is to demonstrate the effectiveness of the machine learning method in optimizing chemo-selective glycosylation reactions by improving the reaction yield and selectivity. To achieve this, the author investigated glycosylations under a series of conditions to get datasets and apply the Gaussian process regression machine learning method to optimize the reaction conditions.

3.1. Donor and Acceptor Synthesis

3.1.1 Synthesis Acceptors 10 and 13

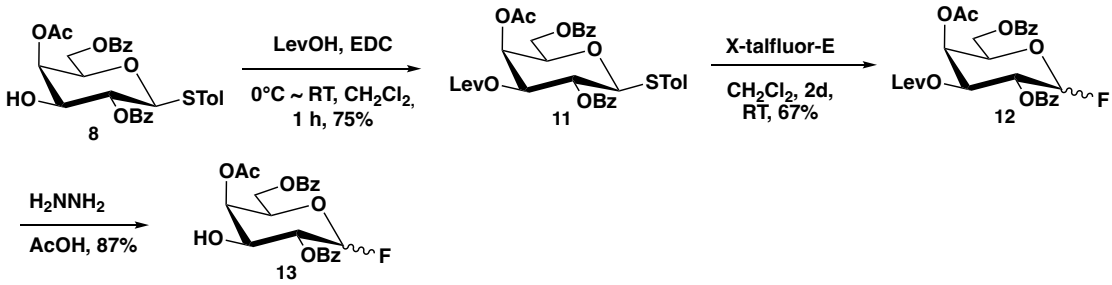
The synthesis of acceptor **10** began with D-galactose as the starting material (Scheme 3.1.1.1). The D-galactose was fully acetylated with acetic anhydride in pyridine, followed by the introduction of *p*-methyl thiophenol at the anomeric position using boron trifluoride diethyl etherate ($BF_3 \cdot Et_2O$) as a Lewis acid, yielding compound **2** with 61% yield. Deprotection of the acetate groups and subsequent isopropylidene protection at the C-3 and C-4 positions with 2,2-dimethoxy propane (2,2-DMP) afforded compound **4**. Benzoylation at the C-2 and C-6 positions of **4**, followed by acid treatment, led to compound **6**, which was further protected as orthoester. Treatment of the obtained **7** with acetic acid resulted in selective ring opening, and after six steps, compound **8** was obtained with 57% yield. TBS protection at the C-3 position of **8** provided **9** with 87% yield, which was then converted to acceptor **10** through a

fluorination reaction with Xtalfluor-E.



Scheme 3.1.1.1. Synthesis of acceptor **10**.

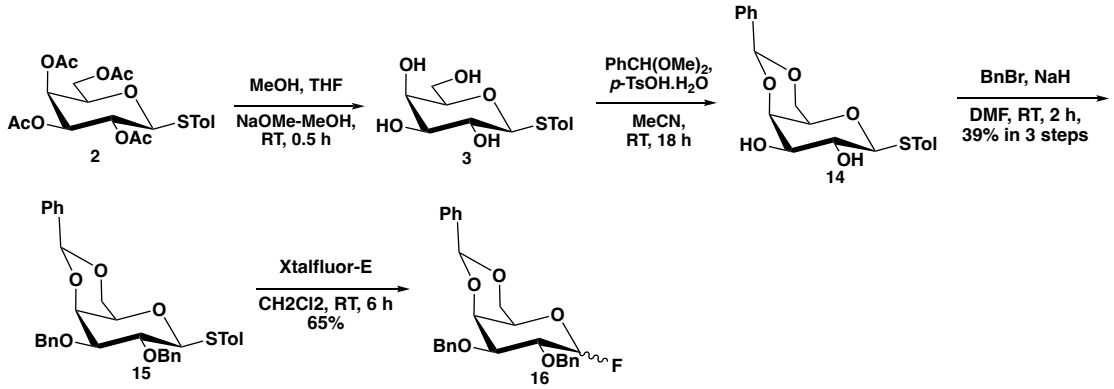
The synthesis of acceptor **13** started with compound **8** (Scheme 3.1.1.2), where the C-3 position was protected with Lev, yielding **11** with 75% yield. A fluorination reaction with Xtalfluor-E provided **12** in 67% yield, and finally, the C-3 position was deprotected with hydrazine in acetic acid to obtain acceptor **13**.



Scheme 3.1.1.2. Synthesis of acceptor **13**.

3.1.2 Synthesis Donor 16

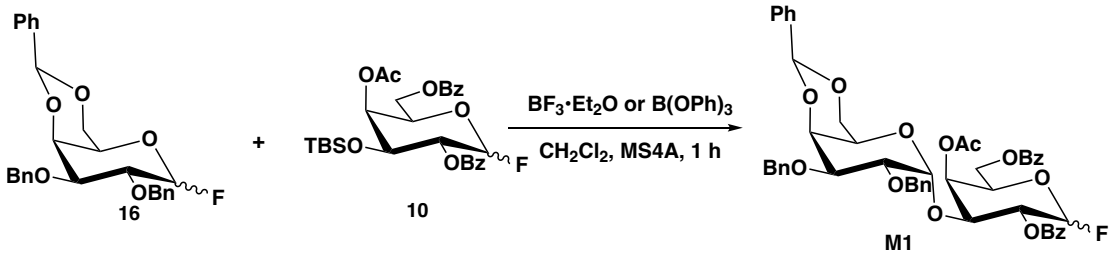
The synthesis of donor **16** began with compound **2** as the starting material (Scheme 3.1.2). After deacetylation, the 4,6-O-benzylidene protection was introduced, affording compound **14**. The remaining hydroxy groups were benzylated by treating **14** with sodium hydride and benzyl bromide in DMF, yielding **15** with 39% yield in three steps. A fluorination reaction with Xtalfluor-E produced glycosyl fluoride donor **16** with 65% yield.



Scheme 3.1.2. Synthesis of glycosyl fluoride donor **16**.

3.2. Conditional Exploration and Data Collection

The author conducted a conditional exploration with employing compound **10** as the glycosylation acceptor and compound **16** as the donor (Scheme 3.2.1, Table 3.2.2). When $BF_3{\cdot}Et_2O$ or $B(OPh)_3$ was used as a catalyst, the disaccharide **M1** yields were relatively low. This could be due to the steric hindrance of the C-3 TBS protecting group in compound **10**, which hampers the attack on the C-1 carbon cation of the donor. To apply the Gaussian process regression machine learning method, an analyzable dataset is required, and a dataset consisting solely of low yields cannot be optimized. Therefore, this dataset is unsuitable for optimizing this glycosylation reaction.



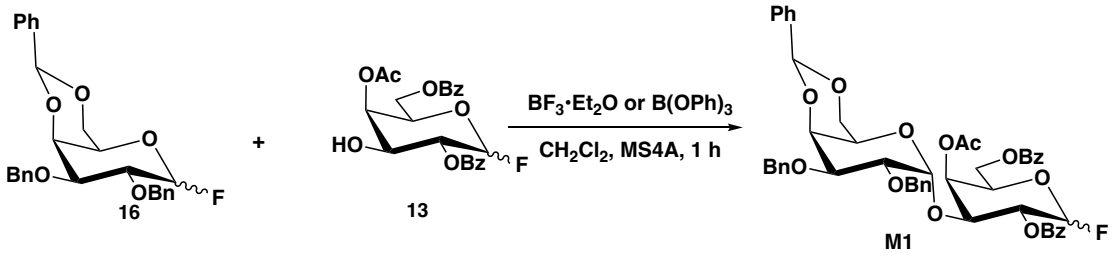
Scheme 3.2.1. Chemo-selective glycosylation by using acceptor **10**.

Table 3.2.2. Condition screening by using acceptor **10**.

Entry	Donor(eq)	$BF_3 \cdot Et_2O$	$B(OPh)_3$	Temp.(℃)	Conc.	Yield(%)
1	1.5	-	0.1eq	-78 to 25	0.02M	trace
2	1.5	-	1.5eq	-78 to 25	0.02M	11
2	1.5+1.5	0.1	-	-78 to 25	0.02M	22
3	1.5	1	-	-78	0.02M	ND
4	1.5	0.1	-	-40	0.02M	~10*
5	1.5	0.1+0.4	-	0	0.02M	~20*
6	3	0.5	-	0	0.02M	~20*
7	1.5	0.1+0.4	-	0	0.1M	~25*
8	1.5	1	-	-40	0.02M	~10*

* Estimated by TLC

Subsequently, the author used compound **13** as the acceptor and compound **16** as the donor for exploring glycosylation reaction conditions (Scheme 3.2.3). When $BF_3 \cdot Et_2O$ and $B(OPh)_3$ was compared, it was found that more by-products were generated by using $B(OPh)_3$. Thus, the author focused on employing $BF_3 \cdot Et_2O$ to investigate the reaction conditions. By altering the catalyst equivalents (0.1 to 1), reaction temperature (-78°C to room temperature), and donor **16** equivalents (1 eq to 1.5 eq), 11 sets of data were obtained (Table 3.2.4, Entries 2-12). Among them, the highest yield, 70%, was achieved at -40°C, with 1 equivalent of the catalyst and 1.5 equivalents of the donor. The data with 1.5 equivalents of donor **16** were selected as the dataset for machine learning (Entries 2-10, 13-15), comprising 12 sets of data.



Scheme 3.2.3. Chemo-selective glycosylation by using acceptor **13**.

Table 3.2.4. Condition screening by using acceptor **13**.

Entry	Catalyst	Eq(Donor)	Eq (Cat.)	Temp.	Yield(%)
1	$B(OPh)_3$	1.5	1	RT	10*
2	$BF_3 \cdot Et_2O$	1.5	0.1	RT	25*
3	$BF_3 \cdot Et_2O$	1.5	0.5	RT	56
4	$BF_3 \cdot Et_2O$	1.5	1	RT	10*
5	$BF_3 \cdot Et_2O$	1.5	0.1	-20℃	25
6	$BF_3 \cdot Et_2O$	1.5	0.5	-20℃	60*
7	$BF_3 \cdot Et_2O$	1.5	1	-20℃	38
8	$BF_3 \cdot Et_2O$	1.5	0.1	-40℃	26
9	$BF_3 \cdot Et_2O$	1.5	0.5	-40℃	66
10	$BF_3 \cdot Et_2O$	1.5	1	-40℃	70
11	$BF_3 \cdot Et_2O$	1.2	1	-40℃	45*
12	$BF_3 \cdot Et_2O$	1	1	-40℃	39
13	$BF_3 \cdot Et_2O$	1.5	1	-60℃	38
14	$BF_3 \cdot Et_2O$	1.5	1	-78℃	26
15	$BF_3 \cdot Et_2O$	1.5	1.5	-78℃	37

* Estimated by TLC

3.3. Machine Learning Optimization and Yield Improvement

The 12 sets of data were used as the dataset and read by the program. The Gaussian process regression machine learning method, established in the model study of Chapter 2, was applied to fit and regress the dataset, optimizing the reaction conditions in a data-driven manner. After three rounds of optimization, the highest yield of 83% was obtained, corresponding to 0.63 equivalents of $BF_3 \cdot Et_2O$ at -36°C. Compared to the highest yield of 70% in the dataset, the glycosylation reaction yield improved by 13% after three rounds of machine learning optimization.

Through this research, the effectiveness of a data-driven glycosylation reaction optimization approach centered on machine learning has been confirmed.

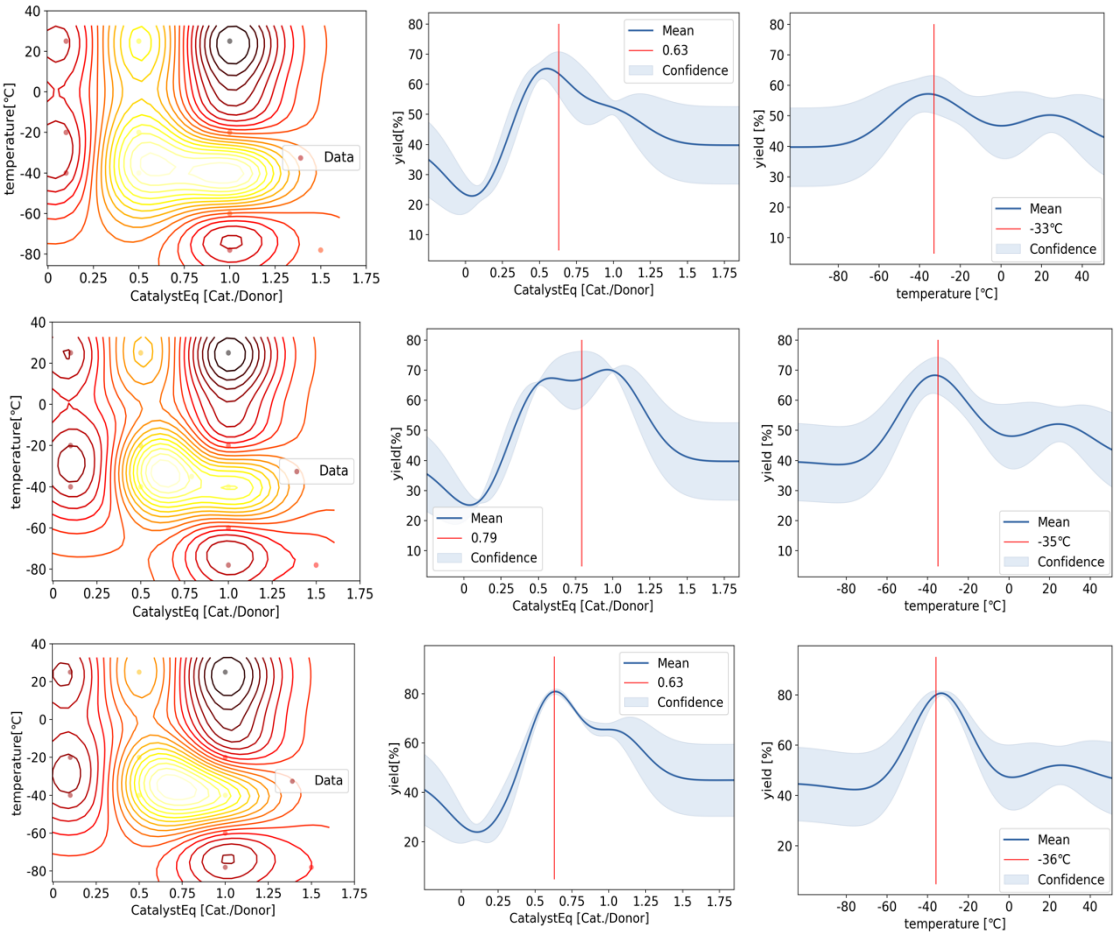


Figure 3.3.1. Simulation result of Gaussian progress regression.

Summary

Table 3.3.2. Machine learning optimization process of chemo-selective glycosylation reaction

Entry	Catalyst	Eq(Donor)	Eq (Cat.)	Temp.	Yield(%)
1	$BF_3 \cdot Et_2O$	1.5	0.63	-33℃	81
2	$BF_3 \cdot Et_2O$	1.5	0.79	-35℃	72
3	$BF_3 \cdot Et_2O$	1.5	0.63	-36℃	83

In Chapter 3, the authors successfully applied the machine learning optimization method developed in Chapter 2 to chemoselective glycosylation reactions. The primary object of this chapter was to demonstrate the effectiveness of the machine learning approach in optimizing chemoselective glycosylation reactions in terms of yield and selectivity. To achieve this, the authors selected a representative chemoselective glycosylation reaction, the synthesis of α-gal, as a case study.

The synthesis of acceptors **10** and **13** and donor **16** was presented in Section 3.1. In Section 3.2, conditional exploration and data collection were conducted, and it was found that acceptor **10** was not suitable for optimization. Consequently, acceptor **13** was used in the glycosylation reaction with donor **16**. A total of 12 sets of data were collected with various reaction conditions, including different catalyst equivalents, temperatures, and donor equivalents.

In Section 3.3, the Gaussian process regression machine learning method was employed to optimize the reaction conditions. After three rounds of optimization, the yield was increased to 83%, with a 13% improvement compared to the highest yield in the dataset. This successful application of the machine learning approach in optimizing chemoselective glycosylation reactions further confirmed its potential in advancing glycochemistry research and the synthesis of complex glycoconjugates.

Chapter 4. α-Selective Glycosylation of Xylose Using Machine Learning

Introduction

Chapter 4 focuses on the application of machine learning to the α-selective xylosylation, fundamental glycosides found in various biomolecules such as hemicellulose and xyloglucans.[36] Xylose is a five-carbon sugar that plays a crucial role in plant cell wall structures and has gained increasing interest in the field of carbohydrate chemistry due to its potential in the production of biofuels, biopolymers, and functional materials.[37] The stereocontrolled synthesis of xylose-containing glycoconjugates is essential for understanding their biological functions and developing novel therapeutic agents.[38]

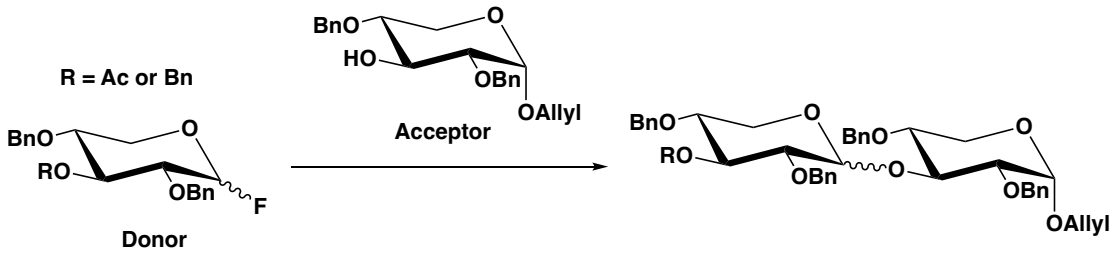
In this chapter, the author implemented the machine learning optimization method, previously established in Chapter 2, to enhance the α-selectivity and yield of xylosylation. By applying the Gaussian process regression machine learning approach to a representative α-selective xylosylation, the authors aim to demonstrate the method's effectiveness in simultaneously optimizing multiple reaction parameters.

The chapter will provide a comprehensive account of the selected xylosylation reaction, the data collection and analysis process, and the machine learning optimization results. The author discusses the impact of various factors, such as temperature, catalyst type, solvent, and protecting groups, on the reaction yield and α-selectivity. The successful application of the machine learning method in optimizing α-selective xylosylation will further validate its potential to advance research in glycochemistry and the synthesis of complex glycoconjugates.

4.1. Initial Batch Analysis and Optimization

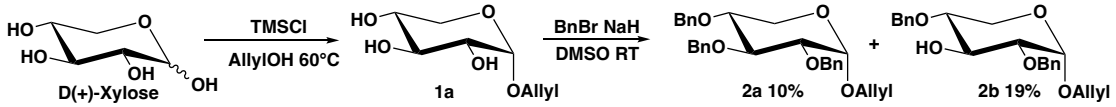
4.1.1 Synthesis of donor and acceptor for xylose glycosylation

In this chapter, α-selective glycosylation of xylose is chosen as the target reaction for investigation. The selected donor and acceptor are shown below (Scheme 4.1.1.1). The synthesis of donor and acceptor is initiated with D-xylose.



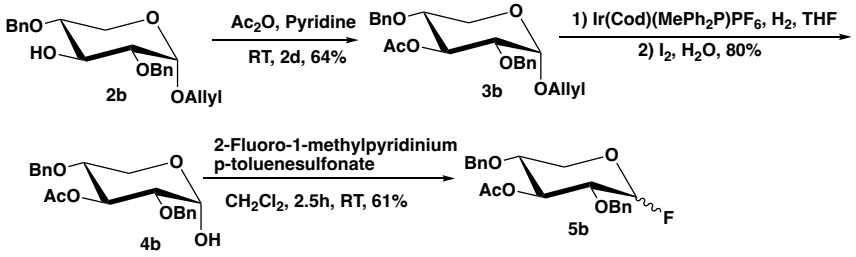
Scheme 4.1.1.1. *α*-Selective xylose glycosylation

Compound **1a** is synthesized by refluxing D-xylose with TMSCl as a catalyst in allyl alcohol, introducing an allyl group at the C-1 position (Scheme 4.1.1.2). The remaining hydroxyl groups are benzylated using sodium hydride and benzyl bromide in DMF to obtain compounds **2a** and **2b** with yields of 10% and 19%, respectively.



Scheme 4.1.1.2 Synthesis of **2a** and acceptor **2b**.

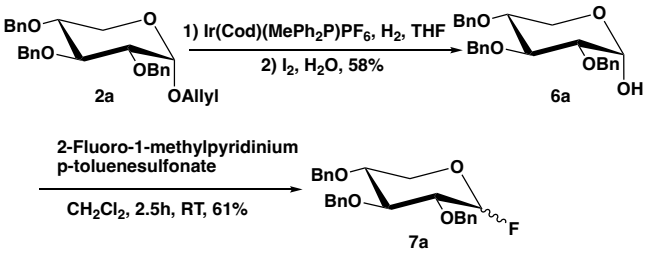
Donor **5b** was synthesized from compound **2b**. Acetylation of the C-3 position affords compound **3b** in 64% yield (Scheme 4.1.1.3). The allyl group at the C-1 position is removed by using an activated Ir (Cod)($MePh_2P$) PF_6 catalyst, yielding compound **4b** in 80% yield. Finally, the C-1 position is fluorinated using 2-fluoro-1-methylpyridinium *p*-toluenesulfonate to obtain donor **5b** with 61% yield.



Scheme 4.1.1.3. Synthesis of donor **5b**.

Donor **7a** was synthesized from compound **2a**. The C-1 allyl group is removed using an activated Ir(Cod)($MePh_2P$)PF_6 catalyst, yielding compound **4b** in 58% yield (Scheme 4.1.1.4). Finally, the C-1 position is fluorinated using 2-fluoro-1-

methylpyridinium *p*-toluenesulfonate to obtain donor **7a** with a 61% yield.



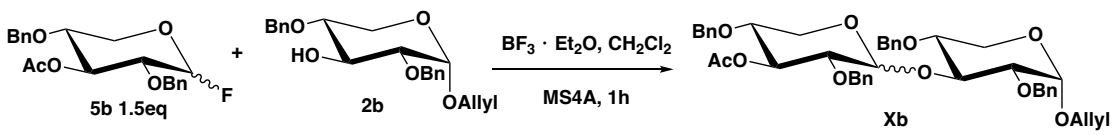
Scheme 4.1.1.4. Synthesis of donor **7a**.

4.1.2 Investigation of xylosylation

Initially, donor **5b** was used in the glycosylation reaction with acceptor **2b**. Reaction condition screening was performed using $BF_3 \cdot Et_2O$ as the catalyst, with temperatures ranging from -30°C to 38°C and catalyst equivalents ranging from 0.2 to 1 eq. Five data points were collected as the dataset, and after two rounds of Gaussian process regression machine learning optimization, no significant improvement in the yield of α-product was observed. The author speculated that low temperatures favor high overall yields but exhibit poor α-selectivity. Although higher temperatures may promote α-product formation, they also increase the likelihood of donor decomposition, leading to lower yields.

Table 4.1.2.1. Xylose glycosylation optimized by using donor **5b**.

No.	Catalyst	Eq（Cat.）	Temp(℃).	Yield(%)	
				α	*β*
1	$BF_3 \cdot Et_2O$	1	-30	26	64
2	$BF_3 \cdot Et_2O$	0.8	10	27	12
3	$BF_3 \cdot Et_2O$	0.5	27	27	16
4	$BF_3 \cdot Et_2O$	0.2	38	34	27
5	$BF_3 \cdot Et_2O$	0.8	38	18	16
6	$BF_3 \cdot Et_2O$	0.25	32	25	40
7	$BF_3 \cdot Et_2O$	0.61	19	32	35



Scheme 4.1.2.1. Xylose glycosylation investigates by using donor **5b**.

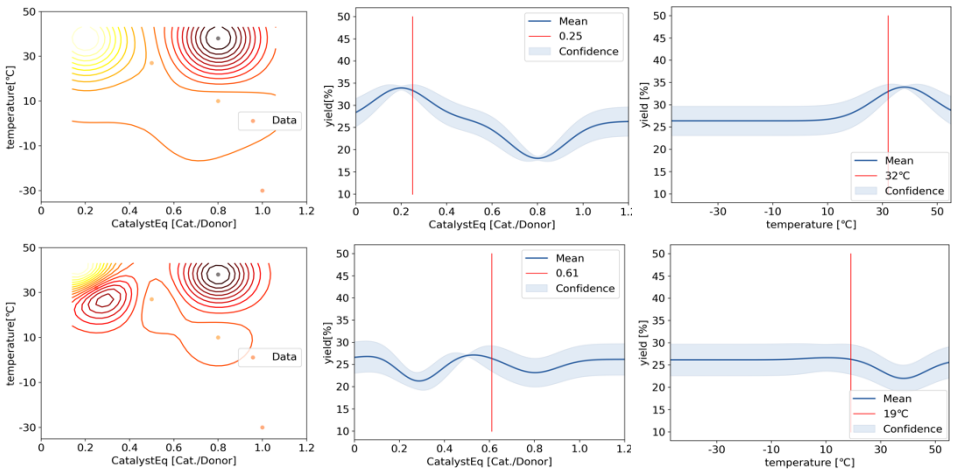
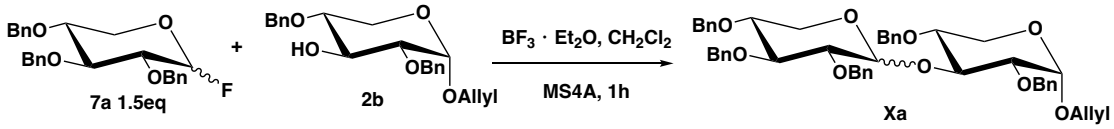


Figure 4.1.2.1. GPR machine learning optimized by using donor **5b**.

Next, the author attempted to use compound **7a** as the donor with $BF_3 \cdot Et_2O$ as the catalyst. The temperature range set between -40 °C and 25 °C, and the catalyst equivalents range from 0.5 to 1.25 eq. Ten data points was collected, and seven of them (Entries 1-7) were used as the input dataset for machine learning optimization. After three rounds of optimization, the highest yield of **α-Xa** was 56.1% with a selectivity of α/β = 1.5/1 (Entry 13) under the conditions of 0.85 eq of catalyst at -30°C. This represents a 6% improvement in yield and enhanced selectivity compared to the highest yield in the input dataset (Entry 1, **α-Xa** yield of 50%, α/β = 1.3/1). In the next chapter, the author introduced the flow chemistry technic to carry out further investigation of this glycosylation reaction.



Scheme 4.1.2.1. Xylose glycosylation investigates by using donor **7a.**

Table 4.1.2.2. Xylose glycosylation optimized by using donor **7a**.

Entry	Catalyst	Eq (Cat.)	Temp (℃).	Yield(%)		α/β*
				α	β	
1	$BF_3 \cdot Et_2O$	0.5	-40	50.3	38.7	1.3/1
2	$BF_3 \cdot Et_2O$	1	-40	49.6	45.0	1.1/1
3	$BF_3 \cdot Et_2O$	0.5	-20	42.0	42.0	1/1
4	$BF_3 \cdot Et_2O$	1	-20	41.9	32.2	1.3/1
5	$BF_3 \cdot Et_2O$	1.25	-20	18.2	52.1	0.35/1
6	$BF_3 \cdot Et_2O$	0.75	0	18.6	18.6	1/1
7	$BF_3 \cdot Et_2O$	1.5	0	9.7	8.8	1.1/1
8	$BF_3 \cdot Et_2O$	0.25	25	11.7	9	1.3/1
9	$BF_3 \cdot Et_2O$	0.5	25	11.8	9.1	1.3/1
10	$BF_3 \cdot Et_2O$	1	25	4.4	3.4	1.3/1
11	$BF_3 \cdot Et_2O$	0.75	-32	54.5	39.0	1.4/1
12	$BF_3 \cdot Et_2O$	0.85	-45	47.5	43.1	1.1/1
13	$BF_3 \cdot Et_2O$	0.85	-30	56.1	37.4	1.5/1

* Estimated by HPLC

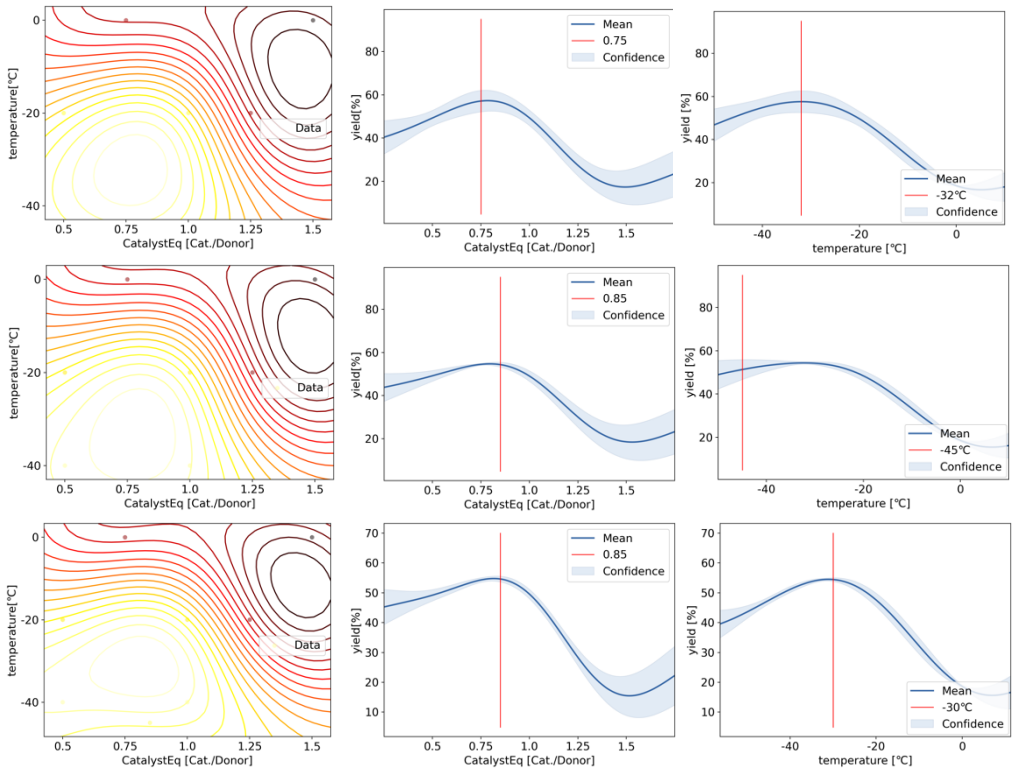


Figure 4.1.2.2. GPR machine learning optimized by using donor **7a**.

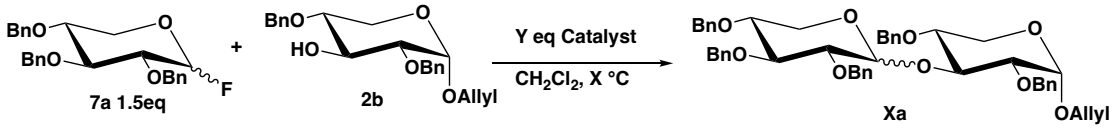
4.2. Flow Chemistry Techniques and Rationale

Flow chemistry, a modern approach to synthetic chemistry, has emerged as a valuable

tool to perform chemical reactions with high efficiency, precise control, and enhanced safety.[39] The technique involves the continuous flow of reactants through a reactor, enabling rapid and efficient mixing, precise control of residence time, and enhanced heat and mass transfer.[40] The application of flow chemistry in glycosylation reactions has demonstrated significant improvements in selectivity and yield, particularly in challenging reactions.[41]

In the context of α-selective xylosylation, flow chemistry offers several advantages over traditional batch reactions. First, it enables better control of reaction conditions such as temperature, which is critical for achieving optimal selectivity and yield.[42] Second, it allows for rapid screening of reaction conditions, accelerating the optimization process and reducing the amount of starting materials and reagents required.[40] Finally, flow chemistry reactions can be easily scaled up, enabling the synthesis of larger quantities of desired products while maintaining the reaction conditions and product quality.[41] Here the author applied flow chemistry techniques to the α-xylosylation to improve the reaction yield and selectivity using the aforementioned advantages.

4.3. Machine Learning Optimization and Selectivity Improvement



Scheme 4.3.1.1. Catalyst selection for ML optimization.

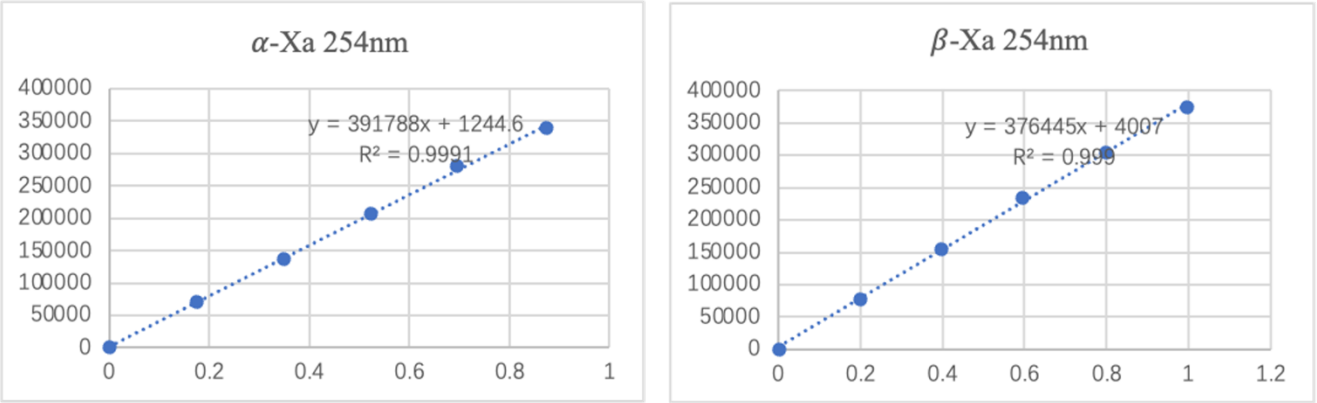
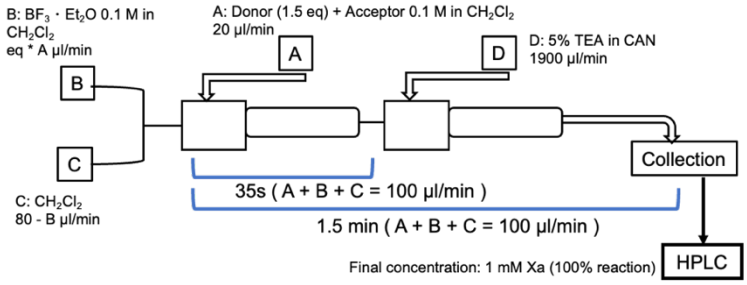


Figure 4.3.1.2. Liner regression of **Xa** by HPLC.



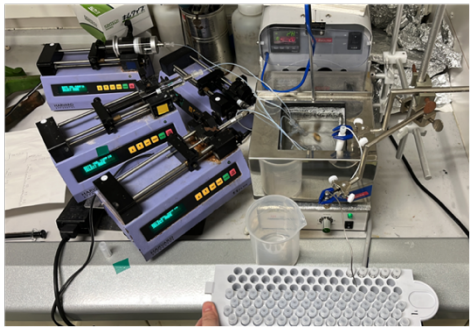


Figure 4.3.1.1. Microfluidic system for optimization.

In this section, the author integrated flow chemistry techniques to optimize the xylosylation reaction. The flow chemistry setup is shown in Figure 4.3.1.1. To efficiently perform high-throughput screening, the author employed HPLC analysis to determine the yields of **α-Xa** and **β-Xa**. The quantification curve was shown in Figure 4.3.1.2.

4.3.1 Catalyst selection

When using a flow chemistry device for xylosylation, donor **7a** (1.5eq) and acceptor **2b** are first dissolved in a solvent to make a 0.1 M solution, which is fed into a Micromixer by pump A. The catalyst and solvent are fed separately by pumps B and C, respectively, to mix with the reactants and initiate the reaction. After 35 seconds, the mixture reaches a second Micromixer where it is quenched by mixing with triethylamine from pump D. The system can adjust the catalyst equivalents by adjusting the flow rates of the pumps. Samples are taken from the final outlet and analyzed by HPLC to determine the yield and ratio of α and β configuration.

The author tested four catalysts, $BF_3 \cdot Et_2O$, $B(C_6F_5)_3$, TfOH, and TMSOTf, and investigated the reaction temperature and catalyst equivalents for each (Scheme 4.3.1.1). With $BF_3 \cdot Et_2O$ as the catalyst, a temperature of -20 ℃ and a catalyst equivalent of 1.0 eq resulted in the highest **α-Xa** yield of 42.5%; however, the selectivity was poor (α/β=1.1/1, Table 4.3.1.1).

Table 4.3.1.1. Xylose glycosylation investigation by using $BF_3 \cdot Et_2O$ as catalyst.

$BF_3 \cdot Et_2O$					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	0.8	1.3	0.7
2	-40	0.5	6.7	17.6	0.4
3	-40	1.0	9.1	24.1	0.4
4	-20	0.1	18.3	21.1	0.9
5	-20	0.5	30.5	45.2	0.7
6	-20	1.0	42.5	40.2	1.1

* Estimated by HPLC

When using $B(C_6F_5)_3$ as the catalyst, the optimal reaction conditions were a temperature of -20 ℃ and a catalyst equivalent of 0.1 eq, which resulted in an **α-Xa** yield of 32.8% and selectivity of α/β=1.1/1(Table 4.3.1.2, Entry 6).

Table 4.3.1.2. Xylose glycosylation investigation by using $B(C_6F_5)_3$ as catalyst.

$B(C_6F_5)_3$					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	0.9	0.5	0.8
2	-40	0.5	14.1	17.1	0.8
3	-40	1.0	20.6	23.9	0.9
4	-20	0.1	6.3	8.2	0.8
5	-20	0.5	27.0	28.9	0.9
6	-20	1.0	32.8	30.1	1.1
7	0	0.1	30.7	29.2	1.1
8	0	0.5	28.7	18.6	1.5
9	0	1.0	14.3	13.2	1.1
10	20	0.1	15.7	11.4	1.3
11	20	0.5	5.2	5.8	0.8
12	20	1.0	2.0	3.4	0.5

* Estimated by HPLC

With TfOH as the catalyst, the optimal conditions were a temperature of 0 ℃ and a catalyst equivalent of 1.0 eq, yielding an **α-Xa** yield of 32.0% and selectivity of α/β=1.5/1(Table 4.3.1.3, Entry 9).

Table 4.3.1.3. Xylose glycosylation investigation by using TfOH as catalyst.

TfOH					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	3.7	2.5	1.4
2	-40	0.5	4.9	3.0	1.5
3	-40	1.0	7.7	5.2	1.4
4	-20	0.1	12.0	10.3	1.2
5	-20	0.5	16.2	13.3	1.2
6	-20	1.0	20.8	16.5	1.3
7	0	0.1	11.1	10.4	1.1
8	0	0.5	25.9	22.0	1.2
9	0	1.0	32.0	22.3	1.5

* Estimated by HPLC

Among the four catalysts, TMSOTf produced the best results, with optimal conditions being a temperature of 0 ℃ and a catalyst equivalent of 1.0 eq, yielding an **α-Xa** yield of 43.1% and selectivity of α/β=1.9/1 (Table 4.3.1.4, Entry 9). Thus, TMSOTf was selected as the catalyst, and the reaction data using this catalyst were used as the dataset for machine learning reaction optimization.

Table 4.3.1.4. Xylose glycosylation investigation by using TMSOTf as catalyst.

TMSOTf					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	4.0	3.7	1.1
2	-40	0.5	4.1	2.3	1.6
3	-40	1.0	5.8	2.8	1.9
4	-20	0.1	8.5	5.2	1.6
5	-20	0.5	15.9	10.9	1.5
6	-20	1.0	18.1	12.4	1.5
7	0	0.1	20.1	15.9	1.3
8	0	0.5	35.3	24.5	1.5
9	0	1.0	43.1	22.8	1.9

* Estimated by HPLC

4.3.2 Machine Learning Optimization

Nine data sets using TMSOTf as the catalyst was used, with temperatures ranging from -40 ℃ to 0 ℃ and catalyst equivalents from 0.1 to 1 eq. The data-driven machine learning optimization method established in Chapter 2 was employed to fit and optimize the dataset (Table 4.3.2.1). However, after one round of optimization, it was found that the temperature range of the nine datasets was insufficient. The authors added six datasets to supplement the original data (Table 4.3.2.1, entries 11-16), expanding the temperature range from -40 ℃ to 20 ℃. A total of 16 datasets, including the results

from the first round of optimization, were used for machine learning optimization. After three rounds of optimization, the best reaction conditions using TMSOTf as the catalyst were determined to be 2 ℃ and 1 eq of catalyst, resulting in an **α-Xa** yield of 42.2% and selectivity of α/β=2.4/1 (Table 4.3.2.1, Entry 19). Compared to the best reaction conditions in the dataset (Table 4.3.2.1, Entry 9), the optimized conditions resulted in a 1% lower **α-Xa** yield but improved selectivity from α/β=1.9/1 to α/β=2.4/1.

Table 4.3.2.1. Xylose glycosylation optimization by ML.

TMSOTf					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	4.0	3.7	1.1
2	-40	0.5	4.1	2.3	1.6
3	-40	1.0	5.8	2.8	1.9
4	-20	0.1	8.5	5.2	1.6
5	-20	0.5	15.9	10.9	1.5
6	-20	1.0	18.1	12.4	1.5
7	0	0.1	20.1	15.9	1.3
8	0	0.5	35.3	24.5	1.5
9	0	1.0	43.1	22.8	1.9
10	10	0.85	33.8	16.2	2.1
11	10	0.1	25.7	22.9	1.1
12	10	0.5	36.1	18.8	1.9
13	10	1.0	36.4	17.2	2.1
14	20	0.1	28.0	22.1	1.3
15	20	0.5	22.3	10.4	2.1
16	20	1.0	10.0	4.9	1.8
17	2	1.1	42.1	17.6	2.4
18	2	0.9	40.3	16.7	2.4
19	2	1.0	42.2	17.2	2.4
20	2	1.5	35.7	14.6	2.4

* Estimated by HPLC

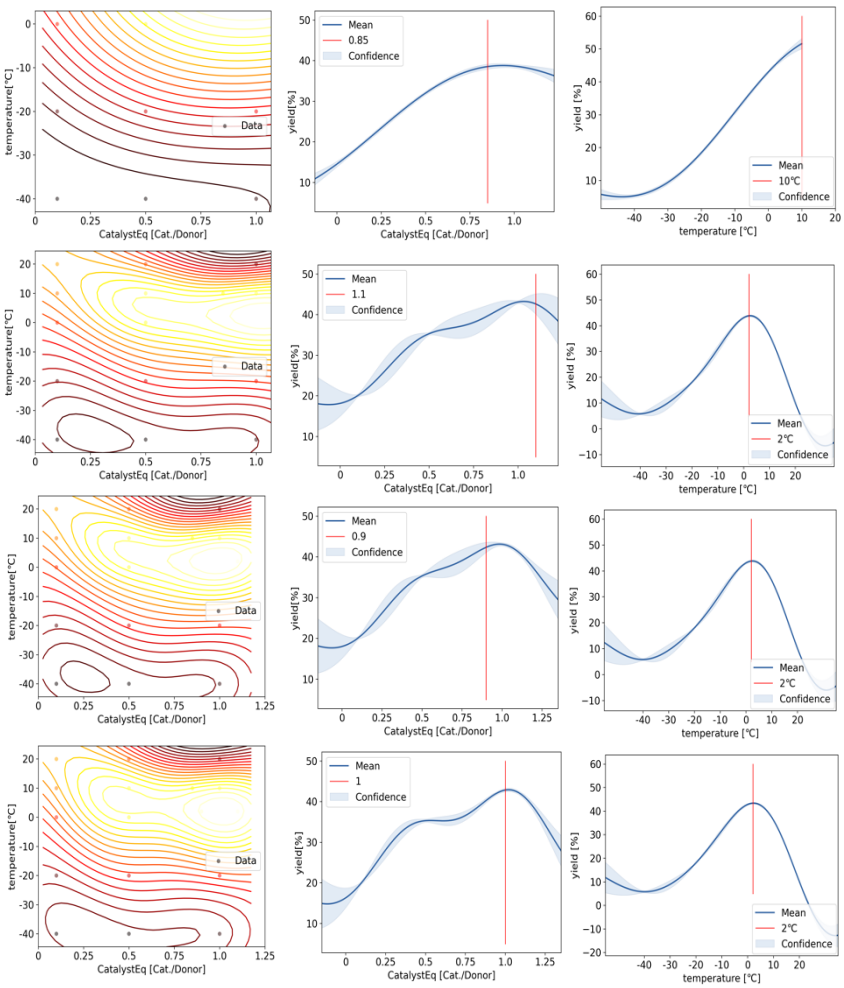


Figure 4.3.2.1. Xylose glycosylation optimization by ML.

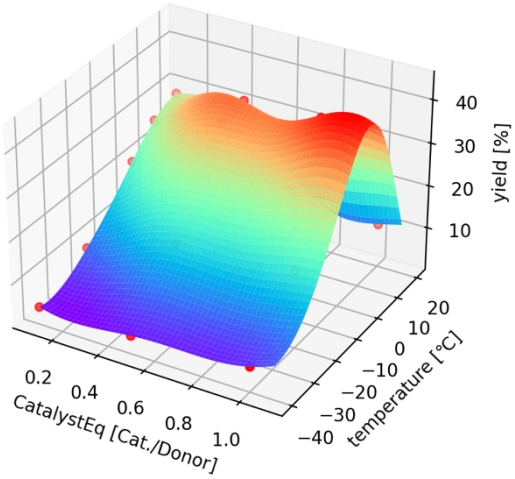


Figure 4.3.2.2. 3D ML model.

Summary

In this chapter, the author explored the α-selective xylosylation, a challenging and complex reaction. The author first synthesized donor and acceptor molecules and screened different catalysts for the glycosylation reaction. After identifying TMSOTf as the most effective catalyst, they expanded the reaction dataset to include a broader temperature range and optimized the reaction conditions through machine learning. The optimization process led to improved selectivity, yielding an **α-Xa** yield of 42.2% and selectivity of α/β=2.4/1.

This chapter demonstrates the potential of using machine learning algorithms and flow chemistry techniques to optimize challenging glycosylation reactions in organic chemistry. The approach not only streamlines the process but also provides valuable insights for enhancing reaction selectivity, which is crucial for the synthesis of complex molecules with multiple stereocenters.

Chapter 5. Multi-Factor Machine Learning Optimization

5.1. Machine Learning Optimization Strategy

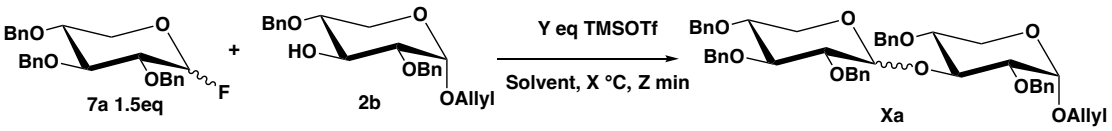
In this section, the author introduces the multi-factor machine learning optimization strategy utilizing Bayesian regression machine learning methods.[19] This approach allows us to investigate the effects of various factors, such as temperature, catalyst equivalents, and time, on the α-selective xylosylation. The use of TMSOTf as the catalyst and flow chemistry techniques further enhance the efficiency of the optimization process.

In recent years, Bayesian regression techniques, such as Gaussian process regression (GPR),[19] have been employed in various optimization problems. GPR is a non-parametric, probabilistic model that provides predictions and uncertainty estimates for function values. This makes it particularly suitable for optimization tasks, as it can guide the search toward promising regions of the parameter space.[19]

In the context of organic chemistry, multi-factor optimization using machine learning methods like GPR can significantly accelerate reaction development and improve selectivity.[43] By integrating flow chemistry techniques, high-throughput experimentation can be conducted, enabling rapid data collection and analysis. This combination of machine learning optimization and flow chemistry has shown great potential in enhancing the discovery and optimization of chemical reactions.[43]

In my study, the multi-factor machine learning optimization strategy is applied to investigate the effects of temperature, catalyst equivalents, and time on the α-xylosylation using TMSOTf as the catalyst. The data collected through flow chemistry experiments are used to train the GPR model, which in turn guides the optimization process to find the most suitable conditions for achieving the desired selectivity.

5.2 Investigating the Effects of Temperature, Catalyst Equivalents, and Time



Scheme 5.2.1. Xylosylation investigation by solvent, and catalyst equivalent.

In this section, the author explored the impact of various solvents (MTBE, Et_2O, ACN, and Toluene) on the α-selective xylosylation using TMSOTf as the catalyst. The reaction temperature and catalyst equivalents were systematically investigated for each solvent (Scheme 5.2.1). When using MTBE and Et_2O, two ether solvents, the reaction activity was weak (Table 5.2.2, Table 5.2.2), potentially due to the fact that ethers can act as Lewis bases and donate electron pairs to a Lewis acid, thus inhibiting the Lewis acid's activity.

Table 5.2.2. Xylose glycosylation investigation by using MTBE as solvent.

TMSOTf in MTBE					
Entry	Temp(℃)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-20	0.5	-	-	-
2	-20	1.0	-	-	-
3	-20	3.0	-	-	-
4	0	0.1	-	-	-
5	0	0.5	-	-	-
6	0	1.0	0.1	-	-
7	0	3.0	0.1	-	-
8	20	0.1	0.2	0.1	0.9
9	20	0.5	0.2	-	-
10	20	1.0	0.1	0.0	0.7
11	20	3.0	0.3	0.1	0.9

* Estimated by HPLC

Table 5.2.3. Xylose glycosylation investigation by using Et_2O as solvent.

TMSOTf in Et_2O					
Entry	Temp(°C)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-40	0.1	0	0	-
2	-40	0.5	0	0	-
3	-40	1.0	0	0	-
4	-40	2.0	0	0	-
5	-40	4.0	0	0	-
6	-20	0.1	0	0	-
7	-20	0.5	0	0	-
8	-20	1.0	0	0	-
9	-20	2.0	0	0	-
10	-20	4.0	0	0.4	-
11	0	0.1	0	0	-
12	0	0.5	0	0	-
13	0	1.0	0.2	0.3	0.6
14	0	2.0	0.5	1.1	0.5
15	0	4.0	1.6	2.7	0.6
16	19	0.1	7.0	12.1	0.6
17	19	0.5	1.1	2.5	0.5
18	19	1.0	1.7	4.2	0.4
19	19	2.0	4.2	9.9	0.4
20	19	4.0	5.8	13.0	0.5
21	0	10	4.7	7.0	0.7
22	0	20	4.9	6.5	0.7
23	0	30	7.4	9.5	0.8
24	20	10	7.1	12.9	0.6
25	20	20	7.5	15.0	0.5
26	20	30	11.2	19.6	0.6
27	30	10	7.3	18.3	0.4
28	30	20	8.3	19.4	0.4
29	30	30	10.9	22.2	0.5

* Estimated by HPLC

The author then used acetonitrile (ACN) as a solvent and optimized the reaction conditions within a temperature range of -20 to 20 °C and catalyst equivalents of 0.1 to 1.5 eq. The optimal conditions were found at 0 °C and 1 eq of catalyst, yielding a 43.7% **α-Xa** yield with a selectivity ratio of α/β = 2.3/1 (Table 5.2.4, Entry 4). Lastly, using toluene as a solvent, the authors investigated a temperature range of -20 to 65 °C and catalyst equivalents from 0.1 to 1.7 eq. The optimal conditions were found at 20 °C and 1.5 eq of catalyst, yielding a 39.9% **α-Xa** yield with a selectivity ratio of α/β = 2.4/1 (Table 5.2.5, Entry 12). The optimal conditions in the solvents ACN and Toluene yielded similar results, with slightly better selectivity observed in Toluene.

Table 5.2.4. Xylose glycosylation investigation by using ACN as solvent.

TMSOTf in ACN					
Entry	Temp (°C)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-20	1.0	33.6	40.4	0.8
2	0	0.1	10.0	19.9	0.5
3	0	0.5	43.5	24.9	1.8
4	0	1.0	43.7	18.9	2.3
5	20	0.1	17.6	19.8	0.9
6	20	0.5	31.0	16.0	2.0
7	20	1.0	11.0	4.2	2.5
8	-2	0.45	20.3	28.0	0.7
9	-2	0.55	30.7	30.1	1.0
10	-2	0.95	41.2	23.2	1.8
11	-2	1.05	42.9	22.5	1.9
12	2	0.45	31.9	29.1	1.1
13	2	0.55	35.8	26.5	1.4
14	2	0.95	40.4	19.5	2.1
15	2	1.05	41.8	19.6	2.1
16	4	0.65	17.6	26.9	0.7
17	4	1.0	27.1	27.0	1.0
18	4	1.5	32.2	22.5	1.4

Table 5.2.5. Xylose glycosylation investigation by using Toluene as solvent.

TMSOTf in Toluene					
Entry	Temp (°C)	Eq	Yield(α)	Yield(β)	α/β(HPLC)
1	-20	0.1	-	0.1	-
2	-20	0.5	2.9	13.3	0.2
3	-20	1.0	4.3	18.9	0.2
4	-10	0.5	3.6	17.8	0.2
5	0	0.1	0.4	2.7	0.2
6	0	0.5	6.0	22.6	0.3
7	0	1.0	10.2	28.1	0.4
8	10	0.5	9.1	26.8	0.3
9	20	0.1	2.6	18.2	0.2
10	20	0.5	18.1	30.6	0.6
11	20	1.0	38.2	27.4	1.4
12	20	1.5	39.9	16.2	2.4
13	35	0.5	18.1	8.1	2.2
14	35	1.0	6.8	3.3	1.9
15	35	1.5	4.7	3.0	1.4
16	50	0.5	7.0	3.1	2.0
17	50	1.0	1.7	1.0	1.3
18	50	1.5	0.6	0.4	0.9
19	65	0.5	0.4	0.1	1.2
20	65	1.0	-	-	-
21	65	1.5	-	-	-
22	15	1.35	37.0	29.0	1.3
23	15	1.58	41.0	22.9	1.8
24	25	1.5	37.6	15.2	2.4
25	22	1.15	34.2	14.4	2.3
26	22	1.72	34.9	14.2	2.4
27	6	1.07	20.1	32.2	0.6
28	6	1.68	27.2	35.2	0.8

5.3. Improved α-Selectivity and Yield

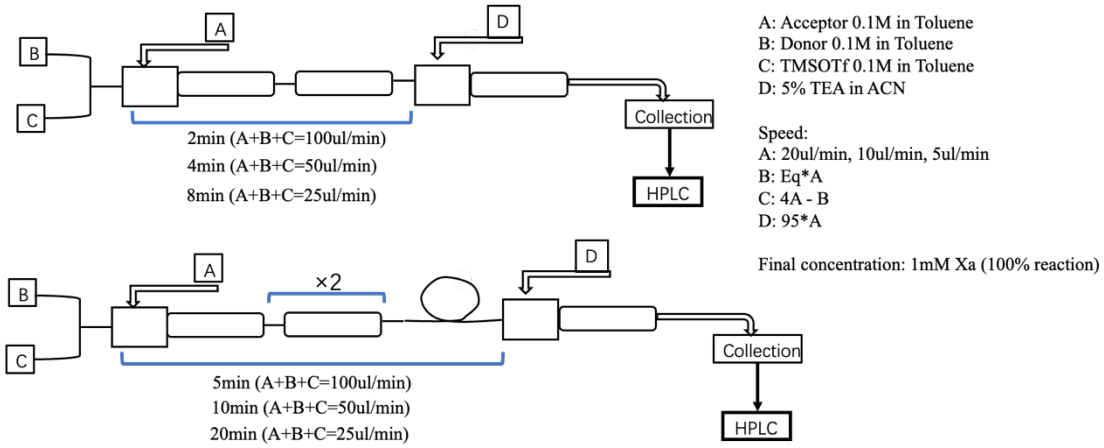


Figure 5.2.6. Microfluidic system for optimization.

The author selected toluene as the solvent and investigated the influence of reaction time on yield and selectivity. The author modified the flow chemistry setup by adding a mixer, extending the pipeline length, and adjusting the flow rate to control the reaction time (Figure 5.2.6). The author explored a range of reaction times (2-20 min, Table 5.2.7) and temperatures (-40 to -10 °C), obtaining 12 sets of data. The optimal reaction conditions were found at -10 °C, 1 eq of TMSOTf, and a reaction time of 20 min, yielding a 56.8% **α-Xa** yield with a selectivity ratio of α/β = 3.9/1 (Table 5.2.7, Entry 12).

Table 5.2.7. Optimizied of temperature and time.

Entry	Temp (°C)	Cat.Eq	Time (min)	Yield(α)	Yield(β)	α/β
			+ 2 Mixer			
1	-40	1	2	0.9	4.3	0.2
2	-40	1	4	2.5	9.2	0.3
3	-40	1	8	4.1	15.4	0.3
4	-20	1	2	3.9	17.4	0.2
5	-20	1	4	5.2	20.7	0.3
6	-20	1	8	8.1	26.6	0.3
			+ 2 Mixer + 1 m tube			
7	-40	1	5	4.1	14.7	0.3
8	-40	1	10	5.0	18.2	0.3
9	-40	1	20	5.8	20.3	0.3
10	-10	1	5	19.3	39.2	0.5
11	-10	1	10	53.5	26.0	2.0
12	-10	1	20	56.8	14.4	3.9

5.3. Improved α-Selectivity and Yield

The 28 data sets from Table 5.2.5 were combined with the new 12 data sets to form 40 data sets, which were used for Bayesian machine learning optimization. The program code is based on the code framework from Kondo, M. *et al*.[12,13] and The GPy authors.[15, 16] The optimization process included time, temperature, and catalyst equivalents as independent variables, and the **α-Xa** yield as the dependent variable (Table 5.3.1, Entries 1-2). After incorporating the optimization results and supplementary data (Table 5.3.1, Entries 3-5), a total of 45 data sets were used for optimization. The optimal reaction conditions were found to be -10 °C, 1.6 eq of TMSOTf, and a reaction time of 20 min, yielding a 60.5% **α-Xa** yield with a selectivity ratio of α/β = 3.2/1 (Table 5.3.1, Entry 6). The reaction was then scaled up (with 50 mg of acceptor 7a), resulting in a selectivity ratio of α/β = 4.0/1 and a 54.5% **α-Xa** yield.

Table. 5.3.1. Multifactor ML optimization of Xylose glycosylation

Entry	Temp (°C)	Cat.Eq	Time (min)	Yield(α)	Yield(β)	α/β
1	-10	0.54	20	40.7	29.6	1.4
2	-10	1.43	20	56.4	16.8	3.3
3	-30	2	20	4.4	12.4	0.4
4	0	1	20	26.3	10.2	2.5
5	0	0.5	10	48.2	18.5	2.6
6	-10	1.6	20	60.5	18.5	3.2
7	-10	2	20	57.9	16.1	3.6

* Estimated by HPLC

Summary

In Chapter 5, the author focused on optimizing the α-selective xylosylation using a multi-factor machine learning method. The author investigated the effects of solvent, time, and catalyst equivalents on the reaction, utilizing TMSOTf as the catalyst. The authors found that the optimal solvent was toluene, with improved α-selectivity and yield achieved under the conditions of -10 °C, 1.6 eq of TMSOTf, and a reaction time of 20 minutes. A Bayesian machine learning approach was employed to analyze the data, optimize the reaction conditions, and achieve higher **α-Xa** yields and selectivity ratios. The results of this chapter provide a better understanding of the factors influencing the α-selective glycosylation reaction and offer valuable insights for further

research and application in organic chemistry.

Summary

1. Summary of Key Findings

This study has demonstrated the effectiveness of machine learning, specifically Gaussian process regression, in optimizing glycosylation reactions. Through various chapters, the machine learning optimization method was applied to different glycosylation reactions, such as the synthesis of the α-gal epitope and the α-selective xylosylation. The key factors affecting yield and selectivity, including temperature, catalyst type, solvent, and protecting groups, were analyzed, and their impact on chemo-selectivity was discussed.

2. Potential Applications and Future Directions

The successful application of machine learning to optimize chemo-selective glycosylation reactions paves the way for several potential applications and future research directions in glycochemistry:

Expansion to other glycosylation reactions: The machine learning optimization method can be applied to other glycosylation reactions, including various donor and acceptor molecules, to enhance the efficiency and selectivity of glycoconjugate synthesis.

Integration with other computational approaches: The machine learning method could be integrated with other computational approaches, such as quantum chemistry calculations and molecular dynamics simulations, to gain a deeper understanding of the reaction mechanisms and improve predictive accuracy.

Automation and high-throughput screening: Machine learning optimization could be combined with automated synthesis platforms and high-throughput screening techniques to accelerate the discovery and optimization of new glycosylation reactions and catalysts.

Development of novel therapeutic agents: Optimized glycosylation reactions can contribute to the synthesis of complex glycoconjugates, which are essential for understanding the biological functions of these molecules and developing novel therapeutic agents, including pharmaceuticals, vaccines, and immunotherapies.

Biofuels and biopolymers: The optimization of xylose glycosylation, for example, can have significant implications for the production of biofuels, biopolymers, and functional materials, as well as for understanding the biological functions of xylose-containing glycoconjugates.

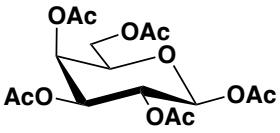
Overall, the application of machine learning optimization in glycochemistry offers promising avenues for advancing research and the development of novel, biologically relevant compounds. By refining and expanding the use of these methods, future studies can continue to unlock the potential of glycosylation reactions and improve our understanding of the complex world of glycoconjugates.

Experimental

General Information

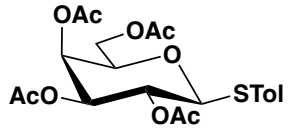
^{1}H NMR spectra were recorded in an indicated solvent with JEOL ECA 500 MHz spectrometer equipped with a cryoprobe. The chemical shifts in $CDCl_3$ are given δ values from TMS as an internal standard. Multiplicities abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiple, br = broad. Chemical purification was carried out using silica-gel column chromatography Keiselgel 60 F_{254} (Merk Co., 0.043-0.063 mm) and Sephadex™ LH-20. TLC analysis of compounds was visualized by UV (254 nm) with p-methoxy benzaldehyde (p-anisaldehyde, 0.03% in the mixture of EtOH; H_2SO_4; AcOH). Anhydrous CH_2Cl_2 was distilled from calcium hydride. Anhydrous DMF was purchased from WAKO (Wako Pure Chemical Industries, Ltd.). The molecular sieve (MS4Å) was activated with microwave heating at 250 °C and dried in vacuo 3 times before use. A chemical reaction using an anhydrous solvent was carried out under an argon atmosphere (Ar). All other commercially available reagents and solvents were used as purchased. HPLC was carried out with a Shimadzu LC-20AD liquid chromatograph and analyzed by HPLC with a Cosmosil 5C18-P (Nacalai Tesque) column [4.6IDX250 mm: solvent: 50 mM ammonium acetate/MeCN; flow rate: 1.0 ml min-^{1}: the peaks were detected with a Shimadzu fluorescence spectrophotometer by using an excitation wavelength of 254 nm.

Synthesis procedure and characterization data



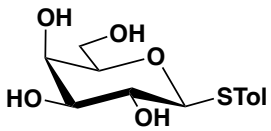
1,2,3,4,6-Penta-O-acetyl-D-Galactopyranoside (1). In an evaporating flask at rt under Ar atmosphere, a solution of D-galactose (50 g, 0.528 mol) was dissolved in pyridine super dehydrate (150 mL, 1.67 mol) and treated with acetic anhydride (150 mL, 1.67 mol). After being stirred for 19 h, the reaction mixture was concentrated in

vacuo and co-evaporated with toluene 4 times affording the crude product **1** as a yellow oil. Rf 0.43 (n-hexane/EtOAc=1/1).



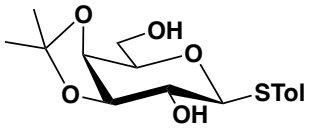
***p*-Tolyl-2,3,4,6-tetra-O-acetyl-1-thio-B-D-Galactopyranoside (2).** In an evaporating flask at r under Ar atmosphere, the yellow oil of compound **1** (0.56 mol) in super dehydrated CH_2Cl_2 (350 mL, 5.48 mol) was added to heated p-toluene thiol (41 mL, 0.34 mol, heated at 60 °C) followed by cooling to 0 °C. After being stirred for 10 minutes, $BF_3 \cdot Et_2O$ (43 mL, 0,34 mol) was slowly added into the mixture at 0 °C under Ar atmosphere for 45 min. The mixture was stirred overnight (17 h) at rt under atmosphere and quenched with sat. aqueous $NaHCO_3$ till pH = 7.0 followed by extraction with $CHCl_3$ 3 times. The organic layer was treated with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo, and co-evaporated with toluene 3 times. The residue was purified by recrystallization using $CHCl_3$/n-hexane yielding compound **2** as a white powder (76.3 g, 60.7% in 2 steps).

^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.41 (d, *J* = 8.2 Hz, 2H, -S-$\underline{CH_2}$-CH_3), 7.13 (d, *J* = 7.9 Hz, 2H, -S-$\underline{CH_2}$-CH_3), 5.41 (d, *J* = 3.4 Hz, 1H, H-4), 5.22 (t, *J* = 9.9 Hz, 1H, H-2), 5.04 (dd, *J* = 10.0, 3.4 Hz, 1H, H-3), 4.65 (d, *J* = 10.0 Hz, 1H, H-1), 4.18 (dd, *J* = 11.4, 6.9 Hz, 1H, H-6), 4.10 (dd, *J* = 11.3, 6.3 Hz, 1H, H-6'), 3.95 – 3.88 (m, 1H, H-5), 2.35 (s, 3H, -C$\underline{H_3}$), 2.12 (s, 3H, OAc), 2.10 (s, 3H, OAc), 1.97 (s, 3H, OAc).

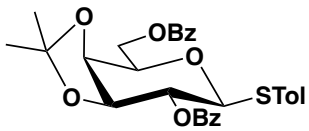


***p*-Tolyl 1-thio-*β*-D-Galactopyranoside (3).** In an evaporating flask at rt under Ar atmosphere, compound **2** (30 g, 66.3 mmol) was dissolved with the mixture of MeOH (300 mL, 8.08 mol) and THF (300mL, 3.98 mol) and cooled to 0 °C. 28% NaOMe-MeOH (12.7 mL, 66.3m mol) was then added to the mixture. After being stirred for 0.5 h at rt under Ar atmosphere, the reaction mixture was slowly quenched with acetyl

chloride till PH = 7 and concentrated in vacuo followed by co-evaporated with toluene for 2 times yielded compound **3** as a crude product.

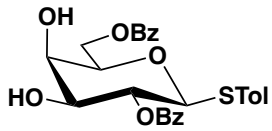


***p*-Tolyl 3,4-*O*-isopropylidine-1-thio-*β*-D-Galactopyranoside (4).** In an evaporating flask at it under Ar atmosphere, to a solution of compound **3** (66.3 mol) in anhydrous CH_2Cl_2 (136 mL, 73.8 mmol) was added 2,2-dimethoxy propane (90.5 mL, 73.8 mmol) and *p*-TsOH.H2O (0.74 g, 3.9 mmol) at 0°C under Ar atmosphere. After being stirred at rt for 18 h, the reaction mixture was slowly quenched with Et_3N till pH = 7.0 and treated with Pyridium p-toluenesulfonate (PPTS, 980 mg, 3.9 mmol). After 3 min the reaction was complete and quenched with Et_3N. The mixture was then concentrated in vacuo, dissolved with EtOAc, treated with satd aq $NaHCO_3$, and washed with brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo followed by being co-evaporated with toluene 3 times. dissolved with EtOAc and treated with PPTS (1.04 g, 4.1 mmol) and MeOH (40ml), After 4.5h add PPTS (0.5 g, 2 mmol), and the reaction was complete and dissolved with EtOAc, treated with sat. aqueous $NaHCO_3$, and washed with brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo affording compound **4** as a crude product, yellow oil.

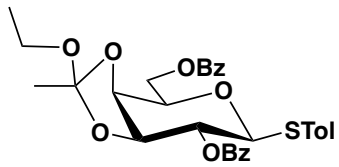


***p*-Tolyl 2,6-di-*O*-benzoyl-3,4-*O*-isopropylidine-1-thio-*β*-D-Galactopyranoside (5).** In an evaporating flask at it under Ar atmosphere, a solution of compound **4** (66.3 mol) in pyridine (442 mL, 5.48 mol) was added benzoyl chloride (30.8 mL, 0.27 mol) and DMAP (809.6 mg, 6.63 mmol) under Ar atmosphere. After being stirred for 10min, the mixture was then evaporated in vacuo. The residue was dissolved with EtOAc and treated with 1M aq HCI. After extracted for 3 times with EtOAc, the mixture was washed with sat. aqueous $NaHCO_3$, and brine. The filtrated was dried

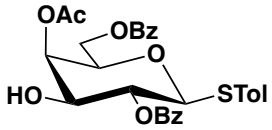
over Na_2SO_4, filtered, and concentrated in vacuo followed by co-evaporated with toluene for 3 times repetitions yielding compound **5** as a yellow oil. Rf 0.75 (n-hexane/EtOAc = 3/1).



***p*-Tolyl 2,6-di-*O*-benzoyl-1-thio-*β*-D-Galactopyranoside (6).** In an evaporating flask at rt under Ar atmosphere, a solution of compound **5** (66.3 mol) in MeOH (132.6 mL, 3.22 mol) and THF (179 mL, 2.23 mol) was added to 1 M aqueous HCI (100.6 mL, 2.85 mol) at rt under Ar atmosphere. After being stirred for 2 h at 60 °C, the reaction mixture was quenched with sat. aqueous $NaHCO_3$ at 0 °C till PH = 7, extracted with EtOAc, and washed with brine. The filtrate was dried over Na_2SO_4, filtered, and concentrated in vacuo yielding compound **6** as a crude product. Rf 0.20 (n-hexane/EtOAc= 3/2).



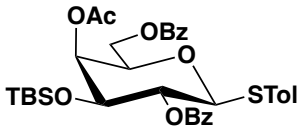
***p*-Tolyl 2,6-di-*O*-benzoyl-3,4-*O*-ethoxyethane-1-thio-*β*-D-Galactopyranoside (7).** In an evaporating flask at rt under Ar atmosphere, compound **6** (66.3 mol) in dry toluene (442ml, 4.2 mol) was added triethyl orthoacetate (49 mL, 0.27 mol) and *p*-TsOH.H2O (1.75 g, 10.0 mmol). After 20 minutes the reaction was complete and quenched with Et_3N till PH = 7. The residue was concentrated in vacuo to give compound **7** as a crude product. Rf 0.60 (n-hexane/E10Ac = 3/2).



***p*-Tolyl 4-*O*-acethyl-2,5-di-*O*-benzoyl-1-thio-*β*-D-Galactopyranoside (8).** In an evaporating flask at rt under Ar atmosphere, compound **7** (66.3 mol) was treated with 67% AcOH (380mL, 7.3 mol). After being stirred for 20 min at 40 °C the reaction was complete as monitored by TLC and concentrated in vacuo. The solid residue was then dissolved in EtOAc and treated with sat. aqueous $NaHCO_3$. After being extracted

with EtOAc for 3 times repetitions, the filtrate was washed with brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo. The crude residue was purified by recrystallization using $CHCl_3$/n-hexane to afford the desired compound **8** (20.2 g, 56.8% in 6 steps). Rf 0.42 (r-hexane/EtOAc = 3/2).

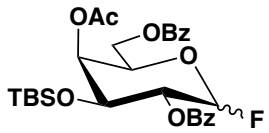
^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 8.11 – 8.01 (m, 4H, Bz), 7.64 – 7.56 (m, 2H, Bz), 7.47 (qd, *J* = 7.1, 1.7 Hz, 4H, Bz), 7.40 – 7.33 (m, 2H, S-$\underline{CH_2}$-CH_3), 6.98 – 6.92 (m, 2H, S-$\underline{CH_2}$-CH_3), 5.50 (d, *J* = 3.6, 1H, H-4), 5.24 (t, *J* = 9.7 Hz, 1H, H-2), 4.82 (d, *J* = 10.0 Hz, 1H, H-1), 4.50 (dd, *J* = 11.5, 7.5 Hz, 1H, H-6), 4.40 (dd, *J* = 11.5, 5.3 Hz, 1H, H-6'), 4.09 – 4.00 (m, 2H, H-3, H-5), 2.27 (s, 3H, S-CH_2-$\underline{CH_3}$), 2.20 (s, 3H, -CH_3).



***p*-Tolyl 4-*O*-acethyl-2,5-di-*O*-benzoyl-3-*O*-tert-Butyldimethylsilyl-1-thio-*β*-D-Galactopyranoside (9).** In an evaporating flask at rt under Ar atmosphere, compound **8** (1 g, 1.86 mmol) and TBSCl (0.56 g, 3.72 mmol) were dissolved by dry DMF (19ml, 0.24 mol), treated Imidazole (0.25 g, 3.72 mmol). After being stirred for 3 h at rt, added TBSCl (0.56 g, 3.72 mmol) and Imidazole (0.25 g, 3.72 mmol), after 2.5 h, added TBSCl (0.56 g, 3.72 mmol) and Imidazole (0.25 g, 3.72 mmol). After being stirred for 24 h the reaction was complete, added n-hexane/EtOAc = 1/1 100ml, washed by H_2O, and the organic layer was dried over Na_2SO_4 and concentrated in vacuo. The solid residue was then dissolved in EtOAc and treated with sat. aqueous $NaHCO_3$. After being extracted with EtOAc for 3 times repetitions, the filtrate was washed with brine. The organic layer was dried over Na2SO4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (n-hexane/EtOAc = 5/1) to give **9** (1.05 g, 87%) as a white powder.

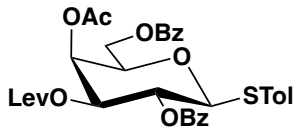
^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 8.21 (ddd, *J* = 8.5, 3.8, 1.4 Hz, 4H, Bz), 7.80 – 7.71 (m, 2H, Bz), 7.62 (td, *J* = 7.8, 6.1 Hz, 4H, Bz), 7.53 – 7.43 (m, 2H, S-$\underline{CH_2}$-CH_3), 7.07 – 7.01 (m, 2H, S-$\underline{CH_2}$-CH_3), 5.66 – 5.57 (m, 2H, H-2, H-4), 4.93 (d, *J* = 10.2 Hz, 1H, H-1), 4.67 – 4.55 (m, 2H, H-3, H-5), 4.23 (ddd, *J* = 8.0, 4.7, 1.1 Hz,

1H, H-6), 4.16 (d, *J* = 8.4 Hz, 1H, H-6'), 2.40 (s, 3H, S-CH_2-$\underline{CH}_3$), 2.34 (s, 3H, -CH_3), 0.87 (s, 9H, Si-C-$C\underline{H}_3$), 0.20 (s, 3H, Si-CH_3), 0.16 (s, 3H, Si-CH_3).



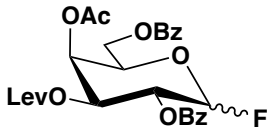
4-*O*-acetyl-2,5-di-*O*-benzoyl-3-*O*-tert-Butyldimethylsilyl-1-fluoride-D-Galactopyranoside (10).

In an evaporating flask at rt under Ar atmosphere, to a solution of **9** (1.0 g, 1.54 mmol) in CH_2Cl_2 (15.4 mL) was added XtalFluor-E ® (0.53 g, 2.37 mmol) at rt. After being stirred for 3.5 h at room temperature, the reaction mixture was added XtalFluor-E ® (0.18 g, 0.8 mmol). After being stirred for 2 h at room temperature, the reaction mixture was added XtalFluor-E ® (0.36 g, 1.6 mmol). After being stirred for 2 h at room temperature, the reaction mixture was added XtalFluor-E ® (0.36 g, 1.6 mmol). After being stirred for 2 h at room temperature, the reaction was quenched by sat. aqueous $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2. The organic layer was washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (toluene/EtOAc = 20/1) to give **10** (358 mg, 43%) as a white powder. ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 8.13 – 8.04 (m, 4H, Bz), 7.65 – 7.56 (m, 2H, Bz), 7.51 – 7.37 (m, 4H, Bz), 7.22 – 7.09 (m, 1H, H-1), 5.57 (dd, *J* = 3.6, 1.3 Hz, 1H, H-4), 5.44 (ddd, *J* = 24.9, 10.0, 2.7 Hz, 1H, H-2), 4.59 – 4.46 (m, 2H, H-3, H-5), 4.43 – 4.34 (m, 2H, H-6, H-6'), 3.51 (s, 9H), 2.39 – 2.31 (m, 1H), 2.20 (s, 3H, -CH_3), 0.96 (s, 3H, Si-CH_3), 0.75 (s, 9H, Si-C-$C\underline{H}_3$), 0.12 (s, 3H, Si-CH_3).

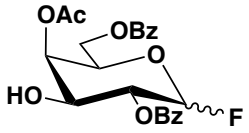


***p*-Tolyl 4-*O*-acethyl-2,5-di-*O*-benzoyl-3-*O*-levulinoyl-1-thio-*β*-D-Galactopyranoside (11).**[21] In an evaporating flask at rt under Ar atmosphere, to a solution of **8** (1 g, 1.86 mmol) in CH_2Cl_2 (31 mL) were added EDC (529 mg, 2.76 mmol), DMAP (22.7 mg, 186 μmol), levulinic acid (320 mg, 2.76 mmol) at rt. After

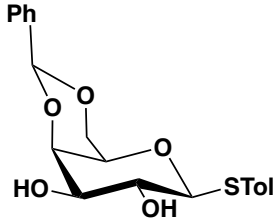
being stirred for 1 h, the reaction was evaporated. The residue was dissolved with EtOAc and washed with sat. aqueous $NaHCO_3$ and brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by recrystallization (benzene/n-hexane) to give **11** (887 mg, 75%) as a white powder. ^1H NMR (500 MHz, CHLOROFORM-*D*) δ 8.02 (ddt, *J* = 6.8, 5.3, 1.3 Hz, 4H, Bz), 7.59 (ddt, *J* = 7.8, 7.0, 1.3 Hz, 2H, Bz), 7.46 (tdd, *J* = 7.3, 4.7, 1.7 Hz, 4H, Bz), 7.39 – 7.31 (m, 2H, S-$\underline{CH_2}$-CH_3), 6.98 – 6.92 (m, 2H S-$\underline{CH_2}$-CH_3), 5.57 (dd, *J* = 3.4, 1.1 Hz, 1H, H-4), 5.52 (t, *J* = 10.0 Hz, 1H, H-2), 5.25 (dd, *J* = 10.0, 3.4 Hz, 1H, H-3,), 4.84 (d, *J* = 10.0 Hz, 1H, H-5), 4.53 (dd, *J* = 11.5, 7.4 Hz, 1H, H-6), 4.36 (dd, *J* = 11.5, 5.6 Hz, 1H, H-6'), 4.13 (ddd, *J* = 7.1, 5.6, 1.2 Hz, 1H), 2.74 (q, *J* = 5.4 Hz, 1H), 2.65 – 2.32 (m, 5H), 2.18 (s, 3H, OAc), 2.01 (s, 3H, OAc).



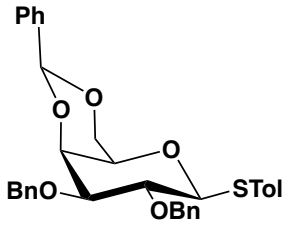
4-*O*-acethyl-2,5-di-*O*-benzoyl-3-*O*-levulinoyl-1-flouride-D-Galactopyranoside (12).[21] In an evaporating flask at rt under Ar atmosphere, to a solution of **11** (60.0 mg, 95.0 μmol) in CH_2Cl_2 (0.95 mL) was added XtalFluor-E® (32.4 mg, 0.126 mmol) at rt. After being stirred for 2.5 h, the reaction was quenched by sat. aqueous $NaHCO_3$ till PH = 7, the aqueous layer was extracted with $CHCl_3$. The organic layer was washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (n-hexane/EtOAc = 10/1) to give **12** (33.6 mg, 67%) as a white solid. ^1H NMR (500 MHz, $CDCL_3$) δ 8.10 – 8.00 (m, 4H, Bz), 7.64 – 7.54 (m, 2H, Bz), 7.53 – 7.38 (m, 4H, Bz), 6.07 – 5.88 (m, 1H), 5.69 (dd, *J* = 3.3, 1.4 Hz, 1H), 5.67 – 5.59 (m, 1H), 5.47 (ddd, *J* = 23.6, 10.8, 2.7 Hz, 1H), 4.62 (td, *J* = 6.5, 1.4 Hz, 1H), 4.50 (dd, *J* = 11.3, 6.7 Hz, 1H, H-6), 4.36 (dd, *J* = 11.4, 6.4 Hz, 1H, H-6'), 2.71 (dt, *J* = 18.3, 7.1 Hz, 1H), 2.66 – 2.48 (m, 2H), 2.41 (dt, *J* = 17.2, 6.5 Hz, 1H), 2.20 (s, 3H, OAc), 2.06 (s, 3H, OAc).



4-*O*-acethyl-2,5-di-*O*-benzoyl-1-flouride-D-Galactopyranoside (13).[21] In an evaporating flask at rt under Ar atmosphere, to a solution of **12** (60.0 mg, 0.113 mmol) in THF/MeOH = 9/1 (5 mL) was added hydrazine acetate (145 mg, 1.53 mmol) at rrt. After being stirred for 30 min, the reaction was evaporated. The residue was dissolved with EtOAc and washed with sat. aqueous $NaHCO_3$ and brine. The organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (n-hexane/EtOAc = 5/1) to give **13** (42.7 mg, 87%) as a white solid. 1H NMR (500 MHz, CDCl3) δ 8.11-8.04 (m, 4H, Bz), 7.63-7.56 (m, 2H, Bz), 7.49-7.44 (m 4H, Bz), 5.90 (dd, J = 49.3, 2.6 Hz, 1H), 5.64 (d, J = 3.3 Hz, 1H), 5.33 (ddd, J = 23.7, 10.3, 2.6 Hz, 1H), 4.58-4.37 (m, 4H), 2.38 (d, J = 5.6 Hz, 1H), 2.22 (s, 3H, OAc).

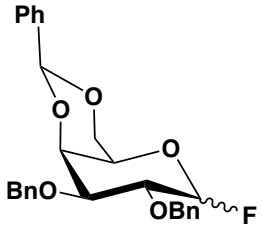


***p*-Tolyl4,6-benzylidene-1-thio-D-galactopyranoside (14).** In an evaporating flask at rt under Ar atmosphere, compound **2** (10g, 22 mmol) was dissolved with the mixture of MeOH (100 mL, 2.7 mol) and THF (100 mL, 1.3 mol) and cooled to 0 °C. 28% NaOMe-MeOH (4.24 mL, 22 mmol) was then added to the mixture. After being stirred for 0.5 h at rt under Ar atmosphere, the reaction mixture was slowly quenched with acetyl chloride till PH = 7 and concentrated in vacuo yielding compound **3** as a crude product, then compound **3** was dissolved in anhydrous MeCN (220mL, 4.2 mol) and benzaldehyde dimethyl acetal (9.9 mL, 66 mmol), p-TsOH.H_2O (0.5 g, 2.7 mmol) was next added into the reaction mixture. After 2 h, added p-TsOH.H_2O (0.5 g, 2.7 mmol), after 2 h, added p-TsOH.H_2O (0.5 g, 2.7 mmol) again, after 1 h, benzaldehyde dimethyl acetal (9.9 mL, 66 mmol), after 13 h, the reaction was complete and quenched with Et_3N till PH = 7. Added $CHCl_3$ 200ml, the organic layers were washed with satd aq $NaHCO_3$, brine, dried over Na_2SO_4, filtered, and concentrated in vacuo to afford compound **14** as a crude product.



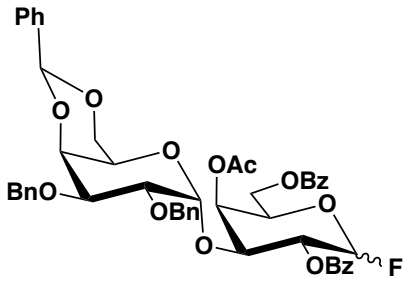
***p*-Tolyl 2,3-di-*O*-benzyl-4-6-*O*-benzylidine-1-thio-*β*-D-galactopyranoside (15).** In an evaporating flask at rt under Ar atmosphere, to a solution of **14** (22 mmol) in anhydrous DMF (220 ml, 2.8 mol) was added benzyl bromide (7.84 mL, 66 mmol) dropwise to the stirring solution at 0 °C. NaH (2.64 g, 66 mmol) was next added to the reaction mixture slowly. After being stirred for 2 h the reaction was complete as monitored by TLC, the reaction mixture was treated with MeOH to quench the excess BnBr, add H_2O 1.5L, extracted with $CHCl_3$. The organic layer was washed with satd aq $NaHCO_3$, brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by recrystallization ($CHCl_3$/n-hexane) to afford **15** 4.8 g (39% in 3 steps) as a yellow crystal.

^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.64 – 7.57 (m, 2H, Ar), 7.55 – 7.48 (m, 2H, Ar), 7.45 – 7.23 (m, 13H, Ar), 7.02 – 6.97 (m, 2H, Ar), 5.47 (s, 1H, -CHPh), 4.80 – 4.66 (m, 4H, CH_2Ph) x2, 4.57 (d, *J* = 9.5 Hz, 1H), 4.37 (dd, *J* = 12.3, 1.6 Hz, 1H), 4.14 (dd, *J* = 3.4, 1.0 Hz, 1H), 3.98 (dd, *J* = 12.3, 1.7 Hz, 1H), 3.84 (t, *J* = 9.4 Hz, 1H), 3.62 (dd, *J* = 9.2, 3.4 Hz, 1H), 3.40 (q, *J* = 1.5 Hz, 1H), 2.30 (s, 3H, -CH_3).



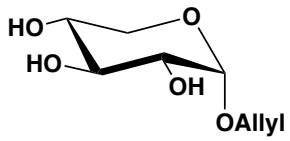
***p*-Tolyl 2,3-di-*O*-benzyl-4-6-*O*-benzylidine-1-fluoride-D-galactopyranoside (16)** To a solution of **15** (2.0 g, 3.6 mmol) in CH_2Cl_2 (36 mL) was added XtalFluor-E ® (1.65 g, 7.20 mmol) at room S3 temperature. After being stirred for 2.5 h at room temperature, to the reaction mixture was added XtalFluor-E ® (0.82 g, 3.60 mmol). After being stirred for 1 h at room temperature, to the reaction mixture was added XtalFluor-E ® (1.65 g, 7.20 mmol). After being stirred for 3.5 h at room temperature, to the reaction mixture was added XtalFluor-E ® (1.65 g, 7.20 mmol). After being stirred for 3 h at

room temperature, the reaction was quenched by sat. aqueous $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2. The organic layer was washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (toluene/EtOAc = 20/1) to give **16** (1.28 g, 79%) as a white powder. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.50 (d, J = 7.8 Hz, 2H, Ar), 7.41-7.24 (m, 13H, Ar), 5.69 (dd, J = 53.3, 1.7 Hz, 1H, -CHPh), 5.48 (s, 1H, -CHPh), 4.89 (d, J = 11.5 Hz, 1H, CH_2Ph), 4.81 (d, J = 12.0 Hz, 1H, CH_2Ph), 4.76 (d, J = 12.0 Hz, 1H, CH_2Ph), 4.72 (d, J = 11.5 Hz, 1H, CH_2Ph), 4.26-4.22 (m, 2H), 4.06 (ddd, J = 24.8, 10.0, 1.7 Hz, 1H), 4.00-3.97 (m, 2H), 3.81 (s, 1H).

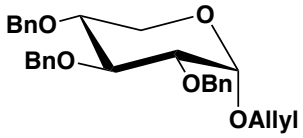


4-*O*-acetyl-2,5-di-*O*-benzoyl-3-*O*-(2,3-di-*O*-benzyl-4-6-*O*-benzylidine-1-fluoride-D-galactopyranosyl)-1-flouride-D-Galactopyranoside (**M1**). [21] In a pear-shaped flask at rt under Ar atmosphere, a suspension of donor **16** (156 mg, 0.348 mmol), acceptor **13** (100 mg. 0.232 mmol), activated MS 4Å in distilled CH_2Cl_2 (11.6 ml) was stirred at -40 °C under Ar atmosphere. $BF_3{\cdot}Et_2O$ (36.4 µl, 0.232 mmol) was added slowly to the stirring mixture. After being stirred for 1 h, then the mixture was quenched by sat. aqueous $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2. The organic layer was washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (toluene/EtOAc = 40/1) to give **M1** (140.6 mg, 70%) as a white powder. ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 8.08 – 8.02 (m, 2H), 8.00 – 7.94 (m, 2H), 7.63 (t, *J* = 7.4 Hz, 1H), 7.58 (t, *J* = 7.4 Hz, 1H), 7.50 – 7.43 (m, 6H), 7.38 – 7.21 (m, 15H), 5.89 (dd, *J* = 53.5, 2.9 Hz, 1H), 5.76 (d, *J* = 3.3 Hz, 1H), 5.58 – 5.48 (m, 1H), 5.28 (s, 1H), 5.21 (d, *J* = 3.4 Hz, 1H), 4.73 (d, *J* = 11.4 Hz, 1H), 4.68 – 4.60 (m, 2H), 4.54 (d, *J* = 12.2 Hz, 1H), 4.47 – 4.40 (m, 2H), 4.35 – 4.27 (m, 1H), 4.08 (d, *J* = 12.5 Hz, 1H), 4.01 (dd, *J* = 10.0, 3.4 Hz, 1H), 3.78 (dd, *J* = 10.0, 3.5 Hz, 1H), 3.74 (d, *J* = 3.5 Hz,

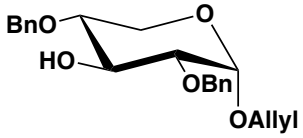
1H), 3.67 (d, *J* = 12.5 Hz, 1H), 3.57 (s, 1H), 1.87 (d, *J* = 1.2 Hz, 3H).



Ally *α*-D-xylopyranoside (1a). In an evaporating flask at rt under Ar atmosphere, D-xylose (5 g, 33.3 mmol) was added Allyl-OH (67.9mL, 1 mol). After being stirred for 5 mins at it under Ar atmosphere, TMSCI (21 mL, 166.5 mmol) was slowly added to the mixture. After being stirred overnight at 60 °C, the reaction mixture was concentrated in vacuo and followed by co-evaporated with toluene for 3 times repetitions to give compound **1a** as a crude product.



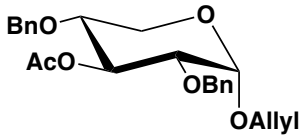
Ally 2, 3, 4-Tri-*O*-benzyl-*α*-D-xylopyranoside (2a)



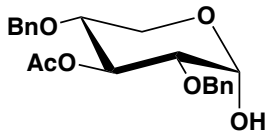
Allyl 2,4-Di-*O*-benzyl-*α*-D-xylopyranoside (2b). To a solution of compound **1a** (6.30g, 33.3 mmol) in DMSO (200 ml) was added NaH (2.0 g, 83.3 mmol) under ice cooling, and the mixture was stirred at room temperature for 30 min. Benzyl bromide (9.9 ml, 83.3 mmol) was added to the mixture under ice cooling. The mixture was stirred at room temperature overnight. 1.5L water was added to the mixture. extracted with $CHCl_3$. The organic layer was washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified by silica-gel column chromatography (Toluene-AcOEt = 92/8 to 90/10 to 88/12) to give an oily product **2a**: Yield 1.49g (10 %) ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.43 – 7.16 (m, 15H, Ar), 6.02 – 5.88 (m, 1H, -C<u>H</u>=CH_2), 5.32 (m, 1H, -CH=C<u>H</u>$_2$), 5.25 – 5.12 (1H, -CH=C<u>H</u>$_2$), 4.96 – 4.81 (m, 2H), 4.80 – 4.69 (m, 2H), 4.68 – 4.48 (m, 2H), 4.42 – 4.32 (m, 1H), 4.20 – 3.81 (m, 3H), 3.65 – 3.56 (m, 2H, H-5), 3.56 – 3.50 (m, 1H), 3.50 – 3.36 (m, 1H).

and **2b**: Yield 2.28 g (19 %) ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.40 – 7.27 (m, 10H, Ar), 5.89 (dddd, *J* = 17.0, 10.3, 6.4, 5.1 Hz, 1H, OCH_2C<u>H</u>=CH_2), 5.31 (dq, *J*

= 17.2, 1.6 Hz, 1H), 5.20 (dq, *J* = 10.4, 1.4 Hz, 1H), 4.79 – 4.58 (m, 6H), 4.13 (ddt, *J* = 12.9, 5.1, 1.5 Hz, 1H), 4.04 (ddd, *J* = 9.8, 8.5, 2.0 Hz, 1H), 3.91 (ddt, *J* = 13.0, 6.4, 1.3 Hz, 1H), 3.59 (dd, *J* = 10.6, 5.5 Hz, 1H), 3.57 – 3.42 (m, 2H), 3.34 (dd, *J* = 9.6, 3.5 Hz, 1H), 2.52 (d, *J* = 2.0 Hz, 1H).

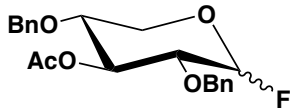


Allyl 3-*O*-acetyl-2,4-Di-*O*-benzyl-*α*-D-xylopyranoside (3b). In an evaporating flask at rt under Ar atmosphere, a solution of **2b** (1 g, 2.7 mmol) was dissolved in pyridine super dehydrate (1.3 mL, 16.2 mmol) and treated with acetic anhydride (0.16 mL, 3.24 mmol). After being stirred for 22 h, added acetic anhydride (0.13 mL, 2.7 mmol). After 1 d, the reaction mixture was concentrated in vacuo and co-evaporated with toluene 4 times. The residue was purified by silica-gel column chromatography (n-hexane-EtOAc = 90/10 to 85/15) to give an oily product **3b**: Yield 708 mg (64 %) ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.39 – 7.24 (m, 10H, Ar), 5.90 (dddd, *J* = 17.0, 10.2, 6.4, 5.1 Hz, 1H $OCH_2C\underline{H}=CH_2$), 5.46 (t, *J* = 9.6 Hz, 1H, H-3), 5.31 (dq, *J* = 17.2, 1.6 Hz, 1H, $OCH_2CH=C\underline{H}_2$), 5.20 (dq, *J* = 10.3, 1.3 Hz, 1H, $OCH_2CH=C\underline{H}_2$), 4.77 (d, *J* = 3.5 Hz, 1H, H-1), 4.58 (s, 2H, $C\underline{H}_2Ph$), 4.56 – 4.49 (m, 2H, $C\underline{H}_2Ph$), 4.15 (ddt, *J* = 13.0, 5.1, 1.5 Hz, 1H, $OC\underline{H}_2CH=CH_2$), 3.95 (ddt, *J* = 12.9, 6.4, 1.3 Hz, 1H, $OC\underline{H}_2CH=CH_2$), 3.62 (d, *J* = 8.2 Hz, 2H, H-5,), 3.54 – 3.45 (m, 1H, H-4), 3.40 (dd, *J* = 10.0, 3.5 Hz, 1H, H-2), 2.01 (s, 3H, Ac).

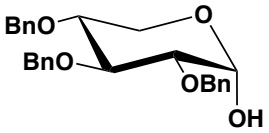


3-*O*-acetyl-2,4-Di-*O*-benzyl-*α*-D-xylopyranoside (4b). In a pear-shaped flask at rt under Ar atmosphere, to a solution of **3b** (600 mg, 1.46 mmol) in anhydrous THF (14.6 ml) was added with the activated solution of $[Ir(cod)(PPh_2Me)_2]PF_6$ (61.7 mg, 0.073 mmol) in anhydrous THF (14.6 ml) after stirred for 5 minutes under H_2 atmosphere (yellow solution) at rt. After being stirred for 30 min at rt under Ar atmosphere. Distilled H_2O (5 mL, 278 mmol) and I_2 (1.1 g, 4.38 mmol) were next

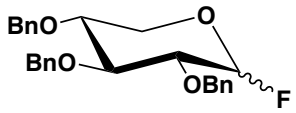
added to the stirring mixture and stirred for 1 h. The reaction was quenched by adding 20% $Na_2S_2O_3$ and extracted with EtOAc 3 times. The organic layer was then treated with sat. aqueous $NaHCO_3$., brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified with silica-gel column chromatography (n-hexane-EtOAc = 90/10 to 85/15 to 80/20) yielded compound **4b** $\alpha/\beta = 1.5/1$ (472 mg, 80%). ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.37 – 7.23 (m, 18H, Ar), 5.39 (t, *J* = 8.8 Hz, 1H, H-3), 5.19 – 5.12 (m, 2H), 4.84 (d, *J* = 11.9 Hz, 1H), 4.70 (dd, *J* = 7.4, 4.3 Hz, 1H), 4.66 – 4.59 (m, 3H), 4.59 – 4.48 (m, 3H), 3.95 (dd, *J* = 11.7, 5.3 Hz, 1H), 3.86 (dd, *J* = 11.4, 9.9 Hz, 1H), 3.69 (dd, *J* = 11.4, 5.2 Hz, 1H), 3.63 (d, *J* = 5.1 Hz, 1H), 3.57 – 3.47 (m, 1H), 3.47 – 3.40 (m, 2H), 3.34 – 3.20 (m, 2H), 2.01 (s, 3H, Ac), 1.94 (s, 2H). MS (ESI-QTOF) for $C_{21}H_{23}NaO_6^+$ [M + Na] $^+$ = 395.1138



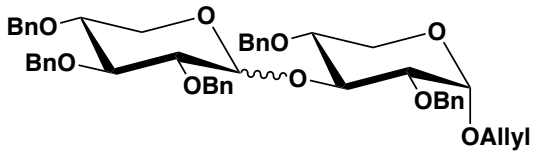
3-*O*-acetyl-2.4-di-*O*-benzyl-*α*-D-xylopyranosyl fluoride(5b) In a pear-shaped flask at rt under Ar atmosphere, to a solution of compound **4b** (100 mg, 0.268 mmol) in CH_2Cl_2 (2.7 ml) were added 2-fluoro-1-methylpyridinium p-toluenesulfonate (152 mg, 0.536 mmol) and triethylamine (0.149 ml, 1.07 mmol). The solution was stirred at room temperature for 3 h. EtOAc and water were added to the solution. Extracted with EtOAc 3 times. The organic layer was then treated with sat. aqueous $NaHCO_3$., brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified with silica-gel column chromatography (n-hexane-EtOAc = 90/10 to 85/15) yielded compound **5b** 61 mg, 61%, $\alpha/\beta = 4/1$. ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.38 – 7.21 (m, 10H, Ar), 5.49 – 5.40 (m, 1H), 4.67 – 4.50 (m, 4H), 3.83 – 3.70 (m, 2H), 3.66 – 3.50 (m, 1H), 3.50 – 3.36 (m, 1H), 2.01 (s, 3H, OAc). MS (ESI-QTOF) for $C_{21}H_{23}NaO_5^+$ [M + Na] $^+$ = 397.1477



2, 3, 4-tri-*O*-benzyl-*α*-D-xylopyranose(6a). In a pear-shaped flask at rt under Ar atmosphere, to a solution of **2a** (800 mg, 1.74 mmol) in anhydrous THF (17.4 ml) was added with the activated solution of $[Ir(cod)(PPh_2Me)_2]PF_6$ (73.5 mg, 0.087 mmol) in anhydrous THF (17.4 ml) after stirred for 5 minutes under H_2 atmosphere (yellow solution) at rt. After being stirred for 30 min at rt under Ar atmosphere. Distilled H_2O (5 mL, 278 mmol) and I_2 (1.3 g, 5.22 mmol) were next added to the stirring mixture and stirred for 1 h. The reaction was quenched by adding 20% $Na_2S_2O_3$ and extracted with EtOAc 3 times. The organic layer was then treated with sat. aqueous $NaHCO_3$., brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified with silica-gel column chromatography (n-hexane-EtOAc = 90/10 to 85/15 to 80/20) yielded compound **6a** (426 mg, 58%). MS (ESI-QTOF) for $C_{26}H_{28}NaO_5^+$ [M + Na] $^+$ = 443.1956



2, 3, 4-tri-*O*-benzyl-*α*-D-xylopyranosyl fluoride(7a). In a pear-shaped flask at rt under Ar atmosphere, to a solution of compound **6a** (100 mg, 0.238 mmol) in CH_2Cl_2 (2.4 ml) was added 2-fluoro-1-methylpyridinium p-toluenesulfonate (135 mg, 0.476 mmol) and triethylamine (0.133 ml, 0.95 mmol). The solution was stirred at room temperature for 3 h. EtOAc and water were added to the solution. Extracted with EtOAc 3 times. The organic layer was then treated with sat. aqueous $NaHCO_3$., brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. The residue was purified with silica-gel column chromatography (n-hexane-EtOAc = 95/5 to 90/10) yielded compound **7a** 62 mg, 61%. ^{1}H NMR (500 MHz, CHLOROFORM-*D*) δ 7.39 – 7.23 (m, 15H, Ar), 5.44 (dd, *J* = 53.1, 2.7 Hz, 1H, H-1), 4.90 – 4.60 (m, 6H, C$\underline{H}_2$Ph), 3.90 (t, *J* = 9.2 Hz, 1H, H-3), 3.76 (dd, *J* = 11.1, 5.7 Hz, 1H, H-4), 3.72 – 3.54 (m, 2H, H-5), 3.45 (ddd, *J* = 25.7, 9.6, 2.7 Hz, 1H, H-2).



Allyl *O*-(2,3,4-Tri-*O*-benzyl-*α*-D-xylopyranosyl)-(1→3)-2,4-di-O-benzyl-*α*-D-

xylopyranoside(*α*-Xa) and Allyl *O*-(2,3,4-Tri-*O*-benzyl-*α*-D-xylopyranosyl)-(1→3)-2,4-di-*O*-benzyl-*β*-D-xylopyranoside(*β*-Xa). In a pear-shaped flask at rt under Ar atmosphere, a suspension of donor **7a** (40 mg, 94.8 µmol), acceptor **2b** (23.4 mg. 63.2 µmol), activated MS 4Å in distilled CH_2Cl_2 (3.16 ml) was stirred at -32 °C under Ar atmosphere. $BF_3{\cdot}Et_2O$ (6 µl, 47.4 µmol) was added slowly to the stirring mixture. After being stirred for 1 h, then the mixture was quenched by Et_3N and filtered through a membrane filter (Fluoro pore®), and concentrated in vacuo. The residue was purified by silica-gel column chromatography (n-hexane/EtOAc = 10/1) to give **Xa** (45.63 mg, 94%) as a white powder. α/β = 1.4/1, Estimated by HPLC, flow rate: 1ml/min, C18-AR-300(4.6ID×250mm), MeCN/50mM NH_4COOCH_3= 85/15. Resident time ***α*-Xa** is 14 min, ***β*-Xa** is 15 min 30 s.

Flow synthesis of Xa

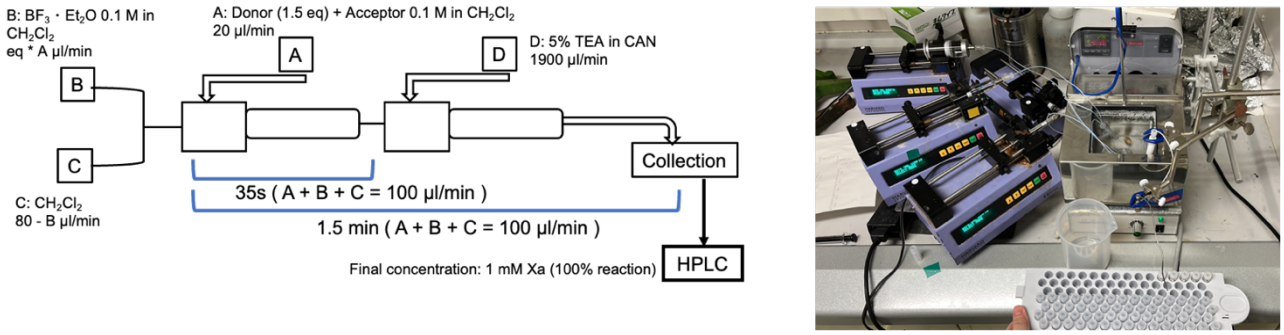


Figure 4.3.1.1. Microfluidic system for optimization.

In a pear-shaped flask at rt under Ar atmosphere, donor **7a** (158.2 mg, 0.375mmol), accepter **2b** (31.46 mg, 0.25 mmol), was dissolved in super dehydrate toluene (2.5 ml) and transferred into 5 mL syringe HAMILTON CO., RENO. NEVADA (pump A). On the other side, pump B was filled out with TMSOTf (54.2µl, 0.3 mmol) in super dehydrate toluene (3 mL) in a 5 mL syringe. Pump C filled out with super dehydrate toluene (8 mL) into a 10 mL syringe, and Pump D filled out 5% TEA in MeCN (50 ml) into a 50 mL syringe. Pump A and B were set up in a syringe pump (HARVARD 11 ELITE single syringe pump) and ran with a flow rate of 20 µL/min at -40 °C. Pump C ran with a flow rate of 60 µL/min, Pump D ran with a flow rate of 1900 µL/min. After 5 mins, the reaction mixture was collected 400 µl as a sample, then change the flow

rate for the next condition. HPLC automatic sample injection analysis of the yield and the ratio of α-Xa and β-Xa.

Machine learning programming code[12-16]

Gaussian process regression

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Mon Feb 20 18:56:59 2023

@author: daichanghao
"""
import csv
import warnings
import GPy
import numpy as np
from matplotlib import pyplot as plt
warnings.filterwarnings('ignore')
# catalystEq, temp
with open('catatemp.csv', encoding='utf-8-sig') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    X = []
    for row in csv_reader:
        X.append([float(row[0]), float(row[1])])
X = np.array(X)
with open('catatemp.csv', encoding='utf-8-sig') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    Y = []
    for row in csv_reader:
        Y.append(float(row[2]))
Y = np.array(Y)[:, np.newaxis]
# standardization
X_std = np.copy(X)
for i in range(X.shape[1]):
    X_std[:, i:i + 1] = (X[:, i:i + 1] - np.mean(X[:, i:i + 1])) / np.sqrt(np.var(X[:, i:i + 1]))

# GPR and visualization
kernel = GPy.kern.RBF(input_dim=2)
model = GPy.models.GPRegression(X_std, Y, kernel=kernel, normalizer=True, noise_var=0.001)
model.optimize(max_iters=3, messages=True)
mean, variance = model.predict(X_std)
top_point = X_std[np.argmax(mean)]
model.plot()
print(model)
temp_prediected = top_point[0]
eq_prediected = top_point[1]
print(temp_prediected)
print(eq_prediected)
print(top_point)

x0_list_law = [1, 2, 3, 4, 5, 6, 7, 8]
x1_list_law = [-40, -30, -20, -10, 0, 10, 20]
x0_list_std = []
x1_list_std = []
x0_list_std = (np.array(x0_list_law) - np.mean(X[:, 0:1])) / np.sqrt(np.var(X[:, 0:1]))
x1_list_std = (np.array(x1_list_law) - np.mean(X[:, 1:2])) / np.sqrt(np.var(X[:, 1:2]))
plt.xticks(x0_list_std, x0_list_law)
plt.yticks(x1_list_std, x1_list_law)
plt.xlabel("CatalystEq [Cat./Donor]", fontsize=15)
plt.ylabel("temperature[℃]", fontsize=15)
plt.tick_params(labelsize=15)
plt.legend(bbox_to_anchor=(1, 0.5), fontsize=15)
plt.show()
from mpl_toolkits.mplot3d import Axes3D
```

```
# Plot original data
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(X[:,0], X[:,1], Y[:,0], c='red', s=20)

# Plot predicted mean
X_mesh, Y_mesh = np.meshgrid(np.linspace(X[:,0].min(), X[:,0].max(), 100),
                             np.linspace(X[:,1].min(), X[:,1].max(), 100))
X_mesh_std = np.column_stack(((X_mesh.flatten() - np.mean(X[:,0])) / np.sqrt(np.var(X[:,0])),
                              (Y_mesh.flatten() - np.mean(X[:,1])) / np.sqrt(np.var(X[:,1]))))
mean_mesh, variance_mesh = model.predict(X_mesh_std)
ax.plot_surface(X_mesh, Y_mesh, mean_mesh.reshape(X_mesh.shape), alpha=1, cmap=plt.cm.rainbow)

# Set axis labels and display plot
ax.set_xlabel('CatalystEq [Cat./Donor]')
ax.set_ylabel('temperature [℃]')
ax.set_zlabel('yield [%]')
plt.show()

model.plot(fixed_inputs=[(1, eq_prediected)], plot_data=False, lower=25, upper=75)
x0_list_law = [1, 2, 3, 4, 5, 6]
x0_list_std = []
x0_list_std = (np.array(x0_list_law) - np.mean(X[:, 0:1])) / np.sqrt(np.var(X[:, 0:1]))
plt.xticks(x0_list_std, x0_list_law)
x0_line = 1.05
x0_line_std = (x0_line - np.mean(X[:, 0:1])) / np.sqrt(np.var(X[:, 0:1]))
plt.plot([x0_line_std, x0_line_std], [5, 60], color='red', linewidth=1, linestyle='-', label=x0_line)
plt.xlabel("CatalystEq [Cat./Donor]", fontsize=15)
plt.ylabel("yield[%]", fontsize=15)
plt.tick_params(labelsize=15)
plt.legend(("Mean", x0_line, "Confidence"), loc="lower right", borderaxespad=0.2, fontsize=15)
plt.show()

model.plot(fixed_inputs=[(0, temp_prediected)], plot_data=False, lower=25, upper=75)
x1_list_law = [-40, -30, -20, -10, 0, 10, 20]
x1_list_std = []
x1_list_std = (np.array(x1_list_law) - np.mean(X[:, 1:2])) / np.sqrt(np.var(X[:, 1:2]))
plt.xticks(x1_list_std, x1_list_law)
x1_line = -2
x1_line_std = (x1_line - np.mean(X[:, 1:2])) / np.sqrt(np.var(X[:, 1:2]))
plt.plot([x1_line_std, x1_line_std], [5, 70], color='red', linewidth=1, linestyle='-', label=x1_line)
plt.xlabel("temperature [℃]", fontsize=15)
plt.ylabel("yield [%]", fontsize=15)
plt.tick_params(labelsize=15)
plt.legend(("Mean", f"{x1_line}℃", "Confidence"), loc="lower right", borderaxespad=0.2, fontsize=15)
plt.show()
```

Bayesian process regression

```
import csv
import numpy as np
import GPyOpt
import warnings
warnings.filterwarnings('ignore')

with open('Flow-Tol-T-B-Con-a-b.csv', encoding='utf-8-sig') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    X = []
    for row in csv_reader:
        X.append([float(row[0]), float(row[1]), float(row[2])])
X = np.array(X)
with open('Flow-Tol-T-B-Con-a-b.csv', encoding='utf-8-sig') as csv_file:
    csv_reader = csv.reader(csv_file, delimiter=',')
    Y = []
    for row in csv_reader:
        Y.append(float(row[3]))
Y = -np.array(Y)[:, np.newaxis]

initial_x = X
initial_y = Y

bounds = [{'name': 'temp', 'type': 'continuous', 'domain': (-78, 70)},
          {'name': 'cat.eq', 'type': 'continuous', 'domain': (0.1, 10)},
          {'name': 'time/min', 'type': 'continuous', 'domain': (0, 100)}]

myBopt = GPyOpt.methods.BayesianOptimization(f=None,
                                             domain=bounds,
                                             X=initial_x,
                                             Y=initial_y,
                                             acquisition_type='EI', )

next_x = myBopt.suggest_next_locations()
print(next_x)
```

Reference:

1, Thomas Hansen, Hidde Elferink, Jacob M. A. van Hengst, Kas J. Houthuijs, Wouter A. Remmerswaal, Alexandra Kromm, Giel Berden, Stefan van der Vorm, Anouk M. Rijs, Hermen S. Overkleeft, Dmitri V. Filippov, Floris P. J. T. Rutjes, Gijsbert A. van der Marel, Jonathan Martens, Jos Oomens, Jeroen D. C. Codée & Thomas J. Boltje. Characterization of glycosyl dioxolenium ions and their role in glycosylation reactions. *Nature Communications* **11**, 2664 (2020).

2, Mana Mohan Mukherjee, Rina Ghosh,corresponding and John A. Hanover. "Recent Advances in Stereoselective Chemical O-Glycosylation Reactions." *Frontiers in Molecular Biosciences*, vol. 9, 2022,1.

3, Moon, S., Chatterjee, S., Seeberger, P. H. & Gilmore, K. Predicting glycosylation stereoselectivity using machine learning. *Chem. Sci.* **12**, 2931–2939 (2021).

4, Toshima, K. Glycosyl Halides. in *Glycoscience: Chemistry and Chemical Biology* (eds. Fraser-Reid, B. O., Tatsuta, K. & Thiem, J.) 429–449 (Springer Berlin Heidelberg, 2008).

5, Edward O. Pyzer-Knapp, Jed W. Pitera, Peter W. J. Staar, Seiji Takeda, Teodoro Laino, Daniel P. Sanders, James Sexton, John R. Smith & Alessandro Curioni. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials* **8**, 84 (2022).

6, Long, Q.; Gao, J.; Yan, N.; Wang, P.; Li, M. $(C_6F_5)_3B\cdot(HF)$ncatalyzed glycosylation of disarmed glycosyl fluorides and reverse glycosyl fluorides. *Org. Chem. Front.* 2021, 8, 3332−3341.

7, Toshima, K. Glycosyl fluorides in glycosidations. *Carbohydrate Research* **327**, 15–26 (2000).

8, Hideki Jona, Hiroki Mandai, Warinthorn Chavasiri, Kazuya Takeuchi, and Teruaki Mukaiyama. Protic Acid Catalyzed Stereoselective Glycosylation Using Glycosyl Fluorides. *Bull. Chem. Soc. Jpn.,* 75, 291–309 (2002)

9, Girish C. Sati, Joshua L. Martin, Yishu Xu, Tanmay Malakar, Paul M. Zimmerman, and John Montgomery. Fluoride Migration Catalysis Enables Simple, Stereoselective, and Iterative Glycosylation. *J. Am. Chem. Soc.* **142**, 7235–7242 (2020).

10, Ahneman, Derek T., Estrada, Jesús G., Lin, Shishi, Dreher, Spencer D., and Abigail G. Doyle. "Predicting reaction performance in C–N cross-coupling using machine learning." *Science*, (2018). Accessed April 15, 2023.

11, Coley, C. W., Rogers, L., Green, W. H. & Jensen, K. F. SCScore: Synthetic Complexity Learned from a Reaction Corpus. *J. Chem. Inf. Model.* **58**, 252–261 (2018).

12, Masaru Kondo, Akimasa Sugizaki, Md. Imrul Khalid, H. D. P. Wathsala, Kazunori Ishikawa, Satoshi Hara, Takayuki Takaai, Takashi Washio, Shinobu Takizawa and Hiroaki Sasai. Energy-, time-, and labor-saving synthesis of α-ketiminophosphonates: machine-learning-assisted simultaneous multiparameter screening for electrochemical oxidation. *Green Chem.* **23**, 5825–5831 (2021).

13, Masaru Kondo, H. D. P. Wathsala, Makoto Sako, Yutaro Hanatani, Kazunori Ishikawa, Satoshi Hara, Takayuki Takaai, Takashi Washio, Shinobu Takizawa and Hiroaki Sasai. ***Chem. Commun.***, 2020,**56**, 1259-1262

14, Masaru Kondo, H. D. P. Wathsala, Makoto Sako, Yutaro Hanatani, Kazunori Ishikawa, Satoshi Hara, Takayuki Takaai, Takashi Washio, Shinobu Takizawa and Hiroaki Sasai. Correction: Exploration of flow reaction conditions using machine-learning for enantioselective organocatalyzed Rauhut–Currier and [3+2] annulation sequence. ***Chem. Commun.*** **56**, 12256–12256 (2020).

15, To carry out the multiparameter screening, GPyOpt was used as a framework in Python, see: https://github.com/SheffieldML/GPyOpt

16, The GPy authors. GPy: A Gaussian process framework in Python. http://github.com/SheffieldML/Gpy

17, Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).

18, Martha M. Flores-Leonar, Luis M. Mejía-Mendoza, Andrés Aguilar-Granda, Benjamin Sanchez-Lengeling, Hermann Tribukait, Carlos Amador-Bedolla, Alán Aspuru-Guzik. Materials Acceleration Platforms: On the way to autonomous experimentation. *Current Opinion in Green and Sustainable Chemistry* **25**, 100370 (2020).

19, Rasmussen, C. E. Gaussian Processes in Machine Learning. in *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures* 63–71

20, Moon, S., Chatterjee, S., Seeberger, P. H. & Gilmore, K. Predicting glycosylation stereoselectivity using machine learning. *Chem. Sci.* **12**, 2931–2939 (2021).

21, Yoshiyuki Manabe, Takuya Matsumoto, Yuka Ikinaga, Yuya Tsutsui, Shota Sasaya, Yuichiro Kadonaga, Akihito Konishi, Makoto Yasuda, Tomoya Uto, Changhao Dai, Kumpei Yano, Atsushi Shimoyama, Ayana Matsuda, Koichi Fukase. Revisiting Glycosylations Using Glycosyl Fluoride by $BF_3 \cdot Et_2O$: Activation of Disarmed Glycosyl Fluorides with High Catalytic Turnover. *Org. Lett.* **24**, 6–10 (2022).

22, Cressie, N. The origins of kriging. *Mathematical Geology* **22**, 239–252 (1990).

23, Stein, M. L. (1999). Interpolation of Spatial Data: Some Theory for Kriging. Springer.

24, SEEGER, MATTHIAS. GAUSSIAN PROCESSES FOR MACHINE LEARNING. *Int. J. Neur. Syst.* **14**, 69–106 (2004).

25, Jerome Sacks, William J. Welch, Toby J. Mitchell, & Henry P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science* **4**, 409–423 (1989).

26, Gramacy, R. B. & Lee, H. K. H. Cases for the nugget in modeling computer experiments. *Statistics and Computing* **22**, 713–722 (2012).

27, Li, Q., Levi, S. M., Wagen, C. C., Wendlandt, A. E. & Jacobsen, E. N. Site-selective, stereocontrolled glycosylation of minimally protected sugars. *Nature* **608**, 74–79 (2022).

28, Solá, R. J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *Journal of Pharmaceutical Sciences* **98**, 1223–1245 (2009).

29, Macher, B. A. & Galili, U. The Galα1,3Galβ1,4GlcNAc-R (α-Gal) epitope: A carbohydrate of unique evolution and clinical relevance. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1780**, 75–88 (2008).

30, Julinton Sianturi, Dr. Yoshiyuki Manabe, Dr. Hao-Sheng Li, Li-Ting Chiu, Dr. Tsung-Che Chang, Kento Tokunaga, Dr. Kazuya Kabayama, Dr. Masahiro Tanemura, Dr. Shinji Takamatsu, Prof. Eiji Miyoshi, Prof. Shang-Cheng Hung, Prof. Koichi Fukase. Development of α-Gal–Antibody Conjugates to Increase Immune Response by Recruiting Natural Antibodies. *Angewandte Chemie International Edition* **58**, 4526–4530 (2019).

31, Galili, U., Shohet, S. B., Kobrin, E., Stults, C. L. & Macher, B. A. Man, apes, and Old World monkeys differ from other mammals in the expression of alpha-galactosyl

epitopes on nucleated cells. *Journal of Biological Chemistry* **263**, 17755–17762 (1988).

32, Galili, U., Rachmilewitz, E. A., Peleg, A. & Flechner, I. A unique natural human IgG antibody with anti-alpha-galactosyl specificity. *Journal of Experimental Medicine* **160**, 1519–1531 (1984).

33, Galili, U., Macher, B. A., Buehler, J. & Shohet, S. B. Human natural anti-alpha-galactosyl IgG. II. The specific recognition of alpha (1----3)-linked galactose residues. *Journal of Experimental Medicine* **162**, 573–582 (1985).

34, Galili, U., Clark, M. R., Shohet, S. B., Buehler, J. & Macher, B. A. Evolutionary relationship between the natural anti-Gal antibody and the Gal alpha 1----3Gal epitope in primates. *Proceedings of the National Academy of Sciences* **84**, 1369–1373 (1987).

35, John E. Yule, Ting C. Wong, Sham S. Gandhi, Dongxu Qiu, Marc A. Riopel, R. Rao Koganty. Steric control of N-acetylgalactosamine in glycosidic bond formation. *Tetrahedron Letters* **36**, 6839–6842 (1995).

36, Scheller, H. V. & Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289 (2010).

37, Ebringerová, A. & Heinze, T. Xylan and xylan derivatives – biopolymers with valuable properties. Naturally occurring xylans structures, isolation procedures and properties. *Macromolecular Rapid Communications* **21**, 542–556 (2000).

38, Fukase, K., Hase, S., Ikenaka, T. & Kusumoto, S. Synthesis of New Serine-Linked Oligosaccharides in Blood-Clotting Factors VII and IX and Protein Z.The Syntheses of O-α-D-Xylopyranosyl-(1→3)-D-glucopyranose, O-α-D-Xylopyranosyl-(1→3)-O-α-D-xylopyranosyl-(1→3)-D-glucopyranose, and Their Conjugates with Serine. *BCSJ* **65**, 436–445 (1992).

39, Jähnisch, K., Hessel, V., Löwe, H. & Baerns, M. Chemistry in Microstructured Reactors. *Angewandte Chemie International Edition* **43**, 406–446 (2004).

40, Hartman, R. L., McMullen, J. P. & Jensen, K. F. Deciding Whether To Go with the Flow: Evaluating the Merits of Flow Reactors for Synthesis. *Angewandte Chemie International Edition* **50**, 7502–7519 (2011).

41, Wiles, C. & Watts, P. Continuous flow reactors: a perspective. *Green Chem.* **14**, 38–54 (2012).

42, Ley, S. V., Fitzpatrick, D. E., Ingham, Richard. J. & Myers, R. M. Organic Synthesis: March of the Machines. *Angewandte Chemie International Edition* **54**, 3449–3464 (2015).
43, Newman, S. G. & Jensen, K. F. The role of flow in green chemistry and engineering. *Green Chem.* **15**, 1456–1472 (2013).

44, Reizman, B. J. & Jensen, K. F. Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **49**, 1786–1796 (2016).

Acknowledgment

I would like to express my deepest gratitude to my supervisor, Professor Koichi Fukase, for his invaluable guidance, continuous support, and encouragement throughout the course of this study at Osaka University. His extensive knowledge, expertise, and passion for research have inspired me to pursue excellence in my work. I am truly grateful for his mentorship and the opportunities he has provided me.

I would also like to thank my co-supervisor, Dr. Yoshiyuki Manabe, for his insightful feedback, constructive suggestions, and unwavering support during my research. His expertise and dedication to my project have significantly contributed to the quality of my work and my growth as a researcher.

I am immensely grateful to the Graduate School of Science Research Assistantship and the Next Generation Challenging Research Fellowship for providing me with the necessary financial support for living expenses and research funding. Their support has enabled me to focus on my studies and pursue my research with dedication.

I would like to convey my heartfelt appreciation to the thesis review committee members, Professor Takayoshi Suzuki and Professor Takashi Kubo, for their invaluable guidance and recommendations. Additionally, I would like to express my profound gratitude to Associate Professor Takizawa Shinobu of the Institute of Scientific and Industrial Research at Osaka University. Their combined expertise and perceptive feedback have significantly contributed to the refinement and enhancement of my work.

I am grateful for the camaraderie and assistance of all the members of our laboratory. Their collaboration, enthusiasm, and friendship have made my research journey an enjoyable and rewarding experience. I would like to express my appreciation for their willingness to share their knowledge, expertise, and time to help me with my project.

Finally, I would like to thank my family and friends for their constant love, encouragement, and support throughout this journey. Their belief in me has been a source of strength and motivation, and I am truly grateful for their presence in my life.

长风破浪会有时 直挂云帆济沧海

Dai Changhao