# Supplementary Materials for

# Decoding the Multi-Dimensional Complexity of Glycosylation Reaction via Machine Learning

**Authors:** Fizza Sabbor[1,2,3], Po-Wei Lu[3,4], Shang-Yi Chen[2,3,5], Sung-Chi Lin[3], Sefunmi Peace Akinseye[3,5,6], Rapelly Venkatesh[3], Tiow-Gan Ong[1,3]*, Cheng-Chung Wang[3,5]*

**Affiliations:**

[1]Molecular Science and Technology (MST), Taiwan International Graduate Program (TIGP), Academia Sinica; Taipei 115, Taiwan.

[2]Department of Chemistry, National Tsing Hua University; Hsinchu, 300, Taiwan.

[3]Institute of Chemistry, Academia Sinica; Taipei, 115, Taiwan.

[4]Institute of Biochemistry and Molecular Biology, National Yang Ming Chiao Tung University; Taipei, Taiwan.

[5]Chemical Biology and Molecular Biophysics (CBMB), Taiwan International Graduate Program (TIGP), Academia Sinica; Taipei, 115, Taiwan.

[6]Institute of Biological Sciences, National Taiwan University; Taipei, 10617, Taiwan.

*Email(s): wangcc@chem.sinica.edu.tw; tgong@gate.sinica.edu.tw

# Table of Contents

# 1 Materials and Methods

## 1.1 General information

All reactions were conducted in flame-dried glassware, under nitrogen atmosphere, unless otherwise noted. All solvents were purified and dried from a safe purification system containing activated $Al_2O_3$. All reagents obtained from commercial sources were used without purification. Air/moisture-sensitive liquids were transferred with the gas-tight syringes fitted with stainless-steel needles. Liquid reagents were added via Hamilton gas-tight microliter syringes. Thin -layer chromatography (TLC) was performed on pre-coated glass plates and aluminum sheets of Silica Gel 60 $F_{254}$ (0.25 mm, E. Merck); detection was executed by spraying with a solution of $Ce(SO_4)_2$, $(NH_4)_6Mo_7O_2$, and $H_2SO_4$ in water and subsequently heating on a hot plate. UV light for TLC analysis was UVGL-25 compact UV lamp (4 watt/ 254 nm), UVP. Flash column chromatography was carried out by Silica Gel Geduran® Si 60 (0.040-0.063 mm, E. Merck). Stir plate: MR Hei-Standard.

## 1.2 Material and reagents

All commercially available reagents were purchased from Sigma-Aldrich, Alfa Aesar, Strem, Oakwood, Acros Organics, Matrix Scientific, TCI, Chem-Impex, Carbosynth, or Ambeed and were used without purification unless otherwise indicated. Solvents for chromatography and extraction were purchased from Sigma Aldrich as ACS grade or above and used without additional purification. Solvents for HPLC were purchased from Sigma Aldrich or Fisher as HPLC grade and used without additional purification. Anhydrous reaction solvents were passed through aluminum oxide on a solvent purification system (LC Technology Solutions) and sparged with Argon for 1 hour before use. Deuterated solvents were purchased from Cambridge Isotope Laboratories and used without prior purification.

## 1.3 Instrumentation

Nuclear magnetic resonance (NMR) spectra such as $^1H$, $^{13}C$ NMR, DEPT, $^1H$-$^1H$ COSY, $^1H$-$^{13}C$ COSY, and NOESY spectra were recorded by Bruker AV400, and AV500 MHz instruments. Chemical shifts are in ppm from $Me_4Si$, generated from the $CDCl_3$ lock signal at δ7.26. Multiplicities are reports by using the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad, ABq = AB quartets. bd = broad doublet; J = coupling constant values in Hertz. IR spectra were taken by a Bruker alpha FT-IR spectrometer using ZnSe plates. Mass spectra were analyzed by a Waters Premier XE instrument with ESI source. Optical rotations were measured with a JASCO P-2000 polarimeter at 25-30 °C, and using a 100 mm cell at 589 nm (Na).

## 1.4 Standard procedure for glycosylation

A flame-dried round-bottom flask was charged with a magnetic bar, glycosylating agent (1 equiv.), anhydrous DCM (0.167 M), and activated 3 Å molecular sieves. The nucleophile (1.5 equiv.) was added to a reaction solution at -40 °C, and the resulting mixture was stirred for 1 h. Subsequently, NIS (1.0 equiv.) and TfOH (0.4 equiv.) were added sequentially at -40 °C and the reaction mixture was stirred overnight at -40 °C. Upon completion, the reaction was quenched by the addition of TEA at same temperature. The mixture was then filtered through a Celite, washing with DCM. The filtrates were concentrated under reduced pressure, and the crude product was

purified by silica gel column chromatography (8:1 hexanes-/EtOAc) to afford the desired glycosylated product.

## 1.5 General procedure for temperature screening

A suspension of glycosylating agent **A-E** (100 mg), molecular sieves (4 Å, 100 mg), and alcohol **51-56** (1.5 equiv.) in DCM (0.167M) was stirred at varied temperature (-40 ℃, -20 ℃, 0 ℃, 25 ℃) under a nitrogen atmosphere for 1 hour. NIS (1.0 equiv.) and TfOH (0.4 equiv.) was then added to the reaction mixture, which was stirred for 12 hours at the same temperature, with progress monitored by TLC (*n*-Hexane/EtOAc, 8:1). The reaction mixture was filtered through Celite, washed with DCM, and the filtrate was concentrated under reduced pressure. The resulting crude product was purified by flash chromatography to afford the desired product.

## 1.6 General procedure for solvent screening

A suspension of glycosylating agent **A-E** (100 mg), molecular sieves (4 Å, 100 mg), and alcohol **51-56** (1.5 equiv.) in varied solvent systems DCM, or THF, or Toluene, DCM/ACN (*v/v* = 1/2), DCM/DMF (*v/v* = 1/2), DCM/*p*-dioxane (*v/v* = 2/1), of which solvent concentration is 0.167M, was stirred at -40 ℃ under a nitrogen atmosphere for 1 hour. NIS (1.0 equiv.) and TfOH (0.4 equiv.) was then added to the reaction mixture, which was stirred for 12 hours at the same temperature, with progress monitored by TLC (*n*-Hexane/EtOAc, 8:1). The reaction mixture was filtered through Celite, washed with DCM, and the filtrate was concentrated under reduced pressure. The resulting crude product was purified by flash chromatography to afford the desired product.

## 1.7 General procedure for concentration screening

A suspension of glycosylating agent **A-E** (100 mg), molecular sieves (4 Å, 100 mg), and alcohol **51-56** (1.5 equiv.) in DCM with varied concentration (0.308M, 0.167M, 0.051M, 0.038M, 0.018M) was stirred at -40 ℃ under a nitrogen atmosphere for 1 hour. NIS (1.0 equiv.) and TfOH (0.4 equiv.) was then added to the reaction mixture, which was stirred for 12 hours at the same temperature, with progress monitored by TLC (n-Hexane/EtOAc, 8:1). The reaction mixture was filtered through Celite, washed with DCM, and the filtrate was concentrated under reduced pressure. The resulting crude product was purified by flash chromatography to afford the desired product.

## 1.8 General procedure for stoichiometry screening

A suspension of glycosylating agent **A-E** (100 mg), molecular sieves (4 Å, 100 mg), and alcohol **51-56** at varied equivalents (1.5 equiv., and 0.66 equiv.) in DCM (0.167 M) was stirred at -40 ℃ under a nitrogen atmosphere for 1 hour. NIS (1.0 equiv.) and TfOH (0.4 equiv.) was then added to the reaction mixture, which was stirred for 12 hours at the same temperature, with progress monitored by TLC (n-Hexane/EtOAc, 8:1). The reaction mixture was filtered through Celite, washed with DCM, and the filtrate was concentrated under reduced pressure. The resulting crude product was purified by flash chromatography to afford the desired product.

## 2 Results and Programming Details

### 2.1 Environmental Factor Impact index (EFI)

EFI quantifies the overall efficiency of glycosylation under given environmental conditions, capturing both mechanistic favorability (α-ratio) and practical productivity (yield) while penalizing unfavorable reactivity, nucleophilicity, and temperature conditions.

- Higher EFI = efficient, selective, practical glycosylation
- Lower EFI = poor selectivity, low yield, or harsh conditions

$$EFI_{General} = \frac{Outcome\ Metrics\ (\text{α-ratio \& yield})}{\Pi_{i=1}^{n} f_i(parameter_i)} \tag{S1}$$

Where;

***Outcome Metrics***= α-ratio, yield, or another measurable reaction outcome that reflects stereochemical efficiency.

***parameter$_i$***= represents each parameter, such as solvent $P_{solvent}$ (table S3), reactivity of glycosylating agent (RRV), alcohol nucleophilicity (Aka), Temperature ($T$), or concentration ($C$).

***$f_i$(parameter$_i$)***= Function to quantify the influence of $parameter_i$ on the reaction. These could be linear or non-linear functions derived from experimental conditions or empirical data.

$\Pi_{i=1}^{n} f_i(parameter_i) = f_1(parameter_1) \times f_2(parameter_2) \times \dots \times f_n(n)$

$\Pi_{i=1}^{n}$ multiplies all the $f_i$ values for the individual parameters

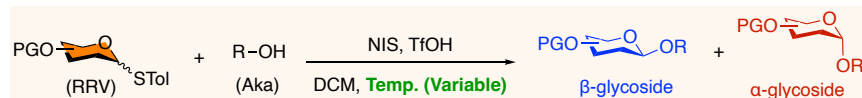**EFI for Temperature**:
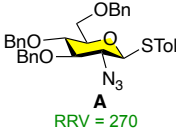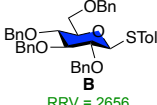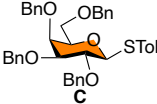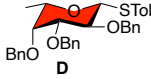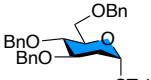
$$EFI_{Temp.} = \frac{\text{α-ratio} \times yield}{RRV \times Aka \times T_{normalized}} \tag{S2}$$

Where, $T_{normalized} = T_{°C} + 273.15$

| An example for $EFI_{Temp.}$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Temp (°C)** | **Solvent** | **RRV** | **Aka** | **α-ratio** | **Yield** | **EFI** | **Efficiency** |
| -40 | DCM | 2656 | 1.36 | 71 | 90 | 0.0068 | Good α-selectivity under mild conditions |

## 2.2 Screening of temperature parameter

**Supplementary Table 1. Screening of temperature effects on glycosylation selectivity across matching partners varying RRV and Aka scale.**

PGO–(RRV)–STol + R–OH (Aka) →[NIS, TfOH; DCM, **Temp. (Variable)**] PGO–O–OR (β-glycoside) + PGO–O–OR (α-glycoside)

| Glycosylating agent | Temp. / Alcohol | -40 °C α-ratio(%) (yield) | -20 °C α-ratio(%) (yield) | 0 °C α-ratio(%) (yield) | 25 °C α-ratio(%) (yield) |
|---|---|---|---|---|---|
| **A** (BnO, BnO, OBn, STol, N₃) RRV = 270 | MeOH **51** | 0% (62%) | 0% (38%) | 0% (7%) | Trace |
| | 6-OH-Glc **53** | 80% (78%) | 77% (85%) | 70% (80%) | 65% (49%) |
| | Ph **55** | 47% (69%) | 47% (67%) | 43% (57%) | Trace |
| | 4-OH-Gal **56** | 68% 62% | 56% (90%) | 53% (74%) | 61% (50%) |
| **B** (BnO, BnO, OBn, STol, BnO) RRV = 2656 | MeOH **51** | 0% (82%) | 27% (69%) | 40% (70%) | 32% (97%) |
| | 6-OH-Glc **53** | 31% (86%) | 30% (85%) | 50% (80%) | 54% (51%) |
| | Ph **55** | 66% (84%) | 62% (86%) | Trace | Trace |
| | 4-OH-Gal **56** | 72% 52% | 71% (78%) | 70% (84%) | 73% (58%) |
| **C** (BnO, OBn, BnO, STol) RRV = 17000 | 1-adm **52** | 40% (87%) | 36% (78%) | 38% (80%) | 31% (92%) |
| **D** (STol, OBn, BnO, OBn) RRV = 72000 | MeOH **51** | 62% (93%) | 40% (91%) | 50% (80%) | 35% (92%) |
| | 6-OH-Glc **53** | 73% (91%) | 64% (87%) | 52% (68%) | 58% (48%) |
| | Ph **55** | 100% (72%) | 100% (92%) | 100% (55%) | 100% (46%) |
| | 4-OH-Gal **56** | 66% 87% | 77% (89%) | 73% (83%) | 77% (48%) |
| **E** (BnO, BnO, OBn, STol) RRV = 10⁶ | MeOH **51** | 83% (38%) | 77% (24%) | 50% (19%) | 100% Trace |
| | 6-OH-Glc **53** | 78% (74%) | 77% (71%) | 72% (60%) | 82% (57%) |
| | Ph **55** | 100% (64%) | 100% (53%) | 100% (52%) | 100% (20%) |
| | 4-OH-Gal **56** | 86% (65%) | 89% (68%) | 84% (48%) | 87% (43%) |

α-ratio(%)= 100- β-ratio (%)

| 100 | +80 | +60 | +50 α(%) | -40 | -60 | -80 | 100 |

Reactions were performed in dichloromethane (DCM) under NIS/TfOH activation. At cryogenic temperatures, glycosylation operates predominantly under kinetic control, favoring β-selectivity. This behavior is particularly evident for low RRV glycosylating agents like glucosamine **A** and **B**, where α-ratios remain negligible (<5%), with EFI nearly 0, reflecting insufficient thermal energy to overcome steric and electronic barriers, where $T_{normalized}$ is °C + 273.15. Temperature critically shapes both selectivity and efficiency in glycosylation. As the temperature increases to 0 °C , the system enters a transitional zone. Partial dissociation of glycosyl intermediates promotes rotational freedom, resulting in an increase in α-selectivity. Glucose **A** shows an α-ratio improvement of +40% with stable yield, indicating balanced kinetic and thermodynamic influences. High RRV glycosylating agents like fucose **D** exhibit enhanced α-selectivity (+35%) with an EFI of 0.00235, reflecting improved efficiency as the system transitions toward thermodynamic control. At higher temperatures, glycosylation shifts to $S_N1$-like pathways, with free rotation of dissociated oxocarbenium ion intermediates favoring the thermodynamically stable α-product, specially the high RRV glycosylating agents, such as fucose **D** and 2-deoxy-glucopyranoside **E**, predominates α-selectivity showcasing the dominance of thermodynamic control, yet EFI values remain modest—largely due to declining yields, particularly for agent E, highlighting the increased competition from side reactions and thermal decomposition.

## 2.2.1  System analysis of quantification of temperature parameter
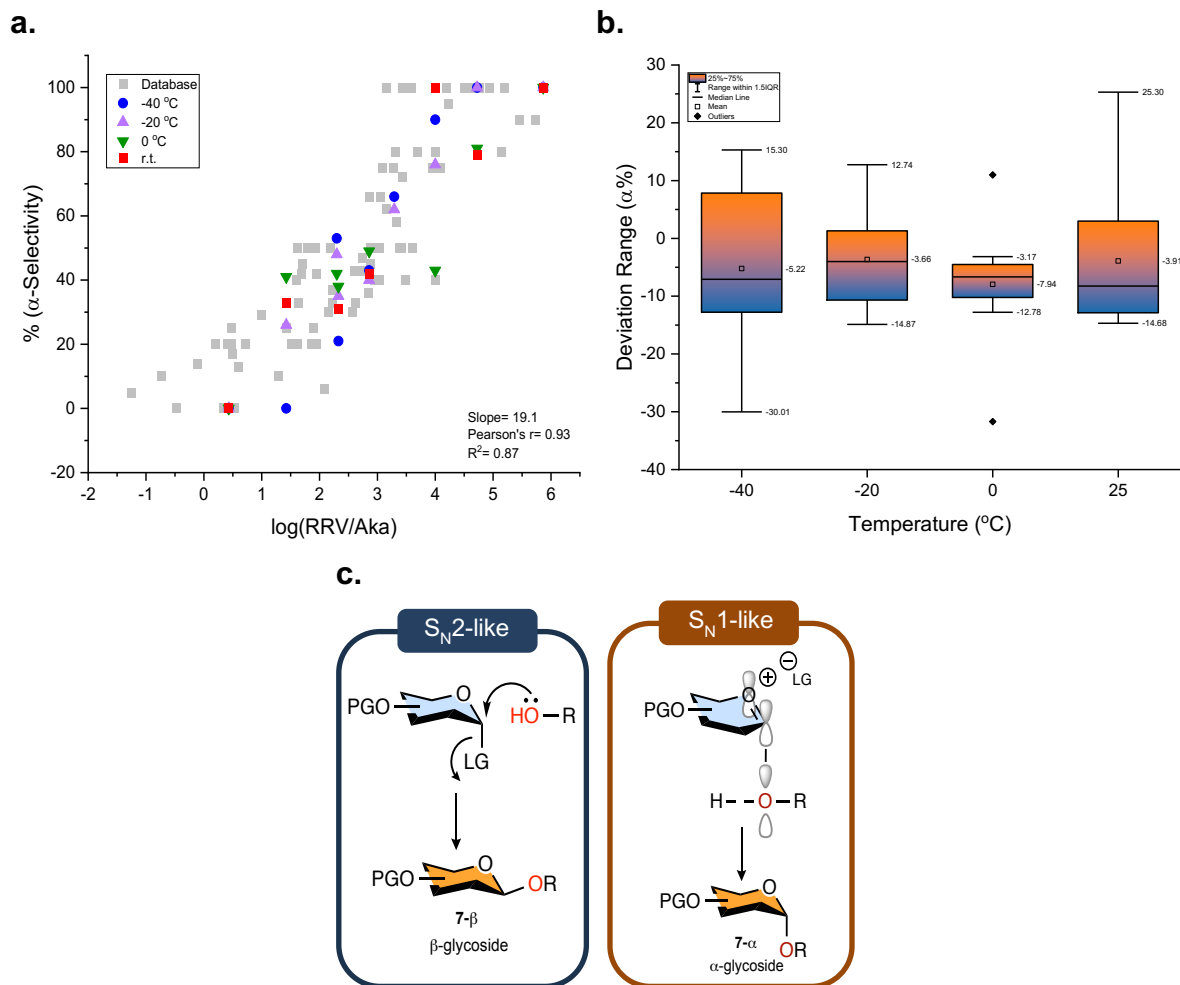


**Supplementary Figure 1. Temperature dependence of glycosylation selectivity and mechanism.** (a) Correlation between log(RRV/Aka) and α-ratio (%) for glycosylation reactions at different temperatures (-40 °C, -20 °C, 0 °C, 25 °C). (b) Box plot showing deviation range in selectivity across temperatures. (c) Proposed mechanistic pathways for β- ($S_N2$-like) and α-glycoside ($S_N1$-like) formation. RRV, relative reactivity value; LG, leaving group; PGO, protecting group; OR, glycoside product.

## 2.3 Screening of solvent parameter

**Supplementary Table 2. Screening of solvent effects on glycosylation selectivity across matching partners varying RRV and Aka scale.**



| Glycosylating agent | Solvent / Alcohol | DCM α-ratio(%) (yield) | THF α-ratio(%) (yield) | Toluene α-ratio(%) (yield) | DCM/ACN[a] α-ratio(%) (yield) | DCM/DMF[a] α-ratio(%) (yield) | DCM/p-Dioxane[b] α-ratio(%) (yield) |
|---|---|---|---|---|---|---|---|
| **A** RRV = 270 | MeOH **51** | 0% (62%) | 0% (44%) | 0% (54%) | 0% (58%) | N.R.[c] | N.R.[c] |
| | 6-OH-Glc **53** | 20% (78%) | 63% (28%) | 75% (22%) | 100% (94%) | 40% (23%) | 60% (52%) |
| | Ph **55** | 47% (69%) | 78% (18%) | 85% (36%) | 62% (21%) | N.R.[c] | 78% (39%) |
| | 4-OH-Gal **56** | 68% 62% | 44% (33%) | 58% (23%) | 55% (87%) | 50% (38%) | 62% (55%) |
| **B** RRV = 2656 | MeOH **51** | 0% (82%) | 23% (92%) | 47% (82%) | 18% (85%) | Trace | 33% (30%) |
| | 6-OH-Glc **53** | 31% (86%) | 30% (55%) | 80% (85%) | 90% (85%) | 50% (40%) | 54% (63%) |
| | Ph **55** | 66% (84%) | N.R.[c] | 70% (73%) | 69% (80%) | N.R.[c] | 47% (80%) |
| | 4-OH-Gal **56** | 72% 52% | 68% (49%) | 61% (47%) | 32% (96%) | 70% (63%) | 78% (68%) |
| **C** RRV = 17000 | 1-adm **52** | 40% (93%) | 52% (89%) | 38% (90%) | 40% (93%) | N.R.[c] | 45% (72%) |
| **D** RRV = 72000 | MeOH **51** | 63% (90%) | 45% (41%) | 20% (75%) | 26% (84%) | 76% (70%) | 40% (46%) |
| | 6-OH-Glc **53** | 73% (91%) | 64% (87%) | 71% (60%) | 79% (87%) | 72% (56%) | 89% (77%) |
| | Ph **55** | 100% (72%) | 100% (68%) | 100% (78%) | 100% (59%) | N.R.[c] | 100% (67%) |
| | 4-OH-Gal **56** | 66% 87% | 77% (89%) | 63% (80%) | 64% (91%) | 66% (50%) | 73% (82%) |
| **E** RRV = 10⁶ | MeOH **51** | 83% (43%) | 47% (39%) | 41% (21%) | 45% (33%) | 73% (58%) | 38% (46%) |
| | 6-OH-Glc **53** | 78% (74%) | 77% (58%) | 72% (60%) | 75% (68%) | 72% (56%) | 90% (55%) |
| | Ph **55** | 100% (64%) | 100% (18%) | 100% (48%) | 100% (31%) | N.R.[c] | N.R.[c] |
| | 4-OH-Gal **56** | 86% (65%) | 89% (68%) | 83% (43%) | 79% (45%) | 84% (38%) | 92% (32%) |

| 100 | +80 | +60 | +50 | α(%) | -40 | -60 | -80 | 100 |

α-ratio(%)= 100- β-ratio (%), a= 1:2 (*v/v*), b= 2:1 (*v/v*), c= No Reaction

9

Reactions were performed at -40 ℃ under NIS/TfOH activation, varying solvents. Solvent polarity and coordination ability significantly influenced both α-ratio and yield. In dichloromethane (DCM), low-reactivity glycosylating agents (e.g., agent **A**, **C**) showed strong β-selectivity, consistent with $S_N2$-like behavior and minimal dissociation of contact ion pairs (CIPs). For higher-reactivity glycosylating agents (e.g., agent **D**), DCM achieved moderate α-selectivity and high yields. THF improves α-ratios by +23%, reflecting an improved mechanistic balance. For the same glycosylating agents, Dimethylformamide (DMF) increases α-ratios by +47%, consistent with enhanced stabilization of solvent-separated ion pairs (SSIPs) and increased front-side attack. The non-reactivity observed in DCM/*p*-Dioxane underscores the importance of solvent compatibility in stabilizing intermediates and enabling effective activation. Conversely, low EFI values (e.g., 0.00 for DCM with low RRV agents) reflect kinetic control dominated by CIPs and $S_N2$-like mechanisms, highlighting the critical role of solvent environment in shaping glycosylation outcome.

**Supplementary Table 3. The polarity indices for common solvents.**

| Solvent | Polarity Index (P') |
|---|---|
| Dichloromethane (DCM) | 3.1 |
| Tetrahydrofuran (THF) | 4.0 |
| Toluene | 2.4 |
| Dimethylformamide (DMF) | 6.4 |
| Acetonitrile (ACN) | 5.8 |
| *p*-Dioxane | 4.8 |

These values are sourced from the solvent polarity chart provided by Louisiana State University (LSU Macro). *P'* were used to obtain EFI values.

### 2.3.1 System analysis of quantification of solvent parameter
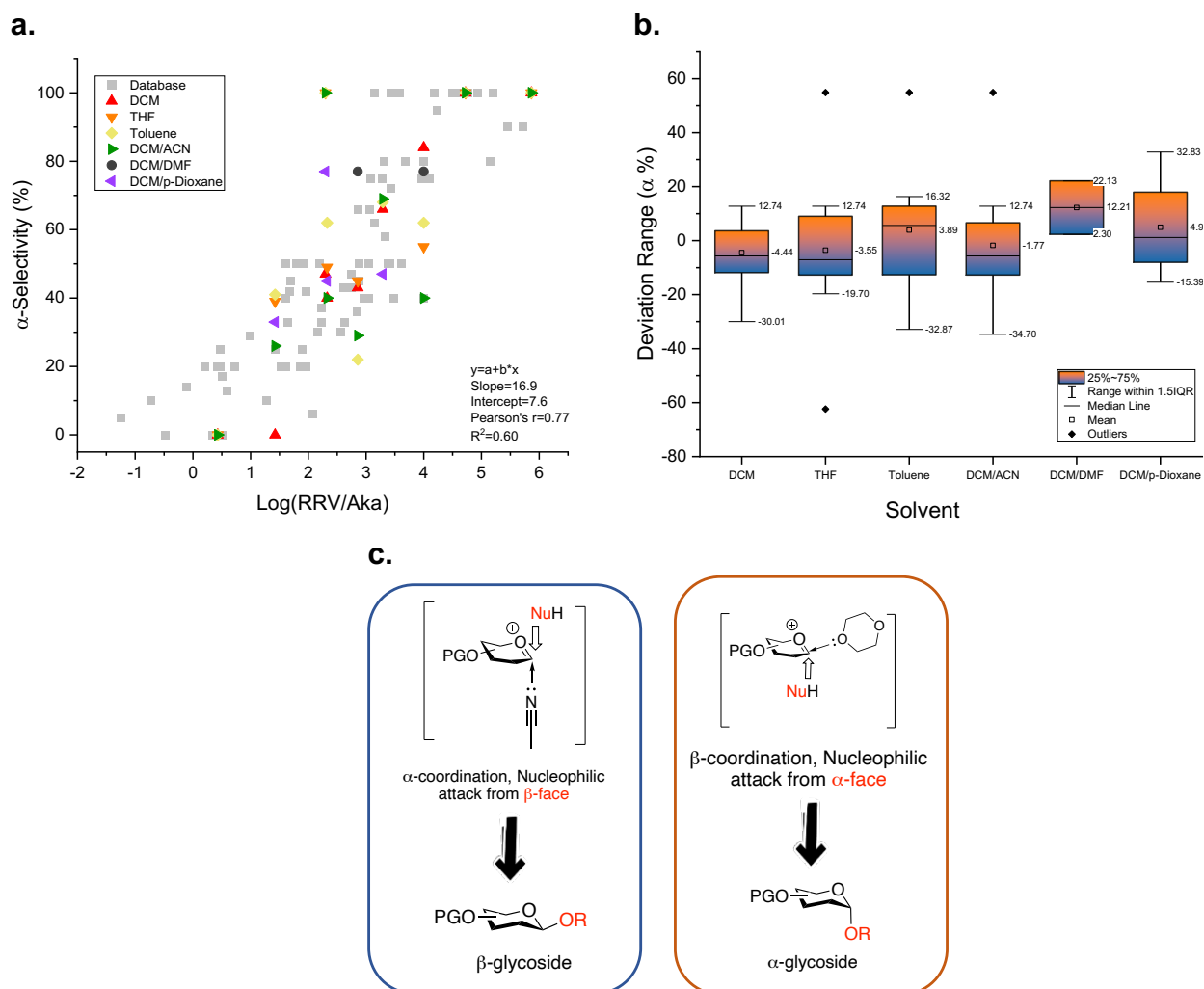


**Supplementary Figure 2. Solvent dependence of glycosylation selectivity and mechanism.** (a) Correlation between log(RRV/Aka) and α-ratio (%) for glycosylation reactions at various solvents (DCM, THF, Toluene, DCM/ACN, DCM/DMF, DCM/*p*-Dioxane). (b) Box plot showing deviation range in selectivity across solvents. (c) Proposed mechanistic pathways for β-and α-glycoside formation. NuH, Nucleophilic alcohol; PGO, protecting group; OR, glycoside product.

## 2.4 Screening of concentration parameter

**Supplementary Table 4. Screening of concentration effect on glycosylation selectivity across matching partners varying RRV and Aka scale.**



| Glycosylating agent | Alcohol | 0.308M α-ratio(%) (yield) | 0.167M α-ratio(%) (yield) | 0.051M α-ratio(%) (yield) | 0.036M α-ratio(%) (yield) | 0.018M α-ratio(%) (yield) |
|---|---|---|---|---|---|---|
| A (RRV = 270) | MeOH 51 | 0% (33%) | 0% (62%) | 0% (64%) | 0% (74%) | 0% (87%) |
| | 6-OH-Glc 53 | 42% (76%) | 20% (78%) | 38% (72%) | 46% (68%) | 50% (53%) |
| | Ph 55 | 59% (20%) | 47% (69%) | 55% (60%) | 55% (88%) | 52% (55%) |
| | 4-OH-Gal 56 | 31% 82% | 28% 62% | 24% 72% | 26% 70% | 30% 68% |
| B (RRV = 2656) | MeOH 51 | 11% (60%) | 0% (82%) | 27% (88%) | 13% (32%) | 12% (20%) |
| | 6-OH-Glc 53 | 38% (62%) | 30%[a] (86%) | 41% (65%) | 32% (61%) | 36% (47%) |
| | Ph 55 | 71% (73%) | 66% (84%) | 71% (77%) | 71% (90%) | 62% (90%) |
| | 4-OH-Gal 56 | 64% (81%) | 72% (52%) | 69% (76%) | 70% (61%) | 68% (65%) |
| C (RRV = 17000) | 1-adm 52 | 15% (75%) | 40% (94%) | 18% (96%) | 20% (61%) | 19% (70%) |
| D (RRV = 72000) | MeOH 51 | 24% (76%) | 59% (90%) | 27% (81%) | 40% (96%) | 27% (72%) |
| | 6-OH-Glc 53 | 69% (87%) | 73% (91%) | 70% (78%) | 66% (74%) | 63% (60%) |
| | Ph 55 | 100% (94%) | 100% (72%) | 100% (78%) | 100% (93%) | 100% (97%) |
| | 4-OH-Gal 56 | 62% (89%) | 66% (87%) | 73% (85%) | 72% (85%) | 70% (80%) |
| E (RRV = $10^6$) | MeOH 51 | H.D.[a] | 83% (43%) | 66% (60%) | 83% (56%) | 91% (60%) |
| | 6-OH-Glc 53 | 75% (63%) | 78%[a] (74%) | 72% (65%) | 74% (69%) | 70% (45%) |
| | Ph 55 | 100% (66%) | 100% (64%) | 100% (68%) | 100% (74%) | 100% (60%) |
| | 4-OH-Gal 56 | 80% (60%) | 86% (65%) | 78% (58%) | 74% (62%) | 77% (55%) |

α-ratio(%)= 100- β-ratio (%), a= Hydrolyzed glycosylating agent

The concentration parameter in glycosylation reactions profoundly impacts stereoselectivity and yield, governed by ion-pair dynamics, acid availability, and reaction time. At high concentrations (e.g., 0.308 M and 0.167 M), kinetic control dominates due to $\alpha$-triflate stabilization and the high availability of activating acids such as triflic acid. This abundance of acid accelerates covalent-triflate formation, favoring $\beta$-selectivity via $S_N2$-like pathway. For low RRV glycosylating agent **A** $\beta$-ratios remain dominant across high concentrations via unimolecular pathway, while yields decline by $-7\%$/M. As concentration decreases (e.g., 0.051 M and 0.036 M), partial CIP dissociation into SSIPs occurs, promoting bimolecular pathways. In these intermediate conditions, reduced acid availability and extended reaction rate enable the formation of both kinetically and thermodynamically controlled products. For glycosylating agent **B**, $\alpha$-ratios improve significantly, increasing by $+20\%$/M, at 0.051 M to 0.036 M, yields remain relatively stable, declining slightly by $-3\%$/M, highlighting a balance between kinetic and thermodynamic control. At low concentrations (e.g., 0.018 M), complete CIP dissociation and longer reaction times favor SSIP formation and thermodynamically stable intermediates, this shift promotes front-side attack via $S_N1$-like pathways, yielding higher $\alpha$-selectivity. For high RRV agents **D** and **E**, $\alpha$-ratios increase by $+18\%$/M, while yields decrease slightly by $-1\%$/M due to side reactions or agent decomposition. These trends underscore the enhanced flexibility of intermediates at low concentrations, favoring thermodynamic products over kinetically controlled ones.
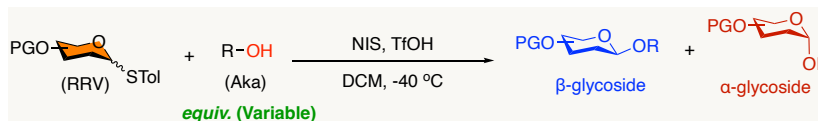
## 2.4.1 System analysis of quantification of concentration parameter

**a.**



**b.**



**c.**
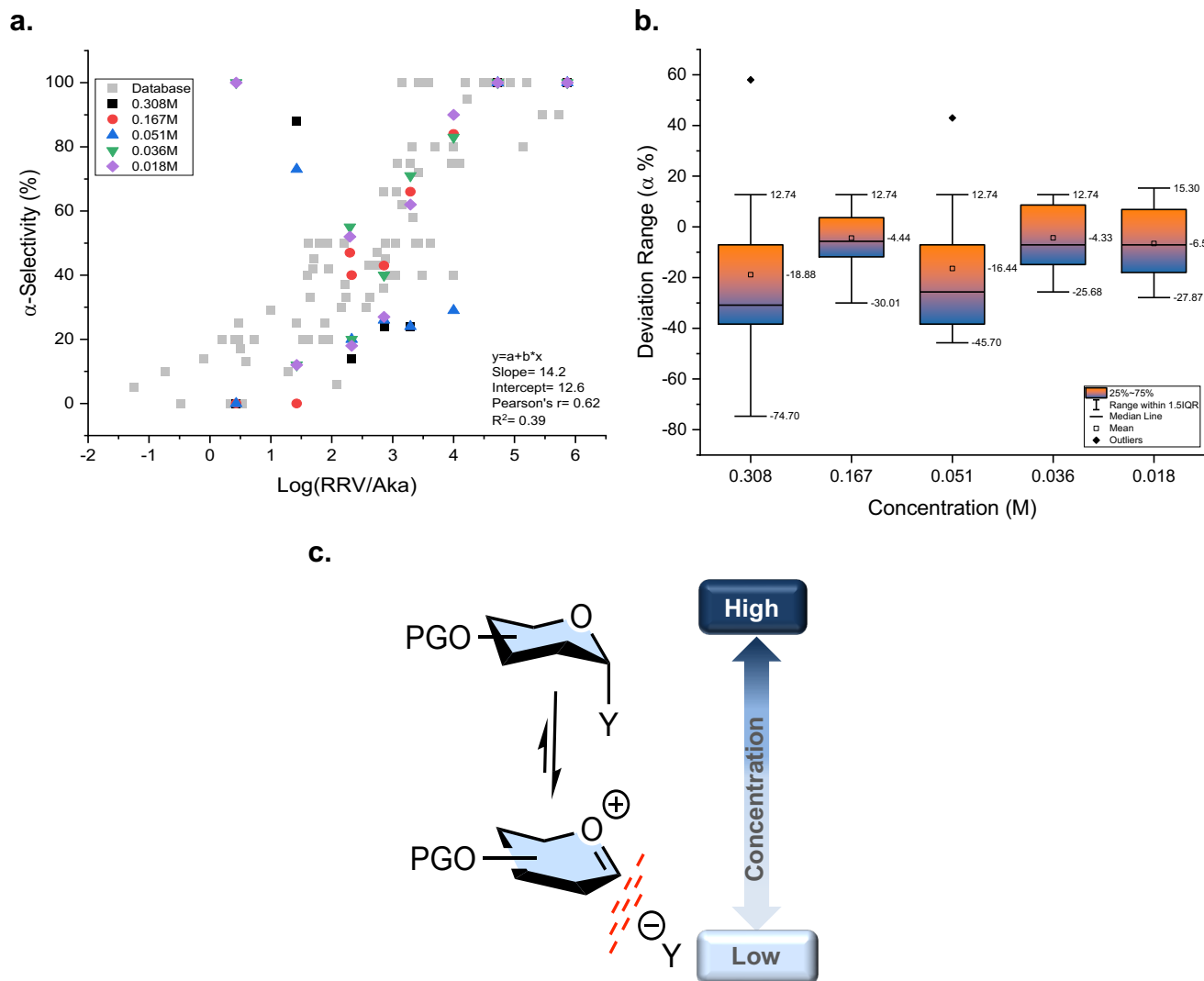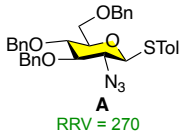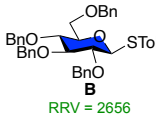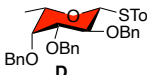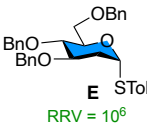


**Supplementary Figure 3. Concentration dependence of glycosylation selectivity and mechanism.** (a) Correlation between log(RRV/Aka) and α-ratio (%) for glycosylation reactions at various concentrations. (b) Box plot showing deviation range in selectivity across various concentrations. (c) Schematic illustration of the mechanistic transition from high to low alcohol concentration, highlighting the shift from contact ion pair to solvent-separated ion pair pathways. Y, any nucleophilic specie presents in reaction environment; PGO, protecting group.

## 2.5  Screening of stoichiometry parameter

**Supplementary Table 5. Screening of stoichiometry effects on glycosylation selectivity across matching partners varying RRV and Aka scale.**



| Glycosylating agent | equiv. | 4-OH-Gal 56 α-ratio(%) (yield) | Ph 55 α-ratio(%) (yield) | 4-OH-Glc 54 α-ratio(%) (yield) | 6-OH-Glc 53 α-ratio(%) (yield) | MeOH 51 α-ratio(%) (yield) |
|---|---|---|---|---|---|---|
| **A** RRV = 270 | 1:0.66 | 34% (75%) | 100% (35%) | 33% (83%) | 35% (75%) | 0% (30%) |
| | 1:1.5 | 28% 62% | 87% (69%) | 25% (66%) | 20% (78%) | 0% (62%) |
| **B** RRV = 2656 | 1:0.66 | 75% (50%) | 77% (68%) | 33% (78%) | 33% (75%) | 0% (52%) |
| | 1:1.5 | 72% (52%) | 66% (84%) | 45% (77%) | 30% (86%) | 0% (82%) |
| **D** RRV = 72000 | 1:0.66 | 77% (75%) | 100% (53%) | 91% (65%) | 78% (66%) | 21% (50%) |
| | 1:1.5 | 66% (87%) | 100% (72%) | 91% (61%) | 73% (91%) | 62% (93%) |
| **E** RRV = 10^6 | 1:0.66 | 91% (52%) | 100% (37%) | 92% (80%) | 83% (71%) | 30% (66%) |
| | 1:1.5 | 86% (65%) | 100% (64%) | 90% (76%) | 78% (74%) | 83% (38%) |

α-ratio(%)= 100- β-ratio (%),  

| 100 | +80 | +60 | +50 | **α(%)** | -40 | -60 | -80 | 100 |

Stoichiometry modulates glycosylation outcomes by influencing the availability of reactive intermediates and the competition between glycosylating agents and alcohols. At low agent-to-alcohol ratios (1:0.66), the reduced agent concentration leads to enhanced alcohol competition and limited intermediate stabilization, favoring β-selectivity through kinetic-controlled pathways. Low RRV agents prevails β-selectivity, attributed to a constrained reaction environment and higher likelihood of backside nucleophilic attack. As the ratio increases, intermediate stabilization improves, and the transition to mixed control emerges, facilitating α-selectivity improvements (+45%) and moderate yield increases (+39%) for mid-RRV glycosylating agents. At higher agent ratios, the excess of agent molecules reduces competition, ensuring efficient oxocarbenium ion formation and enhanced α-product. This mechanistic shift results in maximized α-ratios and yields (+51%), reflecting improved intermediate accessibility and prolonged reaction efficacy.

## 2.6 Comparative performance analysis of ML models in predicting data-driven glycosylation outcomes
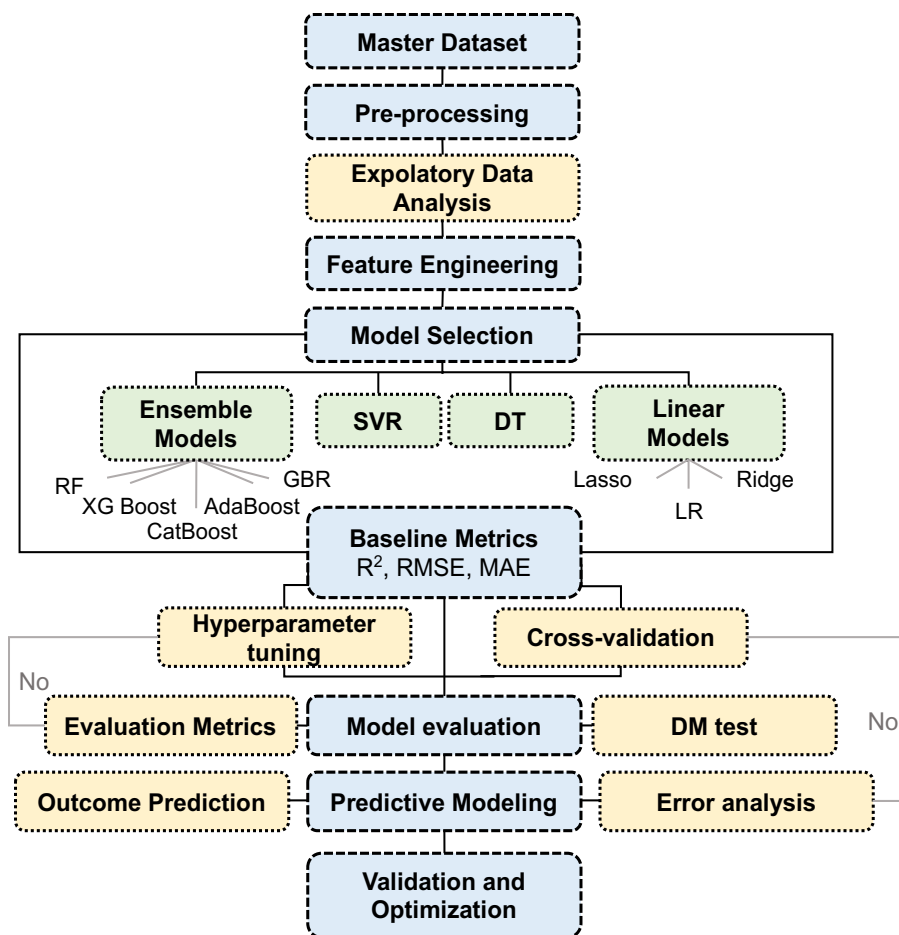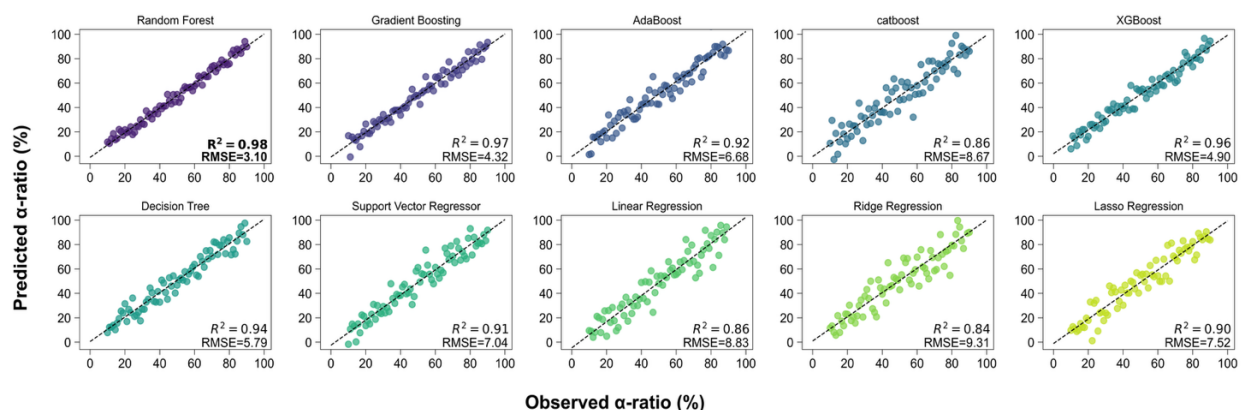


**Supplementary Figure 4. Machine learning pipeline for predictive modeling of glycosylation outcomes.** The schematic outlines the complete workflow adopted for constructing, evaluating, and optimizing machine learning (ML) models aimed at predicting glycosylation reaction outcomes (e.g., $\alpha$-ratio). The process begins with a curated master dataset, followed by standard pre-processing steps to ensure data integrity. Exploratory data analysis (EDA) is performed to uncover underlying patterns and guide feature engineering, where domain-specific variables such as RRV, Aka, EFI, temperature, and solvent are computed or selected. Subsequent model selection involves benchmarking multiple algorithm classes, including support vector regression (SVR), decision trees (DT), linear models, and ensemble methods (e.g., random forest, XGBoost). Models are first trained and evaluated using baseline metrics such as $R^2$, root mean square error (RMSE), and mean absolute error (MAE). Models failing to meet performance thresholds undergo hyperparameter tuning or k-fold cross-validation to improve generalizability. Statistical robustness is validated via the Diebold-Mariano (DM) test and evaluation metrics. Following this, predictive modeling and error analysis are conducted, leading to final model optimization. The output is a robust, interpretable model for accurate glycosylation outcome prediction, with applications in reaction design and optimization.

**Supplemantary Table 6. Molecular and reaction descriptors used to train the models.**

| Variable | Feature Name |
|---|---|
| x1 | C1-functionality_donor |
| x2 | C1_configuration_donor |
| x3 | C2- functionality_donor |
| x4 | C2_configuration_donor |
| x5 | C3-functionality_donor |
| x6 | C4-functionality_donor |
| x7 | C4_configuration_donor |
| x8 | C6-functionality_donor |
| x9 | RRV |
| x10 | C1-functionality_acceptor |
| x11 | C1_configuration_acceptor |
| x12 | C2- functionality_ acceptor |
| x13 | C3-functionality_ acceptor |
| x14 | C4-functionality_ acceptor |
| x15 | C6-functionality_ acceptor |
| x16 | Alcohol_position |
| x17 | Alcohol_type |
| x18 | Aka |
| x19 | Temperature |
| x20 | Solvent |
| x21 | Concentartion |
| x22 | Stoichiometery |
| x23 | Promoter |
| x24 | EFI |
| y1 | α-ratio |
| y2 | Yield |

Donor, glycosylating agent; acceptor, nucleophilic alcohol.

**Supplemantary Figure 5. Prediction of glycosylation stereoselectivity ($\alpha$-ratio, %) using 10 individual machine learning models.** Each subplot compares predicted versus observed $\alpha$-ratio values across the test set. Performance is evaluated using coefficient of determination ($R^2$) and root mean squared error (RMSE). Ensemble tree-based models (e.g., Random Forest, Gradient Boosting) generally outperform linear models followed by XGBoost, reflecting their ability to capture non-linear structure–reactivity relationships.
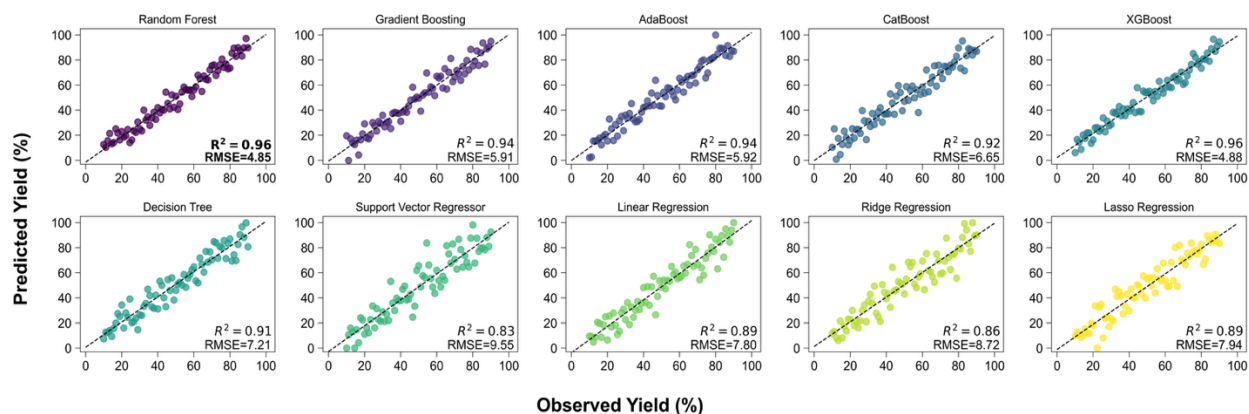


**Supplemantary Figure 6. Prediction of glycosylation reaction yield (%) using 10 individual machine learning models.** Predicted versus observed yield values are shown across the test set for each model. Performance metrics ($R^2$ and RMSE) demonstrate that ensemble tree-based models, particularly Random Forest and XGBoost ($R^2 = 0.96$, RMSE $\approx$ 4.85–4.88), offer superior predictive accuracy compared to linear or kernel-based methods, highlighting their ability to model nonlinear, context-dependent reaction efficiencies.

18
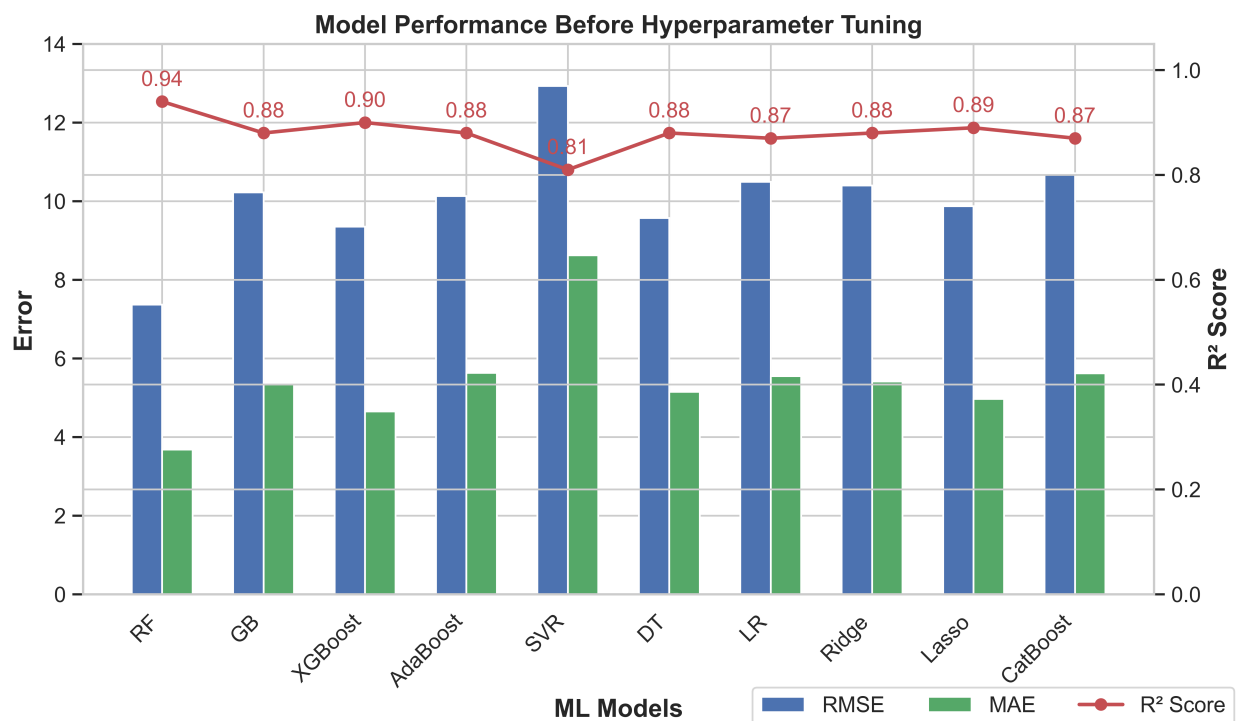
**Supplementary Figure 7. Comparison of machine learning model performance prior to hyperparameter tuning.** Bar plots show the root mean square error (RMSE, blue) and mean absolute error (MAE, green) for each of the 10 machine learning models evaluated on α-ratio prediction. The red line overlay indicates the $R^2$ score for each model. Random Forest (RF), XGBoost, and Linear Regression (LR) achieved the highest $R^2$ values (0.90), while Support Vector Regression (SVR) exhibited the lowest performance ($R^2 = 0.81$) with the highest RMSE and MAE. These baseline results illustrate the relative capacity of each model to capture underlying relationships before optimization of models on the glycosylation dataset.
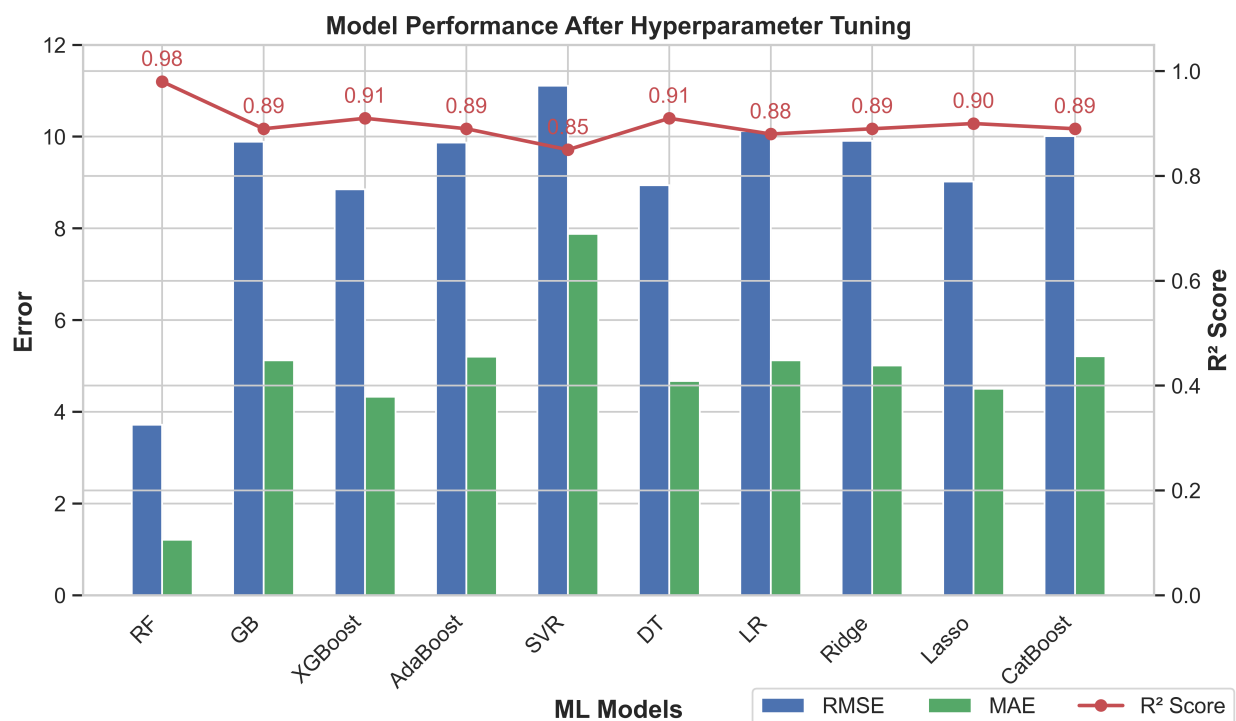
**Supplementary Figure 8. Machine learning model performance after hyperparameter tuning for α-ratio prediction.** Post-tuning evaluation of 10 models reveals improved accuracy across most metrics. Bar plots represent RMSE (blue) and MAE (green), while the red line indicates R² score. Random Forest (RF) shows the best overall performance with the highest R² (0.92) and lowest RMSE and MAE. XGBoost and Decision Tree (DT) also improved significantly after tuning (R² = 0.91). Support Vector Regression (SVR) remains the least performant model (R² = 0.85), though with reduced error. These results highlight the effectiveness of hyperparameter optimization in enhancing model generalization and predictive power.

**Supplementary Table 7. Summary of Hyperparameter Configurations and Tuning Strategies for All Regression Models.**

| Model | Hyperparameter | Value | Tuning Strategy |
|---|---|---|---|
| Linear Regression | Regularization | None | No tuning required |
| Ridge Regression | alpha | 10.0 | Tuned via CV from [0.01, 0.1, 1, 10, 100] |
| Lasso Regression | alpha | 0.1 | Tuned via CV from [0.001, 0.01, 0.1, 1] |
| Support Vector Regressor | kernel | rbf | Best for non-linear data |
| Support Vector Regressor | C | 10 | Tuned via grid search |
| Support Vector Regressor | epsilon | 0.1 | Tested range: [0.01, 0.05, 0.1] |
| Decision Tree | max_depth | 5 | Tuned manually based on validation error |
| Decision Tree | min_samples_split | 20 | Tuned to reduce overfitting |
| Decision Tree | min_samples_leaf | 10 | Tuned for generalization |
| Random Forest | n_estimators | 1000 | Tuned for stable performance |
| Random Forest | max_depth | None | Allow deep trees, tested shallower variants |
| Random Forest | min_samples_leaf | 1 | Default; tested up to 10 |
| Random Forest | max_features | sqrt | Best balance of bias/variance |
| Random Forest | oob_score | TRUE | Used for internal validation |
| Gradient Boosting | n_estimators | 500 | Tuned with early stopping |
| Gradient Boosting | learning_rate | 0.05 | Balanced speed and accuracy |
| Gradient Boosting | max_depth | 5 | Tuned for bias-variance tradeoff |
| XGBoost | n_estimators | 500 | Tuned with early stopping |
| XGBoost | learning_rate | 0.05 | Optimal via validation |
| XGBoost | max_depth | 5 | Avoided overfitting |
| XGBoost | subsample | 0.9 | Improves robustness |
| XGBoost | colsample_bytree | 0.8 | Controls tree diversity |
| XGBoost | reg_alpha | 1 | Regularization term |
| XGBoost | reg_lambda | 1 | Regularization term |
| AdaBoost | n_estimators | 300 | Stable after tuning |
| AdaBoost | learning_rate | 0.5 | Balanced under/overfitting |
| CatBoost | iterations | 500 | Tuned with early stopping |
| CatBoost | depth | 5 | Tested [4–8], best performance |
| CatBoost | learning_rate | 0.05 | Best trade-off |
| CatBoost | l2_leaf_reg | 3 | Controls regularization |

This table lists the key hyperparameters applied to each regression model within the study, along with the corresponding values and the tuning strategies or rationale behind their selection.

Conventional models such as Linear, Ridge, and Lasso regression required minimal tuning, while ensemble and kernel-based models (e.g., Random Forest, XGBoost, CatBoost, SVR) underwent extensive grid or cross-validation–based optimization. For Random Forest, out-of-bag (OOB) validation was used to monitor generalization without relying on an external holdout set. Parameters were tuned to balance model complexity, reduce overfitting, and enhance predictive performance on both α-ratio and yield targets. Early stopping and regularization terms were emphasized for boosting-based methods to avoid overfitting on small or noisy datasets.

$$Hybrid = RF + XGB$$

$$RFFit = RFModel\ (X, y)$$
$$XGB\ Fit\ = XGBoost\ Model\ (X, y)$$

$$\hat{y}_{RF} = RFModel\ (X)$$
$$\hat{y}_{XGB} = XGBoost\ Model\ (X)$$

$$\hat{y}_{Hybrid} = \frac{\hat{y}_{RF} + \hat{y}_{XGB}}{2}$$
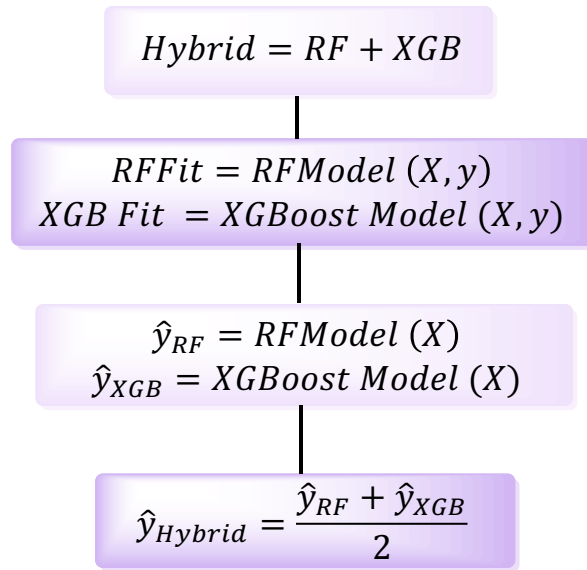
**Supplemantary Figure 9. Flowchart depicting the development of the hybrid model combining Random Forest (RF) and XGBoost.** The process involves initializing the model, training both algorithms, generating individual predictions, and averaging them to form the final unified hybrid prediction.
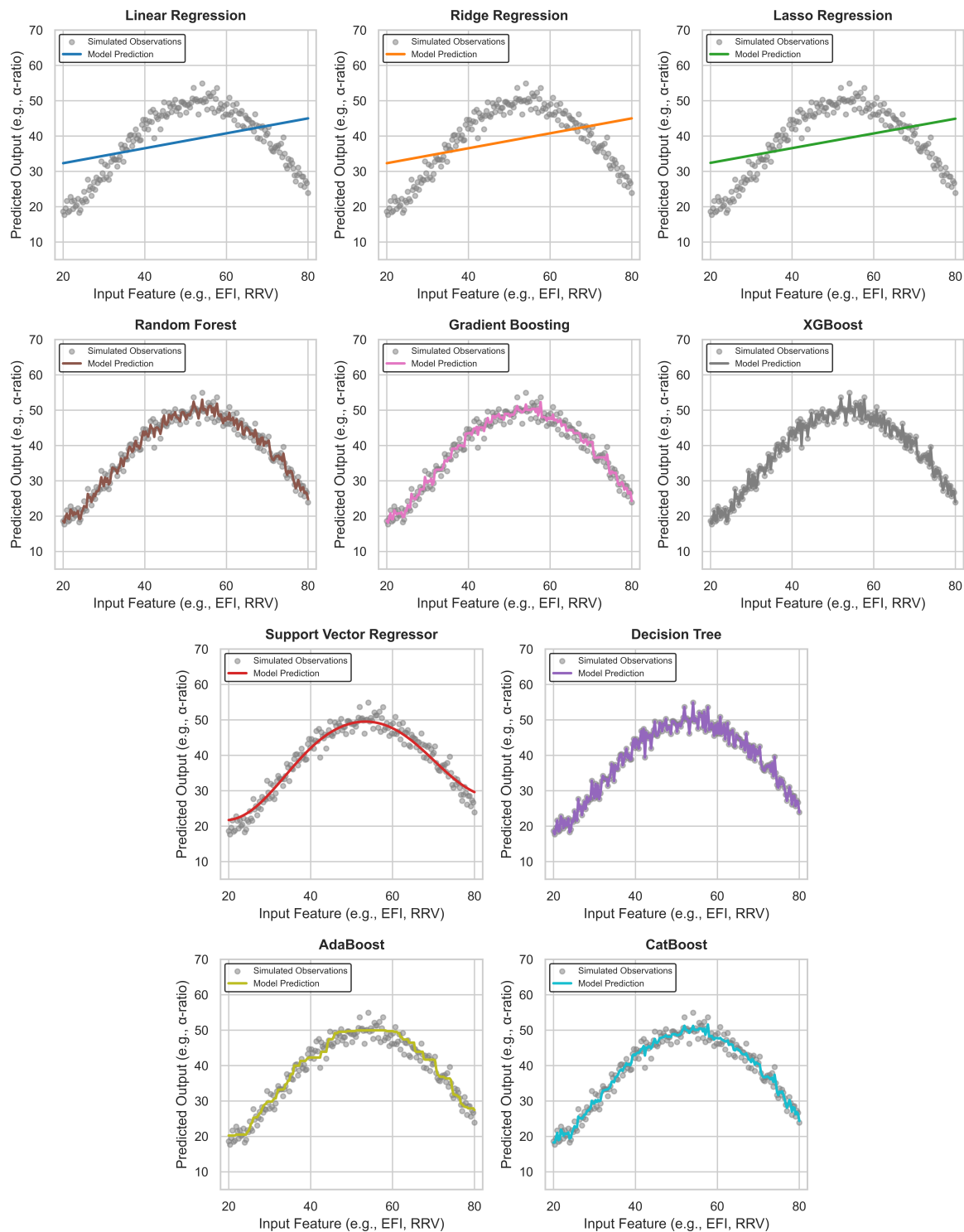
**Supplementary Figure 10. Visualization of model behavior on non-linear data.** Plots show input–output relationships to understand *non-linearity* in dataset, such as solvent or alcohol type to α-ratio. To evaluate how different ML models respond to non-linear trends often observed in

glycosylation data (e.g., solvent or alcohol type influencing α-ratio), we trained ten models on a simulated dataset with a sinusoidal pattern and added noise. The linear models including Linear Regression, Ridge, and Lasso showed strong underfitting, as evident from their inability to capture the curve. These models apply fixed functional forms, making them inadequate for representing complex chemical systems where dependencies between input and output are non-linear. In contrast, non-linear models showed varying degrees of success. Support Vector Regressor (SVR) captured curvature better but lacked the flexibility to model sharper transitions. Decision Tree, while adaptive, exhibited abrupt, step-wise predictions indicating overfitting to local patterns. The ensemble-based models including Random Forest, Gradient Boosting, XGBoost, AdaBoost, and CatBoost offered significant improvements. Among them, Random Forest provided the most consistent and smooth approximation, effectively balancing bias and variance. These results reinforce the suitability of tree-based ensemble learners, particularly Random Forest, in modeling the intricate relationships governing glycosylation.
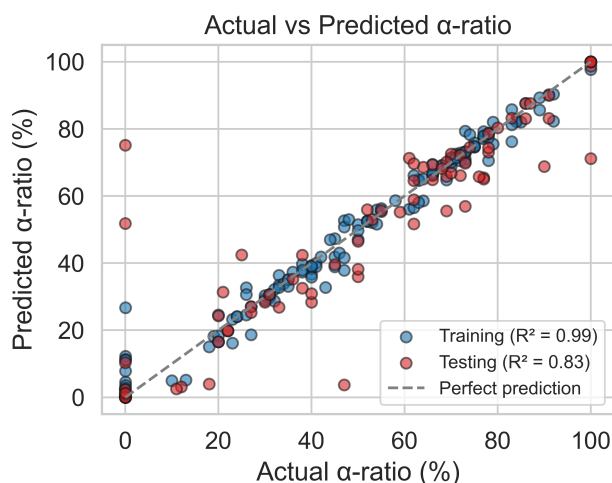


**Supplementary Figure 11. Training and test α-ratio (%) prediction using the Random Forest model for non-linear data.** Blue and red points represent the training and testing datasets with 70/30 (%) data splitting, respectively. The dashed gray line indicates perfect prediction ($y = x$). The model achieved excellent predictive performance, with $R^2 = 0.99$ for the training set and $R^2 = 0.83$ for the testing set, indicating strong generalization and minimal overfitting.
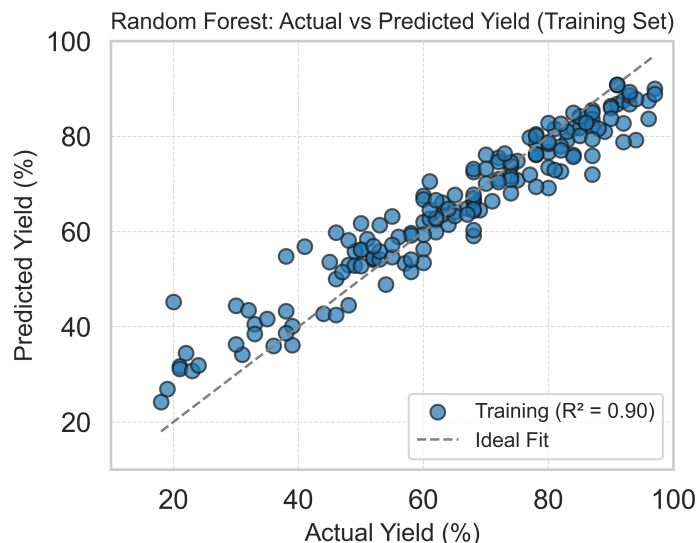
**Supplementary Figure 12. Scatter plot comparing actual and predicted reaction yields (%) for the training dataset using a Random Forest model for non-linear data.** Each blue point represents a training sample with 70/30 (%) data splitting. The gray dashed line indicates perfect prediction ($y = x$). The model achieved an $R^2$ of 0.90 on the training set, indicating a strong fit to the training data with minimal residual error.
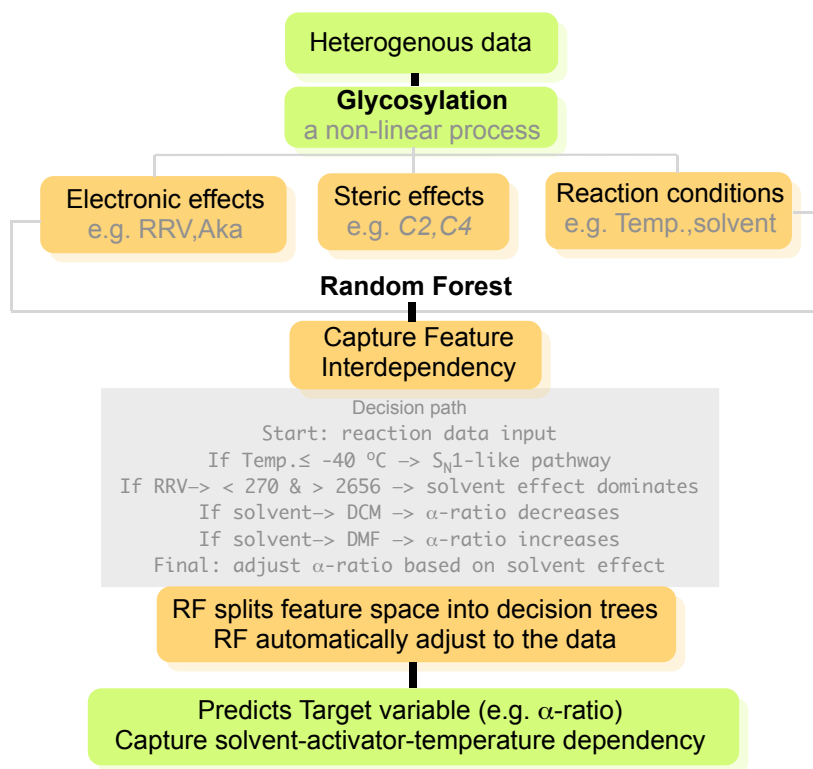


**Supplementary Figure 13. Random Forest model architecture for capturing nonlinear dependencies in glycosylation reactions.** Heterogeneous input features including electronic

parameters (e.g., relative reactivity value, RRV; nucleophilicity constant, Aka), steric descriptors (e.g., C2, C4 substitution), and reaction conditions (e.g., temperature, solvent) are integrated into the RF framework. The model captures feature interdependencies and partitions the feature space into decision trees, adjusting automatically to the data. Example decision paths illustrate how RF encodes mechanistic relationships such as temperature thresholds, solvent polarity, and RRV-based switching to predict glycosylation selectivity (e.g., $\alpha$-ratio), enabling solvent–promoter–temperature dependent modeling.
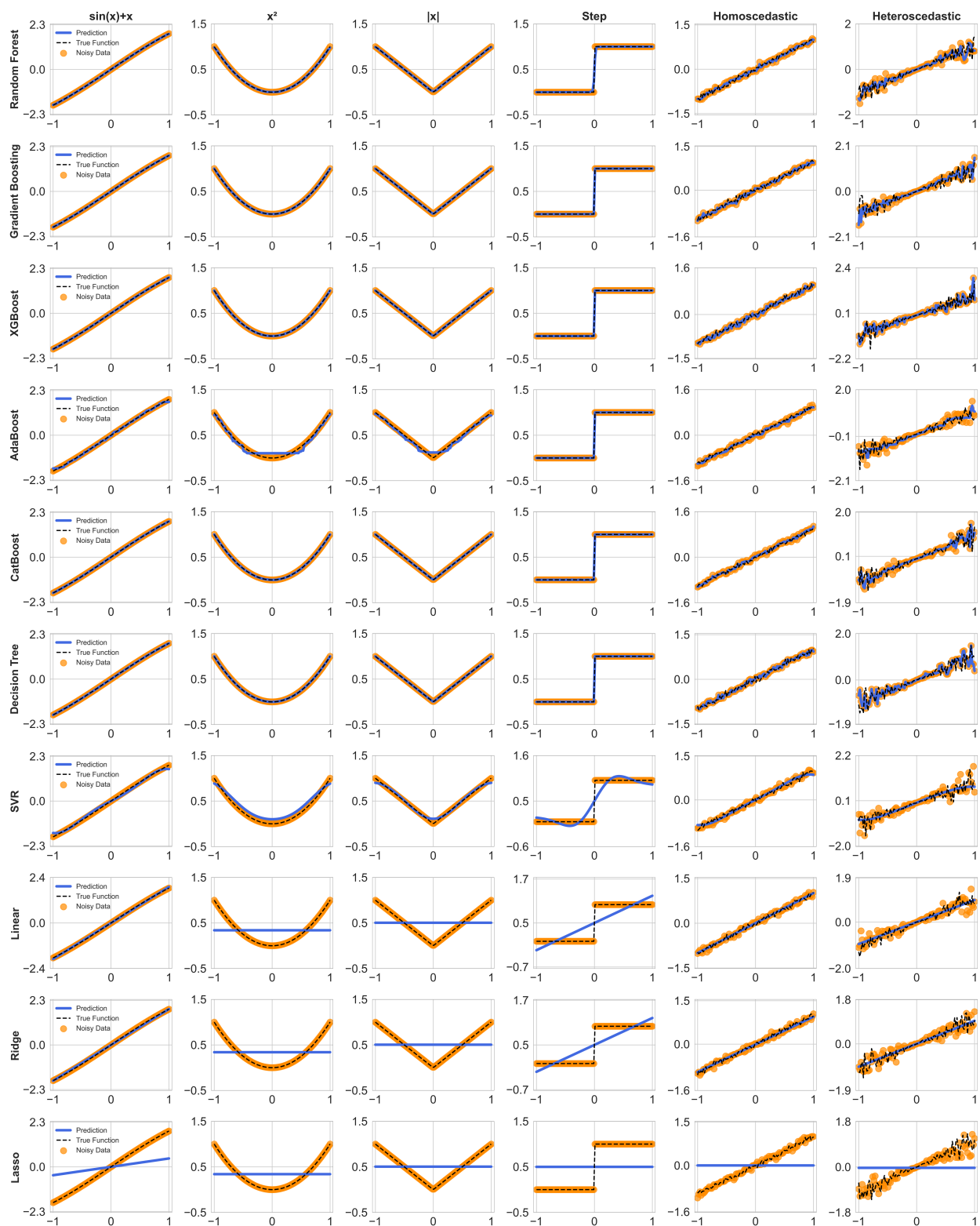
**Supplementary Figure 14. Qualitative analysis of all ML models based on 600 batch glycosylation reactions over 5 noise types.**

## 2.7 Core ML model schematics
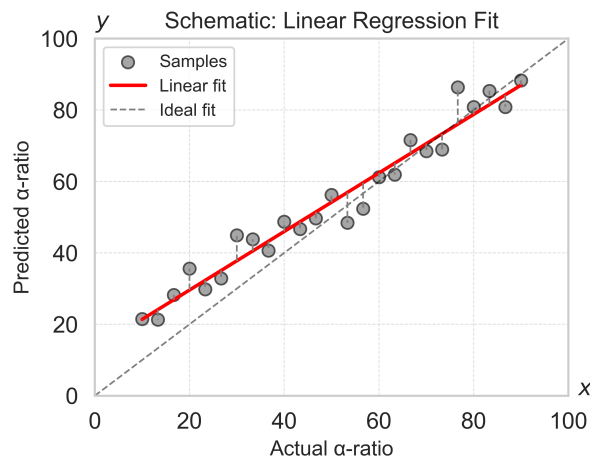


**Supplementary Figure 15. Linear regression applied to glycosylation data.** This figure illustrates a linear regression model applied to glycosylation data. The x-axis represents the composite input features influencing the stereoselectivity (e.g., *actual or observed* α-ratio), while the y-axis represents the experimental output [*predicted* α-ratio (%)], our target value. Each grey point is an individual glycosylation observation. The red line denotes the fitted linear regression model that minimizes squared error between the predicted and actual α-ratio values. Horizontal black bars reflect the conditional mean of the α-ratio within specified feature intervals.
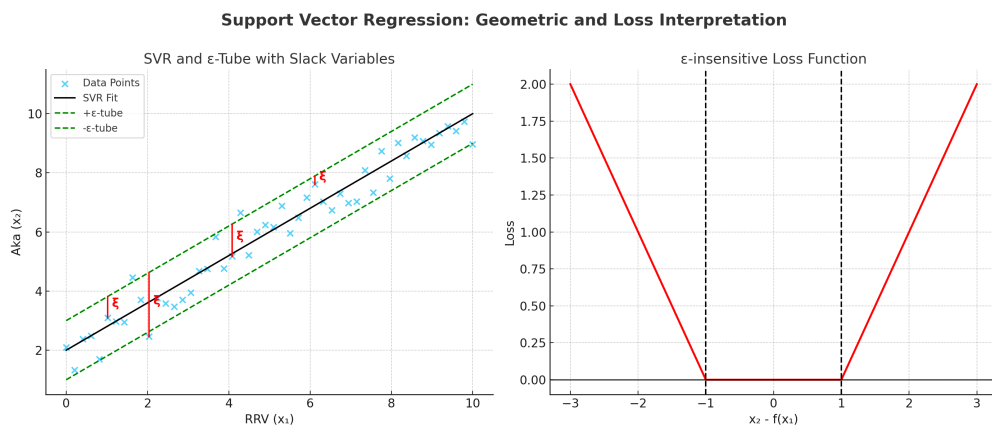


**Supplementary Figure 16. Support vector regression (SVR) and ε-insensitive loss.** Left: The schematic illustrates the SVR concept using a two-dimensional feature space, where the black line represents the regression function. The green boundaries indicate the $\varepsilon$-insensitive region (ε-tube), within which prediction errors are not penalized. Red stars denote observed data points, and those falling outside the ε-tube incur a penalty quantified by slack variables. These slack variables reflect the magnitude by which predictions violate the $\varepsilon$-bound and contribute to the total model loss. SVR seeks to minimize the norm of the weight vector while maintaining prediction errors within

the $\varepsilon$-tube whenever possible, thus achieving a balance between model complexity and prediction accuracy. Right: The $\varepsilon$ -insensitive loss function. The horizontal axis represents the deviation of predicted values from actual values, while the vertical axis denotes loss. The loss remains zero for deviations within the ε-margin (between $-\varepsilon$ and $+\varepsilon$) and increases linearly beyond that range, as defined by the absolute error beyond $\varepsilon$. This loss structure enables SVR to be robust to small fluctuations and noise, which is particularly relevant in chemical datasets where minor variations in experimental conditions may not significantly impact the mechanistic understanding but could otherwise distort a sensitive model.



Decision Tree Regressor for α-Ratio Prediction
(Features: RRV, Aka)
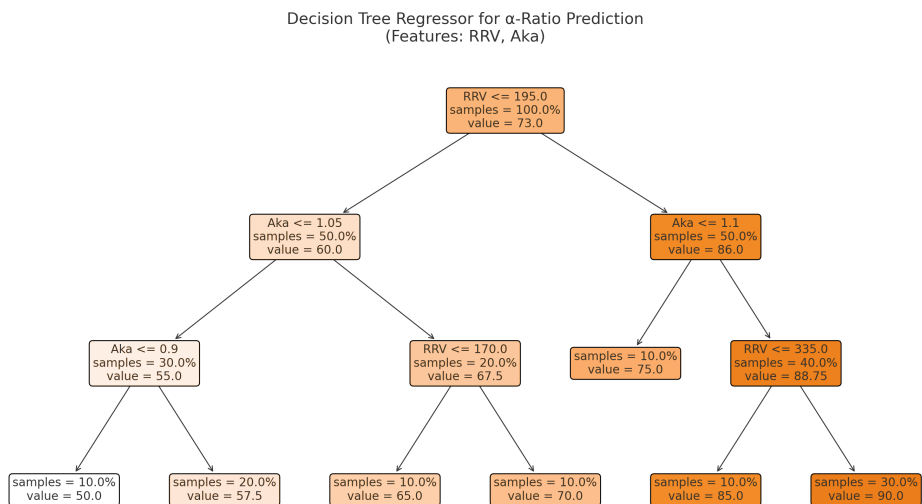
**Supplementary Figure 17. A Decision Tree regression model.** This was constructed to predict α-ratio outcomes from glycosylation reactions based on key reactivity parameters: the Relative Reactivity Value (RRV) of the glycosylating agent and the nucleophilicity constant (Aka) of the alcohol. The Decision Tree (DT) algorithm partitions the data by applying hierarchical, threshold-based splits (e.g., RRV < 2656, Aka < 1.36), resulting in a tree structure where each terminal node corresponds to a predicted α-ratio. This model provides a transparent and interpretable mapping between input descriptors and predicted stereoselectivity. The tree captures non-linear dependencies and interaction effects without requiring prior assumptions on variable distributions, reflecting the context-specific reactivity landscape of glycosylation. While useful for mechanistic interpretation and rational design, the DT model's predictive capacity may be limited by overfitting, especially in sparse or noisy datasets; ensemble models such as Random Forests are often employed to mitigate this.
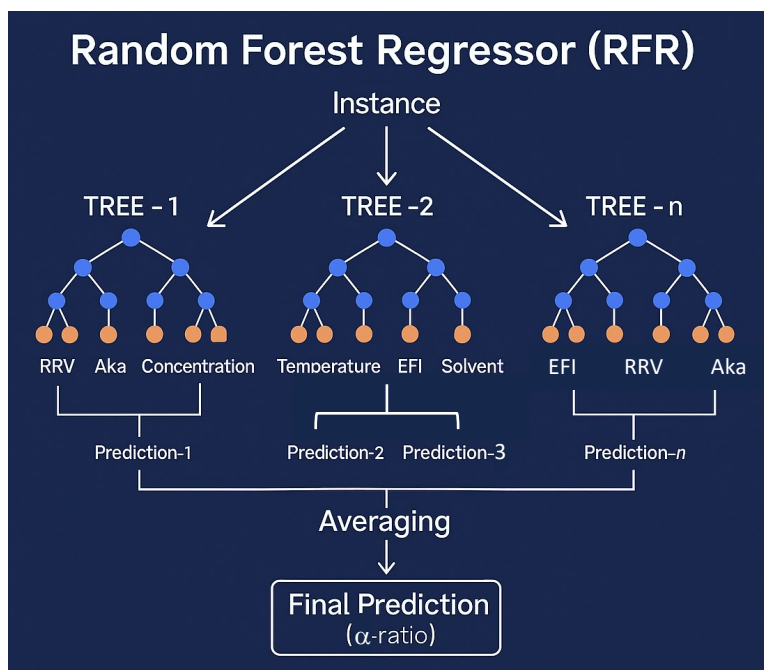
**Supplementary Figure 18. Random Forest Regressor (RFR) architecture.** An ensemble of decision trees (Tree$_1$, Tree$_2$….Tree$_n$) is shown, where each tree receives the same input instance characterized by features such as RRV (relative reactivity value), Aka (acceptor nucleophilicity constant), concentration, temperature, solvent, and EFI (Environmental Factor Impact index). These features serve as decision nodes throughout the trees. Each tree independently produces a prediction (Prediction-1 to Prediction-$n$) based on its structure and selected features. The final output is obtained by averaging all individual predictions to generate a robust and accurate $\alpha$-ratio prediction. This ensemble approach enables the model to capture non-linear and context-dependent interactions among variables while mitigating the risk of overfitting.

## 2.8 Model performance evaluation on train and validation sets for target prediction



Train: Actual vs Predicted α-ratio for All Models
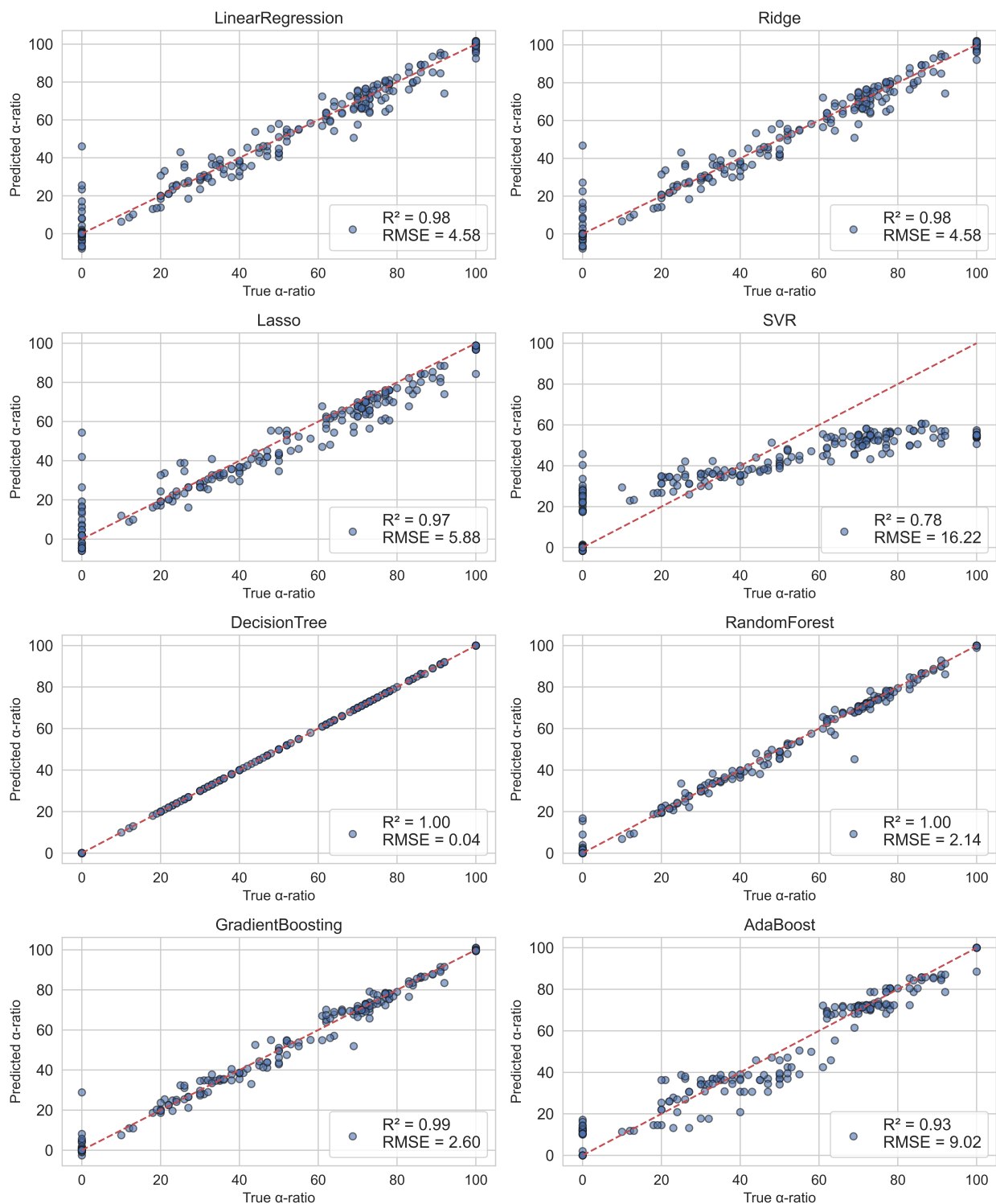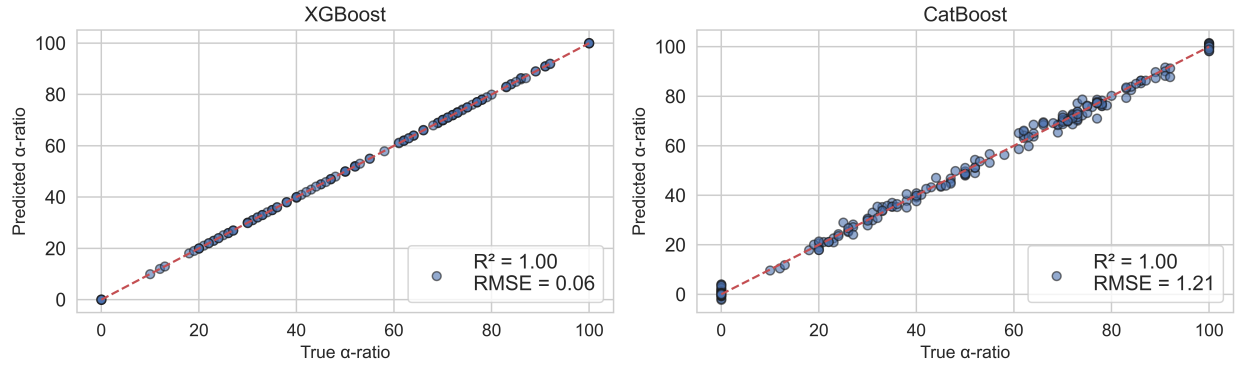
**Supplementary Figure 19. Training set evaluation across ten machine learning models for α-ratio (%) prediction.** Actual versus predicted α-ratio values for the training dataset using ten different regression models: Linear Regression, Ridge, Lasso, SVR, Decision Tree, Random Forest (RF), Gradient Boosting, AdaBoost, XGBoost, and CatBoost. The diagonal dashed red line indicates the ideal 1:1 relationship. Most models, particularly ensemble-based learners such as RF, XGBoost, and CatBoost, show strong alignment between predicted and observed values, with $R^2$ values exceeding 0.97 and low RMSEs. Decision Tree achieves a perfect score ($R^2 = 1.00$, RMSE = 0.04), suggesting that it has overfit the training data by modeling even minor noise and variance. Linear models (e.g., Ridge and Lasso) achieve reasonable fits, but SVR underperforms ($R^2 = 0.78$), highlighting its limited flexibility in capturing complex nonlinear dependencies. These results demonstrate that models capable of modeling feature interactions and nonlinearity can fit α-ratio effectively on training data, but overfitting must be evaluated on validation performance. So next we performed evaluation of models via validation.

Validation: Actual vs Predicted α-ratio for All Models
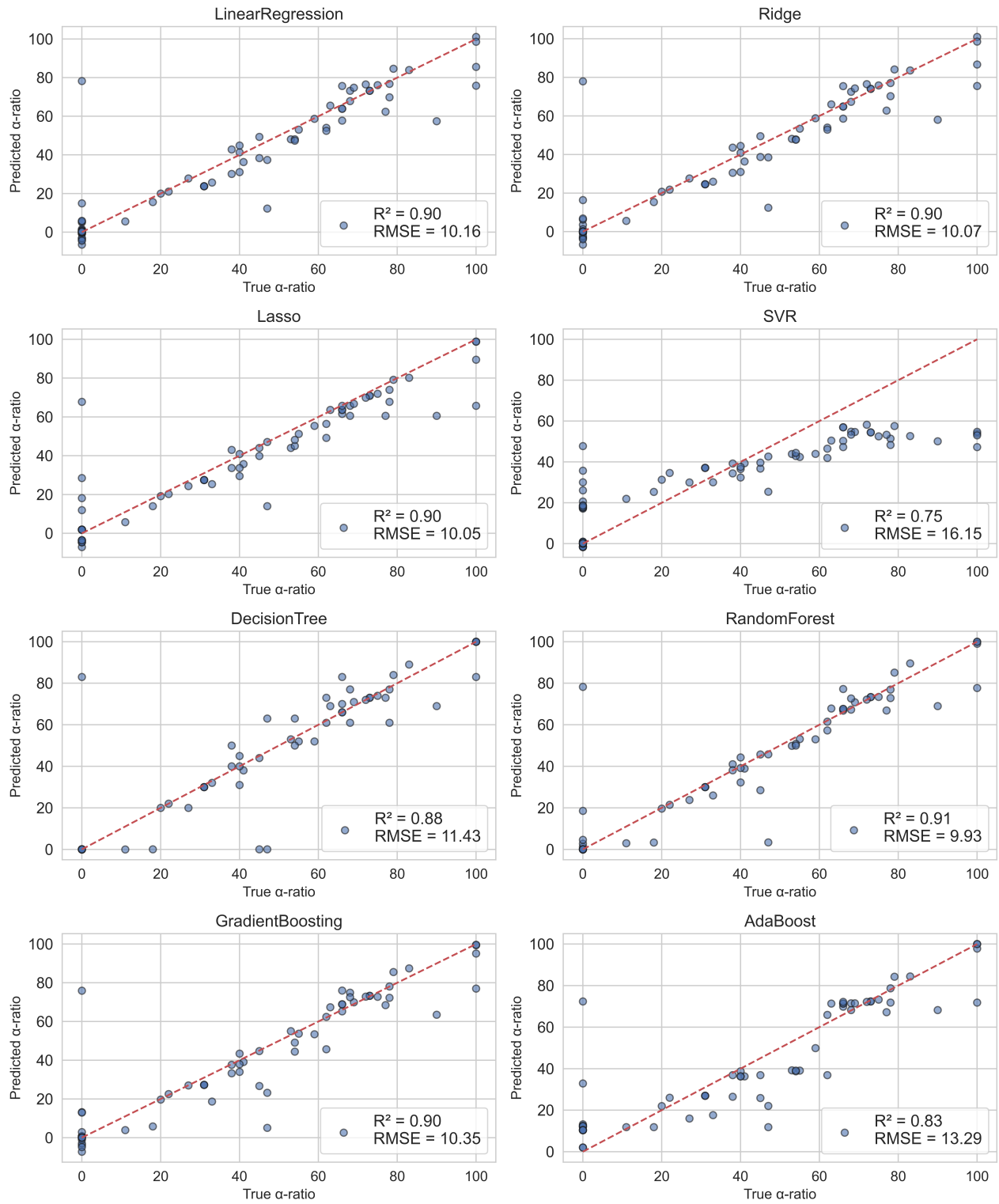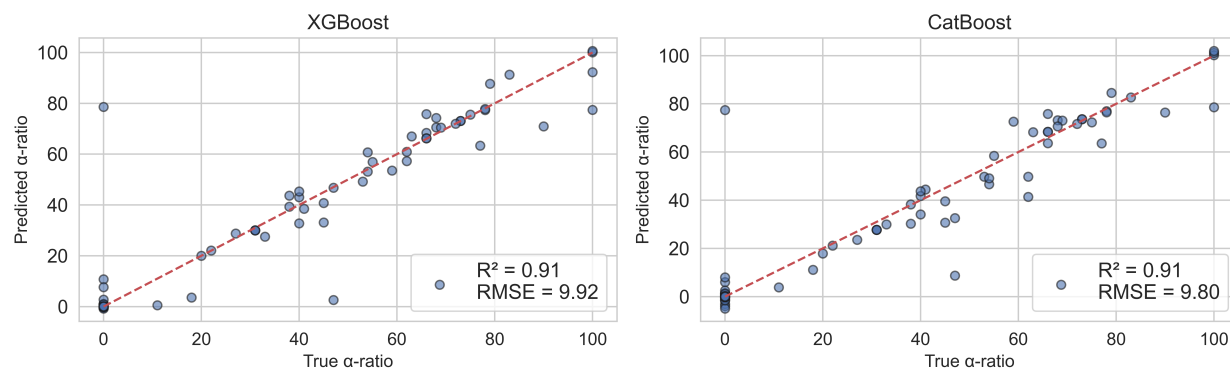
**Supplementary Figure 20. Validation set prediction across ten machine learning models for α-ratio (%) prediction.** Actual versus predicted α-ratio values for the validation set are plotted for the same ten regression models used in training. RF maintains strong predictive performance ($R^2$ = 0.91, RMSE = 9.93), suggesting effective generalization. In contrast, Decision Tree's performance drops significantly ($R^2$ = 0.88, RMSE = 11.43%), reflecting its poor extrapolation capability due to overfitting. Gradient boosting models (XGBoost, CatBoost) show stable performance with minimal degradation compared to their training results. Linear models (e.g., Ridge, Lasso) and SVR show substantial increases in prediction error and lower $R^2$ values, indicating reduced robustness when facing unseen data. These findings validate RF's ability to extract stable, generalizable patterns from the descriptor space, while also emphasizing the importance of model complexity control to avoid overfitting in mechanistically intricate prediction tasks like stereoselectivity.

Training: Actual vs Predicted Yield (y2)
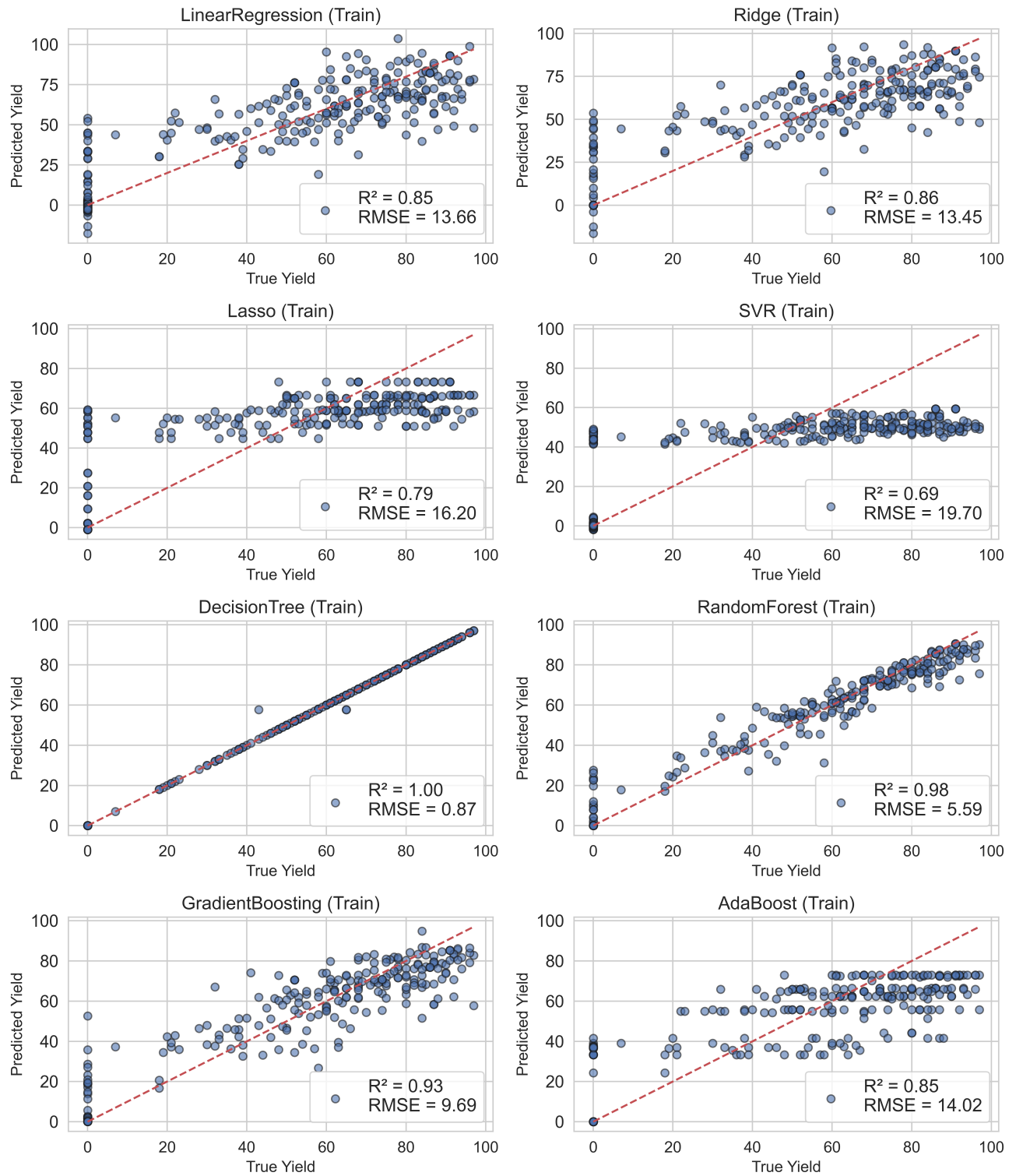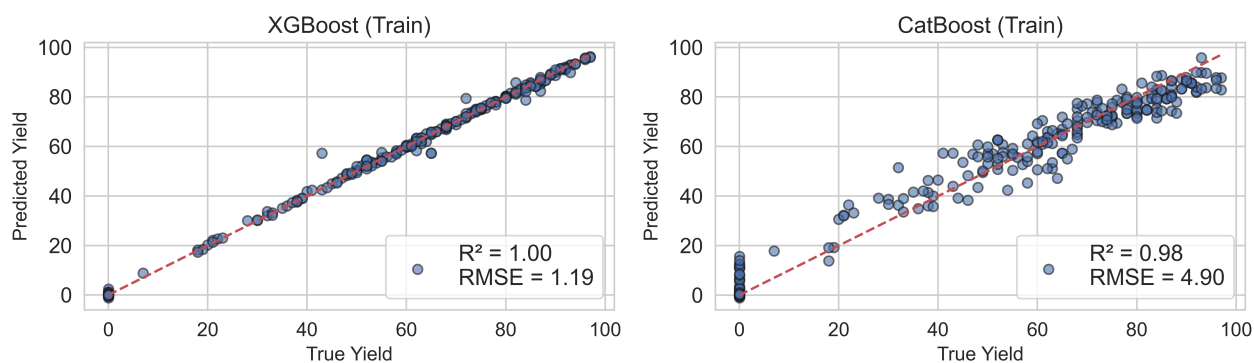
**Supplementary Figure 21. Training set evaluation across ten machine learning models for yield (%) prediction.** This shows training set performance for yield (y2) prediction using the same ten machine learning models. Yield, compared to $\alpha$-ratio, introduces greater noise and variance due to experimental uncertainties such as decomposition, solubility issues, and side reactions. Nevertheless, Decision Tree again achieves perfect accuracy ($R^2 = 1.00$, RMSE = 0.87), strongly suggesting overfitting. RF achieves high accuracy ($R^2 = 0.98$, RMSE = 5.59) while avoiding overfitting, striking a balance between complexity and generalizability. Gradient boosting models perform comparably well. However, models such as SVR, Ridge, and Lasso show increased RMSEs and slightly lower $R^2$ values, underlining their limited capacity to model non-additive relationships and accommodate variance introduced by reaction conditions. These trends suggest that yield prediction is more challenging than selectivity modeling, requiring methods capable of capturing high-dimensional, nonlinear relationships without overfitting.

Validation: Actual vs Predicted Yield (y2)
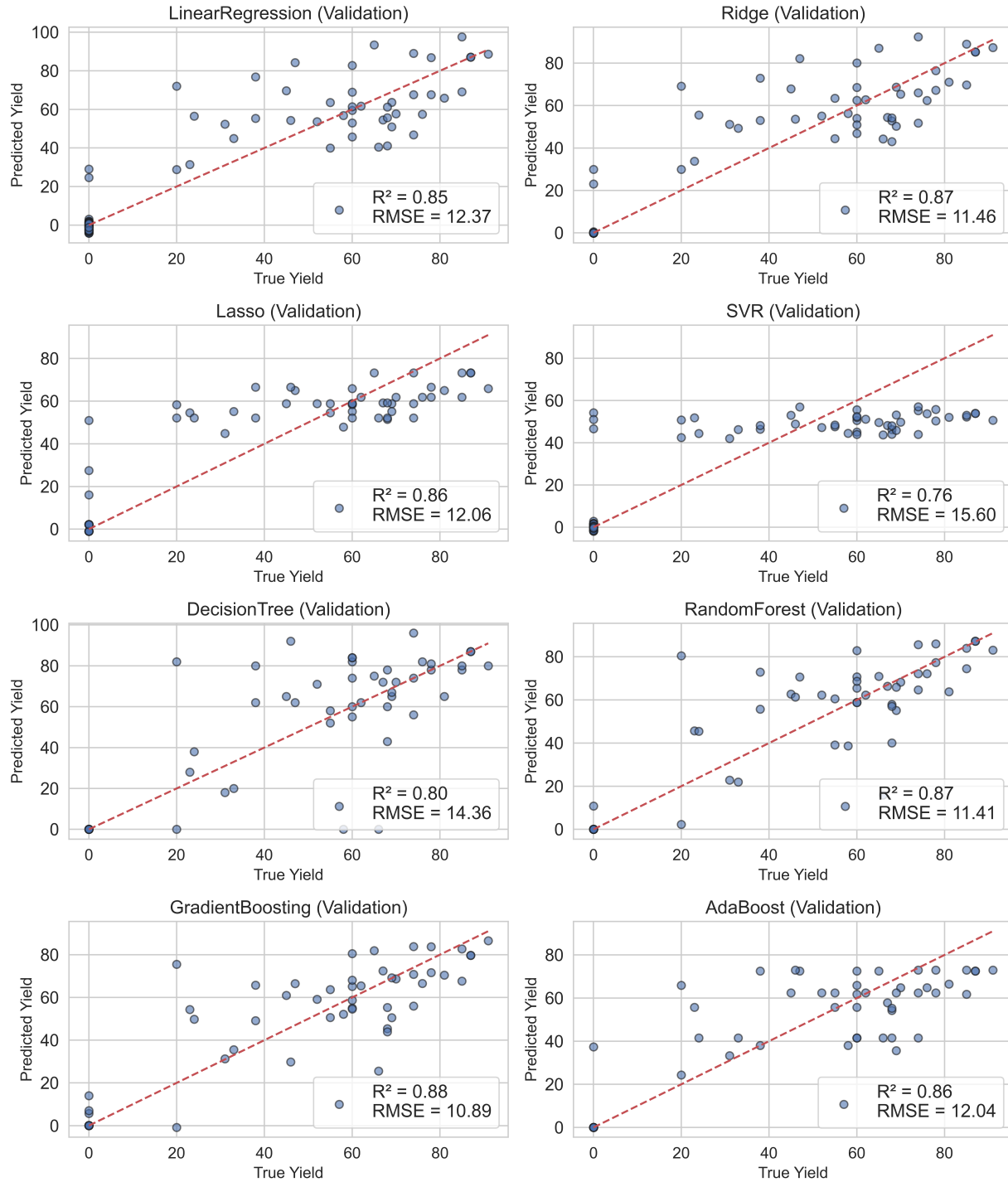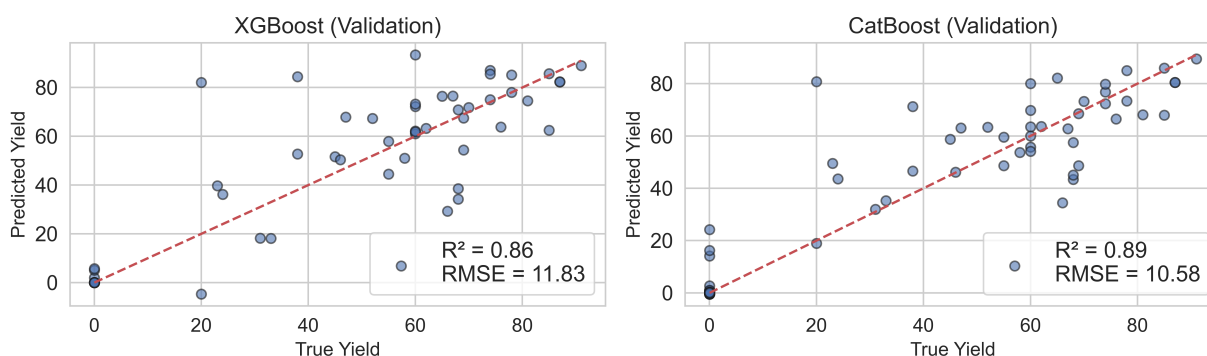
**Supplementary Figure 22. Validation set evaluation across ten machine learning models for yield (%) prediction.** This figure shows the generalization performance of all ten models on the yield (y2) validation set. RF continues to outperform all others ($R^2$ = 0.87, RMSE = 11.41), confirming its robustness in the presence of noisy chemical data. The performance of Decision Tree degrades sharply ($R^2$ = 0.80, RMSE = 14.36), reflecting classic overfitting behavior—relatively lower training error but poor generalization. Boosting methods (e.g., CatBoost and XGBoost) retain strong predictive capabilities, albeit with slightly higher variance. SVR, Ridge, and Linear Regression again underperform, particularly in the presence of yield-specific noise. These findings emphasize that while several ML models may fit training data well, only those that balance complexity with structural regularization (e.g., RF and boosting models) deliver reliable predictions under real-world variability. This underscores the importance of assessing model performance on both training and validation sets to ensure robustness, especially for noisy targets like reaction yield.

Distribution of Predicted α-selectivity
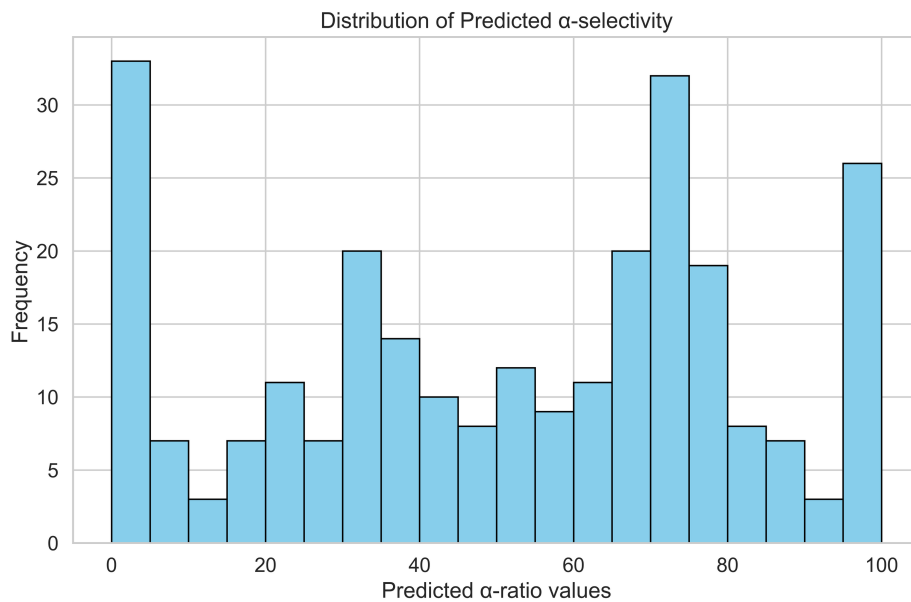
**Supplementary Figure 23. Distribution of predicted α-ratio (%) values for new glycosylation reactions.** Histogram showing the frequency distribution of machine learning–predicted α-selectivity (α-ratio [%]) values for newly generated glycosylating agent–alcohol pairs, illustrating the diversity and range of model-predicted stereoselectivities across the expanded chemical space. The values shows coverage across the diverse range and exhibits a bimodal pattern consistent with mechanistic expectations, while yield predictions are skewed toward higher conversion regions.



Distribution of Predicted Yield (%)
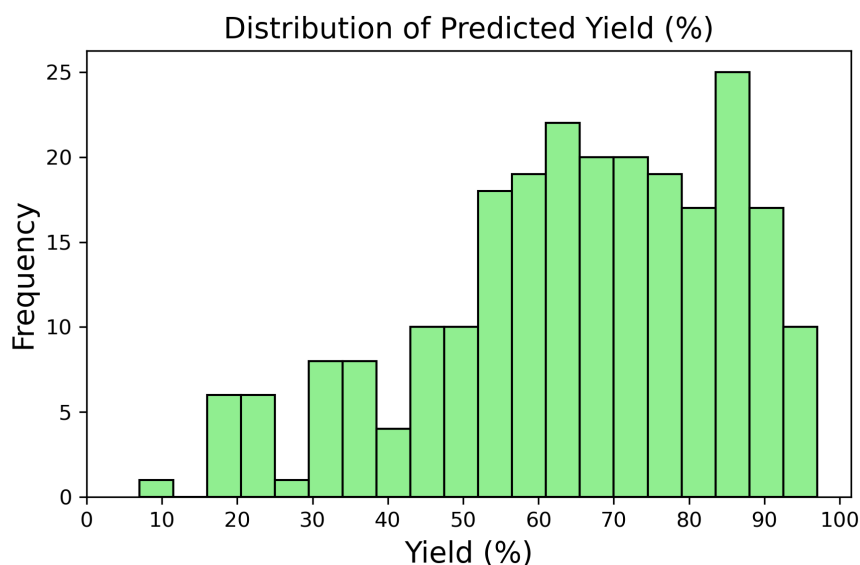
**Supplementary Figure 24. Distribution of predicted yield (%) values for new glycosylation reactions.** This distribution highlights the model's predictions over an expanded chemical space, with most reactions yielding between 50–90%. The broad spread reflects the complexity and variability of reaction efficiency due to subtle interactions among structural features and contextual parameters (e.g., solubility, side-product formation). In contrast to α-selectivity, yield predictions

display greater dispersion, underscoring the importance of incorporating reaction conditions for robust modeling.
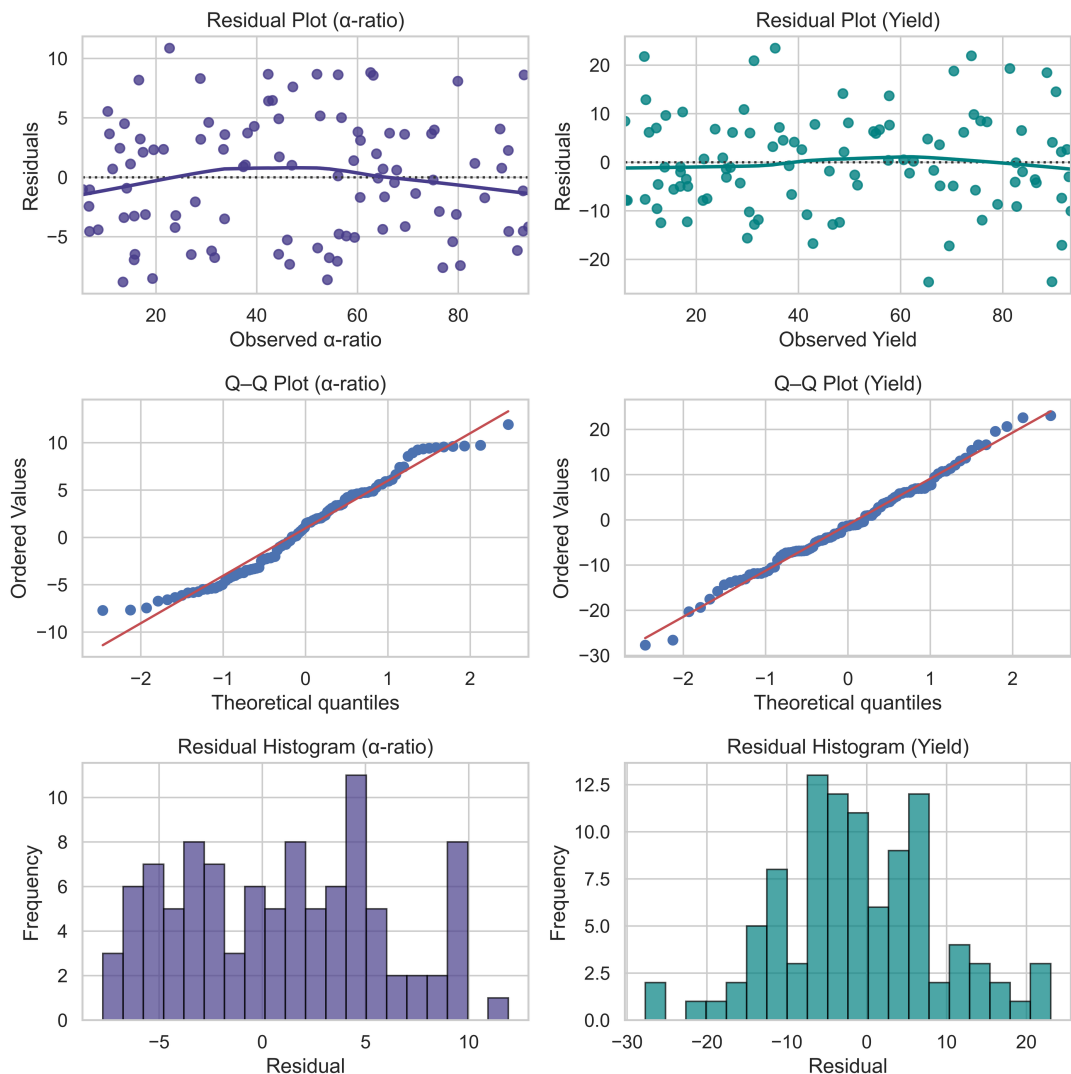


**Supplementary Figure 25. Model residual diagnostics demonstrating generalization capacity.**
To evaluate the generalization capacity of our trained models, we analyzed residual diagnostics for both $\alpha$-ratio and yield on the validation set. The residual plots show no clear heteroscedastic pattern or structural deviation, suggesting that residuals are homoscedastic and randomly distributed — a desirable outcome, particularly for non-linear, heterogeneous datasets such as ours. The Q-Q plots illustrate approximate normality in residual distributions, especially for $\alpha$-ratio, while histograms of residuals appear symmetric and unimodal. This indicates that the trained models are not substantially biased and maintain consistent error behavior across the prediction range. Although the yield model exhibits slightly lower out-of-sample, the residual analysis suggests that this is not due to systematic model bias or variance heterogeneity, but more likely reflects the intrinsic mechanistic complexity and variability in yield observations. Given that yield is highly influenced by both structural and environmental factors, some predictive variance is

expected. Overall, these diagnostics provide strong support for the validity and robustness of our models in capturing the mechanistic structure–function relationships inherent in the dataset.
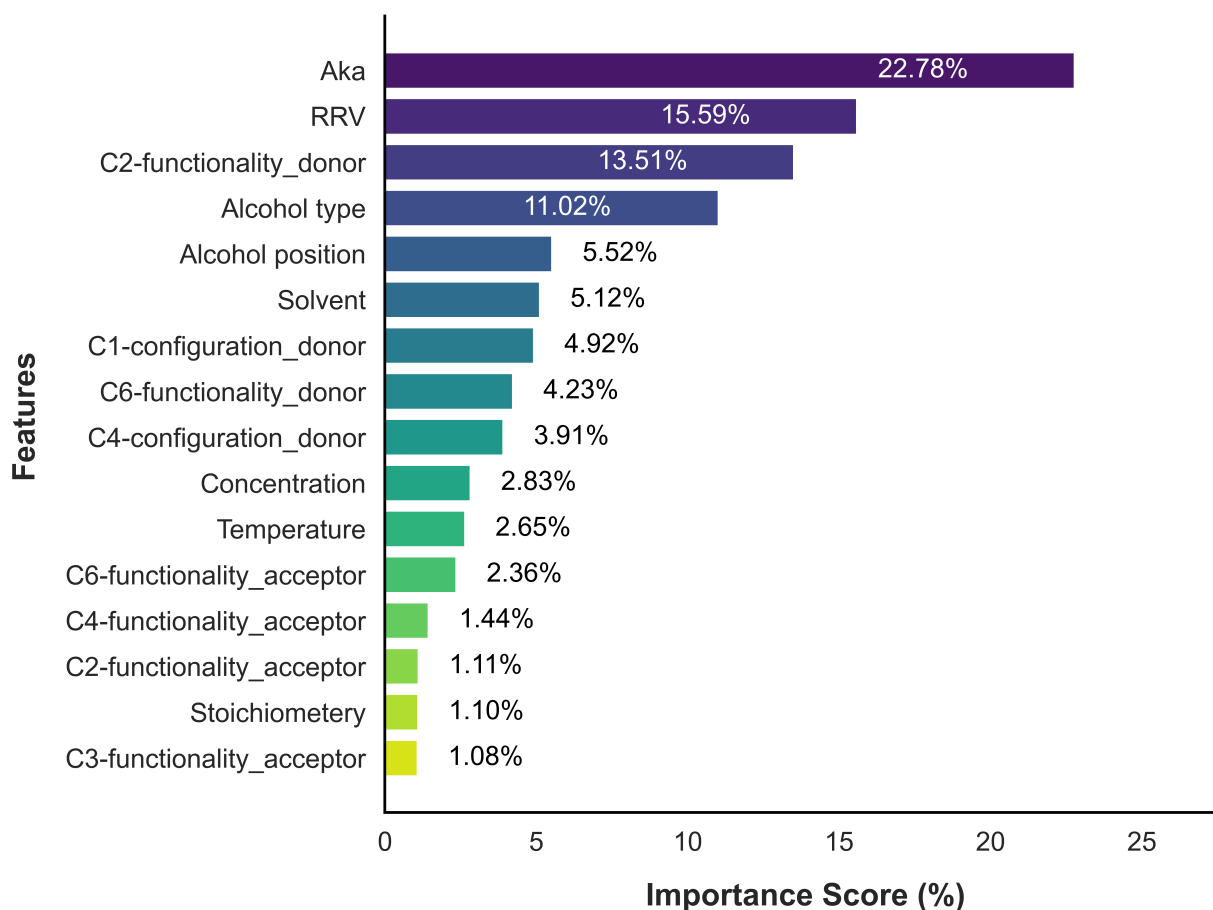


**Supplementary Figure 26. Bar plot ranking the relative contributions of key features to stereoselectivity.** The plot illustrates the relative contributions of molecular and contextual descriptors to the model's prediction for α-selectivity in glycosylation reactions. Acceptor nucleophilicity constant (Aka), glycosylating agent's reactivity (RRV), and agents C2-substitution were the most influential features, followed by alcohol type and EFI, underscoring the critical interplay of electronic, structural, and environmental factors in shaping stereoselectivity. Donor, glycosylating agent; acceptor, alcohol.
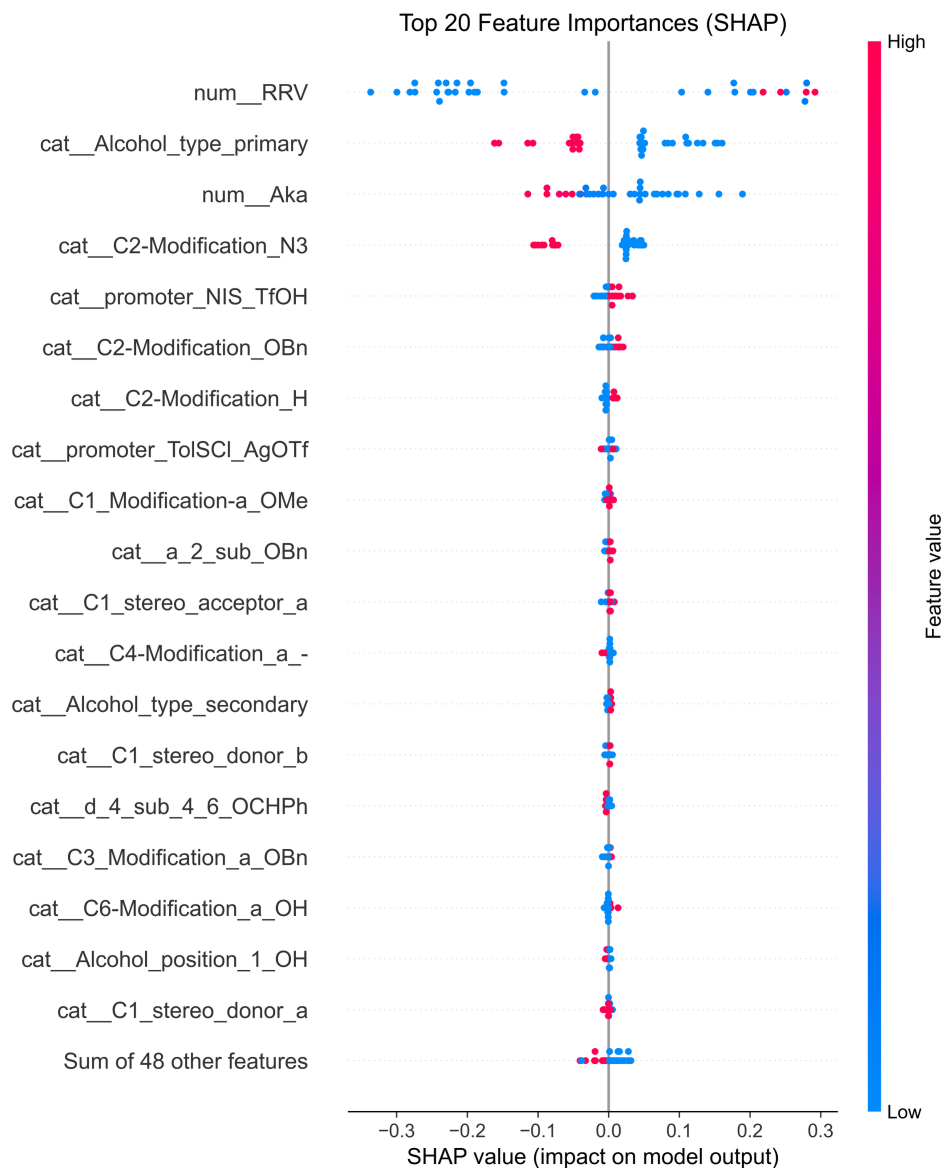
**Figure 27. SHAP summary plot derived from the trained hybrid model under *fixed reaction conditions*.** Illustrating how individual feature values affect predicted stereoselectivity across the nonlinear dataset. Red denotes high feature values, blue denotes low. RRV shows dominancy due to inductive effect.

**Supplementary Table 8. Diebold-Mariano (DM) test results comparing predictive accuracy of machine learning models to Random Forest for $\alpha$-ratio (y1).**

| Model | DM Statistic | p-value | Significance |
|---|---|---|---|
| Gradient Boosting | 2.845 | $4.10 \times 10^{-3}$ | ** |
| AdaBoost | 5.473 | $4.50 \times 10^{-8}$ | *** |
| Decision Tree | 5.915 | $2.80 \times 10^{-9}$ | *** |
| Lasso Regression | 6.198 | $5.70 \times 10^{-10}$ | *** |
| Ridge Regression | 6.252 | $3.95 \times 10^{-10}$ | *** |
| Linear Regression | 6.318 | $2.60 \times 10^{-10}$ | *** |
| CatBoost | 6.629 | $3.40 \times 10^{-11}$ | *** |
| Support Vector Regressor | 7.064 | $1.65 \times 10^{-12}$ | *** |
| XGBoost | 3.025 | $2.95 \times 10^{-3}$ | ** |

A positive and significant DM statistic indicates the respective model performs worse than Random Forest in predicting $\alpha$-ratio. All comparisons are statistically significant, reaffirming the robustness of Random Forest as the top-performing model for this task followed by XGB.

**Supplementary Table 9. Diebold-Mariano (DM) test results for comparing yield (y2) predictive performance of various machine learning models against the Random Forest baseline.**

| Model | DM Statistic | p-value | Significance |
|---|---|---|---|
| Gradient Boosting | 2.765 | $5.25 \times 10^{-3}$ | ** |
| AdaBoost | 5.162 | $8.10 \times 10^{-8}$ | *** |
| Decision Tree | 5.804 | $3.15 \times 10^{-9}$ | *** |
| Lasso Regression | 6.104 | $6.05 \times 10^{-10}$ | *** |
| Ridge Regression | 6.189 | $4.15 \times 10^{-10}$ | *** |
| Linear Regression | 6.278 | $2.95 \times 10^{-10}$ | *** |
| CatBoost | 6.574 | $3.70 \times 10^{-11}$ | *** |
| Support Vector Regressor | 7.012 | $1.75 \times 10^{-12}$ | *** |
| XGBoost | 2.941 | $3.20 \times 10^{-3}$ | ** |
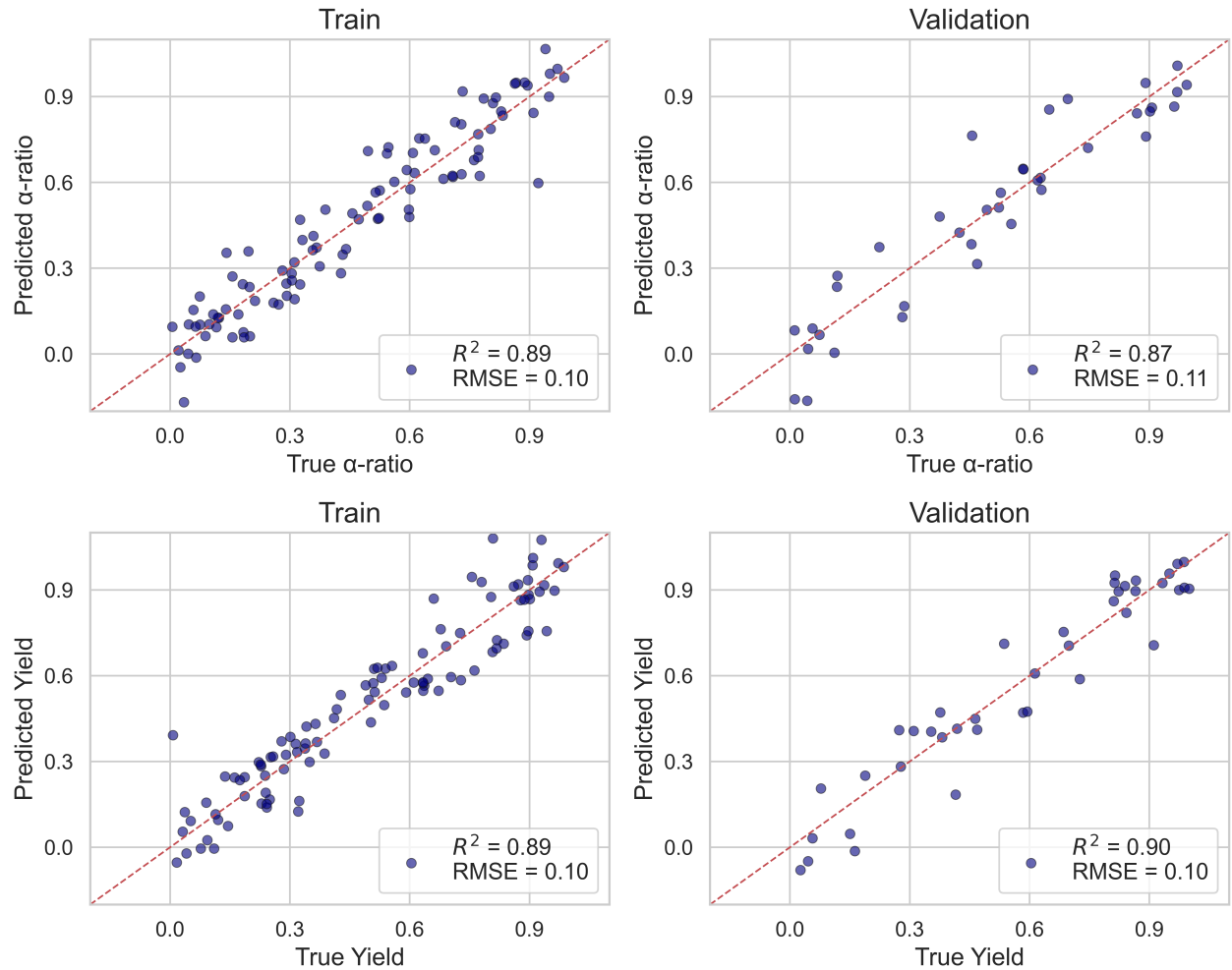
**Supplementary Figure 28. Scatter plots of predicted vs. true values for both α-ratio (y1) and yield (y2) across training and validation sets using the *trained* generalized Random Forest model.** Each subplot includes a diagonal reference line (ideal prediction) and performance metrics (R² and RMSE) shown in the lower-right corner.

## 2.9 Empirical analysis of ML models

**Regression Models**

The first two regression methods we consider are ridge regression and lasso regression, both of which aim to "shrink" or minimize the magnitude of the model's predictor coefficients, while retaining all features. These techniques help prevent overfitting by reducing the influence of less important predictor variables. Ridge regression shrinks coefficients towards zero, whereas lasso regression can set some coefficients exactly to zero, effectively eliminating certain variables from the model. By reducing model variance, these methods improve both accuracy and interpretability. To explain ridge and lasso regression, we start with the baseline model of linear regression, where the conditional expectation of the response variable is approximated as a linear function of the predictors. Ridge and lasso extend this by adding regularization (penalty) terms to control the size of the coefficients, thereby enhancing model stability and preventing overfitting. The baseline linear model is given as:

$$g(\mathbf{y}_{t-1}; \Omega) = \mathbf{y}_t - 1'\Omega \tag{S3}$$

The traditional method for estimating parameters in linear regression is Ordinary Least Squares (OLS). The objective of OLS is to minimize the sum of squared residuals, which represent the differences between the observed and predicted values. This is achieved by minimizing the following objective function:

$$L(\Omega) = \frac{1}{T} \sum_{t=1}^{T} \left( \omega_t - g(\mathbf{y}_{t-1}; \Omega) \right)^2 \tag{S4}$$

Minimizing L ($\Omega$) produces the Ordinary Least Squares (OLS) estimator However, while OLS can fit data well, it may lead to overfitting when dealing with high-dimensional data or multicollinearity. This modification leads to the ridge and lasso regression objective functions. For Ridge function the penalty term, that is the proportional to the sum of square coefficient. This is where ridge and lasso regression come in, adding penalty terms to the OLS objective function to control the size of the coefficients and improve the generalizability of the model.

$$L((\Omega) = \sum_{i=1}^{t} (\omega_i - g_i\beta)^2 + \lambda \sum_{j=1}^{n} |\beta_{j^2}|, \tag{S5}$$

For Lasso function the penalty term which will be the proportional to the sum of absolute value of the coefficient which represents below.

$$L((\Omega) = \sum_{i=1}^{t} (\omega_i - g_i\beta)^2 + \lambda \sum_{j=1}^{n} |\beta_j|, \tag{S6}$$

where, $\lambda$ is the tunning parameters which control the strength of penalty term. And allow to the balance between fitting the data and keep the model coefficient to small and reduce to zero. The modified objective function is expressed as:

$$L(\Omega; \lambda) = L(\Omega) + \phi(\Omega; \lambda) \tag{S7}$$

where the penalty term defines as below.

$$\phi(\Omega; \lambda) = \lambda \|\Omega\|_p \tag{S8}$$

where, $\lambda\|\Omega\|_p$ denotes the $\ell_p$ norm of vector $\Omega$ defined as:

$$\|\Omega\|_p = \sum_{t=1}^{T} |\Omega_t|^p \tag{S9}$$

For lasso, p=1 which results in L1 norm penalty, and for ridge regression, p=2 represents to the L2 norm penalty. While $\lambda$ represents a positive tuning parameter that determine the influence of the penalty term. In the case if $\lambda$=0 the penalty has no impact, as $\lambda \to \infty$ then the estimated coefficient shrinkage to zero.

## Ensemble Models

Ensemble methods, including Decision Trees (DT), Random Forests (RF), Gradient Boosting, and XGBoost Regressor, combine multiple learners to improve regression performance. Random Forests enhance bagging by selecting a random subset of size ($m = \sqrt{p}$), where (p) is the total feature count) at each split, reducing tree correlation and overfitting. Gradient Boosting sequentially builds models, each correcting prior errors, enhancing accuracy and minimizing bias. Predictions are aggregated across (B) trees, typically via averaging for regression, with each tree's contribution weighted to optimize the final model.

$$\hat{g}(y_{t-1}; d, B) = \frac{1}{B} \sum_{b=1}^{B} \hat{g}_b(y_{t-1}; \hat{\Omega}_b, d) \qquad (S10)$$

Where, '$d$' is the total number of splits in each tree which is controlling the complexity of the model. The final model aggregates all the trees to make the prediction, with each tree contributing to the overall prediction as follows:

$$\hat{g}_b(y_{t-1}; B, \vartheta, d) = \sum_{b=1}^{B} \vartheta \hat{f}_b(y_{t-1}) \qquad (S11)$$

Where (b-1) represents the (bth) tree, and ($\varrho$) controls the contribution of each tree to the final model. By focusing on the errors of previous models, gradient boosting progressively enhances prediction accuracy and reduces bias.

Additive combination of weak learners $h_t$, fit sequentially with reweighted data:

$$F_T(x) = \sum_{t=1}^{T} \alpha_t h_t(x) ,$$

where $\alpha_t$ is chosen to minimize the weighted squared error at step $t$. $F_T(x)$ represents ensemble prediction after $t$ boosting iteration. $h_t(x)$ is weak learner fitted at iteration.

## Random Forest Regressor (RFR)

Random Forest (RF) used for predictions of multiple trees trained on the bootstrapped subsets, which reduce the variance and fixed the overfitting, as a result it provides a robustness against the noise in the data. RF combines predictions from multiple decision trees to improve prediction accuracy and control overfitting it shown as by following equation.

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} p_i(x), \qquad (S12)$$

where,

$\bar{y}$ is the forecast value denotes the total number of trees while $p_i(x)$ represents the prediction from the $i^{th}$ decision tree. The study used 100 estimators that is 100 decision trees in the forest and 42 random state which are fixed for reproducibility.

## Gradient Boosting Regressor

Gradient Boosting sequentially fits weak learners to minimize prediction error by focusing on correcting previous errors, resulting in improved the accuracy and reduced bias. It iteratively builds an additive model by sequent fitting trees to residuals are:

$$\bar{y} = F_0(x), \qquad (S13)$$

$$r_i^{(t)} = y_i - F_{t-1}(x_i), \tag{S14}$$

$$h_t \approx \arg \min_h \Sigma_i \left( r_i^{(t)} - h(x_i) \right)^2, \tag{S15}$$

$$F_t(x) = F_{t-1}(x) + v h_t(x), \tag{S16}$$

where,

The model starts with a constant prediction which is equal to the mean of all the target values $\bar{y}$. $r_i^{(t)}$ is the difference between the actual value $y_i$ and the prediction from the previous iteration which is $F_{t-1}(x_i)$. The each itration t, a new base learner $h_t(x)$ is trained to predict the residuals $r_i^{(t)}$ which is a shallow decision tree. $F_t(x)$ adds the new weak learner $h_t(x)$ to ensemble with a scalling factor $v$ (learning rate). $v \in (0,1)$ controls how much influence the new learner has.

## XGBoost Regressor

The objective function of XGBoost consists of two segments. The training loss and a regularization term. The objective function as following:

$$Obj(\alpha) = \Sigma_{i=1}^n l(y_i, \hat{y}_i) + \Sigma_{k=1}^T \left( \gamma T_k + \frac{1}{2} \lambda \Sigma_{j=1}^{T_k} \omega_{kj}^2 \right), \tag{S17}$$

where,

$\hat{y}_i = \Sigma_k fk(x_i)$ is the predicted value for $i$, which is obtained by summing the outputs of all k trees. $l(y_i, \hat{y}_i)$ is the loss function that measuring the difference between the actual and predicted values. $T_k$ is the number of leaves in the $K^{th}$ tree. $W_{kj}$ denotes the weight for $J^{rh}$ leaf in the $K^{th}$ tree. $\gamma$ is the regularization parameter penalizing the complexity of each tree $\lambda$ is the L2 regularization term on the leaf weights. While $\alpha$ represents the total objective function, combining both the training loss and regularization penalties.

The first segment objective function is training loss and the second function is regularization term. The first Segment, $\Sigma l(y_i, \hat{y}_i)$, measures the discrepancy between the actual and predicted values, ensuring that the model fits the data accurately. The second module, $\Sigma_{k=1}^T \left( \gamma T_k + \frac{1}{2} \lambda \Sigma_{j=1}^{T_k} \omega_{kj}^2 \right)$, penalizes the complexity of the model by adding a cost for each additional leaf in the trees and for large leaf weights. This regularization helps to prevent overfitting by favoring simpler, more generalizable models. By iteratively minimizing this combined objective, XGBoost constructs an ensemble of decision trees that progressively refine the predictive performance while maintaining control over model complexity.

## CatBoost Regressor

CatBoost efficiently handles categorical features and leverages ordered target-based encoding, significantly reducing prediction bias and improving generalization.

$$L(F) = \Sigma_{i=1}^n (y_i - F(X_i)^2) + \lambda \Sigma_{j=1}^j \|w_j\|^2, \tag{S18}$$

where,

L(F) is the loss function including a regularization component. $\lambda$ controls regularization function. while $w_j$ denotes the tree parameters.

**Support Vector Regressor (SVR)**

SVR aims to fit the best hyperplane within an acceptable margin of error with support vectors controlling the model's complexity and generalization. Assuming normalized input features represented by the vector $x_i$, and corresponding photovoltaic power outputs denoted by $y_i$ for each data point $i$ in the dataset, the entire dataset can be expressed as a collection of pairs $\{(x_i, y_i)\}^n i = 1$ where N is the total number of data point. Support Vector Regression (SVR) aims to approximate the underlying function using the expression: It solves the following optimization problem:

$$\frac{1}{2}\|w_j\|^2 + C \sum_{i=1}^{n}(\eta_i - \eta_i^*), \tag{S19}$$

Subject to: $y_i - \left(W.\varphi(x_i)\right) + b) \leq \in + \eta_i, \left(W.\varphi(x_i)\right) + b) - y_i \leq \in +\eta_i^*,$ $\qquad$ (S20)

where,

$\in$ denotes tolerance margin C is the penalty for error while $\varphi(x_i)$ is the kernel transformation.

**Decision Tree Regressor**

A decision tree regressor models the output $\hat{y}$ by spliting the inpout space into disjoint regions and assinging a constant value in each region. The prediction function is defined as:

$$\hat{y}(x) = \sum_{k=1}^{k} ck.\| (x \in N_k), \tag{S21}$$

where,

K represents the nomburs of leaf nodes, $N_k$ denotes the region k,which is the subset of input space. $ck$ denotes the constant prediction values in region $N_k$. $\| (x \in N_k)$ indicate the indicator function which is equal to 1 if $x \in N_k$, otherwise 0.

At each node, the model select the best feature K and threshold s that minimizes the sum of squared errors:

$$min_{k,s} \left[ min_{cL} \sum_{xi\in RL_{(K,S)}}(y_i - CL)^2 + min_{CR} \sum_{xi\in RR_{(K,S)}}(y_i - CR + min_{CR} \right], \tag{S22}$$

where,

$R_L(K,S) = \{x_i | x_{ik} \leq s\}$ and $R_R(K,S) = \{x_i | x_{ik} > s\}$ while $C_L$ and $C_R$ represents the predictions in the left and right regions.

For optimal value in each region the best constant prediction value for region $N_k$ is the mean of the target values in that region:

$$C_j = \frac{1}{|N_k|} \sum_{x_i \in R_j} y_i, \tag{S23}$$

where,

$|N_k|$ belongs to number of dta point in refion $N_k$.

After constructiong the tree and determining optimal region $N_k$ and $\hat{y}(x) = C_j$ where x $\in N_k$

Or the full indicator function represent as:

$$\hat{y}(x) = \sum_{k=1}^{k} \left( \frac{1}{|N_k|} \sum_{x_i \in R_j} y_i \right). \| (x \in N_k), \tag{S24}$$

## 2.10   Performance Metrics

**Mean Absolute Error (MAE)**
The MAE is the loss function for regression model, which is the sum of the absolute values of the difference between the target and predicted values. The MAE only measures the average modulus length of the error for the predicted value. The MAE function can be shown as:

$$MAE = \frac{1}{p}\sum_{i=1}^{p}|f_i - y_i| \qquad (S25)$$

where $p$ stands for the numbers of data, $f_i$ refers to the predictive value, and $y_i$ can be described as actual value.

**Root Means Square Error (RMSE)**
The RMSE is a statistical test used in regression tasks to determine the accuracy of a prediction model. In order to give greater error more weight. The lower RMSE values, the better the model performance. The test can be performed as follow:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \,. \qquad (S26)$$

where $\frac{1}{n}$ is the number of data point in the dataset, $y_i$ stand for the value of the dependent variable, while $\hat{y}_i$ refer to the predicted value of the dependent variable.

**The Diebold-Mariano (DM)**

The Diebold-Mariano (DM) test Diebold and Mariano, (1995) evaluates the predictive accuracy of ML models. This test identifies significant differences in model accuracy through quantitative analysis. Suppose that two predictions $f_1, ..., f_n$ and $g_1, ..., g_n$ for a time series are linked with $y_1, ..., y_n$. Let $e_i$ and $r_i$ be the residuals for the two forecasts, i.e.
The forecast residuals are defined as follows:

$$e_i = y_i - f_i, \quad r_i = y_i - g_i, \qquad (S27)$$

Forecast residuals are defined as follows:

$$d_i = e_i^2 - r_i^2 \quad \text{or} \quad d_i = |e_i| - |r_i|, \qquad (S28)$$

and let $d_i$ be defined as one of the following.
The time series is called the loss-differential. The key assumption for using the Diebold-Mariano test is that the loss differential time series $d_i$ is stationary. The first of these formulas is related to the MSE error statistic, and the second is related to the MAE error statistic. Now define Loss-differential mean as:

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \mu = \mathrm{E}[d_i], \qquad (S29)$$

For n > k ≥ 1, define:

$$r_k = \frac{1}{n}\sum_{i=k+1}^{n}(d_i - \overline{d})(d_{i-k} - \overline{d}), \qquad (S30)$$

where autocovariance is at lag $k$.
As described in Autocorrelation Function $r_k$ is the autocovariance at lag k.

$$DM = \frac{\bar{d}}{\sqrt{[r_0 + 2\sum_{k=1}^{h-1} r_k]/n}},$$ $$(S31)$$

For $h \geq 1$, we define the Diebold-Mariano (DM) statistic, where the value $h = n^{1/3} + 1$.

The DM test is based on a standard normal distribution. The null hypothesis indicates that an equal predictive capacity exists between the models. The alternative hypothesis regarding the higher predictability of the model has the lowest value of the loss function.

**Table 10. Performance metrics and model comparison for different regression models used for predicting α-ratio and yield.**

| Model | Mean Normalized \| RMSE (↓) | Mean Normalized \| Spearman (↑) | Mean Normalized \| R² (↑) | Mean Normalized \| MAE (↓) | Mean \|RMSE (↓) | Mean \| Spearman (↑) | Mean \| R² (↑) | Mean \| MAE (↓) | Wins \| RMSE (↓) | Wins \| Spearman (↑) | Wins \| R² (↑) | Wins \| MAE (↓) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hybrid Model | 0.024 | 0.986 | 0.973 | 0.019 | 2.408 | 0.986 | 0.973 | 1.923 | 1 | 1 | 1 | 1 |
| Random Forest | 0.03 | 0.978 | 0.959 | 0.024 | 3.002 | 0.978 | 0.959 | 2.378 | 0 | 0 | 0 | 0 |
| XGBoost | 0.035 | 0.969 | 0.943 | 0.028 | 3.527 | 0.969 | 0.943 | 2.812 | 0 | 0 | 0 | 0 |
| CatBoost | 0.045 | 0.952 | 0.91 | 0.035 | 4.45 | 0.952 | 0.91 | 3.538 | 0 | 0 | 0 | 0 |
| Gradient Boosting | 0.054 | 0.933 | 0.868 | 0.043 | 5.372 | 0.933 | 0.868 | 4.252 | 0 | 0 | 0 | 0 |
| AdaBoost | 0.059 | 0.919 | 0.841 | 0.047 | 5.894 | 0.919 | 0.841 | 4.74 | 0 | 0 | 0 | 0 |
| Ridge Regression | 0.059 | 0.918 | 0.839 | 0.048 | 5.939 | 0.918 | 0.839 | 4.751 | 0 | 0 | 0 | 0 |
| Support Vector Regressor | 0.061 | 0.918 | 0.832 | 0.048 | 6.068 | 0.918 | 0.832 | 4.784 | 0 | 0 | 0 | 0 |
| Decision Tree | 0.066 | 0.901 | 0.798 | 0.054 | 6.639 | 0.901 | 0.798 | 5.388 | 0 | 0 | 0 | 0 |
| Lasso Regression | 0.069 | 0.892 | 0.779 | 0.055 | 6.942 | 0.892 | 0.779 | 5.532 | 0 | 0 | 0 | 0 |

The table presents Mean Normalized and Mean values for key performance indicators: RMSE (↓), Spearman (↑), R² (↑), and MAE (↓), along with the Wins column for each metric.

## 2.11 Mechanistic mapping of reaction space
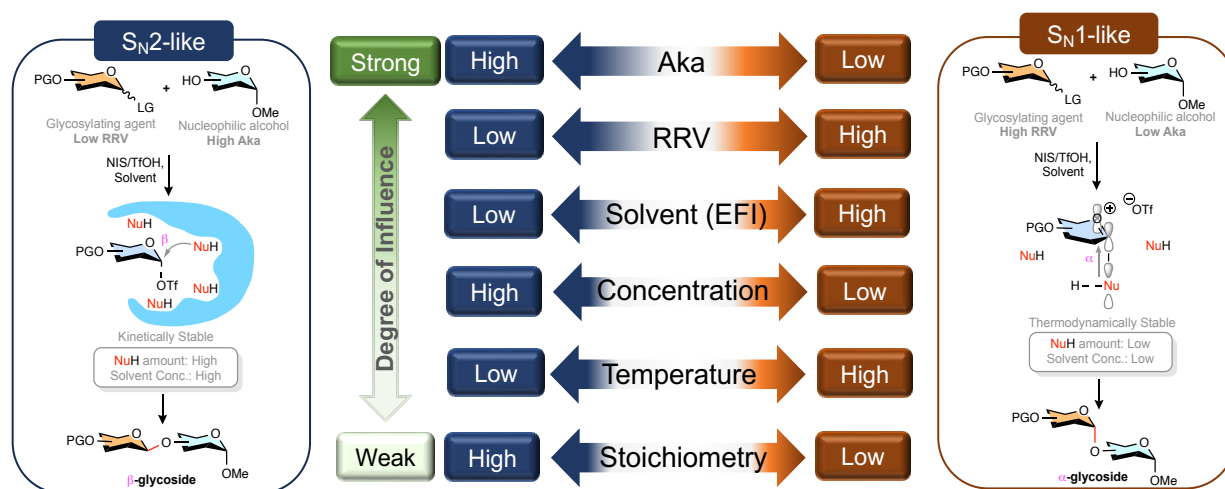


**Supplementary Figure 29. Contextual mapping of stereoselectivity pathways governed by glycosylating agent, alcohol, and environmental parameters.** This schematic summarizes the mechanistic drivers of $\alpha/\beta$-selectivity across the $S_N2$-like (blue) and $S_N1$-like (orange) pathways in glycosylation. On the left, $S_N2$-like reactivity is favored by low RRV donors and highly nucleophilic acceptors (high Aka), especially under high solvent concentration and nucleophile excess, stabilizing kinetically driven transition states. In contrast, $S_N1$-like reactivity on the right is facilitated by high RRV donors and poor nucleophiles (low Aka), with low solvent concentration and reduced nucleophile loading, enabling thermodynamically stable ion pair intermediates. The center gradient ranks the degree of influence of each variable on model predictions, highlighting Aka and RRV as primary drivers, while environmental factors such as EFI (solvent), concentration, and temperature provide secondary modulation. Stoichiometry exerts the weakest relative impact. These trends, identified by the model, were found to be consistent with experimental observations. Green scale shows from strongest (green) to weakest (light green) influence on glycosylation

## 2.12 Programming details

**Software and environment**
- o Anaconda Navigator v2.6.3 (includes Conda v24.9.2, Python 3.12.7)
- o **Jupyter Notebook** v7.2.2 (launched via Navigator)

- ▪ **Required Python packages:**

```
o  pandas

o  numpy

o  scikit-learn

o  pandas_datareader

o  openpyxl

o  matplotlib

o  seaborn
```

- ▪ **Creating and activating your Conda environment**

  1. **Open Anaconda Navigator**

     - o Click the "**Environments**" tab on the left.

     - o Click "**Create**", name it "glycoenv", select **Python 3.8**+.

  2. **Activate** it by selecting "glycoenv" from the list.

- ▪ **Installing Packages**

Open a terminal or Anaconda prompt

Install core data libraries;

```
e.g., Conda install pandas numpy scipy scikit-learn matplotlib
seaborn openpyxl
```

install needed models;

```
pip install optional package (e.g., catboost, xgboost, joblib)
```

55

- **Data Preparation**

Place the primary data file, "`data_ML_FIzZA.csv`", in the same folder as the notebooks. Missing data cleaned, special $y_1$, and $y_2$ values mapped to 0.01 and failed reactions mapped to 0.00- One-hot encoding for `x₁-x₂₄ – y₁/y₂` column converted to float.

- **Running the Analysis**

In Anaconda Navigator (with `glycoenv` active), select **Jupyter Notebook**, then click **Launch**.

- **Open and execute**

**`01_feature_importance.ipynb`**

Reads the data CSV, computes feature-importance scores, and saves

—`feature_importance_plot.png`

**`02_modeling_y1/y2.ipynb`**

Cleans the `y1/y2` target column, trains ten ML pipelines (Random Forest, Gradient Boosting, XGBoost, CatBoost, Support Vector Regressor, AdaBoost, Linear Regression, Ridge regression, Lasso Regression, Decision tree) on a 70/30 or 90/10 train/test split on case/*n*, and evaluates `MAE, RMSE, Pearson's r`, and $R^2$ With `sklearn.metrics`. Hyperparameter tuning with `GridSearchCV.`

—Saves

—`model_performance_results.csv/PNG`

Observed vs. predicted scatter plots and bar charts (displayed inline and exportable as PNGs).

- **Model Optimization**

Used `scipy.optimize.minimize` (L-BFGS-B)

Decoded categorical features with `LabelEncoder`

Displayed optimal settings with `Pandas`

Generated predictions and histograms

Used `matplotlib` for plots

Created download links via `IPython.display.HTML`

**Note:**

- o If you move data "`data_ML_FIZZA.csv`", update the "`file_path`" variable in the first cell of each notebook.
- o Any package-import errors can be resolved by re-running the `Conda install` or `pip install` commands above.

All datasets and code generated during this study are available from the corresponding author upon reasonable request.

# 3 Experimental Section for Synthesis

## 3.1 Synthesis of glycosylating agents

**_p_-Tolyl 2-azido-3,4,6-tri-_O_-benzyl-2-deoxy-β-_D_-thioglucopyranoside (A)**
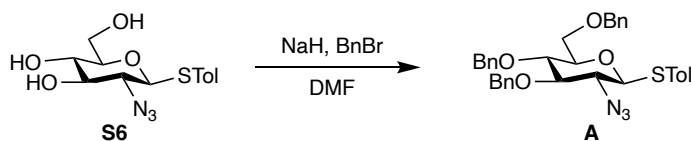


To a solution of compound **S6** (0.8 g, 2.60 mmol) in DMF was added NaH (0.4 g, 60% NaH dispersion in mineral oil, 10.42 mmol) and BnBr (1.6 mL, 13.02 mmol) at 0°C. The reaction mixture was stirred at room temperature for 4 h. The mixture was extracted with EtOAc twice. The organic layer was dried over anhydrous MgSO$_4$, filtered, and concentrated under reduced pressure. The concentrate was purified by silica gel chromatography with hexane/EtOAc (9:1) as eluent. The titled compound **A** was obtained as a colorless oil (57.2 mg, 81%, α:β = 29:71). The RRV of β-anomer referring to previous report is 270 *(13)*. [α]$^{24}$_D +54.0 (*c* 0.6, CHCl$_3$); IR (CHCl$_3$) *v* 2105, 1493, 1453, 1272, 1074, 809, 771, 733, 695 cm$^{-1}$; $^1$H NMR (400 MHz, CDCl$_3$) *δ* 7.51-7.05 (m, 38H, Ar-H), 5.54 (d, *J* = 5.3 Hz, 1H, H-1α), 4.93-4.42 (m, 12H, CH$_2$Ph), 4.39 (ddd, *J* = 10.0, 4.3, 2.1 Hz, 1H, H-5α), 4.35 (d, *J* = 10.1 Hz, 1H, H-1β), 3.93 (dd, *J* = 10.0, 5.3 Hz, 1H, H-2α), 3.82-3.72 (m, 4H, H-4α, H-6$_{ax}$α, H-6$_{ax}$β, H-6$_{eq}$β), 3.64 (dd, 1H, *J* = 10.5, 1.2 Hz, H-6$_{eq}$α), 3.58 (t, *J* = 10.1 hz, 1H, H-4β), 3.50 (t, *J* = 10.1 Hz, 1H, H-3β), 3.46 (ddd, *J* = 9.7, 3.5, 2.1 Hz, 1H, H-5β), 3.30 (t, *J* = 10.1 Hz, 1H, H-2β), 2.31 (s, 6H, ArCH$_3$); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 138.7 (C), 138.3 (C), 137.9 (C), 137.73 (C), 137.65 (C), 134.3 (CH), 132.7 (CH), 130.0 (CH), 129.8 (CH), 129.7 (CH), 128.48 (CH), 128.45 (CH), 128.4 (CH), 128.2 (CH), 128.1 (CH), 128.0 (CH), 127.9 (CH), 127.8 (CH), 127.7 (CH), 127.7 (CH), 127.6 (CH), 87.6 (CH), 85.9 (CH), 85.1 (CH), 81.8 (CH), 79.4 (CH), 78.4 (CH), 77.6 (CH), 75.9 (CH$_2$), 75.7 (CH$_2$), 75.1 (CH$_2$), 73.5 (CH$_2$), 71.7 (CH), 68.8 (CH$_2$), 68.4 (CH$_2$), 64.9 (CH), 64.2 (CH$_2$), 21.1 (CH$_3$); HRMS (ESI) calcd. for C$_{34}$H$_{35}$N$_3$O$_4$NaS [M + Na]$^+$ 604.2246, found 604.2360.

**_p_-Tolyl 2,3,4,6-tetra-_O_-benzyl-1-thio-β-_D_-glucopyranoside (B)**
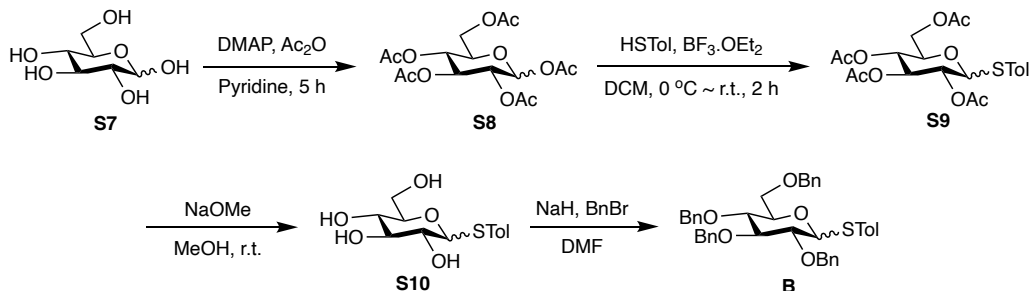


To a solution of compound **S7** (56 mmol, 10 g) in pyridine (60 mL), DMAP (5.4 mmol, 0.67 g) was added. The mixture was cooled to 0 °C, and acetic anhydride (Ac$_2$O) (358 mmol, 36 mL) was added dropwise while maintaining the temperature. The mixture was stirred at room

temperature under $N_2$ for 5 h and then quenched with ethanol (20 mL) at 0 °C. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (300 mL), and washed with water, saturated $NaHCO_3$, 10% aqueous hydrochloric acid, and brine. The organic phase was dried over $MgSO_4$ and then filtered and concentrated. Silica gel column chromatography (2:1 hexanes-/EtOAc) afforded 20.3 g (93%) of 1,2,3,4,6-penta-*O*-acetyl- D-glucopyranoside **S8** as white solid. $[\alpha]^{27}_D$ +9.54 (CHCl$_3$, *c* 0.21); $^1$H NMR (400 MHz, CDCl$_3$) δ 5.7 (d, J = 8.4 Hz, 1H), 5.24 (dd, J = 9.6 and 9.2 Hz, 1H), 5.12 (dd, J = 9.2 and 8.4 Hz, 1H), 5.11 (dd, J = 10.0 and 9.6 Hz, 1H), 4.27 (dd, J = 12.4 and 4.8 Hz, 1H), 4.10 (dd, J = 12.4 and 2.4 Hz, 1H), 3.83 (ddd, J = 10.0, 4.8 and 2.4 Hz, 1H), 2.10 (s, 3H), 2.07 (s, 3H), 2.02 (s, 6H), 1.99 (s, 3H); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 170.5, 170.0, 169.3, 169.2, 168.9, 91.7, 72.7 (2C), 70.2, 67.7, 61.4, 20.7, 20.6, 20.5 (3C).

To a solution of compound **S8** (12.8 mmol, 5 g) in DCM (50 mL), HSTol (15.36 mmol, 1.90 g) was added. The mixture was cooled to 0 °C, and EF$_3$.OEt$_2$ (6.4 mmol, 0.9 g) was added. The mixture was stirred at room temperature under $N_2$. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (300 mL), and washed with water, saturated $NaHCO_3$. Organic layer was dried over $MgSO_4$. White solid *p*-Tolyl 2,3,4,6-tetra-*O*-acetyl-1-thio-β-D-glucopyranoside **S9** was obtained after recrystallization from a mixture of dichloromethane and hexanes, followed by drying under reduced pressure. β-anomer (5.22 g, 90%). $^1$H NMR (500 MHz, CDCl$_3$) δ 7.39 (d, *J* = 7.2 Hz, 2H, Ar-H), 7.12 (d, *J* = 8.1 Hz, 2H, Ar-H), 5.20 (t, *J* = 9.5 Hz, 1H, H-3), 5.02 (t, *J* = 9.5 Hz, 1H, H-4), 4.93 (dd, *J* = 10.2, 9.5 Hz, 1H, H-2), 4.63 (d, *J* = 10.2 Hz, 1H, H-1), 4.20 (td, *J* = 12.3, 4.9 Hz, 1H, H-6a), 4.18 (td, *J* = 12.3, 2.6 Hz, 1H, H-6b), 3.70 (ddd, *J* = 9.5, 4.9, 2.6 Hz, 1H, H-5), 2.35 (s, 3H, CH$_3$), 2.09 (s, 3H, OAc), 2.08 (s, 3H, OAc), 2.01 (s, 3H, OAc), 1.98 (s, 3H, OAc); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 170.6 (C), 170.2 (C), 169.4 (C), 169.2 (C), 138.8 (C), 133.8 (CH), 129.7 (CH), 127.5 (C), 85.8 (CH), 75.7 (CH), 74.0 (CH), 69.9 (CH), 68.2 (CH), 65.1 (CH$_2$), 21.2 (CH$_3$), 20.8 (CH$_3$), 20.7 (CH$_3$), 20.6 (CH$_3$).

Compound **S9** (5.0 g, 11.00 mmol) was dissolved in MeOH (50 mL) and NaOMe (59.4 mg, 1.10 mmol) was added into the mixture. The reaction was processed at room temperature for 12 h. Amberlite 120 (H$^+$) was used to neutralize the reaction. After filtration, MeOH was evaporated, and the desired white-solid **S10** was dried under high vacuum system. The **S10** in DMF (32 ml) was mixed with BnBr (6.4 ml, 52.80 mmol) and stirred for 1 h at 0 °C. NaH (3.5 g, 88.00 mmol) was added into the mixture slowly at 0 °C. The reaction underwent at room temperature for overnight. Water was added into reaction to quench NaH in an ice-bath. The mixture was washed with H$_2$O (10 mL). The aqueous layer was extracted with EtOAc (3 × 5 mL), dried with anhydrous MgSO$_4$, filtered, and concentrated *in vacuo*. The mixture was purified by flash column chromatography (*n*-Hexane/EtOAc 6:1) on silica gel to give white-solid product **B** (5.7 g, 80%). The RRV referring to previous report is 2656 *(13)*. $[\alpha]^{27}_D$ -3.3 (*c* 0.8, CH$_2$Cl$_2$); IR (CH$_2$Cl$_2$) *v* 2864, 1067, 735, 697 cm$^{-1}$; $^1$H NMR (500 MHz, CD$_2$Cl$_2$) δ 7.47 (d, *J* = 8.2 Hz, 2H, Ar-H), 7.41-7.24 (m, 18H, Ph-H), 7.21 (dd, *J* = 7.8, 2.0 Hz, 2H, Ar-H), 7.08 (d, *J* = 7.8 Hz, 2H, Ar-H), 4.89, 4.84 (ABq, *J* = 11.0 Hz, 2H, CH$_2$Ph), 4.89, 4.64 (ABq, *J* = 10.5 Hz, 2H, CH$_2$Ph), 4.82, 4.59 (ABq, *J* = 10.2

Hz, 2H, CH$_2$Ph), 4.63 (d, $J$ = 9.8 Hz, 1H, H-1), 4.59, 4.53 (ABq, $J$ = 11.9 Hz, 2H, CH$_2$Ph), 3.77 (dd, $J$ = 10.8, 1.9 Hz, 1H, H-6a), 4.72 (dd, $J$ = 10.8, 5.7 Hz, 1H, H-6b), 3.68 (t, $J$ = 9.8 Hz, 1H, H-3), 3.61 (t, $J$ = 9.8 Hz, 1H, H-4), 3.48 (ddd, $J$ = 9.8, 5.7, 1.9 Hz, 1H, H-5), 3.46 (dd, $J$ = 9.8, 5.7 Hz, 1H, H-2), 2.32 (s, 3H, CH$_3$); $^{13}$C NMR (125 MHz, CD$_2$Cl$_2$) δ 139.1 (C), 138.83 (C), 138.79 (C), 138.76 (C), 138.1 (C), 132.5 (CH), 130.5 (C), 130.0 (CH), 128.7 (CH), 128.6 (CH), 128.5 (CH), 128.3 (CH), 128.12 (CH), 128.10 (CH), 128.0 (CH), 127.91 (CH), 127.89 (CH), 88.1 (CH), 87.0 (CH), 81.3 (CH), 79.3 (CH), 78.2 (CH), 75.9 (CH$_2$), 75.5 (CH$_2$), 75.2 (CH$_2$), 73.7 (CH$_2$), 69.5 (CH$_2$), 21.2 (CH$_3$); HRMS (ESI) calcd. for C$_{41}$H$_{42}$O$_5$NaS [M + Na]$^+$ 669.2651, found 669.2648.

**$p$-Tolyl 2,3,4,6-tetra-$O$-benzyl-1-thio-β-D-galactopyranoside (C)**
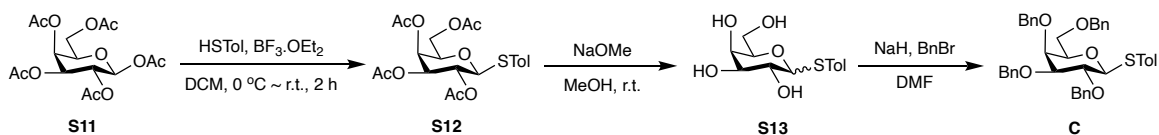


To a solution of compound **S11** (51.2 mmol, 20 g) in DCM (200 mL), HSTol (61.4 mmol, 7.63 g) was added. The mixture was cooled to 0 °C, and EF$_3$.OEt$_2$ (74.6 mmol, 9.32 ml) was added. The mixture was stirred at room temperature under N$_2$. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (300 mL), and washed with water, saturated NaHCO$_3$, organic layer was dried over MgSO$_4$. $p$-Tolyl 2,3,4,6-tetra-$O$-acetyl-1-thio-β-D-galactopyranoside **S12** (white solid) was obtained after recrystallization from a mixture of dichloromethane and hexanes, followed by drying under reduced pressure *(13)*. β-anomer (22.26 g, 96%). $^1$H NMR (400 MHz, CDCl$_3$): δ 7.09 (dd, $J$ = 8.7, 0.6 Hz, 2H), 6.90 (d, $J$ = 8.6 Hz, 2H), 5.47 (dd, $J$ = 9.8, 7.3 Hz, 1H), 5.45 (dd, $J$ = 3.3, 1.1 Hz, 1H), 5.10 (dd, $J$ = 10.5, 3.4 Hz, 1H), 4.99 (d, $J$ = 8.0 Hz, 1H), 4.20 (ddd, $J$ = 27.9, 11.3, 6.7 Hz, 2H), 4.04 (td, $J$ = 6.9, 1.1 Hz, 1H), 2.30 (s, 3H), 2.18 (s, 3H), 2.07 (s, 3H), 2.06 (s, 3H), 2.01 (s, 3H). $^{13}$C NMR (400 MHz, CDCl$_3$): δ 170.48, 170.39, 170.26, 169.52, 155.07, 132.97, 130.12 (2C), 117.09 (2C), 100.24, 71.07, 71.01, 68.85, 67.04, 61.48, 20.87, 20.80, 20.79, 20.73.

Compound **C** was derived from compound **S13** (5.0 g, 11.00 mmol) with the same general procedure as compound **B**. The white-solid product **C** was obtained with the yield of 5.7 g, 80%. The RRV referring to previous report is 17000 *(13)*. [α]$^{27}$$_D$ -1.5 (*c* 0.5, CH$_2$Cl$_2$); IR (CH$_2$Cl$_2$) $v$ 2862, 1089, 1028, 733, 696 cm$^{-1}$; $^1$H NMR (500 MHz, CD$_2$Cl$_2$) δ 7.44 (d, $J$ = 8.1 Hz, 2H, Ar-H), 7.38-7.27 (m, 20H, Ph-H), 7.03 (d, $J$ = 7.9 Hz, 2H, Ar-H), 4.94, 4.57 (ABq, $J$ = 11.2 Hz, 2H, CH$_2$Ph), 4.77, 4.74 (ABq, $J$ = 10.3 Hz, 2H, CH$_2$Ph), 4.76, 4.71 (ABq, $J$ = 11.5 Hz, 2H, CH$_2$Ph), 4.59 (d, $J$ = 9.6 Hz, 1H, H-1), 4.49, 4.44 (ABq, $J$ = 11.7 Hz, 2H, CH$_2$Ph), 3.98 (d, $J$ = 2.6 Hz, 1H, H-4), 3.81 (t, $J$ = 7.7 Hz, 1H, H-2), 3.65-3.59 (m, 4H, H-6a, H-6b, H-5, H-3), 2.32 (s, 3H, CH$_3$); $^{13}$C NMR (125 MHz, CD$_2$Cl$_2$) δ 139.3 (C), 139.1 (C), 138.9 (C), 138.6 (C), 137.7 (C), 132.3 (CH), 130.8 (C), 129.9 (CH), 128.7 (CH), 128.6 (CH), 128.5 (CH), 128.2 (CH), 128.02 (CH), 127.96 (CH), 127.92 (CH), 127.86 (CH), 88.3 (CH), 84.5 (CH), 77.9 (CH), 77.6 (CH), 75.8 (CH$_3$), 75.0

(CH$_3$), 74.4 (CH), 73.8 (CH$_3$), 73.0 (CH$_3$), 69.4 (CH$_3$), 21.2 (CH$_3$); HRMS (ESI) calcd. for C$_{41}$H$_{42}$O$_5$NaS [M + Na]$^+$ 669.2651, found 669.2656.

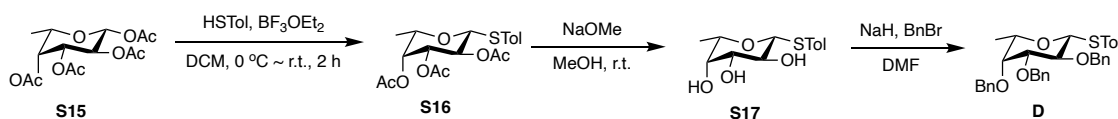### *p*-Tolyl 2,3,4-tri-*O*-benzyl-6-methoxy-1-thio-L-fucopyranoside (D)



To a solution of compound **S15** (56.2 mmol, 20 g) in DCM (200 mL), HSTol (67.3 mmol, 8.37 g) was added. The mixture was cooled to 0 °C, and EF$_3$.OEt$_2$ (82.4 mmol, 10.3 ml) was added. The mixture was stirred at room temperature under N$_2$. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (300 mL), and washed with water, saturated NaHCO$_3$. Organic layer was dried over MgSO$_4$. The obtained crude was recrystallized by EtOAc/hexanes to give *p*-Tolyl 2,3,4-tri-*O*-acetyl-1-thio-β-L-fucopyranoside **S16** as white solid (12. 57 g, 54%). $^1$H NMR (600 MHz, CDCl$_3$): δ 1.15 (d, 3H, $J_{5,6}$ = 6.6 Hz, H-6), 1.90, 2.01, 2.07 (35, 9H, 3 x COCH$_3$), 2.26 (s, 3H, SPhCH$_3$), 3.74 (m, 1H, H-5), 4.58 (d, H, $J_{1,2}$ = 10.2 Hz, H-1), 4.98 (dd, 1H, $J_{2,3}$ = 9.6 Hz, $J_{3,4}$ = 3.0 Hz, H-3), 5.13 (t, 1H, $J$ = 9.6 Hz, H-2), 5.18 (d, 1H, $J_{3,4}$ = 3.0 Hz, H-4), 7.00 - 7.09 (m, 2H), 7.30-7.39 (m, 2H); $^{13}$C NMR (150 MHz, CDCl$_3$): 16.7, 20.8, 20.9, 21.1, 21.3, 67.5, 70.5, 72.6, 73.2, 86.9, 129.3, 129.8, 133.0, 138.3, 169.6, 170.3, 170.8. ESI-MS calcd. for [M + Na]$^+$ C$_{19}$H$_{24}$NaO$_7$S 419.1, found 419.2

Compound **D** was derived from compound **S17** (5.0 g, 11.00 mmol) with the same general procedure as compound **B**. The white-solid product **D** was obtained with the yield of 9.4 g, 75% for two steps. The RRV referring to previous report is 72000 *(13)*. $^1$H NMR (500 MHz, CDCl$_3$) δ 7.50-7.48 (m, 2H), 7.40–7.27 (m, 15H), 7.01 (d, $J$ = 8.1 Hz, 2H), 5.00 (d, $J$ = 11.8 Hz, 1H), 4.80 (d, $J$ = 10.3 Hz, 1H), 4.73–4.71 (m, 3H), 4.66 (d, $J$ = 11.8 Hz, 1H), 4.54 (d, $J$ = 9.6 Hz, 1H), 3.89 (t, $J$ = 9.6 Hz, 1H), 3.62-3.61 (m, 1H), 3.58 (dd, $J$ = 9.6, 3.0 Hz, 1H), 3.52-3.48 (m, 1H), 2.29 (s, 3H), 1.25 (d, $J$ = 6.2 Hz, 3H); HRMS (m/z) calcd. for C$_{34}$H$_{36}$O$_4$SNa [M + Na]$^+$ 563.2226, found 563.2220.

### *p*-Methylphenyl 3,4,6-tri-*O*-benzyl-2-deoxy-1-thio-D-glucopyranoside (E)

To a solution of compound **S18** (30.4 mmol, 5 g) in pyridine (50 mL), DMAP (2.97 mmol, 0.36 g) was added. The mixture was cooled to 0 °C, and acetic anhydride (Ac$_2$O) (194.5 mmol, 18.7 mL) was added dropwise while maintaining the temperature. The mixture was stirred at room temperature under N$_2$ for 5 h and then quenched with ethanol (20 mL) at 0 °C. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (250 mL), and washed with water, saturated NaHCO$_3$, 10% aqueous hydrochloric acid, and brine. The organic phase was dried over MgSO$_4$ and then filtered and concentrated. Silica gel column chromatography (2:1 hexanes-/EtOAc) afforded 8.31 g (82%) of 1,3,4,6-tetra-*O*-acetyl-2-deoxy-D-glucopyranoside **S19** as colorless oil. $^1$H-NMR (400 MHz, CDCl$_3$): δ 5.77 (dd, $J$ = 10.0, 2.20 Hz, 1H), 5.08–4.97 (m, 2H), 4.31–4.26 (m, 1H), 4.01–4.03 (m, 1H), 3.74–3.70 (m, 1H), 2.35–2.29 (m, 1H), 2.09–1.90 (m, 12H). $^{13}$C-NMR (100 MHz, CDCl$_3$): δ 170.71, 170.10, 169.76, 168.79, 91.08, 72.87, 70.15, 68.26, 61.95, 34.73, 20.96, 20.86, 20.77, 20.71. HRMS (*m/z*): calcd. for C$_{14}$H$_{20}$O$_9$ [M + Na]$^+$ 355.0999, found 355.1006.
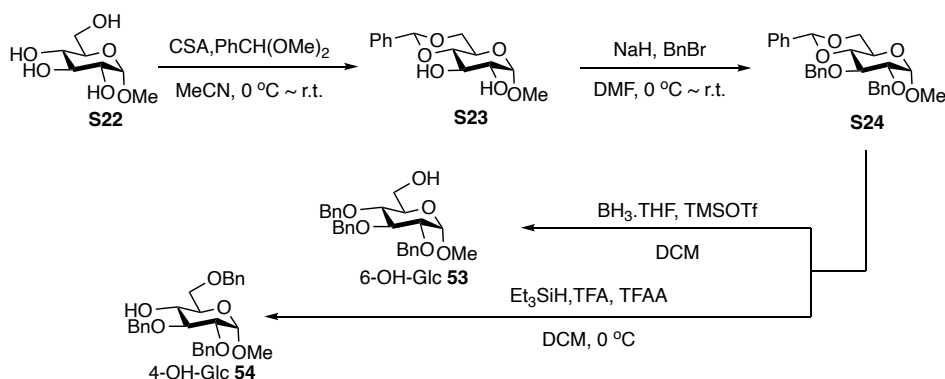
To a solution of compound **S19** (25 mmol, 8.3 g) in DCM (80 mL), HSTol (50 mmol, 6.2 g) was added. The mixture was cooled to 0 °C, and EF$_3$.OEt$_2$ (62.5 mmol, 7.8 ml) was added. The mixture was stirred at room temperature under N$_2$. The mixture was concentrated, and the resulting residue was diluted with ethyl acetate (300 mL), and washed with water, saturated NaHCO$_3$. Organic layer was dried over MgSO$_4$. The organic phase was dried over MgSO$_4$ and then filtered and concentrated Silica gel column chromatography (2:1 hexanes-/EtOAc) afforded 6.9 g (83%) of *p*-Tolyl 3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-D-glucopyranoside **S20** as colorless oil.

Compound **E** was derived from compound **S21** (4.0 g, 14.7 mmol), and compound **S21** was derived from compound **S20** (6.97 g, 17.5 mmol) with the same general procedure as compound **B**. The colorless syrup product **E** was obtained with the yield of 7 g, 88% for two steps *(13)*. α-anomer: Colorless syrup; [α]$^{26}$$_D$ = +186.5° (*c* 1.0, CHCl$_3$); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.36-7.24 (m, 15H, Ph), 7.19 (dd, 2H, $J$ = 7.7 and 1.9 Hz, Ph), 7.02 (d, 2H, $J$ = 8.0 Hz, Ph), 5.59 (d, 1H, $J$ = 5.3 Hz), 4.88 (d, 1H, CH$_2$Ph, $J$ = 10.8 Hz), 4.65 (ABq, 2H, CH$_2$Ph, $J$ = 11.5 Hz), 4.58 (d, 1H, CH$_2$Ph, $J$ = 12.0 Hz), 4.52 (d, 1H, CH$_2$Ph, $J$ = 10.8 Hz), 4.43 (d, 1H, CH$_2$Ph $J$ = 12.0 Hz), 4.30 (ddd, 1H, $J$ = 9.8, 4.0 and 1.8 Hz), 3.95 (ddd, 1H, $J$ = 11.7, 8.8 and 4.8 Hz), 3.80 (dd, 1H, $J$ = 10.6 and 4.9 Hz), 3.64 (dd, 1H, $J$ = 11.0 and 1.9 Hz), 3.61 (t, 1H, $J$ = 9.2 Hz), 2.43 (dd, 1H, $J$ = 12.9 and 4.9 Hz), 2.28 (s, 3H), 2.08 (ddd, 1H, $J$ = 13.2, 11.7 and 5.6 Hz); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 138.7 (C), 138.6 (C), 138.4 (C), 137.4 (C), 132.0 (CH), 131.3 (C), 129.8 (CH), 128.6 (CH), 128.5 (CH), 128.4 (CH), 128.06 (CH), 128.0 (CH), 127.9 (CH), 127.8 (CH), 127.78 (CH), 127.7 (CH), 84.5 (CH), 78.7 (CH), 78.1 (CH), 75.1 (CH$_2$), 73.5 (CH$_2$), 72.06 (CH$_2$), 71.9 (CH), 69.2 (CH$_2$), 36.3 (CH$_2$), 21.2 (CH$_3$); IR (thin film): ν = 3035, 2929, 1496, 1456, 1089, 750, 698, 649 cm$^{-1}$; HRMS (ESI) calcd. for C$_{34}$H$_{36}$NaSO$_4$ [M+Na$^+$] 563.2232, found 563.2227. β-anomer: Colorless syrup; [α]$^{29}$$_D$ = -38.0° (*c* 5.0, CH$_2$Cl$_2$); $^1$H NMR (400 MHz, CDCl$_3$) δ 7.40 (d, 2H, $J$ = 8.1 Hz, Ph), 7.32-7.19 (m, 15H, Ph), 7.01 (d, 2H, $J$ = 8.1 Hz, Ph), 4.86 (d, 1H, CH$_2$Ph, $J$ = 10.9 Hz), 4.65 (dd, 1H, $J$ = 11.7 and 2.0 Hz), 4.658 (d, 1H, CH$_2$Ph, $J$ = 11.7 Hz), 4.57 (d, 2H, CH$_2$Ph, $J$ = 11.6 Hz),

4.56 (d, 1H, CH$_2$Ph, $J$ = 10.9 Hz), 4.51 (d, 1H, CH$_2$Ph, $J$ = 11.8 Hz), 3.77 (dd, 1H, $J$ = 10.9 and 1.8 Hz), 3.70 (dd, 1H, $J$ = 10.9 and 4.6 Hz), 3.65 (ddd, 1H, $J$ = 11.2, 8.1 and 5.3 Hz), 3.47 (t, 1H, $J$ = 9.7 Hz), 3.43 (ddd, 1H, $J$ = 9.6, 4.8 and 1.6 Hz), 2.41 (ddd, 1H, $J$ = 12.8, 5.2 and 1.8 Hz), 2.28 (s, 3H), 1.75 (q, 1H, $J$ = 11.7 Hz); $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 138.7 (C), 138.6 (C), 138.4 (C), 137.7 (C), 132.5 (CH), 130.1 (C), 129.8 (CH), 128.6 (CH), 128.5 (CH), 128.4 (CH), 128.1 (CH), 127.8 (CH), 127.6 (CH), 82.5 (CH), 80.9 (CH), 79.6 (CH), 78.1 (CH), 75.2 (CH$_2$), 73.6 (CH$_2$), 71.8 (CH$_2$), 69.7 (CH$_2$), 37.1 (CH$_2$), 21.3 (CH$_3$); IR (thin film): $\nu$ = 3064, 2887, 1496, 1456, 1276, 1080, 750, 698 cm$^{-1}$; HRMS (ESI) calcd. for C$_{34}$H$_{36}$NaSO$_4$ [M+Na$^+$] 563.2232, found 563.2239.

### 3.2 Synthesis of alcohols

**Methyl 2,3,4-tri-*O*-benzyl-α-D-glucopyranoside (53)**



To a solution of compound methyl 2,3-di-*O*-benzyl-4,6-*O*-benzylidene-α-D-glucopyranoside **S24** (0.3 g, 0.65 mmol) in BH$_3$.THF (0.47 mL, 4.86 mmol). The reaction was stirred for 10 min, and then TMSOTf (0.024 mL, 0.13 mmol) was added. The reaction was stirred at 0 °C gradually to room temperature for 2 h, and quenched by MeOH (5 mL) and evaporated under reduced pressure. The crude product was purified by flash column chromatography on silica gel using EtOAc/ hexane (1:2) as the eluent to obtain product as a colorless oil (0.316 g, 99%). [α]$^{28}_D$ -204.2 (*c* 1.3, CHCl$_3$); IR: $\nu$ 2959, 1069, 1051, 737, 697 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) $\delta$ 7.33-7.26 (m, 15H, Ph-H), 4.98 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 4.85 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 4.82 (d, $J$ = 12.2 Hz, 1H, CH$_2$Ph), 4.68 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 4.66 (d, $J$ = 11.0 Hz, 1H, CH$_2$Ph), 4.59 (d, $J$ = 3.6 Hz, 1H, H-1), 4.03 (dd, $J$ = 9.7, 9.3 Hz, 1H, H-3), 3.79 (dd, $J$ = 11.7, 6.5 Hz, 1H, H-6a), 3.71 (dd, $J$ = 11.7, 3.8 Hz, 1H, H-6b), 3.69 (t, $J$ = 4.2 Hz, 1 H, H-5), 3.54 (dd, $J$ = 9.3, 5.2 Hz, 1H, H-4), 3.52 (dd, $J$ = 9.7, 3.6 Hz, 1H, H-2), 3.35 (s, 3 H, CH$_3$); $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 138.8 (C), 138.2 (C), 138.0 (C), 128.6 (CH), 128.5 (CH), 128.3 (CH), 128.1 (CH), 128.0 (CH), 127.9 (CH), 127.7 (CH), 127.4 (CH), 126.0 (CH), 98.2 (CH), 82.0 (CH), 80.0 (CH), 77.5 (CH), 75.7 (CH$_2$), 75.0 (CH$_2$), 73.4 (CH$_2$), 70.7 (CH), 61.9 (CH), 55.2 (CH$_3$); HRMS (ESI) calcd. for C$_{28}$H$_{32}$O$_6$Na [M + Na]$^+$ 487.2097, found 487.2053.

## Methyl 2,3,6-tri-O-benzyl-α-D-glucopyranoside (54)



4-OH-Glc **54**

Methyl 2,3-di-*O*-benzyl-4,6-*O*-benzylidene-α-D-glucopyranoside **S24** (0.4 g, 0.865 mmol) was dissolved in DCM and then added Et₃SiH (1.24 mL, 7.78 mmol), TFA (0.5 mL, 6.48 mmol) and TFFA (0.025 mL, 0.173 mmol) in the same batch. The reaction was stirred at -40 °C for 6 h and quenched by Et₃N (5 mL). The mixture was extracted with EtOAc (20 mL) and NaHCO₃ (20 mL×2). The organic layers were combined, dried over anhydrous MgSO₄, filtered, and evaporated under reduced pressure. The crude product was purified by flash column chromatography on silica gel using EtOAc/ hexane (1/3) as the eluent to obtain product as a colorless oil (0.365 g, 91 %) *(13)*. [α]$^{28}$$_D$ -290.6 (*c* 0.5, CHCl₃); IR: *v* 2922, 1055, 1028, 736, 697 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl₃) δ 7.35-7.28 (m, 15H, Ar-H), 4.98 (d, *J* =11.4 Hz, 1H, CH₂Ph), 4.78 (d, *J* = 12.1 Hz, 1H, CH₂Ph), 4.75 (d, *J* = 11.4 Hz, 1H, CH₂Ph), 4.67 (d, *J* = 12.1 Hz, 1H, CH₂Ph), 4.64 (d, *J* = 3.6 Hz, 1H, H-1), 4.60 (d, *J* = 12.2 Hz, 1H, CH₂Ph), 4.55 (d, *J* = 12.2 Hz, 1H, CH₂Ph), 3.79 (t, *J* = 9.1 Hz, 1H, H-3), 3.74-3.67 (m, 3H, H-5, H-6a, H-6b), 3.61 (t, *J* = 9.1 Hz, 1H, H-4), 3.54 (dd, *J* = 9.1, 3.6 Hz, 1H, H-2), 3.39 (s, 3H, CH₃), 2.30 (d, *J* = 2.1 Hz, 1H, OH); $^{13}$C NMR (125 MHz, CDCl₃) δ 138.8 (C), 138.2 (C), 138.0 (C), 128.6 (CH), 128.5 (CH), 128.3 (CH), 128.1 (CH), 128.0 (CH), 127.9 (CH), 127.7 (CH), 127.4 (CH), 126.0 (CH), 98.2 (CH), 81.4 (CH), 79.6 (CH), 75.4 (CH₂), 73.6 (CH₂), 73.2 (CH₂), 70.7 (CH), 69.9 (CH), 69.5 (CH) ), 55.2 (CH₃); HRMS (ESI) calcd. for C₂₈H₃₂O₆Na [M + Na]$^+$ 487.2097, found 487.2146.
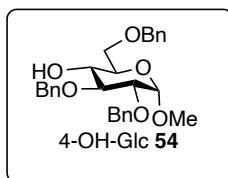
## Methyl 2,3,6-tri-*O*-benzyl-α-D-galactopyranoside (56)
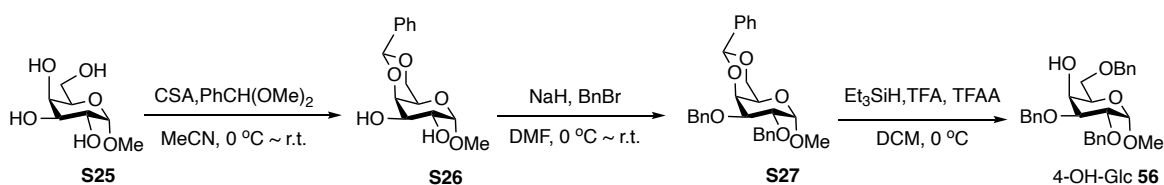


To a solution of methyl 2,3-di-*O*-benzyl-4,6-*O*-benzylidene-α-D-galactopyranoside **S27** (1.65 g, 3.57 mmol) in DCM, Et₃SiH (3.5 mL, 21.42 mmol), TFA (1.4 mL, 17.85 mmol) and TFAA (101 μL, 0.714 mmol) was added. The reaction was stirred at 0 °C for overnight and quenched by Et₃N (1.5 mL). The mixture was extracted with EtOAc (20 mL x 3) and water (20 mL). The organic layers were combined, dried over anhydrous MgSO₄, filtered, and evaporated under reduced pressure. The crude product was purified by flash column chromatography EtOAc/Hexane (1/3) on silica gel to give colorless oil product (1.16 g, 70%) *(13)*. [α]$^{26}$$_D$ +62.7 (*c* 1.0, CH₂Cl₂); IR (CH₂Cl₂) *v* 3476, 3062, 3030, 2910, 1604, 1496, 1454, 1096, 1049, 738, 698 cm$^{-1}$; $^1$H NMR (600 MHz, CDCl₃) δ 7.38-7.28 (m, 15H, Ph-H), 4.82 (ABq, *J* = 12.5 Hz, 1H, CH₂Ph), 4.80 (ABq, *J* =

12.3 Hz, 1H, CH$_2$Ph), 4.71 (ABq, $J$ = 12.3 Hz, 1H, CH$_2$Ph), 4.69 (d, $J$ = 2.8 Hz, 1H, H-1), 4.67 (ABq, $J$ = 12.5 Hz, 1H, CH$_2$Ph), 4.60 (ABq, $J$ = 11.9 Hz, 1H, CH$_2$Ph), 4.56 (ABq, $J$ = 11.9 Hz, 1H, CH$_2$Ph), 4.06 (d, $J$ = 1.0 Hz, 1H, H-4), 3.90 (td, $J$ = 5.6, 1.0 Hz, 1H, H-5), 3.88-3.87 (m, 2H, H-2, H-3), 3.74 (dd, $J$ = 10.1, 5.6 Hz, 1H, H-6a), 3.68 (dd, $J$ = 10.1, 6.2 Hz, 1H, H-6b), 3.39 (s, 3H, OCH$_3$); $^{13}$C NMR (150 MHz, CDCl$_3$) δ 138.3 (C), 138.1 (C), 138.0 (C), 128.5 (CH), 128.4 (CH), 128.0 (CH), 127.82 (CH), 127.79 (CH), 127.77 (CH), 127.6 (CH), 98.6 (CH), 77.6 (CH), 75.7 (CH), 73.6 (CH$_2$), 73.5 (CH$_2$), 72.7 (CH$_2$), 69.6 (CH$_2$), 68.3 (CH), 68.1 (CH), 55.3 (CH$_3$); HRMS (ESI) calcd. for C$_{28}$H$_{32}$O$_6$Na [M + Na]$^+$ 487.2097, found 487.2106.

## Methyl 2-azido-3,4,6-tri-*O*-benzyl-2-deoxy-β-D-glucopyranoside (A51)



[α]$^{25}_D$ -0.2 (*c* 1.2, CHCl$_3$); IR (CHCl$_3$) *v* 2920, 2108, 1453, 1272, 1218, 1057, 772, 697 cm$^{-1}$; $^1$H NMR (400 MHz, CDCl$_3$) δ 7.37-7.16 (m, 15H, Ph-H), 4.90-4.79 (m, 3H, CH$_2$Ph), 4.64-4.53 (m, 3H, CH$_2$Ph), 4.17 (d, 1H, $J$ = 7.4 Hz, H-1), 3.75-3.68 (m, 2H, H-6$_{ax}$, H-6$_{eq}$), 3.63 (t, $J$ = 8.8 Hz, 1H, H-4), 3.58 (s, 3H, OCH$_3$), 3.47-3.35 (m, 3H, H-2, H-6a, H-6b); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 138.0 (C), 137.9 (C), 128.5 (CH), 128.43 (CH), 128.37 (CH), 128.0 (CH), 127.9 (CH), 127.8 (CH), 127.7 (CH), 102.9 (CH), 83.2 (CH), 77.8 (CH), 75.5 (CH), 75.1 (CH$_2$), 75.0 (CH), 73.5 (CH$_2$), 68.6 (CH$_2$), 66.3 (CH), 57.1 (CH$_3$); HRMS (ESI) calcd. for C$_{28}$H$_{31}$N$_3$O$_5$Na [M+Na]$^+$ 512.2161, found 512.2203.

## Methyl 6-*O*-(2-azido-3,4,6-tri-O-benzyl-2-deoxy-β-D-glucopyranosyl)-2,3,4-tri-*O*-benzyl-α-D-glucopyranoside (A53)



[α]$^{19}_D$ = 5.27 (*c* 1.2, CHCl$_3$); IR (CHCl$_3$) *v* 2922, 2108, 1453, 1359, 1060, 735, 696 cm$^{-1}$; 1H NMR (400 MHz, CDCl$_3$) δ 7.38–7.16 (m, 30H, Ph-H), 5.06-4.64 (m, 10H, CH$_2$Ph), 4.63 (s, 1H, H-1), 4.60-4.45 (m, 2H, CH$_2$Ph), 4.21 (d, $J$ = 7.7 Hz, 1H, H-1'), 4.17 (dd, $J$ = 11.1, 2.0 Hz 1H, H-6a), 4.05 (t, $J$ = 9.3 Hz, 1H, H-3), 3.85 (ddd, $J$ = 10.5, 4.2, 1.8 Hz, 1H, H-5), 3.76-3.68 (m, 3H, H-6b, H-6'a, H-6'b), 3.65-3.58 (m, 3H, H-2, H-4, H-4'), 3.53-3.40 (m, 6H, H-2', H-3', H-5', OCH$_3$); $^{13}$C NMR (100 MHz, CDCl$_3$) δ 138.8 (C), 138.4 (C), 138.1 (CH), 138.0 (C), 137.8 (C), 128.4 (CH), 128.3 (CH), 128.1 (CH), 128.0 (CH), 127.9 (CH), 127.82 (CH), 127.80 (CH), 127.7 (CH), 127.60 (CH), 127.55 (CH), 127.50 (CH), 102.1 (CH), 98.1 (CH), 83.2 (CH), 82.0 (CH), 79.8 (CH), 77.72 (CH), 77.70 (CH), 75.6 (CH$_2$), 75.5 (CH$_2$), 75.2 (CH), 74.94 (CH$_2$), 74.88 (CH$_2$), 74.8 (CH$_2$),

74.4 (CH$_2$), 69.7 (CH), 68.6 (CH$_2$), 68.4 (CH$_2$), 66.3 (CH), 55.1 (CH$_3$); HRMS (ESI) calcd. for C$_{55}$H$_{59}$N$_3$O$_{10}$Na [M + Na]$^+$ 944.4098, found 944.4083 *(13)*.

## Methyl 4-*O*-(2-azido-3,4,6-tri-O-benzyl-2-deoxy-β-D-glucopyranosyl)-2,3,6-tri-*O*-benzyl-α-D-glucopyranoside (A54)



Data for β-anomer (**A54β**): [α]$_D$$^{15}$=−11.4° (c 1.09, CHCl$_3$); IR (CHCl$_3$) *v* 3009, 2910, 2870, 2112, 1496, 1454, 1361, 1277, 1087, 750 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 3.18 (ddd, J=1.7, 4.3, 9.6 Hz, 1H, H-5′), 3.24 (dd, J=8.9, 9.8 Hz, 1H, H-3′), 3.31 (dd, J=8.0, 9.8 Hz, 1H, H-2′), 3.38 (s, 3H, OCH$_3$), 3.47–3.50 (m, 2H, H-2, H-6′a), 3.59 (dd, J=8.9, 9.6 Hz, 1H, H-4′), 3.62 (dd, J=1.7, 11.1 Hz, 1H, H-6′b), 3.70 (dd, J=1.6, 10.9 Hz, 1H, H-6a), 3.78 (ddd, J=1.6, 3.2, 9.8 Hz, 1H, H-5), 3.89 (dd, J=8.9, 9.5 Hz, 1H, H-3), 3.92 (dd, J=3.2, 10.9 Hz, 1H, H-6b), 3.96 (dd, J=8.9, 9.8 Hz, 1H, H-4), 4.26 (d, J=8.0 Hz, 1H, H-1′), 4.35 (d, J=12.1 Hz, 1H, O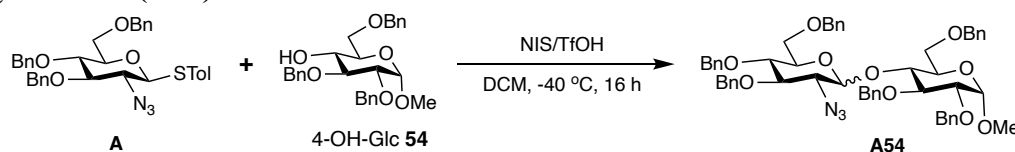CHPh), 4.40 (d, J=12.1 Hz, 1H, OCHPh), 4.47 (d, J=12.1 Hz, 1H, OCHPh), 4.54 (d, J=11.0 Hz, 1H, OCHPh), 4.58 (d, J=12.1 Hz, 1H, OCHPh), 4.59 (d, J=3.6 Hz, 1H, H-1), 4.67 (d, J=12.1 Hz, 1H, OCHPh), 4.74 (d, J=12.1 Hz, 1H, OCHPh), 4.76 (d, J=11.0 Hz, 1H, OCHPh), 4.781 (d, J=11.4 Hz, 1H, OCHPh), 4.782 (d, J=10.7 Hz, 1H, OCHPh), 4.82 (d, J=10.7 Hz, 1H, OCHPh), 5.02 (d, J=11.4 Hz, 1H, OCHPh), 7.17–7.36 (m, 30H, Ar-H); $^{13}$C NMR (126 MHz, CDCl$_3$) δ 55.3, 66.9, 68.2, 68.6, 69.7, 73.3, 73.47, 73.51, 74.7, 75.19, 75.24, 75.4, 77.9, 79.1, 80.3, 83.3, 98.3 (C-1), 100.9 (C-1′), 127.0, 127.4, 127.5, 127.65, 127.67, 127.70, 127.72, 127.8, 127.95, 127.98, 128.1, 128.2, 128.3, 128.36, 128.42, 128.5, 137.8, 137.9, 138.0, 138.3, 138.4, 139.5; HRMS (ESI) calcd. for C$_{55}$H$_{59}$N$_3$O$_{10}$Na [M+Na]$^+$ 944.4089, found 944.4092. Data for α-anomer (**A54α**): [α]$_D$$^{15}$=+39.1° (c 0.78, CHCl$_3$); IR (CHCl$_3$) *v* 3013, 2910, 2870, 2112, 1602, 1454, 1361, 1221, 1049, 713 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 3.27 (dd, J=3.9, 10.4 Hz, 1H, H-2′), 3.34 (brd, J=11.0 Hz, 1H, H-6′a), 3.38 (s, 3H, OCH$_3$), 3.52 (dd, J=1.4, 11.0 Hz, 1H, H-6′b), 3.57 (dd, J=3.6, 9.6 Hz, 1H, H-2), 3.65 (dd, J=1.9, 11.0 Hz, 1H, H-6a), 3.66–3.70 (m, 2H, H-4′, H-5′), 3.72 (dd, J=4.3, 11.0 Hz, 1H, H-6b), 3.79 (ddd, J=1.9, 4.3, 10.0 Hz, 1H, H-5), 3.86 (m, 1H, H-3′), 3.91 (dd, J=8.6, 10.0 Hz, 1H, H-4), 4.08 (dd, J=8.6, 9.6 Hz, 1H, H-3), 4.25 (d, J=12.1 Hz, 1H, OCHPh), 4.45 (d, J=10.9 Hz, 1H, OCHPh), 4.49 (d, J=12.1 Hz, 1H, OCHPh), 4.50 (s, 2H, OCH$_2$Ph), 4.61 (d, J=3.6 Hz, 1H, H-1), 4.62 (d, J=12.2 Hz, 1H, OCHPh), 4.74 (d, J=10.9 Hz, 1H, OCHPh), 4.75 (d, J=12.2 Hz, 1H, OCHPh), 4.83 (d, J=10.9 Hz, 1H, OCHPh), 4.85 (d, J=10.9 Hz, 1H, OCHPh), 4.86 (d, J=10.7 Hz, 1H, OCHPh), 5.10 (d, J=10.7 Hz, 1H, OCHPh), 5.73 (d, J=3.9 Hz, 1H, H-1′), 7.12 (m, 2H, Ar-H), 7.20–7.35 (m, 28H, Ar-H); $^{13}$C NMR (126 MHz, CDCl$_3$) δ 55.3, 63.3, 67.9, 69.3, 69.5, 71.4, 73.27, 73.31, 73.5, 74.9, 75.0, 75.3, 78.1, 80.1, 80.5, 82.0, 97.69 (C-1′), 97.73 (C-1), 127.2, 127.4, 127.45, 127.53, 127.66, 127.71, 127.8, 127.9, 128.1, 128.26, 128.32, 128.4, 128.5, 137.8, 137.9, 138.0, 138.1, 138.2, 138.7; HRMS (ESI) calcd. for C$_{55}$H$_{59}$N$_3$O$_{10}$Na [M+Na]$^+$ 944.4089, found 944.4093.

**Phenyl 2-azido-3,4,6-tri-*O*-benzyl-2-deoxy-D-glucopyranoside (A55)**



Data for **A55**: IR (CHCl3) *v* 2919, 2851, 2108, 1455, 1222, 1056, 755, 696 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.49-7.32 (m, 26.1H, Ar-H), 7.28-7.24 (m, 3.7H, Ar-H), 7.21-7.17 (m, 3.3H, Ar-H), 7.15-7.10 (m, 1.9H, Ar-H), 5.67 (d, *J* = 3.4 Hz, 1H, H-1α), 5.05-4.99 (m, 2.7H, CH$_2$Ph), 4.93-4.84 (m, 3.24H, H-1β, CH$_2$Ph), 4.70-4.58 (m, 4.3H, CH$_2$Ph), 4.52 (d, *J* = 11.95 Hz, 1H, CH$_2$Ph), 4.31 (dd, *J* = 9.5 Hz, 9.5 Hz, 1H, H-3α,), 4.04 (m, 1H, H-5α), 3.94 (dd, *J* = 9.1 Hz, 9.1 Hz, 1H, H-4α), 3.87-3.76 (m, 3.7H, H-6aα, H-2β, H-4β, H-6aβ, H-6bβ ), 3.71-3.68 (dd, *J* = 10.9 Hz, 1.6 Hz, 1H, H-6bα), 3.67-3.64 (m, 1H, H-5β), 3.60-3.56 (m, 1.8H, H-2α, H-3β); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 157.01 (C), 156.26 (C), 137.89 (C), 137.83 (C), 137.81 (C), 137.79 (C), 137.72 (C), 137.64 (C), 134.26 (CH), 129.68 (CH), 129.54 (CH), 129.48 (CH), 128.42 (CH), 128.40 (CH), 128.38 (CH), 128.36 (CH), 128.31 (CH), 128.28 (CH), 127.98 (CH), 127.84 (CH), 127.81 (CH), 127.74 (CH), 127.69 (CH), 127.64 (CH), 127.57 (CH), 122.99 (CH), 122.75 (CH), 116.94 (CH), 116.69 (CH), 100.49 (CH), 96.76 (CH), 82.92 (CH), 80.03 (CH), 77.93 (CH), 77.49 (CH), 75.53 (CH$_2$), 75.39 (CH$_2$), 75.23 (CH), 74.99 (CH$_2$), 73.42 (CH$_2$), 73.39 (CH$_2$), 71.29 (CH), 68.48 (CH$_2$), 68.01 (CH$_2$), 66.05(CH), 63.12 (CH); HRMS (ESI) calcd. for C$_{33}$H$_{33}$N$_3$O$_5$Na [M+Na]$^+$ 574.2312, found 574.2313.

**Methyl 4-*O*-(2-azido-3,4,6-tri-O-benzyl-2-deoxy-β-D-glucopyranosyl)-2,3,6-tri-*O*-benzyl-α-D-galactopyranoside (A56)**
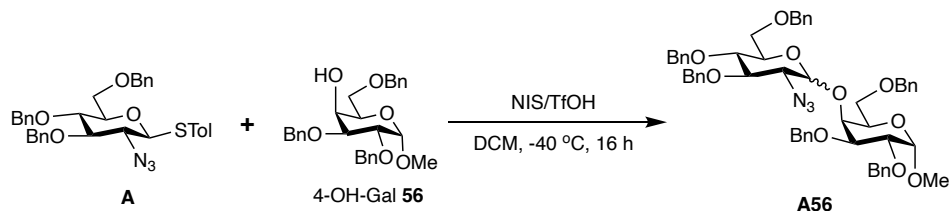


The α- and β-glycosides were separated by flash column chromatography with 15:1 toluene/AcOEt. Data for β-anomer (**A56β**): [α]$_D^{19}$=+3.32° (*c* 0.74, CHCl$_3$); IR (CHCl$_3$) 3020, 2930, 2868, 2114, 1454, 1358, 1277, 1089 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 3.27 (ddd, J=1.9, 3.9, 9.9 Hz, 1H, H-5′), 3.34 (dd, J=8.8, 9.8 Hz, 1H, H-3′), 3.37 (s, 3H, OCH$_3$), 3.40 (dd, J=7.6, 9.8 Hz, 1H, H-2′), 3.58–3.65 (m, 4H, H-6a, H-4′, H-6′a, H-6′b), 3.73 (dd, J=4.7, 10.5 Hz, 1H, H-6b), 3.90–3.94 (m, 2H, H-3, H-5), 4.13 (dd, J=3.7, 10.1 Hz, 1H, H-2), 4.16 (brd, J=3.0 Hz, 1H, H-4), 4.37 (d, J=12.0 Hz, 1H, OCHPh), 4.44 (d, J=12.0 Hz, 1H, OCHPh), 4.48 (d, J=12.1 Hz, 1H, OCHPh), 4.52 (d, J=12.1 Hz, 1H, OCHPh), 4.54 (d, J=12.1 Hz, 1H, OCHPh), 4.64 (d, J=3.7 Hz, 1H, H-1), 4.67 (d, J=12.1 Hz, 1H, OCHPh), 4.69 (d, J=12.1 Hz, 1H, OCHPh), 4.73 (d, J=7.6 Hz, 1H, H-1′), 4.77–4.80 (m, 2H, OCHPh×2), 4.84 (d, J=12.1 Hz, 1H, OCHPh), 4.91 (d, J=10.8 Hz, 1H, OCHPh), 4.92 (d, J=12.1 Hz, 1H, OCHPh), 7.15 (m, 2H, Ar-H), 7.24–7.37 (m, 28H, Ar-

H); $^{13}$C NMR (126 MHz, CDCl$_3$) $\delta$ 55.3, 66.6, 68.9, 69.4, 70.2, 73.2, 73.4, 73.5, 73.8, 74.3, 74.8, 75.0, 75.5, 76.6, 77.7, 78.4, 83.1, 98.8 (C-1), 101.6 (C-1′), 127.2, 127.3, 127.4, 127.5, 127.6, 127.7, 127.80, 127.82, 127.9, 128.0, 128.2, 128.27, 128.32, 128.36, 128.42, 128.44, 137.9, 138.05, 138.14, 138.58, 138.61, 138.9; HRMS (ESI) calcd. for C$_{55}$H$_{59}$N$_3$O$_{10}$Na [M+Na]$^+$ 944.4084, found 944.4093. Data for $\alpha$-anomer (**A56$\alpha$**): [$\alpha$]$_D^{18}$=+44.5° ($c$ 0.21, CHCl$_3$); IR (CHCl$_3$) 3018, 2930, 2870, 2114, 1454, 1358, 1277, 1089 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) $\delta$ 3.03 (dd, $J$=1.7, 11.1 Hz, 1H, H-6′a), 3.28 (dd, $J$=2.0, 11.1 Hz, 1H, H-6′b), 3.35 (dd, $J$=3.6, 9.2 Hz, 1H, H-2′), 3.36 (s, 3H, OC$H_3$), 3.54 (dd, $J$=10.4, 12.9 Hz, 1H, H-5), 3.75 (dd, $J$=9.4, 10.1 Hz, 1H, H-4′), 3.83–3.92 (m, 5H, H-2, H-3, H-6a, H-6b, H-3′), 4.18 (d, $J$=12.2 Hz, 1H, OC$H$Ph), 4.20 (d, $J$=2.9 Hz, 1H, H-4), 4.23 (ddd, $J$=1.7, 2.0, 10.1 Hz, 1H, H-5′), 4.43 (d, $J$=12.2 Hz, 1H, OC$H$Ph), 4.46 (d, $J$=12.2 Hz, 1H, OC$H$Ph), 4.52 (d, $J$=11.8 Hz, 1H, OC$H$Ph), 4.55 (d, $J$=11.8 Hz, 1H, OC$H$Ph), 4.69–4.74 (m, 4H, H-1, OC$H$Ph×3), 4.77 (d, $J$=11.9 Hz, 1H, OC$H$Ph), 4.80–4.82 (m, 2H, OC$H$Ph×2), 4.86 (d, $J$=10.6 Hz, 1H, OC$H$Ph), 4.94 (d, $J$=3.6 Hz, 1H, H-1′), 7.13 (m, 2H, Ar-H), 7.17–7.37 (m, 28H, Ar-H); $^{13}$C NMR (126 MHz, CDCl$_3$) $\delta$ 55.4, 64.1, 67.3, 67.5, 68.9, 70.7, 73.16, 73.19, 73.3, 73.6, 74.8, 74.9, 75.27, 75.33, 78.2, 80.4, 98.5, 98.6, 127.4, 127.5, 127.6, 127.66, 127.73, 127.8, 127.95, 127.99, 128.03, 128.26, 128.31, 128.33, 128.4, 128.5, 137.6, 137.9, 138.1, 138.2, 138.4, 138.7; HRMS (ESI) calcd. for C$_{55}$H$_{59}$N$_3$O$_{10}$Na [M+Na]$^+$ 944.4084, found 944.4093.

## Methyl 2,3,4,6-tetra-*O*-benzyl-D-glucopyranoside (B51)
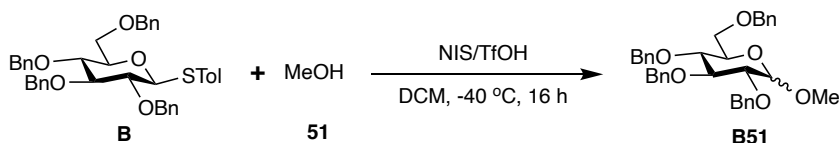


Data for **B51**: [$\alpha$]$^{26}_D$ +31.0 ($c$ 2.9, CH$_2$Cl$_2$); IR (CH$_2$Cl$_2$) $\nu$ 1067, 1046, 1026, 732, 694 cm$^{-1}$; $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 7.41–7.18 (m, 40H, Ar$H$), 5.05–4.05 (m, 16H, ArC$H$), 4.65 (d, $J$ = 3.6 Hz, 1H, H-1$\alpha$), 4.37 (d, $J$ = 8.0 Hz, 1H, H-1$\beta$), 4.04 (t, $J$ = 9.2 Hz, 1H, H-3$\alpha$), 3.82–3.60 (m, 8H, H-3$\beta$, H-4$\alpha$, H-5$\alpha$, H-6$\alpha$, H-6$\beta$, H-4$\beta$), 3.75 (dd, $J$ = 9.2, 3.6 Hz, 1H, H-2$\alpha$), 3.63 (s, 3H, CH$_3$), 3.52 (ddd, $J$ = 9.2, 4.8, 1.6 Hz, 1H, H-5$\beta$), 3.50 (dd, $J$ = 8.4, 8.0 Hz, 1H, H-2$\beta$), 3.42 (s, 3H, CH$_3$) ppm; $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 138.8 (C), 138.6 (C), 138.5 (C), 138.2 (C), 138.1 (C), 137.9 (C), 128.3 (CH), 128.2 (CH), 128.0 (CH), 127.9 (CH),.127.8 (CH), 127.7 (CH), 127.6 (CH), 127.5 (CH), 104.6 (C-$\beta$), 98.1 (C-$\alpha$), 84.6 (CH), 82.2 (CH), 82.0 (CH), 79.8 (CH), 77.8 (CH), 77.6 (CH), 75.6 (CH$_2$), 75.5 (CH$_2$), 74.9 (CH$_2$), 74.8 (CH$_2$), 74.6 (CH), 73.4 (CH$_2$), 70.0 (CH), 68.9 (CH$_2$), 68.5 (CH$_2$), 57.0 (CH$_3$), 55.1 (CH$_3$) ppm; HRMS (ESI-TOF, *m/z*) calcd. for C$_{35}$H$_{38}$O$_6$Na [M + Na]$^+$ 577.2566, found 577.2565 *(13)*.

## Methyl 2,3,4-tri-*O*-benzyl-6-*O*-(2,3,4,6-tetra-*O*-benzyl-D-glucopyranosyl)-$\alpha$-D-glucopyranoside (B53)
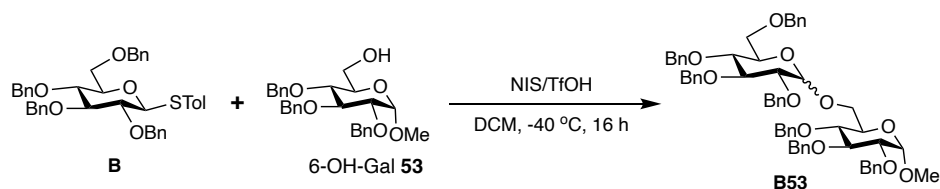
Data for **B53**: $[\alpha]^{26}_D$ +66.9 (*c* 1.2, CH$_2$Cl$_2$); IR (CH$_2$Cl$_2$) ν 1067, 1026, 732, 694 cm$^{-1}$; $^1$H NMR (400 MHz, CDCl$_3$) δ 7.38–7.14 (m, 105H, Ar*H*), 5.03–4.93 (m, 10H), 4.86 (m, 2H), 4.83 (m, 3H), 4.81 (m, 4H), 4.76 (m, 2H), 4.73 (m, 2H), 4.70 (m, 5H), 4.67 (m, 2H), 4.65 (m, 2H), 4.61 (m, 3H), 4.59 (m, 6H), 4.55 (m, 2H), 4.51 (m, 2H), 4.46 (m, 2H), 4.51–4.44 (m, 2H), 4.39 (d, 1H), 4.22 (dd, *J* = 10.8, 2.0 Hz, 0.5 H), 4.02 (m, 5H), 3.84 (m, 5H), 3.75 (m, 4H), 3.69 (m, 5H), 3.66 (m, 3H), 3.63 (m, 2H), 3.59 (m, 6H), 3.55 (m, 2H), 3.48 (m, 3H), 3.39 (s, 6H), 3.36 (s, 3H) ppm; $^{13}$C NMR (100 MHz, CDCl$_3$) δ 138.8 (C), 138.5 (C), 138.4 (C), 138.2 (C), 137.9 (C), 128.4 (CH), 128.3 (CH), 128.2 (CH), 128.1 (CH), 128.0 (CH), 127.9 (CH), 127.8 (CH), 127.7 (CH), 127.6 (CH), 127.5 (CH), 127.4 (CH), 103.8 (CH), 98.0 (CH), 97.9 (CH), 97.2 (CH), 84.8 (CH), 82.1 (CH), 81.9 (CH), 81.6 (CH), 80.1 (CH), 80.0 (CH), 79.8 (CH), 77.9 (CH), 77.8 (CH), 77.6 (CH), 77.3 (CH), 75.6 (CH$_2$), 75.4 (CH$_2$), 75.0 (CH), 74.9 (CH$_2$), 74.8 (CH$_2$), 73.4 (CH$_2$), 73.3 (CH$_2$), 72.3 (CH$_2$), 70.3 (CH), 70.2 (CH), 69.0 (CH$_2$), 68.5 (CH$_2$), 66.0 (CH$_2$), 55.2 (CH$_3$), 55.1 (CH$_3$) ppm; HRMS (ESI-TOF, *m/z*) calcd. for C$_{62}$H$_{66}$O$_{11}$Na [M + Na]$^+$ 1009.4503, found 1009.4511 *(13)*.

## Methyl 2,3,6-tri-*O*-benzyl-4-*O*-(2,3,4,6-tetra-*O*-benzyl-D-glucopyranosyl)-α-D-glucopyranoside (B54)
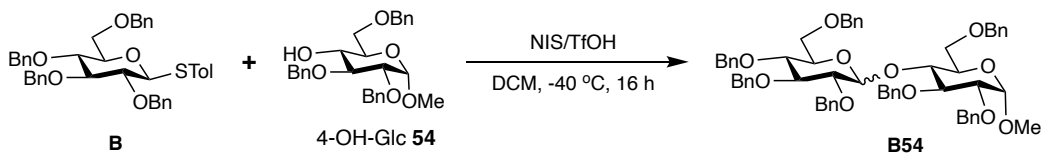


Data for **B54-α**: $^1$H NMR (500 MHz, CDCl$_3$) δ 7.31 – 7.20 (m, 33H), 7.12 (dd, *J* = 7.1, 2.4 Hz, 2H), 5.71 (d, *J* = 3.6 Hz, 1H), 5.05 (d, *J* = 11.6 Hz, 1H), 4.90 (d, *J* = 10.9 Hz, 1H), 4.84 – 4.78 (m, 3H), 4.71 (d, *J* = 12.1 Hz, 1H), 4.63 – 4.57 (m, 3H), 4.57 – 4.50 (m, 4H), 4.44 (d, *J* = 10.9 Hz, 1H), 4.30 (d, *J* = 12.2 Hz, 1H), 4.13 – 4.03 (m, 2H), 3.95 – 3.90 (m, 1H), 3.89 – 3.82 (m, 2H), 3.74 (d, *J* = 10.0 Hz, 1H), 3.69 –3.64 (m, 2H), 3.61 (dd, *J* = 9.2, 3.6 Hz, 1H), 3.51 (dd, *J* = 10.0, 3.5 Hz, 2H), 3.43 – 3.38 (m, 4H). $^{13}$C NMR (126 MHz, CDCl$_3$) δ 139.10, 138.90, 138.63, 138.31, 138.12, 138.07, 128.56, 128.45, 128.42, 128.42, 128.38, 128.35, 128.34, 128.15, 128.05, 127.96, 127.94, 127.83, 127.76, 127.69, 127.62, 127.49, 127.38, 127.21, 126.89, 97.90, 96.77, 82.17, 82.15, 80.34, 79.60, 77.78, 75.67, 75.06, 74.55, 73.58, 73.50, 73.36, 73.27, 72.46, 71.10, 69.66, 69.16, 68.30, 55.28. Data for **B54-β**: $^1$H NMR (600 MHz, CDCl$_3$) δ 7.44 – 7.39 (m, 2H), 7.33 – 7.16 (m, 33H), 5.09 (d, *J* = 11.3 Hz, 1H), 4.87 (d, *J* = 10.9 Hz, 1H), 4.83 – 4.72 (m, 6H), 4.62 – 4.53 (m, 4H), 4.44 (d, *J* = 12.1 Hz, 1H), 4.41 – 4.35 (m, 3H), 3.97 (dd, *J* = 10.0, 8.9 Hz, 1H), 3.88 – 3.81 (m, 2H), 3.72 (dd, *J* = 11.0, 1.9 Hz, 1H), 3.62 – 3.57 (m, 2H), 3.55 (dd, *J* = 11.0, 4.7 Hz, 1H), 3.51 –

3.44 (m, 3H), ), 3.38 – 3.35 (m, 1H), 3.37 (s, 3H), 3.29 (ddd, $J$ = 9.8, 4.7, 1.9 Hz, 1H). $^{13}$C NMR (126 MHz, CDCl$_3$) δ 139.73, 138.75, 138.73, 138.55, 138.47, 137.99, 128.59, 128.51, 128.49, 128.48, 128.38, 128.25, 128.15, 128.12, 127.94, 127.92, 127.90, 127.88, 127.78, 127.73, 127.67, 127.65, 127.45, 127.22, 102.63, 98.57, 85.02, 82.96, 80.57, 78.97, 78.19, 76.75, 75.73, 75.52, 74.93, 73.78, 73.50, 73.48, 70.11, 69.15, 68.00, 55.44. HRMS (ESI) calcd. for C$_{62}$H$_{66}$O$_{11}$Na [M+Na]$^+$ 1009.4197, found 1009.4497.

**Phenyl 2,3,4,6-tetra-*O*-benzyl-D-glucopyranoside (B55)**



Data for **B55-α**: $^1$H-NMR **(**500 MHz, CD$_2$Cl$_2$**)** 3.61 (dd, $J$ = 10.9 Hz, 1H, C-6), 3.72-3.75 (m, 2H, C-4 and C-6), 3.74 (dd, $J$ = 9.6, 3.7 Hz, 1H, C-2), 3.90 (ddd, $J$ = 10.1, 3.8, 1.9 Hz, 1H, C-5), 4.18 (t, $J$ = 9.3 Hz, 1H, C-3), 4.43 (d, $J$ = 11.8 Hz, 1H, CH$_2$Ph), 4.51 (d, $J$ = 11.8 Hz, 1H, CH$_2$Ph), 4.59 (d, $J$ = 11.8 Hz, 1H, CH$_2$Ph), 4.70 (d, $J$ = 11.7 Hz, 1H, CH$_2$Ph), 4.77 (d, $J$ = 11.7 Hz, 1H, CH$_2$Ph), 4.84 (d, $J$ = 11.9 Hz, 1H, CH$_2$Ph), 4.89 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 5.03 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 5.56 (d, $J$ = 3.5 Hz, 1H, C-1), 7.03-7.06 (m, 1H, C- 10), 7.11-7.13 (m, 2H, C-8), 7.22-7.40 (m, 22H, C-9 and Ph). **B54-β**:$^1$H-NMR (500 MHz, CD$_2$Cl$_2$) 3.62-3.65 (m, 1H, C-5), 3.66-3.70 (m, 1H, C-4), 3.70-3.74 (m, 3H, C-2, C-3 and C-6), 3.81 (dd, $J$ = 10.7, 1.9 Hz, 1H, C-6), 4.52 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 4.60 (d, $J$ = 10.9 Hz, 1H, CH$_2$Ph), 4.83-4.87 (m, 2H, CH$_2$Ph), 4.84 (d, $J$ = 11.0 Hz, 1H, CH$_2$Ph), 4.88 (d, $J$ = 10.8 Hz, 1H, CH$_2$Ph), 4.95 (d, $J$ = 11.0 Hz, 1H, CH$_2$Ph), 5.03 (d, $J$ = 11.0 Hz, 1H, CH$_2$Ph), 5.04 (d, $J$ = 6.4 Hz, 1H, C-1), 7.06-7.07 (m, 1H, C-10), 7.08-7.11 (m, 2H, C-8), 7.22-7.40 (m, 22H, C-9 and Ph). $^{13}$C-NMR **(**101 MHz, CD$_2$Cl$_2$**)** = 69.1 (C-6α), 69.4 (C-6β), 71.5 (C-5α), 73.5 (CH$_2$Phα), 73.6 (CH$_2$Phα), 73.7 (CH$_2$Phα), 75.2 (CH$_2$Phβ), 75.31 (CH$_2$Phβ), 75.34 (CH$_2$Phα), 75.4 (C-5β), 75.90 (CH$_2$Phβ), 75.91 (CH$_2$Phβ), 78.0 (C-4α), 78.1 (C-4β), 80.4 (C- 2α), 82.2 (C-3α), 82.3 (C-2β), 85.0 (C-3β), 95.9 (C-1α), 101.8 (C-1β), 117.1 (C-8β), 117.3 (C- 8α), 122.8 (C-10α), 123.0 (C-10β), 127.87 (Ph), 127.91 (Ph), 127.94 (Ph), 127.96 (Ph), 128.02 (Ph), 128.07 (Ph), 128.12 (Ph), 128.16 (Ph), 128.21 (Ph), 128.24 (Ph), 128.28 (Ph), 128.33 (Ph), 128.34 (Ph), 128.6 (Ph), 128.65 (Ph), 128.66 (Ph), 128.67 (Ph), 128.68 (Ph), 128.69 (Ph), 128.8 (Ph), 129.88 (C-9α), 129.93 (C-9β), 138.67 (Ph), 138.68 (Ph), 138.75 (Ph), 138.78 (Ph), 138.95 (Ph), 138.98 (Ph), 139.2 (Ph), 139.5 (Ph), 157.3 (C-7α), 157.8 (C-7β). IR: 3063, 3031, 2917, 2864, 1589, 1598, 1494, 1453, 1360, 1226, 1068, 1027, 733 cm$^{-1}$. HRMS: calcd. for C$_{40}$H$_{40}$NaO$_6$ [M+Na]$^+$ 639.2717, found 639.2699.

**Methyl        2,3,4-tri-*O*-benzyl-4-*O*-(2,3,4,6-tetra-*O*-benzyl-D-glucopyranosyl)-α-D-galactopyranoside (B56)**
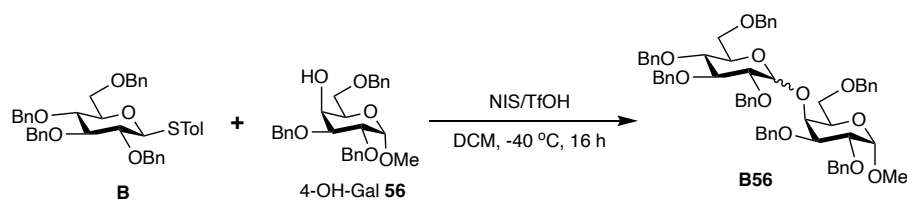
Data for **B56**: IR (CHCl₃) *v* 2919, 1454, 1095, 1071, 1047, 737, 699 cm⁻¹; ¹H NMR (500 MHz, CDCl₃) δ 7.45-7.15 (m, 62.60H, Ar-H), 5.14 (d, *J* = 11.35 Hz, 0.53H, CH₂Ph), 5.07 (d, *J* = 3.4 Hz, 1H, H-1'α), 5.00-4.94 (m, 2.47H, H-1'β, CH₂Ph), 4.90-4.71 (m, 13.84H, H-1α, H-1β, CH₂Ph), 4.65-4.40 (m, 7.44H, CH₂Ph), 4.34-4.16 (m, 6.19H, H-5'α, H-4α, CH₂Ph), 4.04-3.91 (m, 7.61H, H-3'α, H-2α, H-3α, H-5α, H-6aα, H-2β, H-3β, H-4β, H-5β), 3.83-3.56 (m, 8.36H, H-2'α, H-4'α, H-3'β, H-4'β, H-6'aβ, H-6'bβ, H-6bα, H-6aβ, H-6bβ), 3.49-3.38 (m, 7.73H, H-2'β, H-5'β, H-6a'α, OCH₃α, OCH₃β), 3.13 (dd, *J* = 10.65 Hz, *J* = 1.75 Hz 1.03H, H-6b'α); ¹³C NMR (125 MHz, CDCl₃) δ 138.83 (C), 138.76 (C), 138.72 (C), 138.59 (C), 138.56 (C), 138.49 (C), 138.40 (C), 138.26 (C), 138.16 (C), 138.09 (C), 137.89 (C), 129.99 (CH), 129.65 (CH), 128.90 (CH), 128.31 (CH), 128.28 (CH), 128.27 (CH), 128.20 (CH), 128.16 (CH), 128.14 (CH), 128.11 (CH), 127.99 (CH), 127.91 (CH), 127.83 (CH), 127.77 (CH), 127.73 (CH), 127.67 (CH), 127.62 (CH), 127.51 (CH), 127.46 (CH), 127.42 (CH), 127.38 (CH), 127.33 (CH), 127.31 (CH), 127.26 (CH), 102.64 (CH), 99.58 (CH), 98.69 (CH), 98.55 (CH), 84.49 (CH), 82.37 (CH), 82.08 (CH), 80.22 (CH), 78.44 (CH), 77.81 (CH), 77.73 (CH), 77.69 (CH), 76.97 (CH), 76.45 (CH), 75.49 (CH₂), 75.33 (CH₂), 74.94 (CH), 74.87 (CH₂), 74.79 (CH₂), 74.51 (CH), 74.12 (CH₂), 73.83 (CH₂), 73.58 (CH₂), 73.29 (CH₂), 73.26 (CH₂), 73.22 (CH₂), 72.92 (CH₂), 72.81 (CH₂), 70.64 (CH), 70.58 (CH₂), 69.56 (CH) , 69.30 (CH) 69.12 (CH₂), 67.99 (CH₂), 67.54 (CH₂), 55.24 (CH₃), 55.18 (CH₃); HRMS (ESI) calcd. for C₆₂H₆₆O₁₁Na [M+Na]⁺ 1009.4197, found 1009.4497.

### 1- Admantyl 2,3,4,6-tetra-*O*-benzyl-D-galactopyranoside (C52)
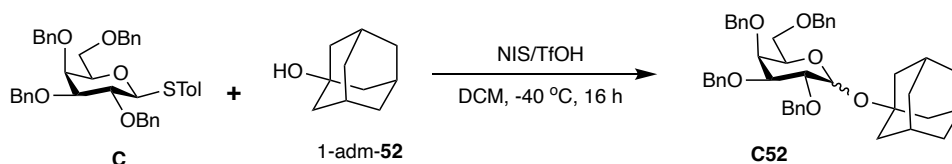


Data for **C52**:IR (CHCl₃) *v* 3826, 2918, 2850, 1453, 1071, 735, 696 cm⁻¹; ¹H NMR (500 MHz, CDCl₃) δ 7.43-7.26 (m, 26.49H, Ar-H), 5.35 (d, *J* = 3.45 Hz, 0.22H, H-1α), 5.03-4.95 (m, 2.38H, CH₂Ph), 4.87 (d, *J* = 11.6 Hz, 0.26H, CH₂Ph), 4.80-4.59 (m, 6.74H, H-1β, CH₂Ph), 4.53-4.43 (m, 2.70H, CH₂Ph), 4.22 (dd, *J* = 6.6 Hz, *J* = 6.6 Hz, 0.26H, H-5α), 4.07-4.00 (m, 0.79H, H-2α, H-3α, H-4α), 3.90 (d, *J* = 2.65 Hz, 1H, H-4β), 3.83 (dd, *J* = 7.85 Hz, *J* = 9.6 Hz, 1.08H, H-2β), 3.63-3.54 (m, 5.10H, H-6aα, H-6bα, H-3β, H-5β, H-6aβ, H-6bβ), 2.16 (m, 4.16H, CH, CH₂), 1.97-1.84 (m, 8.29H, CH₂), 1.74 (d, *J* = 2.45 Hz, 0.29H, CH₂), 1.68-1.62 (m, 8.43H, CH₂); ¹³C NMR (125 MHz, CDCl₃) δ 138.98 (C), 138.84 (C), 138.80 (C), 138.69 (C), 138.66 (C), 138.60 (C), 138.09 (C), 137.99 (C), 128.32 (CH), 128.29 (CH), 128.23 (CH), 128.16 (CH), 128.14 (CH), 128.06 (CH), 128.02 (CH), 127.88 (CH), 127.70 (CH), 127.61 (CH), 127.53 (CH), 127.45 (CH), 127.39 (CH), 127.36 (CH), 127.31 (CH), 96.46 (CH), 90.43 (CH), 82.62 (CH), 79.52 (CH), 79.10 (CH), 76.50 (CH), 75.20 (CH), 75.16 (CH₂), 74.89 (CH), 74.63 (CH₂), 74.33 (CH₂), 74.17 (CH), 73.63 (CH),

73.44 (CH$_2$), 73.35 (CH$_2$), 73.14 (CH), 73.07 (CH$_2$), 72.89 (CH$_2$), 72.83 (CH$_2$), 69.28 (CH$_2$), 69.04 (CH$_2$), 68.63 (CH), 60.32 (CH$_2$), 45.18 (CH$_2$), 42.62 (CH$_2$), 42.36 (CH$_2$), 36.25 (CH$_2$), 30.60 (CH); HRMS (ESI) calcd. for C$_{44}$H$_{50}$O$_6$Na [M+Na]$^+$ 697.3491, found 697.3499.

### Methyl 2,3,4-tri-*O*-benzyl-6-deoxy-L-fucopyranoside (D51)
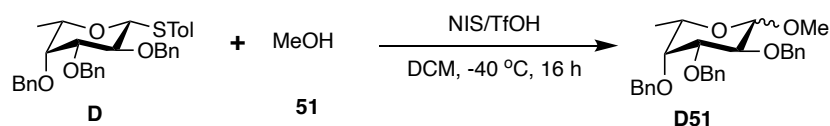


Data for **D51**: IR (CHCl$_3$) $v$ 2910, 2850, 1455, 1103, 735, 698 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.42-7.26 (m, 24H, Ar-H), 5.00-4.69 (m, 9H, CH$_2$Ph), 4.67 (d, *J* = 3.7 Hz, 0.6H, H-1α), 4.25 (d, *J* = 7.7 Hz, 1H, H-1β), 4.05 (dd, *J* = 3.7 Hz, 3.7 Hz, 0.6H, H-2α), 3.94 (dd, *J* = 2.9 Hz, 2.9 Hz, 0.6H, H-3α), 3.85 (q, *J* = 6.5 Hz, 0.6H, H-5α), 3.81 (dd, *J* = 9.7 Hz, 9.7 Hz, 1H, H-2β), 3.65 (d, *J* = 2.0 Hz, 0.6H, H-4α), 3.57 (d, *J* = 2.6 Hz, 1H, H-4β), 3.55 (s, 3H, OCH$_3$β), 3.52 (dd, *J* = 2.9 Hz, 3.0 Hz, 1H, H-3β), 3.47 (q, *J* = 6.4 Hz, 1H, H-5β), 3.37 (s, 1.8H, OCH$_3$α), 1.21 (d, *J* = 6.4 Hz, 3H, CH$_3$β), 1.13 (d, *J* = 6.6 Hz, 1.8H, CH$_3$α); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 138.93 (C), 138.90 (C), 138.60 (C), 138.58 (C), 138.54 (C), 128.39 (CH), 128.33 (CH), 128.28 (CH), 128.21 (CH), 128.14 (CH), 128.06 (CH), 127.62 (CH), 127.51 (CH), 127.46 (CH), 127.43 (CH), 104.85 (CH), 98.78 (CH), 82.46 (CH), 79.48 (CH), 79.38 (CH), 76.74 (CH), 76.31 (CH$_2$), 76.27 (CH$_2$), 75.02 (CH$_2$), 74.79 (CH$_2$), 74.50 (CH$_2$), 73.44 (CH), 73.33 (CH), 73.10 (CH), 70.25 (CH), 66.04 (CH), 56.84 (CH), 55.25 (CH), 16.81 (CH$_3$), 16.56 (CH$_3$); HRMS (ESI) calcd. for C$_{28}$H$_{32}$O$_5$Na [M+Na]$^+$ 471.2133, found 471.2142.

### Methyl 2,3,4-tri-*O*-benzyl-6-*O*-(2,3,4-tri-*O*-benzyl-6-deoxy-L-fucopyranosyl)-α-D-glucopyranoside (D53)
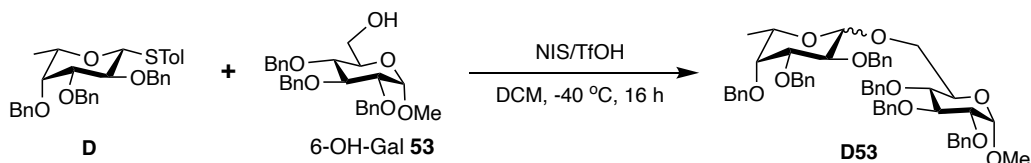


Data for **D53**: IR (CHCl$_3$) $v$ 2921, 2850, 1455, 1095, 1070, 741, 697 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.45-7.24 (m, 37.88H, Ar-H), 5.07-5.03 (m, 3.53H, CH$_2$Ph), 4.97-4.69 (m, 11.31H, H-1'α, CH$_2$Ph), 4.67-4.64 (m, 2.23H, H-1α, H-1β, CH$_2$Ph), 4.49 (d, *J* = 7.6 Hz, 1H, H-1'β), 4.27-4.24 (m, 1H, H-6aβ), 4.12 (dd, *J* = 9.95 Hz, *J* = 3.15 Hz, 0.34H, H-2'α), 4.08-4.04 (m, 1.52H, H-3'α, H-3β), 3.99-3.97 (m, 0.38H, H-5'α), 3.91-3.88 (m, 1.38H, H-2'β, H-6aα), 3.85-3.82 (m, 2.25H, H-3α, H-5β, H-6bβ), 3.76-3.73 (m, 1.61H, H-5α, H-4β, H-6bα ), 3.67 (dd, *J* = 9.75 Hz, *J* = 9 Hz, 0.45H, H-4' α), 3.62-3.58 (m, 2.37H, H-3'β, H-4'β, H-2α, H-4α ), 3.53-3.51 (m, 2.11H, H-5'β, H-2β), 3.40(s, 3.8H, OCH$_3$α, OCH$_3$β), 1.24 (d, *J* = 6.2 Hz, 2.96H, CH$_3$β), 1.18 (d, *J* = 6.35 Hz, 0.76H, CH$_3$α); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 138.85 (C), 138.81 (C), 138.72 (C), 138.70 (C), 138.57 (C), 138.33 (C), 138.20 (C), 138.15 (C), 128.28 (CH), 128.25 (CH), 128.24 (CH), 128.21 (CH), 128.11 (CH), 128.07 (CH), 128.02 (CH), 127.89 (CH), 127.76 (CH), 127.69 (CH), 127.61 (CH), 127.53 (CH), 127.51 (CH), 127.43 (CH), 127.38 (CH), 127.36 (CH), 127.33 (CH), 127.24 (CH), 103.65 (CH), 97.93 (CH), 97.87 (CH), 82.33 (CH), 81.99 (CH), 81.86 (CH), 80.16 (CH),

79.98 (CH), 79.36 (CH), 79.04 (CH), 77.67 (CH), 77.54 (CH), 76.50 (CH), 76.18 (CH), 75.67 (CH$_2$), 75.54 (CH$_2$), 74.95 (CH$_2$), 74.82 (CH$_2$), 74.72 (CH$_2$), 74.49 (CH$_2$), 73.30 (CH$_2$), 73.20 (CH$_2$), 73.07 (CH$_2$), 73.02 (CH$_2$), 72.76 (CH$_2$), 70.22 (CH), 69.97 (CH), 67.38 (CH$_2$), 66.33 (CH$_2$), 66.21 (CH), 54.98 (CH$_3$), 16.77 (CH$_3$), 16.52 (CH$_3$); HRMS (ESI) calcd. for C$_{55}$H$_{60}$O$_{10}$Na [M+Na]$^+$ 903.4076, found 903.4078.
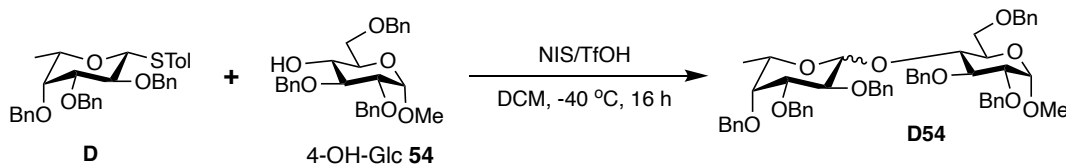
### Methyl 2,3,6-tri-*O*-benzyl-4-*O*-(2,3,4-tri-*O*-benzyl-6-deoxy-L-fucopyranosyl)-α-D-glucopyranoside (D54)



Data for **D54**: IR (CHCl$_3$) *v* 2920, 1454, 1099, 1044, 740, 697 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.48-7.26 (m, 36.3H, Ar-H), 5.13 (d, *J* = 10.9 Hz, 1H, CH$_2$Ph), 5.09 (d, *J* = 3.5 Hz, 1H, H-1'α), 5.07-5.03 (m, 0.47H, CH$_2$Ph), 4.96 (d, *J* = 11.6 Hz, 1H, CH$_2$Ph), 4.92-4.88 (m, 0.53H, H-1'β, CH$_2$Ph), 4.86-4.57 (m, 11.26H, H-1α, H-1β, CH$_2$Ph), 4.43 (q, *J* = 10.85 Hz, 2H, CH$_2$Ph), 4.10-4.03 (m, 2.17H, H-5'α, H-2α, H-4β), 4.00-3.83 (m, 4.54H, H-3'α, H-3α, H-4α, H-5α, H-5β, H-6aβ, H-6bβ), 3.80-3.69 (m, 2.41H, H-2'β, H-6aα, H-6bα), 3.65 (dd, *J* = 9.5 Hz, *J* = 3.55 Hz, 1H, H-2α), 3.60-3.55 (m, 0.60H, H-4'β, H-2β, H-3β), 3.49 (s, 0.55H, OCH$_3$β), 3.47-3.46 (m, 0.12H, H-3'β), 3.42(s, 3.9H, H-4'α, OCH$_3$α), 3.36 (q, *J* = 6.5 Hz, 0.25H, H-5'β), 1.13 (d, *J* = 6.35 Hz, 0.7H, CH$_3$β), 0.77 (d, *J* = 6.4 Hz, 2.89H, CH$_3$α); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 138.79 (C), 138.74 (C), 138.67 (C), 138.57 (C), 138.32 (C), 137.91 (C), 128.39 (CH), 128.32 (CH), 128.28 (CH), 128.20 (CH), 128.16 (CH), 128.08 (CH), 128.05 (CH), 128.01 (CH), 127.91 (CH), 127.81 (CH), 127.69 (CH), 127.55 (CH), 127.49 (CH), 127.45 (CH), 127.41 (CH), 127.38 (CH), 127.32 (CH), 127.23 (CH), 103.60 (CH), 102.81 (CH), 99.76 (CH), 97.75 (CH), 97.68 (CH), 97.50 (CH), 82.60 (CH), 82.20 (CH), 82.11 (CH), 80.48 (CH), 80.13 (CH), 79.87 (CH), 79.61 (CH), 79.32 (CH), 77.51 (CH), 76.24 (CH), 75.45 (CH$_2$), 75.26 (CH$_2$), 75.20 (CH$_2$), 74.63 (CH$_2$), 74.45 (CH$_2$), 74.20 (CH$_2$), 73.75 (CH), 73.66 (CH), 73.28 (CH$_2$), 73.21 (CH$_2$), 73.16 (CH$_2$), 73.10 (CH$_2$), 73.04 (CH$_2$), 72.97 (CH$_2$), 72.62 (CH$_2$), 70.25 (CH), 70.16 (CH), 69.69 (CH), 69.00 (CH$_2$), 68.50 (CH$_2$), 66.56 (CH), 55.15 (CH$_3$), 54.90 (CH$_3$), 16.56 (CH$_3$), 16.14 (CH$_3$); HRMS (ESI) calcd. for C$_{55}$H$_{60}$O$_{10}$Na [M+Na]$^+$ 903.4070, found 903.4079.

### Phenyl 2,3,4-tri-*O*-benzyl-6-deoxy-L-fucopyranoside (D55)
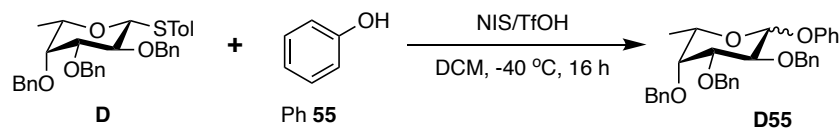


Data for **D55**: IR (CHCl$_3$) *v* 2955, 2924, 2870, 1492, 1453, 1130, 1100, 1028, 755, 734, 695 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.49 (d, *J* = 11.35 Hz 7.35H, Ar-H), 7.44-7.28 (m, 17.94H, Ar-H), 7.15-7.10 (m, 2.89H, Ar-H), 7.05 (dd, *J* = 7.35 Hz, *J* = 7.35 Hz, 1.14H, Ar-H), 5.55 (d, *J* = 3.45 Hz, 1H, H-1α), 5.07 (d, *J* = 11.45 Hz, 1.30H, CH$_2$Ph), 5.02-4.98 (m, 1.16H, H-1β, CH$_2$Ph), 4.92-

4.71 (m, 4.74H, CH$_2$Ph), 4.27-4.19 (m, 1.99H, H-2α, H-3α, H-3β), 4.16 (dd, $J$ = 8.3 Hz, $J$ = 8.3 Hz, 0.18H, H-2β), 4.06 (q, $J$ = 6.45 Hz, 1.01H, H-5α), 3.77 (d, $J$ = 2.15 Hz, 1H, H-4α), 3.67-3.62 (m, 0.48H, H-4β, H-5β), 1.28 (d, $J$ = 6.3 Hz, 0.52H, CH$_3$β), 1.15 (d, $J$ = 6.5 Hz, 3H, CH$_3$α); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 157.56 (C), 157.24 (C), 139.23 (C), 138.85 (C), 138.53 (C), 138.48 (C), 138.36 (C), 137.37 (C), 133.84 (C), 129.73 (CH), 129.34 (CH), 129.30 (CH), 128.49 (CH), 128.27 (CH), 128.22 (CH), 128.17 (CH), 128.12 (CH), 127.87 (CH), 127.59 (CH), 127.58 (CH), 127.54 (CH), 127.52 (CH), 127.47 (CH), 127.43 (CH), 122.29 (CH), 122.02 (CH), 116.90 (CH), 116.85 (CH), 101.85 (CH), 96.29 (CH), 82.35 (CH), 79.22 (CH), 79.03 (CH), 77.62 (CH), 76.10 (CH), 75.27 (CH$_2$), 74.91 (CH$_2$), 74.60 (CH$_2$), 73.29 (CH$_2$), 73.22 (CH$_2$), 73.14 (CH$_2$), 70.62 (CH), 67.14 (CH), 16.89 (CH$_3$), 16.55 (CH$_3$); HRMS (ESI) calcd. for C$_{33}$H$_{34}$O$_5$Na [M+Na]$^+$ 533.2292, found 533.2299.

## Methyl 2,3,6-tri-*O*-benzyl-4-*O*-(2,3,4-tri-*O*-benzyl-6-deoxy-L-fucopyranosyl)-α-D-galactopyranoside (D56)
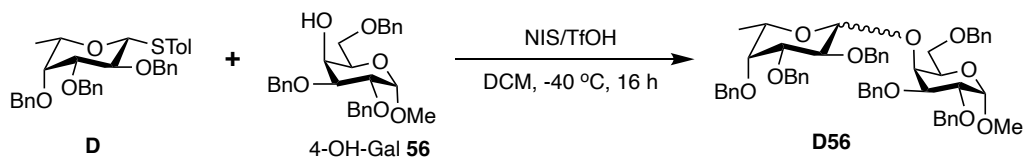


Data for **D56**: IR (CHCl$_3$) $v$ 2916, 2849, 1454, 1099, 1040, 735, 698 cm$^{-1}$; $^1$H NMR (500 MHz, CDCl$_3$) δ 7.51-7.17 (m, 45.9H, Ar-H), 5.84 (d, $J$ = 3.8 Hz, 1H, H-1'α), 5.04-4.68 (m, 13.3H, H-1α, H-1β,CH$_2$Ph), 4.62-4.42 (m, 6.60H, H-1'β, H-5β, CH$_2$Ph), 4.37 (d, $J$ = 11.95 Hz, 1H, CH$_2$Ph), 4.33 (d, $J$ = 1.25 Hz, 1H, H-4α), 4.14 (dd, $J$ = 10.1 Hz, 3.65 Hz, 1H,H-2α,), 4.05-3.90 (m, 6.49H, H-2'α, H-3'α, H-3α, H-5α, H-2'β, H-2β, H-3β, H-4β, H-6aβ), 3.86 (q, $J$ = 6.5 Hz, 1H, H-5'α), 3.65 (d, $J$ = 1.55 Hz, 1H, H-4'α), 3.63-3.61 (m, 2.5H, H-6aα, H-6bα, H-6bβ), 3.56 (d, $J$ = 2.45 Hz, 0.56H, H-4'β), 3.46-3.42 (m, 4.9H, H-3'β, OCH$_3$α, OCH$_3$β ), 3.33 (q, $J$ = 6.15 Hz, 0.5H, H-5'β), 1.19 (d, $J$ = 6.4 Hz, 1.5H, CH$_3$β), 1.05 (d, $J$ = 6.45 Hz, 3H, CH$_3$α); $^{13}$C NMR (125 MHz, CDCl$_3$) δ 139.12 (C), 139.01 (C), 138.87 (C), 138.74 (C), 138.68 (C), 138.52 (C), 138.30 (C), 138.00 (C), 137.88 (C), 128.57 (CH), 128.46 (CH), 128.37 (CH), 128.32 (CH), 128.26 (CH), 128.16 (CH), 128.13 (CH), 128.08 (CH), 127.98 (CH), 127.95 (CH), 127.73 (CH), 127.70 (CH), 127.65 (CH), 127.59 (CH), 127.55 (CH), 127.48 (CH), 127.45 (CH), 127.43 (CH), 127.36 (CH), 127.31 (CH), 127.12 (CH), 127.05 (CH), 126.97 (CH), 103.52 (CH), 98.96 (CH), 98.75 (CH), 96.55 (CH), 82.67 (CH), 79.52 (CH), 79.42 (CH), 78.66 (CH), 77.52 (CH), 76.95 (CH), 76.88 (CH), 76.77 (CH), 76.07 (CH), 75.79 (CH), 74.88 (CH$_2$), 74.71 (CH$_2$), 74.63 (CH$_2$), 73.83 (CH$_2$), 73.49 (CH$_2$), 73.42 (CH$_2$), 73.30 (CH$_2$), 73.12 (CH$_2$), 73.09 (CH$_2$), 72.16 (CH), 71.76 (CH$_2$), 71.12 (CH), 70.46 (CH), 69.99 (CH$_2$), 69.78 (CH), 68.74 (CH), 68.05 (CH$_2$), 66.55 (CH), 55.28 (CH$_3$), 16.92 (CH$_3$), 16.68 (CH$_3$); HRMS (ESI) calcd. for C$_{55}$H$_{60}$O$_{10}$Na [M+Na]$^+$ 903.4075, found 903.4079.

## Methyl 2,3,4-tri-*O*-benzyl-6-*O*-(3,4,6-tri-*O*-benzyl-2-deoxy-D-glucopyranosyl)-α-D-glucopyranoside (E53)
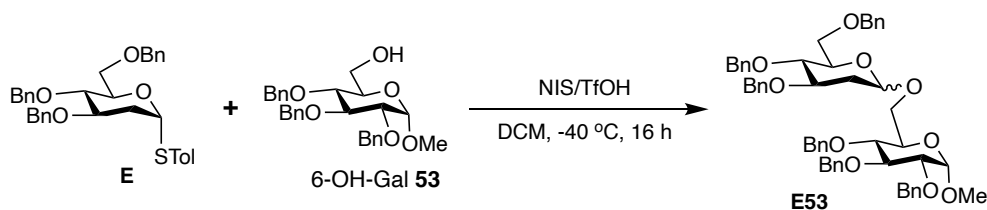
Data for **E53-α**: ¹H NMR (500 MHz, CDCl₃): δ 7.37-7.23 (m, 30H), 7.14-7.12 (m, 2H), 4.98 (d, *J* = 10.6 Hz, 2H), 4.89 (dd, *J* = 23.8, 11.1 Hz, 2H), 4.81- 4.77 (m, 2H), 4.66 (dd, *J* = 13.9, 12.1 Hz, 1H), 4.61-4.59 (m, 2H), 4.55 (t, *J* = 12 Hz, 2H), 4.98 (d, *J* = 10.9 Hz, 1H), 4.91 (d, *J* = 12.1 Hz, 1H), 3.98-3.95 (m, 1H), 3.92 (ddd, *J* = 11.2, 5.0, 3.7 Hz, 1H), 3.81 (dd, *J* = 11.3, 4.3 Hz, 1H), 3.74-3.71 (m, 1H), 3.66 (d, *J* = 9.8 Hz, 1H), 3.61-3.55 (m, 3H), 3.53-3.47 (m, 3H), 3.34 (s, 3H), 2.29 (dd, *J* = 12.9, 5.0 Hz, 1H), 1.70-1.65 (m, 1H); ¹³C NMR (125 MHz, CDCl₃) δ 138.7, 138.6, 138.3, 138.2, 128.4, 128.4, 128.3, 128.3, 128.2, 128.2, 128.0, 128.0, 127.8, 127.7, 127.7, 127.5, 127.4, 114.7, 97.9, 97.8, 82.2, 8.0, 78.1, 77.8, 75.7, 74.8, 74.8, 73.4, 73.2, 71.7, 70.9, 69.7, 68.7, 65.7, 55.1, 35.2, 29.7; LRMS (ESI) clacd. for C₅₅H₆₀O₁₀Na [M+Na]⁺ 904.0517, found 903.4032. Data for **E53-β**: ¹H NMR (500 MHz, CDCl₃): δ 7.34-7.26 (m, 30H), 7.20-7.19 (m, 2H), 4.99 (d, *J* = 10.8 Hz, 1H), 4.88 (d, *J* = 1.5 Hz, 1H), 4.86 (d, *J* = 2.3 Hz, 1H), 4.87-4.77 (m, 2H), 4.65 (d, *J* = 11.8 Hz, 2H), 4.60-4.51 (m, 6H), 4.16 (dd, *J* = 9.6, 1.5 Hz, 1H), 4.07 (dd, *J* = 10.6, 1.7 Hz, 1H), 3.99 (t, *J* = 9.2 Hz, 1H), 3.75-3.71 (m, 2H), 3.44-3.41 (m, 1H), 3.35 (m, 4H), 2.17-2.14 (m, 1H), 1.42 (dd, *J* = 22.1, 9.7 Hz, 1H); ¹³C NMR (125 MHz, CDCl₃) δ 138.8, 138.4, 138.4, 138.4, 138.3, 138.1, 128.4, 128.4, 128.3, 128.3, 128.1, 128.0, 127.9, 127.8, 127.7, 127.6, 127.6, 127.4, 100.0, 98.0, 82.2, 79.8, 79.3, 78.2, 77.4, 75.7, 74.9, 74.7, 73.4, 73.3, 71.4, 69.6, 69.5, 67.6, 55.1, 36.5; LRMS (ESI) clacd. for C₅₅H₆₀O₁₀Na [M+Na]⁺ 903.4084, found 903.4086.

**Methyl        2,3,6-tri-*O*-benzyl-4-*O*-(3,4,6-tri-*O*-benzyl-2-deoxy-α-D-glucopyranosyl)-α-D-glucopyranoside (E54)**
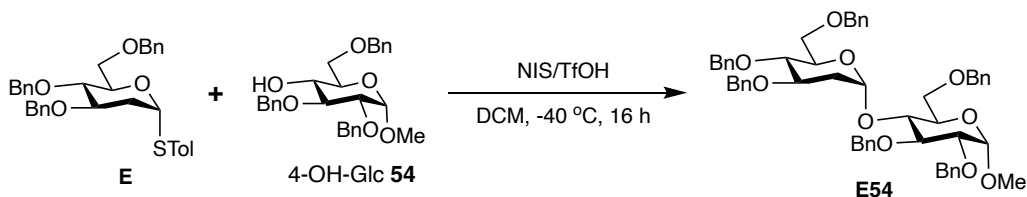


Data for **E54**: [α]²⁶_D = +41.9º (*c* 1.0, CHCl₃); ¹H NMR (400 MHz, CDCl₃) δ 7.36-7.18 (m, 30H, PH), 5.44 (d, 1H, *J* = 2.7 Hz, H-1'), 5.05 (d, 1H, CH₂Ph, *J* = 11.2 Hz), 4.87 (d, 1H, CH₂Ph, *J* = 10.9 Hz), 4.74 (d, 1H, CH₂Ph, *J* = 12.0 Hz), 4.66-4.61 (m, 3H, CH₂Ph), 4.57-4.54 (m, 4H, CH₂Ph), 4.50 (d, 1H, CH₂Ph, *J* = 11.0 Hz), 4.46 (d, 1H, CH₂Ph, *J* = 12.0 Hz), 4.38 (d, 1H, CH₂Ph, *J* = 12.1 Hz), 3.92-3.84 (m, 2H), 3.76-3.67 (m, 5H), 3.65-3.53 (m, 3H), 3.44 (dd, 1H, *J* = 10.3 and 1.7 Hz), 3.41 (s, 3H), 2.09 (ddd, 1H, *J* = 12.9, 5.0, and 1.1 Hz, H-2eq'), 1.58 (td, 1H, *J* = 12.9 and 3.8 Hz, H-2ax'); ¹³C NMR (100 MHz, CDCl₃) δ 138.89 (C), 138.84 (C), 138.8 (C), 138.5 (C), 138.3 (C),

138.2 (C), 128.6 (CH), 128.57 (CH), 128.5 (CH), 128.4 (CH), 128.2 (CH), 128.0 (CH), 127.69 (CH), 127.6 (CH), 127.5 (CH), 99.5 (CH), 97.9 (CH), 82.3 (CH), 80.3 (CH), 78.3 (CH), 77.4 (CH), 76.4 (CH), 75.5 ($CH_2$), 74.9 ($CH_2$), 73.6 ($CH_2$), 73.4 ($CH_2$), 73.3 ($CH_2$), 72.0 (CH), 71.8 ($CH_2$), 70.0 (CH), 69.7 ($CH_2$), 68.9 ($CH_2$), 55.3 ($CH_3$), 35.9 ($CH_2$); IR ν 3039, 2935, 1456, 1367, 1278, 1099, 1047, 750, 738 cm$^{-1}$; HRMS (ESI) calcd for $C_{55}H_{60}NaO_{10}$ [M+Na]$^+$ 903.4084, found 903.4070 *(13)*.

# 4 NMR Spectra
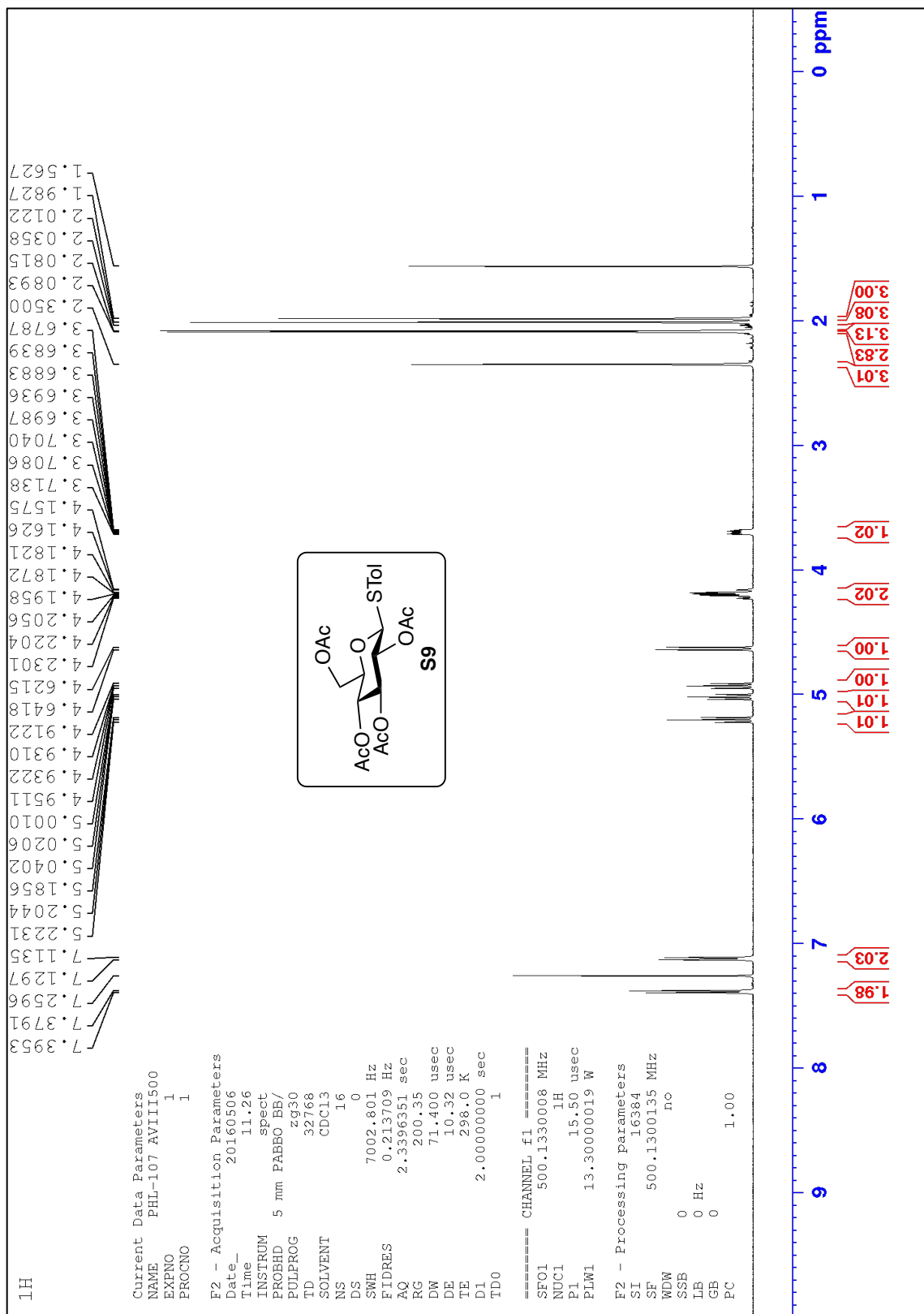


The $^1$H Spectrum in CDCl$_3$ of Compound S9.

The 1D-TOCSY Spectrum in CD$_2$Cl$_2$ of Compound S9.

1D Selective Gradient TOCSY
freq: 4.659ppm

3.679
3.684
3.688
3.694
3.699
3.704
3.709
3.714

4.158
4.163
4.182
4.187
4.196
4.206
4.220
4.230

4.622
4.642

4.912
4.931
4.932
4.951
5.001
5.021
5.040
5.186
5.204
5.223

1H

**The 1D-TOCSY Spectrum in CD$_2$Cl$_2$ of Compound S9.**

DEPT135

DEPT90

13C

20.57
20.73
20.76
21.18

62.12

68.18
69.89

74.01
75.73

85.83

127.52
129.67

133.83

138.80

169.24
169.38
170.20
170.58

ppm
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190

H1

Peak list (ppm):

2.3198
3.5035
3.6269
3.6307
3.6540
3.6572
3.7404
3.7525
3.7635
3.7812
3.7907
3.8025
3.8077
3.8180
3.8273
3.8496
3.9148
3.9281
3.9400
3.9532
4.3386
4.3641
4.3777
4.4257
4.4557
4.5262
4.5544
4.5712
4.5822
4.5992
4.6048
4.6118
4.7685
4.7961
4.8050
4.8346
4.8455
4.8617
4.8883
4.9056
4.9322
5.5311
5.5444
7.0477
7.0671
7.0869
7.1133
7.1871
7.1927
7.2104
7.2157
7.2595
7.2685
7.2972
7.3070
7.3152
7.3248
7.3418
7.3486
7.3672
7.3862
7.4062
7.4889
7.5089

Integrals: 4.47, 0.41, 2.56, 4.20, 1.08, 2.67, 3.73, 4.61, 1.00, 3.29, 3.44, 22.90, 1.02

**The ¹H Spectrum in CDCl₃ of Compound A.**

(Structure labeled A: pyranose ring with OBn, BnO, BnO substituents, O, N₃, and STol group)

80

DEPT135

21.0975

64.1927
64.9485
68.4223
68.7852
71.7662
73.4564
75.0628
75.7002
75.8651
76.6819
76.9999
77.3176
77.5750
78.3762
79.3615
81.7998
85.1022
85.9314
87.6165
127.5645
127.6965
127.7349
127.8322
127.9216
127.9830
128.1148
128.2014
128.3532
128.4509
128.4786
129.6504
129.7564
129.8338
132.6530
134.3128
137.6530
137.7361
137.8940
138.2854
138.6938

DEPT90

C13

190  180  170  160  150  140  130  120  110  100  90  80  70  60  50  40  30  20  ppm

The $^{13}$C Spectrum in CDCl$_3$ of Compound A.

81

The $^1$H Spectrum in CDCl$_3$ of Compound B.

**The 1D-TOCSY Spectrum in CD$_2$Cl$_2$ of Compound B.**

1D Selective Gradient TOCSY
freq: 4.677ppm

1H

B

OBn
O
STol
OBn
BnO
BnO

3.458
3.466
3.469
3.475
3.485
3.489
3.494
3.498
3.593
3.611
3.630
3.663
3.680
3.698
3.704
3.714
3.726
3.735
3.762
3.766
3.784
3.788

4.517
4.541
4.583
4.603
4.607
4.622
4.642
4.728
4.749
4.804
4.826
4.833
4.855
4.876
4.897

5.320

2.04
1.04
0.83
0.49
0.70
1.01
0.05
1.01
1.63
1.43
1.04
1.27
2.79

ppm
3.2  3.4  3.6  3.8  4.0  4.2  4.4  4.6  4.8  5.0  5.2  5.4  5.6  5.8

The $^{13}$C Spectrum in CD$_2$Cl$_2$ of Compound B.

The $^1$H Spectrum in CD$_2$Cl$_2$ of Compound C.

The chemical shift labels on the spectrum (reading top to bottom):

3.594
3.599
3.612
3.620
3.630
3.639
3.650
3.787
3.806
3.824
3.981
3.986

4.424
4.447
4.477
4.501
4.560
4.579
4.583
4.598
4.698
4.721
4.730
4.747
4.751
4.763
4.770
4.784
4.918
4.940

1D Selective Gradient TOCSY
freq: 4.641ppm

1H

Integration values (bottom axis):
0.51
3.54
1.04
1.03
2.06
1.07
1.00
4.13
1.03

ppm axis: 3.2  3.4  3.6  3.8  4.0  4.2  4.4  4.6  4.8  5.0  5.2  5.4  5.6  5.8

Structure label: C

**The 1D-TOCSY Spectrum in CD$_2$Cl$_2$ of Compound C.**

The $^{13}$C Spectrum in CD2Cl$_2$ of Compound C.

21.16

53.36
53.58
53.80
54.01
54.23

69.43
72.97
73.79
74.44
75.04
75.77
77.61
77.86
84.47
88.32

127.86
127.92
127.96
128.02
128.22
128.53
128.57
128.70
129.92
130.75
132.28
137.73
138.63
138.86
139.07
139.25

DEPT135

DEPT90

13C

The $^1$H Spectrum in CD$_2$Cl$_2$ of Compound S16.

The $^1$H Spectrum in CD$_2$Cl$_2$ of Compound D.

**The $^{13}$C Spectrum in CDCl$_3$ of Compound E.**

Peak labels (ppm):
2.311
3.620
3.642
3.666
3.691
3.695
3.813
3.823
3.839
3.850
3.959
3.971
3.981
3.988
3.993
4.000
4.010
4.022
4.316
4.320
4.326
4.330
4.341
4.345
4.350
4.355
4.451
4.481
4.546
4.573
4.602
4.632
4.646
4.675
4.686
4.715
4.905
4.932
5.621
5.634
7.046
7.060
7.214
7.229
7.234
7.259
7.288
7.295
7.315
7.323
7.330
7.348
7.358
7.364

Current Data Parameters
NAME      FS-2-deoxy-AC2-a-AVIII-400
EXPNO     1
PROCNO    1

F2 - Acquisition Parameters
Date_         20231108
Time          14.29
INSTRUM       spect
PROBHD   5 mm PABBO BB-
PULPROG       zg30
TD            32768
SOLVENT       CDCl3
NS            12
DS            0
SWH           7211.539 Hz
FIDRES        0.220079 Hz
AQ            2.2719147 sec
RG            203
DW            69.333 usec
DE            6.50 usec
TE            300.0 K
D1            2.00000000 sec
TD0           1

======== CHANNEL f1 ========
NUC1          1H
P1            15.80 usec
PL1           -1.50 dB
PL1W          14.36994267 W
SFO1          400.1328009 MHz

F2 - Processing parameters
SI            32768
SF            400.1300098 MHz
WDW           EM
SSB           0
LB            0 Hz
GB            0
PC            1.00

Integration values (ppm): 1.01, 3.15, 1.06, 2.13, 1.05, 0.99, 0.99, 1.03, 1.79, 0.64, 1.87, 1.15, 1.00, 2.17, 2.16, 15.34

Structure label: E (OBn, BnO, BnO, STol)

The $^{13}$C Spectrum in CDCl$_3$ of Compound E.

The $^1$H Spectrum in CDCl$_3$ of Compound 54.

DEPT 135

DEPT 90

13C

55.20

69.42
69.84
70.66
73.11
73.53
75.38
79.54
81.40

98.13

127.58
127.80
127.91
127.96
128.08
128.31
128.43
128.54
137.94
138.00
138.74

OBn
O
BnO OMe **54**
HO
BnO
4-OH-Glc **54**

200  190  180  170  160  150  140  130  120  110  100  90  80  70  60  50  40  30  20  ppm
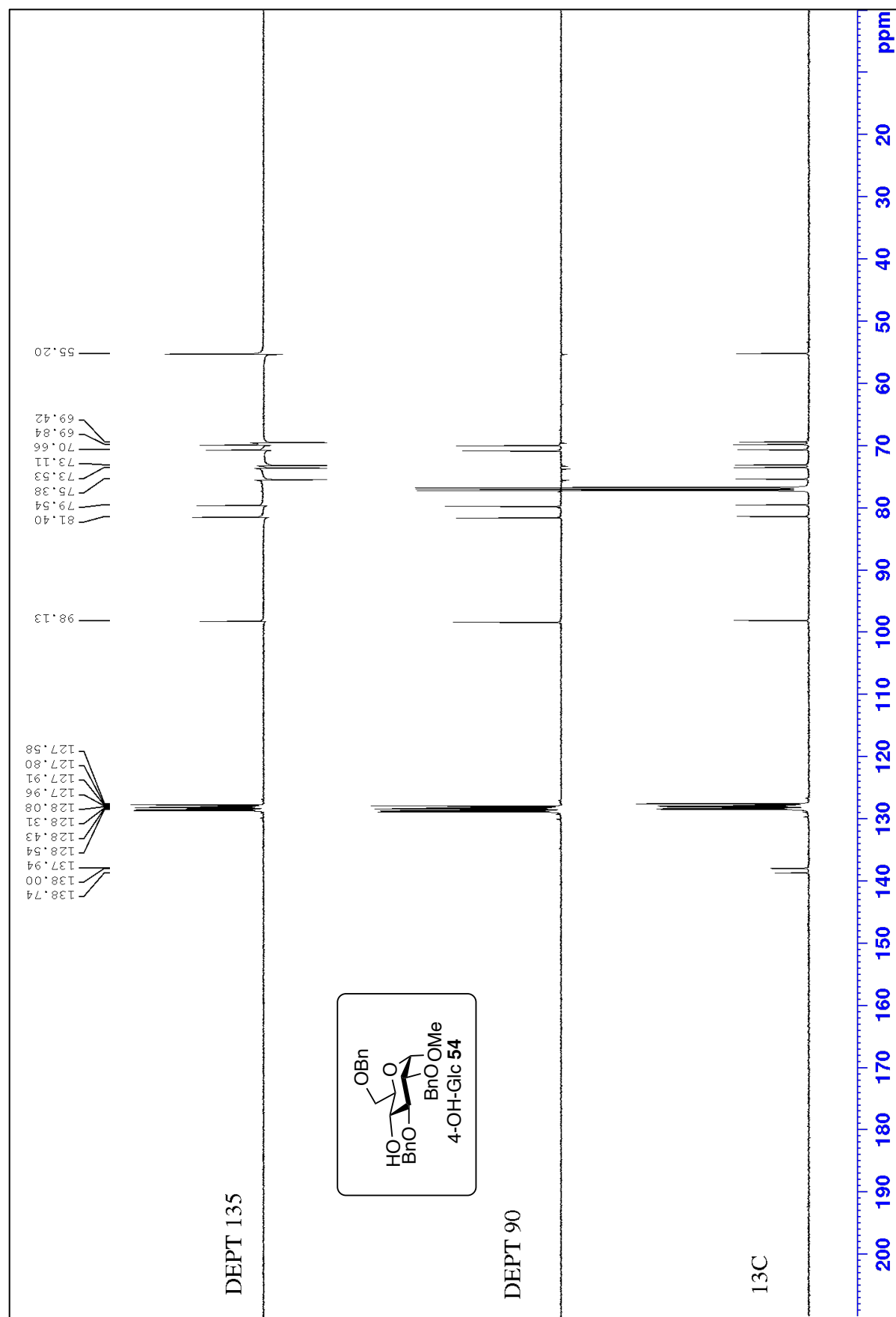
**The $^{13}$C Spectrum in CDCl$_3$ of Compound 54.**

The $^1$H Spectrum in CDCl$_3$ of Compound 53.
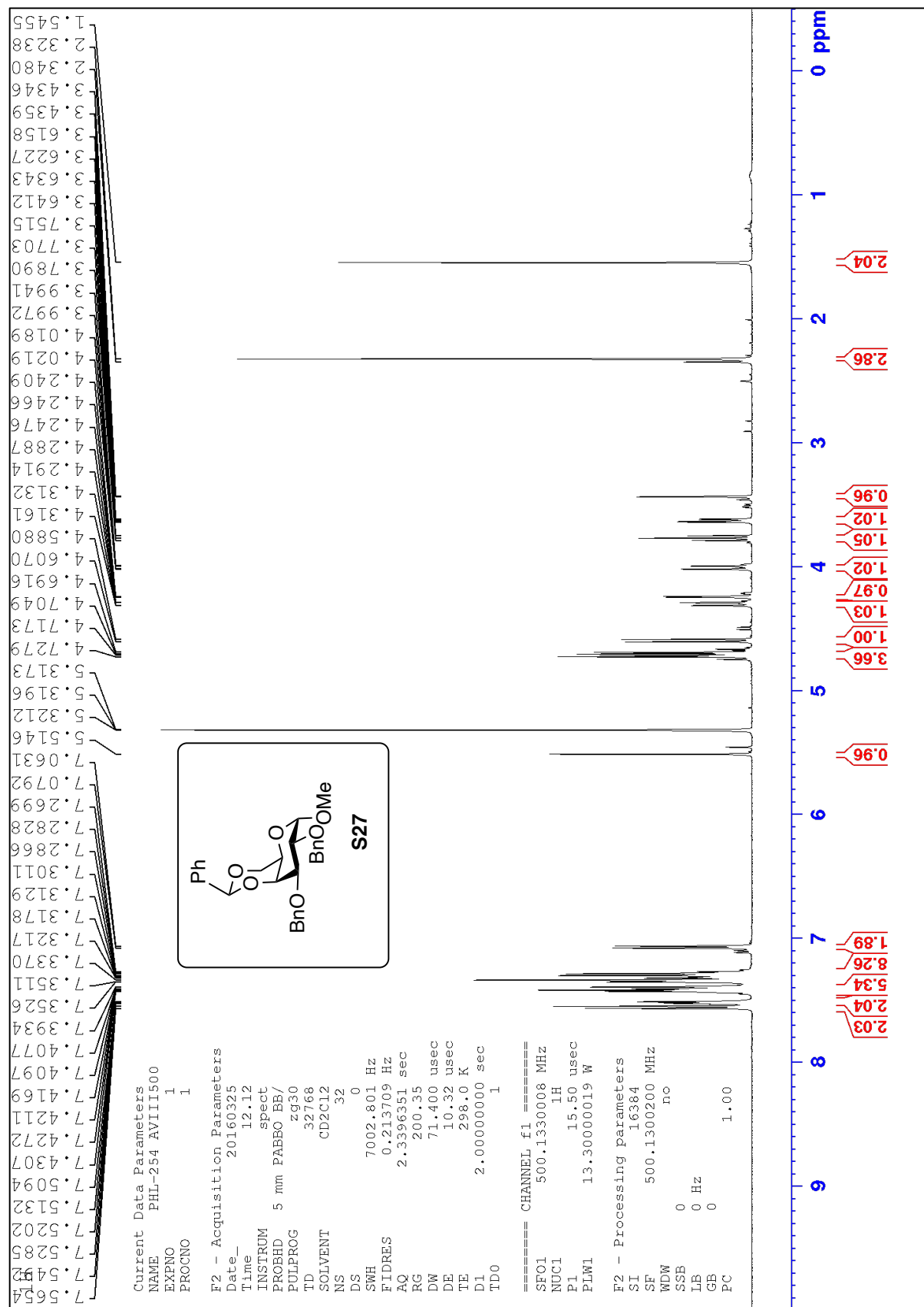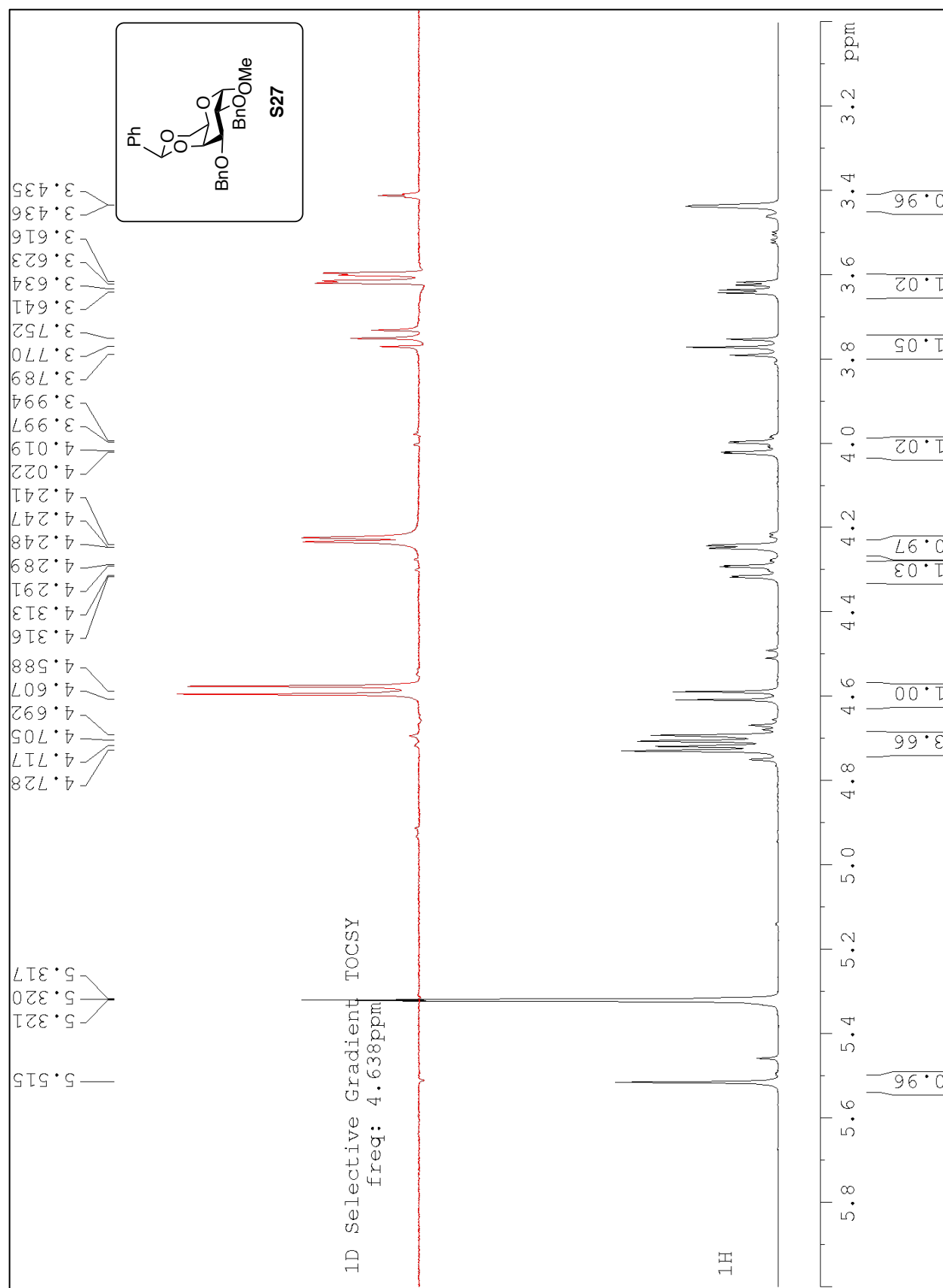
BnO
BnO
OH
O
BnO
OMe
6-OH-Glc 53
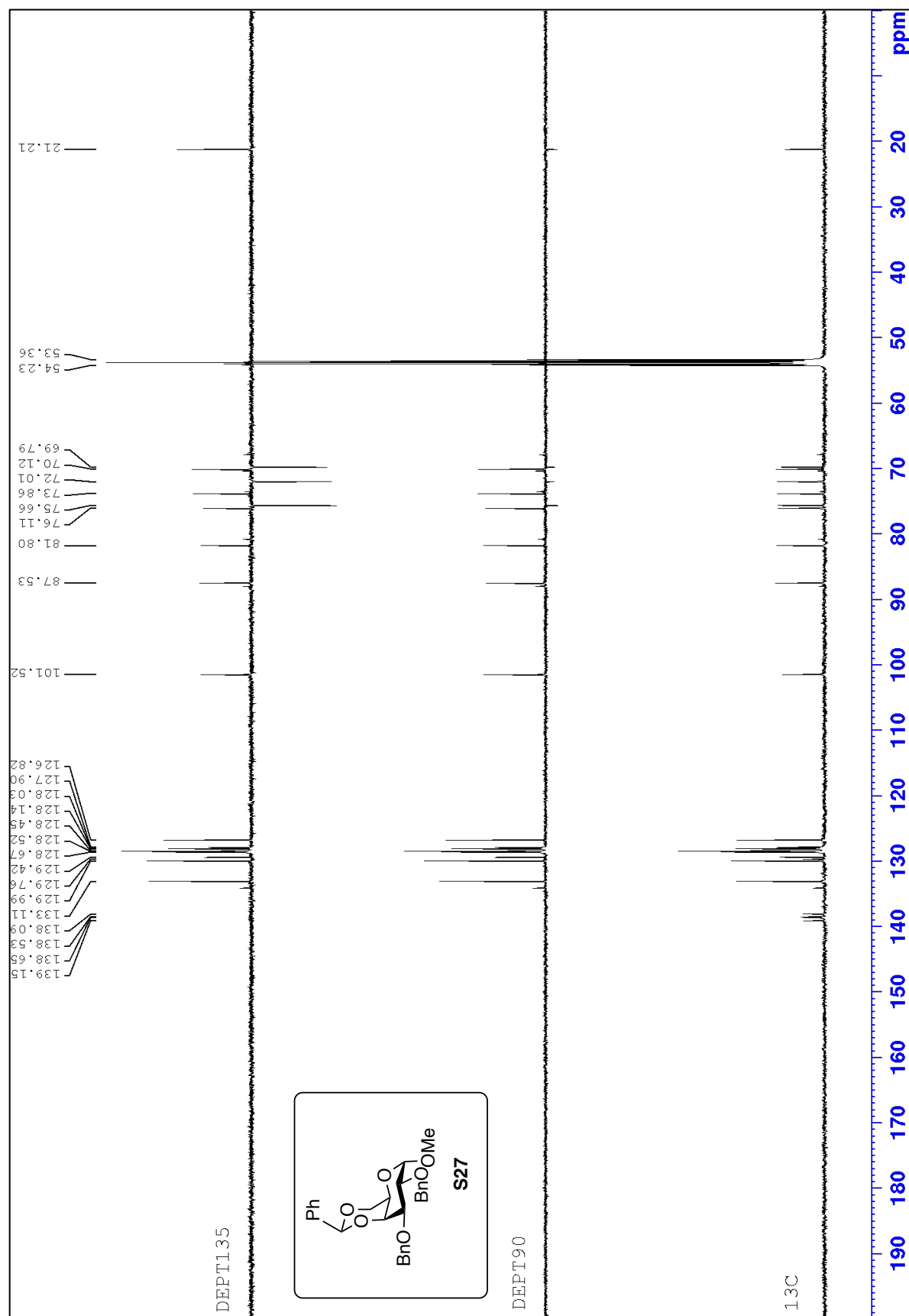
Current Data Parameters
NAME          ypl-6-OH 5.26
EXPNO               1
PROCNO              1

F2 - Acquisition Parameters
Date_          20140526
Time              17.18
INSTRUM           spect
PROBHD     5 mm PABBO BB/
PULPROG           zg30
TD                32768
SOLVENT           CDCl3
NS                   20
DS                    0
SWH          6996.269 Hz
FIDRES      0.213509 Hz
AQ          2.3418698 sec
RG                98.66
DW               71.467 usec
DE                 6.50 usec
TE                298.0 K
D1          2.00000000 sec
TD0                   1

======== CHANNEL f1 ========
SFO1        500.1323476 MHz
NUC1                 1H
P1               15.50 usec
PLW1       13.30000019 W

F2 - Processing parameters
SI                32768
SF          500.1300243 MHz
WDW                  no
SSB                   0
LB                0 Hz
GB                    0
PC                 1.00

1.6954
3.3542
3.4787
3.4858
3.4951
3.4979
3.5051
3.5139
3.5324
3.6231
3.6306
3.6363
3.6427
3.6507
3.6558
3.6648
3.6729
3.6883
3.6963
3.7423
3.7475
3.7658
3.7710
3.9802
3.9987
4.0173
4.5567
4.5638
4.6209
4.6419
4.6650
4.7769
4.8011
4.8183
4.8401
4.8638
4.8859
4.8968
4.9066
4.9906
7.2401
7.2643
7.2739
7.2768
7.2813
7.2877
7.2951
7.2989
7.3048
7.3197
7.3260
7.3279
7.3336
7.3357
7.3453
7.3602

1.61
3.02
2.05
1.04
0.99
1.00
1.00
1.00
2.04
1.03
1.02
1.01
1.00
14.90

The $^{13}$C Spectrum in CDCl$_3$ of Compound 53.

DEPT 135

55.1495
61.8057
70.6207
73.3885
74.4891
75.1720
76.7441
76.9986
77.2525
77.3415
79.9213
81.9073

98.1305

127.5866
127.8381
127.9295
127.9918
128.0842
128.3683
128.4382
138.0695
138.6786

DEPT 90

C 13

OH
O
BnOOMe
BnO
BnO

6-OH-Glc 53

95

The $^1$H Spectrum in CD$_2$Cl$_2$ of Compound S27.

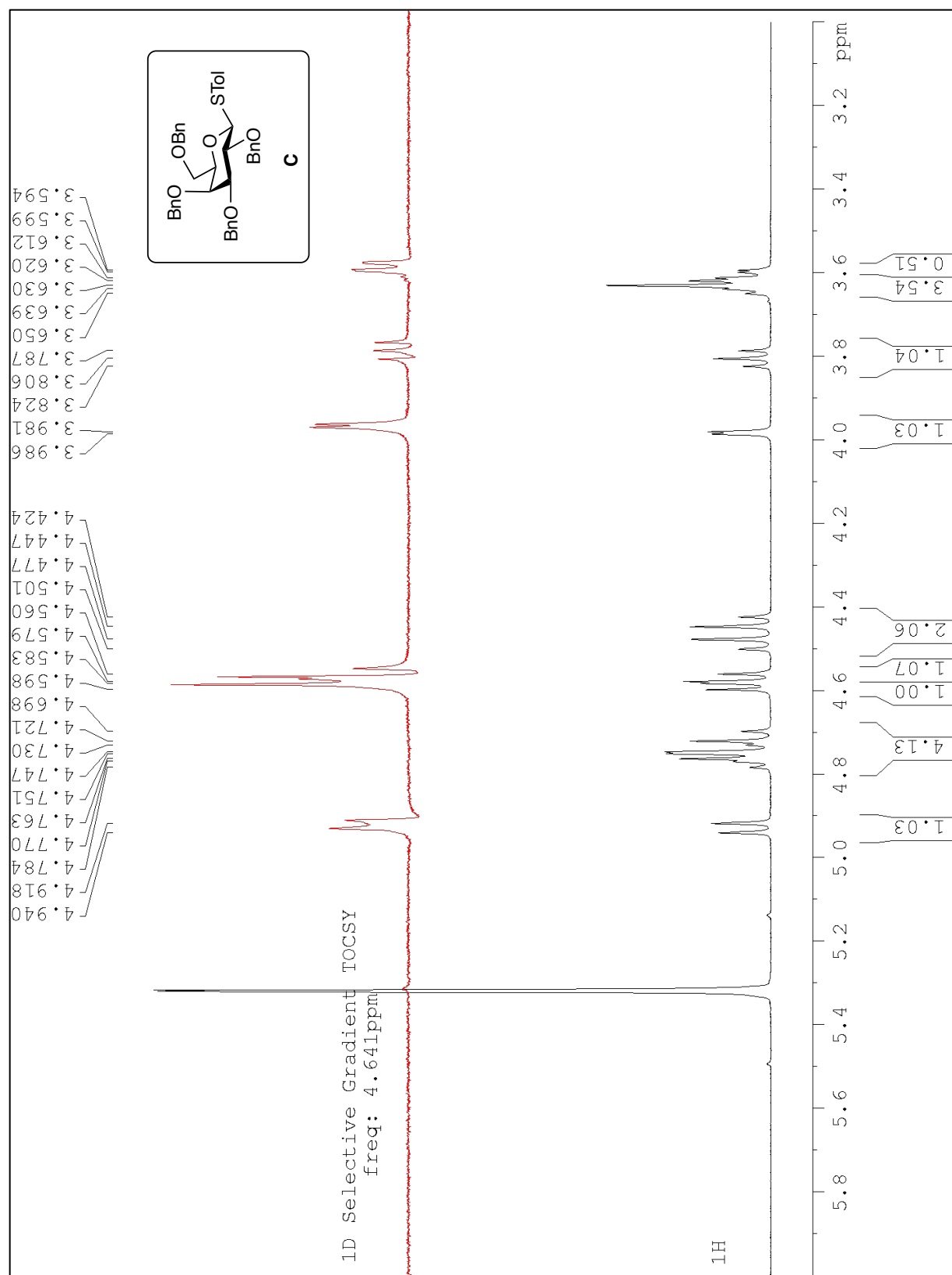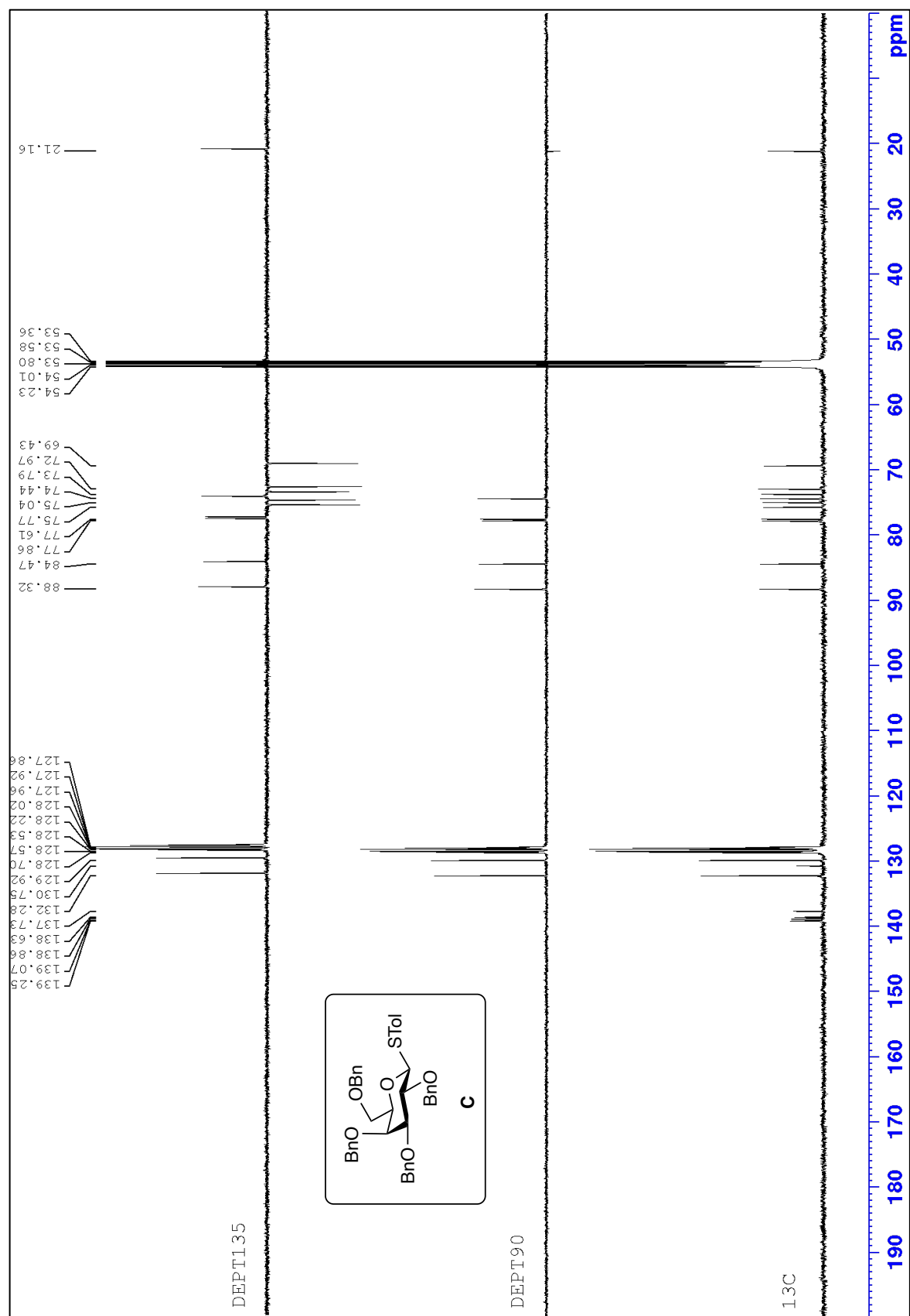**The 1D-TOCSY Spectrum in CD$_2$Cl$_2$ of Compound S27.**

Peak labels: 3.435, 3.436, 3.616, 3.623, 3.634, 3.641, 3.752, 3.770, 3.789, 3.994, 3.997, 4.019, 4.022, 4.241, 4.247, 4.248, 4.289, 4.291, 4.313, 4.316, 4.588, 4.607, 4.692, 4.705, 4.717, 4.728, 5.317, 5.320, 5.321, 5.515

1D Selective Gradient TOCSY
freq: 4.638ppm

1H

S27

Integration values: 0.96, 1.02, 1.05, 1.02, 0.97, 1.03, 1.00, 3.66, 0.96

ppm axis: 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.2, 5.4, 5.6, 5.8

The $^{13}$C Spectrum in CD$_2$Cl$_2$ of Compound S27.

DEPT135

21.21

53.36
54.23

69.79
70.12
72.01
73.86
75.66
76.11
81.80

87.53

101.52

126.82
127.90
128.03
128.14
128.45
128.52
128.67
129.42
129.76
129.99
133.11
138.09
138.53
138.65
139.15

DEPT90

13C

The $^1$H Spectrum in CDCl$_3$ of Compound 56.

**The 1D-TOCSY Spectrum in CDCl₃ of Compound 56.**

Structure label:
HO—OBn
BnO—BnO—OMe
**4-OH-Glc 56**

1D Selective Gradient TOCSY
freq: 4.712 ppm

1D Selective Gradient TOCSY
freq: 3.929 ppm

1H

Peak labels (ppm):
3.3899
3.6651
3.6755
3.6819
3.6923
3.7276
3.7365
3.7443
3.7534
3.8697
3.8707
3.8711
3.8728
3.8739
3.8920
3.8931
3.9025
3.9034
3.9114
3.9127
4.0631
4.0646

4.5484
4.5684
4.5890
4.6089
4.6623
4.6830
4.6877
4.7051
4.7244
4.7888
4.8093
4.8302

Integration values:
3.07
1.00
1.08
1.88
1.09
1.00
1.10
1.12
3.28
2.06

ppm axis: 3.5, 4.0, 4.5, 5.0, 5.5, 6.0

DEPT 135

55.30
68.10
68.32
69.58
72.73
73.51
73.60
75.69
76.79
77.00
77.21
77.59
98.57
127.64
127.77
127.79
127.82
128.02
128.36
128.46
137.96
138.14
138.34

HO    OBn

O

BnO    OOMe

BnO

4-OH-Glc **56**

DEPT 90

13C

200  190  180  170  160  150  140  130  120  110  100  90  80  70  60  50  40  30  20  ppm

The $^{13}$C Spectrum in CDCl$_3$ of Compound 56.

The $^1$H Spectrum in CDCl$_3$ of Compound A51-β.

102

29.6888

57.0876
66.3062
68.5875
73.5370
75.0261
75.0638
75.5307
76.7448
76.9991
77.2532
77.8097
83.2346

102.9181

127.6635
127.7753
127.8429
127.8681
128.0414
128.3706
128.4290
128.4445
137.8689
137.9082
138.0120

DEPT135

DEPT90

C13

**OBn**
**BnO**
**BnO**
**O**
**O**
**CH₃**
**N₃**
**A51**
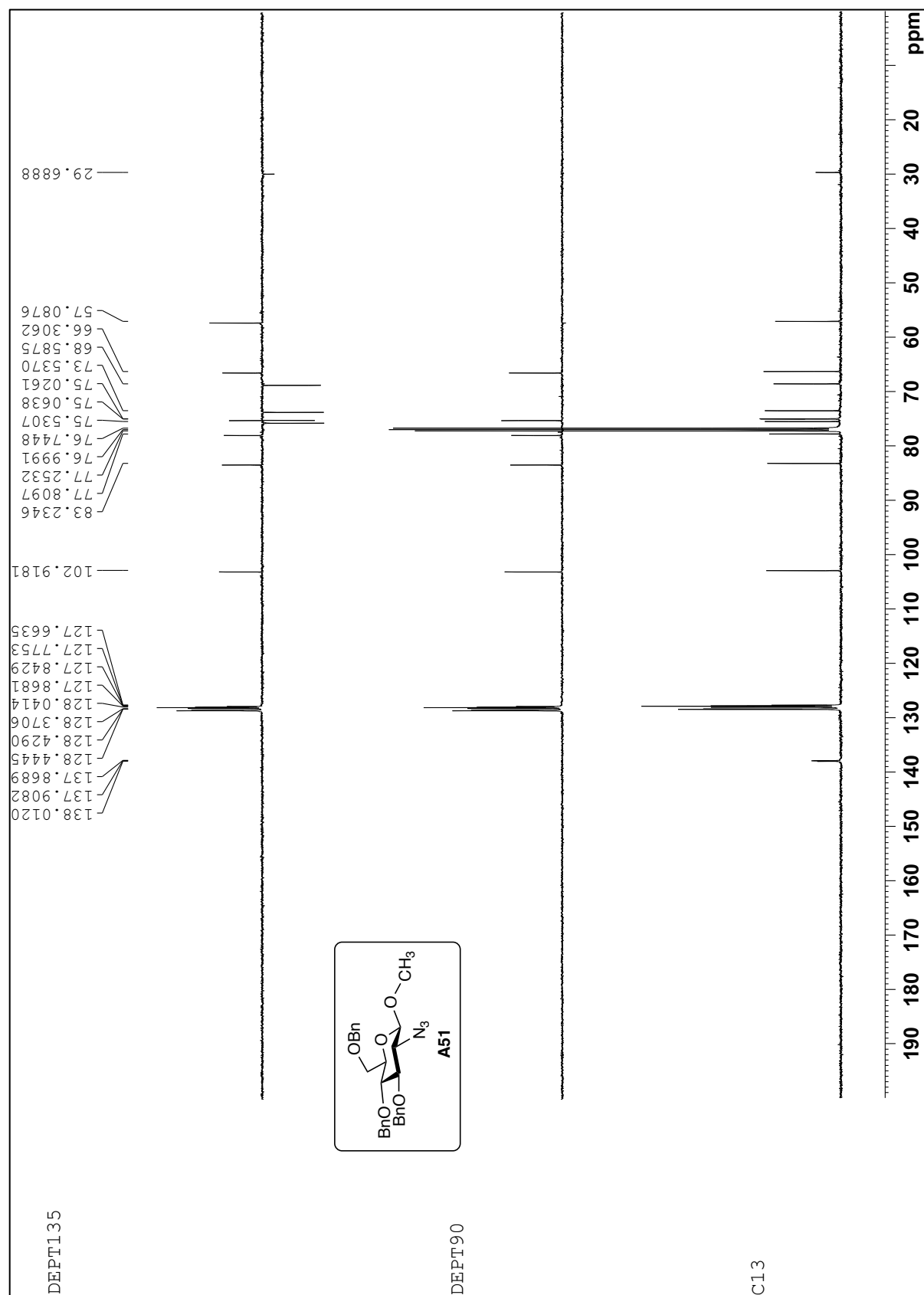
The $^{13}$C Spectrum in CDCl$_3$ of Compound A51-β.

The $^1$H Spectrum in CDCl$_3$ of Compound A53-β.

DEPT135

DEPT90

C13

55.1457
66.3289
68.3823
68.6441
69.7002
73.3744
74.7757
74.8829
74.9436
75.1554
75.5086
75.6323
76.6802
76.9982
77.3163
77.7013
77.7279
79.7825
82.0321
83.2244
98.1243
102.0683
127.5000
127.5543
127.6010
127.7140
127.8032
127.8282
127.9075
128.0066
128.0945
128.3025
128.3780
137.8094
138.0240
138.1143
138.3963
138.7556

A53

The $^{13}$C Spectrum in CDCl3 of Compound A53-β

The $^1$H Spectrum in CDCl$_3$ of Compound A54-α.

The $^{13}$C Spectrum in CDCl3 of Compound A54-α.

DEPT135

DEPT90

C13

55.2910
63.3470
67.9258
69.3210
69.5003
71.4261
73.2827
73.3320
73.5280
74.9081
74.9826
75.3115
76.7487
77.0024
77.2567
78.1053
80.1279
80.5371
81.9912
97.7121
97.7465
127.2684
127.4178
127.4687
127.5424
127.6744
127.7267
127.7935
127.9594
128.0925
128.1494
128.2850
128.3361
128.4258
128.4785
137.8202
137.9650
138.1059
138.1729
138.6908

BnO OBn
O
BnO OMe
OBn
O
N₃
BnO
BnO
BnO
A54-α

ppm
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20
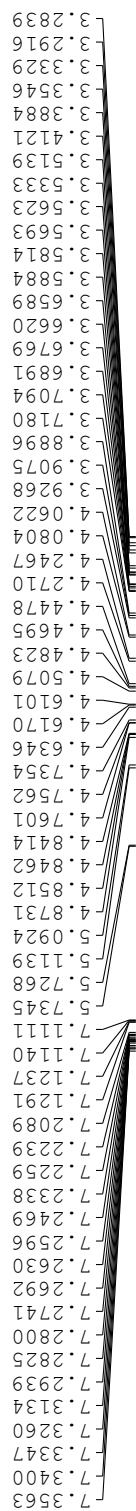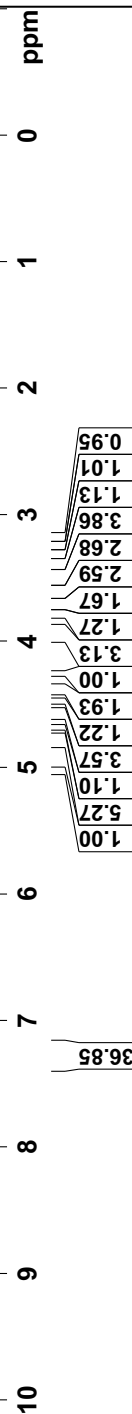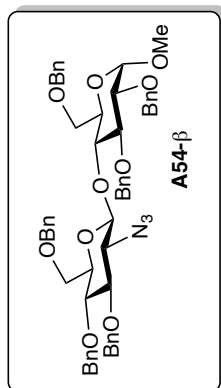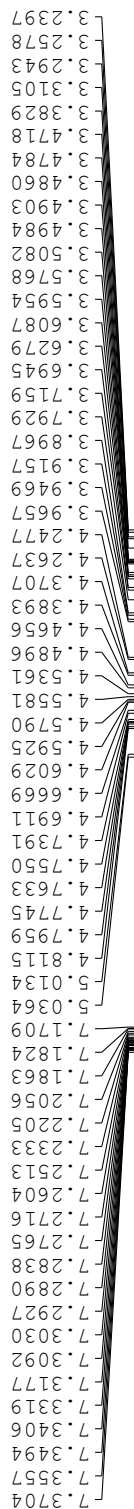
The $^1$H Spectrum in CDCl$_3$ of Compound A54-β.
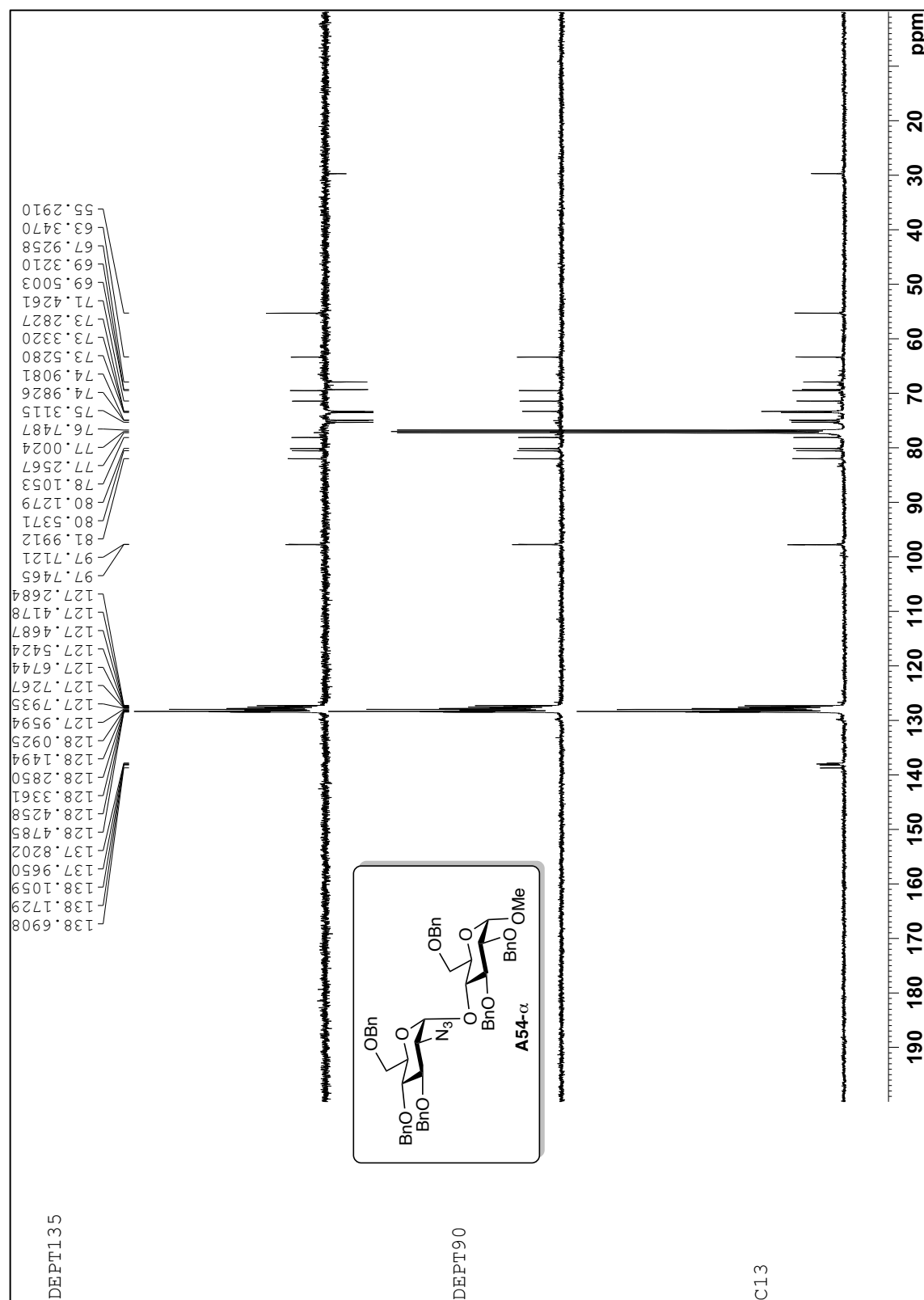
The $^{13}$C Spectrum in CDCl$_3$ of Compound A54-β.

55.2841
66.9141
68.2346
68.5662
69.6999
73.3352
73.4817
73.5329
74.7591
75.1905
75.2655
75.4230
76.7452
76.9986
77.2526
77.8919
79.0982
80.3178
83.3075
98.3103
100.9285
127.0403
127.3834
127.5496
127.6731
127.6924
127.7396
127.8468
127.9765
128.0015
128.0690
128.1358
128.2517
128.3321
128.3777
128.4432
128.5082
137.8234
137.8936
138.0427
138.3057
138.3454
139.5177

DEPT135

DEPT90

13

A54-β

3.579
3.584
3.599
3.605
3.705
3.708
3.775
3.829
3.836

1

0 ppm

**A55**

63.117
66.049
68.012
68.479
71.293
73.392
73.424
74.991
75.230
75.387
75.535
76.741
76.995
77.249
77.488
77.928
80.028
82.921
96.759
100.487
116.688
116.941
122.747
122.990
127.566
127.637
127.690
127.741
127.809
127.883
127.978
128.278
128.309
128.357
128.376
128.400
128.421
129.482
129.537
129.683
134.259
137.640
137.722
137.806
137.826
137.893
156.255
157.011

DEPT 135

DEPT 90

$^{13}$C

190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 ppm

**The $^{13}$C Spectrum in CDCl$_3$ of Compound A55**

The $^1$H Spectrum in CDCl$_3$ of Compound A56

The $^{13}$C Spectrum in CDCl$_3$ of Compound A56

55.251
55.304
66.511
68.757
68.789
69.269
70.105
73.069
73.094
73.156
73.241
73.326
73.375
73.489
73.701
74.253
74.681
74.712
74.781
74.949
75.152
75.260
75.387
76.455
77.556
78.255
80.305
83.020
98.414
98.521
98.715
101.561
127.091
127.242
127.347
127.394
127.466
127.570
127.627
127.652
127.705
127.737
127.757
127.850
127.903
127.967
128.082
128.204
128.235
128.258
128.294
128.362
128.404
137.845
138.041
138.479
138.517
138.853

A56

DEPT 135

DEPT 90

$^{13}$C

190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 ppm

The $^1$H Spectrum in CDCl$_3$ of Compound B51-$\alpha$.

The $^{13}$C Spectrum in CDCl$_3$ of Compound B51-$\alpha$.

29.6895
55.1580
68.4380
70.0187
73.3883
73.4671
75.0238
75.7532
76.7868
76.9987
77.2102
79.7975
82.1176
98.1963
127.5798
127.6762
127.8447
127.8973
127.9674
128.1360
128.3460
128.3826
128.4363
137.8929
138.1465
138.2283
138.7747

OBn
O
BnO
OMe
BnO
BnO
B51-$\alpha$

DEPT135

DEPT90

C13
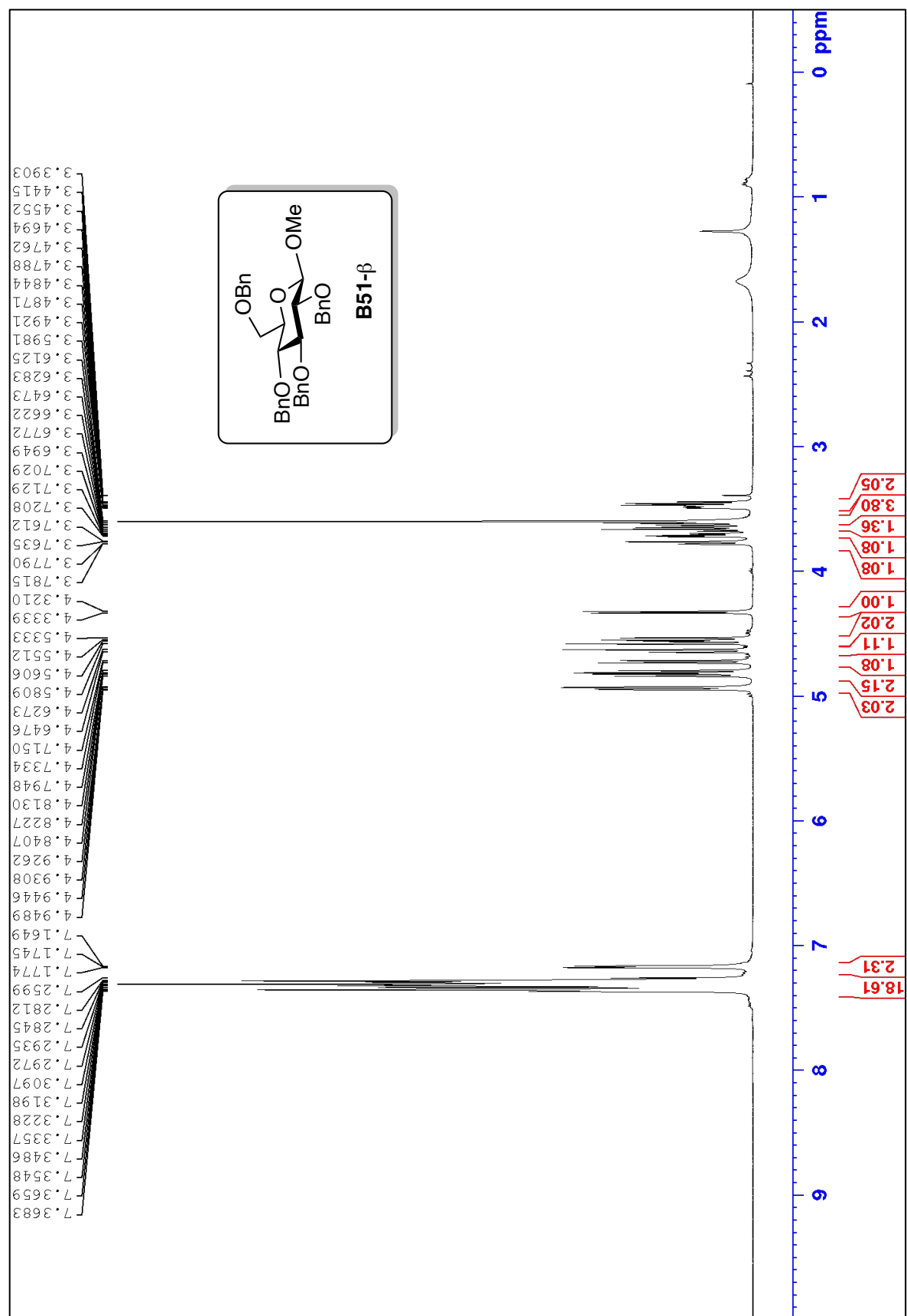
The $^1$H Spectrum in CDCl$_3$ of Compound B51-β.

The $^{13}$C Spectrum in CDCl$_3$ of Compound **B51-β**.

57.0989
68.8833
73.4653
74.7314
74.8064
74.9977
75.6643
76.7872
76.9990
77.2108
77.8334
82.2889
84.6095
104.6718
127.5769
127.6109
127.7359
127.7533
127.8501
127.9509
128.0796
128.3323
128.3659
138.0546
138.1277
138.4888
138.5609

DEPT135

OBn
O
OMe
BnO
BnO
BnO
BnO

**B51-β**

DEPT90

C13

The $^1$H Spectrum in CDCl$_3$ of Compound B53-β.

The 1D-TOCSY Spectrum in CDCl$_3$ of Compound B53-β.

1D Selective Gradient TOCSY
freq: 4.381ppm

1D Selective Gradient TOCSY
freq: 4.209ppm

1H

B53-β

3.3124
3.3374
3.4211
3.4284
3.4493
3.4714
3.4958
3.5113
3.5199
3.5294
3.5518
3.5752
3.5914
3.6136
3.6359
3.6487
3.6625
3.6747
3.6960
3.7190
3.7973
3.8052
3.8228
3.9547
3.9779
4.0000
4.1520
4.1752
4.3245
4.3439
4.4838
4.5065
4.5112
4.5322
4.5616
4.5914
4.5996
4.6208
4.6511
4.6905
4.7178
4.7452
4.7533
4.7816
4.8053
4.8765
4.9038
4.9362
4.9456
4.9634
4.9734

ppm
3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 5.0 5.1 5.2 5.3 5.4 5.5 5.6 5.7 5.8 5.9

2.75
9.15
1.12
1.09
0.97
1.00
12.48
3.21

The $^{13}$C Spectrum in CDCl$_3$ of Compound B53-β.

55.41
68.76
69.21
70.06
73.56
73.64
75.09
75.20
75.25
75.89
75.91
78.11
78.21
80.00
82.18
82.29
85.00
98.26
104.01
127.76
127.79
127.81
127.87
127.96
128.08
128.12
128.15
128.17
128.35
128.55
128.59
128.66
138.32
138.35
138.44
138.56
138.59
138.74
139.05

DEPT135

DEPT90

13C

B53-β

The $^1$H Spectrum in CDCl$_3$ of Compound B54-β.

The 1D-TOCSY Spectrum in CDCl₃ of Compound B54-β.

1D Selective Gradient TOCSY
freq: 3.997ppm

1D Selective Gradient TOCSY
freq: 3.744ppm

1H
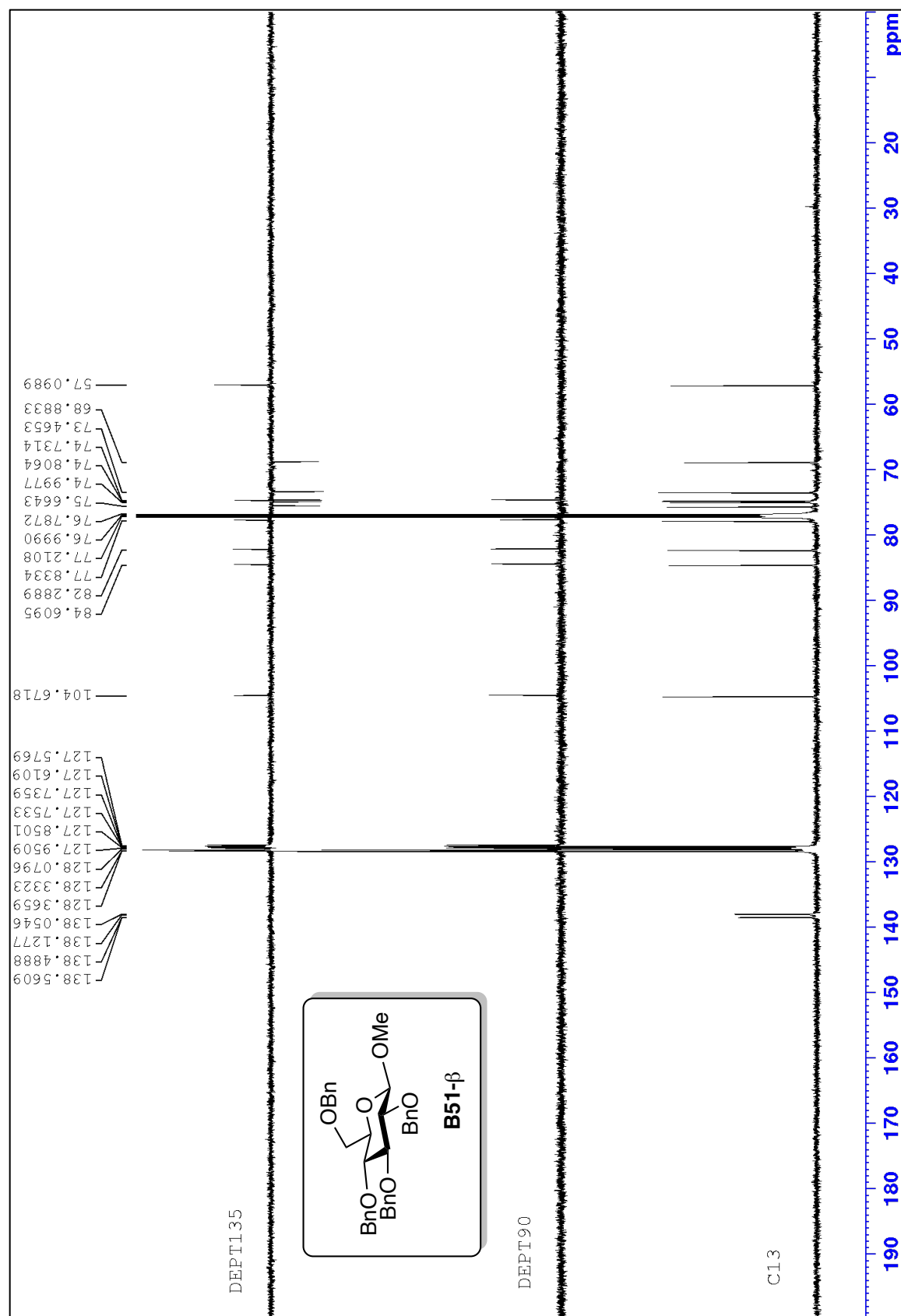
The $^{13}$C Spectrum in CDCl$_3$ of Compound B54-β.

The $^1$H Spectrum in CDCl$_3$ of Compound B56

The $^{13}C$ Spectrum in CDCl$_3$ of Compound B56
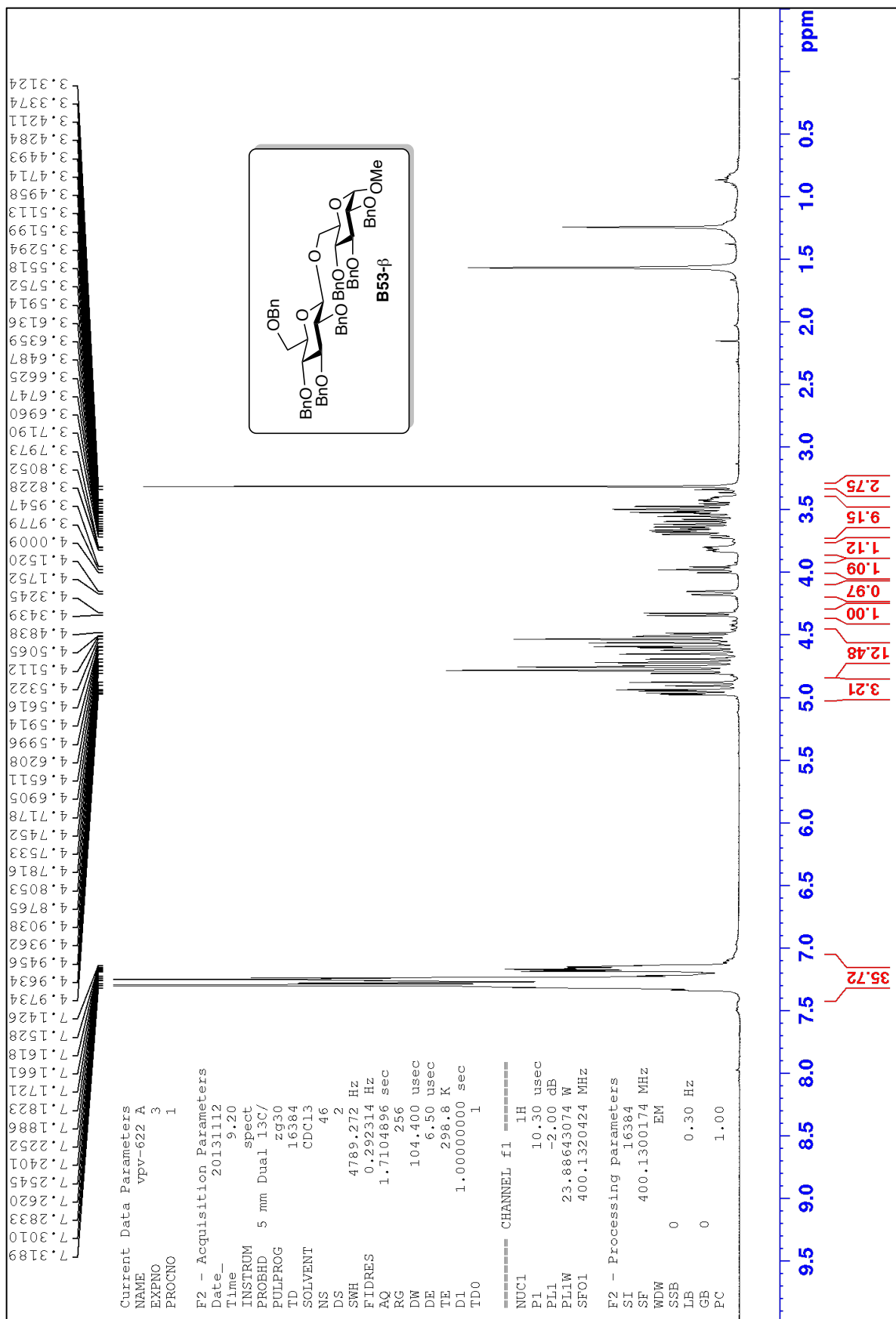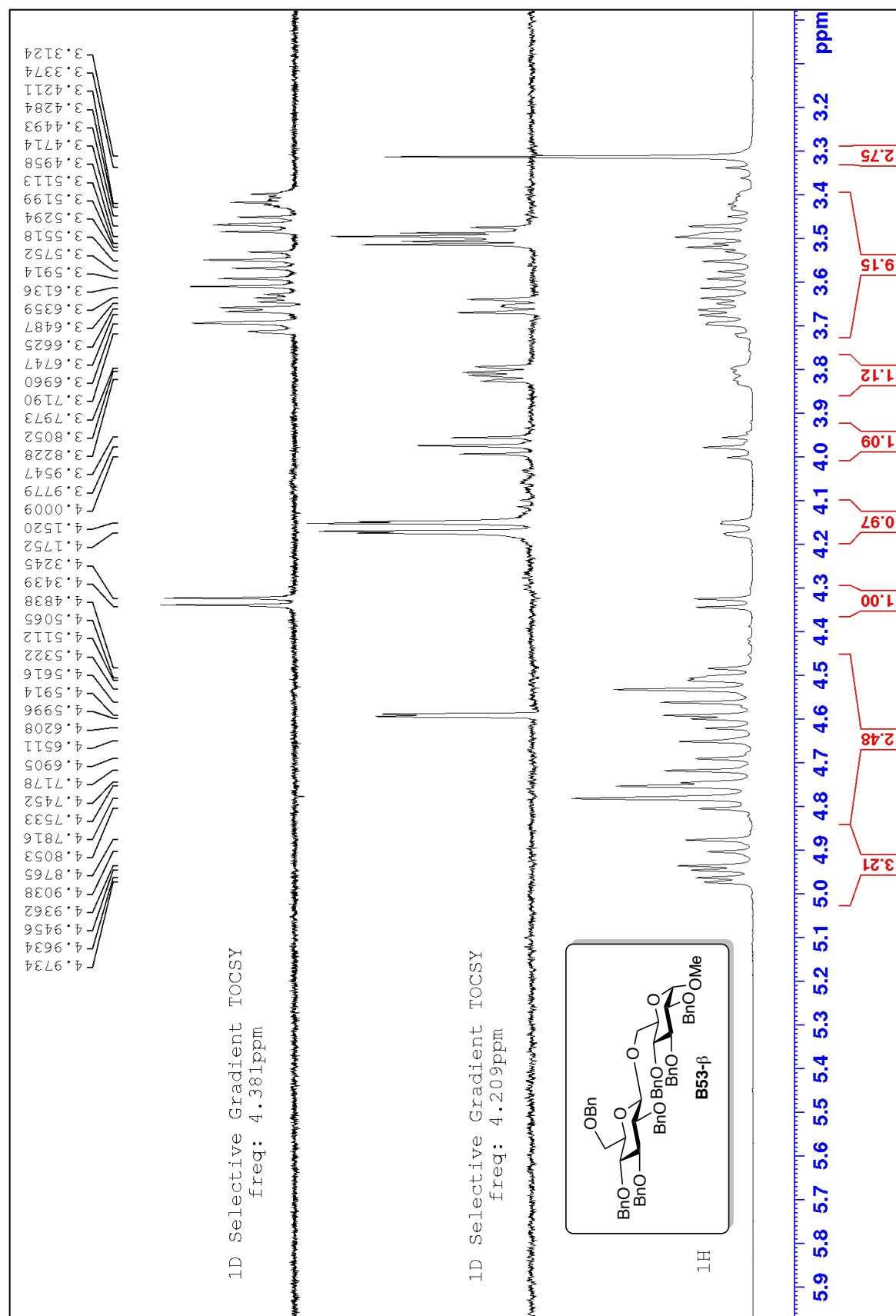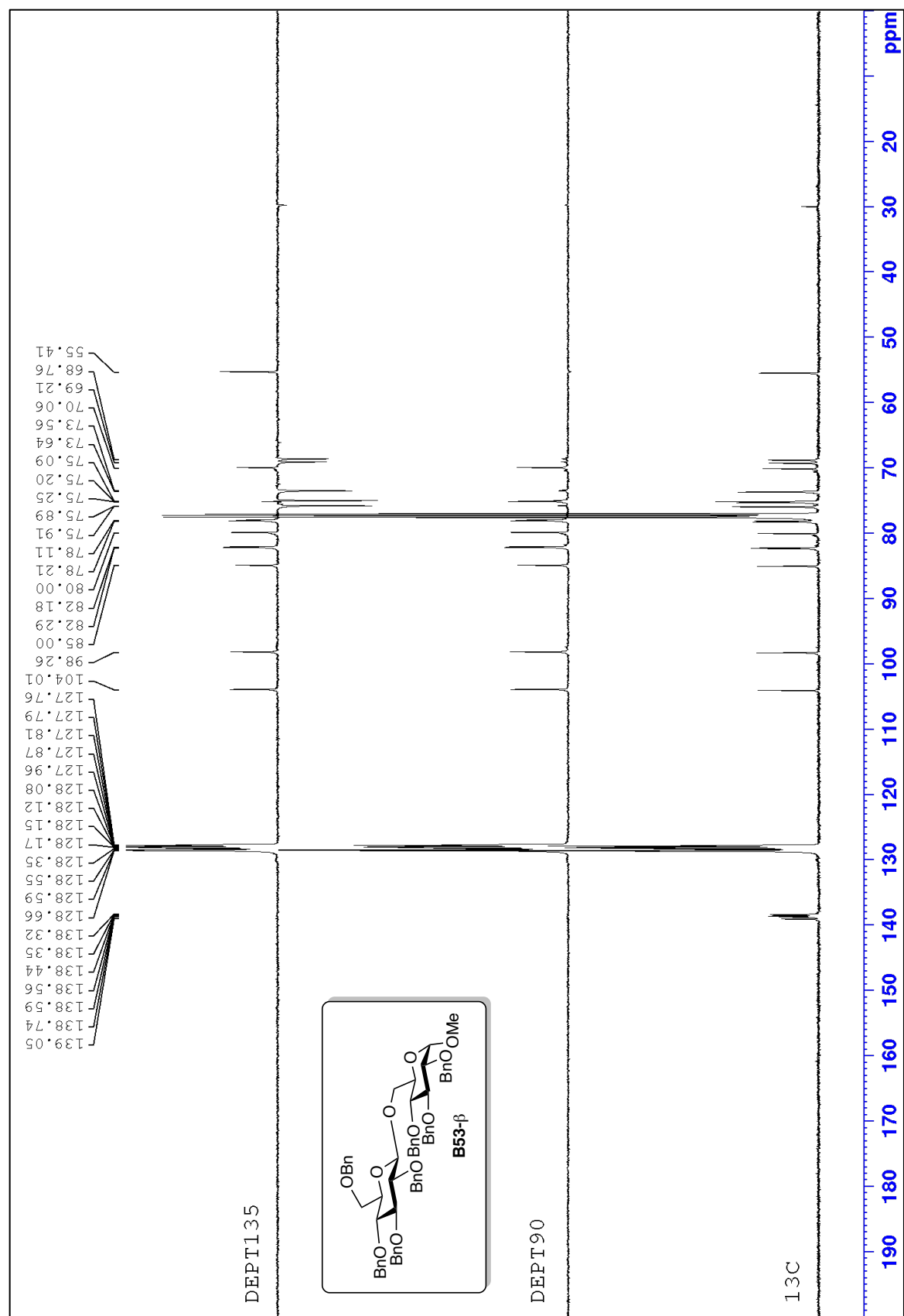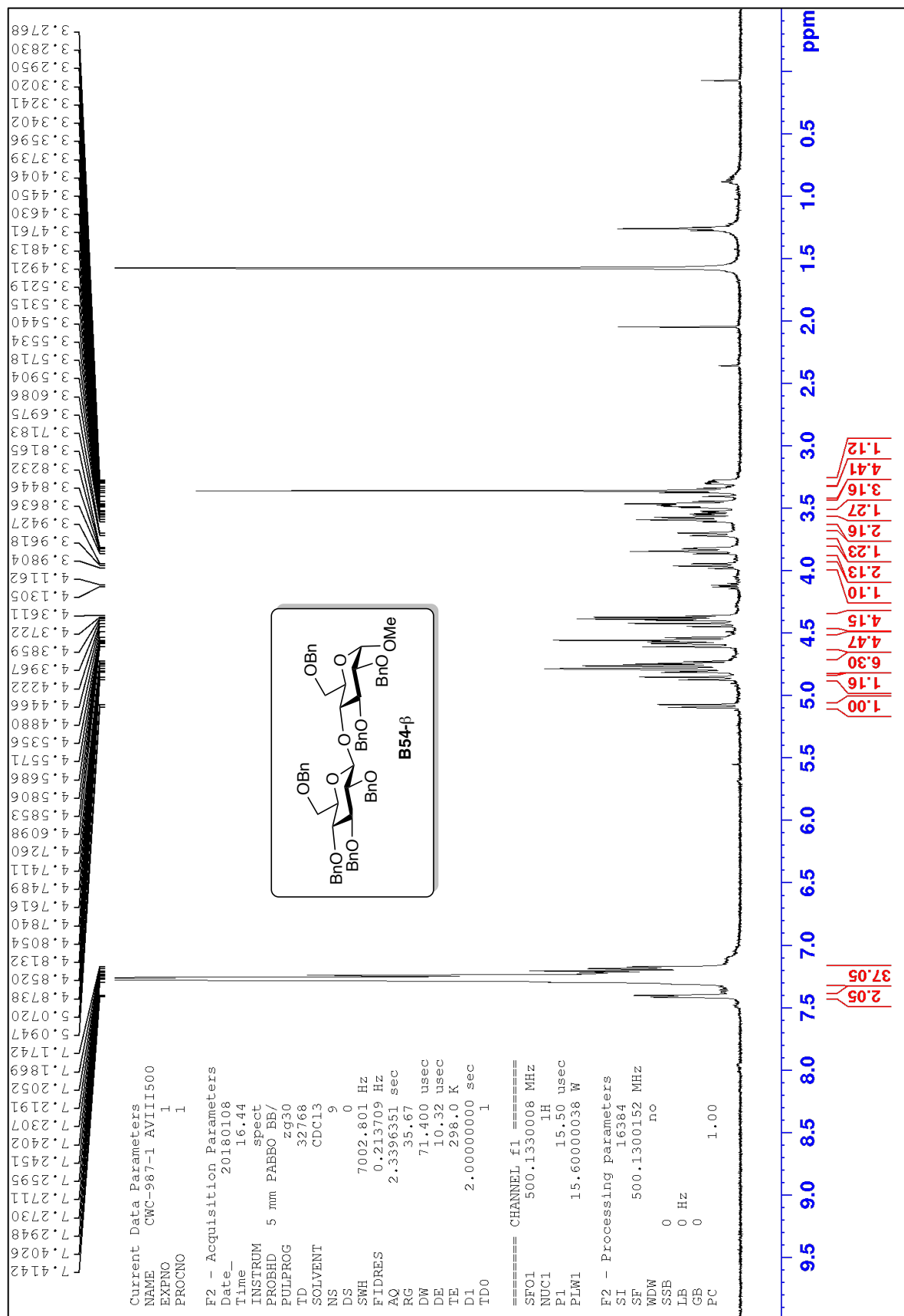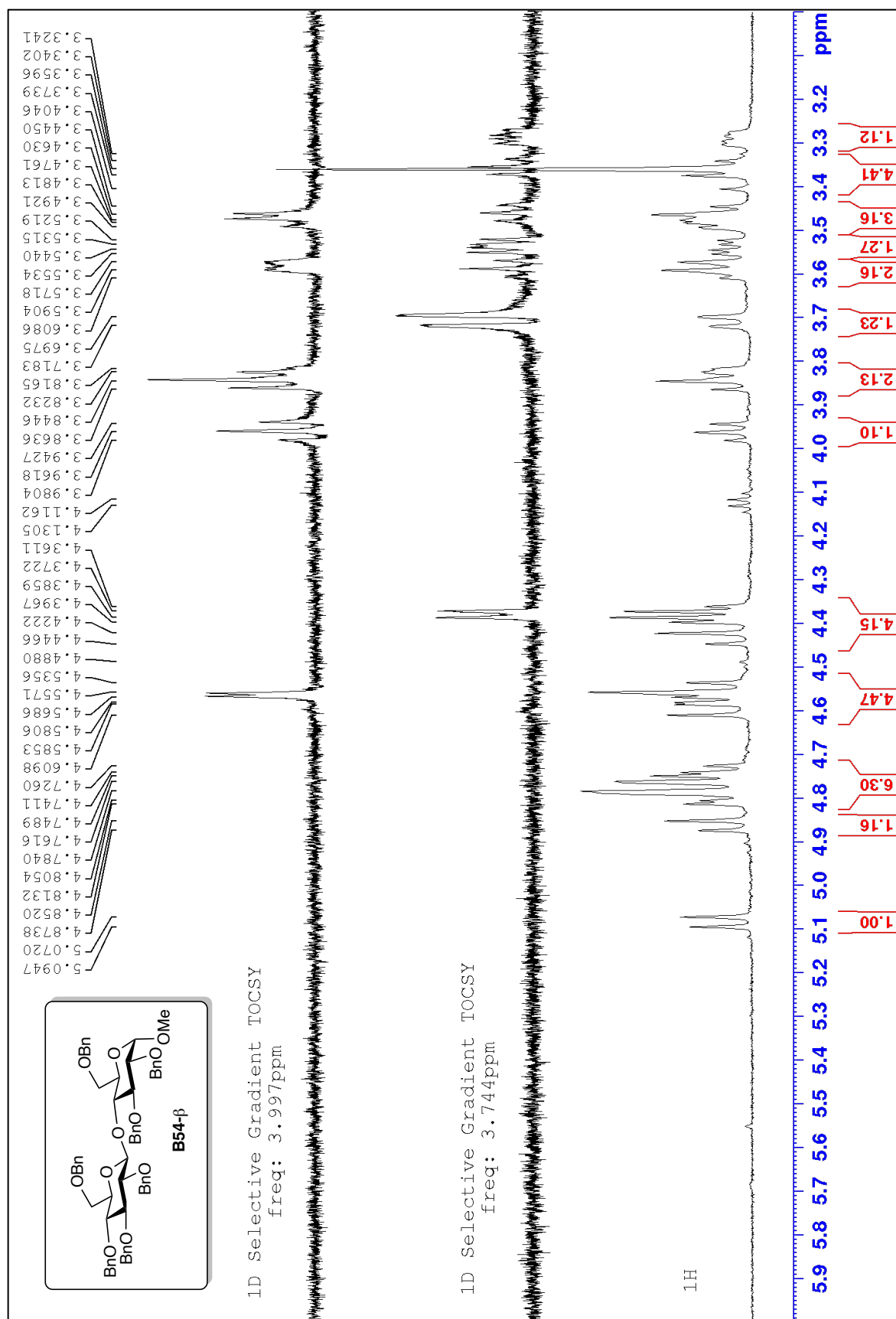
The $^1$H Spectrum in CDCl$_3$ of Compound C52

The ${}^{13}$C Spectrum in CDCl$_3$ of Compound C52

30.603
36.247
42.358
42.623
45.184
60.320
68.630
69.044
69.283
72.833
72.889
73.066
73.144
73.348
73.436
73.629
74.170
74.330
74.625
74.894
75.155
75.202
76.503
79.099
79.523
82.622
90.432
96.460
127.305
127.356
127.387
127.446
127.530
127.609
127.702
127.878
128.020
128.057
128.143
128.191
128.230
128.289
128.322
137.991
138.089
138.597
138.662
138.689
138.800
138.845
138.982

DEPT 135

C52

DEPT 90

${}^{13}$C

126

The $^1$H Spectrum in CDCl$_3$ of Compound D51

D51

Peak labels (ppm):
1.122, 1.135, 1.201, 1.214, 3.368, 3.517, 3.531, 3.537, 3.552, 3.569, 3.574, 3.791, 3.807, 3.826, 4.243, 4.258, 4.665, 4.672, 4.679, 4.692, 4.715, 4.721, 4.745, 4.761, 4.783, 4.788, 4.812, 4.828, 4.877, 4.913, 4.935, 4.981, 5.004, 7.260, 7.270, 7.279, 7.285, 7.288, 7.291, 7.293, 7.301, 7.303, 7.308, 7.312, 7.316, 7.320, 7.323, 7.330, 7.334, 7.339, 7.349, 7.352, 7.355, 7.365, 7.379, 7.389, 7.394, 7.397, 7.405

Integration labels:
1.761, 3.044, 1.729, 1.090, 1.066, 2.984, 1.023, 0.597, 1.589, 0.589, 0.593, 1.000, 1.126, 9.415, 23.694

The $^{13}$C Spectrum in CDCl$_3$ of Compound D51

16.563
16.810

55.250
56.836
66.046
70.250
73.102
73.330
73.442
74.499
74.787
75.024
76.270
76.310
76.740
76.994
77.248
77.839
79.384
79.484
82.463
98.775
104.848

127.427
127.459
127.512
127.620
128.057
128.143
128.211
128.281
128.325
128.387
138.542
138.579
138.600
138.904
138.930

DEPT 90

DEPT 135

$^{13}$C

D51

The $^1$H Spectrum in CDCl$_3$ of Compound D53

The $^{13}$C Spectrum in CDCl$_3$ of Compound D53

The $^1$H Spectrum in CDCl$_3$ of Compound D54

16.140
16.557
54.901
55.146
66.557
68.499
69.691
70.164
70.247
72.624
72.970
73.104
73.163
73.206
73.276
73.655
73.745
74.203
74.625
75.204
75.448
76.243
77.506
79.323
79.614
79.869
80.130
80.477
82.110
82.202
97.505
97.678
97.748
102.810
127.232
127.323
127.376
127.414
127.449
127.486
127.548
127.689
127.812
127.914
128.010
128.052
128.085
128.157
128.205
128.281
128.324
128.390
137.909
138.316
138.574
138.674
138.744
138.789

OBn

O

BnO OMe

BnO

D54

BnO

OBn

O

OBn

BnO

DEPT 135

DEPT 90

$^{13}$C

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 ppm

**The $^{13}$C Spectrum in CDCl$_3$ of Compound D54**
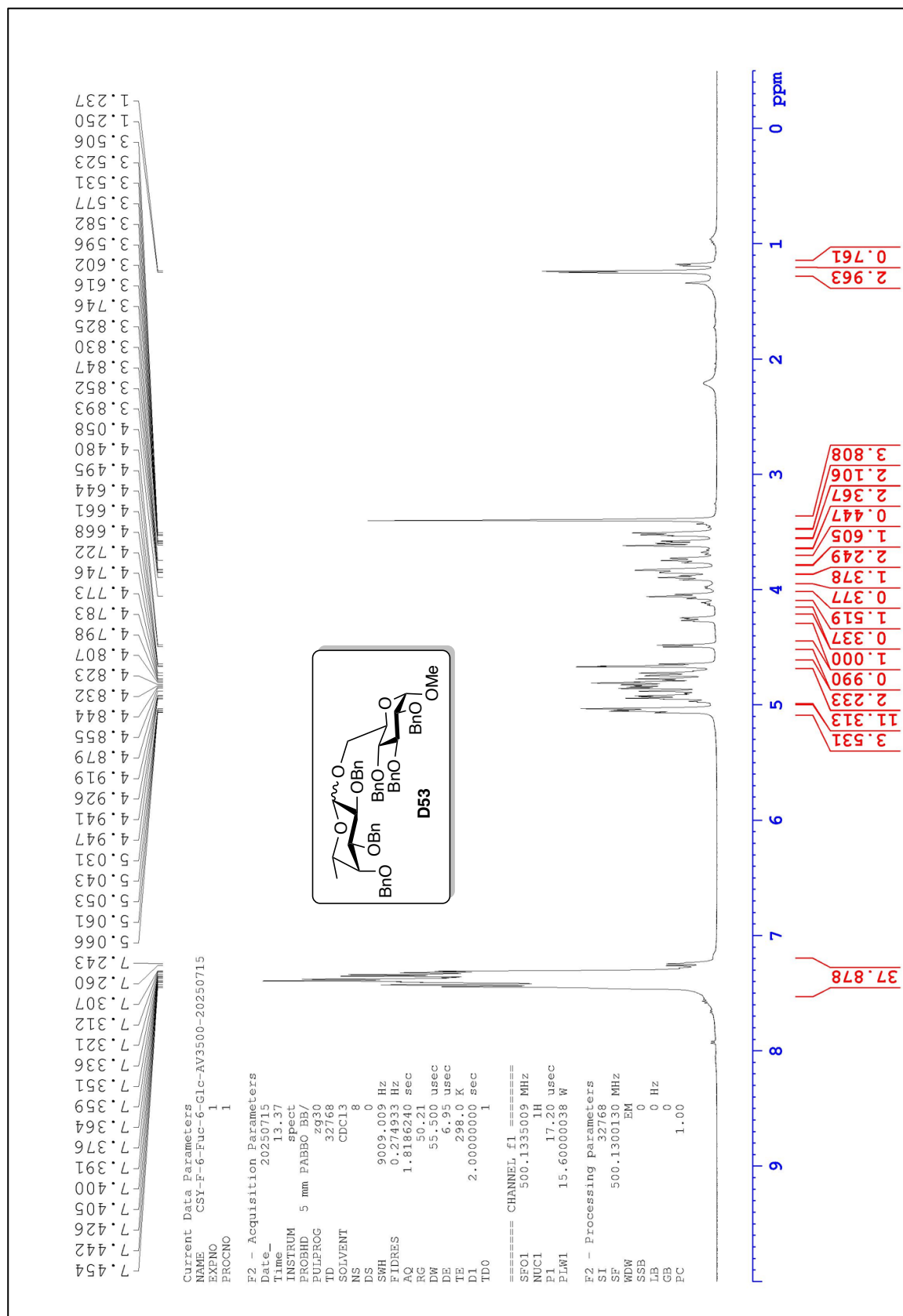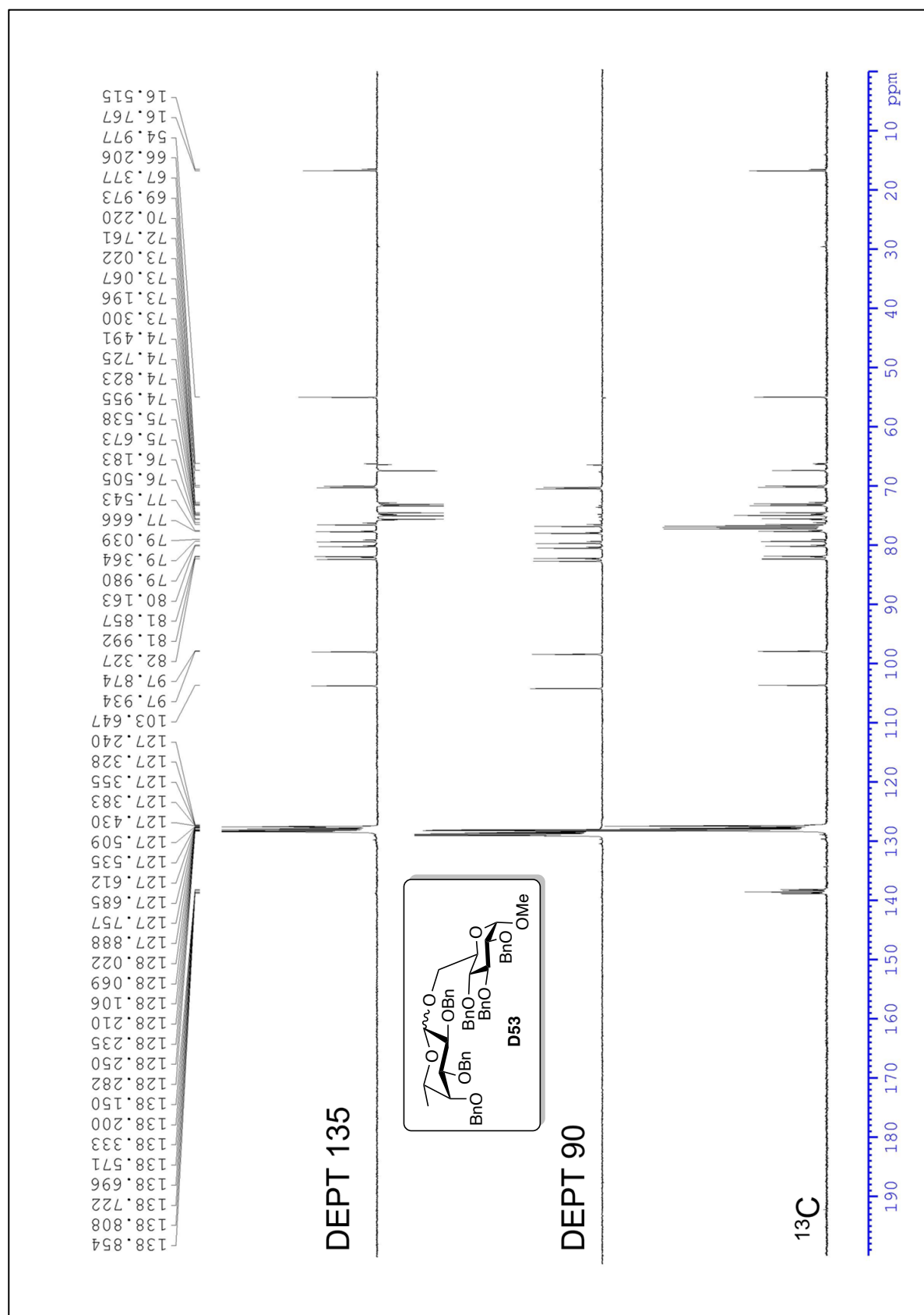
The $^1$H Spectrum in CDCl$_3$ of Compound D55
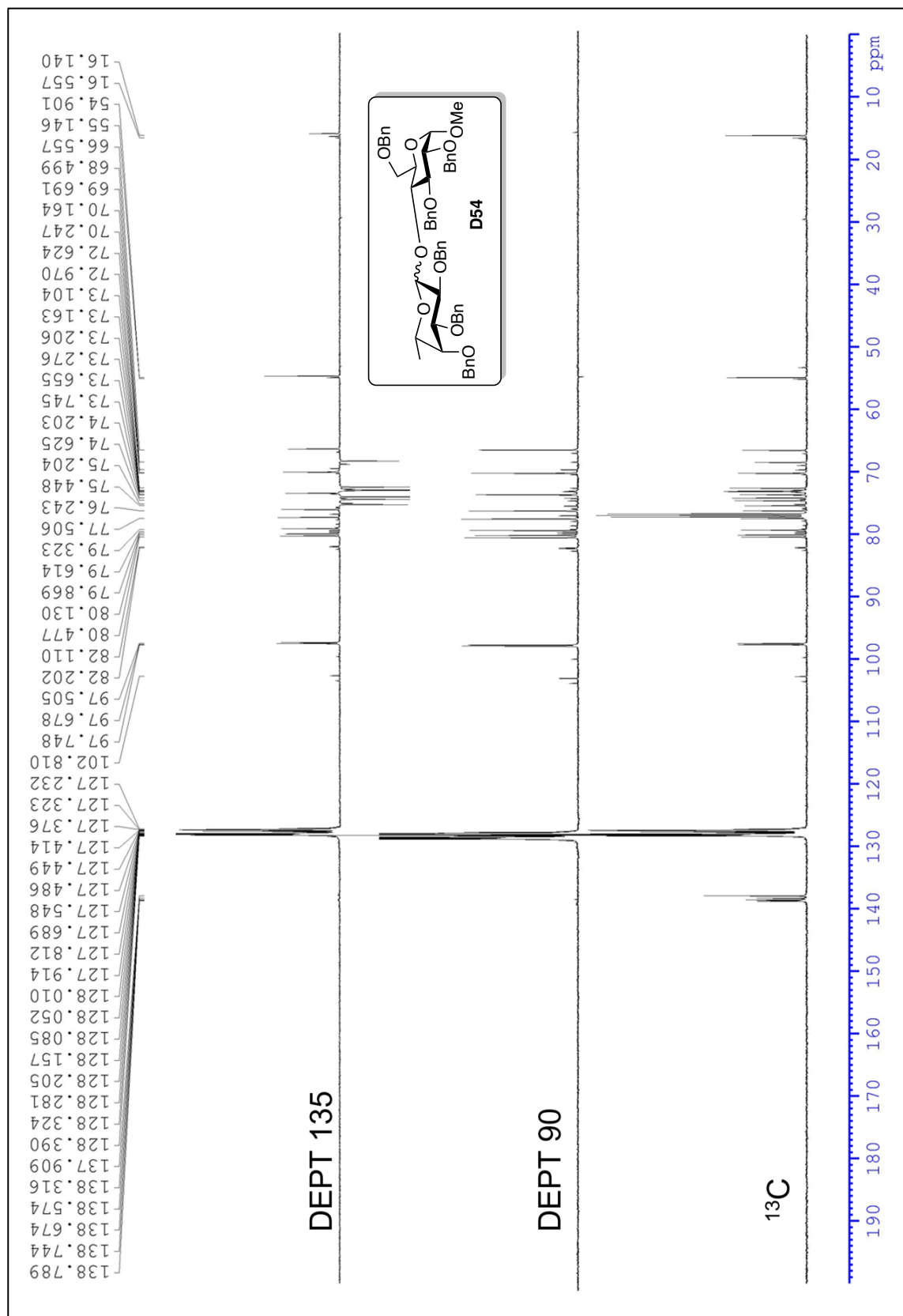


133

The $^{13}$C Spectrum in CDCl$_3$ of Compound D55

16.549
16.888

67.142
70.619
73.140
73.222
73.285
74.595
74.913
75.267
76.096
76.744
76.998
77.252
77.615
79.029
79.221
82.348
96.289
101.852
116.854
116.899
122.025
122.288
127.426
127.472
127.521
127.535
127.579
127.589
127.873
128.124
128.170
128.225
128.274
128.486
129.299
129.336
129.729
133.835
137.372
138.358
138.476
138.534
138.847
139.225
157.236
157.559

DEPT 135

DEPT 90

$^{13}$C

D55

The $^1$H Spectrum in CDCl$_3$ of Compound D56

Peak labels (ppm):
1.043, 1.050, 1.181, 1.198, 3.422, 3.607, 3.619, 3.943, 3.958, 3.964, 4.026, 4.385, 4.498, 4.522, 4.592, 4.675, 4.698, 4.723, 4.731, 4.764, 4.797, 4.818, 4.963, 4.986, 7.171, 7.178, 7.184, 7.208, 7.213, 7.226, 7.238, 7.241, 7.265, 7.268, 7.277, 7.280, 7.285, 7.290, 7.300, 7.309, 7.313, 7.317, 7.328, 7.333, 7.338, 7.344, 7.351, 7.356, 7.359, 7.361, 7.367, 7.369, 7.375, 7.377, 7.392, 7.402, 7.406, 7.442

Integration values:
3.054, 1.509, 0.505, 4.903, 0.559, 2.494, 0.992, 1.119, 6.485, 1.072, 1.018, 1.038, 6.599, 13.307, 1.000, 45.865

The $^{13}$C Spectrum in CDCl$_3$ of Compound D56

16.682
16.919
55.275
66.554
69.780
69.990
71.123
71.760
73.087
73.120
73.299
73.420
73.490
73.833
74.629
74.883
75.785
76.071
76.742
76.885
76.996
77.250
77.523
78.663
79.516
82.673
96.546
98.746
98.955
126.972
127.052
127.117
127.310
127.358
127.432
127.452
127.479
127.550
127.592
127.650
127.696
127.734
127.952
127.983
128.077
128.129
128.159
128.262
128.323
128.369
128.457
128.568
137.877
138.304
138.518
138.678
139.010
139.119

DEPT 135

DEPT 90

$^{13}$C

D56

The $^1$H Spectrum in CDCl$_3$ of Compound E51.

The $^{13}$C Spectrum in CDCl$_3$ of Compound E51.

29.876
35.605
36.814
54.753
56.689
69.271
69.604
70.921
71.668
71.940
73.676
75.055
75.461
76.911
77.228
77.546
77.822
78.491
79.616
98.709
101.002
127.698
127.797
127.882
128.023
128.163
128.472
128.515
138.436
138.569
138.646
138.845
138.939

OBn
O
OMe
E51
BnO
BnO

ppm
ppm
ppm

138
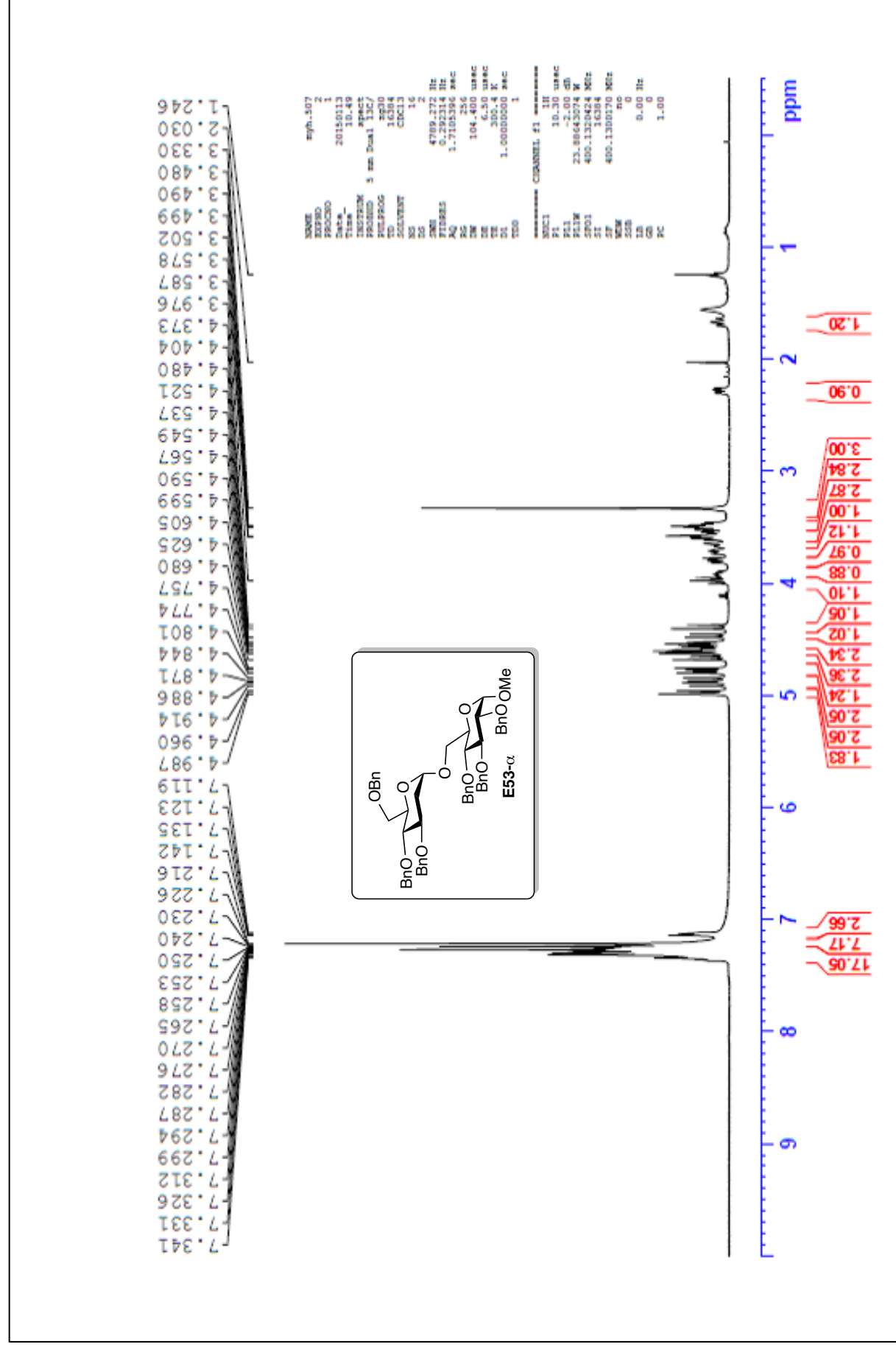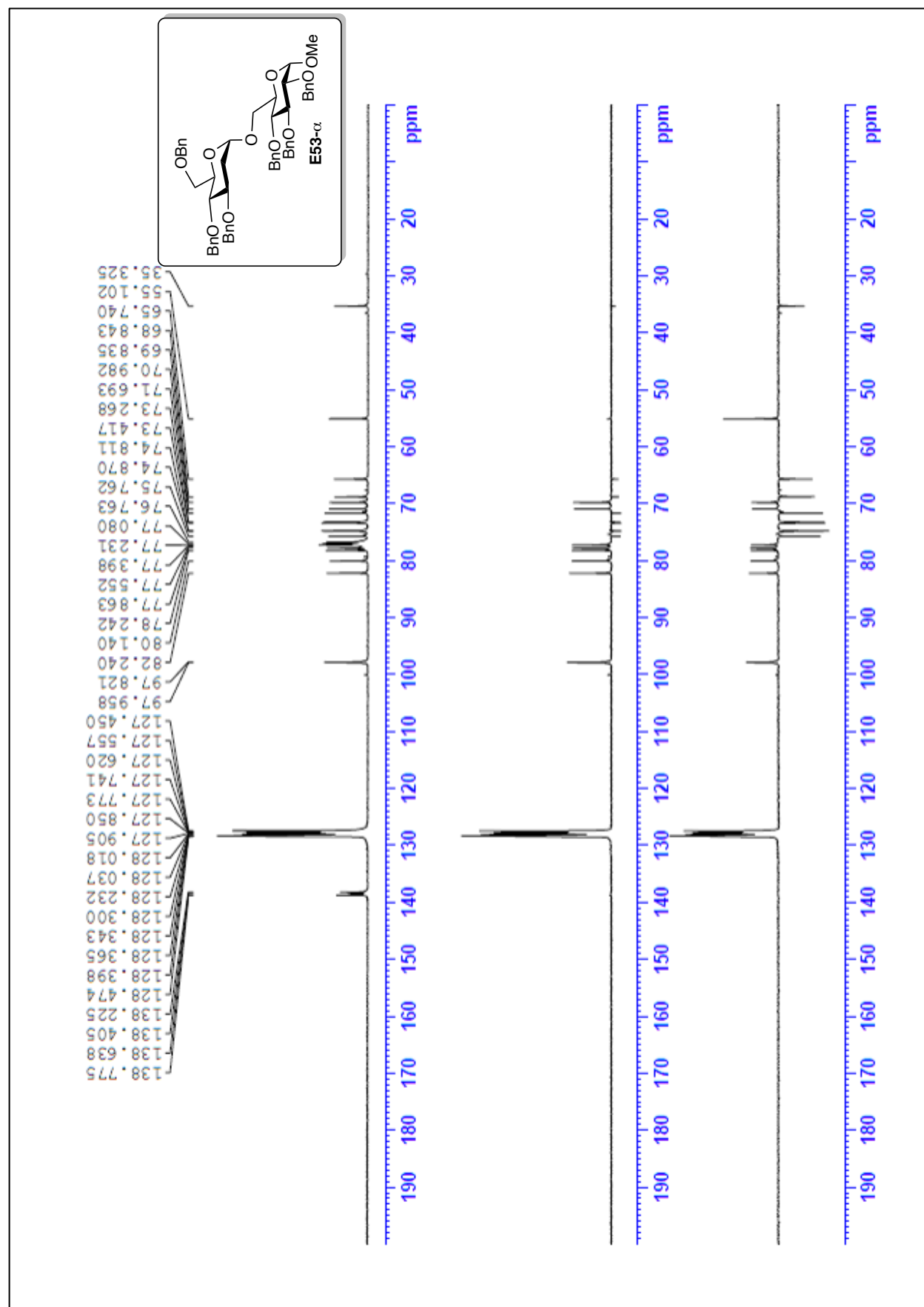
The $^1$H Spectrum in CDCl$_3$ of Compound **E53-α**.

The $^{13}$C Spectrum in CDCl$_3$ of Compound E53-$\alpha$.

35.325
55.102
65.740
68.843
69.835
70.982
71.693
73.268
73.417
74.811
74.870
75.762
76.763
77.080
77.231
77.398
77.552
77.863
78.242
80.140
82.240
97.821
97.958
127.450
127.557
127.620
127.741
127.773
127.850
127.905
128.018
128.037
128.232
128.300
128.343
128.365
128.398
128.474
138.225
138.405
138.638
138.775

The $^1$H Spectrum in CDCl$_3$ of Compound E53-β.

The $^{13}$C Spectrum in CDCl$_3$ of Compound **E53-β**.

36.766
55.337
67.917
69.793
69.960
71.657
73.560
73.650
74.991
75.174
75.624
75.934
76.912
77.229
77.546
77.779
78.472
79.572
80.126
82.481
98.270
100.302
127.680
127.798
127.899
128.087
128.208
128.369
128.513
128.555
128.610
128.649
138.432
138.671
139.089

**E53-β**

The $^1$H Spectrum in CDCl$_3$ of Compound E54-α.

143

The $^{13}$C Spectrum in CDCl$_3$ of Compound E54-$\alpha$.

35.769
55.210
68.695
69.470
69.783
71.652
71.797
73.214
73.455
74.790
75.407
76.167
76.740
77.058
77.229
77.376
78.093
80.122
82.120
97.738
99.385
127.405
127.456
127.561
127.886
128.111
128.256
128.339
128.406
128.442
138.006
138.118
138.315
138.592
138.677

E54-$\alpha$

OBn
OBn
OMe
BnO
BnO
BnO
OBn