

## Glycoscience

## Automated Quantification of Hydroxyl Reactivities: Prediction of Glycosylation Reactions

Chun-Wei Chang, Mei-Huei Lin, Chieh-Kai Chan, Kuan-Yu Su, Chia-Hui Wu, Wei-Chih Lo, Sarah Lam, Yu-Ting Cheng, Pin-Hsuan Liao, Chi-Huey Wong,\* and Cheng-Chung Wang\*

**Abstract:** The stereoselectivity and yield in glycosylation reactions are paramount but unpredictable. We have developed a database of acceptor nucleophilic constants (Aka) to quantify the nucleophilicity of hydroxyl groups in glycosylation influenced by the steric, electronic and structural effects, providing a connection between experiments and computer algorithms. The subtle reactivity differences among the hydroxyl groups on various carbohydrate molecules can be defined by Aka, which is easily accessible by a simple and convenient automation system to assure high reproducibility and accuracy. A diverse range of glycosylation donors and acceptors with well-defined reactivity and promoters were organized and processed by the designed software program “GlycoComputer” for prediction of glycosylation reactions without involving sophisticated computational processing. The importance of Aka was further verified by random forest algorithm, and the applicability was tested by the synthesis of a Lewis A skeleton to show that the stereoselectivity and yield can be accurately estimated.

## Introduction

The advances in synthetic chemistry have made possible access to complex natural or unnatural molecules. These advances are mainly attributed to the development of new chemical reactions and strategies, and the improvement of reaction yield and selectivity. Recently, computer-aided organic synthesis<sup>[1]</sup> and synthesis planning software with algorithm and machine-learning have been developed to

provide more plausible synthetic routes.<sup>[2]</sup> However, their applications to diastereoselective reactions are still rare.

Carbohydrates are ubiquitous biomolecules that mediate numerous biological processes and exhibit important pathogenic effects.<sup>[3]</sup> However, obtaining carbohydrates by extraction from natural sources is impractical and chemical synthesis of glycoconjugates has been hampered by unpredictable glycosylation. Currently, there is no clear rule to direct the configuration of glycosidic linkage as glycosylation reaction often proceeds through a range of mechanisms involving different intermediates, leading to a mixture of  $\alpha$ - and  $\beta$ -glycosides.<sup>[4]</sup> In addition to the reactivities of glycosyl donor **1** and acceptor **2**, the steric effects and use of promoters and solvents also affect the outcome of glycosylation reactions<sup>[5]</sup> (Figure 1 A). It is known that solvents and promoters will influence the counter ion coordination,<sup>[4c,5,6]</sup> and the anomeric selectivity and yield are affected by a combination of all these unquantifiable factors. Tedious trial and error are thus required to optimize each glycosyl linkage, making the chemical synthesis of complex oligosaccharides a major challenge.

In a glycosylation reaction with covalent intermediate, the acceptor tends to attack the anomeric carbon (C1) via the  $S_N2$ -like pathway, while for the reaction via oxocarbenium ion the acceptor attachment generally undergoes the  $S_N1$ -like mechanism either from the top ( $\beta$ ) or the bottom ( $\alpha$ ). The reaction results in a spectrum from  $S_N2$ - to  $S_N1$ -type reaction pathways among various coupling partners<sup>[7]</sup> (Figure 1 B). In general, donors with higher reactivity favor  $\alpha$ -glycosylation, while acceptors with higher reactivity tend to form  $\beta$ -glycosides.<sup>[5,8]</sup> In the absence of C2 neighboring group participation, a less reactive donor gives  $\alpha$ -triflate intermediate **4** of higher stability, and a stronger nucleophile favors the  $S_N2$ -like substitution to yield  $\beta$ -glycoside from the  $\alpha$ -triflate intermediate **4**. In addition, previous studies by Codée, Bennett and us have shown that increasing the reactivity of donor gives more  $\alpha$ -glycosylation,<sup>[8a,c-g]</sup> probably due to the presence of oxocarbenium ion-like intermediate (e.g., **4-SSIP**) which leads to the  $\alpha$ -selective reaction through a unimolecular  $S_N1$  pathway.<sup>[7]</sup> These oxocarbenium ions favor  ${}^3E/{}^3H_4$ -like conformations allowing hyperconjugation between O5 lone pairs and the forming  $\sigma^*$  orbital of  $\alpha$ -glycoside in the transition state.<sup>[9]</sup> Codée et al. further described the conformation of oxocarbenium ion which is highly associated with its functional groups, and these conformational changes directly affect the stereoselectivity of glycosylation reactions. Depending on the shape of ion and the distortion of sugar ring, both the 1,2-*cis* and 1,2-*trans* products can be obtained under

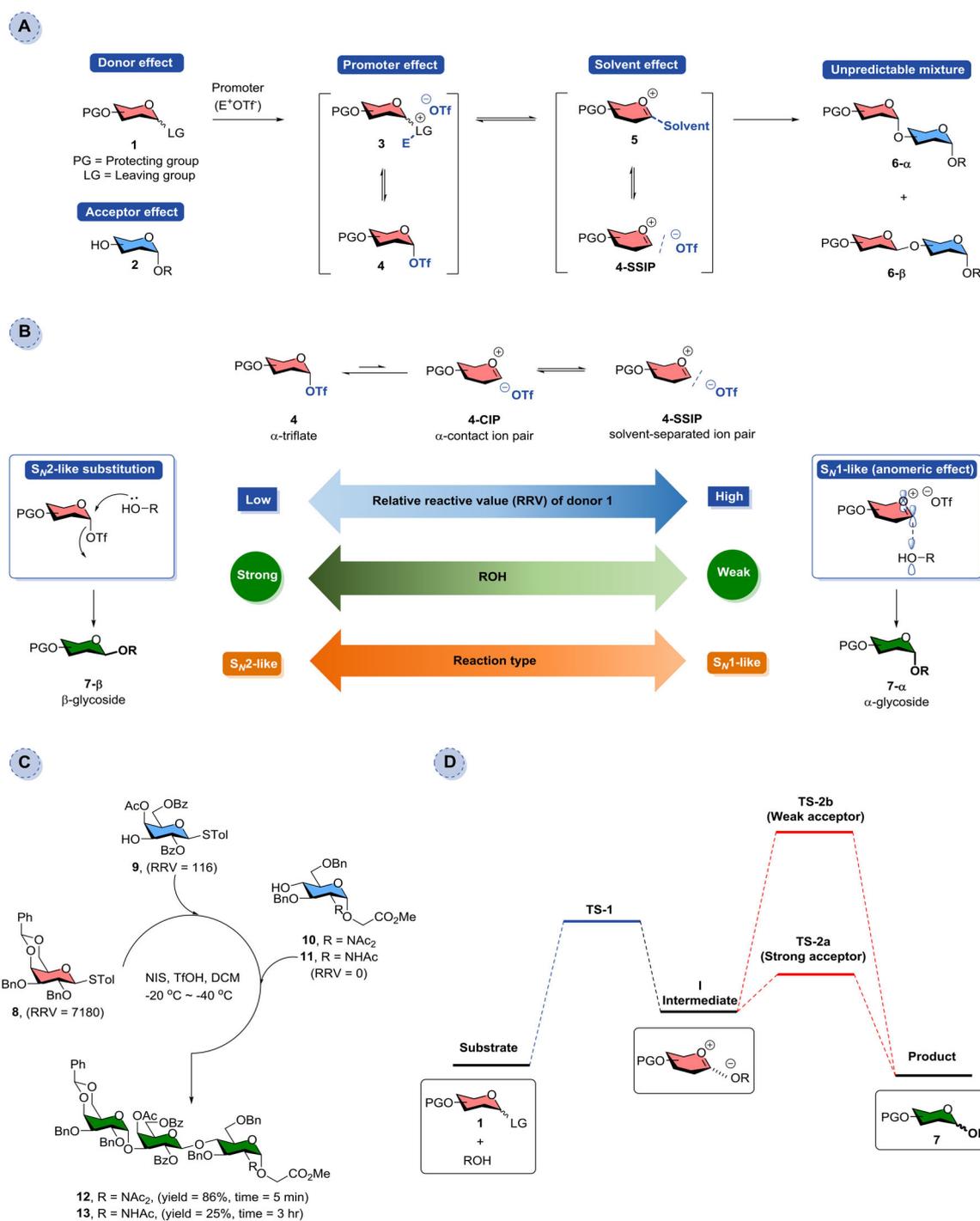
[\*] C.-W. Chang, M.-H. Lin, C.-K. Chan, K.-Y. Su, C.-H. Wu, W.-C. Lo, S. Lam, Y.-T. Cheng, P.-H. Liao, C.-C. Wang  
Institute of Chemistry, Academia Sinica  
Taipei 115 (Taiwan)  
E-mail: wangcc@chem.sinica.edu.tw

C.-H. Wong  
The Genomics Research Center, Academia Sinica  
Taipei 115 (Taiwan)

and  
Department of Chemistry, The Scripps Research Institute  
10550 N Torrey Pines Road, La Jolla, 92037 (USA)  
E-mail: ch Wong@gate.sinica.edu.tw

C.-C. Wang  
Chemical Biology and Molecular Biophysics Program, Taiwan  
International Graduate Program (TIGP), Academia Sinica  
Taipei 115 (Taiwan)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/anie.202013909>.



**Figure 1.** A) The formation of glycosidic bonds via pathways with various transition states and intermediates. B) Reactivities of glycosyl donors and nucleophilic acceptors related to the glycosylation reactions at the S<sub>N</sub>1–S<sub>N</sub>2 interface. C) Acceptor reactivity in glycosylation. D) The diagram represents differences in the rate-determining step.

kinetic conditions. The formation of the glycosidic linkages proceeds under kinetic conditions, and a high energy conformer may be formed initially, after which a ring conformation change can take place in the subsequent step.<sup>[10]</sup> The S<sub>N</sub>1/S<sub>N</sub>2 boundary for a given donor and acceptor pair is typically not well understood, leading to a poor stereoselectivity and reduced efficiency of oligosaccharide synthesis.

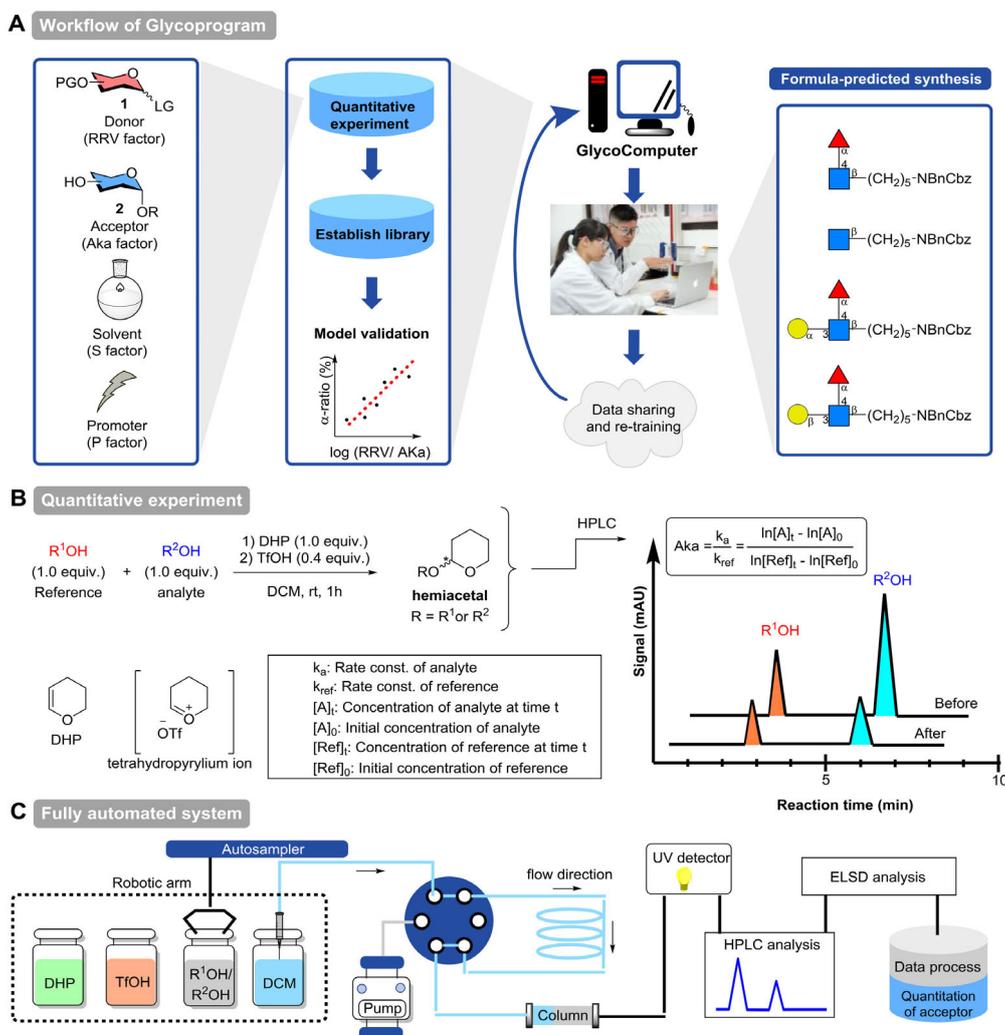
Based on the basic mechanistic principles, different strategies of sequential glycosylations were developed to facilitate the synthesis of oligosaccharides. Inspired by the armed-disarmed concept,<sup>[11]</sup> the Ley group reported the use of protecting groups to tune the anomeric reactivity of glycosyl donors,<sup>[12]</sup> and we developed the programmable one-pot synthesis of oligosaccharides using the relative reactivity

values (RRVs) of thioglycoside building blocks measured by experiments or by prediction through machine learning.<sup>[13]</sup> However, we noticed that although RRV provided a useful indicator for assessing the reactivity of glycosyl donors, this strategy could not work smoothly using acceptors with poor reactivity. For example, Manabe and Fukase et al. reported the one-pot synthesis of a trisaccharide<sup>[14]</sup> (Figure 1 C) using glucosaminyl acceptor **10** to give the desired product **12** in 86% yield in 5 min, but with an unreactive acceptor **11** the yield was significantly reduced to 25% after 3 h. Though the formation of oxocarbenium ion is often the rate-limiting step, which can be quantified by RRV, this result indicated possible existence of two distinct pathways via transition state **TS-2a** or **TS-2b** in glycosylation reactions<sup>[11a,15]</sup> (Figure 1 D). With a strong acceptor, the glycosylated product **7** is predominantly formed in a higher yield since the reaction goes through a rate-limiting transition state 1 (**TS1**) and high-energy oxocarbenium intermediate **I** which reacts with the acceptor quickly to form the product via transition state **TS-2a**. In contrast, with a weak acceptor, the high-energy intermediate **I** reacts with the acceptor through a rate-limiting transition

state **TS-2b** with a much higher energy barrier and lower yield.<sup>[16]</sup> The reactivity of acceptor is highly sensitive to its structural features and protecting groups, so development of a digital index of the acceptor reactivity will make the glycosylation reaction more predictable.

## Results and Discussion

Herein, we developed a GlycoComputer program<sup>[17]</sup> based on the properties of various donors, acceptors, activation systems and solvents to foresee and predict the yield and stereoselectivity of a glycosylation reaction<sup>[17]</sup> (Figure 2 A). The key success of this program is to establish a general comparing system to characterize the reactivity of donor **1** and acceptor **2** by using the RRV<sup>[8a,c,13]</sup> and the acceptor nucleophilic constant (Aka) for statistical analysis. Although researchers have been trying to use high-level quantum chemistry to predict chemical glycosylations,<sup>[10a,b]</sup> our parameters obtained from the competition experiments took both the structural and diverse protecting group effects present on



**Figure 2.** A) The GlycoComputer concept. B) Determination of Aka values. C) Automated system for data collection via upgrading of a commercially available HPLC autosampler.

numerous glycosyl donors and nucleophilic acceptors into account to avoid complicated computational calculation. Next, we programmed a random forest algorithm to test the predictability of our GlycoComputer.<sup>[17]</sup> The variables present in glycosylation reactions were well defined, including substrate factors (donors and acceptors) and environmental factors (solvent and promoter). This advanced programmable system provides potential synthetic solutions for complex glycans as demonstrated in the four successful examples reported here.

Since a statistical measurement of hydroxyl reactivity under an acidic glycosylation condition is absent in literature, we introduced the acceptor reactivity value Aka (Figure 2B; Supporting Information, Figures S2–S7). The acid-catalyzed reaction between a hydroxyl and an achiral electrophile, such as 3,4-dihydropyran (DHP)<sup>[18]</sup> which forms a tetrahydropyrylium ion in the presence of catalytic triflic acid (TfOH) to react with the hydroxyl group, was chosen as a model to simulate common glycosyl reaction. Aka is defined as the relative reactivity difference towards the tetrahydropyrylium ion between the two hydroxyls, R<sup>1</sup>OH and R<sup>2</sup>OH, and is determined by using HPLC. By comparing the HPLC patterns of R<sup>1</sup>OH and R<sup>2</sup>OH before and after the reaction, the Aka values of R<sup>1</sup>OH and R<sup>2</sup>OH were obtained using the second-order rate equation established by our previous works.<sup>[13]</sup>

$$\text{Aka} = \frac{k_a}{k_{\text{ref}}} = \frac{\ln[A]_t - \ln[A]_0}{\ln[\text{Ref}]_t - \ln[\text{Ref}]_0}$$

Since DHP cation is prochiral, the formation of the hemiacetal product would generate a new chiral center. Indeed, both the R/S diastereoisomers can be formed and were further confirmed by NMR (Figures S11–S13). However, according to our previous RRV concept<sup>[13]</sup> the Aka, that quantifies the relative acceptor reactivity, was based on the disappearance of R<sup>1</sup>OH and R<sup>2</sup>OH rather than the appearance of THP product, and therefore the R/S chirality of the corresponding THP hemiacetal product was not taken into account. Therefore it is reasonable to directly measure a relative first-order rate constant by comparing the HPLC signals of R<sup>1</sup>OH and R<sup>2</sup>OH before and after the reaction (Figures S2–S7). The equations (shown in Figure 2B) represented in the relationship between relative rate constant ratios (k) and acceptor concentrations [A] towards the tetrahydropyrylium ion.

The Aka values were acquired from a fully automated system by upgrading a commercially available HPLC autosampler (Figure 2C), which assured high reproducibility and was capable of screening 21 nucleophilic hydroxyls automatically within 24 h. The two required reagents and two substrates were automatically introduced into the sample loop (100  $\mu\text{L}$ ) with HPLC pump. The autosampler mixed the solutions by bubble movement in the loop. The combination allowed the test reaction to occur in the sample loop for 1 h and directly injected the crude solution into the analytical HPLC column. The Aka of each hydroxyl was measured for three independent times and determined after taking their average. Since the minimum injection volume of the auto-

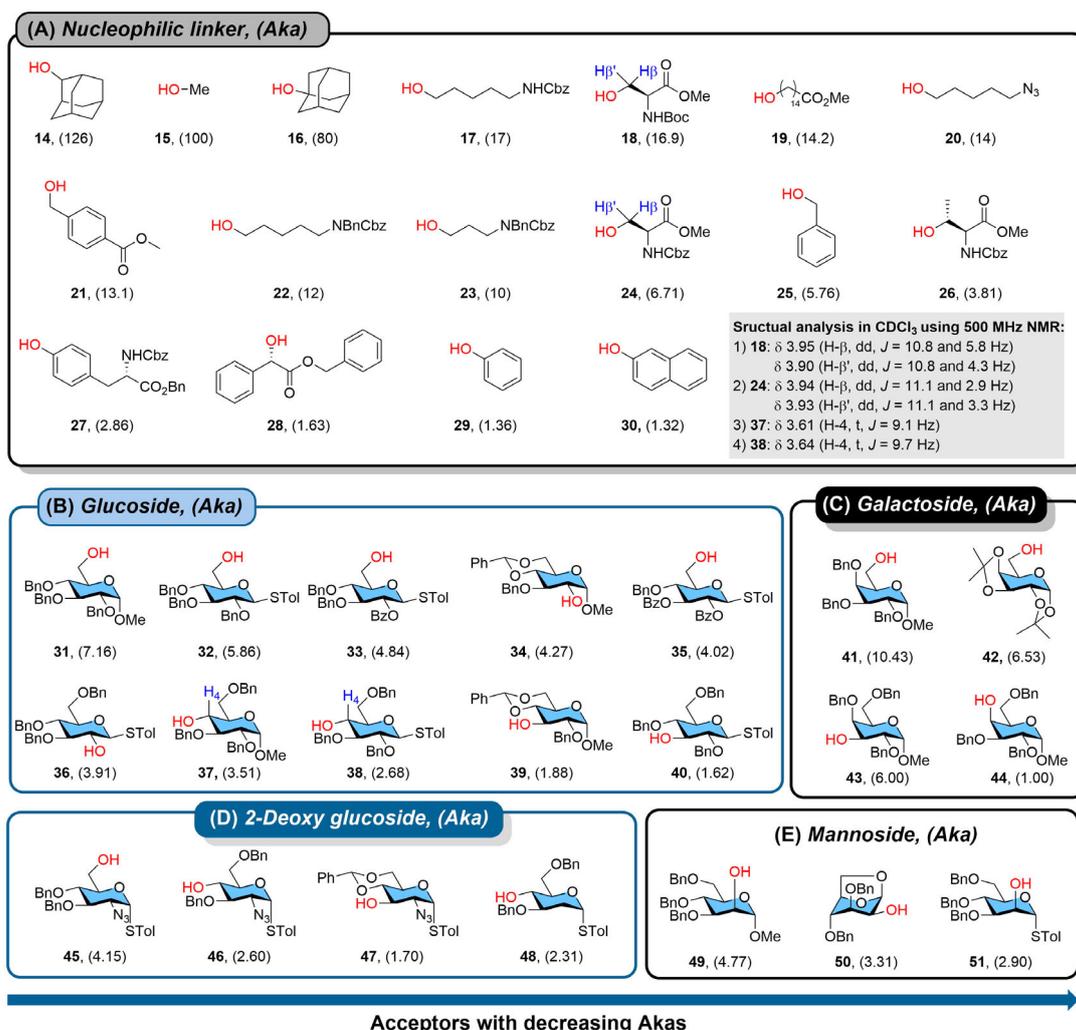
sampler syringe is 0.1  $\mu\text{L}$ , we can significantly reduce the substrate consumption to only 0.5  $\mu\text{mol}$  ( $\approx 0.23$  mg) for each trial. Although tetrahydropyran (THP) migration could happen from one THP hemiacetal to the other alcohol, based on our experiment (Figure S14), this thermodynamic transformation is quite slow. Exemplified by mixing phenol (**29**, Aka = 1.36) and the THP-hemiacetal of benzyl alcohol (**25**, Aka = 5.76) in the presence of TfOH under the same concentration of Aka test, the migration of THP from **25** to **29** was only 0.3 % in 1 h and 0.6 % in 4 h, suggesting a minor influence on the Aka measurement.

The Aka values of various hydroxyls **14–51** were systematically determined for the first time using the least reactive axial 4-OH of galactoside **44** as the reference, 1.00 (Figure 3). In general, small molecules showed much higher reactivities, but surprisingly, the adamantanol **14** and **16** showed the highest Aka of 126 and 80 respectively, despite their commonly known steric hindrance and rigid configuration (Figure 3A). For the positional effect of thioglycosyl derivatives (Figure 3B), the decreasing reactivity is in the order of 6-OH **32** (Aka = 5.86) > 2-OH **36** (Aka = 3.91) > 4-OH **38** (Aka = 2.68) > 3-OH **40** (Aka = 1.62). This is in agreement with the results established by Codée and co-workers,<sup>[8b,d-f,h]</sup> where the primary 6-OH of **32** is the most reactive; but surprisingly, the 2-OH of **36** is the most reactive among the three secondary hydroxyl groups. A similar trend was noticed for methyl glucosides **31**, **34**, **37** and **39**. We also extended our scope to understand the reactivity of acceptors having different functional groups.

When compared to thioglucoside **32**, compounds **33** and **35** exhibited reduced reactivity of 6-OH when increasing the number of electron withdrawing group (OBz). It is also interesting that the primary alcohol **35** is less reactive than the secondary alcohol **34**, suggesting the di-benzoyl (Bz) modification on **35** significantly reduced its nucleophilicity due to the electronic effect. A reduced reactivity was also noted when the C2-position was replaced by a more electron withdrawing substituent such as 2-azidoglucoside **45**. The preliminary trend showed that the deactivating ability of electron withdrawing group was in the order of -OBz > -N<sub>3</sub> > -OBn.

With regard to galactosides (Figure 3C), it became apparent that the 6-OH groups of **41** and **42** show very high Aka of 10.43 and 6.53 individually. The 3-OH of **43** still revealed a very high reactivity, as the corresponding Aka was determined to be 6.00. However, the axial 4-OH of galactoside **44** was the least reactive among all the hydroxyls tested (Aka = 1.00). For the mannose 2-OH acceptors (**49–51**), their reactivities also depended on the identity of the anomeric group.  $\alpha$ -Thiomannoside **51** (Aka = 2.90) was found to be less reactive than  $\alpha$ -methyl mannoside **49** (Aka = 4.77). When the mannose was transformed to 1,6-anhydromannoside **50** to place the 2-OH in the equatorial instead of the usual axial position, the reactivity of **50** (Aka = 3.31) became intermediate, between mannoside **49** and **51**. However, not all of these trends can be explained by basic steric and electronic effects. Therefore, there must be other factors that influence the reactivity of alcohols in glycosylation reactions, such as intramolecular hydrogen bonding,<sup>[8b,19]</sup> conformation<sup>[11c-g,20]</sup>





**Figure 3.** Quantitation of hydroxyl reactivity under acidic condition (Akas). A) Nucleophilic linker. B) Glucosyl acceptors. C) Galactosyl acceptors. D) C-2 Modified glucosyl acceptors. E) 2-OH Mannosyl acceptors. Decreasing trends are shown.

and bond rotation.<sup>[7a,e,21]</sup> With the Aka index, we noticed  $\alpha$ -methyl glycoside usually gives higher nucleophilicity than  $\beta$ -tolyl thioglycoside [for example, 6-OH glucoside: **31** (7.16) vs. **32** (5.86); 4-OH glucoside: **37** (3.51) vs. **38** (2.68); 2-OH mannoside: **49** (4.77) vs. **51** (2.90)]. Analyzing glycosyl acceptor **37** using 500 MHz NMR gave a triplet (t) pattern on H-4 with a  $J_{1,2}$  value of 9.1 Hz (Figure S15). Compound **38**, on the other hand, showed a triplet (t) peak on H-4 with a  $J_{1,2}$  value of 9.7. This clearly indicated that anomeric functionality ( $\beta$ -STol or  $\alpha$ -OMe) significantly impacts the conformation of sugar ring and further influences the reactivity of the alcohols. Besides, it is interesting that different N-substituents on serine changes the Aka values [for example, **18** (16.9); **24** (6.71)]. The distinct reactivities were probably driven by different bond rotation on the primary alcohol.<sup>[7a,21]</sup> This hypothesis was further supported by our NMR analysis (Figure S16). For **18**,<sup>[22]</sup> the  $J$  values are 10.8 and 5.8 Hz on H- $\beta$ , and 10.8 and 4.3 Hz on H- $\beta'$ . Compound **24**,<sup>[23]</sup> on the other hands, shows the  $J$  values of 11.1 and 2.9 Hz on H- $\beta$  and  $J$  values of 11.1 and 3.3 Hz on H- $\beta'$ . A detail investigation to explain these trends is currently undergoing in our group.

Moreover, recently, fluoroalcohols, such as 2-fluoroethanol, 2,2-difluoroethanol, 2,2,2-trifluoroethanol and hexafluoroisopropanol, have been widely used in the modeling studies of glycosylation reactions due to their low reactivities,<sup>[8b,d-f,h,10c]</sup> but we are so far unable to measure their Aka. These molecules are not detectable in HPLC using UV and ELSD (evaporative light scattering detectors) detectors due to their lack of UV chromophore and low boiling points, and the determination using NMR was hampered by the overlap of peaks between the substrate and their THP product (Figures S17–S21). Although gas chromatography-mass spectrophotometer (GC-MS) can be applied to detect and quantify compounds of low polarity (such as menthol, adamantanol; Figures S8–S10), these fluoro-modified molecules are not detectable in GC-MS because of their high polarity. We are currently trying to find a solution for this difficulty.

To predict the stereoselectivity of glycosylation reactions, a database of single-step glycosylation reaction was necessary. We chose the glycosylation in dichloromethane (DCM), as it is the most commonly used solvent in glycosylation. Based on

the numbers of RRV and Aka, we analyzed our data in the previous work to investigate the stereoselective glycosylation with 12 thioglycoside donors and 4 types of hydroxyls in DCM at  $-40^{\circ}\text{C}$  under NIS/TfOH activation<sup>[8a]</sup> (Figure 4). After combining the statistics of 48 glycosylation examples, we found that in the absence of a C2 participation, the stereoselectivity is highly associated with RRV and Aka, and both  $\alpha$ - and  $\beta$ -selectivities can be clearly estimated. The

stereoselectivity was then indicated with colors. The selective  $\beta$ -glycosylations are shown in blue, while the selective  $\alpha$ -glycosylation reactions are highlighted in orange. The higher RRV and lower Aka favor  $\alpha$ -selectivity, but lower RRV and higher Aka favor  $\beta$ -selectivity. A transition from  $\alpha$ -selectivity to  $\beta$ -selectivity was observed from the bottom-right to the top-left corner in the scaffold with RRV values of 2656–4000 and Aka of 3.51–7.16.

Neighboring group effect<sup>[15d,24]</sup> and long-range participation<sup>[10c,21b,25]</sup> are known to significantly influence the stereoselectivity in glycosylation, although the real mechanism and participating level of carbonyl group are at present still unclear. To precisely analyze the correlation between the stereoselectivity and reactivity differences of donor/acceptor counterparts, we studied the trends mathematically on donors **A-M** functionalized with only non-participating groups, and the RRV and Aka were introduced and emphasized as the main parameters. Previously, our laboratory discovered that the  $\alpha$ -selectivity showed a roughly linear relationship with  $\log(\text{RRV})$  on different acceptors in a NIS/TfOH system.<sup>[8a,c]</sup> The  $\alpha$ -selectivity was found to increase with increasing reactivity of donor (RRV), while the  $\beta$ -selectivity was increased with improving nucleophilicity of acceptor (Aka). Capitalizing on this observation, we processed our database by defining  $\log(\text{RRV}/\text{Aka})$  as the X-axis, while the  $\alpha$ -selectivity as the Y-axis (Figure 5). Interestingly, we initially analyzed 37 examples of glycosylation reactions using  $\alpha$ -glycosyl imidates and the most commonly used TMSOTf catalyst system in DCM at temperatures ranging from  $-78^{\circ}\text{C}$  to  $25^{\circ}\text{C}$  (Figure 5A; Supporting Information, Table S1). Linear fitting of data led to an Equation (1) (Table 1).

$$\alpha\text{-ratio} (\%) = 17.3 \times \log(\text{RRV}/\text{Aka}); \text{RMSE} = 15\% \quad (1)$$

$$\alpha\text{-ratio} (\%) (= 100\% - \beta\text{-ratio}), R^2 = 0.76; \text{Pearson}' R = 0.87$$

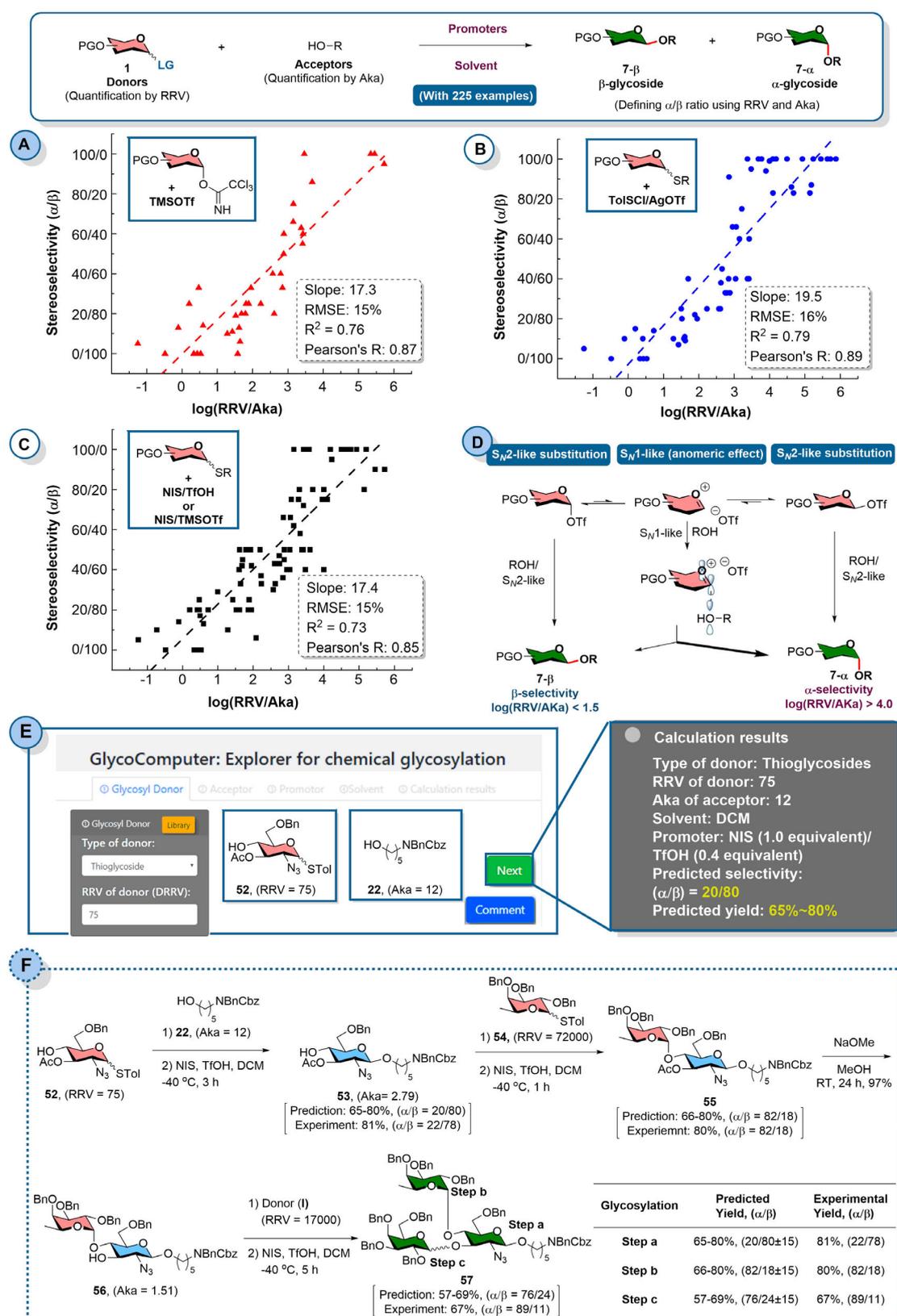
The predicted  $\alpha$ -ratio (%) can only range from 0% to 100%. Therefore, a predicted  $\alpha$ -ratio (%) below 0 (a negative number) is equivalent to exclusive  $\beta$ -selectivity (100%  $\beta$ ). Similarly, a predicted  $\alpha$ -ratio (%) exceeding 100 represents complete  $\alpha$ -selectivity (100%  $\alpha$ ). With the dataset, the RMSE was 15–16%.

Encouraged by these results, we further extended our study to a diverse combination of thioglycoside donors and acceptors in DCM at temperatures ranging from  $-78^{\circ}\text{C}$  to  $25^{\circ}\text{C}$  (Figure 5B,C; Table S1). We included the examples reported previously which were performed using the same solvent (DCM), promoter (TolSCI/AgOTf, NIS/TfOH and NIS/TMSOTf) and acceptor, and thioglycoside donors with the same sugar type and protecting group patterns but had different anomeric thio-functionalities (tolyl thioglycoside, ethyl thioglycoside, phenyl thioglycoside; the RRVs were adapted from their corresponding tolyl thioglycosides; Figure S1) despite the fact that conditions such as concentration and temperature were slightly different. As our expectations, this trend is not limited to the  $\alpha$ -imidate/TMSOTf system. Extended glycosylation studies of thioglycosides also clearly exhibit very similar trends. The RRV/Aka ratio served as a useful indicator to finalize a universal equation for a diverse combination of coupling partners. When we analyzed 176

	MeOH 15 Aka=100	6-OH 31 Aka=7.16	4-OH 37 Aka=3.51	3-OH 39 Aka=1.88
	$\alpha:\beta$ (yield)	$\alpha:\beta$ (yield)	$\alpha:\beta$ (yield)	$\alpha:\beta$ (yield)
<b>A</b> (RRV = 5.6)	1:19 (67%)	1:6 (62%)	1:4 (67%)	1:3 (56%)
<b>B</b> (RRV = 18.6)	1:9 (54%)	1:4 (63%)	1:4 (71%)	1:2.4 (50%)
<b>M</b> (RRV = 270)	0:1 (78%)	1:4 (78%)	1:3 (66%)	1:2.3 (65%)
<b>C</b> (RRV = 295)	1:4 (79%)	1:1 (70%)	1:1 (60%)	1:1 (44%)
<b>D</b> (RRV = 315)	1:5 (70%)	1:2 (66%)	1:1.4 (58%)	1:1.7 (45%)
<b>E</b> (RRV = 2656)	1:3 (95%)	1:2.3 (86%)	1:1.2 (77%)	1.7:1 (52%)
<b>F</b> RRV = 4000	1:1.5 (88%)	1:1.3 (71%)	2:1 (88%)	1.4:1 (64%)
<b>G</b> (RRV = 5000)	1:1.2 (87%)	1:1.8 (73%)	1:0 (66%)	1:0 (70%)
<b>H (8)</b> (RRV = 7180)	1:4 (53%)	1:1.5 (85%)	4:1 (64%)	1:0 (73%)
<b>I</b> (RRV = 17000)	1:2 (88%)	1:1 (90%)	4:1 (85%)	3:1 (77%)
<b>J</b> (RRV = 110000)	1:1.5 (78%)	1:0 (78%)	1:0 (71%)	1:0 (81%)
<b>K</b> (RRV = 300000)	1:1.5 (88%)	1:0 (81%)	1:0 (72%)	1:0 (65%)
<b>L</b> (RRV = 1000000)	1:1.5 (46%)	3.5:1 (74%)	9:1 (76%)	8:1 (58%)

**Figure 4.** Comparison of the stereoselective glycosylation using 12 types of thioglycoside donors (the reactivity was defined by RRV) and four types of hydroxyls (the reactivity was defined by Aka). The  $\beta$ -selectivity is shown in blue, while the  $\alpha$ -selectivity is labeled in orange.





**Figure 5.** Relationship between glycosylation stereoselectivity and RRV/Aka. A) TMSOTf activated  $\alpha$ -glycosyl imidate system in DCM. B) TolSCI/AgOTf activated thioglycoside system in DCM. C) NIS/TfOH activated thioglycoside system in DCM. D) Stereoselective glycosylation through a fast equilibrium among covalent triflate intermediate, contact ion pair, solvent-separated ion pair and oxocarbenium ion. E) GlycoComputer program navigating chemical glycosylation.<sup>[17]</sup> F) Validating the synthesis of Lewis A.

**Table 1:** Linear fitting of the data using RRV/Aka in the analysis of stereoselective glycosylation.

Donor, promoter	Equation <sup>[a]</sup>
$\alpha$ -Glycosyl imidate, TMSOTf	(1) $\alpha$ -ratio (%) = $17.3 \times \log(\text{RRV}/\text{Aka})$ ; $R^2 = 0.76$ ; Pearson's $R = 0.87$ ; RMSE = 15%
Thioglycoside, TolSCL/ AgOTf	(2) $\alpha$ -ratio (%) = $19.5 \times \log(\text{RRV}/\text{Aka}) - 3$ ; $R^2 = 0.79$ ; Pearson's $R = 0.89$ ; RMSE = 16%
Thioglycoside, NIS/ TfOH	(3) $\alpha$ -ratio (%) = $17.4 \times \log(\text{RRV}/\text{Aka}) + 5$ ; $R^2 = 0.73$ ; Pearson's $R = 0.85$ ; RMSE = 15%

[a]  $\alpha$ -Ratio [%] = 100% -  $\beta$ -ratio [%].

glycosylation reactions, of which 53 examples were acquired directly from the literature (Table S1), consistent results related to the changes in stereoselectivity were observed, including the reactions with TolSCL/AgOTf promoted thioglycosides (Figure 5B) and NIS/TfOH activated thioglycosides (Figure 5C) in DCM. The trend of distribution among the three regression lines gave similar root-mean-square error (RMSE) of about 15–16% and the Pearson's  $R$  of about 0.85–0.89. In addition, the corresponding  $R^2$  is in the range between 0.73 and 0.79, implying a general distribution existed, despite the examples from the literature were carried out in slightly different conditions. Therefore, Equation (2) (TolSCL/AgOTf) and Equation (3) (NIS/TfOH) can be introduced, respectively to summarize the stereoselectivity in the thioglycoside activation system (Table 1, entries 2 and 3).

$$\alpha\text{-ratio (\%)} = 19.5 \times \log(\text{RRV}/\text{Aka}) - 3; \text{RMSE} = 16\% \quad (2)$$

$$\alpha\text{-ratio (\%)} (= 100\% - \beta\text{-ratio}), R^2 = 0.79; \text{Pearson}' R = 0.89;$$

$$\alpha\text{-ratio (\%)} = 17.4 \times \log(\text{RRV}/\text{Aka}) + 5; \text{RMSE} = 15\% \quad (3)$$

$$\alpha\text{-ratio (\%)} (= 100\% - \beta\text{-ratio}), R^2 = 0.73; \text{Pearson}' R = 0.85;$$

Despite the differences in promoter systems, similar trends were still observed, but with slightly different sensitivities (slopes) related to the change in the reactant reactivity. It suggested that the promoter could potentially influence the distributions among the covalent intermediate (glycosyl triflate) (**4**), contact-ion pairs (**4-CIP**), solvent-separated ion pair (**4-SSIP**) and oxocarbenium ion (Figure 1B),<sup>[5,26]</sup> which could shift the  $S_N1/S_N2$  interface and leads to the stereoselectivity changes<sup>[5,7,8]</sup> (Figure 5D).

It was observed that both RRVs of donors and Akas of acceptors served as simple and successful indicators that could correlate the saccharide reactivity and stereoselectivity (Figure 5D). A clear transition from the  $\alpha$ -selectivity to the  $\beta$ -selectivity was noted in the scaffold ranging from  $\log(\text{RRV}/\text{Aka})$  of  $-1$  to  $6$ . The most favorable  $\alpha$ -glycosylation reaction ( $\alpha$ -selectivity more than 80%) was performed when  $\log(\text{RRV}/\text{Aka})$  was higher than  $4.0$ .

We then extended our study to the ACN/DCM ( $v/v = 3/1$ ) and Et<sub>2</sub>O/DCM ( $v/v = 3/1$ ) systems to analyze the solvent effect on the formation of glycosides, galactosides and 2-deoxy glucosides (44 examples, Figure S22, Table S2). The preliminary data are aligned with previous work on the study of stereoselective glycosylation affected by solvents.<sup>[6c,27]</sup> But, more data are needed to conclude a linear relationship between stereoselectivity and  $\log(\text{RRV}/\text{Aka})$

when ether (Et<sub>2</sub>O) or nitrile solvent (ACN or EtCN) is used. For glycosylation and galactosylation, in the presence of nitrile co-solvent the glycosylation was expectedly more  $\beta$ -selective than that in pure DCM. Moreover,  $\beta$ -glycosylations ( $\beta$ -selectivity > 80%) were observed when  $\log(\text{RRV}/\text{Aka})$  was lower than  $2.5$  and  $\alpha$ -glycosylation was gradually increased with increasing  $\log(\text{RRV}/\text{Aka})$  from  $2.5$  to  $6$ . In contrast, with Et<sub>2</sub>O as a co-solvent, the  $\alpha$ -selectivity was significantly elevated as compared to that in pure DCM when  $\log(\text{RRV}/\text{Aka})$  was lower than  $2.5$ .

The remote participation effect<sup>[10c,21b,25]</sup> related to stereoselective glycosylation can be preliminarily analyzed in a statistical manner (Figure S24, Table S3), although more data are necessary to elucidate a detailed trend. The results of a series of acetylated 2-azido-2-deoxy-thioglycoside (GlcN<sub>3</sub>)<sup>[25a]</sup> and 2-azido-2-deoxygalactosyl (GalN<sub>3</sub>)<sup>[25b]</sup> donors were obtained from literature. It was observed that the remote acyl participation effect did cause significant deviations from the standard trend, in agreement with the mechanistic study by Pagel, Codée, Boltje and Li.<sup>[10c,25b,c]</sup> The deviation from the non-participation slope can indicate the level of long range participation, and our preliminary statistical analysis showed that so far the C4 long-range participation on GalN<sub>3</sub> enhances the  $\alpha$ -glycosylation most significantly.

The GlycoComputer program<sup>[17]</sup> was designed to include comments from users (Figure 5E). The customer feedback and rating scale can help us understand the deficiency in this RRV/Aka platform. Moreover, human variables from different users are considered, and we would adopt the suggestions and repeat the reaction to further optimize the equation, so the precision and accuracy of the program could be further improved and the data base could be enlarged.

We also verified the GlycoComputer program by synthesizing a Lewis A skeleton **57** which contains a branched core of  $\alpha$ -Fuc-(1 $\rightarrow$ 4)- $\beta$ -GlcNAc and  $\beta$ -Gal-(1 $\rightarrow$ 3)- $\beta$ -GlcNAc with a  $\beta$ -pentyl amino linker (Figure 5F). The predicted stereoselectivities and yields were compared with practical experiments. Our GlycoComputer evaluated each building block by foreseeing the results and obviating trial and error process. According to the GlycoComputer,<sup>[17]</sup> to obtain a  $\beta$ -glucosamine glycoside in high selectivity, the donor should have a low RRV; thus, semi-protected glucosamine donor **52** (RRV = 75) was chosen for the first glycosylation with linker **22** (Aka = 12). Indeed, this NIS/TfOH promoted glycosylation in DCM furnished glucosamine **53** up to 81% with an  $\alpha/\beta$  of 22/78, which agreed well with our prediction (predicted: 65–80% yield and  $\alpha/\beta = 20/80 \pm 15$ ). Next, the subsequent  $\alpha$ -fucosylation at O4 of **53** (Aka = 2.79) required a donor with high RRV; a simple per-*O*-benzylated thiofucoside **54** (RRV = 72000) was used directly. The NIS/TfOH-promoted fucosylation resulted in an  $\alpha$ -disaccharide **55** at 80% ( $\alpha/\beta = 82:18$ ) and again matched with the GlycoComputer predicted value: 66–80%,  $\alpha/\beta = 82/18 \pm 15$ . After removal of the C3 acetyl group to give the 3-OH acceptor **56** (Aka = 1.51), the product was reacted with per-*O*-benzylated donor **I** (RRV = 17000) to give the galactosyl linkage. The GlycoComputer suggested a moderate yield (prediction:



57–69%,  $\alpha/\beta = 76/24 \pm 15$ ) and indeed, the target trisaccharide **57** was isolated in 67% with  $\alpha/\beta = 89/11$ .

Next, we carried out a random forest algorithm to understand the role of reaction counterparts (glycosyl donor and nucleophilic acceptor) involved in controlling the stereoselectivity (Figure 6).<sup>[28]</sup> RRV and Aka were obtained in a competition experiment using the model reactions (with MeOH and DHP respectively) and clear statistical trends could already be observed. Since machine learning algorithm can deal with numerous parameters simultaneously to analyze a set of complicated data, we introduced a number of “descriptors” that took the structural characterization of sugar substrates of both donor and acceptor into account. We applied 20 descriptors to analyze the stereochemistry, functionality and alcohol types (such as primary, secondary, tertiary, benzyl and aromatic) as reported (Figure S25).<sup>[29]</sup> Several additional potential factors were also applied, including selection of promoter and temperature, and a Table summarizing all these parameters is provided. The training set was established by adopting our experimental data (139 examples) in DCM at the temperatures ranging from  $-70^\circ\text{C}$  to  $-40^\circ\text{C}$ , and the result showed an RMSE of 5.4% and an  $R^2$  of 0.97. We also applied five-fold cross validation,<sup>[30]</sup> which is a common statistical approach, to test the performance. The test error rate of the training set was relatively low, of which the  $\alpha$ -selectivity showed an RMSE of 15.1% and an  $R^2$  of 0.82 (Figure 6A). Next, a separate test set (29 examples) was compiled from recent literature to validate our system. The NIS/TMSOTf promoter system was adopted into our standard promoter system of NIS/TfOH. Gladly, the random

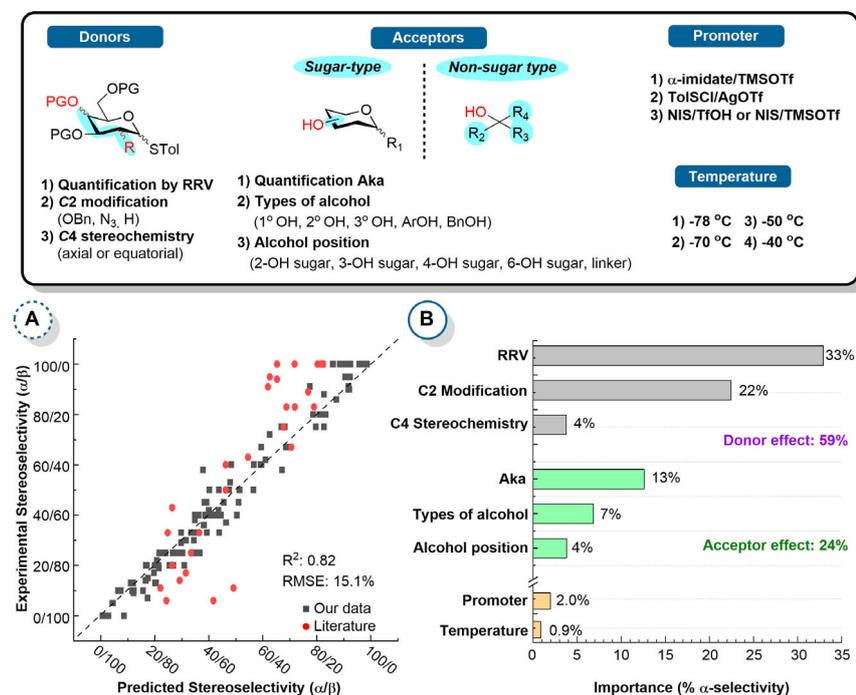
forest algorithm yielded a similar result, giving  $R^2$  of 0.81 and RMSE of 15.8%.

Regarding the importance obtained from random forest algorithm (Figure 6B; Figure S25), both RRV (33%) and Aka (13%) showed the highest impact on donor and acceptor individually, indicating quantitative assessment of reactivity can successfully represent most of the potential factors on the structural character of building block, such as the random contribution of steric, inductive and structural effects. In addition, the sum of donor effect (59%) gave a higher impact than the acceptor (24%). According to the work by Seeberger et al.,<sup>[5]</sup> a significant impact of promoter and temperature ( $-50^\circ\text{C}$  to  $+30^\circ\text{C}$ ), on the stereoselectivity of glycosylation in the automated flow reactor was observed. However, based on our RRV/Aka and random forest analysis, the influence of promoter (2.0%) and temperature (0.9%) was minor at low temperatures ranging from  $-70^\circ\text{C}$  to  $-40^\circ\text{C}$ .

## Conclusion

In summary, chemical synthesis is one of the best tools to access homogeneous oligosaccharides with well-defined configurations. However, optimizing the synthesis is complicated and highly time consuming due to our limited ability to control the glycosylation reactions, for which the stereoselectivity and yield is influenced by numerous factors. Tedious trial and error experimentation in the synthesis of oligosaccharides have long been a rate-limiting advancement in glycoscience. We have established a new index (Aka) to define the reactivity of hydroxyl groups under acidic conditions and this useful indicator can be acquired/determined through competition experiments using an automated system.

With our reactivity database (RRV of donor and Aka of acceptor) as two common numeric parameters, we can now predict the glycosylation reaction for donors bearing non-participating groups (such as benzyl or benzylidene protection). In addition, the assessment of building-block reactivity (RRV of donor and Aka of acceptor) is relatively easy through simple and fast competition experiments, and data regression does not require complicated computational calculations. The RRV and Aka take both the inductive effect and steric effect into account and systematically organize 237 glycosylation reactions in the same comparison system, and the stereoselectivity of glycosylation reaction can be further confirmed by random forest model. In the donor system without participating protecting groups, we try to include as many factors that affect the stereoselectivity and yield in the synthesis of



**Figure 6.** A) The random forest model was built with 20 descriptors, including donor effect (RRV, C2-functionalities and stereochemistry, etc.), acceptor effect (Aka, types of hydroxyls and their position etc.) and environmental effects (promoter and temperature). B) Importance related to  $\alpha$ -selectivity.

oligosaccharides, including benzyl/benzylidene donors, acceptors, activators and solvents, as possible in the predicting system. The building blocks for oligosaccharide synthesis can be more accurately selected using the RRV-Aka database in the GlycoComputer program. Since neighboring group effects,<sup>[15d,24]</sup> long-range participation<sup>[10c,21b,25]</sup> and temperature effects<sup>[5,29]</sup> are all important factors to control the stereoselectivity of glycosylation reactions, the numeric analysis of stereoselectivity change on acylated donors is now underway at various temperatures. We are also trying to connect the relationship of stereoselectivity and intermediate change using linear free energy relationships (LFER).<sup>[31]</sup>

## Acknowledgements

We thank Ping-Yu Lin (IOC, Academia Sinica) for mass measurement and Chun-Lu Cheng (IOC, Academia Sinica) for establishing the GlycoComputer. This work was supported by the Ministry of Science and Technology of Taiwan (MOST 108-2113-M-001-019-; MOST 106-2113-M-001-009-MY2) and Academia Sinica (MOST 108-3114-Y-001-002; AS-SUMMIT-109). C.-W.C. thanks MOST, Taiwan for financial support (MOST 109-2811-M-001-604-).

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** carbohydrates · diastereoselectivity · hydroxyl · glycosylation · predictive algorithms

- [1] a) S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science* **2019**, *363*, eaav2211; b) J. M. Granda, L. Donina, V. Dragone, D. L. Long, L. Cronin, *Nature* **2018**, *559*, 377–381; c) M. Panza, K. J. Stine, A. V. Demchenko, *Chem. Commun.* **2020**, *56*, 1333–1336; d) M. Panza, S. G. Pistorio, K. J. Stine, A. V. Demchenko, *Chem. Rev.* **2018**, *118*, 8105–8150.
- [2] a) D. Caramelli, D. Salley, A. Henson, G. A. Camarasa, S. Sharabi, G. Keenan, L. Cronin, *Nat. Commun.* **2018**, *9*, 3406; b) C. W. Coley, W. H. Green, K. F. Jensen, *Acc. Chem. Res.* **2018**, *51*, 1281–1289; c) K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, *559*, 547–555; d) I. W. Davies, *Nature* **2019**, *570*, 175–181.
- [3] a) S. S. Pinho, C. A. Reis, *Nat. Rev. Cancer* **2015**, *15*, 540–555; b) T. J. Boltje, T. Buskas, G.-J. Boons, *Nat. Chem.* **2009**, *1*, 611; c) J. Munkley, *Endocr.-Relat. Cancer* **2017**, *24*, R49–R64; d) A. Varki, *Glycobiology* **2017**, *27*, 3–49; e) L. Krasnova, C.-H. Wong, *Annu. Rev. Biochem.* **2016**, *85*, 599–630; f) S. J. Danishefsky, Y.-K. Shue, M. N. Chang, C.-H. Wong, *Acc. Chem. Res.* **2015**, *48*, 643–652.
- [4] a) W.-L. Leng, H. Yao, J.-X. He, X.-W. Liu, *Acc. Chem. Res.* **2018**, *51*, 628–639; b) S. S. Kulkarni, C.-C. Wang, N. M. Sabbarvarapu, A. R. Podilapu, P.-H. Liao, S.-C. Hung, *Chem. Rev.* **2018**, *118*, 8025–8104; c) M. M. Nielsen, C. M. Pedersen, *Chem. Rev.* **2018**, *118*, 8285–8358; d) H. Yao, M. D. Vu, X.-W. Liu, *Carbohydr. Res.* **2019**, *473*, 72–81; e) S. S. Park, H. W. Hsieh, J. Gervay-Hague, *Molecules* **2018**, *23*, 1742; f) R. Das, B. Mukhopadhyay, *ChemistryOpen* **2016**, *5*, 401–433; g) Y. Yang, X. Zhang, B. Yu, *Nat. Prod. Rep.* **2015**, *32*, 1331–1355.
- [5] S. Chatterjee, S. Moon, F. Hentschel, K. Gilmore, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 11942–11953.
- [6] a) S. K. Mulani, W.-C. Hung, A. B. Ingle, K.-S. Shiau, K.-K. T. Mong, *Org. Biomol. Chem.* **2014**, *12*, 1184–1197; b) C.-S. Chao, C.-W. Li, M.-C. Chen, S.-S. Chang, K.-K. T. Mong, *Chem. Eur. J.* **2009**, *15*, 10972–10982; c) K.-K. T. Mong, T. Nokami, N. Thi Thanh Tran, P. Be Nhi, *Selective Glycosylations: Synthetic Methods and Catalysts* (Ed.: C. S. Bennett), Wiley, Hoboken, **2017**, chap. 3, pp. 59–77.
- [7] a) P. O. Adero, H. Amarasekara, P. Wen, L. Bohé, D. Crich, *Chem. Rev.* **2018**, *118*, 8242–8284; b) M. Huang, G. E. Garrett, N. Birlirakis, L. Bohé, D. A. Pratt, D. Crich, *Nat. Chem.* **2012**, *4*, 663–667; c) T. G. Frihed, M. Bols, C. M. Pedersen, *Chem. Rev.* **2015**, *115*, 4963–5013; d) P. O. Adero, T. Furukawa, M. Huang, D. Mukherjee, P. Retailleau, L. Bohé, D. Crich, *J. Am. Chem. Soc.* **2015**, *137*, 10336–10345; e) L. Bohé, D. Crich, *Carbohydr. Res.* **2015**, *403*, 48–59; f) A. G. Santana, L. Montalvillo-Jimenez, L. Diaz-Casado, F. Corzana, P. Merino, F. J. Canada, G. Jimenez-Oses, J. Jimenez-Barbero, A. M. Gomez, J. L. Asensio, *J. Am. Chem. Soc.* **2020**, *142*, 12501–12514; g) J. L. Asensio, A. G. Santana, L. Montalvillo-Jimenez, L. Diaz-Casado, E. Mann, J. Jimenez-Barbero, A. M. Gomez, *Chem. Eur. J.* **2020**, *27*, 2032–2042; h) M. Huang, P. Retailleau, L. Bohé, D. Crich, *J. Am. Chem. Soc.* **2012**, *134*, 14746–14749.
- [8] a) C.-W. Chang, C.-H. Wu, M.-H. Lin, P.-H. Liao, C.-C. Chang, H.-H. Chuang, S.-C. Lin, S. Lam, V. P. Verma, C.-P. Hsu, C.-C. Wang, *Angew. Chem. Int. Ed.* **2019**, *58*, 16775–16779; *Angew. Chem.* **2019**, *131*, 16931–16935; b) S. van der Vorm, T. Hansen, J. M. A. van Hengst, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Soc. Rev.* **2019**, *48*, 4688–4706; c) C.-W. Chang, M.-H. Lin, C.-C. Wang, *Chem. Eur. J.* **2021**, *27*, 2556–2568; d) S. van der Vorm, T. Hansen, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Sci.* **2017**, *8*, 1867–1875; e) S. van der Vorm, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *J. Org. Chem.* **2017**, *82*, 4793–4811; f) S. van der Vorm, J. M. A. van Hengst, M. Bakker, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Angew. Chem. Int. Ed.* **2018**, *57*, 8240–8244; *Angew. Chem.* **2018**, *130*, 8372–8376; g) M. H. Zhuo, D. J. Wilbur, E. E. Kwan, C. S. Bennett, *J. Am. Chem. Soc.* **2019**, *141*, 16743–16754; h) T. Hansen, T. P. Ofman, J. G. C. Vlaming, I. A. Gagarinov, J. van Beek, T. A. Gote, J. M. Tichem, G. Ruijgrok, H. S. Overkleeft, D. V. Filippov, G. A. van der Marel, J. D. C. Codée, *Angew. Chem. Int. Ed.* **2021**, *60*, 937–945; *Angew. Chem.* **2021**, *133*, 950–958.
- [9] a) O. Takahashi, K. Yamasaki, Y. Kohno, R. Ohtaki, K. Ueda, H. Suezawa, Y. Umezawa, M. Nishio, *Carbohydr. Res.* **2007**, *342*, 1202–1209; b) I. Cumpstey, *Org. Biomol. Chem.* **2012**, *10*, 2503–2508; c) R. J. Meredith, R. J. Woods, I. Carmichael, A. S. Serianni, *Phys. Chem. Chem. Phys.* **2020**, *22*, 14454–14457.
- [10] a) T. Hansen, L. Lebedel, W. A. Remmerswaal, S. van der Vorm, D. P. A. Wander, M. Somers, H. S. Overkleeft, D. V. Filippov, J. Desire, A. Mingot, Y. Blieriot, G. A. van der Marel, S. Thibaut-deau, J. D. C. Codée, *ACS Cent. Sci.* **2019**, *5*, 781–788; b) S. van der Vorm, T. Hansen, E. R. van Rijssel, R. Dekkers, J. M. Madern, H. S. Overkleeft, D. V. Filippov, G. A. van der Marel, J. D. C. Codee, *Chem. Eur. J.* **2019**, *25*, 7149–7157; c) T. Hansen, H. Elferink, J. M. A. van Hengst, K. J. Houthuijs, W. A. Remmerswaal, A. Kromm, G. Berden, S. van der Vorm, A. M. Rijs, H. S. Overkleeft, D. V. Filippov, F. Rutjes, G. A. van der Marel, J. Martens, J. Oomens, J. D. C. Codée, T. J. Boltje, *Nat. Commun.* **2020**, *11*, 2664.
- [11] a) L. K. Mydock, A. V. Demchenko, *Org. Biomol. Chem.* **2010**, *8*, 497–510; b) P. Konradsson, U. E. Udodong, B. Fraser-Reid, *Tetrahedron Lett.* **1990**, *31*, 4313–4316; c) H. H. Jensen, C. M. Pedersen, M. Bols, *Chem. Eur. J.* **2007**, *13*, 7576–7582; d) C. M.





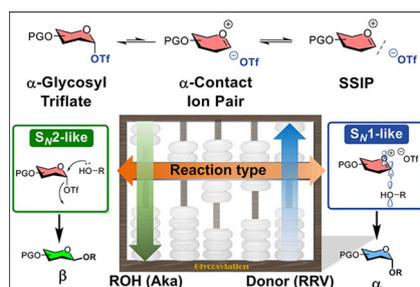
## Research Articles



## Glycoscience

C.-W. Chang, M.-H. Lin, C.-K. Chan,  
K.-Y. Su, C.-H. Wu, W.-C. Lo, S. Lam,  
Y.-T. Cheng, P.-H. Liao, C.-H. Wong,\*  
C.-C. Wang\*     

Automated Quantification of Hydroxyl  
Reactivities: Prediction of Glycosylation  
Reactions



A so-called “GlycoComputer” program has been developed to foresee and predict the yield and stereoselectivity of glycosylation reactions based on the properties of various donors, acceptors, activation systems and solvents. The program statistically analyzes and compares the relative reactivity value (RRV) of donors and the acceptor nucleophilic constant (Aka) of acceptors.