

Cite this: *Chem. Sci.*, 2025, 16, 15056 All publication charges for this article have been paid for by the Royal Society of ChemistryReceived 5th May 2025  
Accepted 7th July 2025

DOI: 10.1039/d5sc03244j

rsc.li/chemical-science

# Discovery of novel glycosylation methods using Bayesian optimization: lithium salt directed stereoselective glycosylations†

Natasha Videcrantz Faurischou  and Christian Marcus Pedersen \*

In recent years, Bayesian optimization has gained increasing interest as a tool for reaction optimization. Here we use Bayesian optimization in a reaction discovery fashion by treating the glycosylation reaction class as a black box function. This provides access to new areas of the glycosylation reaction space and leads to the discovery of novel stereoselective glycosylation methodologies, where stereoselectivity can be directed by the addition of lithium salts in interplay with other reaction conditions. Black box functions are inherently difficult to interpret, but we show how partial dependence plots can be used to infer trends from the obtained data in a similar fashion to the commonly used one-variable-at-time approach.

## Introduction

Reaction discovery and the development of new synthetic methodologies are core topics within organic chemistry. A typical academic workflow for reaction discovery is depicted in Fig. 1. The lead reaction is often found through sheer serendipity or hypotheses based on chemical rationalization. More recently, developments in the field of analytic chemistry, automatization, and artificial intelligence have allowed high-throughput experimentation (HTE) and machine learning (ML) to aid in the search for novel reactivity.<sup>1–6</sup> Despite a constant broadening of our understanding of reaction mechanisms and the influence of various reaction conditions on these, most proposed mechanisms are highly simplified. This makes the rationalization and prediction of undiscovered reactivity challenging and most often mechanisms are therefore rationalized retrospectively. When a lead reaction is discovered it is optimized for yield, selectivity, or other desirable parameters. In academia, the most common strategy for reaction optimization is the one-variable-at-a-time (OVAT) approach, where statistical strategies like design of experiment (DOE) are more widespread in industry.<sup>7</sup> Besides assisting in finding optimal reaction conditions, the OVAT approach is useful for understanding the influence of individual reaction parameters. Since only one reaction parameter is varied at a time, it is easy to analyze trends and try to give them chemical meaning *e.g.* relating a change in the outcome when changing the solvent to the polarity of the solvent. Recently, Bayesian optimization (BO)

has been successfully applied for the optimization of multiple reactions.<sup>7–13</sup> Once the optimal reaction conditions are identified, the scope of the established methodology is explored by testing different combinations of substrates. Lastly, the mechanism is often discussed based on the findings from the reaction optimization and scope exploration, and in some cases, additional experiments will be carried out to gain a deeper mechanistic insight.

New tools for discovering lead reactions for novel methodologies are desirable, especially in cases where rational design can be difficult due to complex mechanisms. An example of a reaction where our understanding of the fundamental reaction mechanism limits the rational design of new methodologies is the glycosylation reaction. One of the main challenges when designing glycosylations is controlling the anomeric selectivity, which is highly important for biological function.<sup>14–16</sup>

Mechanistic understanding of the glycosylations reaction can help predict and guide the anomeric selectivity, and multiple mechanistic studies of glycosylations have been conducted.<sup>17–20</sup> In the simplest scenario, the glycosylation reaction is considered an S<sub>N</sub>1-reaction with formation of a relatively stable oxocarbenium ion (Fig. 2 top). However, it is well-known that this is a very simplified view of the reaction mechanism, and a lot of work has gone into understanding the influence of different reaction conditions and substrate effects.<sup>21</sup> Much work has also gone into trying to identify intermediates, both covalent adducts and ion pairs, formed during the reaction.<sup>17,18,22–25</sup>

Despite many detailed investigations, the general understanding of the glycosylation reaction is limited, and advanced mechanistic scenarios are only described for specific activator/leaving group systems.<sup>17–20</sup> As seen from Fig. 2 (advanced mechanism), the mechanism gets increasingly complicated

Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark. E-mail: cmp@chem.ku.dk

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5sc03244j>



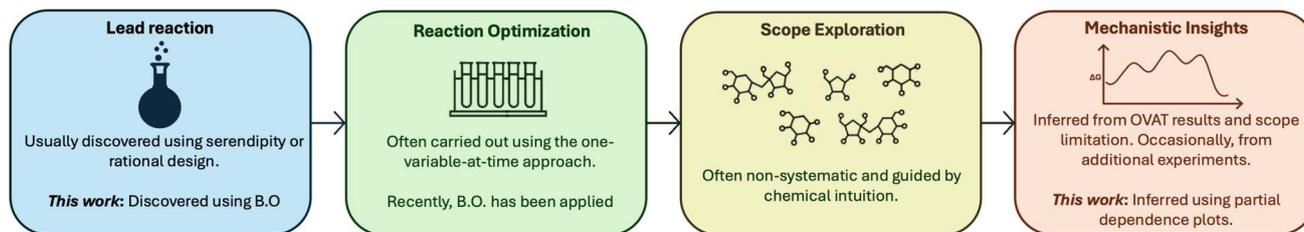


Fig. 1 A typical workflow for reaction discovery. First, a lead reaction is discovered, and then the reaction conditions are optimized to maximize yield, selectivity, etc. Next, the reaction scope for the methodology is explored, and the mechanism is rationalized in hindsight.

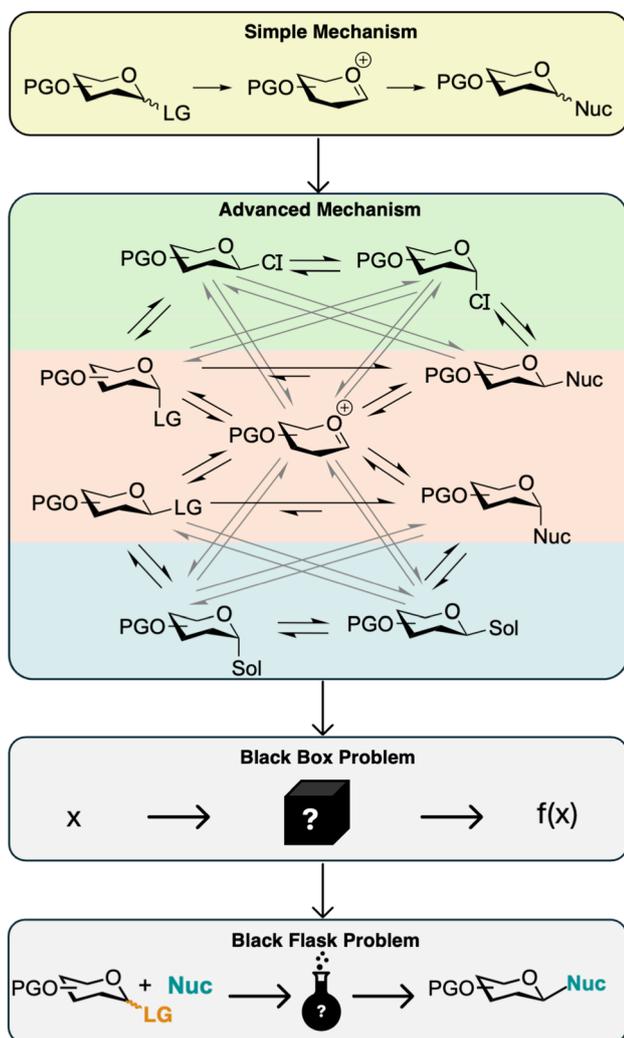


Fig. 2 Top: A simple commonly accepted mechanism for the glycosylation reaction displayed. Below is a more advanced mechanism depicted, which more closely resembles the true reaction path with multiple species involved and all in dynamic equilibria. Both solvent and counter ions (CIs) can participate in the formation of intermediates. However, it is still a simplification and understanding the relationship between these equilibria is extremely difficult. The glycosylation reaction can therefore be viewed as a black box problem, or more aptly, a black flask problem.

when including more possible intermediates. In red is highlighted the “classic” glycosylation mechanism, where the glycosylation reaction is viewed as a nucleophilic substituent reaction proceeding through either a more  $S_N1$ -like mechanism, a more  $S_N2$ -like mechanism, or both in competition. Figuring out where on the  $S_N1/S_N2$ -spectrum a specific glycosylation belongs is in itself challenging, and this will be dependent on both the substrates and conditions.<sup>21,26</sup> In green, the formation of intermediates through reaction with a counter ion of the activator is also considered, here drawn as covalent adducts, but ion pairs are also known to be involved. Examples of such intermediates include glycosyl chlorides<sup>27</sup> and glycosyl triflates.<sup>18</sup> In blue, intermediates formed by reaction with the solvent are included, further complicating the mechanism. The advanced mechanism shown in Fig. 1 is still a simplified picture and does for instance not consider pathways with anchimeric assistance or contact ion pairs. To rationalize the outcome of glycosylations we would have to determine the relationship between all of these equilibria, but as of now, we do not have any way for assessing their individual contribution and co-dependence. Thus, a holistic understanding of the glycosylation mechanism might be impossible given our current tools. The glycosylation reaction can therefore be described as a black box/flask function (Fig. 2), that is, if we put in  $x$  (substrates and reaction condition) we get an outcome,  $f(x)$  (yield and stereo-selectivity), but our understanding of how  $x$  becomes  $f(x)$  is highly limited. We therefore chose to treat the glycosylation reaction and its mechanism as a “black flask” problem and carry out a multiobjective optimization of the glycosylation reaction class by utilizing BO to try to discover new stereo-selective glycosylation methodologies. As mentioned earlier, BO has in recent years been extensively applied to the reaction optimization part of the reaction discovery pipeline and also recently using a more discovery-driven approach for designing new materials<sup>28–30</sup> and new catalysts.<sup>31–34</sup> BO efficiently explores complex, high-dimensional spaces with limited and noisy data, making it an ideal strategy for advanced chemical systems. We envisioned BO could help in designing new glycosylation methodologies, thus shifting the application of BO from pure reaction optimization towards lead discovery in Fig. 2 by identifying new glycosylation strategies. Additionally, we show how trends for specific reaction parameters can be inferred and analyzed from the BO campaign data in a similar fashion to the analyses of OVAT data. This is done using partial dependence



plots, thereby overcoming one of the obstacles of using BO compared to OVAT.

## Results and discussion

### Design of experimental setup

The reaction discovery campaigns were run using a human-in-the-loop setup. A modified version of the Bayesian optimization algorithm ProcessOptimizer<sup>35–37</sup> was used to suggest the experiments. This algorithm has previously been used for reaction optimization<sup>8</sup> and can take both continuous and discrete variables as input. The algorithm has been modified to incorporate variable constraints for multiobjective optimizations. As the GlycoOptimizer is inherently a minimizer, the objectives have been modified accordingly *i.e.* 100 – objective in percentage. This modified algorithm will in the following be referred to as the GlycoOptimizer. Experiments and workup were carried out by hand (details can be found in ESI Section 2.3†). The objectives, *i.e.* yield and anomeric selectivity, were evaluated by NMR analysis using an internal standard. The experimental setup is illustrated in Fig. 3A. The campaign was initiated by a batch of 10 random experiments suggested by the GlycoOptimizer. The results from these were fed to the GlycoOptimizer which then proposed a batch of 5 new experiments. The experiments were proposed either using an estimated Pareto Front<sup>38</sup> (exploitation) or Steinerberger-sampling<sup>39</sup> (exploration), with a chance of Steinerberger-sampling being used of 25%. The results inferred from NMR for the proposed experiments were fed back to the optimizer, which suggested 5 new experiments and so forth. It should be noted that due to

measurement limitations, the conditions under which the experiments were carried out were not always an exact match for the conditions proposed by the GlycoOptimizer with regards to equivalents and concentration and the conditions being fed back to the optimizer were the actual conditions the experiments had been carried out under.

### Design of model reaction and reaction space

Fig. 3B shows the model reaction and reaction space. The reactants are perbenzylated glucosyl trichloroacetimidate (TCA) and *L*-menthol as the glycosyl donor and glycosyl acceptor, respectively. A perbenzylated glycosyl donor was chosen to avoid neighboring group participation (NGP) and remote participation as we were interested in developing a method where the stereoselectivity is reagent-controlled rather than substrate-dependent. The glycosyl donor was chosen to be a TCA as TCAs are easy and cheap to synthesize from the hemiacetal, trichloroacetonitrile, and base catalyst.<sup>40,41</sup> Additionally they are relatively stable and each anomer can be selectively synthesized by the choice of base.<sup>40</sup> *L*-Menthol is a commonly used glycosyl acceptor in model glycosylation reactions,<sup>42–44</sup> as it shares similarities with free secondary alcohol on a monosaccharide.

When selecting the reaction space, we aimed to include as many parameters as possible that influence glycosylation outcomes. In total 11 parameters were chosen as shown in Fig. 3B. All the parameters are either represented as integers or continuous variables.

The TCA-donor configuration,  $\alpha$  or  $\beta$ , was included to take into account that glycosylations can be stereospecific.<sup>45–47</sup> TCAs

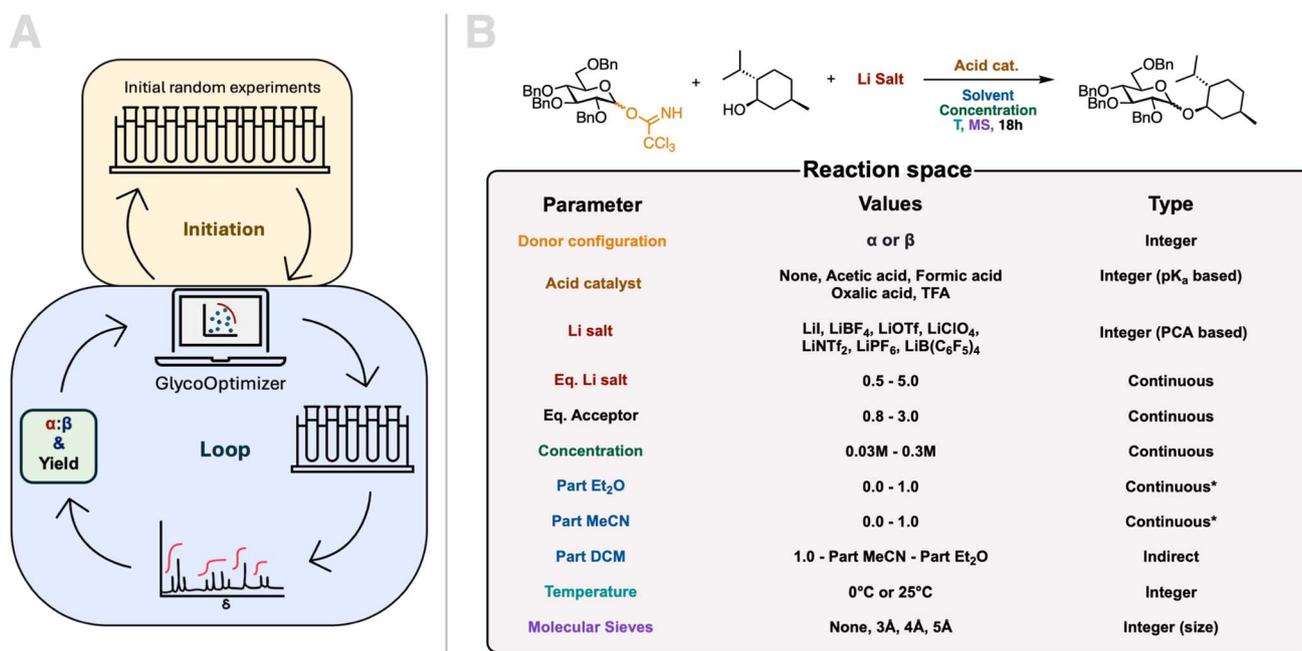


Fig. 3 (A) An illustration of the experimental optimization loop and initiation. The first batch consists of 10 randomly suggested experiments, and the results are fed to the optimizer which proposes a batch of five new experiments using a Bayesian optimization algorithm. The results (anomeric selectivity and yield) are obtained by NMR analysis. (B) Illustration of model reaction and reaction space available for the GlycoOptimizer. The values and the representation of the reaction parameters are indicated.



are most commonly activated by acid catalysis, often using strong acids, but milder acids have also been shown to be sufficient.<sup>45,48,49</sup> We chose to include acids with  $pK_a$ s in the range of 4.8 to 0.2 represented as integers assigned according to acidity, and also with the option of no acid. We avoided stronger acids as we wanted the conditions to be as mild as possible, improving the possibility for upscale and reproducibility by non-experts.

It has been shown that the counterion of the acid can play a role in the outcome of glycosylations with regard to yield and selectivity.<sup>17,50,51</sup> To mimic the counterion effect this a lithium salt was added, and the salts were assigned an integer according to a principle component analysis (PCA). Details on the PCA can be found in ESI (Section 3).<sup>†</sup> Both concentration,<sup>42,52</sup> temperature,<sup>42,53</sup> and solvent<sup>42,51,54</sup> are also known to be important and were included as input parameters. The most well-known solvent effects within carbohydrate chemistry are the ether effect<sup>54</sup> and the nitrile effect.<sup>54</sup> Thus we chose a three-part solvent system to take these into account, with both part Et<sub>2</sub>O and part MeCN being input variables with the sum of these constrained to equal to or less than 1. If the sum is less than one, the remaining part solvent will be DCM, thus part DCM is included as an indirect variable. Temperature is included in the reaction space as a discrete variable and not a continuous variable since each temperature requires a separate reaction station. The reactions were either carried out at 25 °C or in a fridge with a temperature of 0 °C, which are the most common reaction temperatures.<sup>49</sup> The presence and the size of molecular sieves have also been shown to affect the outcome of

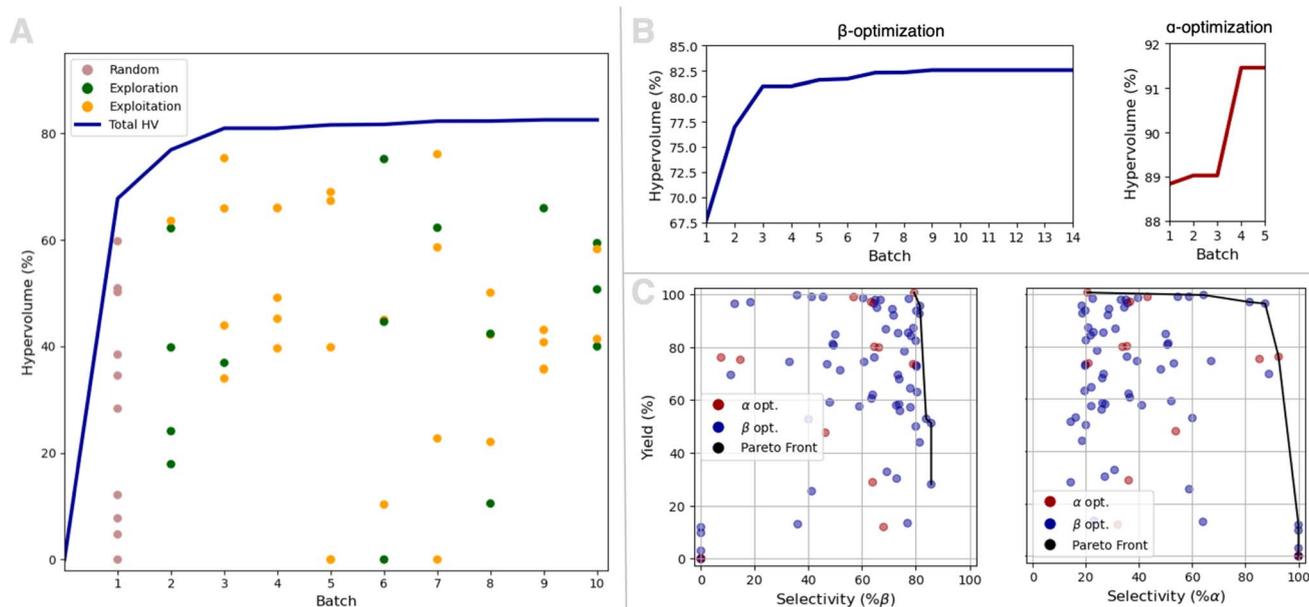
glycosylations,<sup>51</sup> and were therefore also added as an input parameter as integers according to size.

### Yield and stereoselectivity optimization campaigns

The first campaign aimed to optimize the yield and  $\beta$ -selectivity of the glycosylation through multiobjective optimization. In total 10 loops were carried out including the initiation batch with 10 random experiments. The results from each batch are shown in Fig. 4A as the total hypervolume and each experiment's hypervolume contribution. Hypervolumes are a way of evaluating multiobjective optimizations,<sup>55</sup> and a hypervolume contribution of 100% corresponds to 100% yield and 100% stereoselectivity.

It is seen from Fig. 4A that after batch 3 only minor improvements to the total hypervolume are observed. In general, the experiments selected using the exploitative algorithm seem to have the highest hypervolume contributions, while the experiments selected using the more explorative algorithm are more scattered.

After the first 10 batches, it seemed that the optimization was near convergence, but we still envisioned that minor improvements might be possible. However, we were also interested in running a yield and  $\alpha$ -selectivity optimization campaign, to see if we could also find a stereoselective procedure for obtaining the more challenging 1,2-*cis*-glycoside. We, therefore, decided to run a dual optimization campaign still with batches of 5 experiments, but with only two experiments proposed by the yield and  $\beta$ -selectivity optimizer. The last three experiments



**Fig. 4** (A) Results from yield and  $\beta$ -selectivity optimization campaign. The first batch consists of 10 random experiments, and the other batches consist of 5 experiments suggested by the GlycoOptimizer based on the previous experiments either by estimating the Pareto Front (exploitation) or Steinberger sampling (exploration). The blue line shows the total hypervolume for all experiments and the dots indicate the hypervolume contribution for each experiment. (B) Left: convergence plot for yield and  $\beta$ -selectivity optimization with total hypervolume after each batch. Right: convergence plot for yield and  $\alpha$ -selectivity optimization with total hypervolume after each batch and the first batch being all the experiments from the 10 first batches from the  $\beta$ -selectivity campaign. (C) The objectives for both campaigns are plotted against each other with the estimated Pareto front highlighted.



**Table 1** Conditions and results for the experiments carried out during the first optimization campaign optimizing for yield and  $\beta$ -selectivity. Each batch consists of five experiments. Each experiments hypervolume contribution (HV contr.) is given

Exp. no.	Conf.	Li salt	Li salt eq.	Acid	Acceptor eq.	Conc. (M)	Part EtO <sub>2</sub>	Part MeCN	M. S.	Temp (°C)	Yield (%)	Ratio ( $\beta$ %)	HV contr. (%)
1	$\alpha$	LiPF <sub>6</sub>	3.4	Oxalic	1.7	0.18	0.51	0.05	3 Å	25	87	69	60
2	$\beta$	LiI	1.5	Acetic	1.3	0.3	0.29	0.06	4 Å	25	96	13	12
3	$\beta$	LiI	3.5	TFA	2.8	0.26	0.51	0.08	3 Å	0	69	11	8
4	$\alpha$	LiI	1.6	None	1.9	0.27	0.11	0.41	3 Å	25	13	36	5
5	$\alpha$	LiClO <sub>4</sub>	1.7	Acetic	1.4	0.25	0.11	0.81	3 Å	0	64	78	50
6	$\beta$	LiNTf <sub>2</sub>	2.2	Oxalic	1.7	0.19	0.41	0.08	3 Å	0	61	63	38
7	$\beta$	LiClO <sub>4</sub>	1.7	None	1.8	0.18	0.19	0.11	None	0	59	48	28
8	$\beta$	LiClO <sub>4</sub>	3	Acetic	2.2	0.25	0.56	0.16	4 Å	25	74	47	35
9	$\alpha$	LiBF <sub>4</sub>	2.6	Oxalic	3	0.22	0.75	0.18	4 Å	25	69	73	51
10	$\alpha$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	1	Acetic	2.1	0.24	0.65	0.28	5 Å	25	10	0	0
11	$\alpha$	LiPF <sub>6</sub>	2.8	Oxalic	1.5	0.17	0.33	0.04	None	25	97	18	18
12	$\beta$	LiOTf	1.5	Formic	3	0.09	0.44	0.15	None	25	81	49	40
13	$\beta$	LiNTf <sub>2</sub>	3.2	Formic	0.8	0.11	0.15	0.6	4 Å	0	28	86	24
14	$\alpha$	LiOTf	4.6	Oxalic	1.9	0.13	0.05	0.9	4 Å	25	97	64	62
15	$\beta$	LiPF <sub>6</sub>	4.1	Acetic	2.2	0.12	0.8	0.12	3 Å	25	98	65	64
16	$\alpha$	LiNTf <sub>2</sub>	4.8	TFA	1.2	0.12	0.69	0.28	4 Å	0	84	78	66
17	$\alpha$	LiPF <sub>6</sub>	5	Oxalic	2.5	0.03	0.09	0.87	5 Å	25	93	81	75
18	$\alpha$	LiClO <sub>4</sub>	5	Formic	1.2	0.28	0.86	0.12	3 Å	25	71	52	37
19	$\alpha$	LiBF <sub>4</sub>	4.1	Acetic	2.4	0.1	0.98	0.01	5 Å	0	58	59	34
20	$\alpha$	LiNTf <sub>2</sub>	1.8	None	2.8	0.07	0.31	0.15	3 Å	0	51	86	44
21	$\alpha$	LiBF <sub>4</sub>	0.5	Formic	1.1	0.18	0.17	0.53	3 Å	25	82	80	66
22	$\alpha$	LiBF <sub>4</sub>	2.1	Formic	2.3	0.06	0.82	0.04	3 Å	0	62	64	40
23	$\beta$	LiOTf	4	TFA	2.5	0.21	0.38	0.49	5 Å	0	76	65	49
24	$\alpha$	LiPF <sub>6</sub>	4.2	Acetic	2.6	0.15	0.05	0.44	None	0	92	72	66
25	$\alpha$	LiOTf	3.1	Formic	1.7	0.05	0.13	0.23	5 Å	25	74	61	45
26	$\beta$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	2.5	TFA	1.3	0.08	0.1	0.7	5 Å	25	0	0	0
27	$\beta$	LiOTf	3.3	Oxalic	1	0.07	0.52	0.12	4 Å	25	81	49	40
28	$\beta$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	1.5	Formic	2.4	0.16	0.55	0.34	4 Å	0	0	0	0
29	$\beta$	LiBF <sub>4</sub>	4.5	Oxalic	1.4	0.26	0.55	0.38	3 Å	25	94	71	67
30	$\alpha$	LiClO <sub>4</sub>	3.7	Oxalic	2.3	0.23	0.12	0.56	3 Å	25	87	79	69
31	$\beta$	LiClO <sub>4</sub>	3.2	Oxalic	2.5	0.13	0.63	0.34	None	25	99	45	45
32	$\beta$	LiNTf <sub>2</sub>	4.2	Oxalic	2.7	0.28	0.46	0.25	4 Å	25	13	77	10
33	$\beta$	LiNTf <sub>2</sub>	1.2	Oxalic	2.9	0.21	0.04	0.5	4 Å	0	94	80	75
34	$\alpha$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	3	TFA	1	0.19	0.03	0.15	4 Å	25	12	0	0
35	$\beta$	LiPF <sub>6</sub>	4.1	Formic	2.1	0.05	0.57	0.39	5 Å	25	57	78	45
36	$\alpha$	LiOTf	3.6	Acetic	2.5	0.17	0.28	0.27	3 Å	0	95	66	62
37	$\beta$	LiPF <sub>6</sub>	3.1	Acetic	1.8	0.09	0.4	0.55	4 Å	0	98	77	76
38	$\beta$	LiClO <sub>4</sub>	2.5	None	1.3	0.14	0.43	0.49	3 Å	25	3	0	0
39	$\alpha$	LiBF <sub>4</sub>	2.5	Oxalic	2.2	0.14	0.29	0.63	4 Å	0	73	80	59
40	$\alpha$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	1.5	Formic	2.1	0.24	0.87	0.06	4 Å	0	33	69	23
41	$\alpha$	LiClO <sub>4</sub>	0.8	Acetic	2.5	0.27	0.13	0.71	4 Å	25	58	73	42
42	$\alpha$	LiBF <sub>4</sub>	2.2	Acetic	1	0.27	0.54	0.29	5 Å	0	68	74	50
43	$\alpha$	LiOTf	1.8	Acetic	1.9	0.16	0.09	0.47	4 Å	0	30	73	22
44	$\beta$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	1.3	None	2.9	0.24	0.21	0.62	5 Å	0	26	41	11
45	$\beta$	LiClO <sub>4</sub>	1.8	Oxalic	1.7	0.31	0.7	0.27	4 Å	25	85	50	42
46	$\beta$	LiPF <sub>6</sub>	2.7	Formic	2.5	0.2	0.13	0.48	None	25	99	41	41
47	$\alpha$	LiBF <sub>4</sub>	0.8	TFA	1.4	0.09	0.38	0.06	3 Å	25	58	74	43
48	$\alpha$	LiBF <sub>4</sub>	3.5	Acetic	1.4	0.21	0.11	0.73	4 Å	25	85	77	66
49	$\beta$	LiClO <sub>4</sub>	5	Formic	1.5	0.18	0.53	0.21	5 Å	25	100	36	36
50	$\alpha$	LiNTf <sub>2</sub>	3.2	Formic	1	0.22	0.33	0.31	3 Å	25	44	81	36
51	$\alpha$	LiPF <sub>6</sub>	2.7	Acetic	2.3	0.03	0	0.99	5 Å	0	50	80	40
52	$\alpha$	LiPF <sub>6</sub>	1.3	Formic	1.2	0.11	0.03	0.11	3 Å	0	63	80	51
53	$\beta$	LiClO <sub>4</sub>	0.8	Acetic	2.4	0.1	0.45	0.43	3 Å	0	56	74	41
54	$\alpha$	LiBF <sub>4</sub>	3.2	Formic	2.8	0.14	0.57	0.29	3 Å	0	78	76	59
55	$\alpha$	LiBF <sub>4</sub>	1.1	TFA	2.5	0.06	0.02	0.36	4 Å	0	73	80	58

were proposed by a new yield and  $\alpha$ -selectivity optimizer, and after each loop, the results from all 5 experiments were fed to both optimizers. The yield and  $\alpha$ -selectivity optimization was initiated using all the data obtained from the first campaign. All

5 experiments in each batch were chosen using Pareto front sampling. From Fig. 4B it is seen that the total hypervolume for the yield and  $\beta$ -selectivity does not improve after the initial 10 batches *i.e.* does not improve during the second dual campaign.



**Table 2** Conditions and results for the experiments carried out during the second dual optimization campaign. Each batch consists of five experiments. The objectives for the first two experiments in each batch are yield and  $\beta$ -selectivity, whereas the objectives for the remaining three experiments (shaded) are yield and  $\alpha$ -selectivity

Exp. no.	Conf.	Li salt	Li salt eq.	Acid	Acceptor eq.	Conc. (M)	Part EtO <sub>2</sub>	Part MeCN	M. S.	Temp (°C)	Yield (%)	Ratio ( $\beta$ ) (%)	HV <sup>a</sup> contr. (%)
56	$\alpha$	LiPF <sub>6</sub>	4.1	Formic	2.5	0.03	0.01	0.27	None	25	74	33	24 (50)
57	$\alpha$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	2.6	Oxalic	1.4	0.1	0.26	0.22	5 Å	25	0	0	0 (0)
58	$\beta$	LiClO <sub>4</sub>	2.9	Oxalic	2.3	0.28	0.06	0.06	3 Å	25	99	57	56 (43)
59	$\beta$	LiI	4.5	Formic	1.4	0.08	0.31	0.31	4 Å	0	75	15	11 (64)
<b>60</b>	<b><math>\beta</math></b>	<b>LiNTf<sub>2</sub></b>	<b>4.8</b>	<b>Oxalic</b>	<b>1.6</b>	<b>0.05</b>	<b>0.09</b>	<b>0.47</b>	<b>3 Å</b>	<b>25</b>	<b>101</b>	<b>79</b>	<b>80</b> (21)
61	$\alpha$	LiNTf <sub>2</sub>	1	TFA	2	0.17	0.22	0.31	4 Å	0	53	84	44 (8)
62	$\alpha$	LiClO <sub>4</sub>	3.6	TFA	0.9	0.06	0.73	0.2	None	0	53	40	21 (32)
63	$\alpha$	LiB(C <sub>6</sub> F <sub>5</sub> ) <sub>4</sub>	1.6	Oxalic	1.1	0.23	0.48	0.19	4 Å	25	0	0	0 (0)
64	$\alpha$	LiOTf	3.9	TFA	1.9	0.3	0.24	0.66	3 Å	25	80	66	53 (27)
65	$\beta$	LiOTf	4	None	2	0.2	0.44	0.15	4 Å	0	48	46	22 (26)
66	$\alpha$	LiPF <sub>6</sub>	4.5	Acetic	2.1	0.13	0.58	0.1	4 Å	25	85	73	63 (23)
67	$\beta$	LiPF <sub>6</sub>	3	Acetic	2.7	0.18	0.59	0.2	3 Å	0	98	67	65 (32)
<b>68</b>	<b><math>\beta</math></b>	<b>LiI</b>	<b>1.8</b>	<b>Formic</b>	<b>1</b>	<b>0.18</b>	<b>0.3</b>	<b>0.12</b>	<b>4 Å</b>	<b>25</b>	<b>76</b>	<b>7</b>	<b>6</b> (71)
69	$\beta$	LiOTf	4.3	Oxalic	1.1	0.05	0.01	0.59	5 Å	25	12	68	8 (4)
70	$\alpha$	LiPF <sub>6</sub>	2.2	TFA	1.5	0.11	0.35	0.14	4 Å	25	74	79	58 (16)
<b>71</b>	<b><math>\alpha</math></b>	<b>LiPF<sub>6</sub></b>	<b>3.2</b>	<b>Oxalic</b>	<b>1.9</b>	<b>0.06</b>	<b>0.05</b>	<b>0.63</b>	<b>4 Å</b>	<b>25</b>	<b>96</b>	<b>82</b>	<b>78</b> (17)
72	$\beta$	LiBF <sub>4</sub>	4.6	Formic	2.7	0.28	0.33	0.11	3 Å	0	99	60	60 (17)
73	$\alpha$	LiPF <sub>6</sub>	1.3	Formic	2.4	0.23	0.36	0.63	None	25	80	65	52 (40)
74	$\alpha$	LiClO <sub>4</sub>	4.1	Oxalic	1.1	0.17	0.24	0.66	5 Å	25	29	64	18 (10)
75	$\beta$	LiPF <sub>6</sub>	5	Acetic	1.5	0.3	0.44	0.15	3 Å	0	97	63	62 (36)

<sup>a</sup>  $\alpha$ -Selectivity and yield hypervolume contribution in parenthesis.

As seen from Fig. 4B the total hypervolume for the yield and  $\alpha$ -selectivity optimization is  $\sim 89\%$  at the beginning of the optimization, that is only with the experiments from the initial yield and  $\beta$ -selectivity campaign. The dual optimization campaign is terminated once no improvement is observed for yield and  $\beta$ -selectivity nor yield and  $\alpha$ -selectivity.

In Fig. 4C are the yield of all glycosylation plotted against the  $\beta$ -selectivity (left) and the  $\alpha$ -selectivity (right), and the estimated Pareto fronts are highlighted. For the  $\beta$ -selective glycosylations, it seems that the limiting objective is the stereoselectivity, whereas for the  $\alpha$ -selective glycosylation a more classical Pareto front is observed, consisting of a set of non-dominated solutions.

The advantage of using BO instead of the OVAT approach is that it increases the chance of finding the optimal conditions significantly.<sup>56,57</sup> However, a disadvantage is that it is more difficult to infer trends from the data, as multiple reaction parameters are being varied at the time, hence making it difficult to pinpoint the effect of changing a specific parameter. Tables 1 and 2 show the experimental conditions and results from optimization campaigns 1 and 2, respectively. Despite multiple parameters being varied across the experiments, it is possible to infer some general trends. For instance, all glycosylation with LiBF<sub>4</sub> and LiNTf<sub>2</sub> are  $\beta$ -selective, and all glycosylations with LiI are  $\alpha$ -selective, whereas some of the glycosylations with LiPF<sub>6</sub> are  $\beta$ -selective (Exp. no. 1, 17, 24) and some are  $\alpha$ -selective (Exp. no. 11, 56). Interestingly, the presence of molecular sieves seems to be an important factor for the stereoselectivity in some cases. Experiments 1 and 11 are carried out under very similar conditions except for the addition of 3 Å MS to Experiment 1, but a significant difference in

selectivity is observed, 31 : 69 for Experiment 1 and 82 : 18 for Experiment 11. However, the  $\alpha$ -selectivity cannot be ascribed to the presence of LiPF<sub>6</sub> and the absence of molecular sieves alone since experiment 24 also is  $\beta$ -selective (28 : 72). Experiment 24 also does not have any additives, but the major solvent is MeCN and the acid catalyst is acetic acid, rather than Et<sub>2</sub>O and oxalic acid as for Experiments 1 and 11. This suggests that some of the variables are interdependent. For the experiments without any acid catalyst (4, 7, 20, 38, 44, 65) the yields are low to moderate, ranging from 3–59%, indicating that lithium-salts can activate the TCA-donor without any additional catalyst, albeit longer reaction times are required for full conversion. This is in accordance with previous studies.<sup>43,58</sup>

### Partial dependence plots analysis

To get a more systematic understanding of the influence of each parameter we turned to partial dependence plots, which is a way of visualizing the relationship between selected parameters and the predicted outcome, as the plots show the effect of each parameter on each objective when averaging out all other parameters.<sup>59,60</sup> The estimated effect of each parameter on yield,  $\beta$ -selectivity, and  $\alpha$ -selectivity are shown in Fig. 5. It should be noted that the partial dependence of the discrete parameters is illustrated as a continuous function, thus some parts of these graphs do not carry physical meaning. Starting from the top left, it is seen that the anomeric configuration of the glycosyl donor does not influence the yield. However, an inversely correlated effect is seen on the stereoselectivities, indicating that some of the reactions might be stereospecific. For the lithium salts, the identity of the salt influences both the yield and the selectivity.



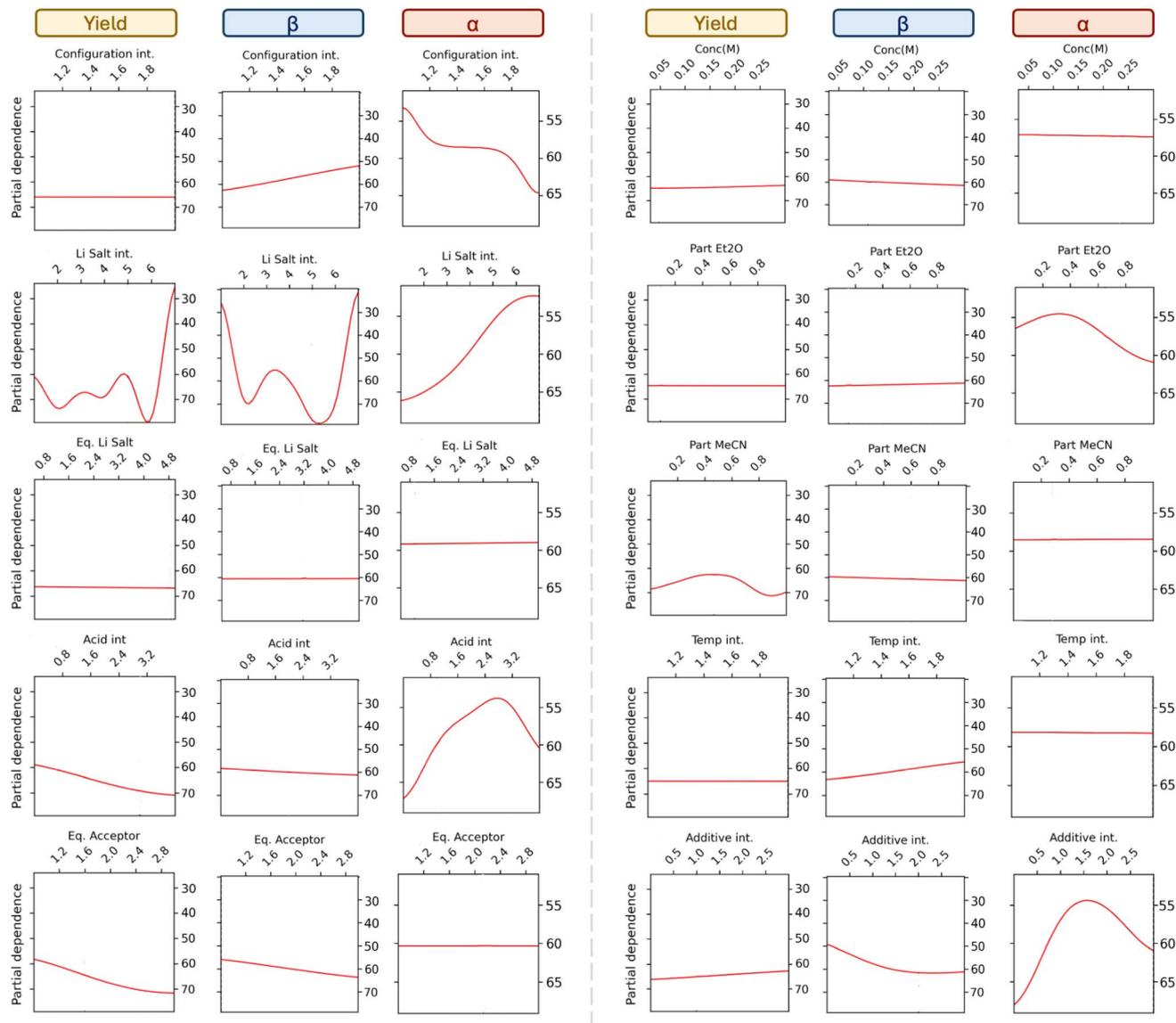


Fig. 5 Partial dependence plot for all features and all objectives. The plot shows how each feature influences the yield, percentage  $\beta$ -anomer, or percentage  $\alpha$ -anomer, while averaging out the effects of all other features. Note that the objective function is approximated as a continuous function even in the case of discrete parameters.

The  $\beta$ -selectivity plot shows the highest  $\beta$ -selectivities for salt 2, 5, and 6, which are  $\text{LiBF}_4$ ,  $\text{LiNTf}_2$ , and  $\text{LiPF}_6$ , respectively. The  $\alpha$ -selectivity plot shows a maximum at lithium salt 1 (LiI). These trends are in line with the observations from the raw data discussed earlier. Interestingly, a close to linear response between the lithium salt PCA integer assignment and the  $\alpha$ -selectivity is observed, indicating that the descriptors used for the PCA are a good measure for  $\alpha$ -selectivity.

The amount of lithium salt does not seem to have an influence on any of the objectives. The acid plots suggest that the stronger the acid, the higher the yield, which might be due to faster reaction times and the absence of rebound product between the glycosyl donor and conjugate base of the acid.<sup>64</sup> The acid also seems to have an impact on the  $\alpha$ -selectivity but with no clear trend, while the influence on the  $\beta$ -selectivity is minor. A higher amount of acceptor results in a higher yield,

which might also be related to faster reaction times. On the top right, the influence of the concentration is shown, which only seems to have a minor impact on all the objectives. The amount

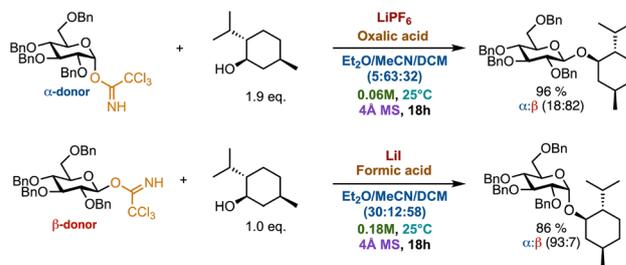


Fig. 6 Picked lead reactions for stereoselective lithium salt directed glycosylations. The top reaction depicts Experiment 71 in Table 2 and the bottom reaction depicts Experiment 68 in Table 2.



of ether solvent improves the  $\alpha$ -selectivity in agreement with known solvent effects. However, interestingly the effect of increasing the amount of acetonitrile in the solvent only shows a very minor increase in  $\beta$ -selectivity, though, this is in agreement with previous observations showing that the presence of other additives diminishes the acetonitrile effect.<sup>62</sup> The partial dependence plot indicates a slight increase in  $\beta$ -selectivity is observed at lower temperatures.

Lastly, the effect of additives is shown. Noticeably, having no additives (additives integer equal to 0) increases the  $\alpha$ -selectivity, which is also supported by the earlier comparison of Experiment 1 and 11 (Table 1). To fully understand the effect of the parameters, the reaction mechanism(s) and involvement of all reaction components would have to be elucidated.

Based on the results we propose the conditions from Experiment 71 in Table 1 as lead for new glycosylation methods for  $\beta$ -selective lithium salt directed glycosylation, and the conditions from Experiment 68 (Table 1) as a new glycosylation method for  $\alpha$ -selective lithium salt directed glycosylation. The lead reactions are depicted in Fig. 6.

From previous studies, it seems plausible that a glycosyl iodide is formed as an intermediate in the LiI-directed glycosylations, which leads to  $\alpha$ -selectivity through Curtin–Hammett kinetics.<sup>63</sup> Similar intermediates and stereoselectivity have been observed for NIS/TfOH-activated glycosylations with thioglycosides.<sup>64</sup> The high  $\beta$ -selectivity observed for 60 and 71 also aligns with the formation of either a covalent adduct or a contact ion pair between the counterion and the putative glycosyl cation. The highly electronegative counterions would favor the axial position due to the anomeric effect leading to attack by the nucleophile on the equatorial position.<sup>50</sup> However, all the counterions are highly electronegative, thus the exact role of the lithium salts, acids, and molecular sieves remains to be elucidated.

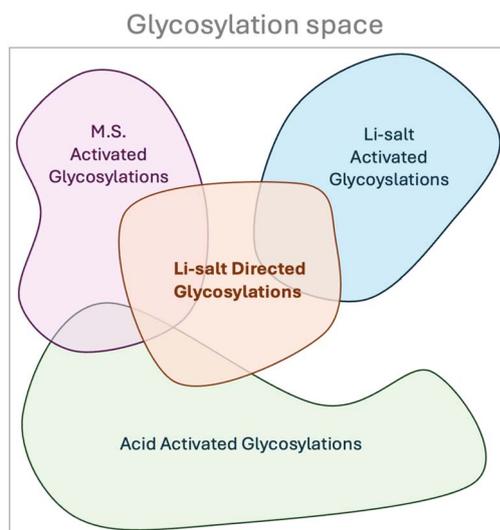


Fig. 7 Depiction of glycosylation space which is a subspace of reaction space. It is illustrated that since both lithium salt, molecular sieves, and acid are important for the outcome of lithium salt-directed glycosylation, these comprise a previously undiscovered part of the glycosylation space.

## Assessing novelty

There is no ubiquitous way of establishing the novelty of a reaction, and the terms new reaction and novel reaction are used ambiguously.<sup>5,65–68</sup> The demand for a reaction to be novel ranges from only one component being new to unprecedented reactivity.<sup>5,65–68</sup> To assess the novelty of our discovery we turned to a definition by Cronin and co-workers,<sup>65</sup> who state, that for a discovery to be novel it has to be repeatable, not observed previously, and non-predictable. We argue that the discovered lithium salt-directed glycosylations fulfill all these demands, as the changes in stereoselectivity based on lithium salts, molecular sieves, *etc.* are non-obvious and unpredictable. However, even with this definition, the term novel reaction is still not entirely unambiguous. The reactions described in this study fall under the known category glycosylation reactions, which in terms of reactivity can by itself not be described as a “novel reaction”, as glycosylation reactions are mostly  $S_N1$  or  $S_N2$ -reactions *i.e.* the reactivity is well-known. We therefore chose to evaluate if our discovery is a novel reaction based on the position in reaction space.

It is clear from the partial dependence plot that both the lithium salt, acid, and additive are important for the outcome of the reaction. Thus the methodologies cannot be classed into well-known procedures like acid-activated glycosylations,<sup>49,69</sup> acid-washed molecular sieves activated glycosylation,<sup>70</sup> or lithium salt activated.<sup>43,58,62,71</sup> Instead, the methodology of lithium salt-directed glycosylation encapsulates a previously unknown part of the glycosylation reaction space as illustrated in Fig. 7. To the best of our knowledge, this is the first example of Bayesian optimization being used for this degree of reaction discovery.

## Conclusion

We demonstrate a new workflow for identifying lead reactions in method development within a broad reaction class. This is done utilizing Bayesian optimization as a tool for discovering novel stereoselective glycosylation methodologies. Specifically, we find that a combination of lithium salt and mild acid promotes the reaction of a glycosyl TCA with *L*-menthol, resulting in high yields. The anomeric selectivity can be directed by the choice of lithium salt and the additional reaction conditions. We also show how partial dependence plots can be used to visualize the influence of each reaction parameter on the yield and stereoselectivity. From the plots, we can infer trends and gain mechanistic insights, in a similar manner to how OVAT data is analyzed.

## Data availability

Natasha Videcrantz Faurschou 2024 GitHub “GlycoTools” <https://github.com/NatashaVF/GlycoTools>.

## Author contributions

NVF designed and performed the experiments and wrote the draft manuscript. CMP secured funding, supervised and



contributed to the discussion of the results and to the revision of the manuscript.

## Conflicts of interest

The authors declare no competing interests.

## Acknowledgements

The authors thank University of Copenhagen for support. PhD Michael Martin Nielsen is acknowledged for his key role in developing the original project idea.

## References

- J. R. Cabrera-Pardo, D. I. Chai, S. Liu, M. Mrksich and S. A. Kozmin, *Nat. Chem.*, 2013, **5**(5), 423–427.
- K. Troshin and J. F. Hartwig, *Science*, 2017, **357**(6347), 175–181.
- D. W. Robbins and J. F. Hartwig, *Science*, 2011, **333**(6048), 1423–1427.
- A. F. Zahrt, Y. Mo, K. Y. Nandiwale, R. Shprints, E. Heid and K. F. Jensen, *J. Am. Chem. Soc.*, 2022, **144**(49), 22599–22610.
- W. Bort, I. I. Baskin, T. Gimadiev, A. Mukanov, R. Nugmanov, P. Sidorov, G. Marcou, D. Horvath, O. Klimchuk, T. Madzhidov, *et al.*, *Sci. Rep.*, 2021, **11**(1), 3178.
- M. N. Hopkinson, A. Gómez-Suárez, M. Teders, B. Sahoo and F. Glorius, *Angew. Chem., Int. Ed.*, 2016, **55**(13), 4361–4366.
- C. J. Taylor, A. Pomberger, K. C. Felton, R. Grainger, M. Barecka, T. W. Chamberlain, R. A. Bourne, C. N. Johnson and A. A. Lapkin, *Chem. Rev.*, 2023, **123**(6), 3089–3126.
- N. V. Faurshou, R. H. Taaning and C. M. Pedersen, *Chem. Sci.*, 2023, **14**(23), 6319–6329.
- M. Christensen, L. P. E. Yunker, F. Adedeji, F. Häse, L. M. Roch, T. Gensch, G. P. Gomes, T. Zepel, M. S. Sigman, A. Aspuru-Guzik and J. E. Hein, *Commun. Chem.*, 2021, **4**, 1–12.
- A. M. K. Nambiar, C. P. Breen, T. Hart, T. Kulesza, T. F. Jamison and K. F. Jensen, *ACS Cent. Sci.*, 2022, **8**(6), 825–836.
- A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne and A. A. Lapkin, *Chem. Eng. J.*, 2018, **352**, 277–282.
- Y. Naito, M. Kondo, Y. Nakamura, N. Shida, K. Ishikawa, T. Washio, S. Takizawa and M. Atobe, *Chem. Commun.*, 2022, **58**, 3893–3896.
- B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, 89–96.
- J. Okuda, *Trends Biochem. Sci.*, 1978, **3**(3), 161–162.
- R. D. Price, M. G. Berry and H. A. Navsaria, *J. Plast. Reconstr. Aesthetic Surg.*, 2007, **60**(10), 1110–1119.
- Y. B. Tewari and R. N. Goldberg, *J. Biol. Chem.*, 1989, **264**(7), 3966–3971.
- C.-W. Chang, M.-H. Lin, T.-Y. Chiang, C.-H. Wu, T.-C. Lin and C.-C. Wang, *Sci. Adv.*, 2023, **9**(42), eadk0531.
- D. Crich, *Acc. Chem. Res.*, 2010, **43**(8), 1144–1153.
- T. Nukada, T. Koyama, M. Sugimoto and K. Fukuyama, *J. Am. Chem. Soc.*, 1998, **120**(11), 2662–2672.
- T. Fang, Y. Gu, W. Huang and G.-J. Boons, *J. Am. Chem. Soc.*, 2016, **138**(9), 3002–3011.
- P. O. Adero, H. Amarasekara, P. Wen, L. Bohé and D. Crich, *Chem. Rev.*, 2018, **118**, 8242–8284.
- Y. Qiao, W. Ge, L. Jia, X. Hou, Y. Wang and C. M. Pedersen, *Chem. Commun.*, 2016, **52**(76), 11418–11421.
- M. M. Nielsen, B. A. Stougaard, M. Bols, E. Glibstrup and C. M. Pedersen, *Eur. J. Org. Chem.*, 2017, **2017**(9), 1281–1284.
- V. Agarkar, A. E. Hart and J. R. Ragains, *J. Carbohydr. Chem.*, 2024, 1–21.
- T. G. Frihed, M. Bols and C. M. Pedersen, *Chem. Rev.*, 2015, **115**(11), 4963–5013.
- T. Hansen, L. Lebedel, W. A. Remmerswaal, S. van Der Vorm, D. P. A. Wander, M. Somers, H. S. Overkleeft, D. V. Filippov, J. Désiré, A. Mingot, *et al.*, *ACS Cent. Sci.*, 2019, **5**(5), 781–788.
- V. P. Verma and C.-C. Wang, *Chem.–Eur. J.*, 2013, **19**(3), 846–851.
- G. Agarwal, H. A. Doan, L. A. Robertson, L. Zhang and R. S. Assary, *Chem. Mater.*, 2021, **33**(20), 8133–8144.
- J. P. Janet, S. Ramesh, C. Duan and H. J. Kulik, *ACS Cent. Sci.*, 2020, **6**(4), 513–524.
- Y. Zhang, D. W. Apley and W. Chen, *Sci. Rep.*, 2020, **10**(1), 4924.
- X. Wang, Y. Huang, X. Xie, Y. Liu, Z. Huo, M. Lin, H. Xin and R. Tong, *Nat. Commun.*, 2023, **14**(1), 3647.
- X. Li, Y. Che, L. Chen, T. Liu, K. Wang, L. Liu, H. Yang, E. O. Pyzer-Knapp and A. I. Cooper, *Nat. Chem.*, 2024, 1–9.
- Z. J. Zhang, S. W. Li, J. C. A. Oliveira, Y. Li, X. Chen, S. Q. Zhang, L. C. Xu, T. Rogge, X. Hong and L. Ackermann, *Nat. Commun.*, 2023, **14**(1), 3149.
- X. Zhou, H. Matsumoto, M. Nagao, S. Hironaka and Y. Miura, *Polym. J.*, 2024, 1–8.
- A. Obdrup, B. E. Nielsen, R. H. Taaning, S. Carlsen, and S. Bertelsen, *ProcessOptimizer v0.7.2*, 2021.
- T. Head, MechCoder, G. Louppe, I. Shcherbatyi, F. Charras, Z. Vinícius, C. M. Malone, C. Schröder, Nel215, N. Campos, T. Young, S. Cereda, T. Fan, Rene-Rex, K. Shi, J. Schwabedal, C. D. Cantos, Hvass-Labs, M. Pak, SoManyUsernamesTaken, F. Callaway, L. Estève, L. Besson, M. Cherti, K. Pfannschmidt, F. Linzberger, C. Cauet, A. Gut, A. Mueller, and A. Fabisch, *scikit-optimize/scikit-optimize: v0.5.2*, zenodo, 2018.
- S. Bertelsen, S. Carlsen, S. Furbo, M. B. Nielsen, A. Obdrup and R. Taaning, *J. Chem. Inf. Model.*, 2025, **65**, 1702–1707.
- K. Deb, A. Pratap, S. Agarwal and T. A. M. T. Meyarivan, *IEEE Trans. Evol. Comput.*, 2002, **6**(2), 182–197.
- S. Steinerberger, *Monatsh. Math.*, 2020, **191**(3), 639–655.
- R. R. Schmidt and J. Michel, *Tetrahedron Lett.*, 1984, **25**, 821–824.
- R. R. Schmidt and J. Michel, *Angew. Chem., Int. Ed. Engl.*, 1980, **19**, 731–732.
- H. H. Trinderup, S. M. Andersen, M. Heuckendorff and H. H. Jensen, *Eur. J. Org. Chem.*, 2021, **2021**(22), 3251–3259.



- 43 N. K. Korber and C. M. Pedersen, *Carbohydr. Res.*, 2022, **511**, 108497.
- 44 H. H. Trinderup, L. Juul-Madsen, L. Press, M. Madsen and H. H. Jensen, *J. Org. Chem.*, 2022, **87**(21), 13763–13789.
- 45 N. V. Faurschou and C. M. Pedersen, *Chem. Rec.*, 2021, **21**(11), 3063–3075.
- 46 S. Moon, S. Chatterjee, P. H. Seeberger and K. Gilmore, *Chem. Sci.*, 2021, **12**(8), 2931–2939.
- 47 H. Yao, M. D. Vu and X.-W. Liu, *Carbohydr. Res.*, 2019, **473**, 72–81.
- 48 R. R. Schmidt, W. Kinzy and D. Horton, *Adv. Carbohydr. Chem. Biochem.*, 1994, **50**, 21–123.
- 49 M. M. Nielsen and C. M. Pedersen, *Chem. Rev.*, 2018, **118**(17), 8285–8358.
- 50 P. R. Andreana and D. Crich, *Guidelines for o-glycoside formation from first principles*, 2021.
- 51 H. Jona, H. Mandai, W. Chavasiri, K. Takeuchi and T. Mukaiyama, *Bull. Chem. Soc. Jpn.*, 2002, **75**(2), 291–309.
- 52 L. O. Kononov, N. N. Malysheva, A. V. Orlova, A. I. Zinin, T. V. Laptinskaya, E. G. Kononova and N. G. Kolotyrkina, *Eur. J. Org. Chem.*, 2012, **2012**(10), 1926–1934.
- 53 O. T. Tuck, E. T. Sletten, J. Danglad-Flores and P. H. Seeberger, *Angew. Chem., Int. Ed.*, 2022, **61**(15), e202115433.
- 54 A. Kafle, J. Liu and L. Cui, *Can. J. Chem.*, 2016, **94**(11), 894–901.
- 55 K. Shang, H. Ishibuchi, L. He and L. M. Pang, *IEEE Trans. Evol. Comput.*, 2020, **25**(1), 1–20.
- 56 R. Liang, X. Duan, J. Zhang and Z. Yuan, *React. Chem. Eng.*, 2022, **7**(3), 590–598.
- 57 W. Xu, Z. Liu, R. T. Piper and J. W. P. Hsu, *Sol. Energy Mater. Sol. Cells*, 2023, **249**, 112055.
- 58 A. Lubineau and B. Drouillat, *J. Carbohydr. Chem.*, 1997, **16**(7), 1179–1186.
- 59 J. H. Friedman, *Ann. Stat.*, 2001, 1189–1232.
- 60 T. Hastie, R. Tibshirani, and J. H. Friedman, Boosting and additive trees, in *The Elements of Statistical Learning*, Springer, New York, NY, USA, 2nd edn, 2009, pp. 337–384.
- 61 N. D. Gould, C. L. Allen, B. C. Nam, A. Schepartz and S. J. Miller, *Carbohydr. Res.*, 2013, **382**, 36–42.
- 62 H. Waldmann, G. Böhm, U. Schmid and H. Rötttele, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**(19), 1944–1946.
- 63 P. J. Meloncelli, A. D. Martin and T. L. Lowary, *Carbohydr. Res.*, 2009, **344**(9), 1110–1122.
- 64 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, *et al.*, *Angew. Chem.*, 2019, **131**(47), 16931–16935.
- 65 P. S. Gromski, A. B. Henson, J. M. Granda and L. Cronin, *Nat. Rev. Chem.*, 2019, **3**(2), 119–128.
- 66 K. D. Collins, T. Gensch and F. Glorius, *Nat. Chem.*, 2014, **6**(10), 859–871.
- 67 M. H. S. Segler and M. P. Waller, *Chem.–Eur. J.*, 2017, **23**(25), 6118–6128.
- 68 X. Wang, C. Yao, Y. Zhang, J. Yu, H. Qiao, C. Zhang, Y. Wu, R. Bai and H. Duan, *J. Cheminf.*, 2022, **14**(1), 60.
- 69 S. C. Ranade and A. V. Demchenko, *J. Carbohydr. Chem.*, 2013, **32**(1), 1–43.
- 70 M. Adinolfi, G. Barone, A. Iadonisi and M. Schiattarella, *Org. Lett.*, 2003, **5**(7), 987–989.
- 71 G. Böhm and H. Waldmann, *Liebigs Ann.*, 1996, **1996**(4), 613–619.

