nature communications

Article

https://doi.org/10.1038/s41467-024-47984-0

In-silico-assisted derivatization of triarylboranes for the catalytic reductive functionalization of aniline-derived amino acids and peptides with H_2

Received: 8 November 2023

Accepted: 16 April 2024

Published online: 07 May 2024

Check for updates

Yusei Hisata[1], Takashi Washio[2], Shinobu Takizawa[3], Sensuke Ogoshi[1] & Yoichi Hoshimoto[1,4]

Cheminformatics-based machine learning (ML) has been employed to determine optimal reaction conditions, including catalyst structures, in the field of synthetic chemistry. However, such ML-focused strategies have remained largely unexplored in the context of catalytic molecular transformations using Lewis-acidic main-group elements, probably due to the absence of a candidate library and effective guidelines (parameters) for the prediction of the activity of main-group elements. Here, the construction of a triarylborane library and its application to an ML-assisted approach for the catalytic reductive alkylation of aniline-derived amino acids and C-terminal-protected peptides with aldehydes and H_2 is reported. A combined theoretical and experimental approach identified the optimal borane, i.e., B(2,3,5,6-Cl_4-C_6H)(2,6-F_2-3,5-$(CF_3)_2$-$C_6H)_2$, which exhibits remarkable functional-group compatibility toward aniline derivatives in the presence of 4-methyltetrahydropyran. The present catalytic system generates H_2O as the sole byproduct.

Catalysis is a fact of our daily lives. A wide variety of important commercial chemical substances are currently produced on both the fine and bulk scales in the presence of molecular catalysts that have been optimized based on specific factors such as efficiency, toxicity, cost, or a combination thereof. Recent advancements in cheminformatics-based machine learning (ML) offer chemists a way to bypass traditional Edisonian empiricism and develop more efficient approaches to optimizing catalysts[1–4]. Several groups have reported successful demonstrations of ML-driven optimizations of homogeneous catalysts such as phosphoric acids and Lewis-basic ligands for metal-based catalysts involving phosphines, *N*-heterocyclic carbenes, and nitrogen-based ligands[5–18].

Recent progress in frustrated Lewis pairs (FLPs)[19,20] has expanded the practical and sustainable application of main-group catalysis, e.g., enabling the hydrogenation of unsaturated molecules without toxic/precious metals[21–26]. In this context, the main-group-catalyzed reductive alkylation of amines with carbonyl compounds and H_2 via the generation of FLP species has been widely accepted as a waste-minimizing process that generates valuable *N*-alkylated amines, whereby H_2O is the only by-product[23,27–31]. Our group[30] and that of Soós[28] have independently shown that triarylboranes effectively catalyze the reductive alkylation of a variety of amines with aldehydes in the presence of H_2. Moreover, we have demonstrated that an FLP that consists of Soós' borane, i.e., B(2,6-Cl_2-C_6H_3)(2,3,5,6-F_4-$C_6H)_2$ ($\mathbf{B^{1a}}$)[27], and tetrahydrofuran (THF) exhibits good functional-group tolerance for aniline derivatives including halogens, hydroxyl, and amide groups (Fig. 1A)[30]. Based on mechanistic studies, we proposed dual catalysis of $\mathbf{B^{1a}}$ in the formation of imine intermediates and the hydrogenation of

[1]Department of Applied Chemistry, Faculty of Engineering, Osaka University, Suita, Osaka 565-0871, Japan. [2]Department of Reasoning for Intelligence and Artificial Intelligence Research Center, SANKEN, Osaka University, Ibaraki, Osaka 567-0047, Japan. [3]Department of Synthetic Organic Chemistry and Artificial Intelligence Research Center, SANKEN, Osaka University, Ibaraki, Osaka 567-0047, Japan. [4]Division of Applied Chemistry, Center for Future Innovation (CFi), Faculty of Engineering, Osaka University, Suita, Osaka 565-0871, Japan. e-mail: hoshimoto@chem.eng.osaka-u.ac.jp

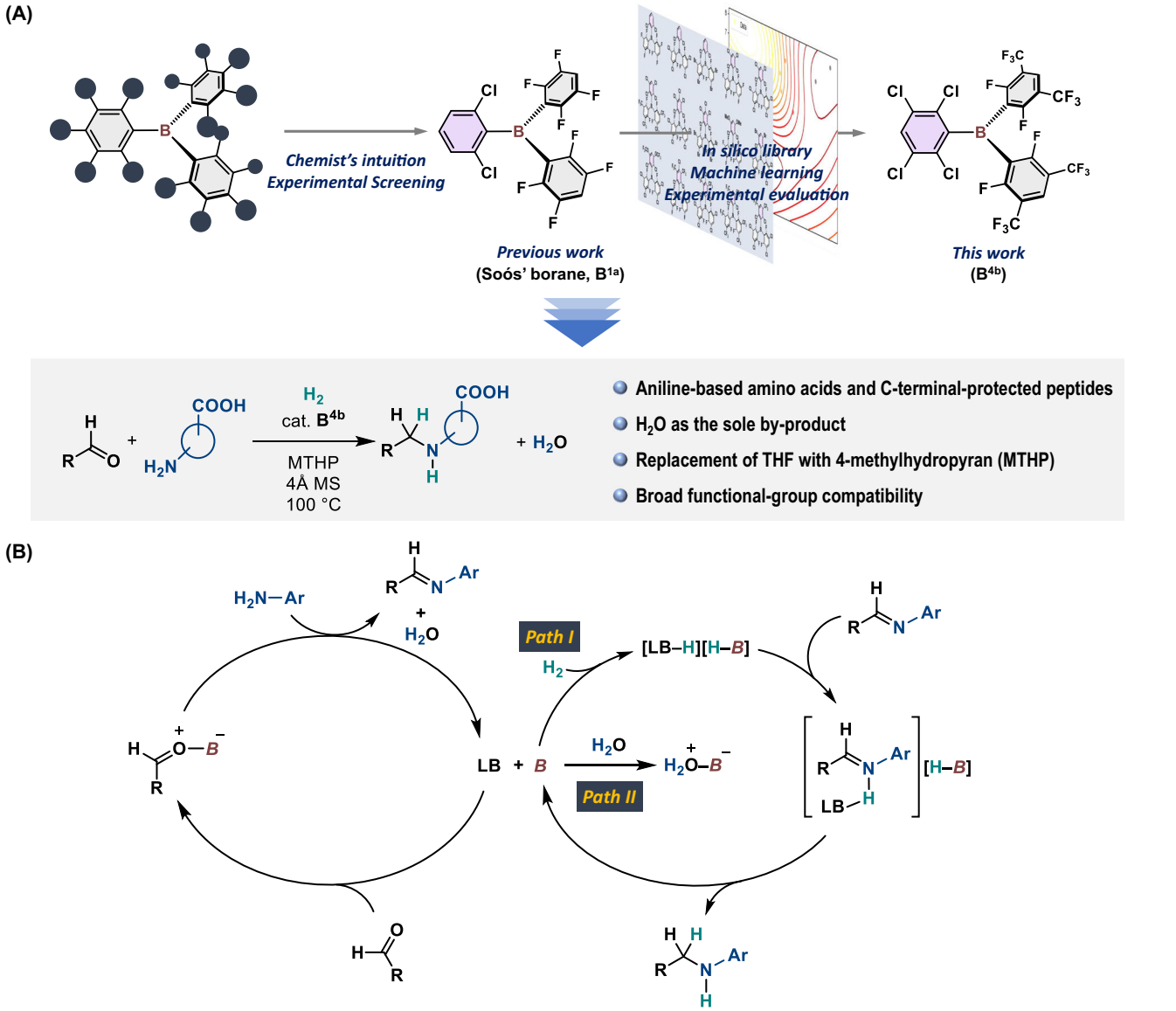


Fig. 1 | Context of this work. A Schematic representation of the concept of this study. **B** Proposed dual catalysis of triarylborane in the catalytic reductive alkylation of amines with aldehydes; ***B*** = triarylborane; **LB** = THF or MTHP.

the imines via the generation of the FLP species (Fig. 1B). However, the direct reductive alkylation of amino acids with H_2 has remained challenging, and such reactions have proceeded in only low-to-moderate yields even under forcing conditions. In terms of toxicity, the solvent THF, which also acts here as a Lewis base to generate FLPs with boranes (Fig. 1B), should be replaced with a less hazardous chemical[32,33]. Given the central role of the reductive alkylation of amines using carbonyl compounds in the synthesis of, e.g., pharmaceuticals, bio-active molecules, and agrochemicals[34–36], the development of a straightforward and greener protocol for derivatizing amino acids and peptides would be worthwhile.

To this end, we envisioned an ML-assisted approach to identify a suitable triarylborane that is able to efficiently catalyze the reductive alkylation of amino acids and peptides with H_2 through the construction of an in-silico library that includes a variety of triarylboranes. The synthesis of triarylboranes with unknown substitution patterns is typically a laborious and time-consuming process that often requires several weeks or even months for optimization. However, once the optimal procedures have been obtained, the optimized conditions can often be extrapolated, which is much faster. Therefore, using an ML-assisted approach to streamlining the selection of triarylborane candidates has the potential to significantly accelerate the optimization process and therefore the entire research process. Moreover, through the construction of this in-silico library, we aimed to contribute to the structural diversification of triarylboranes beyond the archetypical $B(C_6F_5)_3$[37–40], which should expand the utility of this compound class in catalysis, materials science, and other areas. It should be noted that Dyson and Corminboeuf et al. have recently demonstrated the hydrogenation of CO_2 to yield a formate salt [DBU–H][H–COO], which was catalyzed by an FLP consisting of tris(*p*-bromo)tridurylborane and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU)[41]. They identified this combination of borane and DBU using a cheminformatics-assisted approach that profiled the theoretically predicted catalyst activity based on the intrinsic acidity and basicity of the Lewis components.

Herein, we report the construction of an in-silico library with 54 triarylboranes, which was used for the ML-assisted identification of the optimal borane for the catalytic reductive alkylation of aniline-derived amino acids and peptides with aldehydes and H_2 (Fig. 1A). We also explored the functional-group compatibility of the present system using the functional group evaluation (FGE) kit recently proposed by

Morimoto and Oshima et al.[42], which is based on the concept of robustness screening that has been proposed by Glorius et al.[43,44].

Results and discussion

Theoretical and experimental variables collection

We started our investigation with the construction of an in-silico library of triarylboranes using the strategy described below. Generally, we explored triarylboranes that seemed to be synthetically accessible using common procedures[37,40]. Optimization of the gas-phase structures of the 54 boranes shown in Fig. 2A was accomplished using DFT calculations at the ωB97X-D/6-311+G(d,p)//ωB97X-D/6-31G(d,p) level. The 53 explored heteroleptic boranes $\mathbf{B}^{xy}$ (x = **1**–**6**, y = **a**–**w**) include two 2,6-F_2-3,5-R_2-C_6H groups, and their core structures are classified as $\mathbf{B}^{1}$–$\mathbf{B}^{6}$ depending on the R groups, i.e., $\mathbf{B}^{x}$(R) = $\mathbf{B}^{1}$(F), $\mathbf{B}^{2}$(Cl), $\mathbf{B}^{3}$(Br), $\mathbf{B}^{4}$(CF_3), and $\mathbf{B}^{5}$(H), whereas $\mathbf{B}^{6}$ includes 2,6-F_2-3-Cl-C_6H_2 groups. With these core structures, we combined 23 aryl groups (**a**–**w**), and completed the construction of the borane library with the addition of $B(C_6F_5)_3$. It should be noted that boranes $\mathbf{B}^{xa}$ (x = **1**–**3**, **5**, **6**), $\mathbf{B}^{1v}$, and $\mathbf{B}^{1w}$ were known before the construction of this library[27,30,45–47], and their reactivity in the hydrogenation of unsaturated molecules inspired us when designing the library molecules. In fact, $\mathbf{B}^{1a}$ has previously been employed for the catalytic reductive alkylation of aniline derivatives to generate active FLP species with THF[30], and thus, in-silico derivatization of $\mathbf{B}^{1a}$ was carried out via substitution of the *meta* and/or *para* H atoms in the 2,6-Cl_2-C_6H_3 group with Cl, Br, CF_3, OMe, OCF_3 or C_6F_5 groups. We further conducted extensive in-silico derivatization of $\mathbf{B}^{2a}$ and $\mathbf{B}^{3a}$, given that these boranes exhibit far superior catalytic activity than $\mathbf{B}^{1a}$, $\mathbf{B}^{6a}$, and $B(C_6F_5)_3$ in the hydrogenation of *N*-heteroaromatics using a gaseous mixture of H_2/CO/CO_2/CH_4[45]. In these cases, we envisioned that the modulation of the intrinsic Lewis acidity of the triarylboranes, i.e., the energy levels of the LUMO, which includes the p orbital on the boron center[48], as well as the remote back-strain[49] that would influence the stability of the four-coordinated tetrahedral Lewis base–borane adducts. The introduction of 2,6-Br_2-C_6H_3 (**j**) and its derivatives (**k**–**r**) into the $\mathbf{B}^{2}$ core was also explored, as we expected an increase in the front strain that influences the accessibility of the Lewis bases to the boron centers[37].

We subsequently obtained the theoretical parameters. We thought that the use of structural parameters obtained from the gas-phase optimization of $\mathbf{B}^{xy}$ would not play a critical role in predicting the reactivity of the triarylboranes under the chosen conditions, given that no substantial differences were observed among them (Supplementary Table 4). Thus, we obtained the following energetic parameters, calculated at the ωB97X-D/6-311+G(d,p)//ωB97X-D/6-31G(d,p) level: (i) the energy levels of the LUMOs [eV], which include the p orbitals on the boron atoms, (ii) the energy barriers ($\Delta G_H^{\ddagger}$) [kcal mol^{-1}] for the heterolytic cleavage of H_2 with the combination of $\mathbf{B}^{xy}$ and THF, and (iii) the relative Gibbs energy values (ΔG_w°) [kcal mol^{-1}] for the formation of the H_2O–$\mathbf{B}^{xy}$ adducts with respect to [H_2O + $\mathbf{B}^{xy}$]. These theoretical values are shown in Fig. 3A for selected boranes. For parameter (ii), we have previously proposed that the heterolytic cleavage of H_2 by FLPs should be involved in the rate-determining event of the $\mathbf{B}^{1a}$-catalyzed reductive alkylation of amines (Path I in Fig. 1B)[30]. For parameter (iii), H_2O could be a potential quencher of the triarylborane catalysts via the formation of adducts (Path II in Fig. 1B) followed by proto-deboronation[50]. In this context, we envisioned that boranes $\mathbf{B}^{xy}$ that exhibit larger ΔG_w° and smaller $\Delta G_H^{\ddagger}$ values should show superior performance as Lewis acids for the generation of more active FLPs with ethereal components. It should also be mentioned here that we theoretically optimized a structure that included a sole imaginary

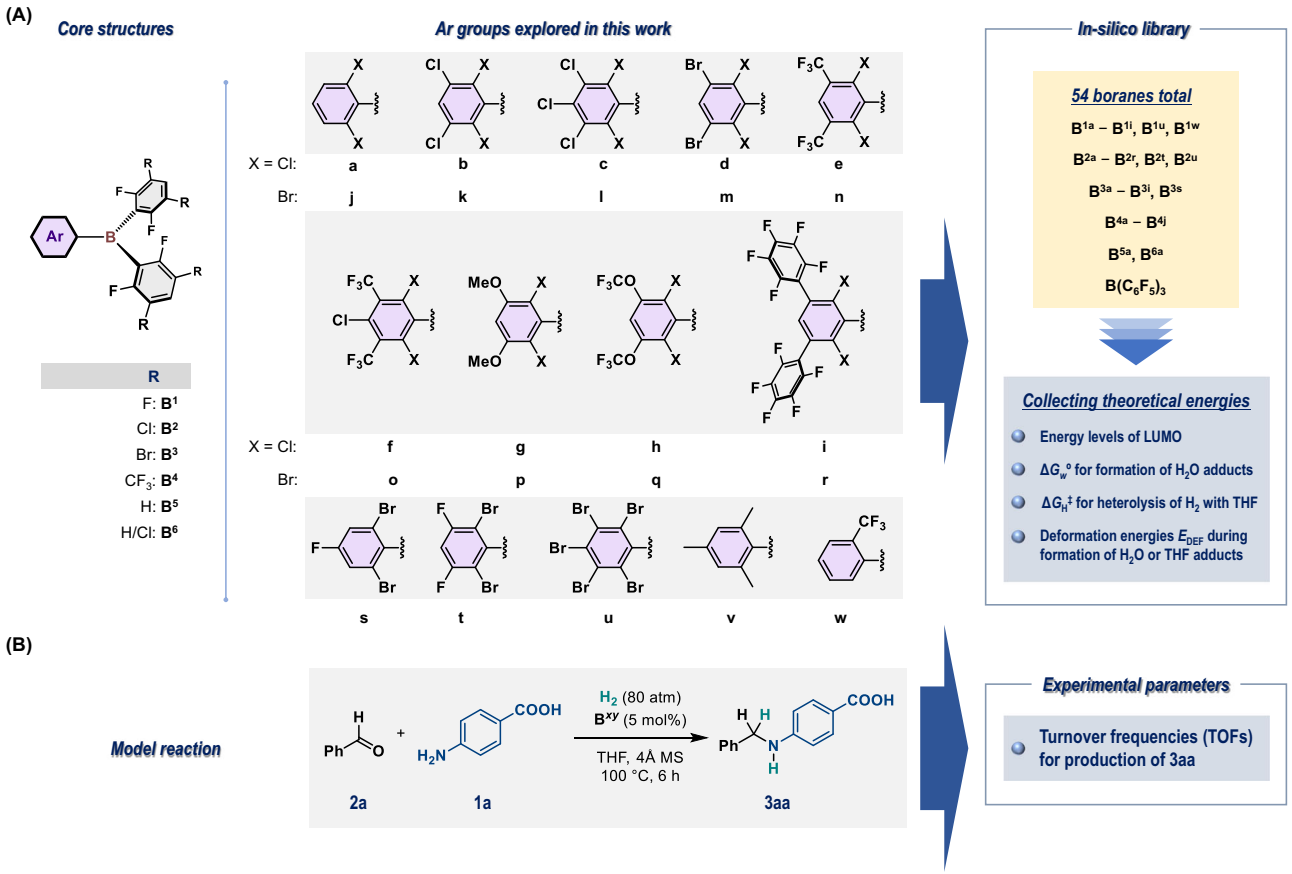


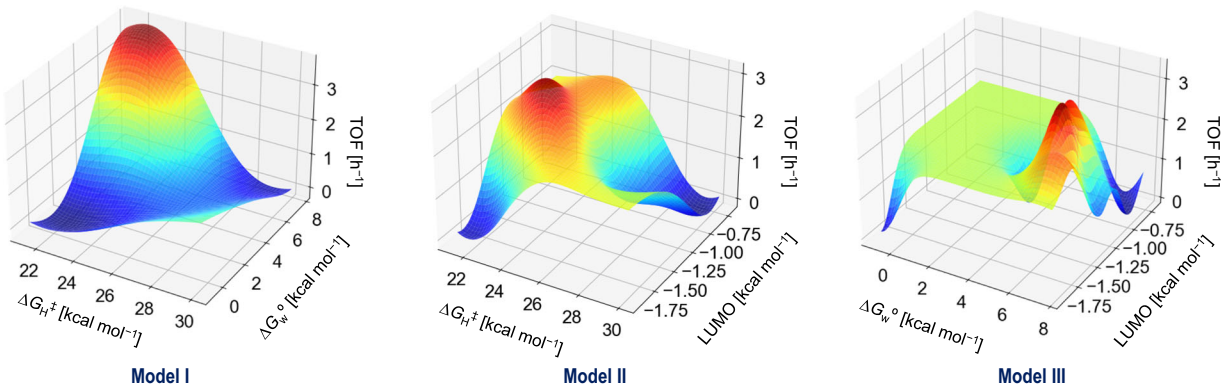
Fig. 2 | Experiment designs. A The in-silico library of triarylboranes explored in this work. The structure of $\mathbf{B}^{6}$ includes 2,6-F_2-3-Cl-C_6H_2 groups. **B** The model reaction for obtaining the experimental parameters (turnover frequencies per hour) used in this work.

https://doi.org/10.1038/s41467-024-47984-0

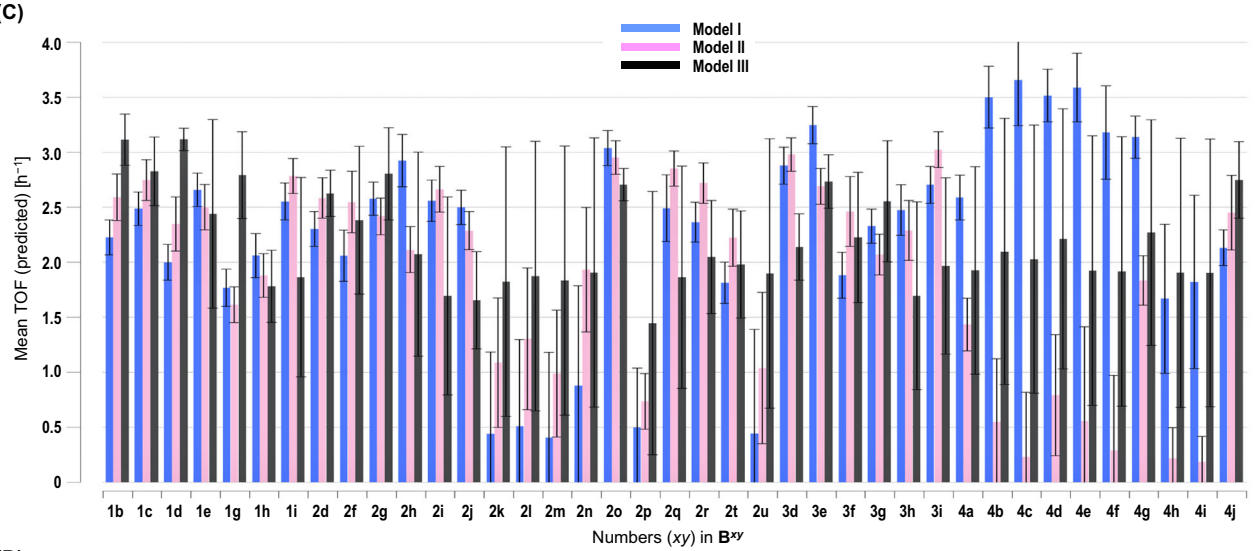
(A)

	B^{1a}	B^{1f}	B^{1v}	B^{1w}	B^{2a}	B^{2b}	B^{2c}	B^{2e}	B^{3a}	B^{3b}	B^{3c}	B^{3s}	B^{5a}	B^{6a}	$B(C_6F_5)_3$
LUMO	−1.33	−1.64	−0.92	−1.25	−1.31	−1.48	−1.53	−1.62	−1.28	−1.45	−1.50	−1.35	−1.03	−0.71	−1.93
$\Delta G_w°$	+5.2	+4.1	+7.7	+4.4	+5.6	+6.3	+6.5	+7.0	+5.5	+6.5	+6.6	+6.9	+6.1	+6.6	−0.9
$\Delta G_H^‡$	+25.5	+23.3	+30.1	+28.7	+24.4	+24.1	+24.5	+24.7	+24.5	+24.5	+24.5	+25.9	+25.4	+28.7	+21.3
TOF	1.87	1.80	0	0.90	2.70	3.20	2.87	2.67	2.70	2.93	2.87	1.63	1.97	0.10	1.33

(B)



(C)



(D)

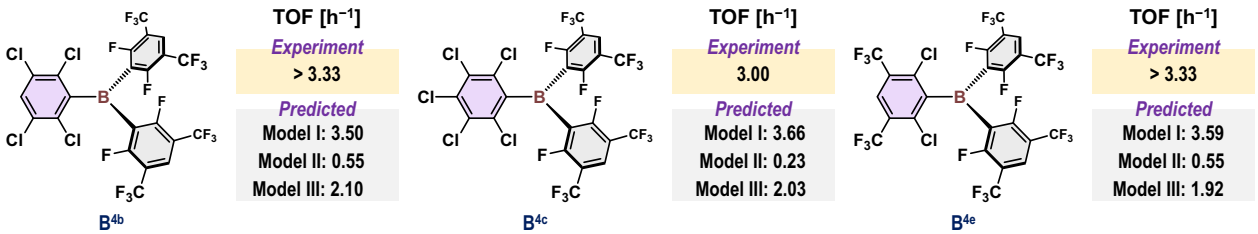


Fig. 3 | Optimization of B^{xy}. A LUMO energy level [eV], $\Delta G_w°$ [kcal mol^{-1}], and $\Delta G_H^‡$ [kcal mol^{-1}] for selected boranes B^{xy}. TOFs [h^{-1}] calculated based on the yield of **3aa** under the model conditions are also shown. **B** Gaussian process regression using the programming library GPy for the prediction of the TOF values. The theoretical values of the parameters $\Delta G_w°$ and $\Delta G_H^‡$ (**Model I**), the LUMO level and $\Delta G_H^‡$ (**Model II**), or the LUMO level and $\Delta G_w°$ (**Model III**) were used. **C** Comparison of the TOF values predicted using **Models I, II**, or **III**. Error bars represent 1σ standard deviation. **D** Comparison of the experimental and predicted TOF values for B^{4b}, B^{4c}, and B^{4e}.

frequency related to the H–H bond cleavage; however, to reduce the calculation costs, we performed an IRC calculation only for selected cases and confirmed their validity as a possible transition state structure.

Next, we turned our attention to collecting experimental data for the reported triarylboranes (B^{xa}, x = **1**–**3**, **5**, **6**; B^{1v}; B^{1w}; $B(C_6F_5)_3$) and newly synthesized boranes (B^{1f}; B^{2y}, y = **b**, **c**, **e**; B^{3y}, y = **b**, **c**, **s**); the latter compounds were used to analyze the influence of the derivatization of

the 2,6-Cl_2-C_6H_3 structure. We obtained the turnover frequency (TOF in h^{-1}; Fig. 3A) as an experimental parameter calculated based on the yield of **3aa** from the reductive alkylation of amino acid **1a** with benzaldehyde (**2a**) and H_2 in the presence of 5 mol% $\mathbf{B}^{xy}$ under the shown conditions as a model reaction (Fig. 2B). The design of this model reaction is based on the potential of $\mathbf{B}^{1a}$ to produce **3aa** in approximately 50% yield after 6 h (i.e., TOF = 1.67 h^{-1}), as this consideration allows for a clearer evaluation of the positive and negative effects of structural derivatization during the optimization of $\mathbf{B}^{xy}$.

Reaction conditions optimization

With the theoretical and experimental parameters in hand, Gaussian process regression (GPR), which is a radial basis function kernel-based statistical-learning algorithm, using GPy (a programming library for GPR)[51], was applied to construct a model to predict the TOF values for the production of **3aa** (it should be noted here that unless stated otherwise, mean values are presented for the theoretically predicted TOF values). GPR using GPy constructs a regression model using a limited number of observed data through ML and searches for a subsequent adequate parameter value of $\mathbf{B}^{xy}$ using the surrogate model. Note that GPR analysis has been widely explored in depth in information science and is characterized by the availability of a wide range of acquisition functions. Hence, its performance and flexibility to address various types of problems can be expected to efficiently promote the exploration of wide-ranging experimental conditions[52]. We evaluated the accuracy of each GPR model based on the coefficient of determination (Q^2) of the leave-one-out (LOO) cross-validation using either training or validation data. It is noteworthy that higher Q^2 values approaching 1 indicate a more robust explanation for the data. We carried out the initial GPR analysis using the experimental TOF values and pairs of two of the three theoretical parameters (LUMO energy level, $\Delta G_w°$, and $\Delta G_H^‡$) obtained for the 15 boranes shown in Fig. 3A as training data. This GPR analysis resulted in three distinct models, labelled **Model I** ($\Delta G_w°$ vs $\Delta G_H^‡$; $Q^2 = 0.75$), **II** (LUMO level vs $\Delta G_H^‡$; $Q^2 = 0.76$), and **III** (LUMO level vs $\Delta G_w°$; $Q^2 = 0.21$) (Fig. 3B). A significantly lower Q^2 of **Model III** would indicate its insufficient reliability. Subsequently, we used the theoretical parameters of the other 39 triarylboranes to predict their TOF values using these three models. We found intriguing inconsistencies among the TOF values for the $\mathbf{B}^4$ derivatives that contain *meta*-CF_3 groups predicted using **Model I** and those predicted using **Model II** (Fig. 3C), albeit that the Q^2 values of **Model I** and **II** were identical at this stage. For example, when **Model I** was applied, the TOF values of $\mathbf{B}^{4b}$, $\mathbf{B}^{4c}$, and $\mathbf{B}^{4e}$ were predicted to be 3.50, 3.66, and 3.59, respectively; however, using **Model II**, the corresponding TOF values were predicted to be 0.55, 0.23, and 0.55 respectively (Fig. 3C, D). A critical difference between **Model I** and **Model II** is the use of the LUMO energy levels in the GPR analysis. Thus, to evaluate whether the LUMO levels can serve here as a critical parameter for the prediction of the TOF for the production of **3aa**, we additionally synthesized $\mathbf{B}^{4b}$, $\mathbf{B}^{4c}$, and $\mathbf{B}^{4e}$, and confirmed that these boranes demonstrate excellent activity for the reductive alkylation of **1a** with **2a** under the model reaction conditions as predicted by **Model I** (Fig. 3D). We updated each model additionally using the experimental TOF values of $\mathbf{B}^{4b}$, $\mathbf{B}^{4c}$, and $\mathbf{B}^{4e}$ to construct **Model I′** ($Q^2 = 0.73$), **II′** ($Q^2 = 0.28$), and **III′** ($Q^2 = 0.28$) (Supplementary Table 3). These results convinced us of the superiority of **Model I′** for the prediction of the catalytic activity of $\mathbf{B}^{xy}$ under the applied conditions. Based on synthetic accessibility considerations, we decided to employ $\mathbf{B}^{4b}$ as the optimal catalyst in the following experiments. These results also suggest that the employment of a combination of theoretical parameters related to the rate-determining step (e.g., $\Delta G_H^‡$ in this work) and the (potential) catalyst deactivation step (e.g., ΔG_w in this work) is decisive, whereas a parameter related to intrinsic Lewis acidity of triarylboranes, such as the LUMO energy level, is inappropriate for the prediction of the catalytic activity of triarylboranes under the applied conditions.

We also aimed to investigate relatively unexplored theoretical parameters for the construction of a regression-based model for the prediction of the catalytic activity of triarylboranes. In this context, we explored the deformation energy (E_{DEF}) [kcal mol^{-1}] that can be used to evaluate the degree of remote back-strain[49], where E_{DEF} represents the energetic penalty associated with the change in the conformation at the boron center from trigonal planar to tetrahedral upon the

(A)

	B^{1a}	B^{1f}	B^{1v}	B^{1w}	B^{2a}	B^{2b}	B^{2c}	B^{2e}	B^{3a}	B^{3b}	B^{3c}	B^{3s}	B^{4b}	B^{4c}	B^{4e}	B^{5a}	B^{6a}	$B(C_6F_5)_3$
$E_{DEF}(H_2O)$	+21.6	+25.4	+17.6	+16.2	+20.6	+22.1	+22.0	+23.6	+20.0	+21.7	+21.7	+21.4	+21.9	+23.3	+22.4	+21.0	+20.3	+18.4
E_{DEF}(THF)	+27.7	+31.4	+27.8	+23.0	+28.6	+29.9	+30.1	+29.8	+28.5	+29.9	+30.1	+31.2	+29.7	+29.9	+29.9	+30.9	+31.2	+25.8
q(B)	+0.31	+0.31	+0.15	+0.25	+0.27	+0.30	+0.33	+0.27	+0.32	+0.28	+0.28	+0.26	+0.27	+0.26	+0.37	+0.24	+0.26	+0.37
q(C)	−0.17	−0.17	−0.14	−0.16	−0.14	−0.19	−0.20	−0.17	−0.16	−0.17	−0.16	−0.16	−0.16	−0.15	−0.17	−0.13	−0.15	−0.21

(B)

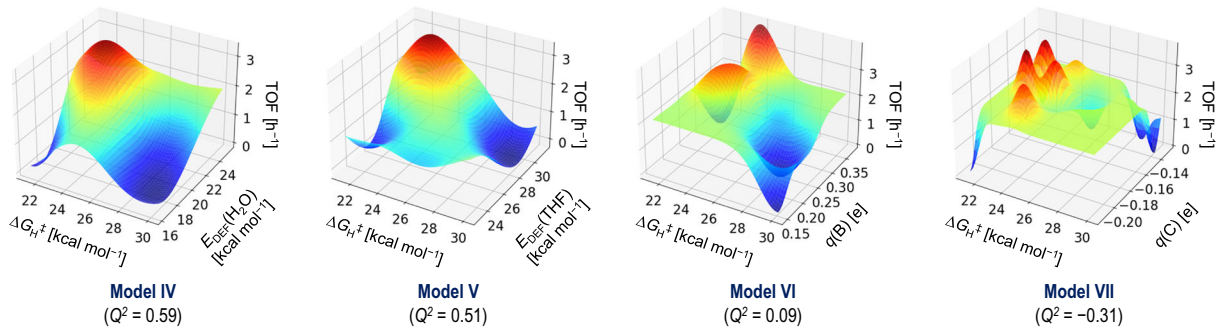


Fig. 4 | Exploring theoretical parameters. A Theoretical parameters calculated for the selected $\mathbf{B}^{xy}$. Deformation energies [kcal mol^{-1}], including $E_{DEF}(H_2O)$ and E_{DEF}(THF), were calculated at the RI-DSD-PBEP86-D3BJ/ma-Def2-QZVPP//PBEh-3c/Def2-SVP level. Mulliken charges [e], q(B) (values on the boron atoms) and q(C) (averaged values of three *ipso*-carbons), were calculated at the PBEh-3c/Def2-SVP level. **B** Gaussian process regression with GPy for the prediction of TOF values [h^{-1}], using $E_{DEF}(H_2O)$ and $\Delta G_H^‡$ (**Model IV**), E_{DEF}(THF) and $\Delta G_H^‡$ (**Model V**), q(B) and $\Delta G_H^‡$ (**Model VI**), and q(C) and $\Delta G_H^‡$ (**Model VII**).

formation of adducts between Lewis bases (LBs) and triarylboranes[48,53,54]. We calculated the E_{DEF}(LB) values for 18 boranes shown in Fig. 4A via the gas-phase optimization of their H_2O or THF adducts (i.e., $E_{DEF}(H_2O)$ and E_{DEF}(THF)) followed by energy-decomposition analysis at the RI-DSD-PBEP86-D3BJ/ma-Def2-QZVPP//PBEh-3c/Def2-SVP level. The GPR analysis with E_{DEF}(LB) and $\Delta G_H^{\ddagger}$ using the experimental TOF values resulted in the construction of **Model IV** (LB = H_2O; $Q^2 = 0.59$) and **V** (LB = THF; $Q^2 = 0.51$) with an acceptable reliability (Fig. 4B). Interestingly, these Q^2 values were found to be insensitive to the differences in H_2O and THF. For further comparison, we also prepared **Model VI** (q(B) vs $\Delta G_H^{\ddagger}$; $Q^2 = 0.09$) and **VII** (q(C) vs $\Delta G_H^{\ddagger}$; $Q^2 = -0.31$), wherein Mulliken charges [e] on the boron atoms and their average values on three *ipso*-carbon atoms in $\mathbf{B^{xy}}$ are given as q(B) and q(C), respectively; however, the reliability of these models was found to be insufficient. These results thus suggest that E_{DEF} should be a valuable parameter to explore during the ML-based optimization of triarylboranes, as the estimation of E_{DEF} through the optimization of structurally simple molecules (i.e., boranes, Lewis bases such as H_2O, and their adducts) is technically easier than the calculation of the activation energies, which is only feasible after the optimization of transition states.

With the optimal triarylborane $\mathbf{B^{4b}}$ in hand, we modified the reaction conditions to reduce its environmental impact and thus establish a greener and more sustainable system. In this context, the use of an alternative reaction solvent that could also act as a Lewis base to generate an FLP with $\mathbf{B^{4b}}$ was initially explored, given the recent demand for the replacement of hazardous THF with alternative ethereal compounds that exhibit lower toxicity combined with high chemical and thermal stability[32]. In the presence of 40 atm H_2, the reductive alkylation of **1a** with **2a** was carried out using THF, 2-methyltetrahydrofuran (2-MeTHF), cyclopentyl methyl ether (CPME), 4-methyltetrahydropyran (MTHP), or 2,2,4-trimethyl-1,3-dioxolane (TMD) (Fig. 5A). While THF provided a superior result (50%) compared to 2-MeTHF (38%) and CPME (18%), **3aa** was generated in 59% yield when MTHP was used. Prolongation of the reaction time to 24 h resulted in the formation of **3aa** in 72% yield; however, the removal of the 4 Å MS caused a decrease in the yield of **3aa** to 42%. Finally, increasing the H_2 pressure to 60 atm resulted in the formation of **3aa** in 95% yield after the period of 24 h. The use of TMD did not furnish any **3aa**. It should be noted that MTHP can be easily separated from water (its solubility in H_2O is ~1.5 wt%) and removed under reduced pressure due to its strong hydrophobicity and low heat of vaporization, although its employment as a greener solvent has been limited in organic synthesis compared with the use of 2-MeTHF and CPME[32,33]. Moreover, to clarify the benefit of using MTHP over THF, we compared the activation energies for the heterolytic cleavage of H_2 by the combination of $\mathbf{B^{4b}}$ and THF or MTHP at the ωB97X-D/6-311+G(d,p)//ωB97X-D/6-31G(d,p) level (Fig. 5B). A possible transition state was found in both cases, and that in the case of MTHP was found to be more stabilized (TS_{MTHP} = +21.8 kcal mol^{-1}) than that in the case involving THF (TS_{THF} = +23.2 kcal mol^{-1}). We attribute this stabilization to the increased structural flexibility of the tetrahydropyrane motif relative to THF, which allows the formation of efficient non-covalent interactions (NCIs) between the F/Cl atoms in $\mathbf{B^{4b}}$ and the H atoms in MTHP. The participation of such NCIs was confirmed using the quantum theory of atoms in molecules (AIM) method (for details, see Supplementary Fig. 13)[55,56].

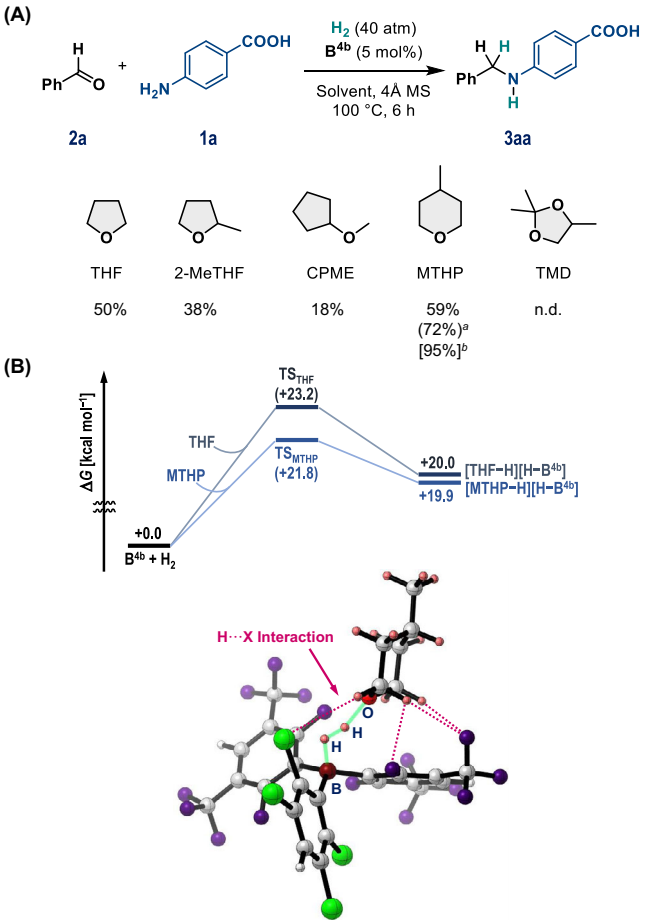
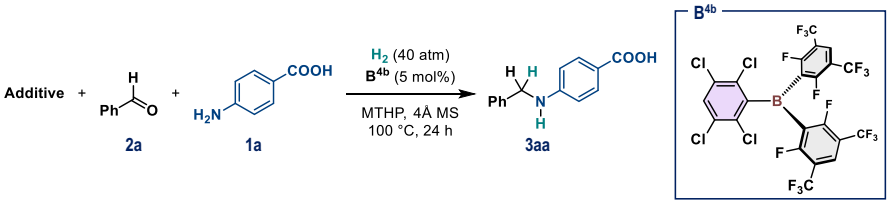


Fig. 5 | Optimization of reaction conditions. A Exploration of greener Lewis-basic solvents. Reaction conditions: **1a** (0.4 mmol, 0.05 M), **2a** (1.0 equiv.), $\mathbf{B^{4b}}$ (5 mol%), and 4 Å MS (100 mg) were mixed in the solvent, followed by pressurization with H_2 (40 atm). The yield of **3aa** was determined via ^{1}H NMR analysis. [a]24 h. [b]60 atm H_2 for 24 h. **B** Relative Gibbs free energies [kcal mol^{-1}] with respect to [$\mathbf{B^{4b}}$ + H_2 + LB], where LB is either MTHP or THF, calculated at the ωB97X-D/6-311+G(d,p)//ωB97X-D/6-31G(d,p) level. The structure of TS_{MTHP} is also shown. Pairs of atoms involved in H···X (F/Cl) interactions that were found using AIM analysis are indicated by dashed lines (H: pink; B: brown; C: gray; O: red; F: purple; Cl: light green).

Functional-group compatibility

The $\mathbf{B^{4b}}$-catalyzed reductive alkylation of **1a** with **2a** in MTHP using H_2 (40 atm) demonstrated remarkable compatibility toward a variety of additives (**A0**–**A21**) (Fig. 6). All these experiments were carried out twice, and mean values [%] are given for the yield of **3aa**, the recovered additive, and the imine intermediate 4-(benzylideneamino)benzoic acid formed in situ through the $\mathbf{B^{4b}}$-catalyzed condensation of **1a** and **2a** (left cycle in Fig. 1B). Initially, we carried out a control experiment using **A0**, which confirmed that the yields of **3aa** and the remaining imine were consistent (72% and 28%, respectively) with those of the reaction conducted without **A0** (Figs. 5A and 6). Relative to the control experiment, the reductive alkylation among **1a, 2a**, and H_2 proceeded without a significant change in the yield of **3aa**, the recovered additives, or imine intermediates for additives with ketone (**A3**/**A13**), primary amide (**A4**), aryl bromide/iodide including alkyl ether (**A6**/**A7**), allylic ether (**A8**), terminal alkyne (**A9**), enone (**A12**), nitrile (**A15**), and ester (**A20**) moieties. It is also noteworthy that additives including sulfhydryl (**A16**) and sulfide (**A17**) moieties did not affect the present reaction, whereas such sulfur-containing compounds can be critical inhibitors in transition-metal-based catalysis and organocatalysis[42]. On the other hand, the hydrogenation of the imine intermediates was suppressed in the presence of aliphatic/aromatic carboxy (**A1**/**A14**), aliphatic hydroxyl (**A2**), bulky silyl ether (**A10**), pinacolatoboryl (**A18**), and *N-tert*-butoxycarbonyl (Boc) (**A21**) moieties, as these functional groups include either a Lewis-basic or -acidic site that can kinetically inhibit the formation of the FLP consisting of $\mathbf{B^{4b}}$ and MTHP. In fact, the quantitative recovery of the additives after a period of 24 h was confirmed, without the generation of any other significant byproduct, i.e., the sum of the yields of **3aa** and the imine was always ~95%. In contrast, the formation of **3aa** was largely

https://doi.org/10.1038/s41467-024-47984-0



Additive	3aa (%)	Additive remaining (%)	Imine (%)
A0	72	>99	28
	A control experiment		
A1	35 ⊗	>99 ±	56
A2	42 −	>99 ±	52
A3	68 ±	90 ±	24
A4	66 ±	94 ±	29
A5	<1 ⊗	99 ±	77
A6	71 ±	>99 ±	27
A7	74 ±	>99 ±	23
A8	72 ±	>99 ±	28
A9	70 ±	99 ±	24
A10	52 −	>99 ±	45

Additive	Scaled yield 3aa (%)	Additive remaining (%)	Imine (%)
A11	13 ⊗	85 ±	59
A12	65 ±	>99 ±	33
A13	61 ±	87 ±	23
A14	53 −	>99 ±	42
A15	72 ±	>99 ±	27
A16	71 ±	97 ±	27
A17	72 ±	91 ±	21
A18	56 −	>99 ±	40
A19	32 −	49 −	10
A20	71 ±	>99 ±	22
A21	50 −	>99 ±	43

± Same level as control

− Statistically significant decrease and equal or more than half of the mean of the control

⊗ Statistically significant decrease and less than half of the mean of the control

Fig. 6 | Exploring functional-group compatibility. Reaction conditions: **1a** (0.4 mmol, 0.05 M), **2a** (1.0 equiv.), additive (1.0 equiv.), **B^{4b}** (5 mol%), and 4 Å MS (100 mg) were mixed in MTHP, followed by pressurization with H_2 (40 atm). The mean values of two experiments are given, which were determined via ^{1}H NMR analysis.

suppressed under reaction conditions including *N*-heteroaromatic moieties such as an imidazole (**A5**) and an indole (**A19**), as these heteroaromatic units can react with **B^{4b}** to form Lewis adducts and/or with aldehydes to complicate the system. In the case of **A11**, which includes an epoxide moiety, **3aa** was only produced in 13% yield, and a significant loss of **A11** was confirmed after the reaction. Given that Lewis-acidic triarylboranes mediate the ring-opening transformation of epoxides[57,58], the **B^{4b}**-catalyzed ring-opening reaction of **A11** to give the corresponding aldehyde can be expected to compete with the targeted reaction (see Supplementary Information for details).

https://doi.org/10.1038/s41467-024-47984-0

Scope study

Finally, we applied the combination of **B**^{4b} and MTHP for the reductive alkylation of aniline-derived amino acids and peptide derivatives in the presence of H_2 (Fig. 7). Aminosalicylic acids **1b** and **1c** were effectively alkylated under the optimized conditions, and **3ba** and **3ca** were obtained in 90% and 93% yield, respectively; 60 atm of H_2 was used in the latter case. For comparison, under a pressure of 80 atm of H_2 in THF, **3ba** and **3ca** were furnished in 47% and 70% yield in the presence

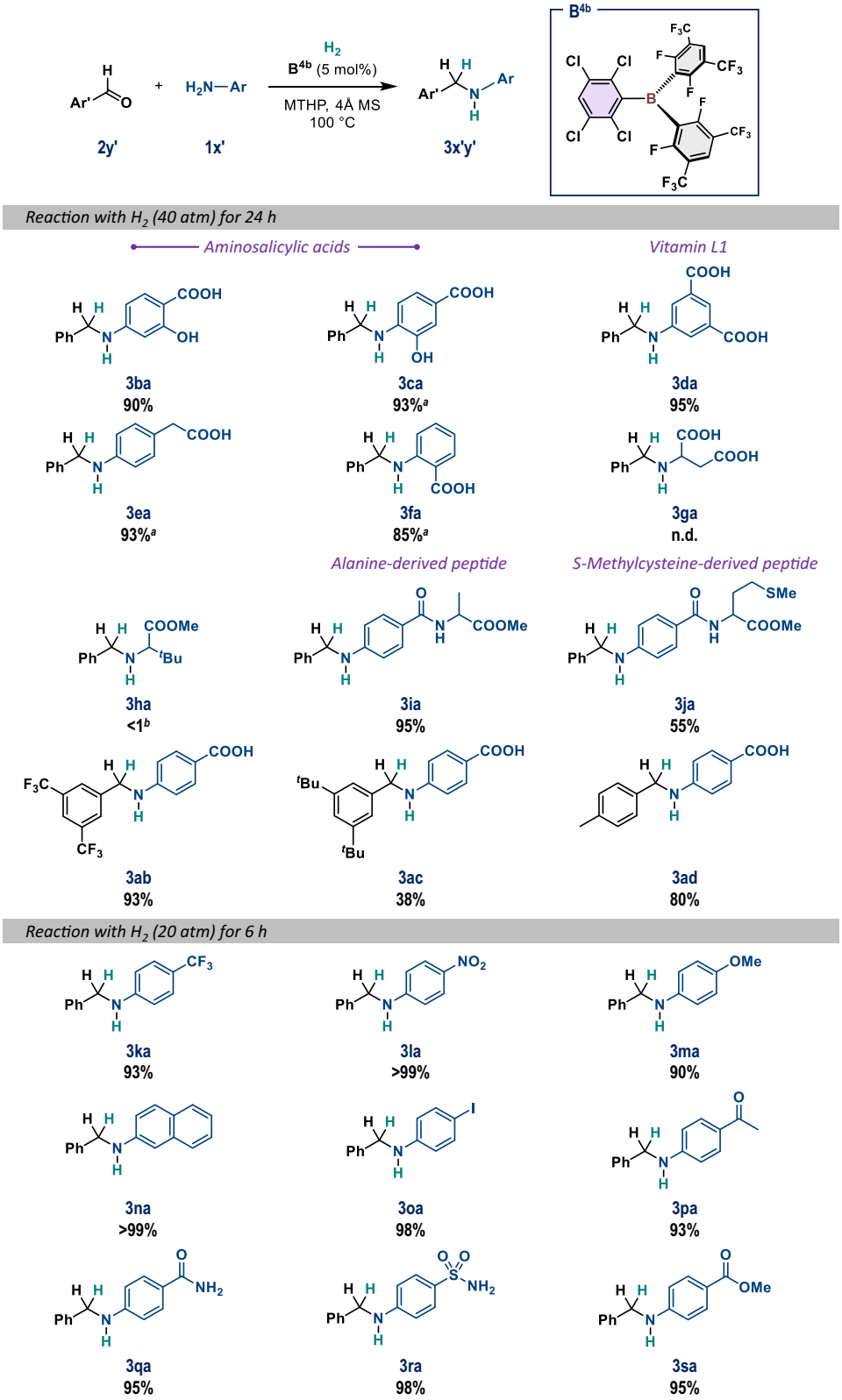


Fig. 7 | The B^{4b}-catalyzed reductive alkylation of aniline derivatives (1b–1s) with aldehydes (2a–d) using H_2 in MTHP. General conditions: **1** (0.4 mmol, 0.05 M), **2** (1.0 equiv.), **B**^{4b} (5 mol%), and 4 Å MS (100 mg) were mixed in MTHP, followed by pressurization with H_2. Yields of isolated products are given. [a]60 atm H_2. [b]The formation of the imine in >99% was confirmed.

of 10 mol% and 15 mol% **B**^{1a}, respectively, which again demonstrates the advantages of the present system using **B**^{4b} and MTHP in terms of synthetic efficiency and sustainability. The reductive alkylation of anthranilic acid (**1d**), which is also known as vitamin L1, 5-aminoisophthalic acid (**1e**), and 4-aminophenylacetic acid (**1f**) afforded **3da, 3ea**, and **3fa** in excellent yield using 40–60 atm of H_2. In contrast, we recognized that aliphatic amino acids (or their imine derivatives) and substrates insoluble in MTHP were not suitable. For example, aspartic acid (**1g**) is insoluble in MTHP, and no reaction took place when **1g** was employed under otherwise identical conditions. Esterification of the carboxy group in 3-amino-4,4-dimethylpentanoic acid effectively improved its solubility in MTHP; however, the hydrogenation of the imine derived from **1h** and **2a** did not occur. Based on these results, we prepared alanine- and *S*-methylcysteine-based peptides **1i** and **1j**, and subjected them to the optimal reaction conditions; alkylated peptides **3ia** and **3ja** were obtained in 95% and 55% yield, respectively. In terms of the scope of aldehydes, 3,5-bis(trifluoromethyl)benzaldehyde (**2b**) furnished **3ab** in 93% yield, but 3,5-di-*tert*-butylbenzaldehyde (**2c**) gave **3ac** in merely 38%. In the latter case, a significant amount of **2c** remained unreacted, indicating difficulties associated with the formation of the imine intermediate due to the decreased electrophilicity of the aldehyde moiety. A comparable result was obtained when *p*-tolualdehyde (**2d**) was used with respect to the case using **2a**, and **3ad** was afforded in 80% yield. As confirmed by the aforementioned robustness screening, the **B**^{4b}/MTHP system exhibited remarkable functional-group compatibility in the reductive alkylation of aniline derivatives to afford **3ka**–**3sa** in excellent yield. Given that harsh conditions, including 80 atm of H_2, 10 mol% borane, and/or a longer reaction time were required for the reactions with **1l, 1 m**, and **1q** in the reported **B**^{1a}/THF system, the present **B**^{4b}/MTHP system clearly demonstrates its advantages.

The present study demonstrates an in-silico-assisted approach to designing triarylboranes that exhibit promising reactivity as main-group catalysts for the reductive alkylation of aniline-derived amino acids with H_2. We have constructed an in-silico library of triarylboranes and obtained their theoretical parameters using DFT calculations. Guided by Gaussian process regression (GPR) using theoretical and experimental parameters, we identified the optimal triarylborane, i.e., B(2,3,5,6-Cl_4-C_6H)(2,6-F_2-3,5-$(CF_3)_2$-C_6H)$_2$ (**B**^{4b}). Through the evaluation of the regression-based models, we confirmed that the use of a parameter related to an intrinsic Lewis acidity of the triarylboranes (e.g., LUMO energy level and Mulliken charge on the boron atom) as one of the variables for the GPR analysis may lead to an underestimation when predicting the catalyst activity (TOF in h^{-1}) under the optimized reaction conditions. Moreover, we propose that the deformation energy (E_{DEF}) may serve as a potentially useful parameter to construct an adequate model. We also identified that 4-methyltetrahydropyran (MTHP) is a superior Lewis-basic solvent for not only the generation of FLP species with **B**^{4b}, but also for the realization of a more practical and less-hazardous reaction system compared to a system using THF. In fact, the **B**^{4b}-catalyzed reductive alkylation using aldehydes as an alkylating reagent and H_2 in MTHP proceeded efficiently even in the presence of a variety of additives, showcasing its broad functional-group compatibility. Aniline-derived amino acids and C-terminal-protected peptides were alkylated in good-to-excellent yields under the optimized conditions with the concomitant generation of H_2O as the sole byproduct.

Methods

General procedures for reductive alkylation of **1x′** with **2y′** affording **3x′y′**: A 30 mL autoclave was charged with **1x′** (0.40 mmol), **2 y′** (0.40 mmol), **B**^{4b} (0.02 mmol), 4 Å MS (100 mg), and MTHP (8 mL). Once sealed, the vessel was pressurized with H_2 (40 or 60 atm), and the reaction mixture was stirred at 100 °C for 24 h. Then, degassed at rt followed by the addition of acetone, the resultant mixture was filtered to remove MS and other solids when generated. Subsequently, all volatiles were removed in vacuo to give **3x′y′**, which was purified by flash column chromatography on silica gel.

Data availability

Data generated or analyzed during this study are provided in full within the published article and its supplementary materials. Metrical data for the solid-state structures are available from the Cambridge Crystallographic Data Centre (CCDC) under reference numbers 2295627 (**B**^{1f}), 2295628 (**B**^{2b}), 2295633 (**B**^{2c}), 2295634 (**B**^{2e}), 2295629 (**B**^{3b}), 2295635 (**B**^{3c}), 2295631 (**B**^{3s}), 2295632 (**B**^{4b}), and 2295630 (**B**^{4e}). These data can be obtained free of charge from the CCDC via www.ccdc.cam.ac.uk/data_request/cif. Coordinates of the optimized structures are provided as source data. All other data are available from the corresponding author. Source data are provided with this paper.

Code availability

The code used in this work can be found in the Zenodo repository at https://doi.org/10.5281/zenodo.8420294.

References

1. Vogiatzis, K. D. et al. Computational approach to molecular catalysis by 3d transition metals: challenges and opportunities. *Chem. Rev.* **119**, 2453–2523 (2019).
2. Toyao, T. et al. Machine learning for catalysis informatics: recent applications and prospects. *ACS Catal.* **10**, 2260–2297 (2020).
3. Foscato, M. & Jensen, V. R. Automated in silico design of homogeneous catalysts. *ACS Catal.* **10**, 2354–2377 (2020).
4. Matsubara, S. Digitization of organic synthesis — how synthetic organic chemists use AI technology —. *Chem. Lett.* **50**, 475–481 (2021).
5. Zahrt, A. F., Athavale, S. V. & Denmark, S. E. Quantitative structure–selectivity relationships in enantioselective catalysis: past, present, and future. *Chem. Rev.* **120**, 1620–1689 (2020).
6. Rinehart, N. I., Zahrt, A. F., Henle, J. J. & Denmark, S. E. Dreams, false starts, dead ends, and redemption: a chronicle of the evolution of a chemoinformatic workflow for the optimization of enantioselective catalysts. *Acc. Chem. Res.* **54**, 2041–2054 (2021).
7. Crawford, J. M., Kingston, C., Toste, F. D. & Sigman, M. S. Data science meets physical organic chemistry. *Acc. Chem. Res.* **54**, 3136–3148 (2021).
8. Liles, J. P. et al. Data science enables the development of a new class of chiral phosphoric acid catalysts. *Chem.* **9**, 1518–1537 (2023).
9. Chen, H. et al. Data-driven catalyst optimization for stereodivergent asymmetric synthesis by iridium/boron hybrid catalysis. *Cell Rep. Phys. Sci.* **2**, 100679 (2021).
10. Gensch, T. et al. A comprehensive discovery platform for organophosphorus ligands for catalysis. *J. Am. Chem. Soc.* **144**, 1205–1217 (2022).
11. Xu, J. et al. Atroposelective Negishi coupling optimization guided by multivariate linear regression analysis: asymmetric synthesis of KRAS G12C covalent inhibitor GDC-6036. *J. Am. Chem. Soc.* **144**, 20955–20963 (2022).
12. Karl, T. M. et al. Machine learning-guided development of trialkylphosphine $Ni^{(I)}$ dimers and applications in site-selective catalysis. *J. Am. Chem. Soc.* **145**, 15414–15424 (2023).
13. Goebel, J. F. et al. Computer-driven development of ylide functionalized phosphines for palladium-catalyzed Hiyama couplings. *Angew. Chem. Int. Ed.* **62**, e202216160 (2023).
14. Huang, H., Zong, H., Bian, G., Yue, H. & Song, L. Correlating the effects of the N-substituent sizes of chiral 1,2-amino phosphinamide ligands on enantioselectivities in catalytic asymmetric Henry reaction using physical steric parameters. *J. Org. Chem.* **79**, 9455–9464 (2014).
15. See, X. Y. et al. Iterative supervised principal component analysis driven ligand design for regioselective Ti-catalyzed pyrrole synthesis. *ACS Catal.* **10**, 13504–13517 (2020).

16. Mukai, M., Nagao, K., Yamaguchi, S. & Ohmiya, H. Molecular field analysis using computational-screening data in asymmetric *N*-heterocyclic carbene-copper catalysis toward data-driven in silico catalyst optimization. *Bull. Chem. Soc. Jpn.* **95**, 271–277 (2022).
17. Wang, J.-W. et al. Nickel-catalyzed remote asymmetric hydroalkylation of alkenyl ethers to access ethers of chiral dialkyl carbinols. *J. Am. Chem. Soc.* **145**, 10411–10421 (2023).
18. Cordova, M., Wodrich, M. D., Meyer, B., Sawatlon, B. & Corminboeuf, C. Data-driven advancement of homogeneous nickel catalyst activity for aryl ether cleavage. *ACS Catal.* **10**, 7021–7031 (2020).
19. Stephan, D. W. & Erker, G. Frustrated Lewis pair chemistry: development and perspectives. *Angew. Chem. Int. Ed.* **54**, 6400–6441 (2015).
20. Jupp, A. R. & Stephan, D. W. New directions for frustrated Lewis pair chemistry. *Trends Chem.* **1**, 35–48 (2019).
21. Weicker, S. A. & Stephan, D. W. Main group Lewis acids in frustrated Lewis pair chemistry: beyond electrophilic boranes. *Bull. Chem. Soc. Jpn.* **88**, 1003–1016 (2015).
22. Scott, D. J., Fuchter, M. J. & Ashley, A. E. Designing effective 'frustrated Lewis pair' hydrogenation catalysts. *Chem. Soc. Rev.* **46**, 5689–5700 (2017).
23. Hoshimoto, Y. & Ogoshi, S. Triarylborane-catalyzed reductive N-alkylation of amines: a perspective. *ACS Catal.* **9**, 5439–5444 (2019).
24. Fasano, V. & Ingleson, M. J. Recent advances in water-tolerance in frustrated Lewis pair chemistry. *Synthesis* **50**, 1783–1795 (2018).
25. Paradies, J. Structure-reactivity relationships in borane-based FLP-catalyzed hydrogenations, dehydrogenations, and cycloisomerizations. *Acc. Chem. Res.* **56**, 821–834 (2022).
26. Feng, X. & Du, H. Metal-free asymmetric hydrogenation and hydrosilylation catalyzed by frustrated Lewis pairs. *Tetrahedron Lett.* **55**, 6959–6964 (2014).
27. Gyömöre, Á. et al. Moisture-tolerant frustrated Lewis pair catalyst for hydrogenation of aldehydes and ketones. *ACS Catal.* **5**, 5366–5372 (2015).
28. Dorkó, É. et al. Expanding the boundaries of water-tolerant frustrated Lewis pair hydrogenation: enhanced back strain in the Lewis acid enables the reductive amination of carbonyls. *Angew. Chem. Int. Ed.* **56**, 9512–9516 (2017).
29. Sapsford, J. S. et al. Direct reductive amination of carbonyl compounds catalyzed by a moisture tolerant tin(IV) Lewis acid. *Adv. Synth. Catal.* **360**, 1066–1071 (2018).
30. Hoshimoto, Y., Kinoshita, T., Hazra, S., Ohashi, M. & Ogoshi, S. Main-group-catalyzed reductive alkylation of multiply substituted amines with aldehydes using H_2. *J. Am. Chem. Soc.* **140**, 7292–7300 (2018).
31. Zhao, J. et al. A theoretical study on the borane-catalyzed reductive amination of aniline and benzaldehyde with dihydrogen: the origins of chemoselectivity. *J. Org. Chem.* **87**, 1194–1207 (2022).
32. Bijoy, R., Agarwala, P., Roy, L. & Thorat, B. N. Unconventional ethereal solvents in organic chemistry: a perspective on applications of 2-methyltetrahydrofuran, cyclopentyl methyl ether, and 4-methyltetrahydropyran. *Org. Process. Res. Dev.* **26**, 480–492 (2022).
33. Samoilov, V. et al. Bio-based solvents and gasoline components from renewable 2,3-butanediol and 1,2-propanediol: Synthesis and characterization. *Molecules* **25**, 1723 (2020).
34. Bryan, M. C. et al. Key green chemistry research areas from a pharmaceutical manufacturers' perspective revisited. *Green Chem.* **20**, 5082–5103 (2018).
35. Murugesan, K. et al. Catalytic reductive aminations using molecular hydrogen for synthesis of different kinds of amines. *Chem. Soc. Rev.* **49**, 6273–6328 (2020).
36. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
37. Carden, J. L., Dasgupta, A. & Melen, R. L. Halogenated triarylboranes: synthesis, properties and applications in catalysis. *Chem. Soc. Rev.* **49**, 1706–1725 (2020).
38. He, J., Rauch, F., Finze, M. & Marder, T. B. (Hetero)arene-fused boroles: a broad spectrum of applications. *Chem. Sci.* **12**, 128–147 (2021).
39. Kumar, G., Roy, S. & Chatterjee, I. Tris(pentafluorophenyl)borane catalyzed C–C and C–heteroatom bond formation. *Org. Biomol. Chem.* **19**, 1230–1267 (2021).
40. Berger, S. M., Ferger, M. & Marder, T. B. Synthetic approaches to triarylboranes from 1885 to 2020. *Chem. Eur. J.* **27**, 7043–7058 (2021).
41. Das, S., Turnell-Ritson, R. C., Dyson, P. J. & Corminboeuf, C. Design of frustrated Lewis pair catalysts for direct hydrogenation of CO_2. *Angew. Chem. Int. Ed.* **61**, e202208987 (2022).
42. Saito, N. et al. Functional group evaluation kit for digitalization of information on the functional group compatibility and chemoselectivity of organic reactions. *Bull. Chem. Soc. Jpn.* **96**, 465–474 (2023).
43. Collins, K. D. & Glorius, F. A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **5**, 597–601 (2013).
44. Collins, K. D. & Glorius, F. Intermolecular reaction screening as a tool for reaction evaluation. *Acc. Chem. Res.* **48**, 619–627 (2015).
45. Hashimoto, T., Asada, T., Ogoshi, S. & Hoshimoto, Y. Main group catalysis for H_2 purification based on liquid organic hydrogen carriers. *Sci. Adv.* **8**, eade0189 (2022).
46. Erös, G. et al. Catalytic hydrogenation with frustrated Lewis pairs: selectivity achieved by size-exclusion design of Lewis acids. *Chem. Eur. J.* **18**, 574–585 (2012).
47. Dorkó, É. et al. Correlating electronic and catalytic properties of frustrated Lewis pairs for imine hydrogenation. *J. Organomet. Chem.* **847**, 258–262 (2017).
48. Erdmann, P. & Greb, L. What distinguishes the strength and the effect of a Lewis acid: analysis of the Gutmann–Beckett method. *Angew. Chem. Int. Ed.* **61**, e202114550 (2022).
49. Sakuraba, M., Morishita, T., Hashimoto, T., Ogoshi, S. & Hoshimoto, Y. Remote back strain: a strategy for modulating the reactivity of triarylboranes. *Synlett* **34**, 2187–2192 (2023).
50. Fasano, V., Radcliffe, J. E. & Ingleson, M. J. $B(C_6F_5)_3$-catalyzed reductive amination using hydrosilanes. *ACS Catal.* **6**, 1793–1798 (2016).
51. The GPy authors. GPy: A Gaussian process framework in python. http://github.com/SheffieldML/Gpy.
52. Kondo, M. et al. Exploration of flow reaction conditions using machine-learning for enantioselective organocatalyzed Rauhut-Currier and [3+2] annulation sequence. *Chem. Commun.* **56**, 1259–1262 (2020).
53. Silva, D. R., Santos, L. A., Freitas, M. P., Guerra, C. F. & Hamlin, T. A. Nature and strength of Lewis acid/base interaction in boron and nitrogen trihalides. *Chem. Asian J.* **15**, 4043–4054 (2020).
54. Timoshkin, A. Y., Davydova, E. I., Sevastianova, T. N., Suvorov, A. V. & Schaefer, H. F. Relationship between the energy of donor–acceptor bond and the reorganization energy in molecular complexes. *Int. J. Quantum Chem.* **88**, 436–440 (2002).
55. Kumar, P. S. V., Raghavendra, V. & Subramanian, V. Bader's theory of atoms in molecules (AIM) and its applications to chemical bonding. *J. Chem. Sci.* **128**, 1527–1536 (2016).
56. Macchi, P. & Sironi, A. Chemical bonding in transition metal carbonyl clusters: complementary analysis of theoretical and experimental electron densities. *Coord. Chem. Rev.* **238–239**, 383–412 (2003).
57. Bhagat, M. N. et al. Enhancing the regioselectivity of $B(C_6F_5)_3$-catalyzed epoxide alcoholysis reactions using hydrogen-bond acceptors. *ACS Catal.* **9**, 9663–9670 (2019).
58. Andrea, K. A. & Kerton, F. M. Triarylborane-catalyzed formation of cyclic organic carbonates and polycarbonates. *ACS Catal.* **9**, 1799–1809 (2019).

Acknowledgements

We thank Kuraray Co., Ltd., for providing 4-methyltetrahydropyran. This project was supported by Grants-in-Aid for Transformative Research Area (A) Digitalization-Driven Transformative Organic Synthesis (JSPS

KAKENHI grants 22H05363 to Y.Ho. and 21H05217 to S.T.) as well as the Environment Research and Technology Development Fund (JPMEERF20211R01 to Y.Ho.) of the Environmental Restoration and Conservation Agency of the Ministry of the Environment of Japan. Part of this work was supported by JST SPRING (grant JPMJSP2138 to Y.Hi.). Parts of the calculations were performed using resources from the Research Center for Computational Science, Okazaki, Japan (22-IMS-C107 and 23-IMS-C094).

Author contributions

Y.Ho. conceived and directed this project. Y.Hi. and Y.Ho. performed the experiments and theoretical calculations. Y.Hi., T.W., and S.T. performed the Gaussian process regression. Y.Hi. prepared an initial draft of the manuscript, which was finalized by Y.Ho. with feedback from T.W., S.T., and S.O.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-47984-0.

Correspondence and requests for materials should be addressed to Yoichi Hoshimoto.

Peer review information *Nature Communications* thanks Tiago Rodrigues, Jiyang Zhao and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.