Direct α-Trifluoromethylthiolation of Carboxylic Acids Enabled by Boron Catalysis

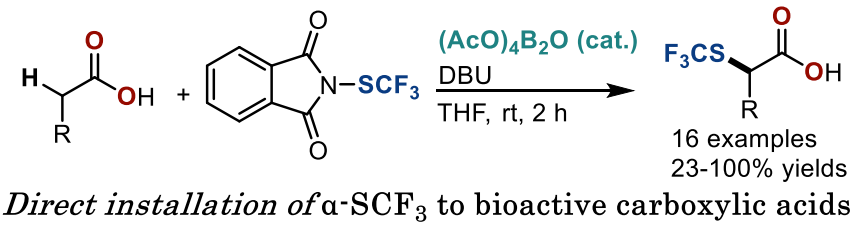
Kai Sun[b]
Chung-Yang (Dennis) Huang*[a]
Masaya Sawamura[a,b]
Yohei Shimizu*[a,b]

[a]Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan

[b]Department of Chemistry, Faculty of Science, Hokkaido University, Kita 10 Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan

dcyhuang@icredd.hokudai.ac.jp
shimizu-y@sci.hokuai.ac.jp

Click here to insert a dedication.



Received:
Accepted:
Published online:
DOI:

Abstract A boron-catalyzed direct α-trifluoromethylthiolation of carboxylic acids was developed. Catalytically generated boron enediolates reacts with electrophilic SCF_3 reagent, *N*-SCF_3-phthalimide, to provide α-SCF_3 carboxylic acids without the need of substrate pre-activation. The method is applicable to direct modification of bioactive carboxylic acids. Data science analyses provided suitable models for substrate classification as well as yield prediction.

Key words Carboxylic acid, Boron catalyst, Trifluoromethylthiolation, Data science, Drug modification, Fluorine compounds

Incorporation of fluorine atoms and fluoro-functional groups in a molecule has been essential for the development of pharmaceuticals and agrochemicals.[1] Due to the unique properties of fluorine atom, such as the highest electronegativity, high lipophilicity, and small van der Waals radius, replacement of hydrogens and oxygens by fluorine atoms can result in improved pharmacokinetic (PK) properties of bioactive compounds, for example, enhanced bioavailability and metabolic stability. Hence, there is a continuous demand to develop novel synthetic methodologies for the introduction of fluoro-functional groups to various molecules of interest.

Trifluoromethylthio (SCF_3) group is one of the most attractive moieties due to its extremely high lipophilicity (Hansch parameter π= 1.44) (Figure 1a).[2] Thus, methods for trifluoromethylthiolation of various molecular motifs have witnessed a significant progress in the past decade, either through nucleophilic, electrophilic, or radical mechanism.[3] In particular, introduction of SCF_3 groups to the α-position of carbonyl groups, as found in cefazaflur, has the potential to greatly enhance their pharmaceutical properties. An ideal approach to construct such motifs is catalytic direct trifluoromethylthiolation via deprotonative enolate formation of carbonyl compounds. The method, however, is only applicable to easily enolizable 1,3-dicarbonyl compounds,[4] ketones,[5] and activated carboxylic acid derivatives.[6] To install an SCF_3 group at the α-position of non-activated carboxylic acid derivatives, such as esters and amides, preparation of α-diazo[7], α-bromo[8], or other pre-activated compounds[9] is necessary (Figure 1b). To provide a more straightforward protocol, we envisioned that a boron-catalyzed carboxylic acid enolate formation approach, previously developed by some of us,[10] would enable direct α-trifluoromethylthiolation of carboxylic acids (Figure 1c).

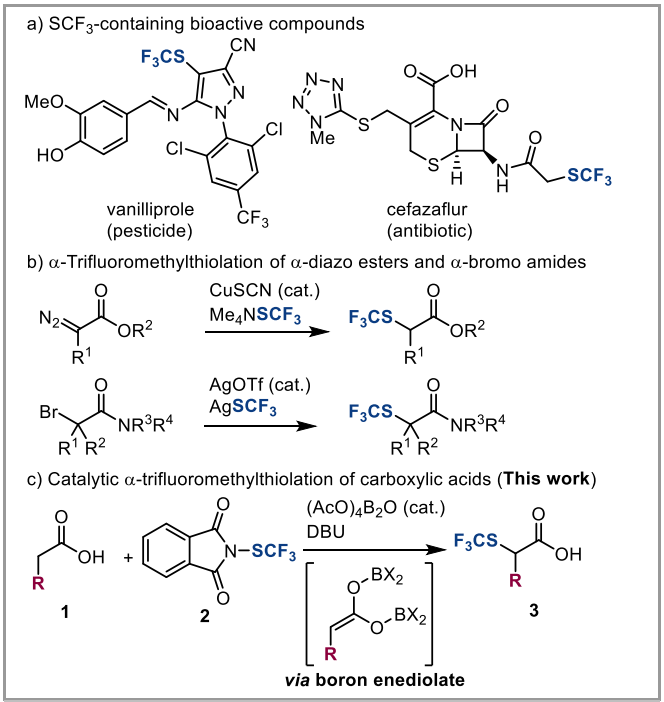


Figure 1 SCF_3-containing bioactive compounds and α-trifluoromethylthiolation of carboxylic acid derivatives.

Indeed, the reaction between 4-methoxyphenylacetic acid **1a** and *N*-SCF_3-phthalimide **2** was promoted by treating the mixture with $(AcO)_4B_2O$ (10 mol%) and DBU (2.4 equiv) in THF at room temperature. The reaction completed within 2 hours, and α-SCF_3 product was obtained in high yield. Scope of the reaction was assessed by a one-pot conversion to methyl ester for the sake of easy isolation (Figure 2). α-Aryl carboxylic acids with an electron-donating substituent, such as methoxy (**1a**), methyl (**1b**) and methylthio (**1c**) groups, at the para position of the aryl groups were competent substrates, giving the corresponding products in high yields. On the other hand, electron-deficient substituents, for example phenyl and chloro groups, exhibited negative effect on the yields, affording **4d** in 44% and **4e** in 41% yields, respectively. Furthermore, **1f** bearing highly electron-deficient pentafluoroaryl group provided **4f** in only 9% yield. It should be noted that the reaction of **1e** produced decarboxylated compound, namely, (4-chlorobenzyl)(trifluoromethyl)sulfane **5e** as the major side product. Sterically demanding *o*-methoxy (**4g**) and *o*-methyl (**4h**) substituents were well tolerated. In addition, carboxylic acids containing heteroaryl groups, thiophene (**1i**) and benzodioxol (**1j**), also showed high reactivity, albeit with low isolated yield for **4i** due to its volatile nature.

The reaction was applicable to a wide variety of bioactive agrochemicals and pharmaceuticals, demonstrating its utility in complex molecule synthesis (Figure 2). Carboxylic acids containing a sterically demanding naphthalene ring, 1-naphthaleneacetic acid (NAA) **1k** and 2-napththaleneacetic acid **1l** (anti-inflammatory naproxen derivative), were reactive, and α-SCF_3 analogues **4k** and **4l** were obtained in moderate yields. Amide-containing carboxylic acids, indomethacin **1m** and actarit **1n**, were also amenable, and SCF_3 groups were introduced selectively at the α position of the carboxy groups. The keto group in isoxepac **1o** remained intact under the reaction conditions, and the product **4o** was obtained in high yield. Despite the absence of α-aryl group, oxaprozin **1p** afforded SCF_3 product **4p** in 23% yield.

Although the current protocol shows broad compatibility, we also identified the limitations in several types of carboxylic acids. Representative examples are as followed (Figure 2). Severe steric hindrance caused by α-mesityl group (**1q**) or α,α-disubstitution (**1r**) completely inhibited the reaction. In addition, carboxylic acids with α-nitrogen substituent (**1s**) and 3-indolyl group bearing free NH moiety (**1t**) were not suitable at all. Other unsuccessful substrates are listed in Supporting Information.

Since decarboxylated products were obtained as the major side products, experiments were performed to gain insight into the mechanism of the decarboxylation. First, α-SCF_3 carboxylic acid **4e** was subjected to several reaction conditions in order to elucidate the requirement for inducing decarboxylation. When **4e** was treated with DBU (2.4 equiv) and $(AcO)_4B_2O$ (10 mol%), the decarboxylation product **5e** was obtained in 30% yield (Table 1, entry 1). Moreover, **5e** was produced in the presence of DBU without boron catalyst (entry 2). On the other hand, no decarboxylation was observed without DBU (entry 3). Thus, the decarboxylation most likely occurs from α-SCF_3 carboxylic acids via carboxylate formation, followed by ionic decarboxylation to form the benzylic anion, which would be stabilized by a strongly electron-withdrawing SCF_3 group and electron-deficient aryl group.

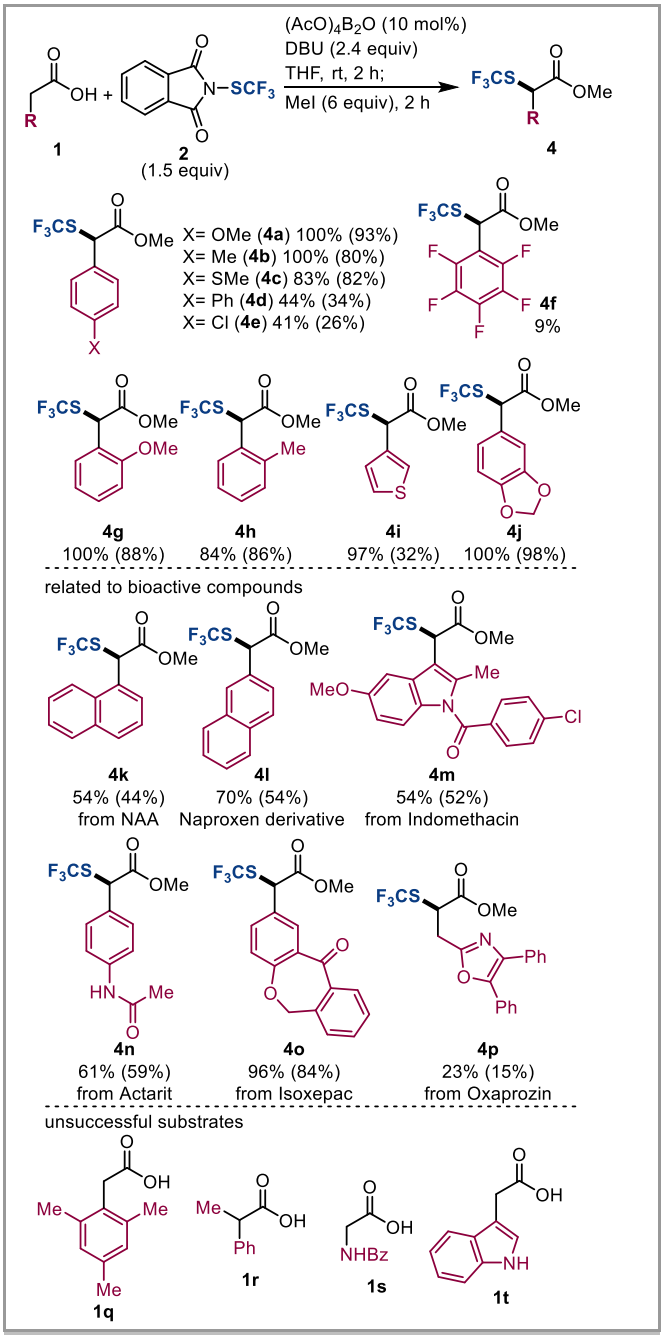
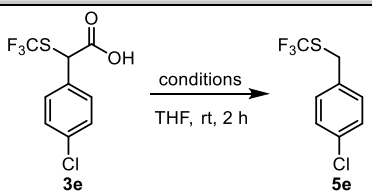


Figure 2 Substrate scope: **1** (0.10 mmol), **2** (0.15 mmol), DBU (0.24 mmol), THF (0.1 M), room temperature, 2h; then MeI (6.0 equiv), 2h. Yields were determined by ^{1}H NMR analysis using tetrabromoethane as the standard and ^{19}F NMR analysis using hexafluorobenzene as the standard. Yields of the isolated products are shown in parentheses.

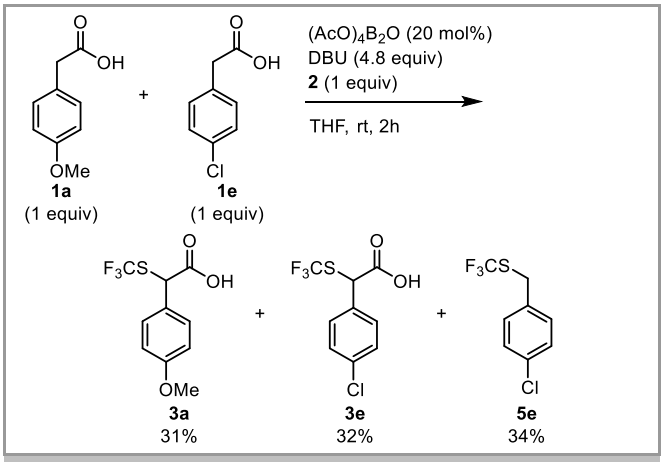
Table 1 Decarboxylation of α-SCF_3 carboxylic acid **3e**



Entry	Conditions	Yield of **5e** (%)[a]
1	DBU (2.4 equiv), $(AcO)_4B_2O$ (10 mol%)	30
2	DBU (2.4 equiv)	26
3	$(AcO)_4B_2O$ (10 mol%)	0

[a]Yields were determined by ^{1}H NMR analysis using tetrabromoethane as the standard and ^{19}F NMR analysis using hexafluorobenzene as the standard.

Next, a competition experiment between **1a** and **1e** was performed (Scheme 1). As a result, equimolar amount of **3a** and **3e** along with same amount of **5e** were obtained, while no decarboxylated product derived from **1a** was obtained. This outcome suggests that the reaction rate of **1e** is higher than that of **1a**, but rapid decarboxylation of **3e** to **5e** diminishes the yield.



Scheme 1 Competition experiment

A preliminary examination of the scope revealed that substrates bearing electron-rich aryl group at the α-position resulted in higher reaction efficiency. To provide further mechanistic insights, we resorted to data science analysis to identify molecular characters of these compounds that are critical to the reaction yields. Specifically, we carried out DFT calculations on the ground-state structures of all the substrates, including compounds that provided only low or no reactivity. Among all, 32 distinct descriptors that cover the electronic and energetic properties of these molecules were selected (see Supporting Information for details).

With these descriptors in hand, we first tested univariate classification to distinguish successful and unsuccessful substrates. This simple yet reliable method has been proven powerful in its application in organic chemistry.[11] By choosing 20% yield as the threshold for successful substrates, we were able to identify 2 descriptors that provide clear classification of odds >3.1 (Figure 3a, b): **NBOC2** representing NBO charge of C2, and **C12D** representing the C1–C2 bond length deviation from 1.508 Å (Figure 4). Mechanistically, it is not surprising that these two descriptors closely related to the reacting α-carbons are the most influential, where the reactivity of the boron-enolates or the stability of the products may play a crucial role.

Next, we tested bivariate classification to see whether more accurate predictions could be achieved. To our delight, the performance of the above classification models was improved by including another descriptor, namely HOMO energy, providing an outstanding odds of 8.7 as a simple classification (Figure 3c, d). Intriguingly, three particular substrates failed in both models to be correctly classified: **1q**, **1t**, and sulindac **S9**. This outcome suggests that for these three substrates, distinct mechanisms not captured by our models may be responsible for their low reactivity. For example, **1q** bearing the most sterically hindered aryl substituent may face difficulty generating the enolate intermediate. The nucleophilic N–H of **1t** on the aryl ring may cause competing side reactions. Furthermore, the enolate formed from **S9** is conjugated with the extended olefin systems that can complicate the reactivity. Nevertheless, we consider these classification models useful in the event of predicting whether a new substrate should be active under this reaction condition.

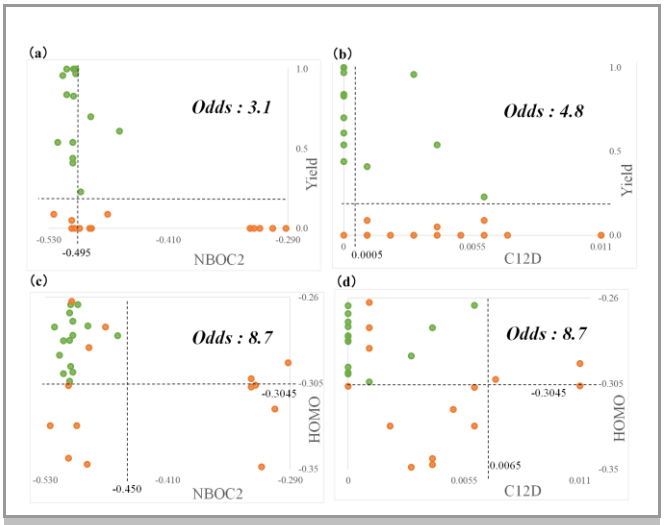


Figure 3 Classification models for successful and unsuccessful substrates. Green data points: successful substrate (>20% yield), orange data points: unsuccessful substrate (<20% yield). **Odds** = (correct classifications)/(incorrect classifications). Borderline: (a) **NBOC2** = -0.495, (b) **C12D** = 0.0005, (c) **NBOC2** = -0.450, **HOMO** = -0.3045. (d) **C12D** = 0.0065, **HOMO** = -0.3045.

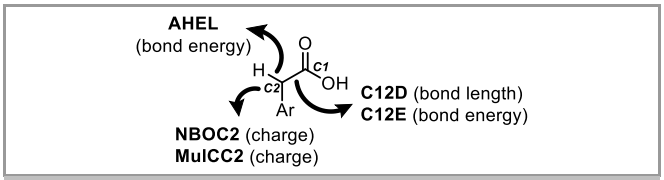


Figure 4. Descriptors for data science analyses. **NBOC2**: NBO charge of α-carbon; **C12D**: absolute deviation of C1-C2 bond length from 1.508 Å; **MulCC2**: Mullikan charge at α-carbon; **AHEL**: lowest α-C-H bond NBO energy; **C12E**: C1-C2 bond NBO energy.

Beyond classification, we were interested in further correlating the yields with molecular descriptors. Among all regression methods, logistic regression, one of the simplest non-linear regressions, was chosen with the limitation to three or less descriptors.[11b,12,13] As shown in Figure 5, moderate performances could be achieved in three different models utilizing either two descriptors among **C12D**, **MulCC2**, **AHEL**, and **C12E** (Figure 5a, b, c: R^2 = 0.55–0.63). Importantly, a more intricate model with a combination of three descriptors, **C12D**, **MulCC2**, and **AHEL**, exhibited an improved level of correlation (Figure 5d, R^2 = 0.75). This method signifies that the charges, energies, and bond lengths related to the reacting α-carbon site can be utilized as quantitative descriptors for yield correlation.

In summary, we developed the first catalytic direct α-trifluoromethylthiolation of carboxylic acids enabled by boron catalysis. Carboxylic acids with electron-rich α-aryl substituents were competent substrates, whereas electron-deficient α-aryl substituents induced decarboxylation from the α-SCF_3 products. Preliminary data science analyses provided suitable models for the classification of successful or unsuccessful substrates and models for predicting the yields. The results demonstrated that it can be utilized in combination with traditional mechanistic

studies to provide deeper insights of the chemical systems, wherein expert knowledge can be useful in guiding the development of new synthetic methodologies. Efforts to apply these models in developing related α-functionalizations of carboxylic acids are currently underway.

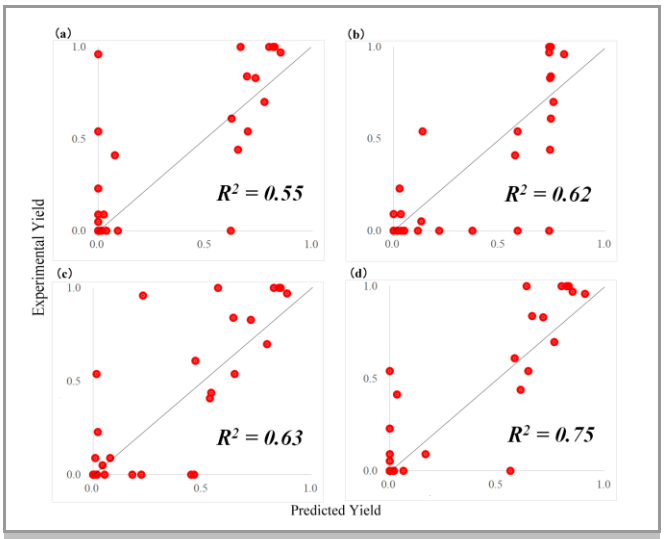


Figure 5. Diagrams of logistic regressions. **Predicted Yield** = 1 / (1 + $e^{(-z)}$), (a) z = -3606.81****C12D*** + 6.15136****MulCC2*** + 3.46992, (b) z = -707.464****C12D*** - 14.2310****AHEL*** + 23.2475****C12E*** + 9.52578, (c) z = -1061.85****C12D*** - 1.39554****MulCC2*** - 16.1425****MulCC2***^{2} + 2.96943 (***MulCC2***^{2} is the square of **MulCC2**), (d) z = 7.37656****MulCC2*** - 4364.12****C12D*** + 368.455****AHEL***$_c^2$ + 3.75559 (***AHEL***$_c^2$ is the square of centralized **AHEL**).

Funding Information

This work was supported by JSPS KAKENHI Grant No. JP22H05329 in Degitalization-driven Transformative Organic Synthesis, Grant-in-Aid for Scientific Research (B) (No. JP20H02729), and Grant-in-Aid for Challenging Reserach (Exploratory, No. JP22K19016) to YS. CYH acknowledged the financial supports by Institute for Chemical Reaction Design and Discovery (WPI-ICReDD) and Hokkaido University.

Supporting Information

YES (this text will be updated with links prior to publication)

Primary Data

NO.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

(1) (a) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37*, 320-330. (b) Kirik, K. L. *J. Fluorine Chem.***2006**, *127*, 1013-1029. (c) Zhou, Y.; Wang, J.; Gu, Z.; Wang, S.; Zhu, W.; Aceña, J. L.; Soloshonok, V. A.; Izawa, K.; Liu, H.; *Chem. Rev.* **2016**, *116*, 422-518. (d) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A.; *J. Med. Chem.* **2015**, *58*, 8315-8359. (e) Zanda, M. *New J. Chem.* **2004**, *28*, 1401–1411. (f) Jeschke, P. *ChemBioChem* **2004**, *5*, 570-589.

(2) (a) Hansch, C.; Leo, A. *Substituent constants for correlation analysis in chemistry and biology*, Wiley, New York, 1979, p. 339. (b) Barata-Vallejo, S.; Bonesi, S.; Postigo, A. *Org. Biomol. Chem*. **2016**, *14*, 7150-7182.

(3) (a) Rossi, S.; Puglisi, A.; Raimondi, L.; Benaglia, M. *ChemCatChem* **2018**, *10*, 2717-2733. (b) Barthelemy, A.-L.; Manier, E.; Dagousset, G. *Synthesis*, **2018**, *50*, 4765-4776. (c) Ghiazza, C.; Billard, T. *Eur. J. Org. Chem.* **2021**, 5571-5584.

(4) (a) Saidalimu, I.; Yoshioka, T.; Liang, Y.; Tokunaga, E.; Shibata, N. *Chem. Commun.* **2018**, *54*, 8761-8764. (b) Kondo, H.; Maeno, M.; Sasaki, K.; Guo, M.; Hashimoto, M.; Shiro, M.; Shibata, N. *Org. Lett.* **2018**, *20*, 7044-7048. (c)Jin, M. Y.; Gu, X.; Deng, M. Wang, C. Wang, J. *Chem. Commun.* **2020**, *56*, 10552-10555.

(5) (a) Alazet, S.; Ismalaj, E.; Glenadel, Q.; Le Bars, D.; Billard, T. *Eur. J. Org. Chem.* **2015**, 4607-4610. (b) Jiang, L.;Yan, Q.; Wang, R.; Ding, T.; Yi, W.; Zhang, W. *Chem. Eur. J.* **2018**, *24*, 18749-18756.

(6) (a) Capaccio, V.; Sicignano, M.; Rodríguez, R. I.; Sala, G. D.; Aléman, J. *Org. Lett.* **2020**, *22*, 219-223. (b) Franco, F.; Meninno, S.; Lattanzi, A.; Puglisi, A. Benaglia, M. *J. Org. Chem.* **2021**, *86*, 14207-14212. (c) Franco, F.; Meninno, S.; Benaglia, M. Lattanzi, A. *Chem. Cummun.* **2020**, *56*, 3073-3076. (d) Zhang, H.; Shen, Q. *Tetrahedron* **2021**, *101*, 132508. (e) Sicignano, M.; Rodríguez, R. I.; Capaccio, V.; Borello, F.; Cano, R.; De Riccardis, F.; Bernardi, L.; Díaz-Tendero, S.; Sala, G. D.; Aléman, J. *Org. Biomol. Chem.* **2020**, *18*, 2914-2920.

(7) (a) Lübcke, M.; Yuan, W.; Szabó, K. J. *Org. Lett.* **2017**, *19*, 4548-4551. (b) Zhang, Z.; Sheng, Z. Yu, W.; Wu, G.; Zhang, R.; Chu, W.-D.; Zhang, Y.; Wang, J. *Nat. Chem.* **2017**, *9*, 970-976. (c)Hu, M.; Rong, J.; Miao, W.; Ni, C.; Han, Y.; Hu, J. *Org. Lett.* **2014**, *16*, 2030-2033. (d) Wang, X.; Zhou, Y. J.; Ji, G.; Wu, G.; Li, M.; Zhang, Y.; Wang, J. *Eur. J. Org. Chem.* **2014**, 3093-3096. (e) Lefebvre, Q.; Fava, E.; Nikolaienko, P.; Rueping, M. *Chem. Commun.* **2014**, *50*, 6617-6619.

(8) (a) Mizuta, S.; Kitamura, K. Morii, Y.; Ishihara, J.; Yamaguchi, T.; Ishikawa, T. *J. Org. Chem.* **2021**, *86*, 18017-18029. (b) Li, S. G.; Zard, S. Z. *Org. Lett.* **2013**, *15*, 5898-5901.

(9) Jiang, X.; Meyer, D.; Baran, D.; Cortés, M. A.; Szabó, K. J. *J. Org. Chem.* **2020**, *85*, 8311-8319.

(10) (a) Shimizu, Y.; Kanai, M. *Chem. Rec.* **2023**, e202200273. (b) Sawamura, M.; Shimizu, Y. *Eur. J. Org. Chem.* **2023**, *26*, e202201249. (c) Sun, K.; Ueno, M.; Imaeda, K.; Ueno, K.; Sawamura, M.; Shimizu, Y. *ACS Catal.* **2021**, *11*, 9722–9728. (c) Morisawa, T.; Sawamura, M.; Shimizu, Y. *Org. Lett.* **2019**, *21*, 7466–7469. (d) Fujita, T.; Yamamoto, T.; Morita, Y.; Chen, H.; Shimizu, Y.; Kanai, M. *J. Am. Chem. Soc.* **2018**, *140*, 5899–5903. (e) Ishizawa, K.; Nagai, H.; Shimizu, Y.; Kanai, M. *Chem. Pharm. Bull.* **2017**, *66*, 231–234. (f) Nagai, H.; Morita, Y.; Shimizu, Y.; Kanai, M. *Org. Lett.* **2016**, *18*, 2076–2079. (g) Morita Y.; Yamamoto T.; Nagai H.; Shimizu Y.; Kanai M. *J. Am. Chem. Soc.* **2015**, *137*, 7075–7078. (h) Chen, H.; Yamaguchi, S.; Morita, Y.; Nakao, H.; Zhai, X.-N.; Shimizu, Y.; Mitsunuma, H.; Kanai, M. *Cell Rep. Phys. Sci.* **2021**, *2*, 100679. (i) Hu, H.; Wang, C.; Wu, X.; Liu, Y. Yue, G. Su, G. Feng, J. *Org. Chem. Front.* **2022**, *9*, 1315-1320.

(11) (a) Mack, K. A.; Sigman, M. S. *J. Am. Chem. Soc.* **2023**, *145*, 110–121. (b) Sigman, M. S.; Doyle, A. G. *Science*, **2021**, *374*, 301–308.

(12) Sigman, M. S.; Davies, H. M. L. *J. Am. Chem. Soc.* **2022**, *144*, 15549–15561.

(13) The program used for logistic regression can be found here: https://github.com/DWs-code-drink/SCF3_Logistic.

(14) **General procedure for α-trifluoromethylthiolation of carboxylic acids**
Carboxylic acid (0.1 mmol, 1.0 equiv) was added to a 4 mL vial and then brought into the glovebox. $(AcO)_4B_2O$ (2.7 mg, 0.1 equiv), THF (1 mL), DBU (35.7 μL, 2.4 equiv) and *N*-(trifluoromethylthio)phthalimide (37.1 mg, 1.5 equiv) were sequentially added. The vial was sealed and removed from the glovebox, and the resulting solution was stirred for 2 h at room temperature (23–28 °C). MeI (37.4 μL, 6.0 equiv) was consequently added, and the reaction mixture was left stirred for another 2 h. Then, the reaction was worked up by adding 2 mL of 2:1:1 solution of brine, water and 4M HCl, followed by extraction with EtOAc (3 times). The combined organic layer was dried over Na_2SO_4, filtered, and concentrated in vacuo. To this crude product was added tetrabromoethane (11.6 μL, 0.1 mmol) and hexafluorobenzene (34.6 μL, 0.3 mmol) as the standards for the determination of NMR

yield. The residue was then purified by silica gel chromatography on a Biotage Isolera One using a dry-load method.