Reinforcing the supply chain of COVID-19 therapeutics with expert-coded retrosynthetic software

Yingfu Lin[1†], Zirong Zhang[1†], Babak Mahjour[1], Di Wang[1], Rui Zhang[2], Eunjae Shim[2], Andrew McGrath[1], Yuning Shen[1], Nadia Brugger[3], Rachel Turnbull[3], Shashi Jasty[3], Sarah Trice[3], Tim Cernak[1,2*]

Supply chains become stressed when demand for essential products increases rapidly in times of crisis. This year, the scourge of coronavirus highlighted the fragility of diverse supply chains, affecting the world's pipeline of hand sanitizer,[1] toilet paper,[2] and pharmaceutical starting materials.[3] Many drug repurposing studies are now underway.[4] If a winning therapeutic emerges, it is unlikely that the existing inventory of the medicine, or even the chemical raw materials needed to synthesize it,[5] will be available in the quantities required to satisfy global demand. We show the use of a retrosynthetic artificial intelligence (AI)[6-10] to navigate multiple parallel synthetic sequences, and arrive at plausible alternate reagent supply chains for twelve investigational COVID-19 therapeutics. In many instances, the AI utilizes C–H functionalization logic,[11-13] and we have experimentally validated several syntheses, including a route to the antiviral umifenovir that requires functionalization of six C–H bonds. This general solution to chemical supply chain reinforcement will be useful during global disruptions, such as during a pandemic.

A diverse array of antiviral and anti-inflammatory drugs is currently under clinical investigation in the hopes that an existing medicine can be repurposed for use against COVID-19.[14] Depending on the dosing regimen of a repurposed drug, multiple grams per person may be needed. There are 18.6 million reported SARS-CoV-2 infections at the time of this writing. The current pharmaceutical supply chain is ill-equipped to deliver specific drugs in this quantity. At the time of our study, the only approved drug for treating COVID-19 was remdesivir (**1**). While its producer, Gilead Sciences, Inc., ramped up production of **1** significantly, there were only enough doses of this antiviral to treat 5,000 patients when the outbreak began.[15] New approaches to supply medicines with speed and from alternate sources are urgently needed.

We realized that the availability of alternative synthetic routes to promising therapeutics, which initiate from alternative starting materials, could alleviate pressure on supply chains during a crisis. At the outset of our investigation, more than a hundred clinical trials exploring the repurposing potential of drugs as COVID-19 treatments were underway.[4] It was therefore necessary to consider multiple therapeutic synthetic targets simultaneously, even though many would eventually prove to be irrelevant. Recently, automated retrosynthesis was used to design contingency plans for the investigational COVID-19 therapeutics hydroxychloroquine and remdesivir (**1**).[16] This reaction-centric computational study identified novel sequences, but the routes initiate with known starting materials, or from starting materials that add significantly to route length or reagent cost. We were specifically interested in a starting material-centric analysis to identify distinct raw materials that were of comparable cost and featured in routes of comparable length to the known routes. Importantly, we sought to experimentally validate selected routes. Among small molecules under clinical investigation (Fig. 1), we focused on a diverse portfolio comprised of remdesivir (**1**), bromhexine (**2**), umifenovir (**3**), favipiravir (**4**), ritonavir (**5**), cobicistat (**6**), ribavirin (**7**), camostat (**8**), darunavir (**9**), nelfinavir (**10**), galidesivir (**11**), and baricitinib (**12**). The simultaneous design of multistep preparative routes to twelve diverse targets, which circumvent the use of established raw materials and patented synthetic routes, presents a considerable data handling challenge wherein multiple existing and proposed synthetic sequences must be considered alongside multiple starting materials, whose appropriateness for large-scale production is heavily influenced by price, availability and, in this case, novelty. This data handling challenge could potentially be addressed through the use of modern computer-assisted synthesis planning software.[6-8, 17-23] We enlisted the SYNTHIA™ retrosynthesis platform[7, 19] to facilitate navigation of requisite parameters including availability, pricing, and novelty of starting materials, route brevity, issues of chemo-, regio-, and stereoselectivity, as well as route visualization, documentation, sharing and storage. We present herein predicted retrosynthetic routes to twelve diverse COVID-19 therapeutic candidates that initiate, as frequently as possible, with starting materials that are distinct from those used in published or patented syntheses. We experimentally validate four new routes to **3**, and a one-step synthesis of **2**.

Our study commenced with a crowd-sourcing approach wherein each member of our lab collected all published and patented synthetic routes for one of the drugs in Fig. 1. The routes were then encoded via their simplified molecular-input line-entry system (SMILES) strings. From this dataset we built an interactive route visualizer, available for free at http://covidroutes.cernaklab.com,[24] to facilitate review of existing routes (Fig. 2). The concatenated list of starting material SMILES from each target was used as an exclusion criterion in each retrosynthetic search. This approach allowed us to rapidly navigate to novel starting materials. A typical search returned 50 route proposals, which were generally optimized to minimize starting material cost. In some instances, a close analog of a known starting material that had a distinct SMILES string would appear in the answer set, for instance if a starting material bearing a methyl ester was excluded but the complementary ethyl ester was proposed; in these cases, the search was repeated to also exclude the analog's SMILES string. Predicted routes were manually reviewed for step count, synthetic feasibility, and ease of execution of proposed

reactions on multikilogram scale. A summary of the visualized routes is shown in Figure 2. In most cases the proposed route has the same number, or fewer, steps than the established routes, and initiates from distinct starting materials.
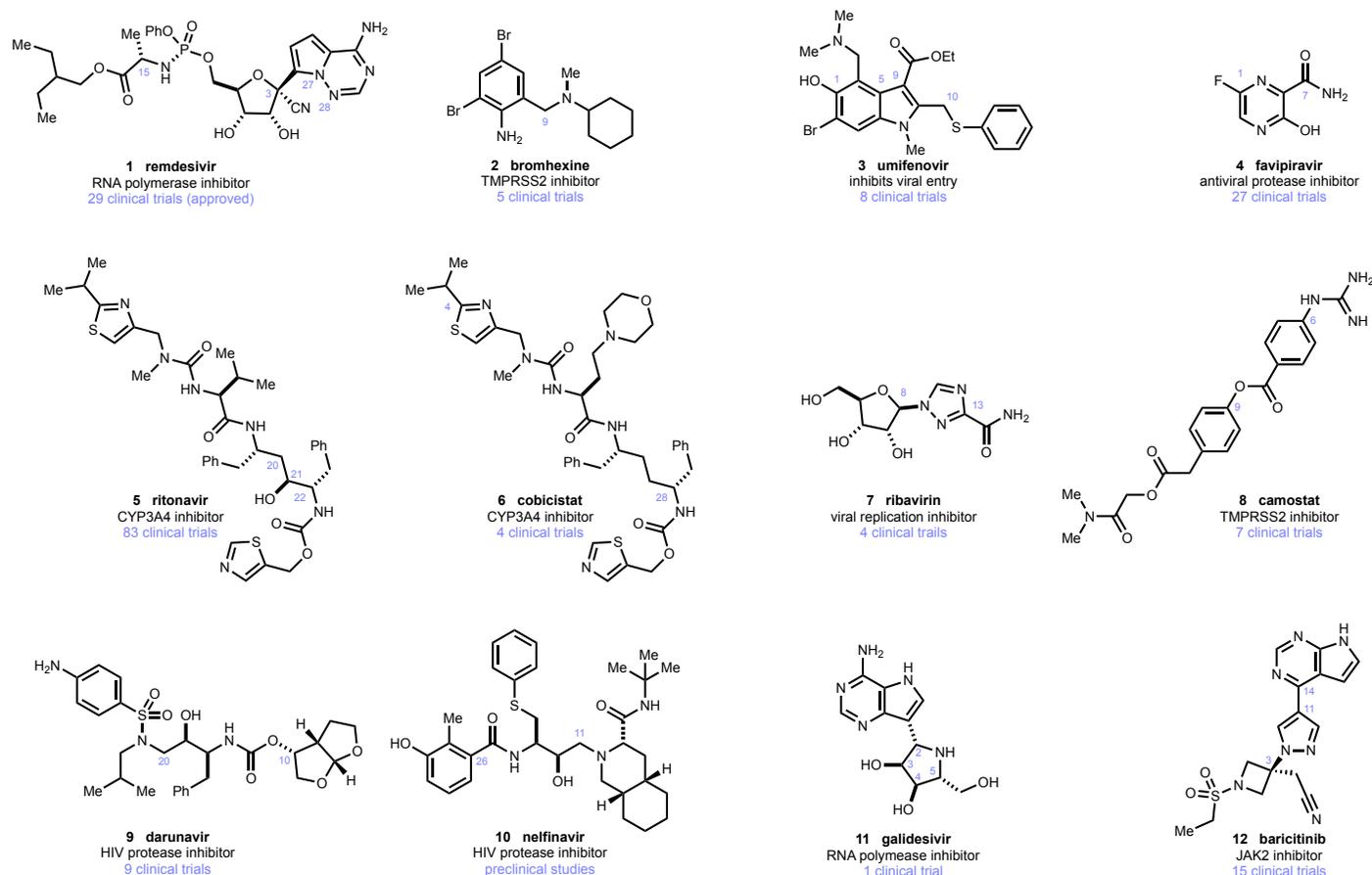
Figure 1. **Investigational COVID-19 therapeutics selected for retrosynthetic analysis.** The number of clinical trials is based on search results for all listed trials, completed, active, or planned, found on www.clinicaltrials.gov.

    Our search for routes to **1** in the SYNTHIA[TM] software yielded a retrosynthetic solution that effectively reproduced a route predicted by the Grzybowski lab,[16] which was very similar to the route published by the drug's maker, Gilead Sciences, Inc.,[25] except for variations in protecting groups. One route we found proposed installation of the C15 stereocenter via amination of a chiral enolate (for details, see Ref. *24*), which is seemingly inefficient given that this moiety is easily installed from (*L*)-alanine. A complementary search in the software's manual mode predicted variations of sequences reliant on the formation of N27–N28 bond through *N*-amination of a pyrrole with *O*-(2,4-dinitrophenyl)-hydroxylamine.[26] Ultimately these proposals were uncompetitive with published routes, in cost and step economy. Given that **1** comprises an esoteric heterocycle, and chiral centers readily available from (*D*)-ribose and (*L*)-alanine, this molecule challenged our ability to identify truly novel starting materials that were competitive with known routes to **1**. This was not the case with the other targets described below.

    The next drug we looked at was bromhexine (**2**), which is currently being investigated in five clinical trials for COVID-19. Most published syntheses of **2** functionalize 2,4-dibromoaniline with formaldehyde, or an equivalent, to install C9, which is primed for *N*-alkylation with *N*-methylcyclohexylamine. A SYNTHIA[TM] search provided alternative routes of comparable length, identifying 2,4-dibromo-6-methylaniline as a novel starting material.[24] The predicted route invoked a C–H oxidation of the benzylic methyl group (C9). This was the first of many instances of proposed C–H functionalization logic.[11-13] We also devised a one-step protocol to couple 2,4,6-tribromoaniline ($0.51/g) with *N,N*-dimethylcyclohexylamine ($0.10/g) via a direct C–H functionalization.[27] This one-step route was suggested by us, rather than the AI, based on a reaction-agnostic search[28] of vendor catalogs for affordable starting materials.

    Another target, umifenovir (**3**) is an antiviral drug developed to combat coronavirus and influenza infections. It is believed to inhibit entry of viruses into human cells, and is currently listed in eight clinical trials as an investigational COVID-19 therapy. Most routes to umifenovir hinge on a Nenitzescu indole coupling[29] between 1,4-benzoquinone and a β-aminocrotonic ester to forge the C5–C9 bond (see Ref. *24*). Indeed, the Nenitzescu reaction using known starting materials featured as a proposal in our search results when default search criteria were used, so the keyword "Nenitzescu" was added

as an exclusion criterion. We then arrived at a series of routes to **3** based on the palladium-catalyzed oxidative cyclization of an aniline with a β-ketoester[30] to install the C5–C9 bond. A key theme that separated the predicted routes from the established palladium-catalyzed indole-forming routes, and enabled the use of distinct starting materials, was the incorporation of a Baeyer-Villiger oxidation to utilize an acetyl group as a surrogate to the requisite hydroxyl group at C1. The proposal of a Baeyer-Villiger oxidation demonstrates that the platform is able to arrive at non-obvious solutions, and this finding was supplemented by the proposal of a C–H oxidation to install the requisite acetyl group from an ethyl group – a proposal which utilizes C–H functionalization logic[11-13] and reduces the cost of the starting material from \$1.15/g to \$0.53/g. In another sequence to **3**, a related indole formation proposed initiating with a halogen coupling handle, instead of a C–H bond, but instead invoked C–H functionalization logic via a Bamberger rearrangement on the hydroxylamine produced from 2,5-dibromonitrobenzene (\$1.73/g). As described below, all four routes were reduced to experimental practice with only minor modifications to reaction conditions and sequences proposed by the software.
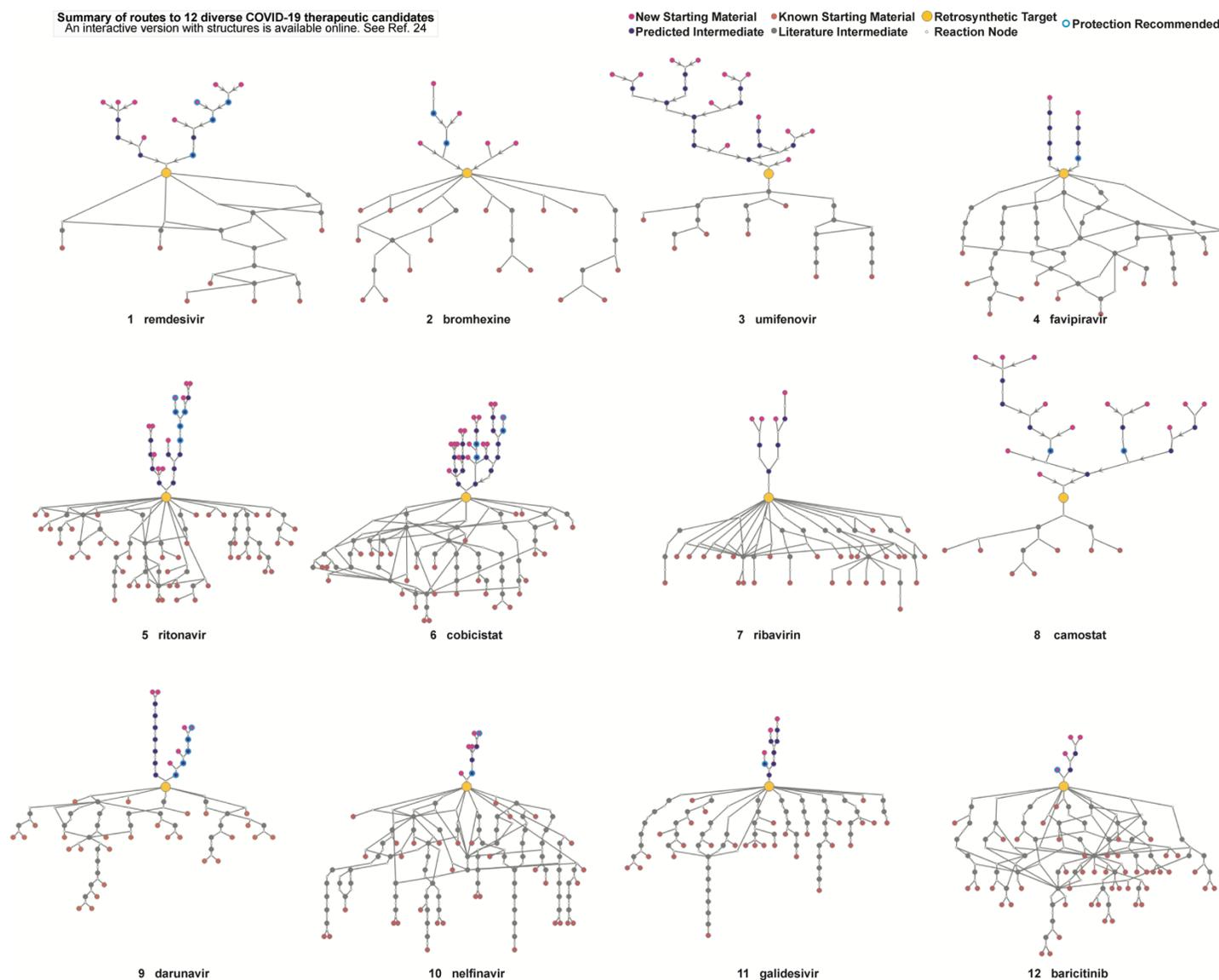
Figure 2. **Network of published and predicted synthetic sequences.** The yellow dot at center is the target molecule (see structures in Fig. 1). Routes in grey and orange below the target are published, and the routes in purple and pink above the target are routes predicted by the AI, with specific routes shown here selected by humans. Crossed lines in literature routes show shared intermediates or starting materials. An interactive version of this diagram, where hovering a mouse cursor over each dot pops up the corresponding molecular structure, is available as an interactive route visualizer.

The remaining targets, **4–12,** represent a diverse cross-section of the small molecule chemical matter under investigation for COVID-19. Some of the key disconnections identified are highlighted with atom numbering in Fig. 2. These route proposals were not experimentally investigated in the current study, but, nonetheless, analysis of these targets yielded multiple starting material proposals that can relieve the fine chemical supply chain. Routes can be reviewed online with an interactive route visualizer (Ref. *24*).

Figure 3. **Experimental demonstration of four predicted routes to 3.** Each route terminates in inexpensive starting materials. **13**, **25**, **29**, **32** and **33** have not been used in a published synthesis of **3**. Frequent use of C–H functionalization logic keeps costs low and adds brevity.
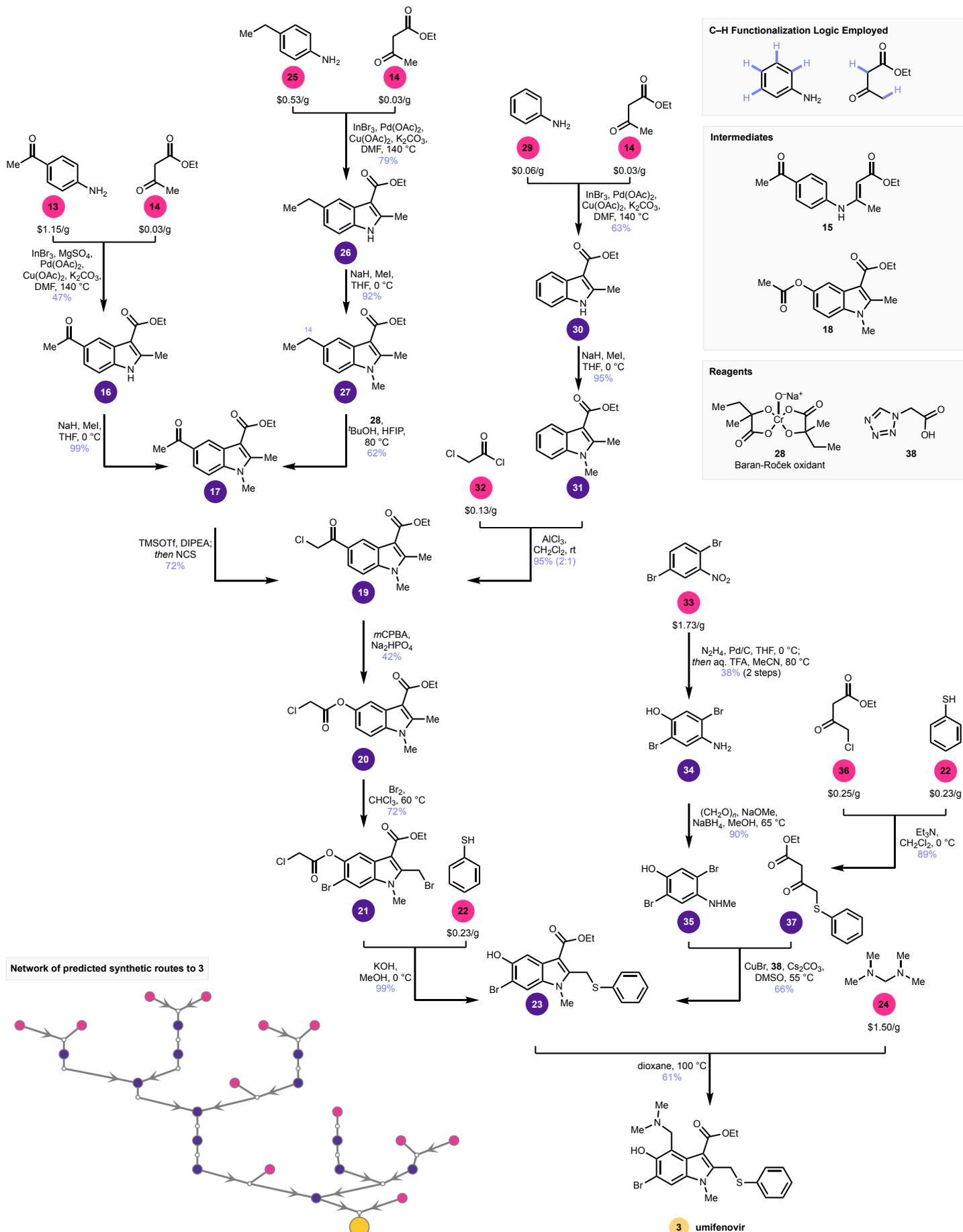
With the routes designed, we initiated experimental syntheses. For **3** (Fig. 3), we first investigated the proposed indole formation[30] from 1-(4-aminophenyl)ethan-1-one (**13**, $1.15/g) and ethyl acetoacetate (**14**, $0.03/g) using oxidative reaction conditions. Pretreatment of **13** and **14** with 1 mol% indium(III) bromide, to form **15** was followed by treatment with palladium(II) acetate and copper(II) acetate to form **16**. While the suggested reaction conditions did provide the desired indole **16**, the yield was only 20%. Using magnesium sulfate to promote the formation of **15** improved the yield of **16** to 47%. *N*-methylation occurred smoothly to produce **17** in 99% yield. An issue was encountered in the experimental realization of the Baeyer-Villiger oxidation using *m*CPBA in that a mixture of oxidation products was obtained. Unstable products we believe to be from oxidation of the indole's double bond accounted for the bulk of the reacted material, and only traces of **18** were isolated. While the AI proposal did accurately predict the formation of **18**, the subtle interplay of electronics that govern the preference for the undesired Prilezhaev oxidation verses the desired Baeyer-Villiger oxidation could not be teased out by the software, and the best modification of reaction conditions we found yielded small amounts of **18** as a mixture with undesired oxidation products. A manual literature search on related indoles revealed that the α-chloroketone **19** should be a viable substrate for Baeyer-Villiger,[31] with the chloro-group acting as a directing group to favor oxidation of the ketone. Indeed, chlorination of **17** by treating first with TMSOTf and diisopropylethylamine and then with *N*-chlorosuccinimide led to **19**, which underwent selective Baeyer-Villiger oxidation to produce **20**. Subsequent bromination produced **21**, which underwent thioetherification with **22** and *in situ* saponification to produce **23**. Here, the route intercepts known syntheses of **3** via alkylation with **24**.[29] Thus, a novel and viable route to **3** was predicted, from starting materials distinct from those existing in the literature. All intermediates predicted by the software were observed, but a modification to incorporate a chlorine directing group was necessary to achieve usable levels of selectivity in the formation of **20**.

The output of a SYNTHIA[TM] search is not a single route proposal, but a ranked list of route proposals. Several other proposed routes to **3** were also experimentally realized. One route, based on a variation of the same indole-formation and Baeyer-Villiger sequence described above, proposed another form of C–H oxidation logic via a benzylic oxidation of ethyl indole **27**. The Glorius indole-synthesis was more productive with **25** than with **13**, yielding **26** in 79% yield, under the exact reaction conditions proposed, likely owing to the increased nucleophilicity of the aniline. Methylation gave **27** in 92% yield. AI-recommended conditions of Oxone® with potassium bromide[32] for the C–H oxidation of **27** were unsuccessful in our hands. An extensive survey of oxidants revealed the recently disclosed Baran-Roček oxidation[33-35] as the solely productive condition we could identify to oxidize C14. Treatment of **27** with commercially available **28** gave **17** in 62% yield, thus intercepting the previous route to **3**. From these studies, it occurred to us that the direct installation of the chloromethyl ketone via a Friedel-Crafts acylation would provide a concise and alternative route to **3**. Indeed, the software had proposed a lower-scoring route that used a Friedel-Crafts acylation. This route was intriguing in that it disconnected the synthesis of **3** back to aniline (**29**), an exceptionally cheap starting material. While the AI proposed a Friedel-Crafts acylation with acetyl chloride, we modified the route to use instead chloroacetyl chloride (**32**, $0.13/g), thus installing the chloride directing group in a single step. In the event, oxidative indole-coupling to form **30** and methylation to form **31** occurred smoothly. Friedel-Crafts acylation of **31** with **32** under influence of aluminum(III) chloride gave **19**, as a 2:1 mixture of regioisomers, intercepting the other routes. The use of C–H functionalization logic by the AI was a recurring theme. With this last route, the software proposed functionalization of six C–H bonds to convert **14** and **29** into **3**.

As a final demonstration of the ability to predict viable routes employing novel starting materials, we demonstrated yet another form of C–H functionalization logic via a Bamberger rearrangement to install the C1–OH. In the event, **33** was partially reduced to the hydroxylamine, and treated with aqueous trifluoroacetic acid to affect the Bamberger rearrangement yielding **34**, and then methylated to arrive at **35**. Copper-catalyzed coupling to **37**, itself obtained through the union of **36** and **22**, produced **23** in 66% yield when **38** was used as a ligand. Subsequent alkylation of **23** with **24** produced **3**.

Finally, we explored the synthesis of bromhexine (**2**, Fig. 4). We found that **39** could be heated in excess **40** with 2 equivalents of *tert*-butylperoxide[27] to produce **2** in 41% yield. While this modest yield, and the use of such a large excess of **40**, are clearly unacceptable for a large-scale production route, further optimization of reaction conditions could be performed. Care must be taken in optimization studies based on peroxide activation, as we observed a violent reaction on a small-scale between neat **40** and benzoylperoxide.
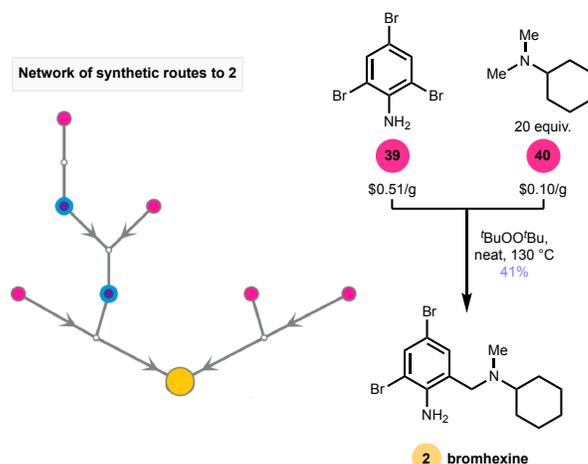
Figure 4. One-step synthesis of bromhexine, the right-side branch of the synthetic network visualization, realized through reaction of **39** with **40** under influence of *tert*-butylperoxide.

The automation of retrosynthesis enabled us to analyze a dozen targets simultaneously from a combinatorial explosion of conceivable predicted reaction sequences and starting material availabilities. We anticipate the workflow described herein may be useful in times of crisis.

AUTHOR INFORMATION
[†] These authors contributed equally. *Corresponding author: tcernak@med.umich.edu
1. Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI, 48109, USA
2. Department of Chemistry, University of Michigan, Ann Arbor, MI, 48109, USA
3. MilliporeSigma, Burlington, MA, 01803, USA

REFERENCES
1. *Food and Drug Administration: Temporary Policy for Preparation of Certain Alcohol-Based Hand Sanitizer Products During the Public Health Emergency (COVID-19) Guidance for Industry*. 2020 06/01/2020; Available from: https://www.fda.gov/media/136289/download.
2. Garbe, L., R. Rau, and T. Toppe, Influence of perceived threat of Covid-19 and HEXACO personality traits on toilet paper stockpiling. *PLOS ONE*, **15**, e0234232 (2020).
3. Ledford, H., Dozens of coronavirus drugs are in development - what happens next? *Nature*, **581**, 247-248 (2020).
4. Guy, R.K., R.S. DiPaola, F. Romanelli, and R.E. Dutch, Rapid repurposing of drugs for COVID-19. *Science*, **368**, 829 (2020).
5. Hardy, M.A., B.A. Wright, J.L. Bachman, T.B. Boit, H.M.S. Haley, R.R. Knapp, R.F. Lusi, T. Okada, V. Tona, N.K. Garg, and R. Sarpong, Treating a Global Health Crisis with a Dose of Synthetic Chemistry. *ACS Central Science*, (2020).
6. Corey, E.J. and W.T. Wipke, Computer-assisted design of complex organic syntheses. *Science*, **166**, 178-92 (1969).
7. Szymkuc, S., E.P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, and B.A. Grzybowski, Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew Chem Int Ed Engl*, **55**, 5904-37 (2016).
8. Segler, M.H.S., M. Preuss, and M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, **555**, 604-610 (2018).
9. Coley, C.W., W.H. Green, and K.F. Jensen, Machine Learning in Computer-Aided Synthesis Planning. *Accounts of Chemical Research*, **51**, 1281-1289 (2018).
10. Lee, A.A., Q. Yang, V. Sresht, P. Bolgar, X. Hou, J.L. Klug-McLeod, and C.R. Butler, Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chemical Communications*, **55**, 12152-12155 (2019).
11. Newhouse, T. and P.S. Baran, If C–H Bonds Could Talk: Selective C H Bond Oxidation. *Angewandte Chemie International Edition*, **50**, 3362-3374 (2011).
12. Cernak, T., K.D. Dykstra, S. Tyagarajan, P. Vachal, and S.W. Krska, The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem Soc Rev*, **45**, 546-76 (2016).
13. Davies, H.M.L. and D. Morton, Collective Approach to Advancing C-H Functionalization. *ACS Cent Sci*, **3**, 936-943 (2017).
14. Fragkou, P.C., D. Belhadi, N. Peiffer-Smadja, C.D. Moschopoulos, F.X. Lescure, H. Janocha, E. Karofylakis, Y. Yazdanpanah, F. Mentré, C. Skevaki, C. Laouénan, and S. Tsiodras, Review of trials currently testing treatment and prevention of COVID-19. *Clinical Microbiology and Infection*, (2020).
15. Gilead Sciences, I., Working to Supply Remdesivir for COVID-19. (2020).
16. Szymkuć, S., E.P. Gajewska, K. Molga, A. Wołos, R. Roszak, W. Beker, B. Moskal, P. Dittwald, and B.A. Grzybowski, Computer-generated "synthetic contingency" plans at times of logistics and supply problems: scenarios for hydroxychloroquine and remdesivir. *Chemical Science*, (2020).
17. Coley, C.W., D.A. Thomas, J.A.M. Lummiss, J.N. Jaworski, C.P. Breen, V. Schultz, T. Hart, J.S. Fishman, L. Rogers, H. Gao, R.W. Hicklin, P.P. Plehiers, J. Byington, J.S. Piotti, W.H. Green, A.J. Hart, T.F. Jamison, and K.F. Jensen, A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, **365**, eaax1566 (2019).
18. Badowski, T., E.P. Gajewska, K. Molga, and B.A. Grzybowski, Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning. *Angew Chem Int Ed Engl*, **59**, 725-730 (2020).
19. Klucznik, T., B. Mikulak-Klucznik, M.P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E.P. Gajewska, A. Toutchkine, P. Dittwald, M.P. Startek, G.J. Kirkovits, R. Roszak, A. Adamski, B. Sieredzińska, M. Mrksich, S.L.J. Trice, and B.A. Grzybowski, Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem*, **4**, 522-532 (2018).
20. Cernak, T., A machine with chemical intuition. *Chem*, **4**, 401-403 (2018).
21. Cadeddu, A., E.K. Wylie, J. Jurczak, M. Wampler-Doty, and B.A. Grzybowski, Organic chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses. *Angew Chem Int Ed Engl*, **53**, 8108-12 (2014).
22. Kowalik, M., C.M. Gothard, A.M. Drews, N.A. Gothard, A. Weckiewicz, P.E. Fuller, B.A. Grzybowski, and K.J. Bishop, Parallel optimization of synthetic pathways within the network of organic chemistry. *Angew Chem Int Ed Engl*, **51**, 7928-32 (2012).
23. Molga, K., P. Dittwald, and B.A. Grzybowski, Navigating around Patented Routes by Preserving Specific Motifs along Computer-Planned Retrosynthetic Pathways. *Chem*, **5**, 460-473 (2019).
24. http://covidroutes.cernaklab.com
25. Siegel, D., H.C. Hui, E. Doerffler, M.O. Clarke, K. Chun, L. Zhang, S. Neville, E. Carra, W. Lew, B. Ross, Q. Wang, L. Wolfe, R. Jordan, V. Soloveva, J. Knox, J. Perry, M. Perron, K.M. Stray, O. Barauskas, J.Y. Feng, Y. Xu, G. Lee, A.L. Rheingold, A.S. Ray, R. Bannister, R. Strickley, S. Swaminathan, W.A. Lee, S. Bavari, T. Cihlar, M.K. Lo, T.K. Warren, and R.L. Mackman, Discovery and Synthesis of a Phosphoramidate Prodrug of a Pyrrolo[2,1-f][triazin-4-amino] Adenine C-Nucleoside (GS-5734) for the Treatment of Ebola and Emerging Viruses. *Journal of Medicinal Chemistry*, **60**, 1648-1661 (2017).
26. Hynes, J., W.W. Doubleday, A.J. Dyckman, J.D. Godfrey, J.A. Grosso, S. Kiau, and K. Leftheris, N-Amination of Pyrrole and Indole Heterocycles with Monochloramine (NH2Cl). *The Journal of Organic Chemistry*, **69**, 1368-1371 (2004).
27. Ueno, R., Y. Ikeda, and E. Shirakawa, tert-Butoxy-Radical-Promoted α-Arylation of Alkylamines with Aryl Halides. *European Journal of Organic Chemistry*, **2017**, 4188-4193 (2017).
28. Mahjour, B., Y. Shen, W. Liu, and T. Cernak, A map of the amine–carboxylic acid coupling system. *Nature*, **580**, 71-75 (2020).

29. Balakin, K.V., R. Filosa, S.N. Lavrenov, A.S. Mkrtchayn, M.B. Nawrozkij, and I.A. Novakov, Arbidol: a quarter-century after. Past, present and future of the original Russian antiviral. *Russian Chemical Reviews*, **87**, 509-552 (2018).

30. Würtz, S., S. Rakshit, J.J. Neumann, T. Dröge, and F. Glorius, Palladium-Catalyzed Oxidative Cyclization of N-Aryl Enamines: From Anilines to Indoles. *Angewandte Chemie International Edition*, **47**, 7230-7233 (2008).

31. Nakatsuka, S.-i.A., Osamu; Ueda, Kazuo; Goto, Toshio, Introduction of a hydroxy group onto 5- and 6-position of indole nucleus by Friedel-Crafts acylation and subsequent Baeyer-Villiger Oxidation. *Heterocycles*, **26**, 1471–1474 (1987).

32. Moriyama, K., M. Takemura, and H. Togo, Direct and Selective Benzylic Oxidation of Alkylarenes via C–H Abstraction Using Alkali Metal Bromides. *Organic Letters*, **14**, 2414-2417 (2012).

33. Kanda, Y., H. Nakamura, S. Umemiya, R.K. Puthukanoori, V.R. Murthy Appala, G.K. Gaddamanugu, B.R. Paraselli, and P.S. Baran, Two-Phase Synthesis of Taxol. *Journal of the American Chemical Society*, **142**, 10526-10533 (2020).

34. Wilde, N.C., M. Isomura, A. Mendoza, and P.S. Baran, Two-Phase Synthesis of (−)-Taxuyunnanine D. *Journal of the American Chemical Society*, **136**, 4909-4912 (2014).

35. Krumpolc, M. and J. Rocek, Synthesis of stable chromium(V) complexes of tertiary hydroxy acids. *Journal of the American Chemical Society*, **101**, 3206-3209 (1979).

Supporting Information for
"Reinforcing the Supply Chain of COVID-19 Therapeutics
with Expert-Coded Retrosynthetic Software"

General Information

All reactions were conducted in oven- or flame-dried glassware under an atmosphere of nitrogen unless stated otherwise. Reactions were set up in an MBraun LABmaster Pro Glove Box (H_2O level <0.1 ppm, O_2 level <0.1 ppm), or using standard Schlenk technique with a glass vacuum manifold connected to an inlet of dry nitrogen gas. Solvents (acetonitrile, tetrahydrofuran, dichloromethane) were purified using a MBraun SPS solvent purification system, by purging with nitrogen, and then passing the solvent through a column of activated alumina. *t*BuOH, 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) were used as received. Methanol, 1,4-dioxane, dimethyl sulfoxide (DMSO), *N,N*-dimethylformamide (DMF) were purchased as the anhydrous solvents and used as received. Reagents [4'-aminoacetophenone, ethyl acetoacetate, $InBr_3$, $MgSO_4$, $Pd(OAc)_2$, $Cu(OAc)_2$, K_2CO_3, NaH, MeI, trimethylsilyl trifluoromethanesulfonate (TMSOTf), *N,N*-diisopropylethylamine (DIPEA), *N*-chlorosuccinimide (NCS), 3-chloroperbenzoic acid (*m*CPBA), Na_2HPO_4, bromine, KOH, thiophenol, 4-ethylaniline, aniline, sodium bis(2-hydroxy-2-methylbutyrato)oxochromate(V), chloroacetyl chloride, $AlCl_3$, hydrazine monohydrate, trifluoroacetic acid (TFA), palladium on activated carbon, paraformaldehyde, $NaBH_4$, CuI, Cs_2CO_3, triethylamine (Et_3N), 2,4,6-tribromoaniline, *N,N*-dimethylcyclohexanamine, di-*tert*-butyl peroxide, 2,5-dibromonitrobenzene, ethyl 4-chloroacetoacetate, *N,N,N',N'*-tetramethyldiaminomethane] were purchased from Sigma Aldrich, Alfa Aesar, Oakwood Chemical, or TCI Chemical. All chemicals were used as received. Glass 2 dram vials (ChemGlass #CG-4912-02) were used as reaction vessels, fitted with a screw-cap with a Teflon-coated silicone septa (CG-4910-02), and magnetic stir bars (Fisher Scientific #14-513-93 or #14-513-65). High temperature reactions were performed in crimp cap vials (Biotage #351521).

Proton nuclear magnetic resonance spectra (^{1}H NMR) were recorded on a Varian MR-500 MHz or Varian MR-400 MHz spectrometer and chemical shifts are reported in parts per million (ppm) using the solvent residual peak as an internal standard ($CDCl_3$ at 7.26 ppm, $(CD_3)_2SO$ at 2.50 ppm, CD_3CN at 1.94 ppm). Data are reported using the abbreviations: app = apparent, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, comp = complex, br = broad. Coupling constant(s) are reported in Hz. Proton-decoupled carbon nuclear magnetic resonance spectra (^{13}C NMR) spectra were recorded on a Varian MR-500 MHz or Varian MR-400 MHz spectrometer and chemical shifts are reported in ppm using the solvent as an internal standard ($CDCl_3$ at 77.16 ppm, $(CD_3)_2SO$ at 39.52 ppm, CD_3CN at 1.32 ppm). High resolution mass spectrometry data (HRMS) was obtained on a Micromass AutoSpec Ultima Magnetic Sector instrument. Reaction analysis was typically performed by thin-layer chromatography on silica gel, or using a Waters I-class ACQUITY UPLC-MS (Waters Corporation, Milford, MA, USA) equipped with in-line photodiode array detector (PDA) and QDa mass detector (ESI positive ionization mode). 0.1 μL sample injections were taken from acetonitrile solutions of reaction mixtures or products (~1 mg/mL). A partial loop injection mode was used with the needle placement at 1.0 mm from bottom of the wells and a 0.2 μL air gap at pre-

aspiration and post-aspiration. Column used: Waters Cortecs UPLC C18+ column, 2.1mm × 50 mm with (Waters #186007114) with Waters Cortecs UPLC C18+ VanGuard Pre-column 2.1mm × 5 mm (Waters #186007125), Mobile Phase A: 0.1 % formic acid in Optima LC/MS-grade water, Mobile Phase B: 0.1% formic acid in Optima LC/MS-grade MeCN. Flow rate: 1 mL/min. Column temperature: 45 °C. The PDA sampling rate was 20 points/sec. The QDa detector monitored m/z 150-750 with a scan time of 0.06 seconds and a cone voltage of 30 V. The PDA detector range was between 210 nm – 400 nm with a resolution of 1.2 nm. 1 minute and 2 minute methods were used. The method gradients are below: 0 min: 0.8 mL/min, 95% 0.1% formic acid in water/5% 0.1% formic acid in acetonitrile; 1.5 min : 0.8 mL/min, 0.1% 0.1% formic acid in water/99.9% 0.1% formic acid in acetonitrile; 1.91 min : 0.8 mL/min, 95% 0.1% formic acid in water/5% 0.1% formic acid in acetonitrile.

Flash chromatography was performed on silica gel (230 – 400 Mesh, Grade 60) under a positive pressure of Nitrogen. Thin Layer Chromatography was performed on 25 μm TLC Silica gel 60 F_{254} glass plates purchased from Fisher Scientific (part number: S07876). Visualization was performed using ultraviolet light (254 nm), potassium permanganate ($KMnO_4$) stain, or Cerium Ammonium Molybdate (CAM) stain.

Experimental
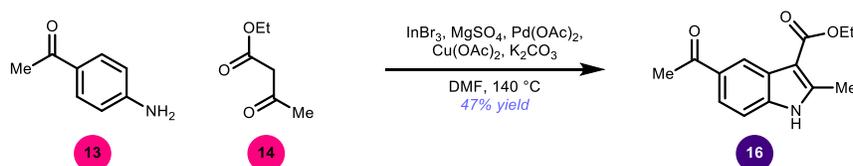
ethyl 5-acetyl-2-methyl-1*H*-indole-3-carboxylate (16).

Following a modified literature procedure,[1] to an oven-dried 2-dram vial with a Teflon-coated stir bar was added 4'-aminoacetophenone (**13**) (1.081 g, 8.00 mmol, 1.00 equiv), ethyl acetoacetate (1.518 mL, 12.00 mmol, 1.50 equiv) and MgSO₄ (0.963 g, 8.00 mmol, 1 equiv). Then the vial was moved into the glovebox, and InBr₃ (28.4 mg, 0.08 mmol, 1 mol%) was added. Then the vial was capped and moved out of the glove box. The reaction mixture was stirred at 80 °C for 2 h. The mixture was transferred with a syringe into a 250 mL round bottom flask containing Pd(OAc)₂ (89.8 mg, 0.40 mmol, 5 mol%), Cu(OAc)₂ (4.359 g, 24.00 mmol, 3.0 equiv), K₂CO₃ (3.317 g, 24.00 mmol, 3.0 equiv), MgSO₄ (0.963 g, 8.00 mmol, 1 equiv) and DMF (70 mL), and the product residues were rinsed into the bigger flask with DMF (2 × 5 mL). The reaction mixture was treated at 140 °C under N₂ atmosphere for 1 h before it was cooled to room temperature, diluted with EtOAc (100 mL) and filtered through a short pad of silica and sea sand. The red-brown solid was washed with EtOAc (2 × 80 mL) and the combined filtrates were concentrated *in vacuo* to yield a crude product, which was purified by flash column chromatography (silica gel, 5% EtOAc/CH₂Cl₂) to yield the product (0.915 g, 3.73 mmol, 47%)

¹H NMR (500 MHz, (CD₃)₂SO) δ 12.12 (s, 1H), 8.59 (d, *J* = 1.7 Hz, 1H), 7.76 (dd, *J* = 8.5, 1.4 Hz, 1H), 7.43 (d, *J* = 8.5 Hz, 1H), 4.30 (q, *J* = 7.1 Hz, 2H), 2.67 (s, 3H), 2.60 (s, 3H), 1.37 (t, *J* = 7.1 Hz, 3H). ¹³C NMR (125 MHz, (CD₃)₂SO) δ 197.41, 164.68, 146.43, 137.50, 130.43, 126.42, 121.97, 121.91, 111.13, 103.96, 59.04, 26.56, 14.37, 13.72. HRMS (ESI): calculated C₁₄H₁₆NO₃ [M+H]⁺: 246.1125, found: 246.1121.
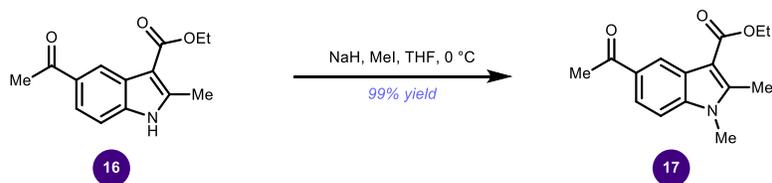
ethyl 5-acetyl-1,2-dimethyl-1*H*-indole-3-carboxylate (17):

To a solution of compound **16** (0.217 g, 0.885 mmol, 1.00 equiv) in anhydrous THF (8.8 mL) was added NaH (60 % dispersion in mineral oil, 42.5 mg, 1.062 mmol, 1.2 equiv) at 0 °C. After 30 min, iodomethane (66.1 µL, 1.062 mmol, 1.2 equiv) was added dropwise at 0 °C. The mixture was stirred at the same temperature for 30 min before it was quenched with water (0.2 mL). After removal of the solvents on a rotary evaporator, the residue was purified by flash column chromatography (silica gel, 2% EtOAc/CH_2Cl_2) to yield the product (0.228 g, 0.879 mmol, 99%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.79 (d, J = 1.8 Hz, 1H), 7.92 (dd, J = 8.7, 1.7 Hz, 1H), 7.33 (d, J = 8.6 Hz, 1H), 4.43 (q, J = 7.1 Hz, 2H), 3.74 (s, 3H), 2.79 (s, 3H), 2.69 (s, 3H), 1.48 (t, J = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 198.16, 165.32, 146.54, 138.72, 131.00, 125.88, 123.29, 121.78, 108.87, 104.92, 59.52, 29.59, 26.45, 14.44, 11.74. HRMS (ESI): calculated $C_{15}H_{18}NO_3$ [M+H]$^+$: 260.1281, found: 260.1278.
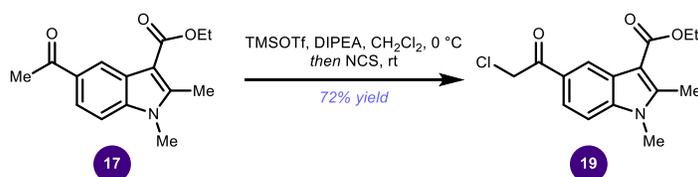
ethyl 5-(2-chloroacetyl)-1,2-dimethyl-1H-indole-3-carboxylate (19):

Following a modified literature procedure,[2] to a solution of compound **17** (25.9 mg, 0.10 mmol, 1.00 equiv.) in dichloromethane (1 mL) was added 4Å molecular sieves (10.0 mg) and N,N-diisopropylethylamine (52.3 µL, 0.30 mmol, 3.00 equiv) at room temperature. The mixture was cooled to 0 °C, then TMSOTf (54.2 µL, 0.30 mmol, 3.00 equiv) was added and stirred for 2 h. The reaction mixture was allowed to warm to room temperature and N-chlorosuccinimide (NCS) (20.0 mg, 0.15 mmol, 1.50 equiv) was added. The reaction was stirred for an additional 2 h, then quenched with saturated aqueous NaHCO$_3$ at room temperature. The crude mixture was extracted with EtOAc (2 × 3 mL). The organic layers were combined, washed with brine, and dried over anhydrous sodium sulfate. The organic layer was then concentrated under reduced pressure and purified by flash column chromatography (silica gel, 2% – 5% EtOAc/CH_2Cl_2) to afford the product (21.1 mg, 0.072 mmol, 72%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.78 (d, J = 1.7 Hz, 1H), 7.92 (dd, J = 8.7, 1.8 Hz, 1H), 7.36 (d, J = 8.7 Hz, 1H), 4.83 (s, 2H), 4.43 (q, J = 7.1 Hz, 2H), 3.75 (s, 3H),

2.80 (s, 3H), 1.49 (t, J = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 191.14, 165.43, 147.18, 139.38, 128.18, 126.22, 123.51, 122.51, 109.60, 105.44, 59.88, 46.36, 29.99, 14.62, 12.06. HRMS (ESI): calculated C$_{15}$H$_{17}$ClNO$_3$ [M+H]$^+$: ^{35}Cl 294.0891, found: 294.0895.
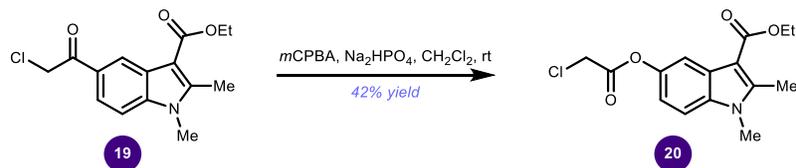
ethyl 5-(2-chloroacetoxy)-1,2-dimethyl-1*H*-indole-3-carboxylate (20):

Following a modified literature procedure,[3] *m*CPBA (70%, 37.0 mg, 0.15 mmol, 3.00 equiv) was added to a mixture of compound **19** (14.7 mg, 0.05 mmol, 1.00 equiv) and Na$_2$HPO$_4$ (21.3 mg, 0.15 mmol, 3.00 equiv) in dichloromethane (0.5 mL) at room temperature. The reaction mixture was stirred at the same temperature for 1 h before quenched with saturated aqueous sodium thiosulfate (1 mL) and saturated aqueous sodium bicarbonate (1 mL). The crude mixture was extracted with EtOAc (2 × 3 mL). The organic layers were combined, washed with brine, and dried over anhydrous sodium sulfate. The organic layer was then concentrated under reduced pressure and purified by flash column chromatography (silica gel, 17% EtOAc/heaxnes) to afford the product (6.5 mg, 0.021 mmol, 42%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.84 (d, J = 2.4 Hz, 1H), 7.27 (d, J = 9.0 Hz, 1H), 6.99 (dd, J = 8.7, 2.3 Hz, 1H), 4.39 (q, J = 7.1 Hz, 2H), 4.35 (s, 2H), 3.70 (s, 3H), 2.77 (s, 3H), 1.43 (t, J = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 166.76, 165.80, 146.69, 145.62, 134.60, 127.12, 115.50, 113.56, 109.69, 104.35, 59.62, 41.17, 29.86, 14.74, 12.05. HRMS (ESI): calculated C$_{15}$H$_{17}$ClNO$_4$ [M+H]$^+$: ^{35}Cl 310.0841, found: 310.0778.
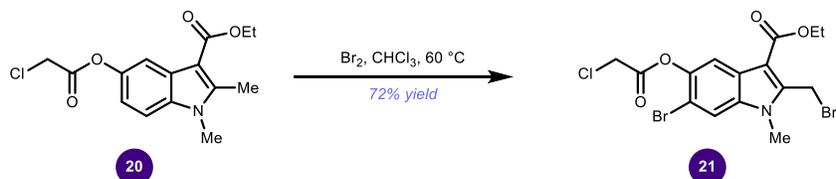
ethyl   6-bromo-2-(bromomethyl)-5-(2-chloroacetoxy)-1-methyl-1*H*-indole-3-

carboxylate (21):

Following a modified literature procedure,[4] to a solution of compound **20** (12.0 mg, 0.039 mmol, 1.00 equiv) in chloroform (0.4 mL) was added bromine (4.0 µL, 0.078 mmol, 2.00 equiv) dropwise at room temperature. After addition, the mixture was treated at 60 °C for 2 h before quenched with saturated aqueous sodium thiosulfate (1 mL) and saturated aqueous sodium bicarbonate (1 mL). The crude mixture was extracted with EtOAc (2 × 3 mL). The organic layers were combined, washed with brine, and dried over anhydrous sodium sulfate. The organic layer was then concentrated under reduced pressure and purified by flash column chromatography (silica gel, 15% EtOAc/heaxnes) to afford the product (13.0 mg, 0.028 mmol, 72%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.93 (s, 1H), 7.61 (s, 1H), 5.12 (s, 2H), 4.48 – 4.38 (m, 4H), 3.79 (s, 3H), 1.45 (t, J = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 165.94, 164.43, 143.06, 142.98, 135.96, 125.86, 116.34, 114.28, 111.53, 106.09, 60.48, 40.94, 30.32, 20.57, 14.65. HRMS (ESI): calculated C$_{15}$H$_{14}$Br$_2$ClNO$_4$Na [M+Na]$^+$: ^{35}Cl/^{79}Br/^{79}Br 487.8870, found: 487.8868.
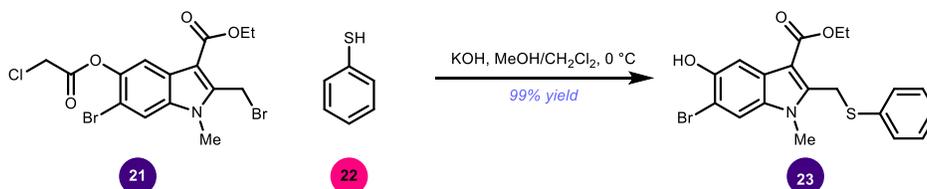
ethyl 6-bromo-5-hydroxy-1-methyl-2-((phenylthio)methyl)-1*H*-indole-3-carboxylate (23):

Following a modified literature procedure,[4] to a solution of potassium hydroxide (85%, 5.1 mg, 0.077 mmol, 3.00 equiv) in methanol (0.1 mL) was added thiophenol (5.3 µL, 0.051 mmol, 2.00 equiv) and the mixture was stirred at room temperature for 15 min before it was cooled in an ice bath. To the above mixture was added a solution of compound **21** (12.0 mg, 0.026 mmol, 1.00 equiv) in dichloromethane (0.25 mL) at 0 °C. The resulting mixture was stirred at the same temperature for 1 h before it was quenched with acetic acid (8 µL). The solvent was removed under reduced pressure and the residue was purified by flash column chromatography (silica gel, 20% EtOAc/hexanes) to yield the product (10.7 mg, 0.026 mmol, 99%).

The characterization data matched reported spectral values.
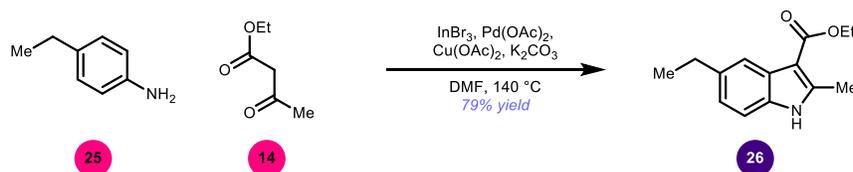
ethyl 5-ethyl-2-methyl-1*H*-indole-3-carboxylate (26):

Following a modified literature procedure,[1] to an oven-dried 2-dram vial with a Teflon-coated stir bar was added 4-ethylaniline (**25**) (0.621 mL, 5.00 mmol, 1.00 equiv), and ethyl acetoacetate (0.696 mL, 5.50 mmol, 1.10 equiv). Then the vial was moved into the glovebox, and InBr$_3$ (17.7 mg, 0.05 mmol, 1 mol%) was added. Then the vial was capped and moved out of the glove box. The reaction mixture was stirred at room temperature for 30 min. The mixture was transferred with a syringe into a 250 mL round bottom flask containing Pd(OAc)$_2$ (56.1 mg, 0.25 mmol, 5 mol%), Cu(OAc)$_2$ (2.724 g, 15.00 mmol, 3.0 equiv), K$_2$CO$_3$ (2.073 g, 15.00 mmol, 3.0 equiv), and DMF (40 mL), and the product residues were rinsed into the bigger flask with DMF (2 × 5 mL). The reaction mixture was treated at 140 °C under N$_2$ atmosphere for 30 min before it was cooled to room temperature, diluted with EtOAc (100 mL) and filtered through a short pad of silica and sea sand. The red-brown solid was washed with EtOAc (2 × 80 mL) and the combined filtrates were concentrated on a rotary evaporator and in high vacuum to yield a crude product, which was purified by flash column chromatography (silica gel, 20% EtOAc/hexanes) to yield the product (0.916 g, 3.96 mmol, 79%)

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.25 (br s, 1H), 7.94 (s, 1H), 7.21 (d, *J* = 8.2 Hz, 1H), 7.05 (d, *J* = 8.2 Hz, 1H), 4.40 (q, *J* = 7.1 Hz, 2H), 2.76 (q, *J* = 7.6 Hz, 2H), 2.72 (s, 3H), 1.45 (t, *J* = 7.1 Hz, 3H), 1.29 (t, *J* = 7.6 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 166.45, 144.10, 137.91, 133.10, 127.62, 122.83, 119.98, 110.41, 104.34, 59.59, 29.31, 16.53, 14.74, 14.43. HRMS (ESI): calculated C$_{14}$H$_{17}$NNaO$_2$ [M+Na]$^+$: 254.1151, found: 254.1148.
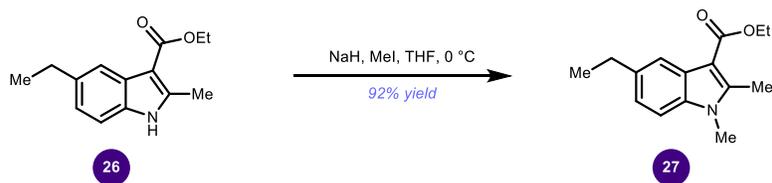
ethyl 5-ethyl-1,2-dimethyl-1*H*-indole-3-carboxylate (27):

To a solution of compound **26** (0.600 g, 2.59 mmol, 1.00 equiv) in anhydrous THF (26 mL) was added NaH (60 % dispersion in mineral oil, 125 mg, 3.11 mmol, 1.2 equiv) at 0 °C. After 30 min, iodomethane (194 µL, 3.11 mmol, 1.2 equiv) was added dropwise at 0 °C. The mixture was stirred at the same temperature for 30 min before it was quenched with water (0.5 mL). After removal of the solvents on a rotary evaporator, the residue was purified by flash column chromatography (silica gel, 15% EtOAc/hexanes) to yield the product (0.583 g, 2.38 mmol, 92%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.96 (s, 1H), 7.20 (d, *J* = 8.3 Hz, 1H), 7.09 (dd, *J* = 8.3, 1.8 Hz, 1H), 4.41 (q, *J* = 7.2 Hz, 2H), 3.67 (s, 3H), 2.83 – 2.69 (m, 5H), 1.46 (t, *J* = 7.1 Hz, 3H), 1.30 (t, *J* = 7.6 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 166.42, 145.26, 137.89, 135.22, 126.99, 122.49, 120.23, 108.92, 103.76, 59.43, 29.73, 29.30, 16.58, 14.78, 12.05. HRMS (ESI): calculated C$_{15}$H$_{20}$NO$_2$ [M+H]$^+$: 246.1489, found: 246.1481.
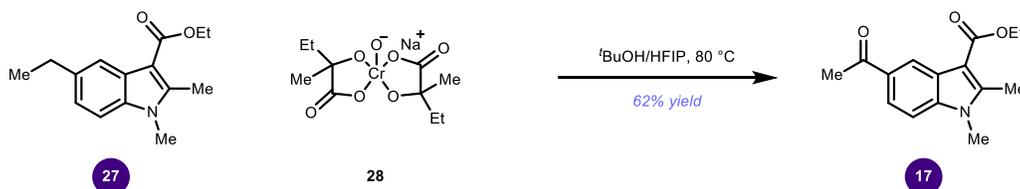
ethyl 5-acetyl-1,2-dimethyl-1*H*-indole-3-carboxylate (17):

Following a modified literature procedure,[5] compound **27** (24.5 mg, 0.10 mmol, 1.00 equiv), and sodium bis(2-hydroxy-2-methylbutyrato)oxochromate(V) (compound **28**, 194 mg, 0.60 mmol, 6.00 equiv) were placed in a 2-dram vial with a stir bar. *t*BuOH (0.10 mL) and HFIP (0.90 mL) were added, and the mixture was then stirred at 80 °C for 5 h, open to air. On complete consumption of compound **27**, silica gel (582 mg, 300 wt% to **28**) was added. After removal of solvent under reduced pressure, the resulting residue was purified by flash chromatography (silica gel, 25% – 30% EtOAc/hexanes) to afford the product (16.0 mg, 0.062 mmol, 62%)

The characterization data matched spectral values from our earlier synthesis of **17**.
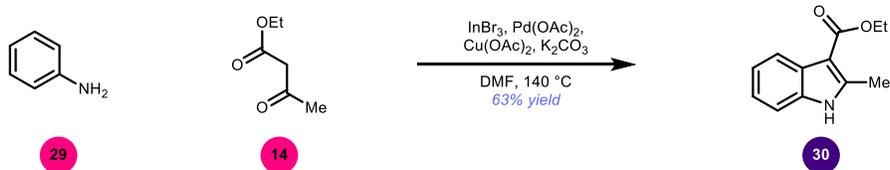
ethyl 2-methyl-1*H*-indole-3-carboxylate (30):

Following a modified literature procedure,[1] to an oven-dried 2-dram vial with a Teflon-coated stir bar was added aniline (**29**) (0.729 mL, 8.00 mmol, 1.00 equiv), and ethyl acetoacetate (1.113 mL, 8.80 mmol, 1.10 equiv). Then the vial was moved into the glovebox, and InBr$_3$ (28.4 mg, 0.08 mmol, 1 mol%) was added. Then the vial was capped and moved out of the glove box. The reaction mixture was stirred at room temperature for 30 min. The mixture was transferred with a syringe into a 250 mL round bottom flask containing Pd(OAc)$_2$ (89.8 mg, 0.40 mmol, 5 mol%), Cu(OAc)$_2$ (4.359 g, 24.00 mmol, 3.0 equiv), K$_2$CO$_3$ (3.317 g, 24.00 mmol, 3.0 equiv), and DMF (70 mL), and the residues were rinsed into the bigger flask with DMF (2 × 5 mL). The reaction mixture was stirred at 140 °C under N$_2$ atmosphere for 30 min before it was cooled to room temperature, diluted with EtOAc (100 mL) and filtered through a short pad of silica and sea sand. The red-brown solid was washed with EtOAc (2 × 80 mL) and the combined filtrates were concentrated *in vacuo* to yield a crude product, which was purified by flash column chromatography (silica gel, 20% EtOAc/hexanes) to yield the product (1.015 g, 5.00 mmol, 63%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.44 (br s, 1H), 8.11 (d, *J* = 7.3 Hz, 1H), 7.30 (dd, *J* = 6.9, 1.3 Hz, 1H), 7.24 – 7.16 (m, 2H), 4.41 (q, *J* = 7.2 Hz, 2H), 2.74 (s, 3H), 1.45 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 166.29, 144.06, 134.62, 127.33, 122.46, 121.81, 121.46, 110.61, 104.80, 59.65, 14.74, 14.36. HRMS (ESI): calculated C$_{12}$H$_{13}$NNaO$_2$ [M+Na]$^+$: 226.0838, found: 226.1582.
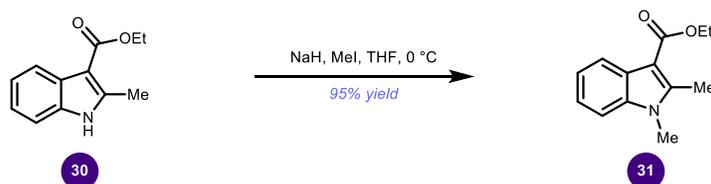
ethyl 1,2-dimethyl-1*H*-indole-3-carboxylate (31):

To a solution of compound **30** (1.003 g, 4.935 mmol, 1.00 equiv) in anhydrous THF (49 mL) was added NaH (60 % dispersion in mineral oil, 237 mg, 5.922 mmol, 1.2 equiv) at 0 °C. After 30 min, iodomethane (369 μL, 5.922 mmol, 1.2

equiv) was added dropwise at 0 °C. The mixture was stirred at the same temperature for 30 min before it was quenched with water (1.0 mL). After removal of the solvents on a rotary evaporator, the residue was purified by flash column chromatography (silica gel, 15% EtOAc/hexanes) to yield the product (1.016 g, 4.676 mmol, 95%).

^{1}H NMR (500 MHz, CDCl$_3$) δ 8.16 – 8.11 (m, 1H), 7.30 – 7.26 (m, 1H), 7.25 – 7.20 (m, 2H), 4.41 (q, *J* = 7.1 Hz, 2H), 3.66 (s, 3H), 2.76 (s, 3H), 1.46 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, CDCl$_3$) δ 166.31, 145.38, 136.60, 126.70, 122.05, 121.70, 121.55, 109.12, 104.03, 59.47, 29.64, 14.74, 11.94. HRMS (ESI): calculated C$_{13}$H$_{15}$NNaO$_2$ [M+Na]$^+$: 240.0995, found: 240.0992.

ethyl 5-(2-chloroacetyl)-1,2-dimethyl-1*H*-indole-3-carboxylate (19):

Following a modified literature procedure,[3] to a suspension of anhydrous AlCl$_3$ (1.600 g, 12.00 mmol, 4.00 equiv) in anhydrous d (15 mL) at 0 °C was added chloroacetyl chloride (1.004 mL, 12.60 mmol, 4.2 equiv) dropwise. The mixture was stirred at 0 °C for 20 min before warmed up to room temperature. To this solution was added compound **31** (0.652 g, 3.00 mmol, 1.00 equiv). After the reaction mixture was stirred at room temperature for 3 h, it was poured into ice-cold water and extracted with EtOAc (3 × 40 mL). The organic layers were combined, washed with brine, and dried over anhydrous sodium sulfate. The organic layer was then concentrated under reduced pressure and purified by flash column chromatography (silica gel, 25% EtOAc/hexanes) to afford a mixture of two regioisomers (2:1, 838 mg, 2.854 mmol, 95%). The major regioisomer is compound **19**.

The characterization data matched with the one obtained from the other route.
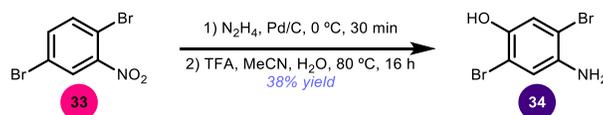
4-amino-2,5-dibromophenol (34):

To a 50 mL round bottom flask, **33** (2.81 g, 10.0 mmol, 1.00 equiv) and palladium on activated carbon (53.4 mg, 0.500 mmol, 0.0500 equiv) were added in

tetrahydrofuran (25.0 mL) and cooled to 0 ºC. Hydrazine monohydrate (0.970 mL, 20 mmol, 2.00 equiv) was added dropwise. The reaction was stirred at 0 °C for 30 minutes and warmed to 25 ºC. The mixture was filtered through a pad of Celite and concentrated under reduced pressure. The crude mixture was used for next step without further purification.[6]

To a 100 mL round bottom flask, *N*-(2,5-dibromophenyl)hydroxylamine from last step and trifluoroacetic acid (3.80 mL, 50.0 mmol, 5.00 equiv) were dissolved in acetonitrile (10.0 mL) and water (40.0 mL) under nitrogen atmosphere. The reaction was stirred and heated at 80 °C for 16 hours. The reaction mixture was neutralized with saturated aqueous sodium bicarbonate solution (pH = 8), extracted with ethyl acetate (3 × 40 mL, dried over magnesium sulfate and concentrated under reduced pressure. The residue was purified with silica gel column chromatography (15% ethyl acetate in hexanes) to afford the product (1.00 g, 38%).[7]

^{1}H NMR (500 MHz, CD$_3$CN) δ 6.99 (s, 1H), 6.94 (s, 1H), 6.68 (s, 1H), 4.14 (s, 2H). ^{13}C NMR (125 MHz, CD$_3$CN) δ 146.26, 140.72, 120.60, 119.66, 110.55, 108.45. HRMS (ESI): calculated C$_6$H$_6$Br$_2$NO [M+H]$^+$: ^{79}Br/^{79}Br 265.8811, found: 265.8806.
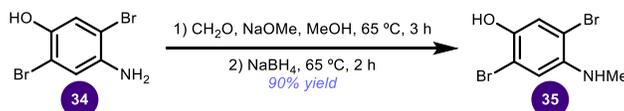
6-bromo-5-hydroxy-1-methyl-2-((phenylthio)methyl)-1*H*-indole-3-carboxylate (35):

To a 10 mL round bottom flask, **34** (187 mg, 0.700 mmol, 1 equiv), sodium methoxide (270 mg, 5.00 mmol, 7.14 equiv) and paraformaldehyde (70.1 mg, 2.5 mmol, 3.57 equiv) were dissolved in methanol (4.00 mL) under nitrogen atmosphere. The reaction mixture was stirred and heated at 65 °C for 3 hours before cooled to 25 ºC. After sodium borohydride (94.6 mg, 2.5 mmol, 3.57 equiv) was added, reaction was heated to 65 ºC and stirred for 2 hours. The reaction mixture was neutralized with saturated aqueous ammonium chloride solution (pH = 7), extracted with ethyl acetate (3 × 5 mL), dried over magnesium sulfate and concentrated under reduced pressure. The residue was purified with silica gel column chromatography (10% ethyl acetate in hexanes) to afford the product (176.4 mg, 90%).[8]

^{1}H NMR (500 MHz, CD$_3$CN) δ 7.08 (s, 1H), 6.76 (s, 1H), 6.62 (s, 1H), 4.27 (s, 1H), 2.75 (s, 3H). ^{13}C NMR (125 MHz, CD$_3$CN) δ 145.35, 142.48, 121.17, 115.02,

110.81, 108.89, 31.20. HRMS (ESI): calculated $C_7H_8Br_2NO$ [M+H]$^+$: $^{79}Br/^{79}Br$ 279.8967, found: 279.8963.
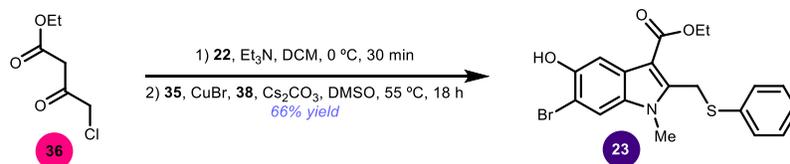
6-bromo-5-hydroxy-1-methyl-2-((phenylthio)methyl)-1*H*-indole-3-carboxylate (23):

To a 2-dram vial, **36** (94.6 µL, 0.700 mmol, 1.00 equiv) and thiophenol (71.4 µL, 0.700 mmol, 1,00 equiv) were dissolved in dichloromethane (4.00 mL). The reaction mixture was cooled to 0 ºC and triethylamine (105 µL, 0.750 mmol, 1.07 equiv) was added dropwise. The reaction was stirred for 30 minutes before warmed to 25 ºC. The reaction mixture was extracted sequentially with saturated aqueous sodium bicarbonate solution (4 mL), hydrochloric acid (1 M, 4 mL) and saturated sodium chloride solution (4 mL), dried over magnesium sulfate and concentrated under reduced pressure. The crude mixture was used for next step without further purification.[9]

To a 2-dram vial, **35** (28.1 mg, 0.100 mmol, 1.00 equiv), ethyl 3-oxo-4-(phenylthio)butanoate from last step, copper (I) bromide (2.9 mg, 0.020 mmol, 0.2 equiv), 2-(1H-tetrazol-1-yl)acetic acid (5.1 mg, 0.040 mmol, 0.4 equiv) and cesium carbonate (97.7 mg, 0.300 mmol, 3.00 equiv) were dissolved in dimethyl sulfoxide (1.00 mL) under nitrogen atmosphere. The reaction mixture was stirred at 55 °C for 18 hours. The reaction mixture was neutralized with saturated aqueous ammonium chloride solution (1 mL), extracted with ethyl acetate (3 × 2 mL), dried over magnesium sulfate and concentrated under reduced pressure. The residue was purified with silica gel column chromatography (10% ethyl acetate in hexanes to afford 6-bromo-5-hydroxy-1-methyl-2-((phenylthio)methyl)-1H-indole-3-carboxylate (27.7 mg, 66%).[10]

^{1}H NMR (500 MHz, (CD$_3$)$_2$SO) δ 9.81 (s, 1H), 7.73 (s, 1H), 7.56 (s, 1H), 7.36 – 7.33 (m, 2H), 7.31 – 7.24 (m, 3H), 4.78 (s, 2H), 4.18 (q, *J* = 7.1 Hz, 2H), 3.66 (s, 3H), 1.29 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, (CD$_3$)$_2$SO) δ 164.18, 149.27, 143.36, 134.22, 131.46, 131.09, 128.96, 127.24, 125.96, 114.22, 106.40, 106.28, 103.06, 59.11, 30.16, 28.26, 14.29. HRMS (ESI): calculated $C_{19}H_{18}BrNNaO_3S$ [M+Na]$^+$: ^{79}Br 442.0083, found: 442.0081.
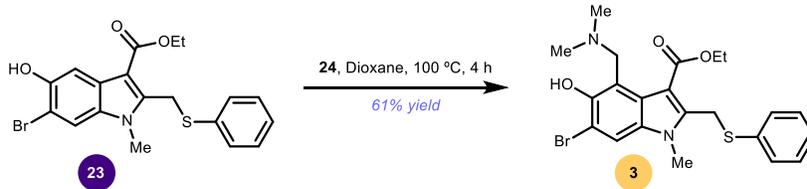
ethyl 6-bromo-4-((dimethylamino)methyl)-5-hydroxy-1-methyl-2-((phenylthio)methyl)-1_H_-indole-3-carboxylate (3):

To a 1-dram vial, **23** (23.8 mg, 0.0566 mmol, 1 equiv) and **24** (20 µL, 0.150 mmol, 2.65 equiv) were dissolved in 1,4-dioxane (0.30 mL). The reaction was stirred and heated to 100 ºC for 4 h before removing the solvent under reduced pressure. The residue was purified with silica gel column chromatography (5% methanol in dichloromethane) to afford the product (14.6 mg, 61%).[11]

^{1}H NMR (500 MHz, (CD$_3$)$_2$SO) δ 7.71 (s, 1H), 7.37 – 7.23 (m, 5H), 4.63 (s, 2H), 4.14 (q, _J_ = 7.1 Hz, 2H), 4.03 (s, 2H), 3.66 (s, 3H), 2.27 (s, 6H), 1.23 (t, _J_ = 7.1 Hz, 3H). ^{13}C NMR (125 MHz, (CD$_3$)$_2$SO) δ 64.80, 150.41, 141.67, 134.37, 131.31, 131.01, 129.06, 127.26, 123.45, 113.08, 112.58, 107.37, 105.23, 59.86, 59.04, 43.54, 30.17, 28.75, 14.04. HRMS (ESI): calculated C$_{22}$H$_{26}$BrN$_2$O$_3$S [M+H]$^+$: [79]Br 477.0842, found: 477.0844.
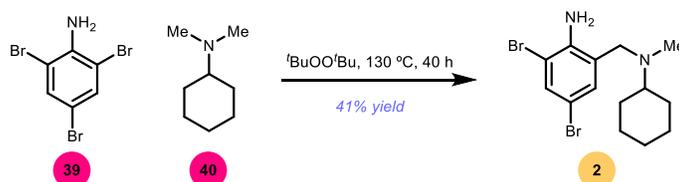
2,4-dibromo-6-((cyclohexyl(methyl)amino)methyl)aniline (2):

To a 2-dram vial, 2,4,6-tribromoaniline (98.9 mg, 0.300 mmol, 1 equiv) and 2-(tert-butylperoxy)-2-methylpropane (110 µL, 0.600 mmol, 2 equiv) were dissolved in N,N-dimethylcyclohexanamine (0.90 mL, 6.0 mmol, 20 equiv) under nitrogen atmosphere. The reaction was stirred and heated at 130 °C for 40 hours. The mixture was diluted with ethyl acetate (3 mL), dried over magnesium sulfate and concentrated under reduced pressure. The residue was purified with silica gel column chromatography (5% ethyl acetate in hexanes) to afford the product (45.9 mg, 41%).[12]

^{1}H NMR (500 MHz, CDCl$_3$) δ 7.46 (d, _J_ = 2.3 Hz, 1H), 7.04 (d, _J_ = 2.3 Hz, 1H), 5.48 (s, 2H), 3.59 (s, 2H), 2.41 (tt, _J_ = 11.3, 3.1 Hz, 1H), 2.11 (s, 3H), 1.83 – 1.77 (m, 4H), 1.68 – 1.59 (m, 1H), 1.36 – 1.17 (m, 4H), 1.09 (qt, _J_ = 12.4, 3.1 Hz, 1H). ^{13}C NMR (125 MHz, CDCl$_3$) δ 144.25, 133.07, 131.80, 126.32, 110.17, 108.19, 62.04, 57.94, 36.51, 28.34, 26.41, 26.04. HRMS (ESI): calculated C$_{14}$H$_{21}$Br$_2$N$_2$ [M+H]$^+$: [79]Br/[79]Br 375.0066, found: 375.0073.
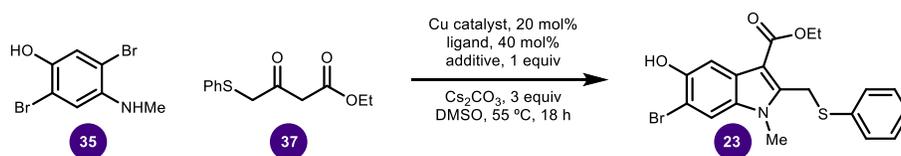
Stock solutions, or suspensions, were prepared as shown in the Table. In an inert atmosphere glovebox, reagents were weighed and dissolved or suspended in anhydrous degassed DMSO to achieve the concentration listed in Table. Stock solutions of reagents were stirred until either a clear solution or a uniform slurry was achieved. A 24-well aluminum microvial plate (Analytical Sales & Services cat. no. 25243) was equipped with oven-dried shell vials (Analytical Sales & Services cat. no. 884001) then moved to the glove box. Stock solutions were dosed to the appropriate shell vials according to the plate map shown in Table using single channel micropipetters. A parylene-coated stir dowel (Analytical Sales & Services cat. no. 13258) was then added to each vial. The microvial plate was sealed, removed from the glove box, and heated to 55 ˚C for 18 h with stirring on a ChemGlass stirring hotplate.

The reactions were quenched by opening the reaction block and adding 100 μL saturated aqueous ammonium chloride solution and 420 μL EtOAc. Reactions were extracted by resealing the plate and shaking manually. From each reaction, an aliquot of the quenched reaction mixture was added into a 96-well polypropylene collection plate (Analytical Sales & Services cat. no. 17P687). The solvent was evaporated by nitrogen blow down on the analytical plate. An acetonitrile solution of caffeine as internal standard (0.005 M, 800 μL) was added, and mixed by pipetting up and down. The reactions were then analyzed by UPLC-MS.The assay yields was produced by measuring the UV absorbance of **23** relative to the caffeine internal standard.

Reagents	C_{stock} (M)	V_{dose} (μL)	Wells
35	0.6	16.7	All
37	3.0	16.7	All
Cs_2CO_3	2.4	16.7	All
Blank	-	16.7	A1,3,5-D1,3,5
$MgSO_4$	0.6	16.7	A2,4,6-D2,4,6
			A1,2-D1,2
2-(1H-tetrazol-1-yl)acetic acid (L1)	0.24	16.7	A3,4-D3,4

2,6-dimethylanilino(oxo)acetic acid (L2)	0.24	16.7	A5,6-D5,6
Copper(I) iodide	0.12	16.7	A1-A6
Copper(I) bromide	0.12	16.7	B1-B6
Tetrakisacetonitrile copper(I) triflate	0.12	16.7	C1-C6
Copper(II) acetate	0.12	16.7	D1-D6

Reagents	Reaction conc. (M)	Color
35	0.100	
37	0.500	
Cs_2CO_3	0.400	
Blank	-	
Blank	-	
$MgSO_4$	0.100	
L1	0.040	
L2	0.040	
CuI	0.020	
CuBr	0.020	
CuOTf	0.020	
$Cu(OAc)_2$	0.020	
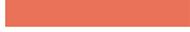
References:

1. Wurtz, S., S. Rakshit, J.J. Neumann, T. Droge, and F. Glorius, Palladium-Catalyzed Oxidative Cyclization of *N*-Aryl Enamines: From Anilines toIndoles. *Angewandte Chemie International Edition*, **47**, 7230-7233 (2008). https://doi.org/10.1002/anie.200802482

2. Rodrigo A. Rodriguez, Chung-Mao Pan, Yuki Yabe, Yu Kawamata, Martin D. Eastgate, and Phil S. Baran, Palau'chlor: A Practical and Reactive Chlorinating Reagent *Journal of the American Chemical Society,* **136**, 6908-6911, (2014). https://doi.org/10.1021/ja5031744

3. Nakatsuka, S.-i.A., Osamu; Ueda, Kazuo; Goto, Toshio, Introduction of a hydroxy group onto 5- and 6-position of indole nucleus by Friedel-Crafts acylation and subsequent Baeyer-Villiger Oxidation. *Heterocycles*, **26**, 1471-1474 (1987).

4. Trofimov, F.A., Tsyshkova, N.G., Zotova, S.A. *et al.* Synthesis of a new antiviral agent, arbidole. *Pharm Chem J* **27,** 75-76 (1993). https://doi.org/10.1007/BF00772858

5. Kanda, Y., H. Nakamura, S. Umemiya, R.K. Puthukanoori, V.R. Murthy Appala, G.K. Gaddamanugu, B.R. Paraselli, and P.S. Baran, Two-Phase Synthesis of Taxol. *Journal of the American Chemical Society*, **142**, 10526-10533 (2020). https://doi.org/10.1021/jacs.0c03592

6. Riesco-Domínguez, A.; van de Wiel, J.; Hamlin, T. A.; van Beek, B.; Lindell, S. D.; Blanco-Ania, D.; Bickelhaupt, F. M.; Rutjes, F. P. J. T. Trifluoromethyl Vinyl Sulfide: A Building Block for the Synthesis of CF$_3$S-

Containing Isoxazolidines. *J. Org. Chem.*, **83**, 1779-1789 (2018). https://doi.org/10.1021/acs.joc.7b02639

7. de Fonzo, N.; Quartarone, G.; Ronchin, L.; Tortato, C.; Vavasori, A. Kinetics and mechanistic study of the Bamberger rearrangement of *N*-phenylhydroxylamine to 4-aminophenol in acetonitrile-trifluoroacetic acid: A substrate acid complex as para selectivity driver. *Applied Catalysis A: General*, **516**, 58–69 (2016). https://doi.org/10.1016/j.apcata.2016.02.020

8. Zhang, W.; Oya, S.; Kung, M.-P.; Hou, C.; Maier, D. L.; Kung, H. F. F-18 Stilbenes as PET Imaging Agents for Detecting β-Amyloid Plaques in the Brain. *J. Med. Chem.*, **48**, 5980–5988 (2005). https://doi.org/10.1021/jm050166g

9. Chen, T.; Benmohamed, R.; Arvanites, A. C.; Ranaivo, H. R.; Morimoto, R. I.; Ferrante, R. J.; Watterson, D. M.; Kirsch, D. R.; Silverman, R. B. Arylsulfanyl pyrazolones block mutant SOD1-G93A aggregation. Potential application for the treatment of amyotrophic lateral sclerosis. *Bioorg. & Med. Chem.*, **19**, 613-622 (2011). https://doi.org/10.1016/j.bmc.2010.10.052

10. Liu X.-G.; Li, Z.-H.; Xie, J.-W.; Liu, P.; Zhang, J.; Dai, B. Copper-catalyzed synthesis of 2,3-disubstituted indoles from ortho-haloanilines and β-keto esters/β-diketone. *Tetrahedron*, **72**, 653-657 (2016). https://doi.org/10.1016/j.tet.2015.12.006

11. Wright, Z. V. E.; Wu, N. C.; Kadam, R. U.; Wilson, I. A.; Wolan, D. W. Structure-based optimization and synthesis of antiviral drug Arbidol analogues with significantly improved affinity to influenza hemagglutinin. *Bioorg. & Med. Chem.*, **27**, 3744-3748 (2017). https://doi.org/10.1016/j.bmcl.2017.06.074

12. Ueno, R.; Ikeda, Y.; Shirakawa, E. *tert*-Butoxy-Radical-Promoted α-Arylation of Alkylamines with Aryl Halides. *Eur. J. Org. Chem.* **2017**, 4188 (2017). https://doi.org/10.1002/ejoc.201700548

Spectra
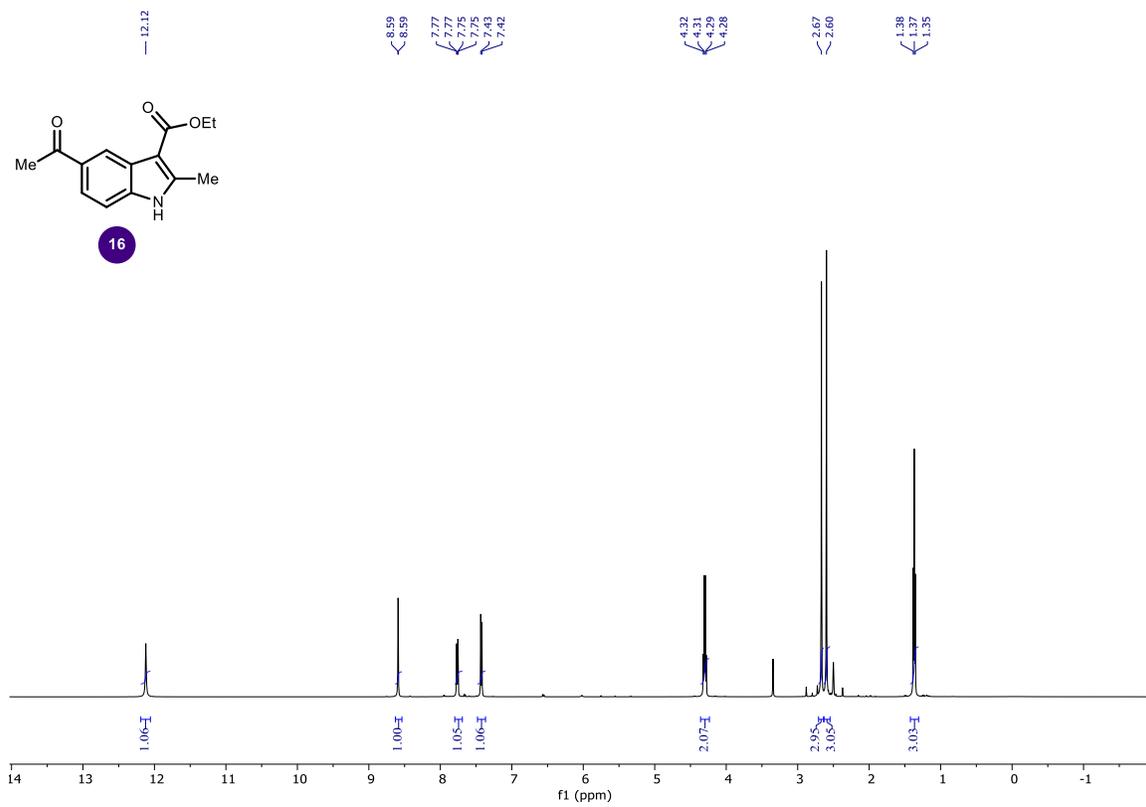
Figure S1. ^{1}H NMR of **16** in $(CD_3)_2SO$ at 25 ºC.

Figure S2. ^{13}C NMR of **16** in $(CD_3)_2SO$ at 25 ºC.
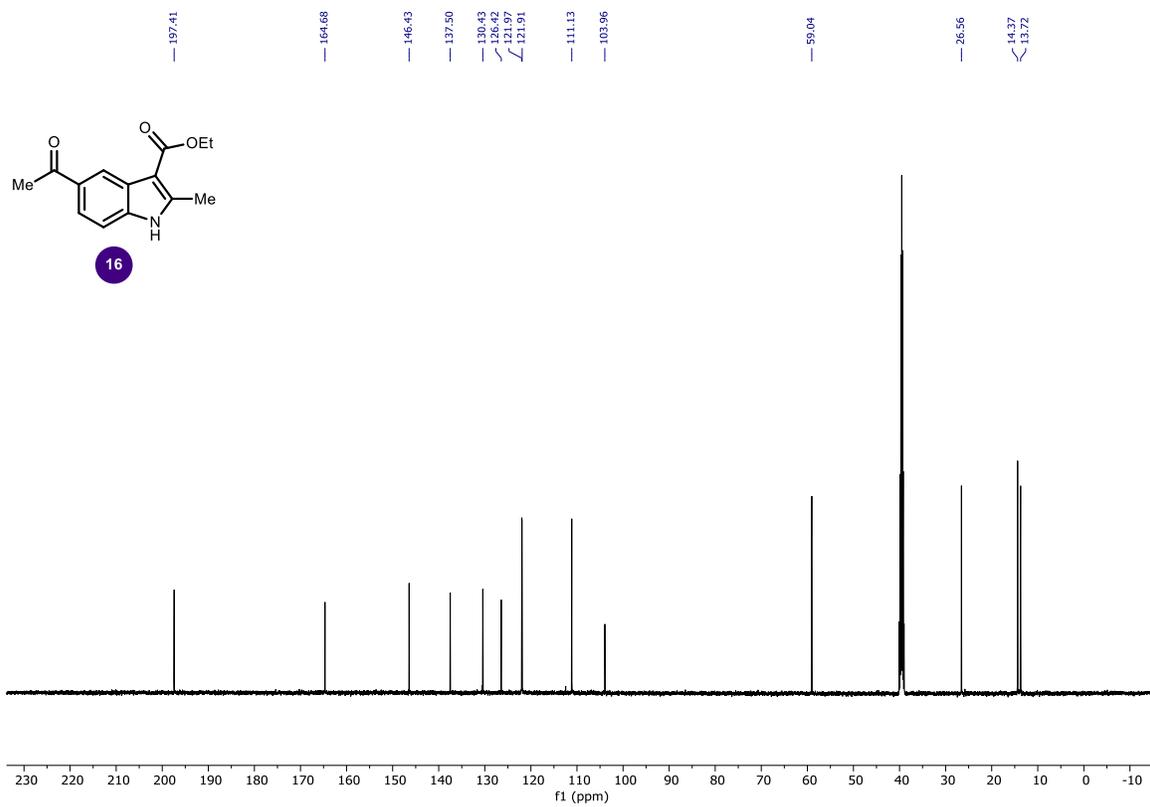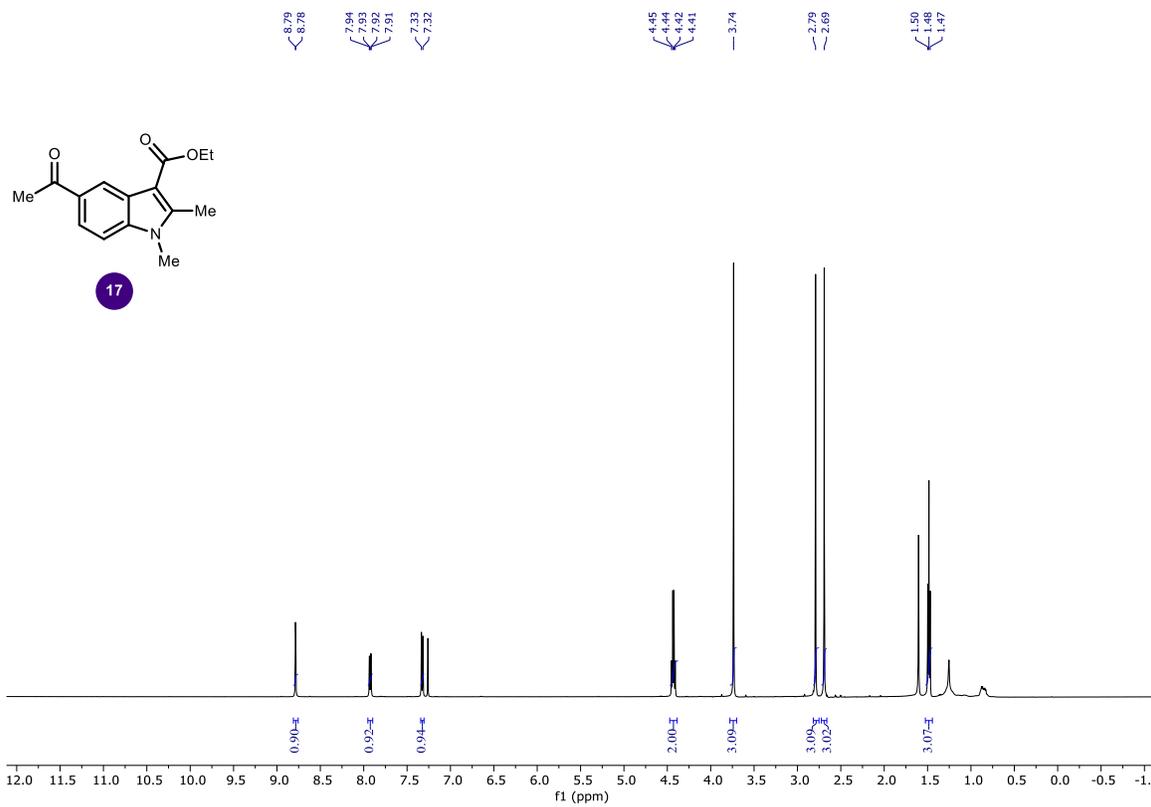
Figure S3. ^{1}H NMR of **17** in CDCl$_3$ at 25 °C.
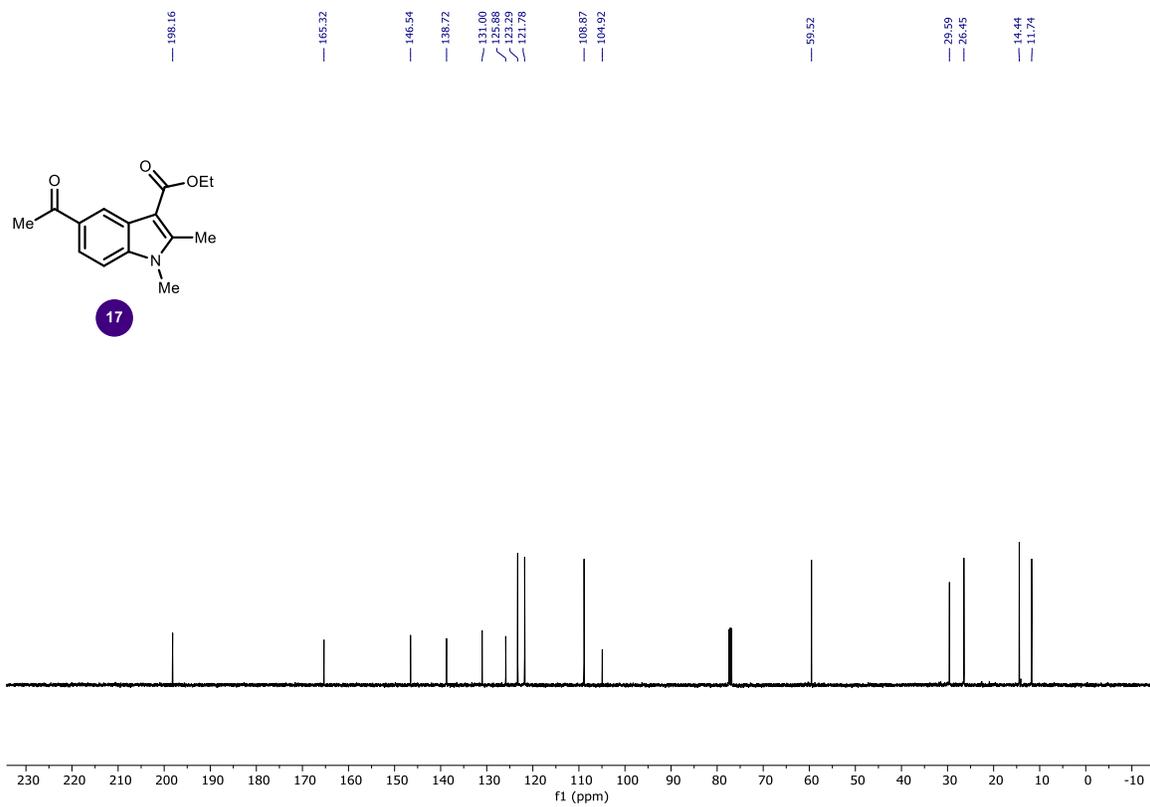
Figure S4. ^{13}C NMR of **17** in CDCl$_3$ at 25 °C.
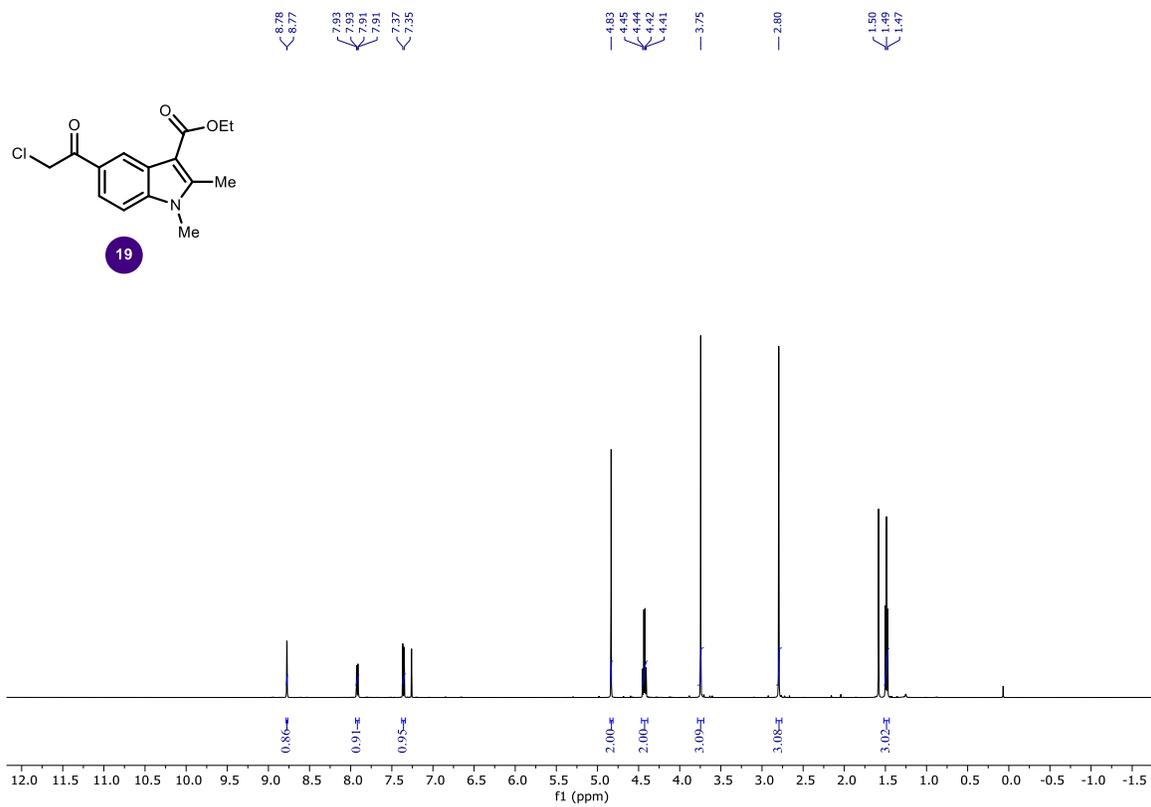
Figure S5. ^{1}H NMR of **19** in CDCl$_3$ at 25 °C.
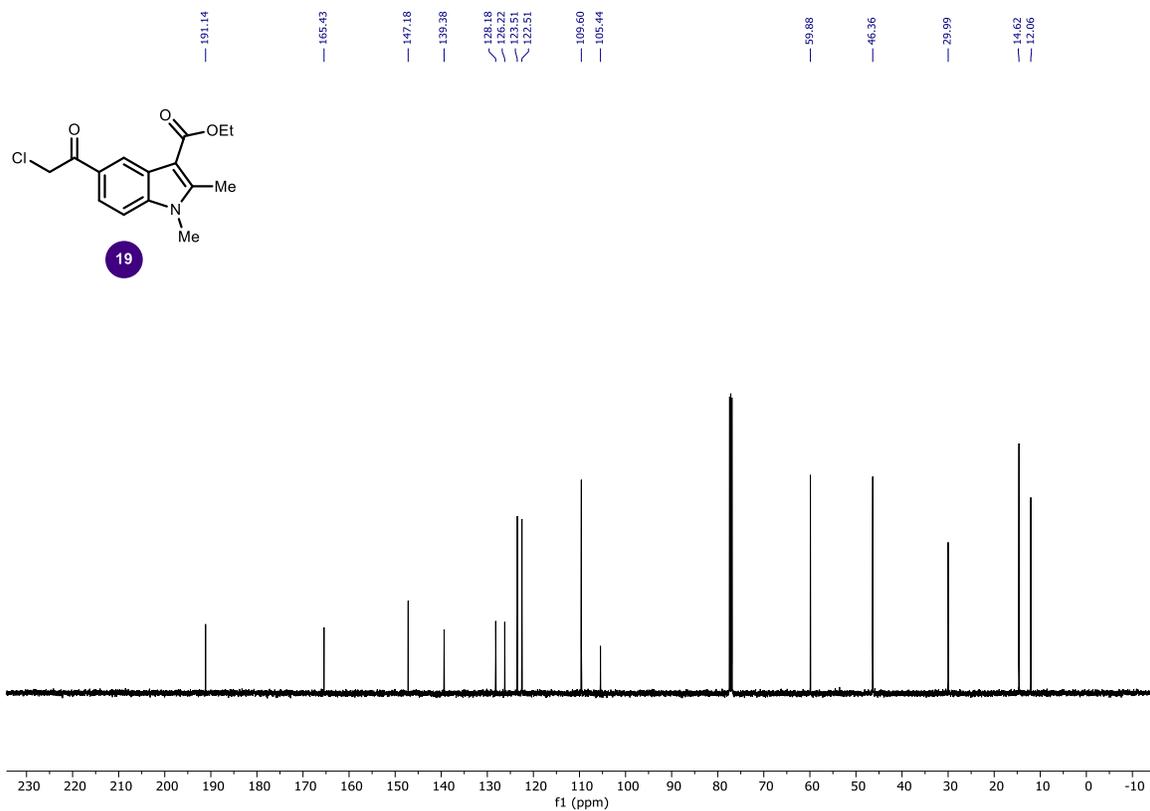
Figure S6. ^{13}C NMR of **19** in CDCl$_3$ at 25 ºC.
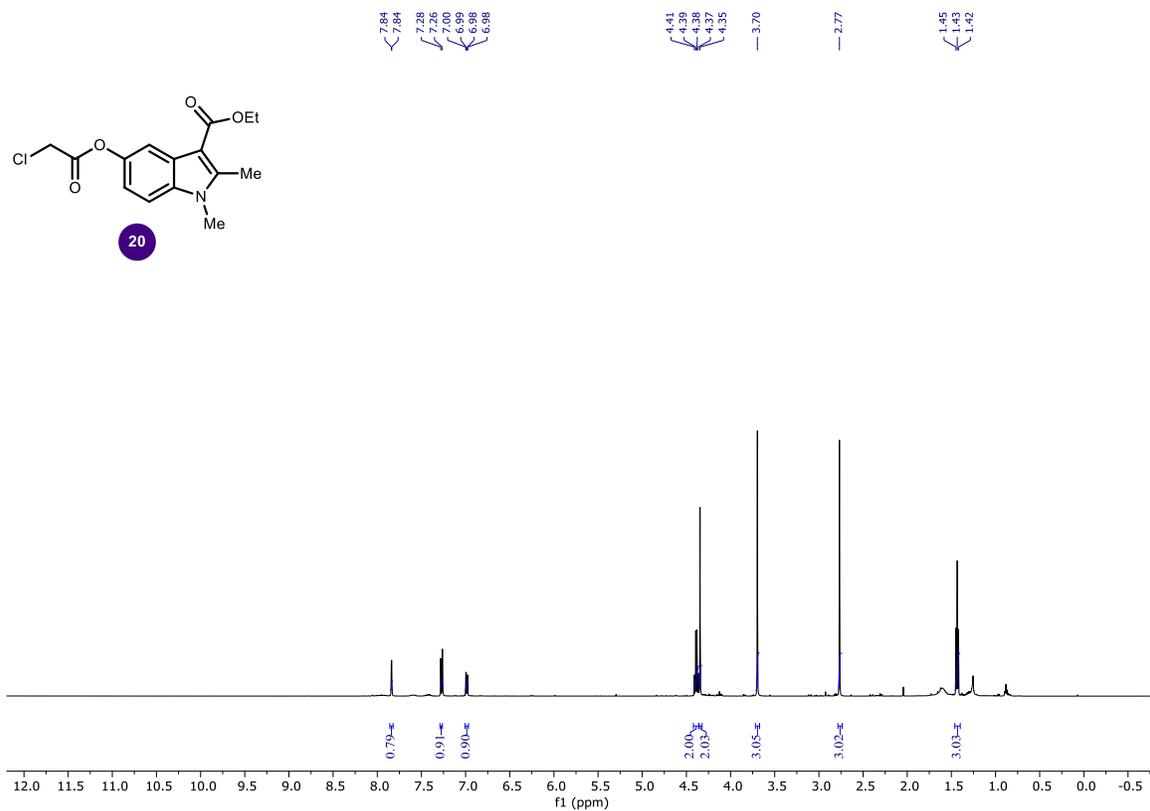
Figure S7. ^{1}H NMR of **20** in CDCl$_3$ at 25 ºC.
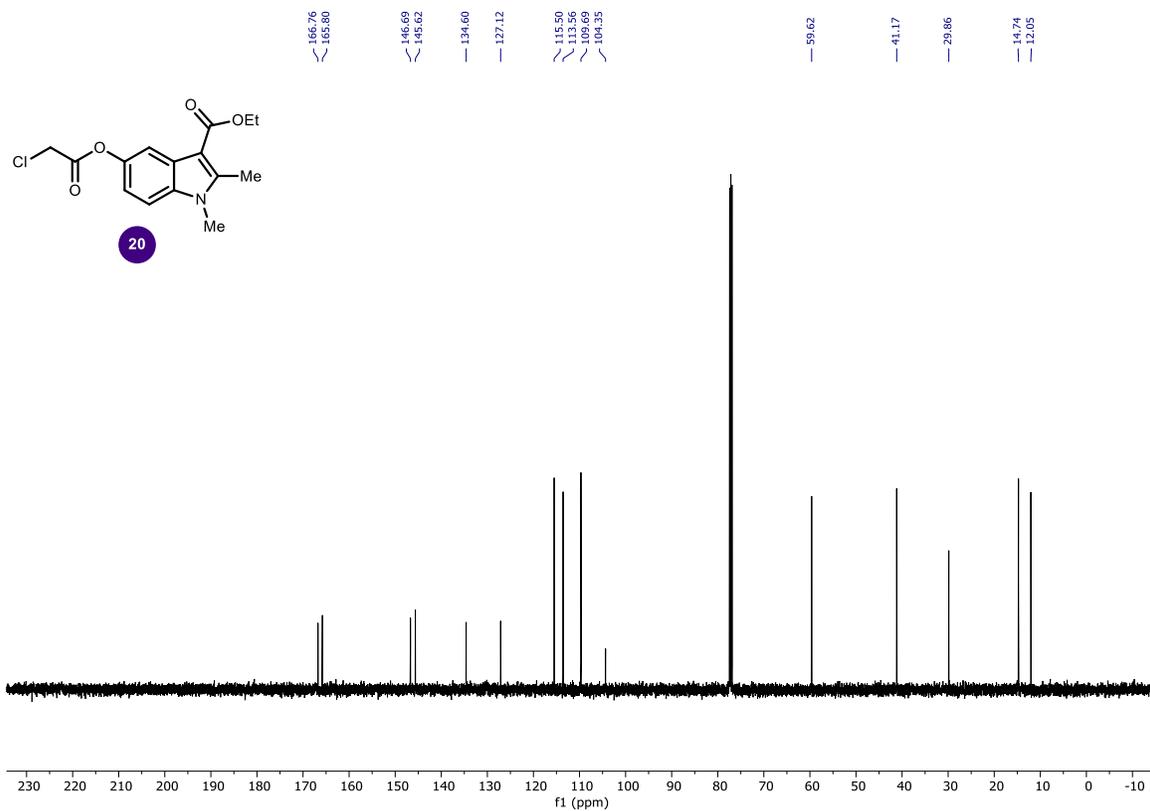
Figure S8. ^{13}C NMR of **20** in CDCl$_3$ at 25 °C.
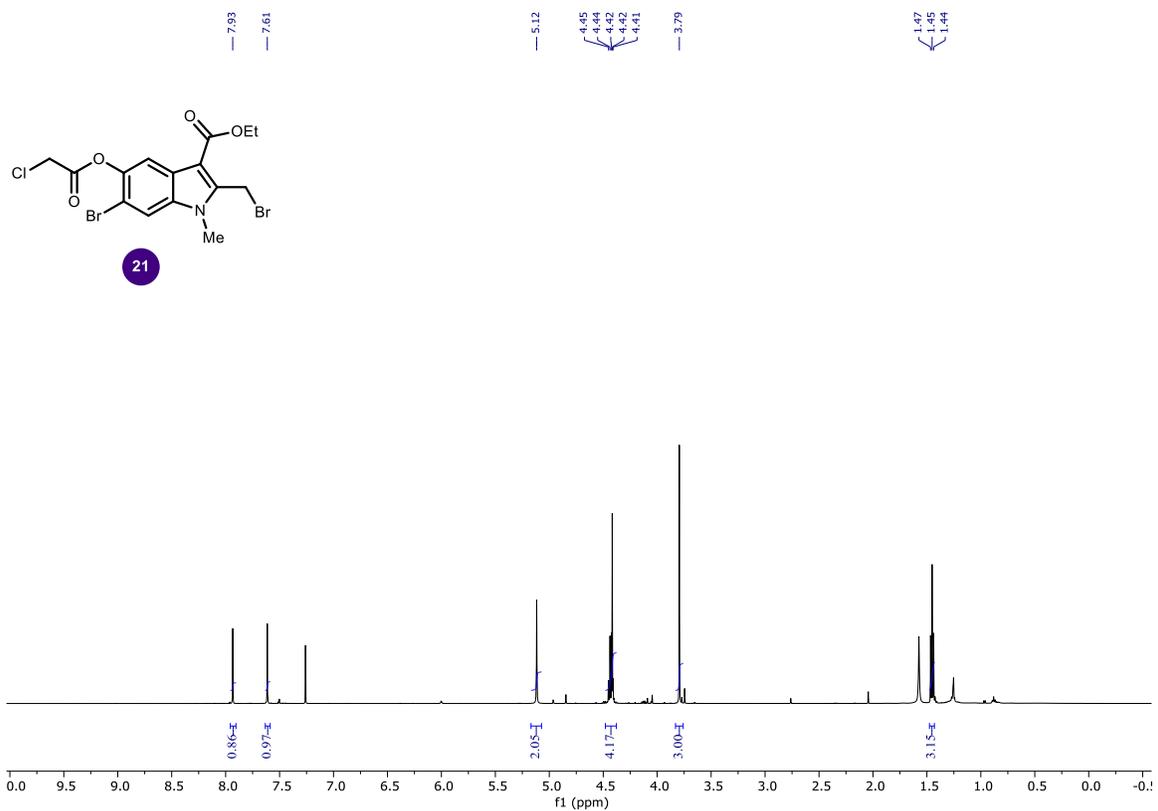
Figure S9. ^{1}H NMR of **21** in CDCl$_3$ at 25 ºC.
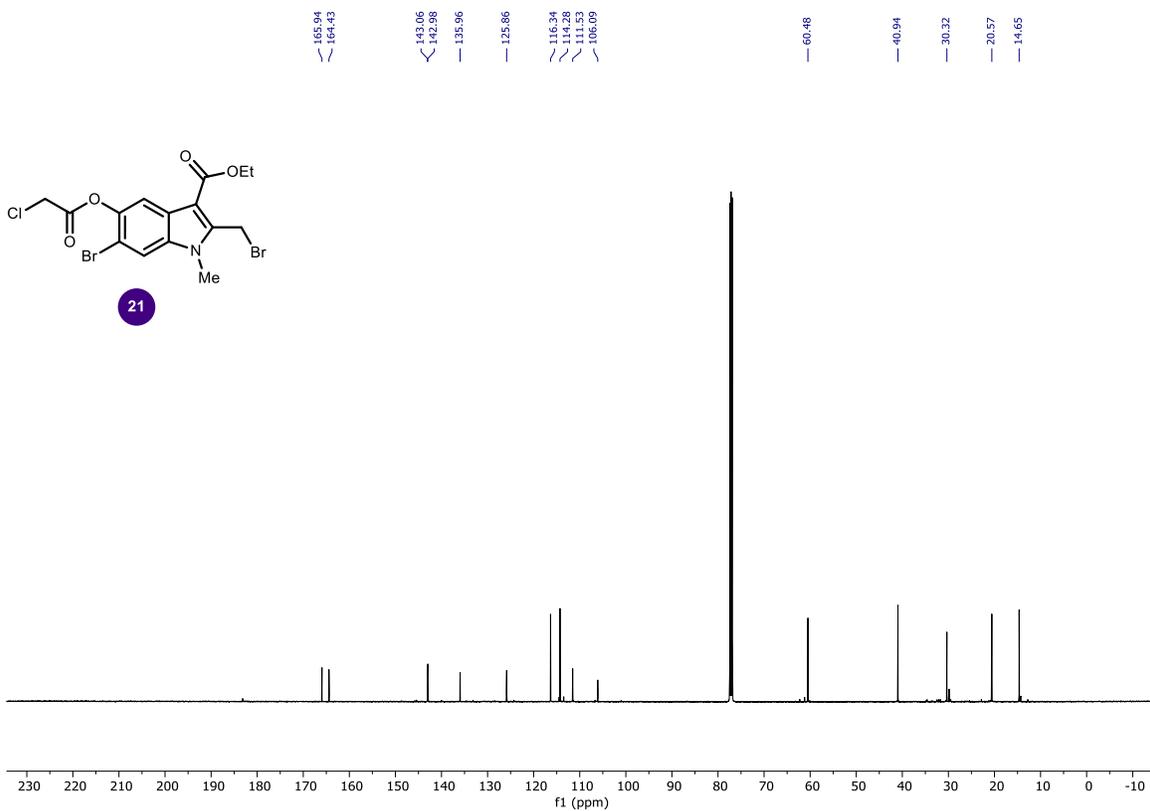
Figure S10. ^{13}C NMR of **21** in CDCl$_3$ at 25 ºC.
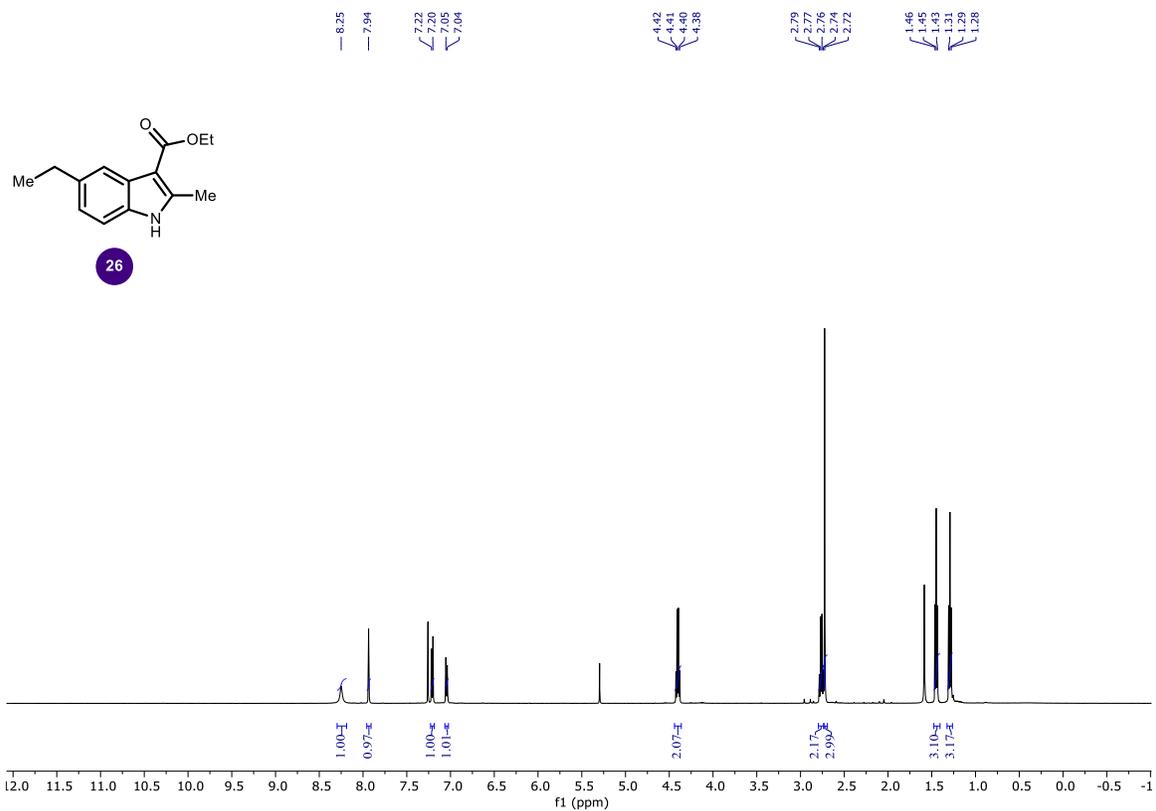
Figure S11. ^{1}H NMR of **26** in CDCl$_3$ at 25 °C.
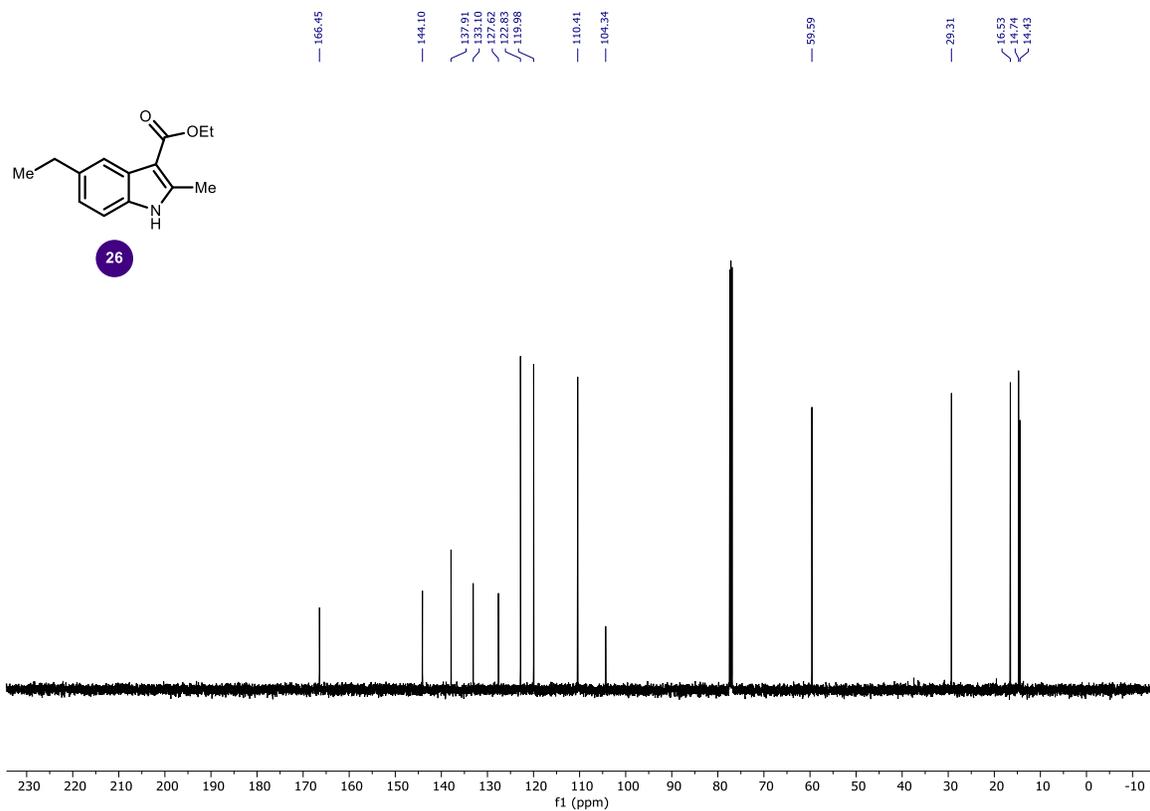
Figure S12. ^{13}C NMR of **26** in CDCl$_3$ at 25 ºC.
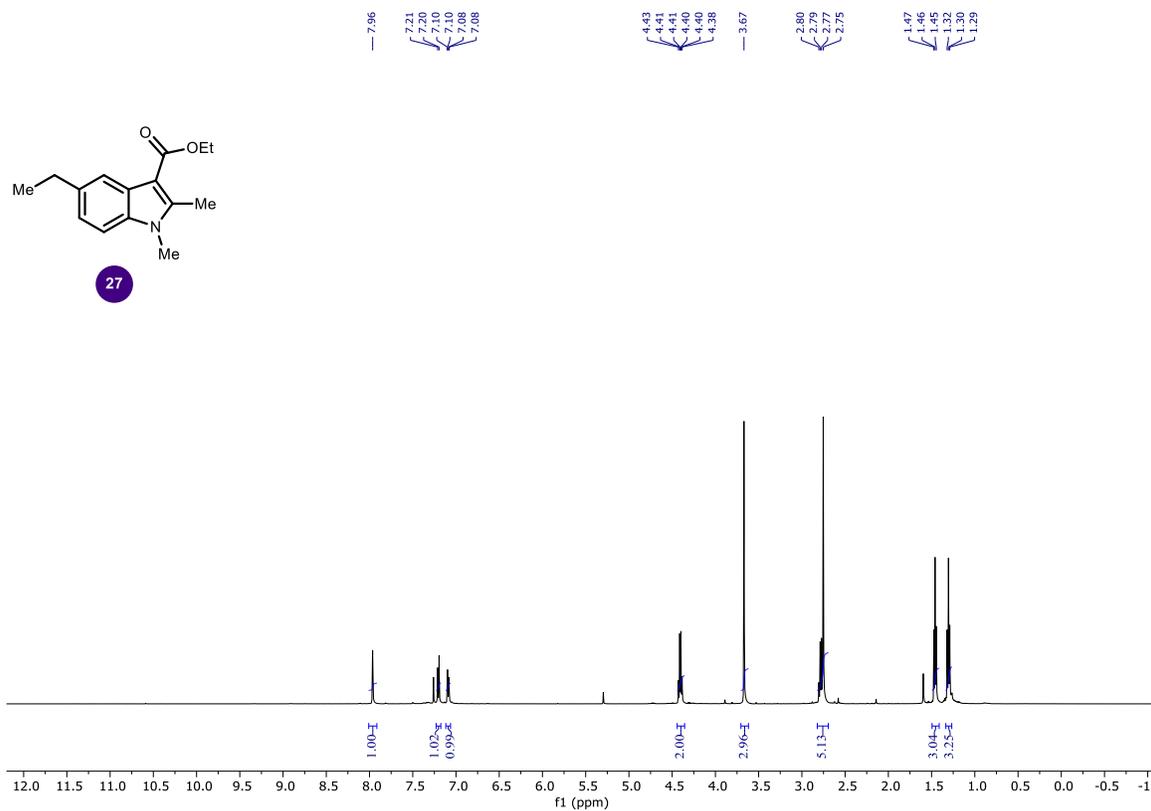
Figure S13. ^{1}H NMR of **27** in CDCl$_3$ at 25 ºC.
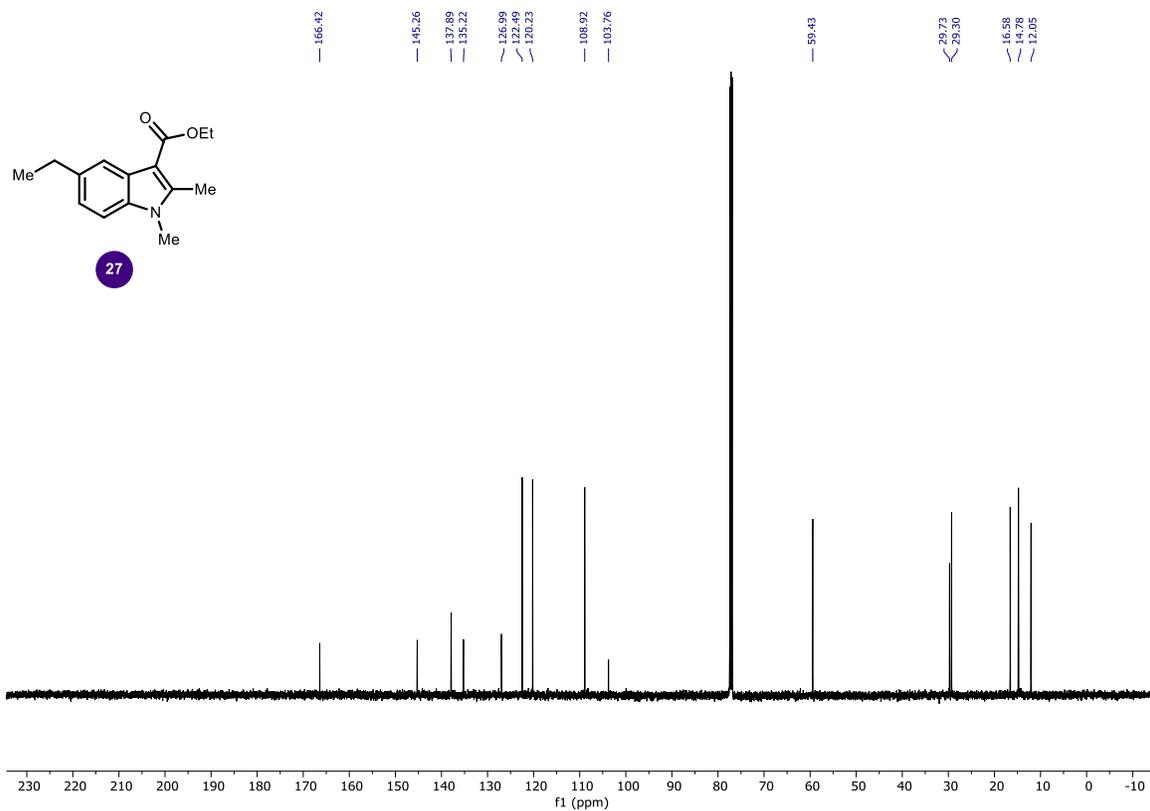
Figure S14. ^{13}C NMR of **27** in CDCl$_3$ at 25 ºC.

Figure S15. ^{1}H NMR of **30** in CDCl$_3$ at 25 ºC.
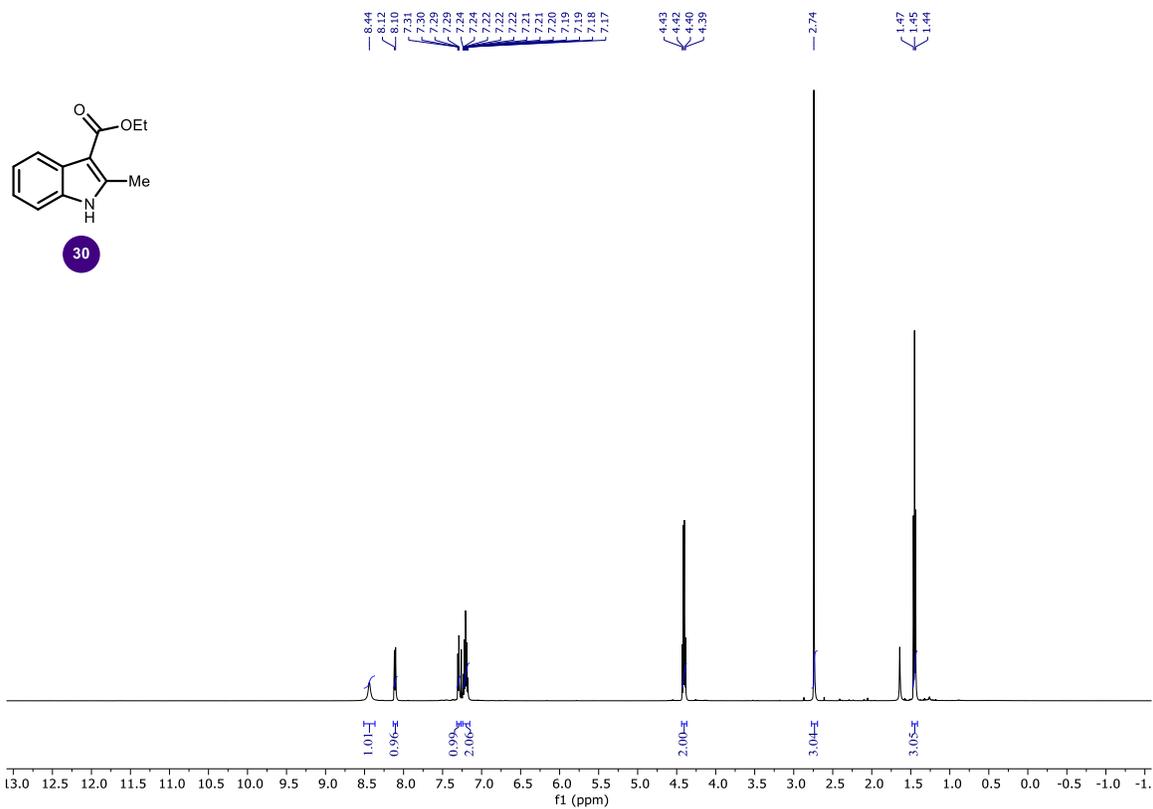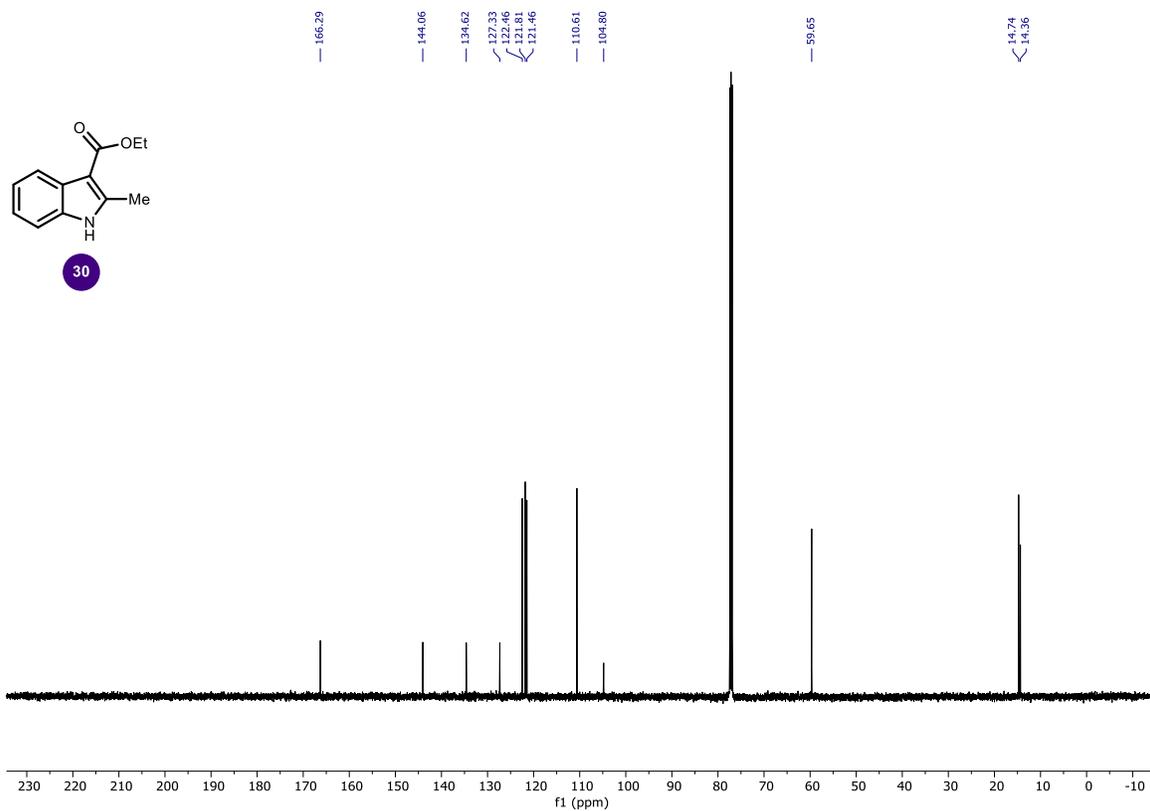
Figure S16. ^{13}C NMR of **30** in CDCl$_3$ at 25 ºC.
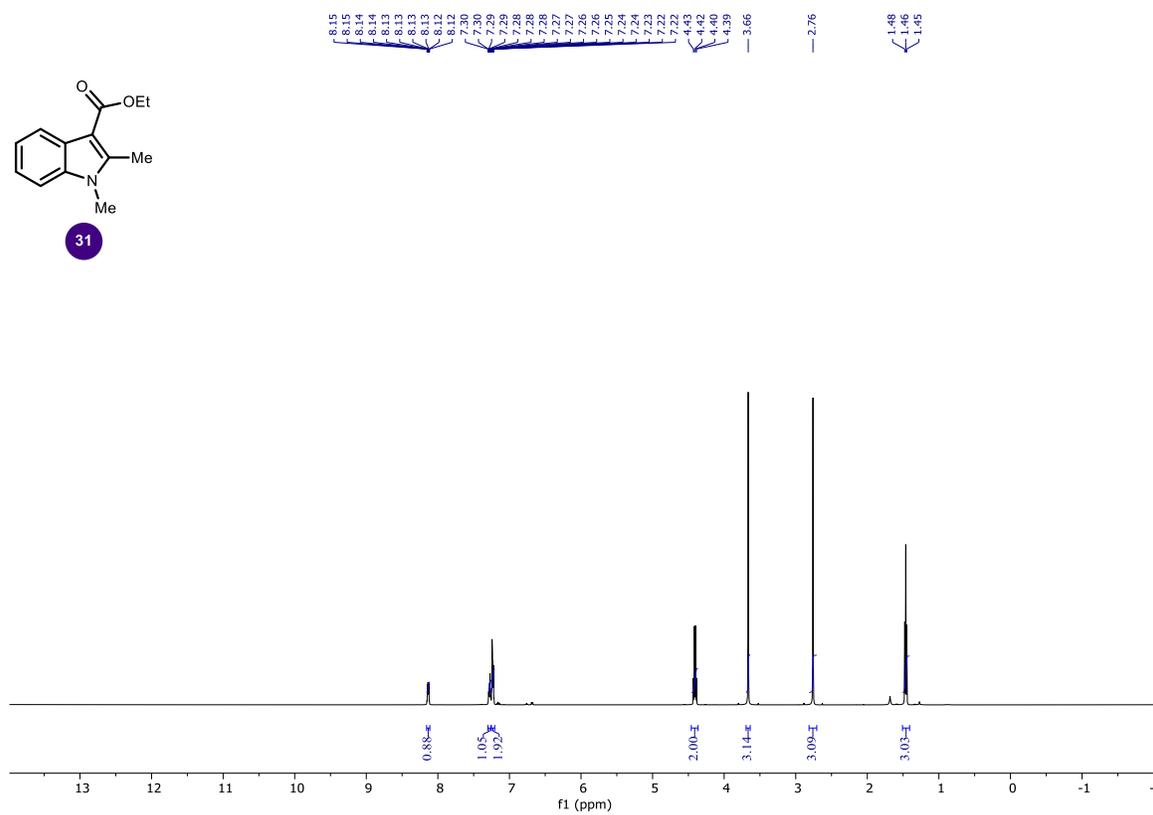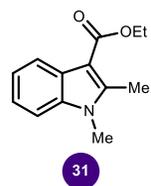
Figure S17. ^{1}H NMR of **31** in CDCl$_3$ at 25 °C.
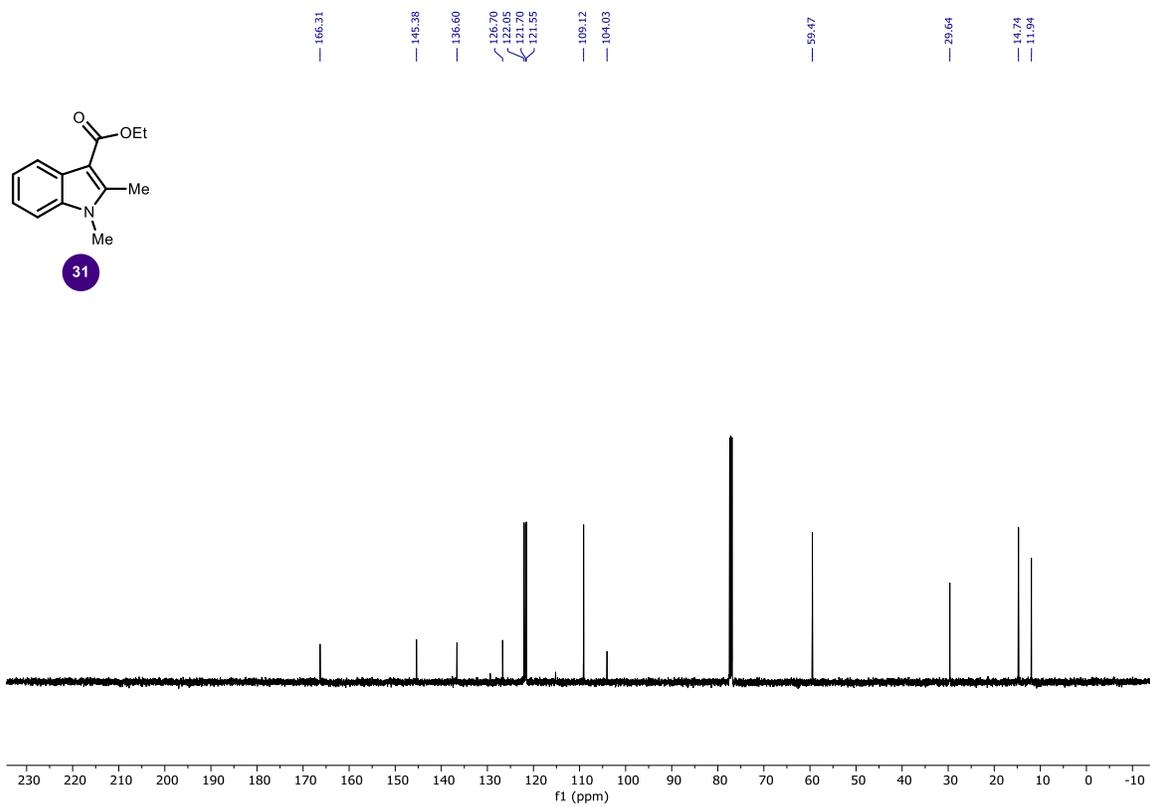
Figure S18. ^{13}C NMR of **31** in CDCl$_3$ at 25 ºC.
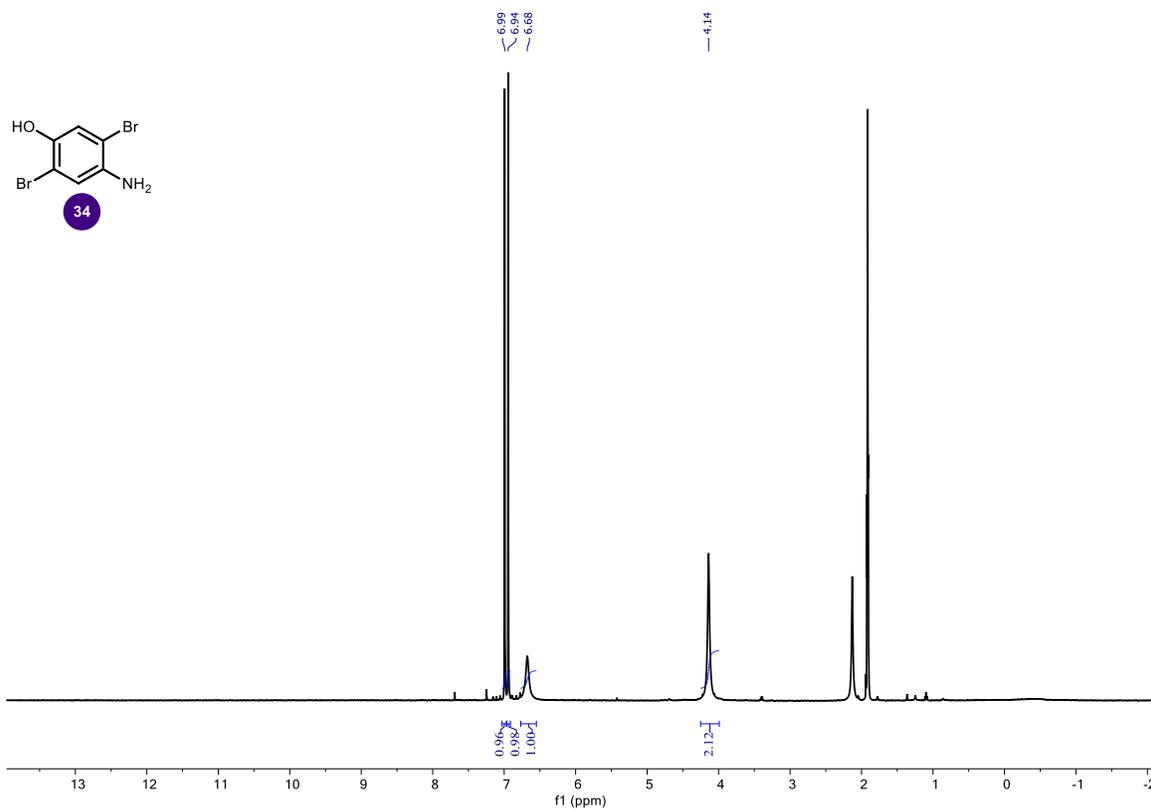
Figure S19. ^{1}H NMR of **34** in CD$_3$CN at 25 ºC.
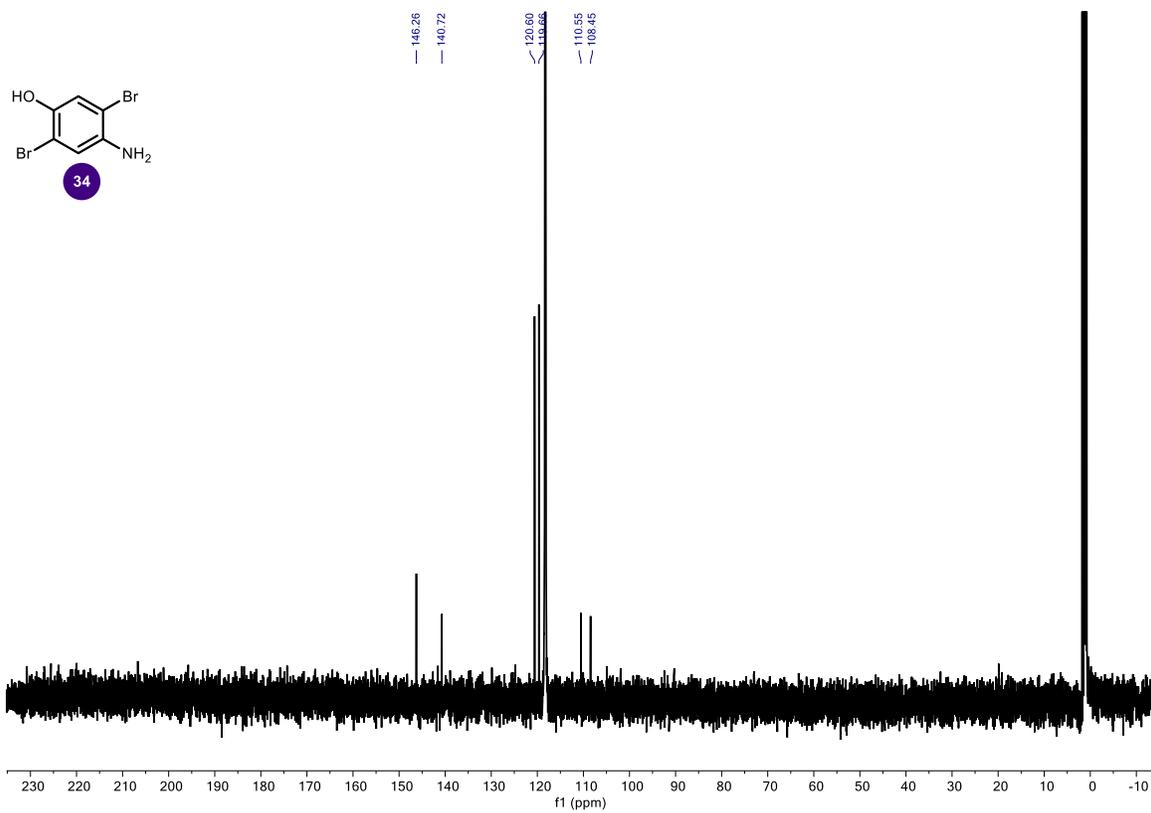
Figure S20. ^{13}C NMR of **34** in CD$_3$CN at 25 ºC.
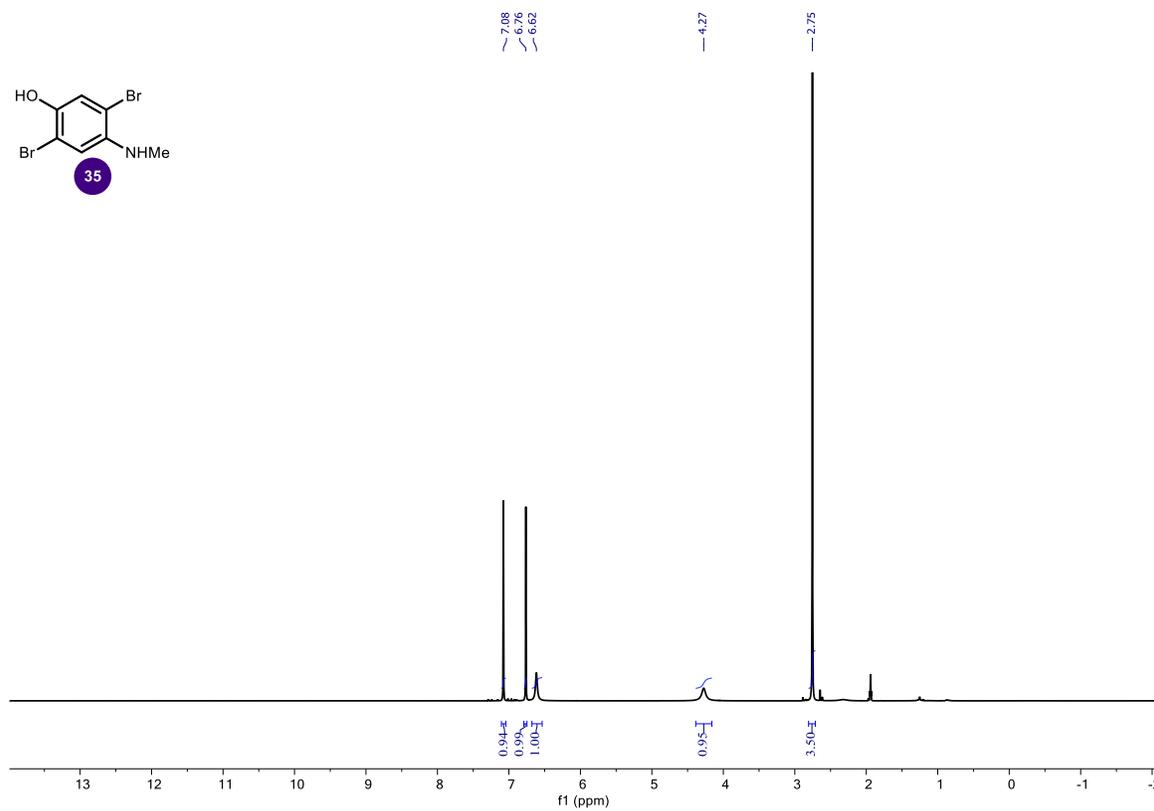
Figure S21. ^{1}H NMR of **35** in CD$_3$CN at 25 ºC.

Figure S22. ^{13}C NMR of **35** in CD$_3$CN at 25 ºC.
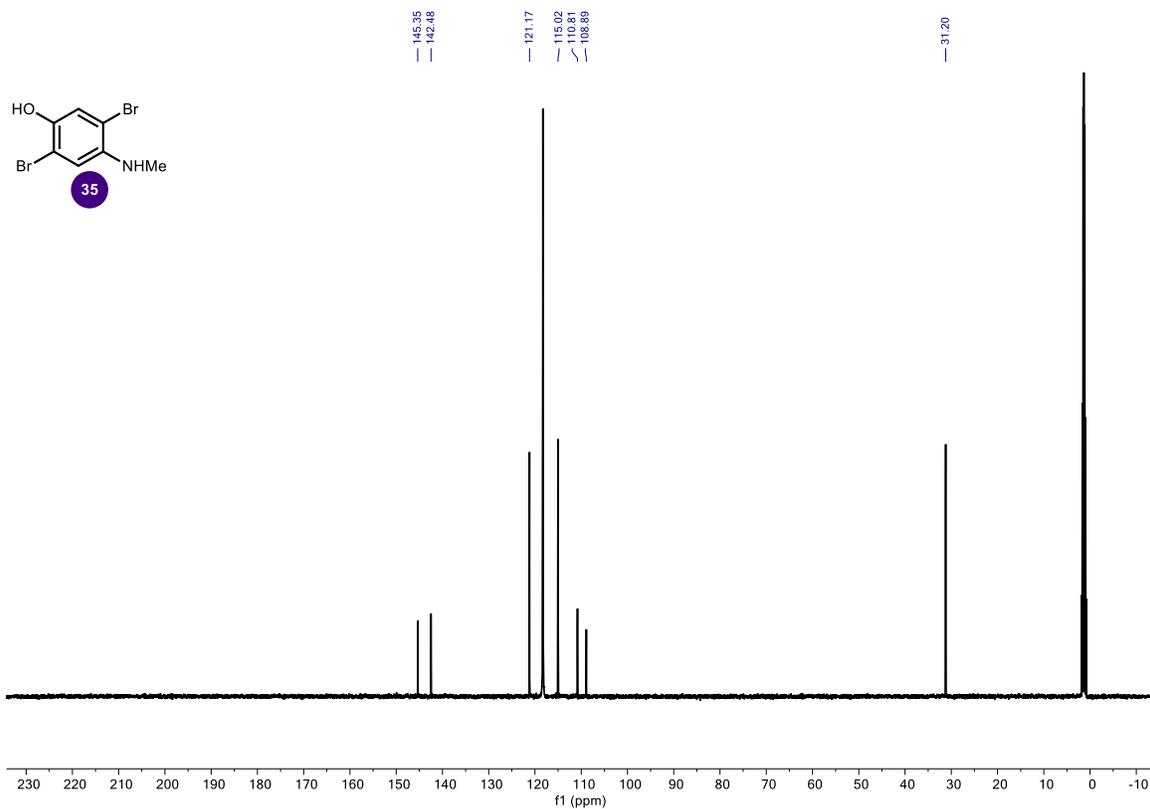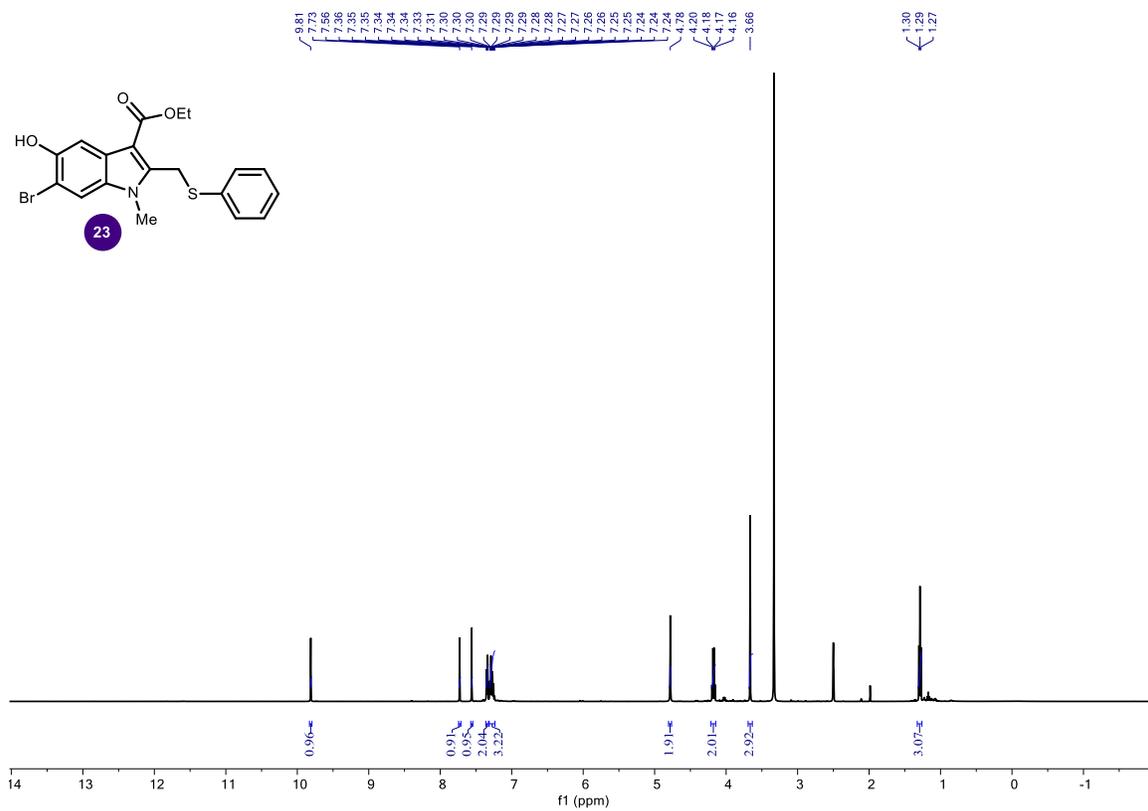
Figure S23. ^{1}H NMR of **23** in $(CD_3)_2SO$ at 25 ºC.
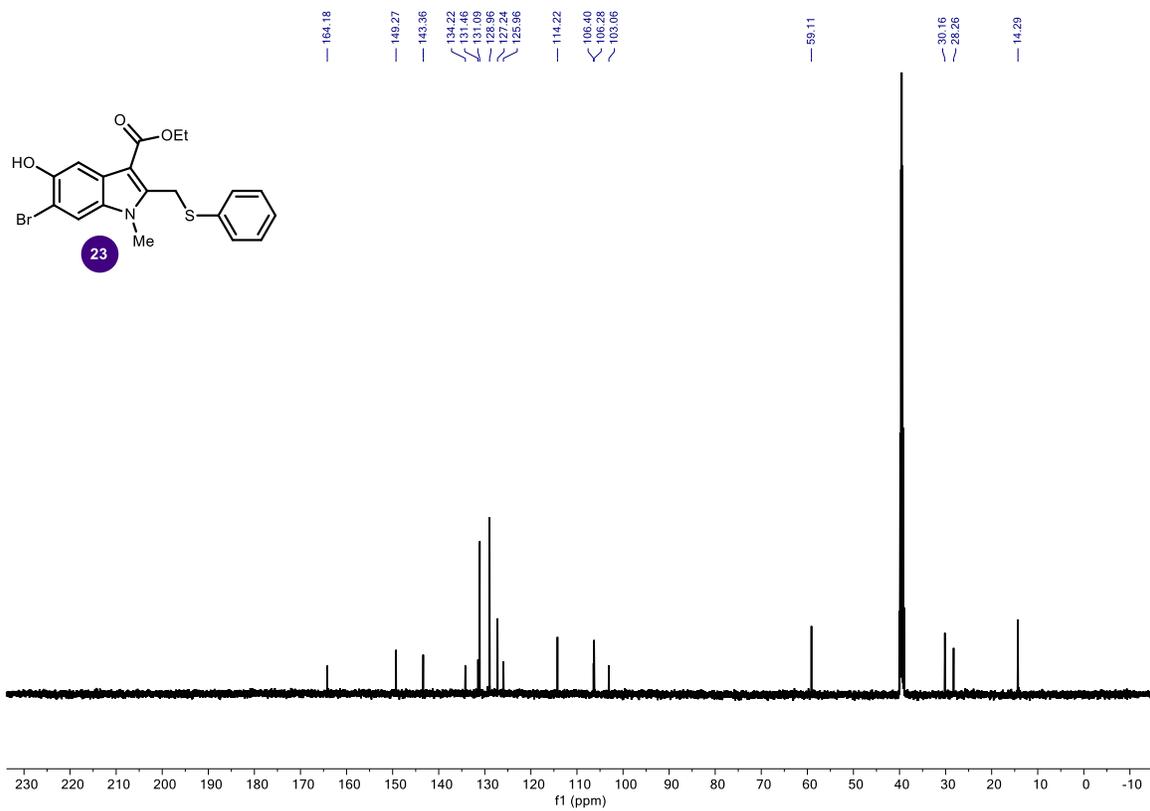
Figure S24. ^{13}C NMR of **23** in $(CD_3)_2SO$ at 25 ºC.
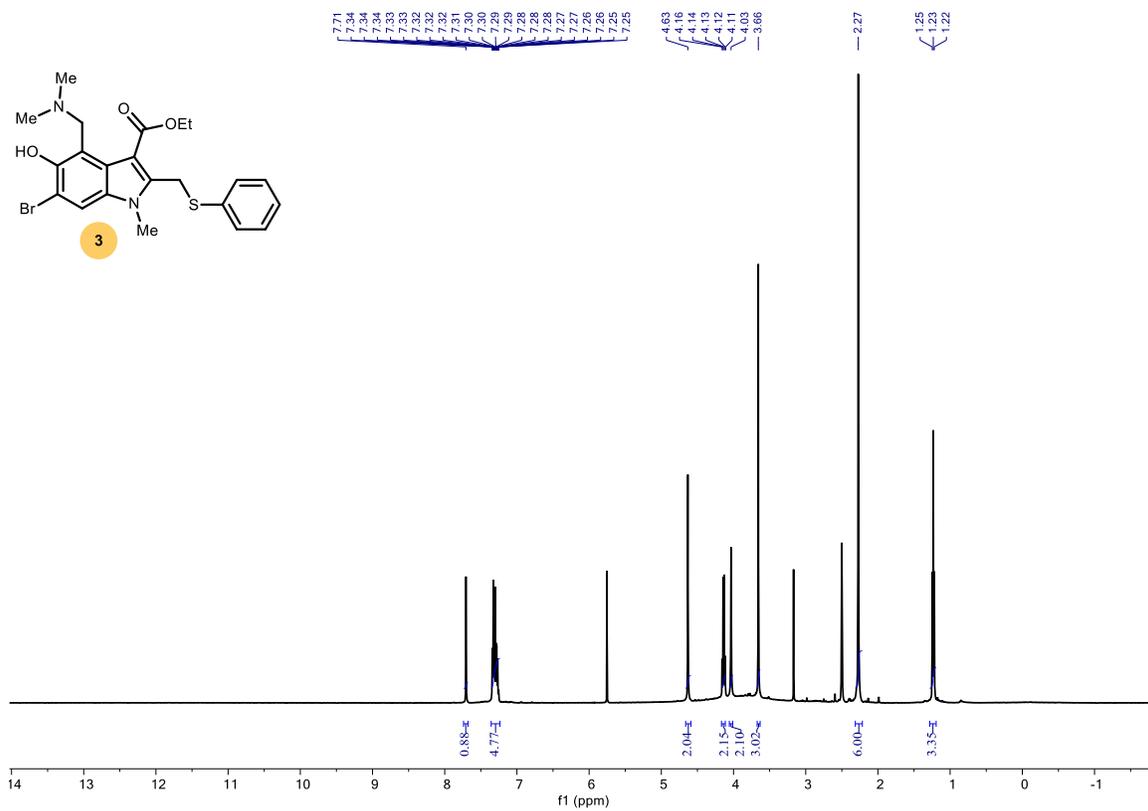
Figure S25. ^{1}H NMR of **3** in $(CD_3)_2SO$ at 25 ºC.
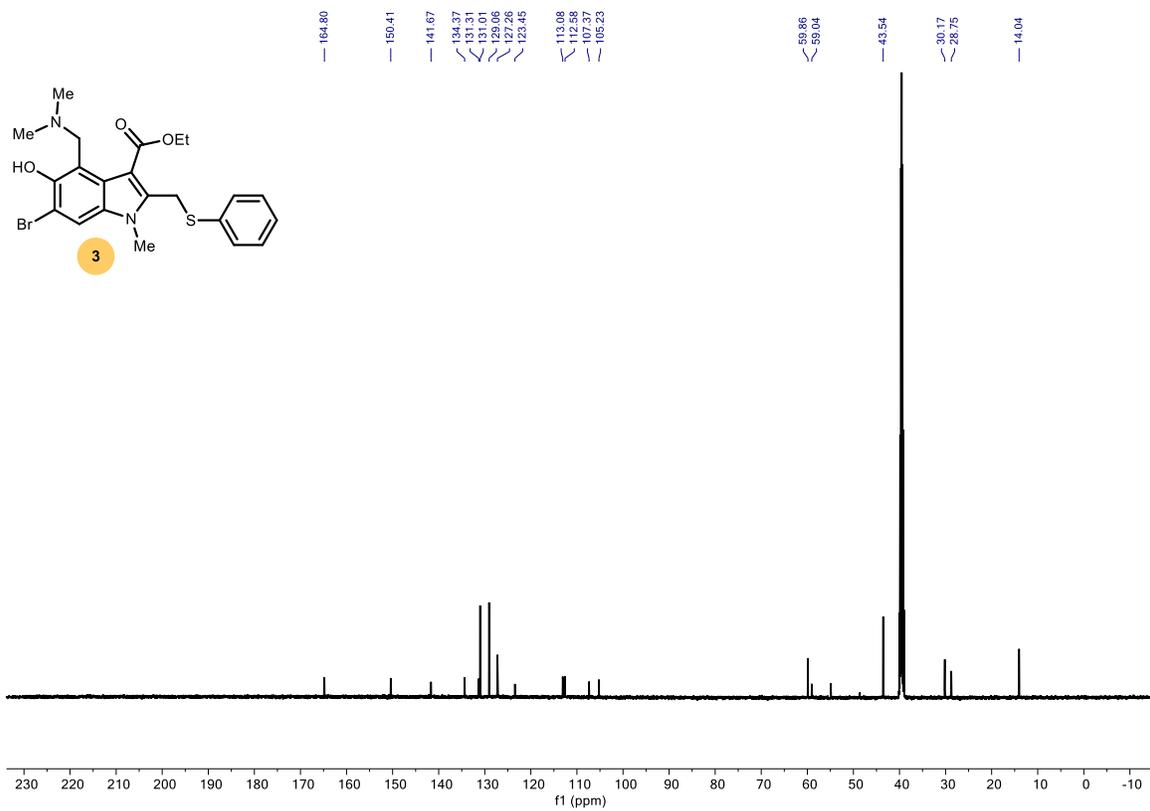
Figure S26. ^{13}C NMR of **3** in $(CD_3)_2SO$ at 25 ºC.

Figure S27. ^{1}H NMR of **2** in CDCl$_3$ at 25 ºC.
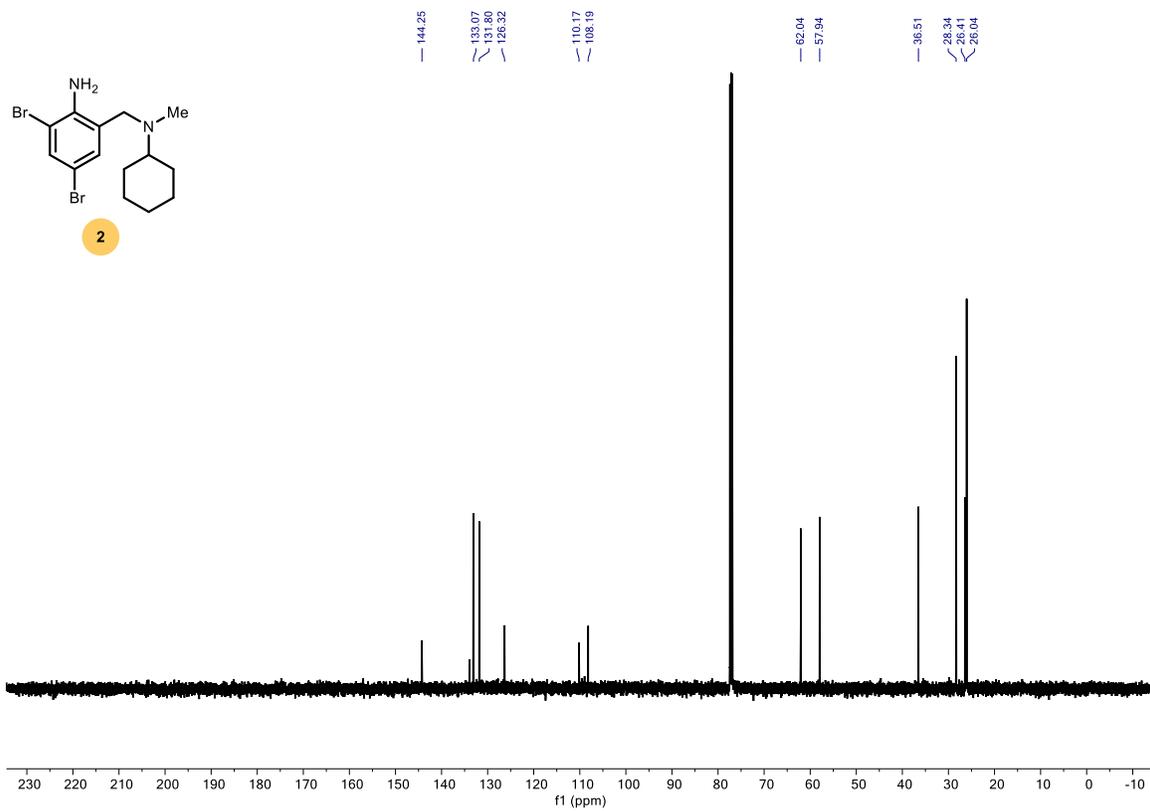
Figure S28. ^{13}C NMR of **2** in CDCl$_3$ at 25 ºC.