Combining Generative Artificial Intelligence and On-Chip Synthesis for De Novo Drug Design

Francesca Grisoni, Berend Huisman, Alexander Button, Michael Moret, Kenneth Atz, Daniel Merk, Gisbert Schneider

Submitted date: 29/12/2020 • Posted date: 30/12/2020
Licence: CC BY 4.0

Citation information: Grisoni, Francesca; Huisman, Berend; Button, Alexander; Moret, Michael; Atz, Kenneth; Merk, Daniel; et al. (2020): Combining Generative Artificial Intelligence and On-Chip Synthesis for De Novo Drug Design. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.13498587.v1

Automation of the molecular design-make-test-analyze cycle speeds up the identification of hit and lead compounds for drug discovery. Using deep learning for computational molecular design and a customized microfluidics platform for on-chip compound synthesis, liver X receptor (LXR) agonists were generated from scratch. The computational pipeline was tuned to explore the chemical space defined by known LXRα agonists, and to suggest structural analogs of known ligands and novel molecular cores. To further the design of lead-like molecules and ensure compatibility with automated on-chip synthesis, this chemical space was confined to the set of virtual products obtainable from 17 different one-step reactions. Overall, 25 de novo generated compounds were successfully synthesized in flow via formation of sulfonamide, amide bond, and ester bond. First-pass in vitro activity screening of the crude reaction products in hybrid Gal4 reporter gene assays revealed 17 (68%) hits, with up to 60-fold LXR activation. The batch re-synthesis, purification, and re-testing of 14 of these compounds confirmed that 12 of them were potent LXRα or LXRβ agonists. These results support the utilization of the proposed design-make-test-analyze framework as a blueprint for automated drug design with artificial intelligence and miniaturized bench-top synthesis.

File list (2)

| File Name | Size   |
|-----------|--------|
| Grisoni_Huisman_MS_2020.pdf | 5.45 MiB |
| Grisoni_Huisman_Supporting_Information.pdf | 6.39 MiB |
Combining generative artificial intelligence and on-chip synthesis for de novo drug design

Francesca Grisoni1*, Berend J.H. Huisman1*, Alexander L. Button1, Michael Moret1, Kenneth Atz1, Daniel Merk2,4* & Gisbert Schneider3*

1Department of Chemistry and Applied Biosciences, RETHINK, ETH Zurich, Zurich, Switzerland; 2Goethe University Frankfurt, Institute of Pharmaceutical Chemistry, Frankfurt, Germany; *These authors contributed equally; *Email: francesca.grisoni@pharma.ethz.ch, merk@pharmchem.uni-frankfurt.de, gisbert@ethz.ch

Automation of the molecular design-make-test-analyze cycle speeds up the identification of hit and lead compounds for drug discovery. Using deep learning for computational molecular design and a customized microfluidics platform for on-chip compound synthesis, liver X receptor (LXR) agonists were generated from scratch. The computational pipeline was tuned to explore the chemical space defined by known LXRα agonists, and to suggest structural analogs of known ligands and novel molecular cores. To further the design of lead-like molecules and ensure compatibility with automated on-chip synthesis, this chemical space was confined to the set of virtual products obtainable from 17 different one-step reactions. Overall, 25 de novo generated compounds were successfully synthesized in flow via formation of sulfonamide, amide bond, and ester bond. First-pass in vitro activity screening of the crude reaction products in hybrid Gal4 reporter gene assays revealed 17 (68%) hits, with up to 60-fold LXR activation. The batch re-synthesis, purification, and re-testing of 14 of these compounds confirmed that 12 of them were potent LXRα or LXRβ agonists. These results support the utilization of the proposed design-make-test-analyze framework as a blueprint for automated drug design with artificial intelligence and miniaturized bench-top synthesis.

Rapid iteration of the molecular design-make-test-analyze (DMTA) cycle has the potential for making ‘better decisions faster’,1,2 with numerous applications in drug discovery and related fields.3,4 Recent advances in chemical reaction monitoring and optimization, computing hardware, and algorithms have boosted the automation of several parts of the drug discovery process, such as robotic synthesis,5–7 computational molecular design8–10 and synthesis planning.11–13 Standardized experimental procedures with robotic assistance increase the reproducibility of results, reduce errors, and decrease the consumption of materials, thereby contributing to ‘green chemistry’.14 Furthermore, reasoning with machine intelligence supports the discovery of novel druglike molecules by freeing the molecular design and optimization process from personal biases.1 Computer-assisted structure generation plays a seminal role in this automation process. Generative deep learning models extend the capabilities of rule-based de novo molecule generators by sampling new molecules from a latent chemical space representation,15–18 without the need for human-crafted construction rules. Recent prospective studies have demonstrated the practical applicability of generative deep learning for de novo molecular design.8,9,19–21

Herein, a recently published generative deep learning model22 was adapted to generate compounds that are synthesizable on a bench-top microfluidic synthesis platform.14,23 We challenged this automated DMTA pipeline to design liver X receptor (LXR) agonists from scratch, with minimal human interference. LXRαs have emerged as promising drug
targets because of their regulatory role in lipid metabolism and inflammation, thereby causing increased reverse cholesterol transport and reduction of atherosclerosis.\textsuperscript{24–27} With 28 molecules successfully synthesized and validated for LXR activation \textit{in vitro}, this present study pioneers the integration of generative artificial intelligence and automated synthesis by designing and experimentally testing the highest number of molecules reported thus far. The proposed modular framework has the potential to speed up the DMTA cycle, thereby addressing one of the main bottlenecks of the pre-clinical drug discovery process.\textsuperscript{28,29}

Results and discussion

Modular design-make-test-analyze platform

The automated molecular design pipeline was composed of three modules (Figure 1):

\textit{Module 1:} A generative deep learning model\textsuperscript{22} based on a recurrent neural network with Long-Short Term Memory (LSTM) cells.\textsuperscript{30} LSTM models were employed for the design of new molecules represented as Simplified Molecular Input Line Entry Systems (SMILES)\textsuperscript{31} strings.\textsuperscript{15,16,32} This LSTM-based ‘chemical language model’ served as the \textit{de novo} structure generator (Figure 1a).

\textit{Module 2:} A virtual reaction filter that captured 17 one-step reactions that were compatible with the microfluidics system (module 3). These reactions were encoded as SMILES ARbitrary Target Specification (SMARTS)\textsuperscript{33} strings. This filtering module selected those generated molecules that were synthetically compatible within the microfluidics platform (Figure 1b).

\textit{Module 3:} A microfluidics platform designed to minimize the amount of manual labor needed to optimize reaction conditions and synthesize small compound libraries. This compact bench-top system combined the automated retrieval of required reagents with the optimization of reaction conditions, online reaction product monitoring by high-performance liquid chromatography-mass spectrometry (HPLC-MS), and the collection of the reaction mixtures (Figure 1c).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Schematic of the modular molecular design pipeline. (a) A generative deep learning model\textsuperscript{22} based on a long-short term memory network\textsuperscript{26} was used to generate putative liver X receptor-\textalpha{} agonists. The two-step network training procedure first trained the model on 656,689 compounds predicted as compatible with the microfluidics system, and then fine-tuned this pre-trained model with 41 known LXR\textalpha{} agonists. (b) The \textit{de novo} generated molecules were filtered on the basis of their predicted synthesis route, using a set of 17 reactions specified in the SMARTS notation ('virtual reaction filter'). Molecular building blocks for synthesis were automatically retrieved from a commercial supplier catalog. (c) A total of 41 \textit{de novo} designs were retained for synthesis on a microfluidics platform, which contained two syringes for the handling of reagents R1 and R2, a Cetoni Qmix element equipped with a Dean Flow microfluidic reactor chip, and a Rheodyne MRA splitter for automated sample transfer to an HPLC-MS system.}
\end{figure}
De novo molecular design with artificial intelligence

The deep learning model (Figure 1a) was pre-trained on the SMILES strings of 656,689 commercially available molecules from four compound vendors (for details, see Materials and Methods), which were predicted by the virtual reaction system to be suitable for the on-chip synthesis within our microfluidics platform. This pre-training step enabled the model to capture the syntax of the SMILES strings. After pre-training, 83 ± 2% of the generated SMILES strings were unique and chemically valid (average of three repetitions, 3000 SMILES strings sampled each time), and corresponded to novel molecules that were not included in the pre-training set. Compared to model pre-training performed in previous studies using bioactive molecules from the ChEMBL database,8,22,34 the approach used here resulted in a significantly higher proportion of de novo molecules that were considered synthesizable in flow (p < 0.001, Kruskal-Wallis test, Figure 2a). This result highlights the capability of the deep learning model to implicitly learn the desired molecular features (here: compound synthesizability as defined by the virtual reaction filter), without the need for explicit, rule-based design constraints.

Figure 2. Automating de novo design with deep learning. (a) Number of de novo designs retained by the virtual reaction filter depending on the utilized pre-training (mean and standard deviation over three replicates and 3000 sampled SMILES strings each). Compared to previous studies using bioactive molecules from ChEMBL8,22,32 (dashed lines) this pre-training strategy (solid line) led to a larger number of compounds retained by the virtual reaction tool (p < 0.001, Kruskal-Wallis test), with up to 255 ± 97 more designs retained in each fine-tuning epoch. The epochs chosen for sampling are highlighted (epochs 15 to 20, grey rectangle). (b) Relative scaffold diversity (i.e., unique scaffolds/total number of scaffolds) of the de novo designs before and after applying the virtual reaction filter. No statistically significant difference in scaffold diversity was observed (Wilcoxon test, α = 0.05). (c) Analysis of 67 de novo designs retained for potential synthesis: 14 compounds were patented LXR modulators annotated in SureChEMBL or Reaxys (22%); 15 compounds existed in PubChem, of which 10 compounds are annotated as commercially available (15% of the total); 4 (6%) compounds lack vendor information; and 2 compounds (3%) were known LXR modulators annotated in ChEMBL27 (IC/EC50 ≤ 2 µM). 37 compounds (55%) were not found in either PubChem, ChEMBL27, SciFinder, SureChEMBL, or Reaxys databases.
After pre-training, the model was fine-tuned with the SMILES strings of 41 LXRα agonists (EC50 < 0.5 μM, Supporting Table S1). This fine-tuning step allowed us to focus the model on features that are shared by a chosen set of compounds. Therefore, it was used to bias the generation of new SMILES strings toward the chemical space of known LXRα agonists. Both the number of fine-tuning epochs for molecule design and the sampling temperature for SMILES string generation were automatically determined to optimize three parameters simultaneously (Supporting Figure S1), namely: the (i) predicted LXR activity, (ii) scaffold diversity, and (iii) pharmacophore similarity to the fine-tuning compounds. From the fine-tuned model (epochs 15 to 20), 3000 SMILES strings were sampled per epoch. Only those generated molecules were retained that were not included in the pre-training and the fine-tuning sets, namely, a total of 3626 de novo designs.

The retrosynthetic route of each generated molecule was predicted using the chosen set of 17 virtual reaction schemes (Figure 1b). The compounds that could be decomposed into suitable reactants were kept (1911 designs). Notably, with a relative scaffold diversity equal to 23% both before and after reaction filtering, this filtering step did not drastically alter the scaffold diversity of the designs. In general, no statistically significant decrease in the relative scaffold diversity due to the application of the virtual reaction filter was observed (Wilcoxon test, α = 0.05, Figure 2b), rendering this method suitable for the design of reaction-focused compound libraries. Whenever the predicted reaction product was compatible with the microfluidics system (i.e., potentially synthesizable following one of the 17 selected reaction schemes), the predicted reactants were automatically retrieved from the Sigma Aldrich catalog extracted from PubChem (2019.02.27). For 67 designs, all of the required reactants were available.

A novelty check of the remaining 67 molecules was performed in PubChem, ChEMBL27, SciFinder, SureChEMBL, and Reaxys compound databases. Of the 67 designs, 17 molecular structures corresponded to patented or otherwise known LXR agonists with EC50 values ranging from 0.2 nM to 2 μM. This result indicated that the deep learning model correctly captured the relevant molecular features for LXR binding and activation. For the remaining 51 de novo designs (Table S5), no information on bioactivity for LXR was available. Of this compound set, 37 molecules were novel, 10 compounds were commercially available, and 4 were described in the PubChem database but were unavailable for purchase (Figure 2c). Overall, 41 compounds were selected for synthesis and three were purchased, while the remaining compounds were discarded due to unavailability or the high price of the respective building blocks.

**Microfluidics-assisted synthesis ‘on-chip’ and first-pass screening**

The 41 selected de novo molecules were synthesized in flow using computationally suggested reactions. Of these compounds, 21 were predicted to be synthesizable by sulfonamide formation, 19 by amide bond formation, and compound 21 by ester bond formation (Table S5). On the basis of the respective HPLC-MS mass peaks, a total of 25 compounds were successfully synthesized on the microfluidics platform (1–25, Figure 3a), corresponding to a 61% success rate. Compounds 26–28 were purchased.

Compounds 1–28 were subjected to preliminary testing for LXRα and LXRβ activation in hybrid Gal4 reporter gene assays with HEK293T cells. This assay relies on chimeric transcription factors composed of the respective human nuclear receptor ligand-binding domain and the DNA-binding domain of the yeast protein Gal4. Gal4-responsive firefly luciferase served as the reporter gene, and a constitutively expressed Renilla luciferase was used for normalization and to monitor the toxicity of the test compound. This assay also captured the cell penetration and cytotoxicity of the test compounds.
Crude samples of all 28 test compounds were analyzed at a single concentration for LXRα and LXRβ activation, with two independent biological duplicates (Figure 2b). The test compound concentration was roughly adjusted to 10 µM based on the HPLC traces of the samples.

Figure 3. Compound structures and first-pass in vitro screening results. (a) Compounds 1–25 were synthesized in flow and compounds 26–28 were purchased. (b) LXRα and LXRβ activation by compounds 1–28, as determined by the hybrid Gal4 reporter gene assays on the crude reaction products (test concentration ~10 µM, n = 2 with two technical replicates each). The numbers and color intensity indicate the fold-activation of LXRα and LXRβ by each compound.
Compounds 1–17 displayed ≥ 3-fold LXR activation in this preliminary screening, potentially corresponding to up to 68% actives among the 25 synthesized molecules. Compounds 6 (52-fold LXRα activation) and 15 (60-fold LXRβ activation) exhibited the strongest response in the primary screening. All compounds showing more than two-fold LXR activation possessed a hexafluoro-2-phenyl-isopropanoyl moiety, suggesting particular relevance of this molecular feature for the observed bioactivity. This feature was also present in 12 of the fine-tuning compounds (29%), with an additional 4 (10%) and 8 (20%) fine-tuning compounds possessing a hexafluoro-2-aryl-isopropanoyl moiety and an aryl-trifluoromethyl motif, respectively. All compounds showing more than 10-fold LXRα activation in the preliminary screening were selected for full characterization of the dose-response curve. Compound 7 was excluded because of its cytotoxicity. Compound 1 (two-fold LXRα activation) was included in the follow-up study because of its novel atomic scaffold,37 which is not present in any molecule annotated for LXRs in ChEMBL27 or in a repository for nuclear receptor bioactivity.46

Bioactivity determination
The selected 14 compounds were prepared in-batch (Supporting Scheme S1), purified, and fully characterized on LXRα and LXRβ (Table 1). Of these compounds, only compounds 2 and 3 were not confirmed to be active in the follow-up screening, suggesting that some other components in the crude reaction product mixture had activated LXRα in the primary screening. This finding indicates that the hexafluoro-2-phenyl-isopropanoyl moiety is not sufficient for LXR activation, despite its ubiquity among de novo designs. The potencies of the remaining 12 LXR modulators were in the range of EC50 = 0.18 – 4.5 µM for LXRα and EC50 = 0.34 – 4.0 µM for LXRβ. In agreement with the primary screening data, compound 6 displayed the highest potency on LXRα, with an EC50 of 0.183 ± 0.006 µM and 32-fold maximum activation. Compound 15 was confirmed as the most potent LXRβ agonist, with an EC50 of 0.34 ± 0.02 µM and 38-fold maximum receptor activation.

Table 1. LXRα and LXRβ modulatory potency of compounds selected for Phase II, as determined in cellular Gal4-based hybrid reporter gene assays. EC50 values and fold activation are reported as mean ± standard error (n = 3, n.d. = not determined).

| ID | LXRα         | LXRβ         | Selectivity for LXRα |
|----|--------------|--------------|----------------------|
|    | EC50 (µM)    | Fold activation | EC50 (µM)      | Fold activation | [LXRβ / LXRα] |
| 1  | 4.5 ± 0.1    | 20.5 ± 0.2   | >10                 | n.d.           | >2.2          |
| 2  | >10          | n.d.         | >10                 | n.d.           | n.d.          |
| 3  | >10          | n.d.         | >10                 | n.d.           | n.d.          |
| 5  | 0.26 ± 0.01  | 18.1 ± 0.1   | 1.30 ± 0.03        | 25.3 ± 0.3    | 5.0 ± 0.2     |
| 6  | 0.183 ± 0.006| 32.4 ± 0.2   | 0.40 ± 0.01        | 23.3 ± 0.1    | 2.19 ± 0.09   |
| 8  | 1.05 ± 0.01  | 15.2 ± 0.1   | 1.72 ± 0.04        | 24.7 ± 0.3    | 1.64 ± 0.04   |
| 9  | 1.68 ± 0.03  | 20.3 ± 0.1   | 4.0 ± 0.1          | 22.2 ± 0.1    | 2.38 ± 0.07   |
| 10 | 1.19 ± 0.01  | 11.2 ± 0.1   | 3.1 ± 0.1          | 20.0 ± 0.4    | 2.61 ± 0.09   |
| 11 | 1.31 ± 0.03  | 23.0 ± 0.2   | 2.37 ± 0.02        | 19.2 ± 0.1    | 1.81 ± 0.04   |
| 12 | 0.8 ± 0.3    | 24.7 ± 0.7   | 1.08 ± 0.02        | 27.9 ± 0.3    | 1.4 ± 0.5     |
| 13 | 1.1 ± 0.1    | 13.5 ± 0.2   | 2.23 ± 0.02        | 19.0 ± 0.1    | 2.0 ± 0.2     |
| 14 | 0.30 ± 0.01  | 26.6 ± 0.1   | 1.41 ± 0.02        | 30.1 ± 0.2    | 4.7 ± 0.2     |
| 15 | 0.24 ± 0.04  | 22 ± 2       | 0.34 ± 0.02        | 38.3 ± 0.3    | 1.4 ± 0.3     |
| 17 | 0.21 ± 0.02  | 18.5 ± 0.1   | 1.25 ± 0.01        | 22.9 ± 0.1    | 6.0 ± 0.6     |
Bioactive compounds 1, 13, and 15 possessed novel atom scaffolds (‘Murcko scaffolds’) compared to the LXRα and LXRβ agonists with EC50 ≤ 50 µM annotated in the ChEMBL27 database (Figure 4a). Furthermore, the scaffolds of compounds 1 and 13 were not present in any of the 15,247 molecules annotated in the Nuclear Receptor Activity (NURA) dataset. Compound 15, the most potent LXRβ agonist, had the lowest fragment similarity to the fine-tuning molecules, as captured by Tanimoto similarity on Morgan molecular fingerprints (Supporting Table S6). These results corroborate the capacity of the computational pipeline to explore narrow regions of the chemical space defined by the known LXR agonists, while at the same time providing hitherto unexplored molecular cores for further compound optimization.

All the confirmed active compounds had a preference for LXRα over the LXRβ subtype (Table 1). This observation reflects the desired effect of fine-tuning the artificial intelligence model with LXRα modulators. Compounds 17 and 5 showed the highest LXRα selectivity, with activity five to six times greater than their LXRβ activation (Table 1). Compound 17 activated LXRα less potently than its closest structural relative among the fine-tuning compounds (EC50 < 0.1 µM, Supporting Table S1). Compound 5 (LXRα: EC50 = 0.26 ± 0.01 µM; LXRβ: EC50 = 1.30 ± 0.03 µM) constituted only a minor structural modification of one of the fine-tuning compounds (compound 29, Figure 4b), which had an EC50 of 0.4 µM on LXRα and an EC50 of 0.18 µM on LXRβ. While compounds 5 and 29 were comparable in their agonistic effects on LXRα, compound 5 was more than seven times less potent than 29 on the beta subtype. As suggested by automated ligand-receptor docking, the preference for LXRα could be ascribed to the different positioning of compounds 5 and 29 in the LXRβ binding pocket (Figure 4c). While compounds 5 and 29 are predicted to adopt similar binding poses within the LXRα pocket (RMSD < 1.8 Å), compound 29 was predicted to engage in an additional CH3-π interaction in the binding

![Figure 4](image-url)
pocket of LXRβ. This hypothesis potentially explains the greater affinity of compound 29 for LXRβ as compared to that of compound 5.

**Conclusions**

With 61% of the computational designs successfully synthesized and 12 novel LXR agonists with low-micromolar to nanomolar activity identified, the integrated de novo design platform shows promise for automating the DMTA cycle in drug discovery. By tailoring the model optimization to the available experimental pipeline, the benefits of ‘rule-free’ de novo design, virtual reaction specification, and automated synthesis were combined. The results further validated the ability of generative molecular design approaches to capture desired molecular properties such as chemical synthesizability and on-target bioactivity, and their potential to support automation. The proposed DMTA framework offers the promise of fast iterations through the molecular design cycle and data-driven compound optimization. Owing to its modular character, the approach can be tailored to other de novo design applications by replacing the computational molecule generator, reactions used for filtering, or synthesis technology. For the purpose of this proof-of-concept study, computational tools were used to obtain a structurally balanced set of bioactive hits, as demonstrated in the design of both the analogs of known LXRα agonists and novel molecular scaffolds. The three novel bioactive scaffolds (with EC_{50} values ranging from 0.24 to 4.5 µM) highlight the capability of the computational pipeline to explore narrow regions of the relevant chemical space, while at the same time providing unexplored molecular cores. To further explore the permissive structural diversity of the designs, several strategies can be adopted; for example, choosing different optimization criteria for the generative deep learning model, or including other artificial intelligence models in the definition of compatible organic reactions. The proposed approach could also fuel active learning, by adding a feedback loop connecting the experimental readout with the de novo molecule generator to refine the deep learning model for iterative molecular design and automated optimization.

**Methods**

**Computational**

Virtual reaction filter. A total of 17 decomposition reactions written in the SMARTS language (in the form “product >> reactant”) were adapted from a recent study to capture feasible synthetic routes for the microfluidics-assisted platform (see Supporting Information, Table S3).

Training data. The pre-training library was obtained from a dataset of 3,383,942 commercially available synthetic compounds, assembled from four providers: Asinex (http://www.asinex.com/libraries-html/ – Elite, Fragments, Gold and Platinum collections), ChemBridge screening compound collection (http://www.chembridge.com), Enamine advanced and HTS collections (http://www.enamine.net), and Specs screening compounds (https://www.specs.net). The library was filtered using the designed reaction SMARTS, leading to a set of 656,689 molecules retained for model pre-training. Fine-tuning was performed by utilizing a set of 41 LXRα agonists (Supporting Table S1) collected from ChEMBL26 (target ID
Molecule pre-processing. The molecular structures were standardized with the MOE “wash” procedure (MOE v.2018.01, default settings) before computing molecular descriptors and performing target prediction. As in our previous study, prior to LSTM training, molecular structures were encoded as canonical SMILES strings using the RDKit package (v.2018.03, www.rdkit.org); stereochemical information was removed and only SMILES strings with a length of up to 140 SMILES characters were retained.

Target prediction and molecular descriptors. Target prediction was performed with the in-house SPIDER software, using MOE2D and Chemically-Advanced Template Search 2 (CATS2) descriptors as input. Only predictions with p < 0.05 were considered. CATS2 descriptors were calculated using in-house software (settings: CorrelationDistance = 10, Scaling = Types, Distance = Euclidean). MOE2D descriptors were calculated with the ‘QSAR descriptors’ node of MOE 2018.01 in a KNIME 3.7.0 environment (charge calculation = MMFF94*).

Model architecture and settings. The chemical language model was implemented in Python (v. 3.6.5) using Keras (https://keras.io/, v2.2.0) with the TensorFlow GPU backend (www.tensorflow.org, v1.9.0) as a recurrent neural network with LSTM cells, as previously published. The neural network used consisted of four layers, for a total of 5,820,515 parameters: (1) BatchNormalization layer; (2) LSTM layer with 1024 units; (3) LSTM layer with 256 units; (4) BatchNormalization layer. The model was trained with SMILES strings encoded as one-hot vectors. SMILES randomization and 10-fold augmentation, as recently published, were used. We used the categorical cross-entropy loss and the Adam optimizer with a learning rate equal to 0.002. After pre-training for 10 epochs, the model was fine-tuned for 40 epochs. The model code can be accessed at the following URL: https://github.com/ETHmodlab/virtual_libraries.

Temperature and sampling epoch choice. Sampling temperature and fine-tuning epochs were automatically determined to optimize three parameters simultaneously: (i) predicted LXR activity by SPIDER, (ii) scaffold diversity, and (iii) distance to the fine-tuning compounds (as encoded by CATS2 descriptors with Euclidean distance). We tested three sampling temperatures (T = 0.2, T = 0.7, T = 1.2, Supporting Figure S1) and fine-tuning epochs 1 to 40 (Supporting Figure S2). Sampling temperature T = 0.70 and fine-tuning epochs in the range 15 – 20 were chosen to generate the final designs, as they resulted in the best compromise between scaffold diversity, number of compounds predicted as LXR modulators (p < 0.05), and CATS2 similarity between the fine-tuning set and de novo design.

Novelty and scaffold analysis. The 67 de novo designs retained after reaction-based filtering were checked for their structural novelty on PubChem, ChEMBL27, SciFinder (version 2019; Chemical Abstracts Service), SureChEMBL, and Reaxys (accessed 4 April 2019). Atomic scaffolds were computed using the RDKit in KNIME (v. 3.6.2).

Similarity analysis. ChEMBL27 compounds were structurally similar to the bioactive hits (as determined by the Jaccard-Tanimoto similarity on RDKit Morgan fingerprints with radius equal to 2 and 1024 bits, Supporting Table S5) and retrieved using the ChEMBL web resource client ‘similarity.filter’ function (https://github.com/chembl/chembl_webresource_client, beta version, 11.06.2020, Python 3.7.7). The same strategy was used to report the two most similar fine-tuning compounds for each bioactive hit (Supporting Table S6).
Automated ligand docking. The crystal structures of LXRα (PDB ID: 3IPS) and LXRβ (PDB ID: 1PQC) were retrieved from the Protein Data Bank (https://www.rcsb.org/) and prepared with MOE v.2019.0102 (QuickPrep module: ‘Preserve Sequence and Neutralize’, ‘Use Protonate 3D for Protonation’ = True; ‘Delete Water Molecules Farther than 4.5 Å from Ligand or Receptor’ = True, Tether Receptor: Strength = 10, Buffer = 0.25; Fix: ‘Atoms Farther than 8 Å from Ligands’, hydrogens close to ligands not fixed; Refine: ‘to RMS Gradient of 0.1 kcal/mol/Å’, ‘Retain QuickPrep Minimization Restraints’ = True). Compounds 5 and 30 were docked with GOLD within MOE v.2019.0102 (Efficiency = default, Score Efficiency = 100; Early Termination = [number:3, RMS = 1.5], PLP scoring, Rigid Receptor, 30 poses per compound) and poses were refined with MOE GBVI/WSA dG (10 refinement poses). Redocking of the crystalized ligand led to root mean square deviation (RMSD) values of 0.8037 Å and 0.3775 Å for 3IPS and 1PQC, respectively.

Synthesis

Chemicals

All chemicals and solvents were reagent grade and used without further purification unless specified otherwise. The building block chemicals were purchased from Sigma-Aldrich (St. Louis, USA, www.sigmaaldrich.com), Apollo Scientific (Cheshire, United Kingdom, www.apolloscientific.co.uk), Alfa Aesar (Kandel, Germany, www.alfa.com), Fluorochem (Derbyshire, United Kingdom, www.fluorochem.co.uk), Acros Organics (Geel, Belgium, www.acros.com), Enamine (Riga, Latvia, www.enamine.net), Maybridge (Waltham, MA, United States, www.fishersci.com), ABCR (Karlsruhe, Germany, www.abcr.de), and ChemDiv (San Diego, CA, United States, www.chemdiv.com). Compounds 26 and 27 were purchased from Enamine (www.enamine.net, compounds Z45510435 and Z45410017, respectively; purity 90%); 28 was purchased from ChemDiv (www.chemdiv.com, compound 8012-4386, purity 90%).

Microfluidics platform

Instruments. Automated synthesis was performed on a Cetoni flow chemistry system (Cetoni GmbH, Korbussen, DE) using two gas tight borosilicate glass syringes (SGE gas tight 2.5 ml, luer lock, Trajan Scientific), a reaction chip (Chip Type Dean Flow A, 16 × 12.5 mm, DFM-A1, 5 µl) and an 800 µl reaction coil of PTFE tubing. The flow through the system was directed by three-way solenoid valves (100T3/S116, BIO-CHEM VALVE™ Inc., Chrom Tech®, Apple Valley, MN, USA). The analysis block consisted of a Rheodyne MRA splitter (MRA100-000, Kinesis, Vernon Hills, IL, USA) coupled with an Advion Expression CMS (Advion, Ithaca, NY, USA) mass spectrometer for in-line mass analysis. This equipment used L-216OU pumps from a VWR LaChrom ULTRA HPLC system (Radnor, PA, USA). An analytical HPLC system (Shimadzu, Kyoto, Japan) equipped with an analytical C18 reverse phase column (Macherey-Nagel, Nucleodur C18 HTec, 5 µm, 150 × 3 mm) was used for follow-up sample analysis. Mass signals were recorded using a Shimadzu LCMS-2020 system (Kyoto, Japan). The automated synthesis system was controlled using the QmixElements software supplied by Cetoni, on an Aspire X3990 PC (i3 Intel Core 2120 CPU, 8GB, 1066 MHz DDR3 RAM, Windows 7 OS).

Protocol. The building blocks were loaded into two 96-well plates and sealed using adhesive slit seal sheets. User input was requested for the number of reactions to be run, the desired residence time, and reaction chip and reaction loop temperatures. Upon reaching the desired reaction conditions, aspiration of the reagent solutions was initiated. Using the
Move-To-Container functions, the 360° rotAXYS arm moved to the location of the first building block and the Dose-Volume function was used to aspirate the dead volume between the syringe and the well plate. This volume was discarded, and the syringe was aspirated with a plug of dissolved building block consisting of 0.75 ml diluted with 0.25 ml of clean solvent. Before moving to the second building block container, the 360° rotAXYS arm moved to a clean solvent container using the ‘Move-To-Container’ function to clean the needle. The same aspiration procedure was used for the second syringe. The volume contained in the syringes was then injected into the reaction chip using the Dose-Volume function at a flow rate calculated from the desired residence time. Once the syringes were empty, they were refilled with 2.5 ml of clean solvent using the Dose-Volume function and solvent was injected until the remaining part of the reaction plug had passed the reaction coil with the correct residence time. Once the reaction plug had passed the analysis unit and reached the end of the tubing, the Move-To-Container function was used to move the second rotAXYS arm to an empty container, where it was allowed to remain until the reaction plug was collected. The arm then moved back to the waste container position using the Move-XY function. HPLC-MS was used as an online monitoring tool. A stream splitting device coupled the ambient-pressure reactor system in an isolated fashion to the high-pressure side required by HPLC-MS.

Automated synthesis

General amide bond synthesis procedure (compounds 1-15, 22). Solutions (0.2 M) of the respective acid chloride building block in THF and the amine building block in MeCN/DMF (9:1 v/v) were prepared, and 1 ml of each solution was loaded into individual wells of a 96-well plate. For automated synthesis, 0.75 ml of each building block solution was used per reaction and diluted with the running solvent (MeCN/THF, 50:50 v/v) to a total volume of 1.0 ml. The residence time was set to 15 min and the temperature to 55 °C. After mixing both solutions in the reaction chip, the reaction mixture was pumped through the coil and 2 ml per sample was collected in another 96-well plate.

General ester bond synthesis procedure (21). Solutions (0.2 M) of the respective acid chloride building block in THF and the amine building block in MeCN/DMF (9:1 v/v) were prepared, and 1 ml of each solution was transferred into individual wells in a 96-well plate. The residence time was set to 5 min and the temperature to 55 °C. A 0.75 ml portion of each solution was aspirated into the syringe pumps and diluted to 1 ml total volume with the running solvent (MeCN/THF, 50:50 v/v). After mixing both solutions in the reaction chip, the reaction mixture was pumped through the coil and 2 ml was collected in a 96-well plate.

General sulfonamide synthesis procedure (compounds 16–20, 23–25). Solutions (0.2 M) of the sulfonyl chloride building block in THF and the amine building block with one equivalent of triethylamine in MeCN/THF (50:50 v/v) were prepared, and 1 ml of each solution was loaded into individual wells of a 96-well plate. For automated synthesis, 0.75 ml of each building block solution was used per reaction and diluted with the running solvent (MeCN/THF, 50:50 v/v) to a total volume of 1.0 ml. The residence time was set to 10 min and the temperature to 55 °C. After mixing both solutions in the reaction chip, the reaction mixture was pumped through the coil and 2 ml per sample was collected in another 96-well plate.
Batch synthesis

Fourteen compounds were synthesized in-batch for dose-response characterization (compounds 1–3, 5–6, 8–15, 17). All compounds had a purity greater than 97% according to the HPLC-UV analysis (λ = 254 nm, λ = 290 nm). For details of the batch synthesis and analytical characterization, see the Supporting Information.

Biological characterization

The Gal4-fusion receptor plasmid (pFA-CMV-hLXRα-LBD$^{44}$ and pFA-CMV-hLXRβ-LBD$^{44}$) codings for the hinge region and ligand-binding domain of the canonical isoform of the respective nuclear receptor have been reported previously. pFR-Luc (Stratagene) was used as the reporter plasmid along with pRL-SV40 (Promega) for normalization of the transfection efficiency and cell growth. HEK293T cells were grown in high-glucose DMEM, supplemented with 10% FCS, sodium pyruvate (1 mM), penicillin (100 U ml$^{-1}$), and streptomycin (100 µg ml$^{-1}$) at 37 °C and 5% CO$_2$. The HEK293T cells were seeded one day before transfection in 96-well plates (3.0 × 10$^4$ cells per well). Before transfection, the medium was changed to Opti-MEM without supplements. Transient transfection was carried out using Lipofectamine LTX reagent (Invitrogen) according to the manufacturer’s protocol, with pFR-Luc (Stratagene), pRL-SV40 (Promega), and the respective pFA-CMV-hNR-LBD plasmid. Five hours after transfection, the medium was changed to Opti-MEM supplemented with penicillin (100 U ml$^{-1}$), streptomycin (100 µg ml$^{-1}$), 0.1% DMSO, and the respective test compounds or 0.1% DMSO alone as an untreated control. During primary screening, the concentration of each sample was roughly adjusted to 10 µM. Each sample was duplicated and tested in two biological repeats. For dose-response characterization of the purified compounds, each concentration was duplicated, and each experiment was repeated independently at least three times. Following overnight (12 – 14 h) incubation with the test compounds, the cells were assayed for luciferase activity using the Dual-Glo™ luciferase assay system (Promega) according to the manufacturer’s protocol. Luminescence was measured using a Tecan Spark luminometer (Tecan Deutschland GmbH, Germany). Normalization of the transfection efficiency and cell growth was performed by dividing the firefly luciferase data by Renilla luciferase data and multiplying the value by 1000, resulting in relative light units (RLU). Fold activation was obtained by dividing the mean RLU of the test compounds at a respective concentration by the mean RLU of the untreated control. T0901317 served as a reference agonist for assay validation and monitoring assay performance.

Author contributions

G.S. conceived the study, with contributions from F.G. and D.M.; F.G. designed the computational workflow and analysis, with contributions from A.B., M.M., and D.M.; B.H. curated the pre-training molecules; D.M. curated the fine-tuning molecules; M.M. trained the LSTM model and generated the designs; F.G. optimized the LSTM settings and performed post-hoc analysis and docking; A.B. performed the retrosynthetic analysis; B.H. developed the microfluidic synthesis platform and performed the synthesis in flow; D.M. designed the in vitro screening procedure and performed the in vitro characterization; K.A. performed the batch synthesis and purification; and F.G. drafted the manuscript, with contributions from D.M., B.H., M.M., and G.S. All authors contributed to manuscript revision.
Acknowledgements

The authors thank Sarah Haller, Benedikt Winkler, Damian Gautschi, and Jan A. Hiss for their technical support. This work was financially supported by the Swiss National Science Foundation (grant no. 205321_182176 to G.S.) and by the RETHINK initiative at ETH Zurich.

Conflict of interest statement

G.S. declares a potential financial conflict of interest as a co-founder of inSili.com LLC, Zurich, and in his role as a consultant to the pharmaceutical industry.

References

1. Schneider, G. Automating drug discovery. Nat. Rev. Drug Discov. 17, 97–113 (2018).
2. King, R. D. et al. The automation of science. Science 324, 85 (2009).
3. Chow, S., Liver, S. & Nelson, A. Streamlining bioactive molecular discovery through integration and automation. Nat. Rev. Chem. 2, 174–183 (2018).
4. Chapman, T. Lab automation and robotics: Automation on the move. Nature 421, 661–663 (2003).
5. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature 559, 377–381 (2018).
6. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science 363, (2019).
7. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science 365, eaax1566 (2019).
8. Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. De novo design of bioactive small molecules by artificial intelligence. Mol. Inf. 37, 1700153 (2018).
9. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat. Biotechnol. 37, 1038–1040 (2019).
10. Nagarajan, D. et al. Computational antimicrobial peptide design and evaluation against multidrug-resistant clinical isolates of bacteria. J. Biol. Chem. 293, 3492–3509 (2018).
11. Davies, I. W. The digitization of organic synthesis. Nature 570, 175–181 (2019).
12. Schwaller, P., Gaudin, T., Lányi, D., Bekas, C. & Laino, T. “Found in translation”: Predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem. Sci. 9, 6091–6098 (2018).
13. Johansson, S. et al. AI-assisted synthesis prediction. Drug Discov. Today Technol. (2020) doi:10.1016/j.ddtec.2020.06.002.
14. Liu, Y. & Jiang, X. Why microfluidics? Merits and trends in chemical synthesis. Lab. Chip 17, 3960–3978 (2017).
15. Segler, M. H., Kogej, T., Tyrchan, C. & Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent. Sci. 4, 120–131 (2018).
16. Olivecrona, M., Blaschke, T., Engkvist, O. & Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminformatics* **9**, 48 (2017).

17. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).

18. Ikebata, H., Hongo, K., Isomura, T., Maezono, R. & Yoshida, R. Bayesian molecular design with a chemical language model. *J. Comput. Aided Mol. Des.* **31**, 379–391 (2017).

19. Merk, D., Grisoni, F., Friedrich, L. & Schneider, G. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. *Commun. Chem.* **1**, 68 (2018).

20. Bruns, D., Merk, D., Santhana Kumar, K., Baumgartner, M. & Schneider, G. Synthetic activators of cell migration designed by constructive machine learning. *ChemistryOpen* **8**, 1303–1308 (2019).

21. Yuan, W. *et al.* Chemical space mimicry for drug discovery. *J. Chem. Inf. Model.* **57**, 875–882 (2017).

22. Moret, M., Friedrich, L., Grisoni, F., Merk, D. & Schneider, G. Generative molecular design in low data regimes. *Nat. Mach. Intell.* **2**, 171–180 (2020).

23. Rodrigues, T., Schneider, P. & Schneider, G. Accessing new chemical entities through microfluidic systems. *Angew. Chem. Int. Ed.* **53**, 5750–5758 (2014).

24. El-Gendy, B. E.-D. M., Goher, S. S., Hegazy, L. S., Arief, M. M. H. & Burris, T. P. Recent advances in the medicinal chemistry of liver X receptors. *J. Med. Chem.* **61**, 10935–10956 (2018).

25. Collins, J. L. Therapeutic opportunities for liver X receptor modulators. *Curr. Opin. Drug Discov. Devel.* **7**, 692–702 (2004).

26. Li, N. *et al.* E17110 promotes reverse cholesterol transport with liver X receptor β agonist activity in vitro. *Acta Pharm. Sin. B* **6**, 198–204 (2016).

27. Hong, C. & Tontonoz, P. Liver X receptors in lipid metabolism: Opportunities for drug discovery. *Nat. Rev. Drug Discov.* **13**, 433–444 (2014).

28. Plowright, A. T. *et al.* Hypothesis driven drug design: Improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discov. Today* **17**, 56–62 (2012).

29. Schneider, P. *et al.* Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2020).

30. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).

31. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).

32. Gupta, A. *et al.* Generative recurrent networks for de novo drug design. *Mol. Inf.* **37**, 1700111 (2018).

33. Daylight Theory: SMARTS - A language for describing molecular patterns. https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (2019).

34. Grisoni, F., Moret, M., Lingwood, R. & Schneider, G. Bidirectional molecule generation with recurrent neural networks. *J. Chem. Inf. Model.* **60**, 1175–1183 (2020).

35. Awale, M., Sirockin, F., Stiefl, N. & Reymond, J.-L. Drug analogs from fragment-based long short-term memory generative neural networks. *J. Chem. Inf. Model.* **59**, 1347–1356 (2019).
36. Reker, D., Rodrigues, T., Schneider, P. & Schneider, G. Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. USA* **111**, 4067 (2014).

37. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **39**, 2887–2893 (1996).

38. Reutlinger, M. *et al.* Chemically Advanced Template Search (CATS) for scaffold-hopping and prospective target prediction for ‘orphan’ molecules. *Mol. Inf.* **32**, 133–138 (2013).

39. Kim, S. *et al.* PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2018).

40. Mendez, D. *et al.* ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2018).

41. Chemical Abstract Service, *SciFinder*, v2019, American Chemical Society, Columbus, OH, USA (2019).

42. Papadatos, G. *et al.* SureChEMBL: A large-scale, chemically annotated patent document database. *Nucleic Acids Res.* **44**, D1220–D1228 (2015).

43. Lawson, A. J., Swienty-Busch, J., Géoui, T. & Evans, D. The making of Reaxys - Towards unobstructed access to relevant chemistry information. in *The Future of the History of Chemical Information*, vol. 1164, pp. 127–148 (American Chemical Society, Columbus, OH, USA, 2014).

44. Heitel, P., Achenbach, J., Moser, D., Proschak, E. & Merk, D. DrugBank screening revealed altiretinoin and bexarotene as liver X receptor modulators. *Bioorg Med Chem Lett* **27**, 1193–1198 (2017).

45. Valsecchi, C., Grisoni, F., Motta, S., Bonati, L. & Ballabio, D. NURA: A curated dataset of nuclear receptor modulators. *Toxicol. Appl. Pharmacol.* **407**, 115244 (2020).

46. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).

47. Aoyama, A. *et al.* Design, synthesis, and biological evaluation of novel transrepression-selective liver X receptor (LXR) ligands with 5,11-dihydro-5-methyl-11-methylene-6H-dibenz[b,e]azepin-6-one skeleton. *J. Med. Chem.* **55**, 7360–7377 (2012).

48. Li, L. *et al.* Discovery and optimization of a novel series of liver X receptor-alpha agonists. *Bioorg. Med. Chem. Lett.* **16**, 1638–1642 (2006).

49. Reker, D. & Schneider, G. Active-learning strategies in computer-assisted drug discovery. *Drug Discov. Today* **20**, 458–465 (2015).

50. Warmuth, M. K. *et al.* Active learning with support vector machines in the drug discovery process. *J. Chem. Inf. Comput. Sci.* **43**, 667–673 (2003).

51. Button, A., Merk, D., Hiss, J. A. & Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nat. Mach. Intell.* **1**, 307–315 (2019).

52. O’Boyle, N. M. Towards a universal SMILES representation - A standard method to generate canonical SMILES based on the InChI. *J. Cheminformatics* **4**, 22 (2012).

53. Berthold, M. R. *et al.* KNIME - The Konstanz information miner: Version 2.0 and beyond. *SIGKDD Explor. Newsl.* **11**, 26–31 (2009).

54. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *ArXiv Prepr. ArXiv14126980* (2014).

55. Fradera, X. *et al.* X-Ray Structures of the LXRα LBD in its Hhomodimeric form and implications for heterodimer signaling. *J. Mol. Biol.* **399**, 120–132 (2010).
56. Färnegårdh, M. et al. The three-dimensional structure of the liver X receptor β reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *J. Biol. Chem.* **278**, 38821–38828 (2003).

57. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein–ligand docking using GOLD. *Proteins Struct. Funct. Bioinform.* **52**, 609–623 (2003).
Combining generative artificial intelligence and on-chip synthesis for de novo drug design

Francesca Grisoni\textsuperscript{1,}\textsuperscript{*}, Berend J.H. Huisman\textsuperscript{1,}, Alexander L. Button\textsuperscript{1,}, Michael Moret\textsuperscript{1,}, Kenneth Atz\textsuperscript{1,}, Daniel Merk\textsuperscript{2,1,}\textsuperscript{*} & Gisbert Schneider\textsuperscript{1,}\textsuperscript{*}

\textsuperscript{1}Department of Chemistry and Applied Biosciences, RETHINK, ETH Zurich, Zurich, Switzerland; \textsuperscript{2}Goethe University Frankfurt, Institute of Pharmaceutical Chemistry, Frankfurt, Germany; \textsuperscript{*}These authors contributed equally; \textsuperscript{*Email: francesca.grisoni@pharma.ethz.ch, merk@pharmchem.uni-frankfurt.de, gisbert@ethz.ch

LSTM training and optimization......................................................................................................................................... 2
Fine-tuning molecules ......................................................................................................................................................... 2
Choice of sampling temperature ....................................................................................................................................... 4
Choice of epochs for sampling ......................................................................................................................................... 5
Reaction SMARTS................................................................................................................................................................. 6
Synthesis ................................................................................................................................................................................. 7
Compounds retained for synthesis ..................................................................................................................................... 7
Flow Synthesis ...................................................................................................................................................................... 7
Similarity analysis ................................................................................................................................................................. 18
Similarity to ChEMBL27 compounds ............................................................................................................................... 18
Similarity to fine-tuning compounds ................................................................................................................................ 23
Chemistry ............................................................................................................................................................................... 28
LSTM training and optimization

Fine-tuning molecules

Table S1. Set of utilized molecules for fine-tuning (identified with SMILES and ChEMBL ID). Compound bioactivities on LXR are reported in Table S2.

| CHEMBL ID   | SMILES                                                                 |
|-------------|------------------------------------------------------------------------|
| CHEMBL1091034 | CS(=O)(=O)c1ccc(-c2ccc(CN(Cc3ccc(F)c3Cl)S(=O)(=O)c3cccc3cc2)c1       |
| CHEMBL1093840 | CS(=O)(=O)c1ccc(-c2ccc(CN(Cc3cc(F)c3cc3Cl)S(=O)(=O)c3cccc3cc2)c1       |
| CHEMBL202902  | CN(=CCc1cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                              |
| CHEMBL203606  | CC(C)OC(=O)(C)c1ccc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                      |
| CHEMBL205937  | O=S(=O)(c1ccccc1)N1CCOc2cc(C(F)(F)F)c1c1c1                             |
| CHEMBL206445  | CN(=CCc1cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                              |
| CHEMBL206626  | CN(=CCc1cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                              |
| CHEMBL206674  | CN(=CCc1cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                              |
| CHEMBL206738  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL215904  | CC(C)CN(c1ccc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1                             |
| CHEMBL228242  | COc1ccc(C(F)c1c1ccccc1)c1ccc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1             |
| CHEMBL3360960 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL3360961 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL3360963 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL3360968 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL3360970 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL3360974 | O=S(=O)(c1ccccc1)N1CCc2cc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1               |
| CHEMBL379225  | CN(CO)c1ccc(C(F)c1c1ccccc1)c1ccc(C(=O)(C(F)(F)F)C(F)(F)F)c1c1       |
| CHEMBL380612  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL380851  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL381478  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL381570  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL383944  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL383945  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL384246  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3890276 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3910597 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3917300 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3926292 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3935187 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3944154 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3945199 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3945820 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3953927 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3960606 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL3980683 | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL59030   | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL595689  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL611735  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL612007  | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
| CHEMBL62136   | CS(=O)(=O)c1ccc(-c2ccc(-c3cc(F)c3cc3Cl)cc2)c1                          |
Table S2. Summary of the bioactivity values of the fine-tuning compounds, as annotated in ChEMBL (n.a. = not available).

| ChEMBL ID | LXRα EC\textsubscript{50} (nM) | n. records | LXRβ EC\textsubscript{50} (nM) | n. records |
|-----------|-------------------------------|------------|-------------------------------|------------|
| CHEMBL59030 | 439±112                       | 15         | 156±44                        | 17         |
| CHEMBL379225 | 365±235                       | 2          | 180                           | 1          |
| CHEMBL62136 | 235±55                        | 28         | 141±26                        | 33         |
| CHEMBL1093840 | 100                           | 1          | 20                            | 1          |
| CHEMBL202902 | 100                           | 1          | n.a.                          | n.a.       |
| CHEMBL203606 | 100                           | 1          | n.a.                          | n.a.       |
| CHEMBL215904 | 100                           | 1          | 30                            | 1          |
| CHEMBL381570 | 100                           | 1          | n.a.                          | n.a.       |
| CHEMBL383944 | 100                           | 1          | 30                            | 1          |
| CHEMBL384246 | 100                           | 1          | 20                            | 1          |
| CHEMBL3960606 | 100                           | 1          | 35±26                         | 2          |
| CHEMBL3890276 | 99                            | 1          | 94±76                         | 2          |
| CHEMBL389345 | 90                            | 1          | 40                            | 1          |
| CHEMBL3360974 | 85                            | 1          | 66                            | 1          |
| CHEMBL206445 | 80                            | 1          | n.a.                          | n.a.       |
| CHEMBL3944154 | 76                            | 1          | 40±32                         | 2          |
| CHEMBL3953927 | 72                            | 1          | 40±32                         | 2          |
| CHEMBL242842 | 70                            | 1          | 6                             | 1          |
| CHEMBL3980683 | 69                            | 1          | 73±67                         | 2          |
| CHEMBL3360968 | 65                            | 1          | 64                            | 1          |
| CHEMBL3926292 | 57                            | 1          | 26±24                         | 2          |
| CHEMBL3360970 | 52                            | 1          | 61                            | 1          |
| CHEMBL205937 | 50                            | 1          | n.a.                          | n.a.       |
| CHEMBL380851 | 40                            | 1          | n.a.                          | n.a.       |
| CHEMBL612007 | 32                            | 1          | 170                           | 1          |
| CHEMBL595689 | 31                            | 1          | 183                           | 1          |
| CHEMBL206674 | 30                            | 1          | n.a.                          | n.a.       |
| CHEMBL381478 | 30                            | 1          | n.a.                          | n.a.       |
| CHEMBL3935187 | 30                            | 1          | 23±20                         | 2          |
| CHEMBL206626 | 20                            | 1          | n.a.                          | n.a.       |
| CHEMBL3360961 | 20                            | 1          | 20                            | 1          |
| CHEMBL611735 | 20                            | 1          | 144                           | 1          |
| CHEMBL3360963 | 18                            | 1          | 14                            | 1          |
| CHEMBL3945820 | 12                            | 1          | 14±12                         | 2          |
| CHEMBL3917300 | 11                            | 1          | 11±10                         | 2          |
| CHEMBL1091034 | 10                            | 1          | 3                             | 1          |
| CHEMBL380612 | 10                            | 1          | n.a.                          | n.a.       |
| CHEMBL206738 | 8                             | 1          | n.a.                          | n.a.       |
| CHEMBL3910597 | 8                             | 1          | 15±13                         | 2          |
| CHEMBL3945199 | 8                             | 1          | 12±11                         | 2          |
| CHEMBL3360960 | 4                             | 1          | 2                             | 1          |
Choice of sampling temperature

Three types of sampling temperatures were chosen, namely T= 0.2, T=0.7 and T=1.2. To check for the optimal settings for sampling, we set out to analyze (i) SPiDER\textsuperscript{1} predictions for LXR (p<0.05), (ii) the number of novel atomic scaffolds\textsuperscript{2} compared to the fine-tuning set and the pre-training library, and (iii) the distance to the fine-tuning compounds, as determined by the Euclidean distance computed on the Chemically Advanced Template Search (CATS)\textsuperscript{3} descriptors. For each fine-tuning epoch, 3000 unique SMILES were sampled, and valid SMILES were used to compute (i) percentage of compounds predicted as active on LXR by SPiDER (the higher, the better), (ii) number of novel scaffolds compared to pre-training and fine-tuning molecules (the higher, the better), and (iii) mean CATS distance to fine-tuning compounds (the lower, the better). The results were analyzed via Principal Component Analysis (PCA, Figure S2), using open-source MATLAB code\textsuperscript{4} (2 components, autoscaling).

![Figure S1](image)

Figure S1. Principal component analysis on the quality metrics for each sampling temperature (T) and each epoch. The first two principal components explain 91.2\% of the total variance (PC1: 66.6\%, PC2: 24.6\%). (a) Score plot. Colors represent the temperature value, where numbers encode the epoch number. (b) Loading plot. Contribution of each quality metric (’%actives’ = percentage of actives as predicted by SPIDER; ’CATSdist’ = mean Euclidean distance on CATS descriptors; ’novel scaffolds’ = number of novel scaffolds that are not present in the training or fine-tuning set).
Choice of epochs for sampling

**Figure S2.** Variation of performance metrics (percentage of actives on LXR as predicted by SPIDER, mean Euclidean distance to fine-tuning molecules using CATS descriptors, and number of novel scaffolds not present in the training or fine-tuning set) as a function of the fine-tuning epochs (3000 SMILES sampled per epoch). Dashed lines represent the values of such metrics on the pre-training and fine-tuning sets. The blue area indicates the selected epochs.
## Reaction SMARTrs

| No. | Reaction                        | Reaction SMARTS                                                                 |
|-----|---------------------------------|-------------------------------------------------------------------------------|
| 1   | Amide formation, Acid Chloride  | ![Reaction SMARTS for Amide formation, Acid Chloride]                          |
| 2   | Aminothiazol formation          | ![Reaction SMARTS for Aminothiazol formation]                                 |
| 3   | Ar-Imidazole formation          | ![Reaction SMARTS for Ar-Imidazole formation]                                 |
| 4   | Ester formation, Acid Chloride  | ![Reaction SMARTS for Ester formation, Acid Chloride]                         |
| 5   | FGI Acyl chloride               | ![Reaction SMARTS for FGI Acyl chloride]                                      |
| 6   | FGI sulfonyl chloride           | ![Reaction SMARTS for FGI sulfonyl chloride]                                  |
| 7   | Fischer indole                  | ![Reaction SMARTS for Fischer indole]                                         |
| 8   | Hantzsch                       | ![Reaction SMARTS for Hantzsch]                                              |
| 9   | Paal-Knorr-pyrole formation     | ![Reaction SMARTS for Paal-Knorr-pyrole formation]                           |
| 10  | Pictet-Spengler-5-membered-ring | ![Reaction SMARTS for Pictet-Spengler-5-membered-ring]                         |
| 11  | Pictet-Spengler-6-membered-ring | ![Reaction SMARTS for Pictet-Spengler-6-membered-ring]                         |
| 12  | Reductive amination-Primary amine-Ketone | ![Reaction SMARTS for Reductive amination-Primary amine-Ketone]           |
| 13  | Sulfonamide formation, Sulfonyl Chloride | ![Reaction SMARTS for Sulfonamide formation, Sulfonyl Chloride]               |
| 14  | Thioster formation, Acid Chloride | ![Reaction SMARTS for Thioster formation, Acid Chloride]                     |
| 15  | Triaryl-imidazol-1 2-diketone   | ![Reaction SMARTS for Triaryl-imidazol-1 2-diketone]                          |
| 16  | Triaryl-imidazol-alpha hydroxy ketone | ![Reaction SMARTS for Triaryl-imidazol-alpha hydroxy ketone]               |
| 17  | UGI-6-ring                      | ![Reaction SMARTS for UGI-6-ring]                                            |

### Table S3: Set of utilized reactions and corresponding decomposition SMARTS.

| No. | Reaction                          | Reaction SMARTs                                                                 |
|-----|-----------------------------------|-------------------------------------------------------------------------------|
| 1   | Amide formation, Acid Chloride   | ![Reaction SMARTS for Amide formation, Acid Chloride]                          |
| 2   | Aminothiazol formation           | ![Reaction SMARTS for Aminothiazol formation]                                 |
| 3   | Ar-Imidazole formation           | ![Reaction SMARTS for Ar-Imidazole formation]                                 |
| 4   | Ester formation, Acid Chloride   | ![Reaction SMARTS for Ester formation, Acid Chloride]                         |
| 5   | FGI Acyl chloride                | ![Reaction SMARTS for FGI Acyl chloride]                                      |
| 6   | FGI sulfonyl chloride            | ![Reaction SMARTS for FGI sulfonyl chloride]                                  |
| 7   | Fischer indole                   | ![Reaction SMARTS for Fischer indole]                                         |
| 8   | Hantzsch                        | ![Reaction SMARTS for Hantzsch]                                              |
| 9   | Paal-Knorr-pyrole formation      | ![Reaction SMARTS for Paal-Knorr-pyrole formation]                           |
| 10  | Pictet-Spengler-5-membered-ring  | ![Reaction SMARTS for Pictet-Spengler-5-membered-ring]                         |
| 11  | Pictet-Spengler-6-membered-ring  | ![Reaction SMARTS for Pictet-Spengler-6-membered-ring]                         |
| 12  | Reductive amination-Primary amine-Ketone | ![Reaction SMARTS for Reductive amination-Primary amine-Ketone]           |
| 13  | Sulfonamide formation, Sulfonyl Chloride | ![Reaction SMARTS for Sulfonamide formation, Sulfonyl Chloride]               |
| 14  | Thioster formation, Acid Chloride | ![Reaction SMARTS for Thioster formation, Acid Chloride]                     |
| 15  | Triaryl-imidazol-1 2-diketone     | ![Reaction SMARTS for Triaryl-imidazol-1 2-diketone]                          |
| 16  | Triaryl-imidazol-alpha hydroxy ketone | ![Reaction SMARTS for Triaryl-imidazol-alpha hydroxy ketone]               |
| 17  | UGI-6-ring                       | ![Reaction SMARTS for UGI-6-ring]                                            |
## Synthesis

**Compounds retained for synthesis**

Table S4. List of all compounds generated by the model that have no reported activity on LXR. The table reports SMILES, predicted reaction (‘Reac’: 1 = Amide bond formation, acid chloride; 2 = sulfonamide formation, sulfonyl chloride, 3 = ester formation, acid chloride), predicted building blocks and type (selected for synthesis, purchased, not available or discarded).

| SMILES | Reac. | Predicted building blocks | Type |
|---------|-------|---------------------------|------|
| CN(C(=O)c1cccs1)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1cccs1.F.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(c1ccc(C(F)(F)F)C(F)(F)F)c1cc1S(=O)(=O)c1cccc(S(N)(=O)=O)c1 | 1 | O=C(Cl)c1cccs1.F.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(C(=O)c1c(Cl)cccc1Cl)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1cccs1.F.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(c1ccc(C(=O)Cl)cc1)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1cccs1.F.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(C(=O)c1ccccc1Cl)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1ccccc1Cl.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1)S(=O)(=O)c1cccc(S(N)(=O)=O)c1 | 1 | O=C(Cl)c1cccs1.F.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(C(=O)c1ccccc1Cl)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1ccccc1Cl.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1)S(=O)(=O)c1cccc(S(N)(=O)=O)c1 | 1 | O=C(Cl)c1ccccc1Cl.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(c1ccc(C(=O)Cl)cc1)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1ccccc1Cl.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |
| CN(C(=O)c1ccccc1Cl)c1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | 1 | O=C(Cl)c1ccccc1Cl.CNc1ccc(C(O)(C(F)(F)F)C(F)(F)F)cc1 | synthesis |

The table reports SMILES, predicted reaction (‘Reac’: 1 = Amide bond formation, acid chloride; 2 = sulfonamide formation, sulfonyl chloride, 3 = ester formation, acid chloride), predicted building blocks and type (selected for synthesis, purchased, not available or discarded).
**Flow Synthesis**

**Starting materials**

1,1,1,3,3,3-Hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol. Synthesized according to WO2015/82533. A mixture of aniline (392 µL) and hexafluoroacetone trihydrate (600 µL) was heated in a sealed tube under microwave irradiation to 150 °C for 2 hours. The crude product was crystallized from hexane with 20 percent ethyl acetate to give 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol as an off-white solid (600 mg). 1H NMR (400 MHz, DMSO-d₆) δ 8.20 (s, 1H), 7.36 (d, J = 8.4 Hz, 2H), 6.59 (d, J = 9.0 Hz, 2H), 6.02 (d, J = 5.4 Hz, 1H), 2.69 (d, J = 4.9 Hz, 3H).

Methyl-2-hydroxyhydratropaat was synthesized by adding p-Toluenesulfonic acid monohydrate (6.0 mg, 32 µmol) to a solution of 2-hydroxy-2-phenylpropanoic acid (50 mg, 300 µmol; 95% ee) in methanol (3 mL). The solution was heated in a sealed tube at 75 °C for 6 h. The solvent was then removed under reduced pressure and the residue was used without further purification.

1-Phenylcyclopentane-1-carbonyl chloride and 4-Sulfamoylbenzoyl chloride. A solution of the respective carboxylic acid (30 mmol) in thionyl chloride (5 ml) with DMF (0.5 mL) was heated to reflux for 2 hours. The reaction mixture was concentrated *in vacuo* and the resulting acid chloride used directly for flow synthesis.

3-Biphenylsulfonyl chloride was synthesized by dissolving 3-Bromobiphenyl (1.5 g, 6.4 mmol) in ether (15 ml) and cooling to -78 °C. T-Butyllithium (1.7 M solution in hexane, 3.8 ml, 6.4 mmol) was added dropwise under constant stirring and an argon atmosphere. The resultant reaction mixture was stirred at -10 °C to -5 °C for 6 h. The reaction mixture was cooled to -78 °C and sulfuric chloride (0.64 ml, 6.4 mmol) was added dropwise. After the completion of the addition, the reaction mixture was allowed to attain ambient temperature slowly and stirred for 1 h. The reaction mixture was diluted with ethyl acetate (50 ml), washed with water and the organic layer dried over anhydrous MgSO₄. The solvent was then removed under reduced pressure and the residue was used without further purification.

**LC-MS analysis of reaction mixtures**

- \(N\)-(4-(1,1,1,3,3,3-Hexafluoro-2-hydroxypropan-2-yl)phenyl)-N-methyl-2,2-diphenylacetamide (1). MW=467.41
- \(N\)-(4-(1,1,1,3,3,3-Hexafluoro-2-hydroxypropan-2-yl)phenyl)-N-methyl-1-phenylcyclopentane-1-carboxamide (2). MW=445.41.
- $\text{N-(4-(1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N'-methylcyclopropanecarboxamide (3). MW}=341.25$

- $\text{N-(4-(1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N'-methyl-2-phenoxyacetamide (4). MW}=407.31$

- $\text{2,4,6-Trifluoro-\text{N-(4-(1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N'-methylbenzamide (5). MW}=431.26}$

- $\text{2-Chloro-6-fluoro-\text{N-(4-(1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N'-methylbenzamide (6). MW}=429.72}$
• **3,4-Dichloro-\(N\{(4\{(1,1,1,3,3,3\text{-hexafluoro-2-hydroxypropan-2-yl)}phenyl\}\)\(N\)methylbenzamide (7). MW=446.17

• **4-Fluoro-\(N\{(4\{(1,1,1,3,3,3\text{-hexafluoro-2-hydroxypropan-2-yl)}phenyl\}\)\(N\)methylbenzamide (8). MW=395.28

• **2-Chloro-4-fluoro-\(N\{(4\{(1,1,1,3,3,3\text{-hexafluoro-2-hydroxypropan-2-yl)}phenyl\}\)\(N\)methylbenzamide (9). MW=429.72

• **\(N\{(4\{(1,1,1,3,3,3\text{-hexafluoro-2-hydroxypropan-2-yl)}phenyl\}\)\(N\)methyl-3-(trifluoromethyl)benzamide (10). MW=445.28
- **2,3-Dichloro-N\(^{(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)}\)-N\(^{methylbenzamide\ (11)}\). MW=446.17

- **3-Fluoro-N\(^{(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)}\)-N\(^{methylbenzamide\ (12)}\). MW=395.28

- **N\(^{(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)}\)-N\(^{methyl-1-naphthamide\ (13)}\). MW=427.35

- **2-Chloro-N\(^{(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)}\)-N\(^{methylbenzamide\ (14)}\). MW=411.73
- $N\{(1,1,3,3,3$-Hexafluoro-2-hydroxypropan-2-yl)phenyl$\}-N$-methylthiophene-2-carboxamide (15). MW=383.31

- $N\{(1,1,3,3,3$-Hexafluoro-2-hydroxypropan-2-yl)phenyl$\}-N$-methyl-2-(trifluoromethyl)benzenesulfonamide (16). MW=481.33 (ESI-)

- $N\{(1,1,3,3,3$-Hexafluoro-2-hydroxypropan-2-yl)phenyl$\}-3$-methoxy-$N$-methylbenzenesulfonamide (17). MW=443.36 (ESI-)

- $N$-Methyl-$N$-phenyl-[1,1'-biphenyl]-3-sulfonamide (18). MW=323.41
- N-Methyl-N-phenyl-3-(trifluoromethyl)benzenesulfonamide (19), MW=315.31

- N-Methyl-N-phenyl-2-(trifluoromethyl)benzenesulfonamide (20), MW=315.31

- 1-Methoxy-1-oxo-2-phenylpropan-2-yl cyclopropanecarboxylate (21), MW=248.28

- (2,6-Dichlorophenyl)(thiazolidin-3-yl)methanone (22), MW=262.15
- 1-[(1,1'-Biphenyl)-3-ylsulfonyl]indoline (23). MW=335.42

- 4-(Phenylsulfonyl)morpholine (24). MW=227.28

- 3-(Phenylsulfonyl)thiazolidine (25). MW=229.31
Batch synthesis

Scheme S1. Batch synthesis of amide derivatives 1, 3, 5, 6, and 8-15.

Scheme S2. Batch synthesis of 2.
Scheme S3. Batch synthesis of sulfonamide derivative 17.
Similarity analysis

**Similarity to ChEMBL27 compounds**

Table S5. Summary of the two most similar ChEMBL27 compounds to the bioactive de novo designs. ChEMBL ID, structure and Jaccard-Tanimoto (ECFP4, 1024 bits) are reported, along with annotated bioactivity in ChEMBL27, if available. ‘FT’ (yes/no) indicates whether the ChEMBL compound was present in the fine-tuning set.

| De novo designs | Most similar compounds (ChEMBL) | Bioactivity (ChEMBL) |
|-----------------|---------------------------------|----------------------|
| ID              | ChEMBL ID Structure Sim. | LXRα [μM] | LXRβ [μM] | FT |
| 1               | CHEMBL206808 | 67.5 | Not reported | Not reported | n |
|                | CHEMBL379225 | 64.29 | EC_{50}=0.4±0.3 | EC_{50}=0.18 | y |
| 1               | CHEMBL1084227 | 79.55 | Not reported | Not reported | n |
| 1               | CHEMBL1084488 | 69.05 | Not reported | Not reported | n |
| 5               | CHEMBL379225 | 60.47 | EC_{50}=0.4±0.3 | EC_{50}=0.18 | n |
| 5               | CHEMBL1575396 | 60 | Not reported | Not reported | n |
| Compound | EC<sub>50</sub> α (μM) | EC<sub>50</sub> β (μM) | Note 1 | Note 2 | Note 3 |
|----------|------------------|------------------|--------|--------|--------|
| CHEMBL379225 | 0.18±0.006 μM | 0.4±0.3 μM | CHEMBL379225 | 58.7 | EC<sub>50</sub>=0.4±0.3 | EC<sub>90</sub>=0.18 | y |
| CHEMBL3321760 | 0.40±0.01 μM | Not reported | CHEMBL3321760 | 56 | Not reported | Not reported | n |
| CHEMBL3322041 | 1.05±0.01 μM | Not reported | CHEMBL3322041 | 78.95 | Not reported | Not reported | n |
| CHEMBL202953 | 1.72±0.04 μM | EC<sub>50</sub>=0.2 μM | CHEMBL202953 | 76.32 | Not reported | Not reported | n |
| CHEMBL3322040 | 1.68±0.03 μM | Not reported | CHEMBL3322040 | 58.33 | Not reported | Not reported | n |
| CHEMBL379225 | 4.0±0.1 μM | Not reported | CHEMBL379225 | 55.32 | EC<sub>50</sub>=0.4±0.3 | EC<sub>90</sub>=0.18 | y |
|   | CHEMBL3321760 | CHEMBL379225 |
|---|--------------|--------------|
| 10 | ![Chemical Structure](image1) | ![Chemical Structure](image2) |
|   | α: EC\(_{50}\) = 1.19±0.01 μM | α: EC\(_{50}\) = 1.31±0.03 μM |
|   | β: EC\(_{50}\) = 3.1±0.1 μM | β: EC\(_{50}\) = 2.37±0.02 μM |
| 11 | ![Chemical Structure](image3) | ![Chemical Structure](image4) |
|   | α: EC\(_{50}\) = 1.31±0.03 μM | α: EC\(_{50}\) = 0.8±0.3 μM |
|   | β: EC\(_{50}\) = 2.37±0.02 μM | β: EC\(_{50}\) = 1.08±0.02 μM |

| CHEMBL3321760 | 76.74 | Not reported | Not reported | n |
| CHEMBL379225 | 73.17 | EC\(_{50}\)=0.4±0.3 | EC\(_{50}\)=0.18 | y |

| CHEMBL3322040 | 68.18 | Not reported | Not reported | n |

| CHEMBL379225 | 61.36 | EC\(_{50}\)=0.4±0.3 | EC\(_{50}\)=0.18 | y |

| CHEMBL3321760 | 76.74 | Not reported | Not reported | n |

| CHEMBL379225 | 73.17 | EC\(_{50}\)=0.4±0.3 | EC\(_{50}\)=0.18 | y |
| Compound | Molecular Structure | EC50 | pEC50 | n
| --- | --- | --- | --- | --- |
| CHEMBL3322040 | ![Molecular Structure](image) | 70.45 | Not reported | n |
| CHEMBL379225 | ![Molecular Structure](image) | 63.64 | EC50 = 0.4±0.3, EC50 = 0.18 | y |
| CHEMBL3322040 | ![Molecular Structure](image) | 73.81 | Not reported | n |
| CHEMBL379225 | ![Molecular Structure](image) | 66.67 | EC50 = 0.4±0.3, EC50 = 0.18 | n |
| CHEMBL3322048 | ![Molecular Structure](image) | 100 | Not reported | n |
| CHEMBL1460916 | ![Molecular Structure](image) | 70.21 | Not reported | n |

**EC50 Values:**
- α: EC50 = 1.1±0.1 μM
- β: EC50 = 2.23±0.02 μM

**EC50 Values:**
- α: EC50 = 0.30±0.01 μM
- β: EC50 = 1.41±0.02 μM

**EC50 Values:**
- α: EC50 = 0.24±0.04 μM
- β: EC50 = 0.34±0.02 μM
| CHEMBL3322061 | CHEMBL3322062 |
|----------------|----------------|
| ![Chemical Structure](image1) | ![Chemical Structure](image2) |
| α: EC₅₀ = 0.21±0.02 μM | β: EC₅₀ = 1.25±0.01 μM |

CHEMBL3322061 | 100 | Not reported | Not reported | n |

CHEMBL3322062 | 73.91 | Not reported | Not reported | n |
### Similarity to fine-tuning compounds

Table S6. Summary of the two most similar fine-tuning compounds to the bioactive de novo designs. ChEMBL ID, structure and Jaccard-Tanimoto (ECFP4, 1024 bits) are reported, along with annotated bioactivity in ChEMBL27, if available.

| De novo designs | Most similar compounds (ChEMBL) | Bioactivity (ChEMBL) |
|-----------------|----------------------------------|----------------------|
| **ID**          | **Structure**                    | **ChEMBL ID**        | **Structure** | **Sim.** | LXRα [μM] | LXRβ [μM]  |
| **α**: EC<sub>50</sub> = 4.5±0.1 μM | ![Structure](image1) | CHEMBL379225 | ![Structure](image2) | 64.3 | EC<sub>50</sub>=0.4±0.3 | EC<sub>50</sub>=0.18 |
| **β**: EC<sub>50</sub> >10 μM  | ![Structure](image3) | CHEMBL203606 | ![Structure](image4) | 52.1 | EC<sub>50</sub>=0.1 | Not reported |
| **α**: EC<sub>50</sub> = 0.26±0.01 μM | ![Structure](image1) | CHEMBL379225 | ![Structure](image2) | 60.5 | EC<sub>50</sub>=0.4±0.3 | EC<sub>50</sub>=0.18 |
| **β**: EC<sub>50</sub> = 1.30±0.03 μM  | ![Structure](image3) | CHEMBL203606 | ![Structure](image4) | 49.0 | EC<sub>50</sub>=0.1 | Not reported |

**5**

![Structure](image1) | ![Structure](image2) | CHEMBL379225 | ![Structure](image3) | ![Structure](image4) | CHEMBL203606 | 49.0 | EC<sub>50</sub>=0.1 | Not reported |

**6**

![Structure](image1) | ![Structure](image2) | CHEMBL379225 | ![Structure](image3) | ![Structure](image4) | 58.7 | EC<sub>50</sub>=0.4±0.3 | EC<sub>50</sub>=0.18 |
|   | α: EC<sub>50</sub> = 0.183±0.006 μM | β: EC<sub>50</sub> = 0.40±0.01 μM |
|---|---------------------------------|---------------------------------|
| 8 | CHEMBL203606 45.3 | EC<sub>50</sub>=0.1 | Not reported |
| 9 | CHEMBL379225 55.3 | EC<sub>50</sub>=0.4±0.3 | EC<sub>50</sub>=0.18 |
| 10| CHEMBL379225 75.0 | EC<sub>50</sub>=0.4±0.3 | EC<sub>50</sub>=0.18 |
α: EC$_{50}$ = 1.19±0.01 μM
β: EC$_{50}$ = 3.1±0.1 μM

CHEMBL206445

EC$_{50}$ = 0.08
Not reported

α: EC$_{50}$ = 1.31±0.03 μM
β: EC$_{50}$ = 2.37±0.02 μM

CHEMBL379225

EC$_{50}$ = 0.4±0.3
EC$_{50}$ = 0.18

α: EC$_{50}$ = 0.8±0.3 μM
β: EC$_{50}$ = 1.08±0.02 μM

CHEMBL203606

EC$_{50}$ = 0.1
Not reported

α: EC$_{50}$ = 0.8±0.3 μM
β: EC$_{50}$ = 1.08±0.02 μM

CHEMBL203606

EC$_{50}$ = 0.1
Not reported

α: EC$_{50}$ = 0.8±0.3 μM
β: EC$_{50}$ = 1.08±0.02 μM

CHEMBL203606

EC$_{50}$ = 0.1
Not reported

α: EC$_{50}$ = 0.8±0.3 μM
β: EC$_{50}$ = 1.08±0.02 μM

CHEMBL203606

EC$_{50}$ = 0.1
Not reported
| No. | Chemical Structure | EC<sub>50</sub> Values | EC<sub>50</sub> Values | 
|-----|-------------------|------------------------|------------------------|
| 14  | ![Chemical Structure](image1) | α: EC50 = 1.1±0.1 μM  
β: EC50 = 2.23±0.02 μM | CHEMBL203606 | 52.0 | EC<sub>50</sub>=0.1 | Not reported |
| 15  | ![Chemical Structure](image2) | α: EC50 = 0.30±0.01 μM  
β: EC50 = 1.41±0.02 μM | CHEMBL203606 | 52.1 | EC<sub>50</sub>=0.1 | Not reported |
| 17  | ![Chemical Structure](image3) | α: EC50 = 0.24±0.04 μM  
β: EC50 = 0.34±0.02 μM | CHEMBL206445 | 70.8 | EC<sub>50</sub>=0.08 | Not reported |
| α: EC<sub>50</sub> = 0.21±0.02 µM | β: EC<sub>50</sub> = 1.25±0.01 µM |
|---------------------------------|----------------------------------|
| CHEMBL206674                    | 70.8                             |
| 70.8                             | EC<sub>50</sub> = 0.03           |
| Not reported                     |                                  |
Chemistry

**Compound synthesis and characterization**

1,1,1,3,3,3-Hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (a). In a microwave tube, 1,1,1,3,3,3-hexafluoropropan-2-one (1.8 g, 1.14 mL, 11 mmol 1 eq.) was dissolved in N-methylpyrrolidone (23.54 g, 23 mL, 220 mmol, 20 eq.). The mixture was stirred at 140 °C under microwave irradiation for 2 h, cooled down to rt, and recrystallized from heptane/AcOEt to give the title compound (1.50 g, 5.49 mmol, 50%) as light brown solid. \(^1\)H NMR (400 MHz, CDCl\(_3\)) \(\delta\) 7.42 (d, \(J = 8.1\) Hz, 2H), 7.37 (d, \(J = 8.1\) Hz, 2H), 2.77 (s, 3H); HRMS (ESI): calculated 274.0661 for \(\text{C}_{16}\text{H}_{22}\text{F}_{6}\text{NO} ([M+H]^+)\), found 274.0658.

\(\text{N}^1\text{-(4-(1,1,1,3,3,3-Hexafluoro-2-hydroxypropan-2-yl)phenyl)-N-methyl-1-phenylcyclopentane-1-carboxamide} (2). 1-phenylcyclopentane-1-carboxylic acid (0.19 g, 1.0 mmol, 1.0 eq.) and 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (0.27 g, 1.0 mmol, 1.0 eq.) were dissolved in CHCl\(_3\) (20 mL), and 4-\(\text{N}^1\text{,}\)N'-dimethylaminopyridine (0.41 g, 3.0 mmol, 3.0 eq.) and EDC*HCl (0.58 g, 3.0 mmol, 3.0 eq.) were added. The solution was stirred under reflux for 24 h. After cooling to room temperature, 20 mL 5% aqueous hydrochloric acid were added, phases were separated, and the aqueous layer was extracted three times with EtOAc (3x20 mL). The combined organic layers were dried over magnesium sulfate and concentrated \textit{in vacuo}. The crude product was purified by flash chromatography (SiO\(_2\), 250 g, 20% AcOEt in hexane, 100% CH\(_2\)Cl\(_2\)) to give the title compound (55 mg, 12%) as a colorless solid. \(^1\)H NMR (500 MHz, CDCl\(_3\)) \(\delta\) 7.42 – 7.39 (m, 2H), 7.35 – 7.30 (m, 2H), 7.29 – 7.24 (m, 1H), 6.58 (d, \(J = 8.6\) Hz, 2H), 6.36 – 6.31 (m, 2H), 2.79 – 2.72 (m, 2H), 2.71 (s, 3H), 1.96 – 1.88 (s, 2H), 1.78 – 1.67 (m, 4H); \(^13\)C NMR (126 MHz, CDCl\(_3\)) \(\delta\) = 170.59, 149.36, 141.60, 128.45, 127.24, 127.27, 115.43, 112.23, 60.39, 59.81, 53.42, 53.53, 30.69, 23.30, 21.05, 14.20; R: TLC: 0.32 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 446.1549 for \(\text{C}_{23}\text{H}_{22}\text{F}_{6}\text{NO}_{2} ([M+H]^+)\), found 446.1541.

\(\text{N}^1\text{-(4-(1,1,1,3,3,3-Hexafluoro-2-hydroxypropan-2-yl)phenyl)-3-methoxy-N-methylbenzenesulfonamide} (17). In a 10 mL round bottom flask 1,1,1,3,3,3-Hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (70 mg, 256 \(\mu\)mol, 1.0 eq.) was dissolved in CH\(_2\)Cl\(_2\) (4 mL). Triethylamine (36 mg, 50 \(\mu\)L, 278 \(\mu\)mol, 1.1 eq.) was added, the reaction mixture was cooled to 0 °C and carboxylic 4-methoxybenzenesulfonyl chloride (64 mg, 307 \(\mu\)mol, 1.2 eq.) was added dropwise. The reaction mixture was allowed to warm to room temperature and stirring was continued for 8 h. The solution was diluted with CH\(_2\)Cl\(_2\), poured onto water, extracted with AcOEt and washed with saturated NaHCO\(_3\) and brine. The combined organic layers were dried over magnesium sulfate and concentrated \textit{in vacuo}.

General procedure A for amide synthesis with 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol and carbonyl chlorides

In a 10 mL round bottom flask 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (1.0 eq.) was dissolved in CH\(_2\)Cl\(_2\) (60 mM). Triethylamine (1.1 eq.) was added, the reaction mixture was cooled to 0 °C and the respective carbonyl acid chloride (1.0 - 1.2 eq.) was added dropwise under stirring. The reaction mixture was allowed to warm to room temperature and stirring was continued for 8 h. The solution was diluted with CH\(_2\)Cl\(_2\), poured onto water, phases were separated, and the aqueous layer was extracted with AcOEt. The combined organic layers were washed with saturated NaHCO\(_3\) and brine, dried over magnesium sulfate and concentrated \textit{in vacuo}. The crude material was purified by flash chromatography (SiO\(_2\), 250 g, 10-80% AcOEt in cyclohexane) to give the title compound.
(methylamino)phenyl)propan-2-ol (70 mg, 256 μmol, 1.0 eq.) in the presence of cyclopropanecarbonyl chloride (26 mg, 307 μmol, 1.0 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH₂Cl₂ (4 mL). ¹H NMR (400 MHz, DMSO-d₆) δ 8.83 (d, J = 2.3 Hz, 1H), 7.76 (d, J = 8.4 Hz, 2H), 7.59 – 7.50 (m, 2H), 3.24 (s, 3H), 1.40 (br, 1H), 0.81 (dt, J = 4.5, 3.3 Hz, 2H), 0.69 (dt, J = 10.2, 3.3 Hz, 2H); ¹³C NMR (101 MHz, DMSO-d₆) δ 172.56, 145.94, 129.55, 128.49, 127.63, 77.48, 37.33, 12.71, 8.64; Rr: TLC: 0.56 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 342.0929 for C₁₇H₁₆F₃NO₂ ([M]+), found 342.0923.

2.4.6-Trifluoro-4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N-methylbenzamide (5). Following general procedure A, compound 5 (80 mg, 186 μmol, 72%) was synthesized as a white solid from 1,1,3,3,hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (70 mg, 256 μmol, 1.0 eq.) in the presence of 2,4,6-trifluorobenzyl chloride (60 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH₂Cl₂ (4 mL). ¹H NMR (400 MHz, DMSO-d₆) δ 8.79 – 8.74 (s, 1H), 7.57 (d, J = 8.5 Hz, 2H), 7.38 (d, J = 8.5 Hz, 2H), 7.32 (dd, J = 8.4, 8.1, 6.3 Hz, 1H), 7.21 (dd, J = 8.1, 1.0, 0.8 Hz, 1H), 7.10 (dd, J = 9.2, 8.4, 1.0 Hz, 1H), 3.43 (s, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 162.52, 159.34, 156.88, 143.98, 132.12, 130.57, 128.07, 126.70, 125.73, 120.00, 114.74, 114.53, 77.02, 38.79; Rr: TLC: 0.29 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 396.0439 for C₁₇H₁₁F₃NO₂ ([M]+), found 396.0428.

2-Chloro-6-fluoro-4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)-N-methylbenzamide (6). Following general procedure A, compound 6 (90 mg, 209 μmol, 82%) was synthesized as a white solid from 1,1,3,3,hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (60 mg, 256 μmol, 1.0 eq.) in the presence of 2-chloro-6-fluorobenzyl chloride (60 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH₂Cl₂ (4 mL). ¹H NMR (400 MHz, DMSO-d₆) δ 8.79 – 8.74 (s, 1H), 7.57 (d, J = 8.5 Hz, 2H), 7.38 (d, J = 8.5 Hz, 2H), 7.32 (dd, J = 8.4, 8.1, 6.3 Hz, 1H), 7.21 (dd, J = 8.1, 1.0, 0.8 Hz, 1H), 7.10 (dd, J = 9.2, 8.4, 1.0 Hz, 1H), 3.43 (s, 3H); ¹³C NMR (101 MHz, DMSO-d₆) δ 161.06, 162.88, 146.44, 132.87, 131.25, 128.91, 128.04, 127.39, 115.36, 77.06, 38.22; Rr: TLC: 0.26 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 342.0636 for C₁₇H₁₁F₃NO₂ ([M]+), found 342.0636.
(methylamino)phenyl)propan-2-ol (60 mg, 256 μmol, 1.0 eq.) in the presence of 3-fluorobenzoyl chloride (49 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH$_2$Cl$_2$ (4 mL). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 8.75 (s, 1H), 7.58 (d, $J = 8.4$ Hz, 2H), 7.35 (d, $J = 8.4$ Hz, 2H), 7.28 (dd, $J = 12.2$, 8.2 Hz 1H), 7.15 (td, $J = 8.6$, 1.8 Hz 1H), 7.13 – 7.03 (m, 2H), 3.40 (s, 3H); $^1$C NMR (101 MHz, DMSO-$d_6$) δ 168.54, 162.99, 160.56, 146.09, 138.67, 130.41, 129.15, 128.04, 127.45, 124.77, 116.90, 115.44, 77.06, 38.10; Rf: TLC: 0.25 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 428.1081 for C$_{17}$H$_{13}$F$_3$NO$_2$ ([M+H]+), found 428.1081.

Following general procedure A, compound 13 (70 mg, 163 μmol, 64%) was synthesized as a white solid from 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (70 mg, 256 μmol, 1.0 eq.) in the presence of naphthalene-1-carbonyl chloride (60 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH$_2$Cl$_2$ (4 mL). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 8.65 (s, 1H), 7.98 (d, $J = 7.9$ Hz, 1H), 7.92 (d, $J = 8.1$ Hz, 1H), 7.87 (s, 1H), 7.64 – 7.50 (m, 2H), 7.46 – 7.39 (m, 5H), 3.43 (s, 3H); $^1$C NMR (101 MHz, DMSO-$d_6$) δ 169.64, 145.62, 134.73, 133.28, 129.15, 127.75, 127.42, 126.84, 126.75, 125.41, 125.20, 124.63, 121.76, 77.01, 37.85; Rf: TLC: 0.48 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 428.1080 for C$_{17}$H$_{13}$F$_3$NO$_2$ ([M+H]+), found 428.1081.

Following general procedure A, compound 14 (55 mg, 133 μmol, 52%) was synthesized as a white solid from 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (60 mg, 256 μmol, 1.0 eq.) in the presence of 2-chlorobenzoyl chloride (54 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH$_2$Cl$_2$ (4 mL). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 8.71 (s, 1H), 7.63 – 7.25 (m, 8H), 3.42 (s, 3H); $^1$C NMR (101 MHz, DMSO-$d_6$) δ 167.13, 144.86, 136.62, 130.78, 130.78, 129.57, 129.83, 127.36, 127.11, 126.15, 124.66, 121.79, 77.06, 36.82; Rf: TLC: 0.23 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 412.0534 for C$_{17}$H$_{13}$F$_3$NO$_2$ ([M+H]+), found 412.0533.

Following general procedure A, compound 15 (50 mg, 130 μmol, 59%) was synthesized as a white solid from 1,1,1,3,3,3-hexafluoro-2-(4-(methylamino)phenyl)propan-2-ol (70 mg, 256 μmol, 1.0 eq.) in the presence of thiophene-2-carbonyl chloride (45 mg, 307 μmol, 1.2 eq.), and triethylamine (36 mg, 50 μL, 278 μmol, 1.1 eq.) in CH$_2$Cl$_2$ (4 mL). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 8.86 (s, 1H), 7.73 (d, $J = 8.4$ Hz, 2H), 7.66 (dd, $J = 5.0$, 1.2 Hz, 1H), 7.55 – 7.47 (d, $J = 8.4$ Hz, 2H), 6.88 (dd, $J = 5.0$, 3.8 Hz, 1H), 6.66 (dd, $J = 3.8$, 1.2 Hz, 1H), 3.37 (s, 3H); $^1$C NMR (101 MHz, DMSO-$d_6$) δ 162.23, 146.08, 138.25, 131.92, 131.81, 130.49, 128.64, 128.45, 127.35, 77.02, 38.88; Rf: TLC: 0.30 (AcOEt/Heptane: 1:2); HRMS (ESI): calculated 384.0487 for C$_{17}$H$_{13}$F$_3$NO$_2$ ([M+H]+), found 384.0481.
