

# De Novo Design of Nurr1 Agonists via Fragment-Augmented Generative Deep Learning in Low-Data Regime

**Citation for published version (APA):**

Ballarotto, M., Willems, S., Stiller, T., Nawa, F., Marschner, J. A., Grisoni, F., & Merk, D. (2023). De Novo Design of Nurr1 Agonists via Fragment-Augmented Generative Deep Learning in Low-Data Regime. *Journal of Medicinal Chemistry*, 66(12), 8170-8177. <https://doi.org/10.1021/acs.jmedchem.3c00485>

**DOI:**

[10.1021/acs.jmedchem.3c00485](https://doi.org/10.1021/acs.jmedchem.3c00485)

**Document status and date:**

Published: 22/06/2023

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# De Novo Design of Nurr1 Agonists via Fragment-Augmented Generative Deep Learning in Low-Data Regime

Marco Ballarotto, Sabine Willems, Tanja Stiller, Felix Nawa, Julian A. Marschner, Francesca Grisoni, and Daniel Merk\*



Cite This: *J. Med. Chem.* 2023, 66, 8170–8177



Read Online

ACCESS |



Metrics & More

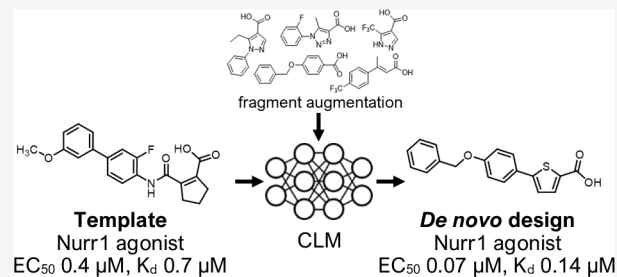


Article Recommendations



Supporting Information

**ABSTRACT:** Generative neural networks trained on SMILES can design innovative bioactive molecules *de novo*. These so-called chemical language models (CLMs) have typically been trained on tens of template molecules for fine-tuning. However, it is challenging to apply CLM to orphan targets with few known ligands. We have fine-tuned a CLM with a single potent Nurr1 agonist as template in a fragment-augmented fashion and obtained novel Nurr1 agonists using sampling frequency for design prioritization. Nanomolar potency and binding affinity of the top-ranking design and its structural novelty compared to available Nurr1 ligands highlight its value as an early chemical tool and as a lead for Nurr1 agonist development, as well as



the applicability of CLM in very low-data scenarios.

## INTRODUCTION

Chemical language models (CLMs)<sup>1–3</sup> are deep learning models trained to generate new molecules from scratch in the form of strings, such as the simplified molecular input line entry system (SMILES).<sup>4</sup> Recently, CLMs have proven to be able to generate novel molecules in a data-driven fashion<sup>5–7</sup> and provide access to new bioactive designs.<sup>1,2,8</sup> CLMs based on recurrent neural networks with long short-term memory<sup>9</sup> have been particularly successful in designing new chemical entities with experimentally confirmed bioactivity on the intended targets.<sup>6,10</sup> This has been achieved *via* transfer learning,<sup>11,12</sup> which uses a model previously trained on a large corpora of data and “fine-tunes” it using a target-focused (and smaller) set of molecules. Despite being successful, the utilized fine-tuning sets typically consisted of several tens of molecules.<sup>6,10,13</sup> However, scenarios of scarce(r) data are common for underexplored targets (e.g., orphan receptors) and have not been investigated for prospective *de novo* design with CLMs, despite their high relevance in medicinal chemistry.

Here, we employ a CLM to design agonists of the orphan nuclear receptor related 1 (Nurr1) *de novo*. This neuroprotective transcription factor emerged as potential target for Alzheimer's disease, Parkinson's disease, and multiple sclerosis treatment.<sup>14,15</sup> Despite the ongoing efforts to develop Nurr1 ligands as chemical tools and drug candidates, potent Nurr1 activators are lacking.<sup>14</sup> We have recently discovered the high-affinity Nurr1 agonist **1** (EC<sub>50</sub> = 0.4 μM, K<sub>d</sub> = 0.7 μM; Figure 1a),<sup>16</sup> which served in this work as template for *de novo* design with a CLM. Due to the very limited availability of Nurr1 agonists, we have “augmented” the set of fine-tuning molecules by adding five weak Nurr1 agonist fragments (1.2–1.8-fold Nurr1 activation at

100 μM), which are structurally related to **1** (Table S1). After a two-stage fine-tuning procedure, six *de novo* designs were synthesized and tested for Nurr1 agonism. Two designs displayed the desired bioactivity, with one of them exhibiting remarkable Nurr1 agonist potency. Our results indicate the potential of CLMs to enable *de novo* design in extremely low-data regimes.

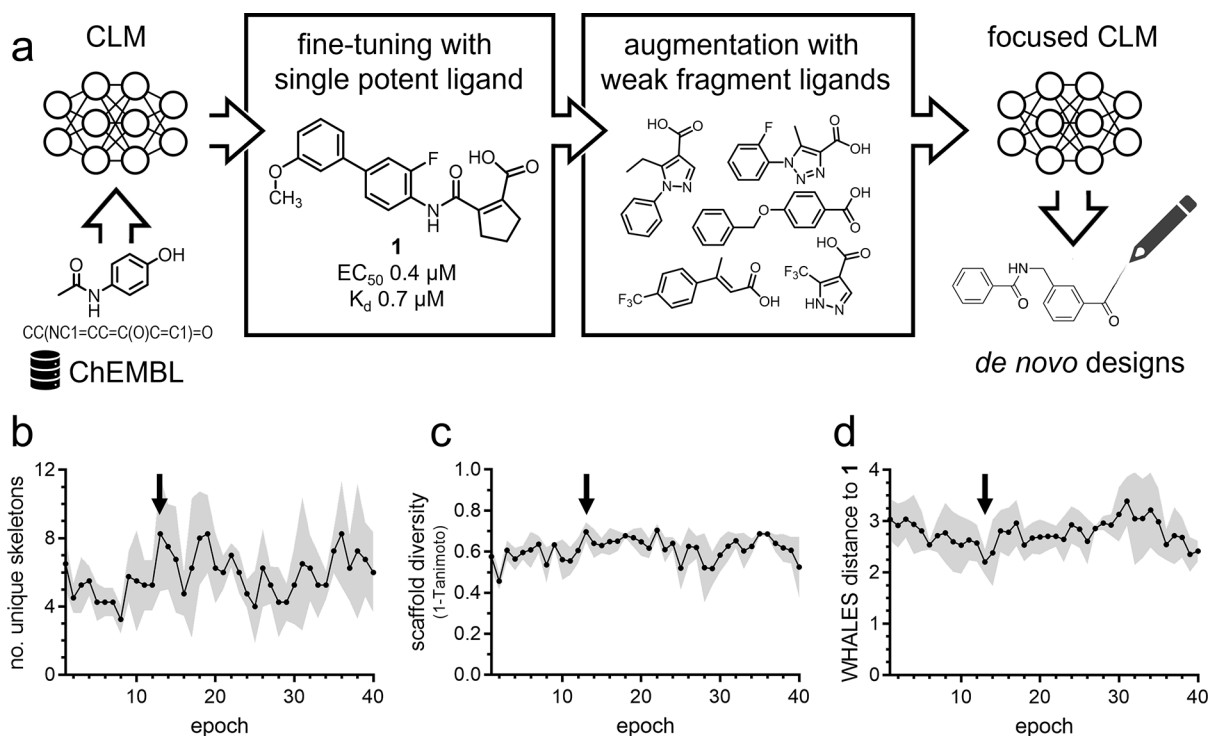
## RESULTS AND DISCUSSION

**CLM Training and De Novo Design.** We employed a two-step fine-tuning procedure, to achieve *de novo* design by CLMs from a very restricted number of known ligands (Figure 1a). Previous studies on CLMs<sup>12</sup> have observed that fine-tuning with a single template reduces the chemical diversity of the designs providing a challenge to our objective. We used a previously published CLM, which was pre-trained on 365k molecules from ChEMBL<sup>17</sup> and captures the “syntax” of SMILES strings (*i.e.*, how to generate chemically valid strings) and some general molecular properties.<sup>10</sup> The model was fine-tuned with **1** for 40 epochs, with a 10-fold SMILES augmentation procedure,<sup>18</sup> *i.e.*, by using ten different SMILES strings representing **1**. We then used beam search<sup>19</sup> to identify epochs at which the model was biased toward designing molecules most similar to **1**. Beam search is a heuristic sampling approach, which progressively

Received: March 18, 2023

Published: May 31, 2023





**Figure 1.** Schematic overview of the CLM training. (a) The starting point was an existing CLM<sup>10</sup> pre-trained on 365k general bioactive compounds from ChEMBL expressed as SMILES as described previously. Fine-tuning of the CLM was then achieved by a two-step procedure. Initial fine-tuning was performed using the potent Nurr1 agonist **1** as single template until analysis of beam search designs indicated that the model had captured features of **1**. Subsequently, the fine-tuning was augmented with weak Nurr1 activating fragments. (b–d) Analysis of beam search designs during the initial fine-tuning indicated epoch 13 as favored in terms of the number of unique skeletons, scaffold diversity, and WHALES distance to **1**, indicating that the model had captured features of the template. Data are the mean (black lines)  $\pm$  SD (gray area) from four repeated trainings.

extends a SMILES string by adding, at each step, the  $k$  characters with the highest conditional probability, given the previous portion of the string (in this work,  $k = 50$ ). We have previously employed beam search to monitor the information captured by a CLM and to devise an optimal training strategy.<sup>10</sup> Beam search sampling from epoch 13 yielded a high number of unique skeletons (graph scaffolds, Figure 1b) and a high average scaffold diversity (1-Tanimoto) based on Morgan fingerprints,<sup>20</sup> which capture common substructures (Figure 1c). At the same time, designs from epoch 13 exhibited the highest similarity to the template **1** based on the Weighted Holistic Atom Localization and Entity Shape (WHALES)<sup>21</sup> descriptors, which encode three-dimensional molecular shape and partial charges (Figure 1d). These results thus indicated that the CLM had captured features of the single template after epoch 13 without a loss of chemical diversity.

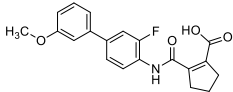
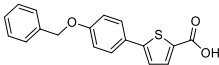
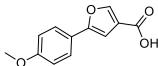
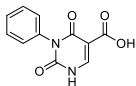
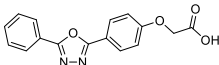
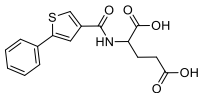
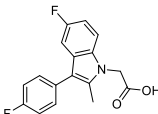
To maintain molecular diversity and explore the chemical space in the vicinity of **1**, we used a “fragment-augmented” approach for further fine-tuning using five weak Nurr1 activating fragments (**2–6**, Table S1) that shared structural features with **1**. The CLM was trained over 100 epochs, and designs were generated for every epoch by beam search, as for the previous step. From this collection of designs, we prioritized molecules for synthesis by two approaches (Table 1). As a first selection criterion, we exploited the sampling frequency of a given molecule as a “CLM-intrinsic” measure of priority for the generated designs. This was done under the assumption that, the higher the frequency of generation, the higher the probability of a given structure for the CLM. This highlighted compounds **7** (50 $\times$  generated, rank 1), **8** (18 $\times$ , rank 2), and **9** (17 $\times$ , rank 4) as the most frequently sampled molecules. The third and fifth

ranked molecules (Table S2) shared the scaffolds of **7** and **8**, respectively, further highlighting them as preferred by the model. Additionally, we selected three molecules (**10–12**) exhibiting high similarity to the template **1** based on physicochemical descriptors and WHALES (Table S3), which have proved suitable for external ranking in previous applications of CLM for *de novo* design.<sup>6,22</sup> Sampling frequency and descriptor similarity prioritized different sets of molecules. The highly frequent designs **7–9** comprised lower similarity to **1**, while the most similar molecules **10–12** were not frequently sampled.

#### Synthesis and *In Vitro* Characterization of Designs.

The computationally selected molecules **7–12** were synthesized over two to four steps according to Scheme 1. **7** was obtained by Suzuki reaction of **13** and **14** to **15** followed by Williamson ether formation with benzyl bromide and ester hydrolysis. For preparation of **8**, **17** was esterified to **18** and then coupled with **19** in a Suzuki reaction to **20** before ester hydrolysis afforded **8**. Synthesis of the pyrimidinedione **9** commenced with the condensation of phenyl urea (**21**), triethyl orthoformate, and diethyl malonate to **22** which cyclized to **9** under basic conditions. Design **10** was prepared by treating 4-hydroxybenzaldehyde (**23**) with ethyl bromoacetate to introduce the side chain in **24**. Subsequent oxidation and cyclization of the obtained free carboxylate with benzohydrazide gave oxadiazole **25** that underwent ester hydrolysis to give **10**. **11** was obtained from methyl 2-bromothiophene-4-carboxylate (**26**) by Suzuki reaction with phenylboronic acid to **27**, ester hydrolysis and amide coupling with dimethyl glutamate to **28** and another ester hydrolysis. Lastly, preparation of **12** succeeded by Fischer indole synthesis of 4-fluorophenylhydrazide hydrochloride (**29**) and 4-

Table 1. Ranking and Biological Activity of *De Novo* Designs 7–12<sup>a</sup>

ID	structure	frequency	distance <sup>b</sup>	EC <sub>50</sub> (Nurr1) [fold act.] <sup>c</sup>		K <sub>d</sub> (Nurr1 LBD)
				Gal4-Nurr1	NBRE	
1		-	-	0.4±0.2 μM [3.1±0.4-fold]	0.3±0.1 μM [3.0±0.2-fold]	0.7 μM
7		50 (rank 1)	0.220 (rank 88)	0.07±0.02 μM [1.7±0.1-fold]	0.04±0.01 μM [1.6±0.1-fold]	0.14 μM
8		18 (rank 2)	0.255 (rank 122)	2.1±0.6 μM [2.5±0.1-fold]	1.7±0.6 μM [4.5±0.4-fold]	2.4 μM
9		17 (rank 4)	0.242 (rank 108)	inactive (100 μM)	n.d.	n.d.
10		2 (rank 66)	0.110 (rank 1)	inactive (50 μM <sup>d</sup> )	n.d.	n.d.
11		7 (rank 18)	0.123 (rank 5)	inactive (100 μM)	n.d.	n.d.
12		1 (rank 90)	0.137 (rank 7)	inactive (30 μM <sup>d</sup> )	n.d.	n.d.

<sup>a</sup>Template 1 for comparison. <sup>b</sup>Euclidean distance to 1 based on WHALES<sup>21</sup> and physicochemical descriptors. <sup>c</sup>Nurr1 agonism was determined in a Gal4 hybrid reporter gene assay and a reporter gene assay for full-length human Nurr1 using the NBRE.<sup>23</sup> Data are shown as mean ± SD, *n* ≥ 3. <sup>d</sup>Highest non-toxic concentration.

fluorophenylacetone (30), *N*-alkylation of the resulting indole 31 with ethyl bromoacetate, and ester hydrolysis. The computational designs 7–12 were hence accessible by standard chemistry and mostly obtained in very good yields, demonstrating that the designs met the key criterion of synthesizability.

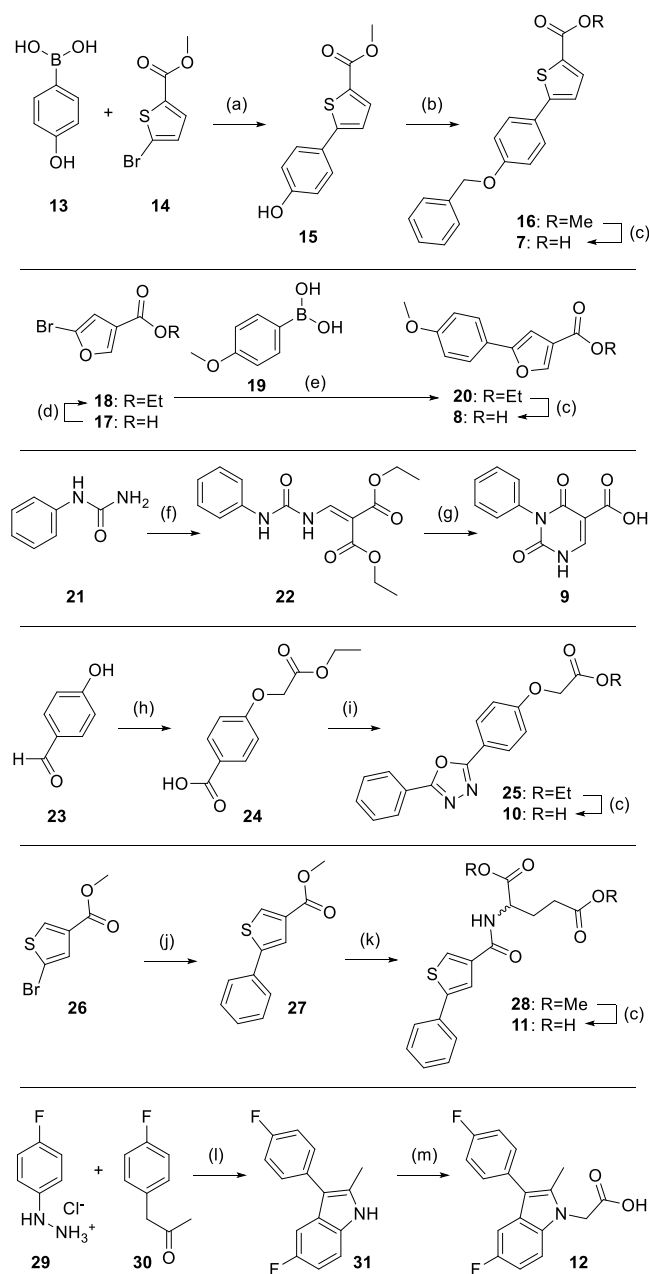
Nurr1 modulation by 7–12 (Table 1) was determined in a Gal4-Nurr1 hybrid reporter gene assay<sup>23</sup> in transiently transfected HEK293T cells with Gal4-responsive firefly luciferase as the reporter and constitutively expressed Renilla luciferase as control gene. The most frequently sampled design 7 exhibited remarkable Nurr1 agonism in this assay with an EC<sub>50</sub> value of 0.07 μM supporting sampling frequency as a valuable measure of design quality. The second-most computationally preferred design 8 activated Nurr1 with low micromolar potency and 9 was inactive. Designs 10–12 which were selected based on descriptor similarity to template 1 showed no Nurr1 agonism. Isothermal titration calorimetry (ITC, Table 1, Figure 2a,b) confirmed binding of 7 and 8 to the recombinant Nurr1 ligand binding domain (LBD) with K<sub>d</sub> values of 0.14 μM (7) and 2.4 μM (8), providing orthogonal validation of on-target activity. Moreover, both Nurr1 agonists activated full-length human Nurr1 on its monomer response element NBRE with comparable potency as observed on the Gal4 hybrid receptor (Table 1). The most active design 7 was additionally profiled for activation of Nurr1 as heterodimer with RXR on the DR5 response element which yielded consistent results (EC<sub>50</sub> = 0.03 ± 0.01 μM, 2.1 ± 0.1-fold activation). In Nurr1 expressing astrocytes (T98G cells), 7 and 8 induced the Nurr1-regulated genes tyrosine hydroxylase (TH) and vesicular amino acid transporter 2 (VMAT2) in a dose-dependent fashion (Figure

2c,d), supporting their value as a chemical tool for studies on Nurr1 biology and as leads.

**Structural Novelty of the *De Novo* Designs.** The *de novo* design 7 is among the most potent Nurr1 agonists discovered so far<sup>14</sup> (Figure 3a) and is characterized by a favorable ligand efficiency. Comparison of 7 with the template set and with known Nurr1 agonists (Figure 3b,c, Table S5) additionally demonstrated high structural novelty in terms of substructure and scaffold similarity. The remarkable biological activity and structural novelty of 7 suggest that the CLM has accessed an uncharted region of Nurr1 ligands in chemical space. The Nurr1 agonist 7 constitutes an important addition to the available collection of ligands for this understudied receptor and appears highly suitable as a lead for systematic optimization toward Nurr1 modulating chemical tools.

## CONCLUSIONS

CLM and other approaches to machine learning-driven *de novo* design<sup>7,24</sup> are a very active field of theoretical work. Prospective experimental applications of AI to design new bioactive compounds are still rare and have mostly focused on target proteins with many known ligands.<sup>6</sup> This is not surprising since deep learning is “data-hungry” and thus requires datasets of sufficient size for training. For this reason, designing mimetics from single-template molecules is more challenging than with rule-driven *de novo* design algorithms.<sup>25</sup> We have employed a CLM to develop agonists for the orphan receptor Nurr1 for which very little ligand knowledge<sup>14</sup> is available. Our pipeline yielded a high-affinity Nurr1 ligand (7) from only one potent template molecule, by a fragment-augmentation procedure. For design prioritization, we confirmed the sampling frequency as a

Scheme 1. Synthesis of *De Novo* Designs 7–12<sup>a</sup>

<sup>a</sup>Reagents and conditions: (a)  $K_2CO_3$ ,  $Pd(PPh_3)_4$ , THF/ $H_2O$  (1:1), reflux, 16 h, then  $H_2SO_4$  (cat.), MeOH, reflux, 47%; (b)  $K_2CO_3$ , BnBr, DMF, r.t., 5 h, quant.; (c) KOH, THF/MeOH/ $H_2O$  (3:2:1), r.t., 16 h, 65–82%; (d)  $H_2SO_4$  (cat.), EtOH, reflux, 24 h, 92%; (e)  $K_2CO_3$ ,  $Pd(PPh_3)_4$ , DME/ $H_2O$  (14:6), 95 °C, 5 h, 89%; (f) diethyl malonate, triethyl orthoformate, neat, 130 °C, 4 h, 27%; (g) KOtBu, tBuOH, 130 °C, 16 h, 11%; (h) ethyl 2-bromoacetate,  $K_2CO_3$ , acetone, reflux, 2 h, then  $NaH_2PO_4$ ,  $NaClO_2$ , MeCN/ $H_2O$  (1:2), 5 °C to r.t., 16 h, 76% over two steps; (i) benzohydrazide, TBTU, DIPEA, MeCN, r.t., 16 h then DIPEA, TsCl, MeCN, r.t., 24 h, 87% over two steps; (j) phenylboronic acid,  $K_2CO_3$ ,  $Pd(PPh_3)_4$ , DME, reflux, 16 h, 95%; (k) KOH, THF/MeOH/ $H_2O$  (3:2:1), r.t., 16 h, then dimethyl glutamate, COMU, DIPEA, DMF, r.t., 16 h, 96% over two steps; (l) conc. HCl, EtOH, reflux, 16 h, 99%; (m) ethyl bromoacetate, NaH, DMF, r.t., 24 h, then KOH, THF/MeOH/ $H_2O$  (3:2:1), r.t., 16 h, 14% over two steps.

good model-intrinsic measure which avoids external ranking. The success of this approach (2/3 designs active) supported its

potential as internal design quality indicator. Our results show the potential of CLMs for *de novo* ligand design for orphan targets with little previous knowledge on ligands and confirm sampling frequency as another approach for model intrinsic design prioritization. Our AI-driven design strategy led to a novel, high-affinity Nurr1 agonist, exceeding available Nurr1 activators in potency, highlighting the still not fully exploited potential of machine learning in the development of chemical tools for underexplored targets.

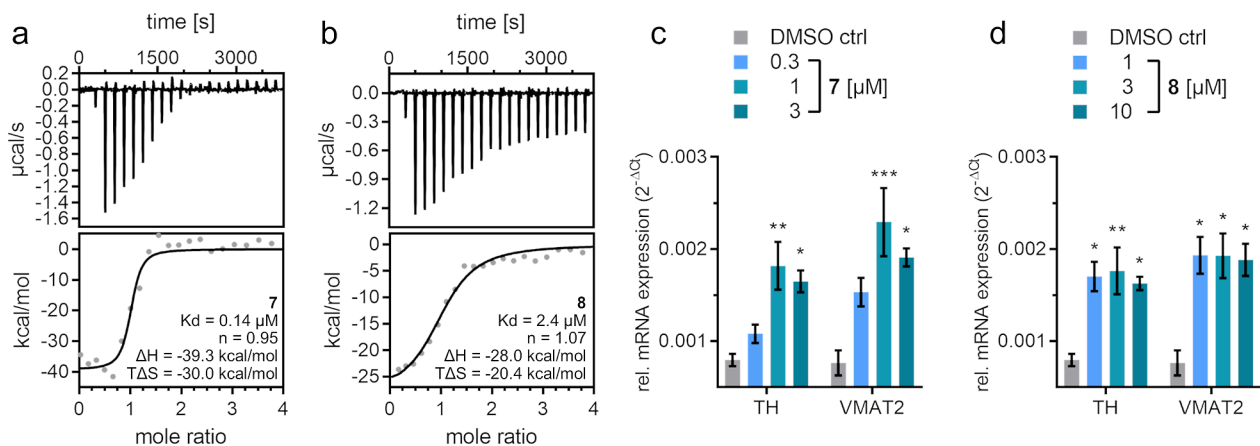
The neuroprotective transcription factor Nurr1 holds great promise as a potential therapeutic target, but Nurr1 modulators are rare. Structural novelty compared to known Nurr1 ligands and superior affinity therefore highlight the value of *de novo* design 7 as a lead for medicinal chemistry to develop potent Nurr1 agonists as chemical tools and drug candidates.

## EXPERIMENTAL SECTION

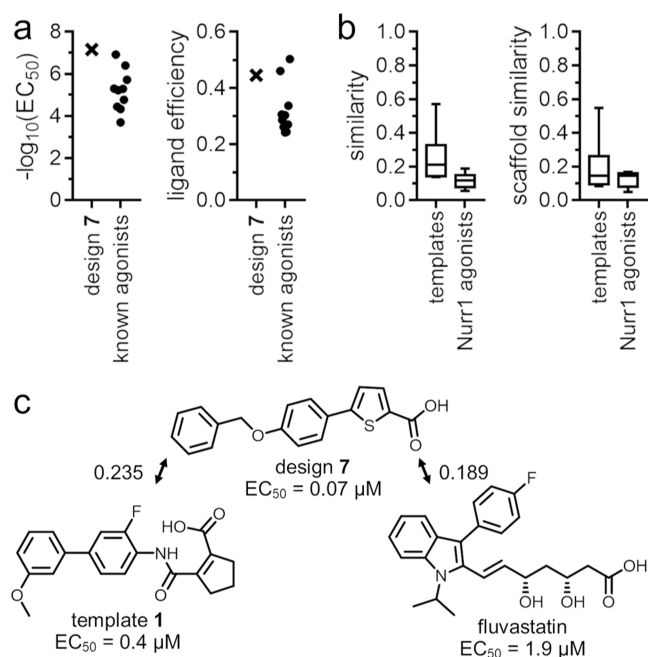
**Chemistry. General.** All final compounds to be tested in the biological evaluation had a purity of >95% according to quantitative NMR. NMR data are provided in the Supporting Information. General procedures as well as synthesis and analytical characterization of 8–12 and precursors are described in the Supporting Information.

**5-(4-(Benzyloxy)phenyl)thiophene-2-carboxylic Acid (7).** Methyl 5-(4-hydroxyphenyl)thiophene-2-carboxylate (15, 0.094 g, 0.4 mmol, 1.0 equiv) and  $K_2CO_3$  (0.17 g, 1.2 mmol, 3.0 equiv) were dissolved in dimethylformamide (DMF, 4 mL, 0.1 M).<sup>26</sup> Benzyl bromide (0.07 mL, 0.6 mmol, 1.5 equiv) was added in one portion, and the mixture was stirred at room temperature for 5 h when thin-layer chromatography (TLC) showed no remaining starting material. The mixture was then diluted with water (30 mL) and 2 M aqueous hydrochloric acid (3 mL) and extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with 5% LiCl solution (1 × 10 mL) and brine (1 × 15 mL) and dried over  $Na_2SO_4$ . The solvent was removed *in vacuo* to give 16 as a colorless solid in quantitative yield. An aqueous solution of KOH (30% w/v, 1 mL) was added to a solution of the methyl ester 16 (0.13 g, 0.4 mmol) in tetrahydrofuran (THF, 3 mL, 0.15 M). MeOH (2 mL) was added, and the mixture was stirred at room temperature overnight.<sup>27</sup> TLC indicated that the reaction was completed. The volatiles were removed *in vacuo*, the residue was diluted with 2 M aqueous hydrochloric acid (8 mL) and water (15 mL), and the mixture was extracted with EtOAc (3 × 15 mL). The combined organic layers were washed with brine and dried over  $Na_2SO_4$ . The solvent was removed *in vacuo* and the residue was purified by reverse-phase column chromatography ( $H_2O$ /MeCN 95:5 → 0:100 over 12 CV) to obtain the title compound 7 as a pale-yellow solid (0.10 g, 0.32 mmol, 81% over 2 steps). mp: 238–239 °C (dec.).  $R_f$  (isohexanes/EtOAc/AcOH, 50:49:1) = 0.47.  $^1H$  NMR (500 MHz,  $DMSO-d_6$ ):  $\delta$  13.05 (br s, 1H), 7.73–7.60 (m, 3H), 7.49–7.43 (m, 3H), 7.43–7.37 (m, 2H), 7.37–7.30 (m, 1H), 7.09 (d,  $J$  = 8.6 Hz, 2H), 5.16 (s, 2H).  $^{13}C$  NMR (126 MHz,  $DMSO-d_6$ ):  $\delta$  162.9, 158.9, 149.8, 136.8, 134.3, 132.3, 128.5, 127.9, 127.7, 127.3, 125.7, 123.4, 115.5, 69.3. HRMS (EI)  $m/z$ : calcd 310.0664 for  $C_{18}H_{14}O_3S^+$ ; found, 310.0652 ( $M^+$ ).

**Methyl 5-(4-Hydroxyphenyl)thiophene-2-carboxylate (15).** Under a nitrogen atmosphere, 4-hydroxyphenylboronic acid (13, 0.33 g, 2.4 mmol, 1.2 equiv), methyl 2-bromothiophene-5-carboxylate (14, 0.45 g, 2.0 mmol, 1.0 equiv), and  $K_2CO_3$  (1.38 g, 10 mmol, 5.0 equiv) were dissolved in a mixture of THF/ $H_2O$  (1:1 v/v, 10 mL total, 0.2 M).<sup>28</sup> The mixture was degassed by purging with nitrogen using a needle submerged into the solution for 10 min. After that,  $Pd(PPh_3)_4$  (0.12 g, 0.1 mmol, 0.05 equiv) was added, and the reaction mixture was heated to reflux overnight. After 18 h, TLC indicated the completion of the reaction, and the mixture was cooled to room temperature, diluted with water (20 mL), adjusted to pH 2–3 with 2 M aqueous hydrochloric acid, and extracted with EtOAc (3 × 20 mL). The combined organic layers were washed with brine (1 × 20 mL) and dried over  $Na_2SO_4$ , and the solvents were removed *in vacuo*. The residue was subjected to a Fischer esterification in MeOH (30 mL) and 5 drops of conc.  $H_2SO_4$  at 80 °C for 24 h. The reaction was then cooled to room temperature and



**Figure 2.** Characterization of *de novo* designs 7 and 8 as Nurr1 agonists. (a,b) ITC demonstrated high-affinity binding of 7 (a) and 8 (b) to the Nurr1 LBD. The upper panels show the isotherms of the compound–protein titrations and the lower panels show the fitting of the blank-corrected heat of binding. (c,d) 7 and 8 induced mRNA expression of the Nurr1-regulated genes TH and VMAT2 in T98G astrocytes. Data are shown as mean  $\pm$  S.E.M.,  $n = 4$ . \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  (vs DMSO-treated cells; ANOVA with Bonferroni correction).



**Figure 3.** (a) *De novo* design 7 is among the most potent available Nurr1 agonists and exhibits favorable ligand efficiency. Known agonists used for comparison are listed in Table S4. (b) *De novo* design 7 is structurally novel compared to the molecules used for CLM fine-tuning (templates) and to known Nurr1 agonists in terms of *Tanimoto* similarity computed on *Morgan fingerprints* and on *Murcko scaffolds* (see Table S5 for individual values). (c) *Tanimoto* similarity computed on *Morgan fingerprints* of the *de novo* design 7 to the template 1 and the most similar known Nurr1 ligand fluvastatin not used for CLM fine-tuning.

the solvent was removed *in vacuo*. The residue was dissolved in 10% (w/v) NaHCO<sub>3</sub> solution (30 mL) and extracted with EtOAc (3  $\times$  20 mL). The combined organic layers were washed with brine and dried over Na<sub>2</sub>SO<sub>4</sub>, and the solvent was removed *in vacuo*. The crude product was purified by automatic flash column chromatography (isohexanes/EtOAc, 80:20) to obtain the title compound 15 as a pale-yellow solid (0.22 g, 0.93 mmol, 47%). mp: 171–172 °C. R<sub>f</sub> (isohexanes/EtOAc 70:30) = 0.57. <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.71 (d,  $J = 4.0$  Hz, 1H), 7.55–7.49 (m, 2H), 7.26 (d,  $J = 4.0$  Hz, 1H), 6.86–6.79 (m, 2H), 3.86 (s, 3H). <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD):  $\delta$  164.3, 159.8, 153.6,

135.8, 131.3, 128.6, 126.1, 123.4, 116.9, 52.6. MS (APCI)  $m/z$ : 234.9 ([M + H]<sup>+</sup>).

**Biological Characterization.** *Gal4-Nurr1 Hybrid Reporter Gene Assay.* Nurr1 modulation was determined in a Gal4 hybrid reporter gene assay in HEK293T cells (German Collection of Microorganisms and Cell Culture GmbH, DSMZ) using pFR-Luc (Stratagene, La Jolla, CA, USA; reporter), pRL-SV40 (Promega, Madison, WI, USA; internal control), and pFA-CMV-hNurr1-LBD,<sup>23</sup> coding for the hinge region and the ligand binding domain of the canonical isoform of human Nurr1. Cells were cultured in Dulbecco's modified Eagle's medium (DMEM), high glucose supplemented with 10% fetal calf serum (FCS), sodium pyruvate (1 mM), penicillin (100 U/mL), and streptomycin (100 μg/mL) at 37 °C and 5% CO<sub>2</sub> and seeded in 96-well plates (3  $\times$  10<sup>4</sup> cells/well). After 24 h, the medium was changed to Opti-MEM without supplements, and cells were transiently transfected using Lipofectamine LTX reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Five hours after transfection, cells were incubated with the test compounds in Opti-MEM supplemented with penicillin (100 U/mL), streptomycin (100 μg/mL), and 0.1% dimethyl sulfoxide (DMSO) for 16 h before luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol on a Tecan Spark luminometer (Tecan Deutschland GmbH, Crailsheim, Germany). Firefly luminescence was divided by Renilla luminescence and multiplied by 1000 resulting in relative light units (RLU) to normalize for transfection efficiency and cell growth. Fold activation was obtained by dividing the mean RLU of the test compound by the mean RLU of the untreated control. All samples were tested in at least three biologically independent experiments in duplicate. For dose–response curve fitting and calculation of EC<sub>50</sub> values, the equation “[Agonist] vs response - variable slope (four parameters)” was used in GraphPad Prism (version 7.00, GraphPad Software, La Jolla, CA, USA).

**Full-Length Nurr1 Reporter Gene Assays.** Activation of full-length human Nurr1 was studied in transiently transfected HEK293T cells using the reporter plasmid pFR-Luc-NBRE<sup>23</sup> or pFR-Luc-DRS,<sup>23</sup> each containing one copy of the respective human Nurr1 response element NBRE N13 (TGA TAT CGA AAA CAA AAG GTC A) or DRS (TGA TAG GTT CAC CGA AAG GTC A). The full-length human nuclear receptor Nurr1 (pcDNA3.1-hNurr1-NE; Addgene plasmid no. 102363) and, for DRS, additionally RXR $\alpha$  (pSG5-hRXR)<sup>29</sup> were overexpressed. pRL-SV40 (Promega) was used for normalization of transfection efficacy and to observe test compound toxicity. Cells were cultured in DMEM, high glucose supplemented with 10% FCS, sodium pyruvate (1 mM), penicillin (100 U/mL), and streptomycin (100 μg/mL) at 37 °C and 5% CO<sub>2</sub> and seeded in 96-well plates (3  $\times$  10<sup>4</sup> cells/well). After 24 h, the medium was changed to Opti-MEM without supplements, and cells were transiently transfected using Lipofectamine

LTX reagent (Invitrogen) according to the manufacturer's protocol. Five hours after transfection, cells were incubated with the test compounds in Opti-MEM supplemented with penicillin (100 U/mL), streptomycin (100  $\mu$ g/mL), and 0.1% DMSO for 16 h before luciferase activity was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's protocol on a Tecan Spark luminometer (Tecan Deutschland GmbH). Firefly luminescence was divided by Renilla luminescence and multiplied by 1000 resulting in relative light units (RLU) to normalize for transfection efficiency and cell growth. Fold activation was obtained by dividing the mean RLU of the test compound by the mean RLU of the untreated control. All samples were tested in at least three biologically independent experiments in duplicate. For dose–response curve fitting and calculation of EC<sub>50</sub> values, the equation “[Agonist] vs response - variable slope (four parameters)” was used in GraphPad Prism (version 7.00, GraphPad Software).

**Recombinant Expression and Purification of Nurr1 LBD Protein.** The Nurr1 LBD (aa 362–598) was subcloned into pNIC28-Bsa4. The recombinant protein containing an N-terminal His<sub>6</sub>-tag was expressed in *Escherichia coli* BL21(DE3)-R3-pRARE2, cultured in TB. Expression was induced with 0.5 mM IPTG at 18 °C overnight. The protein was purified by Ni<sup>2+</sup>-affinity chromatography followed by TEV treatment to remove the histidine tag. The cleaved protein was further purified by reverse Ni<sup>2+</sup>-affinity chromatography and size exclusion chromatography.

**Isothermal Titration Calorimetry.** ITC experiments were conducted on an Affinity ITC instrument (TA Instruments, New Castle, DE) at 25 °C with a stirring rate of 75 rpm. Nurr1 LBD protein (10 or 20  $\mu$ M) in buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 5% glycerol) containing 5% DMSO was titrated with the test compounds (50 or 100  $\mu$ M in the same buffer containing 5% DMSO) in 21 injections (1  $\times$  1 and 20  $\times$  5  $\mu$ L) with an injection interval of 180 s. As control experiments, the test compounds were titrated to the buffer, and the buffer was titrated to the Nurr1 LBD protein under otherwise identical conditions. The heats of the compound–protein titrations were corrected with the heats of the compound–buffer titrations and analyzed using NanoAnalyze software (version 3.11.0, TA Instruments, New Castle, DE) with independent binding models.

**Evaluation of Nurr1 Regulated Gene Expression.** T98G (ATCC CRL-1690) were grown in DMEM, high glucose supplemented with 10% FCS, sodium pyruvate (1 mM), penicillin (100 U/mL), and streptomycin (100  $\mu$ g/mL) at 37 °C and 5% CO<sub>2</sub> and seeded at a density of 2.5  $\times$  10<sup>5</sup> cells/well in a 12-well plate. After 24 h, medium was changed to DMEM, high glucose supplemented with 0.2% FCS, penicillin (100 U/mL), and streptomycin (100  $\mu$ g/mL), and the cells were incubated for another 24 h before stimulation with the test compounds in the same medium additionally containing 0.1% DMSO. After 16 h of incubation, the medium was removed, cells were washed with phosphate-buffered saline, and after full aspiration of residual liquids immediately frozen at –80 °C until further procession. Total RNA was isolated using E.Z.N.A. Total RNA Kit I (Omega Bio-tek, Norcross, GA, USA) following the manufacturer's instructions. The RNA concentration and purity were assessed using a NanoDrop One UV/vis spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) at 260/280 nm. Right before reverse transcription (RT), RNA was linearized at a concentration of 133 ng/ $\mu$ L at 65 °C for 10 min and then immediately incubated on ice for at least 1 min. RT was performed using 2  $\mu$ g total RNA, 20 U Recombinant RNasin Ribonuclease Inhibitor (Promega), 100 U SuperScript IV Reverse Transcriptase including 5 $\times$  first strand buffer and 0.1 M dithiothreitol (Thermo Fisher Scientific), 3.75 ng linear acrylamide, 625 ng random hexamer primers (#11277081001, Merck, Darmstadt, Germany), and 11.25 nmol deoxynucleoside triphosphate mix (2.8 nmol of each ATP, TTP, CTP, and GTP; #R0186, Thermo Fisher Scientific) at a volume of 22.45  $\mu$ L at 50 °C for 10 min and 80 °C for 10 min using a Thermal cycler XT<sup>96</sup> (VWR International, Darmstadt, Germany). Quantitative polymerase chain reaction (qPCR) was conducted using an Applied Biosystems QuantStudio 1 (Waltham, MA, USA) and a SYBR green-based detection method. 0.2  $\mu$ L of prepared cDNA (complementary DNA) was added to 6 pmol of forward and reverse primers,

respectively, 0.8 U Taq DNA Polymerase (#M0267, New England Biolabs, Ipswich, MA, USA), 40 ppm SYBR Green I (#S9430, Sigma-Aldrich, St. Louis, MO, USA), 15 nmol deoxynucleoside triphosphate mix (as indicated above), 60 nmol MgCl<sub>2</sub>, 4  $\mu$ g bovine serum albumin (#B14, Thermo Fisher Scientific), 20% BioStab PCR Optimizer II (#53833, Merck), and 10% Taq buffer without detergents (#B55, Thermo Fisher Scientific) were added and the final volume was made up to 20  $\mu$ L with ddH<sub>2</sub>O. Samples underwent 40 cycles of 15 s denaturation at 95 °C, 15 s of primer annealing at 62.4 °C, and 20 s of elongation at 68 °C. PCR product specificity was evaluated using a melting curve analysis ranging from 65 to 95 °C. TH and VMAT2 mRNA (messenger RNA) expression was normalized to GAPDH mRNA expression per each sample using the  $\Delta$ Ct-method. The following primers were used. VMAT2 (SLC18A2): 5'-GCT ATG CCT TCC TGC TGA TTG C-3' (fw) and 5'-CCA AGG CGA TTC CCA TGA CGT T-3' (rev); TH: 5'-GCT GGA CAA GTG TCA TCA CCT G-3' (fw) and 5'-CCT GTA CTG GAA GGC GAT CTC A-3' (rev); GAPDH: 5'-AGG TCG GAG TCA ACG GAT TT-3' (fw) and 5'-TTC CCG TTC TCA GCC TTG AC-3' (rev). Gene expression data were analyzed with QuantStudio Design & Analysis Software (v1.5.2; Applied Biosystems) and MS Excel. Statistical significance was evaluated by analysis of variance (ANOVA) with Bonferroni correction in R (v4.1.2; The R Foundation for Statistical Computing).

**Computational Procedures. Fine-Tuning Sets.** The first fine-tuning set comprising compound **1** and the second fine-tuning set comprising five Nurr1-activating fragments are shown in Figure 1.

**Canonical SMILES Generation.** The canonical SMILES of molecules used as fine-tuning sets were generated with RDKit (v. 2022.03.1, [www.rdkit.org](http://www.rdkit.org)). The SMILES for CLM fine-tuning with data augmentation are shown in Table S6. All other input data for the CLM have been described previously and are available at [https://github.com/ETHmodlab/molecular\\_design\\_with\\_beam\\_search](https://github.com/ETHmodlab/molecular_design_with_beam_search).

**Chemical Language Model.** This work is based on a previously described CLM<sup>10</sup> which is freely available at [https://github.com/ETHmodlab/molecular\\_design\\_with\\_beam\\_search](https://github.com/ETHmodlab/molecular_design_with_beam_search). We used the available pre-trained model and performed two consecutive rounds of fine-tuning. We used the categorical cross-entropy loss and the Adam optimizer. The first round of fine-tuning with **1** as the fine-tuning set was performed with 10-fold data augmentation for 40 epochs with a learning rate of 10<sup>−4</sup> by keeping the first layer frozen. The batch size was 10. The second round of fine-tuning with **1** and five Nurr1-activating fragments was performed from epoch 13 of the first fine-tuning with 2-fold data augmentation for 50 epochs with a learning rate of 10<sup>−4</sup> by keeping the first layer 1 frozen. The batch size was 50.

**Beam Search Ranking.** The beam search algorithm was applied as previously described.<sup>10</sup> We used a beam search width of 50 and defined the maximum SMILES string length as 140 tokens.

**Molecular Descriptors.** Molecular geometry and WHALES descriptors<sup>21</sup> were computed with the code freely available at [https://github.com/grisoniFr/scaffold\\_hopping\\_whales](https://github.com/grisoniFr/scaffold_hopping_whales) (v1), with default settings, as previously described. Molecular descriptors (molecular weight, total polar surface area, Wildman–Crippen logP, number of rotatable bonds, and Lipinski-type hydrogen bond donors and hydrogen bond acceptors) were generated using RDKit (v. 2022.03.1). The two sets of descriptors for **1** and the beam search designs were concatenated and normalized. The Euclidean distance based on the descriptors was computed between **1** and the beam search designs.

**Sampling Frequency.** The sampling frequencies of the generated beam search designs from the second fine-tuning step were calculated over all 50 epochs based on the generated SMILES.

**Fingerprint Calculation and Data Analysis.** Fingerprint calculation and data analysis was performed in KNIME (v4.4.4, KNIME AG, Zurich, Switzerland) using RDKit and CDK nodes.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jmedchem.3c00485>.

Nurr1 activating fragments related to **1** used for CLM fine-tuning, top-ranking CLM designs based on sampling frequency and molecular descriptor similarity, Nurr1 agonists used for comparison with *de novo* design **7**, Tanimoto similarity of **7** to the fine-tuning molecules **1–6** computed on *Morgan Fingerprints* and *Murcko* scaffolds, SMILES used for CLM fine-tuning, synthetic procedures, and analytical data of **7–17** (PDF)

Molecular formula strings containing chemical structures and biological characterization data of **7–12** (CSV)

## AUTHOR INFORMATION

### Corresponding Author

Daniel Merk – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany; [orcid.org/0000-0002-5359-8128](https://orcid.org/0000-0002-5359-8128); Email: [daniel.merk@cup.lmu.de](mailto:daniel.merk@cup.lmu.de)

### Authors

Marco Ballarotto – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany; Department of Pharmaceutical Sciences, Università degli Studi di Perugia, 06123 Perugia, Italy; [orcid.org/0000-0003-2361-4841](https://orcid.org/0000-0003-2361-4841)

Sabine Willems – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany; [orcid.org/0000-0002-9755-3394](https://orcid.org/0000-0002-9755-3394)

Tanja Stiller – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany

Felix Nawa – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany

Julian A. Marschner – Department of Pharmacy, Ludwig-Maximilians-Universität (LMU) München, 81377 Munich, Germany

Francesca Grisoni – Institute for Complex Molecular Systems, Department of Biomedical Engineering, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands; Centre for Living Technologies, Alliance TU/e, WUR, UU, UMC Utrecht, 3584CB Utrecht, The Netherlands; [orcid.org/0000-0001-8552-6615](https://orcid.org/0000-0001-8552-6615)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jmedchem.3c00485>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was co-funded by the European Union (ERC, NeuRoPROBE, 101040355). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This research was supported by the European Union in the Erasmus+ Traineeship Mobility of M.B. (project no. 2020-1-IT02-KA103-078486).

## ABBREVIATIONS

CLM, chemical language model; ITC, isothermal titration calorimetry; Nurr1, nuclear receptor related 1; TH, tyrosine hydroxylase; VMAT2, vesicular amino acid transporter 2

## REFERENCES

- (1) Skinnider, M. A.; Stacey, R. G.; Wishart, D. S.; Foster, L. J. Deep Generative Models Enable Navigation in Sparsely Populated Chemical Space. *ChemRxiv* **2021**, 13638347.
- (2) Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The Rise of Deep Learning in Drug Discovery. *Drug Discovery Today* **2018**, *23*, 1241–1250.
- (3) Grisoni, F. Chemical Language Models for de Novo Drug Design: Challenges and Opportunities. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102527.
- (4) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (5) Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. Application of Generative Autoencoder in de Novo Molecular Design. **2017**, arXiv:1711.07839.
- (6) Merk, D.; Friedrich, L.; Grisoni, F.; Schneider, G. De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Mol. Inf.* **2018**, *37*, 1700153.
- (7) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4*, 120–131.
- (8) Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug Discovery with Explainable Artificial Intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573–584.
- (9) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780.
- (10) Moret, M.; Helmstädter, M.; Grisoni, F.; Schneider, G.; Merk, D. Beam Search for Automated Design and Scoring of Novel ROR Ligands with Machine Intelligence. *Angew. Chem., Int. Ed.* **2021**, *60*, 19477–19482.
- (11) Awale, M.; Sirokin, F.; Stiefl, N.; Reymond, J. L. Drug Analogs from Fragment-Based Long Short-Term Memory Generative Neural Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1347–1356.
- (12) Moret, M.; Friedrich, L.; Grisoni, F.; Merk, D.; Schneider, G. Generative Molecular Design in Low Data Regimes. *Nat. Mach. Intell.* **2020**, *2*, 171–180.
- (13) Moret, M.; Pachon Angona, I.; Cotos, L.; Yan, S.; Atz, K.; Brunner, C.; Baumgartner, M.; Grisoni, F.; Schneider, G. Leveraging Molecular Structure and Bioactivity with Chemical Language Models for de Novo Drug Design. *Nat. Commun.* **2023**, *14*, 114.
- (14) Willems, S.; Merk, D. Medicinal Chemistry and Chemical Biology of Nurr1 Modulators: An Emerging Strategy in Neurodegeneration. *J. Med. Chem.* **2022**, *65*, 9548–9563.
- (15) Willems, S.; Marschner, J.; Kilu, W.; Faudone, G.; Busch, R.; Duensing-Kropp, S.; Heering, J.; Merk, D. Nurr1 Modulation Mediates Neuroprotective Effects of Statins. *Adv. Sci.* **2022**, *9*, 2104640.
- (16) Vietor, J.; Gege, C.; Stiller, T.; Busch, R.; Schallmayer, E.; Kohlhof, H.; Höfner, G.; Pabel, J.; Marschner, J. A.; Merk, D. Development of a Potent Nurr1 Agonist Tool for in Vivo Applications. *J. Med. Chem.* **2023**, *66*, 6391–6402 DOI: [10.1021/acs.jmedchem.3c00415](https://doi.org/10.1021/acs.jmedchem.3c00415)
- (17) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; et al. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930–D940.
- (18) Bjerrum, E. J.; Sattarov, B. Improving Chemical Autoencoder Latent Space and Molecular de Novo Generation Diversity with Heteroencoders. *Biomolecules* **2018**, *8*, 131.
- (19) Graves, A. Sequence Transduction with Recurrent Neural Networks. **2012**, arXiv:1211.3711.
- (20) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
- (21) Grisoni, F.; Merk, D.; Consonni, V.; Hiss, J. A.; Tagliabue, S. G.; Todeschini, R.; Schneider, G. Scaffold Hopping from Natural Products to Synthetic Mimetics by Holistic Molecular Similarity. *Commun. Chem.* **2018**, *1*, 44.



(22) 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.* **2018**, *1*, 68.

(23) Willems, S.; Kilu, W.; Ni, X.; Chaikuad, A.; Knapp, S.; Heering, J.; Merk, D. The Orphan Nuclear Receptor Nurr1 Is Responsive to Non-Steroidal Anti-Inflammatory Drugs. *Commun. Chem.* **2020**, *3*, 85.

(24) Walters, W. P.; Barzilay, R. Applications of Deep Learning in Molecule Generation and Molecular Property Prediction. *Acc. Chem. Res.* **2021**, *54*, 263–270.

(25) Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. DOGS: Reaction-Driven de Novo Design of Bioactive Compounds. *PLoS Comput. Biol.* **2012**, *8*, No. e1002380.

(26) Brandhofer, T.; Stinglhamer, M.; Derdau, V.; Méndez, M.; Pöverlein, C.; García Mancheño, O. Easy Access to Drug Building-Blocks through Benzylic C–H Functionalization of Phenolic Ethers by Photoredox Catalysis. *Chem. Commun.* **2021**, *57*, 6756–6759.

(27) Ballarotto, M.; Solinas, M.; Temperini, A. A Straightforward Synthesis of Functionalized 6*H*-Benzo[*c*]Chromenes from 3-Alkenyl Chromenes by Intermolecular Diels–Alder/Aromatization Sequence. *Org. Biomol. Chem.* **2021**, *19*, 10359–10375.

(28) Chang, Q.; Ma, T.; Ding, W.; Zhang, L.; Cheng, X. Thiophene-Benzothiadiazole Based Donor–Acceptor–Donor (D–A–D) Bolaamphiphiles, Self-Assembly and Photophysical Properties. *Supramol. Chem.* **2021**, *33*, 174–182.

(29) Seuter, S.; Väisänen, S.; Rådmark, O.; Carlberg, C.; Steinhilber, D. Functional Characterization of Vitamin D Responding Regions in the Human 5-Lipoxygenase Gene. *Biochim. Biophys. Acta* **2007**, *1771*, 864–872.

## Recommended by ACS

### FFLOM: A Flow-Based Autoregressive Model for Fragment-to-Lead Optimization

Jieyu Jin, Yu Kang, *et al.*

JULY 20, 2023  
JOURNAL OF MEDICINAL CHEMISTRY

READ 

### A Simple Way to Incorporate Target Structural Information in Molecular Generative Models

Wenyi Zhang, Jing Huang, *et al.*

JUNE 15, 2023  
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

### Fragtory: Pharmacophore-Focused Design, Synthesis, and Evaluation of an sp<sup>3</sup>-Enriched Fragment Library

Mike Bührmann, Daniel Rauh, *et al.*

MAY 02, 2023  
JOURNAL OF MEDICINAL CHEMISTRY

READ 

### PLANET: A Multi-objective Graph Neural Network Model for Protein–Ligand Binding Affinity Prediction

Xiangying Zhang, Renxiao Wang, *et al.*

JUNE 15, 2023  
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Get More Suggestions >