Article

Prospective de novo drug design with deep interactome learning

Kenneth Atz [1], Leandro Cotos[1], Clemens Isert [1], Maria Håkansson[2], Dorota Focht[2], Mattis Hilleke [1], David F. Nippa [3,4], Michael Iff[1], Jann Ledergerber [1], Carl C. G. Schiebroek [1], Valentina Romeo[3], Jan A. Hiss [1], Daniel Merk [4], Petra Schneider [1], Bernd Kuhn[3], Uwe Grether [3] & Gisbert Schneider [1] ✉

De novo drug design aims to generate molecules from scratch that possess specific chemical and pharmacological properties. We present a computational approach utilizing interactome-based deep learning for ligand- and structure-based generation of drug-like molecules. This method capitalizes on the unique strengths of both graph neural networks and chemical language models, offering an alternative to the need for application-specific reinforcement, transfer, or few-shot learning. It enables the "zero-shot" construction of compound libraries tailored to possess specific bioactivity, synthesizability, and structural novelty. In order to proactively evaluate the deep interactome learning framework for protein structure-based drug design, potential new ligands targeting the binding site of the human peroxisome proliferator-activated receptor (PPAR) subtype gamma are generated. The top-ranking designs are chemically synthesized and computationally, biophysically, and biochemically characterized. Potent PPAR partial agonists are identified, demonstrating favorable activity and the desired selectivity profiles for both nuclear receptors and off-target interactions. Crystal structure determination of the ligand-receptor complex confirms the anticipated binding mode. This successful outcome positively advocates interactome-based de novo design for application in bioorganic and medicinal chemistry, enabling the creation of innovative bioactive molecules.

Computational de novo design encompasses the autonomous generation of new molecules with desired properties from scratch[1,2]. Chemical language models (CLMs) are machine learning techniques designed to process and learn from molecular structures represented as sequences (e.g., simplified molecular input line entry system (SMILES)-strings[3]). CLMs have found numerous applications for the de novo design of novel bioactive molecules[4,5]. Transfer learning, also known as fine-tuning, is one of the most prevalent applications of CLMs in the field of molecular design[6–8]. Transfer learning in the context of CLMs can be conceptualized as a two-step process. In the first step, the CLM undergoes pre-training using a vast data set of bioactive molecules that is not specifically tailored for the task at hand. This initial phase focuses on developing a foundational understanding of chemistry and acquiring knowledge about the characteristics of drug-

[1]ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. [2]SARomics Biostructures AB, Medicon Village, SE-223 81 Lund, Sweden. [3]Roche Pharma Research and Early Development (pRED), Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, CH-4070 Basel, Switzerland. [4]Department of Pharmacy, Ludwig-Maximilians-Universität München, Butenandtstrasse 5, 81377 Munich, Germany. ✉e-mail: gisbert@ethz.ch

like chemical space[9]. In the second step, the pre-trained CLM is fine-tuned using a smaller data set comprising molecules that specifically represent the desired activity and property profile[10]. This process refines the CLM's ability to generate molecules with the desired characteristics. Once trained, the CLM can generate virtual molecular libraries tailored to the specific task at hand[11]. Some CLM approaches integrate reinforcement learning techniques, enabling an additional level of fine-tuning to optimize the properties of the generated molecules[12,13].

However, the utilization of transfer learning and reinforcement learning in CLMs entails additional machine learning steps, which can pose challenges in terms of speed and seamless integration within the design-make-test-analysis cycle in medicinal chemistry[14–16]. Furthermore, transfer learning can be particularly challenging when applied to a single fine-tuning molecule[7,17]. It may also present difficulties in structure-based design applications that rely on explicit information about the protein binding site[18–21]. Although various structure-based de novo design methods have been introduced, their prospective applications have not been extensively explored, highlighting the need to fully assess the potential of these methods in practical scenarios[22,23].

Recent advancements have focused on studying molecular interaction networks, known as interactomes, which encompass various types of interactions such as protein-protein interactions, drug-target interactions, and drug-drug relationships. Analyzing these interactomes enables the prediction of previously unknown interactions and provides insights into the network topology[24–27]. Studying molecular interaction networks as a holistic entity offers a distinct advantage by allowing the analysis of long-range relationships between different nodes that are connected through multiple edges. This approach enables a comprehensive examination of the interconnectedness and dependencies among various components within the network[24].

To address the goal of studying the drug-target interactome comprehensively, we propose an approach that combines a CLM with interactome-based deep learning (Fig. 1a, b). This approach incorporates a neural network architecture consisting of a graph transformer neural network (GTNN) and a CLM utilizing a long-short-term-memory (LSTM) (Fig. 1c, d, e). Herein, the deep learning model resulting from this approach is named DRAGONFLY (Drug-target interActome-based GeneratiON oF noveL biologicallY active molecules). Unlike conventional CLMs that rely on transfer learning with individual molecules, the method leverages interactome-based deep learning, which enables the incorporation of information from both, targets and ligands across multiple nodes. DRAGONFLY is capable of processing small-molecule ligand templates as well as three-dimensional (3D) protein binding site information. It operates on diverse chemical alphabets and does not require fine-tuning through transfer or reinforcement learning specific to a particular application. Furthermore, it enables the incorporation of desired physical and chemical properties into the generation of output molecules. This study introduces the prospective application of DRAGONFLY to structure-based de novo design, specifically for the generation of ligands with desired bioactivity profiles addressing one or multiple specific macromolecular targets (Fig. 1f).

Results
DRAGONFLY enables ligand- and structure-based molecular design
The central component of DRAGONFLY is its drug-target interactome, which captures the connections between small-molecule ligands and their macromolecular targets. This interaction can be depicted as a graph, where nodes represent bioactive ligands and their corresponding macromolecular targets (Fig. 1a). Distinct nodes were used to differentiate between orthosteric and allosteric binding sites within the same target. Edges were established between ligands and proteins that have an annotated binding affinity of less than or equal to 200 nM

(Fig. 1a) (values extracted from the ChEMBL database[28]). As a result of this procedure, an interactome was generated that consisted of ~360,000 ligands, 2989 targets, and around 500,000 bioactivities. This interactome was specifically designed for ligand-based design applications. In the case of structure-based design, only macromolecular targets with known 3D structures were considered, resulting in an interactome containing around 208,000 ligands, 726 targets, and around 263,000 bioactivities. This data structure based on the interactome facilitated the training of two deep learning models, specifically for ligand-based and structure-based de novo design (Fig. 1b).

The neural networks employed in the study accept a molecular graph as their input signal. In particular, a 3D graph was utilized for binding sites, while a 2D molecular graph was used for ligands (Fig. 1c and d). Subsequently, the input graph undergoes a transformation into SMILES-strings, which represent molecules with the desired bioactivity and physicochemical properties. This translation process was achieved by utilizing a graph-to-sequence deep learning model that combines a graph transformer neural network[29–31] with a long-short term memory (LSTM) neural network[32] (Fig. 1e). The selection of the graph-to-sequence architecture was made to facilitate the development of deep learning models capable of supporting both ligand-based and structure-based molecular design.

DRAGONFLY considers synthesizability, novelty, bioactivity, and physicochemical properties for ligand design
The theoretical evaluation of DRAGONFLY focused on investigating the incorporation of specific physical and chemical properties into the DRAGONFLY model, as depicted in Fig. 2a. This evaluation revealed Pearson correlation coefficients (r) greater than or equal to 0.95 for all assessed physical and chemical properties. These properties included molecular weight ($r = 0.99$), rotatable bonds ($r = 0.98$), hydrogen bond acceptors ($r = 0.97$), hydrogen bond donors ($r = 0.96$), polar surface area ($r = 0.96$), and lipophilicity expressed as MolLog$P$[33] ($r = 0.97$). These high correlation coefficients indicate a strong relationship between the desired properties and the actual properties of the generated molecules.

The study also included the evaluation of novelty, synthesizability, and predicted bioactivity for the generated molecules. These criteria were essential in assessing the practicality and potential value of the designed compounds. To quantify molecular novelty, a rule-based algorithm was utilized, which captured both, scaffold and structural novelty. This algorithm, described in Equations (9)–(12), offers a quantitative measure of the uniqueness of each molecule in terms of its chemical structure. Synthesizability was assessed using the retrosynthetic accessibility score (RAScore), a recently published metric that assesses the feasibility of synthesizing a given molecule[34].

Additionally, to estimate the on-target bioactivity of the de novo designs, quantitative structure-activity relationship (QSAR) models were developed. The models utilized kernel ridge regression (KRR)-based machine learning[35], and were trained on three molecular descriptors: ECFP4[36], unscaled CATS[37], and USRCAT[38]. The descriptors used in the study encompassed a wide range of structural, pharmacophore, and shape-based similarities of the molecules, offering a comprehensive representation of their characteristics. A combination of descriptors, including ECFP for structural features as well as CATS and USRCAT for "fuzzy" features, was employed to capture both, specific and general molecular attributes. By incorporating these descriptors, the study aimed to facilitate the identification of molecular similarities between the highly ranked de novo designs and known bioactive compounds. For the majority of the 1265 investigated targets, the mean absolute errors (MAEs) for the predicted pIC$_{50}$ values were equal to or less than 0.6 (Fig. 2b, Fig. S2). Moreover, the KRR models have shown superior performance to decision tree baseline methods including gradient boosting and extreme gradient boosting
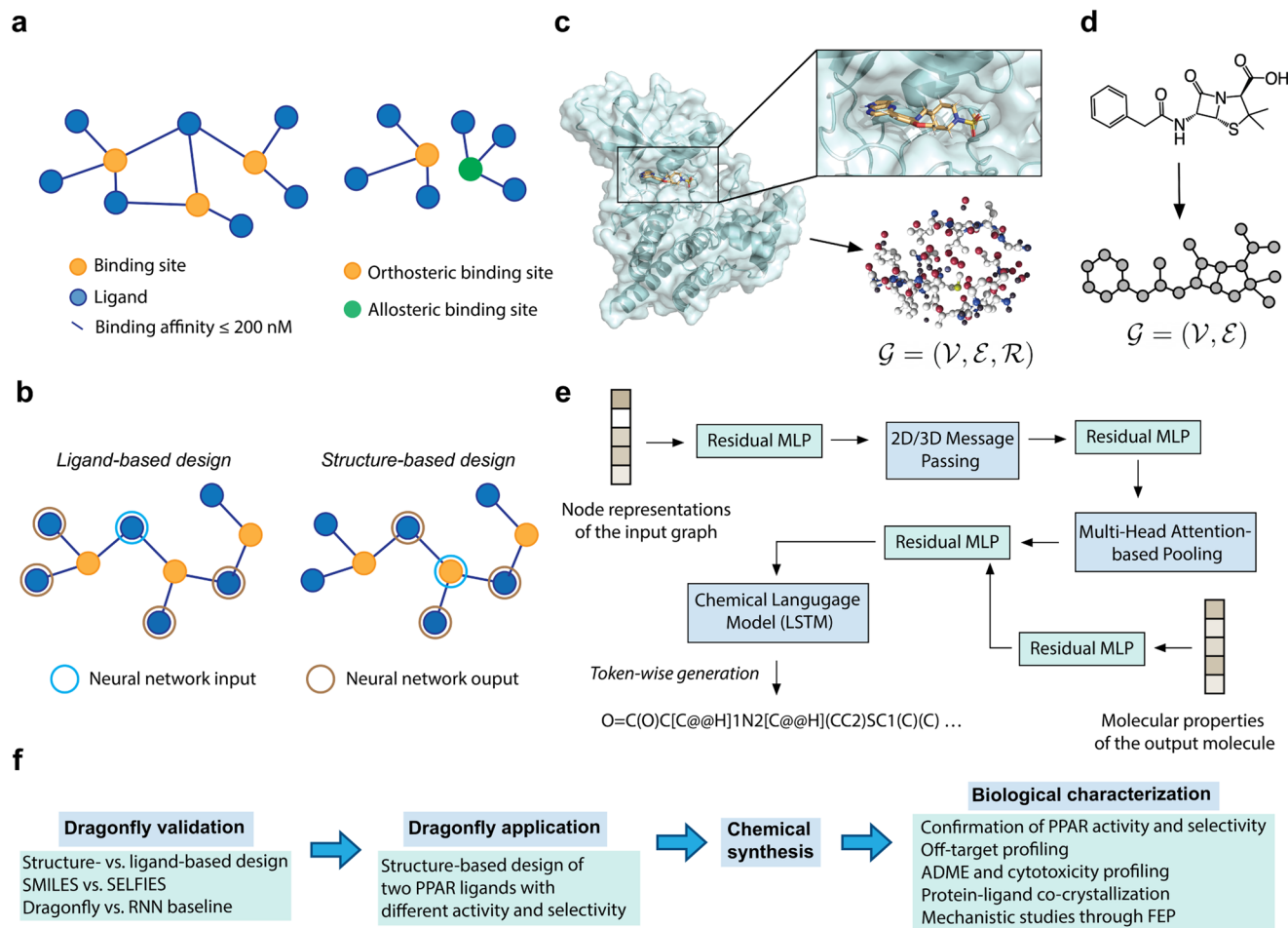
Fig. 1 | DRAGONFLY architecture and workflow. a Left: To construct the drug-target interactome graph, the targets are connected to their corresponding ligands based on reported bioactivities in the ChEMBL database[28]. Specifically, a connection is established between a ligand (blue circle) and its corresponding target (orange circle) if the ligand has been reported with a bioactivity equal to or <200 nM. Right: By representing allosteric and orthosteric binding sites as separate nodes (shown in green and orange, respectively), the drug-target interactome graph captures the specific interactions and relationships associated with each type of binding site. **b** Left: During the training phase for ligand-based design, a ligand molecule (represented as a blue circle) is taken as the input to the model. The desired output molecules (represented as brown circles) are selected based on their connection to the input molecule through a common node, indicating that they share a binding site. Right: For structure-based design, the input for the model is the binding site itself, represented as a blue circle. The desired output molecules, represented as brown circles, are ligands that have been observed to bind to the corresponding binding site. **c** The protein binding site (here: Janus Kinase 2, PDB-ID 6VNK[113]) is represented as a three-dimensional (3D) graph, i.e., $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{R})$ where $\mathcal{G}$ denotes the graph, $\mathcal{V}$ vertices, $\mathcal{E}$ edges and $\mathcal{R}$ the position in 3D space. All protein atoms farther away than 5 Å from any atom of the bound ligand were removed,

yielding a pocket-centric representation of the binding pocket. **d** The ligands are represented as two-dimensional (2D) graphs, i.e., $\mathcal{G}=(\mathcal{V},\mathcal{E})$. **e** In the proposed approach, the node features within the graph are updated through a message passing process. This can be done using either 2D or 3D message passing, depending on the nature of the molecular representation. As a result of the subsequent pooling process, a latent space vector is obtained, which captures the essential characteristics and representations of the molecule. This condensed representation provides a compact encoding of the molecule's features, enabling downstream analysis, prediction, or structure generation tasks. The latent space vector can be optionally concatenated with a wishlist of desired physicochemical properties for the output molecule. This allows for the incorporation of project-specific property constraints or objectives in the de novo molecular design process. MLP denotes Multilayer Perceptron, RNN denotes Recurrent Neural Network, and LSTM refers to a type of RNN with Long Short-Term Memory cell architecture. **f** Workflow of the presented study including DRAGONFLY validation, DRAGONFLY application to peroxisome proliferator-activated receptor (PPAR), chemical synthesis and biological characterization. ADME denotes Absorption, Distribution, Metabolism, and Excretion, and FEP denotes Free Energy Perturbation calculations.

(XGBoost) (SI2.2). These results indicate that the developed models achieved a high level of accuracy in predicting the inhibition constant of novel molecules within similar domains of applicability for the targets studied. Furthermore, the performance of the ECFP and CATS models exhibited a logarithmically decreasing error as the training set size increased. Beyond a certain data set size (-100 molecules), the performance of the USRCAT models reached a plateau, indicating that additional data did not improve their predictive accuracy (Fig. 2b, Fig. S3). These findings underscore the effectiveness of utilizing a combination of descriptors, incorporating both, structural and fuzzy features, in the performance of the KRR models.

DRAGONFLY outperforms standard chemical language models for molecular design

The evaluation criteria, which encompassed synthesizability, novelty, and predicted bioactivity were applied to evaluate virtual libraries generated de novo (Methods for details on metrices). This allowed for a comparison between DRAGONFLY and fine-tuned recurrent neural networks (RNNs). To conduct the comparison, five known ligands each were selected as templates for twenty well-studied macromolecular targets, including nuclear hormone receptors and kinases with over 200 known ligands (Tables S2–S3). DRAGONFLY demonstrated superior performance over the fine-tuned RNNs across the majority of
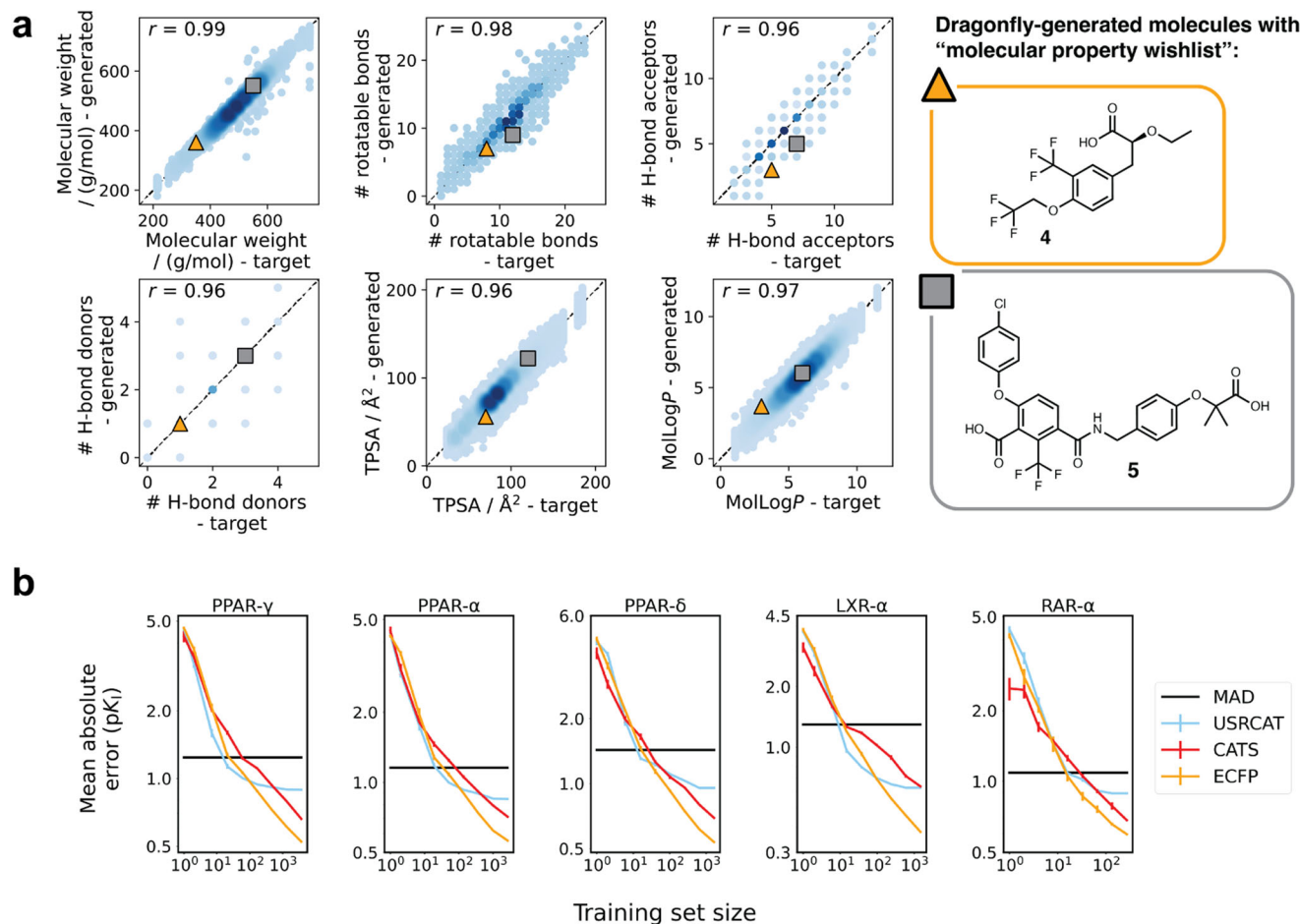
Fig. 2 | Property translation with DRAGONFLY and quantitative structure-activity relationship (QSAR) models. a The scatter plots depict the translation of desired properties into the generated molecules, with squared Pearson correlation coefficients (r^2) >0.95 for all investigated molecular properties. The high correlation suggests a strong relationship between the desired properties and the generated molecules. Properties include molecular weight, rotatable bonds, hydrogen bond acceptors, hydrogen bond donors, polar surface area, and lipophilicity calculated as MolLog$P$ values[33]. Two exemplary molecules **4** and **5** and their position in the six scatter plots are visualized, showing that the desired molecular properties are accurately represented in the generated molecules. **b** The double-logarithmic learning curves depict how the prediction error of ligand-based QSAR models for three descriptors (extended connectivity fingerprint (ECFP4)[36], chemical advanced template search (CATS, absolute values)[37], ultra-fast shape Recognition with atom types (USRCAT)[38]) varies with the size of the data set, focusing on five selected drug targets related to nuclear hormone receptors. Each plot displays a horizontal line representing the mean absolute deviation (MAD) of the training data. The error bars on the plots represent the standard deviation observed during a 10-fold cross-validation. Source data are provided as a Source Data file.

templates and properties examined (Table 1, Tables S4–S6). Furthermore, using the same evaluation criteria, ligand-based design was compared to structure-based design, with ligand-based design applications outperforming structure-based models in all investigated scenarios (Table 2, Tables S4–S6).

To evaluate the potential of DRAGONFLY to generate molecules that extend into new areas of chemical space, we analyzed the similarity of the generated molecules to both the training data set (i.e., a subset of ChEMBL[28]) and an external data set (i.e., PubChem[39], excluding ChEMBL molecules). Our analysis demonstrated that the generated molecules were as similar to the molecules in PubChem when compared to the molecules in ChEMBL (Table S7). While there is a degree of similarity to known molecules, DRAGONFLY also produced a large proportion of molecules with high novelty scores and diverse structures generating higher structural and scaffold novelty than the well-established fine-tuned RNN methods (Table 1). These results suggest that DRAGONFLY is not limited to recapitulating the training data but also has the capacity to explore and generate molecules in previously uncharted regions of chemical space, albeit the extent of this exploration warrants further investigation.

We compared the performance of DRAGONFLY models trained on two widely used chemical alphabets, SMILES-strings[3] and self-referencing embedded strings (SELFIES)[40], to quantify the differences. By employing both string representations in structure- and ligand-based de novo design, we were able to directly compare their performance across various molecular properties (Table 2, Fig. S9). The DRAGONFLY models trained on SELFIES yielded a higher fraction of novel molecules among all of the 20 investigated applications (99.7 ± 0.1% vs. 92.2 ± 0.4%, Table 2) with a greater scaffold diversity (86 ± 1% vs. 53 ± 2%, Table 2) while retaining comparable structural diversity (98.8 ± 0.1% vs. 97.9 ± 0.1%, Table 2). However, the DRAGONFLY models trained on SMILES-strings more accurately fulfilled the property requirements, such as greater synthesizability (93.4 ± 0.6% vs. 84 ± 1%, Table 1), predicted bioactivity (e.g., MAE = 34.7 (± 0.3) vs 31.9 (± 0.1) for PPARγ, Table 1), as well as slightly lower mean absolute errors for physical and chemical properties (e.g., MAE = 0.027 ± 0.005 vs 0.230 ± 0.007 for hydrogen bond donors, Table 3). Overall, the use of the two chemical alphabets resulted in comparable numbers of molecules that were predicted to fulfill all desired properties. Because of the better performance of the SMILES-based models for the objectives of

Table 1 | Comparison of DRAGONFLY with a fine-tuned recurrent neural network (RNN) approach, assessing the percentage of molecules meeting various criteria: (i) Unique and novel, (ii) Novelty score ≥ 0.65, (iii) Retrosynthetic accessibility score (RAScore) ≥ 0.5, (iv) QSAR score ≤ 1 μM, and (v) meeting all four criteria

Template / Method	Unique and novel / %	Novelty score ≥ 0.65 / %	RAScore ≥ 0.5 / %	QSAR score ≤ 1 μM / %	All criteria / %
PPARγ					
RNN-SMILES	75.4 (± 2.7)	28.7 (± 1.1)	67.9 (± 2.3)	29.6 (± 2.4)	5.1 (± 0.2)
DRAGONFLY-SMILES	91.8 (± 0.3)	47.9 (± 1.4)	**86.0 (± 0.3)**	**34.7 (± 0.3)**	9.4 (± 0.0)
DRAGONFLY-SELFIES	**99.8 (± 0.1)**	**77.4 (± 0.1)**	82.2 (± 0.2)	31.9 (± 0.1)	**13.3 (± 0.0)**
LXRβ					
RNN-SMILES	92.4 (± 2.5)	65.9 (± 2.6)	87.9 (± 2.8)	28.6 (± 0.9)	11.3 (± 0.4)
DRAGONFLY-SMILES	94.3 (± 0.5)	80.2 (± 1.2)	**89.1 (± 0.5)**	26.2 (± 0.2)	**11.8 (± 0.1)**
DRAGONFLY-SELFIES	**100 (± 0.0)**	**91.3 (± 0.5)**	84.2 (± 0.3)	**27.9 (± 0.2)**	11.1 (± 0.1)
RARα					
RNN-SMILES	69.7 (± 5.9)	41.9 (± 3.3)	57.2 (± 4.3)	30.1 (± 1.8)	11.1 (± 0.7)
DRAGONFLY-SMILES	92.2 (± 0.4)	62.4 (± 0.7)	75.6 (± 0.5)	**32.4 (± 0.7)**	12.7 (± 0.2)
DRAGONFLY-SELFIES	**99.8 (± 0.0)**	**87.5 (± 0.3)**	**77.1 (± 0.2)**	29.6 (± 0.3)	**14.0 (± 0.1)**
BRAF					
RNN-SMILES	89.2 (± 3.5)	35.1 (± 3.1)	85.9 (± 3.0)	35.0 (± 1.3)	6.7 (± 0.3)
DRAGONFLY-SMILES	87.9 (± 0.6)	46.0 (± 0.8)	**80.9 (± 0.5)**	**42.9 (± 0.5)**	10.7 (± 0.1)
DRAGONFLY-SELFIES	**99.7 (± 0.1)**	**81.1 (± 0.6)**	77.3 (± 0.4)	34.3 (± 0.1)	**12.4 (± 0.0)**
BTK					
RNN-SMILES	82.0 (± 4.4)	64.5 (± 4.1)	61.9 (± 4.7)	20.7 (± 1.8)	4.5 (± 0.2)
DRAGONFLY-SMILES	88.9 (± 0.7)	53.2 (± 0.4)	69.6 (± 0.9)	**36.3 (± 0.7)**	**8.8 (± 0.1)**
DRAGONFLY-SELFIES	**100 (± 0.0)**	**85.8 (± 0.7)**	**68.2 (± 1.0)**	25.8 (± 0.1)	5.8 (± 0.0)
JAK2					
RNN-SMILES	88.8 (± 3.9)	60.2 (± 4.2)	79.9 (± 3.4)	35.0 (± 2.2)	14.5 (± 0.8)
DRAGONFLY-SMILES	84.8 (± 1.0)	39.4 (± 0.9)	69.0 (± 1.0)	**55.9 (± 1.5)**	14.8 (± 0.2)
DRAGONFLY-SELFIES	**99.2 (± 0.0)**	**73.3 (± 0.8)**	**70.5 (± 0.5)**	50.5 (± 1.0)	**18.3 (± 0.2)**

Bold indicates whether the SELFIES- or SMILES-based models achieve a higher value for the investigated property in both structure- and ligand-based models. The values are presented as mean and standard deviation, based on three runs (N = 3), each sampling 2000 SMILES-strings. The complete list of 20 investigated targets can be found in Tables S2–S6. *JAK* Janus kinase, *PPAR* Peroxisome proliferator-activated receptor, *BRAF* Serine/threonine-protein kinase B-Raf (rapidly accelerated fibrosarcoma), *BTK* Bruton's tyrosine kinase, *RAR* Retinoic acid receptor, *LXR* Liver X receptor.

Table 2 | Comparison of four Dragonfly methods, namely ligand-SMILES, ligand-SELFIES, structure-SMILES, and Structure-SELFIES

DRAGONFLY method	Valid and unique molecules / %	Valid, unique and novel molecules / %	RAScore ≥ 0.5 / %	Average Jaccard distance to other molecules
Ligand-SMILES	93.3 (± 0.4)	92.2 (± 0.4)	**93.4 (± 0.6)**	0.778 (± 0.001)
Ligand-SELFIES	**99.9 (± 0.1)**	**99.7 (± 0.1)**	84.0 (± 1.0)	**0.805 (± 0.002)**
Structure-SMILES	90.2 (± 0.8)	87.4 (± 0.9)	**90.0 (± 1.0)**	0.773 (± 0.004)
Structure-SELFIES	**99.9 (± 0.1)**	**99.6 (± 0.1)**	78.0 (± 2.0)	**0.811 (± 0.003)**
	Unique atom scaffolds / %	**Unique and Novel atom scaffolds / %**	**Unique carbon scaffolds / %**	**Unique and novel carbon scaffolds / %**
Ligand-SMILES	85.0 (± 0.1)	53.0 (± 0.2)	98.4 (± 0.3)	58.0 (± 0.2)
Ligand-SELFIES	**96.9 (± 0.4)**	**86.0 (± 0.1)**	**99.8 (± 0.1)**	**83.0 (± 0.1)**
Structure-SMILES	84.0 (± 0.1)	55.0 (± 0.3)	98.3 (± 0.3)	56.0 (± 0.2)
Structure-SELFIES	**96.0 (± 0.1)**	**81.0 (± 0.1)**	**99.9 (± 0.1)**	**83.0 (± 0.2)**

Bold indicates whether SELFIES- or SMILES-based models achieve a higher value for the investigated property in both structure- and ligand-based models. The percentage of molecules is shown that fulfill the desired criteria: (i) valid and unique molecules, (ii) valid, unique, and novel molecules, (iii) fraction of molecules with an RAScore of ≥ 0.5, (iv) average Jaccard distance to other generated molecules from the same run (indicating diversity), and (v)–(viii) various scaffold metrics, including unique and novel carbon and atom scaffolds. The values are presented as mean and standard deviation, based on three Dragonfly runs (N = 3), each sampling 2000 SMILES-strings.

synthetic accessibility, bioactivity, and desired physical and chemical properties, we used these models in the prospective study.

Structure-based design with DRAGONFLY generates potential novel ligands
DRAGONFLY was utilized in a prospective manner for structure-based ligand design targeting human PPARγ (PPARγ, PDB-ID 3G9E[41]). The nuclear hormone receptor PPARγ is one of the three peroxisome proliferator-activated receptors (i.e., PPAR$\gamma/\alpha/\delta$), that have been exploited as drug targets for combating multiple diseases, in particular metabolic syndrome-related disorders and cancer[42–44]. Activation of the PPARs by natural ligands or by synthetic PPAR agonizts triggers the formation of heterodimers with members of the retinoid X receptor (RXR) family[45]. Upon recruitment of specific cofactors, these

Table 3 | Accuracy of the desired physical and chemical properties of molecules generated by DRAGONFLY

DRAGONFLY method	MW	Rot. B.	HBA	HBD	PSA	LogP
Unit	g mol⁻¹	#	#	#	Å²	–
MAD	75.52	2.81	0.981	1.69	27.08	1.25
Ligand-SMILES						
MAE	7.7 (± 0.2)	0.29 (± 0.01)	0.23 (± 0.01)	0.027 (± 0.005)	4.4 (± 0.2)	0.252 (± 0.004)
MAD / MAE	**9.8 (± 0.2)**	**9.9 (± 0.48)**	**4.3 (± 0.21)**	**63 (± 11)**	**6.1 (± 0.2)**	**4.94 (± 0.08)**
Ligand-SELFIES						
MAE	8.0 (± 0.2)	0.88 (± 0.040)	0.40 (± 0.014)	0.230 (± 0.007)	6.8 (± 0.21)	0.380 (± 0.006)
MAD / MAE	9.4 (± 0.2)	3.2 (± 0.14)	2.5 (± 0.09)	7.3 (± 0.3)	4.0 (± 0.1)	3.27 (± 0.05)
Structure-SMILES						
MAE	12.1 (± 0.5)	0.42 (± 0.02)	0.28 (± 0.02)	0.046 (± 0.007)	4.6 (± 0.1)	0.315 (± 0.008)
MAD / MAE	**6.2 (± 0.3)**	**6.7 (± 0.3)**	**3.5 (± 0.2)**	**37 (± 6)**	**5.9 (± 0.2)**	**4.0 (± 0.1)**
Structure-SELFIES						
MAE	15 (± 0.4)	1.12 (± 0.04)	0.50 (± 0.03)	0.27 (± 0.02)	7.4 (± 0.3)	0.426 (± 0.008)
MAD / MAE	5.03 (± 0.1)	2.5 (± 0.09)	2.0 (± 0.1)	6.3 (± 0.4)	3.6 (± 0.1)	2.92 (± 0.05)

Bold indicates if the SELFIES- or the SMILES-based models achieved a higher value for the investigated property. Abbreviations: *MAD* Mean absolute deviation, *MAE* Mean absolute error, *MW* Molecular weight, *Rot. B.* Number of rotatable bonds, *HBA* Hydrogen bond acceptors, *HBD* Hydrogen bond donors, *PSA* Polar surface area. The numbers are presented as the mean and standard deviation, with a sample size of $N = 3$, i.e., 3 DRAGONFLY runs, each sampling 2000 SMILES-strings. MAD / MAE yields a number that indicates by which factor a model is better than the MAD.

heterodimers transactivate PPAR-responsive elements (PPREs) of target genes involved in insulin signaling, lipid and glucose metabolism, immune response, as well as cell cycle and differentiation[46,47]. Several activators with different selectivity for the respective PPAR subtypes have reached advanced clinical trials or were introduced to the market. Aiming to test the method's ability to generalize, the training of DRAGONFLY did not include the protein template PPAR or any closely related structures present in the training data set. Specifically, proteins belonging to the same sub-family (PPARα, PPARγ, PPARδ) or other species were intentionally excluded. The most closely related proteins in the training data were found to be thyroid hormone receptor β-1 and liver X receptor (LXR) β, sharing a sequence identity of 33% and 30%, respectively, with PPARγ (SI8, Table S8).

To obtain a library of candidate ligands, the ligand-binding site of human PPARγ protein (PDB-ID 3G9E[41]) was utilized as the structural template (Fig. 3a). A total of 300 k molecules were generated, and subsequent filtering was performed based on specific criteria. These filters included an upper molecular weight limit of ≤600 g mol⁻¹, a RAScore threshold of ≥0.5, and a novelty score of ≥0.7. The resulting subset of molecules obtained from the filtering process was further ranked using KRR-based QSAR scoring based on the average predicted binding affinity (pK_I or pIC_{50}), combining the ECFP (double weighted contribution), CATS, and USRCAT descriptors. Aiming to explore the potential of the computer-generated molecules for dual-target activity and receptor selectivity, two different scoring procedures were prospectively evaluated. The first procedure focused on selective single-target affinity to PPARγ. The second procedure involved assigning equal weights to dual-target affinity towards both PPARγ and PPARδ. The decision to focus on affinity towards PPAR sub-family members was made to align with their clinical significance[48]. Applying these ranking criteria led to the identification of twice the top-5 molecules (**1, 2, 6-13**), depicted in Fig. 3b.

While exhibiting sufficient structural and scaffold novelty, the generated molecules also possess a topology commonly observed in modulators of the PPAR subfamily, as captured by the three different QSAR models. Specifically, they feature an acidic head group connected to an aromatic core through a linker. Furthermore, this core is linked to another single- or bicyclic aromatic ring system. While the propionic acid head group was predominant among the top-ranking designs, it is worth noting that various other head groups were also present among the 100 highest-scored de novo molecules. Figure 3c highlights a selection of non-carboxylic head groups and secondary amides from this top-100 set. This selection includes a diverse range of secondary amides as well as pyrimidine-diones, i.e., head groups known to promote PPARγ modulation[49,50]. These alternative head groups demonstrate the structural diversity and potential for exploring different chemistries and bioisosters in the design of novel molecules within the top-ranked subset.

Molecules generated with DRAGONFLY potently and selectively activate PPARγ

To test the practical applicability and usefulness of the structure-based molecular design algorithm, the two top-scoring de novo generated designs (**1** and **2**) were chosen for chemical synthesis and subsequent biological characterization. Design **1** was achieved through a convergent synthesis comprising a total of 10 steps, with the longest sequential route consisting of six steps, with an overall yield of 12% (Fig. 4a). Design **2** was synthesized through five steps achieving an overall yield of 0.6% (Fig. 4b). Additionally, regioisomer **3** was isolated during the synthesis of design **2**.

Subsequent biological testing of the three molecules (**1**–**3**) in a cell-based reporter gene assay confirmed the intended activity profiles (SI9). Compound **1** exhibited the desired and predicted dual activity on PPARγ and PPARδ at half maximal effective concentration $EC_{50}(PPARγ) = 1.5 \pm 0.2\,\mu M$, and $EC_{50}(PPARδ) = 0.24 \pm 0.05\,\mu M$, respectively (Fig. 5a, Fig. S13). Moreover, compound **1** was characterized for its affinity to the ligand binding domain of PPARγ by isothermal titration calorimetry (ITC), yielding a measured dissociation constant of $K_D = 0.8 \pm 0.1\,\mu M$, and a molar ratio of one ligand per protein molecule (Fig. 5b). This K_D value of compound **1** confirmed the observed direct receptor modulation in the functional reporter gene assay. Compound **2** demonstrated a noteworthy level of selective activity on PPARγ with an EC_{50} value of $2.3 \pm 0.7\,\mu M$, while displaying no discernible impact on PPARα or PPARδ (Fig. 5c, Fig. S13). This outcome aligns seamlessly with the intended design objective. Compound **3** exhibited a dual, partial agoniztic activity profile, acting on both PPARγ (with an EC_{50} of $1.8 \pm 0.1\,\mu M$) and PPARα (with an EC_{50} of $3.4 \pm 0.3\,\mu M$), while showing no discernible activity towards PPARδ. Furthermore, the predicted selectivity of compounds **1**–**3** towards other nuclear hormone receptor targets was experimentally confirmed for retinoid X receptor (RXR)α, liver X receptor (LXR)α, and retinoic acid receptor (RAR)α (Fig. 5d).

Computer-designed compounds **1** and **2** underwent initial in vitro testing to assess their absorption, distribution, metabolism, and
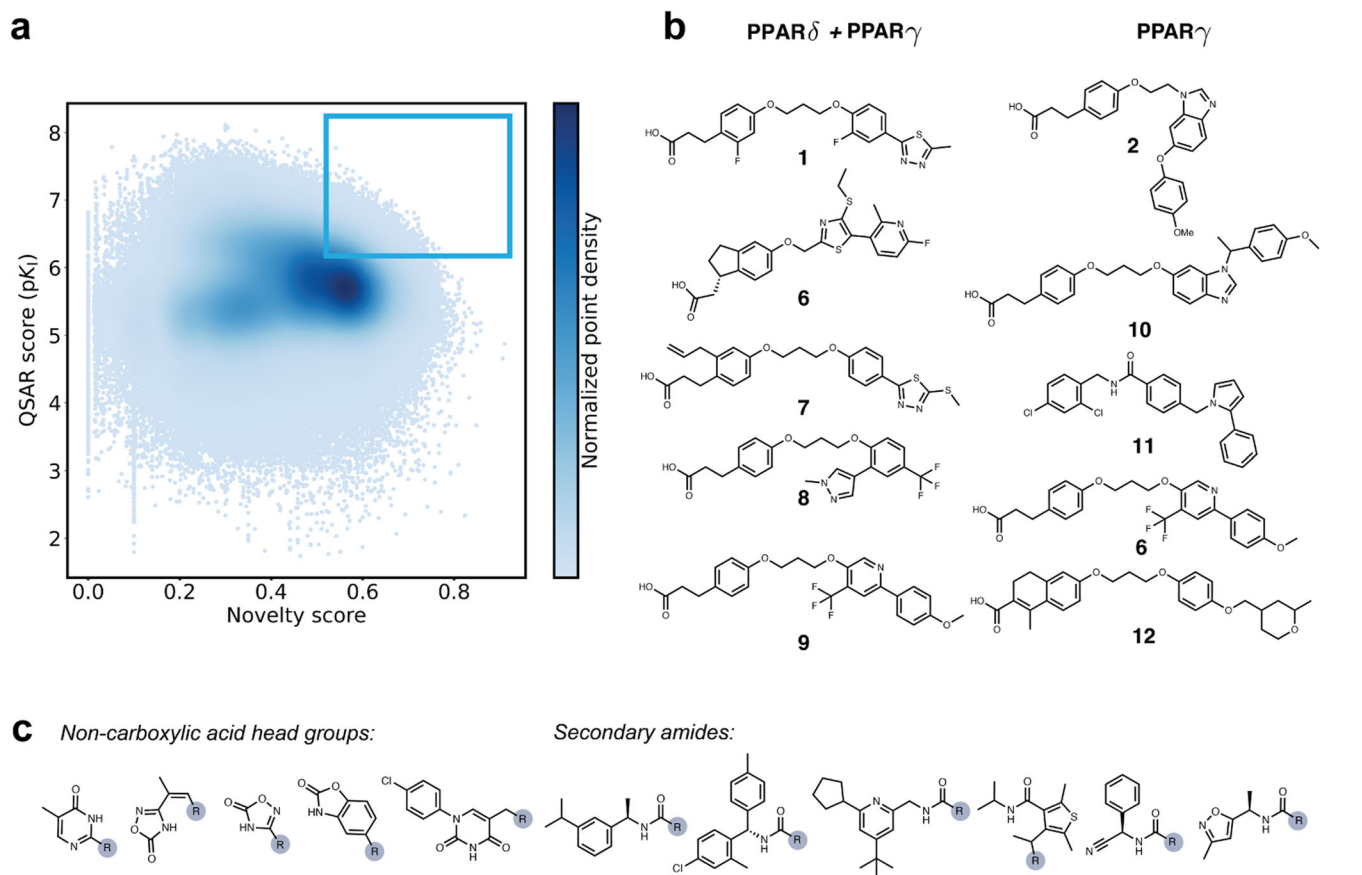
a

b

c

Fig. 3 | Results of structure-based de novo design with DRAGONFLY. a The scatter plot presented showcases the molecules designed de novo by utilizing the human peroxisome-proliferator-activated-receptor (PPAR)γ binding pocket as a template (PDB-ID 3G9E[41]). The plot displays the quantitative structure-activity relationship (QSAR) score representing the predicted binding affinity to PPARγ against the novelty score. The desired region for the generated molecules, which satisfies both, novelty and the predicted bioactivity requirements, is highlighted by a blue box located in the upper right corner of the plot. **b** Molecular structures of the five top-ranking de novo designs. The ranking criteria were PPARγ + PPARδ dual-target affinity and structural novelty (left, **1** & **6**–**9**), or PPARγ single-target affinity and structural novelty (right, **2**, **6** & **10**–**12**). **c** Examples of non-carboxylic head groups and secondary amides from the top-100 molecules ranked for PPARγ, where the gray shaded *R* group represents an aromatic moiety connected to a linker. Source data are provided as a Source Data file.

excretion (ADME) properties, and were compared to the phase three dual PPARα/γ co-agonist aleglitazar[41] (SI11, Table S14). Both ligands exhibited lipophilicity values within the range of 1–2 (logD **1**: 1.5 **2**: 1.7), which falls within the preferred range for orally administered drugs and is comparable to the logD value of aleglitazar (logD: 1.4).

In terms of permeability through membranes, both molecules displayed favorable results in the parallel artificial membrane permeability assay (PAMPA), with permeation coefficients (PAMPA$_{PEFF}$) of 3.9 $cm \cdot s^{-1} \cdot 10^{-6}$ for compound **1** and 14 $cm \cdot s^{-1} \cdot 10^{-6}$ for compound **2**. Achieving sufficient cell permeability is crucial for targeting the PPARγ receptor, located within the cell nucleus. Cellular permeability was confirmed in the P-glycoprotein (Pgp) efflux assay for compounds **1** and **2**, revealing values of 1.6 (15 $nm \cdot sec^{-1}$) and 1.2 (60 $nm \cdot sec^{-1}$), respectively. The observed efflux ratio indicates that compounds **1** and **2** are only interacting weakly with the Pgp transporter and thus, hold high potential to reach multiple cell and tissue types following a systemic application. Moreover, the unbound fractions of compounds **1** and **2** were determined at 0.42% and 0.21%. Such low unbound fractions are attributed to the negatively charged carboxylic acid, similar to other drug-like molecules containing carboxylic acid groups[41]. Furthermore, the clearance values of compounds **1** and **2** in human, rat, and mouse microsomes were consistently low (≤10 $\mu L \cdot min^{-1} \cdot mg^{-1}$ protein) when compared to aleglitazar, suggesting the potential for achieving high oral bioavailability in both humans and rodents for efficacy studies. Compound clearance rates in human hepatocytes

were determined at 19 $\mu L \cdot min^{-1} 10^6 cells^{-1}$ for both compounds **1** and **2** (Table S14). Both metabolic and hepatocyte clearance suggest a sufficient metabolic profile, paving the way for further in vivo pharmacokinetic studies. Compounds **1** and **2** exhibited no interaction with the seven pivotal cytochrome P450 isoenzymes (CYP)−Cyp3A4, Cyp1A2, Cyp2B6, Cyp2C9, Cyp2D6, Cyp2C19, and Cyp2C8 - at dose-response experiments up to 20 μM (Table S15). Moreover, both compounds presented a favorable profile in an expansive panel screen assessing multiple safety-critical off-targets. Importantly, none of the targets exhibited binding or inhibition above 60% at a compound concentration of 10 μM (Tables S13–S16). Furthermore, compounds **1** and **2** did not indicate any cytotoxicity on HEK293T cells at different time points as well as a broad range of cell numbers and compound concentrations (Fig. 5e, Fig. S16). Collectively, the computer-designed compounds **1** and **2** showcase a promising drug-like profile, signifying substantial potential for advancement in further drug development.

To investigate the binding pose of compound **1**, X-ray structure determination of the ligand-protein complex with PPARγ was conducted (Fig. 6a, SI10). Compound **1** was bound to one of two protein molecules in the asymmetric unit. Moreover, the observed binding pose showed how the relevant structural motifs of the design contribute to ligand-receptor interaction. Compound **1** bound in the orthosteric site lined by helices H3 and H11. The buried propionic acid head group is engaged in four intermolecular hydrogen bridges. Three of them are established with the side chains of Tyr[473], His[323], and Ser[289],
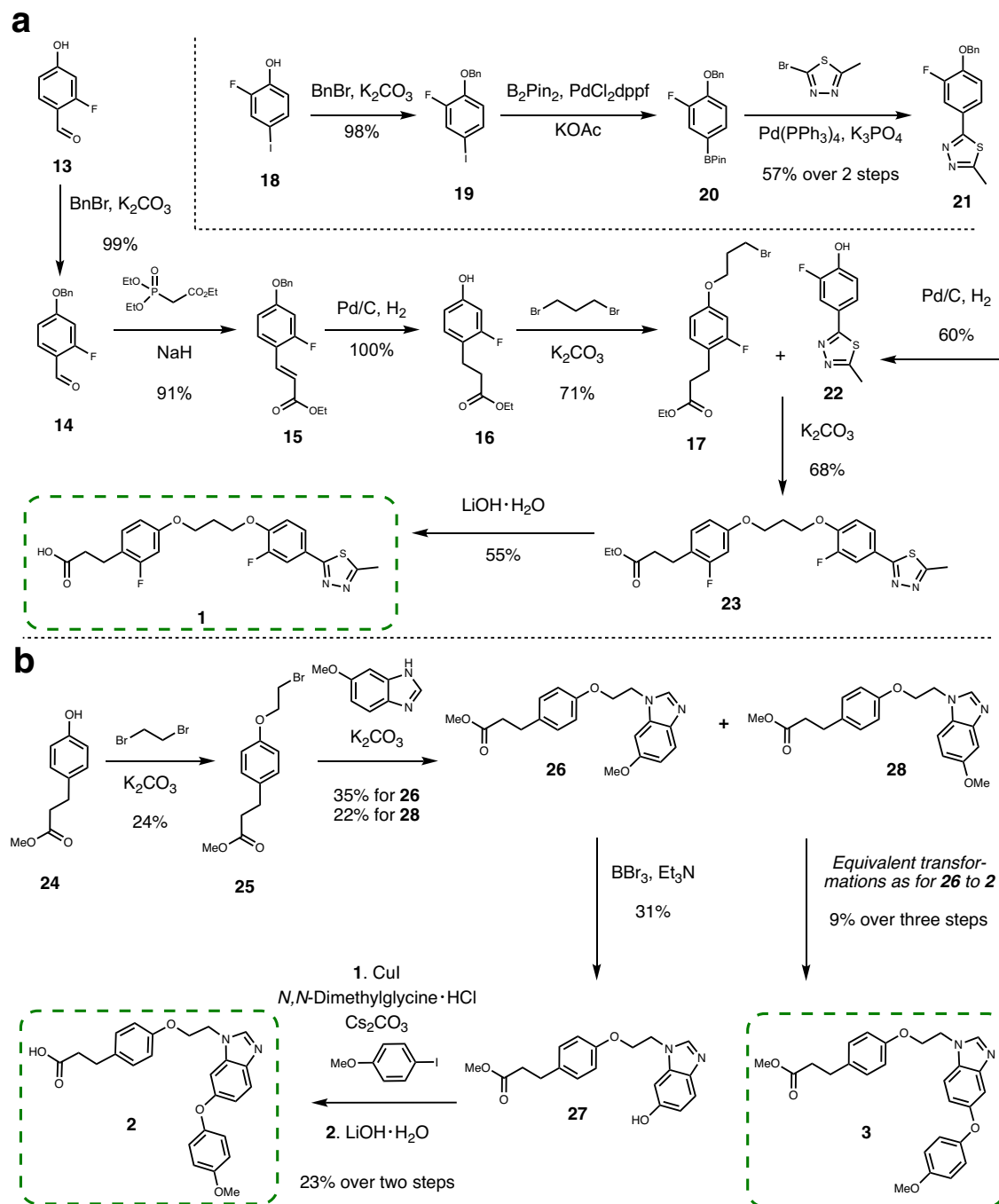
Fig. 4 | Chemical synthesis of compounds 1, 2 and 3. a The synthesis of compound **1** employed a convergent approach, starting from commercially available building blocks **13** and **18**, and spanning a total of 10 steps. The longest sequential route involved 6 steps. The overall yield achieved for the synthesis of **1** was 12%. **b** For the synthesis of compounds **2** as well as its regioisomer **3**, the starting material used was a commercial building block **24**. These compounds were synthesized through a sequential five-step synthesis. The overall yield obtained for compound **2** was 0.6%, while compound **3** was isolated with a yield of 0.5%. For details see SI13.

whereas the fourth one is a water-mediated hydrogen bond with residue His[449]. In the empty PPARγ site, the carboxyl C-terminus of TYR[477] is blocking the site by binding in a similar position as the propionic acid head of the ligand. The ligand's tail moiety is exposed to the solvent, and the propylene glycol-like linker allows the ligand to enter the hydrophobic part of the binding pocket, where the two aromatic ring systems engage in additional interactions with the protein (Fig. 6a).

To computationally assess the binding of compounds **1** and **2** to PPARγ, absolute protein-ligand binding free-energy perturbation (ABFEP) calculations were carried out[51,52]. Compounds **1** and **2** as well as different ligands from ChEMBL with known sub-micromolar PPARγ activity (ChEMBL IDs: ChEMBL391987, ChEMBL241299, ChEMBL213355, ChEMBL212591) were modeled into the PPARγ-aleglitazar X-ray crystal structure (PDB ID: 3G9E)[41]. Compounds **1** and **2** have calculated Δ Gibbs Free Energy (ΔG) values of -20.1 kcal·mol⁻¹ and -19.7 kcal·mol⁻¹, respectively. These values are in the range of other known PPARγ ligands with sub-micromolar activity, further supporting the relevance of the proposed molecules for PPAR modulation (Table 4, Fig. S17).
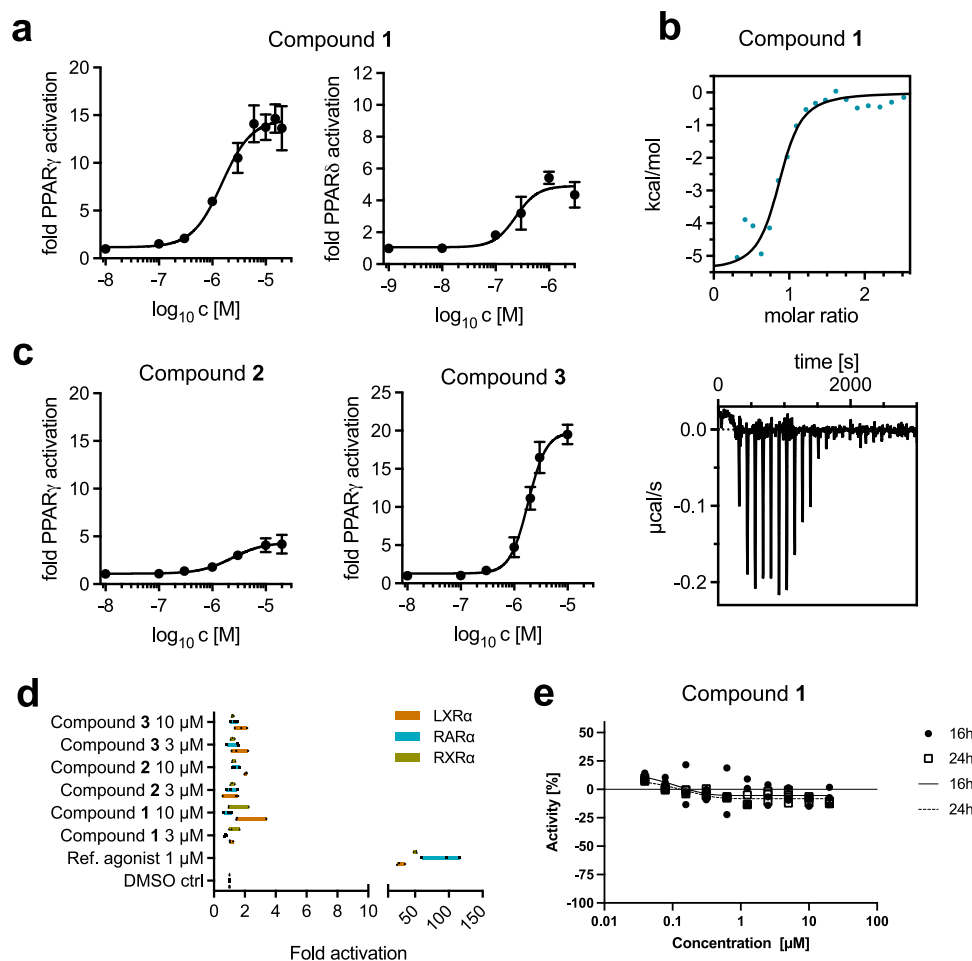
Fig. 5 | Biological characterization of compounds 1–3. The graphs in panels **a** and **c** display a nonlinear fitting curve derived from the mean values of three measurements ($N = 3$). The graph in panel **d** shows mean values of three measurements ($N = 3$). Error bars, represented by whiskers, indicate the standard deviations. **a** Peroxisome proliferator-activated receptor (PPAR) activation by compounds **1, 2** and **3** using a hybrid reporter gene assay. **b** Result of an isothermal titration calorimetry experiment ($N = 2$) measuring direct binding of compound **1** to PPARγ. **c** Dose-response curves from a hybrid reporter gene assay for compounds **2** and **3** measuring PPARγ activation. **d** Receptor selectivity of compounds **1, 2**, and **3** for activation of liver X receptor (LXR)α, retinoic acid receptor (RAR)α, and retinoid X receptor (RXR)α. **e** Cytotoxicity of compound **1** on HEK293T cells for two time points (i.e., 16 h and 24 h), 10 different concentrations (i.e., 0.05–20 μM) and 10000 cells · well^{-1} ($N = 3$). Scale reference: The axes are scaled through neutral control (i.e., Dimethylsulfoxid [DMSO], set to 0) and inhibitor control wells (i.e., 20 μM Staurosporine[114], set to −100). Source data are provided as a Source Data file.

Discussion

The generative deep learning method referred to as DRAGONFLY was evaluated in the context of ligand-based and structure-based molecular design tasks. The collective results specifically highlight the success of structure-based de novo design of potent partial agonizts for PPARγ. These molecules effectively interact with the receptor in a canonical binding mode, while also demonstrating the desired selectivity towards the receptor and favorable ADME properties.

By leveraging an interactome-based deep learning approach and employing a graph-to-sequence neural network architecture, DRAGONFLY addresses certain challenges commonly encountered in generative molecular design methods. This approach demonstrated to (i) achieve similar or even superior results compared to a respectively fine-tuned RNN-based CLM for drug-like ligand templates, (ii) enable structure-based design using 3D protein binding sites, and (iii) effectively incorporate desired physical and chemical properties into the generated molecules. Its ability to combine structure-based and ligand-based approaches, as well as its capacity to incorporate desired properties makes it a potentially useful tool for medicinal chemistry.

The design algorithm has demonstrated its capability to successfully generate molecules with desired properties by incorporating an additional encoding within the input. This encoding allows for the translation of various drug discovery-relevant properties with high accuracy into the generated molecules. Properties such as molecular weight, the number of rotatable bonds, hydrogen-bond acceptors, hydrogen-bond donors, polar surface area, and lipophilicity can be effectively encoded and incorporated into the molecular design process. This means that the algorithm can generate molecules that not only possess the desired structural characteristics but also meet specific physical and chemical property requirements. The ability to accurately translate these user-defined properties into the generated molecules is a potentially substantial advantage of the approach. It enables researchers to identify novel molecules with specific properties and optimize them for desired therapeutic effects, bioavailability, and safety profiles. In an initial assessment, the top-ranking computer-generated molecules revealed favorable in vitro ADME properties.

The results of the study also indicate that ligand-based de novo design outperformed structure-based models for the majority of investigated molecular properties. This performance difference could be attributed to the complexity of the input and the availability of training data. Whereas a small-molecule graph typically comprises up to 200 edges describing covalent bonds, protein binding sites represented by 3D graphs are considerably larger with an average scaling factor of 60. Furthermore, the ligand-based data set used in the study
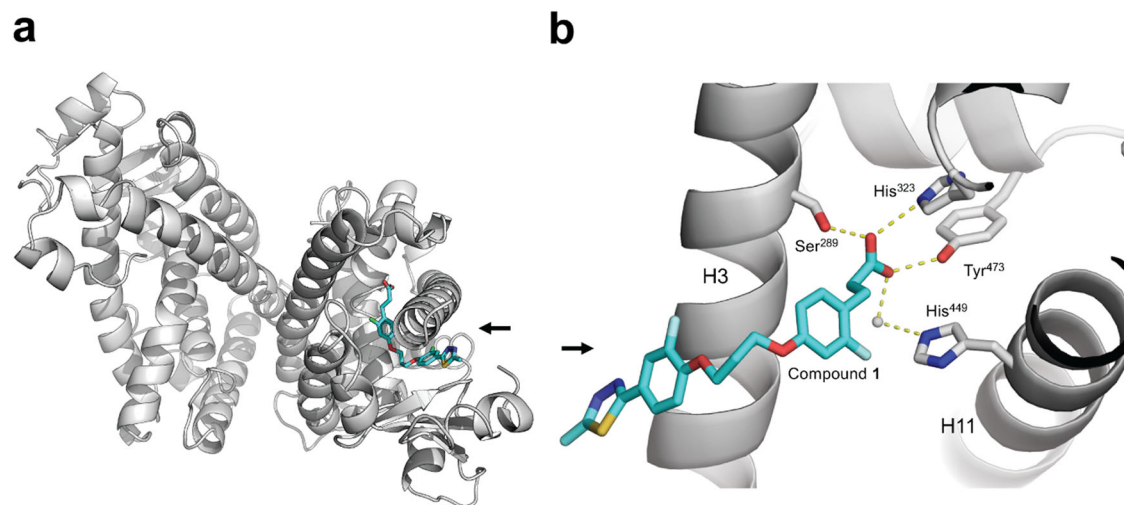
Fig. 6 | Protein-ligand co-crystallization. Crystal structure complex of the ligand-binding domain of human nuclear receptor PPARγ and compound **1**. (**a**) Illustration of the asymmetric unit containing two protein molecules, one of which shows compound **1** bound. Refined cartoon structure of the complex showing compound **1** (stick model in blue) bound to chain A and chain B empty (chain A on the right and chain B on the left). (**b**) Close-up view of the binding pose of compound **1** (blue)

with hydrogen bonds to residues Tyr[473], His[323], and Ser[289] as well as a water-mediated hydrogen bond to His[449] shown as dashed lines (PDB-ID: 8PBO). The black arrowhead marks the entry to the binding pocket of the receptor. To facilitate perceptibility, certain protein residues have been omitted from the illustration. Source data are provided as a Source Data file.

consisted of around 501 k distinct bioaffinity values, whereas only about 236 k bioaffinities were accessible to the structure-based training procedure. This disparity in training data availability may contribute to the superior performance of the ligand-based design models. Nevertheless, it is important to acknowledge that one benefit of structure-based design applications is their flexibility in not mandating exceptionally high-quality query ligands for the generation of molecules. This applicability can be valuable for in silico library design in scenarios where relevant ligand information is limited or unavailable, e.g., for newly identified disease-relevant macromolecular targets.

Such a scenario was emulated in the prospective application to PPARγ. However, it is worth noting that in this study, QSAR models were employed for scoring, and these models were trained on existing ligand activity data. The successful machine learning from the relevant training data is evident in the discovery that compound **1** interacts with the receptor in the canonical binding mode[53], as evidenced in the crystallographic complex. Further studies will be essential to combine DRAGONFLY with scoring functions not involving known active query ligands for bioaffinity assessment. Structure-based scoring for binding pose estimation, bioactivity prediction and virtual screening has indeed shown to be one of the most challenging topics in computational drug design[54]. Prominent among the existing structure-based

scoring models are free energy perturbation (FEP) techniques[51,55], geometric deep learning approaches[56–59], machine-learned force fields[60], and purely statistics-driven models[61–63], which currently receive considerable attention. Moreover, emphasis will be directed towards the utilization of DRAGONFLY applications to create bioactive ligands for protein models derived from apo protein structures (where no ligand is bound), and structure prediction methods like AlphaFold[64]. Understanding the algorithm's performance in these scenarios will provide valuable insights into its applicability and potential limitations in de novo drug design with structure-based ligand scoring and predicted protein structures.

The comparison between DRAGONFLY models trained on SMILES-strings and SELFIES demonstrated similar overall results in terms of different templates and properties. However, certain trends were observed: Libraries generated using SELFIES exhibited a higher level of diversity and novelty, while libraries generated using SMILES-strings achieved higher accuracy in incorporating desired molecular properties such as synthesizability, physical and chemical properties, or predicted bioactivity. These findings unveil specific strengths and trade-offs associated with de novo design approaches based on SELFIES and SMILES strings. Both methodologies have their advantages and limitations, and the choice between them depends on specific requirements and available data.

The combination of DRAGONFLY with scoring functions incorporating compound synthesizability, novelty, and bioactivity towards one or multiple targets using various descriptors was demonstrated to be feasible. By applying a tailored combination of these properties, we achieved promising results, i.e., in the exemplary case of PPARγ. The approach allowed for the identification of top-ranking molecules that exhibited sufficient structural and scaffold novelty, synthesizability, and a desired bioactivity profile across multiple targets. While molecular novelty has been extensively discussed in recent literature, most studies mainly focus on descriptor similarity using structural fingerprints[65,66]. We showed that incorporating additional scaffold criteria can enhance the novelty of top-ranking molecules. By assigning higher weights to different descriptors, distinct outcomes were observed. When higher weights were assigned to the ECFP fingerprint descriptor, the generated molecules exhibited higher structural similarity to known ligands. This approach favored the exploration of the

Table 4 | Absolute protein-ligand binding free-energy perturbation calculations

ChEMBL ID	ΔG / kcal · mol⁻¹	pEC50
Compound **1**	-20.1	5.83
Compound **2**	-19.7	5.64
ChEMBL391987	-22.3	6.94
ChEMBL241299	-18.4	6.51
ChEMBL213355	-19.1	6.82
ChEMBL212591	-19.7	6.64

The numbers represent the calculated absolute protein-ligand binding free-energy perturbation (ABFEP, ΔG) for compounds **1** and **2**, and ligands from the ChEMBL database with known sub-micromolar PPARγ activity (database entries ChEMBL391987, ChEMBL241299, ChEMBL213355, ChEMBL212591). The ABFEP calculations were executed with default parameters over a simulation period of 5 ns. (see Methods).

known chemical space and enabled the design of molecules with recognizable structural features. In contrast, assigning higher weights to the two "fuzzy" descriptors, CATS and USRCAT, resulted in top-ranking molecules that deviated further from the structurally explored space, as expected[67,68]. This latter approach focused on prioritizing novelty and divergence from known ligands, allowing for the exploration of new chemical territories. These findings highlight the flexibility and versatility of the scoring functions and the ability to customize the weights of different descriptors to achieve specific objectives. By balancing the importance of structural similarity, novelty, and other desired properties, we were able to guide the generation of molecules that met the desired criteria for the targeted application, such as PPARγ in this case.

The chemical synthesis of two top-ranking de novo designs, designated as compounds **1** and **2**, along with regioisomer **3**, turned out to be comparably cumbersome, requiring 10 and 5 synthesis steps, respectively. This observation together with low yields point to limitations of the employed scoring function for synthesizability, motivating the development of better measures safeguarding straightforward synthesis of molecules designed with a generative models. For example, a hybrid structure generator for DRAGONFLY that combines rule-based molecule selection with predictive deep learning models could be envisaged[69,70].

Subsequent biological evaluation of compounds **1**–**3** was carried out using the human PPARγ ligand-binding domain. Both, cell-based activity and direct binding, could be confirmed. These investigations led to the identification of novel PPAR modulators that exhibited low micromolar to high nanomolar activity. Importantly, the intended behavior and specificity characteristics for which the two designs were originally prioritized received confirmation through experimental verification. These results highlight their ability to target PPAR with precision, while also sidestepping pronounced influence on closely affiliated nuclear hormone receptors, like RXR, and a sizeable panel of other undesired off-targets. This outcome demonstrates the efficacy of the structure-based DRAGONFLY de novo design approach in generating molecules with the desired properties and biological activities, including selected ADME properties. The observed lack of CYP interaction up to a compound concentration of $10\,\mu$M is a crucial aspect in averting drug-drug interactions, which is of particular relevance for the treatment of metabolic syndrome, where patients frequently require concurrent administration of multiple drugs[71]. These results represent a substantial milestone as they showcase the successful application of a generative deep learning model for molecular de novo design that incorporates ligand activity and selectivity on multiple targets, as well as panel selectivity within the same protein class, exemplified in the context of nuclear hormone receptors.

The concept of interactome-based deep learning was introduced to de novo molecular design as a means to maximize the information learned about interaction networks between drug targets and their ligands. It could be demonstrated that by employing an interactome-based training procedure, some of the limitations encountered by transfer-learning-based CLMs can be mitigated. It is important to note that the concept of interactome-based deep learning is not limited to the specific neural network architecture or drug-target graph presented in this study. There is room for exploration within various frameworks and methodologies. For instance, other neural network architectures, such as sequence-to-sequence models using transformer neural networks[72], or graph-to-graph-based architectures utilizing diffusion-based models[20], could be implemented. These variations would enable additional ligand-based design approaches. The graph-to-graph models would extend the capability to incorporate 3D information for structure-based molecular design, while sequence-to-sequence models would be limited to protein sequence information. Furthermore, the drug-target interactome could be expanded to include additional targets beyond those considered in the presented

study. For example, RNA binding sites, protein surface binders such as molecular glues or certain macrocycles, or protein-protein interactions could be included. This expansion of the interactome would enable the exploration of distinct design possibilities and target-specific applications.

In the presented context, interactome-based deep learning serves as a proof-of-concept for "zero-shot" learning that can be further adapted and customized for specific applications in small molecule drug discovery, ultimately leading to more efficient hit-and-lead discovery in bioorganic and medicinal chemistry. By leveraging data-driven deep learning and interaction networks, this approach offers new avenues for foundation models enabling tailored molecular design strategies and the discovery of innovative drug candidates.

Methods
Neural network architecture
The DRAGONFLY method employs a graph neural network architecture[73–75]. This approach utilizes a GTNN model to encode the input molecular graph, which is represented as a 2D graph for ligands and a 3D graph for protein binding sites. The GTNN transforms the graph into a condensed one-dimensional (1D) feature vector. Subsequently, this feature vector is decoded back into the corresponding molecular string, using a CLM based on an RNN-LSTM[32,76] architecture for the molecule generation process.

Graph transformer neural network. *Message passing*: The atomic features were embedded and transformed using a multilayer Perceptron (MLP) to obtain atomic feature vectors $\mathbf{h}_i^0$. Message passing as suggested by Satorras et al.[77] and used in other 3D-based prediction tasks[78,79] was applied to $L = 3$ layers, iteratively applied over all atomic representations $\mathbf{h}_i^0$. Edges were introduced differently in the 2D and 3D graph representations. In the 2D graph, edges were established between atoms connected by covalent bonds. On the other hand, in the 3D graph, edges were formed between all atoms situated within a radius of 4 Å from each other. This approach ensured that the molecular structures were accurately represented in both 2D and 3D formats, effectively capturing the most relevant interactions occurring between atoms. In each iteration of the message-passing layer, the atomic representations underwent a transformation as described by Equation (1).

$$\mathbf{h}_i^{l+1} = \phi\left(\mathbf{h}_i^l, \sum_{j \in \mathcal{N}(i)} \psi\left(\mathbf{h}_i^l, \mathbf{h}_j^l\right)\right), \quad (1)$$

for 2D graph structures, and Equation (2)

$$\mathbf{h}_i^{l+1} = \phi\left(\mathbf{h}_i^l, \sum_{j \in \mathcal{N}(i)} \psi\left(\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{r}_{i,j}\right)\right), \quad (2)$$

for 3D graph structures.

In Equations (1) and (2) $\mathbf{h}_i^l$ is the atomic representation $\mathbf{h}$ of the i-th atom at the l-th layer; $j \in \mathcal{N}(i)$ is the set of neighboring nodes of atom i connected via edges; $\mathbf{r}_{i,j}$ the inter-atomic distance represented in terms of Fourier features, using a sine- and cosine-based encoding; ψ is an MLP transforming node features into message features $\mathbf{m}_{ij}$: $\mathbf{m}_{ij} = \psi(\mathbf{h}_i^l, \mathbf{h}_j^l, \mathbf{r}_{i,j})$ for 3D graphs, and $\mathbf{m}_{ij} = \psi(\mathbf{h}_i^l, \mathbf{h}_j^l)$ for 2D graphs; $\sum$ denotes the permutation-invariant pooling operator (i.e., sum) transforming $\mathbf{m}_{ij}$ into $\mathbf{m}_i$: $\mathbf{m}_i = \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}$; and ϕ is an MLP transforming $\mathbf{h}_i^l$ and $\mathbf{m}_i$ into $\mathbf{h}_i^{l+1}$. The atomic features from all layers $[\mathbf{h}_i^{l=1}, \mathbf{h}_i^{l=2}, \mathbf{h}_i^{l=3}]$ were concatenated and transformed via an MLP, resulting in final atomic features $\mathbf{H}_i$. The features $\mathbf{H}_i$ were subsequently pooled into a molecular representation via a graph multiset transformer (GMT) and further transformed via two MLPs to the two 1D latent space

representations $\mathbf{l}_{t=0}^1$ and $\mathbf{l}_{t=0}^2$. A detailed description of the GMT module can be found elsewhere[30].

Long-short-term memory neural network. LSTM neural networks represent a specific category of recurrent neural networks renowned for their capacity to understand and produce sequences of characters. Their proficiency in comprehending sequential data and capturing intricate temporal connections renders them suitable for de novo drug design applications. In this context, the LSTM architecture was integrated to convert the acquired hidden states from the GTNN (i.e., $\mathbf{l}t=0^1$ and $\mathbf{l}t=0^2$) into a molecule represented in string form (SMILES or SELFIES). $\mathbf{l}_{t=0}^1$ and $\mathbf{l}_{t=0}^2$ are used as the initial hidden states of the LSTM architecture. At each time step t the next character of the sequence ω_{t+1} is predicted given the two hidden states $\mathbf{l}_t^1$ and $\mathbf{l}_t^2$, the two memory cell states $\mathbf{c}_t^1$ and $\mathbf{c}_t^2$, and the embedding $\mathbf{k}_t$ of the previous character in the sequence ω_t. This transformation is conducted using four non-linear transformations via Equation (3):

$$
\begin{aligned}
\mathbf{g}_i &= \sigma(\mathbf{W}_{ix}\mathbf{k}_t + b_{ix} + \mathbf{W}_{il}\mathbf{l}_{t-1} + b_{il}) \\
\mathbf{g}_f &= \sigma(\mathbf{W}_{fx}\mathbf{k}_t + b_{fx} + \mathbf{W}_{fl}\mathbf{l}_{t-1} + b_{fl}) \\
\mathbf{g}_o &= \sigma(\mathbf{W}_{ox}\mathbf{k}_t + b_{ox} + \mathbf{W}_{ol}\mathbf{l}_{t-1} + b_{ol}) \\
\widetilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_{cx}\mathbf{k}_t + b_{cx} + \mathbf{W}_{cl}\mathbf{l}_{t-1} + b_{cl}) \\
\mathbf{c}_t &= \mathbf{g}_f \odot \mathbf{c}_{t-1} + \mathbf{g}_i \odot \widetilde{\mathbf{c}}_t \\
\mathbf{l}_t &= \mathbf{g}_o \odot \mathbf{c}_t
\end{aligned}
\tag{3}
$$

where $\mathbf{l}_t$ and $\mathbf{c}_t$ represent the hidden state and the memory cell state at time t, respectively. $\mathbf{g}_i$, $\mathbf{g}_f$ and $\mathbf{g}_o$ represent the input, forget, and output gates, respectively. σ and $\odot$ indicate the sigmoid activation function and the Hadamard product[80], respectively. $\widetilde{\mathbf{c}}_t$ represents the candidate memory cell state, which is used to update the previous memory cell state $\mathbf{c}_{t-1}$. $\mathbf{W}$ and b are the weights and biases used for the corresponding linear transformations. The resulting updated hidden state $\mathbf{l}_t$ is then transformed using a softmax activation function to obtain a logit vector $\hat{\mathbf{y}}_t$ (i.e., a vector with the dimension of the alphabet Ω) via Equation (4):

$$
\hat{\mathbf{y}}_t = \text{softmax}(\mathbf{W}_{yl}\mathbf{l}_t + b_{yl})
\tag{4}
$$

Throughout the training phase, the cross-entropy loss was computed based on $\hat{\mathbf{y}}t$ and the ground truth $\mathbf{y}t$. The ground truth vector $\mathbf{y}_t$ was structured with zeros in all positions except for the character's anticipated location, which was assigned a value of 1 for each prediction in the sequence. Subsequently, this calculated loss was backpropagated seamlessly through the LSTM and GTNN networks in an end-to-end manner. The training process involved the application of teacher forcing, as described in the work by Lamb et al.[81].

Molecule sampling
Temperature sampling was employed as a mechanism to facilitate the generation of a diverse array of output molecules using a trained DRAGONFLY model[7], achieved through Equation (5):

$$
P(\hat{\mathbf{y}}_{t+1} = \omega | \hat{\mathbf{y}}_{t=0}, \ldots, \hat{\mathbf{y}}_t) = \frac{\exp(\hat{\mathbf{y}}_t^\omega / T)}{\sum_\omega^\Omega \exp(\hat{\mathbf{y}}_t^\omega / T)}
\tag{5}
$$

where T is the temperature value, and P the probability of the output representation $\hat{\mathbf{y}}_{t+1}$ being the character ω given all previous outputs. The character sampling process was regulated by the temperature parameter T. When T is set to a high value ($T \to \infty$), character probabilities tend to equalize across all characters. Conversely, as T decreases towards 0, the highest likelihood predicted by $\hat{\mathbf{y}}_{t+1}$ approaches 1. In the context of DRAGONFLY applications, four distinct temperature values (0.2, 0.5, 0.8, 1.1) were investigated. A value of

$T = 0.5$ was found to strike the most favorable balance between novelty, diversity, the prediction of active compounds, and synthesizability, as indicated by the outcomes presented in Figs. S9–S10.

Atom featurization
Small molecules: The atomic properties of small-molecule ligands were encoded via the following embeddings: 10 atom types [H, C, N, O, F, P, S, Cl, Br, I], two ring types [True, False], two aromaticity types [True, False], and four hybridization types [sp^3, sp^2, sp, s].

Proteins: The protein binding site was defined by all protein atoms that are within a 5 Å radius to a ligand atom. The atomic properties of the respective protein binding sites were encoded using the following four features: (i) an embedding of the atom types using 22 different embeddings, (ii) an embedding of the combination of amino acid and atom types covering 225 different embeddigs, (iii) the distance to the closest atom of the bound small-molecule ligand, (iv) the calculated B factor, aiming to quantify protein flexibility and intrinsic disorder at the corresponding atom (Section S3).

Bond types: Edges were represented by inter-atomic distance in terms of Fourier features, using a sine- and cosine-based encoding for 3D graphs[82]. No edge features were used for 2D graphs. Edges were introduced between covalently bound atoms for the 2D graphs, and between all atoms within a 4 Å radius from each other for the 3D graphs.

Hyperparameters
The selected hyperparameters for the neural network led to a combined count of trainable parameters amounting to 6.94 million (3.49 million for the GTNN encoder and 3.45 million for the LSTM decoder) for the ligand-based design DRAGONFLY model. Similarly, the structure-based design DRAGONFLY model encompassed 7.01 million trainable parameters (3.56 million for the GTNN encoder and 3.45 million for the LSTM decoder).

Scoring
Quantitative structure-activity relationship. Kernel ridge regression (KRR) was employed to establish QSAR models based on descriptors and fingerprints. Kernel-based machine learning, rooted in the work of Krige[83], resides within the realm of supervised learning techniques and has found application across a spectrum of machine learning investigations[84-87]. The assessment of similarity between two molecules i and j was carried out utilizing the Laplacian Kernel (Eq. (6)):

$$
k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||_1}{\sigma}\right)
\tag{6}
$$

where $\mathbf{x}_i$ is the molecular descriptor or fingerprint of molecule i and σ is the length scale hyperparameter. Herein, σ was set to 51.2, after screening 0.12^i for i in range (1, 20). Three different molecular descriptors were applied in this study, namely, extended-connectivity fingerprints (ECFP, radius = 2, dimension = 512)[36], chemically advanced template search (CATS) with absolute feature frequencies[67,88], and ultrafast shape recognition with pharmacophoric constraints (USRCAT)[38]. Once the kernel matrix $\mathbf{K} = k(\mathbf{x}_i, \mathbf{x}_j)$ was calculated, the fitting coefficients $\boldsymbol{\alpha}$ were computed via the inverse of the kernel matrix $\mathbf{K}$ via Equation (7):

$$
\boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}
\tag{7}
$$

where λ denotes the regularization strength (herein, optimized to 10^{-7}), $\mathbf{I}$ the identity matrix, and $\mathbf{y}$ the labels of the molecules (herein bioactivity to the investigated target). Given a labeled data set with N molecule-label pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, a function was obtained that maps the molecular descriptor of a novel molecule $\mathbf{x}_q$ to its predicted bioactivity

$\hat{y}_q$ via Equation (8):

$$\hat{y}_q(\mathbf{x}_q) = \sum_{i}^{N} \boldsymbol{\alpha}_i \cdot k(\mathbf{x}_i, \mathbf{x}_q) \qquad (8)$$

Molecular novelty. The novelty of the generated molecules was assessed through two distinct metrics: structural novelty score (S_{ECFP}) and scaffold novelty score ($S_{scaffold}$). The structural novelty score (S_{ECFP}) was established based on the Jaccard distance (1 minus Tanimoto similarity[89]) concerning the most similar molecule within the training data set using ECFP[36] descriptors. The Jaccard distance attains a value of 1 between two molecules when they possess no common structural attributes as identified by ECFP (bits within the ECFP vector). Conversely, it reaches a value of 0 when two distinct molecules share identical structural features (identical ECFP vectors). The scaffold novelty score ($S_{scaffold}$) gauges the novelty of both the atom scaffold (commonly referred to as the Murcko scaffold[90]) and the carbon scaffold (also known as the skeleton scaffold[91]) present in a generated molecule. Atom scaffolds were determined by considering the rings and branches of a specific template molecule. In this process, substituents were eliminated, while the identity of atoms and bonds remained unaltered (as detailed in SI2.4). Carbon scaffolds were identified by the carbon framework of a molecule, wherein all non-hydrogen atoms were transformed into carbon atoms and all bonds were replaced by single bonds (illustrated in Fig. S7). The scaffold novelty score was formulated by incorporating both atom and carbon scaffold scores. Each of these scores determined whether the corresponding scaffold was present in any molecule within the training set, as determined by Equations ((9)– (11)).

$$S_{atom} = \begin{cases} 0, & \text{if atom scaffold in training set} \\ 0.1, & \text{otherwise} \end{cases} \qquad (9)$$

$$S_{carbon} = \begin{cases} 0, & \text{if carbon scaffold in training set} \\ 0.1, & \text{otherwise} \end{cases} \qquad (10)$$

$$S_{scaffold} = S_{atom} + S_{carbon} \qquad (11)$$

Both structural and scaffold novelty contribute to the overall novelty score, i.e., Equation (12), ranging from 0 (for molecules very close to molecules the training set) to 1.2 (for molecules with no ECFP overlap with the training set and no shared scaffolds).

$$S_{novelty} = S_{ECFP} + S_{scaffold} \qquad (12)$$

Molecular property analysis
Molecular data sets were generated using a DRAGONFLY model, which was trained on a comprehensive data set excluding proteins and ligands associated with 20 specified targets. These targets are listed in Tables S2 and S3. For each target 2000 random molecules were selected. The physicochemical properties of these molecules were computed and subsequently used as input for the DRAGONFLY model. The properties of the generated molecules were visualized in a scatter plot (Fig. 2a) and summarized in Table 3. The scatter plot illustrates the relationship between the actual and predicted properties of the molecules. The mean absolute errors (MAEs) and Pearson correlation coefficients (*r*) were calculated to assess the predictive performance of the DRAGONFLY model. These statistical measures were derived by comparing the extracted properties of the generated molecules against the properties of the original data set.

Drug-target interactome preprocessing
The data necessary for constructing the drug-target graph, referred to as the "interactome," was sourced from two distinct databases: ChEMBL[28] (Version 29) and PDBBind[92] (Version 2020).

Preprocessing ChEMBL data. To acquire the necessary interactome data, the ChEMBL29 database[28] was queried. Similar to prior studies[93], this data extraction process was divided into two stages: In the initial step, a compilation of biological targets was obtained. Subsequently, compounds were extracted for which specific activities against any of these targets were annotated. Single-protein targets that possessed assay information for a minimum of 10 compounds with unique internal identifiers were retrieved from the ChEMBL database. A series of activity and annotation filters were then applied to these compounds. The molecules underwent neutralization, and any salts and solvents were eliminated. For compounds comprising multiple distinct fragments following this "washing" procedure, all but the fragment with the highest number of heavy atoms were discarded. Furthermore, molecules containing <3 or >100 heavy atoms, as well as radical species, were excluded from the data set. This procedure yielded a data set of 742 k unique SMILES-strings with annotated biologic affinity. Using a cut-off of a binding affinity of 200 nM, removing duplicates, a maximal SMILES-string length of 97 (using the longest SMILES-length from five randomized sampled SMILES-strings) for the ligand, and a minimum number of five ligands per target resulted in a drug-target graph consisting of 501 k unique binding affinities for 360 k unique ligands and 2989 unique target-IDs.

Preprocessing PDBbind. The PDBbind database (Version 2020) was obtained by downloading it from the link http://www.pdbbind.org.cn/download.php, which yielded a collective count of 19,443 protein-ligand structures. After filtering out structures annotated with "incomplete ligand structure", "covalent complex," or "incomplete ligand structure", a total of 19,000 entries remained. Additionally, a more refined filtering process was conducted, excluding structures with ligand molecular weights outside the range of 100–1200 g mol^{-1} and binding affinities >10 μM. This filtration yielded a collection of 17,824 structures. This curated list of entries was then cross-referenced with the target-IDs present within the drug-target graph used for ligand-based design. This specific graph contained 501,000 unique binding affinities encompassing around 360,000 unique molecules and 2989 unique target-IDs. The outcome of this mapping effort revealed a total of 8351 distinct protein structures associated with 744 unique target-IDs. By refining the drug-target graph to exclusively include target-IDs with annotated PDB structures, the modified graph encompassed around 263,000 unique binding affinities spanning around 208,000 unique molecules and 744 unique target-IDs. The connection between PDB-IDs and target-IDs within ChEMBL was facilitated through UNIPROT-IDs, given that both databases provide UNIPROT-IDs for individual proteins.

Numerous drug targets exhibit multiple binding sites, including orthosteric sites and various allosteric sites[94]. Although such details were not present in the ChEMBL database, recognizing these distinct binding sites was deemed essential for effective drug-target interactome learning.molecules known for their allosteric modulation were extracted from the reference cited as Ref. 95. Subsequently, the drug-target graph underwent a modification whereby target-IDs encompassing both allosteric and orthosteric ligands were treated as distinct target-IDs.

Chemical alphabet
DRAGONFLY models underwent training using two distinct chemical alphabets: SMILES strings[3] and SELFIES[40]. To discern the distinct character types in both types of strings, 10 randomly generated SMILES strings were created for each molecule within the data set. For

SMILES strings, all observed characters surrounded by brackets ([]), as well as some frequently occurring functional groups (e.g., sulfoxide, nitro, ketone, nitrile) were encoded as a single token (SI5). In both string types, three supplementary characters were introduced to serve as markers for the beginning, end, and padding of the strings: x, y, and z for SMILES-strings, and [\\X], [\\Y], and [\\Z] for SELFIES. Following this procedure, a SMILES-string alphabet Ω_{SMILES} was established, comprising a total of 57 characters. A SELFIES alphabet $\Omega_{SELFIES}$ was constructed, encompassing a total of 85 characters (as detailed in Table S1).

Absolute free binding energy calculations

Molecules **1** and **2** as well as different ligands from ChEMBL with known PPARγ activity (ChEMBL IDs: ChEMBLl391987, ChEMBL241472, ChEMBL241299, ChEMBL213355, ChEMBL212591) were modeled into the PPARγ-aleglitazar crystal structure (PDB ID: 3G9E)[41]. The chosen reference molecules from the ChEMBL database were selected based on their structural similarity to compounds **1** and **2** (i.e., possessing (i) a carboxylic acid as head group, (ii) an alkyl or polyethylene glycol linker, and (iii) an aromatic tail), and their comparable binding affinity (i.e., EC$_{50}$ values ≤5 μM and ≥100 nM). After structure preparation, ABFEP simulations were carried out with Schrödinger software (release 2023-4) using default settings and a simulation time of 5 ns for both complex and solvent[96]. The lowest calculated free energies were obtained for the co-crystallized ligand aleglitazar (EC$_{50}$ = 21 nM) and ChEMBL241472 (EC$_{50}$ = 140 nM) (Fig. S17).

Cytotoxicity assay on HEK293T cells

HEK293T cells were seeded at the indicated number per well in DMEM-high glucose, complemented with glutamax, pen-strep, and 10% FBS, in a total of 40 μl of medium. The cells were incubated overnight at 37 °C. Compounds were added to the cells at the indicated concentrations, resulting in a final Dimethylsulfoxid (DMSO) concentration of 0.2%. The compounds were incubated on the cells for either 16 h or 24 h. At the specified time point, the medium was carefully removed from the vessel, leaving only 2 μl in the wells. Celltiter-glo (CTG) reagent (G7572, Promega) was prepared according to the manufacturer's instructions. Plates with cells were equilibrated at room temperature for 30 min. Subsequently, 25 μl of CTG reagent was added to the cells. The plates were then shaken for 2 min and incubated for an additional 15 min at room temperature. Luminescence was read afterward with BG Pherastar.

Biological characterization

Compounds **1**–**3** were characterized in a hybrid reporter gene assay for their agoniztic effect on human nuclear receptors PPAR$\alpha/\gamma/\delta$, RXRα, FXRα, RARα in HEK293T cells. Compound **1** was tested in an isothermal titration calorimetry (ITC) assay to measure direct binding affinity to the ligand-binding domain of PPARγ. ADME properties were measured in standardized assays at Roche.

Hybrid reporter gene assays. PPAR activation was determined in uniform Gal4-hybrid reporter gene assays for the PPARα, PPARγ and PPARδ isoforms in HEK293T cells (German Collection of Microorganisms and Cell Culture GmbH, DSMZ) which were transiently transfected with pFR-Luc (Stratagene, La Jolla, CA, USA; reporter) and pRL-SV40 (Promega, Madison, WI, USA; internal control) and one pFA-CMV-hPPAR-LBD[97] clone, coding for the hinge region and ligand binding domain of the canonical isoform of human PPARα, PPARγ, PPARδ or respectively. 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⁻¹), and streptomycin (100 μg · ml⁻¹) at 37 °C and 5% CO$_2$ and seeded in 96-well plates (3 × 10⁴ cells per well). After 24 h, 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⁻¹), streptomycin (100 μg · ml⁻¹) 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, 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 a test compound by the mean RLU of the untreated control. All samples were tested in at least three biologically independent experiments in duplicates. For dose-response curve fitting and calculation of EC$_{50}$ values, the equation "[Agonist] versus response (variable slope−four parameters)" was used in GraphPad Prism (version 7.00, GraphPad Software, La Jolla, CA, USA) with fold activation data. The reference agonizts GW7647 (PPARα)[98,99], pioglitazone (PPARγ)[100,101] and L165,041 (PPARδ)[102,103] were used to validate the assays and to monitor assay performance. Nuclear receptor selectivity profiling was performed with corresponding pFA-CMV-hNR-LBD clones and suitable reference agonizts on RARα (pFA-CMV-hRARα-LBD[104], 1 μM tretinoin), LXRα (pFA-CMV-hLXRα-LBD[104], 1 μM TO901317) and RXRα (pFA-CMV-h RXRα-LBD[105], 1 μM Bexarotene).

Isothermal Titration Calorimetry (ITC). ITC experiments were conducted on an Affinity ITC instrument (TA Instruments, New Castle, DE) at 25 °C with a stirring rate of 75 rpm. PPARγ LBD protein (30 μM, prepared as described previously[106]) in buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol) containing 5% DMSO was titrated with the test compound (**1**) (100 μM in the same buffer containing 5% DMSO) in 21 injections (1 × 1μl and 20 × 5μl) with an injection interval of 120 s. The test compound was titrated into buffer, and the buffer was titrated to the PPARγ LBD proteins under otherwise identical conditions. The ITC results were analyzed using NanoAnalyze software (TA Instruments, New Castle, DE) with an independent binding model.

Protein-ligand co-crystallization

The following construct was used for expression and co-crystallization. PPARγ (L204-Y477) (UniProt ID: P37231-2): MGSS-6His-SG-TEV-(L204-Y477). Molecular weight: 33465 Da. Large-scale expression of human PPARγ was conducted in E. coli BL-21 (DE3) cells (SI10). Subsequently co-crystals of PPARγ were grown using 6 mg · ml⁻¹ protein in buffer: 20 mM Tris-HCl pH 8.0, 1 mM TCEP, 0.5 mM EDTA and 1 mM design **1** mixed with equal amounts of reservoir: 0.1 M Tris-HCl pH 7.5 and 1.6 M ammonium sulfate (Fig. S14). The structure was determination and refinement yielding the elucidated co-crystal structure with a resolution of 1.85 Å as depicted in Fig. 5 (Table S13 and Fig. S15).

Off-target screening

To test the specificity of compound **1** and **2**, both were subject to panel screen against 50 safety-relevant off-targets[107]. Both compounds have shown a clear profile not reaching ≥50% inhibition or binding at a concentration of 10 μM with the exception for PPARγ (Tables S9–S12).

Chemical synthesis

Compounds **1**–**3** were synthesized starting from commercial building blocks. The synthesis and the full analytical characterization of the final compounds and intermediates are described in SI13.

Co-crystallization

Compound **1** was co-crystalized with the ligand binding domain of human PPARγ (Leu²⁰⁴–Tyr⁴⁷⁷) (UniProt ID: P37231-2). The crystallographic structure is accessible from the Protein Data Bank[108] (PDB ID: 8PBO). Details about construct design, protein expression and

purification, crystallization, data collection, and structure determination and refinement can be found in SI10.

Data availability

Source data is provided in Source_Data.zip and available on Figshare, https://doi.org/10.6084/m9.figshare.25234159, represented by https://doi.org/10.6084/m9.figshare.25234159[109]. The individual files in the ZIP file are named according to their location in the manuscript, for example, Figure_2_MolLogP.csv or Figure_6.pdb. Source data are provided with this paper.

Code availability

A reference implementation of the DRAGONFLY method based on PyTorch[110] and PyTorch Geometric[111] is available at https://github.com/ETHmodlab/dragonfly_gen(rep. https://doi.org/10.5281/zenodo.10671327, https://zenodo.org/record/10671327)[112].

References

1. Schneider, G. & Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).
2. Schneider, G. & Clark, D. E. Automated de novo drug design: are we nearly there yet? *Angew. Chem. Int. Ed.* **58**, 10792–10803 (2019).
3. 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).
4. Yuan, W. et al. Chemical space mimicry for drug discovery. *J. Chem. Inf. Model.* **57**, 875–882 (2017).
5. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
6. Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. De novo design of bioactive small molecules by artificial intelligence. *Mol. Inf.* **37**, 1700153 (2018).
7. Moret, M., Friedrich, L., Grisoni, F., Merk, D. & Schneider, G. Generative molecular design in low data regimes. *Nat. Mach. Intell.* **2**, 171–180 (2020).
8. Moret, M. et al. Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat. Commun.* **14**, 114 (2023).
9. Grisoni, F., Moret, M., Lingwood, R. & Schneider, G. Bidirectional molecule generation with recurrent neural networks. *J. Chem. Inf. Model.* **60**, 1175–1183 (2020).
10. Grisoni, F. & Schneider, G. Molecular design with long short-term memory networks. *J. Comput. Aided Mol. Des.* **20**, 35–42 (2019).
11. Skinnider, M. A., Stacey, R. G., Wishart, D. S. & Foster, L. J. Chemical language models enable navigation in sparsely populated chemical space. *Nat. Mach. Intell.* **3**, 759–770 (2021).
12. Blaschke, T. et al. Reinvent 2.0: an ai tool for de novo drug design. *J. Chem. Inf. Model.* **60**, 5918–5922 (2020).
13. Fialková, V. et al. Libinvent: reaction-based generative scaffold decoration for in silico library design. *J Chem. Inf. Model.* **62**, 2046–2063 (2021).
14. Schneider, P. et al. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2020).
15. Ilnicka, A. & Schneider, G. Designing molecules with autoencoder networks. *Nat. Comp. Sci.* **3**, 922–933 (2023).
16. Tropsha, A., Isayev, O., Varnek, A., Schneider, G. & Cherkasov, A. Integrating qsar modelling and deep learning in drug discovery: the emergence of deep qsar. *Nat. Rev. Drug Discov.* **23**, 141–155 (2024).
17. Ballarotto, M. et al. De novo design of nurr1 agonists via fragment-augmented generative deep learning in low-data regime. *J. Med. Chem.* **66**, 8170–8177 (2023).
18. Skalic, M., Jiménez, J., Sabbadin, D. & De Fabritiis, G. Shape-based generative modeling for de novo drug design. *J. Chem. Inf. Model.* **59**, 1205–1214 (2019).
19. Wang, M. et al. Relation: A deep generative model for structure-based de novo drug design. *J. Med. Chem.* **65**, 9478–9492 (2022).
20. Schneuing, A. et al. Structure-based drug design with equivariant diffusion models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2210.13695 (2022).
21. Igashov, I. et al. Equivariant 3d-conditional diffusion models for molecular linker design. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2210.05274 (2022).
22. Atz, K., Grisoni, F. & Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **3**, 1023–1032 (2021).
23. Isert, C., Atz, K. & Schneider, G. Structure-based drug design with geometric deep learning. *Curr. Opin. Struct. Biol.* **79**, 102548 (2023).
24. Zhang, J. D. & Wiemann, S. Kegggraph: a graph approach to kegg pathway in r and bioconductor. *Bioinformatics* **25**, 1470–1471 (2009).
25. Sun, J., Wu, Y., Xu, H. & Zhao, Z. Dtome: A web-based tool for drug-target interactome construction. *BMC Bioinform.* **13**, S7 (2012).
26. Li, Z.-C. et al. Identification of drug–target interaction from inter-actome network with 'guilt-by-association'principle and topology features. *Bioinformatics* **32**, 1057–1064 (2016).
27. Crouzet, S. J. et al. G-plip: Knowledge graph neural network for structure-free protein-ligand bioactivity prediction. Preprint at *bioRxiv* https://doi.org/10.1101/2023.09.01.555977 (2023).
28. Mendez, D. et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **47**, D930–D940 (2019).
29. Satorras, V. G., Hoogeboom, E., Fuchs, F. B., Posner, I. & Welling, M. E (n) equivariant normalizing flows. *Advances in Neural Information Processing Systems (NeurIPS)*, **34**, 4181–4192 (2021).
30. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
31. Nippa, D. F. et al. Enabling late-stage drug diversification by high-throughput experimentation with geometric deep learning. *Nat. Chem.* **16**, 239–248 (2024).
32. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
33. Wildman, S. A. & Crippen, G. M. Prediction of physicochemical parameters by atomic contributions. *J. Chem. Inf. Comp. Sci.* **39**, 868–873 (1999).
34. Thakkar, A., Chadimová, V., Bjerrum, E. J., Engkvist, O. & Reymond, J.-L. Retrosynthetic accessibility score (rascore)–rapid machine learned synthesizability classification from ai driven retrosynthetic planning. *Chem. Sci.* **12**, 3339–3349 (2021).
35. Rupp, M., Tkatchenko, A., Müller, K.-R & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
36. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
37. Reutlinger, M. et al. Chemically advanced template search (cats) for scaffold-hopping and prospective target prediction for 'orphan' molecules. *Mol. Inf.* **32**, 133 (2013).
38. Schreyer, A. M. & Blundell, T. Usrcat: real-time ultrafast shape recognition with pharmacophoric constraints. *J. Cheminform.* **4**, 27 (2012).
39. Nakata, M., Maeda, T., Shimazaki, T. & Hashimoto, M. The Pub-ChemQC Project. *J. Chem. Inf. Model.* **57**, 1300–1308 (2017).
40. Krenn, M., Häse, F., Nigam, A., Friederich, P. & Aspuru-Guzik, A. Self-referencing embedded strings (selfies): a 100% robust molecular string representation. *Mach. Learn.: Sci. Technol.* **1**,

045024 (2020).

41. Bénardeau, A. et al. Aleglitazar, a new, potent, and balanced dual pparα/γ agonist for the treatment of type ii diabetes. *Bioorg. Med. Chem. Lett.* **19**, 2468–2473 (2009).

42. Cheatham, W. W. Peroxisome proliferator-activated receptor translational research and clinical experience. *Am. J. Clin. Nutr.* **91**, 262S–266S (2010).

43. Savkur, R. S. & Miller, A. R. Investigational ppar-γ agonists for the treatment of type 2 diabetes. *Expert. Opin. Investig. Drugs* **15**, 763–778 (2006).

44. Wang, Y. et al. Peroxisome proliferator-activated receptors as therapeutic target for cancer. *J. Cell. Mol. Med.* **28**, e17931 (2023).

45. Kodera, Y. et al. Ligand type-specific interactions of peroxisome proliferator-activated receptor γ with transcriptional coactivators. *J. Biol. Chem.* **275**, 33201–33204 (2000).

46. Berger, J. & Moller, D. E. The mechanisms of action of ppars. *Annu. Rev. Med.* **53**, 409–435 (2002).

47. Botta, M. et al. Ppar agonists and metabolic syndrome: an established role? *Int. J. Mol. Sci.* **19**, 1197 (2018).

48. Balakumar, P., Rose, M., Ganti, S. S., Krishan, P. & Singh, M. Ppar dual agonists: are they opening Pandora's box? *Pharmacol. Res.* **56**, 91–98 (2007).

49. Ma, L. et al. Synthesis and biological evaluation of novel 5-benzylidenethiazolidine-2, 4-dione derivatives for the treatment of inflammatory diseases. *J. Med. Chem.* **54**, 2060–2068 (2011).

50. Sime, M. et al. Discovery of gsk1997132b a novel centrally penetrant benzimidazole pparγ partial agonist. *Bioorg. Med. Chem. Lett.* **21**, 5568–5572 (2011).

51. Wang, L. et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).

52. Chen, W. et al. Enhancing hit discovery in virtual screening through absolute protein–ligand binding free-energy calculations. *J. Chem. Inf. Model.* **63**, 3171–3185 (2023).

53. Kuhn, B. et al. Structure-based design of indole propionic acids as novel pparα/γ co-agonists. *Bioorg. Med. Chem. Lett.* **16**, 4016–4020 (2006).

54. Volkov, M. et al. On the frustration to predict binding affinities from protein–ligand structures with deep neural networks. *J. Med. Chem.* **65**, 7946–7958 (2022).

55. Schindler, C. E. et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.* **60**, 5457–5474 (2020).

56. Isert, C., Atz, K., Riniker, S. & Schneider, G. Exploring protein-ligand binding affinity prediction with electron density-based geometric deep learning. *RSC Adv.* **14**, 4492–4502 (2024).

57. Corso, G., Jing, B., Barzilay, R., Jaakkola, T. et al. Diffdock: Diffusion steps, twists, and turns for molecular docking. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2210.01776 (2023).

58. Harris, C. et al. Benchmarking generated poses: How rational is structure-based drug design with generative models? Preprint at *arXiv* https://doi.org/10.48550/arXiv.2308.07413 (2023).

59. Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.*, **15**, 3130–3139 (2024).

60. Unke, O. T. et al. Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments. *Sci. Adv.* **10**, eadn4397 (2024).

61. Tosstorff, A., Cole, J. C., Taylor, R., Harris, S. F. & Kuhn, B. Identification of noncompetitive protein–ligand interactions for structural optimization. *J. Chem. Inf. Model.* **60**, 6595–6611 (2020).

62. Tosstorff, A., Cole, J. C., Bartelt, R. & Kuhn, B. Augmenting structure-based design with experimental protein-ligand interaction data: Molecular recognition, interactive visualization, and rescoring. *ChemMedChem* **16**, 3428–3438 (2021).

63. Tosstorff, A. et al. A high quality, industrial data set for binding affinity prediction: performance comparison in different early drug discovery scenarios. *J. Comput. Aided Mol. Des.* **36**, 753–765 (2022).

64. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).

65. Elton, D. C., Boukouvalas, Z., Fuge, M. D. & Chung, P. W. Deep learning for molecular design-a review of the state of the art. *Mol. Syst. Des. Eng.* **4**, 828–849 (2019).

66. Kutchukian, P. S. & Shakhnovich, E. I. De novo design: balancing novelty and confined chemical space. *Expert Opin. Drug Discov.* **5**, 789–812 (2010).

67. Renner, S. & Schneider, G. Scaffold-hopping potential of ligand-based similarity concepts. *ChemMedChem* **1**, 181–185 (2006).

68. Schneider, G. De novo design - hop(p)ing against hope. *Drug Discov. Today* **10**, e453–e460 (2013).

69. 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).

70. Grisoni, F. et al. Combining generative artificial intelligence and on-chip synthesis for de novo drug design. *Sci. Adv.* **7**, eabg3338 (2021).

71. Barau, C., Ghaleh, B., Berdeaux, A. & Morin, D. Cytochrome p450 and myocardial ischemia: potential pharmacological implication for cardioprotection. *Fundam. Clin. Pharmacol.* **29**, 1–9 (2015).

72. Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Sci. Rep.* **11**, 321 (2021).

73. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Int. Conf. Mach. Learn. (IMCL)* **70**, 1263–1272 (2017).

74. Unke, O. T. & Meuwly, M. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).

75. Isert, C., Kromann, J. C., Stiefl, N., Schneider, G. & Lewis, R. A. Machine learning for fast, quantum mechanics-based approximation of drug lipophilicity. *ACS Omega* **8**, 2046–2056 (2023).

76. Giles, C. L., Kuhn, G. M. & Williams, R. J. Dynamic recurrent neural networks: theory and applications. *IEEE Transactions on Neural Networks* **5**, 153–156 (1994).

77. Satorras, V. G., Hoogeboom, E. & Welling, M. E (n) equivariant graph neural networks. *Int. Conf. Mach. Learn. (IMCL)* **139**, 9323–9332 (2021).

78. Atz, K., Isert, C., Böcker, M. N., Jiménez-Luna, J. & Schneider, G. Δ-Quantum machine-learning for medicinal chemistry. *Phys. Chem. Chem. Phys.* **24**, 10775–10783 (2022).

79. Nippa, D. F. et al. Identifying opportunities for late-stage ch alkylation with high-throughput experimentation and in silico reaction screening. *Commun. Chem.* **6**, 256 (2023).

80. Kim, J.-H. et al. Hadamard product for low-rank bilinear pooling. *Proceedings of KIIS Spring Conference*, Vol. 26, 165–166 (2016).

81. Lamb, A. et al. Professor forcing: a new algorithm for training recurrent networks. *Adv. Neural. Inf. Process. Syst.* Vol. 29 (2016).

82. Vaswani, A. et al. Attention is all you need. *Adv. Neural. Inf. Process. Syst.* Vol. 30, 5998–6008 (2017).

83. Krige, D. G. A statistical approach to some basic mine valuation problems on the witwatersrand. *J. South. Afr. Inst. Min. Metall.* **52**, 119–139 (1951).

84. Rupp, M. & Schneider, G. Graph kernels for molecular similarity. *Mol. Inf.* **29**, 266–273 (2009).

85. Christensen, A. S., Bratholm, L. A., Faber, F. A. & von Lilienfeld, A. O. FCHL revisited: faster and more accurate quantum machine learning. *J. Chem. Phys.* **152**, 044107 (2020).

86. Heinen, S., Schwilk, M., von Rudorff, G. F. & von Lilienfeld, O. A. Machine learning the computational cost of quantum chemistry. *Mach. Learn. Sci. Technol.* **1**, 025002 (2020).

87. Lemm, D., von Rudorff, G. F. & von Lilienfeld, O. A. Machine learning based energy-free structure predictions of molecules, transition states, and solids. *Nat. Commun.* **12**, 4468 (2021).

88. Schneider, G., Neidhart, W., Giller, T. & Schmid, G. "scaffold-hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.* **38**, 2894–2896 (1999).

89. Bajusz, D., Rácz, A. & Héberger, K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 1–13 (2015).

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

91. Xu, Y.-J. & Johnson, M. Algorithm for naming molecular equivalence classes represented by labeled pseudographs. *J. Chem. Inf. Model* **41**, 181–185 (2001).

92. Wang, R., Fang, X., Lu, Y. & Wang, S. The PDBbind database: collection of binding affinities for protein- ligand complexes with known three-dimensional structures. *J. Med. Chem.* **47**, 2977–2980 (2004).

93. Isert, C., Atz, K., Jiménez-Luna, J. & Schneider, G. QMugs, quantum mechanical properties of drug-like molecules. *Sci. Data* **9**, 273 (2022).

94. Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discov.* **1**, 198–210 (2002).

95. Burggraaff, L. et al. Annotation of allosteric compounds to enhance bioactivity modeling for class a gpcrs. *J. Chem. Inf. Model.* **60**, 4664–4672 (2020).

96. Schrödinger. Schrödinger Release 2023-4: FEP+, (Schrödinger, LLC, 2023).

97. Rau, O. et al. Carnosic acid and carnosol, phenolic diterpene compounds of the labiate herbs rosemary and sage, are activators of the human peroxisome proliferator-activated receptor gamma. *Planta Med.* **72**, 881–887 (2006).

98. Brown, K. K. et al. A novel n-aryl tyrosine activator of peroxisome proliferator-activated receptor-gamma reverses the diabetic phenotype of the zucker diabetic fatty rat. *Diabetes* **48**, 1415–1424 (1999).

99. Brown, P. J. et al. Identification of a subtype selective human pparα agonist through parallel-array synthesis. *Bioorg. Med. Chem. Lett.* **11**, 1225–1227 (2001).

100. Young, P. W. et al. Identification of high-affinity binding sites for the insulin sensitizer rosiglitazone (brl-49653) in rodent and human adipocytes using a radioiodinated ligand for peroxisomal proliferator-activated receptor γ. *J. Pharmacol. Exp. Ther.* **284**, 751–759 (1998).

101. Sakamoto, J. et al. Activation of human peroxisome proliferator-activated receptor (ppar) subtypes by pioglitazone. *Biochem. Biophys. Res. Commun.* **278**, 704–711 (2000).

102. Berger, J. et al. Novel peroxisome proliferator-activated receptor (ppar) γ and pparδ ligands produce distinct biological effects. *J. Biol. Chem.* **274**, 6718–6725 (1999).

103. Willson, T. M., Brown, P. J., Sternbach, D. D. & Henke, B. R. The ppars: from orphan receptors to drug discovery. *J. Med. Chem.* **43**, 527–550 (2000).

104. Flesch, D. et al. Nonacidic farnesoid x receptor modulators. *J. Med. Chem.* **60**, 7199–7205 (2017).

105. Farol, L. T. & Hymes, K. B. Bexarotene: a clinical review. *Expert Rev. Anticancer Ther.* **4**, 180–188 (2004).

106. Willems, S. et al. Endogenous vitamin e metabolites mediate allosteric pparγ activation with unprecedented co-regulatory interactions. *Cell Chem. Biol.* **28**, 1489–1500 (2021).

107. Bendels, S. et al. Safety screening in early drug discovery: an optimized assay panel. *J. Pharm. Tox. Meth.* **99**, 106609 (2019).

108. Berman, H. M. et al. The protein data bank. *Acta Crystallogr. D* **58**, 899–907 (2002).

109. Atz, K. et al. Prospective de novo drug design with deep interactome learning. *figshare* https://doi.org/10.6084/m9.figshare.25234159 (2024).

110. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural. Inf. Process. Syst.* **32**, 8026–8037 (2019).

111. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch geometric. In *International Conference on Learning Representations,* Vol. 7 (2019).

112. Atz, K. et al. Prospective de novo drug design with deep interactome learning. *Zenodo* https://doi.org/10.5281/zenodo.10671327 (2024).

113. Davis, R. R. et al. Structural insights into jak2 inhibition by ruxolitinib, fedratinib, and derivatives thereof. *J. Med. Chem.* **64**, 2228–2241 (2021).

114. Tamaoki, T. et al. Staurosporine, a potent inhibitor of phospholipidca++ dependent protein kinase. *Biochem. Biophys. Rep.* **135**, 397–402 (1986).

Acknowledgements

Author contributions

Correspondence: Gisbert Schneider. K.A.: Conceptualization, data curation, formal analysis, investigation, methodology, initial software development, deep learning architecture design, model training and validation, visualization, writing—original draft. L.C.: Chemical synthesis, isolation, and characterization of compounds **1–3**, writing of the experimental procedure. C.I.: Conceptualization, formal analysis, investigation, methodology, software development, validation, visualization, writing—original draft. M.Hå.: Protein-ligand co-crystallization, writing—review and editing. D.F.: Protein-ligand co-crystallization, writing—review and editing M.Hi.: Methodology, software validation, visualization, writing—review and editing. D.F.N.: Conceptualization, methodology, writing—review and editing. M.I.: Methodology—validation of the ranking and scoring procedure. J.L.: Methodology—library creation with the RNN baseline model. C.C.G.S.: Methodology—validation of the ranking and scoring procedure. V.R.: Biological characterization, cytotoxicity assays, writing—review and editing. J.A.H.: Formal analysis, investigation, methodology, supervision, writing—review and editing. D.M.: Biological characterization, binding and functional assay, writing—review and editing. P.S.: Formal analysis, investigation, methodology, supervision, writing—review and editing. B.K.: ABFEP calculations, formal analysis, investigation, visualization, methodology, writing—review and editing. U.G.: Formal analysis, investigation, methodology, supervision, project administration, writing—original draft. G.S.: Formal analysis, investigation, methodology, supervision, funding acquisition, project administration, writing—original draft.

Competing interests

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-47613-w.

Correspondence and requests for materials should be addressed to Gisbert Schneider.

Peer review information *Nature Communications* thanks Pravir Kumar, Dong-qing Wei, Yunyun Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.