


RESEARCH ARTICLE

Chemical space exploration with Molpher: Generating and assessing a glucocorticoid receptor ligand library

M. Isabel Agea¹  | Ivan Čmelo¹ | Wim Dehaen^{1,2} | Ya Chen^{3,4} | Johannes Kirchmair^{3,4} | David Sedlák⁵ | Petr Bartůnek⁵ | Martin Šicho¹ | Daniel Svozil^{1,5}

¹Department of Informatics and Chemistry & CZ-OPENSREEN: National Infrastructure for Chemical Biology
Faculty of Chemical Technology
University of Chemistry and Technology
Prague, Czech Republic

²Department of Organic Chemistry
Faculty of Chemical Technology
University of Chemistry and Technology
Prague, Czech Republic

³Center for Bioinformatics (ZBH)
Department of Informatics
Faculty of Mathematics, Informatics and Natural Sciences
Universität Hamburg
Hamburg, Germany

⁴Division of Pharmaceutical Chemistry
Department of Pharmaceutical Sciences
Faculty of Life Sciences
University of Vienna
Vienna, Austria

⁵CZ-OPENSREEN: National Infrastructure for Chemical Biology
Institute of Molecular Genetics of the Czech Academy of Sciences
Prague, Czech Republic.

Correspondence

Daniel Svozil, Department of Informatics and Chemistry & CZ-OPENSREEN: National Infrastructure for Chemical Biology Faculty of Chemical Technology University of Chemistry and Technology Prague 16628, Czech Republic.
Email: svozild@vscht.cz

Funding information

Ministry of Education, Youth and Sports of the Czech Republic under the projects

Abstract

Computational exploration of chemical space is crucial in modern cheminformatics research for accelerating the discovery of new biologically active compounds. In this study, we present a detailed analysis of the chemical library of potential glucocorticoid receptor (GR) ligands generated by the molecular generator, Molpher. To generate the targeted GR library and construct the classification models, structures from the ChEMBL database as well as from the internal IMG library, which was experimentally screened for biological activity in the primary luciferase reporter cell assay, were utilized. The composition of the targeted GR ligand library was compared with a reference library that randomly samples chemical space. A random forest model was used to determine the biological activity of ligands, incorporating its applicability domain using conformal prediction. It was demonstrated that the GR library is significantly enriched with GR ligands compared to the random library. Furthermore, a prospective analysis demonstrated that Molpher successfully designed compounds, which were subsequently experimentally confirmed to be active on the GR. A collection of 34 potential new GR ligands was also identified. Moreover, an important contribution of this study is the establishment of a comprehensive workflow for evaluating computationally generated ligands, particularly those with potential activity against targets that are challenging to dock.

KEYWORDS

chemical space, de novo design, glucocorticoid receptor, molecular generation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Molecular Informatics* published by Wiley-VCH GmbH.

CESNET, Grant/Award Number: LM2015042; CERIT-Scientific Cloud, Grant/Award Number: LM2015085; Metacentrum e-INFRA CZ, Grant/Award Number: 90254; Ministry of Education, Youth and Sports of the Czech Republic, Grant/Award Number: LM2023052; RVO, Grant/Award Number: 68378050-KAV-NPUI; New Technologies for Translational Research in Pharmaceutical Sciences/NETPHARM, Grant/Award Number: CZ.02.01.01/00/22_008/0004607; China Scholarship Council, Grant/Award Number: 201606010345

1 | INTRODUCTION

Chemical space is understood as a multi-dimensional space that represents the functional and structural properties of molecules, along with their interrelationships [1]. Historically, this concept has been predominantly applied [2] to small organic compounds [3] emphasizing the vastness of this space [4], estimated at approximately 10^{33} compounds [4a] conforming to Lipinski's rule of five [5]. However, the application of the chemical space concept extends beyond organic molecules to encompass a variety of molecular types, demonstrating its practical utility in fields ranging from drug discovery [6] to natural products [7], material science [8], food [9] and flavor [10] chemicals, peptides [6c, 11] or metal complexes [12], thereby highlighting its diverse relevance across various scientific disciplines.

Two main computational approaches of *de novo* exploration of the chemical space of organic molecules have been well-established in cheminformatics research [13]: genetic algorithms (GAs) and deep generative modeling. GAs [14] take a combinatorial approach that involves mixing scaffolds, functional groups, and relevant fragments to obtain the desired molecular compound. GAs can be further classified as either structure-based or ligand-based methods, depending on the use of a 3D protein target structure [14a, 15]. Deep generative modeling [16] uses various architectures of deep neural networks to generate new compounds based on learned patterns.

A primary objective in generating a new virtual library is to design biologically active compounds. Therefore, the biological activity of these compounds must be predicted as reliably as possible. Quantitative Structure-Activity Relationship (QSAR) models are widely used for this purpose [17], but they can only reliably predict biological activity for new molecules structurally similar to those used in the training sets. The region of chemical space where reliable predictions can be made is defined by the model's applicability domain (AD) [18]. This

concept is particularly relevant for molecular generation, which aims to introduce new chemotypes. Accurately predicting the biological activity of new chemotypes is challenging because they often reside outside the AD of QSAR models. Therefore, it is essential to integrate AD into QSAR predictions for generated molecules to mitigate the occurrence of false positives.

In a previous study [19], we developed Molpher, a ligand-based GA approach for the systematic exploration of chemical space. This process involves creating a molecular path between start and target structures through the iterative application of molecular morphing operators (Figure 1). To ascertain the synthetic accessibility of the morphs [20], the Synthetic Accessibility Score (SAscore) [21] is integrated within the morphing process, facilitating the immediate rejection of synthetically non-feasible structures. Molpher was utilized to construct a library of hard-to-synthesize compounds [22], from which the SYBA synthetic accessibility score was subsequently derived [23]. Molpher is available as a Python open-source library *molpher-lib* [24] facilitating widespread use by the broader scientific community.

Given that the start and target structures are active at the same molecular target, their intermediate molecules encountered along the morphing path, known as morphs, should form a focused virtual library enriched in active compounds. In this study, the potential of Molpher to propose biologically active compounds was evaluated by generating two virtual libraries: a library of glucocorticoid receptor (GR) ligands (GRML library) and a library of random compounds (RML library).

The glucocorticoid receptor [25] is a compelling drug target [26] due to its pivotal role in modulating inflammatory and immune responses, making it crucial for treating a range of conditions, including autoimmune diseases, allergies, and certain cancers [27]. Its therapeutic relevance is underscored by its ability to influence gene expression, offering potential pathways for developing novel treatments that can precisely target

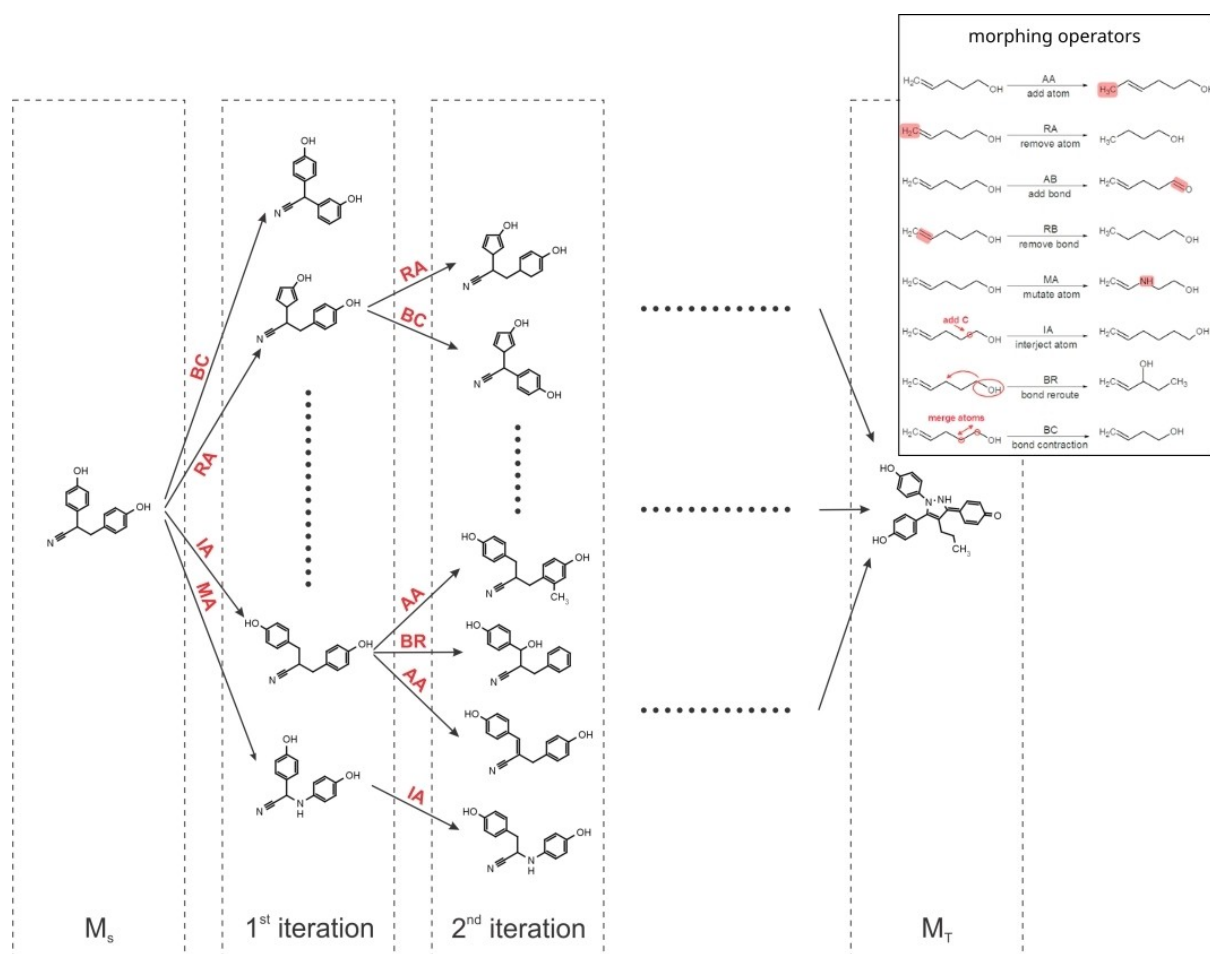


FIGURE 1 Principles of the molecular morphing approach. Molecular morphing generates a path in chemical space consisting of structures (referred to as morphs) lying between a start molecule M_S and a target molecule M_T . In each iteration, a set of morphs is generated by randomly applying morphing operators, that represent simple structural modifications (mutations), such as the addition or removal of an atom or bond, at molecules from the previous iteration. In each iteration, the morphs are accepted for the next iteration with the probability derived from their distance to the target molecule.

cellular processes involved in disease progression [28]. In this study, the GR was selected as the primary focus, aligning with our long-standing interest and extensive experience in steroid receptors, evidenced by the development, optimization, and establishment of specialized bioassays for their profiling [29]. In addition, the GR's structurally inherent flexibility, posing significant challenges in ligand docking [30], presents a unique opportunity to apply and validate our newly developed workflow.

The quality of both GRML and RML libraries was assessed by analyzing the structures and scaffolds of compounds predicted to be active by the QSAR model. Given its robust performance in various molecular machine learning tasks, random forest (RF) with compounds encoded by ECFP fingerprints was chosen for QSAR modeling [31]. The AD of the RF models was determined by Mondrian Conformal prediction (CP) [32]. Two QSAR models were developed: one using data from the

ChEMBL17 database and the other from the ChEMBL33 database, with both training datasets enriched by additional compounds from the internal library of the Institute of Molecular Genetics (IMG) in Prague. The GRML and RML libraries were compared with the input compounds used to generate them, as well as retrospectively with GR actives that appeared in the ChEMBL database between its versions 17 and 33. Additionally, a set of new potentially active compounds with novel chemical scaffolds was proposed that are, thus, worthy of subsequent experimental validation. The results demonstrate the usefulness of Molpher in generating virtual libraries enriched in active ligands. Moreover, an important contribution of this study is the establishment of a comprehensive workflow for evaluating computationally generated ligands, particularly those with potential activity against targets that are challenging to dock.

2 | DATA AND METHODS

2.1 | Data sources

In this study, the following two sources of bioactivity data were used (Figure 2):

- 1 ChEMBL [33], a publicly available, manually curated database of bioactive molecules, was utilized to select ligands of the human glucocorticoid receptor (ChEMBL ID: CHEMBL2034), encompassing both binding and functional assays. The decision to analyze these diverse assays collectively was made due to their complementary insights into compound interactions with the GR. The rationale for merging data from binding and functional assays is reinforced by the EC₅₀ values from the low nanomolar range for both dexamethasone and prednisolone across all assay types. The compounds with an EC₅₀ of 1 μM or lower were considered active. For a compound with several EC₅₀ values, their average and standard deviation were calculated. The compound was considered active if the average EC₅₀ was 1 μM or lower and the standard deviation was less than 0.5. If the standard deviation was higher than 0.5 μM, the compound was manually examined, and its inclusion or exclusion was determined based on the findings reported in the corresponding publications. Two versions of the ChEMBL database were utilized: version 17 (released in August 2013) and version 33 (released in May 2023).
- 2 IMG library of GR ligands was constructed by screening a large library of 24,511 compounds. This extensive collection consisted of a diverse set, bioactive set, and proprietary set.

The *diverse set*, obtained from ChemBridge [34], contained 9,845 drug-like compounds strictly conforming to Lipinski's rule-of-five [5a], thus exhibiting molecular weights of 500 Dalton or less, a maximum of 10 hydrogen bond acceptors, no more than 5 hydrogen bond donors, and a cLogP value of 5 or below. Additionally, the topological polar surface area (TPSA) [35] of these molecules was below 100. Compounds with undesirable chemical groups, including Michael acceptors, crown ethers and their analogs, disulfides, and epoxides, were deliberately excluded from this set. The *bioactive set* of 3,111 compounds was represented by carefully selected sets from established sources such as Prestwick Chemical Library [36], Sigma LOPAC Library of Pharmacologically Active Compounds [37], and the NIH Clinical Collection [38], containing pharmacologically active compounds, many of which were approved as drugs or have been tested in human clinical trials. The

proprietary set consisted of a diverse set of 11,555 compounds obtained from various academic institutions.

The deck of 24,511 compounds was tested in the primary luciferase reporter cell assay. The generation of the GR-LBD U2OS stable reporter cell line was described earlier [29]. Cells were grown for 48 h preceding the experiment in the phenol red-free DMEM supplemented with 4% C/D FBS and 4 mM glutamine. After 48 h, cells were trypsinized, counted, and seeded at a density of 10,000 cells/well in white opaque cell culture 384-well plates (Corning Inc., NY, USA). Compounds were diluted in DMSO and stored in the polypropylene 384-well plates. They were transferred to cells by JANUS Automated Workstation (PerkinElmer, Inc.) equipped with Pin tool (V&P Scientific, Inc., San Diego, CA, USA) at 1 μM final concentration. After incubation for 18 h, luciferase activity was determined using the One-Glo Luciferase Assay System according to the manufacturer's protocol. Luminescence was recorded by an EnVision (PerkinElmer, Inc.) plate reader using 1 s of signal integration, and data were analyzed using proprietary LIMS software ScreenX [39].

The B-score [40], a normalization technique employed in high throughput screening, was utilized to assess the activity of compounds in the primary screen. By adjusting for spatial and systematic errors across assay plates, this method aids in the identification of true biological activity, mitigating the influence of plate position and other systematic biases. A compound was considered active if a B-score of 5 or more was exhibited. The screening resulted in the IMG library of 154 GR ligands, comprising 26 bioactive compounds, 84 diverse compounds, and 45 proprietary compounds. Because the compounds were tested at only one concentration in the primary screen, no EC₅₀ values are available. Therefore, if multiple compounds represented a Bemis-Murcko (BM) scaffold, one was randomly included in the GRML input set.

2.2 | Project workflow

The scheme of the project workflow (Figure 2) is provided here, with individual steps elaborated in more depth in the following sections.

- 1 *Data generation*. Molpher [19] was used to create two virtual libraries. The Glucocorticoid Morphing Library (GRML) was generated from active compounds (termed *GR inputs*) with unique Bemis-Murcko (BM) scaffolds [41] from the ChEMBL17 and IMG libraries. As a baseline, the Random Morphing Library (RML) was generated from compounds with unique BM

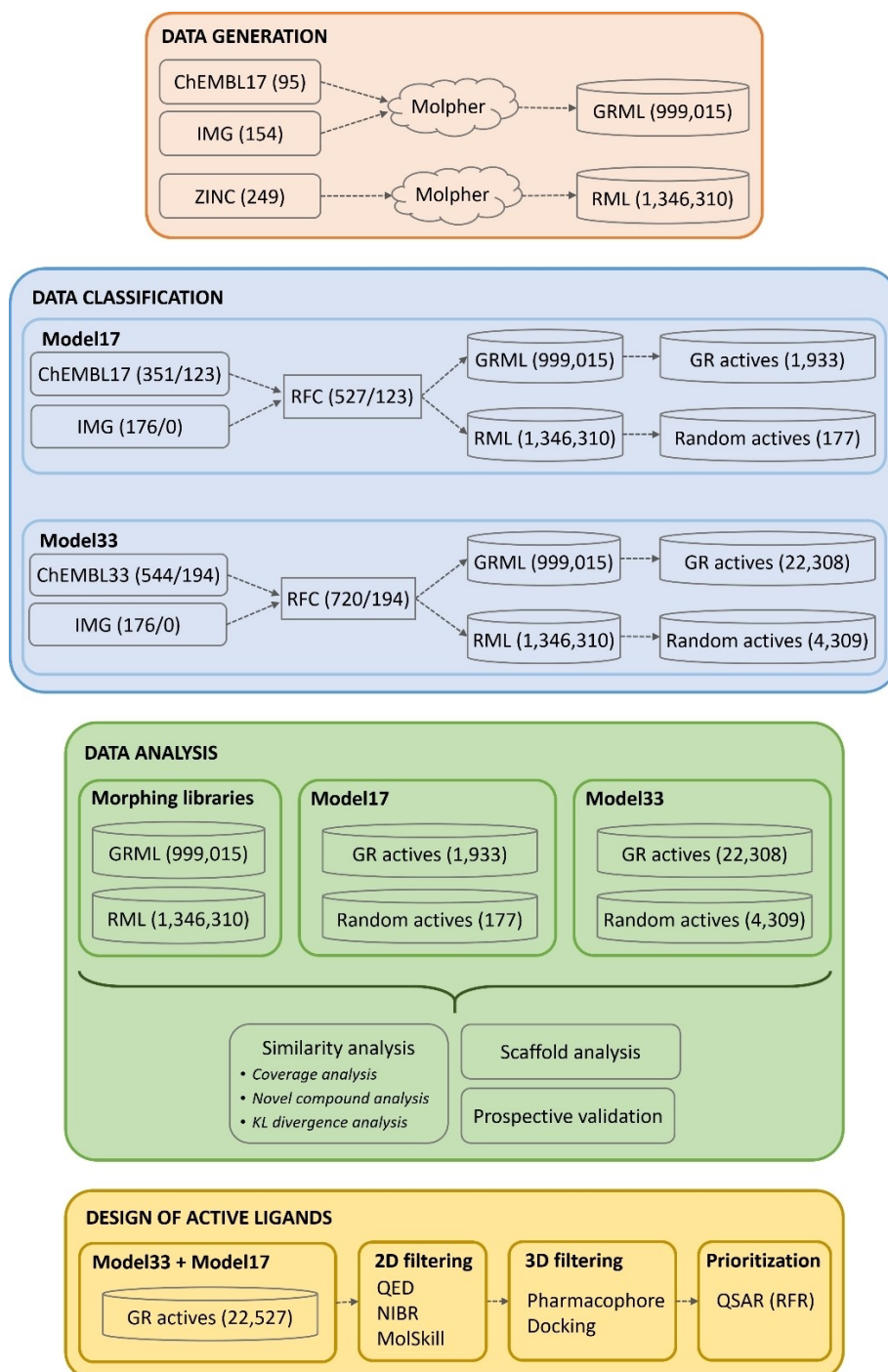


FIGURE 2 Project workflow. The glucocorticoid receptor (GR) morphing library (GRML) and the Random morphing library (RML) were generated from their respective input libraries (termed GR inputs and Random inputs) using the Molpher algorithm. A random forest classifier, trained on two distinct datasets (the numbers in brackets indicate the number of GR actives/inactives), was employed to predict active compounds in both GRML (termed GR actives) and RML (termed Random actives) libraries. Subsequent analysis of both morphing libraries, as well as the GR and Random actives, was conducted through similarity analysis, scaffold analysis, and prospective validation. For the design of active ligands, the number of GR actives was narrowed down using 2D (structural) filtering, followed by pharmacophore modeling and molecular docking. The final set of potential GR ligands was prioritized using a random forest regression QSAR model to predict their biological activity (pEC50).

scaffolds (termed *Random inputs*) randomly selected from the ZINC database [42]. Input compounds were

encoded as 1,024-bit Morgan fingerprints with a radius of 2 (i.e., ECFP4 fingerprints [43]).

- 2 Data *classification*. Virtual compounds from both GRML and RML libraries were classified as active or inactive using a random forest classification (RFC) model, with its applicability domain (AD) [44] evaluated by conformal prediction (CP) [32b]. Compounds were encoded as 1,024-bit Morgan fingerprints with radius 2 [43]. For RFC training and testing, active ligands from the ChEMBL were merged with the active ligands from the IMG library. By merging two datasets, more data for training was obtained enriching, thus, the number and diversity of compounds. Despite mixing EC50 threshold from ChEMBL and B-score threshold from IMG libraries, the critical aspect is the maintained clear distinction between active and inactive compounds, ensuring the reliability of the classifiers.
- 3 Compounds predicted as active in the GRML library are termed *GR actives*, and compounds predicted as active in the RML library are termed *Random actives*. Two RFC models were constructed, differing only in training data:
- 4 *Model17*, trained on 527 active and 123 inactive GR ligands obtained from the ChEMBL17 and IMG library (Figure 2), was deployed for a retrospective validation. It was applied to the GRML library, and the compounds predicted to be active were compared to experimentally validated actives that were added to the ChEMBL between its versions 17 and 33. Restricting the model to an older ChEMBL dataset enforced an actual time-split scenario, representing a more realistic use case: training the model on data available up to a certain point and testing its predictive abilities against data that become available later.
- 5 *Model33*, trained on 720 active and 194 inactive GR ligands obtained from the ChEMBL33 and IMG library (Figure 2), was established to evaluate the impact of a more comprehensive training set on predictive accuracy for in both the GRML and RML libraries and it provides insights into the predictive enhancements achievable through a more extensive and more diverse dataset.
- 6 Data *analysis*. The quality of the GRML and RML libraries was evaluated by analyzing the GR and Random actives considering both their complete structures and BM scaffolds.
- 7 Design of *active ligands*. A computational workflow was developed to prioritize the generated GR actives for further use. The workflow consists of several filtering steps, both 2D (structural and drug-like filters) and 3D (pharmacophore modeling, molecular docking), followed by compound ranking based on pEC50 predicted by a random forest regression QSAR model.

2.3 | Data generation

The generation of the GRML and RML virtual libraries involved the following steps:

- 1 The first step was the construction of the GR and Random inputs, from which both morphing libraries were generated. The GRML library was generated from input compounds with unique BM scaffolds [41] biologically active on human GR. These compounds were obtained from the ChEMBL17 and IMG libraries. Their BM scaffolds were constructed, and one compound was included in the GR inputs for each scaffold. Whenever multiple compounds shared the same BM scaffold, the one with the best biological activity was selected to represent the scaffold. The RML library was generated from compounds randomly selected from the ZINC database [42] with unique BM scaffolds not present in the GR inputs. The number of compounds in the Random inputs matched those in the GR inputs.
- 2 In each input set, all possible compound pairs were formed. For each compound pair, two morphing paths were generated: the first path started from compound #1 and ended with compound #2, while the second path ran in the opposite direction, beginning with compound #2 and ending with compound #1.
- 3 Only morphs dissimilar to their starting or target compounds were included in the morphing library. Two compounds were considered different if their Tanimoto coefficient T_c [45], calculated for compounds encoded as 1,024-bit Morgan fingerprints with a radius of 2 [43], was equal to or less than 0.7.

2.4 | Data classification

Molecular generation focuses on the design of new compounds, many of which fall outside the applicability domain (AD) of QSAR models [18]. Consequently, the biological activity of such compounds can be challenging to predict, emphasizing the importance of considering the AD of any used model. This study assessed the AD of the binary classification RFC model using Conformal Prediction (CP) [32b]. In binary classification, compounds are labeled either as active or inactive. However, CP modifies this approach by predicting regions containing all possible label combinations. In the binary CP classification, the following four possible regions are formed: *active* (contains only the active label), *inactive* (contains only the inactive label), *both* (contains both active and inactive labels), or *empty* (contains no label). The *active* and *inactive* regions indicate the parts of chemical space where the model is confident in its predictions at a user-

defined confidence level. If a compound is assigned to both regions, the model is less confident about the prediction, indicating that the compound shares similarities with both active and inactive classes. In such cases, additional information is required to make a single class prediction [32e]. If a compound falls into the *empty* region, the model is unsure about the prediction, suggesting that the compound is dissimilar to active and inactive classes. These compounds are considered to lie outside the model's AD.

Two forms of CP exist: transductive CP (TCP) [46] and inductive CP (ICP) [32b,47]. TCP is an online approach where the model is retrained for each new object [48]. For QSAR modeling, the offline ICP approach is more suitable. In the ICP, data are divided into three sets, each containing both active and inactive compounds (Figure 3):

1 *Proper training set* (PTrS). The model is trained on this data.

2 *Calibration set* (CS). Each compound in this set is predicted using the trained model and further characterized by its nonconformity score α , which quantifies the novelty of the compound compared to the training data. In the Mondrian Conformal Prediction (MCP) [32e,49] used in this study, the nonconformity scores are generated for each class separately, making the MCP suitable for imbalanced data sets.

3 *Test set* (TS) is used to evaluate the model's predictive performance. For each test compound, its p-value is calculated. The p-value represents the proportion of compounds in the calibration set with nonconformity scores equal to or lower than the test compound's α . In the MCP, two p-values are obtained and compared with the user-defined significance level ϵ . Note that the concept of p-value in the CP is not equivalent to the one used in inferential statistics.

Under the CP framework, prediction errors are straightforward to interpret. For instance, with a

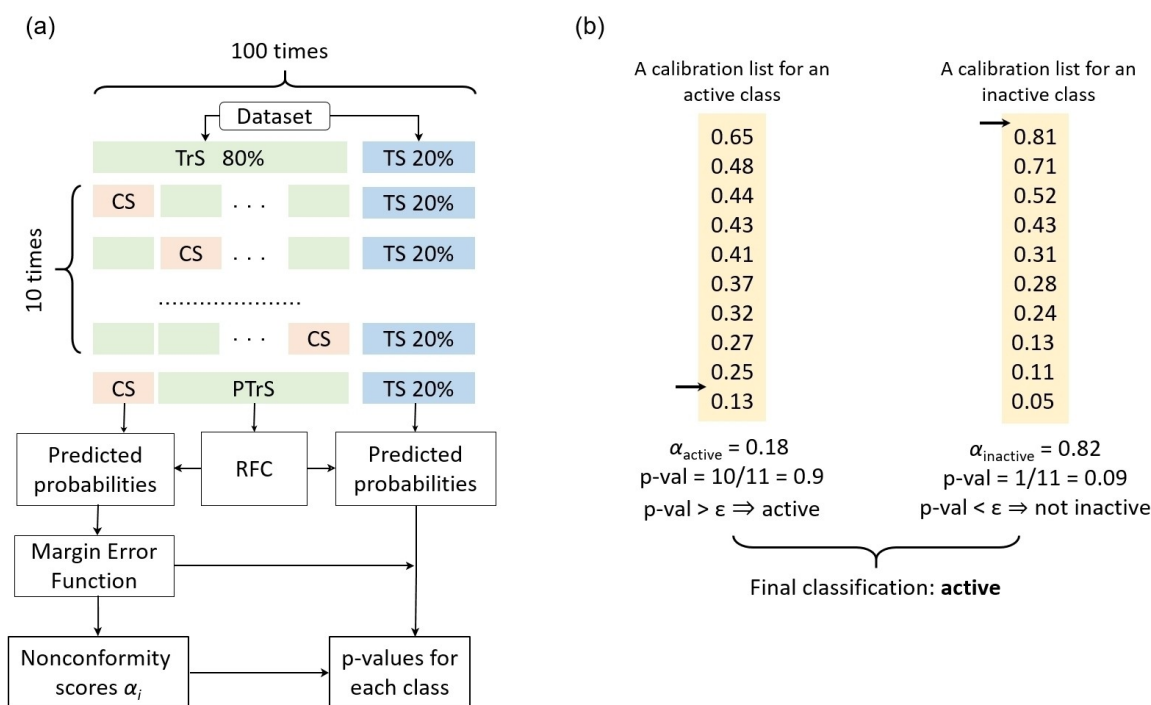


FIGURE 3 Conformal prediction. (a) The input dataset was partitioned into a training set (TrS) and a test set (TS). The training set was subsequently divided into 10 folds. Each fold was used as the calibration set (CS), while the remaining nine folds constituted the proper training set (PTrS). The RFC model was trained using the PTrS and then employed to predict instances of the CS, producing their class membership probabilities. Using these probabilities, the margin error function yielded the nonconformity scores α_i . The nonconformity scores α_i of the CS compounds then formed two calibration lists, one for each class. (b) An illustrative example of the computation of p-values for 'new instances' based on predicted probabilities from the RF model. The nonconformity score α of the new instance for the active class equals 0.18, and for the inactive class equals 0.82. Arrows indicate the placement of the nonconformity scores of the new instance within both active and inactive calibration lists. The p-value for a class is computed as the proportion of instances in the calibration set that have a higher nonconformity score than the new instance when it's assumed to belong to that class. A high p-value for a class means the new instance fits well with that class, while a low p-value suggests it's an unusual member. The new instance is assigned to the class with a p-value higher than a predefined significance level ϵ ($\epsilon = 0.1$ in this example). If none of the classes meets this threshold, the prediction can be labeled as empty; if both classes meet this threshold, the prediction is marked as both.

significance level $\epsilon = 0.1$, the true label will fall within the predicted confidence regions in at least 90% of cases, where 90% is the confidence level defined as $1 - 1\epsilon$. This property, known as validity, is underpinned by the mathematical principles of the CP [32a]. Another critical aspect of the CP is efficiency [50], the proportion of correct single-class predictions. A model is more efficient if fewer non-informative predictions (i.e., *both* or *empty* regions) are obtained. Both the validity and efficiency of models are assessed with test set data.

In this study, a variant of the ICP, known as the Cross-Conformal Prediction (CCP) [51], was used. The dataset was randomly split into the training (TrS) and test (TS) set in an 80:20 ratio preserving the active/inactive ratio (Figure 3). In each iteration, the TrS was divided into ten non-overlapping folds. Each fold was then used as the calibration set (CS), with the remaining 9/10 of data serving as the proper training set (PTrS). In this way, it was ensured that all labeled data were used for training. Two calibration nonconformity score lists, one for each label (*active* and *inactive*), were calculated for every fold. For each TS compound in every fold, two p-values, corresponding to the position of the TS compound within the *active* and *inactive* nonconformity score lists, were calculated. Two final compound's p-values were obtained as the average across all ten folds.

Nonconformity scores α_i were calculated using the margin error function [52] defined as

$$\alpha_i = 0.5 - \frac{\widehat{P}(y_i|x_i) - \max_{y \neq y_i} \widehat{P}(y|x_i)}{2} \quad (1)$$

where $\widehat{P}(y_i|x_i)$ is the probability obtained by the underlying RFC model for the *active* label and $\max_{y \neq y_i} \widehat{P}(y|x_i)$ is the maximum probability obtained for the *inactive* label. The CCP was implemented using the nonconformist package, version 2.1.0 [53].

Morphs in the GRML and RML libraries were classified as active or inactive using an RFC QSAR model. Each compound was encoded as a 1,024-bit Morgan fingerprint with a radius of 2 (i.e., ECFP4 fingerprint) [43]. The following hyperparameters were optimized via Bayesian optimization: the number of trees in the forest (ranging from 10 to 500), the number of features (ranging from 10 to 1,024), the depth of the tree (ranging from 1 to 200) and the split criterion (gini or entropy). All other hyperparameters were set to their implicit values. The objective function for the optimization was model accuracy evaluated on the test set (Acc_{test}).

$$Acc_{test} = \frac{\text{number of correct predictions}}{\text{total number of predictions}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. The accuracy was preferred over balanced accuracy for hyperparameter optimization because the dataset's imbalance was mild [54] (approx. 1:4 in favor of the active class) and the primary interest was in correctly classifying the majority (active) class. The RFC is quite robust to such mild imbalance due to its ensemble nature and the randomness in selecting samples and features. Moreover, models optimized for accuracy surprisingly yielded higher balanced accuracy scores (data not shown), suggesting that optimizing for accuracy aligned better with the project's objectives.

To ensure a reliable model error calculation and to minimize random effects, the stratified training/test set split was performed 100 times for each model. The RFC was implemented in Scikit-learn version 1.3.0 [55]. The hyperparameters were optimized using the hyperopt library, version 0.2.7 [56]. All cheminformatics calculations were conducted using the RDKit toolkit version 2023.3.2 [57].

2.5 | Data analysis

The quality of the GRML library was evaluated using similarity, scaffold, and prospective validation analyses (Table 1).

2.5.1 | Similarity analysis

The similarity analysis evaluated the novelty of the GRML library by comparing the structures of its compounds to those in the RML library. Three levels of the

TABLE 1 The number of active morphs in the GRML and RML libraries predicted by RFC Model17 and Model33. A confidence level of 90% was used. Compounds predicted as active in at least one split were considered active. The GRML and RML library comprised 999,015 compounds and 1,346,310 compounds, respectively. Compound counts are provided in parentheses.

	GRML	RML
Model17	24.68% (246,642)	10.19% (137,216)
Model33	43.27% (432,296)	24.49% (329,763)

analysis were used, addressing various aspects of similarity dependencies:

- the coverage analysis.
- the novel compound analysis.
- the KL divergence analysis.

The *coverage analysis* evaluated chemical space covered by the GRML and RML libraries by comparing the structures of morphs with the GR and Random inputs. Tc was calculated between each morph predicted to be active and its most similar compound from the input set. The compounds were encoded as 1,024-bit Morgan fingerprints with a radius of 2. The highest Tc distributions of the GR actives, Random actives, GR inputs, and Random inputs were compared.

In the *novel compound analysis*, actives predicted by Model17 and Model33 were analyzed in relation to the GR inputs. A new dataset termed *Novel set*, consisting of active compounds added to the ChEMBL database between its versions 17 and 33, was created. The Novel set was divided into three subsets based on the Tc between each active in the Novel set and its most similar active in the ChEMBL17: low similarity subset ($T_c < 0.4$), medium similarity subset ($0.4 \leq T_c \leq 0.7$), and high similarity subset ($T_c > 0.7$). For each of these subsets, the highest Tanimoto coefficients between the subset actives and GR actives predicted by either Model17 or Model33 were used to establish Tc distributions (Figure 4).

The assumption is that the active compounds predicted by Model17 will be closer to the high similarity subset than the low similarity one and that the active compounds predicted by Model33 will be roughly equally distant from the low similarity, medium similarity, and high similarity subsets. Furthermore, Model33, containing the Novel set as the part of its training set, should have a higher average Tc than Model17 for every subset.

The *KL divergence analysis* aimed to determine whether Molpher generates ligands from a subspace known to be enriched with active ligands. This was achieved by comparing the Novel set, containing newly characterized active ligands not used as molecular morphing inputs, with the GRML and RML morphing libraries. The assumption was that the distribution of the highest Tc between the Novel set and the RML or GRML libraries would be shifted towards higher Tc values for the GRML compared to the RML library. The difference between these two distributions was quantified using Kullback-Leibler (KL) divergence [58], an information theory-based measure commonly used to assess how one probability distribution diverges from a second (i.e., reference) probability distribution. To circumvent zero

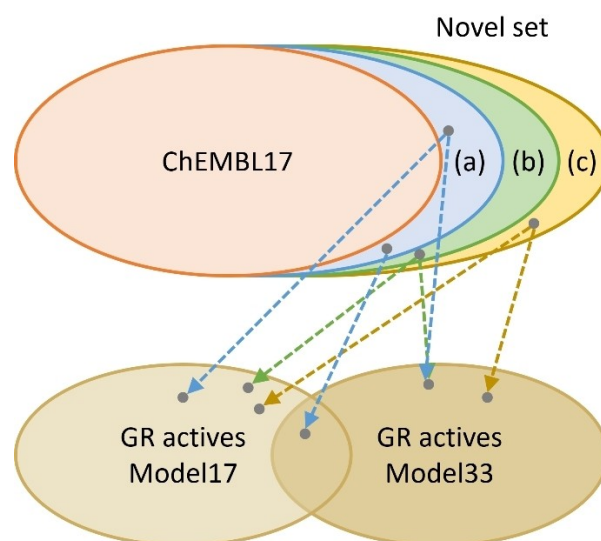


FIGURE 4 Novel compound analysis. The Novel set was formed by the GR actives added to the ChEMBL database between its versions 17 and 33. Based on the highest Tc to the ChEMBL17 GR actives, the Novel set was divided into three subsets: (a) low similarity subset ($T_c < 0.4$), (b) medium similarity subset ($0.4 \leq T_c \leq 0.7$), and (c) high similarity subset ($T_c > 0.7$). For illustrative purposes, a few examples of the highest Tcs between Novel subset compounds and the GR actives predicted by either Model17 or Model33 are depicted by dashed lines. The distributions of these Tcs were constructed for each Novel subset.

values in KL divergence calculations, Laplace smoothing [59] was employed. KL divergence was calculated using the `entropy()` function from the SciPy library version 1.11.1. [60], with the logarithm of base 2.

2.5.2 | Scaffold analysis

The scaffold analysis aimed to assess the diversity of chemical scaffolds generated by Molpher. This was achieved by comparing BM scaffolds [41] from the GR and Random input sets, GRML and RML libraries, and GR and Random active sets.

2.5.3 | Prospective validation

The goal of the prospective validation was to identify novel BM scaffolds that were rediscovered through the morphing process and were annotated as actives in the ChEMBL33 database yet were not used as inputs for the morphing. For this purpose, the BM scaffolds of Model17 and Model33 inputs were juxtaposed with those of GR and Random actives, and the results were analyzed.

2.5.4 | Design of active ligands

Although the active morphs were classified as such with a high degree of confidence, there was a need to further prioritize them based on their structure and other properties important for ligand-GR binding. Thus, a multi-step virtual screening workflow was established (Figure 5). Its steps are described in more detail in the following sections.

- 1 In the *2D filtering* step, compounds with undesirable substructures and compounds that could not be considered drug-like were removed.
- 2 In the pharmacophore *modeling* step, a pharmacophore model was constructed, and only compounds that conformed to the pharmacophore model were passed to the subsequent step. This pharmacophore filtering ensured that any potential ligand would likely bind to the GR through the established binding mode. While this method might reject compounds exhibiting novel binding modes, its inclusion was justified by the increased confidence in the subsequent active compound identification and prioritization.
- 3 In the *molecular docking* step, compounds that passed the previous step were docked into the ligand-binding domain of the GR. Once their optimum poses were established, the pharmacophore model developed in the last step was applied once more. The replacement of the pharmacophore modeling before the docking in step 2. offers significant advantages. It establishes a good initial pose placement, thus accelerating the docking procedure considerably, as it obviates the necessity for conformer generation and advanced pose placement strategies. This acceleration is particularly beneficial in systems such as the GR, which exhibit suboptimal docking performance in terms of AUC [30].
- 4 In the *random forest regression* step, compounds were ranked by their pEC50 predicted by a random forest regression (RFR) model. This model's training set comprised 89 ChEMBL33 compounds that conformed to the constructed pharmacophore model.

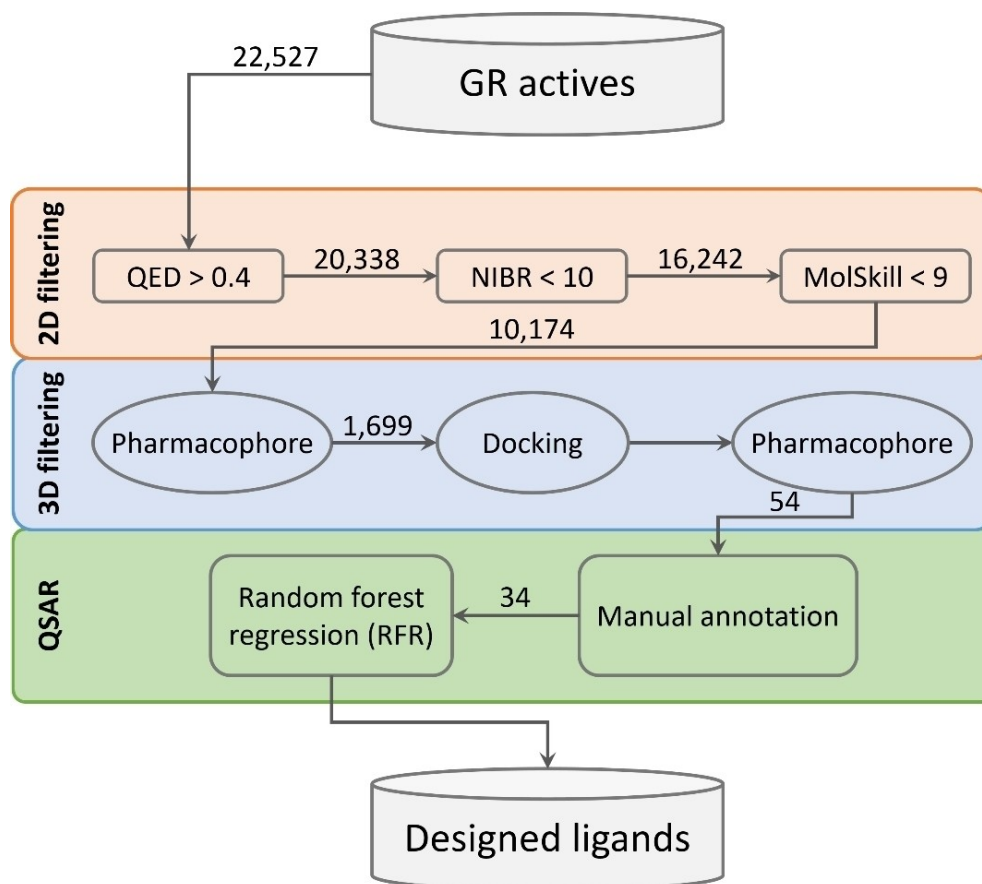


FIGURE 5 The workflow of the design of active ligands. The workflow consists of several consecutive steps during which proposed ligands are gradually filtered based on their drug-like, structural, and 3D properties. Subsequently, ligands are ranked based on their biological activity predicted by a random forest regression model and manually annotated by an experienced organic chemist.

2.5.5 | 2D filtering

Compounds with undesirable substructures were excluded using NIBR substructure filters [61] designed by Novartis to validate and prioritize potential active compounds. In addition, the use of NIBR filters is recommended before the application of MolSkill [62]. Compound drug-likeness was assessed by two methods: QED [63] and MolSkill [62]. QED is a well-established method based on molecular properties and structural descriptors correlating with favorable pharmacokinetics and safety profiles. MolSkill is a recently developed approach based on utilizing artificial intelligence learning-to-rank techniques on feedback obtained from 35 chemists at Novartis over several months. All filters were used with the following thresholds recommended by their respective authors: NIBR severity score < 10, QED > 0.4 and MolSkill < 9.

2.5.6 | Pharmacophore modeling

The GR pharmacophore model was developed in the Pharmacophore Query Editor from the Molecular Operating Environment 2019 (MOE) [64]. Though multiple pharmacophore models have been developed for the GR [65], consensus among them remains elusive. Thus, a new pharmacophore model was constructed to utilize the latest published data and our internal GR agonist information. To ensure the generality of the pharmacophore model, PDB structures were selected based on the 2D structural diversity of GR agonists. Five distinct agonist structures were used: one steroidal agonist (dexamethasone, PDB ID: 1P93 [66]) and four non-steroidal agonists each representing a unique structural class (pyrazole-based agonist – PDB ID: 3E7C [67], indazole-based agonist – PDB ID: 4CSJ [68], pyrimidine-based agonist – PDB ID: 6EL7 [69], and next-gen indazole-based agonist – PDB ID: 7PRX [70]). Using the Alignment/Superpose tool in the MOE 2019, 3D protein-agonist complexes were superimposed, and the aligned coordinates of the ligands were retained for subsequent steps without alterations. Conformers for all input molecules were generated with the Stochastic Conformation Search method in the MOE 2019 using the MMFF94x force field and employing the following implicit settings: a rejection limit of 200, an iteration limit of 200, an energy window of 100, and a conformation limit of 200.

The pharmacophore model was validated using the GR data from the DUD–E benchmark set [30], which contains 468 active molecules and 15,150 inactive (decoy) compounds with similar physicochemical properties to the active ones. Pharmacophore model quality was

evaluated using the positive likelihood ratio ($LR+$) and model sensitivity (Sen). $LR+$ measures how much more likely a pharmacophore model is to return a positive result for an active compound than for an inactive compound. The higher the $LR+$, the better the pharmacophore model is. $LR+$ is defined as

$$LR+ = \frac{Sen}{1 - Spe} = \frac{TPR}{1 - TNR} \quad (3)$$

where specificity Spe (True Negative Rate, TNR) is the proportion of known inactive compounds correctly identified as not fitting the pharmacophore model and sensitivity Sen (True Positive Rate, TPR) is the proportion of known active compounds correctly identified by the pharmacophore model

$$Sen = \frac{TP}{TP + FN} \quad (4)$$

A high sensitivity is essential to ensure the pharmacophore model does not falsely reject active compounds. The combined use of the $LR+$ and Sen offers a comprehensive perspective on the performance of pharmacophore models. There is a trade-off between these metrics: while a model with high Sen may detect most of the actives, it may also produce more false positives, which $LR+$ helps to account for. Partial matching was employed in the present work to achieve an optimal balance between $LR+$ and Sen . Partial matching refers to the concept that a molecule needs to match only a subset of all the features of a pharmacophore model to be considered a hit. This approach provides flexibility, recognizing that not all features identified in a pharmacophore model may be required for activity. It can also help to identify molecules with novel structures or scaffolds that might not precisely match the used pharmacophore.

2.5.7 | Molecular docking

Using MOE 2019 [64], each conformation that passed the previous step was docked within the ligand binding domain of the crystal structure of dexamethasone bound to the GR (PDB ID: 1P93). [66] The Triangle Matcher method and the London dG scoring function were used for placement, with five poses retained. Refinement was conducted using the Rigid Receptor method and the GBVI/WSA dG scoring, resulting in a single pose. For each optimal pose, the previously developed pharmacophore model was reapplied. In this step, absolute

positions were used, with molecules assumed to be pre-aligned and stationary in space.

2.5.8 | Random forest regression

Since the morphs were classified as active or inactive with a high (90%) confidence level, their pEC₅₀ was subsequently estimated using the RFR model without conformal prediction. To build the RFR model, the following set consisting of 89 compounds was constructed:

- 1 GR actives from the ChEMBL33 database. 558 ligands with EC₅₀ values of 2 μM or less were included in the data set. This limit was chosen because experimental groups typically report ligands as inactive above this concentration without providing their respective EC₅₀ values. If a compound had multiple EC₅₀ values, its standard deviation was calculated. If the standard deviation was less than or equal to 0.5 μM and the average was less than 2 μM, the average was used as the compound's EC₅₀. Otherwise, the publications in which the compound activity measurements were reported were checked manually, and the decision of whether or not to include the compound in the data set was made based on the obtained insights.
- 2 The previously developed pharmacophore model was applied to the compounds assembled in the previous step, filtering out 469 and retaining 89 ligands.

To provide an unbiased and robust evaluation of RFR performance, 100 random training-test splits with a ratio of 80:20 were performed. Similarly to RFC, each input compound was encoded as a 1,024-bit Morgan fingerprint with a radius of 2, and the following hyperparameters were optimized: the number of trees in the forest (ranging from 10 to 500), the number of features (ranging from 10 to 1,024), the depth of the tree (ranging from 1 to 200) and the split criterion (squared error, absolute error, friedman mse or poisson). All other hyperparameters were set to their implicit values. The objective function for RFR hyperparameter optimization was R_{test}^2 defined as

$$R_{test}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where y_i is the true test set value, \hat{y}_i is the predicted test set value, and \bar{y} is the mean of true test set values. The RFR was implemented in the Scikit-learn version 1.3.0 [55], and the hyperparameters were optimized via the Bayesian optimization using the hyperopt library,

version 0.2.7/[56]. All cheminformatics calculations were conducted using the RDKit toolkit version 2023.3.2 [57].

3 | RESULTS AND DISCUSSION

3.1 | Data generation

The GR input set consists of 249 active compounds with unique BM scaffolds originating from the ChEMBL17 [33a] (95 ligands) and IMG (154 ligands) libraries. Similarly, the Random input set comprises 249 compounds randomly selected from the ZINC20 database [42]. From the GR and Random input sets, the GRML virtual library consisting of 999,015 compounds and the RML virtual library of 1,346,310 compounds (Figure 2) were generated by molecular morphing.

3.2 | Data classification

The best hyperparameters of Model17 were 242 trees, 327 features, a tree depth of 122, and Gini impurity as the split criterion. Similarly, the optimum hyperparameters of Model33 were 178 trees, 325 features, a tree depth of 198, and entropy as the split criterion. The AD of optimized models was considered by utilizing the cross-conformal prediction. Though in typical CP classification problems, an 80% confidence level is conventionally adopted [71], in this study, a 90% confidence level (i.e., a significance level $\epsilon=0.1$) was employed to reduce the potential for false positives and to provide a more conservative estimate, thus reinforcing the reliability and robustness of the model's predictions. A CP model is considered valid if its validity is higher than its confidence level [32b]. For the chosen confidence level of 90%, Model17 was valid in 99 out of 100 splits, with an average validity of 0.97. Similarly, Model33 was valid in 92 out of 100 splits with an average validity of 0.92 (Figure 6). These high validity scores underscore the robustness and reliability of both Model17 and Model33. However, the trade-off between validity and efficiency [32b,e,72] resulted in relatively low average efficiency for both models: 0.11 and 0.24 for Model17 and Model33, respectively.

A comparison of the RML and GRML libraries revealed significant differences in the number of predicted actives, depending on the model (Table 1). When Model17 was used, the GRML library contained approximately 2.5 times more predicted actives than the RML library. In contrast, when predictions were made with Model33, the GRML library had about 1.8 times more predicted actives than the RML library. These findings

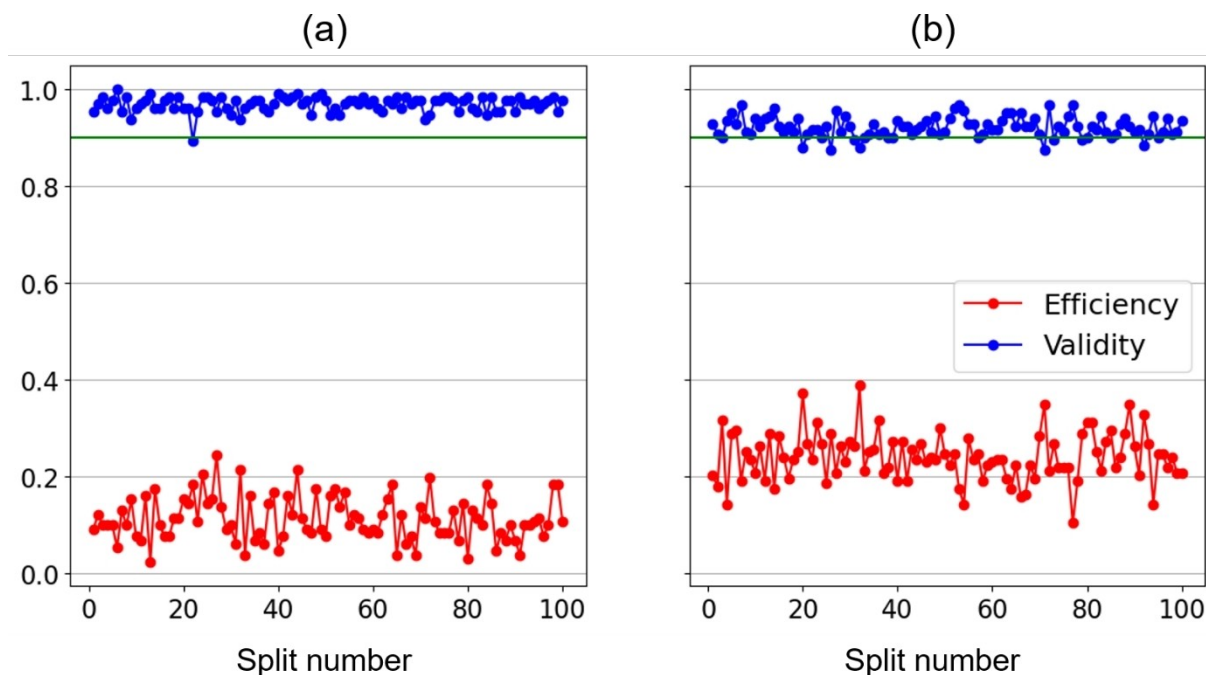


FIGURE 6 The efficiency and validity of Model17 (a) and Model33 (b) at the 90% confidence level (green, horizontal line) in 100 splits. A model is considered valid if its validity is higher than its confidence level.

suggest that the GRML library is enriched with more predicted actives than the RML library. Furthermore, Model33, trained on a larger dataset, inherently encompasses a greater scope of information than Model17. As a result, the GRML library showed a 1.7-fold increase in predicted actives when Model33 was used compared to Model17. Similarly, the RML library showed a 2.5-fold increase in predicted actives when classified with Model33. These results underscore the significant influence of the complexity of training data on a model's prediction accuracy.

The number of splits is another factor that influences the confidence of predictions. The more splits in which a compound is predicted as active, the higher the confidence associated with that prediction. For example, if a compound is predicted as active in all 100 splits, this indicates the highest confidence level. However, setting such a stringent criterion can be overly restrictive. The interplay between the number of splits and the confidence in predicted activity displays an inverse relationship with the number of compounds identified as active. In essence, as the confidence, assessed by the number of splits, in prediction increases, the number of predicted actives tends to decrease. An analysis was conducted to find the optimal balance between these two effects (Table 2). A threshold of a minimum of 50 splits was determined as the optimal criterion, ensuring both a sufficient number of predicted actives and a reasonable degree of confidence in their predicted activity. At this threshold, approximately 12 times more actives were

predicted in the GRML library with Model33 than with Model17. Similarly, the RML library showed a 10-fold increase in the number of actives predicted with Model33. These findings highlight the importance of considering the number of splits when setting a threshold for the confidence of predictions.

The GRML and RML morphs predicted by both Model17 and Model33 as active at a confidence level of 90% and in at least 50 splits are available as Supporting Information.

3.3 | Data analysis

3.3.1 | Similarity analysis

3.3.1.1 | Coverage analysis

The highest Tc distributions between the active and input sets for both Model17 and Model33 are shown in Figure 7. Based on the Tc, each compound was assigned to one of three subsets: low similarity subset ($T_c < 0.4$), medium similarity subset ($0.4 \leq T_c \leq 0.7$) and high similarity subset ($T_c > 0.7$).

The distribution of similarities between the GR actives and GR inputs is right-skewed and does not contain highly similar compounds (Figure 7a and Figure 7b). Most GR actives can be found in the low similarity subset (Table 3). Specifically, for Model17, 68% (1,318 out of 1,933) of the GR actives have low similarity to the GR inputs, and for Model33, 76% (17,044 out of 22,308) of

TABLE 2 The number of morphs predicted as active across various splits. Actives in both the GRML and RML libraries were predicted by RFC Model17 and Model33 using a 90% confidence level. The optimum level of 50 splits in bold. Compound counts are provided in parentheses.

# splits	–GRML		–RML	
	Model17	Model33	Model17	Model33
100	0% (0)	0.01% (81)	0% (0)	0.0002% (3)
90	0.003% (28)	0.17% (1,665)	0% (1)	0.007% (105)
80	0.012% (127)	0.42% (4,273)	0.0005% (7)	0.03% (411)
70	0.04% (382)	0.83% (7,813)	0.002% (28)	0.08% (1,082)
60	0.09% (939)	1.39% (13,951)	0.006% (87)	0.17% (2,293)
50	0.19% (1,933)	2.23% (22,308)	0.01% (177)	0.32% (4,309)
40	0.39% (3,918)	3.42% (34,208)	0.03% (439)	0.59% (7,946)
30	0.79% (7,922)	5.24% (52,331)	0.079% (1,067)	1.07% (14,421)
20	1.69% (16,957)	8.29% (82,835)	0.21% (2,828)	2.03% (27,388)
10	4.29% (42,913)	14.55% (145,391)	0.75% (10,156)	4.59% (61,842)
0	24.68% (246,642)	43.27% (432,296)	10.19% (137,216)	24.49% (329,763)

TABLE 3 The counts of morphs predicted to be active by Model17 and Model33 in various highest Tc distributions. GRi – GR inputs, Ri – Random inputs, GRa – GR actives, Ra – Random actives. Compounds were encoded as 1,024 bits long Morgan2 fingerprints.

	Distribution	Low similarity	Medium similarity	High similarity	Exact match
		(Tc < 0.4)	(0.4 ≤ Tc ≤ 0.7)	(Tc > 0.7)	
Model17	GRa vs. GRi	1,318	598	17	2
	Ra vs. GRi	149	27	1	0
	GRa vs. Ri	1,834	99	0	0
	Ra vs. Ri	148	29	0	0
Model33	GRa vs. GRi	17,044	5,179	85	6
	Ra vs. GRi	4,095	212	2	0
	GRa vs. Ri	21,788	517	3	0
	Ra vs. Ri	4,156	153	0	0

the GR actives display the same degree of similarity. These results indicate that, despite utilizing a conformal classifier, which inherently restricts predictions to within the model's applicability domain, Molpher can generate chemically novel actives beyond the conventional similarity search range.

Comparing the GR actives vs. the Random inputs (Figure 7c and Figure 7d) provides an insight into the distance of the morphing library from the general synthetically accessible chemical background. The high presence of compounds in the low similarity subset (1,834 out of 1,933 for Model17 and 21,788 out of 22,308 for Model33) confirms the assumption that the GRML actives are distant from the general chemical background, as morphing paths between the GR inputs are focused on the narrower part of chemical space.

Comparing the Random actives vs. the GR inputs (Figure 7e and Figure 7f), 84% (149 out of 177) of compounds are found to exhibit low similarity for Model17,

and 95% (4,095 out of 4,309) for Model33. This indicates that the Random actives are more dissimilar to the GR inputs than the GR actives, which can be attributed to the Random actives originating from different regions of chemical space than the GR actives.

Comparing the Random actives vs. the Random inputs (Figure 7g and Figure 7h), the number of the low-similarity actives produced by morphing between the Random inputs (148 for Model17 and 4,156 for Model33) is significantly lower than those produced by morphing between the GR inputs (1,318 for Model17 and 17,044 for Model33, Figure 7a and Figure 7b). This suggests that they likely originate from distinct regions of chemical space. Moreover, from the comparison of Random actives vs. Random inputs (Figure 7g and Figure 7h, 148 low-similarity compounds for Model17 and 4,156 low-similarity compounds for Model33) with Random actives vs. GR inputs (Figure 7e and Figure 7f, 149 low-similarity compounds for Model17 and 4,095 low-similarity

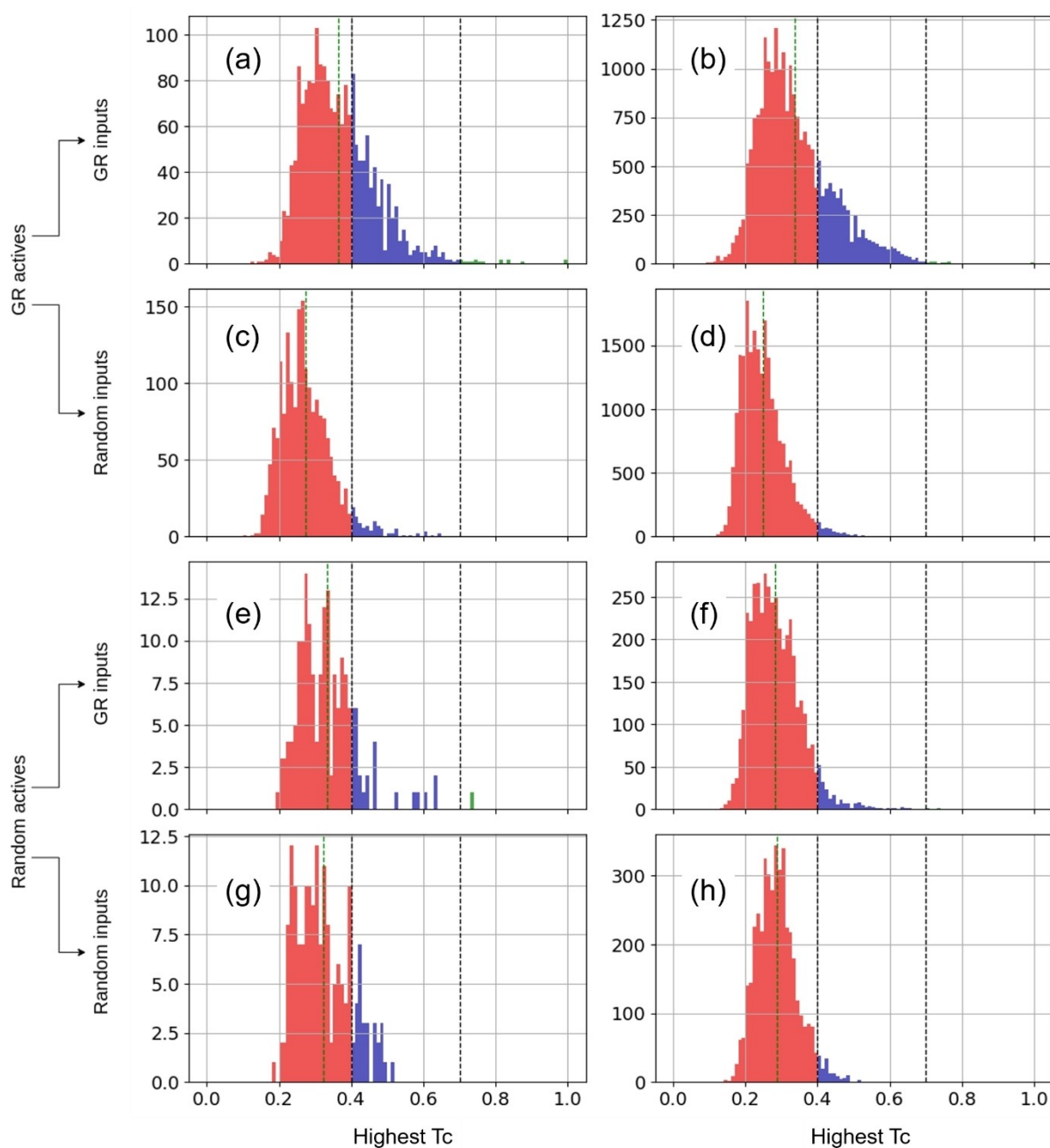


FIGURE 7 The highest Tc distributions between the active and input sets for Model17 and Model33. Dashed lines separate individual similarity subsets. Compound counts with Tanimoto coefficients of the compound pairs from the high similarity subset ($T_c > 0.7$) are depicted in green, from the medium similarity subset ($0.4 \leq T_c \leq 0.7$) in blue, and from the low similarity subset ($T_c < 0.4$) in red. Compounds were encoded as 1,024 bits long Morgan2 fingerprints.

compounds for Model33) it can be inferred that the random actives have low similarity to both the GR inputs and Random inputs, leading to the conclusion that morphing is capable of generating diverse sets of compounds.

In the GR actives, compounds directly corresponding to experimentally verified biologically active ligands were also identified (Table 3, Figure 8). Specifically, Model17 revealed two compounds that are present in the

IMG library, and another four active compounds were discovered by Model33, three of them from the ChEMBL17 and one from the IMG library.

These rediscoveries underscore Molpher's utility for compound generation and the importance of synergizing with predictive models in active compound identification.

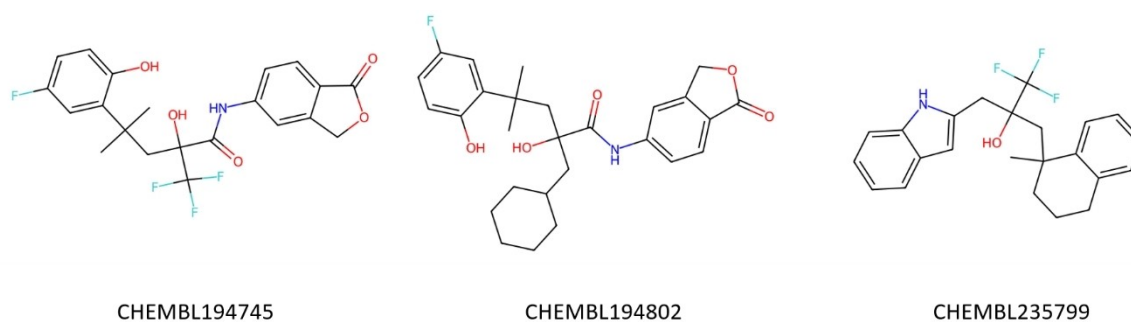


FIGURE 8 ChEMBL17 compounds (given by their ChEMBL IDs) rediscovered by Molpher and predicted to be active by Model33.

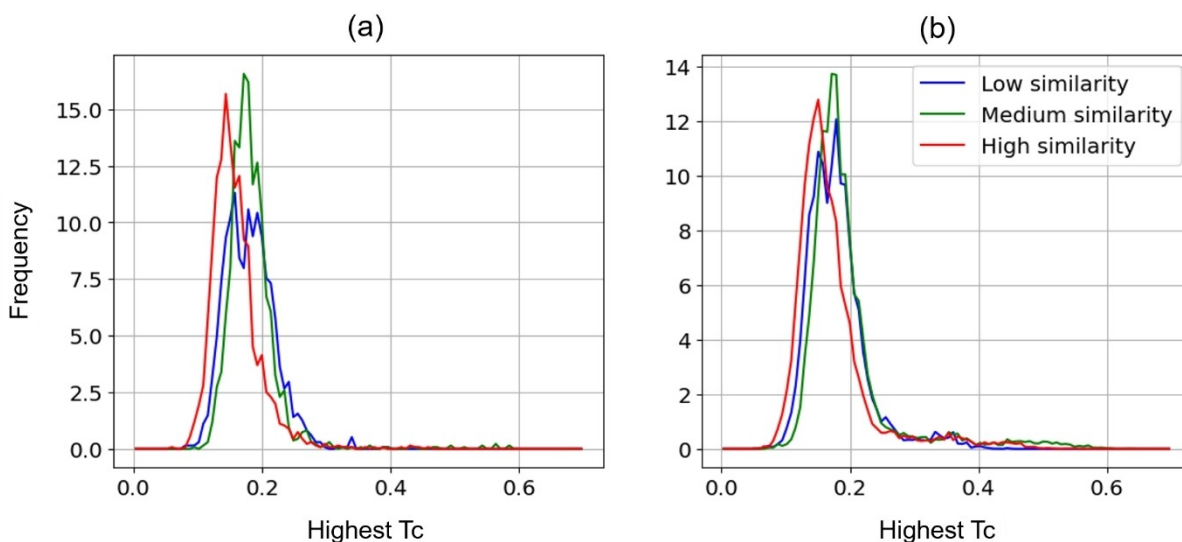


FIGURE 9 The distributions of the highest Tc between the GR actives predicted by Model17 (a) and Model33 (b) and ligands in the individual subsets of the Novel set. The Novel set was divided into three subsets based on the Tc between each active in the Novel set and its most similar active in the ChEMBL17: the low similarity subset ($Tc < 0.4$), medium similarity subset ($0.4 \leq Tc \leq 0.7$), and high similarity subset ($Tc > 0.7$).

3.3.1.2 | Novel compound analysis

252 compounds from the Novel set were divided into three subsets based on their similarity to the ChEMBL17 GR actives: the low-similarity subset ($Tc < 0.4$) with 97 compounds, medium-similarity subset ($0.4 \leq Tc \leq 0.7$) with 127 compounds, and high-similarity subset ($Tc > 0.7$) with 28 compounds. The distributions of the highest Tc between the GR actives predicted by Model17 and Model33 and ligands in the individual subsets of the Novel set are illustrated in Figure 9, and the distribution means are listed in Table 4.

Surprisingly, the average Tc of the predicted actives remained essentially constant for both Model17 and Model33, regardless of the similarity subset of the Novel set they were compared to. This suggests that the assumption of constant distance of actives predicted by Model33 from the Novel set subsets was confirmed, but the same was not true for Model17, where a higher similarity of actives to the high-similarity subset than to the

TABLE 4 The distribution means of the highest Tc between the GR actives predicted by Model17 and Model33 and ligands in the individual subsets of the Novel set. Compounds were encoded using 1,024-bit long Morgan2 fingerprints.

	Low similarity ($Tc < 0.4$)	Medium similarity ($0.4 \leq Tc \leq 0.7$)	High similarity ($Tc > 0.7$)
Model17	0.18	0.18	0.16
Model33	0.18	0.20	0.17

low-similarity subset was expected. Additionally, the results disproved the assumption that Model33 would exhibit a higher average Tc than Model17. These findings suggest that the predicted actives do not necessarily resemble the new ligands that are introduced into the training set, regardless of how much diversity they introduce. Therefore, the difference in the number of actives predicted by Model17 and Model33, respectively, is

based solely on the number of new compounds used for their training and is not influenced by compound diversity.

3.3.1.3 | *KL divergence analysis*

The distributions of the highest Tc between the Novel set and the GRML or RML libraries shown in Figure 10 reveal a significant difference. This observation is further supported by the high KL divergence value of 14.5, which was calculated with the distribution of the highest Tc between the Novel set and the RML library as a reference set. This value indicates a notable dissimilarity between these two distributions further reinforcing the notion that the GRML and RML libraries cover different parts of chemical space.

3.3.2 | Scaffold analysis

The BM scaffolds of the GR and Random inputs and actives were compared with those of the GRML and RML libraries. Out of the 510,180 unique BM scaffolds in the RML library, 1,842 were predicted to be active, of which 158 were also predicted as active in the GRML library (Figure 11). Moreover, 220 RML BM scaffolds appeared in the GRML library but were not predicted as active. Notably, while the active BM scaffolds make up merely

0.36% of the RML scaffolds, they account for 2.6% of the GRML scaffolds, highlighting Molpher's efficiency in exploring the chemical space of GR-active compounds.

While a comprehensive visual inspection of all active BM scaffolds is impractical, the most prevalent ones are presented in Figure 12. Benzene was the most common BM scaffold across the GRML and RML libraries, as identified by both Model17 and Model33. The dominance of this scaffold in drug-like molecules is well known [41] and it has also been called a "trivial framework" [73]. The benzylideneaniline (Figure 12a2) scaffold in the GRML actives predicted by Model17 ranked in the top three in all other sets except the Random actives predicted by Model17. The benzylidenebenzohydrazide (Figure 12a3) scaffold from the GR actives predicted by Model17 resembles the benzylidene-4-phenylbutanehydrazide scaffold (Figure 12d3) from the Random actives predicted by Model33. Further exploration of the phenylindazole (Figure 12b2) scaffold from the GR actives predicted by Model33 is reserved for the following prospective validation section. In summary, the scaffold analysis revealed a significantly larger number of BM scaffolds predicted to be active in the GRML library than in the RML library, which indicates that directed Molpher runs generated a set of predicted actives that is not only larger, but also more diverse than the

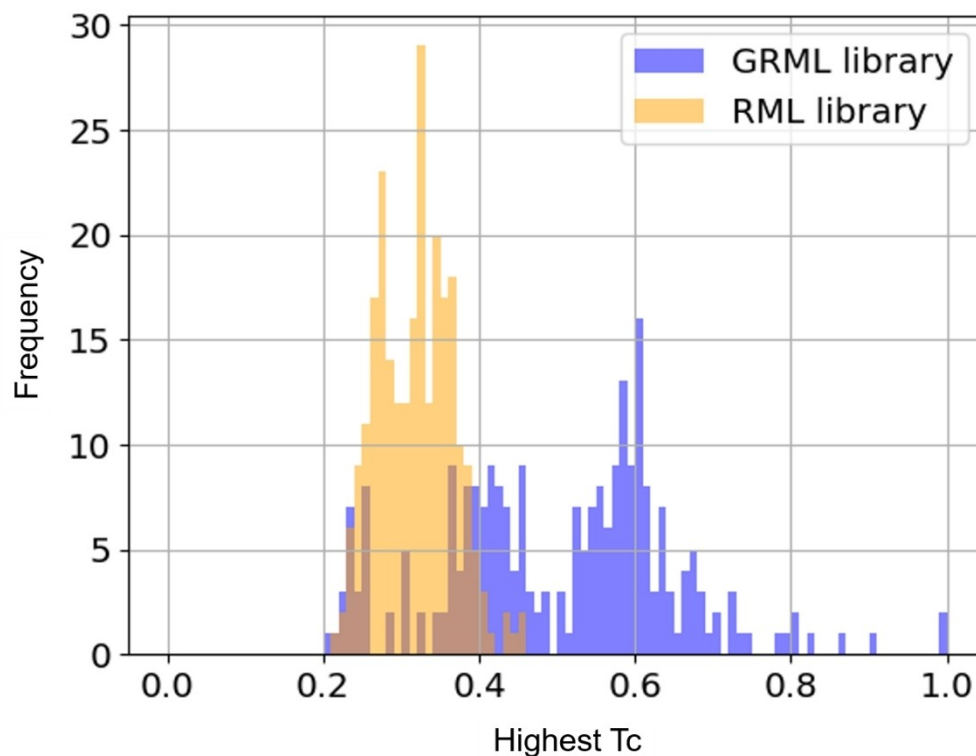


FIGURE 10 The distribution of the highest Tc between the Novel Set and the GRML (blue) or RML (orange) libraries. The histograms are plotted with 100 bins.

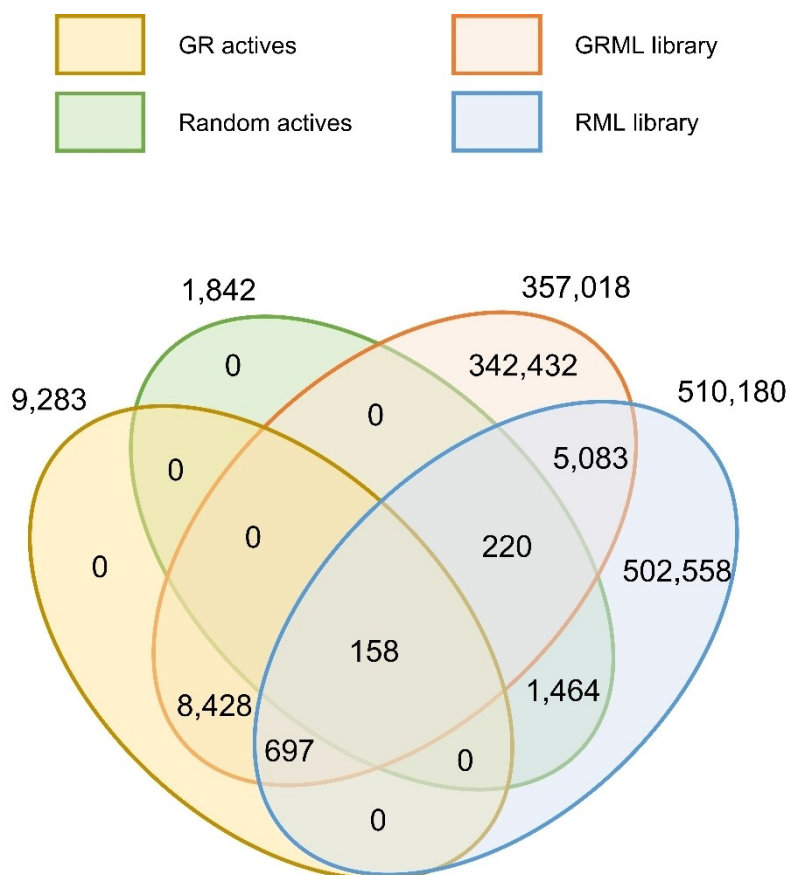


FIGURE 11 Venn diagram illustrating the overlaps between the GRML and RML libraries, as well as the GR and Random actives. The diagram was generated using the venny4py software, version 1.0.0.

reference set. The most common predicted active BM scaffolds were non-steroid.

3.3.3 | Prospective validation

Figure 13 illustrates the relationships between the number of BM scaffolds in the individual sets. 116 new BM scaffolds were found in the Novel set, meaning that they were experimentally identified as active between the ChEMBL17 and ChEMBL33. Two of these were predicted as active in the GRML library (Figure 13). Because the GRML library was generated from unique actives from the ChEMBL17 and IMG libraries, these two BM scaffolds can be considered recovered by Molpher and prospectively validated as active. One of these BM scaffolds was exclusively found in the GRML library (Figure 13b), while the other was found in both the RML and GRML libraries (Figure 13c). This evidence not only confirms Molpher's efficacy in generating active GR scaffolds but also, considering the challenging benchmark imposed by the temporal split [74], the rediscovery of two active compounds is a noteworthy achievement.

3.4 | Design of active ligands

3.4.1 | Pharmacophore model

A pharmacophore model (Figure 14) was constructed containing two hydrogen bond acceptor features (radius = 1.5 Å), two lipophilic features (radius = 1.0 Å), and one essential hydrogen bond acceptor plus donor feature (radius = 1.0 Å). The model also included a volume constraint defined by a radius of 2.0 Å around every heavy atom in ligand superposition.

For the DUD-E benchmark set, both complete matching, using all five features, and partial matching, utilizing either three or four features, were evaluated (Table 5). When all five features were used, only 2.99% of actives and 0.00% of decoys were recognized, excluding a substantial number of active compounds. Employing the four-feature pharmacophore model identified 22.0% of active compounds and 1.14% of decoys (i.e., false positives), producing the $LR+$ of 19.30 and Sen of 0.22. Upon implementing a three-feature partial match, there was a reduction in the $LR+$ to 6.36, accompanied by a slight rise in the Sen to 0.31. Given the reduced discriminative ability of models with five and three

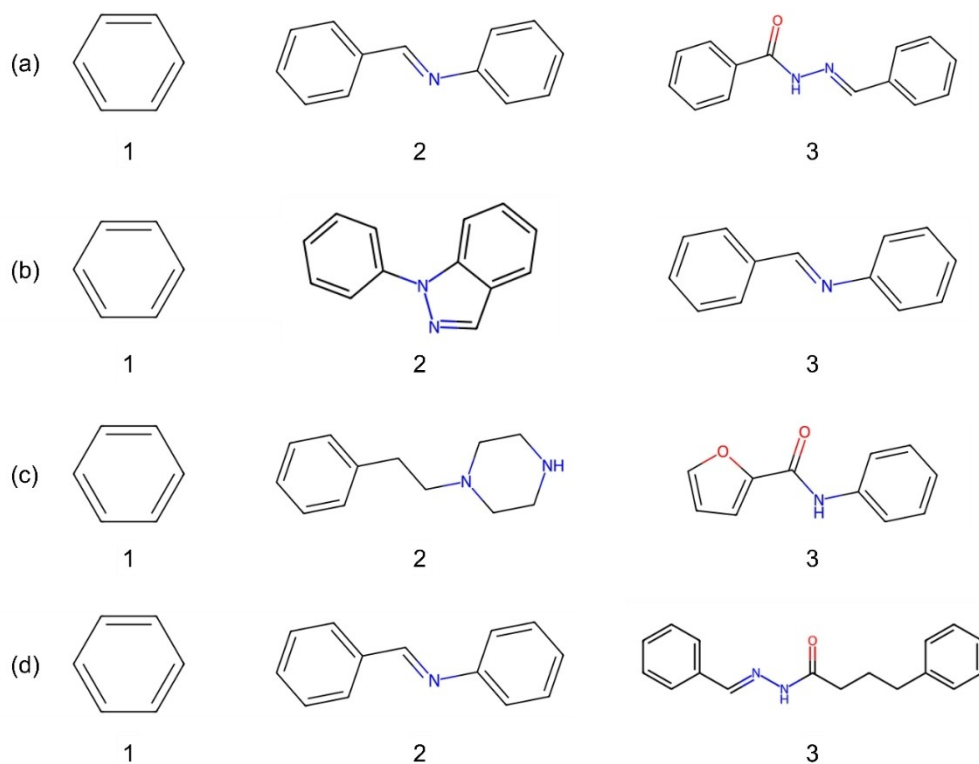


FIGURE 12 Examples of the most common active predicted scaffolds by: (a) Model17 at GRML, (b) Model33 at GRML, (c) Model17 at RML, and (d) Model33 at RML.

(a)

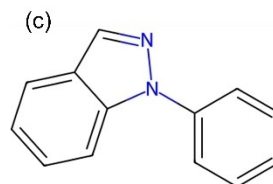
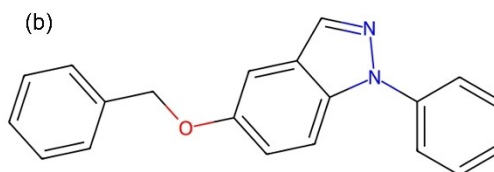
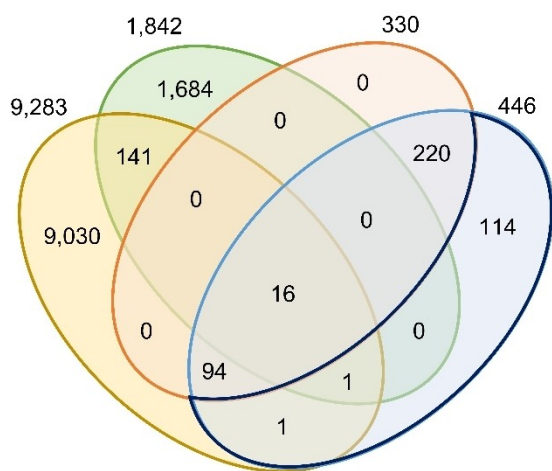
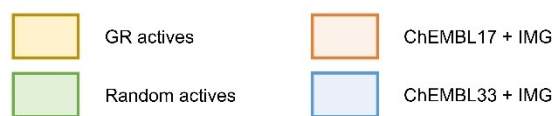


FIGURE 13 (a) Venn diagram illustrating the overlaps in the number of BM scaffolds between the GR and Random actives and two input sets. The Novel set is outlined in dark blue. (b) The scaffold from the GRML library that was experimentally validated as active in the Novel set. (c) The scaffold from both the GRML and RML libraries that was experimentally validated as active in the Novel set. The Venn diagram was generated using the venny4py software, version 1.0.0.

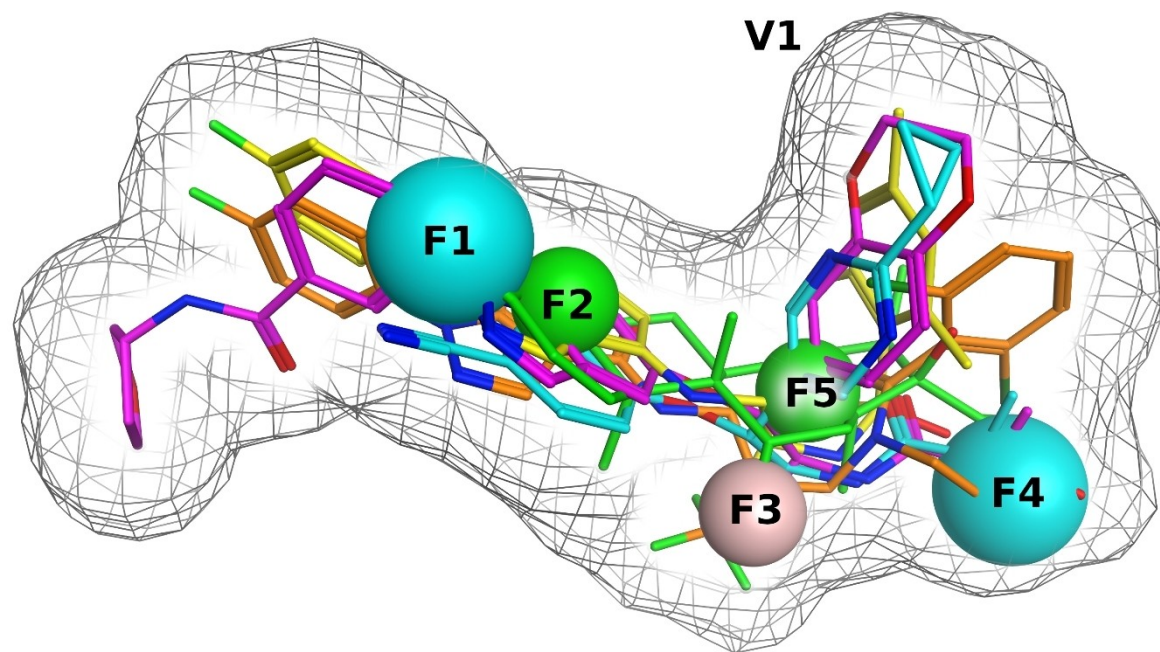


FIGURE 14 Structure-based five-feature pharmacophore model of a GR agonist. Five distinct agonist/GR complexes were used to construct the model: dexamethasone (PDB ID: 1P93 [66]), pyrazole-based agonist (PDB ID: 3E7 C [67]), indazole-based agonist (PDB ID: 4CSJ [68]), pyrimidine-based agonist (PDB ID: 6EL7 [69]), and next-gen indazole-based agonist (PDB ID: 7PRX [70]). The following features were used: F1 and F4 (cyan): hydrogen bond acceptors with a radius of 1.5 Å; F2 and F5 (green): lipophilic features with a radius of 1.0 Å; F3 (salmon): a hydrogen bond acceptor and donor, feature with the radius of 1.0 Å. V1 (grey) is the volume constraint with a radius of 2.0 Å, a hit conformation must lie within this mesh. This figure was created using Molecular Operating Environment 2019.

TABLE 5 Counts of hits identified in the DUD-E benchmark set by the three-, four- and five-feature pharmacophore model. Percentages are shown in brackets. Pharmacophore model quality was evaluated using the positive likelihood ratio (LR+) and model sensitivity (Sen). Four-feature pharmacophore (in bold) yielded the optimal balance between the Sen and LR+.

	Three-feature pharmacophore	Four-feature pharmacophore	Five-feature pharmacophore
Actives	145/468 (31.0%)	103/468 (22.0%)	14/468 (2.99%)
Decoys	737/15150 (4.86%)	173/15150 (1.14%)	0/15150 (0.00%)
LR+	6.36	19.30	Inf
Sen	0.31	0.22	0.03

features, the four-feature pharmacophore model was selected for subsequent investigations due to its optimal balance between the *Sen* and *LR+*.

3.4.2 | Identification and prioritization of active ligands

Following the design of active ligands workflow (Figure 5), a total of 22,527 unique morphs were predicted as active by Model33 or Model17 (Table 2). Of these, 20,338 morphs had a Quantitative Estimation of Drug-likeness (QED) greater than 0.4, 16,242 morphs had a NIBR Severity Score below 10, and 10,174 morphs had a MolSkill score less than 9. Conformations were

generated for these 10,174 morphs, and 23,652 conformations passed the four-feature pharmacophore. These conformations corresponded to 1,699 morphs. Subsequent docking and another use of the four-feature pharmacophore model reduced the number of conformations to 4,009, representing 54 morphs (Supporting information) comprising 34 distinct scaffolds. Upon visual inspection, 34 morphs (Supporting information) with 20 unique scaffolds (Figure 15) were determined to be stable. Of these, 12 molecules (with 8 scaffolds, Figure 15b) were flagged with warnings because of issues like strained ring systems, while 22 morphs with 15 unique scaffolds (Figure 15c) had no warnings. Intriguingly, 13 of these scaffolds differed from known GR

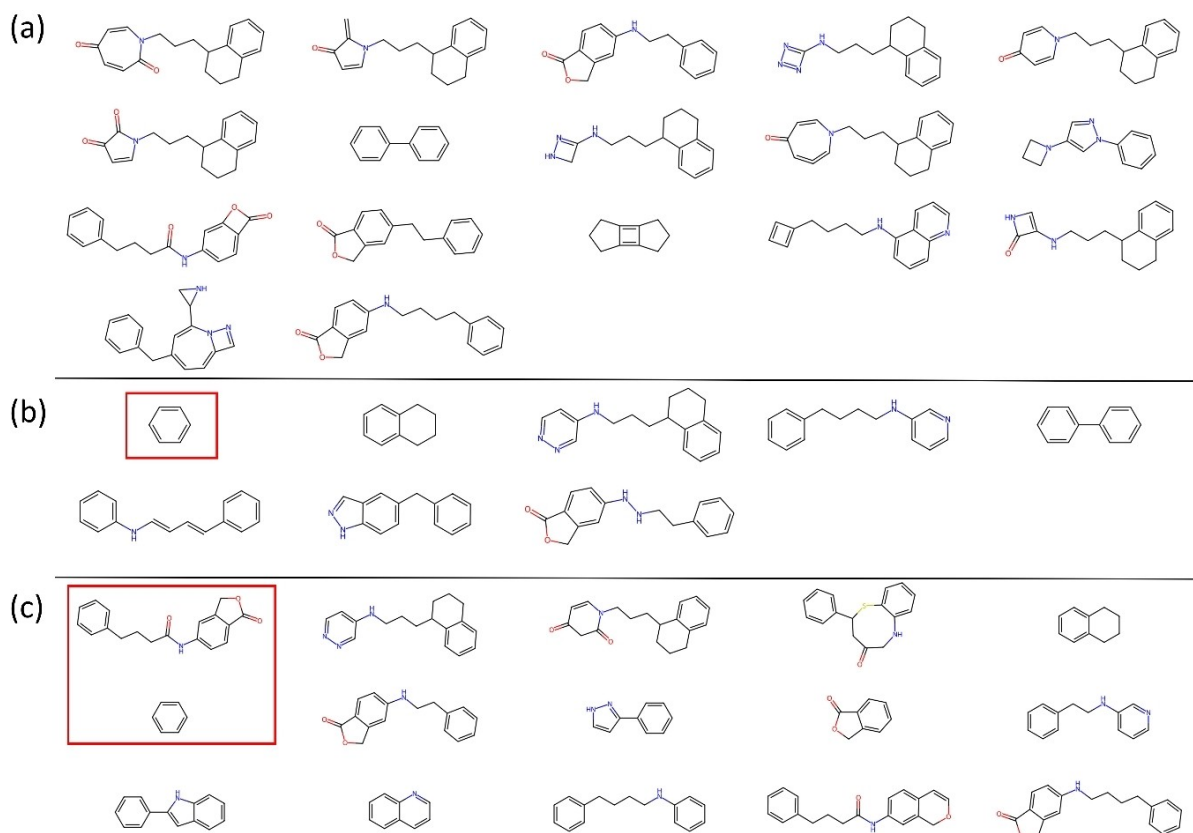


FIGURE 15 BM scaffolds of 34 morphs (Supporting information) predicted to be active on the GR. (a) BM scaffolds of compounds that were manually annotated as not stable. (b) BM scaffolds of compounds that were manually annotated with warnings. (c) BM scaffolds of compounds that were manually annotated without warnings. Two scaffolds already present in ChEMBL33 are outlined in red.

ligands, while two (including the trivial benzene scaffold [73]) were already present in ChEMBL33 (Figure 15c).

Finally, the filtered morphs were prioritized using the rank by rank consensus (RbR) method. This method aggregates the rankings from 100 individual RFR models to determine a consensus rank for each morph. Specifically, the consensus rank for morph i , denoted as RbR_i , is calculated using the formula $RbR_i = \frac{1}{n} \sum_j r_i^j$, where n is the total number of RFR models (i.e., 100), and r_i^j represents the rank of morph i as assigned by the j -th RFR model. These ranks are assigned based on the sorting of predicted pEC50 values from each model. This consensus approach ensures that the final ranking of morphs reflects a comprehensive evaluation across all predictive models, enhancing the reliability of the prioritization process.

4 | CONCLUSIONS

In this study, the capability of Molpher to generate biologically active ligands for the glucocorticoid receptor (GR) was assessed. The ChEMBL database and an internal IMG library, comprising 24,511 chemically diverse

compounds experimentally tested for GR activity in a primary luciferase reporter cell assay, served as the compound sources for Molpher and random forest models. Two virtual libraries were generated using Molpher: one by exploring chemical space between known GR ligands (GRML) and another by exploring space between random compounds (RML). The compounds from these libraries were then classified as either active or inactive by a random forest classifier, with its applicability domain established through Mondrian conformal prediction. An extensive comparison of the GRML and RML libraries indicated that the GRML library was richer in predicted actives than the RML library, and the actives in the GRML library were distinct from the chemical background represented by the RML. By prospective analysis, it was demonstrated that Molpher has the potential to generate compounds later confirmed experimentally as active on the GR. A set of 34 novel potential GR actives has been identified, with their synthesis and subsequent testing planned for future studies. Furthermore, the methodology outlined in this manuscript can serve as a general workflow for assessing computationally generated ligands, especially for those potentially active against challenging-to-dock targets like the GR.

AUTHOR CONTRIBUTIONS

D.S. and P.B. supervised the research project. M.I.A. designed the work, collected the data, constructed models, evaluated their performance, and performed data analysis. M.Š. developed the molecular morphing Python library. I.Č and M.Š. designed and performed molecular morphing experiments. W.D. constructed the pharmacophore model. J.K. designed and Y.Ch. and M.I.A. performed molecular docking experiments. P.B. supervised the experimental screening, D.Se. performed screening experiments. D.S. and M.I.A. wrote the manuscript. All authors read the manuscript and approved the final version.

ACKNOWLEDGMENTS

Computational resources were supplied by the Ministry of Education, Youth and Sports of the Czech Republic under the projects CESNET (Project No. LM2015042), CERIT-Scientific Cloud (Project No. LM2015085) and Metacentrum e-INFRA CZ project (ID:90254). D.Sv., P.B., D.Se., W.D., I.Č and M.I.A. were supported by the Ministry of Education, Youth and Sports of the Czech Republic – project number LM2023052. D.Sv., P.B., D.Se. were supported by RVO 68378050-KAV-NPUI. D.Sv. was further supported by the project New Technologies for Translational Research in Pharmaceutical Sciences/NET-PHARM, project ID CZ.02.01.01/00/22 008/0004607, co-funded by the European Union. Y.C. was supported by the China Scholarship Council (201606010345). Open access publishing facilitated by Vysoka skola chemicko-technologicka v Praze, as part of the Wiley - CzechELib agreement.

DATA AVAILABILITY STATEMENT

Data may be requested via the authors.

ORCID

M. Isabel Agea  <http://orcid.org/0000-0002-3017-7742>

REFERENCES

1. J. L. Medina-Franco, A. L. Chavez-Hernandez, E. Lopez-Lopez, F. I. Saldivar-Gonzalez, *Mol. Inform.* **2022**, *41*, e2200116.
2. a) J. L. Medina-Franco, K. Martinez-Mayorga, N. Meurice, *Expert Opin. Drug. Discov.* **2013**, *23*, 565–574; b) B. Nisius, J. Bajorath, *Expert Opin. Drug. Discov.* **2011**, *6*, 1–7; c) A. De la Vega de Leon, J. Bajorath, *Future Med. Chem.* **2016**, *8*, 1769–1778; d) M. Awale, R. Visini, D. Probst, J. Arus-Pous, J. L. Reymond, *Chimia (Aarau)* **2017**, *71*, 661–666; e) G. M. Maggiora, J. Bajorath, *J. Comput. Aided Mol. Des.* **2014**, *28*, 795–802; f) G. Opassi, A. Gesu, A. Massarotti, *Drug Discov. Today* **2018**, *23*, 565–574.
3. a) C. M. Dobson, *Nature* **2004**, *432*, 824–828; b) J.-L. Reymond, R. van Deursen, L. C. Blum, L. Ruddigkeit, *MedChemComm* **2010**, *1*, 30–38; c) C. Lipinski, A. Hopkins, *Nature* **2004**, *432*, 855–861.
4. a) P. G. Polishchuk, T. I. Madzhidov, A. Varnek, *J. Comput. Aided Mol. Des.* **2013**, *27*, 675–679; b) P. Ertl, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380; c) W. P. Walters, M. T. Stahl, M. A. Murcko, *Drug Discov. Today* **1998**, *3*, 160–178.
5. a) C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26; b) M. A. Abdelgawad, A. M. Hayallah, S. N. A. Bukhari, A. Musa, M. Elmowafy, H. M. Abdel-Rahman, M. K. Abd El-Gaber, *Pharmaceuticals (Basel)* **2022**, *15*, 1–25.
6. a) J. L. Reymond, M. Awale, *ACS Chem Neurosci* **2012**, *3*, 649–657; b) Z. L. Deng, C. X. Du, X. Li, B. Hu, Z. K. Kuang, R. Wang, S. Y. Feng, H. Y. Zhang, D. X. Kong, *J. Chem. Inf. Model* **2013**, *53*, 2820–2828; c) B. I. Diaz-Eufracio, O. Palomino-Hernandez, R. A. Houghten, J. L. Medina-Franco, *Mol. Divers.* **2018**, *22*, 259–267; d) E. Arnold, *Prog. Biophys. Mol. Biol.* **2014**, *116*, 81; e) C. S. Dampalla, Y. Kim, N. Bickmeier, A. D. Rathnayake, H. N. Nguyen, J. Zheng, M. M. Kashipathy, M. A. Baird, K. P. Battaile, S. Lovell, S. Perlman, K. O. Chang, W. C. Groutas, *J. Med. Chem.* **2021**, *64*, 10047–10058.
7. F. I. Saldivar-Gonzalez, B. A. Pilon-Jiménez, J. L. Medina-Franco, *Phys. Sci. Rev.* **2019**, *4*, 20180103.
8. a) C. Y. Cheng, J. E. Campbell, G. M. Day, *Chem. Sci.* **2020**, *11*, 4922–4933; b) Z. Allahyari, A. R. Oganov, *J. Phys. Chem. C* **2020**, *124*, 23867–23878.
9. J. J. Naveja, M. P. Rico-Hidalgo, J. L. Medina-Franco, *F1000Res* **2018**, *7*.
10. D. Gaytan-Hernandez, A. L. Chavez-Hernandez, E. Lopez-Lopez, J. Miranda-Salas, F. I. Saldivar-Gonzalez, J. L. Medina-Franco, *J. Cheminf.* **2023**, *15*, 100.
11. A. Capecchi, A. Zhang, J. L. Reymond, *J. Chem. Inf. Model.* **2020**, *60*, 121–132.
12. a) E. Meggers, *Curr. Opin. Chem. Biol.* **2007**, *11*, 287–292; b) J. P. Janet, C. Duan, A. Nandy, F. Liu, H. J. Kulik, *Acc. Chem. Res.* **2021**, *54*, 532–545; c) A. Nandy, M. G. Taylor, H. J. Kulik, *J. Phys. Chem. Lett.* **2023**, *14*, 5798–5804.
13. a) X. Liu, I. J. A. P. G. J. P. van Westen, *Methods Mol. Biol.* **2021**, *2190*, 139–165; b) D. C. Elton, Z. Boukouvalas, M. D. Fuge, P. W. Chung, *Mol. Syst. Des. Eng.* **2019**, *4*, 828–849; c) P. B. Jorgensen, M. N. Schmidt, O. Winther, *Mol. Inform.* **2018**, *37*; d) C. W. Coley, *Trends Chem* **2021**, *3*, 133–145.
14. a) G. Schneider, U. Fechner, *Nat. Rev. Drug Discov.* **2005**, *4*, 649–663; b) B. Pirard, *Expert Opin. Drug. Discov.* **2011**, *6*, 225–231; c) C. Grebner, H. Matter, A. T. Plowright, G. Hessler, *J. Med. Chem.* **2020**, *63*, 8809–8823.
15. C. A. Nicolaou, N. Brown, *Drug Discov Today Technol* **2013**, *10*, e427–e435.
16. a) N. De Cao, T. Kipf, **2018**, p. arXiv:1805.11973; b) A. Kadurin, S. Nikolenko, K. Khrabrov, A. Aliper, A. Zhavoronkov, *Mol. Pharm.* **2017**, *14*, 3098–3104; c) A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zhulus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, *Nat. Biotechnol.* **2019**, *37*, 1038–1040; d) H. Kim, J. Na, W. B. Lee, *J. Chem. Inf. Model.* **2021**, *61*, 5804–5814.
17. a) A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, Baskin, II, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R.

- Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* **2014**, *57*, 4977–5010; b) D. E. Dawson, B. L. Ingle, K. A. Phillips, J. W. Nichols, J. F. Wambaugh, R. Tornero-Velez, *Environ. Sci. Technol.* **2021**, *55*, 6505–6517; c) V. H. Masand, V. Rastija, M. K. Patil, A. Gandhi, A. Chapolikar, *SAR QSAR Environ. Res.* **2020**, *31*, 643–654.
18. a) T. I. Netzeva, A. Worth, T. Aldenberg, R. Benigni, M. T. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. Roberts, T. Schultz, D. W. Stanton, J. J. van de Sandt, W. Tong, G. Veith, C. Yang, *Altern. Lab. Anim.* **2005**, *33*, 155–173; b) I. V. Tetko, P. Bruneau, H. W. Mewes, D. C. Rohrer, G. I. Poda, *Drug Discov. Today* **2006**, *11*, 700–707.
 19. D. Hoksza, P. Skoda, M. Vorsilak, D. Svozil, *J. Cheminformatics* **2014**, *6*, 1–19.
 20. W. Gao, C. W. Coley, *J. Chem. Inf. Model.* **2020**, *60*, 5714–5723.
 21. P. Ertl, A. Schuffenhauer, *J. Cheminf.* **2009**, *1*, 8.
 22. M. Vorsilak, D. Svozil, *J. Cheminf.* **2017**, *9*, 20.
 23. M. Vorsilak, M. Kolar, I. Cmelo, D. Svozil, *J. Cheminf.* **2020**, *12*, 35.
 24. molpher-lib, <https://github.com/lich-uct/molpher-lib>, accessed in January 2024.
 25. S. Vandevyver, L. Dejager, C. Libert, *Endocr. Rev.* **2014**, *35*, 671–693.
 26. E. A. Lesovaya, D. Chudakova, G. Baida, E. M. Zhidkova, K. I. Kirsanov, M. G. Yakubovskaya, I. V. Budunova, *Oncotarget* **2022**, *13*, 408–424.
 27. a) R. A. Quax, L. Manenschijn, J. W. Koper, J. M. Hazes, S. W. Lamberts, E. F. van Rossum, R. A. Feelders, *Nat. Rev. Endocrinol.* **2013**, *9*, 670–686; b) M. Kadmiel, J. A. Cidlowski, *Trends Pharmacol. Sci.* **2013**, *34*, 518–530.
 28. N. C. Nicolaides, G. P. Chrousos, *Int. J. Mol. Sci.* **2023**, *24*, 11030.
 29. D. Sedlak, A. Paguio, P. Bartunek, *Comb. Chem. High Throughput Screening* **2011**, *14*, 248–266.
 30. M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2012**, *55*, 6582–6594.
 31. a) J. Deng, Z. Yang, H. Wang, I. Ojima, D. Samaras, F. Wang, *Nat. Commun.* **2023**, *14*, 6395; b) S. Liu, M. Alnammi, S. S. Ericksen, A. F. Voter, G. E. Ananiev, J. L. Keck, F. M. Hoffmann, S. A. Wildman, A. Gitter, *J. Chem. Inf. Model.* **2019**, *59*, 282–293; c) C. Skuta, I. Cortes-Ciriano, W. Dehaen, P. Kriz, G. J. P. van Westen, I. V. Tetko, A. Bender, D. Svozil, *J. Cheminf.* **2020**, *12*, 39; d) I. Cortes-Ciriano, C. Skuta, A. Bender, D. Svozil, *J. Cheminf.* **2020**, *12*, 41.
 32. a) V. Vovk, A. Gammerman, G. Shafer, *Algorithmic learning in a random world*, Springer, New York, **2005**; b) U. Norinder, L. Carlsson, S. Boyer, M. Eklund, *J. Chem. Inf. Model.* **2014**, *54*, 1596–1603; c) G. Shafer, V. Vovk, *J. Mach. Learn. Res.* **2008**, *9*, 371–421; d) A. Morger, F. Svensson, S. Arvidsson McShane, N. Gauraha, U. Norinder, O. Spjuth, A. Volkamer, *J. Cheminf.* **2021**, *13*, 35; e) A. Wilm, U. Norinder, M. I. Agea, C. de Bruyn Kops, C. Stork, J. Kuhn, J. Kirchmair, *Chem. Res. Toxicol.* **2021**, *34*, 330–344; f) A. Forreryd, U. Norinder, T. Lindberg, M. Lindstedt, *Toxicol. In Vitro* **2018**, *48*, 179–187.
 33. a) A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magarinos, J. P. Overington, G. Papadatos, I. Smit, A. R. Leach, *Nucleic Acids Res.* **2017**, *45*, D945–D954; b) D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Felix, M. P. Magarinos, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Maranon, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, A. R. Leach, *Nucleic Acids Res.* **2019**, *47*, D930–D940.
 34. ChemBridge Diversity Library, <https://chembridge.com/diversity-and-pre-plated-libraries/diversity-libraries/>, accessed in October 2023.
 35. P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **2000**, *43*, 3714–3717.
 36. Prestwick Chemical Library, <https://www.prestwickchemical.com/screening-libraries/prestwick-chemical-library/>, accessed in October 2023.
 37. Sigma LOPAC Library of Pharmacologically Active Compounds, <https://www.sigmaaldrich.com/CZ/en/product/sigma/lo1280>, accessed in October 2023.
 38. NIH Clinical Collection, <https://commonfund.nih.gov/Molecularlibraries/tools>, accessed in October 2023.
 39. T. Müller, D. Sedlák, P. Bartuněk, *Chem. Listy* **2017**, *111*, 766–771.
 40. C. Brideau, B. Gunter, B. Pikounis, A. Liaw, *J. Biomol. Screen.* **2003**, *8*, 634–647.
 41. G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1996**, *39*, 2887–2893.
 42. a) T. Sterling, J. J. Irwin, *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337; b) J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
 43. D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
 44. S. Weaver, M. P. Gleeson, *J. Mol. Graphics Modell.* **2008**, *26*, 1315–1326.
 45. J. D. Holliday, C. Y. Hu, P. Willett, *Comb. Chem. High Throughput Screening* **2002**, *5*, 155–166.
 46. V. Vovk, *Ijip Adv. Inf. Comm. Te.* **2013**, *412*, 348–360.
 47. V. Vovk, *Mach. Learn.* **2013**, *92*, 349–376.
 48. a) H. Papadopoulos, in *Tools in Artificial Intelligence* (Ed.: P. Fritzsche), IntechOpen, **2008**; b) L. Carlsson, C. Bendtsen, E. Ahlberg, in *Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications, Vol. 60* (Eds.: G. Alex, V. Vladimir, L. Zhiyuan, P. Hahrris), PMLR, Proceedings of Machine Learning Research, **2017**, pp. 201–212.
 49. J. Sun, L. Carlsson, E. Ahlberg, U. Norinder, O. Engkvist, H. Chen, *J. Chem. Inf. Model.* **2017**, *57*, 1591–1598.
 50. a) S. Lampa, J. Alvarsson, S. Arvidsson McShane, A. Berg, E. Ahlberg, O. Spjuth, *Front Pharmacol* **2018**, *9*, 1256; b) V. Vovk, V. Fedorova, I. Nouredinov, A. Gammerman, in *Conformal and Probabilistic Prediction with Applications* (Eds.: A. Gammerman, Z. Luo, J. Vega, V. Vovk), Springer International Publishing, Cham, **2016**, pp. 23–39.
 51. V. Vovk, **2012**, p. arXiv:1208.0806.
 52. U. Johansson, H. Linusson, T. Löfström, H. Boström, *IEEE IJCNN* **2017**, 2072–2079.
 53. Nonconformist package, <https://github.com/donlnz/nonconformist>, accessed in October 2023.
 54. a) A. Mellor, S. Boukir, A. Haywood, S. Jones, *ISPRS J. Photogramm.* **2015**, *105*, 155–168; b) A. S. More, D. P. Rana, *2017 1st International Conference on Intelligent Systems and Information Management (Icिसim)* **2017**, 72–78.

55. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
56. J. Bergstra, D. Yasmins, D. D. Cox, *Proceedings of the 30th International Conference on Machine Learning (ICML 2013)* **2013**, 28.
57. RDKit: Open-Source Cheminformatics, <http://www.rdkit.org/>, accessed in October 2023.
58. S. Kullback, R. A. Leibler, *Ann. Math. Stat.* **1951**, *22*, 79–86.
59. P.-S. Laplace, *Theorie analytique des probabilités*, Courcier, **1812**, Paris, France.
60. SciPy, <https://scipy.org/>, accessed in October 2023.
61. A. Schuffenhauer, N. Schneider, S. Hintermann, D. Auld, J. Blank, S. Cotesta, C. Engeloch, N. Fechner, C. Gaul, J. Giovannoni, J. Jansen, J. Joslin, P. Krastel, E. Lounkine, J. Manchester, L. G. Monovich, A. P. Pelliccioli, M. Schwarze, M. D. Shultz, N. Stiefl, D. K. Baeschlin, *J. Med. Chem.* **2020**, *63*, 14425–14447.
62. O.-H. Choung, R. Vianello, M. Segler, N. Stiefl, J. Jiménez-Luna, *Nat. Commun.* **2023**, *6651*, 10.
63. G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, *Nat. Chem.* **2012**, *4*, 90–98.
64. Molecular Operating Environment (MOE) 2019, Chemical Computing Group ULC, 910–1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.
65. a) C. Potamitis, D. Siakouli, K. D. Papavasileiou, A. Boulaka, V. Ganou, M. Roussaki, T. Calogeropoulou, P. Zoumpoulakis, M. N. Alexis, M. Zervou, D. J. Mitsiou, *J. Steroid Biochem. Mol. Biol.* **2019**, *186*, 142–153; b) R. Metin, E. D. Akten, *J. Biomol. Struct. Dyn.* **2022**, *40*, 11418–11433; c) V. Onnis, G. K. Kinsella, G. Carta, W. N. Jagoe, T. Price, D. C. Williams, D. Fayne, D. G. Lloyd, *J. Med. Chem.* **2010**, *53*, 3065–3074.
66. B. Kauppi, C. Jakob, M. Farnegardh, J. Yang, H. Ahola, M. Alarcon, K. Calles, O. Engstrom, J. Harlan, S. Muchmore, A. K. Ramqvist, S. Thorell, L. Ohman, J. Greer, J. A. Gustafsson, J. Carlstedt-Duke, M. Carlquist, *J. Biol. Chem.* **2003**, *278*, 22748–22754.
67. K. P. Madauss, R. K. Bledsoe, I. McLay, E. L. Stewart, I. J. Uings, G. Weingarten, S. P. Williams, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6097–6099.
68. K. Edman, R. Ahlgren, M. Bengtsson, H. Bladh, S. Backstrom, J. Dahmen, K. Henriksson, P. Hillertz, V. Hulikal, A. Jerre, L. Kinchin, C. Kase, M. Lepisto, I. Mile, S. Nilsson, A. Smailagic, J. Taylor, A. Tjornebo, L. Wissler, T. Hansson, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2571–2577.
69. L. Ripa, K. Edman, M. Dearman, G. Edenro, R. Hendrickx, V. Ullah, H. F. Chang, M. Lepisto, D. Chapman, S. Geschwindner, L. Wissler, P. Svanberg, K. Lawitz, J. Malmberg, A. Nikitidis, R. I. Olsson, J. Bird, A. Llinas, T. Hegelund-Myrback, M. Berger, P. Thorne, R. Harrison, C. Kohler, T. Drmota, *J. Med. Chem.* **2018**, *61*, 1785–1799.
70. S. Postel, L. Wissler, C. A. Johansson, A. Gunnarsson, E. Gordon, B. Collins, M. Castaldo, C. Kohler, D. Oling, P. Johansson, L. Froderberg Roth, B. Beinsteiner, I. Dainty, S. Delaney, B. P. Klaholz, I. M. L. Billas, K. Edman, *Nat. Struct. Mol. Biol.* **2023**, *30*, 286–295.
71. a) F. Svensson, N. Aniceto, U. Norinder, I. Cortes-Ciriano, O. Spjuth, L. Carlsson, A. Bender, *J. Chem. Inf. Model.* **2018**, *58*, 1132–1140; b) N. Bosc, F. Atkinson, E. Felix, A. Gaulton, A. Hersey, A. R. Leach, *J. Cheminf.* **2019**, *11*, 4.
72. J. Zhang, U. Norinder, F. Svensson, *J. Chem. Inf. Model.* **2021**, *61*, 2648–2657.
73. Ajay, G. W. Bemis, M. A. Murcko, *J. Med. Chem.* **1999**, *42*, 4942–4951.
74. A. Morger, M. Garcia de Lomana, U. Norinder, F. Svensson, J. Kirchmair, M. Mathea, A. Volkamer, *Sci. Rep.* **2022**, *12*, 7244.

SUPPORTING INFORMATION

Data and models used in the experiments are freely available on Zenodo (<https://zenodo.org/records/11001950>). Python code to reproduce the workflow is freely available in the GitHub repository (<https://github.com/Iagea/GRML> analyses).

Following additional supporting information can be found online in the Supporting Information section at the end of this article: The file designed ligands.xlsx containing the list and structures of 54 GR ligands with their QED, NIBR severity score, MolSkill score, RBR consensus ranking score, the result of the manual annotation and remarks, if available.

How to cite this article: M. I. Agea, I. Čmelo, W. Dehaen, Y. Chen, J. Kirchmair, D. Sedlák, P. Bartůněk, M. Šícho, D. Svozil, *Molecular Informatics* **2024**, *43*, e202300316. <https://doi.org/10.1002/minf.202300316>