



Article

Rational Method for Structural Simplification as Key Step in Hit Discovery: The Case of FGFR2 and IGF1R Dual Inhibitors

Endika Torres-Urtizberea , José I. Borrell , Raimon Puig de la Bellacasa and Roger Estrada-Tejedor *

Grup de Química Farmacèutica, IQS School of Engineering, Universitat Ramon Llull, Via Augusta 390, E-08017 Barcelona, Spain; endika.torres@iqs.url.edu (E.T.-U.); jose.borrell@iqs.url.edu (J.I.B.); raimon.puig@iqs.url.edu (R.P.d.l.B.)

* Correspondence: roger.estrada@iqs.url.edu

Abstract: In the classic medicinal chemistry hit discovery procedure, large virtual libraries undergo different filtering and prediction steps until a small group of molecules is selected for their subsequent synthesis and biological testing. The starting molecular libraries can easily be composed of millions of molecules, hindering the selection of the most representative and promising compounds. Moreover, the resulting molecular systems tend to be overcomplex structures, hardly attainable, and often involve extrapolations of the prediction models used. We present a rational-based method to reduce the structural complexity of molecular candidates without compromising their biological activity, improving the attainability and efficiency of hit discovery. This approach has been successfully applied to identify potential tyrosine kinase dual inhibitors against Fibroblast Growth Factor Receptor 2 (FGFR2) and Insulin-Like Growth Factor 1 Receptor (IGF1R), a set of overexpressed proteins in different cancers, such as pancreatic ductal adenocarcinoma (PDAC).

Keywords: QSAR; structure complexity; hit discovery



Academic Editor: Ana Amić

Received: 22 March 2025

Revised: 1 May 2025

Accepted: 6 May 2025

Published: 7 May 2025

Citation: Torres-Urtizberea, E.; Borrell, J.I.; Puig de la Bellacasa, R.; Estrada-Tejedor, R. Rational Method for Structural Simplification as Key Step in Hit Discovery: The Case of FGFR2 and IGF1R Dual Inhibitors. *Int. J. Mol. Sci.* **2025**, *26*, 4457. <https://doi.org/10.3390/ijms26094457>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the drug development process, choosing the strategy for lead discovery is often a crucial step, as it will determine the starting point of the whole process. Such strategies are often based on high-throughput screening (HTS) [1], structure-based drug design (SBDD) [2], fragment-based lead generation (FBLG) [3] or the most used, knowledge-based lead generation (KBLG) [4]. While the selection of a strategy is driven by several factors, KBLG has been the most used method overall, as it allows one to skip the hit identification step, increasing the process' efficacy while reducing the costs (Figure 1).

Although KBLG has been successful in identifying new leads [5], it requires a large number of very detailed, high-quality data, which, in many cases, are not available in open databases. For this reason, in-house databases are commonly used for hit selection, allowing the generation of tailored structures that may align more closely with the research strategy of a given research group. At the same time, this ensures the data's sensitivity and that the intellectual property remains within the group, while increasing the novelty of future discoveries. The generation of such hit selection often begins by synthesizing and testing a set of molecules [6]. These, ideally, would be easily synthesizable while having a high probability of showing the desired activity, generating as many hits as possible in the minimum time possible. For this, the selection of the structures to be synthesized should consider both structural diversity and synthetic feasibility, as they will set the foundations for lead generation.

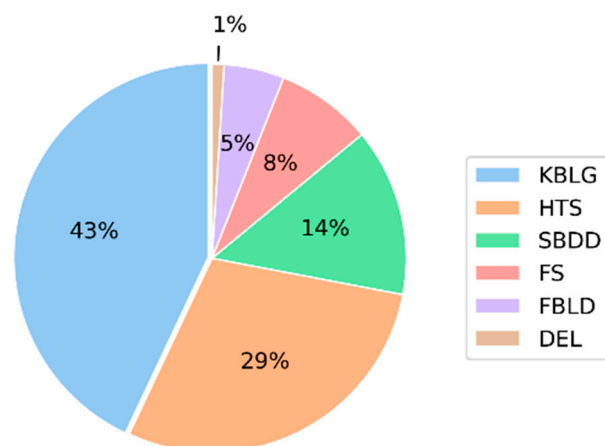


Figure 1. Percentage of lead generation sources published in J. Med. Chem between 2016 and 2017 [5]. Knowledge-based lead generation (KBLG), high-throughput screening (HTS), structure-based drug design (SBDD), focused screening (FS), fragment-based lead generation (FBLG), DNA encoded libraries (DEL).

A popular means of generating hit candidates is ligand-based drug design (LBDD) approaches [7], where predictive models allow the determination of the best candidates to be synthesized from a previously generated pool of structures. This pool is generated by centering the structure generation around a privileged scaffold [8], which, thanks to its various substitution points, eases the creation of numerous and diverse structures by applying combinatorial approaches in which an independent number of diverse substituents is proposed for each position amenable to functionalization (Figure 2). However, these approaches are often affected by the generation of highly complex structures, which hinders the progression of the hit discovery process [9], as they would require longer synthetic paths. This situation leads to the selection of less complex structures, often consisting of non-relevant molecules, neglecting relevant information from complex structures.

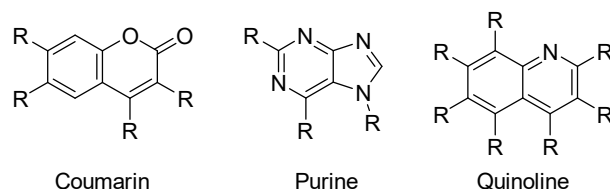


Figure 2. Examples of some privileged scaffolds and their positions amenable to functionalization.

In this work, we present a rational methodology to reduce the structural complexity of drug candidates, removing unnecessary structural features that may pose synthetic challenges while keeping relevant structural information by considering their effects on the predicted activity. This would expedite greatly the hit discovery process in small research groups, not only reaching the hit-to-lead stage sooner but also allowing the use of in-house KBLG once the hit library grows sufficiently. Moreover, the suggested methodology lacks human bias, allowing the selection of structures that may not have been explored otherwise, increasing the chemical space coverage while maintaining high predicted activity.

The use and validation of the method is demonstrated within the context of a project focused on developing dual tyrosine kinase inhibitors (TKI) targeting Fibroblast Growth Factor Receptor 2 (FGFR2) and Insulin-Like Growth Factor 1 Receptor (IGF1R), two proteins that have previously been identified as promising targets against PDAC [10,11]. The simultaneous inhibition of these proteins has been shown to reduce both tumor progression and drug resistance development [12–14]. At the same time, the pyrido[2,3-*d*]pyrimidine

scaffold was used as the core for the generation of the combinatorial library, given its high degree of tunability and biologically active examples [15–21] (Figure 3).

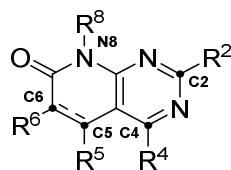


Figure 3. Main substitution points in the pyrido[2,3-*d*]pyrimidine scaffold.

2. Results and Discussion

2.1. QSAR Model Performance

Following the methodology stated at Materials and Methods, the created prediction models performed outstandingly, achieving Q^2 values of above 0.7 and requiring less than 70 features (Table 1).

Table 1. Correlation coefficient values for the validation of the SVM model against the target proteins, in the training set (R^2 self) and the validation set: LOO (Q^2), CV5 (R^2 CV5), CV10 (R^2 CV10) and y-randomization (y-rand).

Target	R^2 (Self)	Q^2	R^2 (CV5)	R^2 (CV10)	y-Rand	Features
FGFR2	0.946	0.733	0.698	0.720	−0.285	69
IGF1R	0.925	0.739	0.695	0.710	−0.352	51

Application of the Prediction Model on the Test Set

The generated model was applied on a test set composed of the combinatorial library obtained by substituting the pyrido[2,3-*d*]pyrimidine scaffold's diversifiable points (Figure 4). The final combinatorial library included a total number of 682,500 different molecules.

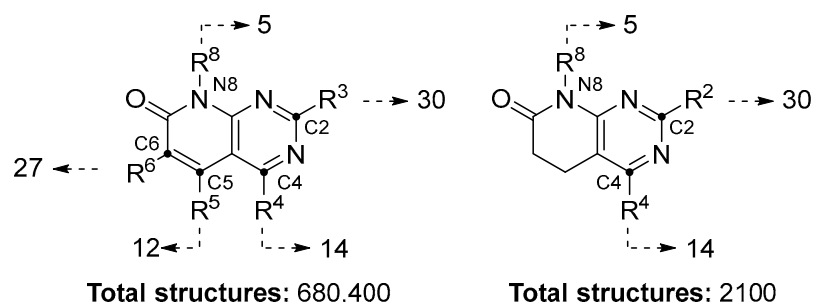


Figure 4. Number of substituents per pyrido[2,3-*d*]pyrimidine scaffold position.

Due to the structural diversity of the compounds included in the combinatorial library, some of them fell outside the applicability range of the prediction models, as their feature values did not fit within the margins set by the training dataset.

The database was consequently divided into two groups: those whose prediction was considered extrapolated and those whose predictions were not. Therefore, the predicted activity distribution of each group was plotted (Figure 5), which showed that extrapolated predictions were found throughout the whole observed activity value range, indicating a lack of extrapolation bias. Considering this, using the whole unfiltered database would have given as much information as possible without compromising the quality of the simplification.

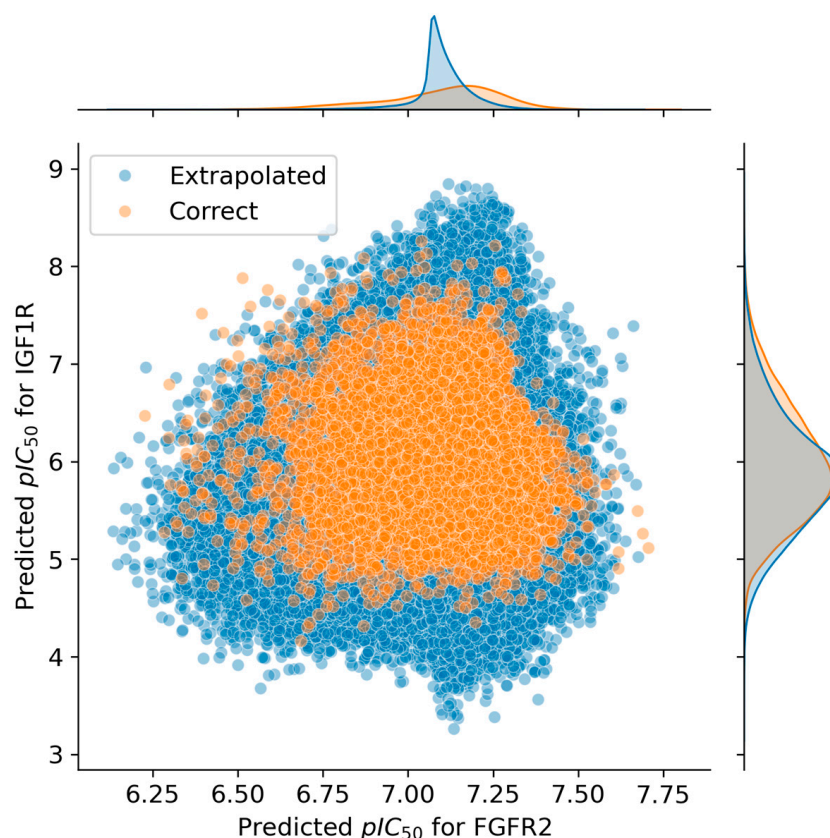


Figure 5. Distribution analysis of the predicted pIC_{50} values against FGFR2 and IGF1R. Molecules with features included in the range of the training set are shown in orange, while compounds that are supposedly extrapolating are shown in blue.

A closer examination of the most active molecules confirmed the first part of the proposed hypothesis: due to the nature of the combinatorial structure generation, the best predicted structures showed great complexity due to their high degree of substitution (Figure 6).

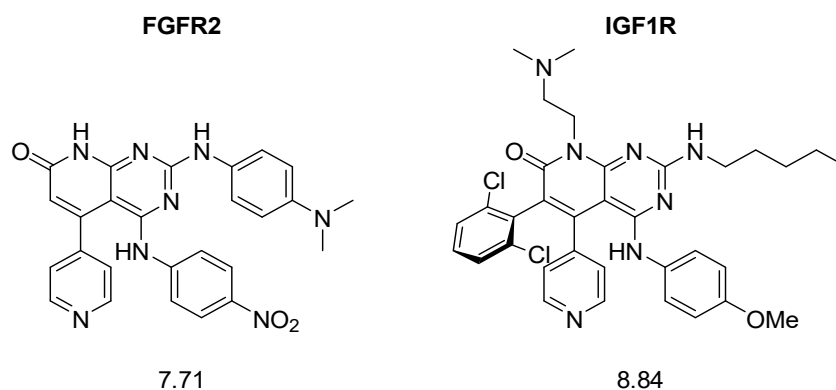


Figure 6. Example structures with high predicted pIC_{50} values for each protein.

The synthesis of these structures would require, in the best case, more than 8 steps, becoming a very costly means of generating hit molecules.

Following our hypothesis, if we were able to rank the best structures by both synthetic feasibility and predicted activity, much more efficient candidates could be selected. The problem with this type of metric is its lack of consistency, being very dependent on the type of structure used (Figure 7).

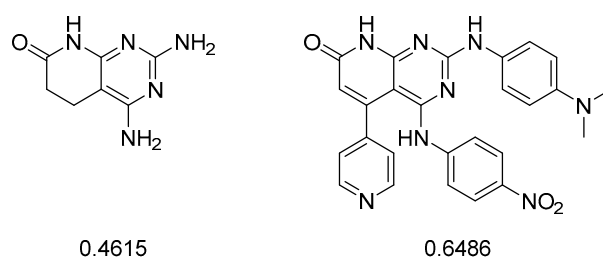


Figure 7. The synthetic feasibility molecular descriptor (*rsynth*), calculated in MOE, leads to a value between 0 and 1 (where 0 corresponds to a not synthetically feasible structure and 1 means that it is fully feasible). For instance, this descriptor considered less feasible a pyrido[2,3-*d*]pyrimidine structure that was obtained in a single step compared to another that required more than eight steps.

It is challenging to define synthetic feasibility as it should consider the reaction steps, reaction complexity and reagent availability, among others [22]. While these aspects are strongly dependent on the used scaffold and substituents, generally, one can understand that the more decorated a scaffold is, the longer the synthetic path will likely be. Based on this, we defined synthetic feasibility solely as a function of the number of substitutions in the scaffold, meaning that reducing the number of substitutions would increase the synthetic feasibility.

In order to conduct the simplification of the structural complexity in a rational way, we proposed taking advantage of SAR analysis tools, as they would allow us to eliminate those substituents with weaker relations to the predicted activity.

2.2. SAR-Based Structural Simplification

SAR methodologies are the foundation of drug design [23], allowing a systematic approach to understanding how the chemical structures of molecules influence their biological activity.

Although this type of methodology is not usually used with predicted activity values, it could help to assess the influence of the substituents on the prediction model. SAR analysis could be therefore used to identify the most frequent substituents among the compounds with higher predicted activity. Thus, one could reduce the structural complexity by focusing on structurally simple structures bearing the privileged substituents identified from compounds with the highest predicted activity.

This procedure would yield structures with reduced predicted activity, but it should also allow the fast exploration of several substituents that could point towards more complex structures requiring longer synthetic paths.

At the same time, while this method did not consider the synergistic effect between multiple substitutions, since the substituents were chosen based on the frequency of appearance, we expected that these substituents could maintain their activity, even when different substituents were present in the scaffold, making them privileged. Structures derived from these could be considered good hit candidates as they would still maintain the possibility of further modification in the following hit-to-lead steps.

Finding Privileged Substituents

Using MOE's SAR analysis tool (SARReport), the 4000 structures with the best average predicted activity against both FGFR2 and IGF1R were analyzed, defining $\text{pIC}_{50} > 7$ as the threshold to identify the most active molecules.

The results showed that no significant differences existed between the privileged substituents found among both the extrapolating and the non-extrapolating compounds. This further supported the idea that extrapolation was not an issue in this case (Figure 8).

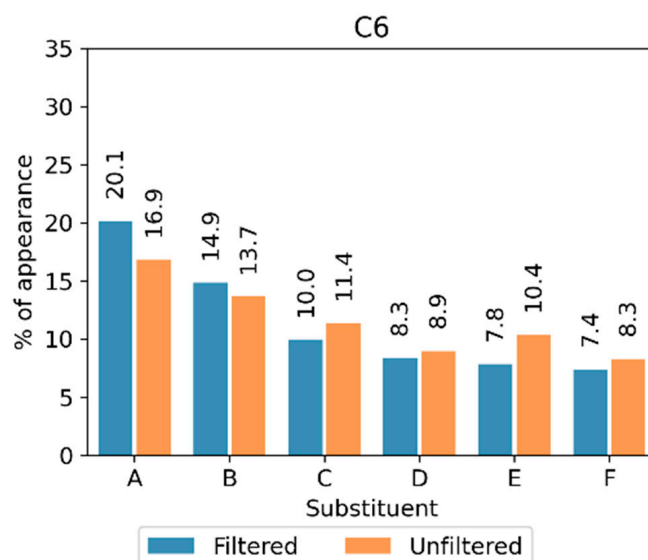


Figure 8. Percentage of appearance of the most popular substituents at the C6 position when removing the compounds that permit the extrapolation of the prediction model (filtered) or analyzing the entire chemical library (unfiltered). The different structural fragments are labeled with capital letters (A–F).

Using the privileged substituents for each position, one or more simplified pyrido[2,3-*d*]pyrimidines were generated, where only the given substituent was maintained in the structure (Figure 9).

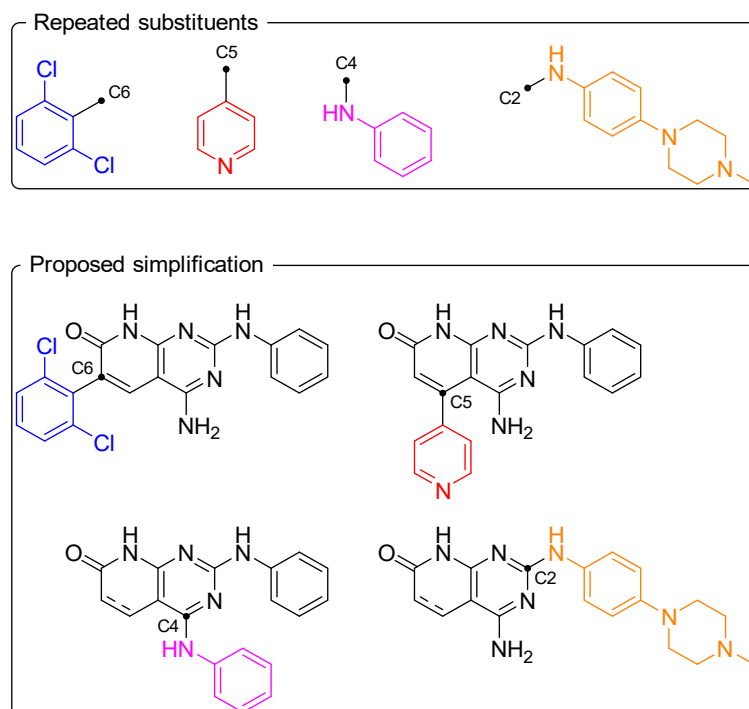


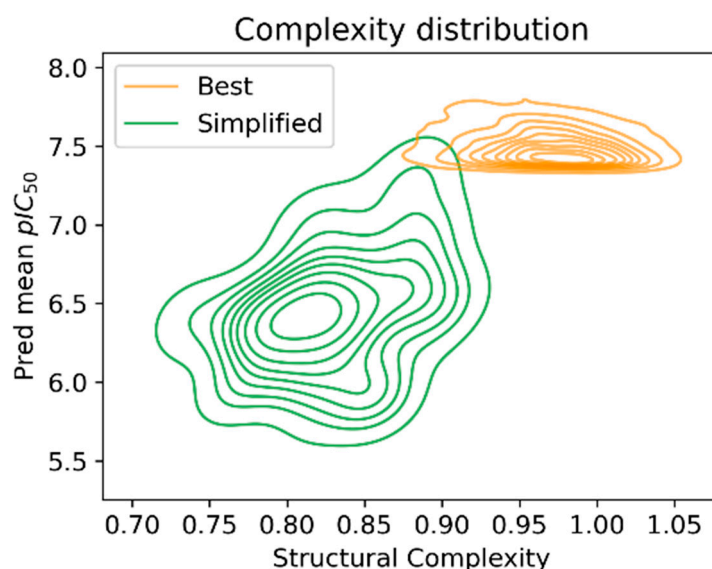
Figure 9. Visual representation of the simplified structure generation.

In this way, 94 simplified structures were generated. While analyzing the number of molecules and substituents throughout the whole process, the initial molecule number was reduced by 99.98%, and, among the different substituents generating the combinatorial database, only 25% were considered privileged (Table 2).

Table 2. Summary of the number of substituents and molecules in the simplification process.

	Starting Database	SARepor	Simplification
C2 substituents	30	20	5
C4 substituents	14	9	4
C5 substituents	12	10	5
C6 substituents	27	25	7
N8 substituents	5	5	4
TOTAL	680,400	4000	94

To quantify the complexity reduction achieved by the suggested strategy, the complexity parameter defined by DataWarrior [24] was calculated. These values are represented in the kernel density estimation plot (Figure 10), where the reduced structural complexity of the simplified set of molecules is apparent. At the same time, as was expected, this complexity reduction came with a cost in the predicted pIC_{50} ; still, their reduced synthetic paths should compensate for this, as, once their real activity is measured, it would allow us to identify which complex structures are worth investing resources in.

**Figure 10.** Simplification distribution of the simplified structures and the 4000 best predicted ones.

To further analyze these results, the t-distributed stochastic neighbor embedding (TSNE) algorithm was used to plot the chemical spaces of the three molecule sets: the initial combinatorial database, the 4000 best compounds used in the SARepor and the simplified 94 molecules (Figure 11).

It can be seen how most of the 94 simplified structures are located in areas covered by the best 4000 compounds, indicating that they still maintained a favorable activity profile. At the same time, their peripheral location relative to the orange molecules shows also their structural differences.

2.3. Final Candidate Selection

The 94 simplified structures were ranked based on Pareto optimization metrics. This method is widely used for the optimization of multi-objective virtual screening [25], where molecules that form or are close to the Pareto front receive higher scores, balancing the predicted value against each target.

Based on this, the highest-scoring 17 compounds were selected, taking into account the privileged substituent diversity, thus selecting as many different substitutions as possible (Table 3).

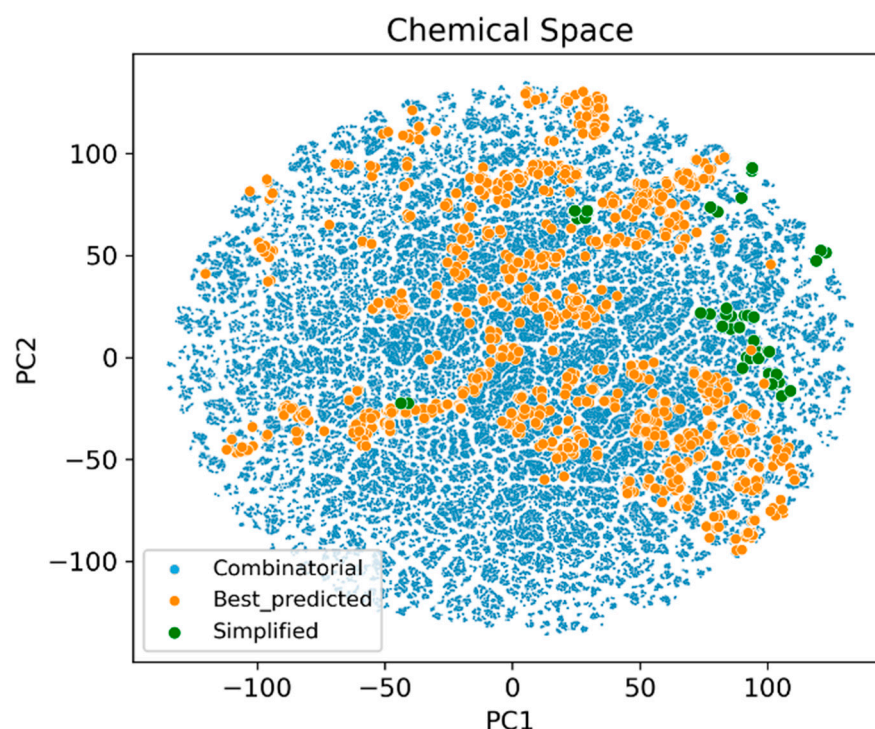


Figure 11. TSNE plot of the different molecule sets used in the structural simplification process.

Table 3. Summary of the selected candidates after the structure simplification process. The last two columns show their ranks among the final 94 simplified compounds and the original combinatorial database based on their mean predicted pIC_{50} .

Name	IGF1R	FGFR2	Mean	Pareto Value	Rank Combinatorial
IQS227	7.81	7.08	7.44	92	1978
IQS228	7.55	6.96	7.25	71	5479
IQS229	7.53	6.95	7.24	68	5943
IQS236	7.11	6.95	7.03	63	22,997
IQS222	6.12	7.11	6.61	59	214,246
IQS226	6.46	6.84	6.65	48	110,333
IQS231	6.82	6.76	6.79	40	49,587
IQS232	6.02	6.86	6.44	39	258,649
IQS230	6.40	6.77	6.58	37	125,595
IQS240	6.01	6.83	6.42	36	263,397
IQS237	6.23	6.78	6.51	34	174,159
IQS224	7.48	6.66	7.07	30	6987
IQS235	6.94	6.66	6.80	27	36,502
IQS233	6.25	6.69	6.47	26	168,728
IQS223	6.49	6.58	6.49	21	104,524
IQS234	6.38	6.49	6.43	13	128,756

From this compound selection, it could be seen that the SAR-based structural simplification led us towards structures that would not have been selected otherwise, due to them having too low activity or being hidden behind a huge database, as none of them were among the best 1000 compounds.

For the external validation of the hypothesis, the two best-predicted non-simplified compounds were selected (Table 4).

Table 4. Predicted pIC₅₀ values of the two best non-simplified compounds.

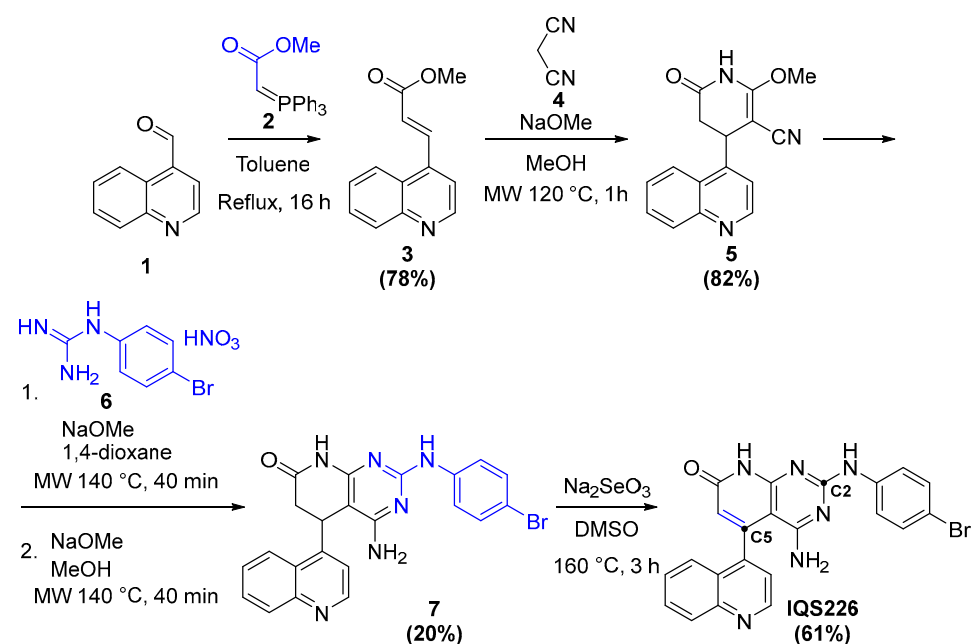
Name	IGF1R	FGFR2	Mean	Rank Combinatorial
IQS238	8.79	7.23	8.01	1
IQS239	8.82	7.17	8.00	2

2.4. Synthesis of the Candidates

The chosen candidates were synthesized following different synthetic paths, depending on the substituted position of the pyrido[2,3-*d*]pyrimidines, based on methodologies developed in the group. To illustrate this, the synthesis of three key examples, IQS224, IQS226 and IQS229, is disclosed.

2.4.1. Synthesis of IQS226

Generally, C2- and C5-substituted compounds begin with the formation of a substituent bearing acrylate **3** by a Wittig reaction; it is then heated with malononitrile (**4**) and NaOMe assisted by microwave irradiation. The formed pyridone **5** is then reacted with a 4-bromophenylguanidine nitrate (**6**), which, over two steps [19], yields the pyrido[2,3-*d*]pyrimidine scaffold **7**. Finally, IQS226 is obtained after the dehydrogenation of **7** and its reaction with Na₂SeO₃ at high temperatures [26] (Figure 12).

**Figure 12.** The followed synthetic path towards IQS226.

2.4.2. Synthesis of IQS229

In the case of C2 and C6 substitution, the synthesis of the common intermediate **10** allowed a much more efficient route towards the candidates. Molecule **10** is obtained by the path shown in Figure 13, where, after the microwave-assisted one-pot formation of the pyrido[2,3-*d*]scaffold **9** [27], its reaction with NIS leads to dehydrogenation and iodination, introducing a key iodine in the C6 position.

From this point, different organometallic reactions allow the late-stage introduction of diversity into the structure [27]. Specifically, the selective palladium-catalyzed Suzuki cross-coupling depicted in Figure 14 gave IQS229 in a single step.

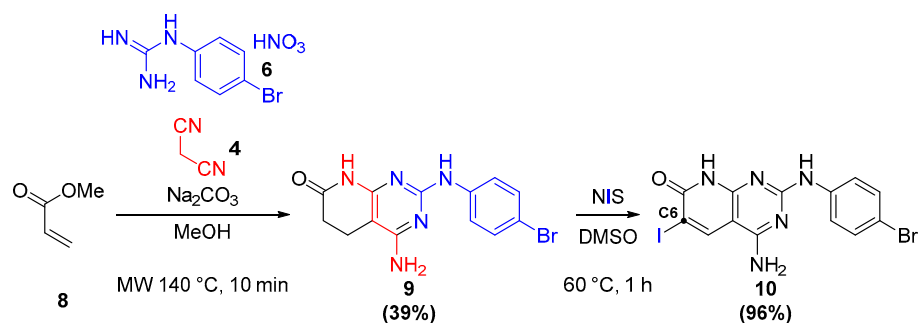


Figure 13. Formation of the common intermediate 10.

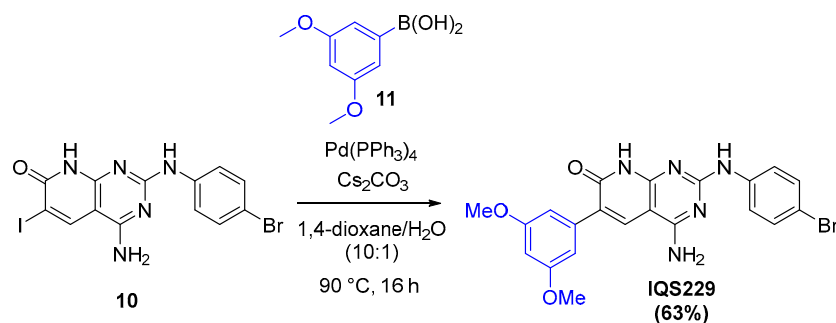


Figure 14. Suzuki cross-coupling reaction for the formation of IQS229.

2.4.3. Synthesis of IQS224

C2- and C4-substituted structures follow a similar strategy as the one above, where a leaving group is introduced in the C4 position by the reactions shown in Figure 15. Here, the previously obtained 9 would be reacted using Sandmeyer-like conditions, forming compound 13. This would be reacted with triflic anhydride (14), allowing the introduction of the pseudohalogen at the C4 position.

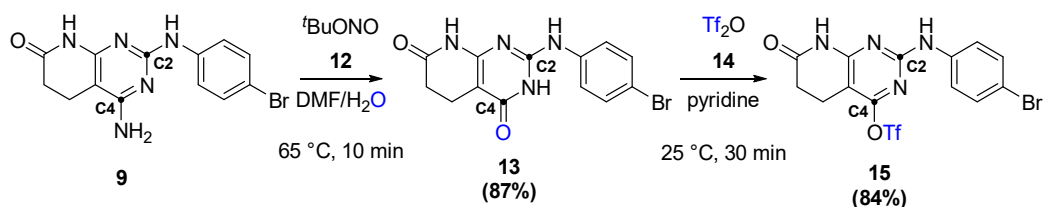


Figure 15. Introduction of a leaving group at the C4 position of the pyrido[2,3-*d*]pyrimidine.

This leaving group would then be substituted by either organometallic reactions [20] or, in the case of IQS224, by a direct S_NAr (Figure 16).

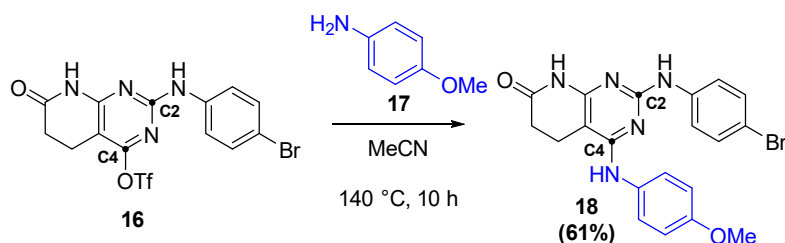


Figure 16. Formation of IQS224 by a S_NAr reaction.

2.5. Biological Evaluation

The validation of the process was performed by measuring each molecule's TKI activity through an enzymatic assay against our targeted proteins (Figure 17).

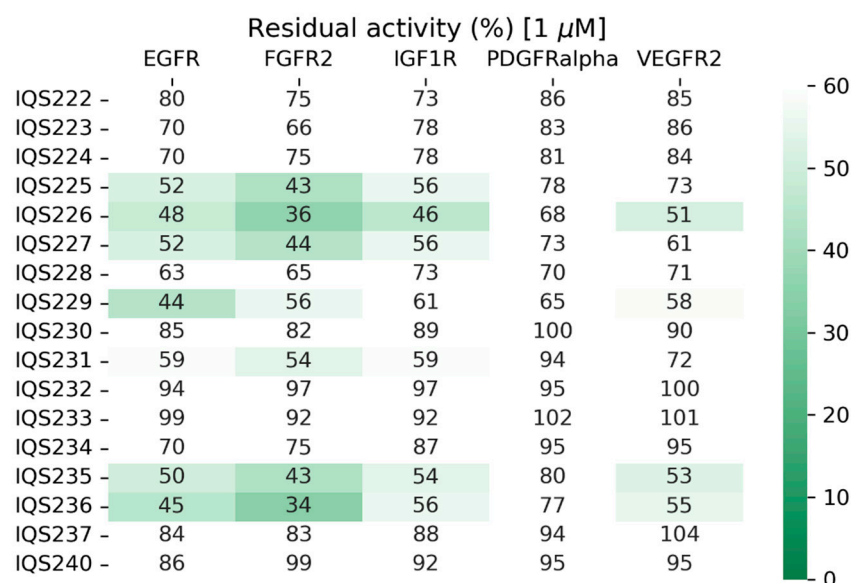


Figure 17. Residual activity percentages of the target proteins when treated with 1 μ M of the simplified inhibitors in an enzymatic assay.

From the 17 synthesized molecules, seven of them inhibited both FGFR2 and IGF1R with an IC_{50} value close to 1 μ M (residual activity below or very close to 50%) and thus were considered acceptable hits. This meant that the success rate of the methodology described was close to 50% in both proteins, making it a very effective method of hit selection.

In addition, in the case of IGF1R, the predicted and measured activity values achieved high Spearman correlation values ($\rho = 0.64$), further supporting the applicability of the methodology.

Both IQS226 and IQS229, disclosed previously, are among the seven active candidates (Figure 18).

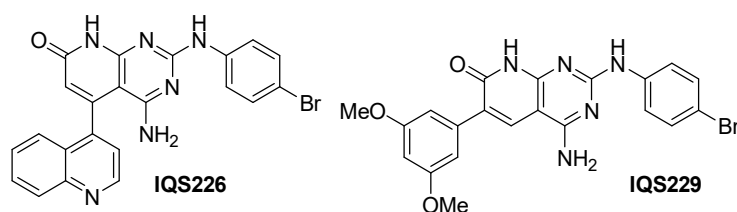


Figure 18. Example structures of the simplified active candidates.

To further validate the hypothesis of this work, two non-simplified candidates (Figure 19) were synthesized following non-published methodologies.

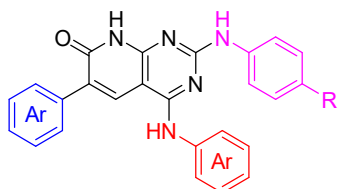


Figure 19. General structures of the non-simplified candidates. Each structure has a molecular weight above 570 Da.

These candidates (IQS238 and IQS239) showed very similar activity (Figure 20) while requiring up to eight synthetic steps to obtain and having more limited structural diversity.

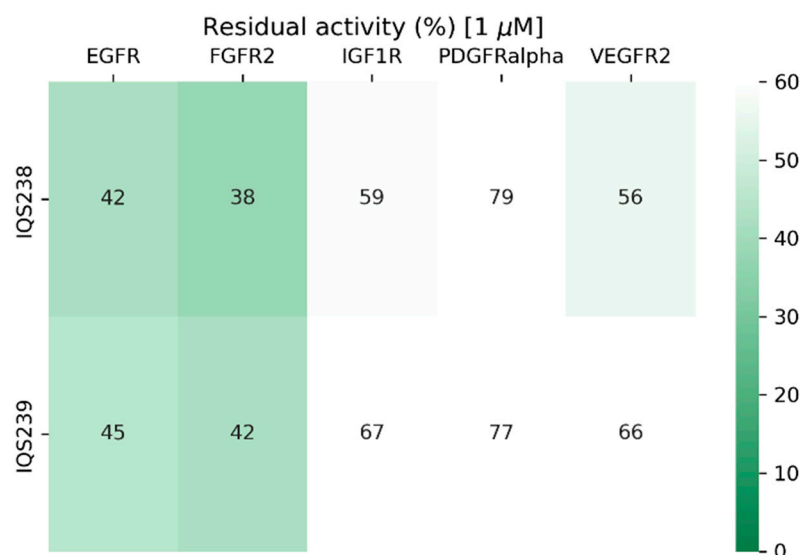


Figure 20. Residual activity percentages of the target proteins when treated with 1 μ M of complex inhibitor (IQS238 and IQS239) in an enzymatic assay.

Considering these results, in an early hit discovery scenario, synthesizing a diverse group of molecules is preferred even if the success rate is lower, as both active and inactive molecules will serve as information for more active structures.

3. Computational Methodology

3.1. Datasets

QSAR models were trained using the previously curated ChEMBL database [28] ($n_{\text{FGFR2}} = 687$, $n_{\text{IGF1R}} = 816$), with known pIC_{50} values (measured in single protein format assays), where duplicate information and non-exact or missing activity values were removed.

Smilib v2.0 [29] was used to create the combinatorial library used as a test set. Independent SMILES libraries for the different substituents were created considering the synthetic feasibility of each independent substitution, the structural diversity of the groups and the availability of reagents (the list of the different substituents is available in the Supplementary Materials).

3.2. Feature Selection

All structural-related calculations were carried out using the Molecular Operating Environment (MOE) [30]. All 2D molecular descriptors ($n = 206$) were calculated and used as starting features.

The in-house feature selection method consisted of merging three well-described feature selection methods, compensating for each one's limitations while yielding an efficient and effective new hybrid feature selection methodology.

The method consisted of correlation-based feature elimination (CFE), which reduced the initial set of features based on their correlations with the pIC_{50} value, removing those with a correlation below 0.3. The remaining features were further reduced by recursive feature elimination (RFE), where all those features whose elimination would increase the model's performance were removed. Finally, all removed features were retested using recursive feature addition (RFA), where each was added to the remaining set of features, maintaining those that improved the model's performance (Figure 21).

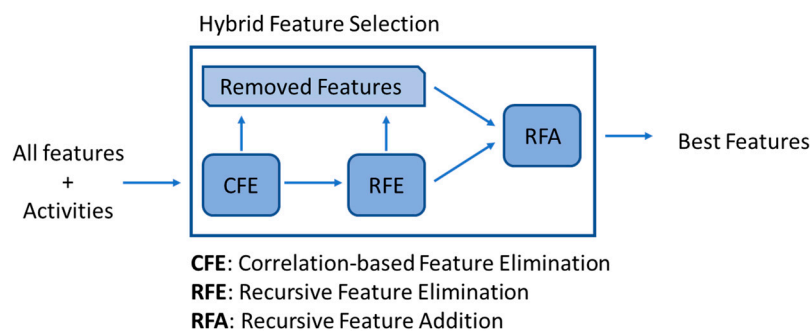


Figure 21. In-house feature selection methodology, where CFE, RFE and RFA are combined.

This method was able to outperform the three algorithms that it was based on in terms of feature reduction, the obtained precision and the calculation times.

3.3. QSAR Model Creation Methodology

The quantitative structure–activity relationship (QSAR) prediction models were generated using Python’s Scikit-Learn library [31], where the support vector machines (SVM) algorithm was chosen based on previous group experience.

Its hyperparameters were selected using grid search optimization, using the selected features (Supplementary Materials).

The models’ precision was measured using leave one out cross-validation (LOO-CV) and validated using 5-fold and 10-fold cross-validation (CV5 and CV10) and y-randomization (Supplementary Materials).

3.4. Structural Simplification Method

The molecules with the best predicted pIC_{50} values were simplified following a multistep method, where an automatized structure–activity relationship (SAR) analysis tool (SARreport, available in MOE 2022 [30]) allowed us to highlight the structural decorations responsible for the predicted activity, which served as the basis for the generation of a new group of structures bearing fewer substituents. Among these, those with the best predicted activity were selected, yielding a selection of simplified molecules (Figure 22).

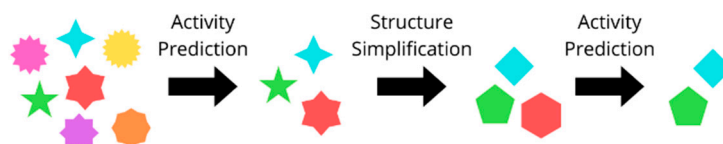


Figure 22. Visual representation of the structure simplification procedure. Each polyhedron depicts a molecule, where a reduction in vertex exemplifies the simplification process.

4. Conclusions

Starting with a combinatorial library containing close to 700,000 structures, we developed a methodology that allowed the rational selection of 17 simplified molecules. Compared to non-simplified candidates, these were obtained using divergent and efficient synthetic methodologies, yielding a moderate number of molecules in a fast manner, while maintaining a high rate of success. The proposed methodology was validated using two well-established kinases, FGFR2 and IGF1R, whose roles have been linked in various pathological conditions (e.g., pancreatic ductal adenocarcinoma, PDAC). This allowed us to discover seven novel active inhibitors, while uncovering various privileged functional groups, such as the disclosed 2,4-methoxyphenyl group in IQS229, key in the discovery of Futibatinib, a commercial FGFR inhibitor [32]. These results not only show the efficacy of this structural simplification methodology but also demonstrate its potential to be a

standard practice in hit discovery programs, as it allows the fast identification and development of bioactive compounds. At the same time, the fact that it was used to develop multitarget inhibitors shows the process' flexibility, as it could be applied to diverse drug design campaigns.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26094457/s1> [33,34].

Author Contributions: Conceptualization, R.E.-T. and E.T.-U.; methodology, R.P.d.I.B., R.E.-T. and E.T.-U.; software, E.T.-U.; formal analysis, R.E.-T. and E.T.-U.; investigation, E.T.-U.; writing—original draft preparation, E.T.-U.; writing—review and editing, J.I.B., R.P.d.I.B., R.E.-T. and E.T.-U.; supervision, R.P.d.I.B. and R.E.-T.; project administration, J.I.B. and R.E.-T.; funding acquisition, J.I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Government, Ministerio de Ciencia e Innovación, grant number RTI2018-096455-B-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy.

Acknowledgments: We gratefully acknowledge the technical support of Margalida Artigues Cladera, IQS-SCIEX Demo lab. E.T. acknowledges the Joan Oró grant (2023 FI-3 00106) funded by the Departament de Recerca i Universitats of the Generalitat de Catalunya and by Fons Socials Europeus.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CFE	Correlation-Based Feature Elimination
CV	Cross-Validation
DEL	DNA-Encoded Libraries
DMF	Dimethylformamide
FBLG	Fragment-Based Lead Generation
FGFR2	Fibroblast Growth Factor Receptor 2
FS	Focused Screening
HTS	High-Throughput Screening
IGF1R	Insulin-Like Growth Factor 1 Receptor
KBLG	Knowledge-Based Lead Generation
LBDD	Ligand-Based Drug Design
LOO	Leave One Out
NIS	<i>N</i> -Iodosuccinimide
MOE	Molecular Operating Environment
MW	Microwave
PDAC	Pancreatic Ductal Adenocarcinoma
QSAR	Quantitative Structure–Activity Relationship
RFA	Recursive Feature Addition
RFE	Recursive Feature Elimination
SAR	Structure–Activity Relationship
SBDD	Structure-Based Drug Design
SVM	Support Vector Machines
TKI	Tyrosine Kinase Inhibitor

References

1. Blay, V.; Tolani, B.; Ho, S.P.; Arkin, M.R. High-Throughput Screening: Today's Biochemical and Cell-Based Approaches. *Drug Discov. Today* **2020**, *25*, 1807–1821. [\[CrossRef\]](#)
2. Moitessier, N.; Pottel, J.; Therrien, E.; Englebienne, P.; Liu, Z.; Tomberg, A.; Corbeil, C.R. Medicinal Chemistry Projects Requiring Imaginative Structure-Based Drug Design Methods. *Acc. Chem. Res.* **2016**, *49*, 1646–1657. [\[CrossRef\]](#)
3. Erlanson, D.A.; Fesik, S.W.; Hubbard, R.E.; Jahnke, W.; Jhoti, H. Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* **2016**, *15*, 605–619. [\[CrossRef\]](#)
4. Jiang, Y.; Liu, Z.; Holenz, J.; Yang, H. Competitive Intelligence-Based Lead Generation and Fast Follower Approaches. In *Lead Generation*; Holenz, J., Ed.; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; pp. 183–220.
5. Brown, D.G.; Boström, J. Where Do Recent Small Molecule Clinical Development Candidates Come From? *J. Med. Chem.* **2018**, *61*, 9442–9468. [\[CrossRef\]](#)
6. Hughes, J.P.; Rees, S.S.; Kalindjian, S.B.; Philpott, K.L. Principles of Early Drug Discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249. [\[CrossRef\]](#)
7. Ajjarapu, S.M.; Tiwari, A.; Ramteke, P.W.; Singh, D.B.; Kumar, S. Ligand-Based Drug Designing. In *Bioinformatics: Methods and Applications*; Academic Press: Cambridge, MA, USA, 2022; pp. 233–252, ISBN 9780323897754.
8. Welsch, M.E.; Snyder, S.A.; Stockwell, B.R. Privileged Scaffolds for Library Design and Drug Discovery. *Curr. Opin. Chem. Biol.* **2010**, *14*, 347–361. [\[CrossRef\]](#)
9. Méndez-Lucio, O.; Medina-Franco, J.L. The Many Roles of Molecular Complexity in Drug Discovery. *Drug Discov. Today* **2017**, *22*, 120–126. [\[CrossRef\]](#)
10. Matsuda, Y.; Yoshimura, H.; Suzuki, T.; Uchida, E.; Naito, Z.; Ishiwata, T. Inhibition of Fibroblast Growth Factor Receptor 2 Attenuates Proliferation and Invasion of Pancreatic Cancer. *Cancer Sci.* **2014**, *105*, 1212–1219. [\[CrossRef\]](#)
11. Urtasun, N.; Vidal-Pla, A.; Pérez-Torras, S.; Mazo, A. Human Pancreatic Cancer Stem Cells Are Sensitive to Dual Inhibition of IGF-IR and ErbB Receptors. *BMC Cancer* **2015**, *15*, 223. [\[CrossRef\]](#)
12. Raghavendra, N.M.; Pingili, D.; Kadasi, S.; Mettu, A.; Prasad, S.V.U.M. Dual or Multi-Targeting Inhibitors: The next Generation Anticancer Agents. *Eur. J. Med. Chem.* **2018**, *143*, 1277–1300. [\[CrossRef\]](#)
13. Holohan, C.; Van Schaeybroeck, S.; Longley, D.B.; Johnston, P.G. Cancer Drug Resistance: An Evolving Paradigm. *Nat. Rev. Cancer* **2013**, *13*, 714–726. [\[CrossRef\]](#)
14. Chand, S.; O'Hayer, K.; Blanco, F.F.; Winter, J.M.; Brody, J.R. The Landscape of Pancreatic Cancer Therapeutic Resistance Mechanisms. *Int. J. Biol. Sci.* **2016**, *12*, 273–283. [\[CrossRef\]](#)
15. Kammasud, N.; Boonyarat, C.; Sanphanya, K.; Utsintong, M.; Tsunoda, S.; Sakurai, H.; Saiki, I.; André, I.; Grierson, D.S.; Vajragupta, O. 5-Substituted Pyrido[2,3-d]Pyrimidine, an Inhibitor against Three Receptor Tyrosine Kinases. *Bioorg Med. Chem. Lett.* **2009**, *19*, 745–750. [\[CrossRef\]](#)
16. Edupuganti, R.; Wang, Q.; Tavares, C.D.J.; Chitjian, C.A.; Bachman, J.L.; Ren, P.; Anslyn, E.V.; Dalby, K.N. Synthesis and Biological Evaluation of Pyrido[2,3-d]Pyrimidine-2,4-Dione Derivatives as EEF-2K Inhibitors. *Bioorg. Med. Chem.* **2014**, *22*, 4910–4916. [\[CrossRef\]](#)
17. Elzahabi, H.S.A.; Nossier, E.S.; Khalifa, N.M.; Alasfoury, R.A.; El-Manawaty, M.A. Anticancer Evaluation and Molecular Modeling of Multi-Targeted Kinase Inhibitors Based Pyrido[2,3-d]Pyrimidine Scaffold. *J. Enzyme Inhib. Med. Chem.* **2018**, *33*, 546–557. [\[CrossRef\]](#)
18. Anderson, K.; Chen, Y.; Chen, Z.; Dominique, R.; Glenn, K.; He, Y.; Janson, C.; Luk, K.C.; Lukacs, C.; Polonskaia, A.; et al. Pyrido[2,3-d]Pyrimidines: Discovery and Preliminary SAR of a Novel Series of DYRK1B and DYRK1A Inhibitors. *Bioorg Med. Chem. Lett.* **2013**, *23*, 6610–6615. [\[CrossRef\]](#)
19. Puig De La Bellacasa, R.; Roué, G.; Balsas, P.; Pérez-Galán, P.; Teixidó, J.; Colomer, D.; Borrell, J.I. 4-Amino-2-Arylamino-6-(2,6-Dichlorophenyl)-Pyrido[2,3-d]Pyrimidin-7-(8H)-Ones as BCR Kinase Inhibitors for B Lymphoid Malignancies. *Eur. J. Med. Chem.* **2014**, *86*, 664–675. [\[CrossRef\]](#)
20. Masip, V.; Lirio, Á.; Sánchez-López, A.; Cuenca, A.B.; Puig De La Bellacasa, R.; Abrisqueta, P.; Teixidó, J.; Borrell, J.I.; Gibert, A.; Estrada-Tejedor, R.; et al. Expanding the Diversity at the C-4 Position of Pyrido[2,3-d]Pyrimidin-7(8H)-Ones to Achieve Biological Activity against ZAP-70. *Pharmaceuticals* **2021**, *14*, 1311. [\[CrossRef\]](#)
21. Camarasa, M.; De La Bellacasa, R.P.; González, À.L.; Ondoño, R.; Estrada, R.; Franco, S.; Badia, R.; Esté, J.; Martínez, M.Á.; Teixidó, J.; et al. Design, Synthesis and Biological Evaluation of Pyrido[2,3-d]Pyrimidin-7-(8H)-Ones as HCV Inhibitors. *Eur. J. Med. Chem.* **2016**, *115*, 463–483. [\[CrossRef\]](#)
22. Oprea, T.I.; Bologa, C. Molecular Complexity: You Know It When You See It. *J. Med. Chem.* **2023**, *66*, 12710–12714. [\[CrossRef\]](#)
23. Guha, R. On Exploring Structure-Activity Relationships. In *Methods in Molecular Biology*; Humana Press Inc.: Totowa, NJ, USA, 2013; Volume 993, pp. 81–94, ISBN 9781627033411.
24. Sander, T.; Freyss, J.; Von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program for Chemistry Aware Data Visualization and Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. [\[CrossRef\]](#)

25. Fromer, J.C.; Graff, D.E.; Coley, C.W. Pareto Optimization to Accelerate Multi-Objective Virtual Screening. *Digit. Discov.* **2024**, *3*, 467–481. [[CrossRef](#)]
26. Perez-Pi, I.; Berzonsa, X.; Galve, I.; Teixido, J.; Borrel, J.I. Dehydrogenation of 5,6-Dihydropyrido[2,3-d]Pyridimin-7(8H)-Ones: A Convenient Last Step for a Synthesis of Pyrido[2,3-d]Pyrimidin-7(8H)-Ones. *Heterocycles* **2010**, *82*, 581–591.
27. Galve, I.; Ondoño, R.; De Rocafiguera, C.; Puig De La Bellacasa, R.; Batllori, X.; Puigjaner, C.; Font-Bardia, M.; Vallcorba, O.; Teixidó, J.; Borrell, J.I. A Captured Room Temperature Stable Wheland Intermediate as a Key Structure for the Orthogonal Decoration of 4-Amino-Pyrido[2,3-d]Pyrimidin-7(8H)-Ones. *Org. Biomol. Chem.* **2020**, *18*, 9810–9815. [[CrossRef](#)]
28. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [[CrossRef](#)]
29. Schüller, A.; Hähnke, V.; Schneider, G. SMIlib v2.0: A Java-Based Tool for Rapid Combinatorial Library Enumeration. *QSAR Comb. Sci.* **2007**, *26*, 407–410. [[CrossRef](#)]
30. Chemical Computing Group ULC. *Molecular Operating Environment (MOE)*, 2022.0; Chemical Computing Group ULC: Montreal, QC, Canada, 2022.
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
32. Ito, S.; Otsuki, S.; Ohsawa, H.; Hirano, A.; Kazuno, H.; Yamashita, S.; Egami, K.; Shibata, Y.; Yamamiya, I.; Yamashita, F.; et al. Discovery of Futibatinib: The First Covalent FGFR Kinase Inhibitor in Clinical Use. *ACS Med. Chem. Lett.* **2023**, *14*, 396–404. [[CrossRef](#)]
33. Boers, R.B.; Randulfe, Y.P.; Van Der Haas, H.N.S.; Van Rossum-Baan, M.; Lugtenburg, J. Synthesis and Spectroscopic Characterization of 1-¹³C- and 4-¹³C-Plastoquinone-9. *Eur. J. Org. Chem.* **2002**, *2002*, 2094–2108. [[CrossRef](#)]
34. El-Batta, A.; Jiang, C.; Zhao, W.; Anness, R.; Cooksy, A.L.; Bergdahl, M. Wittig Reactions in Water Media Employing Stabilized Ylides with Aldehydes. Synthesis of α,β -Unsaturated Esters from Mixing Aldehydes, α -Bromoesters, and Ph_3P in Aqueous NaHCO_3 . *J. Org. Chem.* **2007**, *72*, 5244–5259. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.