



Article

Frag2Drugs: A Novel In Silico Fragment-Based Approach to the Discovery of Kinase Inhibitors

Gautier Peyrat [†], Colin Bournez [†], Pascal Krezel, José-Manuel Gally, Stéphane Bourg , Samia Aci-Sèche ^{*} and Pascal Bonnet ^{*}

Université d'Orléans, CNRS, ICOA, UMR 7311, Orléans, France; stephane.bourg@cnrs.fr (S.B.)

^{*} Correspondence: samia.aci-seche@univ-orleans.fr (S.A.-S.); pascal.bonnet@univ-orleans.fr (P.B.);

Tel.: +33-2-38-41-99-02 (S.A.-S.); +33-2-38-41-72-54 (P.B.)

[†] These authors contributed equally to this work.

Abstract

Background/Objectives: Fragment-based approaches in the field of drug discovery and design have been widely developed and employed in both academia and industry. We present here an innovative in silico fragment-based drug design approach aimed at designing new inhibitors in the ATP-binding site of protein kinases. **Methods:** This tool, named Frags2Drugs (F2D), relies on a three-dimensional fragment library obtained from co-crystallized ligands. This library is stored in a graph-oriented database containing the required information to link fragments together. F2D builds every possible molecule that fits into the given cavity on a minute scale. Molecules are then filtered to keep those presenting the best predicted affinity. Several specific molecular filters can be applied, including protein kinase inhibitor-like filters. **Results:** We validated our method by reconstructing existing co-crystallized ligands and known kinase inhibitors. In this study, we provide several examples of its use to retrieve known or design new type I, type II/2, type II, and macrocyclic inhibitors on several protein kinases. **Conclusions:** We have developed an in silico fragment-based ligand design tool able to identify novel kinase inhibitors by growing any scaffolds positioned in the ATP-binding site..

Keywords: kinase research; fragment growing; fragment-based ligand design (FBLD); network graph; protein kinase inhibitor



Academic Editors: Alessia Bono and Antonino Lauria

Received: 15 January 2026

Revised: 6 February 2026

Accepted: 10 February 2026

Published: 12 February 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In recent decades, fragment-based drug design (FBDD) has emerged to become a major approach in the discovery of new chemical compounds [1]. The first success from FBDD dates from 2011, with U.S. Food and Drug Administration (FDA) approval of vemurafenib (Zelboraf), a drug targeting the mutated protein kinase BRAF V600E [2]. Nowadays, several drugs approved or in advanced clinical trials have been discovered by using FBDD approaches [3,4]. Recent examples are erdafetinib (Balversa) [5], developed by Janssen and Astex and approved by the FDA in 2019; capivasertib (Truqap), an AKT kinase inhibitor approved in 2023 and developed by AstraZeneca; and sevabertinib, the last approved drug targeting the kinase domain of ERBB2.

Fragments are an interesting alternative to traditional compounds used in high-throughput screenings: they are smaller (usually < 300 Da) and cover a larger chemical space [6]. Once identified during a screening campaign, fragment hits are optimized by

medicinal chemists using, when available, structural approaches such as X-ray crystallography or Nuclear Magnetic Resonance (NMR) spectroscopy. Computational chemistry may also play a substantial role in this process. Multiple methods have been recently developed in the field of *in silico* FBDD [7] that can be classified into three main categories: growing, linking, and merging.

Growing methods are the most commonly used, by adding new fragments sequentially from an initial fragment, usually called the seed. Software programs such as AutoGrow [8–10], FOG [11], and Open-Growth [12] are based on this strategy. Other tools use atoms rather than fragments to optimize the seed, such as LEGEND [13] and Genstar [14]. However, atom-based methods remain less employed nowadays. The growth is usually performed in the binding site, using the three-dimensional coordinates of the original hit found experimentally or by docking methods. Available space surrounding the attachment atom and physicochemical properties of the residues in the binding site are important features for the optimization of the seed.

The second strategy, fragment linking, starts from multiple fragments rather than from a unique seed, with the aim of connecting them together directly or using a linker. This method is equivalent to the experimental Structure Activity Relationships (SAR) by NMR approach [15]. The protein target needs to present distinct pockets in its binding site, allowing the starting fragments to bind at different key interaction sites. The fragment linking method is notably employed in the tools LUDI [16], LEA3D [17], GANDI [18], and Fragment-Shuffling [19]. More recently, DeLinker [20], a deep learning tool, was developed to generate linkers between fragments using three-dimensional information.

The third and perhaps less employed strategy is fragment merging. It relies on the superimposition of multiple fragments presenting overlapping structural groups. Common structural features of the fragments are combined to create novel inhibitors.

One of the main challenges of computational FBDD is to design synthesizable compounds. Therefore, some software applications, such as LigBuilder [21] or PINGUI [22], include an estimation of the synthetic accessibility in their algorithms with the help of chemical rules applied during the growth and/or with retrosynthetic analysis.

Most of the FBDD programs spend a major part of their calculation time in the systematic exploration of the available space in the binding site, as well as in energy minimization of the generated compounds to find an equilibrium position. At the same time, the number of experimental protein structures with co-crystallized ligand is continuously expanding, providing more available structural information and data to be harnessed. From this observation, it appears relevant to develop a method exploiting these data so that the time dedicated to cavity exploration and energy minimization is considerably reduced. Frags2Drugs (F2D), presented in this study, uses fragments from co-crystallized ligands or from docking experiments and searches for possible combinations between them to construct new molecules. F2D is a method based on both fragment growing and linking strategies. The growing is performed by linking fragments already in the binding site and placing them close to each other. F2D requires 3D fragments, generated from the fragmentation of crystal structures of ligands, in close vicinity, to create covalent bonds while keeping their initial position. During this work, we were particularly interested in a specific class of protein, named protein kinases, having common ATP binding sites. Protein kinases are enzymes that catalyze the transfer of a γ -phosphate group of ATP to a protein substrate or the same protein during autophosphorylation. They are critical for intracellular transduction, and deregulation of a member of this protein family may lead to diverse diseases, including cancer, autoimmune, or neurological troubles [23]. Their determining role in transduction pathways continues to make kinases targets of major interest for pharmaceutical industries now and over the last decades [24,25]. From a structural point of

view, protein kinases share a common catalytic domain composed of two distinct lobes: an N-terminal lobe (N-lobe) and a C-terminal lobe (C-lobe) connected by a hinge region. The cleft between these two lobes is the binding site of the cofactor, ATP, which forms hydrogen bonds with the hinge region, stabilizing its position in the catalytic site (Figure 1). Most kinase inhibitors in development or already on the market target the ATP binding site [26], but other druggable pockets are known, such as the allosteric site or the back pocket [27]. For computational chemists working on protein kinases, the wealth of information coming from structural data is a strong opportunity: 7786 structures of human kinase domains are deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein DataBank [28] (PDB, <https://www.rcsb.org/>, accessed on 15 December 2025).

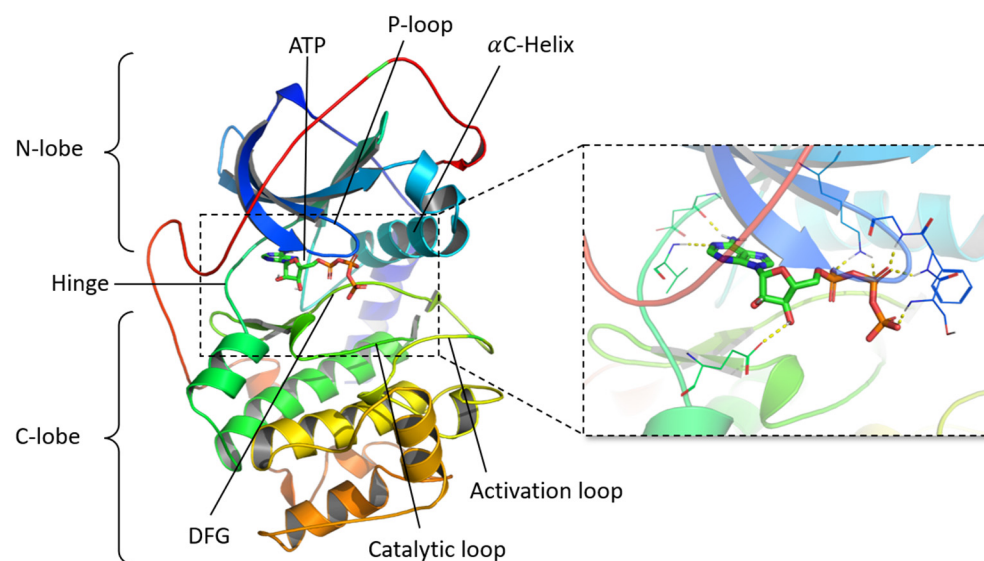


Figure 1. Cartoon representation of a protein kinase crystallographic structure (PDB ID 1ATP, chain E) with a focus on the ATP binding site, highlighting the interactions between ATP and the residues of the catalytic site (yellow dotted lines). N-lobe: N-terminal lobe; C-lobe: C-terminal lobe; ATP: Adenosine triphosphate; P-loop: phosphate-binding loop; DFG: Asp-Phe-Gly.

Protein kinase inhibitors (PKIs) are divided into seven types depending on their binding modes [29]. The first three types are non-covalent ATP competitive inhibitors. Type I inhibitors bind in DFG-in kinase conformation, type I1/2 in DFG-in conformation with access to an adjacent hydrophobic pocket, and type II inhibitors bind kinases in DFG-out conformation, lying from the ATP site to the adjacent allosteric pocket. Other types are allosteric inhibitors (types III and IV), bivalent inhibitors (type V), and covalent inhibitors (type VI).

2. Results

2.1. F2D Implementation

F2D is an FBDD tool relying on 3D experimental data. Indeed, fragments obtained from 3D crystallographic structures of protein kinase-ligand complexes fill the database used by F2D to generate novel ligands.

To create the fragment library used in F2D, we first extracted protein kinase crystallographic structures available from the RCSB protein databank by only keeping the kinase domain. We then performed a customized structural alignment of all crystal structures on the reference structure: the first crystallized structure of the protein kinase bound with ATP [30] (PDB ID 1ATP, chain E, Figure 2). Once aligned, structures were cleaned to remove ions, co-factors, crystallization agents [31], and solvent molecules. Ligands were afterward

extracted and fragmented using fragmentation tools (see Materials and Methods) while keeping their 3D coordinates, so their placements in the aligned active site were kept and recorded in a database.

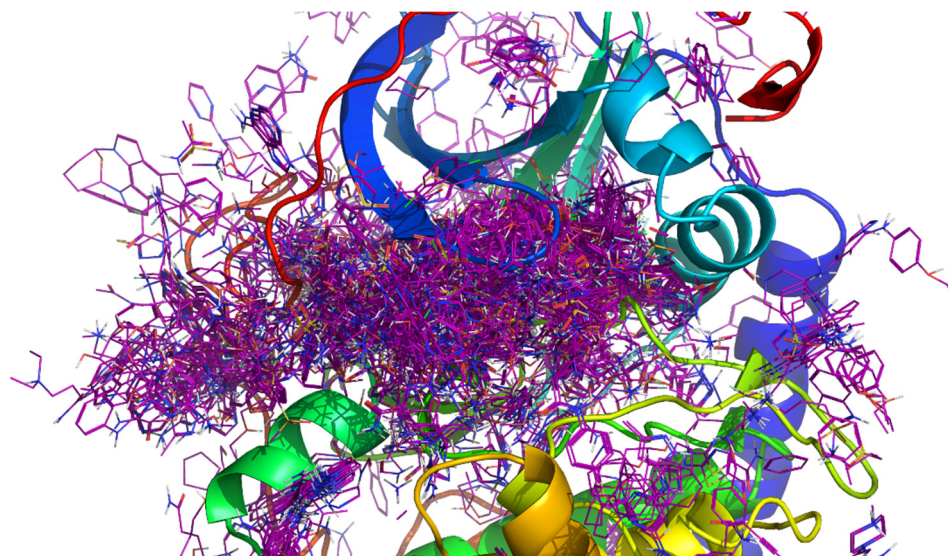


Figure 2. Graphical representation of all fragments after superimposition of all protein kinases on a common reference (PDB ID 1ATP, chain E). Fragments are colored in purple.

A recent study applied a similar approach to the one used for the creation of our F2D database. The Kinase–Ligand Interaction Fingerprints and Structure database (KLIFS) is composed of an alignment of 85 key residues from kinase binding sites [32]. Based on fragments from KLIFS, co-crystallized ligands were collected and fragmented. This collection of fragments, called KinFragLib [33], has been used to generate novel kinase inhibitors. Because it is based solely on KLIFS, this work is limited to kinase–ligand complexes only from human and mouse species. In addition, fragments obtained in KinFragLib are restricted to only type I and I1/2 inhibitors. In F2D, we extended the database to all species available in the RSCB PDB, which allowed the creation of type II inhibitors in addition to types I and I1/2.

The aim of F2D is to link several fragments together in a target protein kinase to create potential new bioactive molecules. The structure of the protein kinase used by F2D could be a holo form where the co-crystallized ligand is removed, an apo form for which no co-crystallized ligand was identified, or a homology model of an unknown crystal structure of a protein kinase. We decided to store each fragment in a graph database along with its 3D position and the information of the neighboring fragments. Indeed, a graph network is particularly suited to store information on connectable entities and, thus, perfectly fulfills the aim of F2D with linkable fragments. By browsing a graph of connected fragments, those fragments can be linked together to form a possible PKI.

2.2. Creation of the 3D Fragment Network

Fragments are represented by nodes in a 3D fragment network and are linked by edges. The graph database of F2D is composed of two types of nodes (fragment and protein) linked by three types of edges: inclusion, exclusion, and compatibility. A sample of the F2D graph database with only nodes and edges from the PDB ID 3OG7 and vemurafenib fragments is shown in Figure 3.

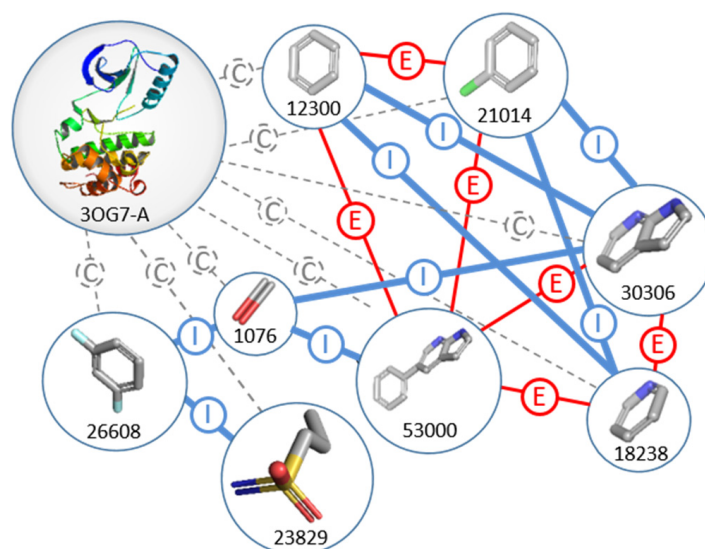


Figure 3. Representation of the F2D graph database, focused on fragments obtained after fragmentation of vemurafenib. Nodes are 3D fragment structures and 3D target proteins. Edges are colored according to their type: exclusion (E in red), inclusion (I in blue), and compatibility (C in gray). Fragment IDs are mentioned.

The first step to create the graph database consists of calculating the relations between all the fragments. These relations belong to two types for ligand information only: inclusion or exclusion, and one type for protein and ligand information: compatibility. An inclusion relation means that the two considered fragments satisfy all the following conditions to bind together: an atom pair exists, one from each fragment, having incomplete valences and separated by a distance that meets a given allowed interval. Moreover, the out-of-plane (OOP), dihedral, and torsion angles with the direct neighbors also meet predefined allowed intervals. The definition of allowed intervals is based on MMFF94 [34–38] force field values by taking into account the type and the hybridization of both atoms involved in the link.

If the two atoms of each fragment are too close or if there is a non-realistic angle value between them, an exclusion relation is established between them. It means they cannot be linked. Furthermore, they cannot belong to the same built molecule, even if they have inclusion relations with a common intermediate fragment. If the two atoms are too far from each other, no relation is established between them. These rules are summarized in Figure 4. Moreover, to avoid suggesting unfeasible molecules to medicinal chemists, additional connection rules were added, such as prohibiting the creation of undesired substructures, detailed in Materials and Methods.

The second step of the creation of the graph database is the calculation of compatibility relations between the fragments and all the residues in the binding site of the protein kinase target. Fragments are compatible if their atoms fulfill an allowed distance between each atom of the fragment and an atom of the residue of the binding site. When this condition is met, there is no steric hindrance between the fragment and the protein; the fragment is kept; otherwise, it is removed for the fragment linking step.

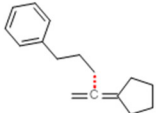
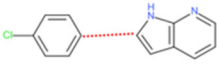
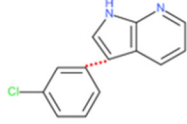
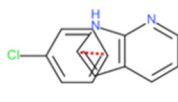
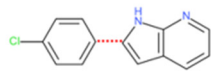
Examples					
Valence	✗	✓	✓	✓	✓
Distance	✓	✗	✓	✓	✓
Angles (torsion, dihedral and OOP)	✓	✓	✗	✓	✓
No steric hindrance	✓	✓	✓	✗	✓
Bond creation	✗	✗	✗	✗	✓

Figure 4. Bond formation conditions between two fragments. Molecules are drawn in 2D for visual purposes, but they are stored in 3D in the F2D database. In each example, the red dash lines represent the potential bond between both fragments. OOP: out of plane. Red crosses correspond to invalid bonds, green ticks correspond to valid bond.

In the graph database, additional information is encoded in nodes and edges. The attributed molecular weight is assigned to fragment nodes and used to stop the addition of new fragments during molecule generation. The inclusion edges, between fragment nodes that can be linked, contain complementary information. When an inclusion relation can be established between two fragments, each linkable atom's information is saved, as well as the type of bond (simple, double, or triple) that will be made. Linking fragments from molecules bound to different protein kinases could result in a little distortion of their angles and distance compared to reference values from the force field MMFF94 [34]. We thus defined four parameters to set the thresholds of allowed distortions: three parameters for angles (dihedral, out of plane, and torsion) and one for distance. The lower the parameter values, the less the generated molecules are distorted.

2.3. Molecular Generation and Filtering Process

F2D molecular generation is initialized with a fragment, referred to as the seed, and a PDB structure or a homology model of the targeted kinase. In a drug design project, a seed can be drawn in 2D by a medicinal chemist, then added, in 3D, to an F2D database after a molecular docking. This is useful when a medicinal chemist is working on a novel scaffold. Starting from the seed, F2D uses only fragments compatible with the chosen protein kinase. Then, F2D searches all possible paths through bonding relations and extracts the corresponding graph. To save time in further treatments, this graph is split into two subgraphs: the Graph of inclusions (G_i) of nodes linked by inclusion relations and the Graph of exclusions (G_e) with only exclusion relations between nodes. In G_i , F2D browses all possible paths starting from the seed. For each obtained path, the tool checks if G_e contains one of its edges (meaning an exclusion relation and therefore an incompatibility for fragments belonging to the same molecule). If so, the corresponding path is discarded. Once all possible paths are obtained and validated, the construction of the corresponding molecules is performed.

In order to build molecules of reasonable size, F2D stops the growing step until a molecular weight (MW) of 650 Da is reached. The value is based on the upper limit of MW published PKI (Table 1).

Table 1. PKI-like physicochemical properties used to filter molecules built by Frags2Drugs (F2D).

	Lower Limit	Upper Limit
MW (Da)	314	613
TPSA ($\text{\AA}^2$)	55	138
ClogP	0.7	6.3
HBA	3	10
HBD	0	4
NRB	1	11
NAR	1	5
NCA	0	2

MW: molecular weight; TPSA: topological polar surface area; HBA: hydrogen bond acceptors; HBD: hydrogen bond donors; NRB: number of rotatable bonds; NAR: number of aromatic rings; NCA: Number of chiral atoms.

At this stage, several thousand molecules can be created, which is too large for all to be synthesized. However, due to the presence of the same fragment belonging to the same ligands but co-crystallized in different crystal structures (like imatinib bound to Abl or Src, PDB ID 2HYX, 2OIQ, 3G6G, and 4CSV), we remove duplicate molecules based on their 3D conformation. We removed molecules having a Root Mean Square Deviation (RMSD) $\leq 1.0 \text{ \AA}$ between two identical created molecules. This also means that several conformations of the same molecule are kept if their RMSD is greater than $1.0 \text{ \AA}$. Next, we implemented additional filters to keep only compounds that fulfill the PKI chemical space. Some filters already exist, especially for oral drug assessment, such as Lipinski's Rule of Five (Ro5) [39]. According to this rule, a compound will probably not be efficiently absorbed orally if there are at least two violations of the following constraints: $\text{MW} \leq 500$, calculated logP (ClogP) ≤ 5 , number of hydrogen bond acceptors (HBA) ≤ 10 , number of hydrogen bond donors (HBD) ≤ 5 . Other physicochemical properties, such as topological polar surface area (TPSA) and number of rotatable bonds (NRB), can also be used to predict a reliable oral bioavailability (TPSA $\leq 140 \text{ \AA}^2$ and NRB ≤ 10) [40]. By analyzing a large collection of PKIs in the protein kinase inhibitor database (PKIDB) [41,42], we identified the upper and lower limits of various molecular physicochemical parameters (Table 1). We also allowed the violation of 2 of the 8 parameters, so molecules created by F2D are kept if they fulfil at least 6 of these parameters.

Compounds with undesired substructures are then filtered out. Once all the PKI-like filters are applied, the results obtained by F2D are downloadable as an SDF file.

In F2D, the molecules are built directly inside the binding site of the protein kinase target. No minimization or conformational optimization is performed during the building process of 3D molecules inside the cavity. Therefore, to make sure that the molecules adopt a reliable conformation, we docked each created molecule in the binding site. After the docking step, the resulting docking poses are ranked in ascending order of their difference in RMSD with the original created molecule. F2D only keeps conformations of the created molecule if the RMSD is less than $2.5 \text{ \AA}$, and the conformation with the lowest RMSD is first provided to the user.

Other post-processing tools were developed to reduce and optimize the number of suggested compounds by F2D, but as they depend on the target to which F2D is applied, they are non-compulsory. For instance, molecules having Pan-Assay Interference Compounds (PAIS) [43] substructures can be identified and removed. To evaluate the synthetic feasibility of created molecules, a synthetic accessibility (SA) score [44] is calculated. Molecules obtaining an SA score < 4 are thus favored. In the same way, the central nervous system multiparameter optimization (CNS MPO) score [45] estimates the ability of molecules to

pass through the blood–brain barrier (BBB). An upper threshold value of 4 is set to keep CNS-like compounds. Finally, as the objectives of F2D are to find active and innovative compounds, we implemented a similarity search to compounds with known bioactivities in the ChEMBL [46] database, and to purchasable compounds available in ZINC [47] or Ambinter [48]. We extended this similarity search to protein kinase inhibitors in clinical trials available in PKIDB. Figure 5 summarizes the general F2D workflow. At the end of the process, an SDF file containing all the information about the origin of each fragment forming each molecule is returned.

After the implementation of the F2D molecule generation process, we validated the F2D methodology.

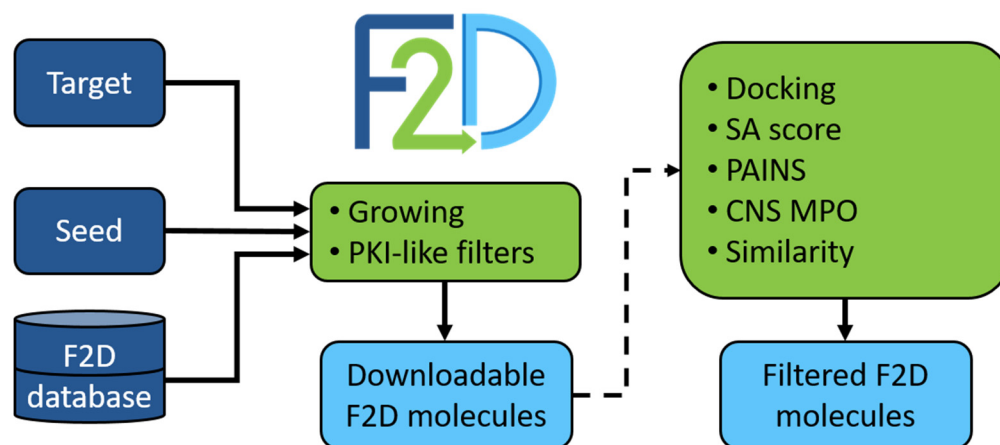


Figure 5. General F2D workflow. Three inputs are required to run F2D: a target and a seed (which can be either picked from the F2D database or a new docked fragment), and the F2D database. Inputs are colored in dark blue, F2D functions are colored in light green, and outputs are colored in light blue. The dotted arrow indicates optional filtering steps.

2.4. Validation of F2D Method

As a first validation step of the F2D methodology, we aimed to build 3218 co-crystallized structures of ligands bound to a protein kinase from the RCSB protein databank. Each ligand was fragmented to create a 3D fragment network. Then, F2D started the growing process from the fragment bound to the hinge region. The tool succeeded in reconstructing 88.13% of the 3218 co-crystallized kinase inhibitors. As an example of F2D successes, Figure 6 shows the reconstruction of vemurafenib starting from the 7-azaindole seed.

A total of 382 ligands could not be reconstructed by the program. However, after in-depth analysis, these failures are not due to the F2D process itself but to the ligand fragmentation. Indeed, we observed in several cases the absence of fragments in the database, required to build the final ligand. Fragmentation methods employed in F2D sometimes lead to fragments with molecular weights > 300 Da, and those fragments are removed from the F2D database.

After this first validation step, we applied F2D to several protein kinase targets to ensure its ability to find innovative and active compounds. Firstly, we searched for BCR-ABL inhibitors, secondly for BRAF V600E inhibitors, thirdly for type I and type II inhibitors of MELK, and, finally, for ALK macrocyclic inhibitors. These four examples of F2D are detailed below.

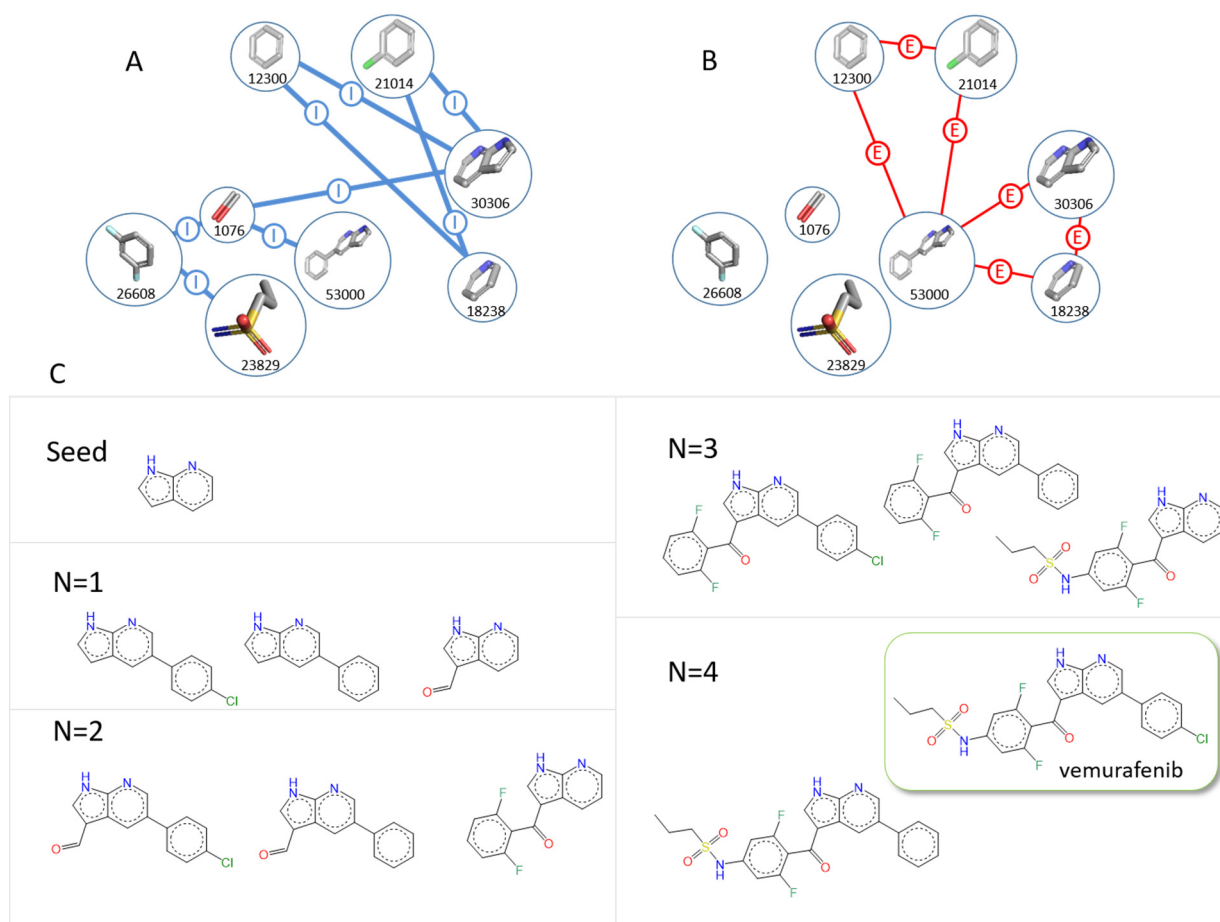


Figure 6. Example of validation of the reconstruction of vemurafenib (PDB ID 3OG7). **(A)** Detailed of a graph of inclusion (Gi) relations between nodes (blue). **(B)** Detailed of a graph of exclusion (Ge) relations between nodes (red). **(C)** Representation of created molecules for each addition (N) of a new fragment, starting from the seed 7-azaindole. For better visualization, molecules are shown in 2D.

2.4.1. Discovery of BCR-ABL Inhibitors

Breakpoint Cluster Region-Abelson oncogene (BCR-ABL) is a member of the tyrosine kinase family, and its deregulation may lead to chronic myeloid leukemia (CML) [49]. Before discovering its involvement in CML, classical anticancer treatments were given to patients, such as busulfan and hydroxyurea [50]. Due to the cytotoxic effects of these compounds, recombinant interferon-alfa (rIFN- α) has also been employed [51]. However, thanks to increasing knowledge of CML and with the availability of structural data [52], the first inhibitor targeting specifically BCR-ABL, imatinib [53], reached the market in 2001, becoming the reference treatment of CML. Due to the resistance to this drug observed in several patients, a second generation of Abl inhibitors has been developed since then, such as dasatinib, nilotinib, and bosutinib [54]. As BCR-ABL is today one of the most studied kinases, we chose this protein kinase for the first F2D study. The structure used to assess the F2D process is the wild-type ABL kinase domain co-crystallized with imatinib (PDB ID 2HYY, chain A). The pyridine group of imatinib, which is bound to the hinge region, was selected as a seed (Figure 7A). From this seed, we aimed to grow toward the hydrophobic pocket of the cavity. Therefore, after a visual inspection of the seed positioned in the 3D crystal structure, we selected the C atom in the meta position relative to the N atom as a unique starting point, which corresponds to position 3 of the seed (Figure 7).

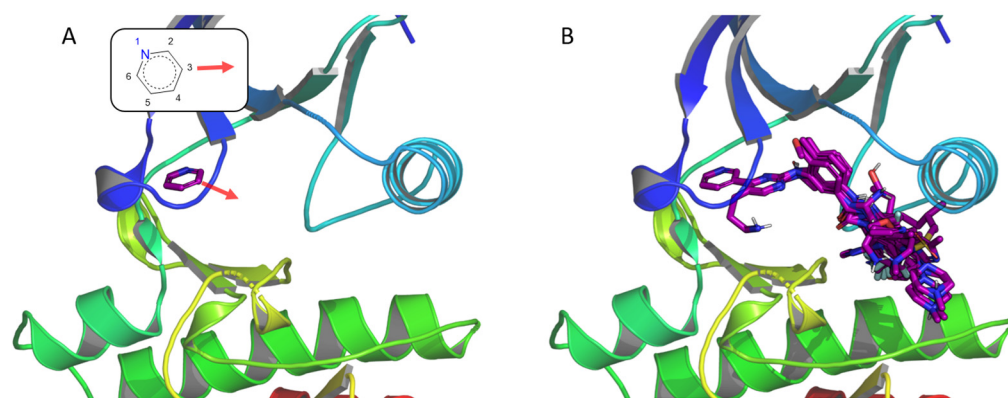


Figure 7. Example of F2D applied to BCR-ABL (PDB ID 2HYY, chain (A)). (A) Initial position of the seed from the co-crystallized ligand, the red arrow indicates the starting atom and the growing direction. 2D representation of the fragment is shown. (B) Representation of the 50 molecules obtained in the active site of BCR-ABL.

F2D generated 2595 molecules and clustered them ($\text{RMSD} \leq 1.0 \text{ \AA}$) to keep 1128 of them, 514 of which fulfil PKI-like filters as described in Table 1. Then, we applied the filtering process of the F2D workflow. Among the 514 molecules, 505 do not have a PAINS substructure, and of those 505 molecules, 495 molecules have a SA score < 4 . The 495 remaining molecules were docked to check if the combination of several fragments from different crystallized ligands could form favorable interactions with the active site of Abl (PDB ID 2HYY). A total of 138 molecules have an RMSD value lower than $2.5 \text{ \AA}$. Unsurprisingly, as it is present in many Abl inhibitors, we found the N-phenyl-4-(3-pyridyl)pyrimidin-2-amine moiety as a common substructure.

We then searched for similar molecules in the ChEMBL, ZINC, Ambinter, and PKIDBs to check the novelty of the built compounds and found 50 new molecules among the 138 retained. The 2D structures of these compounds are provided in S1, and the 3D superimposition inside the binding site is shown in Figure 8B. The 5 best molecules, according to the Quantitative Estimate of Drug-likeness (QED) score [55], are shown in Figure 8A. Their QED scores range from 0.63 to 0.79. They have a higher probability of becoming a hit than the other molecules generated by F2D. The expertise of medicinal chemists would be required to further optimize these molecules.

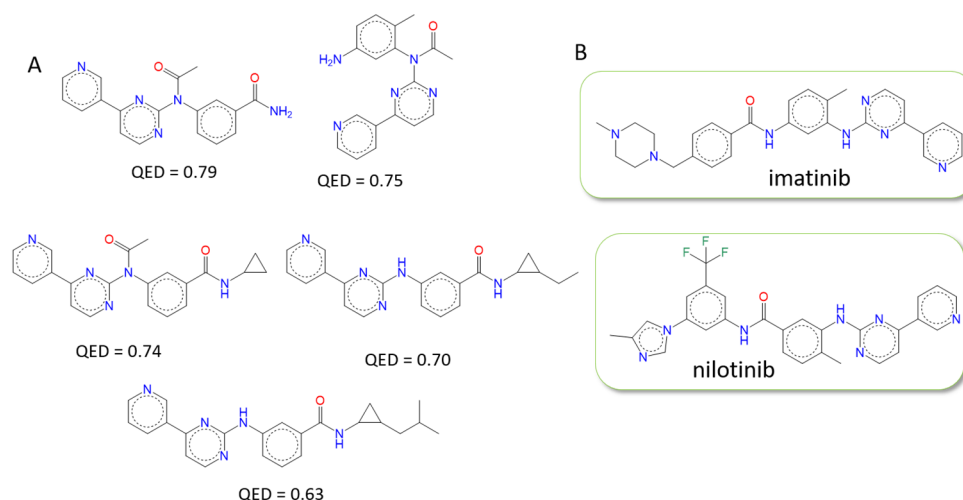


Figure 8. Structures of molecules obtained by generating molecules with F2D on BCR-ABL from the pyridine moiety as a seed. (A) Top 5 of the 50 new molecules sorted by QED score. (B) Two PKI found among the results: imatinib and nilotinib.

The 88 other molecules present similarities with at least one published molecule, including imatinib, nilotinib, radotinib, flumatinib, bafetinib, and masitinib, by considering a Tanimoto coefficient (Tc) of 0.7. Two of them correspond to the exact structure of known drugs found by F2D: imatinib and nilotinib (Figure 8B).

2.4.2. Discovery of BRAF V600E Inhibitors

B-Rapidly Accelerated Fibrosarcoma (BRAF) is a serine/threonine kinase and a member of the RAF family [56]. BRAF is mutated in several cancers [57], the common mutation V600E is mostly related to melanoma [58]. Vemurafenib (Zelboraf), targeting BRAF V600E [2], was found by using structural data from sorafenib (Nexavar), a CRAF inhibitor [59] having low inhibition on BRAF V600E. As vemurafenib is the first approved drug derived from FBDD, we chose BRAF as a second target example for assessing F2D. The structure provided to F2D is chain A of PDB ID 3OG7, which is the structure of BRAF V600E co-crystallized with vemurafenib. The 7-azaindole bound to the hinge, already used in the validation example of F2D (Figure 6), was selected as a seed (Figure S1A), but here the complete F2D database was used in the process.

We obtained 69,868 unique PKI-like molecules. We applied all F2D filters, removed molecules with PAINS substructures (305 molecules), and molecules having a SA score < 4, leading to 12,378 final molecules. After docking, only 525 docked molecules showed an RMSD value lower than 2.5 Å compared to the molecules generated by F2D.

Among the 525 molecules, 510 found no match by similarity search (Tc > 0.7) in ChEMBL, ZINC, Ambinter, or PKIDBs (Figure S1B). The only common substructure they share is the pyrrolo[2,3-b]pyridine seed. The higher number of new molecules found on the BRAF V600E target compared to BCR-ABL can be explained by the four growing atoms in the 7-azaindole seed. A total of 15 molecules among the 510 have a QED $\geq$ 0.75 and are shown in Figure S2.

In total, 510 molecules were new, and 15 molecules have at least one similar molecule found in one of the four databases, including vemurafenib, which was also reconstructed by F2D. The 2D structures of the 510 new molecules are depicted in the Supplementary Information.

2.4.3. Discovery of MELK Inhibitors

Maternal Embryonic Leucine zipper Kinase (MELK) is another serine/threonine kinase, a member of the CAMK family [60]. Its deregulation seems to be implicated in many cancers [61], and this subject was challenged in the literature [62]. Type I and type II inhibitors have already been discovered for MELK [63], also from FBDD approaches [64,65].

Type I MELK Inhibitors

We first searched for new type I inhibitors of MELK. The target and the seed were chosen from the PDB ID 4UMQ structure, which is MELK co-crystallized with 3-{5-[(3-hydroxy-5-methoxyphenyl) amino]-2-(phenylcarbamoyl) phenoxy} propan-1-aminium. This structure has a DFG motif and an α C-helix in the in-conformation. The seed positioned in the protein active site and the starting growing directions are shown in Figure S3A.

87,961 molecules were generated, reduced to 86,669 molecules without PAINS substructures, and only 3934 molecules with a SA score < 4. After the docking step, this number decreased to 41 molecules presenting the closest docking pose with a RMSD value < 2.5 Å. Their 3D structures are presented in Figure S3B. Visualization of those molecules confirms they are type I inhibitors occupying only the ATP-binding pocket and not the allosteric pocket.

All the 41 molecules, selected after applying F2D filters, are new molecules. They have no similar compounds in ChEMBL, Ambinter, ZINC, or PKIDB. Molecules with the

best QED score are drawn in Figure S4. The 41 molecules are provided in the Supplementary Information.

Type II MELK Inhibitors

Secondly, we focused on the discovery of new type 2 inhibitors of MELK. We selected the PDB ID 4UMT, which is the structure of MELK co-crystallized with 1-(4-([3-(isoquinolin-7-yl)prop-2-yn-1-yl]oxy)-2-methoxybenzyl)piperazinediium [65]. This conformation of the target has a DFG-in motif and an out-form of the α C helix. We started the fragment growing from the 3-(isoquinolin-7-yl)prop-2-yn-1-ol moiety seed (Figure S5A).

We found 63 unique PKI-like molecules. Then, we applied F2D filters, and 3 molecules were removed because they contained PAINS substructures, and 18 molecules were kept with SA scores < 4. After docking, 14 molecules had an RMSD value < 2.5 Å. From these 14 molecules, 6 molecules were similar ($T_c > 0.7$) to at least one compound present in ChEMBL, and none to the other databases. Five of the 6 ChEMBL molecules were synthesized and biologically tested on MELK [65]. Their respective ChEMBL IDs and MELK activity are ChEMBL3355060 (MELK IC₅₀ = 24 μ M), ChEMBL3355061 (MELK IC₅₀ = 5.1 μ M), ChEMBL3355062 (MELK IC₅₀ = 0.65 μ M), ChEMBL3355063 (MELK IC₅₀ = 0.52 μ M), ChEMBL3355066 (MELK %inhibition = 52% at 300 μ M), and ChEMBL215098 (not tested on MELK).

The 8 other molecules have low similarity ($T_c < 0.7$) to the compounds in the ChEMBL, ZINC, Ambinter, or PKIDBs. Figure S5B represents one of these 8 new molecules. Their QED scores vary from 0.45 to 0.73. The 2D structures of the molecules with the best QED score are shown in Figure S6. The 2D structures of these 8 new molecules are given in the Supplementary Information.

The type II inhibitors found on MELK provide better selectivity on crystallized human protein kinases than type I inhibitors.

2.4.4. Discovery of Macrocyclic Inhibitors Targeting ALK

Macrocycles represent an interesting class of compounds able to inhibit challenging targets [66,67]. Even if there are more macrocyclic natural products on the market than synthetic macrocycles, the latter provide a better kinase affinity and selectivity than other compounds [66]. Fragmentation and generation of macrocycles is not straightforward in *in silico* FBDD. Fragmentation methods often fail to accurately cut the correct bonds in macrocycles. To manage this, we used several fragmentation methods on the crystallized PKI used to generate the F2D fragment database. However, some parts of the ligands still could not be fragmented and provided fragments with a molecular weight > 300 Da, which were deleted as found also during the validation step of F2D (see above validation of F2D method). Regardless, F2D was able to generate macrocycles from fragments contained in the graph database.

We focused on Anaplastic Lymphoma Kinase (ALK) target and its macrocyclic inhibitor lorlatinib. ALK is a tyrosine kinase belonging to the insulin receptor subfamily implicated in the nervous system's development and function [68]. Chromosomal translocations involving ALK result in several fusion proteins identified in cancers [69–71] as well as the amplification and mutations of the full ALK gene [72–75].

Several ALK inhibitors have already been described, such as crizotinib [76] and lorlatinib, which is a macrocycle targeting wild-type or L1196M mutant ALK [77]. We selected the PDB ID 4CLI as the target for F2D. We chose the 3 carbon atoms in ortho, meta, and para positions from the amine of the aminopyridine moiety seed to start the growing (Figure 9A).

Macrocycles are characterized by a ring composed of at least 12 atoms [66]. However, to maximize the number of macrocycles after F2D execution, we selected molecules with at least 11 atoms in their largest rings. We obtained 592 molecules having macrocycles with lengths from 11 to 15 atoms. A total of 35 molecules were removed because of PAINS substructures. Because the synthesis of macrocycles is more complex than for other molecules, we increased the threshold of the SA score used to retain the more feasible ones. Thus, we obtained 174 molecules having an SA score < 5 . After the docking experiment, 153 macrocycles show an RMSD value < 2.5 Å between the docked poses and the molecules generated by F2D.

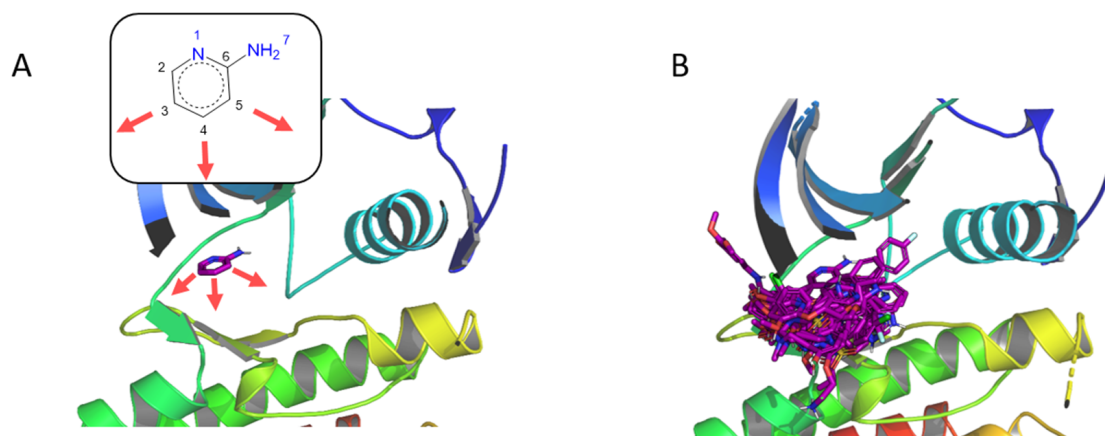


Figure 9. Example of F2D on ALK (PDB ID 4CLI, chain A). **(A)** Initial position of the aminopyridine seed in the active site, the red arrows indicate the starting atoms and the growing directions. 2D representation of the fragment is shown. **(B)** New macrocyclic molecules obtained in the active site of ALK.

No similar compound with a $T_c > 0.7$ was found in the ChEMBL, Ambinter, ZINC, or PKIDBs, leading to 153 new macrocycles identified by F2D (Figure 9B).

Figure 10 represents the 2D structures of 9 macrocycles having a QED score ≥ 0.7 . All 2D structures obtained are listed in the Supplementary Information. By visual inspection, we found that one compound seemed to be very close to lorlatinib, even if its Tanimoto coefficient (T_c) is only 0.41 (Figure 11).

2.4.5. F2D Website

F2D was created to help researchers (medicinal chemists or biologists) find new protein kinase inhibitors. Thus, we provide an online interface to facilitate the use of F2D. This interface is freely accessible at <http://frags2drugs.icoa.fr> (accessed on 14 January 2026). F2D needs several input data: a protein kinase structure, a seed structure, and, optionally, the definition of atoms for growing. This leads to four initial steps required to launch F2D, detailed in a user guide. Firstly, the user has to select the desired target from its protein gene name (e.g., BRAF). A table of the PDB ID structures corresponding to this protein gene name is listed, and the user can choose the desired target structure. Finally, the user chooses a seed and selects the atoms for growing. The four selection steps are provided with a user-friendly interface.

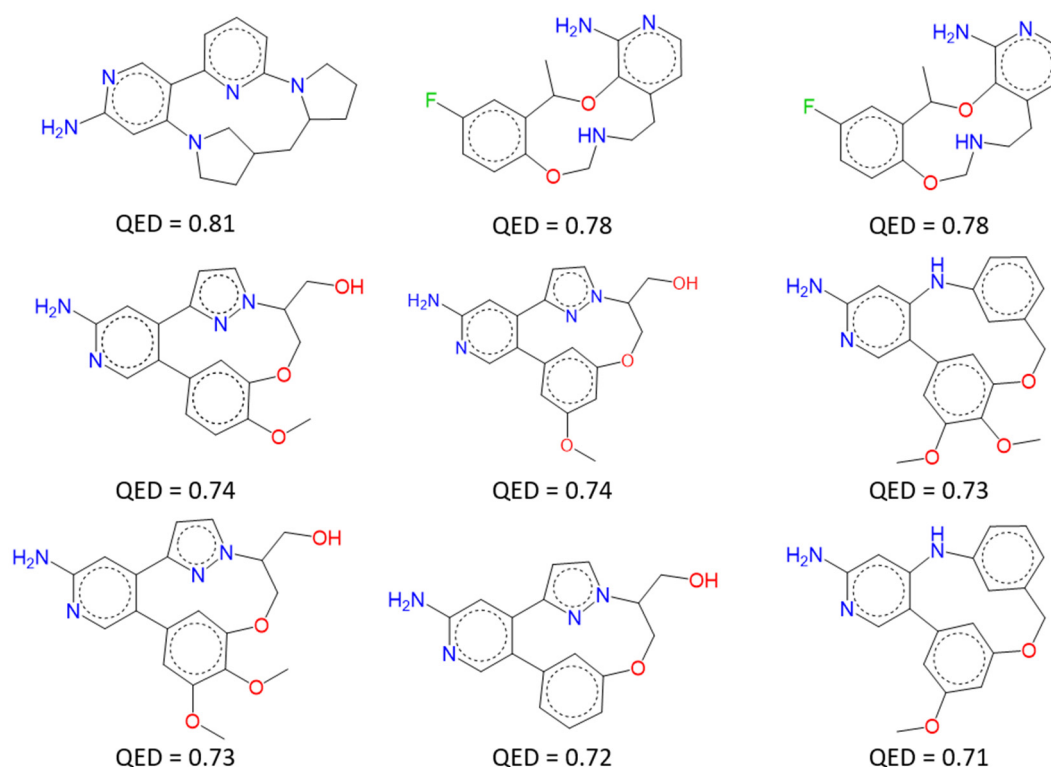


Figure 10. Structures of the 9 macromolecules having the highest QED scores generated with F2D on ALK (PDB ID 4CLI) from an amino-pyridine moiety seed.

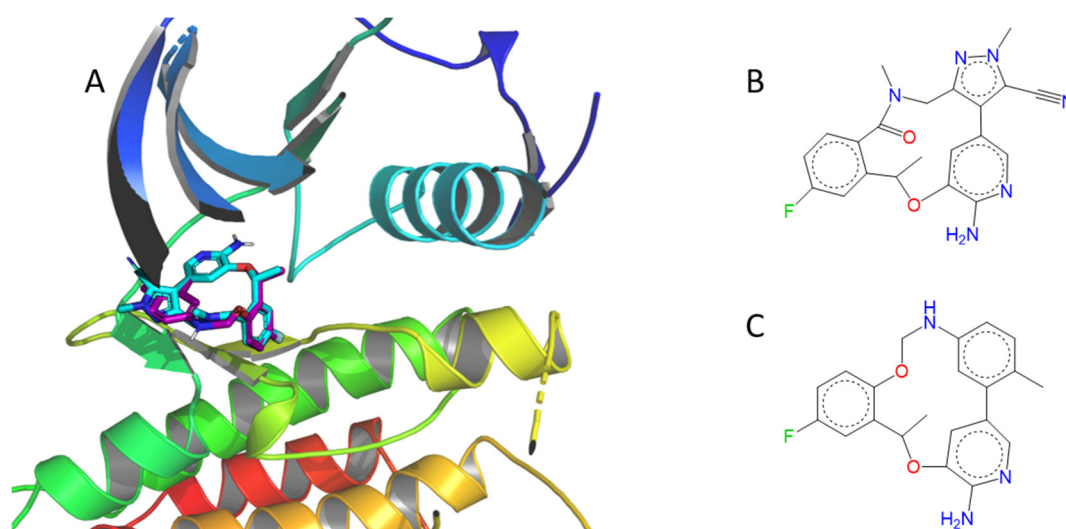


Figure 11. Comparison between lorlatinib and the most similar expected inhibitor found in Frags2Drugs. (A) 3D structures of co-crystallized lorlatinib (in cyan) and the F2D inhibitor (in purple) bound in the ALK active site (PDB ID 4CLI). (B) 2D structure of lorlatinib. (C) 2D structure of the most similar F2D inhibitor compared to lorlatinib (Tc = 0.41).

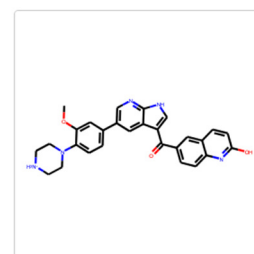
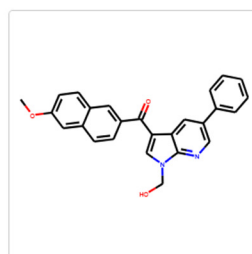
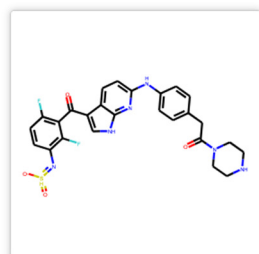
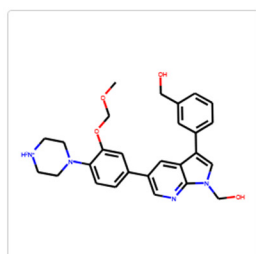
The search for new molecules starts. F2D will then provide a web page link to retrieve results. Once available, the results are displayed as shown in Figure 12. Furthermore, the parameters used and the number of generated molecules is detailed in a table. The 2D representation of built molecules is displayed in an 8-by-8 matrix. The user may select a molecule to get additional details. Results are downloadable as an SDF file to allow researchers to further process them. All results obtained on the F2D website are accessible for one month.

Molecules generated by Frags2Drugs

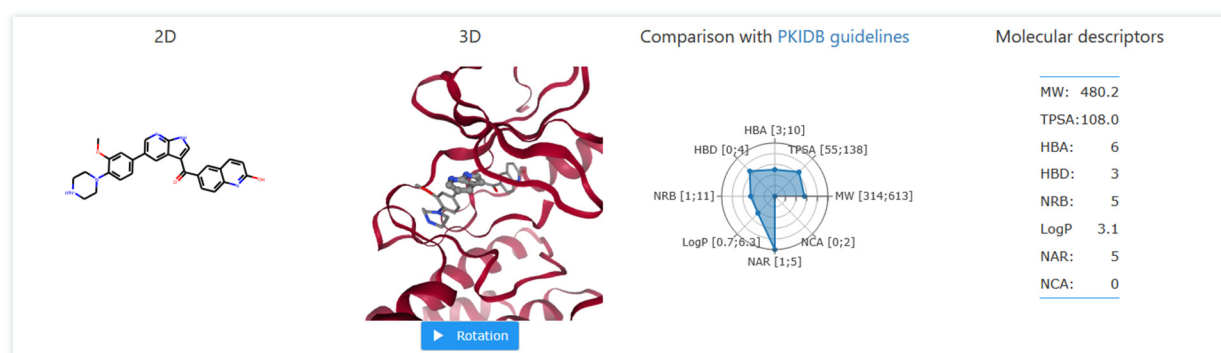
[Download SDF](#)

Date	Protein	PDB ID	Seed	Starting atoms	Number of generated molecules
2026-02-11	BRAF	3OG7	3OG7_A_L1F003	[0, 1, 2, 4, 5, 6]	334

Select a molecule to get more details



Molecule details

[Download SDF](#)


Fragment	Fragment name	Protein gene name of origin	PDB Code of origin	Chain
	4BGG-ALI-L1F003	ACVR1	4BGG	B
	3OG7A-ALI-L1F003	BRAF	3OG7	A

Figure 12. Example of results obtained from the Frags2Drugs website. This figure shows 4 of the 334 created molecules. On the website, all the molecules are available, displayed in an 8-by-8 matrix. The table showing the details for each fragment used in the selected molecule is also provided.

3. Discussion

We present here Frags2Drugs (F2D), a new *in silico* fragment-based ligand design tool built on a 3D fragment network and a growing-linking approach to find novel protein kinase inhibitors. Several PKI-like filters are applied, and a synthetic accessibility score is provided at the end of the process. We have demonstrated its success on several protein kinase examples for which we generated type I and type II inhibitors. In addition, we presented an example of finding novel macrocyclic inhibitors for the ALK protein kinase. Interestingly, we were able to identify compounds similar to known active molecules

or drugs. The main advantage of F2D is the use of 3D experimental structures without the need for energy minimization after fragment growing and linking, because it uses empirical rules of the MMFF94 force field when creating covalent bonds. Furthermore, no systematic cavity exploration is needed; thus, the results can be rapidly obtained. F2D is freely accessible at <http://frags2drugs.icoa.fr/> (accessed on 14 January 2026). Once the user has selected the target of interest, the seed, and the anchoring points, the tool provides a set of novel molecules bound inside the active site of the target.

F2D is able to create type I, type I1/2, and type II inhibitors of protein kinases, but can also generate more complex molecules such as macrocycles. Further study is ongoing to apply F2D to covalent kinase inhibitors.

F2D currently focuses on the design of protein kinase inhibitors. A great improvement would be to broaden the application of F2D to any protein family. Because F2D is based on the sequence and structure alignment of protein kinase-ligand complexes, this method cannot be applied to other protein families. In order to expand our application, several procedures may be envisaged, such as aligning cavities from different protein families [78] or moving fragments from a binding site to another using subpockets [79] or shape-based descriptors [80].

A key element of the creation of an F2D database is the sequence and structure alignment of all protein kinases so that the binding sites of crystallized structures are well superimposed. The seed position needs to be defined and positioned into the binding site to enable the creation of type I, type I1/2, and type II inhibitors. Therefore, allosteric inhibitors outside the active site cannot be created by F2D. Again, F2D could be improved by adding more fragments in allosteric sites, by using docking methods of fragments, or by repositioning fragments from similar subpockets.

The web interface of F2D is restricted to proteins with available crystallographic structures. The entire protein kinase space could be covered by using bioinformatics methods such as homology or AI modeling to predict the 3D structure of protein kinases.

The fragmentation of molecules is another key point of database creation. However, despite the use of published fragmentation algorithms to achieve this step, several known inhibitors could not be reconstructed because of these issues. We thus have to improve the implemented fragmentation methods to overcome this limitation. This could be achieved by rewriting our own method or using other established methods such as DAIM [81] or Ftrees-FS [82].

We have implemented a filtering procedure to reduce the number of molecules given by F2D, based on structural and physicochemical parameters of compounds. However, the predicted bioactivity of the created molecules is not currently considered. To guide the choice of the most interesting compounds, it could be wise to implement a deep learning model to predict binding affinities [83].

4. Materials and Methods

4.1. Protein Kinase Superimposition and Fragment Dataset

PDB ID of protein kinases was retrieved by using a request based on specific PFAM [84] identifiers: the PFAM id “PF07714” for protein tyrosine kinase and the PFAM id “PF00069” for protein kinase domain. These PFAM accession numbers gave us access to 3809 structures, regardless of the organism’s origin (July 2016). These structures were then split by chain using Biopython tools [85] (version 1.72). Structures having alternative locations for a few atoms were duplicated to treat cases independently. The alignment of the protein kinases on the reference protein kinase with the PDB ID 1ATP chain E was performed with Molecular Operating Environment (MOE, Chemical Computing Group, Montreal, QC, Canada, 2016_0802) [86].

Several algorithms already implemented in KNIME [87] were used to fragment the ligands: RECAP [88], BRICS [89], and Scaffold Tree [90]. The three methods were used to provide a larger number of different fragments, so further novel molecules will be created. Fragments were then standardized using VSPrep [91], a KNIME workflow. At this stage, a filter was applied to fragments. Fragment having a MW > 300 Da, a phosphate, a carbohydrate, an ester, an acyclic N-N, N-O, or O-O substructure was removed because of potential pharmacokinetic issues. The next step consisted of removing duplicate fragments. Here, we considered a fragment as a chemical group in a particular 3D orientation. So, duplicates were not only 2D identical compounds, such as SMILES, but also compounds having both the same SMILES and 3D identical coordinates. To manage this, we grouped the same fragments together, in clusters, based on their SMILES. Then, for each cluster, the full RMSD matrix was calculated between all members. For each pair of fragments having an RMSD value less than 0.25 Å, only the first one was kept. The database, at the end, contained several 3D orientations for the fragments.

Starting from 6204 initial ligands, the database finally contained 72,480 fragments. Each of these fragments was labeled with a unique identifier and information about its origin (PDB ID, chain, alternative location, Uniprot gene name, and species).

4.2. Creation of the 3D Fragment Network

We used Python (3.6.8), neo4j (3.5.11), and mainly the Python library networkx (2.2) to create the fragment network.

Calculation of the compatibility relationship was performed in two steps. Firstly, for each protein, its binding site cavity and the surrounding atoms were identified using FPocket [92], an open-source cavity detection tool based on Voronoï tessellation. Secondly, a fragment was defined as compatible with a cavity if the distances between its atoms and the atoms of the protein ranged between 1.4 Å and 3.5 Å. Thus, we avoided both steric hindrances with the fragment and the protein and built compounds far from the active site.

When two fragments were linked, the four calculated distortion parameter values ranged from 0 to 45, with 0 meaning there is no distortion compared to the reference value of the force field. For angles, increasing the value by one means authorizing a difference of one additional degree, while for distance, the value corresponded to a percentage of allowed deviation from the reference value of the force field, so the maximum distortion allowed for a distance was 45% of the reference value.

4.3. Molecular Generation and Filtering Process

Nodes from the graph database were associated with corresponding structures of fragments with the use of Pandas dataframe (1.0.5) and RDKit [93] (2020.03.3). RDKit was also used to link fragments together to generate molecules.

During the filtering process, substructure recognition relies on SMARTS patterns [94] and was applied by substructure match from RDKit. The PKI-like filter removed molecules containing undesired physicochemical parameters.

The molecular docking step was carried out with rDock [95] (2013.1) software. Before proceeding to docking, molecules were standardized and protonated. New conformations were generated using the RDKit method of conformation generation ETKDG [96] coupled to MMFF94 force field refinement [34]. New conformations allowed us to avoid the bias consisting of giving the docking software a conformation similar to that provided by F2D. Root Mean Square Deviation (RMSD) between molecules from F2D and docking poses was calculated with RDKit. We chose a threshold of 2.5 Å from the minimal RMSD values observed by applying the same procedure on all protein kinase-ligand complexes. Because rDock failed in generating conformations of macrocycles, the conformational

search was performed with the MOE (Chemical Computing Group, Montreal, Quebec, Canada, 2019_0101) suite [97]. SA score calculation and PAINS substructure removal were achieved using RDKit modules. CNS MPO is calculated using ChemAxon's Calculator plugins, Marvin 15.2.9, 2015, ChemAxon (<https://www.chemaxon.com>), and RDKit.

Similarity searches performed on whole databases used FPSim2 [98]. The similarity search was based on the Tanimoto coefficient [99] with a similarity threshold of 0.7. The versions of the four databases were ChEMBL 27 (accessed June 2020), ZINC 15 (accessed June 2020), Ambinter February 2020, and PKIDB 9 December 2020 <http://www.icoa.fr/pk-idb/> (accessed on 14 January 2021).

4.4. Validation of F2D Methodology

From 6204 initial ligands, we first selected the 4338 PKI-like ligands and removed staurosporine-like molecules (PDB ligand IDs STU, UCN, and LY4). Indeed, these compounds contain an indolocarbazole, which is not fragmented by the algorithms. Due to its MW > 300 Da, this group is too heavy to be kept in the database. We also removed the ligands having chemical issues in their registered structures in the PDB. This led to 3218 ligands being reconstructed to validate the F2D method.

For each of the 3218 ligands, we selected its corresponding nodes and edges from the graph database, and we applied F2D using each fragment alternatively as a seed. After each generation of molecules, we checked the presence of the original ligand among the molecules created. If the ligand was not found, we repeated the process using another seed.

4.5. F2D Website

F2D is freely accessible from <http://frags2drugs.icoa.fr> (accessed on 14 January 2026). This website is based on several Docker 20.10.5 containers communicating through docker-compose 1.28.6. The Python web framework Django LTS 2.2 and Celery 4.4.6 are used to run the website and F2D calculations in a task queue. The database relies on PostgreSQL 12.2 and the RDKit cartridge 2020.03.

4.6. Discovery of Kinase Inhibitors

The 3D position of the seeds used in the four examples (ABL, BRAF, MELK, and ALK) was obtained by removing all atoms, except those of the seed, from the co-crystallized molecules.

The four distortion parameters used while growing molecules in the examples presented in this article are shown in Table 2. The values vary depending on the number of generated molecules. If one wants to generate more molecules, the distortion parameters will be increased.

Table 2. Distortion parameters used in F2D during fragment growing examples.

	Torsion Angle (deg)	Out of Plane Angle (deg)	Dihedral Angle (deg)	Distance (%)
BCR-ABL	15	15	15	10
BRAF V600E	10	10	10	10
MELK Type I	10	10	10	10
MELK Type II	14	14	14	12
ALK	10	10	10	10

During the search for macrocycles, no restriction was applied on the distortion parameters to maximize the number of results.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ph19020308/s1>, Figure S1: Example of F2D applied to BRAF V600E (PDB ID 3OG7, chain A); Figure S2: Structures of the 15 molecules with the highest QED score generated by F2D on BRAF V600E from the pyrrolo[2,3-b]pyridine moiety seed; Figure S3: Example of F2D to the discovery of MELK type I inhibitors (PDB ID 4UMQ, chain A); Figure S4: Structures of the 9 predicted type I inhibitors, having the highest QED scores, generated by F2D on MELK from a benzamide seed; Figure S5. Example of F2D execution to discover type II inhibitors on MELK (PDB ID 4UMT, chain A); Figure S6. Structures of the 6 type II inhibitors, having the highest QED score, generated by F2D on MELK from a 3-(isoquinolin-7-yl)prop-2-yn-1-ol moiety seed. The sdf files contain the 50, 510, 41, 8 and 153 molecules found by F2D for BCR-ABL, BRAF V600E, MELK (type I), MELK (type II), and ALK respectively. The mol2 files of the proteins, the sdf files of co-crystallized ligands and the seeds are provided.

Author Contributions: Conceptualization, S.A.-S. and P.B.; methodology, G.P., C.B., P.K., J.-M.G., S.A.-S. and P.B.; software, G.P., C.B., P.K. and J.-M.G.; validation, G.P., C.B., P.K. and J.-M.G.; formal analysis and investigation, all authors; resources, P.B.; data curation, all authors; writing—original draft preparation, G.P., C.B. and J.-M.G.; writing—review and editing, S.B., S.A.-S. and P.B.; visualization, G.P., C.B., P.K. and J.-M.G.; supervision, S.A.-S. and P.B.; project administration, P.B.; funding acquisition, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: Authors gratefully acknowledge financial support from Région Centre Val de Loire (APR-IR Kinometric and isNatProd) and from ANR (uF2D ANR-21-CE18-0056 and LabEx SYNORG ANR-11-LABX-0029).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Material. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors thank the projects CHemBio (FEDER-FSE 2014-2020-EX003677), Val-biocosm (FEDER-FSE 2014-2020-EX003202), Techsab (FEDER-FSE 2014-2020-EX011313), QUALICHIM (APR-IA-PF 2021-00149467), the RTR Motivhealth (2019-00131403), and the Labex programs IRON (ANR-11-LABX-0018-01) for their financial support of ICOA, UMR 7311, University of Orléans, CNRS. Part of this study has been published in the thesis of G.P.: Gautier Peyrat, Conception d'inhibiteurs de protéines kinases à partir de méthodes in silico basées sur les fragments, Université d'Orléans, Orléans, France, 2021. The authors thank Karen Plé for the careful grammar correction of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ALK	Anaplastic Lymphoma Kinase
BBB	Blood–Brain Barrier
BRAF	B–Rapidly Accelerated Fibrosarcoma
BCR-ABL	Breakpoint Cluster Region–Abelson oncogene
ClogP	Calculated logP
CNS MPO	Central Nervous System Multiparameter Optimization
CML	Chronic Myeloid Leukemia
FBDD	Fragment-Based Drug Design
F2D	Fragments2Drugs
Ge	Graph of exclusions
Gi	Graph of inclusions
rIFN- α	interferon-alfa
KLIFS	Kinase–Ligand Interaction Fingerprints and Structure database

Ro5	Lipinski's Rule of Five
MELK	Maternal Embryonic Leucine zipper Kinase
MOE	Molecular Operating Environment
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NAR	Number of Aromatic Rings
NCA	Number of Chiral Atoms
HBA	Number of Hydrogen Bond Acceptors
HBD	Number of Hydrogen Bond Donors
NRB	Number of Rotatable Bonds
OOP	Out of Plane
PAINS	Pan-Assay Interference Compounds
PDB	Protein DataBank
PKIDB	Protein Kinase Inhibitor Database
PKI	Protein Kinase Inhibitors
QED	Quantitative Estimate of Druglikeness
RCSB	Research Collaboratory for Structural Bioinformatics
RMSD	Root Mean Square Deviation
SAR	Structure Activity Relationships
SA	Synthetic Accessibility
3D	Three Dimensional
TPSA	Topological Polar Surface Area
FDA	U.S. Food and Drug Administration

References

1. Scott, D.E.; Coyne, A.G.; Hudson, S.A.; Abell, C. Fragment-Based Approaches in Drug Discovery and Chemical Biology. *Biochemistry* **2012**, *51*, 4990–5003. [[CrossRef](#)] [[PubMed](#)]
2. Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P. Vemurafenib: The First Drug Approved for *BRAF*-Mutant Cancer. *Nat. Rev. Drug Discov.* **2012**, *11*, 873–886. [[CrossRef](#)] [[PubMed](#)]
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. Xu, W.; Kang, C. Fragment-Based Drug Design: From Then until Now, and Toward the Future. *J. Med. Chem.* **2025**, *68*, 5000–5004. [[CrossRef](#)]
5. Perera, T.P.S.; Jovcheva, E.; Mevellec, L.; Vialard, J.; Lange, D.D.; Verhulst, T.; Paulussen, C.; Ven, K.V.D.; King, P.; Freyne, E.; et al. Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol. Cancer Ther.* **2017**, *16*, 1010–1020. [[CrossRef](#)]
6. Hall, R.J.; Mortenson, P.N.; Murray, C.W. Efficient Exploration of Chemical Space by Fragment-Based Screening. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 82–91. [[CrossRef](#)]
7. Grove, L.E.; Vajda, S.; Kozakov, D. Computational Methods to Support Fragment-Based Drug Discovery. In *Fragment-Based Drug Discovery Lessons and Outlook*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; pp. 197–222. [[CrossRef](#)]
8. Durrant, J.D.; Amaro, R.E.; McCammon, J.A. AutoGrow: A Novel Algorithm for Protein Inhibitor Design. *Chem. Biol. Drug Des.* **2009**, *73*, 168–178. [[CrossRef](#)] [[PubMed](#)]
9. Durrant, J.D.; Lindert, S.; McCammon, J.A. AutoGrow 3.0: An Improved Algorithm for Chemically Tractable, Semi-Automated Protein Inhibitor Design. *J. Mol. Graph. Model.* **2013**, *44*, 104–112. [[CrossRef](#)]
10. Spiegel, J.O.; Durrant, J.D. AutoGrow4: An Open-Source Genetic Algorithm for de Novo Drug Design and Lead Optimization. *J. Cheminform.* **2020**, *12*, 25. [[CrossRef](#)]
11. Kutchukian, P.S.; Lou, D.; Shakhnovich, E.I. FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules Occupying Druglike Chemical Space. *J. Chem. Inf. Model.* **2009**, *49*, 1630–1642. [[CrossRef](#)]
12. Chéron, N.; Jasty, N.; Shakhnovich, E.I. OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands. *J. Med. Chem.* **2016**, *59*, 4171–4188. [[CrossRef](#)] [[PubMed](#)]
13. Nishibata, Y.; Itai, A. Automatic Creation of Drug Candidate Structures Based on Receptor Structure. Starting Point for Artificial Lead Generation. *Tetrahedron* **1991**, *47*, 8985–8990. [[CrossRef](#)]
14. Rotstein, S.H.; Murcko, M.A. GenStar: A Method for de Novo Drug Design. *J. Comput. Aided Mol. Des.* **1993**, *7*, 23–43. [[CrossRef](#)]

15. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534. [[CrossRef](#)]
16. Böhm, H.-J. The Computer Program LUDI: A New Method for the de Novo Design of Enzyme Inhibitors. *J. Comput. Aided Mol. Des.* **1992**, *6*, 61–78. [[CrossRef](#)]
17. Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: A Computer-Aided Ligand Design for Structure-Based Drug Design. *J. Med. Chem.* **2005**, *48*, 2457–2468. [[CrossRef](#)]
18. Dey, F.; Cafilisch, A. Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf. Model.* **2008**, *48*, 679–690. [[CrossRef](#)]
19. Nisius, B.; Rester, U. Fragment Shuffling: An Automated Workflow for Three-Dimensional Fragment-Based Ligand Design. *J. Chem. Inf. Model.* **2009**, *49*, 1211–1222. [[CrossRef](#)]
20. Imrie, F.; Bradley, A.R.; van der Schaar, M.; Deane, C.M. Deep Generative Models for 3D Linker Design. *J. Chem. Inf. Model.* **2020**, *60*, 1983–1995. [[CrossRef](#)] [[PubMed](#)]
21. Yuan, Y.; Pei, J.; Lai, L. LigBuilder 2: A Practical de Novo Drug Design Approach. *J. Chem. Inf. Model.* **2011**, *51*, 1083–1091. [[CrossRef](#)]
22. Chevillard, F.; Rimmer, H.; Betti, C.; Pardon, E.; Ballet, S.; van Hilten, N.; Steyaert, J.; Diederich, W.E.; Kolb, P. Binding-Site Compatible Fragment Growing Applied to the Design of B2-Adrenergic Receptor Ligands. *J. Med. Chem.* **2018**, *61*, 1118–1129. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, J.; Yang, P.L.; Gray, N.S. Targeting Cancer with Small Molecule Kinase Inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28–39. [[CrossRef](#)]
24. Cohen, P. Protein Kinases—The Major Drug Targets of the Twenty-First Century? *Nat. Rev. Drug Discov.* **2002**, *1*, 309–315. [[CrossRef](#)] [[PubMed](#)]
25. Attwood, M.M.; Fabbro, D.; Sokolov, A.V.; Knapp, S.; Schiöth, H.B. Trends in Kinase Drug Discovery: Targets, Indications and Inhibitor Design. *Nat. Rev. Drug Discov.* **2021**, *20*, 839–861. [[CrossRef](#)] [[PubMed](#)]
26. Noble, M.E.M.; Endicott, J.A.; Johnson, L.N. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **2004**, *303*, 1800–1805. [[CrossRef](#)]
27. Wu, P.; Clausen, M.H.; Nielsen, T.E. Allosteric small-molecule kinase inhibitors. *Pharmacol. Ther.* **2015**, *156*, 59–68. [[CrossRef](#)]
28. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [[CrossRef](#)]
29. Roskoski, R. Classification of Small Molecule Protein Kinase Inhibitors Based upon the Structures of Their Drug-Enzyme Complexes. *Pharmacol. Res.* **2016**, *103*, 26–48. [[CrossRef](#)]
30. Zheng, J.; Trafny, E.A.; Knighton, D.R.; Xuong, N.; Taylor, S.S.; Ten Eyck, L.F.; Sowadski, J.M. 2.2 Å Refined Crystal Structure of the Catalytic Subunit of cAMP-Dependent Protein Kinase Complexed with MnATP and a Peptide Inhibitor. *Acta Crystallogr. Sect. D Struct. Biol.* **1993**, *49*, 362–365. [[CrossRef](#)]
31. Fogha, J.; Diharce, J.; Obled, A.; Aci-Sèche, S.; Bonnet, P. Computational Analysis of Crystallization Additives for the Identification of New Allosteric Sites. *ACS Omega* **2020**, *5*, 2114–2122. [[CrossRef](#)]
32. Kooistra, A.J.; Kanev, G.K.; van Linden, O.P.J.; Leurs, R.; de Esch, I.J.P.; de Graaf, C. KLIFS: A Structural Kinase-Ligand Interaction Database. *Nucleic Acids Res.* **2016**, *44*, D365–D371. [[CrossRef](#)] [[PubMed](#)]
33. Sydow, D.; Schmiel, P.; Mortier, J.; Volkamer, A. KinFragLib: Exploring the Kinase Inhibitor Space Using Subpocket-Focused Fragmentation and Recombination. *J. Chem. Inf. Model.* **2020**, *60*, 6081–6094. [[CrossRef](#)]
34. Halgren, T.A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519. [[CrossRef](#)]
35. Halgren, T.A. Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520–552. [[CrossRef](#)]
36. Halgren, T.A. Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586. [[CrossRef](#)]
37. Halgren, T.A. Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* **1996**, *17*, 616–641. [[CrossRef](#)]
38. Halgren, T.A.; Nachbar, R.B. Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comput. Chem.* **1996**, *17*, 587–615. [[CrossRef](#)]
39. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development settings1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 3–25.1. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [[CrossRef](#)]
40. Veber, D.F.; Johnson, S.R.; Cheng, H.-Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [[CrossRef](#)]

41. Carles, F.; Bourg, S.; Meyer, C.; Bonnet, P. PKIDB: A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials. *Molecules* **2018**, *23*, 908. [CrossRef]
42. Bournez, C.; Carles, F.; Peyrat, G.; Aci-Sèche, S.; Bourg, S.; Meyer, C.; Bonnet, P. Comparative Assessment of Protein Kinase Inhibitors in Public Databases and in PKIDB. *Molecules* **2020**, *25*, 3226. [CrossRef]
43. Baell, J.B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [CrossRef] [PubMed]
44. Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminform.* **2009**, *1*, 8. [CrossRef] [PubMed]
45. Wager, T.T.; Hou, X.; Verhoest, P.R.; Villalobos, A. Moving beyond Rules: The Development of a Central Nervous System Multiparameter Optimization (CNS MPO) Approach To Enable Alignment of Druglike Properties. *ACS Chem. Neurosci.* **2010**, *1*, 435–449. [CrossRef]
46. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef]
47. Sterling, T.; Irwin, J.J. ZINC 15—Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [CrossRef]
48. Ambinter. Available online: <http://www.ambinter.com/> (accessed on 7 April 2021).
49. Deininger, M.W.N.; Goldman, J.M.; Melo, J.V. The Molecular Biology of Chronic Myeloid Leukemia. *Blood* **2000**, *96*, 3343–3356. [CrossRef] [PubMed]
50. Bolin, R.W.; Robinson, W.A.; Sutherland, J.; Hamman, R.F. Busulfan versus Hydroxyurea in Long-Term Therapy of Chronic Myelogenous Leukemia. *Cancer* **1982**, *50*, 1683–1686. [CrossRef]
51. An, X.; Tiwari, A.K.; Sun, Y.; Ding, P.-R.; Ashby, C.R.; Chen, Z.-S. BCR-ABL Tyrosine Kinase Inhibitors in the Treatment of Philadelphia Chromosome Positive Chronic Myeloid Leukemia: A Review. *Leuk. Res.* **2010**, *34*, 1255–1268. [CrossRef]
52. Cowan-Jacob, S.W.; Fendrich, G.; Floersheimer, A.; Furet, P.; Liebetanz, J.; Rummel, G.; Rheinberger, P.; Centeleghe, M.; Fabbro, D.; Manley, P.W. Structural Biology Contributions to the Discovery of Drugs to Treat Chronic Myelogenous Leukaemia. *Acta Crystallographica. Sect. D Biol. Crystallogr.* **2007**, *63*, 80–93. [CrossRef] [PubMed]
53. Savage, D.G.; Antman, K.H. Imatinib Mesylate—A New Oral Targeted Therapy. *N. Engl. J. Med.* **2002**, *346*, 683–693. [CrossRef]
54. Bitencourt, R.; Zalcberg, I.; Louro, I.D. Imatinib Resistance: A Review of Alternative Inhibitors in Chronic Myeloid Leukemia. *Rev. Bras. Hematol. Hemoter.* **2011**, *33*, 470–475. [CrossRef]
55. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4*, 90–98. [CrossRef]
56. Ikawa, S.; Fukui, M.; Ueyama, Y.; Tamaoki, N.; Yamamoto, T.; Toyoshima, K. B-Raf, a New Member of the Raf Family, Is Activated by DNA Rearrangement. *Mol. Cell. Biol.* **1988**, *8*, 2651–2654.
57. Davies, H.; Bignell, G.R.; Cox, C.; Stephens, P.; Edkins, S.; Clegg, S.; Teague, J.; Woffendin, H.; Garnett, M.J.; Bottomley, W.; et al. Mutations of the BRAF Gene in Human Cancer. *Nature* **2002**, *417*, 949–954. [CrossRef]
58. Pollock, P.M.; Meltzer, P.S. A Genome-Based Strategy Uncovers Frequent BRAF Mutations in Melanoma. *Cancer Cell* **2002**, *2*, 5–7. [CrossRef] [PubMed]
59. Wilhelm, S.M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; et al. BAY 43-9006 Exhibits Broad Spectrum Oral Antitumor Activity and Targets the RAF/MEK/ERK Pathway and Receptor Tyrosine Kinases Involved in Tumor Progression and Angiogenesis. *Cancer Res.* **2004**, *64*, 7099–7109. [CrossRef] [PubMed]
60. Gil, M.; Yang, Y.; Lee, Y.; Choi, I.; Ha, H. Cloning and Expression of a cDNA Encoding a Novel Protein Serine/Threonine Kinase Predominantly Expressed in Hematopoietic Cells. *Gene* **1997**, *195*, 295–301. [CrossRef]
61. Gray, D.; Jubb, A.M.; Hogue, D.; Dowd, P.; Kljavin, N.; Yi, S.; Bai, W.; Frantz, G.; Zhang, Z.; Koeppen, H.; et al. Maternal Embryonic Leucine Zipper Kinase/Murine Protein Serine-Threonine Kinase 38 Is a Promising Therapeutic Target for Multiple Cancers. *Cancer Res.* **2005**, *65*, 9751–9761. [CrossRef] [PubMed]
62. Settleman, J.; Sawyers, C.L.; Hunter, T. Challenges in Validating Candidate Therapeutic Targets in Cancer. *eLife* **2018**, *7*, e32402. [CrossRef]
63. Chung, S.; Suzuki, H.; Miyamoto, T.; Takamatsu, N.; Tatsuguchi, A.; Ueda, K.; Kijima, K.; Nakamura, Y.; Matsuo, Y. Development of an Orally-Administrative MELK-Targeting Inhibitor That Suppresses the Growth of Various Types of Human Cancer. *Oncotarget* **2012**, *3*, 1629–1640. [CrossRef]
64. Johnson, C.N.; Berdini, V.; Beke, L.; Bonnet, P.; Brehmer, D.; Coyle, J.E.; Day, P.J.; Frederickson, M.; Freyne, E.J.E.; Gilissen, R.A.H.J.; et al. Fragment-Based Discovery of Type I Inhibitors of Maternal Embryonic Leucine Zipper Kinase. *ACS Med. Chem. Lett.* **2015**, *6*, 25–30. [CrossRef]
65. Johnson, C.N.; Adelinet, C.; Berdini, V.; Beke, L.; Bonnet, P.; Brehmer, D.; Calo, F.; Coyle, J.E.; Day, P.J.; Frederickson, M.; et al. Structure-Based Design of Type II Inhibitors Applied to Maternal Embryonic Leucine Zipper Kinase. *ACS Med. Chem. Lett.* **2015**, *6*, 31–36. [CrossRef]

66. Driggers, E.M.; Hale, S.P.; Lee, J.; Terrett, N.K. The Exploration of Macrocycles for Drug Discovery—An Underexploited Structural Class. *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624. [\[CrossRef\]](#)
67. Marsault, E.; Peterson, M.L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **2011**, *54*, 1961–2004. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Morris, S.W.; Naeve, C.; Mathew, P.; James, P.L.; Kirstein, M.N.; Cui, X.; Witte, D.P. ALK, the Chromosome 2 Gene Locus Altered by the t(2;5) in Non-Hodgkin's Lymphoma, Encodes a Novel Neural Receptor Tyrosine Kinase That Is Highly Related to Leukocyte Tyrosine Kinase (LTK). *Oncogene* **1997**, *14*, 2175–2188. [\[CrossRef\]](#)
69. Morris, S.W.; Kirstein, M.N.; Valentine, M.B.; Dittmer, K.G.; Shapiro, D.N.; Saltman, D.L.; Look, A.T. Fusion of a Kinase Gene, ALK, to a Nucleolar Protein Gene, NPM, in Non-Hodgkin's Lymphoma. *Science* **1994**, *263*, 1281–1284. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Mano, H. Non-Solid Oncogenes in Solid Tumors: EML4–ALK Fusion Genes in Lung Cancer. *Cancer Sci.* **2008**, *99*, 2349–2355. [\[CrossRef\]](#)
71. Griffin, C.A.; Hawkins, A.L.; Dvorak, C.; Henkle, C.; Ellingham, T.; Perlman, E.J. Recurrent Involvement of 2p23 in Inflammatory Myofibroblastic Tumors. *Cancer Res.* **1999**, *59*, 2776–2780. [\[PubMed\]](#)
72. Tuma, R.S. ALK Gene Amplified in Most Inflammatory Breast Cancers. *JNCI J. Natl. Cancer Inst.* **2012**, *104*, 87–88. [\[CrossRef\]](#)
73. Ren, H.; Tan, Z.-P.; Zhu, X.; Crosby, K.; Haack, H.; Ren, J.-M.; Beausoleil, S.; Moritz, A.; Innocenti, G.; Rush, J.; et al. Identification of Anaplastic Lymphoma Kinase as a Potential Therapeutic Target in Ovarian Cancer. *Cancer Res.* **2012**, *72*, 3312–3323. [\[CrossRef\]](#)
74. Azarova, A.M.; Gautam, G.; George, R.E. Emerging Importance of ALK in Neuroblastoma. *Semin. Cancer Biol.* **2011**, *21*, 267–275. [\[CrossRef\]](#)
75. Bergethon, K.; Shaw, A.T.; Ignatius Ou, S.-H.; Katayama, R.; Lovly, C.M.; McDonald, N.T.; Massion, P.P.; Siwak-Tapp, C.; Gonzalez, A.; Fang, R.; et al. ROS1 Rearrangements Define a Unique Molecular Class of Lung Cancers. *J. Clin. Oncol.* **2012**, *30*, 863–870. [\[CrossRef\]](#)
76. Cui, J.J.; Tran-Dubé, M.; Shen, H.; Nambu, M.; Kung, P.-P.; Pairish, M.; Jia, L.; Meng, J.; Funk, L.; Botrous, I.; et al. Structure Based Drug Design of Crizotinib (PF-02341066), a Potent and Selective Dual Inhibitor of Mesenchymal–Epithelial Transition Factor (c-MET) Kinase and Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* **2011**, *54*, 6342–6363. [\[CrossRef\]](#)
77. Johnson, T.W.; Richardson, P.F.; Bailey, S.; Brooun, A.; Burke, B.J.; Collins, M.R.; Cui, J.J.; Deal, J.G.; Deng, Y.-L.; Dinh, D.; et al. Discovery of (10R)-7-Amino-12-Fluoro-2,10,16-Trimethyl-15-Oxo-10,15,16,17-Tetrahydro-2H-8,4-(Metheno)Pyrazolo [4,3-h][2,5,11]-Benzoxadiazacyclotetradecine-3-Carbonitrile (PF-06463922), a Macrocyclic Inhibitor of Anaplastic Lymphoma Kinase (ALK) and c-Ros Oncogene 1 (ROS1) with Preclinical Brain Exposure and Broad-Spectrum Potency against ALK-Resistant Mutations. *J. Med. Chem.* **2014**, *57*, 4720–4744. [\[CrossRef\]](#)
78. Eguida, M.; Rognan, D. A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. *J. Med. Chem.* **2020**, *63*, 7127–7142. [\[CrossRef\]](#)
79. Kalliokoski, T.; Olsson, T.S.G.; Vulpetti, A. Subpocket Analysis Method for Fragment-Based Drug Discovery. *J. Chem. Inf. Model.* **2013**, *53*, 131–141. [\[CrossRef\]](#)
80. Penner, P.; Martiny, V.; Gohier, A.; Gastreich, M.; Ducrot, P.; Brown, D.; Rarey, M. Shape-Based Descriptors for Efficient Structure-Based Fragment Growing. *J. Chem. Inf. Model.* **2020**, *60*, 6269–6281. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking. *J. Med. Chem.* **2006**, *49*, 7384–7392. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Rarey, M.; Stahl, M. Similarity Searching in Large Combinatorial Chemistry Spaces. *J. Comput. Aided Mol. Des.* **2001**, *15*, 497–520. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Stepniewska-Dziubinska, M.M.; Zielenkiewicz, P.; Siedlecki, P. Development and Evaluation of a Deep Learning Model for Protein–Ligand Binding Affinity Prediction. *Bioinformatics* **2018**, *34*, 3666–3674. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Finn, R.D.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [\[CrossRef\]](#)
85. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; et al. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423. [\[CrossRef\]](#) [\[PubMed\]](#)
86. *Molecular Operating Environment (MOE)*; 2016_0802; Chemical Computing Group ULC: Montreal, QC, Canada, 2016.
87. Berthold, M.R.; Cebon, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications*; Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R., Eds.; Springer Berlin/Heidelberg: Berlin/Heidelberg, Germany, 2008; pp. 319–326.
88. Lewell, X.Q.; Judd, D.B.; Watson, S.P.; Hann, M.M. RECAPRetrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522. [\[CrossRef\]](#) [\[PubMed\]](#)

89. Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the Art of Compiling and Using “Drug-Like” Chemical Fragment Spaces. *ChemMedChem* **2008**, *3*, 1503–1507. [[CrossRef](#)]
90. Schuffenhauer, A.; Ertl, P.; Roggo, S.; Wetzel, S.; Koch, M.A.; Waldmann, H. The Scaffold Tree – Visualization of the Scaffold Universe by Hierarchical Scaffold Classification. *J. Chem. Inf. Model.* **2007**, *47*, 47–58. [[CrossRef](#)]
91. Gally, J.M.; Bourg, S.; Do, Q.T.; Aci-Sèche, S.; Bonnet, P. VSPrep: A General KNIME Workflow for the Preparation of Molecules for Virtual Screening. *Mol. Inform.* **2017**, *36*, 1700023. [[CrossRef](#)]
92. Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinform.* **2009**, *10*, 168. [[CrossRef](#)]
93. RDKit: Open-Source Cheminformatics. Available online: <https://www.rdkit.org/> (accessed on 6 November 2020).
94. Daylight Theory: SMARTS—A Language for Describing Molecular Patterns. Available online: <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html#RTFTToC35> (accessed on 6 November 2020).
95. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A.B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R.E.; Morley, S.D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571. [[CrossRef](#)]
96. Riniker, S.; Landrum, G.A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562–2574. [[CrossRef](#)]
97. *Molecular Operating Environment (MOE)*; 2019_0101; Chemical Computing Group ULC: Montreal, QC, Canada, 2019.
98. *ChEMBL/FPsim2*; The ChEMBL Group: Cambridge, UK, 2021.
99. Tanimoto, T.T. *An Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corporation: New York, NY, USA, 1958.

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.