Searching Drug Space Using QuADD Best-In-Class Mode

Introduction

At the outset of a drug design project, there are persistent challenges in applying rational design principles to the process of creating a novel screening library. The eye of computational and medicinal chemists is limited by both a tendency to process structure-activity relationships in two dimensions and by the relatively small amount of chemical/structural space that can be kept at the forefront of mind. Furthermore, many drug design projects, especially those projects with recently identified or challenging proteins, do not have extensive publications to inform these strategies. In order to create a screening library, we designed a tool that achieves several requirements at the same time: Design novel and synthesizable starting structures that fit well into the binding pocket, retain key interactions, and also include a sampling of the chemical space around those structures to give us a reasonably diverse library of 2k to 5k compounds. Time and computational power are also an issue for many drug design projects, so it is often not possible to use methods such as computational docking to churn through extremely large databases to find possible structures.

In order to speed up this step in drug discovery, we translated chemistry into a form that a quantum annealing computer could use to generate a high-quality screening library. The tool we developed is Quantum-Aided Drug Design (QuADD), a Software-as-a-Service (SaaS) platform for lead-like library generation. QuADD uses quantum annealing computers to find and assemble optimal fragments inside a binding pocket while searching unprecedentedly large chemical spaces extremely rapidly. From an input 3D protein structure with a bound ligand, QuADD generates a ~10^{30} structure search space to find lead-like hits that are novel, bioavailable, and synthesizable. The number of molecules returned by QuADD is tractable for processing in traditional computer-aided drug design programs or by high-throughput screening. QuADD includes key parts of our established pipeline that have been described in Ars Technica, Nature Biopharma Dealmakers [1], and Quantum Insider.

The drug space that can be searched using a template structure, characterized by the size and shape of an idealized Lipinski structure, is approximately 10^{30} using PolarisQB's in-house fragment library. QuADD currently has two curated fragment libraries for constructing molecular structures. Drug Space (*Fragment library 1*) is decomposed from pharmaceutical structures in the World Drug Index [2] (WDI) and

PubChem [3] databases and is used in cases where a more conservative drug space search is desired. Drug-like Space (*Fragment library 2*) is decomposed from a drug-like subset of the ZINC-20 [4] database and is used where the novelty of structures is more of a priority. Both libraries are decomposed using a set of retrosynthetic rules designed to ensure the reconstruction of synthesizable molecules, and both can be searched simultaneously.

QuADD provides four modes of operation: Best-in-Class, First-in-Class, Standard, and Scaffold Hopping. Best-In-Class mode is intended to be used to recover structures close to the template structure, extending outward into drug space from that point. By prioritizing the template molecule's size, shape, and chemical feature characteristics, the user is able to search the drug space at the edges of intellectual property. First-In-Class mode uses the template structure as a marker of the binding pocket and searches drug space based on the physicochemical features of the binding site region itself. It is used where the novelty of structure is a higher priority. The Standard mode of operation strikes a balance between reliable drug-likeness and novelty through consideration of both the template structure and the binding site. Scaffold Hopping mode, as the name suggests, allows the user to retain certain portions of the template structure during library generation. It allows not only the retention of "decorations" in the search for novel cores, it also allows for the retention of any given fragment or set of fragments in its search of drug space. In the following case study, the Best-In-Class mode was used to probe the drug space around sorafenib to find structures that appear in published literature or may be covered by existing patents.

Retrieval of Known Structures from QuADD Best-In-Class

In this test of QuADD's Best-In-Class mode, we chose to examine sorafenib (VEGFR, PDGFR, and RAF kinase inhibitor) in complex with p38 MAP kinase (PDB ID: 3GCS [5], Figure 1). This drug structure was selected as a test case of our study both because it is a marketed drug (Nexavar, *Bayer*) and because it has been used as a test case in the literature describing the BRICS [6] (Breaking of Retro–synthetically Interesting Chemical Substructures) fragmentation rules.

A set of constraints is applied at the outset of every QuADD run to make sure that the set of structures it returns are, by definition, drug-like according to the user. The constraints used are: topological polar surface area (TPSA), number of rotatable bonds (NRot), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), molecular weight (MW), partition coefficient octanol/water (logP), and solubility

(logS). These constraints were set to be sorafenib-like and are shown below in Table 1. An interesting and useful feature of QuADD's drug-likeness constraints lies in that they are soft filters, *e.g.* if too many structures are being rejected with property values close to a cutoff, the cutoff will be extended incrementally.

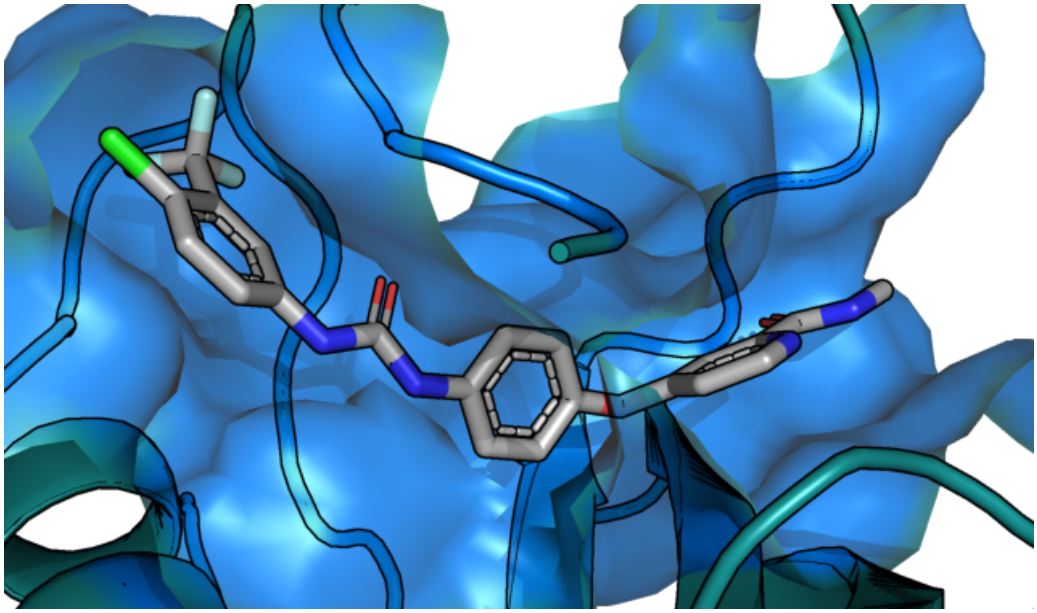
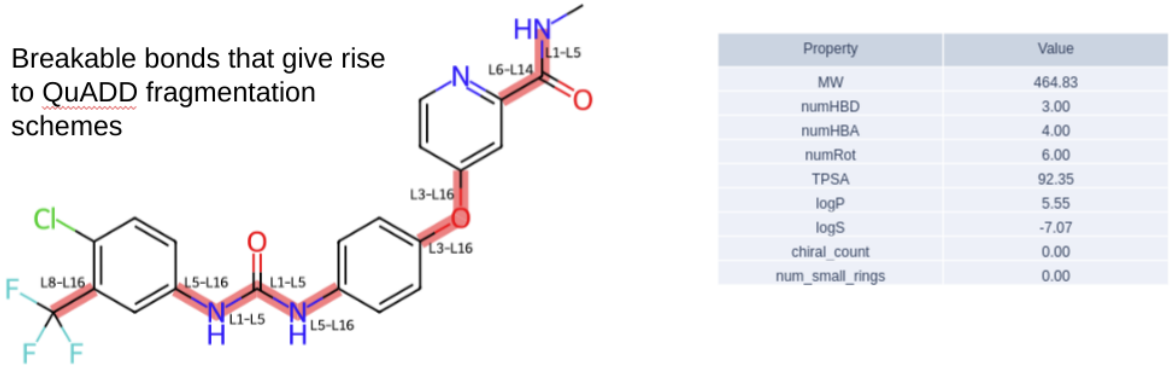


Figure 1. Sorafenib in the ATP-binding site of p38 MAPK

Table 1. Drug-likeness constraints that were used in the case study.

Property	HBA	HBD	nRot	TPSA	MW	logP	logS
Value	≤10	≤5	≤9	≤140	≤500	≤5	≥-7

QuADD generated a set of fragmentation schemes (Figure 2) of the template ligand: four schemes with 3 subregions, four with 4 subregions, and three with 6 subregions. Next, it searched for molecular fragments with optimal properties in the Drug Space fragment library, assembled the set of optimal fragments inside the binding pocket, and applied the drug-likeness constraints, as well as a set of stability and synthesizability filters. The filtered structures are then geometry optimized in the binding site and returned as a library of compounds.



Property	Value
MW	464.83
numHBD	3.00
numHBA	4.00
numRot	6.00
TPSA	92.35
logP	5.55
logS	-7.07
chiral_count	0.00
num_small_rings	0.00

Figure 2. QuADD fragmentation of sorafenib showing template property values

After running QuADD, we recovered a final library of 3,310 structures optimally fitted to the binding site and which passed all filters. This set includes the original sorafenib structure reconstructed from library fragments. Additionally, two QuADD structures appeared in SureChEMBL [7] patent searches and a third structure appeared with a ChEMBL identifier connected to a journal article of Ras kinase inhibitors, shown below in Figure 3.

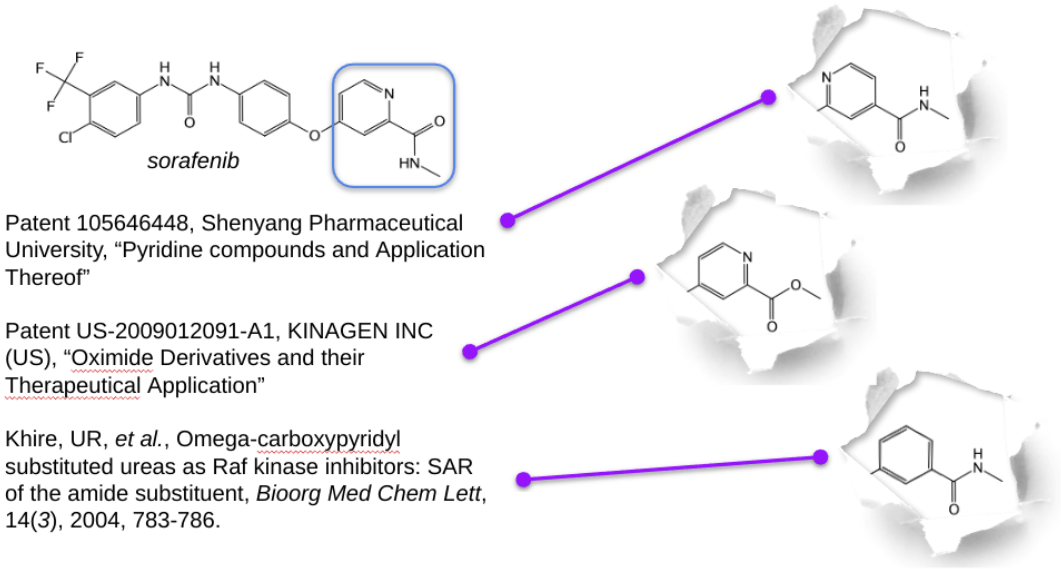


Figure 3. QuADD library structures also reported in literature and patent searches for sorafenib. These structures differed only in the fragment shown in the blue box.

The SureChEMBL search, starting with structures whose binding mode most closely resembled that of the crystal structure revealed one structure covered by multiple patents where the only change to the structure was the conversion of a terminal

N-methyl amide to a methyl ester (the Kinagen, Inc. compound). A second structure was covered by an international patent and differed only in the placement of a ring nitrogen (Shenyang Pharmaceutical University). The third compound, with the pyridine ring replaced by benzene, has an associated series of *N*-alkyl substituted compounds described in the publication [8] with ChEMBL ID: CHEMBL1147437.

QuADD Compounds Are Drug-like

To evaluate the drug-likeness profile of the molecules in the Final QuADD library, we compared the set of properties to that of a representative set of MAPK binders reported in the literature, extracted from the ChEMBL [9] database. We filtered the database finding a set of 190 MAPK binders with IC_{50}/K_i values ≤ 35 nM and calculated their property values. The comparison with the QuADD library is shown below in Figure 4.

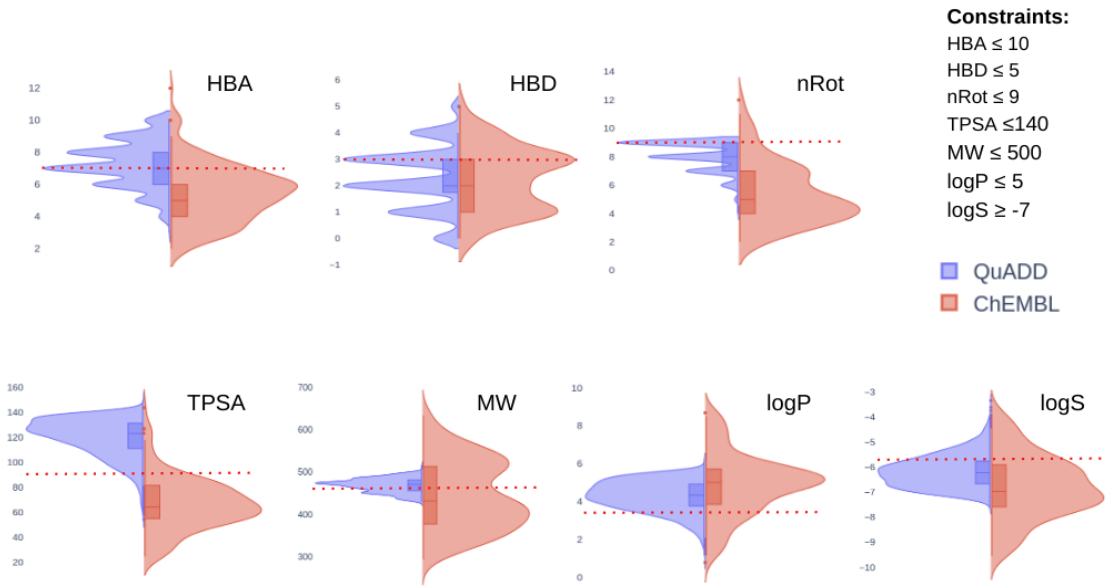


Figure 4. Drug-like properties comparison between 190 ChEMBL ligands and the Final QuADD Best-In-Class library (*sorafenib property values are shown as dotted red lines*)

As illustrated in the violin plots of the various drug-likeness properties, the QuADD molecules remain within the specified constraints and are close to the values calculated for sorafenib (shown as dotted red lines), with the exception of TPSA, which was set to the cutoff of 140 Å^2, commonly used for drug design projects.

QuADD Compounds Bind in a Similar Way as the Template

We studied the protein-ligand interaction profile of the QuADD molecules and found that many bind in a similar manner as the template molecule (co-crystallized ligand). When QuADD molecules undergo further geometry optimization inside the binding site, substantial rearrangements can occur for some structures. To assess the degree to which library structures retain the template binding mode, we calculated Protein-Ligand Interaction Fingerprints (PLIFs) implemented in MOE [10] for the QuADD molecules and for the template molecule. We then computed the Tanimoto coefficients (TCs) between all molecules and the template. The TC can span values from 1 (total similarity) to 0 (total dissimilarity). The histogram of TC values for the QuADD library (Figure 5) shows what appears to be three overlapping distributions of binding modes. There were 50 molecules with TC values equal to 1.00 in the QuADD library.

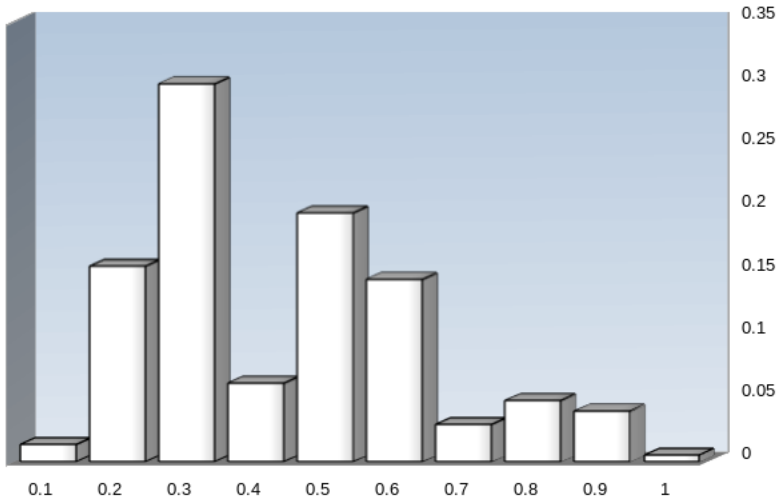


Figure 5. Histogram of Tanimoto coefficients for PLIF similarity to the sorafenib binding mode in the 3GCS crystal structure

In order to find molecules that lie immediately outside of IP, we performed a patent search by sorting a Best-In-Class library by PLIF TC values and working down the list until a match is found.

Interestingly, the molecules found in the SureChEMBL search had TC values lower than 0.90, which we attributed to improved hydrophobic interactions with Val38 after the geometry optimization. As such, these three structures were not included in the set of 50 that we examine more closely in the following sections.

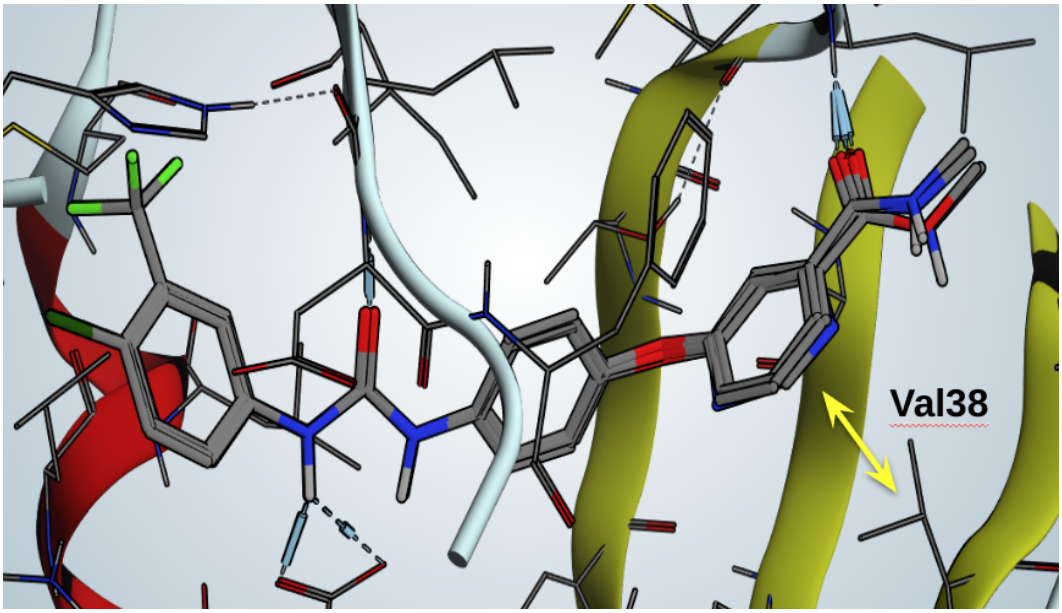


Figure 6. Overlay of the three previously described structures discovered in the QuADD library with sorafenib, showing the small discrepancies in location with respect to Val38 that give rise to PLIF TC < 1.00

QuADD Library Has Good Binders

Next, we compared the predicted binding affinities of QuADD molecules with a set of MAPK binders from the ChEML database using molecular docking with the MOE GB/VI scoring function. Whereas the ChEMBL structures were docked into the binding pocket and scored, the QuADD molecules were minimized in place using the same force field and scored with the same scoring function. As a control for the known weak correlation between docking and experimentally determined activities, we evaluated 2,000 random decoys from the DUD-E [11] database with similar drug-like properties (using the same set of descriptors as the QuADD constraints) and docked those to compare the scores with both sets of binders. Histograms of the GB/VI scores for each set of structures are shown below in Figure 7.

As expected, the carefully curated set of decoys exhibited a similar distribution of binding affinities to both those of the ChEMBL set of known binders and the QuADD library of compounds, which serves to highlight the approximate nature of docking scoring to distinguish strong binders from weak binders. The distribution of binding affinities in the set of QuADD structures is bimodal, with a primary distribution that overlaps with the other sets and a second distribution with a lower average binding

affinity. If docking scores are reliable, this would mean that the best QuADD structures (those in the leftmost distribution) would assay in the low nanomolar range, alongside the best binders reported in the literature, with the most energetic GB/VI scores approaching -10 kcal/mol.

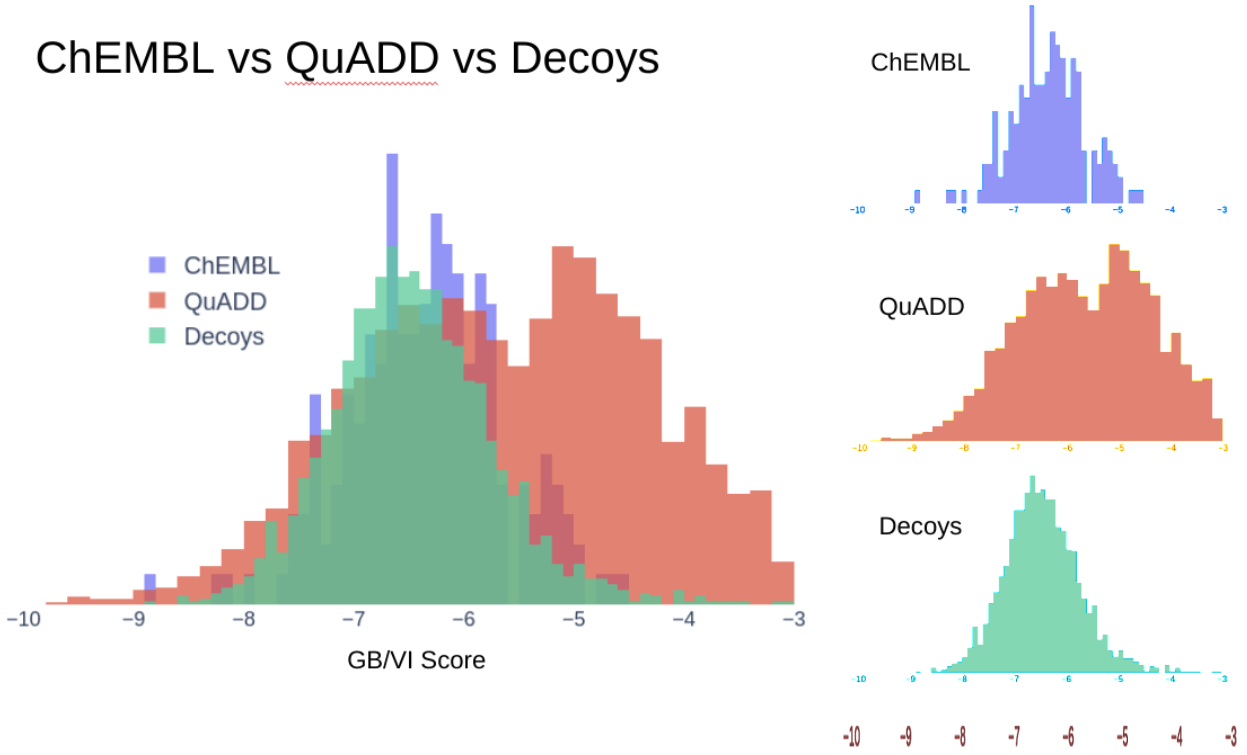


Figure 7. Histograms of predicted binding affinities for QuADD structures, ChEMBL binders, and the set of decoys

Interestingly, when looking at the set of 50 QuADD structures with PLIF TC values of 1.00, the distribution peaks around a GB/VI value of ~ -7.5 kcal/mol and skews left, toward the more energetic end, as shown in the histogram in Figure 8. For comparison, the GB/VI score in place for sorafenib from the crystal structure is -8.19 kcal/mol.

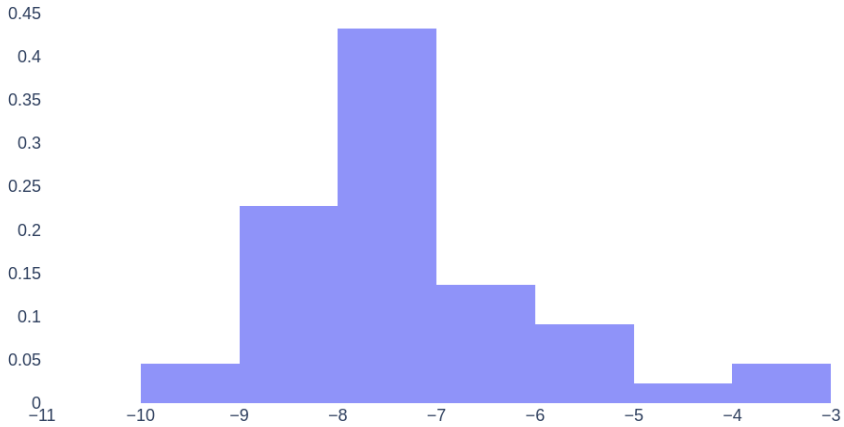


Figure 8. Histogram of GB/VI scores for the set of 50 QuADD structures with a PLIF Tanimoto coefficient of 1.00

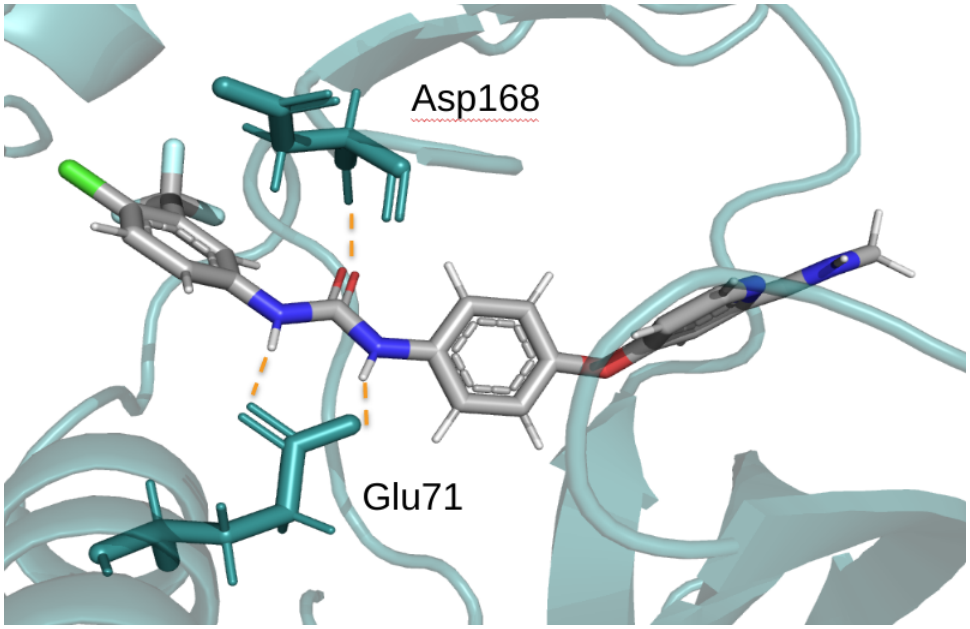


Figure 9. The urea-binding residues, Asp168 and Glu71, in the ATP binding site of p38 MAP kinase with sorafenib bound

A closer examination of the structures in the less energetic (rightmost) distribution reveals that incremental changes in local structure (through bioisosteric replacement) resulted in an accumulation of conformational deformations in the fully assembled structures that precluded optimal interactions with the key residues Asp168 and Glu71. This is the highly energetic region of the MAPK ATP binding site in which the urea moiety of sorafenib interacts (Figure 9). This is the subject of an upcoming PolarisQB case study

that examines the impact of bioisosteric replacement upon the final geometry of optimized structures within QuADD across several targets.

QuADD Best-In-Class Compounds Are Novel

We also measured the similarity/diversity between the QuADD library and the template molecule sorafenib. To do this, we calculated MACCS fingerprints, composed of 166 structural keys, and commonly used to compare molecular similarity [12]. We then computed the Tanimoto coefficients between each pair of fingerprints, taking into account each QuADD molecule against sorafenib. The QuADD library showed significant structural diversity against the template molecule. The majority of the QuADD molecules (99%) yielded TC values against the template between 0.55 and 0.85. Only 0.36% of the QuADD library was similar to the template with a TC value greater or equal to 0.90 (Figure 10). We consistently find that, while QuADD molecules present different scaffolds and bioisosteres when compared with the template, similar 3D pharmacophoric features emerge that allow analogous binding interactions.

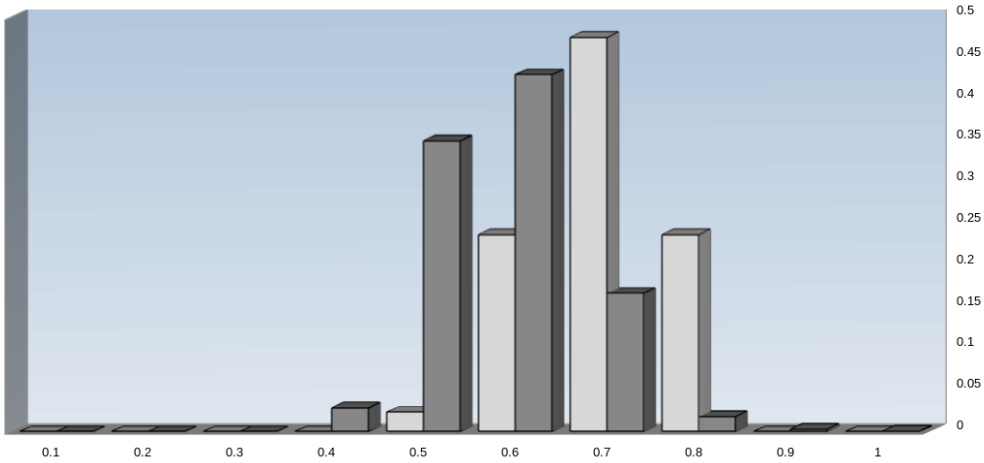


Figure 10. Similarity to sorafenib: histograms of the MACCS fingerprint TCs for the entire QuADD library (*dark gray*) and the set of 50 structures with a binding mode most similar to sorafenib (*light gray*)

Additionally, we computed the TCs for all the pairs of molecules within the QuaDD library and plotted the results in a cluster heatmap (Figure 11). The cluster analysis showed multiple subsets within the library with a high structural similarity, indicating a tighter focus around certain scaffolds and bioisosteres. Typically, QuADD libraries exhibit much lower similarity to the template and are more structurally diverse. Because this study was performed using Best-In-Class mode, we sought to replicate a search of drug

space in the vicinity of the template molecule, sorafenib.

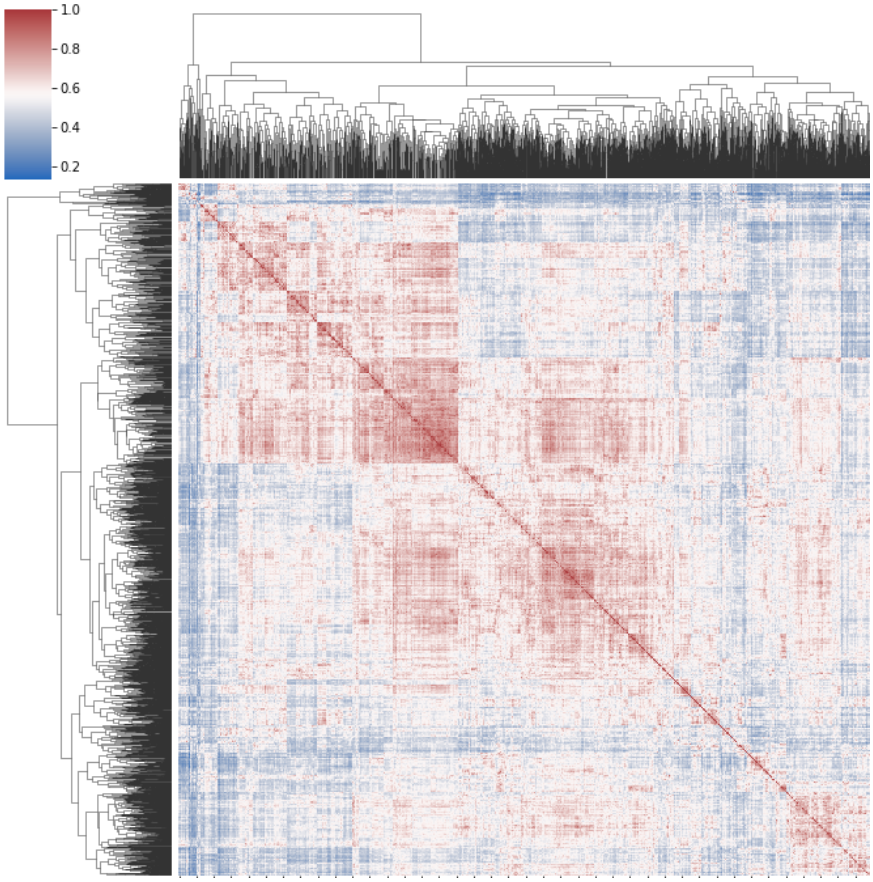


Figure 11. Heatmap of structural similarity within the QuADD Best-In-Class library for sorafenib

And yet, when considering the 50 molecules that reproduce most closely the binding mode of sorafenib in the crystal structure, the majority of molecules within this set are strikingly dissimilar (as shown in the heatmap below in Figure 12). Despite reproducing the crystal structure binding mode with high fidelity, none of these compounds returned any information from our patent searches. Given that the Best-In-Class mode is intended for searching around the perimeters of intellectual property to find novel, yet closely related, structures as a starting point for leads for future best-in-class drugs, any or all of these 50 QuADD molecules satisfy the intended search for appropriate leads.

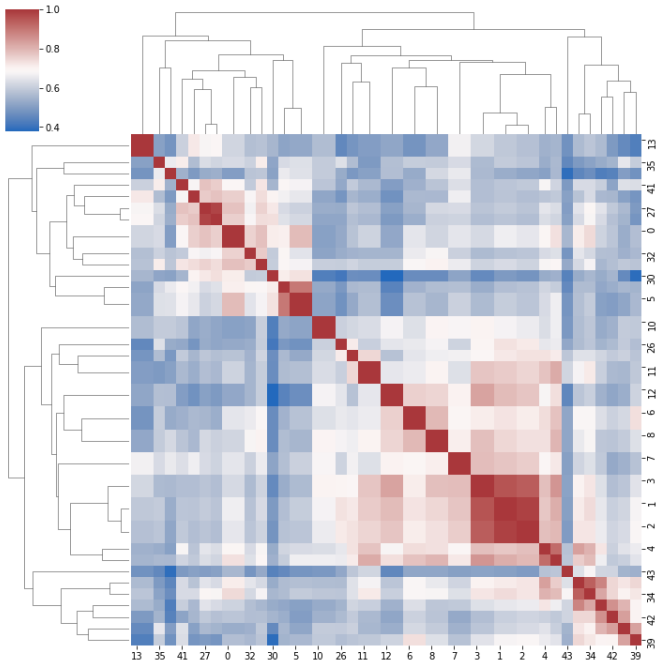


Figure 12. Heatmap of structural similarity for the 50 structures with PLIF similarity scores of 1.00 within the QuADD Best-In-Class library

Conclusions

This case study demonstrates the strength of QuADD's Best-In-Class mode to rapidly search a vast drug space to generate a library of lead-like molecules that bind in a highly similar manner, with similar binding affinities, to a template structure. These molecules offer just enough diversity to exist within and outside the published IP of a given template structure. The performance of QuADD was assessed from multiple perspectives, including binding affinity, geometrical binding profile, similarity/diversity, and drug-like properties. QuADD clearly yielded the best set of binders in a much shorter time frame. In summary, the QuADD Best-In-Class library is composed of molecules with optimal drug-like properties and similar binding profiles as the template but with distinct structural characteristics and maintaining an appropriate degree of diversity - ideal for performing searches around the periphery of intellectual property. These characteristics make QuADD an excellent tool to facilitate drug discovery and development, de-risking preclinical stages and decreasing the time to reach clinical programs.

References

[1] POLARISqb. Harnessing the power of quantum computing for drug discovery. https://www.nature.com/articles/d43747-023-00021-3.

[2] *World Drug Index*, Version 2004, Thomson, Philadelphia, PA (USA), 2004.

[3] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. PubChem 2023 update. *Nucleic Acids Res.* 2023, Jan 6, 51(D1), D1373-D1380.

[4] Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Moroz YS, Mayfield J, Sayle RA. ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J Chem Inf Model.* 2020, 60(12), 6065-6073.

[5] Simard JR, Getlik M, Grütter C, Pawar V, Wulfert S, Rabiller M, Rauh D. Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J Am Chem Soc.* 2009, 131(37), 13286-96.

[6] Degen J, Wegscheid-Gerlach C, Zaliani A, Rarey M. On the art of compiling and using 'drug-like' chemical fragment spaces. *ChemMedChem.* 2008, 3(10), 1503-1507.

[7] Papadatos G, Davies M, Dedman N, Chambers J, Gaulton A, Siddle J, Koks R, Irvine SA, Pettersson J, Goncharoff N, Hersey A, Overington JP. SureChEMBL: a large-scale, chemically annotated patent document database. *Nucleic Acids Res.* 2016 Jan 4, 44(D1), D1220-8.

[8] Khire UR, Bankston D, Barbosa J, Brittelli DR, Caringal Y, Carlson R, Dumas J, Gane T, Heald SL, Hibner B, Johnson JS, Katz ME, Kennure N, Kingery-Wood J, Lee W, Liu X-G, Lowinger TB, McAlexander I, Monahan M-K, Natero R, Renick J, Riedl B, Rong H, Sibley RN, Smith RA, Wolanin D. Omega-carboxypyridyl substituted ureas as Raf kinase inhibitors: SAR of the amide substituent. *Bioorg & Med Chem Let.* 2004, 14(3), 783-786.

[9] ChEMBL Database - EMBL-EBI. European Bioinformatics Institute.

[10] *Molecular Operating Environment* (MOE), 2022.02 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, Canada, 2023.

[11] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J Med Chem.* 2012, 55(14), 6582–6594.

[12] Daylight Chemical Information Systems, Inc. https://www.daylight.com/dayhtml/doc/theory/theory.finger.html