JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Application Note

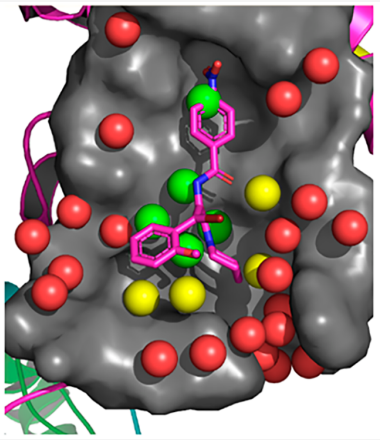
BMaps: A Web Application for Fragment-Based Drug Design and Compound Binding Evaluation

Daniel R. Bryan, John L. Kulp, Jr., Manoj K. Mahapatra, Richard L. Bryan, Usha Viswanathan, Micah N. Carlisle, Surim Kim, William D. Schutte, Kevaughn V. Clarke, Tony T. Doan, and John L. Kulp, III*

Cite This: https://doi.org/10.1021/acs.jcim.3c00209 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Fragment-based drug design uses data about where, and how strongly, small chemical fragments bind to proteins, to assemble new drug molecules. Over the past decade, we have been successfully using fragment data, derived from thermodynamically rigorous Monte Carlo fragment−protein binding simulations, in dozens of preclinical drug programs. However, this approach has not been available to the broader research community because of the cost and complexity of doing simulations and using design tools. We have developed a web application, called BMaps, to make fragment-based drug design widely available with greatly simplified user interfaces. BMaps provides access to a large repository (>550) of proteins with 100s of precomputed fragment maps, druggable hot spots, and high-quality water maps. Users can also employ their own structures or those from the Protein Data Bank and AlphaFold DB. Multigigabyte data sets are searched to find fragments in bondable orientations, ranked by a binding-free energy metric. The designers use this to select modifications that improve affinity and other properties. BMaps is unique in combining conventional tools such as docking and energy minimization with fragment-based design, in a very easy to use and automated web application. The service is available at https://www.boltzmannmaps.com.



INTRODUCTION

In the last two decades, *in silico* fragment-based drug design (FBDD) has emerged as an alternative for high throughput screening (HTS) in early phase drug discovery and lead optimization programs. HTS involves large-scale screening of compound libraries but is not efficient enough to cover a large part of the drug-like region of the chemical space. FBDD offers numerous advantages over HTS and virtual screening (VS) methods. These include reliably producing fragment hits that are less complex but with high binding efficiency; straightforward modification and optimization of fragment hits to develop drug-like compounds with favorable properties; low molecular weight; and sampling a larger portion of chemical space.[1] Experimental detection of low-affinity fragment hits in various biophysical/biochemical screening makes FBDD a complementary strategy for structure-based drug discovery.[2] It has been applied for hard-to-drug therapeutic targets where other methods met with limited success. The adoption of *in silico* protein−ligand modeling techniques, especially fragment-based methods, has grown steadily over the last 25 years, in both academia and industrial research projects. An online survey shows an increasing trend of FBDD practice by industrial as well as nonprofit users in the past 10 years.[3] We fully expect the percentage of projects involving *in silico* methods will continue to increase. However, the expertise required and the cost of simulations and tools supporting *in silico* methods have been limiting. Our goal is to dramatically change that with an innovative web application for FBDD.

Chemical fragments, as considered here, are low molecular weight (150−300 Da) chemical components that can be synthetically combined into larger drug molecules. They typically have low-affinity binding at many sites yet possess high ligand efficiency (ratio of biological activity to number of heavy atoms). By studying how and where these fragments bind to various sites on the protein surface, the chemist can develop ideas for the druggability of new target proteins, identify inhibitory sites of protein−protein interactions (PPI), and locate hot spots supporting binding of diverse chemistry. Off-target interactions—when a drug binds to proteins other than those it was meant for—can cause toxicity. Toxicity is a leading cause of attrition in clinical trials.[4] Undesired

Received: February 9, 2023

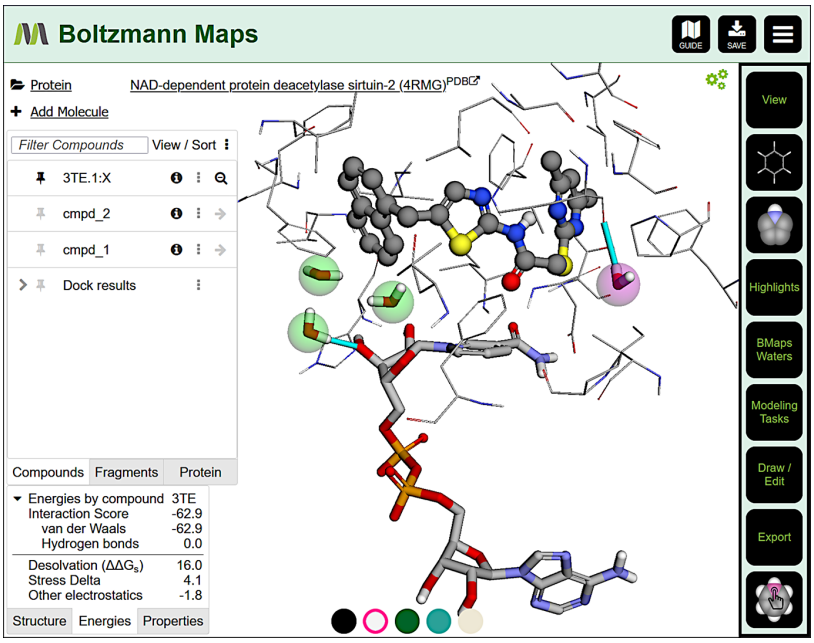


Figure 1. BMaps user interface showing the 4RMG binding site, energy table, and SACP waters colored by chemical potential.

interaction at off-target sites can also be studied and optimized by fragment-based analysis of a ligand and its constituent fragments. FBDD has made significant contributions to the discovery of four approved drugs and over 40 clinical trial candidates.[5−7] These promising cases have compelled researchers to explore more challenging aspects of protein−protein and protein−DNA interactions.[6]

When fragment binding poses are determined via computational methods, a key need is to evaluate and rank the poses. Ranking of fragment poses by interaction energy or empirical scoring usually suffers from suboptimal accuracy resulting in an excessive percentage of false positive predictions. This motivated us to pursue more accurate and robust techniques such as Grand Canonical Monte Carlo (GCMC) simulations. Such simulations offer a thermodynamically rigorous and efficient method to derive statistical distributions of fragment poses ranked by a binding free energy expressed as excess chemical potential (average free energy per molecule). An ensemble of fragment poses is generated across the protein surface by using the Monte Carlo algorithm[8] which provides comprehensive sampling. The algorithm accurately predicts a fragment binding free energy, including entropy, from a Boltzmann distribution as a function of excess chemical potential. The Simulated Annealing of Chemical Potential (SACP) technique forcefully inserts fragments into all the binding sites of the protein. It then removes them, with gradual annealing of chemical potential, to identify the binding sites where fragment and/or water molecules are tightly bound.[9] Thus, the combined GCMC-SACP technique provides an approach from first-principles for generating accurate, reliable fragment maps ranked by a free energy metric.

The GCMC-SACP technique is especially valuable for calculating robust water maps. Water mapping tools for predicting structured water molecules usually deploy scoring methods, grid-based sampling, or molecular dynamics simulation methods, with less accuracy.[10] By applying GCMC-SACP, our research group has developed a water map tool that can discover multibody water−macromolecule interactions and provide accurate prediction and ranking of structured water molecules based on their free energy.

Applications of GCMC include numerous targets for drug design by academic and industrial researchers, as recorded by Bodnarchuk et al.[11] GCMC-SACP has been successfully utilized for distinguishing hydration propensity of water molecules in major and minor grooves of DNA,[12] discovery of a potent non-ATP p38 inhibitor,[13] detecting PPI (MDM2/MDM4-p53), locating binding subpockets in elastase enzyme, and accurately predicting multibody water triplet in Bovine pancreatic trypsin inhibitor.[9] When the traditional FBDD approach was not successful in designing or screening any compounds to break the complex interactions between PCSK9 (proprotein convertase subtilisin/kexin type 9) and LDLR (low-density lipoprotein receptor), SACP quickly located the high-affinity binding sites in the complex PPIs between PCSK9 and LDLR. Using the insights from the calculations, we were able to design potent, small molecule inhibitors by fragment merging.[14] Designing a small molecule renin inhibitor with good oral bioavailability, a challenge for structure-based or fragment-based design, was successfully carried out by utilizing SACP augmented with the constrained-fragment annealing (CFA) method. Many small molecule renin inhibitors were designed, synthesized, and found to possess good oral bioavailability and physiochemical parameters.[15]

Various fragment-based designing tools such as LUDI,[16] CCLD,[17] MCSS,[18] GANDI,[19] PRO_Ligand,[20] HOOK,[21] UCSF DOCK,[22] SEED,[23] Skelgen,[24] eHits,[25] BREED,[26] CrystalDock,[27] Lea3D,[28] and FragFEATURE[29] are available, differing from each other in the type of fragments and scoring functions used. To use fragment techniques in a simpler way, several web services have also been developed. These can be classified as geometry-based (e.g., HybridSim-VS,[30] FragRep,[31] e-LEA3D[32]), energy-based (e.g., ACFIS,[33] FragRep,[31]

https://doi.org/10.1021/acs.jcim.3c00209
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

FTMap[34]), grid-based (e.g., FTMap), Artificial Intelligence (AI)-based (e.g., DEEPScreen,[35] DeepScreening,[36] DeepFrag[37]), and docking-based (e.g., CHARMMing,[38] Probis-CHARMMing,[39] e-LEA3D[32]). Although many open source and commercial tools exist for *in silico* drug design, barriers to entry may include complexity of use, restricted access and high cost, the need for adequate storage and compute capacity, or having to download and configure tools. These barriers could be overcome by integrating and facilitating the smooth usage of the various required *in silico* tools in a single web application.

To address these needs, we have developed BMaps: a comprehensive fragment-based web application supported by a large repository of simulated fragment binding data, a geometric search algorithm to find the best binding poses for compound modification, accurate water molecule mapping technology, and a high-performance Monte Carlo simulator for modeling protein−ligand and protein−water interactions using the SACP technique. The web service supports computations for docking, energy minimization, lead optimization, fragment mapping, fragment search, hot spot analysis, and, of particular note, water binding maps. BMaps allows users to carry out fragment/water simulations on structures and fragments of their choice and then apply the insights to the design and optimization of new drug leads. This new tool brings *in silico* FBDD techniques directly into the compound design and evaluation process.

RESULTS

The BMaps web application provides a unified platform for a range of FBDD activities, including modeling targets, compounds, and fragments; evaluating and modifying compounds; running fragment and water simulations; and visualizing simulation results and applying them to drug design. The main components of the user interface are discussed in this section, followed by an example workflow for reproducibility (Figure 1).

A detailed technical explanation of how BMaps imports, modifies, evaluates, and exports molecular data can be found in Supporting Information Section 1 − Methods.

1. Visualizing and Evaluating Compound Binding.

1.1. Visualizing Protein Targets. A user can choose a target protein from the repository of proteins with precomputed fragment binding data, import from the Protein Data Bank (PDB)[40] or AlphaFold,[41] or upload their own structure (Figure S1, Supporting Information Section 2 − User Interface Figures). The precomputed repository includes over 550 therapeutically relevant structures which were manually optimized and simulated with fragments. These prepared structures each have precomputed druggable hot spots, water maps, and more than 100 fragment maps.

If the user's target has not yet been prepared, the user can directly import a structure from the PDB. In the absence of a 3D crystal structure, a predicted structure may be imported from AlphaFold, the AI-based predicted protein database, by using its UniProt ID.[42] The highest-confidence AlphaFold residues will be colored dark blue and are a reasonable structure to design against. When users need to work with proprietary structures, they may upload in PDB or CIF formats.

For all imported structures, Amber force field parameters and AM1-BCC[43] charges are assigned by AmberTools,[44] and any crystallization artifacts, cofactors, ions, and crystal waters are interpreted. These structures are initially devoid of fragment binding data, so the user would then run fragment or water simulations to get information about hot spots, fragments, and water maps.

1.2. Adding Compounds. Compounds may be imported by file drag-and-drop, pasting compound data into the workspace, drawn with the onboard compound editor (2D sketcher), or added from a user's library at CDD Vault[58] (Figure S2). If necessary, 3D coordinates will be generated for imported compounds. The user chooses how to position the compounds: by aligning to a cocrystal ligand or another compound, placing outside the protein to be docked later, or keeping existing 3D coordinates. The compounds are automatically prepared for energy calculations. The system can detect steric clashes arising between the ligand−protein system and display a warning. In this situation, energy minimization or redocking can help find a comfortable pose.

Docking: After importing compounds, the user may dock them against the target, specifying a grid box centered around a crystal ligand, a hot spot, or any selected atoms. By default, the grid box is optimally sized for the docked compound (Figure S3), based on a technique published by Feinstein et al.[45] Autodock Vina[46] is used to carry out the docking. After completion, individual binding poses can be visualized along with their energy values.

1.3. Evaluating Compounds. Computation of ligand binding energy is important in guiding fragment-based designs, optimization, and binding affinity prediction. Therefore, an interaction energy calculation is made available for proteins and compounds with a single click. Interaction energies are computed with OpenMM,[47] parsed into intuitive components, and presented with our own energy calculations for desolvation[48] and hydrogen bonds. Energy calculation helps users observe the effects of changes made during fragment design and modifications. The energy window separately displays an interaction score (combination of hydrogen bonds and van der Waals energy contribution), desolvation cost ($\Delta\Delta G_s$), stress delta due to change in internal energy between bound and unbound conformation, and other electrostatic contributions. As an aid to intuition, semiquantitative assessments of hydrogen bonds, desolvation, and hydrophobic interaction energies are also available in the form of 3D visualizations. After fragment modification, before-and-after comparison of energy values and other physio-chemical properties aid in better judgment for selecting a suitable modification. In Figure S4, the cocrystal ligand (ligand ID: 4K6) of the factor IXa structure (PDB ID 4YZU) has been modified with a benzisoxazole fragment, resulting in a new compound (4K6+benzisoxazole) with a better interaction score than the former.

2. Applying Fragment Binding Data to Compound Design. The availability of fragment binding data for a protein target enables new insights for designing and optimizing compounds.

2.1. Hot Spots. Hot spots—sites on the biomolecule structure that support strong binding of diverse chemical motifs—are derived by applying cluster analysis to the simulated fragment binding data. The three principles of cluster analysis are (i) diverse chemical fragments, (ii) strong binding, and (iii) water exclusion (do not compete strongly with water molecules).[9,49] GCMC-SACP fragment maps accurately predict the location of hot spots and druggable sites governing various protein−ligand, protein−protein, and

https://doi.org/10.1021/acs.jcim.3c00209
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

protein−DNA/RNA interactions. Our hot spot identification technique has been successfully applied for characterizing a number of biomolecular recognition events involving RecA, PDF, DHFR, Elastase, p53-MDM2, and HIV TAR.[9] Figure S5 depicts the hot spot residues of the SIRT2 protein along with various fragment clusters ranked according to their average excess chemical potential.

2.2. Water Maps. Understanding the spatial distributions and thermodynamic properties of water molecules can provide valuable insights into ligand binding and can motivate compound changes that could increase affinity by displacing a weakly bound water molecule or by avoiding tightly bound waters. GCMC simulations are also known for accurately predicting multibody water interactions, which is a limitation of other water mapping technologies.[50] Figure S6 illustrates the water map and multibody water interactions around the binding site of factor IXa (4YZU) with crystal ligand 4K6. A semiquantitative visualization of hydrogen bonds allows distinguishing the hydrogen bonds in terms of their strength.

The simulated waters are highlighted to indicate the cost of desolvation. Displacing bulk water (green halos) will give an entropic benefit. Displacing tightly bound or structured waters (red halos) also provides an entropic benefit, but the compound will need sufficient interaction properties to overcome the cost of desolvation. When viewing the water map in the full protein view, only the red halos are shown, providing a step toward identifying water networks.

All water maps generated on BMaps prepared structures have been run *without* the crystal ligand. However, it is possible to run new water simulations with a compound in place.

2.3. Fragment Growing and Searching. Fragment data include numerous fragment poses all around a protein surface. During compound design, these fragment poses can be searched to find high-affinity poses of suitable geometry for substitution into a compound. Fragment search results are displayed in a table with the binding score (binding free energy metric incorporating configurational entropy and desolvation cost) and other properties (Figure S7). Visualization of fragment binding poses, overlapping with the compound or a part of it, provides a clear picture to carry out fragment-based modifications. "Grow with Distribution" displays the cloud of simulated poses and gives an opportunity to select a pose from a distribution with a desired excess chemical potential value. When the user chooses a fragment to merge into the compound, the assembly happens automatically, and the resulting compound is brought into the workspace.

Individual fragment poses can also be brought individually into the workspace using the "Search Nearby" feature (Figure S8). This is similar to "Fragment Grow," but it does not modify the compound; it just inserts the fragment itself. For projects involving de novo design, a Search Nearby in a hot spot can be used as a starting point to find suitable fragment hits, which can then be elaborated to design high-affinity ligands.

2.4. Fragment Map Summary. This is a composite, high-level view of all the binding data for a fragment, where the best pose (highest affinity) is retained from a distribution with lowest chemical potential. Summary fragment poses can be filtered by free energy value (Figure S9).

3. Generating New Fragment Binding Data. BMaps allows users to run new fragment simulations, so they can apply insights from fragment binding data to their own targets of interest.

3.1. Simulation Workflow. The protein system is first prepared by specifying the structural features to be included in the simulation—chains, ions, cofactors, and for water maps, an optional compound. Then the fragments are chosen, and the simulation job is submitted. After completion, the structure can be loaded into the design environment, providing a visualization of hot spots, water maps, and fragment maps.

3.2. Specifying Fragments. The decision about which fragments to run depends on whether the objective is to identify hot spots, to find general common fragment growing opportunities, or to probe the binding site for specific fragment growing opportunities. To support a wide range of uses, BMaps incorporates several fragment collections from literature and industry into a database of more than 4000 chemically diverse fragments, plus other small molecules. The collections include BMaps' own fragments, "MiniFrags"[51] and "Rings in Drugs"[52] literature sets, and subsets of Maybridge fragments[53] and other libraries.[54−56] In the "Library search" tool (Figure S10), fragment libraries may be browsed or filtered based on name, chemical properties, element symbols, or specific motif.

To get started with fragment-based design on new structures, several recommended fragment sets are provided: "BMaps Starter Fragments", "BMaps Clustering Fragments," and "BMaps Favorites." The details of the built-in fragment collections can be found in Supporting Information Section 3 − BMaps Fragment Listing.

3.3. User-Defined Fragments. User-defined fragments may be configured by drawing in the 2D sketcher or by importing an SDF file. A bioactive molecule or a compound of interest can also be automatically broken down to generate new smaller fragments. Fragments of interest can be grouped to create customized fragment sets to be used for fragment searching, modification/optimization, and running simulations. This enables medicinal chemists to organize their fragments per desired chemical moiety and customized features.

4. Example Workflow. The following workflow demonstrates how BMaps can be used to generate and utilize fragment data for hit-to-lead projects. This example concerns PDB ID 4YZU, a crystal structure of the human factor IXa responsible for blood coagulation. Instructions for reproducing this workflow are available in Supporting Information Section 4 − BMaps Workflow.

Choose the simulation structure: Import the original 4YZU structure from the PDB and stage for simulation.

Choose fragments to simulate:

- "BMaps Clustering Fragments", for hot spots
- Water, for water map
- Benzisoxazole, for fragment growing

Run the simulation and load results: The fragments should all finish within 12−48 h, with waters running the longest.

View hot spots and water map: The "Hot Spots" highlight displays the hot spots determined via automatic clustering analysis. Precomputed BMaps structures have had manual tuning of the cluster analysis (radius, diversity, etc.). In the 4YZU case, the automatic clustering parameters find the hot spot around the benzimidazole. These hot spots are good places to look for compound modifications.

"BMaps Waters" button displays the simulated water map, with waters highlighted in red (structured waters) or green (bulk waters) to indicate the relative cost of desolvation. The

https://doi.org/10.1021/acs.jcim.3c00209
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

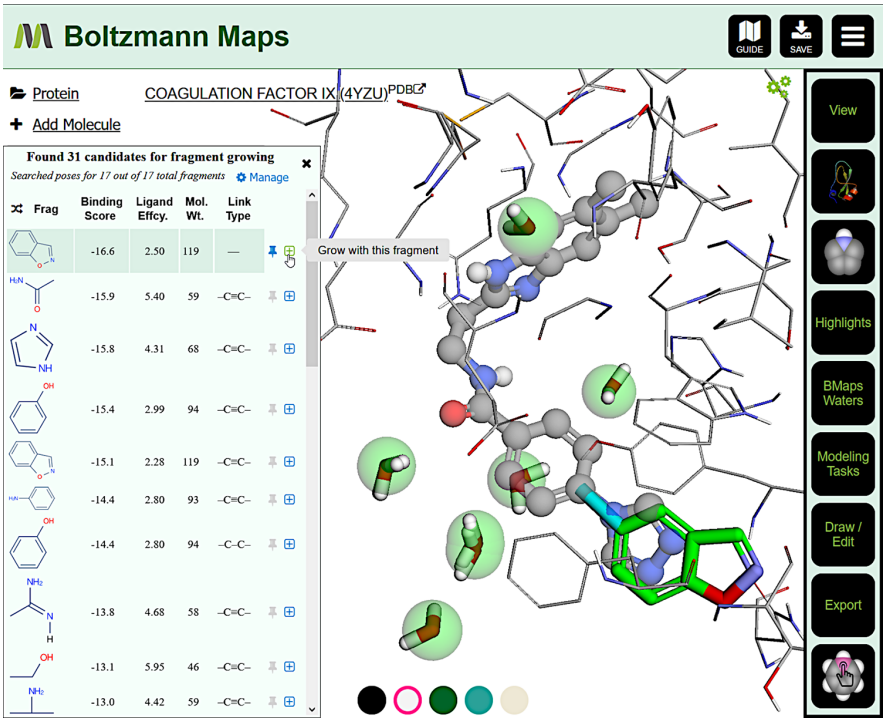


Figure 2. Fragment Grow suggests benzisoxazole is a good option to replace triazole in 4YZU. Hovering on the substitution option reveals the position of the fragment to substitute and the very good bond angle. Simulated waters are also seen.

position, highlight color, and hydrogen bonds of binding site waters may give new ideas to modify the ligand (Figure S11).

Use a "fragment grow": This is to modify the compound. When searching for fragments along the triazole vector, benzisoxazole with a single bond link type has the top binding score and a very good bond angle (Figure 2, Figure S11).

Minimize and compare energies: After the modified compound has been assembled, energy minimization can find the optimal pose and report the change in interaction energy. When viewed alongside the ligand in the energy table, we can see that the benzisoxazole fragment substitution has an interaction energy score of −75.5 which is more favorable compared to the triazole's −67.4 (Figure S12) and would be a suitable replacement.

Dock the modified ligand: Do this to see if the compound changes pose. When the dock results finish and are minimized, we see that, in this example, the top docking result has the exact pose as the modified ligand (Figure S12).

This simple example illustrates how fragment data can be generated and used to find compound modifications that lead to improved energies, potency, or novelty.

DISCUSSION

By offering an integrated web-based environment for these powerful tools, BMaps aims to bring *in silico* FBDD techniques to a wider range of researchers.

A key goal is to simplify compound evaluation workflows. The design environment provides a single interface to visualize a binding site, dock and modify compounds, and view interaction energies. Backend tooling handles the data management for the different kinds of computation—assigning force field parameters with AmberTools,[44] docking via Autodock Vina,[46] energies via OpenMM,[57] and fragment simulation with our simulator. The resulting compound or fragment data can be exported in various formats for use with other tools. BMaps also uses interfaces provided by several external web services. CDD Vault[58] integration allows users to transfer compounds between BMaps and their storage at CDD Vault. Reaxys,[59] PubChem,[60] and Pharmit pharmacophore screening[61] can all receive searches from BMaps, so researchers can easily access information from those services.

Supplementing these compound evaluation features, BMaps provides statistically robust and accurate water and fragment maps, with hot spot analysis and specific fragment-based tools. The GCMC algorithm is a thermodynamically principled way of generating an ensemble of fragment binding poses and ranking the distributions according to configurational free energy. This approach is more reliable than conventional docking or probing methods. We combine GCMC with SACP—including an adaptive annealing strategy—to generate fragment maps with greater efficiency and accuracy. This enables the researcher to target a broader range of biomolecular interactions with minimal effort.

Fragment binding data can offer various kinds of insight into design: suggestions for starting points in de novo projects, modifications in hit-to-lead optimization, or even hints about the selectivity of a ligand or off-target toxicities associated with mutations. To explore selectivity, a drug can be broken down into various fragments, and those fragments simulated against the structures of interest. In simulation, the fragments bind to and thus explore different binding cavities of a protein and show possible interactions. Having fragment maps on multiple proteins can assist in the binding site analysis of the fragments, giving evidence about selectivity. Moreover, selected fragments can be analyzed around the hot spots which govern various

https://doi.org/10.1021/acs.jcim.3c00209
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

protein−protein and protein−DNA/RNA interactions, to provide important clues for mutational avoidance. Fragment maps have been successfully applied in over a dozen lead identification and optimization programs where novel, low molecular weight fragment hits with <25 μM activity were discovered by testing only 7−20 fragments. In four of the projects, leads with <100 nM potency were identified by synthesizing and testing only 17−40 compounds per project (unpublished research, Supporting Information Section 5 − Internal Research Summary). Small molecule renin inhibitors with nanomolar potency and good oral bioavailability were achieved by preparing only a few dozen compounds.[13]

To broadly apply these insights from fragment simulations, BMaps offers precomputed data for over 550 therapeutically relevant targets. For each target, protein structures were prepared and run with 133 fragment and 3 water simulations, and hot spots were identified using our previously published method.[49] These targets were selected from approved drugs or clinical trial candidates obtained from fragment-based drug discovery programs, drugs listed on the PDB[40] or Binding DB[62] web sites, targets being pursued by venture capital-funded pharmaceutical and biotech companies, targets for prospective customers, and many SARS-CoV-2 proteins. To the best of our knowledge, no other group has developed a public repository of fragment and water map data of such magnitude. Until now, the best fragment maps have been confined to proprietary enterprises, due to the various barriers previously mentioned, but BMaps aims to democratize the availability of accurate fragment and water maps. We will continue to augment our repository of open access fragment maps.

The COVID-19 pandemic situation demanded accelerated research for discovering effective SARS-CoV-2 inhibitors. Hence, our group generated fragment binding data for various SARS-CoV-2 proteins on an emergency basis. In addition, a number of commercially available sample compounds were identified as high-affinity starting points for inhibitors targeting the hot spots of the concerned target structures.

FUTURE DIRECTIONS

Important next steps for BMaps include integrating tools and techniques for free energy calculations and expanding our capabilities for visualizing and evaluating the selectivity of ligands over a range of targets. We will continuously expand the fragment mapping data while making it easier to derive useful information for designing drug leads. We hope to continue working with practitioners to identify additional fragment collections and other associated data that could become valuable in expanding access to these important techniques. Longer term, incorporating protein flexibility into BMaps' use of modeling tools will be important.

CONCLUSION

The BMaps system, an *in silico* fragment-based drug design platform, is now broadly available to the research community through a web application backed by cloud computing. It provides access to an unrivaled repository of fragment binding data from thermodynamically principled GCMC-SACP fragment−protein simulations, including precomputed fragment maps, druggable hot spots, and water maps. This fragment data help to understand the binding pocket and can motivate compound modifications, while the integrated computational tools help explain the energetic details of compound binding and physiochemical properties. The BMaps platform will benefit casual users, medicinal chemists who are new to *in silico* methods, and experienced computational chemists alike.

ASSOCIATED CONTENT

Data Availability Statement

The BMaps web application is publicly available at www.boltzmannmaps.com. Visualization-only access to the 550+ precomputed structures is available without login at www.boltzmannmaps.com/preview. This includes viewing water maps, hot spots, and fragment summary maps, but not computation-oriented features like fragment growing, energy minimization, or docking. Access to all computation features is available with an open access account, but operations that spin up new compute-cloud resources (docking and fragment simulations) have monthly limits. Additional account types are available for users who wish to perform more simulations.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c00209.

Methods, user interface figures, BMaps fragments listing, BMaps workflow, and internal research summary (PDF)

AUTHOR INFORMATION

Corresponding Author

John L. Kulp, III − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States;* orcid.org/0000-0003-2586-9207; Email: john.kulp@coniferpoint.com

Authors

Daniel R. Bryan − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States;* orcid.org/0009-0000-9208-5109

John L. Kulp, Jr. − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States; Zymergen, Inc., New York, New York 10016, United States*

Manoj K. Mahapatra − *Kanak Manjari Institute of Pharmaceutical Sciences, Rourkela 769015 Odisha, India*

Richard L. Bryan − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Usha Viswanathan − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Micah N. Carlisle − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Surim Kim − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States; Zymergen, Inc., New York, New York 10016, United States*

William D. Schutte − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Kevaughn V. Clarke − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Tony T. Doan − *Conifer Point Pharmaceuticals, Doylestown, Pennsylvania 18902, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c00209

Funding

The BMaps web application was developed through NIH SBIR Grants R43GM109549 and R44GM109549.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Frank Guarnieri for his collaboration in the invention and implementation of the GCMC-SACP algorithm. Frank Hollinger and Dimitris Gazgalis provided helpful discussions. We also recognize technical support from Charlie Weatherall at Collaborative Drug Discovery, Matthew Clark at Elsevier, and Doug Koes at the University of Pittsburgh. Pat Walters' integrated ASAP platform at Vertex provided inspiration for a tool that integrates services by easy sharing of compound structures. Thanks to Jacob Chappell for his help with the manuscript edits.

■ REFERENCES

(1) Mortier, J.; Rakers, C.; Frederick, R.; Wolber, G. Computational Tools for in Silico Fragment-Based Drug Design. *Curr. Top Med. Chem.* **2012**, *12*, 1935−1943.
(2) de Souza Neto, L. R.; Moreira-Filho, J. T.; Neves, B. J.; Maidana, R. L. B. R.; Guimarães, A. C. R.; Furnham, N.; Andrade, C. H.; Silva, F. P. In Silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery. *Front Chem.* **2020**, *8*, Review.
(3) *Practical Fragments*. http://practicalfragments.blogspot.com/2019/12/poll-results-affiliation-and-fragment.html (accessed 2022/12/22).
(4) Lin, A.; Giuliano, C. J.; Palladino, A.; John, K. M.; Abramowicz, C.; Yuan, M. L.; Sausville, E. L.; Lukow, D. A.; Liu, L.; Chait, A. R. Off-Target Toxicity Is a Common Mechanism of Action of Cancer Drugs Undergoing Clinical Trials. *Sci. Transl Med.* **2019**, *11*, na DOI: 10.1126/scitranslmed.aaw8412.
(5) Murray, C. W.; Rees, D. C. The Rise of Fragment-Based Drug Discovery. *Nat. Chem.* **2009**, *1*, 187−192.
(6) 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.
(7) Erlanson, D. Fragments in the Clinic, 2020. *Practical Fragments*. https://practicalfragments.blogspot.com/2020/03/fragments-in-clinic-2020-edition.html (accessed 2023/05/01).
(8) Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J. Grand Canonical Monte Carlo Simulation of Ligand-Protein Binding. *J. Chem. Inf Model* **2006**, *46*, 231−242.
(9) Kulp, J. L., III; Cloudsdale, I. S.; Kulp, J. L., Jr; Guarnieri, F. Hot-Spot Identification on a Broad Class of Proteins and Rna Suggest Unifying Principles of Molecular Recognition. *PloS One* **2017**, *12*, No. e0183327.
(10) Nittinger, E.; Gibbons, P.; Eigenbrot, C.; Davies, D. R.; Maurer, B.; Yu, C. L.; Kiefer, J. R.; Kuglstatter, A.; Murray, J.; Ortwine, D. F.; et al. Water Molecules in Protein−Ligand Interfaces. Evaluation of Software Tools and Sar Comparison. *J. Comput-Aided Mol. Des* **2019**, *33*, 307−330.
(11) Bodnarchuk, M. S.; Packer, M. J.; Haywood, A. Utilizing Grand Canonical Monte Carlo Methods in Drug Discovery. *ACS Med. Chem. Lett.* **2020**, *11*, 77−82.
(12) Guarnieri, F.; Mezei, M. Simulated Annealing of Chemical Potential: A General Procedure for Locating Bound Waters. Application to the Study of the Differential Hydration Propensities of the Major and Minor Grooves of DNA. *J. Am. Chem. Soc.* **1996**, *118*, 8493−8494.
(13) Guarnieri, F. Designing an Orally Available Nontoxic P38 Inhibitor with a Fragment-Based Strategy. In *Fragment-Based Methods in Drug Discovery*; Klon, A. E., Ed.; Springer New York, 2015; pp 211−226.
(14) Guarnieri, F.; Kulp, J. L., Jr; Kulp, J. L., III; Cloudsdale, I. S. Fragment-Based Design of Small Molecule Pcsk9 Inhibitors Using Simulated Annealing of Chemical Potential Simulations. *PloS One* **2019**, *14*, No. e0225780.
(15) Cloudsdale, I. S.; Dickson, J. K., Jr; Barta, T. E.; Grella, B. S.; Smith, E. D.; Kulp, J. L., 3rd; Guarnieri, F.; Kulp, J. L., Jr Design, Synthesis and Biological Evaluation of Renin Inhibitors Guided by Simulated Annealing of Chemical Potential Simulations. *Bioorg. Med. Chem.* **2017**, *25*, 3947−3963.
(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.
(17) Caflisch, A. Computational Combinatorial Ligand Design: Application to Human A-Thrombin. *J. Comput. Aided Mol. Des* **1996**, *10*, 372−396.
(18) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins* **1991**, *11*, 29−34.
(19) Dey, F.; Caflisch, A. Fragment-Based De Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf Model* **2008**, *48*, 679−690.
(20) Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. Pro-Ligand: An Approach to De Novo Molecular Design. 1. Application to the Design of Organic Molecules. *J. Comput. Aided Mol. Des* **1995**, *9*, 13−32.
(21) Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. Hook: A Program for Finding Novel Molecular Architectures That Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site. *Proteins* **1994**, *19*, 199−221.
(22) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. Dock 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36*, 1132−1156.
(23) Marchand, J. R.; Caflisch, A. In Silico Fragment-Based Drug Design with Seed. *Eur. J. Med. Chem.* **2018**, *156*, 907−917.
(24) Todorov, N. P.; Dean, P. M. Evaluation of a Method for Controlling Molecular Scaffold Diversity in De Novo Ligand Design. *J. Comput-Aided Mol. Des* **1997**, *11*, 175−192.
(25) Zsoldos, Z. A Fragment-Based Docking Engine: Ehits. In *Library Design, Search Methods, and Applications of Fragment-Based Drug Design*; Bienstock, R., Ed.; ACS Symposium Series, Vol. *1076*; American Chemical Society, 2011; pp 91−130.
(26) Pierce, A. C.; Rao, G.; Bemis, G. W. Breed: Generating Novel Inhibitors through Hybridization of Known Ligands. Application to Cdk2, P38, and Hiv Protease. *J. Med. Chem.* **2004**, *47*, 2768−2775.
(27) Durrant, J. D.; Friedman, A. J.; McCammon, J. A. Crystaldock: A Novel Approach to Fragment-Based Drug Design. *J. Chem. Inf Model* **2011**, *51*, 2573−2580.
(28) 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.
(29) Tang, G. W.; Altman, R. B. Knowledge-Based Fragment Binding Prediction. *PLoS Comput. Biol.* **2014**, *10*, No. e1003589.
(30) Shang, J.; Dai, X.; Li, Y.; Pistolozzi, M.; Wang, L. Hybridsim-Vs: A Web Server for Large-Scale Ligand-Based Virtual Screening Using Hybrid Similarity Recognition Techniques. *Bioinformatics (Oxford, England)* **2017**, *33*, 3480−3481.
(31) Shan, J.; Pan, X.; Wang, X.; Xiao, X.; Ji, C. Fragrep: A Web Server for Structure-Based Drug Design by Fragment Replacement. *J. Chem. Inf Model* **2020**, *60*, 5900−5906.
(32) Douguet, D. E-Lea3d: A Computational-Aided Drug Design Web Server. *Nucleic Acids Res.* **2010**, *38*, W615−621.
(33) Hao, G. F.; Jiang, W.; Ye, Y. N.; Wu, F. X.; Zhu, X. L.; Guo, F. B.; Yang, G. F. Acfis: A Web Server for Fragment-Based Drug Discovery. *Nucleic Acids Res.* **2016**, *44*, W550−556.
(34) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The Ftmap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nat. Protoc* **2015**, *10*, 733−755.
(35) Rifaioglu, A. S.; Nalbat, E.; Atalay, V.; Martin, M. J.; Cetin-Atalay, R.; Doğan, T. Deepscreen: High Performance Drug−Target Interaction Prediction with Convolutional Neural Networks Using 2-

D Structural Compound Representations. *Chem. Sci.* **2020**, *11*, 2531−2557.
(36) Liu, Z.; Du, J.; Fang, J.; Yin, Y.; Xu, G.; Xie, L. Deepscreening: A Deep Learning-Based Screening Web Server for Accelerating Drug Discovery. *Database: J. Bio. Databases Curation* **2019**, *2019*, baz104.
(37) Green, H.; Koes, D. R.; Durrant, J. D. Deepfrag: A Deep Convolutional Neural Network for Fragment-Based Lead Optimization. *Chemical Science* **2021**, *12*, 8036−8047.
(38) Pevzner, Y.; Frugier, E.; Schalk, V.; Caflisch, A.; Woodcock, H. L. Fragment-Based Docking: Development of the Charmming Web User Interface as a Platform for Computer-Aided Drug Design. *J. Chem. Inf Model* **2014**, *54*, 2612−2620.
(39) Konc, J.; Miller, B. T.; Štular, T.; Lešnik, S.; Woodcock, H. L.; Brooks, B. R.; Janežič, D. Probis-Charmming: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *J. Chem. Inf Model* **2015**, *55*, 2308−2314.
(40) 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.
(41) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with Alphafold. *Nature* **2021**, *596*, 583−589.
(42) The UniProt Consortium. Uniprot: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480−D489.
(43) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. Am1-Bcc Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(44) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2021*; University of California: San Francisco,**2021**.
(45) Feinstein, W. P.; Brylinski, M. Calculating an Optimal Box Size for Ligand Docking and Virtual Screening against Experimental and Predicted Binding Pockets. *J. Cheminform* **2015**, *7*, 18.
(46) Trott, O.; Olson, A. J. Autodock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.
(47) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. Openmm 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(48) Boyer, R. D.; Bryan, R. L. Fast Estimation of Solvation Free Energies for Diverse Chemical Species. *J. Phys. Chem. B* **2012**, *116*, 3772−3779.
(49) Kulp, J. L., III; Kulp, J. L., Jr; Pompliano, D. L.; Guarnieri, F. Diverse Fragment Clustering and Water Exclusion Identify Protein Hot Spots. *J. Am. Chem. Soc.* **2011**, *133*, 10740−10743.
(50) 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.
(51) O'Reilly, M.; Cleasby, A.; Davies, T. G.; Hall, R. J.; Ludlow, R. F.; Murray, C. W.; Tisi, D.; Jhoti, H. Crystallographic Screening Using Ultra-Low-Molecular-Weight Ligands to Guide Drug Design. *Drug Discov Today* **2019**, *24*, 1081−1086.
(52) Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845−5859.
(53) Maybridge Fragment Libraries. *ThermoFisher Scientific*. www.maybridge.com (accessed 2023/05/01).
(54) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. Chembl: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−1107.
(55) *eMolecules*. www.emolecules.com (accessed 2023/04/21).
(56) Chemdiv. www.chemdiv.com (accessed 2023/04/21).
(57) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. Openmm 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(58) Hohman, M.; Gregory, K.; Chibale, K.; Smith, P. J.; Ekins, S.; Bunin, B. Novel Web-Based Tools Combining Chemistry Informatics, Biology and Social Networks for Drug Discovery. *Drug Discov Today* **2009**, *14*, 261−270.
(59) Lawson, A.; Swienty-Busch, J.; Geoui, T.; Evans, D. The Making of Reaxys—Towards Unobstructed Access to Relevant Chemistry Information. In The Future of the History of Chemical Information; McEwen, L. R., Buntrock, R. E., Eds.; ACS Symposiumm Series, Vol. *1164*; American Chemical Society, 2014; pp 127−148.
(60) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; et al. Pubchem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202−1213.
(61) Sunseri, J.; Koes, D. R. Pharmit: Interactive Exploration of Chemical Space. *Nucleic Acids Res.* **2016**, *44*, W442−W448.
(62) *Binding DB*. https://www.bindingdb.org/rwd/bind/index.jsp (accessed 2023/06/15).

https://doi.org/10.1021/acs.jcim.3c00209
J. Chem. Inf. Model. XXXX, XXX, XXX−XXX