

# bbSelect – An Open-Source Tool for Performing a 3D Pharmacophore-Driven Diverse Selection of R- groups.

*Francesco Rianjongdee<sup>†\*#</sup>, David Palmer<sup>§</sup>, Stephen D. Pickett<sup>†</sup>, Peter Pogány<sup>†</sup>, Nicholas C. O.  
Tomkinson<sup>§</sup>, Darren V. S. Green<sup>†</sup>*

<sup>†</sup>GSK Medicines Research Centre, Stevenage, Hertfordshire, SG1 2NY, UK

<sup>§</sup>Department for Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street,  
Glasgow, G1 1XL, UK

## KEYWORDS.

Medicinal Chemistry, Chemical Space, Pharmacophores, Diverse Selection, Hit-to-Lead,  
Chemical Descriptors, Partitioning.

## **Abstract**

The design of compounds during hit-to-lead often seeks to explore a vector from a core scaffold to form additional interactions with the target protein. A rational approach to this is to probe the region of a protein accessed by a vector with a systematic placement of pharmacophore features in 3D, particularly when bound structures are not available. Herein, we present bbSelect, an open-source tool built to map the placements of pharmacophore features in 3D Euclidean space from a library of R-groups, employing partitioning to drive a diverse and systematic selection to a user-defined size. An evaluation of bbSelect against established methods exemplified the superiority of bbSelect in its ability to perform diverse selections, achieving high levels of pharmacophore feature placement coverage with selection sizes of a fraction of the total set and without the introduction of excess complexity. bbSelect also reports visualisations and rationale to enable users to understand and interrogate results. This provides a tool for the drug discovery community to guide their hit-to-lead activities.

## **Introduction**

During the hit-to-lead phase of small-molecule drug discovery, it is important to build an understanding of the structure-activity relationships (SAR) of a hit to optimise its affinity against a desired target, whilst also keeping within desirable physicochemical, pharmacokinetic and selectivity boundaries. Regions of a hit molecule, referred to as vectors, are usually identified which are thought to have the potential to form interactions with the target protein, perhaps through an unexplored pocket of the protein. The groups attached to a vector are referred to as “R-groups” (Figure 1). An understanding of the preferable groups to place along a vector is key to progression of the chemotype towards further optimisation.<sup>1,2</sup>



**Figure 1.** Simplified representation of a core scaffold for a chemotype (green) residing in a protein binding pocket (white pocket within a blue protein) with a vector (arrow) extending to an unexplored region of a protein. An R-group is placed along this vector to access the region.

The approach used to discover preferable groups to place in a vector depends on the information available. If 3D structures of ligands bound to the protein can be obtained, the exploration can be guided using rational structure-based drug design (SBDD), a technique which is constantly incorporating new technologies.<sup>3-5</sup> Bound ligand structures, however, cannot always fully account for SAR due to protein conformational flexibility, the artificial, static environment used to produce crystal structures, and errors in binding predictions, providing reasons for limitations.<sup>6,7</sup> When protein structures are not available, or where a target protein is unknown, for example when optimising against a phenotypic response assay, it is common for the investigation of a vector to be performed by an exploration of chemical space: testing diverse analogues of the hit chemotype to seek additional interactions, without the introduction of unnecessary complexity.<sup>1,8</sup>

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

When exploring chemical space, it is important to consider the nature of ligand binding, which is partially driven by the formation of intermolecular interactions between the ligand and the protein in question. The chemical features that can perform intermolecular interactions can be generalised to groups such as hydrogen-bond donors, hydrogen-bond acceptors, positively and negatively charged groups, aromatic rings, and hydrophobic groups, referred to as pharmacophore features. Therefore, the 3D placements of these pharmacophore features are pertinent to the binding of a molecule to a protein.<sup>9</sup> As such, pharmacophore methods have been implemented for virtual screening campaigns, to match ligands to the pharmacophores of known active ligands or to complement the protein binding pocket, quantitative structure-activity relationship modelling, and diversity assessment.<sup>10-18</sup> Herein, we explore 3D pharmacophore methods for the diverse selection of R-groups.

A rational process to follow in the design of diverse analogues would be to systematically probe the 3D Euclidean space extending from a vector with different pharmacophore features, aiming to pick up interactions with the target protein. This process is non-trivial and requires a description of an R-group's placement of pharmacophore features in 3D space, and a method to systematically select a diverse set to cover said space.

Algorithms to perform diverse molecular selection, traditionally using dissimilarity, clustering, or partitioning approaches, have been reviewed and discussed previously in the literature.<sup>19-30</sup> Commonly used methods are the MaxMin dissimilarity-based selection, Sphere Exclusion clustering, and k-Means clustering.<sup>26,31</sup> However, these methods are not typically used for R-groups and tend to use topological descriptors which do not explicitly describe the placements of pharmacophore features in 3D Euclidean space, so are not well suited for the purpose outlined here. Furthermore, dissimilarity-based approaches have been found to select compounds from the

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups ‘edges’ of chemical space, resulting in sets with high complexity.<sup>21</sup> 3D pharmacophore approaches have been explored for building blocks, for example by using 3-point pharmacophore keys to describe relative pharmacophore feature placement. These pharmacophore keys are generally not aligned to an attachment point, and therefore do not describe locations in 3D Euclidean space. Furthermore, they utilise maximum-dissimilarity algorithms, again leading to complex selections.<sup>32,33</sup> It is important to note that fragment growth, which is a related use case to the one described here, has also had methods published which utilise pharmacophore or shape methods.<sup>34,35</sup> These methods, however, typically require crystal structures to guide fragment growth. As such, there are no published, open-source tools available to perform a systematic search of pharmacophore feature placements in 3D from a set of R-groups.

Our group previously published a methodology named Monomer Gridding and Partitioning (mGAP), which used gridding to quantify the pharmacophore feature placements of a building block (referred to as a monomer at the time) from an attachment point.<sup>36</sup> This was built to inform building block acquisition for combinatorial chemistry, utilising a greedy algorithm to prioritise building blocks which place pharmacophore features in regions of 3D space that were under-explored by an existing collection. The challenge with this method is that highly functionalised and flexible building blocks tend to be prioritised due to their ability to provide a high level of pharmacophore feature placement coverage, leading to molecular complexity in the selections. Furthermore, mGAP was not built to select a set which could sample an input space, instead to select compounds iteratively to maximise overall pharmacophore feature placements of a set. Nonetheless, the mGAP methodology provided a starting point for the development of a new methodology herein, due to its ability to describe the 3D placement of pharmacophore features from an attachment point.

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

Our aims for the development of a new tool, named “bbSelect”, were to describe and utilise the placement of pharmacophore features in 3D Euclidean space to provide a selection of R-groups which efficiently sampled the overall placements of an input set. We wanted bbSelect to make efficient selections, in terms of size, flexibility, and level of functionality of the building block. In effect, for each selection to be the best R-group which could place pharmacophore  $x$  in region  $y$  of 3D space. We also wanted to permit bbSelect to take a user-defined selection size, and report visualisations giving insight into the information generated by the tool and quantify how well a selection covered the input space.

## **Methodology**

bbSelect was built to process building block-like R-groups. These groups are intended to be small; therefore, some simplifications were utilised in comparison to 3D pharmacophore methods used for full-sized drug-like molecules.<sup>36</sup>

First, because small building-block sized molecules tend to have a small conformational space, extensive 3D geometry optimisation of R-groups was not incorporated. The use of fast methods to perform low energy conformer sampling to generate conformer ensembles, such as the OMEGA software from OpenEye, or the distance geometry methods implemented in RDKit (which is included in the code base), was deemed sufficient to cover likely conformers of an R-group and reduce the computational cost of running the algorithm.<sup>37</sup> Furthermore, the capture of the independent placements of single pharmacophore features relative to the attachment point was deemed sufficient to describe an R-group, which should be less functionalised than a full drug molecule. Finally, because the R-groups were assumed to be attached to a core scaffold at the same position, direct comparisons of the pharmacophore feature placements of R-groups could be

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups performed by alignment at the attachment point, rather than attempting to perform a full molecular alignment.

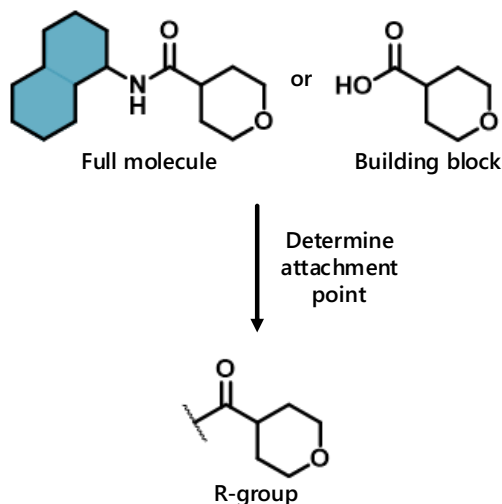
To illustrate and evaluate the bbSelect methodology, an example set of commercially available carboxylic acids from Enamine was used. This set was generously filtered to building block-like properties by imposing the following filters: heavy atom count < 14, hydrogen bond donors < 4, hydrogen bond acceptors < 5, rotatable bonds < 5 and molecular weight < 250 Daltons. Following this, a set of published substructure filters were applied to remove unwanted substructures (see Table S1 of reference 26).<sup>38,39</sup> The resulting size of this set was 19,253 carboxylic acids. Simple properties were also calculated for the molecules in the set using RDKit.

OpenEye OMEGA 3.1.1.2 was used to generate the conformers used in this report, utilising the flipper functionality to enumerate all stereoisomers of racemic molecules. The default MMFF94s\_NoEstat forcefield was used, which includes all MMFF94s terms except Coulombic interactions. A maximum ensemble size of 200 conformers was used. In some cases, the conformer generation method failed on more complex ring systems or where there was excessive strain, which is accounted for in the resulting total of 19,253 carboxylic acids. bbSelect is written in Python 3.6.13 using RDKit 2022.03.1. The open-source code can be found at <https://github.com/f-rianjongdee/bbSelect> or <https://zenodo.org/doi/10.5281/zenodo.10804697>, along with the Enamine carboxylic acid test set, and the notebooks and data that were used to produce the figures in this text.

## **R-group preparation**

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

**Scheme 1.** Illustration of how R-groups can be obtained from either full molecules or building blocks.



bbSelect requires the input of R-groups. These can be obtained from the results of an R-group decomposition for a set of whole molecules, or from building blocks which have been clipped to assign where they will be attached to a core scaffold (Scheme 1). The example provided in the code base uses RDKit to clip the acid building blocks by replacement of the hydroxyl group in the acid, as shown in Scheme 1. bbSelect uses a  $^{15}\text{CH}_3$  group to represent an attachment point internally, which enables compatibility of the clipped R-group with conformer generation tools whilst ensuring the attachment atom can be identified for alignment.

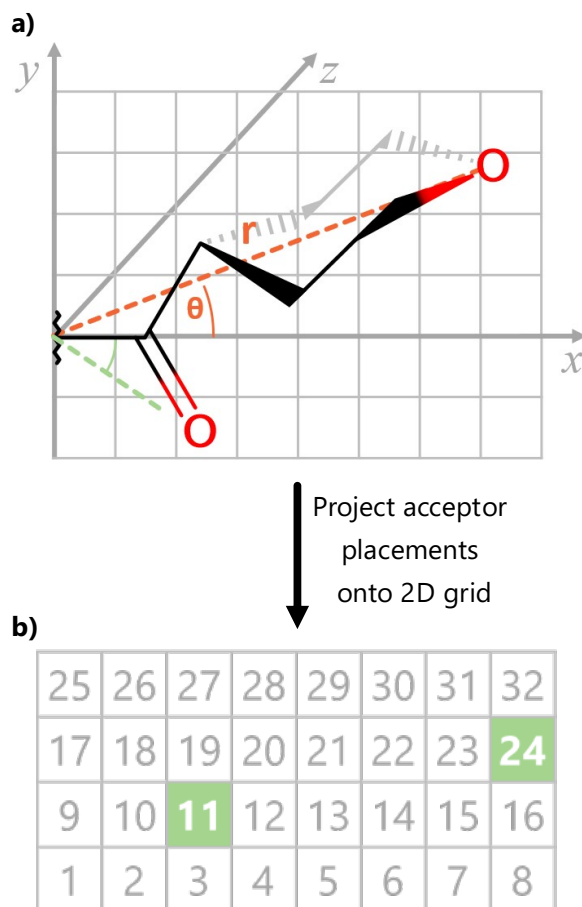
### **Capturing pharmacophore feature placements**

This section details how bbSelect describes an R-group's placement of pharmacophore features in 3D and stores this information in a fingerprint.

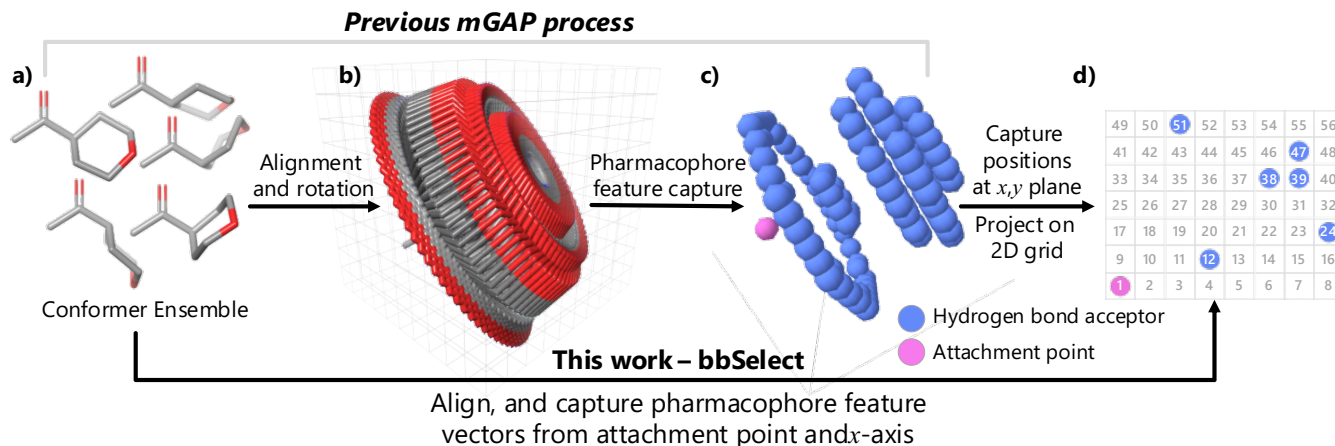
To begin this process, a conformer ensemble is required for each R-group. To prepare the examples shown here, we use OpenEye OMEGA to generate conformer ensembles, enumerating

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups all stereoisomers of racemic input molecules and otherwise using default parameters.<sup>37,40</sup> An example RDKit conformer generation script using the ETKDGv3 algorithm is also provided in the code base to facilitate open-source use.<sup>41,42</sup> In our hands, however, this took several orders of magnitude longer to generate the conformers against the same test set, both utilising 20 CPUs. Any conformer ensemble generation method may be used to generate input for bbSelect. Users may generate, or filter to, preferred conformers at this stage. The choice of conformer generation tool is not key to this work. Benchmarking has been conducted previously in the literature.<sup>43</sup>

Conformers are then aligned in 3D Euclidean space by rotation and translation, such that the attachment atom is placed at the Cartesian coordinate origin (0, 0, 0), and the adjacent atom is placed on the *x*-axis in the positive direction (Figure 2). In the mGAP method, the molecule was then freely rotated about the *x*-axis, and the placement of pharmacophore features captured within a numbered 3D grid.<sup>36</sup> The 3D grid was then stored in a fingerprint, where the corresponding bit for a numbered cell in the 3D grid was turned “on” if that cell was occupied by a pharmacophore feature (Figure 3).



**Figure 2.** a) Illustration of the bbSelect process of measuring the distance ( $r$ ) and angle ( $\theta$ ) from the  $x$ -axis between a pharmacophore feature and the attachment point in 3D. This information is then used to project the placements onto 2D coordinates, which is then used to assign a cell number on a 2D grid. The placements of two separate pharmacophore points are illustrated with their measurements in green and orange lines. b) Result of projection of 3D pharmacophore placements onto 2D coordinates, and illustration of cell numbering.



**Figure 3.** Schematic showing the stages of the mGAP fingerprint generation with a comparison to the bbSelect fingerprint generation. a) A representative conformer ensemble. b) Result of alignment of all conformers and rotation about the  $x$ -axis. c) Capture of pharmacophore feature placement in 3D. d) Result from capturing positions at 2D plane  $z = 0, y > 0$ .

The free rotation about the  $x$ -axis used in the mGAP approach results in a 3D representation (Figure 3c) with axial symmetry about the  $x$ -axis. Therefore, a radial slice captured at  $z = 0$  and  $y > 0$  captures all the information of the 3D representation within a 2D grid with complete fidelity (Figure 3d). This 2D representation, shown in Figure 3d, is what bbSelect utilises to describe pharmacophore feature placement.

bbSelect obtains the 2D representation shown in Figure 3d directly from the aligned conformer by measuring the distance ( $r$ ) and angle from the  $x$ -axis ( $\theta$ ) of each pharmacophore feature from the attachment point in 3D (Figure 2a). This vector is then projected as 2D Cartesian coordinates and transformed to a cell number using configurable parameters for cell size, in Angstroms squared, and number of cells in the grid (Figure 2b). By default, a cell size of  $1 \text{ \AA}$  is used, with a capture area of 40 cells along the  $x$ -axis, the origin in the middle between the 20<sup>th</sup> and 21<sup>st</sup> cell,

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups and 20 cells along the *y*-axis only in the positive direction. Users need to ensure that these parameters will capture all the pharmacophore features present in the R-group conformers supplied to bbSelect.

The occupied cells within the 2D grid are transformed into a fingerprint for data storage, where the bits in the fingerprint corresponding to occupied cells are set to “on”. Each pharmacophore feature receives its own fingerprint, which are concatenated to form the final fingerprint. This, in effect, creates a 2D fingerprint which is stored as a 1D bit string. Pharmacophore features are captured using SMARTS patterns, which need to account for tautomers. bbSelect implements the in-built pharmacophore feature factory in RDKit, and the RDKit pharmacophore feature definitions. Some modifications were applied to the RDKit pharmacophore feature definitions, removing fluorine as a hydrogen-bond acceptor, as these are atypical, and removing tertiary amines as hydrogen-bond donors, as this would depend on the protonation state. These modifications can be changed depending on the preferences of the user. The modified definitions can be found in the code base.<sup>44</sup>

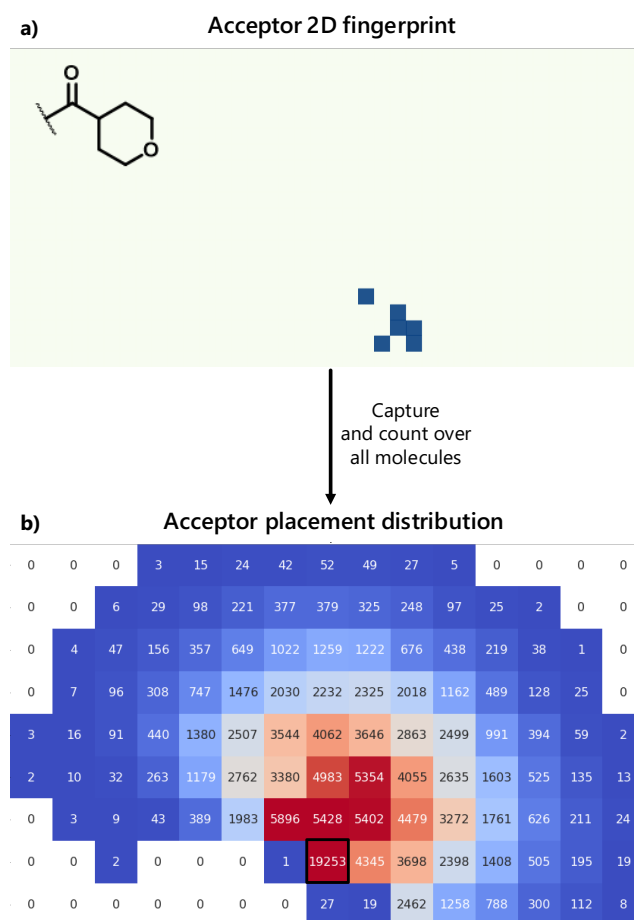
### **Analysis of the pharmacophore feature placement over a set of R-groups.**

Following pharmacophore feature placement capture, a holistic analysis for the overall distributions of pharmacophore feature placements for a set of R-groups can be performed.

To do this, the occupancies of each cell in the fingerprint are counted across the total set of molecules. This is illustrated in Figure 4, where the hydrogen bond acceptor (“Acceptor”) fingerprint is shown in for the R-group in Figure 4a. When the cell occupancies are counted across the full set of R-groups, the distribution of hydrogen bond acceptor placements across the set of 19,253 carboxylic acids can be visualised in 2D, giving what we term the “overall coverage map”

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

(Figure 4b, truncated for clarity. The untruncated overall coverage map is shown in Figure S1, and the overall coverage maps for all pharmacophore features is shown in Figure S2). The numbers within each cell represent the number of compounds which place a hydrogen bond acceptor in that respective region of 3D Euclidean space. The cell highlighted with an occupancy count of 19,253 is the cell where the carbonyl oxygen from the parent carboxylic acid is placed.



**Figure 4.** a) Visualisation of the 2D fingerprint generated for the R-group shown in Figure 2. The cells which are occupied by a hydrogen bond acceptor are shown in blue, against an off-white background showing all the cells that are present during capture b) Plot showing the distribution of hydrogen bond acceptor placements across the example set of Enamine carboxylic acids, where

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

the number of molecules which place a hydrogen bond acceptor in a cell is labelled. The highlighted cell indicates the 3D placement of the carbonyl oxygen present in all the molecules.

### **Partitioning the space**

With the overall coverage maps describing the overall pharmacophore feature placements in hand, the occupied space is then partitioned to guide a selection.

bbSelect requires a selection size to be provided, reflecting the number of compounds desired for synthesis. For example, a set of 96 compounds can be prepared simultaneously in a 96-well plate, and therefore 96 compounds would be the selection size for that experiment. The selection size is distributed evenly across the pharmacophore features captured to define the number of partitions to make. For example, a selection size of 96 over 6 pharmacophore features results in 16 selections for each pharmacophore feature, and 16 partitions of each pharmacophore feature space. If a selection size cannot be evenly distributed, bbSelect also permits specific selection sizes from each pharmacophore to be defined.

Two methods for partitioning are implemented in bbSelect. A classical partitioning method aims to divide the occupied space as evenly as possible using rectangular partitions with a threshold ratio between the lengths of the two edges of 0.4. Each partition is given a label which is inherited by the compound that is selected from it. The labels are 2D coordinates, which describe the arrangement of the partition with respect to others.

An example partitioning by 16 for the hydrogen bond acceptor placement overall coverage map, shown in Figure 4b, using the classical method is shown in Figure 5a. In this case, it was not possible to divide the given space into 16 evenly, so the partitions are slightly different in shape. This highlights one of the drawbacks of the classical method, where partition dimensions may need

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups to differ to fit the space, and there may be some, for example (0, 0), which contain only a single cell. Furthermore, whilst the partitions are of similar sizes, the number of compounds which place their pharmacophore features in each of the partitions varies widely, leading to an under-sampling of the distribution in that area (Figure S3). While that may be desirable for an unbiased selection, there may be cases where the user may want to perform a selection that reflects the distribution of the input set.

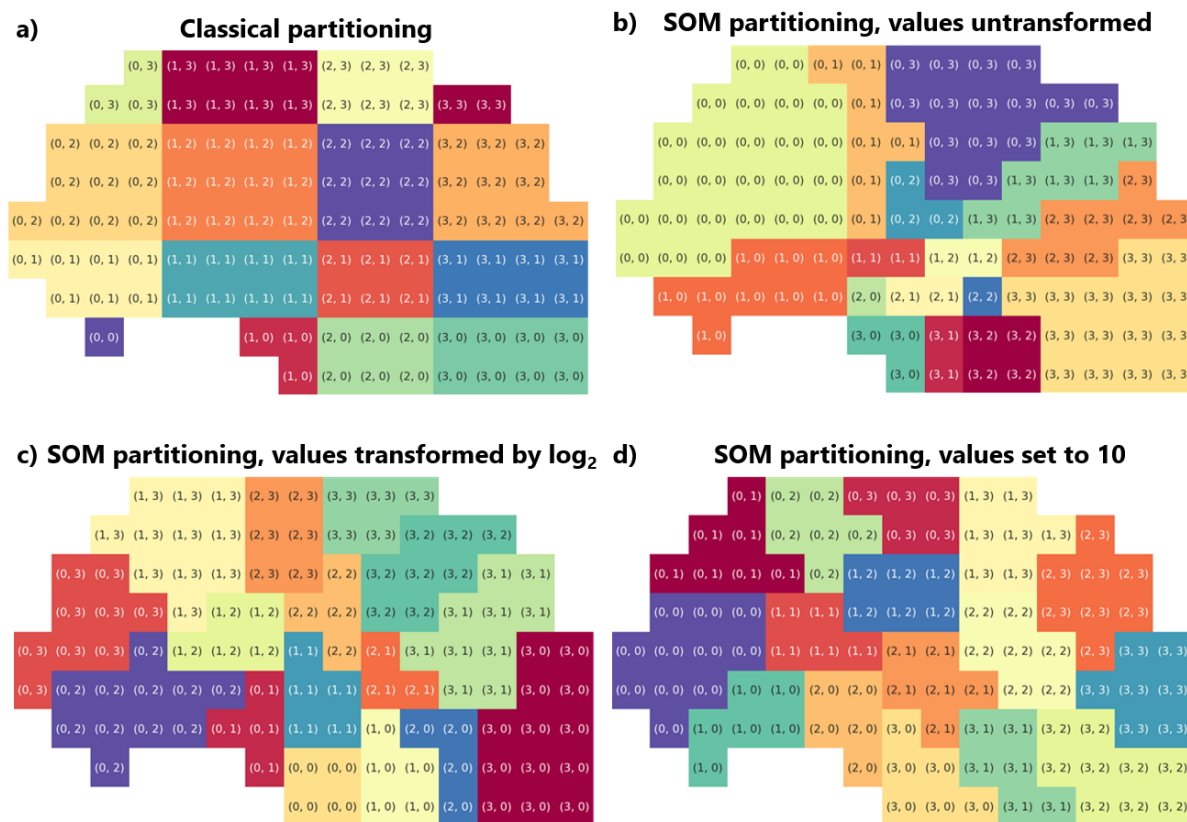
To address this, a second method was implemented to perform a heuristic partitioning, not requiring that a rectangular-shaped partition be formed. For this, self-organising maps (SOMs) were employed.<sup>45</sup> SOMs are a type of neural network where the weight vector for the neurons reflects the dimensionality of the input set. The neurons can therefore be seen as discrete data points in the same space as the input data. During training, the weights of the neurons are updated such that the neurons distribute within the input space to mimic the distribution of the input data. This technique is often used for dimensionality reduction, as SOM neurons maintain a 2D topology between them, however this was not the case for bbSelect, where the input data already existed in two-dimensional space. Following training, when input data is clustered to their closest neuron, a similar number of datapoints should be associated with each neuron. In this case, the number of neurons would equal the number of partitions required, and the partitions should be smaller in denser regions of the overall coverage map, and larger in the sparse regions. bbSelect implements the open-source “minisom” python package to construct SOMs.<sup>46</sup>

SOMs require discrete data to be passed as training data, so the overall coverage maps were discretised by creating a random scatter of data points in 2D Euclidean space within the Cartesian boundaries of each cell. The number of data points within each cell could either equal the exact occupancies of that cell or could be transformed to reduce the variance in the distribution. Due to

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

the several orders of magnitude difference between the occupancies of the cells close to the attachment point and on the edges of the covered space, training the SOM on untransformed data led to most of the neurons localising near the attachment point (Figure 5b, Figure S4). Therefore, only transformed data is used by default in bbSelect. Two transformations were explored to discretise the occupancy data before training the SOM.

To reduce the variance of the data whilst preserving the topology of the distribution, we looked to transform the data logarithmically, arbitrarily to base 2. This transformation resulted in a partitioning which correctly created smaller partitions in the more occupied regions of space, whilst creating larger partitions in the more sparsely occupied regions. This is shown in Figure 5c. A second transformation was also employed, which used a binary transformation, setting the value of a cell to 10 if it was occupied, and 0 if it was unoccupied. This creates data which does not preserve the topology of the distribution. The value of 10 was chosen to boost the number of data points being used to train the SOM. This resulted in more evenly sized partitions, analogous to that performed by the classical method. This is shown in Figure 5d. Further detail is provided in Figure S4, Figure S5, and Figure S6, showing the locations of the neurons and histograms describing the number of compounds placing their pharmacophore features and the area covered by each partition.



**Figure 5.** Illustration of the different bbSelect partitioning methods used on overall coverage map shown in Figure 4b to create 16 partitions. Each cell is labelled with the parent partition. a) Results from classical partitioning. b) Results from SOM partitioning without value transformation. c) Results from SOM partitioning where values are transformed logarithmically to base 2. d) Results from SOM partitioning where each value is set to 10.

### Selection of compounds

Following partitioning, a compound is then selected from each partition. To begin this process, compounds are matched to partitions by creating dummy fingerprints where the bits corresponding to the cells associated to a partition are set to “on”. An “any” bit operation is then performed against all the molecular fingerprints to determine which compounds place the desired pharmacophore feature in the region covered by the partition in question. This associates the

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups partitions with the compounds which place pharmacophore features in that region of space. To note, compounds may appear in multiple partitions, given that they may have multiple pharmacophore features or multiple conformers.

To determine which compound to select from each partition, the algorithm requires a user-defined prioritisation. This can be provided using calculated molecular properties of the building blocks, such as the molecular weight, or using properties of the full molecules calculated before R-group decomposition, for example a QED score.<sup>47</sup>

For the example provided in this text, a composite score for building block complexity was created, referred to as the “bbComplexity score”, analogous to a multi-parameter optimisation score.<sup>48,49</sup> This score incorporated simple calculated parameters for the carboxylic acids, scaling each parameter between 1 and 0 based on a scaling function. Two scaling functions were implemented: a linear scaling function, and a gaussian scaling function. The linear scaling function returned a scaled value between 1 and 0 between a high and a low threshold, whereas the gaussian scaling function returned a value of 1 at the apex, and 0 past a defined deviation. Anything beyond the boundaries is attributed the boundary score. A weighted average of these scaled values generated the bbComplexity score, where a higher score indicates a more complex building block.

The thresholds and weightings for the bbComplexity score are shown in Table 1. The values chosen for the boundaries and the weights aimed to prioritise smaller, more rigid, and less functionalised building blocks based on the distribution of the input set. Therefore, bbSelect will prioritise the R-group with the lowest score to select from each partition, provided the R-group has not been selected previously.

**Table 1.** Details of the parameters, scaling functions, boundaries, and weights used in the bbComplexity score.

Parameter	Scaling function	Low boundary	High boundary	Mode	Weight
Rotatable bond count	Linear	0	19	Minimise	2
Molecular Weight	Linear	60	300	Minimise	1
Heavy atom count	Linear	5	20	Minimise	1
Heteroatom count	Linear	2	4	Minimise	1
Hydrogen bond acceptor count	Linear	2	4	Minimise	1
Hydrogen bond donor count	Linear	1	4	Minimise	1
Racemic centre count	Linear	0	2	Minimise	1
		<b>Centre</b>	<b>Deviation</b>		
Ring count	Gaussian	1	2	Maximise	0.5

An equation for how the bbComplexity score was calculated is shown below, where  $x_i$  represents the value for parameter  $i$  in the set,  $f(x_i)$  is the scaling function applied to  $x_i$  to scale the value between 0 and 1,  $w_i$  is the weight assigned to each scaled parameter, and  $n$  is the number of parameters. The code used to calculate this score is included in the code base.

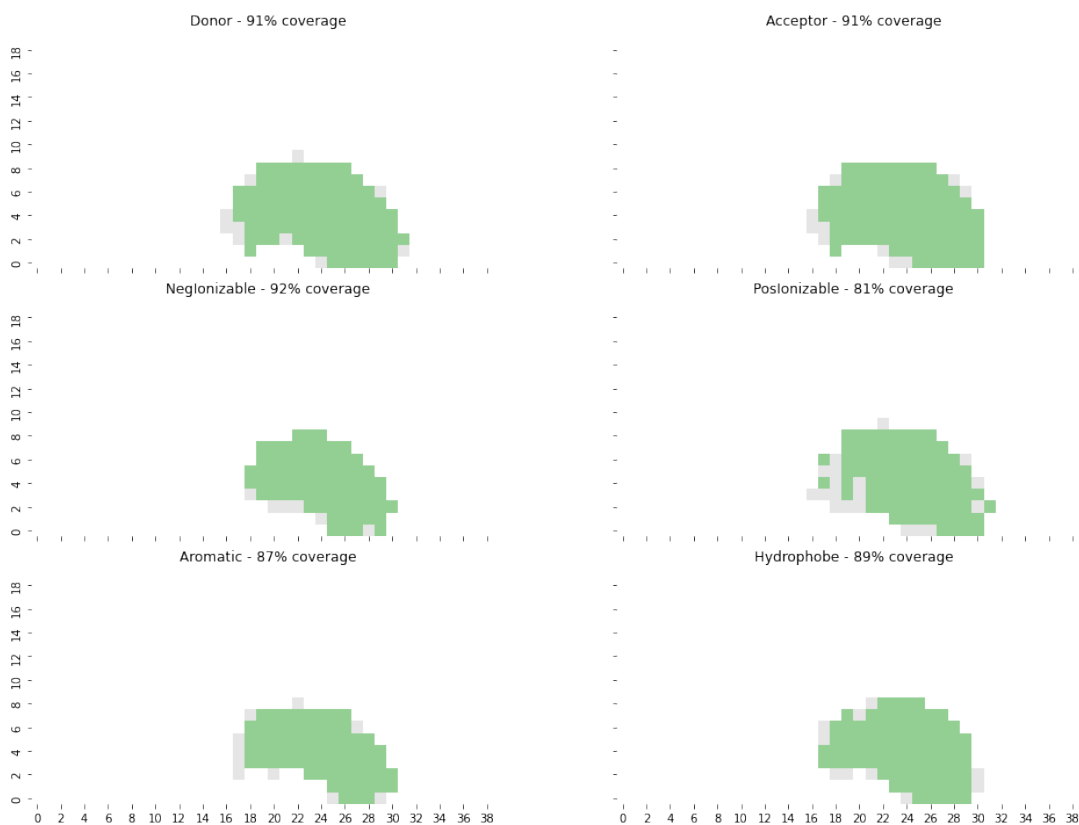
$$bbComplexityScore = \frac{\sum_{i=1}^n w_i \cdot f(x_i)}{\sum_{i=1}^n w_i}$$

To increase the topological diversity of the selection, a Tanimoto threshold of similarity, calculated on Morgan fingerprints with a radius of 2, is implemented in bbSelect. A new selection must be below this threshold when compared to previously selected compounds. Furthermore, an optional filter was implemented which requires that each selected molecule introduces a new bit into the overall aggregated bits for the selected set, thereby increasing the overall pharmacophore coverage of the selected set. This is referred to as the “coverage” filter.

Following selection, the aggregated pharmacophore feature placement fingerprints for selected molecules could be compared to the overall coverage map to calculate a percentage coverage of

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

the set. An example of this for a selection of 96 acids from the Enamine acids test set, using bbSelect with classic partitioning, is shown in Figure 6. The pharmacophore feature space covered by the entire set is shaded in grey, and the coverage from the selection of 96 in green. These areas are used to calculate a percentage coverage by the selected set.



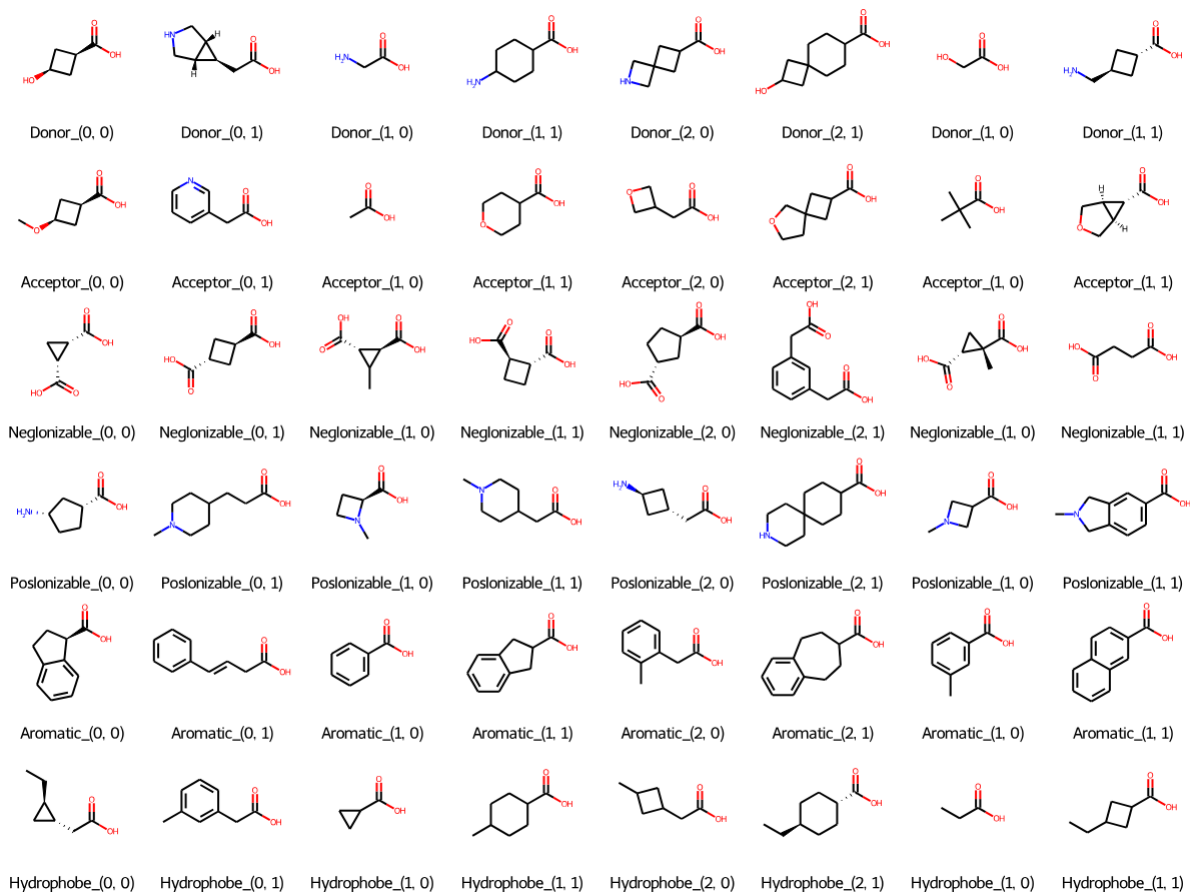
**Figure 6.** Visualisations provided by bbSelect showing the overall coverage of the pharmacophore feature space by a selected set of 96 compounds. The space covered by the total set of compounds is shaded in grey, and the space covered by the selected compounds is shaded in green. The percentage coverage is shown with the pharmacophore features in the titles.

An alternative selection methodology is also implemented into bbSelect, analogous to the greedy, space-filling approach used in the mGAP methodology. This method, termed “full

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups coverage”, iterates through the R-groups, sorted by the score, and selects any compound that adds to the overall coverage until full coverage is achieved. This algorithm does not utilise a selection size, instead returns the number and identity of compounds required to complete full coverage of the pharmacophore feature placements of the input set through the greedy process.

## **Results and discussion**

An example selection of 48 compounds given by the bbSelect methodology using classical partitioning, with a Tanimoto similarity threshold of 0.9, and without the coverage filter, is shown in Figure 7. Annotations for each molecule indicate the pharmacophore feature and the partition which the selected R-group represents. These were implemented to provide the user with the context behind a compound selection and allows users to reference a partitioning map, as shown in Figure 5, to understand where a compound is placing its pharmacophore feature. The selections demonstrate the ability of bbSelect to select simple representations of molecules able to place pharmacophore features in different regions of 3D space.



**Figure 7.** Selection of 48 compounds using classic partitioning in bbSelect. The annotations describe which pharmacophore feature a compound was selected for, and for which partition.

An example selection of 48 compounds was also performed using the MaxMin algorithm implementation in RDKit, and is shown in Figure 8.<sup>31</sup> MaxMin is a common dissimilarity-based selection algorithm where compounds are selected on the merit of dissimilarity from the previously selected set of compounds. The selection given in this example illustrates the higher complexity of molecules compared to the bbSelect selections, and, of course, do not include annotations or rationales associated with the selections. Example selections were also performed using the SOM partitioning, random selection, sphere exclusion, and k-Means clustering for comparative purposes, which are shown in Figure S7, Figure S8, Figure S9, and Figure S10, respectively. The



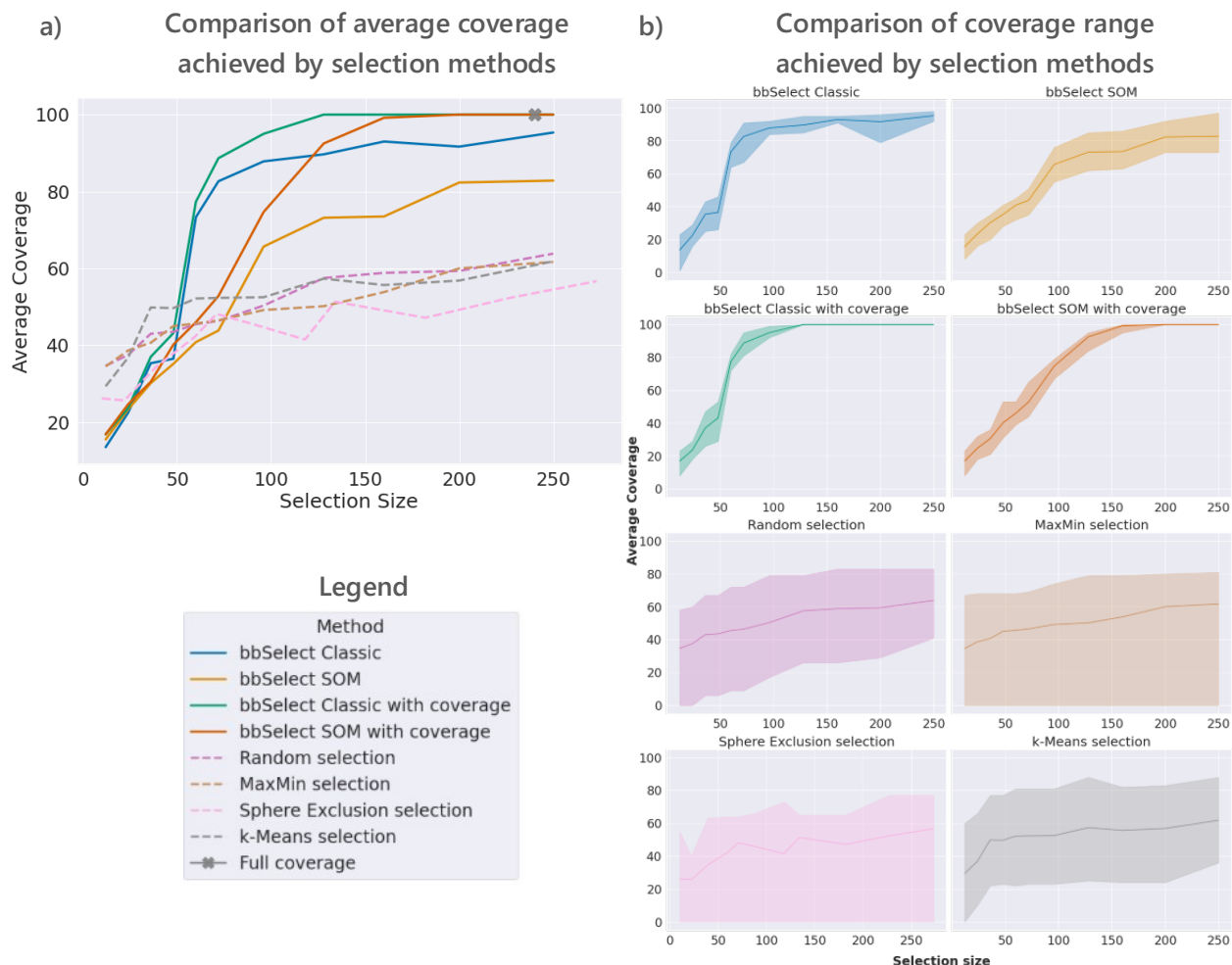
bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups in a particular region of chemical space, and this activity would therefore limit the chemical space from which to select. This is accounted for by the prioritisation step in bbSelect.

To perform a holistic analysis of the ability of bbSelect to cover the pharmacophore feature placement space, selections were performed across different selection sizes, and the coverage, and average bbComplexity score of the selections, were captured. Samples of coverage and average bbComplexity score were taken at selection sizes of 12, 24, 36, 48, 60, 72, 96, 128, 160, 200, and 250. The bbSelect methods, both with and without the coverage filter, were analysed. As a side note, transforming the data used to train the SOM logarithmically or transforming the cell values to 10 did not affect the coverage performance significantly, therefore only the logarithmic transformation is shown here as it offered the greatest distinction from the classic partitioning method (Figure S11). Alongside these, a random selection, MaxMin selection, Sphere exclusion selection, and k-means selection were also performed. For the sphere-exclusion selection, Tanimoto thresholds were set to give similar numbers of results as the selection sizes, but these weren't identical. The number of compounds returned using the "full coverage" method using the implemented method in bbSelect was also calculated, requiring 240 compounds to fill the space.

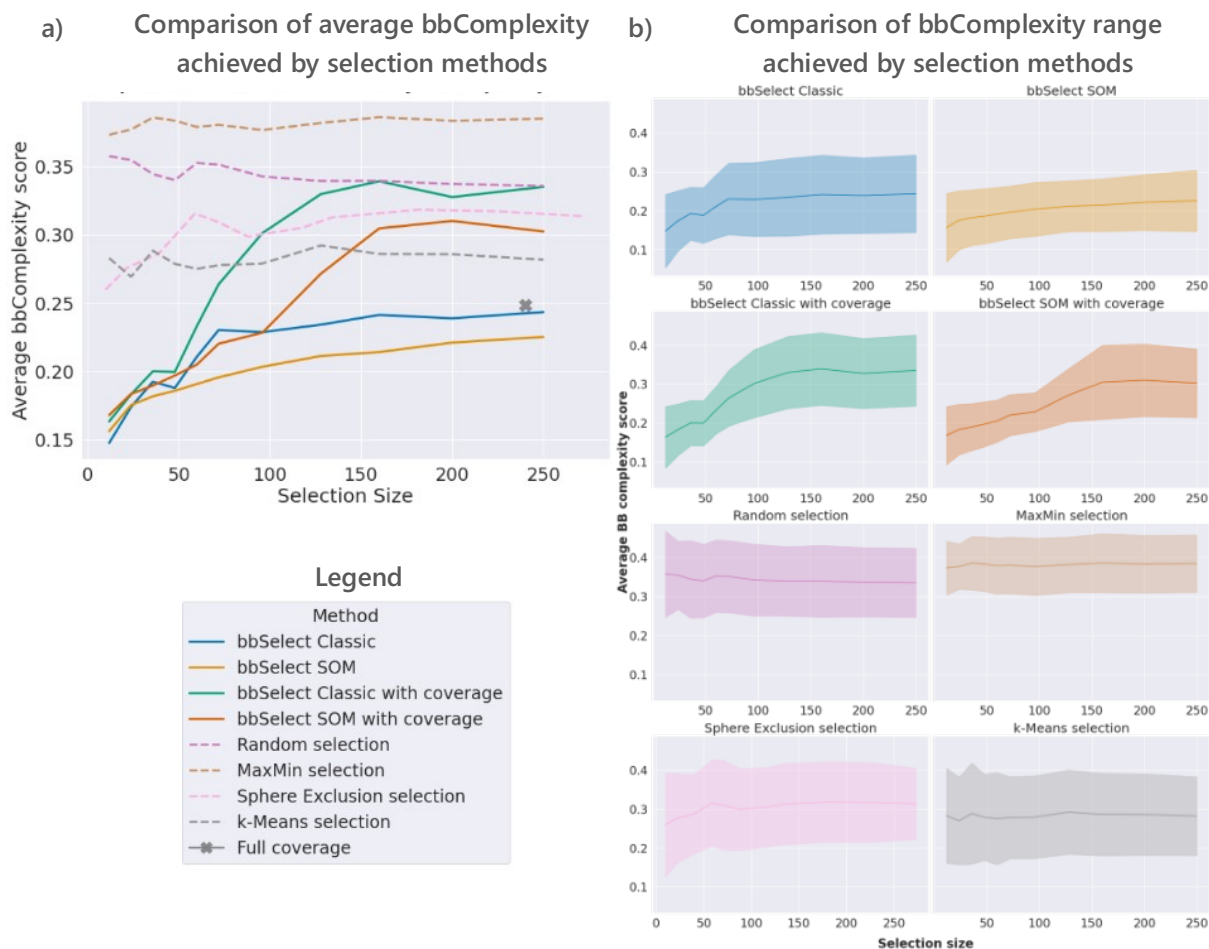
Figure 9a shows an overlay of the average percentage space coverage across the pharmacophore features, for each of the selection methods, with the number returned by the "full coverage" method indicated with an "x" symbol. Where the "coverage" filter has been applied, it has been noted in the bbSelect method name in the legend. Figure 9b shows the distribution of coverage, where the shaded areas show the maximum and minimum coverage across the pharmacophore features, over each selection size sample. Figure 10a shows the same analyses using the average bbComplexity scores. The shaded areas in Figure 10b show the maximum and minimum bbComplexity scores across the selected compounds.

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

A caveat to these analyses is that bbSelect was built to cover the pharmacophore feature placement space, and to prioritise selections based on the bbComplexity score, so this was not a perfect comparison; the other selection algorithms were not prioritising against these evaluation parameters. However, as there are no objective means of comparing diverse subsets with different selection methodologies, these analyses were deemed sufficient to compare methodologies for the task at hand.



**Figure 9.** Analyses of the pharmacophore feature placement coverage achieved by different selection methods. a) Overlay of average coverage across different methods and selection sizes. The grey cross indicates the number required for full coverage using the mGAP analogous selection algorithm. b) Average coverage shown for individual selection methods. The shaded areas represent the maximum and minimum coverage obtained across the pharmacophore features.



**Figure 10.** Analyses of the bbComplexity score achieved by different selection methods. a) Overlay of average bbComplexity score across different methods and selection sizes. The grey cross indicates the average score required for full coverage using the mGAP analogous selection algorithm. b) Average score shown for individual selection methods. The shaded areas represent the maximum and minimum scores obtained across the pharmacophore features.

In general, the bbSelect methods were found to have a superior ability to cover the pharmacophore feature placement space compared to the other methods. At selection sizes below 72, other methods were able to cover more space, however at the cost of complexity, which can be seen by comparing Figure 9a with Figure 10a. This highlights that complexity and coverage

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups often come hand in hand with diverse selection methodologies, as complex molecules will be more flexible and functionalised, placing pharmacophore features in more regions of 3D space. This is further exemplified when bbSelect was run using the “coverage” filter. Here, 100% coverage was achieved with just 128 compounds using classical partitioning, yet the average complexities of the molecules became akin to the other selection methodologies. Still, this represents a full coverage of the pharmacophore feature placements of the set of 19,253 carboxylic acids with a 0.66% sample. Above selection sizes of 72, all the selections performed by bbSelect methods achieved better coverage of the space with lower average scores for bbComplexity. Furthermore, the use of partitioning enabled smaller selection sizes to be used to sample the input space compared to the full coverage method, which required 240 compounds to reach full coverage. This demonstrates how bbSelect was able to select compounds efficiently, keeping the complexity of its selections low whilst covering the space. The bbSelect methods were also able to consistently cover all the pharmacophore features evenly, whereas the MaxMin and sphere exclusion methods consistently missed entire pharmacophore features from their selections, as shown by the shaded areas in Figure 9b.

Notably, the use of classical partitioning was able to consistently achieve a higher percentage coverage over the SOM method of partitioning. This was unsurprising, as the SOM partitioning biased the partitions to perform a selection which was more representative of the input set, at the cost of exploring the less-occupied regions of the space.

To exemplify the orthogonality of bbSelect to the other selection approaches, the overlap between the selections at a selection size of 96 are shown in Table 2. There were very few overlaps between the bbSelect methods and the comparator methods, demonstrating a high level of orthogonality in the approaches.

**Table 2.** The number of overlapping compounds between each selection method at a selection size of 96. To simplify the table, self-overlaps and overlaps covered in lower rows are not shown.

Set overlap	SOM	Classic	SOM with coverage	Classic with coverage	Full coverage	Sphere exclusion	MaxMin
Classic	46	-	-	-	-	-	-
SOM with coverage	33	28	-	-	-	-	-
Classic with Coverage	15	32	39	-	-	-	-
Full coverage	73	75	70	75	-	-	-
Sphere exclusion	4	3	2	3	6	-	-
MaxMin	0	0	0	0	0	2	-
k-Means	10	10	4	2	11	0	0

In summary, bbSelect was shown to make selections that were diverse in terms of pharmacophore feature placement, yet simple in terms of our bbComplexity score, compared to existing diverse selection methodologies. High levels of pharmacophore feature placement coverage could be achieved using bbSelect at selection sizes less than 1% of the size of the exemplary input set. Orthogonality was also demonstrated to comparator methods. As a selection methodology, bbSelect has shown its value for the selection of diverse compounds and provides another tool for drug discovery scientists performing hit-to-lead chemistry to use.

## Conclusion

The ability to perform a diverse selection of compounds to cover chemical space is crucial across different stages of medicinal chemistry, and is a topic often discussed in the literature. By concentrating on the purpose of exploring an unknown vector in the hit-to-lead stage of drug discovery, several simplifications could be introduced to typical pharmacophore-based approaches to suit small, building block-sized R-groups. To this effect, bbSelect was developed as a tool to

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups guide medicinal chemists during the design of analogues which probe pharmacophore feature placements in 3D with the aim to pick up interactions with a target protein. bbSelect evolved from concepts introduced by mGAP, drawing inspiration to fit the purpose of hit-to-lead. By creating a novel 2D fingerprint that describes the placement of pharmacophore features in 3D, bbSelect informs the user on the space that is available to them and partitions the space to perform a selection to a user-defined size. Through comparisons to established methods, bbSelect was shown to select compounds with diverse placements of pharmacophore features with high efficiency against our defined prioritisation criteria. The open-source code for bbSelect has been released to provide the medicinal chemistry community with an additional method for their toolbox to accelerate hit-to-lead in their projects.

### **Data Availability**

The source code, data, and notebooks used to generate figures used in this text have been deposited on GitHub at <https://github.com/f-rianjongdee/bbSelect>. A permanent record of the version of the repository used for this publication has been stored in Zenodo at <https://zenodo.org/doi/10.5281/zenodo.10804697>.

### ASSOCIATED CONTENT

#### **Supporting Information**

The following files are available free of charge.

Supplementary figures for bbSelect.pdf (PDF) Contains additional figures referenced in the text.

### AUTHOR INFORMATION

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

## **Corresponding Author**

Francesco Rianjongdee\*: E-mail: francesco@charmtx.com

## **Present Addresses**

†#Francesco Rianjongdee: Charm Therapeutics, London, UK

## **Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## **Funding Sources**

GlaxoSmithKline, EPSRC Prosperity Partnership EP/S035990/1.

## **Notes**

The authors declare the following competing financial interest(s): Authors were full-time employees at GlaxoSmithKline or the University of Strathclyde when this study was performed.

## **ACKNOWLEDGMENT**

We are grateful to the University of Strathclyde, GSK and EPSRC for funding via Prosperity Partnership EP/S035990/1. We would like to thank Enamine for giving us their permission to use and publish their commercially available carboxylic acids as a test set for this work and in the GitHub.

## **ABBREVIATIONS**

SAR, Structure-activity relationships; SBDD, Structure-based drug design; mGAP, monomer Gridding and Partitioning; CPU, Central processing unit.

## **REFERENCES**

- (1) Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J Chem Inf Comput Sci* **2001**, *41* (3), 856–864. <https://doi.org/10.1021/ci000403i>.
- (2) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *Br J Pharmacol* **2011**, *162* (6), 1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>.
- (3) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Med Res Rev* **1996**, *16* (1), 3–50. [https://doi.org/10.1002/\(SICI\)1098-1128\(199601\)16:1<3::AID-MED1>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6).
- (4) Batool, M.; Ahmad, B.; Choi, S. A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019**, *20* (11), 2783–2801. <https://doi.org/10.3390/ijms20112783>.
- (5) Sadybekov, A. V.; Katritch, V. Computational Approaches Streamlining Drug Discovery. *Nature* **2023**, *616* (7958), 673–685. <https://doi.org/10.1038/s41586-023-05905-z>.
- (6) Verlinde, C. L.; Hol, W. G. Structure-Based Drug Design: Progress, Results and Challenges. *Structure* **1994**, *2* (7), 577–587. [https://doi.org/10.1016/s0969-2126\(00\)00060-5](https://doi.org/10.1016/s0969-2126(00)00060-5).
- (7) Fraser, J. S.; Murcko, M. A. Structure Is Beauty, but Not Always Truth. *Cell* **2024**, *187* (3), 517–520. <https://doi.org/10.1016/j.cell.2024.01.003>.
- (8) Keseru, G. M.; Makara, G. M. Hit Discovery and Hit-to-Lead Approaches. *Drug Discov Today* **2006**, *11* (15–16), 741–748. <https://doi.org/10.1016/j.drudis.2006.06.016>.
- (9) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J Med Chem* **2010**, *53* (2), 539–558. <https://doi.org/10.1021/jm900817u>.
- (10) Spitzer, G. M.; Heiss, M.; Mangold, M.; Markt, P.; Kirchmair, J.; Wolber, G.; Liedl, K. R. One Concept, Three Implementations of 3D Pharmacophore-Based Virtual Screening: Distinct Coverage of Chemical Search Space. *J Chem Inf Model* **2010**, *50* (7), 1241–1247. <https://doi.org/10.1021/ci100136b>.
- (11) Meyenburg, C.; Dolfus, U.; Briem, H.; Rarey, M. Galileo: Three-Dimensional Searching in Large Combinatorial Fragment Spaces on the Example of Pharmacophores. *J Comput Aided Mol Des* **2023**, *37* (1), 1–16. <https://doi.org/10.1007/s10822-022-00485-y>.
- (12) Schaller, D.; Šribar, D.; Noonan, T.; Deng, L.; Nguyen, T. N.; Pach, S.; Machalz, D.; Bermudez, M.; Wolber, G. Next Generation 3D Pharmacophore Modeling. *Wiley Interdiscip Rev Comput Mol Sci* **2020**, *10* (4), e1468. <https://doi.org/10.1002/wcms.1468>.
- (13) Fan, F.; Warshaviak, D. T.; Hamadeh, H. K.; Dunn, R. T. The Integration of Pharmacophore-Based 3D QSAR Modeling and Virtual Screening in Safety Profiling: A Case Study to Identify Antagonistic Activities against Adenosine Receptor, A2A, Using 1,897 Known Drugs. *PLoS One* **2019**, *14* (1), e0204378. <https://doi.org/10.1371/journal.pone.0204378>.

- (14) Pickett, S. D.; Mason, J. S.; McLay, L. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ). *J Chem Inf Comput Sci* **1996**, *36* (6), 1214–1223. <https://doi.org/10.1021/ci960039g>.
- (15) Cramer, R. D. Topomer CoMFA: A Design Methodology for Rapid Lead Optimization. *J Med Chem* **2003**, *46* (3), 374–388. <https://doi.org/10.1021/jm020194o>.
- (16) Dixon, S. L.; Smondjrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J Comput Aided Mol Des* **2006**, *20* (10–11), 647–671. <https://doi.org/10.1007/s10822-006-9087-6/>.
- (17) Kohlbacher, S. M.; Langer, T.; Seidel, T. QPHAR: Quantitative Pharmacophore Activity Relationship: Method and Validation. *J Cheminform* **2021**, *13* (1), 1–14. <https://doi.org/10.1186/S13321-021-00537-9>.
- (18) Pickett, S. D. The Biophore Concept. In *Methods and Principles in Medicinal Chemistry*; John Wiley & Sons, Inc., 2003; pp 73–105.
- (19) Taylor, R. Simulation Analysis of Experimental Design Strategies for Screening Random Compounds as Potential New Drugs and Agrochemicals. *J Chem Inf Comput Sci* **1995**, *35* (1), 59–67. <https://doi.org/10.1021/ci00023a009>.
- (20) Lewis, R. A.; Mason, J. S.; McLay, L. M. Similarity Measures for Rational Set Selection and Analysis of Combinatorial Libraries: The Diverse Property-Derived (DPD) Approach. *J Chem Inf Comput Sci* **1997**, *37* (3), 599–614. <https://doi.org/10.1021/ci60471y>.
- (21) Snarey, M.; Terrett, N. K.; Willett, P.; Wilton, D. J. Comparison of Algorithms for Dissimilarity-Based Compound Selection. *J Mol Graph Model* **1997**, *15* (6), 372–385. [https://doi.org/10.1016/s1093-3263\(98\)00008-4](https://doi.org/10.1016/s1093-3263(98)00008-4).
- (22) Pearlman, R. S.; Smith, K. M. Novel Software Tools for Chemical Diversity. *Perspectives in Drug Discovery and Design* **1998**, *9*, 339–353. [https://doi.org/10.1007/0-306-46857-3\\_18](https://doi.org/10.1007/0-306-46857-3_18).
- (23) Bayley, M. J.; Willett, P. Binning Schemes for Partition-Based Compound Selection. *J Mol Graph Model* **1999**, *17* (1), 10–18. [https://doi.org/10.1016/s1093-3263\(99\)00016-9](https://doi.org/10.1016/s1093-3263(99)00016-9).
- (24) Bayada, D. M.; Hamersma, H.; Van Geerestein, V. J. Molecular Diversity and Representativity in Chemical Databases. *J Chem Inf Comput Sci* **1999**, *39* (1), 1–10. <https://doi.org/10.1021/ci980109e>.
- (25) Schnur, D. Design and Diversity Analysis of Large Combinatorial Libraries Using Cell-Based Methods. *J Chem Inf Comput Sci* **1999**, *39* (1), 36–45. <https://doi.org/10.1021/ci980138p>.
- (26) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J Chem Inf Comput Sci* **1999**, *39* (4), 747–750. <https://doi.org/10.1021/ci9803381>

- (27) Lajiness, M.; Watson, I. Dissimilarity-Based Approaches to Compound Acquisition. *Curr Opin Chem Biol* **2008**, *12* (3), 366–371. <https://doi.org/10.1016/j.cbpa.2008.03.010>.
- (28) Khanna, V.; Ranganathan, S. Molecular Similarity and Diversity Approaches in Chemoinformatics. *Drug Dev Res* **2011**, *72* (1), 74–84. <https://doi.org/10.1002/ddr.20404>.
- (29) Brown, N. *In Silico Medicinal Chemistry*; Theoretical and Computational Chemistry Series; The Royal Society of Chemistry, 2016. <https://doi.org/10.1039/9781782622604>.
- (30) Nakamura, T.; Sakaue, S.; Fujii, K.; Harabuchi, Y.; Maeda, S.; Iwata, S. Selecting Molecules with Diverse Structures and Properties by Maximizing Submodular Functions of Descriptors Learned with Graph Neural Networks. *Scientific Reports* **2022**, *12*:1 **2022**, *12* (1), 1–18. <https://doi.org/10.1038/s41598-022-04967-9>.
- (31) Ashton, M.; Barnard, J.; Casset, F.; Charlton, M.; Downs, G.; Gorse, D.; Holliday, J.; Lahana, R.; Willett, P. Identification of Diverse Database Subsets Using Property-Based and Fragment-Based Molecular Descriptions. *Quantitative Structure-Activity Relationships* **2002**, *21* (6), 598–604. <https://doi.org/10.1002/qsar.200290002>.
- (32) Pickett, S. D.; Mason, J. S.; McLay, L. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries (PDQ). *J Chem Inf Comput Sci* **1996**, *36* (6), 1214–1223. <https://doi.org/10.1021/ci960039g>.
- (33) Pickett, S. D.; Luttmann, C.; Guerin, V.; Laoui, A.; James, E. DIVSEL and COMPLIB - Strategies for the Design and Comparison of Combinatorial Libraries Using Pharmacophoric Descriptors. *J Chem Inf Comput Sci* **1998**, *38* (2), 144–150. <https://doi.org/10.1021/ci970060x>.
- (34) 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* (12), 6269–6281. <https://doi.org/10.1021/acs.jcim.0C00920>.
- (35) Hadfield, T. E.; Imrie, F.; Merritt, A.; Birchall, K.; Deane, C. M. Incorporating Target-Specific Pharmacophoric Information into Deep Generative Models for Fragment Elaboration. *J Chem Inf Model* **2021**. <https://doi.org/10.1021/acs.jcim.1c01311>.
- (36) Leach, A. R.; Green, D. V. S.; Hann, M. M.; Judd, D. B.; Good, A. C. Where Are the GaPs? A Rational Approach to Monomer Acquisition and Selection. *J Chem Inf Comput Sci* **2000**, *40* (5), 1262–1269. <https://doi.org/10.1021/ci0003855>.
- (37) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J Chem Inf Model* **2010**, *50* (4), 572–584. <https://doi.org/10.1021/ci100031x>.
- (38) Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I. H.; Frearson, J.; Wyatt, P. G. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem* **2008**, *3* (3), 435–444. <https://doi.org/10.1002/cmdc.200700139>.

- (39) Sydow, D.; Morger, A.; Driller, M.; Volkamer, A. TeachopenCadd: A Teaching Platform for Computer-Aided Drug Design Using Open Source Packages and Data. *J Cheminform* **2019**, *11* (1), 1–7. <https://doi.org/10.1186/s13321-019-0351-x>.
- (40) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J Chem Inf Model* **2012**, *52* (11), 2919–2936. <https://doi.org/10.1021/ci300314k>.
- (41) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J Chem Inf Model* **2015**, *55* (12), 2562–2574. <https://doi.org/10.1021/acs.jcim.5b00654>.
- (42) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J Chem Inf Model* **2020**, *60* (4), 2044–2058. <https://doi.org/10.1021/acs.jcim.0c00025>.
- (43) Friedrich, N. O.; De Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J Chem Inf Model* **2017**, *57* (11), 2719–2728. <https://doi.org/10.1021/acs.jcim.7b00505>.
- (44) *Pharmacophore definitions can be found at this address:* [https://github.com/franjongdee/bbSelect/blob/master/data/BaseFeatures\\_bbGAP.fdef](https://github.com/franjongdee/bbSelect/blob/master/data/BaseFeatures_bbGAP.fdef).
- (45) Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biol Cybern* **1982**, *43* (1), 59–69. <https://doi.org/10.1007/bf00337288>.
- (46) Vettigli, G. *MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map*. <https://github.com/JustGlowing/minisom/> (accessed 2021-07-21).
- (47) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat Chem* **2012**, *4* (2), 90–98. <https://doi.org/10.1038/nchem.1243>.
- (48) 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* (6), 435–449. <https://doi.org/10.1021/cn100008c>.
- (49) D. Segall, M. Multi-Parameter Optimization: Identifying High Quality Compounds with a Balance of Properties. *Curr Pharm Des* **2012**, *18* (9), 1292–1310. <https://doi.org/10.2174/138161212799436430>.

bbSelect - an open-source tool for performing a 3D pharmacophore-driven diverse selection of R-groups

## TOC graphic

