

SOFTWARE

Open Access



# Bioisoldentifier: an online free tool to investigate local structural replacements from PDB

Tinghao Zhang<sup>1†</sup>, Shaohua Sun<sup>1†</sup>, Runzhou Wang<sup>2</sup>, Ting Li<sup>1</sup>, Bicheng Gan<sup>3</sup> and Yuezhou Zhang<sup>1,4\*</sup>

## Abstract

Within the realm of contemporary medicinal chemistry, bioisosteres are empirically used to enhance potency and selectivity, improve adsorption, distribution, metabolism, excretion and toxicity profiles of drug candidates. It is believed that bioisosteric know-how may help bypass granted patents or generate novel intellectual property for commercialization. Beside the synthetic expertise, the drug discovery process also depends on efficient *in silico* tools. We hereby present Bioisoldentifier (BII), a web server aiming to uncover bioisosteric information for specific fragment. Using the Protein Data Bank as source, and specific substructures that the user attempt to surrogate as input, BII tries to find suitable fragments that fit well within the local protein active site. BII is a powerful computational tool that offers the ligand design ideas for bioisosteric replacing. For the validation of BII, catechol is conceived as model fragment attempted to be replaced, and many ideas are successfully offered. These outputs are hierarchically grouped according to structural similarity, and clustered based on unsupervised machine learning algorithms. In summary, we constructed a user-friendly interface to enable the viewing of top-ranking molecules for further experimental exploration. This makes BII a highly valuable tool for drug discovery. The BII web server is freely available to researchers and can be accessed at <http://www.aifordrugs.cn/index/>. Scientific Contribution: By designing a more optimal computational process for mining bioisosteric replacements from the publicly accessible PDB database, then deployed on a web server for thoroughly free access for researchers. Additionally, machine learning methods are applied to cluster the bioisosteric replacements searched by the platform, making a scientific contribution to facilitate chemists' selection of appropriate bioisosteric replacements. The number of bioisosteric replacements obtained using BII is significantly larger than the currently available platforms, which expanding the search space for effective local structural replacements.

<sup>†</sup>Tinghao Zhang and Shaohua Sun contributed equally to this work.

\*Correspondence:

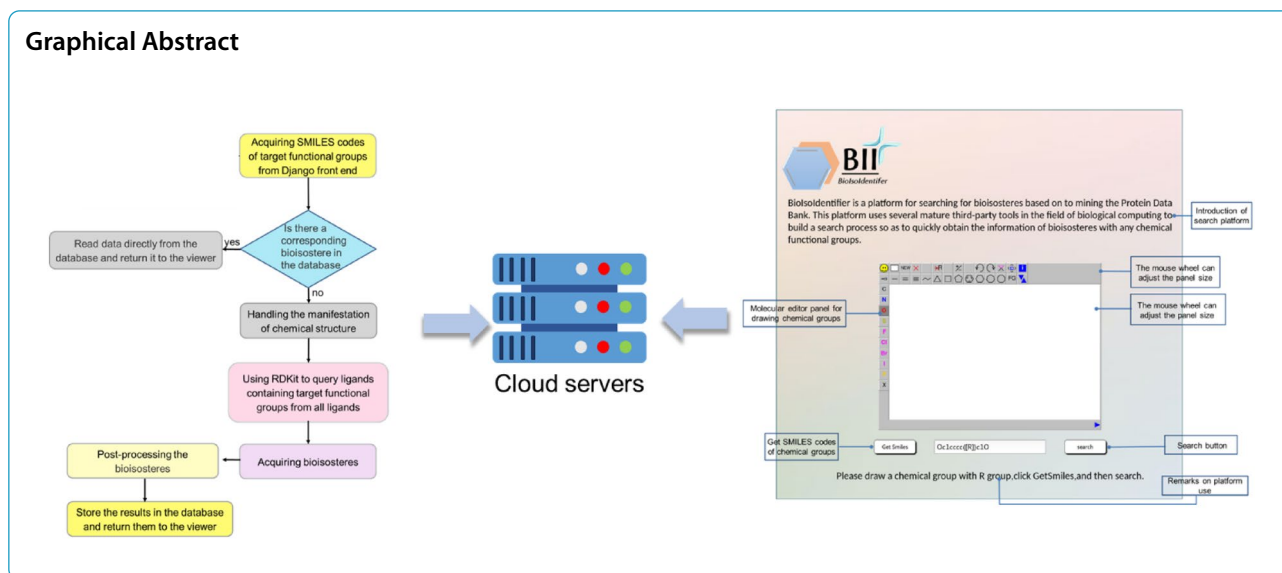
Yuezhou Zhang

iamyzzhang@nwpu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



## Introduction

It is essential to view databases not only as repositories of experimental results but also as valuable resources for data exploration and exploitation, particularly when mining data from publicly accessible databases. Among these, the Protein Data Bank (PDB), Cambridge Structural Database (CSD), and ChEMBL all contain rich implicit information that can be leveraged for drug discovery. ChEMBL, which aggregates chemical, bioactivity, and genomic data, is a meticulously curated database of bioactive molecules with drug-like properties [1]. EMBL-EBI recently released ChEMBL 30, which includes approximately 2.2 million compounds, 1.5 million assays, and 43,000 indications, all deposited and well-archived. Both CSD and PDB consist of ASCII files containing three-dimensional (3D) atomic coordinates of molecules, although they differ in terms of molecule size. Established in 1965, CSD serves as the global repository for organic crystal structures of small molecules, managed by the Cambridge Crystallographic Data Centre and updated thrice annually. As part of this commercialized project, several tools, including the CSD System, DASH, Mercury Menu, GOLD, and SuperStar, have been developed to provide comprehensive knowledge derived from CSD, making it widely utilized by the research and industrial communities.

Established in 1971 by the structural biology community as a central repository for macromolecular structure data, the PDB has consistently upheld a culture of open access and is now widely employed in fundamental biology, with millions of users leveraging its data to advance biomedical research [2]. Structural biology and structural bioinformatics have profoundly influenced

our understanding of the mechanisms and functions of biological macromolecules. The PDB serves as a custodian for all this data, representing the repository for the vast majority of accomplishments and milestones in the structural biology community. It also offers numerous additional sequence and structural annotations, along with tools for pairwise and multiple structure comparisons, including those for the analysis of ligands and their interactions. Therefore, PDB has the potential to be further utilized for specific applications. The cheminformatics and bioinformatics knowledge within PDB can be extracted through in-silico parsing of textual files. For instance Borrel et al. characterized the frequency, type, and density of the salt bridges during the ligand-receptor recognition [3], which can greatly benefit drug design. However, the development of tools and applications based on PDB data has fallen short of expectations, not to mention commercialized products.

A key challenge for medicinal chemists is to modulate the potency and selectivity of small therapeutics toward their biological targets and some believe that biososteric replacement is an effective strategy to expedite the process of identifying analogues with improved potency, intending to bypass existing patents [4]. Biosostericism, described as functional group exchanges to achieve similar biological outcomes, has garnered significant attention among practitioners. Biososteric replaceability relies on broader structural similarities to elicit the desired biological effects, rather than adhering strictly to physical or electronic mimicry. Typically, in medicinal chemistry, one modifies a promising pharmacophore by replacing specific functional groups with the aim of achieving the same biological response. Examples have

demonstrated that bioisosterism is a powerful tool for guiding successful drug development projects [5]. The replacement of the amide moiety and benzene ring of the phase II clinical candidate GSK'772 led to the discovery of more potent compounds with EC<sub>50</sub> values of 2.8 nM toward the target [6]. The surrogation of L-proline in melanostatin with 3-furoic acid has afforded two potent analogues with 2- and 4.3-fold improved EC<sub>50</sub> to dopamine D<sub>2</sub> receptors, respectively [7]. Instead of improving the potency of parent ligands by using local structural replacement approach, a brand-new molecule can also be created. Starting with a kinase inhibitor, Grigori et al. searched for commercially available replacements of the individual building blocks that constitute the parent ligand, then determined which fragments were suitable for merging into new compounds with a high binding affinity [8]. Referring to bioisosteric replacements strategy, Yang et al. developed DrugSpaceX database which dramatically diversified the modifications of the molecular framework thereby extended drug space [9]. Bioisosteric replacement as a tool for either anti-HIV drug design [10] or specific chemical moieties, including amide [11], phenyl [12] has been reviewed.

From a molecular perspective, bioisosteric replacement enable the conservative interactions between a ligand and a target protein [13] and this mutual recognition can be depicted *in silico*. Nowadays, computational tools have become indispensable in drug discovery process and have emerged to accelerate the acquisition of bioisosteric information from bio- or/and cheminformatic database. Analysing data from the PDB, the investigation into tetrazole-carboxylic acid bioisosterism revealed that protein binding site needs to be flexible enough to establish robust hydrogen bonds with tetrazolate ligands, especially when compared to carboxylate counterparts [14]. In a computational lead optimization process using bioisosterism, structural data of the target protein–ligand complex are leveraged [15] to modify the parent scaffold, following the principle of ensuring a suitable fit and interaction compatibility within the specific subpocket of the target protein [16]. Other than the extraction of bioisosteric information through computational tools, the identification of appropriate bioisosteres heavily relies on the experience of individual practitioners, making it subjective and potentially influenced by personal biases. While these semiempirical methods have been praised for offering alternatives, they frequently fall short in elucidating the underlying interaction mechanisms, particularly in how the bioisostere in question consistently interacts with the receptor in comparison to the reference moiety. Furthermore, having an excessive number of bioisosteres to choose from without proper organization and categorization could lead to the pitfalls of trial-and-error

screening, frustrating researchers who prefer a clear ranking of top candidates. As drug development costs rise, there is a growing need for a user-friendly, readily applicable system for bioisosteric information. However, it is currently lacking in this regard.

Due to the discrepancy between the vast, but under-used data repository and the increasing demand of medicinal chemists for valuable bioisosteres, especially those with implicit characteristics that are difficult to imagine or have not been previously experienced, there is a pressing need for computational methods that can efficiently traverse the database for such information. SwissBioisostere, hosted by the Swiss Institute of Bioinformatics and being accessible via a web interface [17], uses the ChEMBL database as a primary data source to identify matched molecular pairs by applying the Hussain and Rea algorithm after data curation. scPDB-Frag [18], differentiating from ligand based scaffold hopping, searches bioisosteric replacements from the protein–ligand interaction pattern. In contrast, KRIPPO [19], quantifies the similarities of binding site subpockets not only intra- but also interprotein family, broadening the application spectrum of bioisosterism. Seddon et al. fragmented the ligands for a given target using the BRICS scheme, then considered a pair of extracted moieties to be bioisosteric if they occupy a similar volume of the protein binding site [20].

A web tool to automate bioisosteric functional groups identification was developed by Novartis through the calculation of electronic, hydrophobic, steric, and hydrogen bonding properties as well as by the drug-likeness index of about 8.5 million unique organic substituents [21]. The web server MolOpt assists in drug design using bioisosteric transformations, with rules derived from data mining, deep generative machine learning, and similarity comparisons [22]. After the input of a protein and a ligand structure and users' selection of specific substructures which intended to replace, computational tool FragRep [23] tried to find suitable fragments that simultaneously match the geometric requirements of the remaining part of the ligand and well complementary with local protein environments. One crucial aspect of structure-based drug design is the use of GRID software to identify potential chemical modifications that can be made to known ligands. Recently Cross et al. proposed FragExplorer approach aiming to show users which fragments would best match the GRID molecular interaction fields in a protein binding pocket [24]. Craig Plot 2.0 fragmented ChEMBL database bioactive molecules, determined Hammett  $\sigma$  and Hansch-Fujita  $\pi$  values for their substituents, and grouped them by root or atom type, aiding in the selection of bioisosteric analogs [25].

Successful application of bioisosteric transformation hinges upon a thorough understanding of the physicochemical attributes of frequently encountered substituents, which can be accurately represented. For example, R-group descriptors encoding the distribution of atomic properties at increasing distances from a substituent's point-of-attachment to a central ring scaffold for identifying structurally similar pairs of substituents were reported by Holliday et al. [26] 3D descriptors Flexsim-R were calculated based on docking of small building blocks drug-like molecules into a reference panel of protein binding sites for bioisosteric functional groups [27]. So far, the acquisition of the bioisosteric information depending on (1) the experience of medicinal chemists working many years in the field; (2) mining the medicinal chemistry literature and extracting information by querying an internal library containing bioisosteric families [28]; (3) similarity in molecular physicochemical properties, including size, hydrophobicity, 3D substituents [29] or electron-donating profiles and (4) deep neural network trained on experimentally validated analogues extracted from medicinal chemistry literature [30].

The structural replacement of phosphate [31] and ribose [32] group identification was executed using our previously developed computational workflow, yielding some intriguing results. This protocol can be streamlined and led to the development of a user-friendly web server, BioIsoIdentifier (BII), equipped with fragment sketching tools. The process involves drawing the replacement fragment, converting it into Simplified Molecular Input Line Entry System (SMILES) code, and then processing it through the main program (Python and R). The program interfaces with third-party software, including Blastp, US-align, and RDKit, to organize individual PDB files. In this virtual system, spherical probes (2.5 Å radius) are created, targeting atoms within the reference ligand's chemical moiety for replacement as centroids. The sensed atoms serve as structural replacements for the reference fragment. To enhance output visualization, potential bioisosteric moieties are clustered based on structural similarity or unsupervised machine learning.

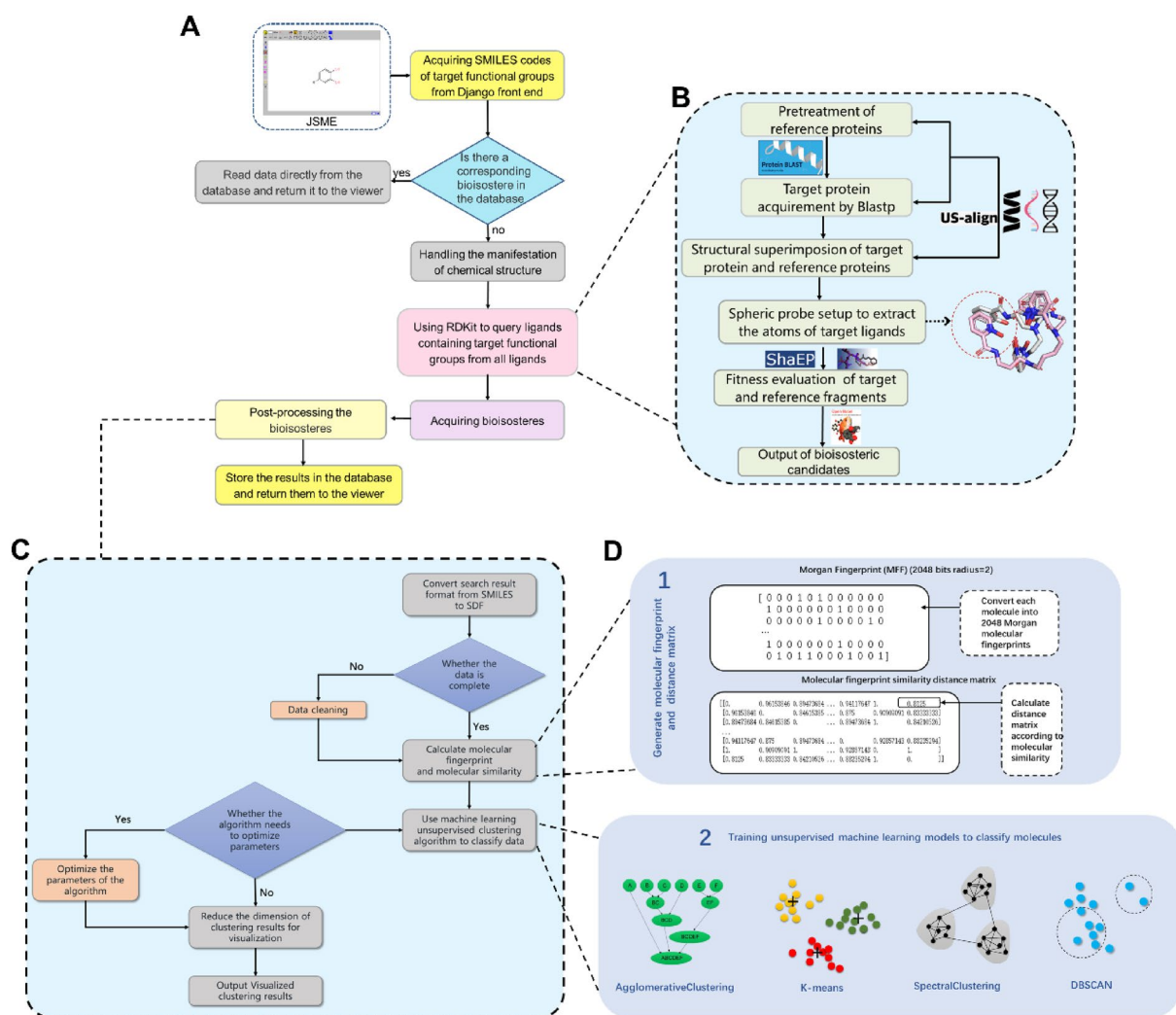
## Method

### Workflow of BII

BII identifies bioisosteres in six steps, as illustrated in Fig. 1. Users sketch the target functional group using JSME in the Django frontend and obtain the SMILES code, which is transmitted to the backend. The backend searches the database for stored bioisosteres based on the provided SMILES code. If found, results are directly retrieved. If not, further processing occurs, with ligands containing the target functional group queried from the PDB using RDKit's substructure search. These reference

ligands undergo a sequential search to obtain and save bioisosteres. The notable benefit of this approach arises from its ability to be explained through a molecular interaction perspective, leveraging information derived from PDB data to uncover details about local structural replacements. Figure 1B illustrates the specific calculation process.

1. PDB download: RCSB PDB provides a shell script, named "batch\_download.sh" (in S1), which can download multiple PDB archive files by providing a file containing a comma-separated list of PDB IDs. An essential prerequisite for running this script is to have the 'curl' tool installed. However, during our attempts to acquire the PDB archive, we encountered slow download speeds. Therefore, we developed a Python-based web crawler to swiftly retrieve the data.
2. Pretreatment of target protein: The small-molecule ligands with substructures intended to be bioisosterically replaced are selected from the PDB archive, with the macromolecular structures containing these ligands serving as reference proteins. We obtain the FASTA sequences of these proteins and input them into Blastp [33] to compare them with the sequences in the PDB, then output protein homologues with very close or identical structure.
3. Protein structure superimposition: Protein homologues exhibiting remarkably similar or identical structures are meticulously superimposed onto the reference protein using TM-align [34]. Subsequently, these alignments are further refined through the application of US-align [35] to achieve a more precise protein structure alignment.
4. Local structure extraction: Upon the successful alignment of these protein homologues, the atomic coordinates of the reference fragment earmarked for replacement within the reference protein are extracted. Each atom of the fragment functions as the centroid of a sphere with a radius of 2.5 Å. These spheres are employed to explore target ligand fragments, capturing atoms that come into contact, which are subsequently extracted and regarded as potential bioisosteric replacements for the reference substructure.
5. Fitness evaluation of extracted fragment with reference substructure: To assess the extent of overlap between the extracted fragments and the reference moiety, we utilized ShaEP [36], a tool designed for evaluating the similarity of ligand-sized molecules in terms of both shape and electrostatic potential. As per its definition, the fitness of a molecule pair based on ShaEP falls within the range of [0,1], with 1 signi-



**Fig. 1** The workflow of Bioisidentifier (BII) to identify the local structural replacements (LSR). **A** The complete workflow of BII; **B** the calculation process of obtaining LSR; **C** the process of LSR clustering with unsupervised algorithms; **D** calculation of molecular fingerprint, molecular similarity, and conduct unsupervised clustering

ifying a perfect match. In this context, we established a threshold of 0.2 based on empirical rules and experience.

- Output of extracted fragment with SMILES code: While computers are well-suited for processing textual strings, the human brain often finds graphical information more intuitive and comfortable to work with. To address both of these requirements, Open Babel [37], which enables the interconversion of more than 100 formats of chemical structures, was employed to specifically convert the SMILES string into an output fragment graph.

To classify the structural isosteres of the 3-substituted catechol, a clustering post-processing step was employed,

utilizing unsupervised machine learning. In this regard, several algorithms were experimented with and underwent parameter adjustments to optimize each one individually. The detailed process is illustrated in Fig. 1C and is described as follows:

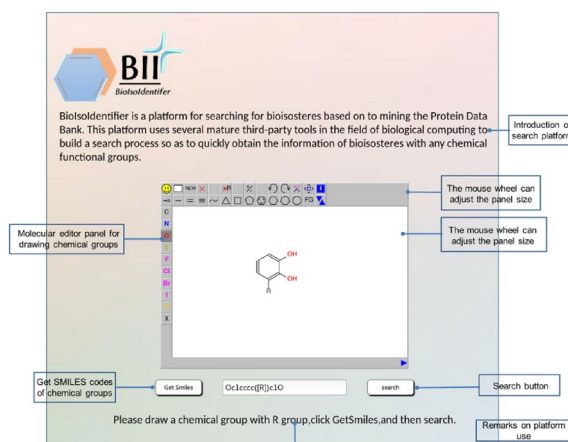
- Search result format conversion: To calculate molecular similarity for the subsequent calculations, the format of all search results was converted from SMILES to SDF format using custom-written code. Converting from SMILES to SDF format can result in potential loss of information. As a precaution, it is necessary to clean the data, which involves removing entries with missing content and eliminating duplicates.

- Molecular fingerprint and molecular similarity calculation: The molecular Morgan fingerprints were calculated at first, and then the RDKit tool was used to calculate the molecular similarity matrix through Tanimoto distance, as depicted in the zoomed-in view in Fig. 1D1
- Data classification by using machine learning unsupervised clustering algorithms: we explored the application of various unsupervised clustering algorithms, as illustrated in Fig. 1D2. These algorithms can be broadly categorized into two groups. The first category comprises algorithms like K-means and DbSCAN, which necessitate specifying the hyperparameter for the number of clusters. In contrast, the second category includes algorithms such as AgglomerativeClustering and AffinityPropagation, which do not require specifying the number of clusters.
- Optimization of algorithms parameters: For algorithms that necessitate the specification of additional hyperparameters, including the number of clusters, we employed techniques like the elbow method, silhouette coefficient method, and hyperparameter random search to optimize the clustering results by searching for the best parameters.
- Dimension reduction of clustering results for visualization: As previously mentioned, data points are stored in the form of 2048-bit MFF, which makes it challenging to effectively visualize clustering results in such high-dimensional space. Therefore, we employ principal component analysis (PCA) to reduce the data dimension from 2048 dimensions to 2D or 3D. We utilize the matplotlib tool to create visual representations and display the clustering results graphically.

## Web server

### Interface features and usage

Figure 2 displays a screenshot of the BII homepage, featuring a concise introduction and a web server input interface. Users can draw the chemical structure of the target functional groups in the molecular editor JSME. The 'R' denotes the vertex where the target functional group bifurcates, indicating that only the sketched core substructure requires replacement. The input fragment is always assumed to be complete. Once the structural construction is complete, users can obtain the SMILES code corresponding to the target functional group by clicking the "Get Smiles" button on the page. Subsequently, they can initiate the LSR search by clicking the "search" button.



**Fig. 2** The interface of BII

## Implementation

The Django web framework and Python code are employed to develop the interface functionality of the web server and execute MySQL database queries for ligand substructure replacement. RDKit [38] is utilized to facilitate fragment database construction, calculate molecular descriptors, and depict 2D molecular structures.

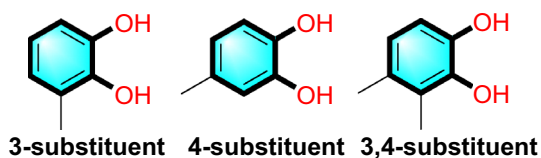
## Case study

Catechol, an unsaturated six-carbon ring (phenolic group) with two hydroxyl groups attached to adjacent carbons (dihydroxyphenol), is a widely observed group in neurotransmitters such as dopamine and noradrenaline. The nitrocatechol based compounds tolcapone and entacapone are successfully used as adjuncts to treat Parkinson's Disease. Meanwhile, bisubstrate and non-nitro hydroxypyridone catechol *O*-methyltransferase (COMT) inhibitors have also been reported for the same disease. However, tolcapone and entacapone mainly act peripherally and poorly penetrate brain as centrally acting drugs. Besides, phenolic compounds are prone to high metabolic clearance due to their acidity and polarity. Therefore, next generation COMT inhibitor prefer replace catechol with corresponding bioisostere [39]. This need has drawn our attention to explore catechol bioisosteres, which we present as a case study. Apart from the two contact points of the hydroxyl group in the benzene ring, four other positions are available for ligand extension, representing three types (Fig. 3) of possible catechol containing ligands.

## Results and discussion

### The LSR of catechol

When inputting a 3-substituted catechol encoded as *Oc1ccc([R])c1O* into the server, it suggests over 496 replacement ideas, all of which are displayed in a table,



**Fig. 3** Three possible catechol containing ligands

paginated for convenience. Figure 4 provides a snapshot of the first page, showcasing the clustering results represented in both two-dimensional and three-dimensional structures. The remaining replacements are documented in Additional file 1: Figure S2. Each entry in the table includes valuable information such as SMILE codes, 2D and 3D representations, a similarity index, as well as the associated reference protein complex and its corresponding ligand PDB ID, along with details of the target protein complex and its related ligand PDB ID.

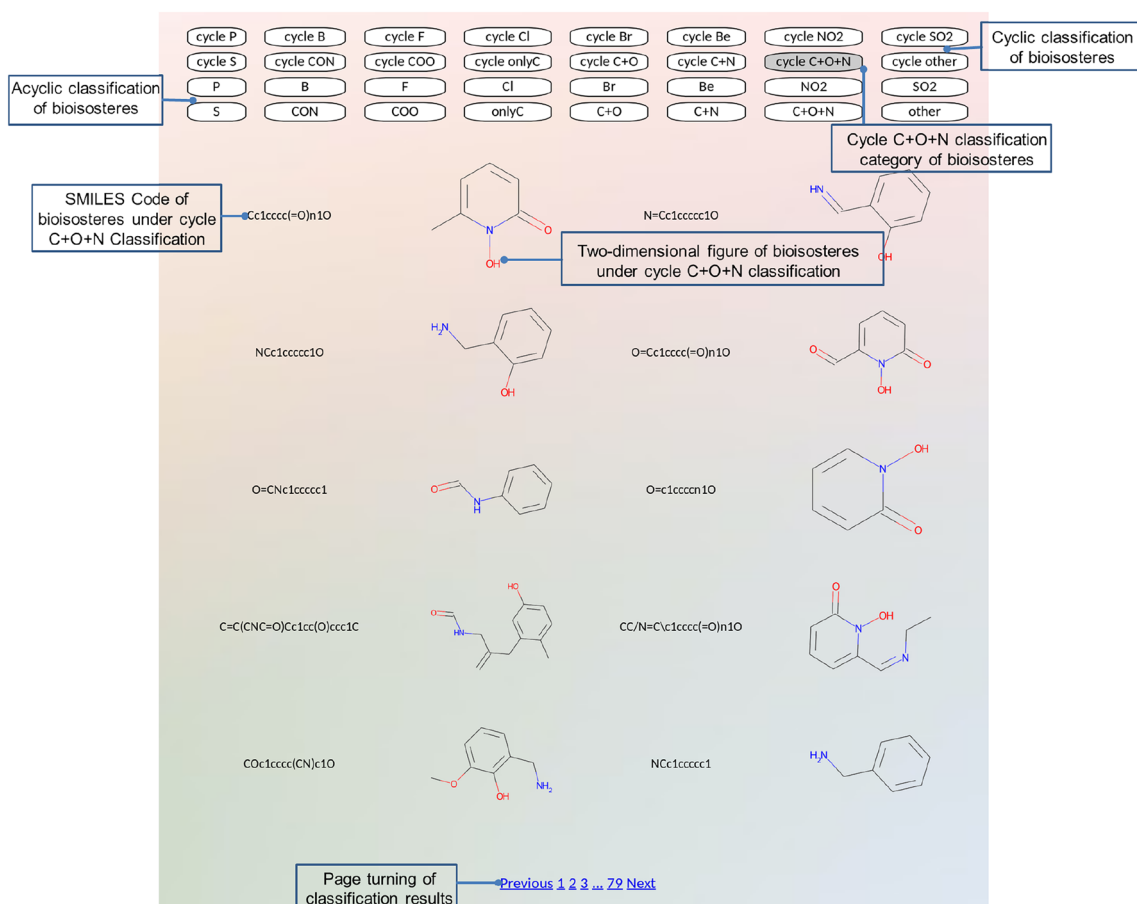
The LSR of 3-substituted catechol are first sorted according to their ShaEP index and subsequently

recorded in a table. Based on their structural similarity, they are then hierarchically classified into 32 distinct groups. Users can easily visualize this classification by clicking on the “Classification” tab. For a more detailed view, specific LSR included in the “C+O+N” group are exemplified in Fig. 5, accessible by clicking the corresponding group name. Moreover, unsupervised learning algorithms have been employed to further refine and narrow down the number of subgroups.

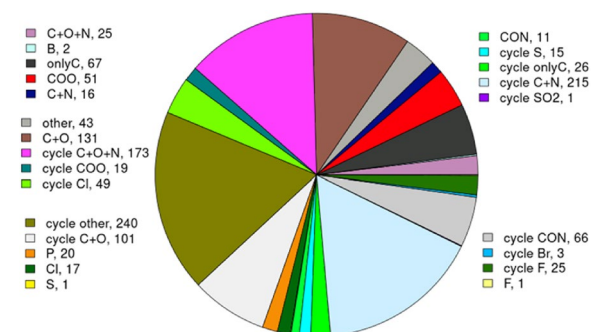
Figure 6 illustrates the categorization of LSR for 3-substituted catechol recognized using BII. They are sorted into 24 categories based on the SMILES code. Among these, 240 bioisosteres, although belonging to cyclic structures, do not fall into any predefined category; therefore, they are grouped under [cycle other], making it the largest family. This is followed by 215 members categorized under [cycle C+N], and there is only one bioisostere in the [F] category. For further insights, bioisosteres of 4-substituted and 3,4-substituted catechol are also presented individually in Additional file 1: Figure S3 and S4. Notably, the primary focus of this work is

SMILES code of bioisosteres	SMILES	Bioisostere_2D	Bioisostere_3D	ShaEP_score	Ref_complex	Ref_ligand	Ref_similar_complex	Ref_similar_ligand
	<chem>Cc1ccc(F)c(O)c1</chem>			0.91	5EN4	5Q6	5O42	9JW
	<chem>Nc1ccccc1</chem>			0.9	3K3L	MCK	3T2S	PHU
	<chem>N=Cc1cccc(=O)n1O</chem>			0.89	3I0A	MCJ	4ZHH	4OL
	<chem>CN(C)c1ccc(F)c(O)c1</chem>			0.87	5EN4	5Q6	5O43	9JQ
	<chem>NC(=O)c1cccc1O</chem>			0.87	1MDB	DBH	3O83	IXN
	<chem>Cc1c[nH]c2ccccc12</chem>			0.87	7KK0	CAQ	7KFO	IAC
	<chem>O=Cc1ccc(F)c(O)c1</chem>			0.86	5EN4	5Q6	6ZR6	F45
	<chem>C=Nc1ccccc1</chem>			0.85	3I0A	3ET	3T2S	PHU
	<chem>Cc1cccc(O)c1C(O-)</chem>			0.85	4IZ6	1HZ	5WM4	B5M
	<chem>Oc1c(F)ccc(Cl)c1F</chem>			0.85	5EN4	5Q6	6ZT2	QPK

**Fig. 4** The LSR list of 3-substituted catechol



**Fig. 5** The LSR subgroup of 3-substituent catechol categorized as cycle C+O+N



**Fig. 6** Distribution of the data set into categories assigned based on the SMILES codes of the structural isosteres of 3-substituent catechol

on the conservativity of interactions between the parent ligand moiety and the protein, without explicitly discriminating between the replacement of the moiety and the generation of entirely new molecules. While BII may suggest local structural replacements for specific moieties

in the catechol example, our goal is to identify bioisosteric replacements with greater stringency. Our approach involves superimposing proteins with identical groups but accommodating different ligands. We then concentrate on the space where the intended moiety is to be replaced. The docking of replacement moieties into the original catechol's position may induce a shape change in the binding pocket due to its flexibility. Importantly, our approach can be applied to scaffold hopping and the generation of combinatorial libraries to a certain extent.

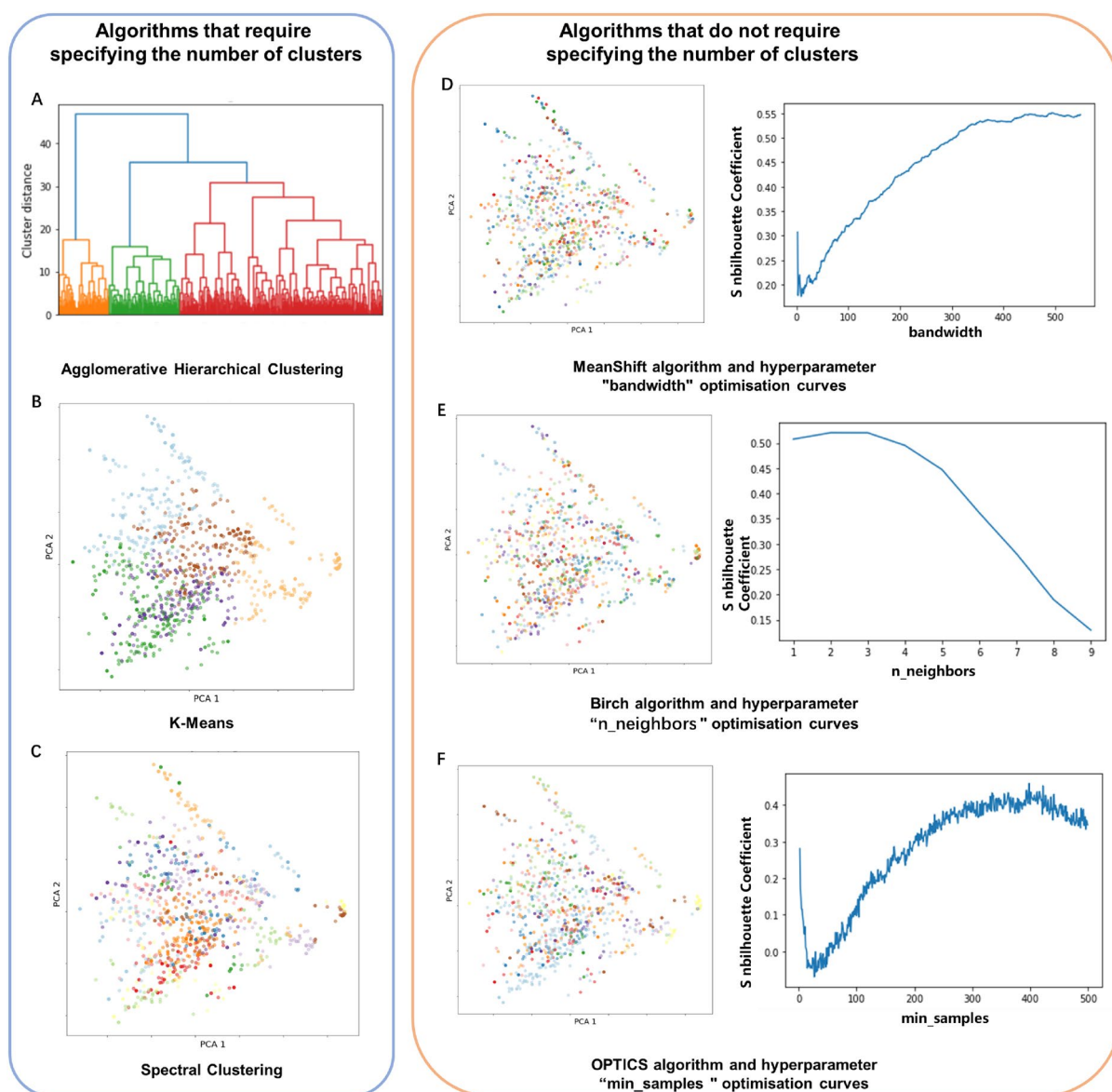
Unsupervised clustering methods are employed to categorize structural replacements of 3-substituent catechol into fewer categories, utilizing the SMILES encoding approach. This unsupervised clustering unveils latent similarities among these structural replacements, thereby simplifying data complexity and enhancing comprehensibility and visualization. This simplification streamlines the selection of representative samples from each cluster, facilitating in-depth research and, consequently, enhancing screening efficiency. In Fig. 7, you can observe the results obtained from the application of various



algorithms and their respective optimization techniques. The algorithms are divided into two categories based on the necessity of pre-specifying the number of clusters, each category employing unique hyperparameter optimization strategies. For algorithms where pre-specifying the cluster number is unnecessary, as exemplified by the MeanShift algorithm, we construct an optimization curve that correlates the “bandwidth” hyperparameter with the silhouette coefficient to determine the optimal “bandwidth” value of 446. This corresponds to a cluster count of 47 with an average silhouette coefficient of 0.561. The

Birch clustering algorithm employs a similar approach to ascertain the optimal “n\_neighbors” hyperparameter value, achieving the highest silhouette coefficient of 0.519 when “n\_neighbors” equals 3. In the case of algorithms requiring a predefined number of cluster groups, a more intricate method is employed to determine the optimal cluster count.

Figure 8 illustrates the process of determining the optimal number of clusters for the K-Means algorithm. The optimal number of clusters was determined using the elbow rule and the silhouette coefficient method,

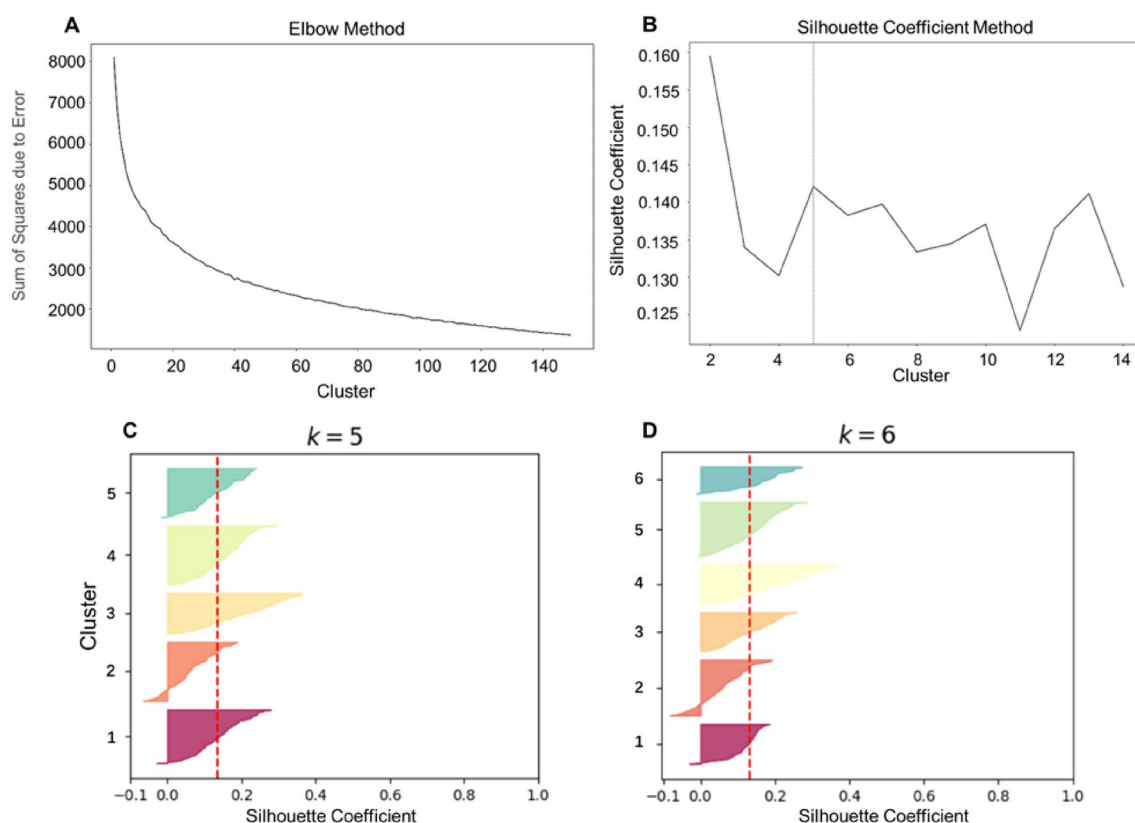


**Fig. 7** Results using algorithms without hyperparameters and algorithms requiring hyperparameters. **A** Agglomerative Hierarchical Clustering; **B** K-Means Clustering; **C** spectral clustering; **D** MeanShift algorithm and hyperparameter “bandwidth” optimisation curves; **E** Birch algorithm and hyperparameter “n\_neighbors” optimisation curves; **F** OPTICS algorithm and hyperparameter “min\_samples” optimisation curves

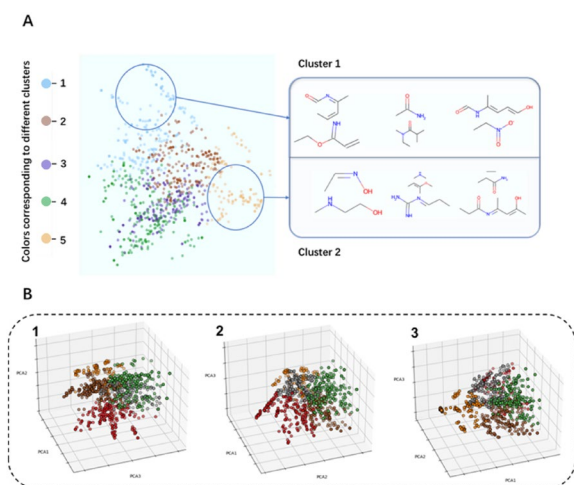
individually for rational segregation of the structural replacements in the chemical space. The elbow method and silhouette coefficient method are used to determine the optimal number of clusters. Figure 8A shows that the elbow of the sum of squares due to error (SSE) sharply drops when the number of classes is less than 15. It can be observed that the largest value of  $k$  for the contour coefficient is 2. However, the elbow diagram of  $k$  and SSE reveals that the SSE is still relatively large when  $k$  is taken as 2. This is due to that the contour coefficient takes into account the degree of separation, and so it is an irrational number of clusters for  $k=2$ . Therefore, retreating to the second largest value of  $k$  for the contour coefficient, we consider the second largest value of  $k$  for the contour coefficient. Further analysis of the relationship between the silhouette coefficient and the number of clusters (Fig. 8B) reveals that the best cluster number (the number of clusters with the maximum silhouette coefficient) is 5. To verify this conclusion, silhouette coefficient diagrams for each class were plotted separately for clustering with 5 and 6 classes, and the average silhouette coefficients of the clustering results are indicated by the red dashed line. As shown in Figs. 8C and D, each class

was more uniformly distributed when the cluster number was 5, supporting the empirical division of the LSR of 3-substituent catechol into 5 groups accordingly. It should be noted that the presented computational results are illustrative of our computational process using 3-substituted catechol as an example, which is why some algorithms may have lower silhouette scores.

To provide a detailed view of the clustering results of 3-substituted catechol LSR, principal component analysis (PCA) was employed to reduce the dimensionality of the 2048-dimensional data to 2D or 3D, as demonstrated in Fig. 9A for 2D visualization and Fig. 9B for additional perspectives on the 2D and 3D visualization, which are summarized in Additional file 1: Figure S5. In Fig. 9, dots of the same color represent a category, and two categories are chosen as examples to present a list of classified molecules. The acidity dissociation constants for catechol are  $pK_{a1}$  of 9.25 and  $pK_{a2}$  of 13.0 [40], suggested that the catechol is slightly acidic at biological environment of pH 7.4, it is therefore thought acidic groups are intrinsic biosisosteres of catechol to conserve molecular interactions where possible. However, we envision it is likely that basic groups might be



**Fig. 8** The strategy for determining the optimal number of clusters for the K-Means algorithm. **A** The optimal cluster number determination using elbow rule; **B** the optimal cluster number determination using contour coefficient; **C, D** determine the optimal number of clusters by comparing the contour coefficients of different clusters



**Fig. 9** K-Means algorithm clustering results visualised by PCA (principal component analysis) for dimensionality reduction. **A** Two-dimensional visualization clustering space; **B** 3D visualization clustering space; B1 front view; B2 left view; B3 vertical view

suggested by our BII tool. It is not surprise since our previous investigation revealed that basic  $-\text{CH}_2\text{NH}_3^+$  replaced acidic phosphate group and a  $\text{Mg}^{2+}$  concurrently [31]. The metal cations hence may play an important role during local structure replacement of catechol since they can readily coordinate.

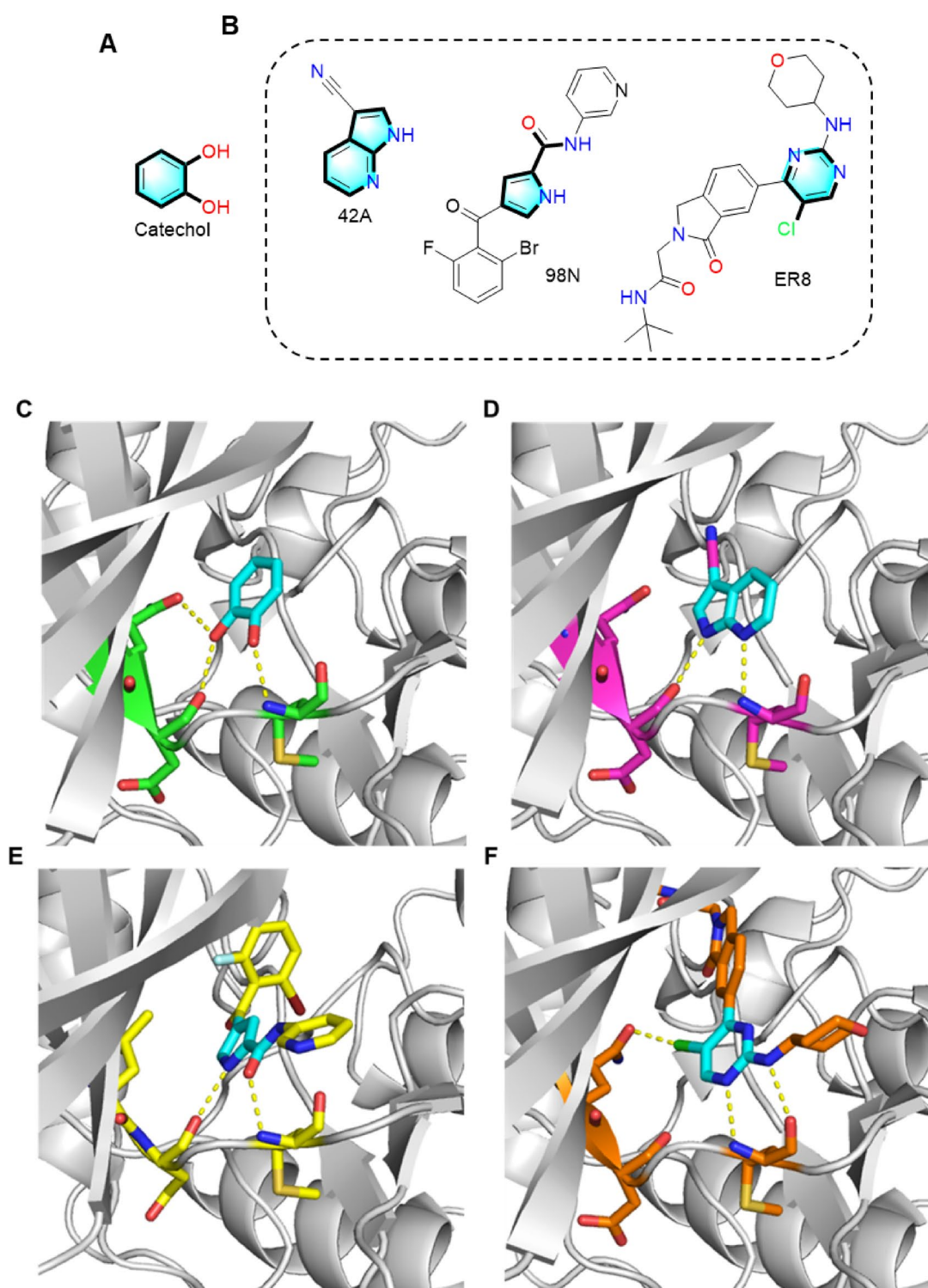
Three optional LSRs of catechol are displayed in Fig. 10, where it can be observed that these newly identified substructures exhibit similarities in shape to catechol. To elucidate structure–activity relationship of catechol and corresponding replacements, the structural and biological data are compiled from reference publications. In addition, we leveraged the structure diversification of identified new chemicals with activity change toward a selected target, discussed how substitutes deletion or protrusion impacts the biological activity of resulting molecules. The therapeutic impact of catechol in lung cancer treatment was achieved by inhibiting the activity of extracellular signal-regulated kinase 2 (ERK2), and its direct binding to the active site of ERK2 (PDB code: 4ZXT) was confirmed through X-ray crystallography [41]. Catechol was anchored to the hinge loop of the ATP-binding site of ERK2, with its hydroxyl groups interacting with the main chain of Asp106, Met108, and the side chain of Gln105, all located on the hinge loop. The azaindole ligand (compound 3 in Ref. [42] PDB code: 42A) occupied the same binding site where catechol was positioned in ERK2. In detail, the pyrrole NH of 7-azaindole formed a strong hydrogen bond ( $d=2.8 \text{ \AA}$ ) with the backbone carboxyl oxygen of Asp104, and the pyridine nitrogen served as a hydrogen bond acceptor ( $d=3.0 \text{ \AA}$ ) for the Met106

backbone NH. The ligand (compound 46 in Ref. [43] PDB code: 9N8) binds in the ATP-binding site of ERK5.

The pyrrole NH and amide carbonyl formed hydrogen bonds ( $d=2.8 \text{ \AA}$ ,  $d=2.7 \text{ \AA}$ ) with the backbone carbonyl of Asp138 and amide of Met140 in the ERK5 hinge-region, respectively. Noticeably, the pyrrole-2-carboxamide took the position of catechol. The chloro-substituted aminopyrimidine moiety of ER8 (compound 15 in Ref. [44]) took the space of catechol as so that halogen bond ( $d=2.7 \text{ \AA}$ ) between the the chloro atom and amide residue oxygen of gatekeeper Gln105. Hydrogen bonds ( $d=3.1 \text{ \AA}$ ,  $d=2.9 \text{ \AA}$ ) were observed between the ligand's pyrimidine N, amino NH and the backbone NH, C=O of hinge residue Met108 respectively. C=O of hinge residue Met108 respectively. The p38 $\alpha$ MAPK inhibitor hit (compound 3 in Ref. [45] PDB code: MWL) occupied the active site space of p38 $\alpha$ MAPK.

The pyridine ring nitrogen allowed for hydrogen bonding ( $d=2.8 \text{ \AA}$ ) with the peptide backbone of Met109 from the hinge region. In this context, the pyridine moiety can be considered a structural replacement for the C=O of hinge residue Met108, effectively taking the place of catechol. The idea bioisosteres by definition, entails both steric and but electronic conservatism. However, achieving a perfect match for both criteria simultaneously can be challenging and may require some degree of compromise. It's conceivable that an imperfect match in electronic conservativity could be compensated for by a precise steric fit, thereby maintaining overall binding affinity. It should be acknowledged that the inability of BII to distinguish between hydrogen bond donors and acceptors, as it primarily focuses on the conservativity of the interaction itself. For instance, the hydroxyl group in catechol serves as a hydrogen bond receptor in the reference, whereas the  $-\text{C}=\text{O}$  group of the carboxamide in ligand 9N8 can only function as a hydrogen bond (HD) acceptor due to its electron-rich nature. The same applies to the cationic  $-\text{N}(\text{CH}_3)-$  group, which acts as a HD acceptor.

The human enzyme 17 $\beta$ -hydroxysteroid dehydrogenase 14 (17 $\beta$ -HSD14), using  $\text{NAD}^+$  as cofactor, oxidizes estradiol and 5-androstenediol. The human HSD17B14 gene is widely expressed in major organs, such as brain, liver and kidney. It has also been identified in breast cancer tissue, but the physiological function of this enzyme was poorly understood. The use of inhibitors can be important tools to study the physiological role of 17 $\beta$ -HSD14 in vivo. The methanone compound 1 (compound 12 in Ref. [46] PDB code: 5Q6) inhibits the activity of 17 $\beta$ -HSD14 with  $K_i$  of 64 nM. The hydroxyl residue of Tyr154 forms two hydrogen bonds bifurcately ( $d=2.5 \text{ \AA}$ ,  $d=3.1 \text{ \AA}$ ) with hydroxyl groups of the catechol moiety. Besides, the 4-OH hydrogen bond ( $d=2.5 \text{ \AA}$ ) also extends toward Ser141 hydroxyl



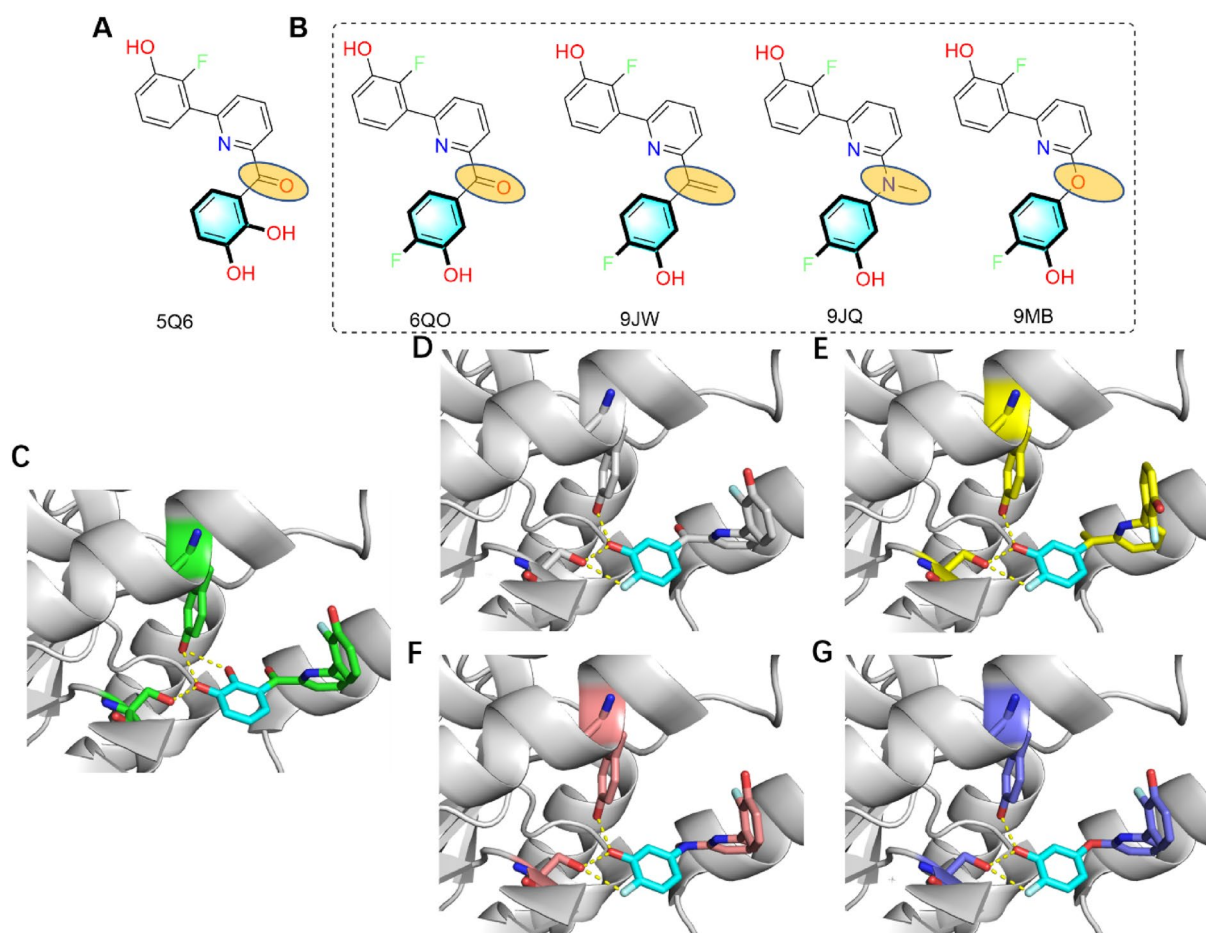
**Fig. 10** **A** Structure of Catechol. **B** Three active ERK2 inhibitors suggested from a BII search. **C–F** Interaction networks between 42A and 98N, ER8, MWL and ERK2/MAPK, respectively. In this figure, the ligands are named according to their PDB 3-lettercodes, and the proteins are named according to PDB4-lettercodes. Putative hydrogen bonds are shown as yellow dotted lines and the distance is labelled. The carbon atoms of structural replacements in the target ligand are highlighted in cyan, while others are shown in green, purple, yellow and brown, respectively

residue (Fig. 11A). Four of 5Q6's optional analogues are shown in Fig. 11B and suggested that 4-fluoro-3-phenol is the bioisostere of the 3-substituent catechol, offering a ligand (compound 9 in Ref. [46] PDB code: 6QO) with increased affinity (a  $K_i$  of 13 nM). The 3-OH groups at the C-ring of 9 and compound 12 in Ref. [46] interact through remarkably short H-bond interactions with the side chain of Tyr154 (9,  $d=2.3$  Å, 12,  $d=2.5$  Å) and the side chain of Ser141 (9,  $d=2.5$  Å, 12,  $d=2.5$  Å) from the catalytic triad. The 4-F group at the C-ring of 9 is possibly involved in forming a halogen bond ( $d=2.8$  Å) with Ser141 hydroxyl side residue. The 3-OH groups at the C-ring of 12 hydrogen bond toward the side chain of Tyr154 ( $d=3.1$  Å). The replacement of the ketone linker of compound 9 with ethenyl resulted in an eightfold more potent inhibitor (compound 5 in ref. PDB code: 9JW) with a  $K_i$  of 1.5 nM; while methylamine (compound 4 in ref. PDB code: 9JQ) and ether (compound 2 in reference

PDB code: 9 MB) surrogate each individually deteriorated the binding affinity to a  $K_i$  of 42 and 58 nM. Keeping the B and C ring of 6QO unchanged, the equipotent quinoline base inhibitor (compound 9 in Ref. [47], PDB code: 9ME), and a two folds more active naphthalene derivative (compound 10 in Ref. [47]) were obtained, but the quinoline analog was found to be four times more soluble than the naphthalene compound. Herein, we rather than concentrate on the structural replacement of catechol, where it is replaced by a 4-fluoro-3-hydroxyphenyl moiety, instead emphasize that the linker connecting replacements to other parts can vary. However, it's crucial to acknowledge that the choice of linker may impact the physicochemical properties of the ligand.

#### Comparison with other tools

The fundamental of isostere replacement lies matching of protein moieties, but sometimes this concept of



**Fig. 11** **A** Structure of 5Q6. **B** Four active 17β-HSD14 inhibitors suggested from a BII search. **C–G** Interaction networks between 6QO, 9JW, 9JQ and 9 MB and 17β-HSD14, respectively. In this figure, the ligands are named according to their PDB 3-lettercodes, and the proteins are named according to PDB4-lettercodes. Putative hydrogen bonds are shown as yellow dotted lines and the distance is labelled. The carbon atoms of structural replacements in the target ligand are highlighted in cyan, while others are shown in green, white, yellow, bronze and blue respectively

replacement not aligned with the intended objective of functional group/ring/core replacement for a ligand. Therefore, BII was compared with other bioisosteric search tools, such as the SwissBioisostere database and the MolOpt network server. The SwissBioisostere database is a comprehensive resource containing information about molecular substitutions and their performance in biochemical analysis. This data is obtained by matching molecular pairs and mining biological activity data from the ChEMBL database. Notably, SwissBioisostere not only provides information about molecular substitutions but also offers interactive analysis capabilities. On the other hand, the MolOpt network server is constructed through a combination of data mining, cheminformatics similarity comparison, and machine learning techniques. Users have the flexibility to query for bioisosteres of specific molecular substructures and even generate entirely new molecular alternatives.

To perform a comparative analysis, three distinct substructures, namely the 3-substituent, 4-substituent, and 3,4-substituent, were input into each of the three search tools. Consequently, users can access the corresponding bioisosteric data for their chosen substructures. In Table 1, we have summarized the number of bioisosteres identified by SwissBioisostere, MolOpt, and BII. Additionally, it's important to note that MolOpt offers four distinct bioisosteric replacement rules. MolOpt-1 is based on data mining principles, MolOpt-2 utilizes similarity comparison, MolOpt-3 incorporates data mining techniques, and MolOpt-4 is designed around a deep generative model. It becomes evident that when compared to the SwissBioisostere database and the MolOpt web server, BII excels in providing a more extensive array of bioisosteric ideas, making it a valuable resource for medicinal chemistry research. The bioisosteres with the top-ten rankings from each tool are depicted in Fig. 12, illustrating consistent results. The chemical accessibility represents an important concern indeed for the novel structure generated based on this tool, but we want to emphasize that BII focus on local structural replacements yet did not consider how to incorporate suggested moieties into new ligands, but definitely it will be put

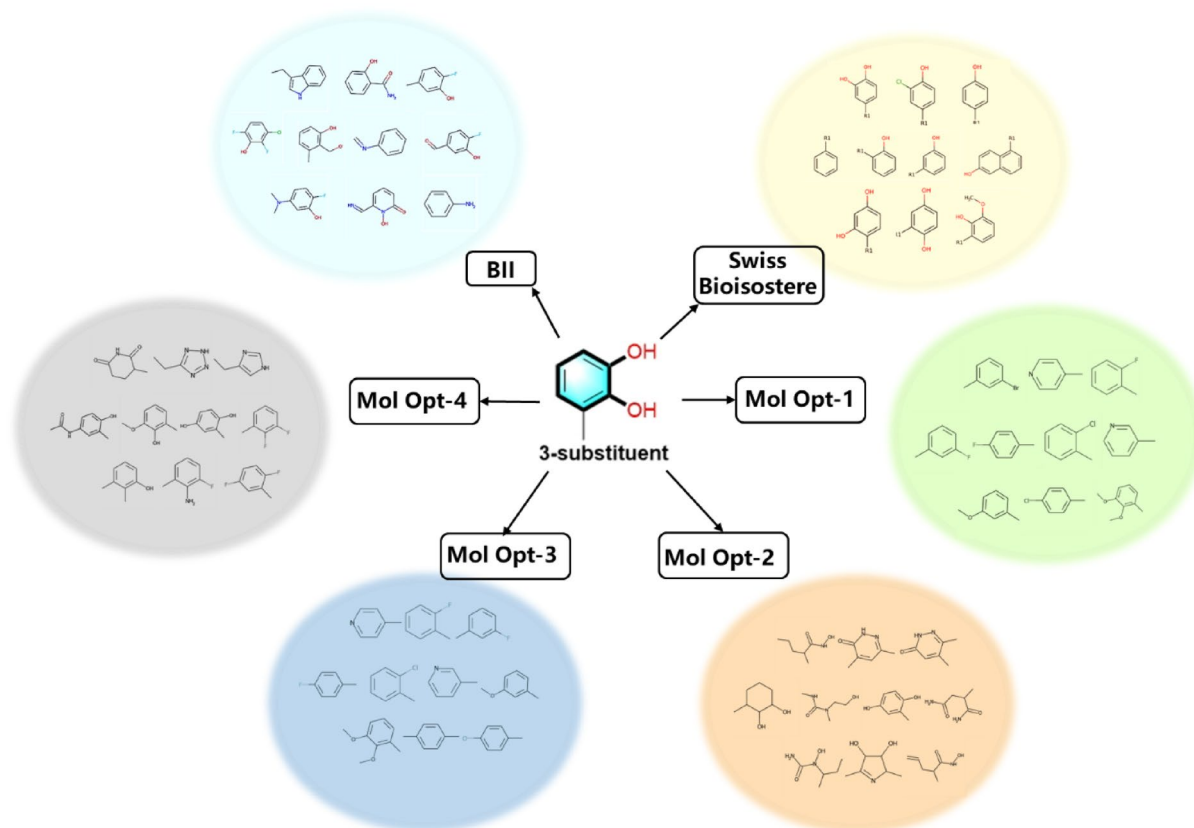
into consideration as a filter of replacement moieties in updated BII version. In addition, we recognized that a retrospective validation is not satisfactory to launch BII since experimental validation in any case is a benchmark of computational tool. In fact, we conducted both wet lab synthetic and bioassay experiments in-house. It has been demonstrated that a squaryldiamide or an amide group is the bioisosteric replacement of phosphate moiety [48], NH in the urea serves as isostere of carboxylic acid [49]. After previous computational investigation of phosphate [31], ribose [32] bioisosteric replacement, the bioisosterism of these moieties have been verified. Consequently, we think it is necessitated to develop a generic tool to facilitate bioisostere identification of any chemical fragment, which pillars the basement of our current attempt.

## Conclusions

To optimize the efficiency of BII, we integrated the extended multiprocessing library of Python into the code. BII stands out as a user-friendly and robust tool for generating innovative ligand replacement ideas. The substructure replacement identification process for a specific single task typically takes about two to eleven hours using a machine with a CPU of 24 processors. Notably, the web server is designed to be accessible without the need for computational or programming skills, a feature particularly advantageous for medicinal chemists. These results affirm BII's capability to identify suitable LSR where the chemical structure differs, yet the interaction patterns with the protein pocket remain conserved. Moreover, our application of BII has led to the rediscovery of scaffold hopping ideas, underscoring the utility of our web server in providing valuable insights for ligand design. In essence, BII serves as a valuable tool to assist medicinal chemists during the hit/lead optimization process, aiding in the search for appropriate molecular fragments. As part of our commitment to ongoing improvement, the BII server will receive regular updates as new data and advancements become available. We are pleased to offer this service freely to the public at <http://www.aifordrugs.cn/index/>.

**Table 1** Comparison of query results of different search platforms

Target functional group	The number of bioisosteres found by different search tools					
	SwissBioisostere	MolOpt-1	MolOpt-2	MolOpt-3	MolOpt-4	BII
3-substituent	161	100	200	121	200	496
4-substituent	631	100	200	200	200	2559
3,4-substituent	56	9	200	9	200	3322



**Fig. 12** Top-10 ranked bioisosteres of 3-substituent catechol suggested by different tools

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-024-00801-8>.

**Additional file 1:** S1:batch\_download.sh # Python script to download the PDB database code: S2. Taking 3-substituent as the target functional group, the bioelectronic isoplatoon was searched in BII, and the results were as follows, a total of 50 pages of data. S3: The LSR subgroup of 4-substituent catechol categorized as cycle C+O+N. S4 The LSR subgroup of 3,4-substituent catechol categorized as cycle C+O+N. S5 Visualization of the data clustering.

## Acknowledgements

This research was sponsored by the Joint Research Funds of Department of Science & Technology of Shaanxi Province, Northwestern Polytechnical University (No. 2020GXLH-Z-017), funded by Ningbo Natural Science Foundation (No. 202003N4006) and the key research program of Ningbo (No. 2023Z210).

## Author contributions

The study was designed and conceptualized by YZZ and RZW. The workflow was developed by THZ and TL. The deployment and operation of cloud services was performed by SHS and BCG. The results were discussed and interpreted by all authors. The manuscript was written by YZZ and advanced by all authors.

## Funding

This study was supported by Ningbo Natural Science Foundation (202003N4006), the key research program of Ningbo (2023Z210), the Joint Research Funds of Department of Science & Technology of Shaanxi Province.

## Availability of data and materials

The focus of our manuscript is on the online webserver development computational to identify local structural replacements/bioisosteres for drug design. ChemDraw 19.0 was used to sketch the structure of ligands. The PyMol 1.8.x used in this work to visualize and demonstrate the interactions between ligand and receptor is free and open-source software. All code, data and deployment environments for this work have been uploaded to Zeodo and can be accessed via the following link: <https://doi.org/https://doi.org/10.5281/zenodo.8215113>.

## Declarations

### Competing interests

There are no conflicts to declare.

### Author details

<sup>1</sup>Xi'an Institute of Flexible Electronics (IFE) and Xi'an Institute of Biomedical Materials & Engineering (IBME), Northwestern Polytechnical University, 127 West Youyi Road, Xi'an 710072, China. <sup>2</sup>School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China. <sup>3</sup>College of Petroleum Engineering, Northeast Petroleum University, Daqing 163318,

Heilongjiang, China. <sup>4</sup>Ningbo Institute of Northwestern Polytechnical University, Frontiers Science Center for Flexible Electronics (FSCFE), Key Laboratory of Flexible Electronics of Zhejiang Province, Ningbo Institute of Northwestern Polytechnical University, 218 Qingyi Road, Ningbo 315103, China.

Received: 24 October 2023 Accepted: 7 January 2024

Published online: 13 January 2024

## References

- Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E, Davies M, Dedman N, Karlsson A, Magariños MP, Overington JP, Papadatos G, Smit I, Leach AR (2017) The ChEMBL database in 2017. *Nucleic Acids Res* 45:D945–D954
- Bonvin AMJJ (2021) 50 years of PDB: a catalyst in structural biology. *Nat Methods* 18:448–449
- Borrel A, Camproux A-C, Xhaard H (2017) Characterization of ionizable groups' environments in proteins and protein-ligand complexes through a statistical analysis of the protein data bank. *ACS Omega* 2:7359–7374
- Brown N (2014) Bioisosteres and scaffold hopping in medicinal chemistry. *Mol Inform* 33:458–462
- Agnew-Francis KA, Williams CM (2020) Squaramides as bioisosteres in contemporary drug design. *Chem Rev* 120:11616–11650
- Xia C, Yao Z, Xu L, Zhang W, Chen H, Zhuang C (2021) Structure-based bioisosterism design of thio-benzoxazepinones as novel necroptosis inhibitors. *Eur J Med Chem* 220:113484
- Sampaio-Dias IE, Reis-Mendes A, Costa VM, García-Mera X, Brea J, Loza MI, Pires-Lima BL, Alcoholido C, Algarra M, Rodríguez-Borges JE (2021) Discovery of new potent positive allosteric modulators of dopamine D<sub>2</sub> receptors: insights into the bioisosteric replacement of proline to 3-furoic acid in the melanostatin neuropeptide. *J Med Chem* 64:6209–6220
- Andrianov GV, Gabriel Ong WJ, Serebriiskii I, Karanicolas J (2021) Efficient hit-to-lead searching of kinase inhibitor chemical space via computational fragment merging. *J Chem Inf Model* 61:5967–5987
- Yang T, Li Z, Chen Y, Feng D, Wang G, Fu Z, Ding X, Tan X, Zhao J, Luo X, Chen K, Jiang H, Zheng M (2021) DrugSpaceX: a large screenable and synthetically tractable database extending drug space. *Nucleic Acids Res* 49:D1170–D1178
- Dick A, Cocklin S (2020) Bioisosteric Replacement as a tool in anti-HIV drug design. *Pharmaceuticals (Basel)* 13:36
- Kumari S, Carmona AV, Tiwari AK, Trippier PC (2020) Amide bond bioisosteres: strategies, synthesis, and successes. *J Med Chem* 63:12290–12358
- Ratni H, Baumann K, Bellotti P, Cook XA, Green LG, Luebberts T, Reutlinger M, Stephan AF, Vifian W (2021) Phenyl bioisosteres in medicinal chemistry: discovery of novel  $\gamma$ -secretase modulators as a potential treatment for Alzheimer's disease. *RSC Med Chem* 12:758–766
- Jayashree BS, Nikhil PS, Paul S (2022) Bioisosterism in drug discovery and development-an overview. *Med Chem* 18:915–925
- Allen FH, Groom CR, Liebeschuetz JW, Bardwell DA, Olsson TS, Wood PA (2012) The hydrogen bond environments of 1H-tetrazole and tetrazolate rings: the structural basis for tetrazole-carboxylic acid bioisosterism. *J Chem Inf Model* 52:857–866
- Langdon SR, Ertl P, Brown N (2010) Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol Inform* 29:366–385
- Oebbecke M, Siefker C, Wagner B, Heine A, Klebe G (2021) Fragment binding to kinase hinge: if charge distribution and local pK(a) shifts mislead popular bioisosterism concepts. *Angew Chem Int Ed* 60:252–258
- Wirth M, Zoete V, Michielin O, Sauer WH (2013) SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res* 41:D1137–D1143
- Desaphy J, Rognan D (2014) sc-PDB-Frag: a database of protein-ligand interaction patterns for Bioisosteric replacements. *J Chem Inf Model* 54:1908–1918
- Wood DJ, de Vlieg J, Wagener M, Ritschel T (2012) Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. *J Chem Inf Model* 52:2031–2043
- Seddon MP, Cosgrove DA, Gillet VJ (2018) Bioisosteric replacements extracted from high-quality structures in the protein databank. *ChemMedChem* 13:607–613
- Ertl P (2003) Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J Chem Inf Comput Sci* 43:374–380
- Shan J, Ji C (2020) MolOpt: a web server for drug design using bioisosteric transformation. *Curr Comput Aided Drug Des* 16:460–466
- Shan J, Pan X, Wang X, Xiao X, Ji C (2020) FragRep: a web server for structure-based drug design by fragment replacement. *J Chem Inf Model* 60:5900–5906
- Cross S, Cruciani G (2022) FragExplorer: GRID-based fragment growing and replacement. *J Chem Inf Model* 62:1224–1235
- Ertl P, Craig plot 2.0: an interactive navigation in the substituent bioisosteric space. *J. Cheminformatics* 2020, 12, 8.
- Holliday JD, Jelfs SP, Willett P, Gedeck P (2003) Calculation of Intersubstituent similarity using R-group descriptors. *J Chem Inf Comput Sci* 43:406–411
- Weber A, Teckentrup A, Briem H (2002) Flexsim-R: a virtual affinity fingerprint descriptor to calculate similarities of functional groups. *J Comput Aided Mol Des* 16:903–916
- Elias TC, de Oliveira HCB, da Silveira NJF (2018) MB-Isoster: a software for bioisosterism simulation. *J Comput Chem* 39:2481–2487
- Watson P, Willett P, Gillet VJ, Verdonk ML (2001) Calculating the knowledge-based similarity of functional groups using crystallographic data. *J Comput Aided Mol Des* 15:835–857
- Ertl P (2020) Identification of bioisosteric substituents by a deep neural network. *J Chem Inf Model* 60:3369–3375
- Zhang Y, Borrel A, Ghemto L, Regad L, Boije af Gennäs G, Camproux A-C, Yli-Kauhaluoma J, Xhaard H (2017) Structural isosteres of phosphate groups in the protein data bank. *J Chem Inf Model* 57:499–516
- Zhang T, Jiang S, Li T, Liu Y, Zhang Y (2023) Identified isosteric replacements of ligands' glycosyl domain by data mining. *ACS Omega* 8:25165–25184
- Boratyn GM, Camacho C, Cooper PS, Coulouris G, Fong A, Ma N, Madden TL, Matten WT, McGinnis SD, Merezhuk Y, Raytselis Y, Sayers EW, Tao T, Ye J, Zaretskaya I (2013) BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* 41:W29–W33
- Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309
- Zhang C, Shine M, Pyle AM, Zhang Y (2022) US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat Methods* 19:1109–1115
- Vainio MJ, Puranen JS, Johnson MS (2009) ShaEP: molecular overlay based on shape and electrostatic potential. *J Chem Inf Model* 49:492–502
- O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. *J Cheminformatics* 3:33
- Landrum, G. A. In *RDKit: Open-source cheminformatics. Release 2014.03.1, 2014*; 2014.
- Lerner C, Jakob-Roetne R, Buettelmann B, Ehler A, Rudolph M, Rodríguez Sarmiento RM (2016) Design of potent and druglike nonphenolic inhibitors for catechol O-methyltransferase derived from a fragment screening approach targeting the S-adenosyl-L-methionine pocket. *J Med Chem* 59:10163–10175
- Schweigert N, Zehnder AJ, Eggen RI (2001) Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. *Environ Microbiol* 3:81–91
- do Lim Y, Shin SH, Lee MH, Malakhova M, Kurinov I, Wu Q, Xu J, Jiang Y, Dong Z, Liu K, Lee KY, Bae KB, Choi BY, Deng Y, Bode A, Dong Z (2016) A natural small molecule, catechol, induces c-Myc degradation by directly targeting ERK2 in lung cancer. *Oncotarget* 7:35001–35014
- Gelin M, Delfosse V, Allemand F, Hoh F, Sallaz-Damaz Y, Pirocchi M, Bourguet W, Ferrer JL, Labesse G, Guichou JF (2015) Combining "dry" co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. *Acta Crystallogr D* 71:1777–1787
- Myers SM, Miller DC, Molyneux L, Arasta M, Bawn RH, Blackburn TJ, Cook SJ, Edwards N, Endicott JA, Golding BT, Griffin RJ, Hammonds T, Hardcastle IR, Harnor SJ, Heptinstall AB, Lochhead PA, Martin MP, Martin NC, Newell DR, Owen PJ, Pang LC, Reuillon T, Rigoreau LJM, Thomas HD, Tucker JA, Wang L-Z, Wong A-C, Noble MEM, Wedge SR, Cano C (2019) Identification of a novel orally bioavailable ERK5 inhibitor with selectivity over p38 $\alpha$  and BRD4. *Eur J Med Chem* 178:530–543



44. Heightman TD, Berdini V, Braithwaite H, Buck IM, Cassidy M, Castro J, Courtin A, Day JEH, East C, Fazal L, Graham B, Griffiths-Jones CM, Lyons JF, Martins V, Muench S, Munck JM, Norton D, O'Reilly M, Palmer N, Pathuri P, Reader M, Rees DC, Rich SJ, Richardson C, Saini H, Thompson NT, Wallis NG, Walton H, Wilsher NE, Woolford AJA, Cooke M, Cousin D, Onions S, Shannon J, Watts J, Murray CW (2018) Fragment-based discovery of a potent, orally bioavailable inhibitor that modulates the phosphorylation and catalytic activity of ERK1/2. *J Med Chem* 61:4978–4992
45. Roy SM, Minasov G, Arancio O, Chico LW, Van Eldik LJ, Anderson WF, Pelletier JC, Watterson DM (2019) A selective and brain penetrant p38 $\alpha$ MAPK inhibitor candidate for neurologic and neuropsychiatric disorders that attenuates neuroinflammation and cognitive dysfunction. *J Med Chem* 62:5298–5311
46. Braun F, Bertolotti N, Möller G, Adamski J, Steinmetzer T, Salah M, Abdelsamie AS, van Koppen CJ, Heine A, Klebe G, Marchais-Oberwinkler S (2016) First structure-activity relationship of 17 $\beta$ -hydroxysteroid dehydrogenase type 14 nonsteroidal inhibitors and crystal structures in complex with the enzyme. *J Med Chem* 59:10719–10737
47. Braun F, Bertolotti N, Möller G, Adamski J, Frotscher M, Guragossian N, Madeira Girio PA, Le Borgne M, Ettouati L, Falsone P, Müller S, Vollmer G, Heine A, Klebe G, Marchais-Oberwinkler S (2018) Structure-based design and profiling of novel 17 $\beta$ -HSD14 inhibitors. *Eur J Med Chem* 155:61–76
48. Zhang Y, Jumppanen M, Maksimainen MM, Auno S, Awol Z, Ghemtio L, Venkannagari H, Lehtiö L, Yli-Kauhaluoma J, Xhaard H, Boije Af Gennäs G (2018) Adenosine analogs bearing phosphate isosteres as human MDO1 ligands. *Bioorg Med Chem* 26:1588–1597
49. Ruan B, Zhang Y, Tadesse S, Preston S, Taki AC, Jabbar A, Hofmann A, Jiao Y, Garcia-Bustos J, Harjani J, Le TG, Varghese S, Teguh S, Xie Y, Odiba J, Hu M, Gasser RB, Baell J (2020) Synthesis and structure-activity relationship study of pyrrolidine-oxadiazoles as anthelmintics against *Haemonchus contortus*. *Eur J Med Chem* 190:112100

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.