

# SwissBioisostere 2021: updated structural, bioactivity and physicochemical data delivered by a reshaped web interface

Alessandro Cuzzo<sup>1,3</sup>, Antoine Daina<sup>1</sup>, Marta A.S. Perez<sup>1</sup>, Olivier Michielin<sup>1,2,\*</sup> and Vincent Zoete<sup>1,3,\*</sup>

<sup>1</sup>Molecular Modeling Group, SIB Swiss Institute of Bioinformatics, University of Lausanne, Quartier UNIL-Sorge, Bâtiment Amphipole, CH-1015 Lausanne, Switzerland, <sup>2</sup>Department of Oncology, Precision Oncology Center, University Hospital of Lausanne, CH-1011 Lausanne, Switzerland and <sup>3</sup>Department of Oncology UNIL-CHUV, University of Lausanne, Ludwig Institute for Cancer Research, Route de la Corniche 9A, CH-1066 Epalinges, Switzerland

Received September 15, 2021; Revised October 08, 2021; Editorial Decision October 15, 2021; Accepted November 04, 2021

## ABSTRACT

At several stages of drug discovery, bioisosteric replacement is a common and efficient practice to find new bioactive chemotypes or to optimize series of molecules toward drug candidates. The critical steps consisting in selecting which molecular moiety should be replaced by which other chemical fragment is often relying on the expertise of specialists. Nowadays, valuable support can be obtained through the wealth of dedicated structural and knowledge data. The present article details the update of SwissBioisostere, a database of >25 millions of unique molecular replacements with data on bioactivity, physicochemistry, chemical and biological contexts extracted from the literature and related resources. The content of the database together with analysis and visualization capacities is freely available at [www.swissbioisostere.ch](http://www.swissbioisostere.ch).

## INTRODUCTION

The rationale for isosteric modifications finds its root back to the work by Langmuir in 1919 on the similarities of properties between chemical entities like atoms, groups or molecules (1). Much latter was added the idea of similar biological properties, instigating the concept of bioisosterism (2). Bioisosteric replacement, i.e. exchanging part of a bioactive molecule by a fragment to generate a new one, was driven by the *similarity principle* stating that similar molecules are prone to have similar activities (3) and demonstrated later (4). Consequently, *classical bioisosterism* focuses on keeping stable physicochemical properties,

like size, polarity or lipophilicity (5). Modern drug discovery has made the concept evolve to the more qualitative and pragmatic notion of *non-classical bioisosterism*. A successful bioisosteric replacement may be defined as exchanging a fragment from the molecule by another fragment that results in a novel compound with similar or more potent bioactivity and at the same time in improved properties important for a drug (6). As drug discovery is highly multi-objective, the flaws to be corrected can be very diverse, spanning from toxicity, improper pharmacokinetics or lack of specificity for the target, to synthesis or intellectual property issues, just to cite some obvious ones (7).

Bioisosteric replacement is a routine practice for *lead optimization*. At this step, medicinal chemists apply numerous small modifications at the periphery of validated chemotypes to adjust properties and finally generate drug candidates (8). Here, the replacement concerns mainly ‘side chains’, i.e. fragments with a single attachment point to the rest of the molecule.

*Hit finding* is the discovery of novel bioactive chemotypes typically achieved by high-throughput screening or through literature and patent analysis. Here the strategy is to exchange the central core with a potential bioisosteric moiety yet keeping the pharmacophore, with the hope of retaining biological activity (9). Following this ‘scaffold-hopping’ approach (10), replacements concern ‘scaffolds’ or ‘linkers’ fragments with three and two attachment points to the rest of the molecule, respectively.

Bioisosteric replacements reflect the intuitive strategy relying on the experience of the chemists. Chances of success increase when applied in a systematic and rational fashion (11,12). Efficient computer support is based on (i) the quality, comprehensiveness and relevance of the proposed

\*To whom correspondence should be addressed. Tel: +41 21 692 59 07; Email: [vincent.zoete@unil.ch](mailto:vincent.zoete@unil.ch)

Correspondence may also be addressed to Olivier Michielin. Tel: +41 21 692 40 53; Email: [olivier.michielin@sib.swiss](mailto:olivier.michielin@sib.swiss)

**DEDICATION:** This article is dedicated to the memory of our colleague Wolfgang H. B. Sauer, a brilliant mind, and a wonderful person.

replacements, and (ii) an informed choice among the proposals thanks to analysis capacities.

A pioneer effort of analyzing literature resulted in a closed database of 27 000 structural replacements called Bioster and accessible through commercial products, such as StartDrop ([www.optibrium.com/project/bioster/](http://www.optibrium.com/project/bioster/)).

BoBER ([www.bober.insilab.org](http://www.bober.insilab.org)) gives access to replacements resulting from mining part of the Protein Data Bank by superimposing binding sites showing 3D similarity in residues arrangement and then extracting ligand information from complexes and classical bioisosteric rules (13). Although the interface includes interesting fragmentation features, user misses a description of the database content.

Another small structural collection relies on high-quality tridimensional data from the Protein Data Bank (14). The analysis focused on 1458 molecules to point out, among 55 000 replacements, those considered ‘bioisosteres’ because they occupy similar volume within a binding site of 121 protein targets.

scPDBFrag ([www.bioinfo-pharma.u-strasbg.fr/scPDBFrag/](http://www.bioinfo-pharma.u-strasbg.fr/scPDBFrag/)) is a database including 12 000 fragments from 8077 ligand–protein complexes that aims at supporting the selection of the most probable bioisosteric moieties (15). Replacements are scored according to how likely they bind in a similar fashion by analyzing conservation in interaction pattern graphs after alignment.

Our structural and knowledge database SwissBioisostere has been made freely browsable online since 2012 (16). It supports mainly but not exclusively *hit-finding* and *lead optimization* through a user-friendly Web interface to avoid methodological pitfalls for non-experts and tedious technical efforts for specialist users. The candidate fragments for replacement are easily yet thoroughly evaluated according to diverse criteria acquired from highly curated bioactivity and physicochemical data. Information from the chemical and biological contexts is provided to user from the root of the replacement—that is the pairs of compounds baring the replaced and replacing groups with their difference in potency on classified protein targets—and in some key properties. External links to ChEMBL compounds and assays ([www.ebi.ac.uk/chembl/](http://www.ebi.ac.uk/chembl/)), along with PubMed links to publications of origin, are provided ([pubchem.ncbi.nlm.nih.gov](http://pubchem.ncbi.nlm.nih.gov)). All these capacities are meant for the end users to wisely select the most probable and suitable bioisosteres fostering drug discovery.

The SwissBioisostere website has been increasingly visited and the database increasingly requested. In 2020, 8800 unique visitors opened 13 500 sessions to submit 17 152 queries. Compared to 2019, this represents an increase of 33%, 32% and 6%, respectively. As a result, at the time of writing the manuscript, we recorded total numbers reaching 45 900 unique visitors having opened 70 200 sessions to submit 103 000 queries since 2012. Geographically, users are from 163 countries (top 10 is USA, India, Japan, China, UK, Germany, Brazil, Switzerland, France and South Korea).

These numbers have fostered the major update detailed in the present article and released on-line in September 2021. The novelties primarily regard the dataset to build the structural collection. It relies on the bioactivity data of ChEMBL28 (17) and thus describes more molecules, more

protein targets and more assays than the former version (based on ChEMBL17).

Finally, although keeping a look-and-feel comparable to the trusted previous version, the new Web interface includes numerous improvements. The reader is encouraged to employ the 2021 version of SwissBioisostere ([www.swissbioisostere.ch](http://www.swissbioisostere.ch)). However, the previous version will remain accessible for at least one year ([old.swissbioisostere.ch](http://old.swissbioisostere.ch)) to ensure the continuity of user’s ongoing projects.

## UPDATES

The overall process behind SwissBioisostere (16) consists in (i) retrieving and filtering high quality bioactivity data from ChEMBL (17), (ii) applying the Molecular Matched Pair algorithm (MMP) to pairs of molecules tested in the same assay to find molecular replacements and (iii) store those replacements and associated data in a SQL database. SwissBioisostere is available under the CC-BY 4.0 Creative Commons 4.0 International License.

## Technologies

The MMP engine of SwissBioisostere was mainly coded in Python 2.7. Important improvements were implemented for chemical handling, which is now performed by JChem Microservices ([www.chemaxon.com](http://www.chemaxon.com), version 21.3), OpenBabel ([www.openbabel.org](http://www.openbabel.org), version 2.3.2) (18) and Pybel (19), included in the latter. Murcko Scaffolds were generated with the Python RDKit API package ([www.rdkit.org](http://www.rdkit.org), version 2016.03.5). We relied on version 2.7 of Python for compatibility reasons with some functionalities with OpenBabel and Pybel.

The website was written using HTML5, PHP (version 7.2.24), JavaScript and MySQL (version 5.7.35). Most significant novelties are

- the full integration of MarvinJS Web services to MarvinJS sketcher ([www.chemaxon.com](http://www.chemaxon.com), version 21.2.0) for providing multiple ways of inputting molecules and to handle SMILES inputs,
- the use of the DataTables plug-in (version 1.10.16) for dynamic output tables and advanced export options (CSV, Excel and PDF files, clipboard and printing),
- dynamic pie charts of R groups chemical environment for all cases of replacements (which were absent for scaffolds in the previous version of SwissBioisostere) generated with flot (version 0.7, [www.flotcharts.org](http://www.flotcharts.org)),
- riddance of multiple dependencies (Java, Indigo, CDK, Tomcat),
- SMILES synchronization in both directions between text-boxes and sketchers on the input page,
- a dynamic and more user-friendly input page and FAQ,
- the full interoperability with other SwissDrugDesign tools (i.e. SwissTargetPrediction, SwissADME and SwissSimilarity),
- a static help page,
- video tutorials explaining how to use the website and
- an option to receive by e-mail a link with the user job ID to easily access and share the results.

## DATABASE CONSTRUCTION AND FEATURES

### Experimental data retrieval and preparation

The ChEMBL database was used as primary data source for experimental assay records. The version 28 (2021) was downloaded ([www.chembl.gitbook.io/chembl-interface-documentation/downloads](http://www.ebi.ac.uk/chembl-interface-documentation/downloads)) and stored in a local MySQL database. Strict filtering criteria led to extract high-quality interaction data for a considerable amount of target-driven active compounds.

The input data for the MMP was obtained by extracting ChEMBL compounds, with molecular weight <800 g/mol, from binding or functional assays reported with an activity of  $0 < \text{standard value} < 100\,000$  nM on a molecular target. Standard activity types were either  $IC_{50}$ ,  $EC_{50}$ ,  $K_i$  or  $K_d$  and standard relations were strict (=) with a confidence score > 7. Custom target classes were made by concatenating class levels 1 and 2 from the target protein family tree classification.

The new threshold of 800 g/mol is stricter than on previous version of SwissBioisostere to avoid overly large compounds that would result in excessive number of cuts, computational time and storage. Despite this limitation, the number of entries in SwissBioisostere, both replacements and fragments, dramatically increased (see section *Swiss-Bioisostere database content*).

Each compound was standardized keeping only the main fragment, itself dearomatized, dehydrogenized and neutralized using the standardizer engine included in JChem Microservices ([www.chemaxon.com](http://www.chemaxon.com), version 21.3).

### Identification of matched molecular pairs

Matched molecular pairs (MMPs) were identified between molecules tested in the same assay by implementing the Hussain and Rea algorithm (20) in Python using the OpenBabel API ([www.openbabel.org](http://www.openbabel.org)). The algorithm fragments molecular structures by cutting up to three bonds simultaneously. Cutable bonds were defined as single bonds not part of a cycle or aromatic system, including at least one carbon atom but excluding hydrogen, and outside any chemical function. Furthermore, additional rules were implemented by us to avoid irrelevant cuts and reduce time storage. Since sugars and aliphatic tails are most frequently considered as a whole in medicinal chemistry (i.e. as indivisible moieties), one rule was not to cut within cyclic hexoses defined using SMARTS patterns ([www.daylight.com/dayhtml/doc/theory/theory.smarts.html](http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html)) for glucose and fructose. Another rule was added to prevent cuts inside an unbranched aliphatic carbon chain (SMARTS:  $[CH2 \times 4 \&!R, CH1 \times 3 \&!R, CX2 \&!R] \sim [CH2 \times 4 \&!R, CH1 \times 3 \&!R, CX2 \&!R]$ ).

The stereochemistry was treated differently than in the previous version of SwissBioisostere. In the new MMP implementation, the asymmetry of carbon atoms in fragments was neglected but considered within the whole compounds in which they were found.

When the fragmentation of two different compounds found in the same ChEMBL assay leads to one fragment being different while all others (up to three) are identical, a replacement has been found. Replacements are encoded as

**Table 1.** Comparative volume of bioactivity data between versions 2012 and 2021 of SwissBioisostere (extracted from ChEMBL17 and ChEMBL28, respectively).

Number of	SwissBioisostere 2012 (based on ChEMBL17 data)	SwissBioisostere 2021 (based on ChEMBL28 data)
Compounds	321 463	483 927
Assays	39 393	61 199
Targets	1552	2036
Target classes <sup>a</sup>	32	35
Unique replacements	5 586 462	25 305 017
Unique exchanged fragments	510 063	1 216 118

<sup>a</sup>Concatenation of levels 1 and 2 from ChEMBL protein family classification.

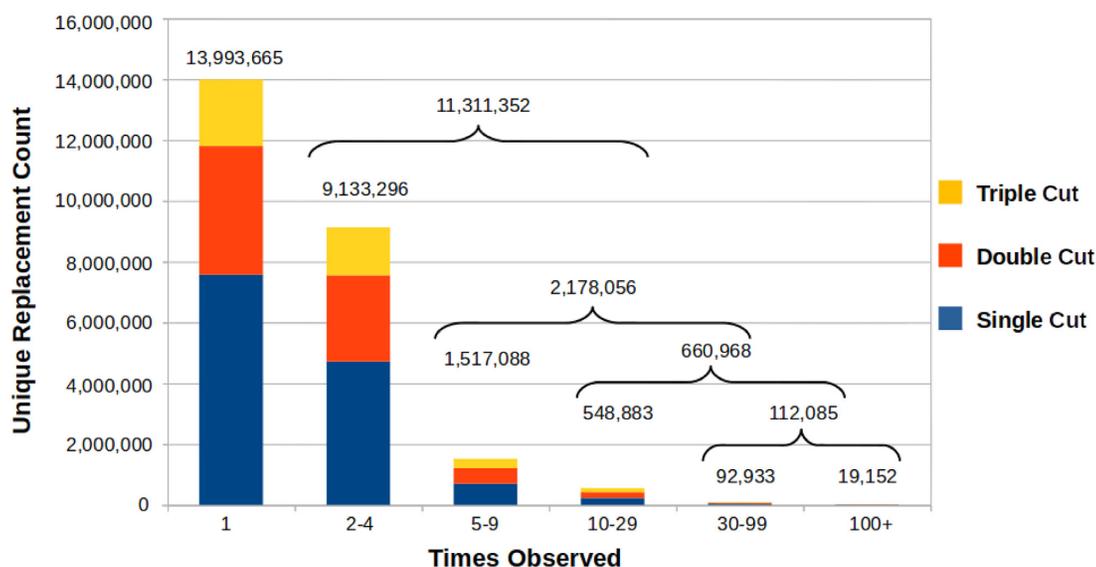
canonicalized SMIRKS ([www.daylight.com/dayhtml/doc/theory/theory.smirks.html](http://www.daylight.com/dayhtml/doc/theory/theory.smirks.html)). The size thresholds of fragments were increased, compared to the 12 heavy atoms applicable to all types of fragments in the previous SwissBioisostere version. This enables replacements of larger molecular parts. In the current version, for compounds having differing fragments with double or triple cuts (defined as *linkers* and *scaffolds*, respectively), a maximum threshold of 15 heavy atoms per fragment was set. For differing fragments obtained by simple cut (defined as *side chains*), the fragment has to be smaller than the common core of the two compounds.

Such distinctions between *side chains*, *linkers* and *scaffolds* were established since their diverse main applications in the drug discovery context. Side chains replacements are mostly performed during *lead optimization*, while replacing linkers or scaffolds mainly occurs as *scaffold-hopping* in the *hit-finding* phase. Thus, SwissBioisostere MMP deals with useful cases for distinct purposes and objectives along the drug discovery process, all in one go (refer to *Use cases* section for detailed examples).

### SwissBioisostere database content

As shown in Table 1, the wealth of available bioactivity information regarding compounds has greatly increased compared to the 2012 version of SwissBioisostere. Overall, the dataset includes 25 305 017 unique replacements, which constitutes an increase of 453% compared to the previous version. 13 280 691 of these replacements involve side chains exchange, 7 804 296 linkers exchange and 4 220 030 scaffolds exchange. These replacements occur between 1 216 118 unique fragments, corresponding to an increase of 238% compared to the previous version of SwissBioisostere. 356 459 of these fragments correspond to side chains, 448 859 to linkers and 410 800 to scaffolds. In total, 65 098 550 data points—defined as one given replacement between two fragments within a pair of compounds tested in a well identified assay—can be accessed via the graphical interface.

More than a half of all unique replacements (13 993 665; 55%) are singletons, i.e. have been observed only once (Figure 1). About 44.7% (11 311 352) replacements have been observed at least twice, 8.6% (2 178 056) at least five times, 2.6% (660 968) found at least ten times, 0.5% (112 085) at



**Figure 1.** Unique replacements stratified by the type of cut that has been performed. Single Cut for *side chains*, Double Cut for *linkers* and Triple Cut for *scaffolds*.

least thirty times and finally, and 0.1% (19 152) were found at least a hundred times.

The content of SwissBioisostere database comes either directly from data stored in ChEMBL database (compounds, assays, assay types, confidence, PubMed IDs, log *P*, tPSA, bioactivity, molecular weight, targets and target classes), or is further computed during the building of the database (fragments, replacements, molecules constant parts, murcko scaffolds, the attachment point contexts,  $\Delta$  log *P*,  $\Delta$  tPSA,  $\Delta$  bioactivity and  $\Delta$  molecular weight). In case a replacement was found more than once, the mean of the differences over all observed MMPs is reported.

### Graphical interface (user experience)

The new web interface of SwissBioisostere keeps a look-and-feel similar to the previous version, yet with substantial additions and significant improvements made to enrich the user experience. Importantly, as several processes are performed through different web services, client-side internet connection and browser must allow binding to Web services.

### Input

User journey begins at the input page (Figure 2) freely accessible without log in nor registration at URL [swissbioisostere.ch](http://swissbioisostere.ch). In the Home main page (Figure 2), two tabs enable the user to choose between either querying information regarding the possible molecular replacements for a fragment (default), or regarding a specific replacement (by clicking on the right tab).

If the first query option is chosen, the fragment can be drawn in the left-hand MarvinJS molecular sketcher or inputted in SMILES format in the left-hand dedicated text box. The SMILES (or a molecule name, if any) can also be directly pasted inside the canvas (main window) of the sketcher. Importantly, the new implementation of MarvinJS allows the import from a molecule file (locally or

through the network) or by name (chemical or common, if any). Note that the sketcher has now copy/paste capabilities and is synchronized with the SMILES text box. Input fragments must be linked to exactly one, two or three R-groups—representing attachment points to the constant parts of the molecule—for side chains, linkers and scaffolds, respectively (Figure 2).

If the second query option is chosen, an additional sketcher with text-box appear on the right-hand side, so the user can input the fragment to replace in sketcher 1 (left) and the substitute fragment in sketcher 2 (right).

Input examples are available for side chain, linker and scaffold for both query options.

Optionally, user provides an electronic address in the dedicated box, to receive a link to the results by e-mail, including a unique job ID.

Once all R-groups are set, a request can be submitted by clicking on the ‘Query Database’ button.

### Output

**Fragment.** A typical query takes from a couple of seconds to a minute, depending on the number of possible replacements to process. When the query is completed, the results are displayed in a first output page (Figure 3).

A graph produced using the Flot library (version 0.7, [www.flotcharts.org](http://www.flotcharts.org)) on the top-right describes the physicochemical space of possible replacements as  $\Delta$  tPSA versus  $\Delta$  log *P*. Importantly, an area of interest within this plot can be selected to dynamically filter the result table lower in the page, and described below. A blue ‘info’ icon below the query structure (middle-left) gives informative details for proper interpretation.

The principal information presented in this result page consists in the tabulated list of candidate fragments that were found in at least one replacement occurrence (pair of compounds in the same assay). Rows are listed according to (i) the number of occurrences for the replacement and

The screenshot shows the SwissBioisostere web interface. At the top, there is a navigation bar with icons for various tools: SwissDeck, SwissParam, SwissSidechain, SwissBioisostere (highlighted), SwissTargetPrediction, SwissADME, and SwissSimilarity. Below this is the header with the SwissBioisostere logo and navigation links (Home, About, FAQ, Help, Tutorials, Citing, Contact). The main content area has two tabs: 'I want to search for possible replacements of a fragment' (selected) and 'I want to get information on a given molecular replacement'. Below the tabs are two sketcher windows, 'Fragment 1' and 'Fragment 2'. Fragment 1 shows a chemical structure with R<sub>1</sub> and R<sub>2</sub> groups. Fragment 2 shows a chemical structure with R<sub>1</sub> and R<sub>2</sub> groups. Below the sketchers are input fields for SMILES, an optional email address field, and a 'Query Database' button. At the bottom, there is a note: 'Fill with an example: side chain, linker scaffold'.

Callouts in the image point to the following features:

- Links to online CADD tools developed by the SIB Molecular Modelling Group
- Clear and hide the right sketcher for a fragment query
- Show the right sketcher for a replacement query
- Import query fragment by name or from file
- Sketcher for fragment 1
- "Smart R-group" button
- Sketcher for fragment 2
- Input query fragment as SMILES
- Clear text box and sketcher
- Input E-mail address (optional)
- "Query Database" button
- Apply examples for side chain, linker or scaffold fragment/replacement

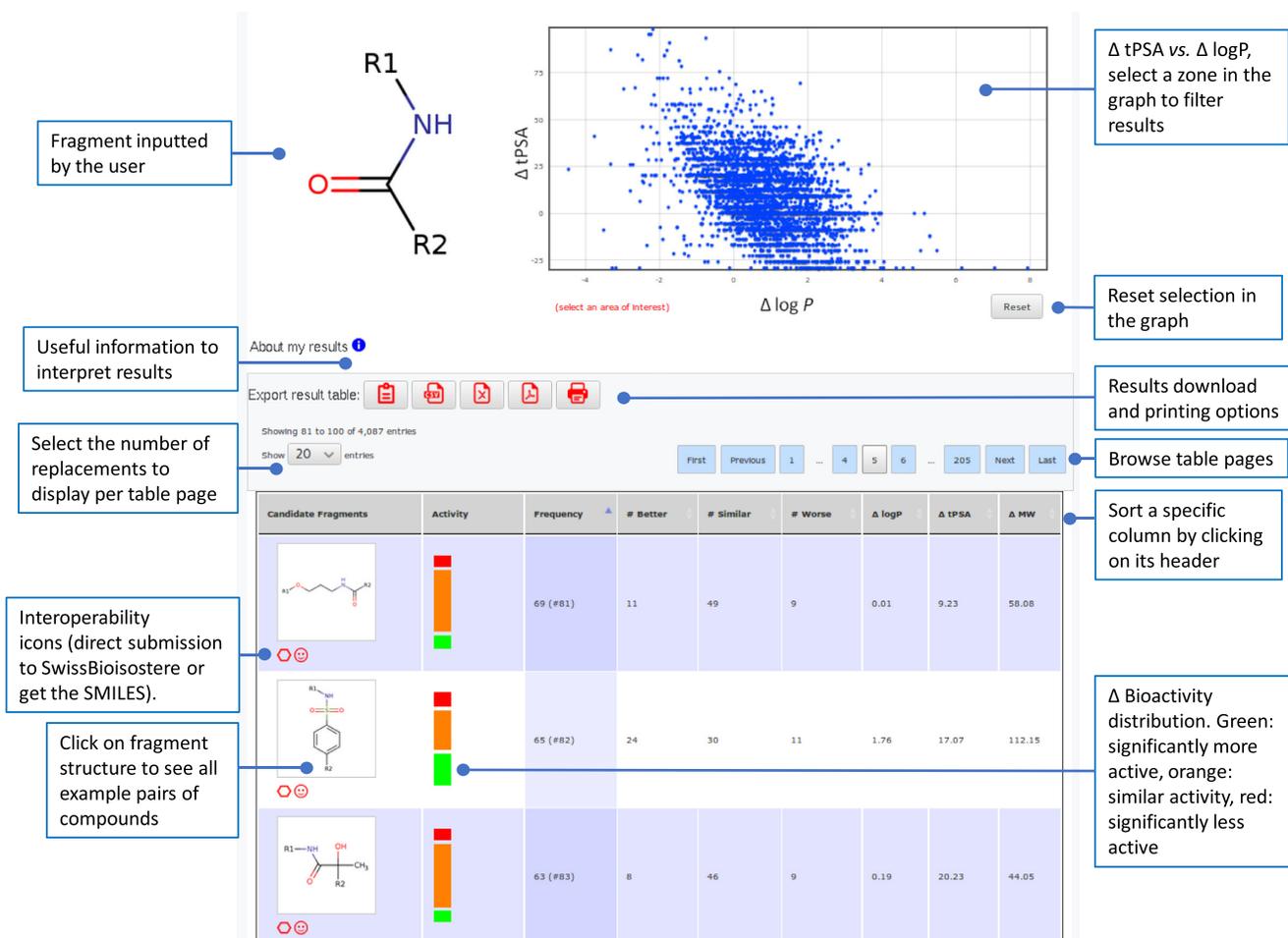
**Figure 2.** SwissBioisostere input page. The upper icons banner provides links to other SwissDrugDesign CADD online tools. FAQ, help, tutorials, contact and other pages are in the header menu. The query fragment(s) can be inputted either as SMILES or through MarvinJS sketcher(s) for two query options: (i) possible replacements of a fragment and (ii) information on a given replacement (two fragments). 'Smart R-group' (MarvinJS left toolbars) enables automated numbering (R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>) of mandatory attachment points. A 'Clear' button for each sketcher allows resetting all inputs. Providing e-mail address (optional) allows the user to receive links with unique job ID, to replicate the same query and obtain identical results. For both query options, examples can be loaded below the 'Query Database' button, either for side chain, linker or scaffold.

(ii) the number of occurrences showing a significant bioactivity increase ( $> 0.5$  log). Of note, this default listing order was chosen to start with replacements for which the largest amount of experimental information is available. Therefore, this should not be interpreted as a score of bioisosterism or relevance. Indeed, for instance, the number of occurrences for a replacement is biased by its age and is not necessarily an expression of its usefulness as bioisostere. On each row, a three-color bar sums up the impact of the replacement on bioactivity for the example pairs of compounds ( $> 0.5$  log: 'green' significant increase;  $< -0.5$  log: 'red' significant decrease; or 'orange' similar bioactivity). Two so-called interoperability icons appear below the chemical structure of every fragment in the table. The 'hexagon' submits the fragment back to SwissBioisostere. The 'happy face' generates the SMILES. The table can be sorted according to any relevant tabulated property by clicking on the column header. Moreover, advanced export options are provided to the users with dedicated red icons. The entire or a filtered table can be downloaded in various formats (PDF, CSV or Excel), copied into the clipboard or directly printed.

Clicking on a fragment chemical structure launches another request and leads to another output page with detailed information about the specific replacement (Figure 4).

**Replacement.** Upon querying information regarding a precise replacement from either the input page (option with two sketchers) or the fragment output page (click on a candidate fragment structure), results are displayed in the final output page (Figure 4). The latter lists data for all occurrences, which have been retrieved from medicinal chemistry literature. This consists in pairs of compounds tested on the same assay. By default, the table rows are ranked according to the difference in bioactivity between both compounds. Compounds and assays ChEMBL IDs are mentioned and hyperlinked to their corresponding pages on the ChEMBL website. A link to the publication entry in PubMed is also provided when available.

General statistics for a given replacement are provided together with a graph showing the distribution of activity differences for all occurrences. An additional scrollable area with multiple pie charts shows the activity difference by attachment point chemical context (aromatic ring, aliphatic ring or aliphatic linker). Activity measurement filters can be applied from a box on the middle-left. Those are activity level (from 10  $\mu$ M to 1 pM), assay type (binding or functional), confidence (high = 8, highest = 9, according to ChEMBL confidence score) and result type (IC<sub>50</sub>, EC<sub>50</sub>, K<sub>i</sub> or K<sub>d</sub>). The table and graphs are dynamically updated by these criteria. The whole or a filtered



**Figure 3.** Fragment query output page (all possible replacements for a given fragment). Upper-left: recall of the query fragment. Upper-right: plot of the  $\Delta$  tPSA versus  $\Delta$  log  $P$  for each candidate fragment. Region of interest on this graph can be selected to filter the result and dynamically adapt the table below. Middle-left: information about how to interpret the results (blue info icon) and other icons allowing downloading the results as PDF, CSV or Excel files, as well as printing or copying them to the clipboard. Bottom: main table listing the candidate fragments that were found in at least one example pair of molecules. By default, 20 fragments are sorted according to Frequency. User can display more rows and sort for other columns (by clicking on header). The three-color bar sums up the global impact on bioactivity measurements on all pairs of molecules ( $>0.5$  log, significant increase, 'green';  $<-0.5$  log, significant decrease, 'red'; or similar bioactivity, 'orange'). Clicking on a chemical structure launches a query for the replacement of the input fragment by the candidate fragment, leading to the output page of this specific replacement (see Figure 4). One major novelty is the interoperability icons for straightforward submission of any fragment back to SwissBioisostere Home page ('hexagon') or generate the SMILES ('happy face').

table can be downloaded in various formats (PDF, CSV or Excel) or copied into the clipboard as well as directly printed.

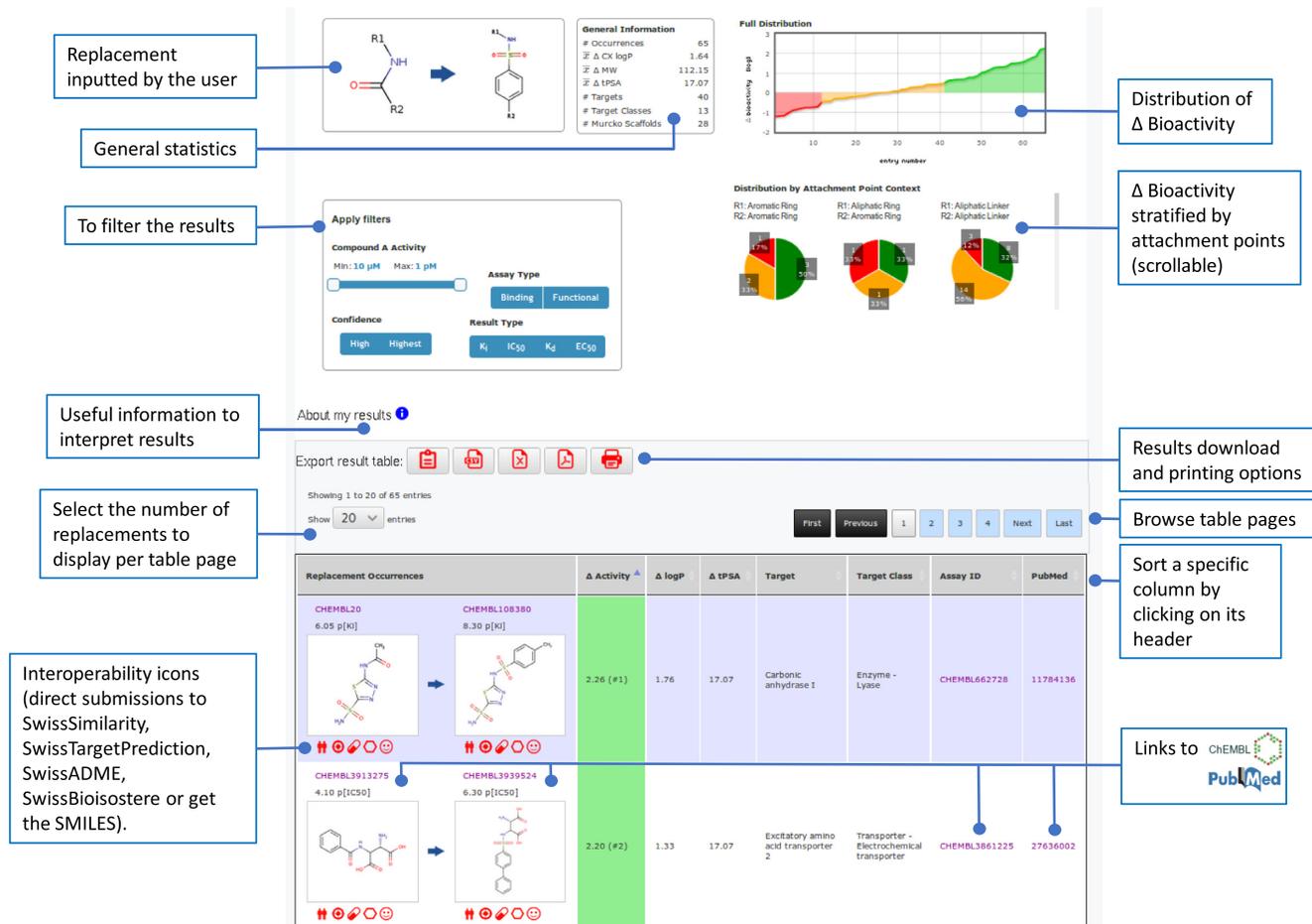
Five so-called interoperability icons appear below the chemical structure of every compound in the table and provide the user with the possibility to straightforwardly submit the corresponding molecule to different CADD Web tools, developed by the Molecular Modeling Group of the SIB Swiss Institute of Bioinformatics. At the time of writing the present manuscript, four CADD tools are interoperable both ways: SwissSimilarity (21), SwissTarget-Prediction (22), SwissADME (23) and SwissBioisostere itself. The planned extension of interoperability by including other SwissDrugDesign tools ([www.molecular-modelling.ch/swiss-drug-design.html](http://www.molecular-modelling.ch/swiss-drug-design.html)) is an on-going work and therefore additional icons will be added in these different websites in the future.

### Other pages

Support for using the website and analysing the results is provided by renewed FAQ, Help and Tutorials pages accessible through the header menu. The latter contains videos explaining in detail how to use SwissBioisostere Web interface. Should users need more assistance, a contact form is provided.

### Use cases

*Hit finding example.* The discovery of Usmarapride, a 5-HT<sub>4</sub> partial agonist useful in Alzheimer's disease, was recently described (24). Three scaffold hopping steps were crucial to generate chemotypes with different pharmacodynamics, along with improved pharmacokinetics and toxicity profiles. The first replacement was from a imidazo[1,5-a]pyridine (fragment 1) to an indazole (fragment 2) core



**Figure 4.** Replacement query output page. Important information includes: upper-middle: general statistics of all occurrences (pairs of compounds bearing the query replacement); Upper-right: distribution plot of the impact on bioactivity differences; Middle-right: scrollable area with pie charts showing activity differences by attachment point context; Middle-left: activity-related filters for the table and graphs; optionally displayed information about result interpretation (blue 'info' icon) and other (red) icons for downloading the results as PDF, CSV or Excel files, as well as printing or copying to the clipboard; Bottom: main table with the retrieved occurrences from medicinal chemistry literature. For each replacement are listed target, target class, ChEMBL IDs of both compounds and assay hyperlinked to their corresponding pages in ChEMBL. A link to the publication in PubMed is also provided when applicable. By default, the number of displayed rows is 20, but the user can choose 50, 100, 250 or all. By default, the occurrences are sorted by  $\Delta$  Activity, but the user can sort by other properties by clicking on the respective column header. One major novelty is the interoperability icons for straightforward submission of any result molecule to another SwissDrugDesign tool: SwissSimilarity ('twins'), SwissTargetPrediction ('target'), SwissADME ('pill'), SwissBioisostere ('hexagon') or get the SMILES ('happy face').

and this replacement is suggested by SwissBioisostere. By inputting the corresponding fragment 1 in SwissBioisostere (see Supplementary Figure S1A), 77 possible replacements can be retrieved, among which fragment 2 is in 20th row with two examples of increased bioactivity (green bar, see Supplementary Figure S1B). By clicking on the chemical structure of fragment 2 (see Supplementary Figure S1B), the user is informed that this relies to the same pair of molecules (CHEMBL4205172 and CHEMBL4212052) tested on the MAP3K14 kinase. Thus, whereas the chemical context can be considered as similar, the biological context is clearly different. Further details on the bioassays can be accessed through the respective ChEMBL links (CHEMBL4195748 and CHEMBL4195749) and in the publication (PubMed link 29940120) (see Supplementary Figure S1C). The second hopping step was the exchange of a 3-azabicyclo[3.1.0]hexane by a piperidine linker, which

is found in the SwissBioisostere database as the most frequent out of 52 possible replacements. Some of the 45 occurrences of this replacement show pairs of compounds bearing this change in close chemical and biological contexts. The third important step was the cyclization of a N-ethylamide to form a 1,3,4-oxadiazole linker. This specific replacement can be retrieved 19 times in SwissBioisostere in various biological contexts providing an ample source of information.

**Linker analysis: amide bioisosteres.** The amide linker have been extensively evaluated by medicinal chemists because of its high frequency in druglike and bio-molecules, of its convenience for automated synthesis, and for peptidomimetic design (25). The oxadiazole reported in the previous paragraph about Usmarapride is part of the most important heterocycle category of amide bioisostere, as defined by

Trippier *et al.* (26). Their conclusion is a detailed description of the 23 most common amide bioisosteres.

With the exceptions of diketopiperazine and phosphoramidate (not bioisosteres *stricto sensu* since the rest of the molecules is not constant), all other 21 common amide bioisosteres can be retrieved from the SwissBioisostere database (user can perform this request through the *linker* example in the submission page). The inverted amide is the first one, when sorted according to the Frequency column, with 2078 browsable examples within broad chemical and biological spaces. On the other end, the 1,2,3-thiazolodiazole moiety is involved in two occurrences only. In a few clicks, one can retrieve SwissBioisostere data for these 21 replacements based on 8449 pairs of molecules with measured activity on 586 proteins of 31 target classes published in 1414 articles referenced in PubMed. Also, 1625 of the datapoints (19.2%) are not obtained from literature but from additional sources included in ChEMBL, mainly high-throughput screening campaigns. 2079 replacement occurrences show a significant improvement of activity ( $\log$  activity  $\geq 0.5$ ), 1794 occurrences show significant reduction of activity ( $\log$  activity  $< 0.5$ ) and 4,576 occurrences show broadly similar bioactivity. With overall 24.6% and 54.2% of 'green' and 'orange' classes for bioactivity, respectively (see Input section in Graphical interface), the 21 replacing fragments described by Trippier *et al.* (26) can arguably be considered as *true* amide bioisosteres. Noteworthy, this results from the exact definition of both fragments, excluding elongation of chains or any substitution but including every appropriate tautomer and isomer. Of course, SwissBioisostere as any computer-aided tool is not intended to replace the thorough analysis capabilities of knowledgeable experts in drug discovery. However, our experience dictates that such tools can be of valuable support. For instance, SwissBioisostere Web users can access a first level of knowledge for 4087 possible replacements of amide linker. This extended bioisosteric space useful for guiding drug design is generated from 42 430 pairs of molecules with measured activity on 812 proteins of 33 target classes published in 3317 articles referenced in PubMed and 12 584 datapoints (29.7%) from other sources. Summing up all occurrences of possible amide replacements in SwissBioisostere, one finds 27.8% and 50.8% of 'green' and 'orange' classes for bioactivity, respectively. The proposition of improved and similar bioactivity is comparable to the 21 well described bioisosteres but on a much larger space expanding the opportunities for bioisosteric design.

*Side chain analysis: carboxylic acid bioisosteres.* The tetrazole heterocycle considered among the most common bioisosteres of amide by Trippier *et al.* and found in the data described in the previous paragraph is better known as a typical carboxylic acid bioisostere (27). Accordingly, SwissBioisostere (see Supplementary Figure S2A) returns both tautomers of tetrazole at position #12 and #20, respectively, when all 5093 possible replacement fragments of carboxylic acid are sorted by the Frequency column, with a total of 824 examples of pairs of molecules active on 195 proteins of 24 target classes. About 24.6% of the molecule with tetrazole instead of carboxylate are more active (green class) and 55.6% are similarly potent (orange class). Tetra-

zole is among the 36 bioisosteres of carboxylic acid studied by Ballatore *et al.* for physicochemical and permeation properties (28). A set of molecules was built on bioisosteric criteria but also following technical and diversity considerations. Most examples are present in SwissBioisostere. Only sulfinic acid, thiazolidinedione and other dione moieties, some urea derivatives and some specific phenol substitution patterns were not found. However, the search of thiazolidinedione, an important moiety for PPAR $\gamma$  agonist antidiabetic drugs, returned known replacements of diverse carboxylic acid containing longer branched side chains (user can perform this request through the *side chain* example in the submission page). In some cases, multiple requests are useful to behold properly the bioisosteric space of a chemical function or of a family of derivatives. Nonetheless, the 5093 possible fragments returned by SwissBioisostere as surrogate of carboxylic acid (strict definition) are based on 43 699 pairs of molecules active on 762 proteins of 32 target classes published in 2466 articles referenced in PubMed and 12 733 datapoints (29.13%) from other sources.

Ballatore *et al.* describe the unexpected low permeability of tetrazole bioisosteres failing at improving this deficiency of carboxylate molecules. Tetrazole is generally considered as more lipophilic and computed so by standard  $\log P$  methods. Compared to carboxylic acid, the CX Log  $P$  values ([www.chemaxon.com](http://www.chemaxon.com)) obtained from ChEMBL estimate a variation of + 0.68 and a decrease -0.27 for the  $\tau$ - and  $\pi$ -tautomers, respectively. This can be questioned by the higher tPSA for tetrazole than for carboxylic acid. Further analyses are available to SwissBioisostere users thanks to the interoperability icons (see Figure 4). Clicking on the pill icon directs to SwissADME Web tool (23), which includes iLOGP, a physics-based method to estimate  $\log P$  (29). Averaging calculated iLOGP values of tautomers of tetrazole gives 0.32 whereas iLOGP value for carboxylic acid is 0.63. High apparent polarity and low lipophilicity are properties impairing passive membrane crossing. 1788 out of the 5093 proposed fragments by SwissBioisostere combine both criteria for good permeability, i.e. higher  $\log P$  and lower tPSA than carboxylic acid (see Supplementary Figure S2B). This illustrates the complementarity of wisely chosen computational methods, experimental and expert knowledge for an accurate analysis of bioisosteric replacements allowing the best possible choice for molecular design.

*Scaffold analysis* available in Supplementary Data.

## CONCLUSIVE REMARKS

With major improvements, regarding the Match Molecular Pair machinery to find replacements, the backend and the frontend of the Web interface ([www.swissbioisostere.ch](http://www.swissbioisostere.ch)), this updated version of SwissBioisostere provides not only more numerous molecular replacements but also advanced and interactive analysis capacities. This is intended to support scientists working in computational biology, cheminformatics, medicinal chemistry and drug research for making the most informed choice on chemical fragments as potential bioisosteres to be considered for their drug discovery endeavors.

SwissBioisostere is part of the SwissDrugDesign project, a collection of tools and databases for computer-aided drug

design. The increased interoperability allows now Swiss-Bioisostere users to submit in one click any result molecule as input to SwissSimilarity for virtual screening, to SwissTargetPrediction for estimation of protein targets or to SwissADME for computing physicochemical and pharmacokinetic parameters. This represents a further step toward an integrated and free Web workspace for computer-aided drug design.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

Chemaxon is acknowledged for the licensing agreement. We thank SIB Swiss Institute of Bioinformatics for funding.

## FUNDING

SIB Swiss Institute of Bioinformatics. Funding for open access charge: own funds.

*Conflict of interest statement.* None declared.

## REFERENCES

- Langmuir, I. (1919) The structure of atoms and the octet theory of valence. *Proc. Natl. Acad. Sci.*, **5**, 252–259.
- Thorner, C.W. (1979) Isosterism and molecular modification in drug design. *Chem. Soc. Rev.*, **8**, 563.
- Maggiore, G., Vogt, M., Stumpfe, D. and Bajorath, J. (2014) Molecular similarity in medicinal chemistry. *J. Med. Chem.*, **57**, 3186–3204.
- Daina, A. and Zoete, V. (2019) Application of the SwissDrugDesign online resources in virtual screening. *Int. J. Mol. Sci.*, **20**, 1805–1831.
- Veale, C.G.L. (2021) Into the Fray! A Beginner's Guide to Medicinal Chemistry. *ChemMedChem*, **16**, 1199–1225.
- Bunch, L. (2013) Bioisosteres in Medicinal Chemistry. Edited by Nathan Brown. *ChemMedChem*, **8**, 1012–1012.
- Langdon, S.R., Ertl, P. and Brown, N. (2010) Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol. Inf.*, **29**, 366–385.
- Hughes, J.P., Rees, S., Kalindjian, S.B. and Philpott, K.L. (2011) Principles of early drug discovery. *Br. J. Pharmacol.*, **162**, 1239–1249.
- Hu, Y., Stumpfe, D. and Bajorath, J. (2017) Recent advances in scaffold hopping. *J. Med. Chem.*, **60**, 1238–1246.
- Schneider, G., Neidhart, W., Giller, T. and Schmid, G. (1999) 'Scaffold-Hopping' by Topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed. Engl.*, **38**, 2894–2896.
- Meanwell, N.A. (2011) Synopsis of some recent tactical application of bioisosteres in drug design. *J. Med. Chem.*, **54**, 2529–2591.
- Patani, G.A. and LaVoie, E.J. (1996) Bioisosterism: a rational approach in drug design. *Chem. Rev.*, **96**, 3147–3176.
- Lešnik, S., Škrlić, B., Eržen, N., Bren, U., Gobec, S., Konc, J. and Janežič, D. (2017) BoBER: web interface to the base of bioisosterically exchangeable replacements. *J. Cheminform.*, **9**, 62.
- Seddon, M.P., Cosgrove, D.A. and Gillet, V.J. (2018) Bioisosteric replacements extracted from high-quality structures in the Protein Databank. *ChemMedChem*, **13**, 607–613.
- Desaphy, J. and Rognan, D. (2014) sc-PDB-Frag: a database of protein-ligand interaction patterns for Bioisosteric replacements. *J. Chem. Inf. Comput. Sci.*, **54**, 1908–1918.
- Wirth, M., Zoete, V., Michielin, O. and Sauer, W.H.B. (2013) SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.*, **41**, D1137–D1143.
- Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M. et al. (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
- O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R. (2011) OpenBabel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.
- O'Boyle, N.M., Morley, C. and Hutchison, G.R. (2008) Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.*, **2**, 5.
- Hussain, J. and Rea, C. (2010) Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *J. Chem. Inf. Comput. Sci.*, **50**, 339–348.
- Zoete, V., Daina, A., Bovigny, C. and Michielin, O. (2016) SwissSimilarity: a web tool for low to ultra high throughput ligand-based virtual screening. *J. Chem. Inf. Comput. Sci.*, **56**, 1399–1404.
- Daina, A., Michielin, O. and Zoete, V. (2019) SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res.*, **47**, W357–W364.
- Daina, A., Michielin, O. and Zoete, V. (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.*, **7**, 42717.
- Nirogi, R., Mohammed, A.R., Shinde, A.K., Gagginapally, S.R., Kancharla, D.M., Ravella, S.R., Bogaraju, N., Middekadi, V.R., Subramanian, R., Palacharla, R.C. et al. (2021) Discovery and preclinical characterization of usmarapride (SUVN-D4010): a potent, selective 5-HT<sub>4</sub> receptor partial agonist for the treatment of cognitive deficits associated with Alzheimer's disease. *J. Med. Chem.*, **64**, 10641–10665.
- Pattabiraman, V.R. and Bode, J.W. (2011) Rethinking amide bond synthesis. *Nature*, **480**, 471–479.
- Kumari, S., Carmona, A.V., Tiwari, A.K. and Trippier, P.C. (2020) Amide bond bioisosteres: strategies, synthesis, and successes. *J. Med. Chem.*, **63**, 12290–12358.
- Herr, R.J. (2002) 5-Substituted-1H-tetrazoles as carboxylic acid isosteres: medicinal chemistry and synthetic methods. *Bioorg. Med. Chem.*, **10**, 3379–3393.
- Lassalas, P., Gay, B., Lasfargeas, C., James, M.J., Tran, V., Vijayendran, K.G., Brunden, K.R., Kozlowski, M.C., Thomas, C.J., Smith, A.B. et al. (2016) Structure property relationships of carboxylic acid isosteres. *J. Med. Chem.*, **59**, 3183–3203.
- Daina, A., Michielin, O. and Zoete, V. (2014) iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Comput. Sci.*, **54**, 3284–3301.