Current Computer-Aided Drug Design, 2020, *16*, 460-466

RESEARCH ARTICLE

MolOpt: A Web Server for Drug Design using Bioisosteric Transformation

Jinwen Shan[1,2] and Changge Ji[1,2,*]

[1]*Shanghai Engineering Research Center for Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062 P.R. China;* [2]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, 200062 P.R. China*

ARTICLE HISTORY

Received: March 13, 2019
Revised: May 12, 2019
Accepted: June 13, 2019

DOI:
10.2174/1573409915666190704093400

Abstract: ***Background*:** Bioisosteric replacement is widely used in drug design for lead optimization. However, the identification of a suitable bioisosteric group is not an easy task.

***Methods*:** In this work, we present MolOpt, a web server for *in silico* drug design using bioisosteric transformation. Potential bioisosteric transformation rules were derived from data mining, deep generative machine learning and similarity comparison. MolOpt tries to assist the medicinal chemist in his/her search for what to make next.

***Results and Discussion*:** By replacing molecular substructures with similar chemical groups, MolOpt automatically generates lists of analogues. MolOpt also evaluates forty important pharmacokinetic and toxic properties for each newly designed molecule. The transformed analogues can be assessed for possible future study.

***Conclusion*:** MolOpt is useful for the identification of suitable lead optimization ideas. The MolOpt Server is freely available for use on the web at http://xundrug.cn/molopt.

Keywords: Bioisosteric replacement, lead optimization, *in silico* drug design, data mining, machine learning, web server.

1. INTRODUCTION

Bioisosteric replacement plays a critical role in the process of drug optimization [1]. A bioisosteric replacement transforms an active compound into another compound by replacing part of a molecule with other substructures that share similar polar groups, volumes, shape and physiochemical properties [2-10]. Bioisosteric replacement is broadly used in medicinal chemistry to improve potency and selectivity, unravel issues related to drug pharmacokinetics and to get rid of undesirable side effects such as toxicity and metabolic liabilities [11-25].

Identification of a suitable bioisosteric group is not an easy task and requires large amounts of trial and error [1]. Computational strategies can provide very valuable help for the identification of suitable bioisosteric transformation ideas [26-28]. In this work, we present MolOpt, a publicly available web server for drug design using bioisosteric replacement rules. Three kinds of transformation rules were used in developing the server. One is based on the mining of bioactivity database through matched molecular pairs analysis method [16, 29-34]. The second one is based on a sequence-based generative model. The third one is based on a similarity comparison of putative bioisosteric substructure with topological pharmacophore fingerprints. MolOpt also evaluates important Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) related properties [35] for each newly designed molecule using various predictive models [36-42].

2. METHODS

2.1. Overview

MolOpt is a web server for *in silico* hit/lead optimization using bioisosteric transformation methods. When the user inputs a molecule and selects a substructure he/she wants to optimize, the server would recommend potential fragments for replacing and generating analogs. The program is composed of three modules; including a database of bioisosteric transformation rules, analog generation module and molecular property prediction module. The heart of the program is the bioisosteric replacement rule database which is designed by cheminformatics techniques, data mining approaches, and machine learning methods. Details of the methods used for bioisosteric replacement rule database construction are explained in the following sections. The MolOpt server takes

*Address correspondence to this author at the Shanghai Engineering Research Center for Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062 P.R. China; and NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, 200062 P.R. China; Tel: 86-21-20596013; Fax: 86-21-62232403; E-mail: chicago.ji@gmail.com

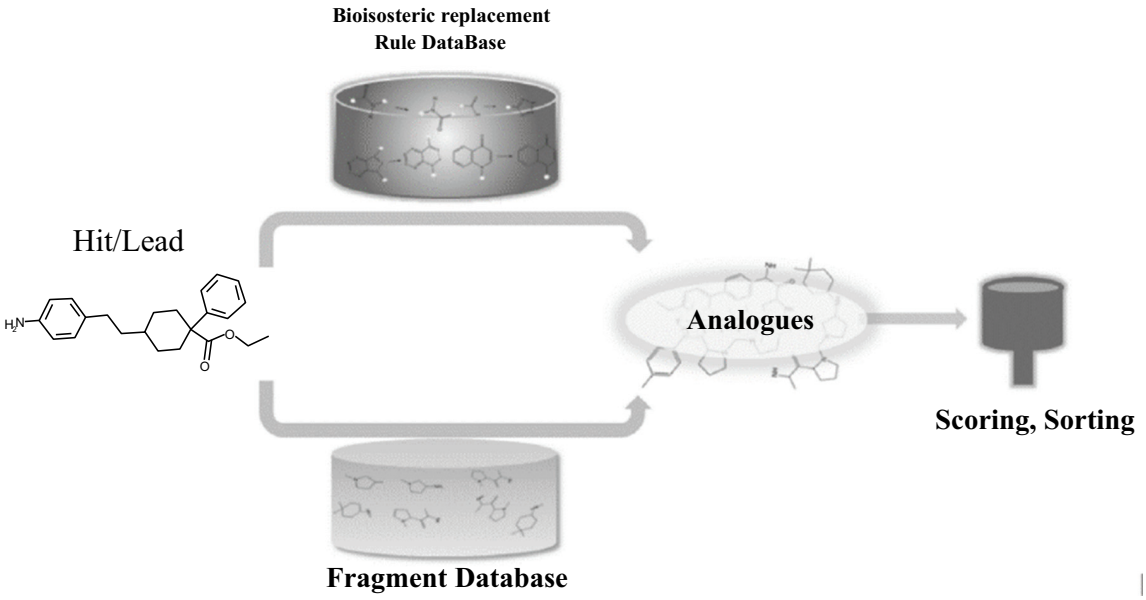


Fig. (1). Overview of the MolOpt webserver architecture.

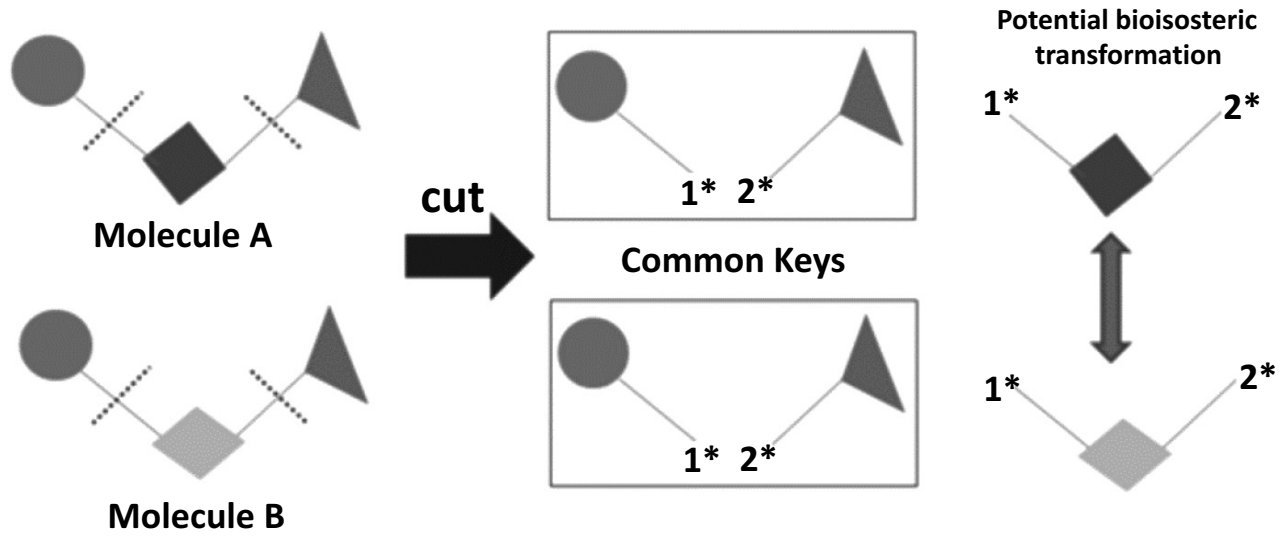


Fig. (2). Schematic workflow of the fragment indexing Matched Molecular Pair algorithm.

four steps to complete the hit/lead optimization recommendation task. Firstly, the program would divide the molecule into fragments and pick the fragment which is chosen by the user for data searching. Secondly, the program searches replaceable fragments according to the query fragment from the bioisosteric replacement rule database. Thirdly, it sorts the new fragments according to the topological similarity and other properties. Finally, analogs were generated by connecting the new fragments and the remaining parts of the molecule (Fig. **1**).

2.2. Mining Bioisosteric Groups from CHEMBL Database

CHEMBL is a manually curated chemical database of bioactive molecules with drug-like properties [43]. CHEMBL contains bioactivity data against drug targets extracted from more than 60000 medicinal chemistry publications. Compounds were associated with each of the journal article's records showing information about how medicinal chemists optimize a lead compound. We obtained sets of analogues from individual journal articles in the Chembl database (v. 23) and analyzed each set of analogues through the Matched Molecular Pairs Analysis [23] method to derive historical bioisosteric transformation rules used by medicinal chemists.

Hussain and Rea developed a computationally efficient algorithm [44] to identify matched molecular pairs in large data sets which are suitable for our data mining job. The method works by fragmenting and appropriately indexing the compounds under analysis (Fig. **2**). Firstly, all the compounds in the input data are fragmented on all its acyclic single bonds iteratively. Secondly, appropriately all the enumerated complementary fragments are stored and indexed. Finally, for each pair molecule, the fragments having a common pair are found followed by the corresponding bioisosteric transformation. In our current implementation, we only considered transformations smaller than 20 non-hydrogen atoms. The transformation rules were encoded in SMIRKS format and stored in a mysql database.

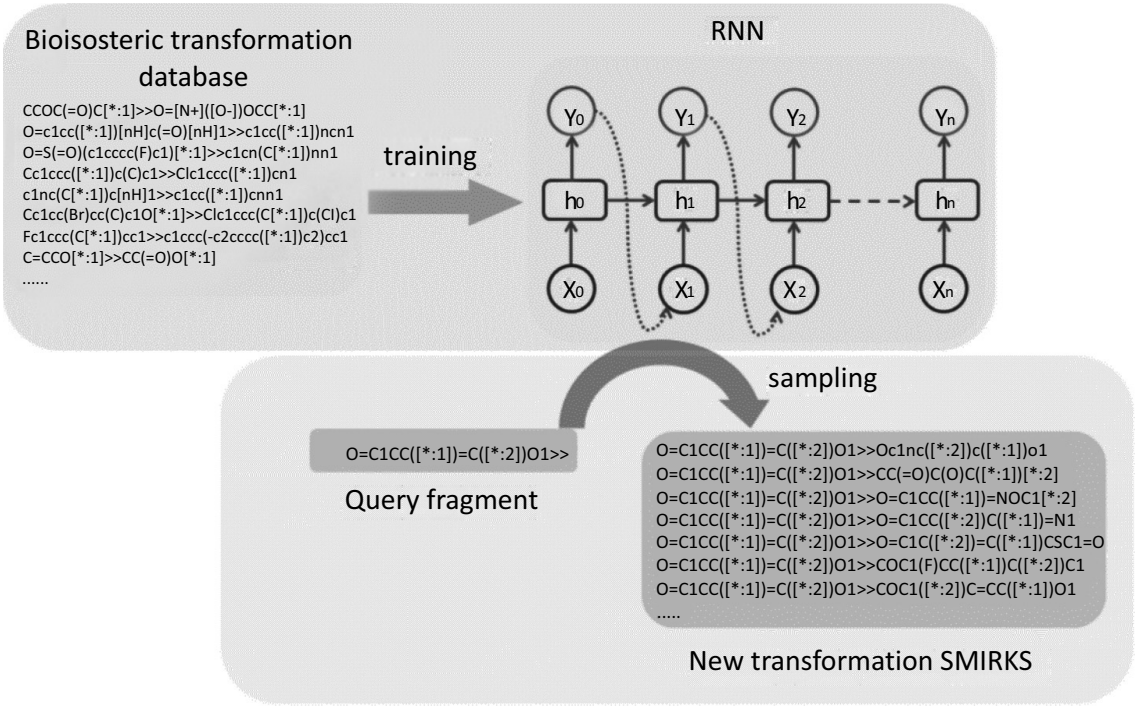


Fig. (3). Sequence-based generative model for molecular optimization.

2.3. Deep Generative Model

In the past several years, there was a growing trend in using deep neural networks for the generation of new chemical structures. Many comprehensive generative models [45-47] were constructed using SMILES as a representation, partly driven by the adoption of natural language processing deep learning tools. SMIRKS representation is widely used for storing chemical transformation rules. Similar to SMILES, SMIRKS representation is also suitable for deep generative model building using recurrent neural networks. Similar to the generative method described in a previous work by Olivecrona and coworkers [46], we applied a Recurrent Neural Network (RNN) to bioisosteric transformations represented as SMIRKS strings to obtain generative models which learn the probability distribution of characters in a SMIRKS string. The SMIRKS is tokenized according to a vocabulary list (Text **S2**). PyTorch was used for model building and training. The RNN is composed of 3 layers with 512 Gated Recurrent Units in each layer. The RNN was trained on SMIRKS of 1.4 million bioisosteric transformation rules derived from MMP analysis in section 2.1. The resulting models were capable of generating new strings which correspond to chemically meaningful SMIRKS. After training, 92% of the sequences generated by the model are recognized as valid structures according to RDKit parsing. When using the model for compound optimization, SMARTS of a query molecular fragment and the transformation syntax ('>>') is fed to the network as the initial sequence. The network would generate new fragment SMARTS according to the initial input sequence and the conditional probability distribution learned by previous training (Fig. **3**).

2.4. Identify Bioisosteric Groups by Similarity Comparison

Another approach to determine bioisosteric fragments is to compare similarity. This approach does not use historical lead optimization data, it is more likely to find out some novel bioisostere. Many molecular representations have been used for similarity comparison of putative bioisosteric substructures including molecular properties and descriptors such as size, shape, hydrophobicity, topological fingerprints, molecular field, H-bond capabilities or pharmacophore. In this study, we used topological pharmacophore fingerprints [14] to calculate the similarity between fragments to identify bioisosteres. Each fragment is abstracted as a combination of attachment points, hydrogen bond donors and acceptors, hydrophobic atoms, and non-hydrogen atoms. The distance relation for the underlying pharmacophore feature set is encoded as fingerprints for similarity calculation. Tanimoto coefficient is employed to measure similarity through fingerprints.

2.5. ADMET Prediction

Various types of pharmacokinetic and toxic data are generated *in vitro* and *in vivo,* and this data can now be used to predict ADMET properties of a drug candidate at an early stage. In our current work, we have collected different types of data from public databases and literature, including Caco-2 permeability assay, transporter inhibition assay, CYP450 inhibition/activation assay, cytotoxicity assay, acute oral toxicity assay and so on. A total of 40 data sets were extracted, details of which are listed in Table **S1**. 2048 bit ECFP_4 circular Morgan fingerprints and 196 different physiochemical descriptors (Text **S1**) generated by RDKit [48] were

Analog structure and property:

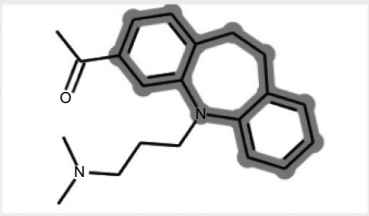
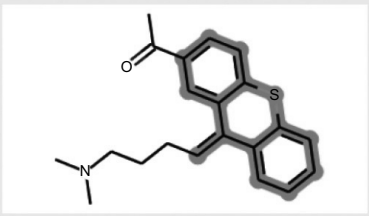
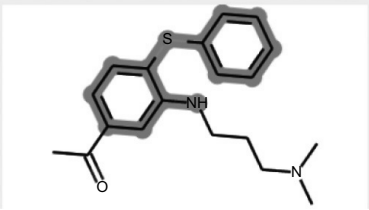
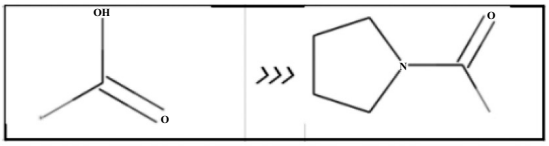
Index	Analog	Synthetic Accessibility	cLogP	TPSA	Caco-2 Permeability (logPapp)	Download Molecule	Rule Example
000		2.13	4.078	23.55	-4.574	smi	history
001		2.485	5.127	20.31	-4.459	smi	history
002		2.109	4.404	32.34	-4.476	smi	history

Fig. (4). Example figure on result table from MolOpt.

Transformation In Previous Studies



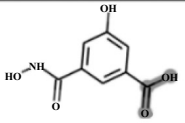
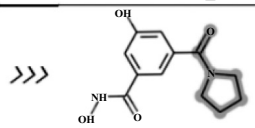
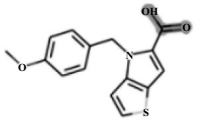
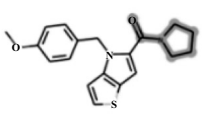
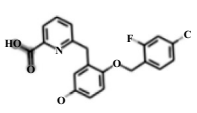
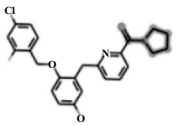
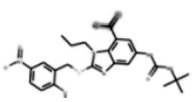
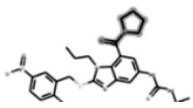
Index	Molecule_A		Molecule_B	Reference
0				*Bioorg. Med. Chem Lett., 2007, 17, 964-968*
1				*J. Med. Chem, 2015, 58, 9196-9213*
2				*Bioorg. Med Chem Lett., 2013, 23, 37-41*
3				*Bioorg. Med Chem Lett., 2005, 15, 2265-2269*

Fig. (5). Example xlsx file of historical examples using a specific transformation rule.

used for building each of the prediction model. The Random Forest (RF) module of the Scikit-Learn package [49] was used as an underlying machine learning model to build binary classification RF models and regression RF models. The performance of each prediction model is estimated by the fivefold cross valuation and is listed in Table **S1**.

2.6. Web Implementation

The web server uses node.js code to run the interface functionality and python code to perform SQL database query and similarity comparison. RDKit [48] is used for analyzing matched molecular pairs, generating molecular topological pharmacophore fingerprints, matching substructures and drawing molecule picture. Standardizer [50] of ChemAxon was used for structure canonicalization. Chemical Identifier Resolver is used for converting molecular name and CAS id to SMILES strings.

3. APPLICATION

The MolOpt web server offers the user many ways to submit query molecules, namely SMILES strings, IUPAC name, commercial name, CAS ID, InChI key or by drawing a molecule. Usage of the MolOpt Sever takes two steps: Firstly, a molecule is given as input followed by the selection of a transformation rule method(using historical bioisosteric transformation rules or using similarity-based methods). Secondly, the server will generate all replaceable substructures on the input molecule, and the users choose the substructure they want to optimize and submit. The results page of MolOpt is composed of two main sections: Firstly, a table of all the analogues generated by applying bioisosteric replacements to the query molecule, together with the calculated synthetic accessibility [51] and properties (Fig. **4**). The replaced substructure of the analog is highlighted for visualization convenience. Each molecule can be downloaded directly as the SMILES format. Secondly, the full prediction results are provided as an xlsx file for downloading, including SMILES and calculated ADMET properties for all analogues. The SMILES can be used for further *in silico* evaluation of potency, safety and other properties of these analogues. In addition, the server also provides historical examples (Fig. **5**) using a specific transformation rule which is useful for further analysis and decision making.

CONCLUSION

MolOpt is a user-friendly and robust tool to navigate the historical bioisosteric group space and identify new bioisosteric transformation ideas. The analogue generation process for a single group optimization task usually takes about one to two minutes. The web server does not require computational or programming skills from the user which is favorable to medicinal chemists. MolOpt tries to assist the medicinal chemist in his/her searching process. The MolOpt server will be continuously updated as new data become available. This service is freely available to the public at http://xundrug.cn/molopt.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used in the study that is the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary material.

FUNDING

We thank the National Key Research and Development Plan (2016YFA0501700), National Natural Science Foundation of China (Grants 21003048 and 21433004), Shanghai Natural Science Foundation (Grant 14ZR1411900) and China Scholarship Council for financial support.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

We thank the Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development R&D Program.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Langdon, S.R.; Ertl, P.; Brown, N. Bioisosteric replacement and scaffold hopping in lead generation and optimization. *Mol. Inform.,* **2010**, *29*(5), 366-385. http://dx.doi.org/10.1002/minf.201000019 PMID: 27463193

[2] Seddon, M.P.; Cosgrove, D.A.; Gillet, V.J. Bioisosteric replacements extracted from high-quality structures in the protein databank. *ChemMedChem,* **2018**, *13*(6), 607-613. http://dx.doi.org/10.1002/cmdc.201700679 PMID: 29314719

[3] Lange, J.H.M.; van Stuivenberg, H.H.; Coolen, H.K.A.C.; Adolfs, T.J.P.; McCreary, A.C.; Keizer, H.G.; Wals, H.C.; Veerman, W.; Borst, A.J.M.; de Looff, W.; Verveer, P.C.; Kruse, C.G. Bioisosteric replacements of the pyrazole moiety of rimonabant: synthesis, biological properties, and molecular modeling investigations of thiazoles, triazoles, and imidazoles as potent and selective CB1 cannabinoid receptor antagonists. *J. Med. Chem.,* **2005**, *48*(6), 1823-1838. http://dx.doi.org/10.1021/jm040843r PMID: 15771428

[4] Bhandare, R.R.; Canney, D.J. Bioisosteric replacement and related analogs in the design, synthesis and evaluation of ligands for muscarinic acetylcholine receptors. *Med. Chem.,* **2014**, *10*(4), 361-375. http://dx.doi.org/10.2174/1573406411309666004 3 PMID: 24021160

[5] Rombouts, F.J.R.; Tovar, F.; Austin, N.; Tresadern, G.; Trabanco, A.A. Benzazaborinines as novel bioisosteric replacements of naphthalene: propranolol as an example. *J. Med. Chem.,* **2015**, *58*(23), 9287-9295. http://dx.doi.org/10.1021/acs.jmedchem.5b01088 PMID: 26565745

[6] Goldberg, K.; Groombridge, S.; Hudson, J.; Leach, A.G.; MacFaul, P.A.; Pickup, A.; Poultney, R.; Scott, J.S.; Svensson, P.H.; Sweeney, J. Oxadiazole isomers: all bioisosteres are not created equal. *MedChemComm,* **2012**, *3*(5), 600-604. http://dx.doi.org/10.1039/c2md20054f

[7] Elliott, T.S.; Slowey, A.; Ye, Y.; Conway, S.J. The use of phosphate bioisosteres in medicinal chemistry and chemical biology. *MedChemComm,* **2012**, *3*(7), 735-751. http://dx.doi.org/10.1039/c2md20079a

[8] Marson, C.M. New and unusual scaffolds in medicinal chemistry. *Chem. Soc. Rev.,* **2011**, *40*(11), 5514-5533. http://dx.doi.org/10.1039/c1cs15119c PMID: 21837344

[9] Mykhailiuk, P.K. Saturated bioisosteres of benzene: where to go next? *Org. Biomol. Chem.,* **2019**, *17*(11), 2839-2849. http://dx.doi.org/10.1039/C8OB02812E PMID: 30672560

[10] Kenny, P.W.; Montanari, C.A.; Prokopczyk, I.M.; Sala, F.A.; Sartori, G.R. Automated molecule editing in molecular design. *J. Comput. Aided Mol. Des.,* **2013**, *27*(8), 655-664. http://dx.doi.org/10.1007/s10822-013-9676-0 PMID: 24002455

[11] Meanwell, N.A. Synopsis of some recent tactical application of bioisosteres in drug design. *J. Med. Chem.,* **2011**, *54*(8), 2529-2591. http://dx.doi.org/10.1021/jm1013693 PMID: 21413808

[12] Patani, G.A.; LaVoie, E.J. Bioisosterism: a rational approach in drug design. *Chem. Rev.,* **1996**, *96*(8), 3147-3176. http://dx.doi.org/10.1021/cr950066q PMID: 11848856

[13] Diana, G.D.; Volkots, D.L.; Nitz, T.J.; Bailey, T.R.; Long, M.A.; Vescio, N.; Aldous, S.; Pevear, D.C.; Dutko, F.J. Oxadiazoles as ester bioisosteric replacements in compounds related to disoxaril. Antirhinovirus activity. *J. Med. Chem.,* **1994**, *37*(15), 2421-2436. http://dx.doi.org/10.1021/jm00041a022 PMID: 8057290

[14] Wagener, M.; Lommerse, J.P.M. The quest for bioisosteric replacements. *J. Chem. Inf. Model.,* **2006**, *46*(2), 677-685. http://dx.doi.org/10.1021/ci0503964 PMID: 16562998

[15] Desaphy, J.; Rognan, D. sc-PDB-Frag: a database of protein-ligand interaction patterns for Bioisosteric replacements. *J. Chem. Inf. Model.,* **2014**, *54*(7), 1908-1918. http://dx.doi.org/10.1021/ci500282c PMID: 24991975

[16] Weber, J.; Achenbach, J.; Moser, D.; Proschak, E. VAMMPIRE: a matched molecular pairs database for structure-based drug design and optimization. *J. Med. Chem.,* **2013**, *56*(12), 5203-5207. http://dx.doi.org/10.1021/jm400223y PMID: 23734609

[17] Krier, M.; Hutter, M.C. Bioisosteric similarity of molecules based on structural alignment and observed chemical replacements in drugs. *J. Chem. Inf. Model.,* **2009**, *49*(5), 1280-1297. http://dx.doi.org/10.1021/ci8003418 PMID: 19402687

[18] Dick, B.L.; Cohen, S.M. Metal-binding isosteres as new scaffolds for metalloenzyme inhibitors. *Inorg. Chem.,* **2018**, *57*(15), 9538-9543. http://dx.doi.org/10.1021/acs.inorgchem.8b01632 PMID: 30009599

[19] Meanwell, N.A. Fluorine and fluorinated motifs in the design and application of bioisosteres for drug design. *J. Med. Chem.,* **2018**, *61*(14), 5822-5880. http://dx.doi.org/10.1021/acs.jmedchem.7b01788 PMID: 29400967

[20] Zhang, Y.; Borrel, A.; Ghemtio, L.; Regad, L.; Boije Af Gennäs, G.; Camproux, A-C.; Yli-Kauhaluoma, J.; Xhaard, H. Structural isosteres of phosphate groups in the protein data bank. *J. Chem. Inf. Model.,* **2017**, *57*(3), 499-516. http://dx.doi.org/10.1021/acs.jcim.6b00519 PMID: 28234462

[21] Zafrani, Y.; Yeffet, D.; Sod-Moriah, G.; Berliner, A.; Amir, D.; Marciano, D.; Gershonov, E.; Saphier, S. Difluoromethyl bioisostere: examining the "lipophilic hydrogen bond donor" Concept. *J. Med. Chem.,* **2017**, *60*(2), 797-804. http://dx.doi.org/10.1021/acs.jmedchem.6b01691 PMID: 28051859

[22] 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., III; Huryn, D.M.; Ballatore, C. Structure property relationships of carboxylic acid isosteres. *J. Med. Chem.,* **2016**, *59*(7), 3183-3203. http://dx.doi.org/10.1021/acs.jmedchem.5b01963 PMID: 26967507

[23] Griffen, E.; Leach, A.G.; Robb, G.R.; Warner, D.J. Matched molecular pairs as a medicinal chemistry tool. *J. Med. Chem.,* **2011**, *54*(22), 7739-7750. http://dx.doi.org/10.1021/jm200452d PMID: 21936582

[24] Poulie, C.B.M.; Bunch, L. Heterocycles as nonclassical bioisosteres of α-amino acids. *ChemMedChem,* **2013**, *8*(2), 205-215. http://dx.doi.org/10.1002/cmdc.201200436 PMID: 23322633

[25] Dudkin, V.Y. Bioisosteric equivalence of five-membered heterocycles. *Chem. Heterocycl. Compd. (N. Y., NY, U. S.),* **2012**, *48*(1), 27-32. http://dx.doi.org/10.1007/s10593-012-0964-8

[26] Stewart, K.D.; Shiroda, M.; James, C.A. Drug Guru: a computer software program for drug design using medicinal chemistry rules. *Bioorg. Med. Chem.,* **2006**, *14*(20), 7011-7022. http://dx.doi.org/10.1016/j.bmc.2006.06.024 PMID: 16870456

[27] Wirth, M.; Zoete, V.; Michielin, O.; Sauer, W.H.B. SwissBioisostere: a database of molecular replacements for ligand design. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D1137-D1143. http://dx.doi.org/10.1093/nar/gks1059 PMID: 23161688

[28] Ertl, P.; Lewis, R. IADE: a system for intelligent automatic design of bioisosteric analogs. *J. Comput. Aided Mol. Des.,* **2012**, *26*(11), 1207-1215. http://dx.doi.org/10.1007/s10822-012-9609-3 PMID: 23053736

[29] Tyrchan, C.; Evertsson, E. Matched molecular pair analysis in short: algorithms, applications and limitations. *Comput. Struct. Biotechnol. J.,* **2016**, *15*, 86-90. http://dx.doi.org/10.1016/j.csbj.2016.12.003 PMID: 28066532

[30] Leach, A.G.; Jones, H.D.; Cosgrove, D.A.; Kenny, P.W.; Ruston, L.; MacFaul, P.; Wood, J.M.; Colclough, N.; Law, B. Matched molecular pairs as a guide in the optimization of pharmaceutical properties; a study of aqueous solubility, plasma protein binding and oral exposure. *J. Med. Chem.,* **2006**, *49*(23), 6672-6682. http://dx.doi.org/10.1021/jm0605233 PMID: 17154498

[31] Turk, S.; Merget, B.; Rippmann, F.; Fulle, S. Coupling matched molecular pairs with machine learning for virtual compound optimization. *J. Chem. Inf. Model.,* **2017**, *57*(12), 3079-3085. http://dx.doi.org/10.1021/acs.jcim.7b00298 PMID: 29131617

[32] Kramer, C.; Fuchs, J.E.; Whitebread, S.; Gedeck, P.; Liedl, K.R. Matched molecular pair analysis: significance and the impact of experimental uncertainty. *J. Med. Chem.,* **2014**, *57*(9), 3786-3802. http://dx.doi.org/10.1021/jm500317a PMID: 24738976

[33] Dossetter, A.G.; Griffen, E.J.; Leach, A.G. Matched molecular pair analysis in drug discovery. *Drug Discov. Today,* **2013**, *18*(15-16), 724-731. http://dx.doi.org/10.1016/j.drudis.2013.03.003 PMID: 23557664

[34] de la Vega de León, A.; Bajorath, J. Matched molecular pairs derived by retrosynthetic fragmentation. *MedChemComm,* **2014**, *5*(1), 64-67. http://dx.doi.org/10.1039/C3MD00259D

[35] Ji, C.; Svensson, F.; Zoufir, A.; Bender, A. eMolTox: prediction of molecular toxicity with confidence. *Bioinformatics,* **2018**, *34*(14), 2508-2509. http://dx.doi.org/10.1093/bioinformatics/bty135 PMID: 29522123

[36] Cheng, F.; Li, W.; Liu, G.; Tang, Y. *In silico* ADMET prediction: recent advances, current challenges and future trends. *Curr. Top. Med. Chem.,* **2013**, *13*(11), 1273-1289. http://dx.doi.org/10.2174/15680266113139990033 PMID: 23675935

[37] Yan, Y.; Wang, W.; Sun, Z.; Zhang, J.Z.H.; Ji, C. Protein-ligand empirical interaction components for virtual screening. *J. Chem. Inf. Model.,* **2017**, *57*(8), 1793-1806. http://dx.doi.org/10.1021/acs.jcim.7b00017 PMID: 28678484

[38] van de Waterbeemd, H.; Gifford, E. ADMET *in silico* modelling: towards prediction paradise? *Nat. Rev. Drug Discov.,* **2003**, *2*(3), 192-204. http://dx.doi.org/10.1038/nrd1032 PMID: 12612645

[39] Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; Liu, G.; Tang, Y. ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. *MedChemComm,* **2018**, *10*(1), 148-157. http://dx.doi.org/10.1039/C8MD00472B PMID: 30774861

[40] Dearden, J.C. *In silico* prediction of ADMET properties: how far have we come? *Expert Opin. Drug Metab. Toxicol.,* **2007**, *3*(5), 635-639.

http://dx.doi.org/10.1517/17425255.3.5.635 PMID: 17916052

[41] Yang, H.; Sun, L.; Li, W.; Liu, G.; Tang, Y. *In silico* prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Front Chem.,* **2018**, *6*(30)

[42] Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. Deeptox: toxicity prediction using deep learning. *Front. Environ. Sci.,* **2016**, *3*(80).

[43] Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J.P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.,* **2012**, *40*(Database issue), D1100-D1107.
http://dx.doi.org/10.1093/nar/gkr777 PMID: 21948594

[44] Hussain, J.; Rea, C. Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *J. Chem. Inf. Model.,* **2010**, *50*(3), 339-348.
http://dx.doi.org/10.1021/ci900450m PMID: 20121045

[45] Gómez-Bombarelli, R.; Wei, J.N.; Duvenaud, D.; Hernández-Lobato, J.M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T.D.; Adams, R.P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.,* **2018**, *4*(2), 268-276.
http://dx.doi.org/10.1021/acscentsci.7b00572 PMID: 29532027

[46] Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.,* **2017**, *9*(1), 48.
http://dx.doi.org/10.1186/s13321-017-0235-x PMID: 29086083

[47] Segler, M.H.S.; Kogej, T.; Tyrchan, C.; Waller, M.P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.,* **2018**, *4*(1), 120-131.
http://dx.doi.org/10.1021/acscentsci.7b00512 PMID: 29392184

[48] Landrum, G. *Landrum, G., RDKit: Open-source cheminformatics,* **2006**.

[49] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikitlearn: Machine Learning in Python. *J. Mach. Learn. Res.,* **2011**, *12*, 2825-2830.

[50] ChemAxon, JChem (Version 16.10.10), ChemAxon, 2016 http://www.chemaxon.com

[51] Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.,* **2009**, *1*(1), 8.
http://dx.doi.org/10.1186/1758-2946-1-8 PMID: 20298526