Journal of
CHEMICAL INFORMATION
AND MODELING

Article

pubs.acs.org/jcim

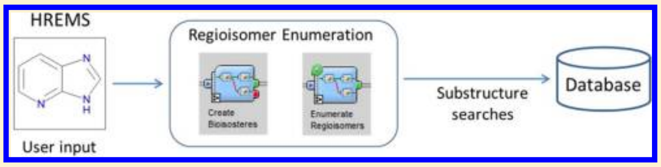
Heterocyclic Regioisomer Enumeration (HREMS): A Cheminformatics Design Tool

Sriram Tyagarajan,*[†] Christopher T. Lowden,*[‡] Zhengwei Peng,[†] Kevin D. Dykstra,[†] Edward C. Sherer,[†] and Shane W. Krska[†]

[†]Merck Research Laboratories, Merck & Co., Inc., P.O. Box 2000, Rahway, New Jersey 07065, United States

[‡]Workflow Informatics Corp., 7014 Englehardt Drive, Raleigh, North Carolina 27617, United States

Supporting Information



ABSTRACT: We report the development and implementation of a cheminformatics tool which aids in the design of compounds during exploratory chemistry and lead optimization. The Heterocyclic Regioisomer Enumeration and MDDR Search (HREMS) tool allows medicinal chemists to build greater structural diversity into their synthetic planning by enabling a systematic, automated enumeration of heterocyclic regioisomers of target structures. To help chemists overcome biases arising from past experience or synthetic accessibility, the HREMS tool further provides statistics on clinical testing for each enumerated regioisomer substructure using an automated search of a commercial database. Ready access to this type of information can help chemists make informed choices on the targets they will pursue being mindful of past experience with these structures in drug development. This tool and its components can be incorporated into other cheminformatics workflows to leverage their capabilities in triaging and in silico compound enumeration.

■ INTRODUCTION

The road to selection of a preclinical candidate in drug discovery involves iterative design−make−test cycles in order to explore structure−activity relationships (SAR) and optimize physical and biological properties. In the design phase, medicinal chemists traditionally map out a limited target space to be explored in the next iteration. Recognizing the limitations of such an approach in terms of generating sufficient chemical diversity and making efficient progress toward an optimized structure, in recent years the drug discovery community has embraced the use of in silico compound library enumeration and prospective property prediction. Exploration of virtual chemical space can be done in several ways but typically involves the in silico coupling of a given core molecule with a selection of chemist-defined monomers according to a specific reaction type to generate a virtual compound library. In many cases the cores in these exercises consist of heterocycles, which constitute the backbone of most pharmaceutical compounds.[1] Heterocyclic compounds of a given molecular formula often may exist as distinct regioisomers, each having different positioning of the constituent atoms. The positioning of heteroatoms in heterocyclic scaffolds plays a crucial role in determining both the physical and biological properties of the overall molecule, including pK_a, lipophilicity, polar surface area, hydrogen bonding donors and acceptors, potency, metabolic stability, and various off-target activities.[1,2] Thus, a thorough exploration of core regioisomers ought be a key component of any virtual chemical space enumeration exercise.

For most chemists, the choice of which regioisomer to synthesize within a given class of heterocycles is often influenced by previous experience, commercial availability, and/or ease of synthesis. If the design strategy with respect to heterocyclic regioisomers is not exhaustive, it is possible that the design plans could be missing regioisomers with better profiles compared to those selected for initial exploration. A cheminformatics tool that would allow the design chemist to observe and triage all regioisomeric possibilities on a dashboard-like interface would help to remove some of these biases and enable a more informed approach. Furthermore, this type of cheminformatics tool could be incorporated into a larger molecular design workbench to leverage additional computational methods such as physical property predictions or structure based design workflows which allow for other considerations such as conformation, intramolecular hydrogen bonding or binding interactions.

Chemists currently use various computational tools to identify novel scaffolds for a given chemical target.[3−5] As an example, the DBMAKER program generates compound libraries based on various user-defined parameters and constraints—part of this algorithm involves the enumeration of heterocyclic regioisomers.[6] Other groups have used computational approaches to build prospective databases of heterocycles, including associated regioisomeric forms, for use in mapping unexplored chemical space,[7] identifying diverse sets of synthetic building blocks[8] and scaffold-hopping/bioisostere identification.[9,10] What these tools have in common is the generation of prepopulated heterocycle databases which allow one to search over multiple ring systems by defining parameters which control the breadth of output heterocyclic diversity.

Received: March 26, 2015
Published: May 27, 2015

DOI: 10.1021/acs.jcim.5b00162

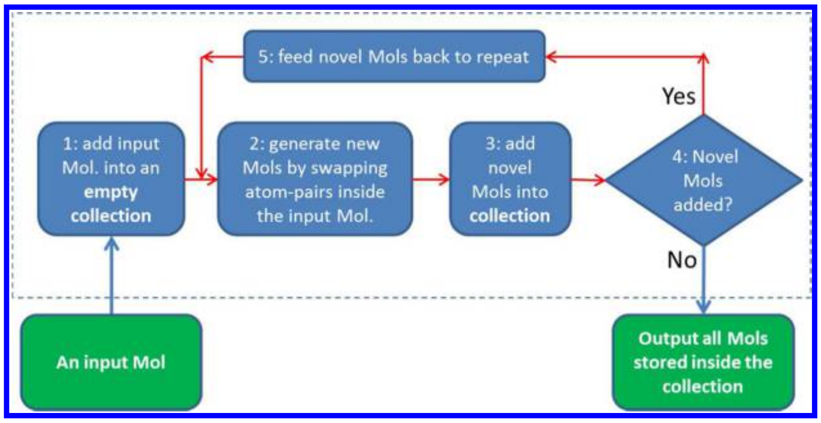

5: feed novel Mols back to repeat
Yes
1: add input Mol. into an empty collection
2: generate new Mols by swapping atom-pairs inside the input Mol.
3: add novel Mols into collection
4: Novel Mols added?
No
An input Mol
Output all Mols stored inside the collection

Figure 1. Regioisomer generation via atom-pair swapping.

In contrast, a complementary approach would take a target structure supplied by the medicinal chemist as a starting point and utilize a simple algorithm to generate a complete enumerated list of all possible regioisomers for the input heterocycle. This approach could be made even more powerful by retrieving information on how each enumerated regioisomer has fared in past clinical experience. The MDL Drug Data Report (MDDR, available from Biovia) is a very popular database used by medicinal chemists, pharmacologists and molecular modelers to create SAR tables and mine information about various pharmacologically relevant molecular entities. The database furnishes pertinent clinical information for a given molecular entity, such as therapeutic area, mechanism of action, patents and status in the clinical pipeline. Each heterocycle generated in the algorithm described above could be used as input for a substructural search of the MDDR, which enables linkage of information between retrieved structures and input molecules. From this derivative information, comparisons could be made among regioisomers based upon how each regioisomer has progressed in the clinical arena (Preclinical, Phase I, II, and III, and discontinued). With this information in hand, a chemist might choose to prioritize synthesis of a given regioisomer which has obtained marketed drug status over one which has not reached Phase I (of course, maintaining on-target activity is essential in addition to the pursuit of more favorable molecular properties). The information regarding molecular fate enables chemists to start thinking about the clinical impact of decisions made in the early phase of drug discovery. Furthermore, a regioisomer enumeration tool may facilitate protection of intellectual property for a particular scaffold in a patent. Patent informatics is a rapidly growing field, and this tool could be helpful in analyzing patents for gaps in a proprietary or competitor's patent and addressing those gaps.[11] While the tool described does not account for synthetic accessibility, we note the output of the tool could be passed through such a secondary filter where available.

Despite the obvious utility of such an approach, to date no computational tools have been reported which provide enumeration of heterocycle regioisomers of an input heterocycle partnered with subsequent searches over connected databases. This publication highlights a method to address such a gap and details the construction of a tool called HREMS (Heterocyclic Regioisomer Enumeration and MDDR Search) which enables chemists to optimize their design strategy based on cheminformatics analysis to facilitate drug discovery efforts and maximize potential for clinical success. This tool is made publically available in the Supporting Information.

MATERIALS AND METHODS

The HREMS tool is implemented as a web application based on Biovia's AEP framework (also called Pipeline Pilot, http://www.3ds.com/products-services/biovia/).

Molecular Preparation. Molecules entering from the input form are transformed to canonical SMILES and converted to a molecular data record. If bioisosteric enumeration is specified, bioisosteres are generated (See Bioisostere Enumeration). Each resulting molecule is subjected to a filtration process to ensure that molecules contain at least one carbon atom and at least one heteroatom (N, O, or S).

Regioisomer Generation. The basic steps for regioisomer generation via atom-pair swapping are depicted in Figure 1. Step 1 initializes the process by adding the input molecule (a heterocyclic ring system within the context of this report) into an empty collection. Step 2 permutes (swaps) all atom-pairs inside the input molecule when feasible according to specific rules to create new regioisomers of the input molecule. Step 3 identifies and adds novel molecules not yet inside the collection. Step 4 checks if any new novel molecules have been encountered. If novel molecules are still being generated, they will be fed back into Step 2 to be expanded further into additional regioisomers. Finally when no additional novel molecules are generated, this regioisomer generation process terminates and all of the molecules stored inside the collection are exported as output.

To ensure that valid molecules are created as regioisomers of the input molecule, certain rules are applied during the atom-pair swapping step. Rule 1: swap a pair of atoms only when the two atoms are of different element types; this is a trivial rule intended to avoid unnecessary swaps which can lead to lower performance. Rule 2: the bond orders for two swapped atoms need to be the same so that the swap leads to a valid molecule and a regioisomer of the input molecule. Rule 3: the neighbor counts for both atoms must be the same; this rule has been enforced inside the published work from other groups.[8] As an example, the swap seen inside Figure 2 would be rejected if all rules were enforced, even though the molecule after the swap is a valid molecule and a regioisomer of the original one. In our work, Rules 1 and 2 are fully enforced. We created an option for users to enforce Rule 3 if desired.

After the generation of regioisomers, the canonical tautomers of each regioisomer are enumerated using the *Enumerate*

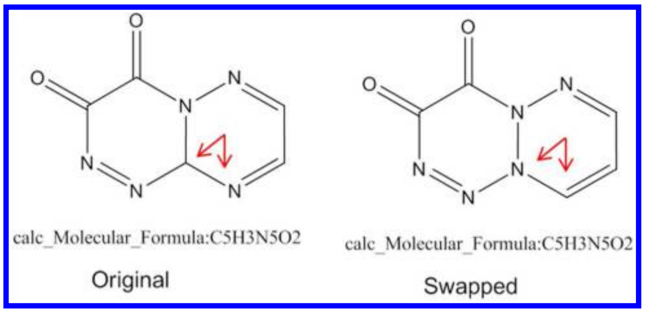

calc_Molecular_Formula:C5H3N5O2
Original
calc_Molecular_Formula:C5H3N5O2
Swapped

Figure 2. Example swap without enforcing the neighbor count constraint. The nitrogen atom has two heavy atom neighbors before the swap but has three afterward.

Tautomers component in Pipeline Pilot, followed by removal of all duplicate molecules.

The regioisomer generation workflow was implemented as a reusable Pipeline Pilot component. The step involving atom-pair swapping was implemented as a pilot script (Figure S1, Supporting Information) utilizing the AEP's Chemistry Toolkit.

Bioisostere Enumeration. Bioisostere enumeration uses a modified version of the *Enumerate Bioisosteres (Similarity)* component in Pipeline Pilot. The native form of this Pipeline Pilot component uses a set of molecular fragments created from commercially available molecules. For our purposes, only fragments containing heterocycles (and their alpha exocyclic atoms) were selected from this set. This selected set was then expanded by enumerating all possible regioisomers as described above. The expanded set of heterocyclic fragments was stored in a fragment cache.

For each user entered molecule, the *N* most similar fragments from the cache are selected. Default settings use FCFP_6, PHFC_2 fingerprints, and AlogP to evaluate similarity, but end-users may specify different properties. Similarities for fingerprints are evaluated using Tanimoto, and similarities of numeric properties are evaluated using a Euclidean distance function.[12,13]

MDDR Query. Regioisomers and bioisosteres generated based on the user's input molecule were used as substructure queries to search against the MDDR database (version 2013.2, hosted inside Biovia's Direct Chemical Cartridge). The initial query step is designed to count the number of substructural matched molecules in all possible phases of drug development (i.e., Launched, Phases I−III, Preclinical, etc.). This query only returns hit counts and IDs for each query structure in order to optimize search performance. The returned tabular results of the initial query step include hyperlinks on hit counts which allow users to drill down and retrieve structural and other data for individual phases of a given heterocyclic ring. The full search workflow is depicted in Figure 3.

Performance and Validation. To test the performance of the regioisomer generating component, we used the 24 867 one- and two-fused ring systems created by Pitt et al. as input.[7] For the first test, we applied Rule 3, consistent with the way Pitt et al. intended. This test resulted in the generation of 190 399 ring systems by the regioisomer generating component in ∼200 s on a Windows server using only one CPU. On average, ∼120 input molecules were processed per second, and ∼8 regioisomers were generated for each input molecule.

Since the set of one- and two-fused ring systems was created by Pitt et al. in a comprehensive and exhaustive manner, we also checked to see if there were any novel rings among the 190 299 output set not yet covered by the 24 867 input set. Our result confirmed that every record in the 190 299 output set is already present inside the input set of 24 867. Such an outcome is what should be expected if the set created by Pitt et al. is truly comprehensive and our atom-swapping method is generating valid regioisomers.

We performed a test using the same 24 867 input set but without the enforcement of Rule 3. This time 868 893 molecules were generated, of which 678 494 were unique

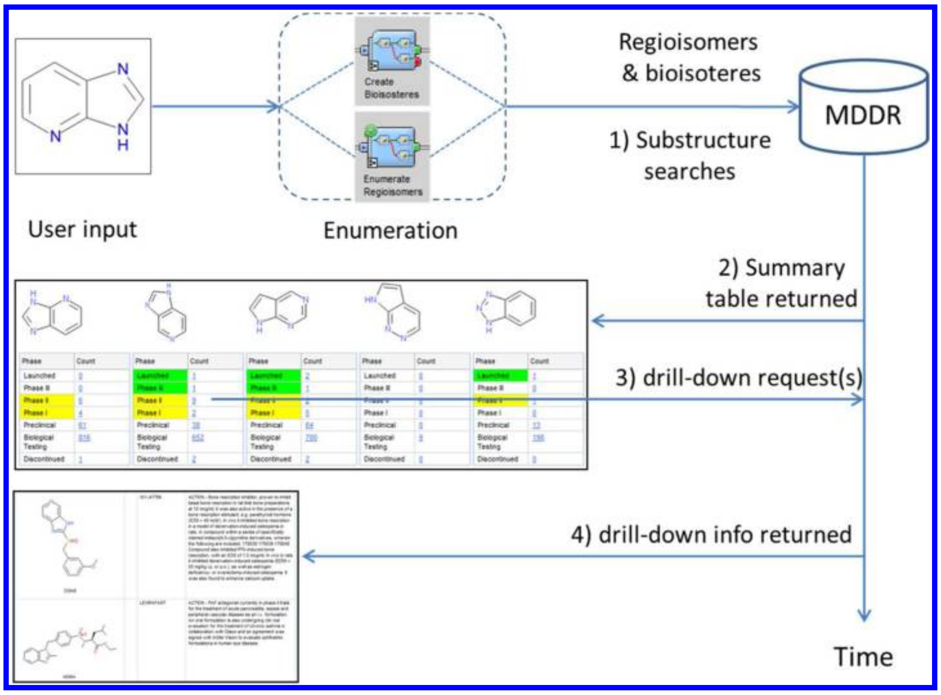
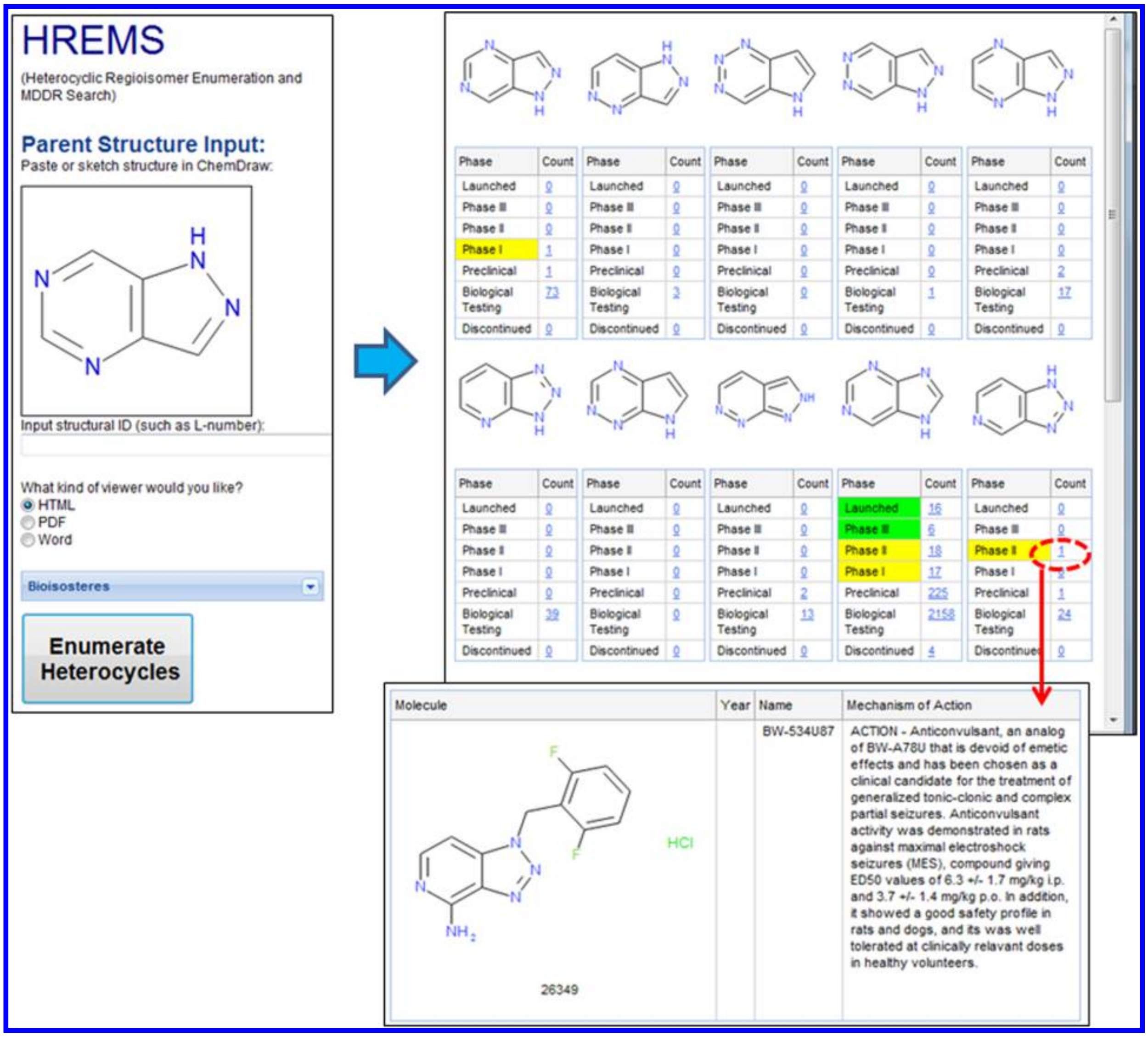


Figure 3. Cheminformatics workflow implemented by the HREMS tool.



HREMS

(Heterocyclic Regioisomer Enumeration and MDDR Search)

Parent Structure Input:

Paste or sketch structure in ChemDraw:

Input structural ID (such as L-number):

What kind of viewer would you like?

● HTML
○ PDF
○ Word

Bioisosteres ▾

Enumerate Heterocycles

Phase	Count	Phase	Count	Phase	Count	Phase	Count	Phase	Count
Launched	0	Launched	0	Launched	0	Launched	0	Launched	0
Phase III	0	Phase III	0	Phase III	0	Phase III	0	Phase III	0
Phase II	0	Phase II	0	Phase II	0	Phase II	0	Phase II	0
Phase I	1	Phase I	0	Phase I	0	Phase I	0	Phase I	0
Preclinical	1	Preclinical	0	Preclinical	0	Preclinical	0	Preclinical	2
Biological Testing	73	Biological Testing	3	Biological Testing	0	Biological Testing	1	Biological Testing	17
Discontinued	0	Discontinued	0	Discontinued	0	Discontinued	0	Discontinued	0

Phase	Count	Phase	Count	Phase	Count	Phase	Count	Phase	Count
Launched	0	Launched	0	Launched	0	Launched	16	Launched	0
Phase III	0	Phase III	0	Phase III	0	Phase III	6	Phase III	0
Phase II	0	Phase II	0	Phase II	0	Phase II	18	Phase II	1
Phase I	0	Phase I	0	Phase I	0	Phase I	17	Phase I	2
Preclinical	0	Preclinical	0	Preclinical	2	Preclinical	225	Preclinical	1
Biological Testing	39	Biological Testing	0	Biological Testing	13	Biological Testing	2158	Biological Testing	24
Discontinued	0	Discontinued	0	Discontinued	0	Discontinued	4	Discontinued	0

Molecule	Year	Name	Mechanism of Action
HCl 26349		BW-534U87	ACTION - Anticonvulsant, an analog of BW-A78U that is devoid of emetic effects and has been chosen as a clinical candidate for the treatment of generalized tonic-clonic and complex partial seizures. Anticonvulsant activity was demonstrated in rats against maximal electroshock seizures (MES), compound giving ED50 values of 6.3 +/- 1.7 mg/kg i.p. and 3.7 +/- 1.4 mg/kg p.o. In addition, it showed a good safety profile in rats and dogs, and its was well tolerated at clinically relavant doses in healthy volunteers.

Figure 4. Example output tables for a regioisomer search where the input structure and dialogue are shown on the left. A dropdown menu controls selection of bioisosteres. The table at the upper right provides counts of the phases of development for each regioisomer. As an example of the drill-down capability, the image on the lower right results from clicking on the hyperlink circled which then displays specifics such as the fully enumerated drug molecule(s).

(data not shown). This dramatic increase from 24 867 to 678 494 highlights the significance of imposing Rule 3.

The overall performance of the HREMS tool is satisfactory in the hands of users. As depicted inside Figure 3, the summary table and all subsequent drill-down tables are generated within a few seconds.

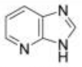
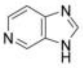
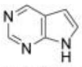
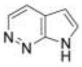
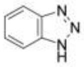
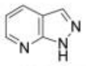
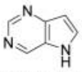
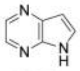
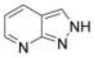
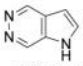
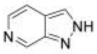
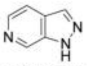
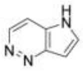
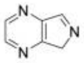
■ RESULTS AND DISCUSSION

The HREMS application is based on the pursuit of rational drug design and development. It is built on the concept that replacement of the molecular substructures with similar molecular substructures can potentially lead to a molecule with more desirable properties. The proposed workflow of the system is shown in Figure 3.

For a given input structure, the tool enumerates regioisomers for the heterocyclic rings inside the input molecule, and if requested, will provide potential bioisosteric replacements for the input structure. The resulting regioisomers and bioisosteres can be queried against any molecular database, exemplified here using MDDR to determine their frequency of use in various stages of drug development. The resultant data is reported to the user in a tabular format, highlighting the frequency counts greater than 0 for clinical trials and launched compounds. The data can be generated as a Word, PDF, or SD file. It allows the user to drill down on selected developmental phase combinations to see the individual structures contained within that combination. The user can further drill down upon these results to obtain the basic patent information on specific structures provided that the information is available in MDDR.

A snapshot of the results from a user's point of view is shown in Figure 4. Given an input structure shown at the left in Figure 4, the enumeration produces a series of regioisomeric structures shown on the right of the figure. The results of these regioisomers queried against the MDDR are shown on the right. Any input greater than 0 is color coded. The colors associated with the various stages of development are as follows: green (Phase III and Launched), yellow (Phase I and II) and red (discontinued). Clicking on the color coded region will pull up the complete structure and the patent associated

Table 1. Example of Generated Regioisomers

Regio-Isomer	Launched	Phase III	Phase II	Phase I	Preclinical	Biological Testing	Discontinued
 3*H*-imidazo[4,5-*b*]pyridine (1)	0	0	6	4	61	816	1
 3*H*-imidazo[4,5-*c*]pyridine (2)	1	1	9	2	38	652	2
 7*H*-pyrrolo[2,3-*d*]pyrimidine (3)	2	1	2	5	64	780	2
 7*H*-pyrrolo[2,3-c]pyridazine (4)	0	0	0	0	0	9	0
 1*H*-benzo[d][1,2,3]triazole (5)	1	0	1	0	13	198	0
 1*H*-pyrazolo[3,4-*b*]pyridine (6)	0	0	0	2	42	585	0
 5*H*-pyrrolo[3,2-*d*]pyrimidine (7)	0	0	1	0	10	149	0
 5*H*-pyrrolo[2,3-*b*]pyrazine (8)	0	0	1	0	7	98	0
 2*H*-pyrazolo[3,4-*b*]pyridine (9)	0	0	1	0	1	59	9
 1*H*-pyrrolo[2,3-*d*]pyridazine (10)	0	0	9	1	2	33	0
 2*H*-pyrazolo[3,4-c]pyridine (11)	0	0	0	0	3	69	0
 1*H*-pyrazolo[3,4-c]pyridine (12)	0	0	0	0	3	73	0
 5H-pyrrolo[3,2-c]pyridazine (13)	0	0	0	0	0	2	0
 5*H*-pyrrolo[3,4-*b*]pyrazine (14)	0	0	0	0	1	4	0

with that core in the MDDR database. As shown in Figure 4, the patent details include the mechanism of action found in the MDDR database; of course, output fields are completely customizable to support various uses. This output provides grounds for the chemist to make a rational decision on which regioisomer(s) to pursue.

We further demonstrate the utility of the tool by exploring the regioisomer enumeration of a fused bicyclic heterocycle 3*H*-imidazo[4,5-*b*]pyridine (1, Table 1). This example illustrates the scope and utility of the atom-swapping method to generate valid regioisomers. The output results from the automated search generated 14 fused bicyclic regioisomers that are categorized in various stages of testing (Launched, Phases I–III, Preclinical, Biological testing, or Discontinued status). Coupled with the bioisostere enumeration feature, a highly diversified library can be assembled that incorporates the best performing regioisomers and bioisosteres, based on historical clinical data which could facilitate synthetic strategy and maximize the potential for clinical success. Detailed information with respect to an API (active pharmaceutical ingredient) that has previously been incorporated in a specific regioisomer can be accessed through a drill down tool which provides the structure, the mechanisms of action, corresponding patent numbers, as well as manufacturer sources, in any phase of testing.

Ranking of comparative regioisomers can be particularly useful as a filter for selecting a subset of monomers for library selection. If only imidazopyridine regioisomers were targeted for enumeration, the corresponding 3*H*-imidazo[4,5-*c*]-pyridines (2) have historically outperformed the 3*H*-imidazo-[4,5-*b*]pyridine regioisomers (1) in the later Phase testing (12 versus 10 in Phases I−III; with 1 versus no launched drugs, respectively). If non-3*H*-imidazo pyridines were considered as potential monomers, an additional seven regioisomers were identified in later Phase testing (24 in Phases I−III; with 3 launched drugs). In particular, 7*H*-pyrrolo[2,3-*d*]pyrimidine (3) stands out as a highly desirable regioisomer that may offer unique chemical and physical properties with eight APIs in Phases I−III and two launched drugs to its credit. On the other hand, 2*H*-pyrazolo[3,4-*b*]pyrazine (4) regioisomers have a more disappointing profile with nine discontinued APIs and would require further assessment to determine if adverse effects or other liabilities were associated with a particular regioisomer's fate. With this knowledge in hand the medicinal chemist can make better choices during the regioisomer selection process to increase probability of success. We postulate that incorporation of the HREMS routine into a cheminformatics workflow can further leverage QSAR (quantitative structure−activity relationship) calculations of ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties or other in silico modeling. We additionally propose that the tool can help in analyzing patents for gaps in coverage of structural diversity.

■ CONCLUSION

A novel regioisomer enumeration web tool based on the concept of rational drug design and development was developed to assist chemists in the exploration of heterocyclic ring systems. For a given scaffold, the tool can enumerate all possible regioisomers, illustrate how these cores have fared in clinical testing using categories like Preclinical, Phases I, II, and III, Launched, and Discontinued. The tool can also enumerate bioisosteres for these cores. This tool is highly portable and can be further extended and incorporated into other molecular design tools. All of this data provides the chemist with the information needed to make prudent decisions in the selection of regioisomers with an expected higher probability of success.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Short Pipeline Pilot protocol (text file: save text file as .xml). It illustrates the basic workflow of regioisomer generation via atom-pair swapping as depicted inside Figure S1. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00162.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: sriram_tyagarajan@merck.com (S.T.).
*E-mail: chris.lowden@workflowinformatics.com (C.T.L.).

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the Structural Diversity, Substitution Patterns, and Frequency of Nitrogen Heterocycles among US FDA Approved Pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257−10274.

(2) Dalvie, D., Kang, P., Loi, C. M., Goulet, L., Nair, S., Eds. *Influence of Heteroaromatic Rings on ADME Properties of Drugs*; RSC Publishing: Cambridge, United Kingdom, 2010; pp 328−369.

(3) Lexwell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; McLay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface A Tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257−3274.

(4) 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*, 7011−7022.

(5) Ujvary, I.; Gyorffy, W.; Lopata, A. Fragment-Based Drug Design Using Stereoisomers. A Case Study of Analogues of the Phenol Group in the Bioster Database. *Acta. Pharm. Hung.* **2003**, *73*, 163−169.

(6) Ho, C. M.; Marshall, G. R. DBMAKER: A Set of Programs to Generate Three-Dimensional Databases Based Upon User-Specified Criteria. *J. Comput. Aided. Mol. Des.* **1995**, *9*, 65−86.

(7) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952−2963.

(8) Ward, R. A.; Kettle, J. G. Systematic Enumeration of Heteroaromatic Ring Systems As Reagents For Use In Medicinal Chemistry. *J. Med. Chem.* **2011**, *54*, 4670−4677.

(9) Tu, M.; Rai, B. K.; Mathiowetz, A. M.; Didiuk, M.; Pfefferkorn, J. A.; Guzman-Perez, A.; Benbow, J.; Guimaraes, C. R.; Mente, S.; Hayward, M. M.; Liras, S. Exploring Aromatic Chemical Space With NEAT: Novel And Electronically Equivalent Aromatic Template. *J. Chem. Inf. Model.* **2012**, *52*, 1114−1123.

(10) Rabal, O.; Amr, F. I.; Oyarzabal, J. Novel Scaffold Fingerprint (SFP): Applications in Scaffold Hopping and Scaffold-Based Selection of Diverse Compounds. *J. Chem. Inf. Model.* **2015**, *55*, 1−18.

(11) Rabal, O.; Oyarzabal, J. Biologically Relevant Chemical Space Navigator: From Patent And Structure-Activity Relationship Analysis To Library Acquisition And Design. *J. Chem. Inf. Model.* **2012**, *52*, 3123−3137.

(12) Gower, J. C., Ed. *Measures of Similarity, Dissimilarity, and Distance*; John Wiley and Sons: New York, 1985; Vol. *5*, pp 397−405.

(13) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983−996.