JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Exploring Aromatic Chemical Space with NEAT: Novel and Electronically Equivalent Aromatic Template

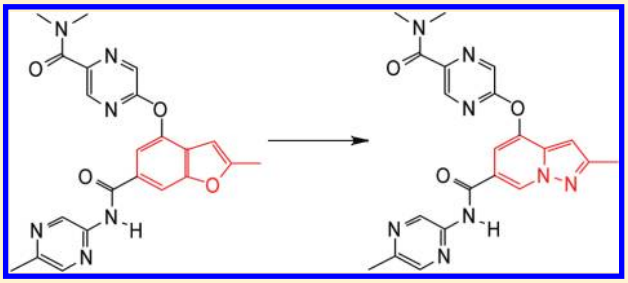
Meihua Tu,*[,†] Brajesh K. Rai,[‡] Alan M. Mathiowetz,[†] Mary Didiuk,[‡] Jeffrey A. Pfefferkorn,[†] Angel Guzman-Perez,[‡] John Benbow,[‡] Cristiano R. W. Guimarães,[†] Scot Mente,[‡] Matthew M. Hayward,[‡] and Spiros Liras[†]

[†]Pfizer Worldwide Research and Development, 620 Memorial Drive, Cambridge, Massachusetts 02139, United States

[‡]Pfizer Worldwide Research and Development, Eastern Point Road, Groton, Connecticut 06340, United States

Supporting Information

ABSTRACT: In this paper, we describe a lead transformation tool, NEAT (Novel and Electronically equivalent Aromatic Template), which can help identify novel aromatic rings that are estimated to have similar electrostatic potentials, dipoles, and hydrogen bonding capabilities to a query template; hence, they may offer similar bioactivity profiles. In this work, we built a comprehensive heteroaryl database, and precalculated high-level quantum mechanical (QM) properties, including electrostatic potential charges, hydrogen bonding ability, dipole moments, chemical reactivity, and othe properties. NEAT bioisosteric similarities are based on the electrostatic potential surface calculated by Brood, using the precalculated QM ESP charges and other QM properties. Compared with existing commercial lead transformation software, (1) NEAT is the only one that covers the comprehensive heteroaryl chemical space, and (2) NEAT offers a better characterization of novel aryl cores by using high-evel QM properties that are relevant to molecular interactions. NEAT provides unique value to medicinal chemists quickly exploring the largely uncharted aromatic chemical space, and one successful example of its application is discussed herein.



INTRODUCTION

The high attrition rate of clinical compounds in pharmaceutical research and development necessitates that drug discovery project teams deliver multiple differentiated clinical candidates with orthogonal risk factors. Therefore, the need to identify novel chemotypes is an important objective for many drug discovery project teams. Even though the size of small-molecule chemical space is immense (estimated to be greater than 10^{60} molecules),[1] the size of biologically relevant chemical space is much smaller. It is becoming increasingly difficult to identify novel and diverse chemical matter for highly competitive drug targets. This, for example, is evident for certain oncology kinase targets.[2]

Traditionally, the need for novel chemotypes was addressed through high-throughput screening of corporate compound collections, lead diversification by medicinal chemistry design practices such as bioisosteric replacement, and via virtual screening. In recent years, computational lead transformation (or lead hopping) tools have been successfully used to generate new lead matter.[3,4] These tools typically apply predefined medicinal chemistry design rules to transform a lead molecule to numerous possible follow-up molecules,[5] or replace specific moieties in a lead with viable isosteric replacements stored in a database.[6]

Aromatic rings play important roles in biochemical processes and are found in the majority of drug molecules.[7−9] In a drug molecule, an aryl ring can serve as a scaffold where variable substitution patterns can be exploited to interact with the biological target. An aryl ring can also interact directly with its biological target, through π interactions, hydrogen bond interaction, or van der Waals interactions. Such interactions are usually essential for a therapeutic molecule to achieve its efficacy and safety profile. It has been estimated that more than two-thirds of marketed drugs contain at least one aromatic core.[9] While aryl rings are heavily represented in drug molecules, the total number of unique aryl ring structures is relatively small. It has been estimated that there are only 120 unique aryl rings in all marketed drug molecules.[9] Considering that the overall aryl space is ~25 000 (*vide infra*),[9] it is clear that the aromatic space is very sparsely covered by marketed drugs.

Using heteroaryl rings as bioisosteres to lower lipophilicity is a documented practice in drug discovery,[10] driven by a deeper understanding of the liabilities associated with lipophilicity.[11] Recently, researchers from Pfizer reported that compounds with reduced lipophilicity (ClogP < 3, TPSA > 75) had significantly improved odds of surviving rigorous preclinical toxicology evaluation.[12] The impact of lipophilicity on compound survival has been carefully studied by medicinal

Received: January 17, 2012
Published: April 9, 2012

   dx.doi.org/10.1021/ci300031s | *J. Chem. Inf. Model.* 2012, 52, 1114−1123

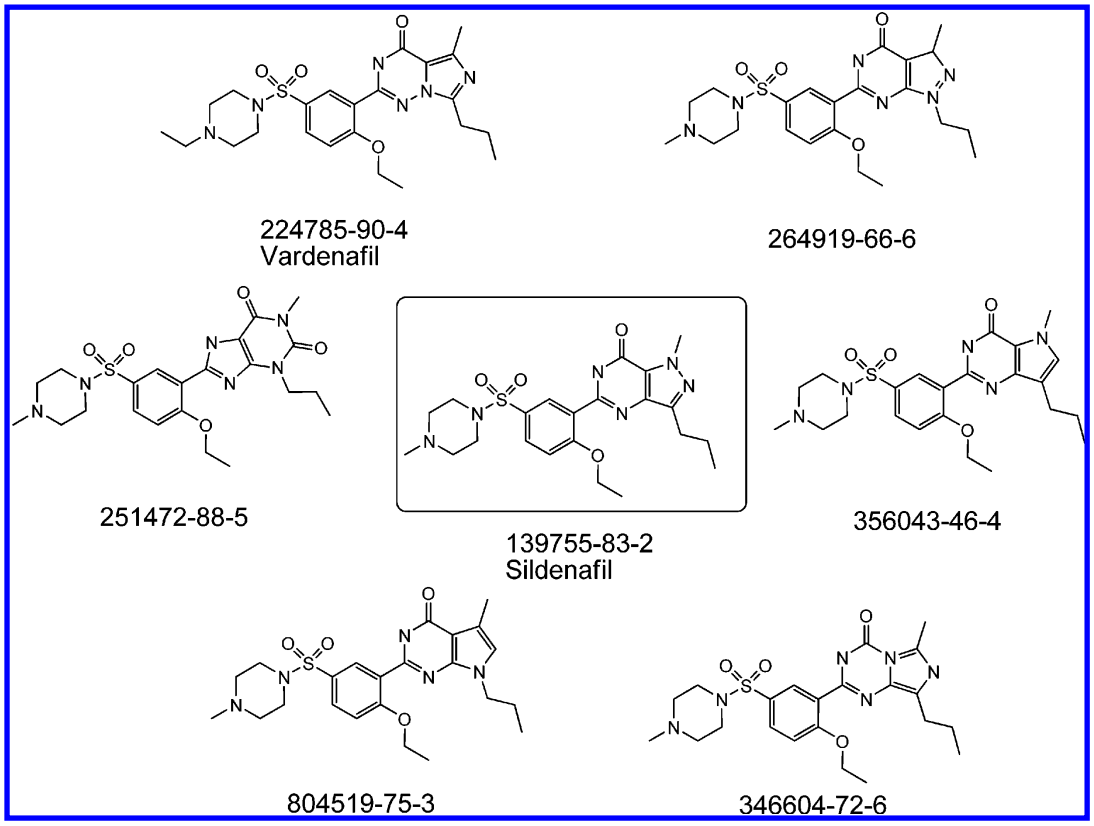


Figure 1. Heteroaryl bioisosteres of Sildenafil in the literature. The CAS registration number is indicated.

chemists[13,14] and useful concepts, such as ligand lipophilicity efficiency (LLE)[15] have emerged. In addition, it has been clearly demonstrated that reduced lipophilicity translates into improved pharmacokinetic properties such as clearance and other properties such as solubility.[10] Taken together, these concepts illustrate the importance of constructing molecules with reduced lipophilicity. Hence, the design and synthesis of diverse heterocyclic and heteroaromatic moieties is a key objective in contemporary medicinal chemistry, especially when considered in the context of conclusions that investment in heterocyclic synthetic chemistry is decreasing.[9,16]

An additional benefit of heteroaryl replacement is to increase structural novelty. One well-known example was the replacement of the pyrazolopyrimidine ring in Sildenafil with other aromatic bioisosteres (see Figure 1). This approach led to the discovery of a second drug, Vardenafil, and enabled the delivery of several others to the clinic.[17]

Even though aromatic ring structures are simple and their topological shapes are similar, their physicochemical properties are not. It is very common for two aryl rings of the same size, but differing by a single atom, to possess very different physicochemical and biological properties. Traditional similarity measurements based on structural fingerprints are not adequate to quantify such differences. The assessment of such differences through empirical means is usually challenging even to an experienced physical organic chemist. In order to find a successful heteroaryl bioisostere of an existing aromatic template, a drug discovery team often mus perform a systematic heteroatom walk on the template, until the desired heteroaryl bioisostere is found. The new heteroaryl rings are often very challenging synthetically, and many do not end up providing the desired properties, so the systematic approach can be very resource-intensive. Therefore, an efficient way to identify novel heteroaryl bioisosteres of an existing aryl template would have the potential to accelerate drug discovery. To our knowledge, there is no report in the literature of a method to achieve this.

In this paper, we describe a computational lead transformation tool—NEAT (**N**ovel and **E**lectronically equivalent **A**romatic **T**emplate)—that can help identify novel aromatic rings that are estimated to have similar electrostatic potentials, dipoles, and hydrogen-bonding capabilities to a query template, and, hence, may offer similar bioactivity profiles. NEAT uses the Brood engine (OpenEye Scientific Software, Inc., Santa Fe, NM) and precalculated quantum mechanical (QM) charges to conduct electrostatic potential calculations and comparisons. Unlike the commercially available Brood software, NEAT comprehensively covers the heteroaryl chemical space. In addition, it provides a better bioisosteric similarity measure of aromatic rings than Brood, through the use of QM properties.

METHODS

A Comprehensive Heteroaryl Ring Database. There have been several reports in the literature that describe the enumeration of chemical space. One well-known example is the enumeration of the GDB database.[18,19] It includes all feasible chemical structures up to 11 atoms, using the atoms C, N, O, and F. The database was generated by exhaustively enumerating mathematical graphs corresponding to simple hydrocarbons into which unsaturations and heteroatoms were introduced combinatorially. The database contains 26 million compounds (not including tautomers and stereoisomers), many of which are manageable with current computing power and storage. Since the GDB database is a comprehensive coverage of chemical space up to 11 atoms, it should include all possible monocyclic and bicyclic aryl cores. Hence, we decided to extract our heteroaryl database from the GDB. This approach is

dx.doi.org/10.1021/ci300031s | J. Chem. Inf. Model. 2012, 52, 1114−1123

different from the generation of the VEHICLe heteroaryl database.[9] All monocyclic (five- and six-membered) and bicyclic (five−five, five−six, and six−six) aromatic rings were retrieved via a substructure search of the GDB database using an in-house tool developed with OpenEye toolkits (OpenEye Scientific Software, Inc.) (see Table 1). Rings with substituents were excluded, except those with *N*-methyl and an exocyclic carbonyl group, since these substituents are expected to impart dramatically different properties relative to the parent N−H and des-carbonyl parent structures, respectively. Because the GDB database did not include sulfur atoms, we added sulfur-containing rings by computationally changing endocyclic oxygen (O) atoms to sulfur (S) atoms, and exocyclic carbonyl (C═O) groups to sulfone (SO_2) groups. We then removed rings that were flagged as "Reactive groups" or "Risky groups" using in-house computational filters,[20] for example, acyclic imines, pyridine oxides, etc. We also removed rings with <20% carbon content (except for tetrazole), in terms of the number of heavy atoms. After that, we generated all tautomers of the remaining rings using "Enumerate Tautomers" protocol in Pipeline Pilot (V8.0, Accelrys Software, Inc., San Diego, CA). Our final database included 25 131 unique heteroaryl rings. All structures in the database were first minimized in the Maestro software package (Version 9.0, Schrodinger, Inc., New York) with a molecular mechanics force field (OPLS_2005), then in Jaguar with a HF 3-21G* basis set.

Table 1. Substructure Search of the 26 Million Compound GDB Database

SMARTS query	maximum number of heavy atoms	number of hits	final hits after cleanup and transformation	comments
a1aaaa1	7	1299	123	5-member aryl rings
a1aaaaa1	8	1391	192	6-member aryl rings
a1aaa2aaaa21	10	18529	3809	5,5-bicyclic aryl rings
a1aaa2aaaaa21	11	32045	16212	5,6-bicyclic aryl rings
a1aaaa2aaaaa21	11[a]	2089	4856	6,6-bicyclic aryl rings

[a]The maximum number of atoms in a structure in GDB is 11.

We compared the heteroaryl database with Pfizer's compound collection and an external compound collection (which was the combination of 40 million compounds from the PubChem database[21] and 8 million compounds from the IBM patent database[22]). Of the 25 131 unique heteroaryl rings, only 2444 of them are represented in the Pfizer database, and only 3475 were found in the external compound collections (Figure 2). The low number of unique heteroaryl rings in known compounds is consistent with what has been reported,[9] even though a significantly larger known-compound database was used in our analysis. Interestingly, there are 1294 heteroaryl rings existing in the external compound collection that do not exist in Pfizer's compound collection. If Pfizer and external compound databases are representative of typical corporate compound collections, then the comprehensive aryl database here clearly demonstrates that aromatic chemical space is under-represented in most companies' sample collections by a large degree.

Descriptor Calculations. While an aryl ring structure is relatively simple (any conjugated ring system that follows Hückel's $4n + 2$ π-electron rule), its physicochemical properties are usually nonintuitive. This is largely due to the delocalization of the π electrons. Such electronic contributions to physicochemical properties can be well-characterized by high-level QM descriptors.[23] We calculated the following atomic (and molecular) properties, which may be relevant to an aryl ring's molecular interactions with biological systems.

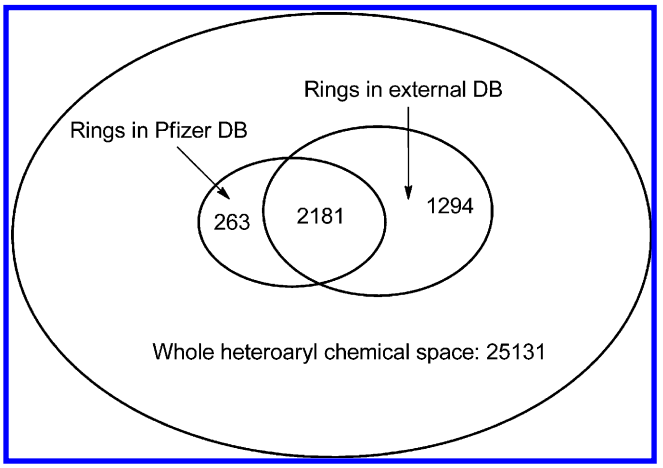


Figure 2. Structure comparison of the comprehensive heteroaryl database, Pfizer's compound collection, and the external compound database.

Electrostatic Potential (ESP) Charge and Dipole Moment. The electrostatic potential (ESP) gives the interaction energy of a molecule with a unit positive charge in three-dimensional (3D) space. ESP charges approximate the effective charge on each atom by fitting atom-centered point charges to the quantum mechanical electrostatic potential.[24] A dipole moment is a measure of the separation of positive and negative charges in a molecule. It is relevant to a molecule's overall polarity. ESP charges and dipole moments may provide a quantitative measure of the electrostatic interactions of a drug with its protein target.

ESP charges and dipole moments of each heteroaryl ring were calculated using the Jaguar module in the Maestro software package (Version 9.0, Schrodinger, Inc., New York) at the HF/6-31G* level of theory. Figure 3a shows a histogram for the dipole moments of all heteroaryl rings. It can be readily seen that most heteroaryl rings have dipole moment values between 3.0 debye and 7.0 debye. The median dipole moment of the heteroaryl database is 4.7 debye. Based on this calculation, for example, pyrrole has a dipole moment of 1.8 debye and imidazole has a dipole of 3.8 debye. Dipole moments for the entire database are available in the Supporting Information.

Hydrogen Bond Strength. Hydrogen-bond interactions play important roles in molecule recognition in a biological system. In order to achieve high binding affinity and selectivity in a biological system, a heteroraryl ring must overcome its desolvation penalty by forming favorable hydrogen-bond interactions with its biomolecular target. Hence, quantification of hydrogen-bond strength is of considerable practical relevance in helping to achieve an optimal hydrogen-bond interaction.

Several factors affect hydrogen-bond strength, including electrostatic interactions (ES), polarizability (PL), and charge transfer (CT).[25] There are many theoretical and empirical methods measuring relative hydrogen-bond donor/acceptor strengths.[25,26] In our approach, we use atomic hydrogen bond moments from the COSMOTherm program (COSMOlogic

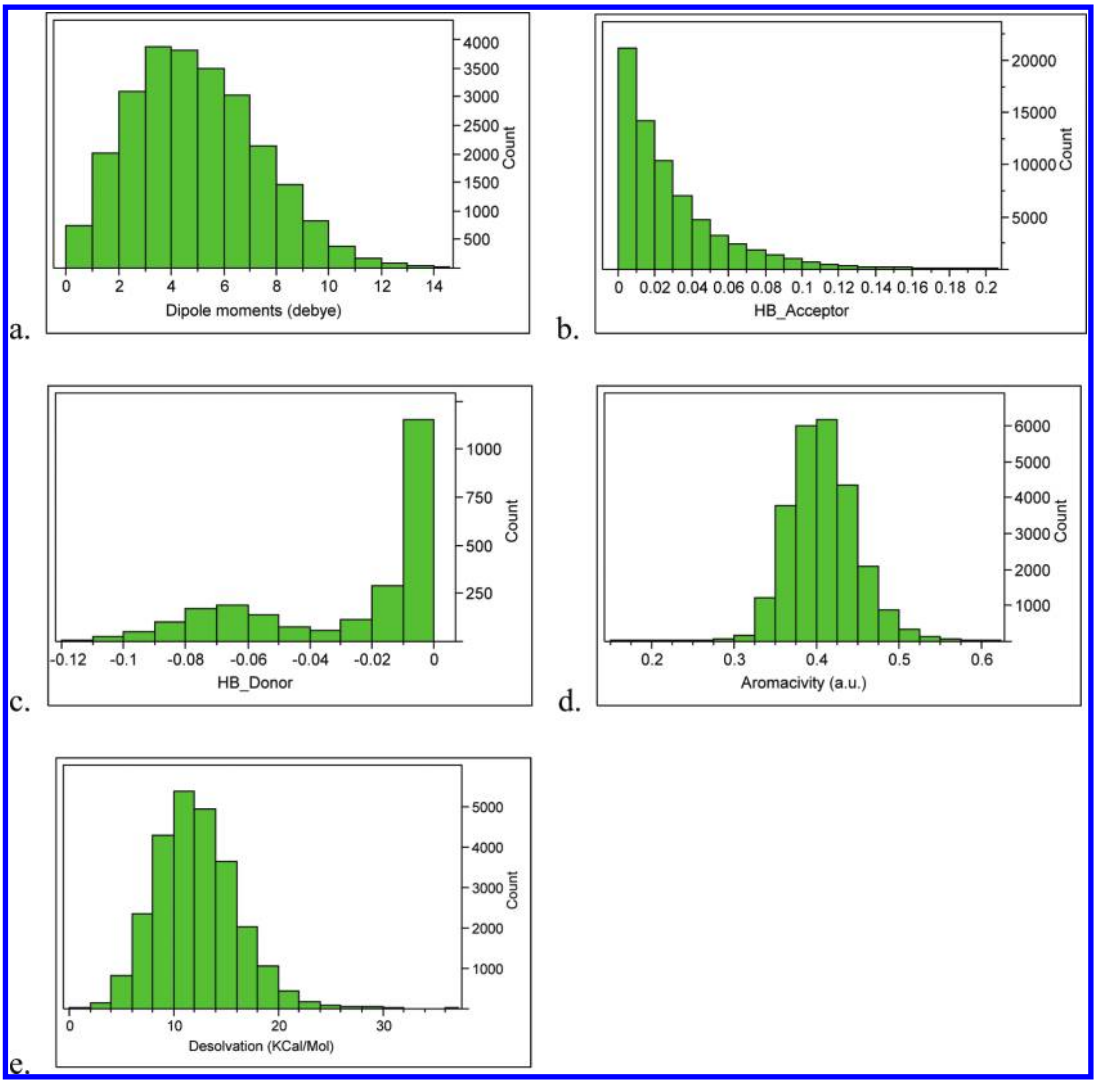


Figure 3. Histograms of calculated properties: (a) dipole moments, (b) hydrogen-bond acceptor moments, (c) hydrogen-bond donor moments, (d) aromaticity, and (e) desolvation energy.

GmbH & Co. KG, Leverkusen, Germany). Hydrogen-bond moments incorporate solvent effects into the density functional calculations.[27] They provide better quantitative measures of the hydrogen-bond acceptor and donor capacities in our internal evaluation.

Atomic hydrogen-bond moments were calculated with the COSMOTherm program (V. C21_0108), using the BP-TZVP-COSMO level of theory. Figures 3b and 3c are histograms for the calculated hydrogen-bond acceptor and donor moments of all hetero atoms in the database, respectively. A negative sign is assigned to a hydrogen-bond donor moment in order to differentiate it from a hydrogen-bond acceptor moment. A larger absolute value of the hydrogen-bond moment corresponds to a stronger hydrogen-bond donor or acceptor capability. It is readily seen from the graphs that the number of hydrogen-bond acceptors is far greater than the number of hydrogen-bond donors. Most hydrogen-bond acceptor moments are less than 0.2. Most hydrogen-bond donor moments are greater than −0.02. From our calculation, for example, N1 in an imidazole ring has a hydrogen-bond acceptor moment of 0.078 and N3 has a hydrogen-bond donor moment of −0.052. In comparison, O1 in an oxazole ring has a hydrogen-bond acceptor moment of 0.0, and N3 has a hydrogen-bond acceptor moment of 0.042. Hydrogen-bond moments of all hetero atoms in the database are available in the Supporting Material.

Fukui Indices. Fukui indices were derived from density functional theory (DFT) based on frontier orbital theory[28] by Parr and Yang.[29] Fukui indices are widely used in estimating chemical reactivity and metabolic stability.[30,31] In an N-electron system with the electron density $\rho(r)$ at constant external potential $\nu(r)$, the Fukui function is defined as[31]

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N}\right)_{\nu(r)} \quad (1)$$

The Fukui function provides three reaction indices, which correspond to the electrophilic reactivity index $(f^{+}_{(r)})$, the nucleophilic reactivity index $(f^{-}_{(r)})$, and the free-radical reactivity index $(f^{0}_{(r)})$, respectively.

The three Fukui indices of each molecule in the heteroaryl database were calculated using the DMOL3 module in PipelinePilot (V8.0, Accelrys Software, Inc., San Diego, CA), using the BLYP functional. The data are available in the Supporting Material.

Aromaticity. The HOMO (highest occupied molecular orbital)−LUMO (lowest unoccupied molecular orbital) energy

dx.doi.org/10.1021/ci300031s | J. Chem. Inf. Model. 2012, 52, 1114−1123

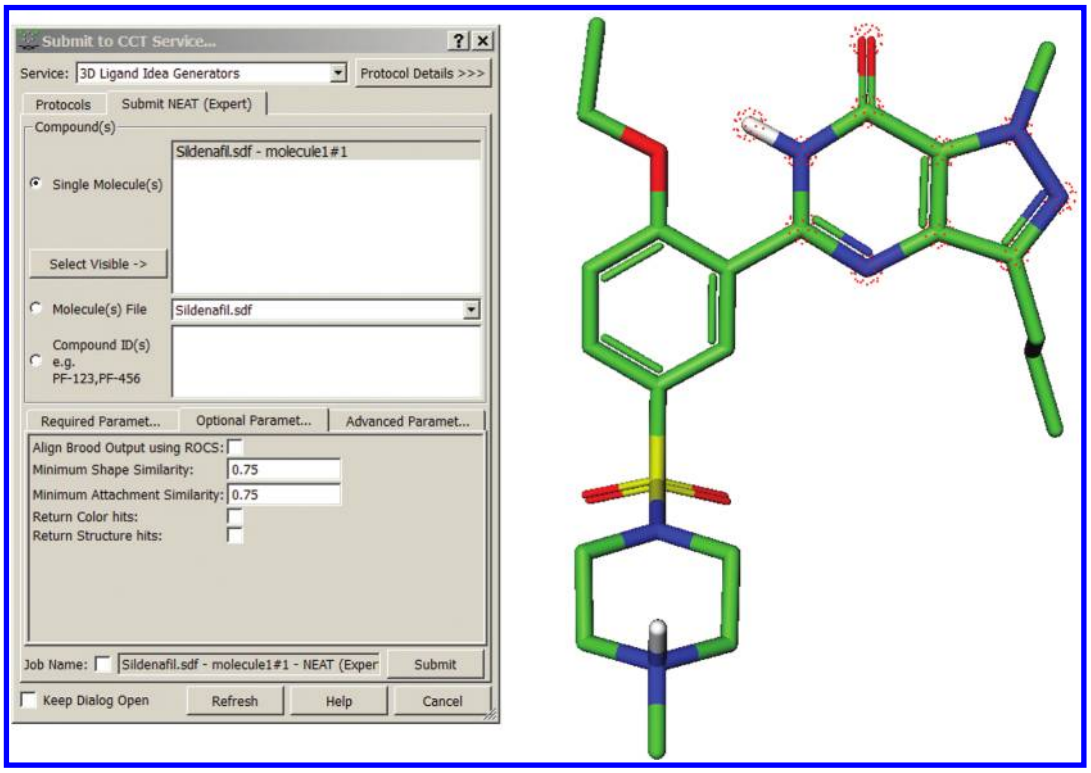


Figure 4. Graphical user interface (GUI) of NEAT (Novel and Electronically equivalent Aromatic Template). Query atoms are denoted by red dots, and hydrogen-bond requirements are denoted in yellow. Additional parameters from Brood can also be entered through this GUI interface.

gap is a measure of the energy needed to activate a molecule from the ground (most stable) state into an excited state. Frank et al. have demonstrated its connection to the thermodynamic and kinetic criteria of the aromatic character.[32] The HOMO−LUMO gap has also been commonly used as a measure of resonance energy in an aromatic system. A large HOMO−LUMO gap has been demonstrated to be associated with stable structures.[33] In our implementation, HOMO and LUMO energies of each aromatic ring were calculated with the Jaguar module in the Maestro software package, using the HF/6-31G* level of theory.

Figure 3d is a histogram of the calculated aromaticity in the database. It is apparent from the graph that most aryl rings have aromaticity values between 0.3 au and 0.5 au. As a reference, a benzene ring has an aromaticity value of 0.480 au and a thiophene has an aromaticity value of 0.465 au. Aromaticity values of all heteroaryl rings are available in the Supporting Material.

Desolvation Energy. Desolvation energy is the energy required to move a drug molecule from solvent (usually water) into vacuum. When a drug molecule binds to its protein target, it first must overcome such an energy barrier, before it interacts with its molecular target. Therefore, desolvation energy is a good measure of a molecule's availability in the biological medium. In our implementation, molecules in the heteroaryl database were minimized using the OPLS_2005 force field,[34,35] combined with the Generalized Born/Surface Area (GB/SA) continuum solvation model,[36] using the MacroModel module in the Maestro software package. The desolvation energies for the heterocycles were taken as the negative of their GB/SA solvation energies. Figure 3e is a histogram of the calculated desolvation energies for the database. It is readily seen from the graph that heteroaryl rings have a wide range of desolvation

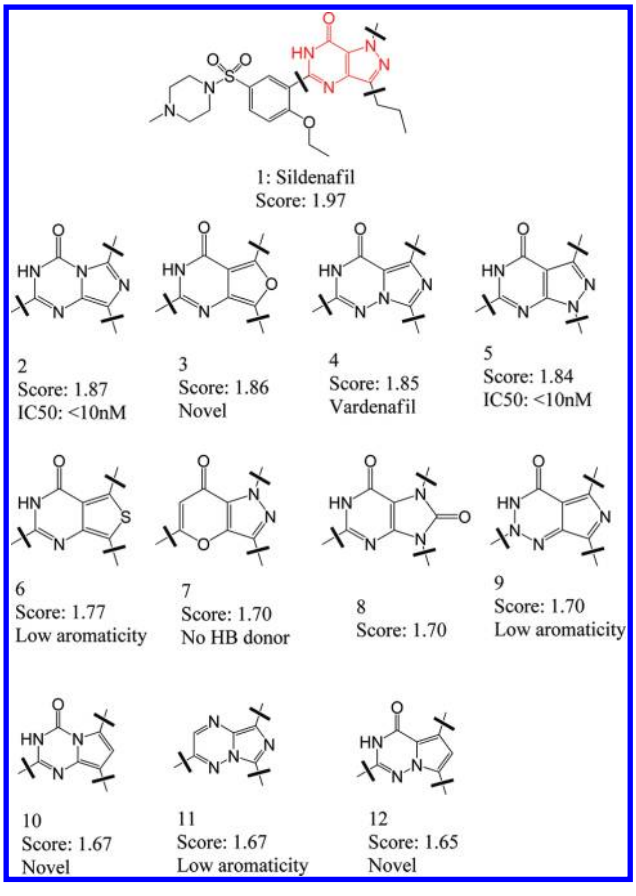


Figure 5. Top 12 hits from NEAT.

dx.doi.org/10.1021/ci300031s | J. Chem. Inf. Model. 2012, 52, 1114−1123

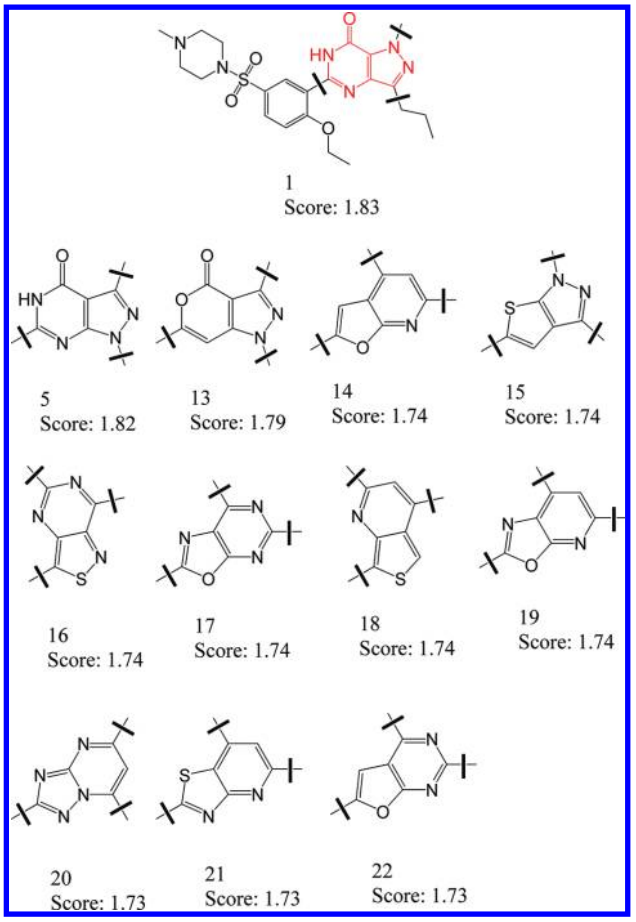


Figure 6. Top 12 hits from Brood.

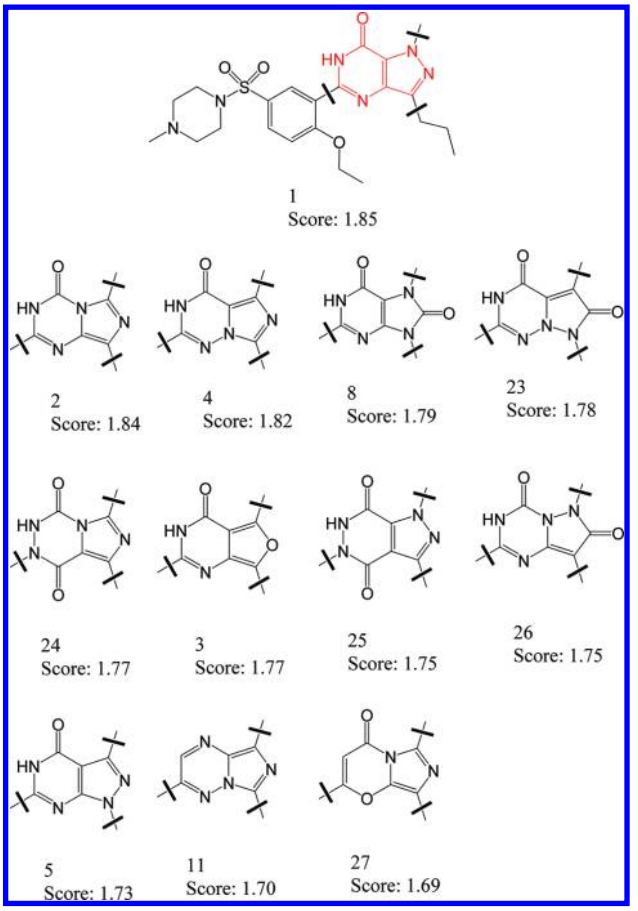


Figure 7. Top 12 hits from NEAT with MMFF charges.

energies that span from 1 kcal/mol to 30 kcal/mol. Such an energy difference is much larger than the binding free energy window that a typical biological assay covers.[37] This clearly demonstrates the importance of considering desolvation energy in bioisostere replacement. From our calculations, as an example, a benzene ring has a desolvation energy of 0.84 kcal/mol and an imidazole ring has a desolvation energy of 11.18 kcal/mol. Desolvation energy values of all heteroaryl rings are available in the Supporting Material.

Electrostatic Potential Similarity and Implementation. One of the most important characteristics of an aromatic ring is the negative electrostatic potential on both faces of its planar ring. This negative potential is one of the determining factors of the strength of its interactions with other molecules. Therefore, comparing such potentials is critical in finding a bioisostere of an aromatic ring. Here, NEAT uses the Brood engine from OpenEye to perform compound alignments and conduct electrostatic potential surface calculations and comparisons. Brood carries out a Poisson−Boltzmann calculation to generate the electrostatic potential of each molecule. The similarity between the query fragment and the database is measured by the overlap of electrostatic potentials and of the attachment points on the fragment. By default, Brood uses molecular mechanical force field (MMFF) partial charges in the electrostatic potential calculation. Such a charge model might not work well with novel aromatic ring structures, as demonstrated in the examples below. For the NEAT method, we decided to use high-level QM ESP charges in conjunction with the Brood electrostatic potential overlap calculations.

Internally, at Pfizer, we have made this method available to our colleagues as a desktop workflow, which leverages our internal desktop design and visualization applications as well as our scientific computing infrastructure. As shown in Figure 4, a query is set up by specifying the fragment for bioisosteric replacement. In addition, the GUI of this workflow enables users to specify hydrogen-bond donor and/or acceptor atoms on the selected fragment that may be important for activity. While the specification of hydrogen-bonding atoms has no influence on the search process, it can be used to prioritize the database hits. For example, in some cases, it might be necessary to retain or strengthen the hydrogen-bonding characteristics of an atom at certain position on the fragment. By specifying such atoms, this workflow produces additional output comparing the relative strengths of hydrogen bonds of atoms in the database versus the query fragment, thereby enabling post-processing of the hit list. In addition, the search parameters can be specified using the same GUI. Input files and parameters are preprocessed using in-house python scripts, before using the Brood software program for searching the database. Molecules are aligned by Brood shape- and color-based optimization. An initial hit list is produced based on Brood shape and attachment point similarity cutoffs of 0.75. The hit list compounds will be further evaluated for their electrostatic similarity to the query fragment using the precomputed ESP atomic charges. The hits identified by this step are rank-ordered by decreasing electrostatic similarity as indicated by the "etcombo score" (the sum of the shape and the electrostatic Tanimoto scores) produced by Brood. An "etcombo score" of 2.0 means a perfect match between query and hit. Based on our visual inspection of

the results, a minimal "etcombo score" of 1.5 is necessary for an appealing bioisostere. This workflow also returns additional data which are used to post-process the top hits by "etcombo score". The additional data consists of (1) the relative magnitude and angle between the electrical dipole moments of the query and the database fragments, (2) the relative hydrogen-bonding potential of the user-specified donor and acceptor atoms, (3) the Fukui index of the atoms of interest, (4) the aromaticity of the hits, and (5) the desolvation energy of the entire molecule.

The workflow described above masks the complexity associated with the electronic structure similarity assessments and enables such calculations to be performed by a wide range of scientists.

RESULTS AND DISCUSSION

Evaluation of NEAT Methodology. There are two distinctive features in NEAT that make it better for identifying heteroaryl bioisosteres. One is the comprehensive coverage of heteroaryl chemical space; and the other is a better way of characterizing aromatic similarities using QM properties. To test the merits of our methodology, we used the pyrazolopyrimidine core (denoted in red in Figure 4) in Sildenafil as a query. We compared NEAT results to those from commercial Brood software (Version 1.1.2 with Brood large database, OpenEye Scientific Software, Inc., Santa Fe, NM), and to those from NEAT without using QM properties (i.e., only using MMFF charges). The goal was to determine (1) if NEAT would retrieve more heteroaryl cores than Brood that had been shown to successfully replace pyrazolopyrimidine, (2) if NEAT could identify other novel heteroaryl cores that had not been published in the literature or predicted by Brood, and (3) if QM charges (and other QM properties) helped NEAT in finding heteroaryl bioisosteres.

Figures 5, 6, and 7 show the top 12 bioisosteres of the pyrazolopyrimidine core in Sildenafil from NEAT, Brood, and NEAT with MMFF charges, respectively. (All hits with scores greater than 1.5 are available in the Supporting Material.) In Figure 5, the top scoring compound, compound **1**, is Sildenafil itself. A Vardenafil analog (compound **4**) is the fourth-best compound in the list. Compounds **2** (second) and **5** (fifth) were reported with an IC_{50} of <10 nM in a PDE5 inhibition assay.[17] In Figure 6, Brood only identified two known active PDE5 inhibitors (compounds **1** and **5**) and missed the other two, probably because of the lack of corresponding cores in its database. In Figure 7, without using QM charges, NEAT still retrieved the four known active PDE5 inhibitors: compounds **1**, **2**, **4**, and **5**). However, Compound **5** ranked very low. Another difference between using QM charges and MMFF charges is that many more hits (with scores greater than 1.5) were retrieved from using MMFF charges than those from using QM charges. These results clearly demonstrate the value of a comprehensive heteroaryl database in retrieving all known active pyrazolopyrimidone bioisosteres. The results also show that QM charges do a slightly better job in differentiating heteroaryl bioisosteres than MMFF charges.

The other eight top hits (compounds **3** and **6**–**12**) in Figure 5 have not been reported in the literature as PDE5 inhibitors, based on a substructure search of SciFinder; therefore, they might have potential as novel PDE5 inhibitors.

The eight potential novel hits can be further rank-ordered based on other QM properties from NEAT (see Table 2). One key property is the stability of the heteroaryl ring based on aromaticity values. Compounds **6**, **9**, and **11** have relative low aromaticity values; therefore, they might have chemical stability (and synthetic) issues. Another important factor to consider is hydrogen-bond interactions. Based on the protein X-ray structure,[38] the pyrazolopyrimidine core forms a pair of hydrogen-bond interactions with the PDE5 protein. Hence, we might eliminate compounds **7** and compound **11**, since they do not have the required hydrogen-bond donor. This would leave compounds **3**, **10**, and **12** as viable bioisosteres of Sildenafil. The heteroaryl core in compound **10** has never been reported in any compound in the external database previously mentioned. This suggests a high potential for novelty or possibly increased synthetic challenges. In contrast, the heteroaryl cores in compounds **3** and **12** have been made internally or externally, so the risk of synthetic difficulty and chemical stability is lower. There are other properties that one might use as a consideration to select the final synthetic target to pursue, such as dipole moment difference and desolvation energy.

Table 2. Other Atomic/Molecular Properties of the Top 12 Hits from NEAT

compound	score[a]	HB[b]	PF[c]	Ext[d]	Aro[e]	dipole[f]	DA[g]
1	1.97	1	Y	Y	0.422	2.75	0.31
2	1.87	1.21	Y	Y	0.412	4.50	8.09
3	*1.86*	*0.93*	*N*	*Y*	*0.407*	*1.95*	*26.3*
4	1.85	1.08	Y	Y	0.437	4.81	12.9
5	1.84	0.97	Y	Y	0.436	4.27	27.8
6	1.77	0.92	Y	Y	0.388	1.79	33.2
7	1.70	0.00	N	N	0.426	2.88	42.4
8	1.70	0.99	Y	Y	0.402	5.74	7.24
9	1.70	0.00	N	Y	0.349	3.69	35.2
10	*1.67*	*1.29*	*N*	*N*	*0.401*	*7.24*	*12*
11	1.67	1.09	Y	Y	0.351	2.18	11.9
12	*1.65*	*0.98*	*Y*	*Y*	*0.430*	*3.02*	*39.4*

[a]The term "Score" denotes the "etcombo score" from Brood. [b]"HB" represents the relative hydrogen-bond strength to the query molecular. It was calculated as the ratio of hydrogen bond strength between the hit and the query. [c]The term "PF" denotes whether or not the core existed in the Pfizer compound collection. [d]The term "Ext" denotes whether or not the core existed in the external compound database. [e]The term "Aro" represents aromaticity; it was calculated as LUMO energy less the HOMO energy (in a.u.). [f]The term "Dipole" represents the relative scale of dipole moments, relative to the query molecule. It was calculated as the ratio of dipole moments (in debye) between the hit and the query. [g]The term "DA" denotes the angle of the dipole moments between the query molecule and the hit. Molecule alignment was done using Brood shape and color based optimization.

All these additional properties in NEAT make it possible to further differentiate the heteroaryl bioisosteres, which are not available in other software.

Application Example: Replacing the Benzofuran Core in the Glucokinase Activator Project. In an effort to replace the benzofuran core of compound **28**[39] in our Glucokinase activator project, we initially tried benzothiophene, benzoxazole, and reversed benzofuran templates; unfortunately, all of these resulted in a significant loss of potency (see Figure 8).

Using the benzofuran core (i.e., compound **28**) as an input to NEAT, we were able to identify over 30 rings with a score greater than 1.5 (Figure 9 has the top 12 of these). Not surprisingly, some of the top scoring templates had already been being investigated by the team. These include benzothiophene (compound **29**) and indazole (compound **31**) cores. Indeed, the indazole analog demonstrated comparable potency and superior physicochemical properties. However, there were also a couple of

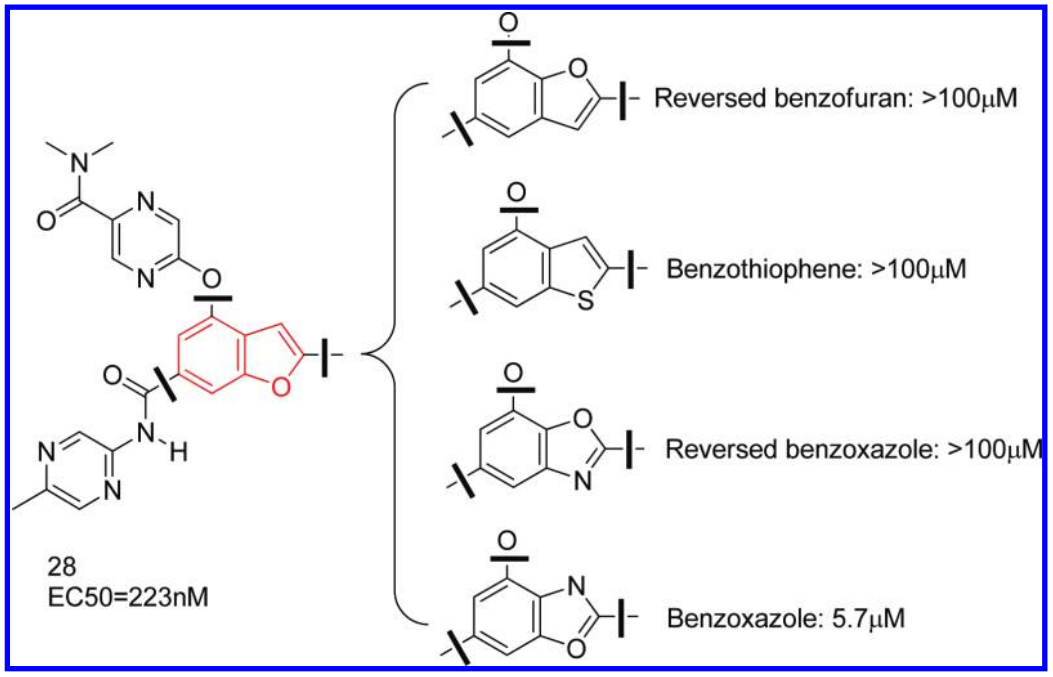


Figure 8. Benzofuran replacements.

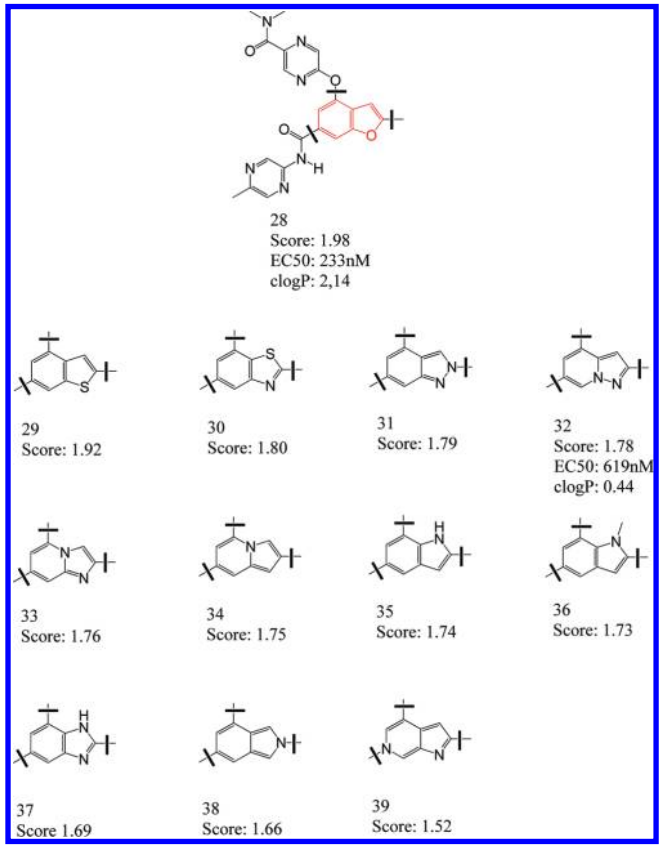


Figure 9. Top 12 hits from NEAT for the benzofuran core.

new templates that had not been actively considered, such as pyrazolopyridine (compound **32**) and imidazopyridine (compound **33**). The team decided to pursue these two templates because of their high electrostatic potential similarity to the benzofuran core and better calculated properties. The pyrrazolopyridine analog was synthesized and was found to retain appreciable affinity for the target while exhibiting much lower lipophilicity (compound **32** in Figure 9). The synthesis of imidazopyridine was unsuccessful due to stability issues. This is consistent with the imidazopyridine core having a lower aromaticity value (0.387 au) than that of the pyrazolopyridine core (0.397). The instability of the imidazopyridine analog might also be due to the substitution effects on the core which are not currently considered in the NEAT methodology.

CONCLUSION

While aromatic rings are prevalent in drug molecules, the number of unique aryl rings in marketed drugs is relatively small. The comprehensive heteroaryl database discussed here provides drug discovery team opportunities to access this unexplored chemical space. Aryl ring structures often look simple and similar. However, their properties can be substantially different chemically and biologically. A small change in the heteroatoms on a ring may dramatically impact its synthetic feasibility, physicochemical properties, and biological activities. We have shown that NEAT (*N*ovel and Electronically equivalent Aromatic Template), which performs quantitative rank ordering of aryl rings, using a comprehensive description of its electronic character, has the potential to facilitate efficient exploration of the under-represented heteroaryl chemical space.

Because of their consistency and conformer-independency, molecular mechanics force field (MMFF) charges have been demonstrated by Openeye to be the best for general purpose Brood isobiosteric search. However, when comparing aromatic rings, conformational variation is no longer an issue. A charge model that can better capture physical properties is more important. Our example demonstrates that high-level quantum mechanical (QM) charges and QM properties work better than MMFF charge only in rank-ordering aromatic ring's electrostatic potential. It is important to mention that both QM and MMFF charge-based electrostatic potential are approximations of the true electrostatic potential of molecular orbitals. Sometimes, an *ab initio* calculation is necessary to truly characterize a novel aryl ring's electronic properties.

dx.doi.org/10.1021/ci300031s | *J. Chem. Inf. Model.* 2012, 52, 1114−1123

It should also be noted that NEAT was specifically designed to find aromatic replacements. In the case of finding non-aromatic bioisosteres, readers are encouraged to look into other commercial packages, such as Brood, Pipeline Pilot, and MOE (Chemical Computing Group Inc., Montreal, QC, Canada).

The productivity challenge[40] in pharmaceutical research and development calls for innovative approaches. NEAT is one tool now available to scientists undertaking the herculean effort of identifying compounds with optimal physicochemical properties and biological activities.

■ ASSOCIATED CONTENT

Supporting Information

The comprehensive heteroaryl database is provided in the format of a zipped SDF file. The rest of the top scored hits in Figures 5−7 are provided in the format of SMILES code. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: 617-5513548. E-mail: meihua.tu@pfizer.com.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3−50.

(2) Bamborough, P.; Drewry, D.; Harper, G.; Smith, G. K.; Schneider, K. Assessment of Chemical Coverage of Kinome Space and Its Implications for Kinase Drug Discovery. *J. Med. Chem.* **2008**, *51*, 7898−7914.

(3) Lewell, 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) 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*, 171−178.

(5) 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.

(6) Corkery, J. J.; Schmidt, K. E.; Kelley, B.; Skillman, A. G. In *Abstracts of Papers, 239th ACS National Meeting*, San Francisco, CA, March 21−25, 2010; American Chemical Society: Washington, DC, 2010; Paper No. CINF-71.

(7) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.

(8) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. Aromatic interactions. *J. Chem. Soc., Perkin Trans.* **2001**, *2*, 651−669.

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

(10) Dalvie, D.; Kang, P.; Loi, C. M.; Goulet, L.; Nair, S. Influence of Heteroaromatic Rings on ADME Properties of Drugs. In *Metabolism, Pharmacokinetics and Toxicity of Functional Groups*; Smith, D. A., Ed.; RSC Publishing: Cambridge, United Kingdom, 2010; pp 328−369.

(11) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A Comparison of Physiochemical Property Profiles of Development and Marketed Oral Drugs. *J. Med. Chem.* **2003**, *46*, 1250−1256.

(12) Hughes, J. D.; Blagg, J.; Price, D. A.; Bailey, S.; Decrescenzo, G. A.; Devraj, R. V.; Ellsworth, E.; Fobian, Y. M.; Gibbs, M. E.; Gilles, R. W.; Greene, N.; Huang, E.; Krieger-Burke, T.; Loesel, J.; Wager, T.; Whiteley, L.; Zhang, Y. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872−4875.

(13) Edwards, M. P.; Price, D. A. Role of Physicochemical Properties and Ligand Lipophilicity Efficiency in Addressing Drug Safety Risks. *Annu. Rep. Med. Chem.* **2010**, *45*, 381−391.

(14) Leeson, P. D.; Empfield, J. R. Reducing the Risk of Drug Attrition Associated with Physicochemical Properties. *Annu. Rep. Med. Chem.* **2010**, *45*, 393−407.

(15) Leeson, P. D.; Springthorpe, B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discovery* **2007**, *6*, 881−890.

(16) Ward, E. A.; Kettle, J. G. Systematic enumeration of heteroaromatic ring systems as reagents for use in medicinal chemistry. *J. Med. Chem.* **2011**, *54*, 4670−4677.

(17) Haning, H.; Niewoehner, U.; Schenke, T.; Lampe, T.; Hillisch, A.; Bischoff, E. Comparison of different heterocyclic scaffolds as substrate analog PDE5 inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3900−3907.

(18) Blum, L. C.; Reymond, J. L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *J. Am. Chem. Soc.* **2009**, *131*, 8732−8733.

(19) Fink, T.; Reymond, J. L. Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery. *J. Chem. Inf. Model.* **2007**, *47*, 342−353.

(20) Blake, J. F. Identification and Evaluation of Molecular Properties Related to Preclinical Optimization and Clinical Fate. *Med. Chem.* **2005**, *1*, 649−655.

(21) Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217−241.

(22) Chen, Y.; Spangler, W.; Kreulen, J. SIMPLE: A Strategic Information Mining Platform for Licensing and Execution. In *IEEE International Conference on Data Mining Workshops*, CPS: Miami, FL, 2009; pp 270−275.

(23) Alonso, M.; Herradon, B. A Universal Scale of Aromaticity for π-Organic Compounds. *J. Comput. Chem.* **2009**, *31*, 917−928.

(24) Chirlian, L. E.; Francl, M. M. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* **1987**, *8*, 894−905.

(25) Schwobel, J.; Ebert, R. U.; Kuhne, R.; Schuurmann, G. Prediction of the Intrinsic Hydrogen Bond Acceptor Strength of Organic Compounds by Local Molecular Parameters. *J. Chem. Inf. Model.* **2009**, *49*, 956−962.

(26) Abraham, M. H. Scales of Solute Hydrogen-Bonding: Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22*, 73−83.

(27) Eckert, F.; Klamt, A. Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE J.* **2002**, *48*, 369−385.

(28) Fukui, K. Role of Frontier Orbitals in Chemical Reactions. *Science* **1982**, *218*, 747−754.

(29) Parr, R. G.; Yang, W. Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity. *J. Am. Chem. Soc.* **1984**, *106*, 4049−4050.

(30) Beck, M. E. Do Fukui Function Maxima Relate to Sites of Metabolism? A Critical Case Study. *J. Chem. Inf. Model.* **2005**, *45*, 273−282.

(31) Contreras, R. R.; Fuentealba, P.; Galvan, M.; Perez, P. A direct evaluation of regional Fukui functions in molecules. *Chem. Phys. Lett.* **1999**, *304*, 405−413.

(32) De Proft, F.; Geerlings, P. Conceptual and Computational DFT in the Study of Aromaticity. *Chem. Rev.* **2001**, *101*, 1451−1464.

(33) Parr, R. G.; Chattaraj, P. K. Principle of Maximum Hardness. *J. Am. Chem. Soc.* **1991**, *113*, 1854−1855.

(34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of OPLS all-atom force field on conformational energetics

and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11235.
(35) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. J. Evaluation and reparametrization of the OPLS−AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(36) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(37) Tirado-Rives, J.; Jorgensen, W. L. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Energies for Protein−Ligand Binding. *J. Med. Chem.* **2006**, *49*, 5880−5884.
(38) Wang, H.; Liu, Y.; Huai, Q.; Cai, J.; Zoraghi, R.; Francis, S. H.; Corbin, J. D.; Robinson, H.; Xin, Z.; Lin, G.; Ke, H. Multiple conformations of phosphodiesterase-5: Implications for enzyme function and drug development. *J. Biol. Chem.* **2006**, *281*, 21469−21479.
(39) Pfefferkorn, J. A.; Guzman-Perez, A.; Oates, P. J.; Litchfield, J.; Aspnes, G.; Basak, A.; Benbow, J.; Berliner, M.; Bian, J.; Choi, C.; Freeman-Cook, K.; Corbett, J. W.; Didiuk, M.; Dunetz, J. D.; Filipski, K. J.; Hungerford, W. M.; Jones, C. S.; Karki, K.; Ling, A.; Li, J.; Patel, L.; Perreault, C.; Risley, H.; Saenz, J.; Song, W.; Tu, M.; Aiello, R.; Atkinson, K.; Barucci, N.; Beebe, D.; Bourassa, P.; Bourbounais, F.; Brodeur, A. M.; Burbey, R.; Chen, J.; D'Aquila, T.; Derksen, D. R.; Haddish-Berhane, N.; Huang, C.; Landro, J.; Lapworth, A. L.; MacDougall, M.; Perregaux, D.; Pettersen, J.; Robertson, A.; Tan, B.; Treadway, J. L.; Liu, S.; Qiu, X.; Knafels, J.; Ammirati, M.; Song, X.; DaSilva-Jardine, P.; Liras, S.; Sweet, L.; Rolph, T. P. Designing Glucokinase Activators with Reduced Hypoglycemia Risk: Discovery of *N*,*N*-dimethyl-5-(2-methyl-6-((5-methylpyrazin-2-yl)-carbamoyl)-benzofuran-4-yloxy)pyrimidine-2-carboxamide as a Clinical Candidate for the Treatment of Type 2 Diabetes Mellitus. *MedChemComm* **2011**, *2*, 828−839.
(40) Pammolli, F.; Magazzini, L.; Riccaboni, M. The productivity crisis in pharmaceutical R&D. *Nat. Rev. Drug Discovery* **2011**, *10*, 428−438.