JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

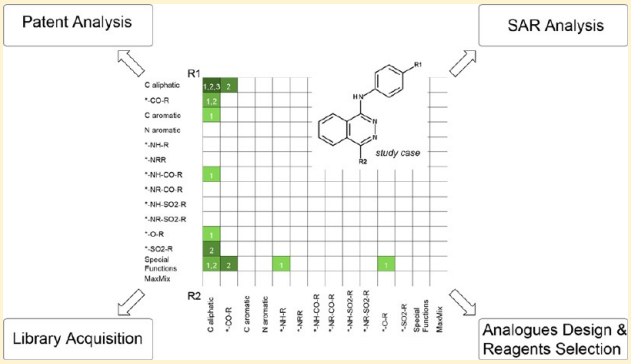
Biologically Relevant Chemical Space Navigator: From Patent and Structure–Activity Relationship Analysis to Library Acquisition and Design

Obdulia Rabal[†] and Julen Oyarzabal*[,†]

[†]Small Molecule Discovery Platform, Center for Applied Medical Research (CIMA), University of Navarra, Avda. Pio XII 55, E-31008 Pamplona, Spain

Supporting Information

ABSTRACT: A new and versatile visualization tool, based on a descriptor accounting for ligand–receptor interactions (LiRIf), is introduced for guiding medicinal chemists in analyzing the R-groups from a congeneric series. Analysis is performed in a reference-independent scenario where the whole biologically relevant chemical space (BRCS) is represented. Using a real project-based data set, we show the impact of this tool on four key navigation strategies for the drug discovery process. First, this navigator analyzes competitors' patents, including a comparison of patents coverage and the identification of the most frequent fragments. Second, the tool analyzes the structure–activity relationship (SAR) leading to the representation of reference-independent activity landscapes that enable the identification not only of critical ligand–receptor interactions (LRI) and substructural features but also of activity cliffs. Third, this navigator enables comparison of libraries, thus selecting commercially available molecules that complement unexplored spaces or areas of interest. Finally, this tool also enables the design of new analogues, which is based on reaction types and the exploration purpose (focused or diverse), selecting the most appropriate reagents.



INTRODUCTION

Considering that "much of chemical space contains nothing of biological interest",[1] recent reports have suggested that exploring and screening within the bioactivity-relevant space may provide compounds that are likely to be recognized by biological targets,[2−4] thereby increasing the probability of finding hits. Therefore, to conduct an efficient drug discovery project, we must clearly direct our navigation strategies to the overlapping region where chemistry and therapeutically relevant targets reside: "the biologically relevant chemical space".[5]

In a project-based context, once one initial hit is discovered, its exploration and optimization represent critical phases for the iterative drug discovery process. Thus, a variety of navigation strategies are required by medicinal chemists to achieve an optimized lead and enter the preclinical phase, including patent analysis, library comparison, compound acquisition, selection of molecules for focused screening campaigns, library design and synthesis, and structure–activity relationship (SAR), and structure–property relationship (SPR) analysis. From a pragmatic perspective, technology that supports medicinal chemists in congeneric series analyses, navigates the "biologically relevant chemical space" (BRCS) and aids the subsequent decision-making is required. Therefore, our efforts were devoted to implementing an interactive visualization technique to properly navigate the BRCS according to project needs for an analogue series: the BRCS navigator.

Our BRCS representation is predefined,[6] is based on established ligand–receptor interaction (LRI) types, and remains constant regardless of the specific structural features shown by the studied compounds, thus ready for any comparison analyses. The BRCS is populated according to the recently published ligand–receptor interaction fingerprint (LiRIf)[6] generated for each analyzed structural motif, in which any chemical structure can be represented. This methodology has been previously applied to noncongeneric R-group analyses;[6] however, considering that lead drug discovery projects are focused on analogue series, here we concentrate on the R-groups branching off from the common central scaffold.

The quality of corporate compound libraries is a key aspect of the drug discovery process because the discovery of hits from costly screening campaigns will depend upon their constituting molecules. When a novel target is identified, a validation process is pursued before any medicinal chemistry is started, and chemical probe availability is mandatory. Therefore, investment in library acquisition is of great importance for research groups working with novel targets where no

Received: September 21, 2012

   dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

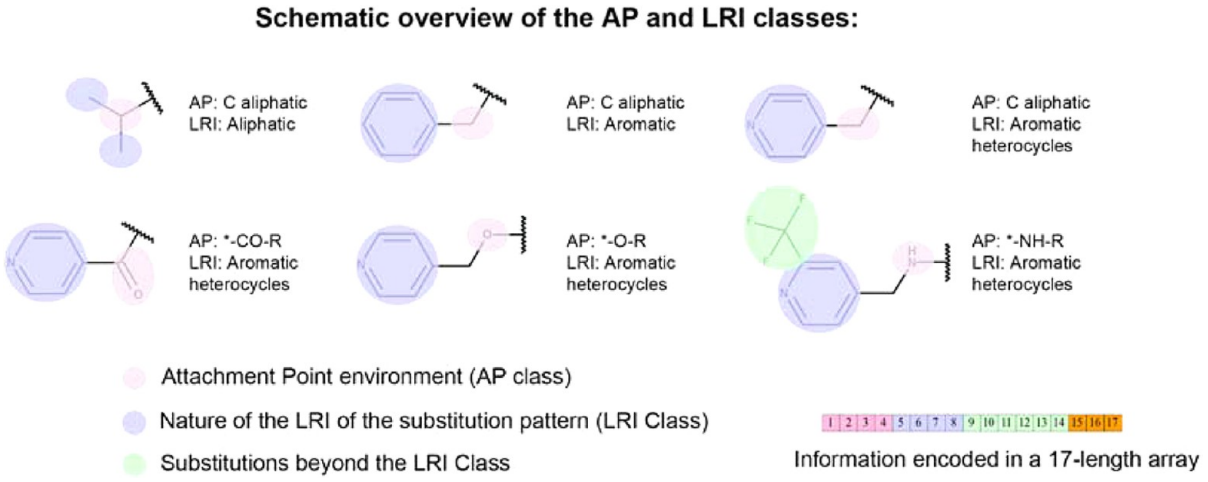


Figure 1. Schematic representation of the meaning of the AP class and the LRI class for a selected set of R-groups. All this information is encoded as a 17-length array, as detailed in ref 6.

information (structure- or ligand-based) may guide the design of novel compounds binding these targets.

A guidance tool for optimal selection and purchase may be very useful for optimizing the investment. The BRCS navigator is not suited for scaffold analyses, but there are some tools to perform this task.[7−9] However, once the central cores are selected (either from external providers or by an in-house "diversity" design),[10,11] this navigation tool provides an added value by selecting representative molecules for these scaffolds. Considering that the main aim, from the chemical biology perspective, is the identification of the chemical probe(s), providing diversity in the LRI might increase the probability of identifying hits while navigating the biologically relevant chemical space. Thus, proper selection should involve exploring as much of the BRCS as possible while avoiding redundant compounds covering the same AP and/or LRI. Conversely, after the discovery of a hit, an initial focused exploration around the area of interest where the hit is located may be required from the medicinal chemistry perspective. This navigation tool also enables a focused selection of molecules populating the area(s) of interest to confirm or discard the LRI held by the R-groups from identified hit(s), thus involving no synthetic effort.

Similarly, as a complement to the previous strategy, the BRCS navigator can also be utilized for library design to determine which compounds should be synthesized, providing an optimal balance between synthetic effort and coverage. To guide medicinal chemists in this latter task, from a synthetic feasibility perspective, the BRCS navigator disposes of the LiRIf-processed lists of commercial reagents that are ready to be selected according to the LRI of interest (diverse or focused) and synthesize the corresponding target molecules with the optimal R-groups in the analyzed growing vector(s).

The study of structure−activity relationships (SAR) and structure−property relationships (SPR) is one of the central themes in medicinal chemistry; it is a critical task for the iterative drug discovery process. In fact, medicinal chemists are still challenged by most hit explosion and lead optimization projects to explore individual SARs on a case-by-case basis.[12] Thus, in recent years, visualization techniques that view SAR features have become popular;[13,14] for instance, local SAR representations monitor individual analogue series, which include graphical extensions of traditional R-group decomposition schemes[15] and a network-like representation.[16,17] The BRCS navigator provides an added value for SAR navigation because three SAR extraction methods (out of the four, individually implemented into software tools)[18] are simultaneously utilized herein: dimensionality reduction, substructure annotation, and clustering (quantitative similarity comparison is not performed with this tool). This new visualization tool leads to interactive activity landscapes that are represented on a reference independent space based on the LRI-coordinates, independently of the analyzed compounds data sets and are, therefore, comparable. Navigating the BRCS allows for the localization of the "LRI activity cliffs"; further exploration to deeper levels within the corresponding LRI cluster of interest leads to the identification of the substructural changes that are responsible for this "activity cliff" and the detection of "matched molecular pairs". In addition, an explicit representation of the analyzed R-groups within information-rich SAR maps, including heat maps, which also display informative labels, is provided for user-selected LRI cluster(s) at any navigation level. The R-groups are, by default, sorted according to the LRI, thus providing a fast and intuitive analysis tool for medicinal chemists.

Patent informatics is a growing area; in fact, analyzing a competitor's chemical patent application is key for researchers working in the same or related research areas. Thus, when a new patent application is published, medicinal chemists generally read the patent document to identify key example compounds and to retrieve as much information as possible.[19,20] Predicting key compounds in the patent application is important work for drug discovery programs; the structural modification of competitors' key compounds is one of the strategies used to obtain novel lead compounds.[20−22] In addition, fragment occurrence analysis has also been used to transfer knowledge between chemical series, for example, in the context of fragment or scaffold hopping.[21−23] Finally, this visualization tool may also provide a positive impact on patent informatics by extracting the most relevant information for an analogue series, including the SAR and R-group frequencies as well as the sampled space comparisons, not only to identify unexemplified areas in the BRCS but also to learn from the evolution of competitors' patents.

In summary, based on a very recently reported novel descriptor that we developed, LiRIf, and the representation of the BRCS,[6] we report a new and intuitive visualization tool for

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Figure 2. Main interface of the Web tool to navigate the BRCS.

congeneric series that adds value to the following key tasks along the iterative drug discovery process: i) patent analysis, ii) library acquisition, iii) library design (together with reagent selection), and iv) SAR/SPR analysis. This versatile tool facilitates good decision-making by medicinal chemists. The practical applications of the BRCS navigator that show the impact of these four tasks are described herein (along with Supporting Information and a video demonstrating its use).

METHODS AND MATERIALS

R-Group Analysis and Characterization. Initially, an SD file containing a data set with compounds and all their available corresponding annotated information (biological activities, ADME data, *in silico* profiling, and so forth) is required as input; there is no limit on matrix dimensions. The target scaffold for the chemical series to be analyzed is explicitly provided by the user via a structure drawing applet. Thus, the analogue series from large, medium, or small hetero/homogeneous compound data sets are ready to be processed in detail. Then, the R-group analysis, using the *Generate SAR Information* component of the Pipeline Pilot,[24] is applied to identify all substituents or R-groups around the user-defined substructure (scaffold). Next, each R-group at each variation site is characterized and annotated according to its LiRIf descriptor. The LiRIf is a variable length array that captures information on the plausible ligand−receptor interactions that the R-group may establish as well as on its topology. For visualization purposes, the LiRIf is translated into a 17-length array, a predefined hierarchical classification schema consisting of 17 dimensions that provides information on i) the chemical nature of its attachment point (AP) and ii) the nature of the potential LRI that the substitution pattern beyond the attachment point may establish. A detailed description of the LiRIf calculation and its dimensionality reduction has recently been published;[6] therefore, here we will focus on the two main dimensions that define the highest-level representation for the BRCS: the AP and LRI classes, schematically depicted in Figure 1. Based on the chemical environment of its cleavage site (A0), an R-group is assigned to any of the following 14 AP classes: A0-$C_{aliphatic}$-R, A0-CO-R (A0-CS-R), A0-$C_{aromatic}$-R, A0-$N_{aromatic}$-R, A0-NH-R, A0-NR′-R, A0-NH-CO-R, A0-NR′-CO-R, A0-NH-SO2-R, A0-NR′-SO2-R, A0-O-R (A0-S-R), A0-SO2-R (A0-SO-R), the small groups class (A0-R, A0-CN) or the MaxMix class for more complex chemical moieties. In contrast, depending upon the chemical moieties beyond the attachment point, 17 LRI classes are distinguished: Hydrogen, Aliphatic, Aryl, Aromatic Heterocycles, Primary/Secondary Amines, Tertiary Amines, Hydroxyl/Thiol, (Thio)Ethers, (Thio)-Ketones/Sulfones, Amides/Carbamates/Ureide/Sulfonamide, Esters, Acids, Aliphatic Heterocycles, Charged Groups, Small Special Groups (nitrile and halogens), Rare functional groups/Phosphorus, and the MaxMix Class. The class definition is in agreement with the definition of the organic functional groups, facilitating intuitive user navigation. The MaxMix class includes substitution patterns bearing two or more branched LRI patterns beyond the attachment point. Both the AP and LRI classes are further divided into additional hierarchical subclasses, which are assignable to their corresponding dimension in the 17-length final array characterizing the R-group.[6] For example, the AP class A0-O-R encompasses, among many others, ethers, oxygen-linked esters, and the corresponding sulfur-linked equivalents. Similarly, the LRI class named "Amides" is a group of different functional groups that are equivalent from the viewpoint of ligand−receptor interactions (e.g., amides, sulfonamides, sulfinamides, and carbamates). Finally, as shown later for the study cases, other R-group levels from the 17-length array (moving to the BRCS lower levels) incorporate information about the following features: the presence of additional patterns of substitution for a given LRI class, including unsubstituted, monosubstituted, and polysubstituted; the nature of this additional substitution (if present); the type of ring assembly for the following (hetero)aromatic rings: the presence of monocycles, fused to (hetero)aromatic rings or fused to nonaromatic carbocycles; and the topological distances from the attachment point to the LRI class.

BRCS Navigator. A web tool for interactive navigation through the BRCS was developed using Pipeline Pilot.[24] It allows the user to input a data set of molecules as well as invariant molecular core structures. Then the customizable BRCS displays organized interactive heat maps representing R-

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

groups from selected chemical series (Figure 2). In a heat map, a cluster of items for one of the 17 available dimensions of an R-group (e.g., the LRI class) is depicted along its *y*-axis together with the clustering of another R-group dimension along its *x*-axis (e.g., the AP class). As shown later in the paper, these classes are further navigable in detail according to different environments (e.g., regarding the AP class, A0-O-R: O-linked esters and O-linked ethers). Moreover, if necessary, these subclasses can be easily upgraded to form a new AP class at the highest level of the BRCS; thus, a fully customizable representation of the BRCS is possible (e.g., Figure 7b). In addition, the depicted dimensions may correspond to the same diversity point of the scaffold (e.g., R1 AP class versus R1 LRI class) or to different sites (e.g., R1 AP class versus R2 AP class). For each axis, two dropdown boxes allow the user to select which diversity point and R-group dimension to display. Thus, by combining distinct R-group dimensions, the exploration can be hierarchically guided in a 'forward−backward' manner from the highest level of BRCS examination (AP versus LRI) to more detailed levels (e.g., degree of substitution of (hetero)-aromatic ring LRI classes versus type of (hetero)aromatic assembly). Each cluster of the heat map contains all of the compounds having the combination of R-groups matching the cluster defined by the combination of the corresponding *x*- and *y*-axes. This map can be color-coded according to the following different criteria: cluster population (number or percentage of compounds in a cell), biological activity (e.g., best value within each cell), or any other numerical property annotated in the input list of chemical structures (e.g., compound source if assigned numerically) or that can be rapidly calculated on the fly (e.g., R-group frequency, Polar Surface Area, and AlogP). At any level of BRCS exploration, compounds belonging to relevant cells i) can be viewed and exported to third-party programs (e.g., SD files), ii) can be separately partitioned/filtered out and imported into a new heat map window for independent BRCS examination, or iii) can be visualized in the form of standard SAR tables,[15] in which R-groups are sorted according to their LiRIf. These features are illustrated in Supplementary Movie 1.

When the library design and, consequently, the synthesis of novel compounds are required to explore undersampled areas of the BRCS or to focus on clusters of interest during the iterative optimization process, a common question is which compounds should be synthesized next. Once the LRI cluster of interest is identified (from a diversity or focused perspective) and the synthetic reaction to functionalize the selected diversity point is established (e.g., Suzuki coupling reaction for R2), then one must select among all of the commercially available building blocks for those driving toward synthetically feasible target compounds (e.g., boronic acids or boronates) bearing R-groups located in the desired cluster of the BRCS. To this end, we incorporated a module to map the lists of commercial reagents onto the BRCS representation of the data set being analyzed. To fully exploit this module, the user must specify the expected attachment point (AP), which is derived from the underlying reaction between the reactive group of the chosen reagent list and the scaffold under study. Building blocks catalogs were downloaded from the Zinc database[25] and classified according to their reactivity into different lists (e.g., alcohols, carboxylic acids, alkyl halides, acyl halides, primary amines, secondary amines, boronates, boronic acids, isocyanates, and aldehydes). These lists were preprocessed by clipping off their particular leaving group, and the 17 length array, defining the remaining R-group, was computed and stored internally. If accessible, a better alternative would be a real-time connection to an established database of reagents, such as ACD[26] or Zinc,[25] so that these lists can be quickly processed on the fly.

Data Sets. Compounds from proprietary patents of Novartis and Amgen covering the 1-anilinophthalazine substructure **1** were collected from the Kinase Knowledge Database (KKB).[27] The Novartis data set, compiled from 3 different patents,[28−30] consists of a total of 183 unique structures among which 52 have IC50 values that are annotated against the Flt1 VEGF receptor tyrosine kinase. For the Amgen proprietary series, a total of 767 nonrepeated structures were extracted from 4 patents.[31−34] The IC50 values against Aurora A and B isoforms are available for a total of 492 and 506 compounds, respectively (490 have both values), as shown in Table 1. The structures were tagged with increasing integers

Table 1. Data Set Composition

origin, *1-anilinophthalazine series*	# compounds	target[a]	target[a]
Novartis, 183 molecules		**Flt1**	
WO 98/35958[28]	90	5.07−7.38 (47)	
WO 00/59509[29]	78	5.31−6.50 (3)	
WO 02/090346[30]	15	5.72−5.72 (2)	
Amgen, 767 molecules		**Aurora A**	**Aurora B**
WO 2007/087276[31]	244	5.63−8.49 (22)	6.18−9.40 (22)
WO 2008/124083[32]	486	4.60−8.40 (467)	4.60−9.00 (469)
WO 2009/117157[33]	20	7.00−7.00 (6)	7.00−7.00 (15)
WO 2010/017240[34]	17	----	----

[a]Biological activity data reported in the format of pIC_{50}; the activity ranges covered in each patent are described. Values in parentheses indicate the number of compounds from each patent with available biological data.

according to their corresponding patent application priority dates to track the originating patent as well as the appearance order. As expected, both sets contain nonrepeated novel structures.

For the library comparison analysis, 959 commercially available compounds mapping to 1-anilinophthalazines **1** were downloaded from the Zinc database;[25] in addition to the structural information, the supplier and catalogue number were kept in the corresponding SD file.

RESULTS AND DISCUSSION

The 1-anilinophthalazine substructure **1** (Figure 3) is very versatile for kinase inhibition and is found in the following compounds: Vatalanib **2**,[35] a vascular endothelial growth factor receptor (VEGFR) inhibitor; and AMG900 **3**,[36] a pan-Aurora inhibitor, which are both currently undergoing clinical development for oncology indications. The discovery of these candidates was achieved after an iterative lead optimization process; a detailed SAR exploration is thus available. To illustrate the utility of this tool, we therefore selected the proprietary patent applications of Novartis and Amgen on Vatalanib and AMG900 substructures, respectively. Two separate sets of 183 (Novartis) and 767 (Amgen) unique structures were collected from exemplary compounds in patents, together with their biological data. Here, we will focus our discussion on the R1 groups (the substituents attached to the *para* position of the aniline moiety placed in 1)

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

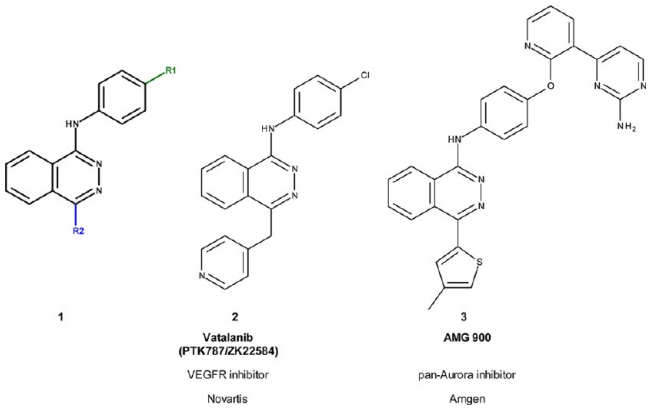


Figure 3. 1-Anilinophthalazine core **1** and structures of two clinical candidates: Vatalanib[35] and AMG 900.[36]

and the R2 groups (the substituents attached to position 4 of the phthalazine ring) as major exploration efforts in both sets were directed toward these diversity points.

Patent Analyses. This application case focuses on how to quickly one can mine the chemical structural information that is contained in patents and, in addition, takes advantage of the reference-independent nature of the BRCS to compare the mined chemical data from different data sets. In particular, we illustrate the potential of our tool to i) examine compound progression across subsequent patents with the goal of predicting key compounds, ii) analyze R-group frequency to identify preferred substitution patterns, and iii) establish comparisons between different competitors, although ideally, one would compare them with the in-house compound database. Exemplary compounds can be extracted from specialized databases, as in this paper, using text-mining techniques or other tools (Scifinder).[20,37] We note that this kind of analysis exclusively attempts to ease the visualization and analysis of exemplary compounds extracted from patents and that no definitive conclusion can be stated on patent claims.

The first step in our procedure is the examination of the different linkers at each substitution site. A heat map representation of the coverage of the 14 AP classes by the R1 substituents (*y*-axis) and the R2 substituents (*x*-axis) is shown for the Novartis (Figure 4a) and Amgen (Figure 4b) exemplifications. Trends in the chemical space sampling across subsequent patent applications are readily accessible by color-coding the clusters according to the patent membership of the different R-groups. The Novartis preference evolved toward the R1 and R2 substituents that are linked to the core via aliphatic carbons, with at least one novel compound from three subsequent patent applications falling within this cluster. Other linkers at these positions (e.g., *-NH-R/*-O-R at R2 or *-NH-CO-R/aromatic carbons at R1) are exclusive of novel compounds exemplified in the first patent application (clusters labeled with "1" in Figure 4a). A similar trend, in accordance with a selection patent strategy, is observed for the Amgen program. Here, the exemplification surrounds compounds bearing O-linked substituents at R1 and carbon-bonded aromatic rings at R2 (cluster labeled with "1, 2, 3, 4" in Figure 4b, which means the four Amgen patents exemplified this cluster); in fact, the third and fourth patents were exclusively focused on this cluster.

When repeating this highest BRCS level representation for the set obtained after merging both lists (Figure 4c, 950 structures, hereafter referred to as the "proprietary set"), the differences in chemical space protection between the companies are visually recognized using the color-coding according to the compound origin. The number of representatives is shown in the cell labels. As observed in Figure 4c, the only overlapping region corresponds to the 22-membered yellow green cell (color gradient indicating Novartis predominance) assignable to compounds derivatized with groups attached via an aliphatic carbon to R2 and oxygen-linked groups at R1. Obviously, the clusters with the most concentrated exploration efforts over time are the most populated (122- and 634-membered clusters from Novartis and Amgen sets, respectively).

The next step, after analyzing the nature of the different linkers, is to examine the chemical moieties of the R-groups beyond each attachment point. For each substitution site of the 950-proprietary set, the LRI class (*y*-axis) was plotted against its corresponding AP class (*x*-axis) for the R1 (Figure 4d) and R2 (Figure 4e) substituents. Reflective of the companies' respective clinical candidates in Figure 3, the exploration at R1 greatly differs between Novartis (green cells in Figure 4d) and Amgen (blue cells), with only one compound from each company assignable to the same cluster (carbon-linked aliphatic heterocycles). While Novartis focused on small groups (hydrogen, aliphatic chains, and halogens), a large number of Amgen compounds contain heteroaromatic rings attached to R1 through different linkers, although with a strong preference for oxygen linkers, as for the case of AMG900. For R2, both companies concentrated on hydrophobic planar substituents (aromatic and heteroaromatic LRI classes) that were coupled either directly to the 1-anilinophthalazine core **1** (Amgen) or via an aliphatic linker (Novartis). To a certain point, one would assume (without inferring anything from the SAR data against their respective targets or knowing the historical background of the series)[38] that Amgen (with posterior patent application dates) took advantage of the initially unexplored 206-membered cluster from Novartis. The corresponding heat maps for each separate set are shown in Figure S1 of the Supporting Information.

As part of the patents analysis strategy performed with this new visualization tool, the R-group frequency is computed when generating the R-group pool. This calculation enables the identification of the most frequently occurring R-groups at each particular diversity point, which may be utilized for SAR transfer assuming that the binding modes between the different chemical series remain constant.[23] These frequently occurring R-groups are likely to be present in the key compounds in the patents.

For example, focusing on R1, a visual inspection of Figure 4d highlights that the most populated cluster contains 748 molecules (such as AP, A0-O-R; and LRI, heteroaromatic), where the most frequently occurring R-group appears 132 times (Supporting Information Figure S2a). Further navigation within this cluster of interest, which focuses on the ring assembly type of the heteroaromatic moiety and its degree of substitution (*y*-axis and *x*-axis, respectively, in Supporting Information Figure S2b), pinpoints the presence of the two preferred fragments **4** and **5** (Figure 5) with frequencies of 132 and 73 on two different clusters comprised of 276 and 158 compounds, respectively. In fact, the clinical candidate AMG900 holds fragment **5**.

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

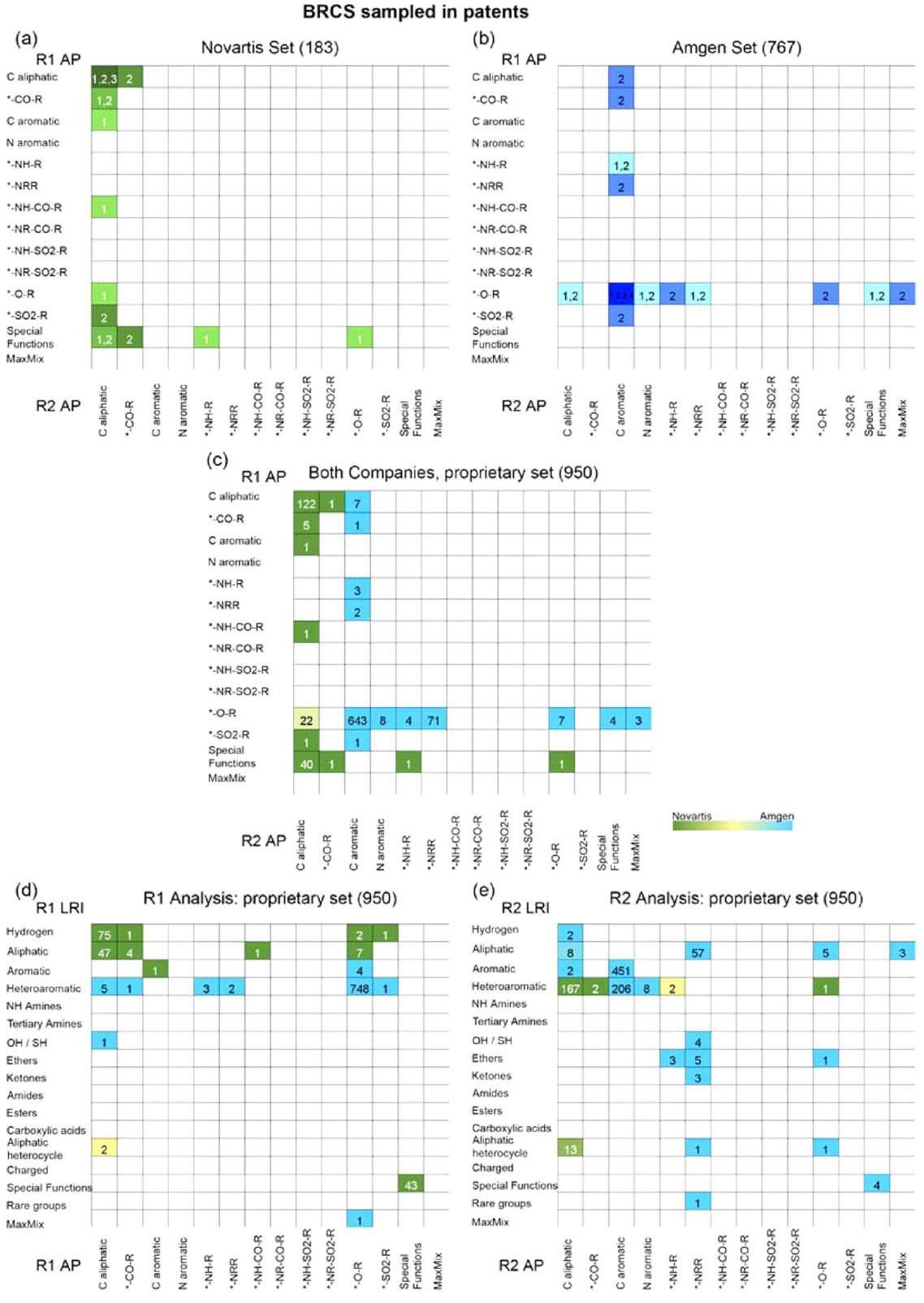


Figure 4. Heat maps corresponding to the BRCS projection of the R-groups at the R1 and R2 positions as extracted from the Novartis (green) and Amgen (blue) exemplary compounds. (a-c) Analysis of the coverage of the 14 AP classes by the R2 groups (*x*-axis) against the R1 groups (*y*-axis). In parts (a) and (b), the labels indicate the patent application(s), which are assigned increasing integers according to their priority dates, for which at least one novel compound falls within a cell. Each combination of patent application(s) was assigned a different green (Novartis, part a) or blue color (Amgen, part b), where the intensities correlate with the priority dates of the compounds within the cluster (a higher intensity represents more recent dates). (c) Combination of both sets. The total number of representatives within each cell is shown. (d-e) The highest level of the BRCS analysis of the R1 (part d) and R2 (part e) positions obtained by projecting the compounds from both sets onto the 238 clusters defined by the 14 AP classes (*x*-axis) and the 17 LRI classes (*y*-axis). The heat maps are color coded by the proprietary company as in part c and labeled with the total number of compounds. Light gray classes correspond to the 58 nonexistent pairwise combinations because they are unrealistic by definition.[6]

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

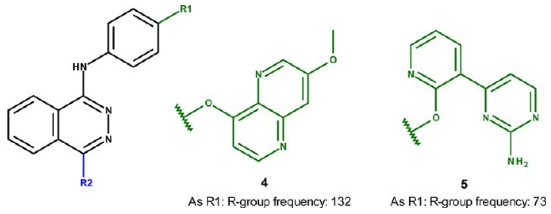


Figure 5. Patent analysis focused on R1 diversity point from core structure **1** identifies fragments **4** and **5** as the most common R-groups for examples reported in Amgen patents.

Library Acquisition. Our BRCS navigator leads to an adequate scenario for library acquisition from two perspectives: i) selection of the most representative candidates from a novel chemical series for the corporate collection, which suggests that the selection strategy covers as much of the BRCS as possible and provides diversity in the LRI; and ii) a complementing chemical series that is already available in the corporate collection. In the latter case, after performing the proper comparison between the compounds from the proprietary database and those that are commercially available, the purchased molecules should cover the unexplored BRCS space to increase the chance of finding hits. A focused application based on the same principles will be discussed in the next section (SAR/SPR analysis).

As an example of a screening library, 959 commercial compounds mapping to substructure **1** were downloaded from the Zinc Database.[25] In Figure 6a, the corresponding heat map for the projection of the R1 and R2 AP classes sampled by this commercial set is shown (color coded by occupancy). Combining this set with the set of 950 proprietary 1-anilinophthalazines (our "internal" library) found only 3 structures in common; the visual inspection of the resulting heat map in Figure 6b allows the user to quickly identify novel, unexplored chemistry around the linker region of both variation sites (violet cells in Figure 6b) as well as alternatives to

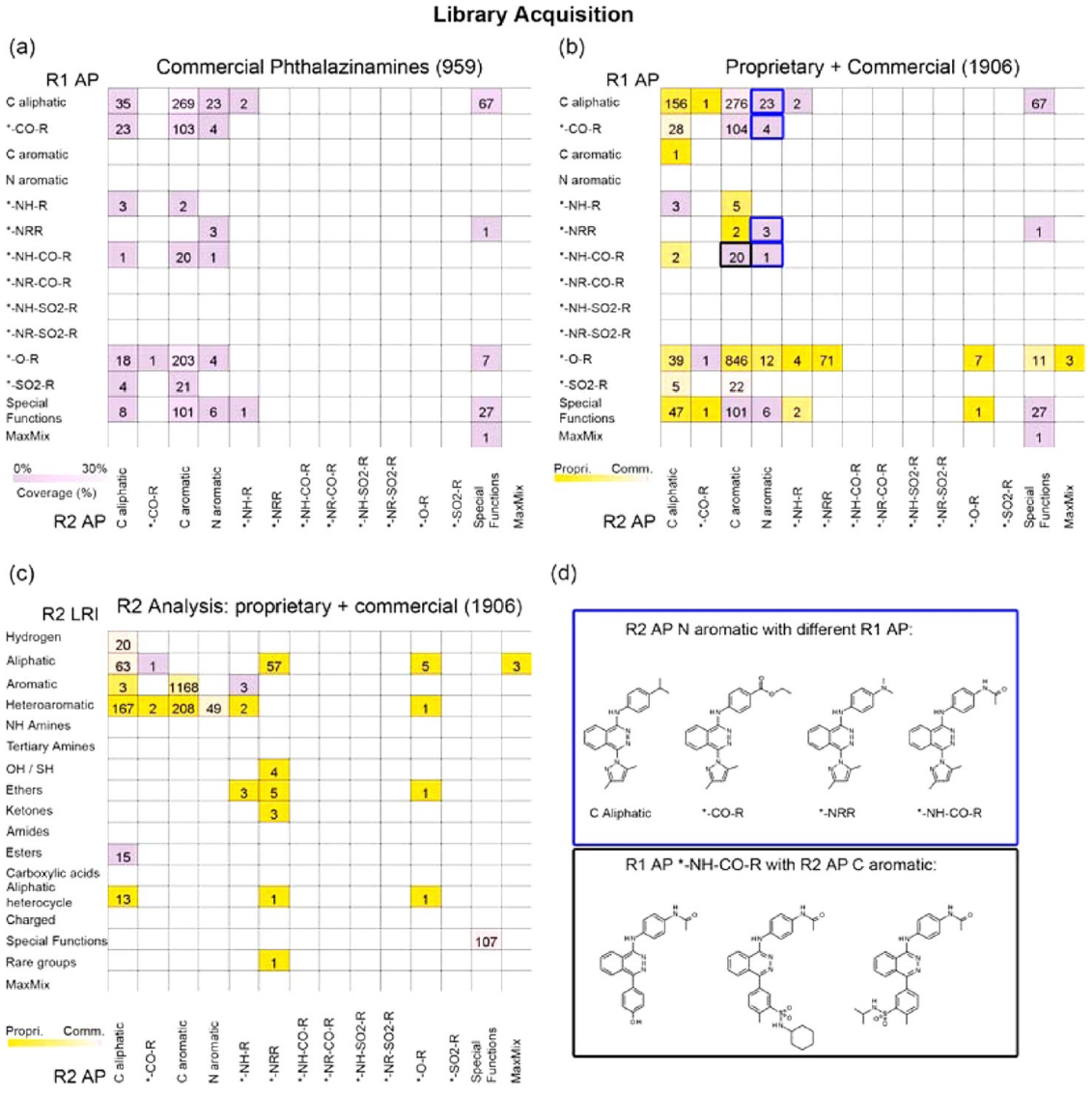


Figure 6. BRCS projection of 959 commercial phthalazines (violet) alone (a) or in combination with the 950 compounds extracted from proprietary patents (b-c), color coded in yellow). (a-b) Analysis of the coverage of the 14 AP classes by R2 groups (*x*-axis) against R1 groups (*y*-axis). (c) Highest level BRCS analysis of the R2 substituents from both sets (AP class against LRI class). Labels show the number of compounds. (d) Exemplary compounds from clusters bordered in blue and black in part b.

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

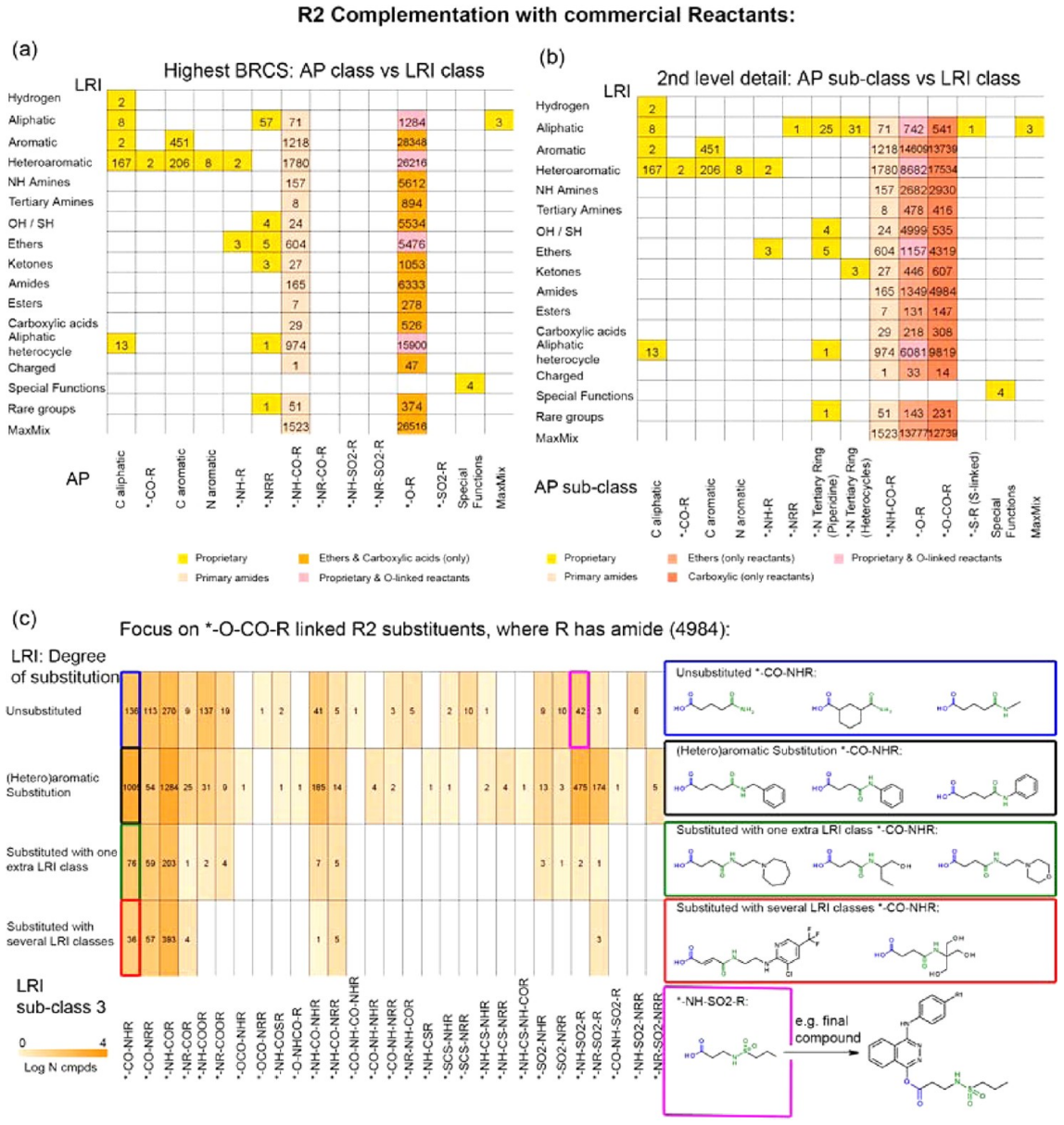


Figure 7. Diversity-guided complementation of the space sampled at the **R2** position by the 950 proprietary compounds with different lists of selected reactants: primary amides that provide unexplored *-NH-CO-linked substituents and ethers and carboxylic acids that yield novel oxygen-linked **R2** substituents. (a) highest level BRCS: AP class (*x*-axis) versus LRI class (*y*-axis). Cells are color coded as follows according to the **R2** substituent source: described in the proprietary set (yellow), new cells covered by available primary amides (salmon), new cells covered by commercial ethers and carboxylic acids (orange), and cells already sampled by the proprietary library that can be complemented with chosen reagents (pink). (b) Heat map obtained for refined AP class information: AP subclass (*x*-axis) versus LRI class (*y*-axis). Color pattern as in part a; here, ethers and carboxylic acids are differentiated by orange intensity. (c) The next level of BRCS: the heat map of a focused subset of 4984 reagents bearing the amide-like moiety (LRI class amide, green in displayed examples) to be linked by reacting carboxylic acids (*-O-CO-R AP subclass, blue in displayed examples) at the **R2** position. The classes are distinguished by the degree of substitution of the amide-like moiety (*y*-axis) and the specific functional group providing the amide-like interaction pattern (*x*-axis). The heat map in part c is color coded by cluster coverage (logarithmic scale). In all heat maps, the labels correspond to the number of structures. To the right of part c, some selected examples that are representative of the classes bordered with the same color as in the heat map are depicted.

complement the undersampled clusters (salmon cells). At a first glance, the new compounds are detectable from the commercially available screening library, which is functionalized with aromatic nitrogens bonded to the R2 position (imidazoles) that bear different combinations of linkers at R1 (cells bordered in blue in Figure 6b) or new compounds with nitrogen-linked amide substituents at R1 that are functionalized with aromatic rings at R2 (cell bordered in black in Figure 6b). Exemplary compounds from each of these clusters are shown in Figure 6d.

The same procedure is applicable at any BRCS level, as shown in Figure 6c for the particular examination of the R2 substituents (AP class against LRI class) from the 950-proprietary (yellow) and 959-commercial (violet) sets. With the exception of the previously mentioned nitrogen-bonded heteroaromatic rings, 15 esters linked through aliphatic chains, 3 aromatic rings bonded via NH, and 1 keto-linked aliphatic

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

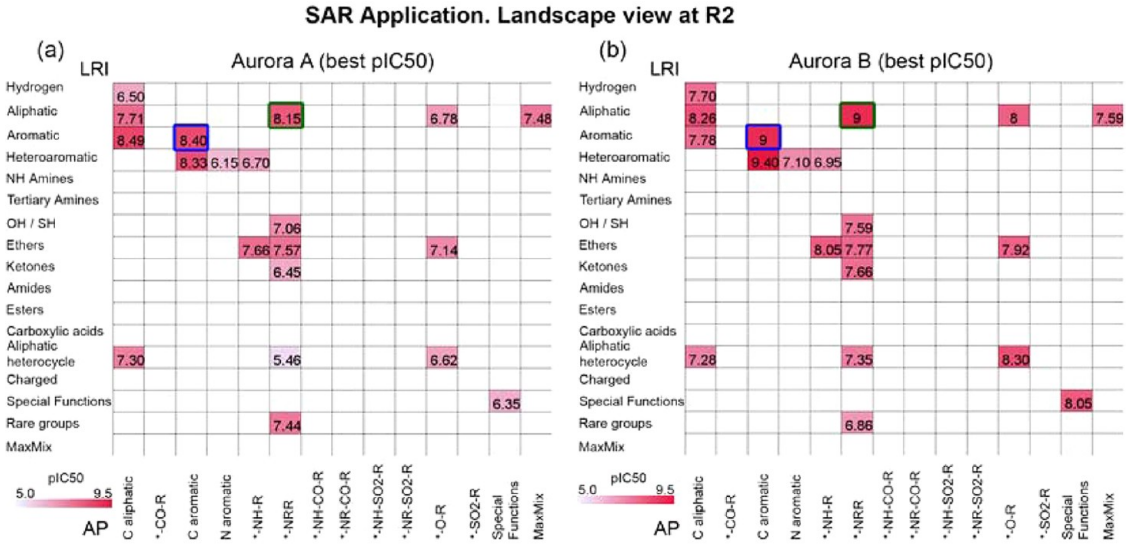


Figure 8. The highest level BRCS representation of the R-groups at the R2 position from the 767 proprietary Amgen compounds: the R2 activity landscape. Cells are color coded and labeled according to the best pIC50 value found among all of the compounds in a class against Aurora A (a) and Aurora B (b). The light gray classes correspond to the 58 nonexistent pairwise combinations because they are unrealistic by definition.[6] The green- and blue-bordered cells correspond to the classes analyzed in depth in Figures 10 and 11, respectively.

group, most of the commercial R2 substituents covering unexplored clusters are in fact small groups (hydrogen, halogen, or nonfunctionalized alkyl chains). Also of note, 717 commercial compounds share the aromatic substitution pattern that is covered by the Amgen compounds (1,168-membered cluster in Figure 6c); at first glance, one could therefore assume that they are likely to already be IP protected, although a deeper BRCS analysis of this cluster should be performed to draw conclusions on the novelty of the substituents (data not shown).

Library Design. For the cases where commercially available compounds and proprietary molecules do not explore certain areas of interest in the BRCS, we must design molecules according to our aim (focused or diverse) and plan their synthesis.

Initially, we were focused on a library design that covered diversity, complementing the BRCS unexplored by the 950 proprietary compounds, particularly those areas where the 959 commercially available molecules were not able to explore. To illustrate this capability, we propose the design of compounds bearing two new attachment point classes at the R2 position that remained either unexplored (*-NH-CO-R column in Figures 4e and 6c) or undersampled (*-O-R column in Figures 4e and 6c, with only 8 representatives) by the 950 proprietary data sets and by those 959 commercial 1-anilinophthalazines. Moreover, for the *-O-R class, we distinguished between the ethers (*-O-R) and the oxygen-linked esters (*-O-CO-R) by fixing the AP subclass dimension of the R-group description. Thus, after the retrosynthetic analyses, primary amides, alcohols, and carboxylic acids were retrieved as reagents bearing the functional groups that may drive, through different synthetic strategies, the compounds containing each of the proposed AP environments: *-NH-CO-R, *-O-R, and *-O-CO-R. In Figure 7a, the highest level of the BRCS projection of 131,022 commercially available reagents (see Methods) onto the proprietary set of 950 is shown: a high diversity of the LRI is provided, and, in fact, all possible LRI clusters for each new attachment may be explored. Visual inspection of all substituents within these clusters may be tedious because many are highly populated, so a better alternative is to depict the corresponding AP and LRI subclasses, as shown in Figure 7b (AP subclass against the LRI class) or in Figure S3 of the Supporting Information (AP subclass against the LRI subclass). An interactive navigation within these heat maps is aided by the tooltips displayed when the mouse passes over a cell. As observed on the *x*-axis of Figure 7b, a detailed description is provided for the tertiary nitrogen-bonded classes borne by the commercially available 1-anilinophthalazines described in Figure 7a (noncyclic tertiary amines, piperidine rings and other nonaromatic nitrogen-linked rings); in-depth detail is achieved for the oxygen-linked AP (ethers, esters, and thiols). If, for example, we focus our interest on the oxygen-linked esters with an amide LRI class (the 4984-membered cell in Figure 7b), then we can further refine the selection by asking for the different third level subclasses that are contained within the amide LRI class (*x*-axis in Figure 7c) and their degree of substitution (*y*-axis in Figure 7c). By clicking on each cluster, the corresponding reagents are retrieved, including details about the supplier and catalogue number, as shown for a set of selected examples in Figure 7c.

Once the SAR analysis is performed and the most relevant area of the BRCS to achieve the activity and selectivity aims is identified, we will apply identical rationale for a focused library design, i.e., focused on certain LRI clusters (details below, just after the SAR/SPR analysis).

SAR/SPR Analysis. Finally, we show the application of this navigation tool for investigating SAR/SPR landscapes within a compound analogue series, in particular for the fast recognition of the R-group patterns that are SAR/SPR determinants and can be used as cluster seeds for further chemical exploration. In addition, the interactive landscape visualized in this reference-independent BRCS representation may also lead to the identification of LRI "activity cliffs" and, from there, to the detection of matched molecular pairs. As a study case, we consider the Amgen data set because of the amount of available inhibition data for Aurora A (AURKA, 492 points) and Aurora B (AURKB, 506 points), which cover a wide potency range (pIC50 from 4.6 to 8.5 against Aurora A and pIC50 from 4.6 to

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

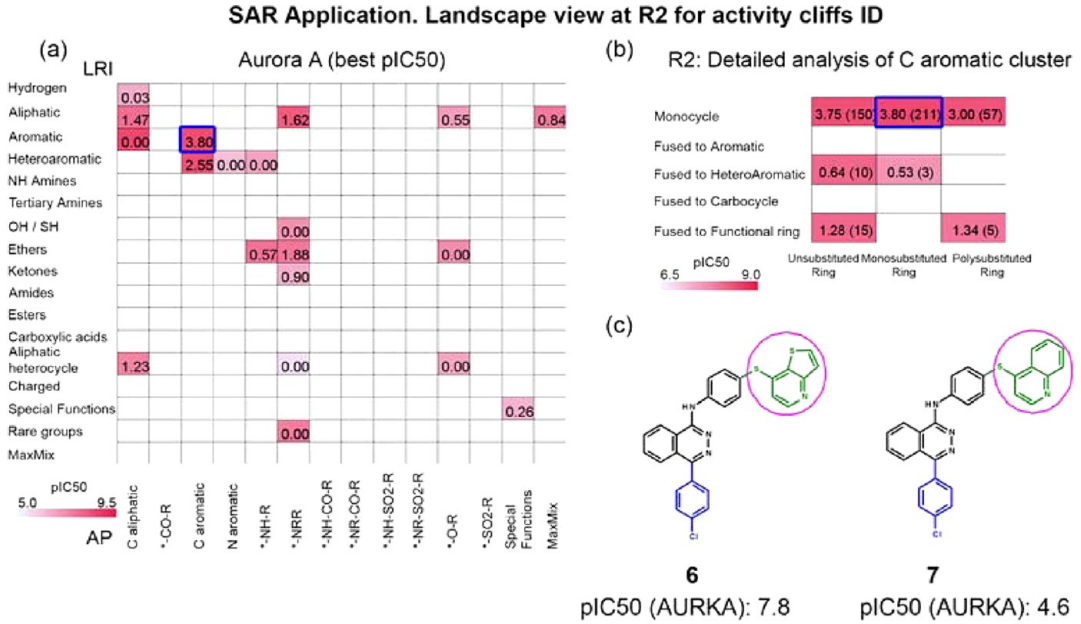


Figure 9. a) The highest level BRCS representation for the identification of the activity cliff at the R1 site from the 767 proprietary Amgen compounds. Cluster labels indicate the potency range against Aurora A covered by the compounds within each cluster. b) Detailed analysis of the aromatic substituents directly bonded to the R2 site (blue-bordered cluster in part a). The cells are labeled as in part a, with the total number of compounds within each cluster in parentheses. c) Matched molecular pairs identified from the "LRI activity cliff" after visual inspection of the interactive activity landscape representation.

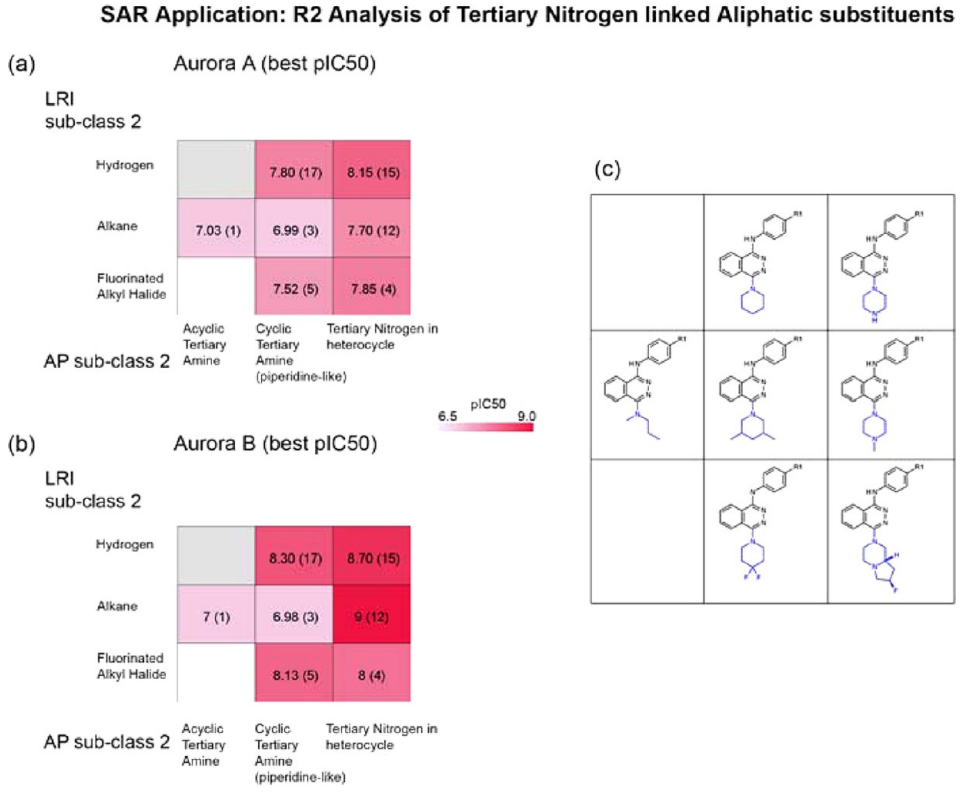


Figure 10. Detailed analysis of the tertiary nitrogen-linked substituents bearing aliphatic LRI classes at R2, corresponding to the cluster bordered in green in Figure 8. The AP subclass level (*x*-axis) is plotted against the LRI subclass level (*y*-axis). The cells are color coded and labeled (together with cluster occupancy) according to the best pIC50 value found among all of the compounds in a class against Aurora A (a) and Aurora B (b). The light gray class corresponds to the nonexistent pairwise subclasses combinations, which are already covered in the highest level BRCS (corresponding to A0-NR′-R, where both substituents are hydrogens). (c) Exemplary R-groups from each cell.

9.4 against Aurora B). Novartis compounds are reported to bind to a different target and with a different binding mode;[35,38] so, this SAR/SPR study case is only focused on the Amgen data set.

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

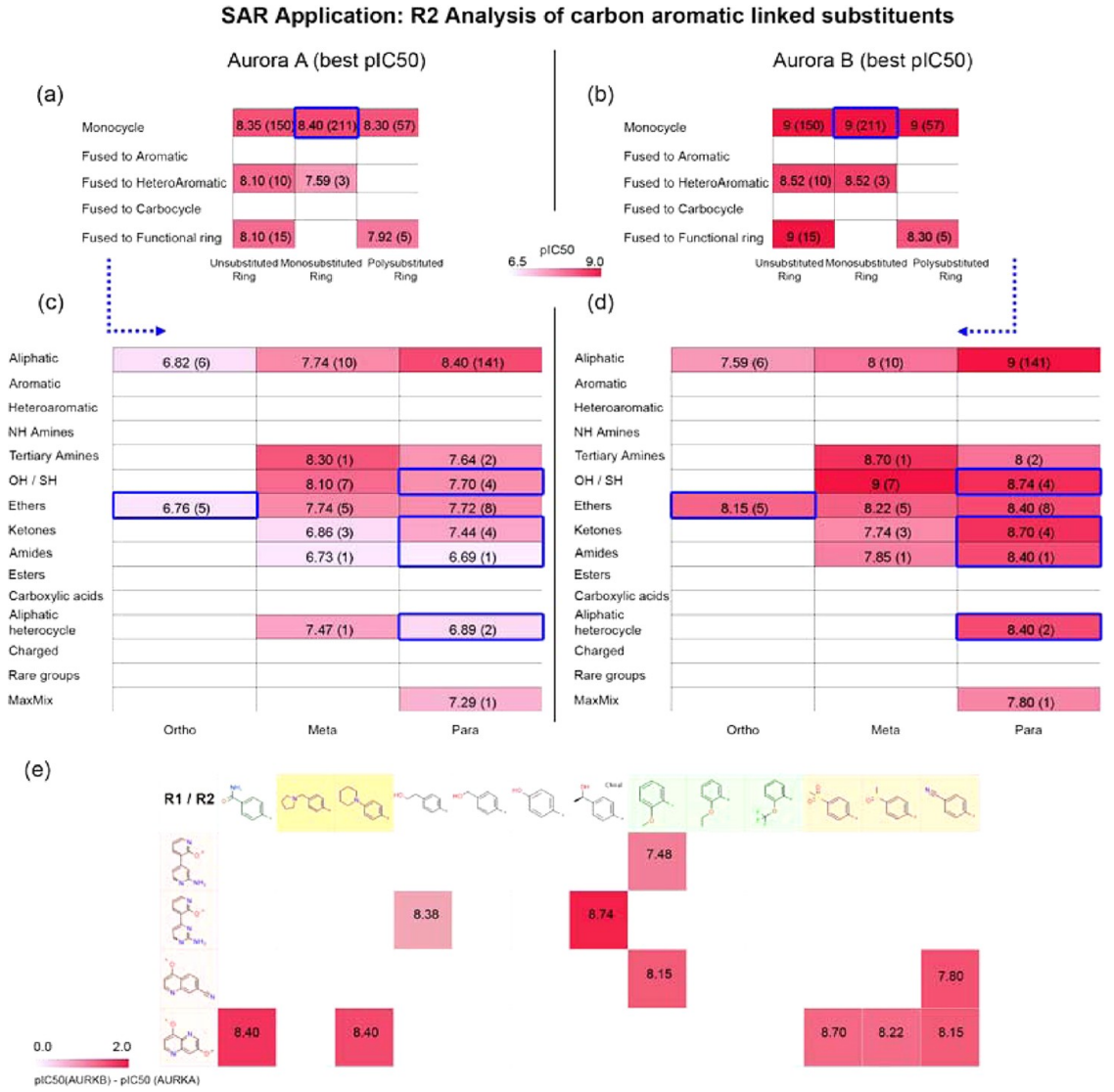


Figure 11. Detailed analysis of the aromatic substituents directly bonded to the R2 site, corresponding to the cluster bordered in blue in Figure 8. First, the ring assembly type (*y*-axis) is plotted against its degree of substitution (*x*-axis) in part a (best pIC50 values against Aurora A) and part b (Aurora B). Second, the separate representation of the compounds in the blue-squared cluster in parts a and b, corresponding to the benzene ring, enables their differentiation by the LRI class assignable to the substitution pattern of this ring (*y*-axis) as well as their ring position (*x*-axis) in parts c and d. The cells are color coded and labeled (together with cluster occupancy) according to the best pIC50 value found among all of the compounds in a class against Aurora A (a, c) and Aurora B (b, d). (e) The SAR map obtained by plotting the structures of the R2 groups (*x*-axis) against the structures of the R1 groups (*y*-axis) for the clusters selected in parts c and d. The log of the ratio of Aurora B (AURKB) IC50 (shown in the label) to Aurora A IC50 is used as the color-coding parameter. Untested compounds are gray.

Figure 8 shows the highest BRCS representation for the substituents at R2, color coded according to the best pIC50 value found within a cluster against Aurora A (Figure 8a) and Aurora B (Figure 8b); thus, these figures represent the activity landscapes for the R2 diversity point. Because of the retrospective character of this example, when the projects were finished, the clusters containing the most potent R2 groups against both kinases, mainly (hetero)aromatic rings directly coupled to the scaffold, were also generally the most explored (corresponding occupancy shown in Figure 4e). Alternatively, it is possible to opt for other color metrics: the worst value within a cluster or the mean or mode value over all of the molecules. Thus, with a few simple mouse clicks and basic monitoring of the breaks in the color trend, the medicinal chemist is able to identify the clusters with steep changes in potency against a particular target or clusters, thus providing opportunities to improve the selectivity against other targets.

Alternatively, color coding or labeling the potency range within a cluster (the difference between the best and the worst activity values) enables the fast visual identification of "activity cliffs". As shown in Figure 9a, the IC50 against Aurora A of the Amgen compounds bearing aromatic moieties at R2 spans 3.8 log units. Detailed disclosure of this cluster (Figure 9b) restricts this potency change to molecules where R2 is a monosubstituted phenyl ring. Visual examination of the molecules within this cluster reveals an interesting "matched molecular pair" consisting of the replacement of a phenyl ring by thiophene (compounds **6** and **7** in Figure 9c). While typically established as a bioisosteric replacement, in this particular case this structural modification involves an increment of 3.2 log units against Aurora A.

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

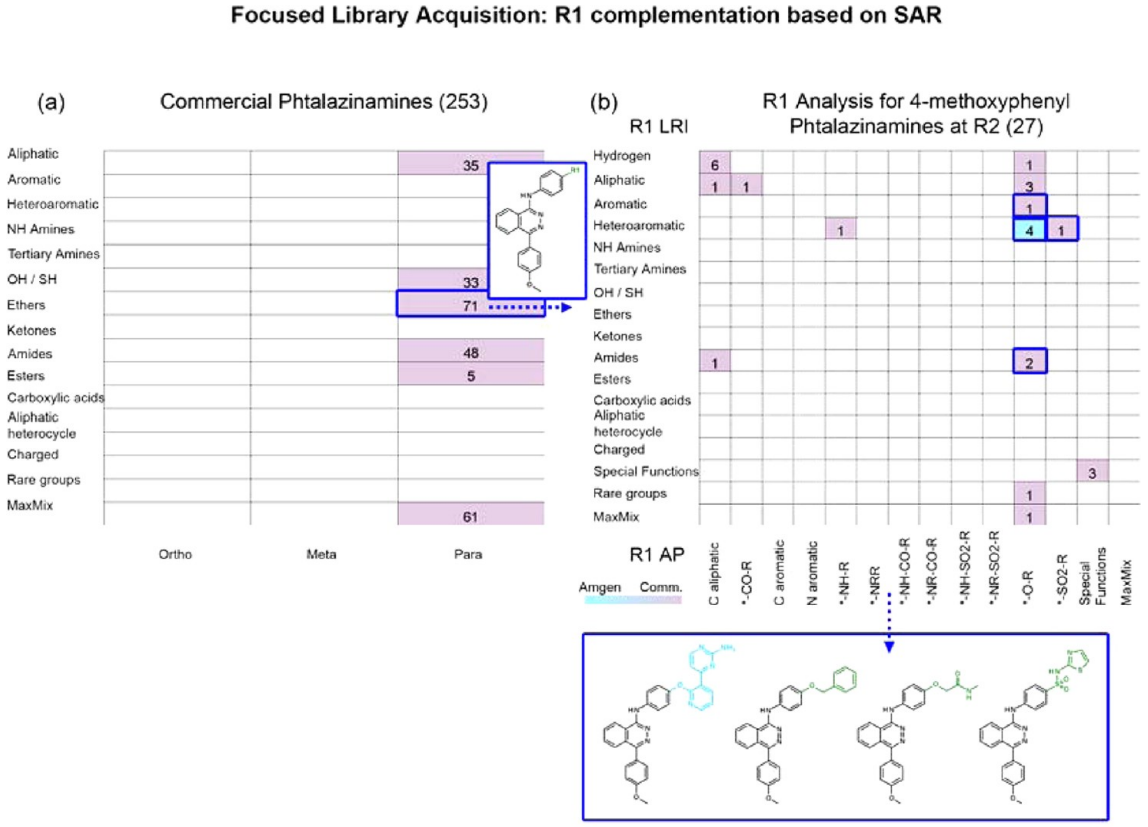


Figure 12. Focused library acquisition. Alternative compounds that complement the interesting regions identified with SAR analysis in Figures 11c-11d. (a) Projection of the purchasable phthalazinamines. (b) The highest level BRCS analysis of the R1 substituents of the commercial (violet) and Amgen compounds (blue) bearing a para methoxy phenyl group at R2. Structures from the selected clusters (squared in blue in part b) are shown below in the blue-squared area. Clusters are color coded according to the compound origin.

To demonstrate this SAR/SPR analysis, we now focus on two different LRI moieties: a nonaromatic class and an aromatic class, both containing highly potent phthalazines as well as more than 10 representatives (occupancies shown in Figure 4e and Figure S1d of the Supporting Information). We start with the nonaromatic class analysis by selecting the potent tertiary nitrogen-linked R2 substituents within the aliphatic LRI class (green-bordered cells in heat maps in Figure 8). In Figure 10, the corresponding heat maps obtained by plotting the AP subclass (*x*-axis) against the LRI subclass (*y*-axis) levels are shown. Again, the clusters are color coded by their best pIC50 values against each target Aurora A (Figure 10a) and Aurora B (Figure 10b). Exemplary R-groups from each cell are shown in Figure 10c. Although the subclass consisting of acyclic tertiary nitrogen-linked groups is uniquely represented by a single compound, a preference for cyclic rings (having one or more heteroatoms) is observed for increased potency, especially against Aurora B. A direct side-by-side visual comparison of both maps reveals that heterocycles with a second tertiary nitrogen substituted with alkyl chains present some Aurora B selectivity against Aurora A. Because of the simplicity of the aliphatic LRI class, which comprises hydrogen (unsubstituted heterocyclic rings, part of the attachment point), alkyl, or fluorinated chains without any other interacting moiety,[6] the analysis of the R2 position finishes here for this cluster, and selected examples in the clusters may be re-examined for the analysis of the remaining positions (e.g., R1 for this compound series) to test their influence on the activity/selectivity patterns.

For the aromatic class example, the corresponding heat maps in Figures 11a (Aurora A) and 11b (Aurora B) were created after selecting compounds with their aromatic assemblies attached to R2 (the cluster bordered in blue in Figure 8) and opening a new window tab where the ring assembly type (*y*-axis) is represented against its degree of substitution (*x*-axis). Then, by repeating the same exportation process for the 211 monosubstituted monocycles (benzene rings, bordered in blue in Figures 11a and 11b), a new heat map is requested to plot the LRI type of the substitution on the benzene ring against its ring position (Figures 11c and 11d for the Aurora A and B pIC50 values, respectively). Ortho-ethers and some substructures at the para position (alcohols, amides, ketones, and aliphatic heterocycles) present a ∼1.2-fold selectivity for Aurora B over Aurora A. Aliphatic chains at the para position is the preferred class, with the highest potency values against both isoforms.

BRCS navigator is provided with a standard module to generate SAR maps of compounds from selected clusters, with the novelty that the R-groups are sorted according to their calculated LiRIf.[6] From our viewpoint, this sorting is in good agreement with the classification that is traditionally amenable to medicinal chemists, which is typically manually configured and based on experience. For example, a table of the R-substituents of compounds from selected clusters in the heat maps in Figures 11c and 11d (blue-squared) is shown in Figure 11e. The R2 substituents are displayed horizontally, with the cells colored according to the LRI pattern of the benzene substitution (as in the heat map in Figure 11c and d). The R1

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

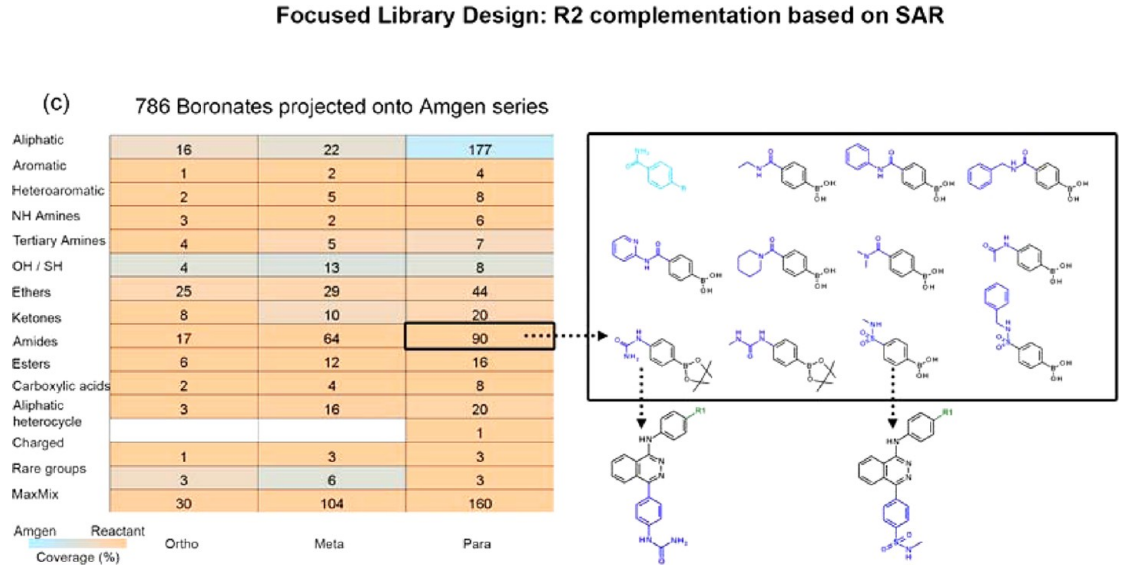


Figure 13. Focused library design. Projection of 786 aromatic boronic acids and boronates onto the 211-membered Amgen set from Figures 11c-11d. Exemplary reactants of representative subclasses from the black-squared cluster corresponding to phenyl rings monosubstituted with amide-like interaction patterns at the *para* position are shown on the right.

substituents, displayed vertically, are colored according to the AP class of the attachment point (oxygen-linked). The cell color coding corresponds to the selectivity values, and the gray cells correspond to the synthesized compounds without any activity value.

Once the SAR/SPR analysis is performed, the BRCS navigator goes one step beyond and enables medicinal chemists to explore those areas of interest that are identified through the SAR analysis and to determine which compound(s) to "synthesize" next. Two different options may be available: i) focused library acquisition or/and ii) focused library design.

A side-by-side library comparison can also be useful for compound acquisition when biological data are available. In Figure 12a, a heat map equivalent to those in Figures 11c and 11d is shown for the 959 commercial phthalazinamines after repeating the same filtering steps explained above for the Amgen data set and retaining the 253 structures with benzene substituents attached to the R2 position. As shown in this map, the diversity of the substitution patterns is very restricted. Focusing on a given cluster with an intermediate potency range (Figure 11c-d), for example, the ethers at the *para* position, R-group substituents common to the proprietary set and the purchasable library (e.g., a para methoxy phenyl group) can be identified. After selecting the 23 compounds that have this *para* methoxy phenyl group at R2, the corresponding highest level of the BRCS heat map for the exploration of the R1 position can be obtained (Figure 12b). For comparison, the 4 compounds from the Amgen set containing this substructure at R2 are included (blue-colored cell in Figure 12b). The screening compounds that provide alternatives to the heteroaromatic LRI pattern of the Amgen compounds are depicted in the blue square region. As a reference, the compound having the same R1 substitution as AMG 900 is included (blue-colored R1 group). A final decision on whether to purchase these commercially available compounds should be established based on the conclusions drawn from the total SAR analysis of the R1 position (data not shown) or according to the known binding mode (if available).[38]

Alternatively, the module implemented in the BRCS navigator to map lists of reagents can also be applied at lower BRCS levels to suggest synthetic alternatives. This tool enables the design of a focused library that is based on commercially available reagents that are selected after the corresponding retrosynthetic analysis to ensure chemical feasibility and ability to obtain target compounds, with the goal of achieving deeper knowledge around the clusters of interest that are unexplored by previous proprietary compounds and/or commercially available molecules. For example, in Figure 13, the projection of the 211 compounds shown in Figure 11c-d is complemented with available boronates and boronic acids. At least one commercial boronate or boronic acid is accessible to almost all of the LRI substitution patterns, with the exception of aliphatic heterocycles. As an example, the analyst may review the 89 reagents that specifically have the amide-like interaction pattern in the *para* position (with only 1 proprietary compound by Amgen, Figure 11c-d, corresponding to the primary amide), as exemplified in the black squared area close to the heat map.

CONCLUSION

In this report, we have introduced a new and intuitive visualization tool that is based upon established ligand–receptor interaction (LRI) types: the BRCS navigator, which aids in conducting a more efficient drug discovery project and avoiding misleading navigation through infertile areas of the chemical universe.

The BRCS navigator analyzes and consistently represents entire congeneric series in a context-independent space, regardless of the number of analogues, substitution sites, and annotated or calculated fields. This representation enables unprecedented data set comparisons, with applications in library acquisition and patent analysis that provide a clear positive impact on the drug discovery process. Commercially available reagent lists, covering a variety of reaction types, are also capitalized by LiRIf; therefore, their representation in the BRCS also leads to the comparison and selection of those reagents that fit the project requirements, whether diverse or

dx.doi.org/10.1021/ci3004539 | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

focused. Thus, when also taking chemical feasibility into account, this navigator may drive an optimal library design.

Traditional SAR navigation is enhanced by integrating, in a simultaneous application, three SAR extraction methods: dimensionality reduction, substructure annotation, and clustering. Thus, the BRCS navigator leads to interactive information-rich heat maps, which are reference-independent activity landscapes that enable the identification of critical ligand−receptor interactions (LRI) and substructural features as well as activity cliffs. Classical SAR tables can be depicted from selected cluster(s) at any navigation level; furthermore, R-groups placed on the *x*- and *y*-axis are sorted according to their calculated LiRIf. From our viewpoint, this sorting is in good agreement with the classification that is traditionally amenable to medicinal chemists and clearly eases its interpretation. Once the SAR/SPR analysis is performed, the BRCS navigator goes one step farther and enables medicinal chemists to explore those areas of interest and design which compound(s) to synthesize next. In addition, mining the structural information that is contained in competitors' patents leads also to the identification of preferred substitution patterns, which is useful for SAR transfer.

Exemplary applications based on a real, project-based data set for a congeneric series has been conducted to illustrate the positive impact of the versatile BRCS navigator on i) patent analyses, ii) library acquisition, iii) SAR analysis, and iv) library design. These illustrated cases suggest that user customization, interactivity, and ease of interpretation are very attractive aspects of the BRCS navigator as a guidance tool for medicinal chemists in the iterative drug discovery process. However, plugging into large databases, in which chemical-biological information is annotated or/and data from patents is included (e.g., SciFinder,[37] KKB,[27] pharmaceutical companies, ChEMBL,[39] and PubChem[40]), as well as into repositories where commercially available screening compound collections and reagents are annotated (e.g., ZINC),[25] may make the BRCS navigator an extremely useful tool.

The BRCS navigator is routinely utilized in our drug discovery process; nevertheless, there are still several points to implement and/or improve, e.g., i) permit a third R-group to be simultaneously visualized together with the other two (currently, this is tedious); ii) use the LiRIf for quantitative similarity or SAR (QSAR) analyses; and iii) allow scaffold analysis (mainly focused on library acquisition purposes), specially as the current version involves manual revision to define the scaffold of interest when comparing different data sets. For this latter point, we may utilize known techniques, such as Scaffold Hunter[7] or Scaffold Explorer,[8] together with the BRCS navigator. We are currently in the process of further extending the LiRIf descriptor to capitalize on all of the key features from scaffolds. Upon completion of this extension, the method will be made available; however, interested readers in the current version of the BRCS navigator may contact the authors to get access.

■ ASSOCIATED CONTENT

Supporting Information

A live demonstration video of the BRCS navigator is provided as an avi file, in which some exemplary applications reported in the manuscript are shown. Figure S1 describes separately the highest level of the BRCS analysis of the R-groups, from the R1 and R2 diversity points, extracted from the Novartis and Amgen proprietary data sets. Figure S2 reports the fragment frequency analysis of Amgen's patents. Figure S3 reports the detailed representation of the BRCS at deeper levels and utilized for a library design that is focused on R2, including reagent selection (950 proprietary compounds complemented with different lists of reagents). Finally, a mol file containing the data set reported in Table 1 is also provided. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +34 948 194700. E-mail: julenoyarzabal@unav.es.

Funding

The authors declare no competing financial interest.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Foundation for Applied Medical Research (FIMA) of the University of Navarra for financial support as well as Alvaro Bonet and Santiago Echevarria from the School of Communication (University of Navarra) for their help in editing the video. In addition, we thank Dr. Steven Muskal, Eidogen-Sertanty Inc., for providing access to the KKB database and permission to distribute the data set (mol file). This work was partially supported by MINECO and FSE (Inncorpora-Torres Quevedo grant), PTQ-1-04781.

■ REFERENCES

(1) Kirkpatrick, P.; Ellis, C. Chemical Space. *Nature* **2004**, *432*, 823.

(2) Hert, J.; Irwin, J. J.; Laggner, C.; Keiser, M. J.; Shoichet, B. K. Quantifying Biogenic Bias in Screening Libraries. *Nat. Chem. Biol.* **2009**, *5*, 479−483.

(3) Wetzel, S.; Bon, R. S.; Kumar, K.; Waldmann, H. Biology Oriented Synthesis. *Angew. Chem., Int. Ed.* **2011**, *50*, 10800−10826.

(4) Hartenfeller, M.; Eberle, M.; Meier, P.; Nieto-Oberhuber, C.; Altmann, K. H.; Schneider, G.; Jacoby, E.; Renner, S. Probing the Bioactivity-Relevant Chemical Space of Robust Reactions and Common Molecular Building Blocks. *J. Chem. Inf. Model.* **2012**, *52*, 1167−1178.

(5) Dobson, C. M. Chemical Space and Biology. *Nature* **2004**, *432*, 824−828.

(6) Rabal, O.; Oyarzabal, J. Using Novel Descriptor Accounting for Ligand−Receptor Interactions To Define and Visually Explore Biologically Relevant Chemical Space. *J. Chem. Inf. Model.* **2012**, *52*, 1086−1102.

(7) Wetzel, S.; Klein, K.; Renner, S.; Rauh, D.; Oprea, T. I.; Mutzel, P.; Waldmann, H. Interactive Exploration of Chemical Space with Scaffold Hunter. *Nat. Chem. Biol.* **2009**, *5*, 581−583.

(8) Agrafiotis, D. K.; Wiener, J. J. M. Scaffold Explorer: An Interactive Tool for Organizing and Mining Structure−Activity Data Spanning Multiple Chemotypes. *J. Med. Chem.* **2010**, *53*, 5002−5011.

(9) Gupta-Ostermann, D.; Hu, Y.; Bajorath, J. Introducing the LASSO Graph for Compound Data Set Representation and Structure-Activity Relationship Analysis. *J. Med. Chem.* **2012**, *55*, 5546−5553.

(10) Kettle, J. G.; Ward, R. A. Toward the Comprehensive Systematic Enumeration and Synthesis of Novel Kinase Inhibitors Based on a 4-Anilinoquinazoline Binding Mode. *J. Chem. Inf. Model.* **2010**, *50*, 525−533.

(11) 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.

(12) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. Navigating Structure-Activity Landscapes. *Drug Discovery Today* **2009**, *14*, 698−705.

(13) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure−Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209−8223.

(14) Stumpfe, D.; Bajorath, J. Methods for SAR Visualization. *RSC Adv.* **2012**, *2*, 369−378.
(15) Agrafiotis, D. K.; Shemanarev, M.; Connolly, P. J.; Farnum, M.; Lobanov, V. S. SAR Maps: A New SAR Visualization Technique for Medicinal Chemists. *J. Med. Chem.* **2007**, *50*, 5926−5937.
(16) Wassermann, A. M.; Peltason, L.; Bajorath, J. Computational Analysis of Multi-Target Structure−Activity Relationships To Derive Preference Orders for Chemical Modifications toward Target Selectivity. *ChemMedChem* **2010**, *5*, 847−858.
(17) Wassermann, A. M.; Bajorath, J. Directed R-group Combination Graph: A Methodology to Uncover Structure−Activity Relationship Patterns in Series of Analogues. *J. Med. Chem.* **2012**, *55*, 1215−1226.
(18) Wawer, M.; Lounkine, E.; Wassermann, A. M.; Bajorath, J. Data Structures and Computational Tools for the Extraction of SAR Information From Large Compound Sets. *Drug Discovery Today* **2010**, *15*, 630−639.
(19) Tyrchan, C.; Boström, J.; Giordanetto, F.; Winter, J.; Muresan, S. Exploiting Structural Information in Patent Specifications for Key Compound Prediction. *J. Chem. Inf. Model.* **2012**, *52*, 1480−1489.
(20) Hattori, K.; Wakabayashi, H.; Tamaki, K. Predicting Key Example Compounds in Competitors' Patent Applications Using Structural Information Alone. *J. Chem. Inf. Model.* **2008**, *48*, 135−142.
(21) Oyarzabal, J.; Howe, T.; Alcazar, J.; Andres, J. I.; Alvarez, R. M.; Dautzenberg, F.; Iturrino, L.; Martinez, S.; Van der Linden, I. Novel Approach for Chemotype Hopping Based on Annotated Databases of Chemically Feasible Fragments and a Prospective Case Study: New Melanin Concentrating Hormone Antagonists. *J. Med. Chem.* **2009**, *52*, 2076−2089.
(22) Saluste, G.; Albarran, M. A.; Alvarez, R. M.; Rabal, O.; Ortega, M. A.; Blanco, C.; Kurz, G.; Salgado, A.; Pevarello, P.; Bischoff, J. R.; Pastor, J.; Oyarzabal, J. Fragment-Hopping-Based Discovery of a Novel Chemical Series of Protooncogene PIM-1 Kinase Inhibitors. *PLoS One* **2012**, *7*, e45964.
(23) Rabal, O.; Urbano-Cuadrado, M.; Oyarzabal, J. Computational Medicinal Chemistry in Fragment-Based Drug Discovery: What, How and When. *Future Med. Chem.* **2011**, *3*, 95−134.
(24) *Pipeline Pilot*, version 8.5; Accelrys, Inc.: San Diego, CA, 2011.
(25) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC - A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(26) *Accelrys Available Chemicals Directory (ACD)*; Accelrys, Inc.: San Diego, CA, 2012.
(27) *Kinase Knowledgebase (KKB)*; Eidogen-Sertanty, Inc.: San Diego, CA, 2011.
(28) Bold, G.; Frei, J.; Traxler, P.; Altmann, K. H.; Mett, H.; Stover, D. R.; Wood, J. M. Phthalazines with angiogenesis inhibiting activity. PCT WO 98/35958 (A1), August 20, 1998. 113 pp.
(29) Bold, G.; Dawson, K. J.; Frei, J.; Heng, R.; Manley, P. W.; Wietfeld, B.; Wood, J. M. Phthalazine derivatives for treating inflammatory diseases. PCT WO 00/59509 (A1), October 12, 2000. 127 pp.
(30) Bold, G.; Manley, P. W. Phthalazines with angiogenesis inhibiting activity. PCT WO 02/090346 (A1), November 14, 2002. 67 pp.
(31) Cee, V. J.; Deak, H. L.; Du, B.; Geuns-Meyer, S. D.; Hodous, B. L.; Nguyen, H. N.; Olivieri, P. R.; Patel, V. F.; Romero, K.; Schenkel, L. Aurora Kinase Modulators and Method of Use. PCT WO 2007/087276 (A1), January 22, 2007.189 pp.
(32) Cee, V. J.; Deak, H. L.; Geuns-Meyer, S. D.; Du, B.; Hodous, B. L.; Martin, M. W.; Nguyen, H. N.; Olivieri, P. R.; Panter, K.; Romero, K.; Schenkel, L.; White, R. Aurora Kinase Modulators and Method of Use. PCT WO 2008/124083 (A2), October 16, 2008. 211 pp.
(33) White, R.; Human, J. B. Aurora Kinase Modulators and Method of Use. PCT WO 2009/117157 (A1), September 24, 2009. 94 pp.
(34) Deak, H. L.; Geuns-Meyer, S. D.; Human, J. B.; Martin, M.; Marx, I. Aurora Kinase Modulators and Method of Use. PCT WO 2010/017240 (A2), February 11, 2010. 77 pp.
(35) Wood, J. M.; Bold, G.; Buchdunger, E.; Cozens, R.; Ferrari, S.; Frei, J.; Hofmann, F.; Mestan, J.; Mett, H.; O'Reilly, T.; Persohn, E.; Rösel, J.; Schnell, C.; Stover, D.; Theuer, A.; Towbin, H.; Wenger, F.; Woods-Cook, K.; Menrad, A.; Siemeister, G.; Schirner, M.; Thierauch, K. H.; Schneider, M. R.; Drevs, J.; Martiny-Baron, G.; Totzke, F. PTK787/ZK 222584, a Novel and Potent Inhibitor of Vascular Endothelial Growth Factor Receptor Tyrosine Kinases, Impairs Vascular Endothelial Growth Factor-induced Responses and Tumor Growth after Oral Administration. *Cancer Res.* **2000**, *60*, 2178−2189.
(36) Payton, M.; Bush, T. L.; Chung, G.; Ziegler, B.; Eden, P.; McElroy, P.; Ross, S.; Cee, V. J.; Deak, H. L.; Hodous, B. L.; Nguyen, H. N.; Olivieri, P. R.; Romero, K.; Schenkel, L. B.; Bak, A.; Stanton, M.; Dussault, I.; Patel, V. F.; Geuns-Meyer, S.; Radinsky, R.; Kendall, R. L. Preclinical Evaluation of AMG 900, a Novel Potent and Highly Selective Pan-Aurora Kinase Inhibitor with Activity in Taxane-Resistant Tumor Cell Lines. *Cancer Res.* **2010**, *70*, 9846−9854.
(37) *Scifinder, web version*; Chemical Abstracts Service: Columbus, OH, 2012.
(38) Cee, V. J.; Schenkel, L. B.; Hodous, B. L.; Deak, H. L.; Nguyen, H. N.; Olivieri, P. R.; Romero, K.; Bak, A.; Be, X.; Bellon, S.; Bush, T. L.; Cheng, A. C.; Chung, G.; Coats, S.; Eden, P. M.; Hanestad, K.; Gallant, P. L.; Gu, Y.; Huang, X.; Kendall, R. L.; Lin, M. H.; Morrison, M. J.; Patel, V. F.; Radinsky, R.; Rose, P. E.; Ross, S.; Sun, J. R.; Tang, J.; Zhao, H.; Payton, M.; Geuns-Meyer, S. D. Discovery of a Potent, Selective, and Orally Bioavailable Pyridinyl-Pyrimidine Phthalazine Aurora Kinase Inhibitor. *J. Med. Chem.* **2010**, *53*, 6368−6377.
(39) *ChEMBL*; European Bioinformatics Institute (EBI): Cambridge, UK, 2010. http://www.ebi.ac.uk/chembl (accessed July 31, 2012).
(40) *PubChem Compound Database*; National Center for Biotechnology Information (NCBI). http://pubchem.ncbi.nlm.nih.gov/ (accessed July 31, 2012).