Kinase Patent Space Visualization Using Chemical Replacements

Noel T. Southall* and Ajay

Informatics Department, Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850

Received November 29, 2005

Here we present a methodology for characterizing the structure of patented chemical space. This approach identifies those chemical replacements that can connect sets of exemplified compounds in individual patents. Chemists can then search these replacements to help them discover the architecture within their patent space of interest. To demonstrate the utility of such an approach, we characterize a set of kinase inhibitors from patents and literature and find that many companies' patents can be understood to be straightforward modifications of competitors' patents. By reapplying these same chemical themes to other related compound series, novel, biologically active compounds can be discovered.

Introduction

A key asset of any commercial drug discovery program is the patent protection of its active pharmaceutical ingredients. Patent protection gives the patent holder the right to exclude competitors from marketing the same product and time to recoup the research and development costs of bringing a drug to market. Companies also file patents for a variety of strategic reasons including for use as a negotiation tool, to block competitors, to prevent lawsuits, as a means to improve apparent productivity, etc.[1] Given the value of patent protection, companies try to file for patent protection early in a program's life cycle, as soon as a potential drug activity is identified. This is usually well before the true safety and efficacy profile of any compound is understood, and given the difficulty of discovering a safe and efficacious drug, it is extremely rare that the initially identified compound will actually be the one brought to market. Thus, companies will usually try to protect not only one specific composition of matter but a series of related compounds, or even several patents of related chemical series, to hedge their bets that a commercially useful compound will be contained within that patent space. It is not uncommon to find that the patent space around a given compound is crowded with several companies each having filed multiple patents.

However, chemical space is so vast that there are often "holes" or closely related compounds that possess the same biological activity but are still available to be patented. And it can be extremely profitable to identify these holes in another's patent coverage, for example, identifying vardenafil (compound **2**, the active pharmaceutical ingredient in Bayer's Levitra) given Pfizer's sildenafil (compound **1**, Viagra) patent space (Figure 1). Me-too and "fast follower" drugs such as Levitra utilize lessons from their predecessors about how to effect a biological response, and fold in their own understanding about holes in the patent space to rapidly create value and improved drug properties.[2] Understanding how to exploit holes in patent space is also important when screening hits are evaluated. Each screening hit entering follow-up requires that a defendable intellectual property position be identified. For a naïve chemistry team following up on a screening hit in a crowded patent space, this is a challenging task.

Several tools exist to help search patented chemical space, and there are many ways to characterize the patent space around a given compound. There is extensive work based on text mining approaches that help categorize and search patents, like DOLPHIN[3] and DerwentAnalytics.[4] Obviously, compound structural searches are also a core part of examining patents. These include searching by substructure, by Markush structure, or by similarity.[5−7] The structure space around a compound is often visualized by similarity and clustering approaches to characterize how crowded that space actually is. There are, of course, multiple definitions of similarity[8] and many clustering methods available to organize compounds into sensible groups, and a range of tools exist to organize and browse such collections. One such method based on simple definitions of scaffolds has been found to be useful in organizing and browsing a large number of compounds and creating new molecules that are synthesizable.[9,10] A parallel approach is to project compounds onto 2D space with the idea that such projections maintain the "true" distance between molecules typically calculated in very high dimensional space. Statistical techniques such as linear and nonlinear multidimensional scaling (MDS) and Sammon maps have also been used.[11] Here the idea is that such 2D projections will point out "holes" in patent space which chemists can fill. One way to utilize such a tool is to take an idea for a new molecule, project it onto this 2D space and if distances are preserved it can give the chemist a clue toward whether she is moving in directions that overlap with other patented molecules.[12]

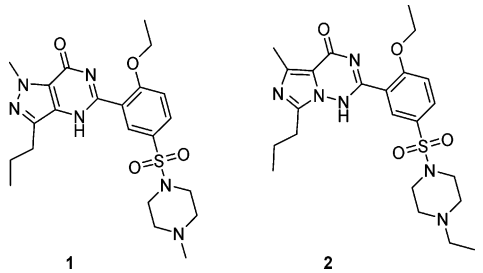

Figure 1. Chemical structures of sildenafil (**1**) and vardenafil (**2**), closely related but different enough to warrant coverage by separate patents.

However, translating this information into knowledge about an available chemical space remains an art. In part, this is because it requires a patent agent to parse the often obfuscated text of a patent, interpret its coverage, and assess obviousness.[6] But the other problem is that these structure-based search tools and manipulations that follow do not try to characterize how

* Corresponding author: phone 240-453-3743; fax 240-453-3303; e-mail noel.southall@celera.com.

10.1021/jm051201m CCC: $33.50 © 2006 American Chemical Society
Published on Web 02/23/2006

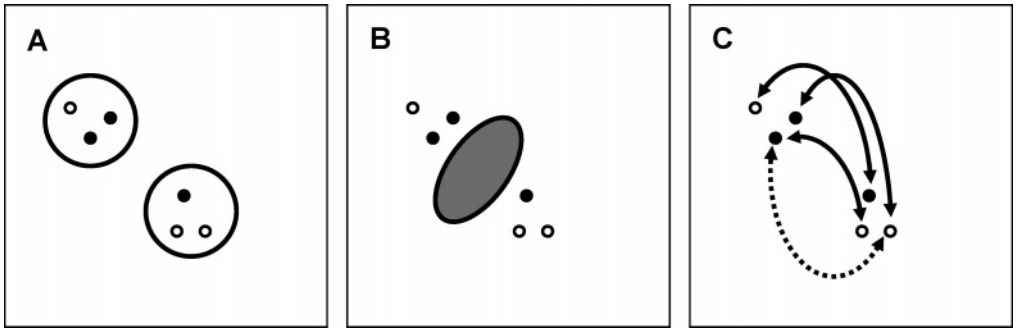

Figure 2. Pictorial representation that compares the current approach to previous methods. Compounds are represented as dots in two dimensions; their location is determined by projecting compounds in similarity space onto a two-dimensional area: (●) compounds from patent 1; (○) compounds from patent 2. (A) Similar compounds are clustered together (enclosed in black circles). Notice that a mix of compounds from patents 1 and 2 fill both clusters; the patents appear to be on top of one another. (B) 'Holes' in patent space (shaded oval) are detected by demarcating regions of similarity space devoid of any compounds. Whether any truly patentable, synthetically accessible, and biologically active compounds exist within this space is not known. (C) Present approach. Related compounds are mapped between patents. Three examples of one chemical replacement (solid arrows) and one other replacement (dashed arrow) are observed. The solid arrow replacement is the most important theme in that it maps all of the compounds in one patent series to compounds in the other series. In other words, unlike in panel A, the solid arrow is what relates the two patents to one another.

patents are related to one another. These tools are very useful for addressing the legal status of a given compound (the compound is/is not covered) but are less useful for vetting a chemical theme, or how a compound series relates to others in the literature. They can identify when patents appear to be closely related (by similarity) but leave it to the chemist to identify the chemical story that separates them. For example, there are often constraints imposed by a synthetic route or the availability of key reagents that narrow the set of claimed compounds. If one can identify such limitations and ways to avoid them, then they can generate unique, patentable compounds. Another related strategy that also appears frequently in the patent (and other) literature is use of isosterism, taking each of the compounds in a series and substituting part of their structure with an isosteric group, and then patenting this novel series. Of course, to protect against this, most patent applications will try to cover the most obvious isosteres within a series. Nonetheless, these patterns of regular substitutions into a chemical series do appear often (as with **1** and **2** above) and provide an interesting type of signal to train a computer to identify. Together, these chemical stories provide a useful way of understanding how patent space is organized.

Here we present a method for characterizing the structure of patent space by identifying recurring chemical stories across patents and companies, organizing a database of compounds by cataloging the chemical replacements that relate pairs of molecules. In particular, we want to discover those cases where, for instance, patent A has 20 compounds that differ from patent B's 20 compounds by substitution of a nitrogen into the scaffold in a regular way, for example, pyrimidine to a triazine. It is not enough to identify that one compound from patent B is similar to a compound from patent A and should be plotted next to it on a graph; rather, we want to characterize the patent strategies that *differentiate* the two patents by identifying the recurring chemical stories that connect the two patents. Presenting these types of stories to the chemist helps them to understand and explore patent space, and ultimately can help them devise a way to avoid covered regions of chemical space.

Methods

As mentioned above, clustering-based procedures provide the most obvious route to grouping compounds. Some clustering methods such as hierarchical clustering also show the relationships between clusters. Molecules in this space do not necessarily have distance relationships maintained; however, cluster distances are knowable. On the other hand, every pair of molecules (lines connecting two molecules) in the 2D MDS-type projections maintain their distance relationships. This type of representation also gives you a visual handle on "holes" in this 2D space that can be filled with appropriately selected molecules added to the database.[13] In the present approach, we organize a database of compounds by cataloging the chemical replacements that relate pairs of molecules. In this representation, the lines connecting molecules (edges) represent a transformation (no implicit directionality is present) that can be applied to one molecule to produce the other. As is obvious from Figure 2, it is difficult to denote all computed edges on such graphs and retain clarity, and we do not produce such graphs here. We note, however, that various navigation aids such as visually organizing the transformations, using color to group lines that are the same transformation, and other graph layout algorithms could be used to generate more useful, interactive representations for the chemist.

Overview of Method. The results presented in this work are for a set of kinase inhibitors obtained from patents and journal articles. A pairwise Tanimoto similarity is calculated for all molecules based on topological torsions.[14] This Tanimoto similarity is used to prioritize maximum common substructure (MCS) calculations, which are performed on a compute farm. The details of our implementation of the MCS algorithm are similar to other published algorithms,[15,16] apart from some heuristic differences, which will be published elsewhere.[17] After maximum common substructures are identified, a canonical chemical replacement is calculated, and a database is populated with these data. Chemical replacements can then be retrieved by querying with a replacement or piece of a replacement, a compound structure or substructure, by patent, or by company name.

Patented Kinase Inhibitors. The kinase patent literature is notoriously complicated and provides an ideal test for the kind of tool presented here. A set of patent and literature kinase inhibitors, assay values, and references was purchased from GVK Biosciences.[18] The set contains 116 550 unique structures of compounds exemplified in patents and also from primary literature. Each structure was mapped to a single company and patent by identifying the earliest publication of a compound and the company connected with that publication. Obviously, this was not always accurate as the patent coverage was incomplete, and the company affiliation of applicants was not always obvious. Approximately 90% of compounds were assigned to a patent in this way. This automatic assignment of a compound to a patent was checked for accuracy by comparing individual assignments with other sources, such as Prous' Integrity database,[19] and by direct mining of the patent literature. In addition, a lookup table was generated to translate original assignees to a probable current assignee to reflect current

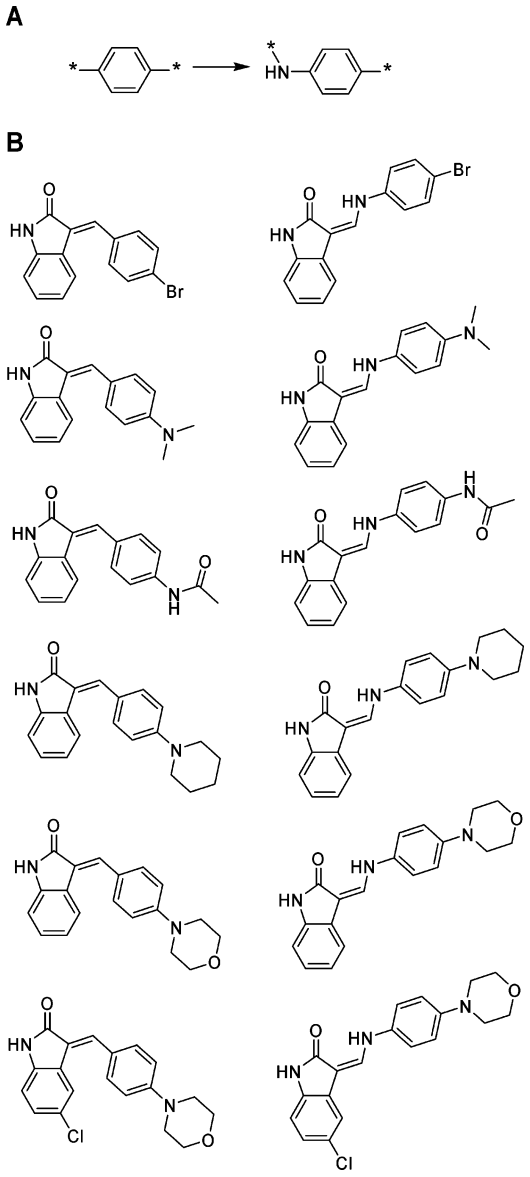

Figure 3. (A) Example of a chemical replacement. In the example, a nitrogen has been inserted into the scaffold of the other series. The starred atoms reflect connections back to isomeric scaffold atoms. (B) This replacement connects these 12 molecules from the chemical series of two patents, Sugen's (now Pfizer's) U.S. 6225335 B (compounds on left) and Bristol-Myers Squibb's WO 03/27102 A1 (compounds on right). It would appear that Bristol-Myers Squibb's patented compounds were influenced or are related to Sugen's published series.

patent rights. This table essentially contains a list of recent pharmaceutical company mergers, as well as several collaborations.

Maximum Common Substructure. Several recent articles have published algorithms for calculating MCS and heuristics for speeding up such a calculation. We have implemented one such algorithm here.[17] The implementation can calculate both the largest connected and disconnected isomorphisms between graphs. The former is much faster than the latter and is useful for identifying chemical replacements at the exterior of a molecule, whereas the latter is needed to identify small changes to the core of a molecule (in the worst-case scenario, the largest connected isomorphism would be only half of the molecule, even though only a single atom might be different between the two molecules; for example, see Figure 3). MCS has been used in the literature for multiple purposes, including as a similarity metric[15] and for pharmacophore perception type problems,[20] and Sheridan[21] has recently used it to catalog a set of common replacements found in the MDDR[22] (MDL Drug Data Report) database.

Even with recent improvements in the speed of these algorithms, such a calculation can still be quite slow, especially for certain pathological cases. To avoid these pathological cases, we abort a calculation after 30 s of calculation time. The vast majority of calculations took far less than this; in fact, only about 8400 out of roughly 5 million comparisons were aborted in this way. Of course, it is still not possible to calculate the maximum common substructure for all pairs of molecules from a set of 100 000 (5 billion comparisons). But because we are generally uninterested in large (nearly as large as the molecule) chemical replacements we can avoid doing the calculation at all on very dissimilar molecules. In practice, we find that using a 0.5 similarity in topological torsions identifies 90% of the molecule pairs with identifiable chemical replacements. In addition, only comparisons between molecules from different companies were performed. Discovering replacements within a company's patented series is interesting, but we do not focus on this problem here. These assumptions reduced the computational load to 5 million comparisons, which took 2 weeks on a 30, 2.4 GHz AMD Opteron processor compute farm to complete.

Chemical Replacement Perception. In his paper on common chemical replacements from the MDDR database, Sheridan[21] outlined a method for discovering recurring medicinal chemistry themes within a chemical database. For pairs of molecules in the database, he calculates the maximum common substructure between them, and he defines the remaining differences in structure as chemical replacements. By going through the MDDR database in this pairwise manner, he observed that certain transforms appeared repeatedly. For example, he found 188 pairs of molecules in the MDDR whose only difference was between a thiophene and a furan ring. The final histogram of all transforms observed in the MDDR reflected many basic medicinal chemistry strategies such as heteroatom substitution into an aromatic ring and identified many well-known bioisosteric groups. Although it is computationally expensive to calculate such a large number of maximum common substructures between molecules, in the end this approach was very successful at identifying common medicinal chemistry strategies and themes and was quite a coup for a computer program. Here, we apply a similar methodology to store, query, and organize a subset of the kinase patent space. There are two chief differences between that effort and the present one. First, instead of focusing on bioisosterism and counting the gross number of times that a chemical transform appears in any context, we focus on patent space and finding those transforms that can connect compound series between two patents or generally between two companies. Second and more costly, we need to calculate many more maximum common substructures. In the Sheridan work, comparisons were calculated between molecules only if they had the same activity label. Here we utilize the entire set of literature and patented kinase inhibitors, and in addition, we use a lower similarity threshold to discover more subtle relationships between compounds.

A chemical replacement was calculated from the maximum common substructure between two compounds. Backtracking was done to include whole rings and other groups as per the algorithm specified in Sheridan.[21] Chemical replacements were only calculated if the common substructure was at least as large as half of both molecules. See Figure 3 for an example of a chemical replacement. Once a replacement is identified, its representation is canonicalized into the SMILES notation by use of ChemAxon's JChem package,[23] and these canonical smiles are used as the unique key into the database.

Results and Discussion

For the set of 116 550 kinase compounds, roughly 5 million MCSs were calculated. From these calculations, 820 000 chemical replacements were identified that relate 40 000 compounds. However, no replacements were identified for nearly 80 000 of the compounds in this set. Apart from concerns about whether this kinase compound set comprehensively covers the patent and primary literature, this is most likely due to the case where

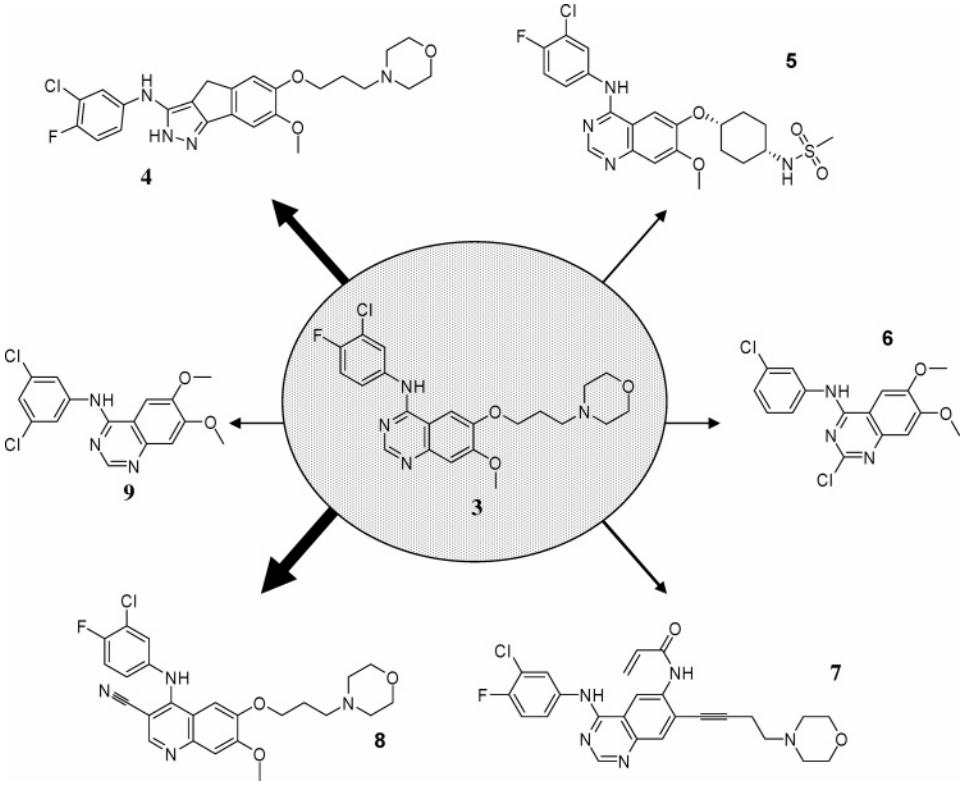

Figure 4. Example of some of the compounds related to gefitinib (AstraZeneca's Iressa, compound **3**) by frequent chemical replacements. Arrow edge widths are proportional to the number of replacement examples connecting compounds in AstraZeneca's patent for gefitinib to other companies' patents. Compounds were cited as follows: **3**, AstraZeneca WO 96/33980 A1; **4**, Johnson & Johnson U.S. 2004/0082639 A1; **5**, Boehringer Ingelheim WO 03/082290 A1; **6**, Pfizer US 5981569 A; **7**, Mitsubishi WO 02/066445 A1; **8**, Wyeth U.S. 6002008 A; **9**, Sugen (now Pfizer) *Bioorg. Med. Chem.* **1996**, *4*, 1203.

no other company has attempted a fast-follower strategy for these series. Because replacements were not calculated for compounds from the same company, and these series are not related to other companies' series, our algorithm fails to identify any replacements for these molecules. On the other hand, for compounds where a replacement was identified, that compound was also connected to 40 other compounds, on average. Thus, where a fast-follower strategy might be employed, we obtained a richly nested set of replacements connecting these molecules.

Figure 4 nicely illustrates this latter case, where several competitors may be found close by in patent space. Given a starting compound (gefitinib, compound **3**), we ask, what related patents are adjacent to this compound? A simple answer to this question might be found by generating a list of similar compounds. Here, we frame the answer by looking for series of chemicals that are related by chemical replacements, which hopefully will provide better insights into how patents are organized and are related to one another. Practically, the maximum common substructure approach is successful at identifying these motifs because medicinal chemistry is a component-based methodology for exploring chemical space. Stepwise reactions and preferred coupling points naturally lead to delineated units of structure, otherwise known as scaffolds, that are invariant across a series and are easily identified by chemists and this algorithm alike. Similarity approaches, in contrast, rob the analysis of any structural context by translating their results into a numerical representation. Similarity has been found to be useful in organizing molecules. Here we argue for a parallel approach where a database of compounds is stored along with the relationships between molecules. We can then track how often a replacement occurs in the database and support querying of replacements by structure.

Interestingly, of the 820 000 replacements identified, less than 70 000 replacements were found more than once. This means that even for a highly nested set of connections between related compound series, there are a much small number of replacements that represent the major themes connecting these series. This can be seen in Figure 4, with the arrow widths representing the occurrence of the different replacements. While there are replacements connecting **3** to many compounds, the replacements connecting it to **8** and **4** emerge as the major theme connecting AstraZeneca's compounds to related series from Wyeth and Johnson & Johnson. Cataloging how often certain replacements occur helps resolve patent strategies from more random associations between compounds. If two patents' compound series are related, then by luck, it will often be possible to construct some type of contorted replacement to turn a compound of one series into any of the compounds in the other series. However, it is much more powerful to identify the single replacement that pairs off all of the compounds in the two series.

If we compare our results with those of Sheridan and his work with the MDDR, we see that there is an important difference in the types of chemical replacements identified. The most common chemical replacement he identified was changing a single chlorine to a methyl in a molecule. For our purposes, these are the most uninteresting replacements. In contrast, consider the most common replacement found in Figure 5. By focusing on comparisons between molecules of different companies and patents, we obtain only those replacements worthy of having a new patent filed. In general, this means that we identify core changes to a molecule instead of peripheral ones, and the resulting replacements are nonobvious, much more

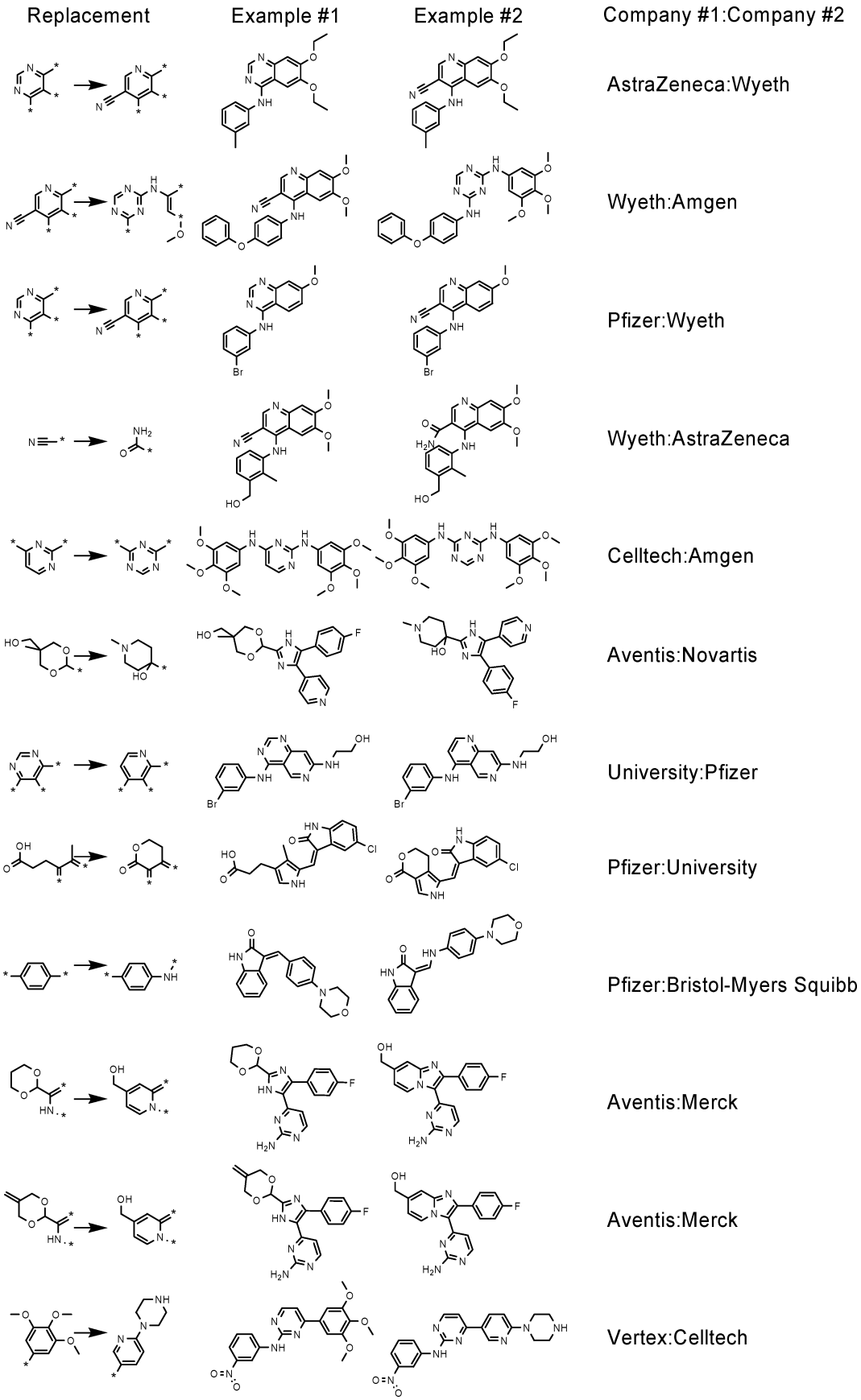

Figure 5. Survey of the most frequent replacements between companies. Replacements are ranked by the total number of examples that connected the two companies. The chemical replacement, a typical compound from the first company, a typical compound from the second company, and the company names are given. The starred atoms reflect connections back to isomeric scaffold atoms. The most common replacement observed (first row) connects 70 molecules from AstraZeneca (left) and Wyeth (right) including chemical series across several patents (AstraZeneca's U.S. 5457105 A and U.S. 5580870 A and Wyeth's U.S. 6002008 A and WO 00/18761 A1). This replacement appears to be a key strategy used by Wyeth (right) to differentiate itself from AstraZeneca (left) and the well-known quinazoline series.

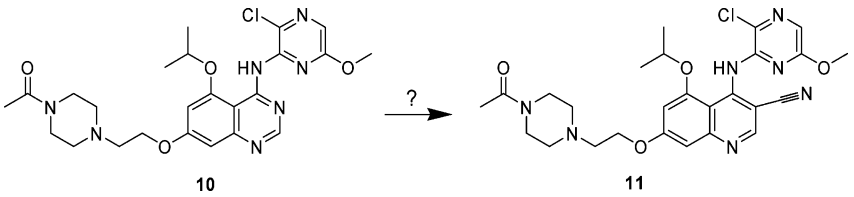

Figure 6. Potent, selective inhibitor of c-Src, **10**, and a related compound, **11**, constructed by applying the most common replacement from Figure 5 to this molecule. Applying common replacements to known compounds may help one to discover other novel, biologically active compounds.

inspiring to the medicinal chemist, and likely to be more relevant for a patent agent.

As one surveys the results closely, it becomes apparent that some replacements are conservative and would not be expected to affect biological activity in a profound way, while other replacements are quite profound and it is not obvious that biological activity would be retained. Yet, it is still an interesting exercise to at least consider the possibility that some of these profound replacements conserve biological activity, that is, that both series bind kinases in the same way. For instance, consider whether **3** could bind in a similar mode to a kinase as **4**. The insertion of an extra ring into the quinazoline scaffold is not a conservative change, but for this particular example we know that activity of this series against several kinases and cell lines is retained, though the spectrum of activity may have changed.[24,25] Alternatively, consider whether the second set (row) of compounds in Figure 5 both bind in the same way. Both series exhibit activity against multiple kinases.[26,27] And even if there is no biological activity relationship between the chemical series, the fact remains that these related chemical spaces are covered by patents and need to be considered in the development of a patenting strategy.

The replacements identified here can also be employed to construct novel, biologically active compounds. Consider the compound series recently identified by a group at AstraZeneca that is potent but selective for c-Src kinase, compound **10**.[28] The quinazoline scaffold on which this compound is based is known to bind many different kinases. However, utilizing certain 5-substituted quinazolines seems to convey c-Src selectivity. In Figure 6, we then apply replacements that we have observed with other quinazolines, such as the first replacement of Figure 5, to construct a novel compound, **11**, which potentially has the same biological activity as the parent compound, **10**. Of course, simply applying a replacement does not guarantee novelty—this question can only be addressed by the expert patent agent—and additional modeling may be necessary before deciding whether to proceed with synthesis, but it does provide a very useful starting point for ideas.

The present approach does suffer from several limitations. First, patent coverage is not complete. The tool provides a useful exploration and hypothesis generation functionality but cannot be the ultimate arbitrator of patent strategy. This role remains reserved for the expert patent agent. Moreover, if every patent was covered, and in addition, not only exemplified compounds but all potential compounds were included in our compound set, then the number of compound comparisons required to perform such an analysis would probably be prohibitive. Of course, there are probably clever algorithmic improvements one could make to prioritize the most useful calculations, so it is probably more accurate to say that it would be difficult. One could focus exclusively on scaffold replacements, perhaps first clustering by scaffold and then comparing the scaffolds themselves. Also, we do not predict whether the biological activity of related series is the same. If the task was to predict whether a chemical series did have a similar biological activity as another, one could use other data and modeling approaches to hone in on more probable candidates, such as utilizing actual activity data from literature and patents or building structural models of activity.

In general, because these replacements have been observed in real molecules, they could be a useful basis set for de novo chemistry tools (better ensuring synthetic accessibility). Organizing a database of compounds from the perspective of replacements can potentially be profitably applied to other problems in chemistry, such as detecting toxicophores (chemical replacements that confer a toxic liability to a molecule) or characterizing SAR development (how related chemical series' activities are; see also ref 29). Indeed, clustering by chemical replacement seems to be a powerful algorithm for identifying chemical themes.

Conclusions

A methodology is presented for characterizing the structure of patented chemical space. We identify chemical replacements by calculating the maximum common substructure between similar pairs of molecules from different patents. We propose and exemplify how a catalogue of replacements that occur in the database is a useful way to summarize the structural information in the database. This strategy helps with the discovery of chemical themes others have used to design patents within a space of interest. To demonstrate the utility of such an approach, we have characterized a set of kinase inhibitors from patents and literature and find that many companies' patents can be understood to be straightforward modifications of competitors' patents. In the future, we hope that this approach might be extended to identify chemical themes in other contexts.

We have shown that encoding patent space in terms of patent strategies is useful as it ameliorates the problem of having to enumerate all the compounds that are covered by a patent (as enumerated in specific claims), and those that are potentially problematic (as covered in generic claims). One specific example of this is the following. Because we store patent strategies instead of enumerating specific points in chemical space, we avoid the problem of every follow-up question about a specific compound having to be passed back to the expert patent agent for clarification. We continue to work on other useful visual representations of these transforms.

References

(1) Cohen, W. M., et al. *NBER Work. Pap. Ser.* **2000** (w7552).
(2) Doggrell, S. A. Comparison of clinical trials with sildenafil, vardenafil and tadalafil in erectile dysfunction. *Expert Opin. Pharmacother.* **2005**, *6*, 75−84.
(3) DOLPHIN, http://scientific.thomson.com/currentpatents/.
(4) DerwentAnalytics, http://scientific.thomson.com/products/derwentanalytics/.
(5) Multiple articles in *J. Chem. Inf. Comput. Sci.* **1991**, *31* (1).
(6) Lynch, M. F.; Holliday, J. D. The Sheffield Generic Structures Project—a Retrospective Review *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 930−936.
(7) Downs, G. M.; Barnard, J. M. Techniques for Generating Descriptive Fingerprints in Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 59−61.

(8) Kitchen, D. B.; Stahura, F. L.; Bajorath, J. Computational techniques for diversity analysis and compound classification. *Mini Rev. Med. Chem.* **2004**, *4*, 1029−39.
(9) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−93.
(10) Ajay; Bemis, G. W.; Murcko, M. A. Designing libraries with CNS activity. *J. Med. Chem.* **1999**, *42*, 4942−51.
(11) Gedeck, P.; Willett, P. Visual and computational analysis of structure−activity relationships in high-throughput screening data. *Curr. Opin. Chem. Bio.* **2001**, *5*, 389−395.
(12) Walters, P. The impact of in silico methods on lead discovery. PharmaDiscovery 2005, Washington, DC, May 10−12, 2005.
(13) Agrafiotis, D. K. Stochastic proximity embedding. *J. Comput. Chem.* **2003**, *24*, 1215−21.
(14) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82−85.
(15) Raymond, J. W.; Gardiner, E. J.; Willett, P. Heuristics for Similarity Searching of Chemical Graphs Using a Maximum Common Edge Subgraph Algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 305−316.
(16) García, G. C.; Ruiz, I. L.; Gomez-Nieto, M. A. Step-by-Step Calculation of All Maximum Common Substructures through a Constraint Satisfaction Based Algorithm. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 30−41.
(17) Lippert, R.; Fain, B.; Nguyen, T.; Southall, N. T.; Ajay. Graph matching and Connected and Disconnected Maximum Common Substructure Calculations. Manuscript in preparation.
(18) GVK Biosciences, http://www.gvkbio.com.
(19) Prous Integrity, http://www.prous.com/integrity/.
(20) Willett, P.; Winterman, V.; Bawden, B. Implementation of nonhierarchic cluster analysis methods in chemical information systems: Selection of compounds for biological testing and clustering of substructure search output. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 109−118.
(21) Sheridan, R. P. The Most Common Chemical Replacements in Drug-Like Compounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103−108.
(22) Molecular Design Drug Data Report, version 99.1, distributed by Molecular Design Ltd., San Leandro, CA.
(23) ChemAxon, http://www.chemaxon.com/.
(24) Koizumi, F., et al. Synergistic interaction between the EGFR tyrosine kinase inhibitor gefitinib ("Iressa") and the DNA topoisomerase I inhibitor CPT-11 (irinotecan) in human colorectal cancer cells. *Int. J. Cancer* **2004**, *108*, 464−72.
(25) Ho, C. Y., et al. N-substituted tricyclic 3-aminopyrazoles as inhibitors for the treatment of cell proliferative disorders. U.S. Pat. Appl. 2004/0082639 A1.
(26) Zhang, N., et al. MEK (MAPKK) inhibitors. Part 2: structure−activity relationships of 4-anilino-3-cyano-6,7-dialkoxyquinolines. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1407−1410.
(27) Armistead, D. M., et al. Triazine kinase inhibitors. PCT Int. Appl. 1996, WO 01/25220 A1.
(28) Barlaam, B., et al. New heterocyclic analogues of 4-(2-chloro-5-methoxyanilino)quinazolines as potent and selective c-Src kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5446−5449.
(29) Sheridan, R. P., et al. Molecular transformations as a way of finding and exploiting consistent local QSAR. *J. Chem. Inf. Model.* **2005**, published online, http://dx.doi.org/10.1021/ci0503208.

JM051201M