JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Library Enhancement through the Wisdom of Crowds

Michael D. Hack,[†] Dmitrii N. Rassokhin,[‡] Christophe Buyck,[§] Mark Seierstad,[†] Andrew Skalkin,[‡] Peter ten Holte,[§] Todd K. Jones,[∥,†] Taraneh Mirzadegan,[*,†] and Dimitris K. Agrafiotis[*,‡]

[†]Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 3210 Merryfield Row, San Diego, California 92121, United States

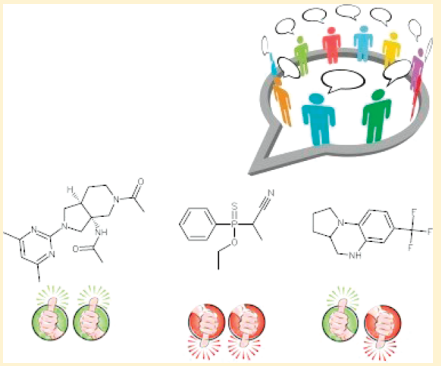
[‡]Johnson & Johnson Pharmaceutical Research & Development, L.L.C., Welsh & McKean Roads, Spring House, Pennsylvania 19477, United States

[§]Janssen Research & Development, Division of Janssen Pharmaceutica NV, Turnhoutseweg 30, B-2340 Beerse, Belgium

[∥]Todd Jones Consulting, San Diego, California.

Supporting Information

ABSTRACT: We present a novel approach for enhancing the diversity of a chemical library rooted on the theory of the wisdom of crowds. Our approach was motivated by a desire to tap into the collective experience of our global medicinal chemistry community and involved four basic steps: (1) Candidate compounds for acquisition were screened using various structural and property filters in order to eliminate clearly nondrug-like matter. (2) The remaining compounds were clustered together with our in-house collection using a novel fingerprint-based clustering algorithm that emphasizes common substructures and works with millions of molecules. (3) Clusters populated exclusively by external compounds were identified as "diversity holes," and representative members of these clusters were presented to our global medicinal chemistry community, who were asked to specify which ones they liked, disliked, or were indifferent to using a simple point-and-click interface. (4) The resulting votes were used to rank the clusters from most to least desirable, and to prioritize which ones should be targeted for acquisition. Analysis of the voting results reveals interesting voter behaviors and distinct preferences for certain molecular property ranges that are fully consistent with lead-like profiles established through systematic analysis of large historical databases.



INTRODUCTION

A pharmaceutical company's compound library is one of its most prized assets.[1] A well-designed library must be diverse and drug-like and must maximize the probability of finding novel hits that can be turned into sustainable, differentiated leads. Compound libraries are typically augmented through internal synthesis or external acquisition. The latter has become increasingly popular in recent years, driven by a rapid growth in the number and diversity of compounds that are available from commercial vendors and improvements in price, availability, quality, delivery time, and logistical support. Compound acquisitions often entail major capital investments, and many pharmaceutical companies have established safe-guarding mechanisms to maximize the utility of the acquired chemicals in relation to their own internal efforts.[2,3] Such mechanisms involve extensive use of chemoinformatic techniques, including substructure and duplicate screening,[4−6] similarity, diversity and QSAR analysis,[7−15] and lead- or drug-like profiling.[16] However, what is often lacking is an effective means for utilizing the collective experience and intuition of the medicinal chemists working within the organization. How to capture that experience while minimizing subjectivity is the subject of the present work.

In his 2004 book "The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations",[17] Surowiecki argues that decisions made by a group of individuals are often better than those made by a single expert. From estimating the weight of an ox in a county fair to predicting future oil prices or offering audience advice to contestants on *Who Wants to be a Millionaire*, there is ample empirical evidence that a collection of people with different points of view but the same motivation to make a good guess can produce predictions that are, on aggregate, more accurate than those of any single individual, no matter how knowledgeable or intelligent. Even when the individual members are not particularly well-informed or unbiased, the group can still collectively arrive at a good decision. Surowiecki attributes this observation to our imperfect nature as decision makers:

> We generally have less information than we would like. We have limited foresight into the future. Most of us lack the ability — and the desire — to make sophisticated cost−benefit calculations. Instead of insisting on finding the best possible decision, we will often accept one that seems good

Received: September 20, 2011
Published: October 31, 2011

> enough. And we often let emotion affect our judgment. Yet despite all these limitations, when our imperfect judgments are aggregated in the right way, our collective intelligence is often excellent.

Of course, not all crowds make good decisions. They can often turn into mindless herds or irrational mobs, following the latest fashion trend, creating stock market bubbles, or trampling each other in football stadiums and department stores. There are four essential characteristics that distinguish intelligent crowds: (1) diversity of opinion, (2) independence, (3) decentralization, and (4) aggregation. For a crowd to be wise, members must have private information, even if it is just a personal interpretation of commonly known facts, must be allowed to form their opinion without influence from others around them, must be able to draw on local experience and knowledge, and must have a mechanism for combining their individual opinions in order to reach a collective judgment.

Although relatively new to the popular literature, the principle of crowd intelligence has a long history in computational statistics and machine learning. Indeed, it is known that the accuracy of classification or regression methods can be significantly improved through aggregation of individual predictors. So-called ensemble techniques, such as bagging,[18] boosting,[19] and stacking,[20] combine multiple models to achieve better predictive performance than could be obtained from any of the constituent models. Obviously, combining the output of multiple predictors is useful only if there is disagreement between them, so much of the work in this field has been devoted to methods for introducing diversity into the model ensemble. Ensemble techniques are becoming increasingly prevalent in computational drug design, where they are used to improve the prediction accuracy of QSAR and QSPR models,[21−23] the quality of similarity searches,[24−27] the scoring of poses in receptor−ligand docking,[28,29] and the identification of structure−activity cliffs.[30]

Diversity is particularly important in drug discovery, where empiricism and intuition still play a dominant role. In 2004, Lajiness and co-workers published a study that challenged a widely held belief that most chemists have a common understanding of what makes a molecule drug- or nondrug-like.[31] The authors asked a group of medicinal chemists with a broad range of experience to sift through ∼22 000 compounds that were candidates for acquisition and identify molecules that were unsuitable for purchase for any reason. This larger pool was divided into several smaller groups, and each group was spiked with molecules that had been previously rejected by an acknowledged expert. The results were striking: not only did the chemists disagree with each other but they often contradicted their own choices. Indeed, reviewers agreed to reject the same compounds only 28% of the time, and when a reviewer looked at the same set of compounds repeatedly, they rejected the same compounds just 50% of the time. These findings have important implications in lead generation and optimization because the choice of which compounds to screen or move forward depends on who is doing the analysis. To paraphrase a familiar adage, Lajiness and his colleagues concluded that drug-likeness is in the eyes of the beholder and that the process of predicting problematic compounds is highly subjective and inconsistent (similar conclusions were reached in a more recent evaluation by a panel of 11 experts of 64 chemical probes identified through the NIH Molecular Libraries Program[32]). Yet, no one would question that experience and knowledge are essential in designing a good screening deck.

Here, we describe a crowd-based strategy that we followed in a recent library expansion initiative at Johnson & Johnson Pharmaceutical Research and Development, LLC (J&JPRD). Our approach combines chemoinformatics analysis to eliminate clearly nondrug-like matter and organize the remaining compounds into clusters of closely related molecules, with community voting to prioritize which of these clusters should be targeted for acquisition. Apart from the crowd component, much of the novelty in the present work lies in the information and the way it was presented to our medicinal chemists to allow them to make an informed decision while keeping the process as efficient as possible.

Our approach differs from Lajiness et al.'s work in several important respects: (1) Chemists were asked to vote on clusters rather than individual compounds. (2) They were allowed to cast not just negative (reject) votes but also positive (must have) and neutral (no preference) ones. (3) They had convenient access to the existing corporate library as a reference. (4) The experiment was conducted on a much larger scale and involved our entire global medicinal chemistry community, spanning six research sites and four different countries.

In the remaining sections, we provide a description of the computational methodology and implementation of the overall system, followed by a detailed analysis of the results and statistical trends. While the ultimate value of our approach cannot be quantified at the present time (in terms of hit rates, probability of producing sustainable leads, utility across target families, etc.), the results confirm that we met our primary goal, which was to enhance the diversity, drug-likeness, and "interestingness" of our existing library. More importantly, our approach did not only yield library enrichment but organizational enrichment as well. Some of these "softer" benefits include (1) employee engagement through a democratic process that made no distinction based on experience or organizational status, (2) personal connection to and ownership of the outcome, (3) global collaboration, (4) transparency in decision making and open scientific debate, and (5) educational value by exposing our medicinal chemists to the wealth of available chemistry in a systematic, efficient, and unbiased way.

METHODS

1. Conceptual Framework. The key challenge in the present work is essentially an identification problem: find holes in the drug-like or lead-like space in our corporate collection and identify the smallest number of compounds that can adequately fill as many holes as possible. Naturally, concepts such as "hole," "adequate," "drug-like," and "lead-like" do not have universally accepted definitions. Furthermore, the problem of looking at a collection of molecules and identifying missing chemotypes is nontrivial. In the present case, however, the self-imposed restriction to fill holes with molecules available for purchase led to a convenient approach.

That approach was to assemble all in-house structures that were either part of our existing screening collection or could potentially be added to it, ignoring any compounds that were depleted from our inventory. To this collection of molecules we added the set of all compounds available for purchase (5 261 676 compounds from 43 vendors). This combined set was clustered with a method described below into very large, loosely related groups of molecules. Any such group that contained no in-house molecules defined a potential "hole," and molecules contained therein were considered to be of particular interest. These molecules

were then filtered for "drug-likeness" or "lead-likeness," reclustered into smaller clusters of more closely related compounds, and submitted to our medicinal chemists for voting, along with relevant information such as the size of the cluster, the structures of the nearest neighbors in our existing library, etc. We must point out that natural products were excluded from this particular exercise because they violate many of the structural and property rules that apply to synthetic drugs. A separate voting initiative specifically tailored to natural products is currently underway and will be described in a subsequent publication.

2. Databases and Database Preparation. Like many others, our group has an ongoing project to define and maintain a set of rules for filtering compounds before they are subjected to scrutiny by medicinal chemists. The goal of these rules is to take advantage of those areas where there is clear consensus, such as very reactive compounds (e.g., acyl halides and isocyanates) and chemical structures that are uniformly recognized as nondrug-like (e.g., hydrocarbons). The rules used in this particular exercise are listed in the Supporting Information.

We must point out that even this conservative set of rules was not wholly without controversy. For example, we chose to reject any compounds containing two or more nitro groups but did not reject compounds containing a single nitro group. In some ways, this is a consequence of the fact that the criteria employed depend on the specific problem at hand; the question "should we buy this compound?" can have an answer that differs from "should we screen this compound that is in our inventory?" and "should we follow up on this screening hit?"

The following procedure was used to prepare molecules for the initial clustering. Input structures were read in SDF format or generated from SMILES strings in Pipeline Pilot.[33] Any associated salts were removed, and ionized groups were neutralized by protonation (to restore neutral acids) or deprotonation (to restore neutral bases). For each structure, a canonical tautomer was generated with the "Enumerate Tautomers" component in Pipeline Pilot, and a canonical SMILES was generated for that tautomer. This SMILES string was then used as a key for merging, so that duplicate structures in different databases could be identified. For each unique canonical SMILES, SPFP_6 fingerprints (Scitegic's implementation of Daylight-style path-based fingerprints of length 6 with Sybyl atom types) were generated and converted to fixed-length hex strings of length 1024, which were then converted to binary files for input into the clustering algorithm.

3. Clustering. As described above, clustering was used for two different purposes in this work: first, to identify holes in our corporate library and, second, to reduce the number of molecules that our voters had to review while making their selection.

Our clustering methodology was based on the concept of maximum common fingerprints (MCF), described in detail elsewhere.[34] In summary, a cluster of molecules was defined in terms of the common set of features shared by every member of the cluster (MCF), analogous to the maximum common substructure (MCS). This is similar in spirit to the concept of a modal fingerprint described by Shemetulskis et al.,[35] where a fingerprint was derived from a group of structurally related molecules with some common biological activity and then used to search a database and identify other potentially active compounds. In our case, we began with fingerprints for each molecule and organized the molecules into groups with rigorous structural conservation. While this approach is applicable to any type of binary fingerprints, we chose SPFP_6 because path-based fingerprints have been shown to correspond more closely with the concept of a conserved structure than, for example, extended connectivity fingerprints.[36]

The partitioning was carried out using a hierarchical agglomerative method involving an iterative sequence of excluded sphere clusterings, in which the threshold used for determining cluster membership was loosened slightly in each iteration, allowing clusters that were distinct in earlier rounds to merge together in subsequent rounds.

This clustering methodology does not produce a specific definition for a "centroid". However, in Ward's method,[37] where a centroid is formally defined, the molecules closest to the centroid are those that have the fewest "on" bits set. Therefore, one reasonable procedure for selecting a cluster representative using the MCF methodology is to pick the molecule with the smallest number of "on" bits. This is consistent with the definition of each cluster as sharing a set of common features; the member with the smallest number of "on" bits will be the one that has the smallest number of nonshared features and therefore is most representative of the conserved features of that cluster. In the current work, we approximated this strategy by selecting as a representative for each cluster the molecule in that cluster with the lowest molecular weight.

4. Voting. The voting system was implemented as a plug-in component of Third Dimension Explorer (3DX), a desktop application designed to address a broad range of data analysis and visualization needs in drug discovery. 3DX is part of a broader platform known as ABCD,[38] which aims to connect disparate pieces of chemical and pharmacological data into a unifying whole and provide discovery scientists with tools that allow them to make informed, data-driven decisions. 3DX is a table-oriented application, similar in concept to the ubiquitous Microsoft Excel. Much of 3DX's analytical power comes from its ability to handle very large data sets through its embedded database technology, to associate custom cell renderers with each data type in the spreadsheet, and to visualize the entire data set using a variety of custom viewers.[39−42] The program offers the full gamut of navigation and selection options, augmented through linked visualizations and interactive filtering and querying. The versatility of 3DX as a general purpose visualization tool has been demonstrated in a variety of other domains beyond discovery research.[43]

The voting system was designed with several goals in mind: (1) ease of use and maximum efficiency in sifting through potentially thousands of candidate clusters, (2) a flexible presentation layer that could provide customizable views of the data, (3) support for concurrency, (4) support for multiple sessions that would allow the chemists to stop and resume the voting at their convenience, (5) randomized presentation of clusters to minimize potential bias, (6) flexible architecture that could accommodate alternative content and presentation styles, and (7) reusable service that could be leveraged and repurposed for future applications.

In a typical voting session, the voter is presented a random subset of records (clusters) selected from the entire voting set, excluding any records that he or she has already voted on in the same campaign. The records are presented as a grid (or matrix) of structures with additional information to help the user make an informed choice. This information includes the size of the cluster, the structures of its representative (the molecule with the lowest molecular weight) and up to four additional members (the molecule with the highest molecular weight and 1−3 other randomly chosen members), the five most similar compounds to the

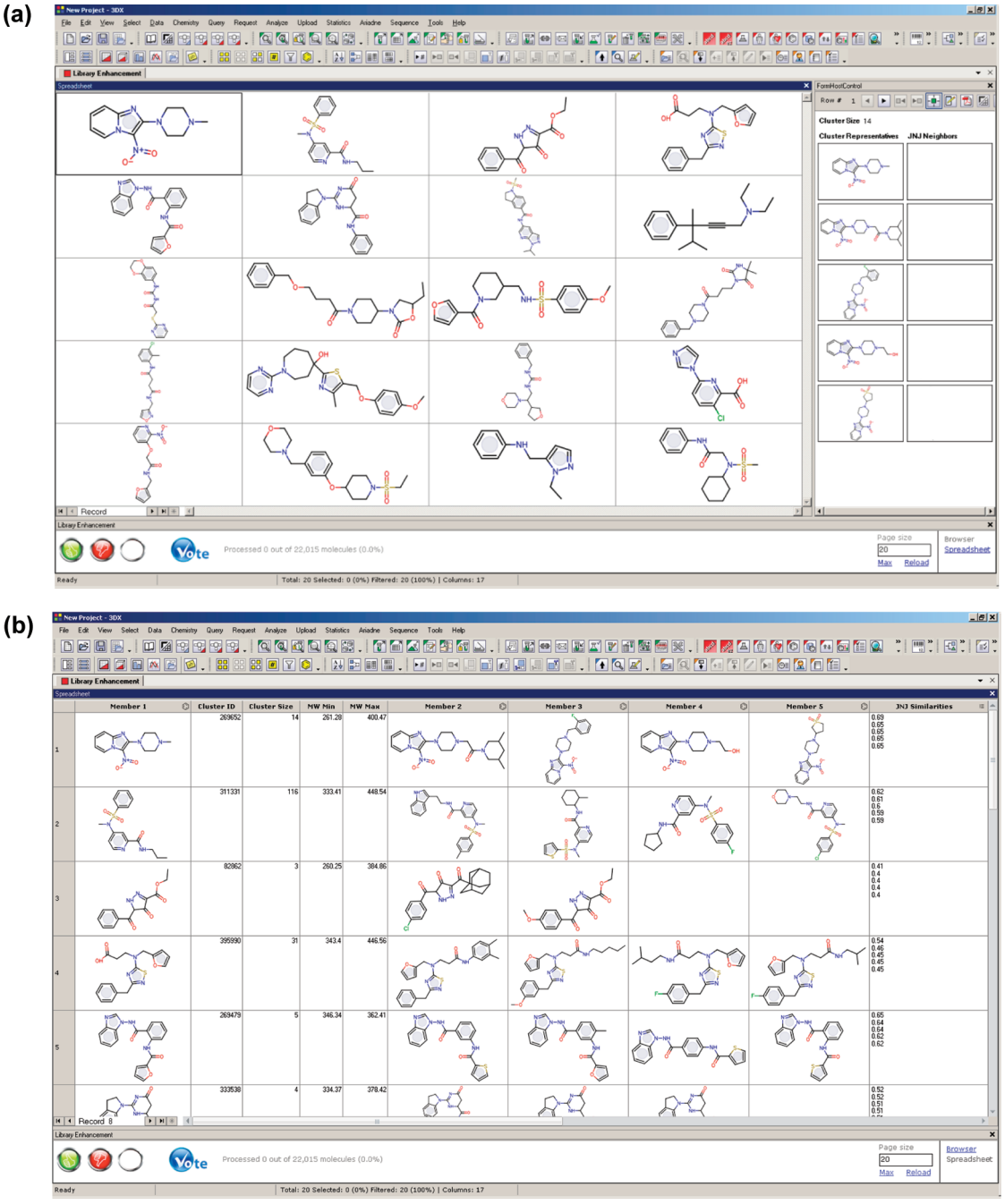


Figure 1. User interface of the 3DX voting plug-in. The two screenshots display the information in browser (top) and spreadsheet (bottom) modes, which can be switched interactively using the corresponding links on the bottom right of the voting panel.

representative in the existing J&JPRD library as determined by fingerprint similarity, and some relevant physicochemical properties such as molecular weight. The maximum size of the subset is configurable and usually selected so that all records can be viewed without scrolling. The voter then proceeds to marking records as "desirable" (thumbs up), "undesirable" (thumbs down), or neutral and submitting the votes to the system. Thus, the system supports three types of votes: good (must have), bad (must not have), and indifferent (could have).

From a software architecture point of view, the voting system was designed following a three-tier model. The presentation layer was implemented in 3DX and is shown in Figure 1. It consists of a menu item to invoke the plug-in and a user interface that allows the user to retrieve the records, review the data, and cast the votes.

Once the user submits the votes on the current subset, the next randomly selected subset is automatically retrieved and presented. The user may stop and restart the process at any time, during the same or different 3DX sessions.

The business layer was implemented as a Web service exposing a remote procedure call interface. It is based on Microsoft .NET Windows Communication Foundation and hosted by IIS. The service's programmatic interface consists of methods to retrieve information about currently available voting sets, to request a subset of records to vote on, and to submit the votes and

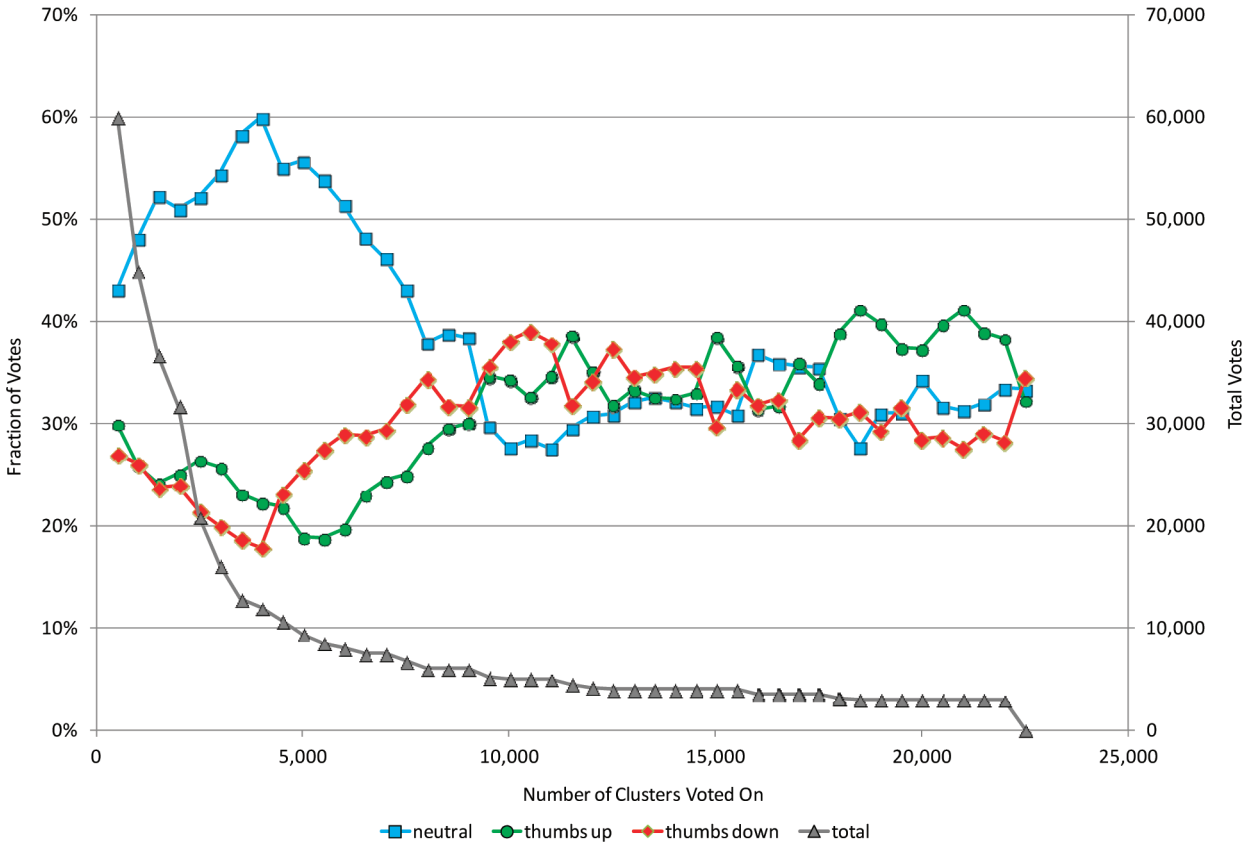


Figure 2. Average percentage of positive, negative, and neutral votes (primary *y* axis) cast by the chemists at each site every 500 clusters, superimposed with the total number of votes (secondary *y* axis). The *x* axis essentially imposes temporal ordering aligned along the number of clusters processed.

retrieve aggregated voting campaign results. The Web service communicates with the data layer (described below) by sending SQL queries and retrieving the results via the protocol supported by the database (abstracted away by the standard .NET Framework data access components).

Finally, the data layer is a Microsoft SQL Server 2008 database with tables containing information about voting sets, voters, and submitted voting results. A design feature worth mentioning is that the database stores only meta-information about the voting sets, while the actual data (chemical structures and compound properties) are stored in 3DX files residing on the server computer. The database table containing information about voting sets stores the path to the file for each set, and the records corresponding to each molecule contain the voting set ID and the row ID inside the corresponding 3DX file. When a subset of records is requested for voting, the Web service opens the 3DX file containing the set's data, extracts a subset of rows, packages it as another 3DX file containing the same table as the original one but only the selected subset of records, and sends it to the client (3DX plug-in) as a byte stream. The 3DX plug-in opens the file, imports the table into the current 3DX session, and presents the table to the user. This design insulates the database model from changes in content and presentation and provides maximum flexibility for future use of this plug-in.

The overall architecture maximizes the amount of information that can be processed by a medicinal chemist in a single voting operation, optimizes the visual layout of that information, and minimizes the number of round trips to the database, thus greatly improving efficiency.

RESULTS AND DISCUSSION

To enable proper interpretation of the results, a general overview of our research footprint and therapeutic area focus can be helpful. When the voting took place, small molecule drug discovery at our company was carried out in six different research sites in the United States and Europe: La Jolla, California; Spring House, Pennsylvania; Beerse, Belgium; Mechelen, Belgium; Toledo, Spain; and Val de Reuil, France. Our research activities focus on five broad therapeutic areas: neuroscience, oncology, immunology, cardiovascular/metabolism, and infectious diseases. We must point out that some of these sites represent consolidation of several other sites with different historical foci, and many of our medicinal chemists have experience in multiple disease areas, both within and outside our company.

As shown in Figure 2 (gray triangles, with values plotted on the secondary *y* axis), voting behavior followed the expected exponential decay curve with the vast majority of medicinal chemists voting only on the first few hundred to a few thousand clusters (data shown in increments of 500 clusters). We remind the reader that the voting was optional and that chemists were advised to vote on a minimum of 2000 clusters, so that we could record a reasonable number of votes (10−20) for each cluster in the collection. (Participation was purely voluntary and did not involve any specific incentives, financial or otherwise. Occasional e-mail reminders were sent out by medicinal chemistry management and the project's steering committee through e-mail and a popular online forum.) As seen in this graph, not everybody did so. Of the 145 chemists who participated in this exercise, only 60 (41%) reached the recommended 2000 vote mark, 10 (7%) went

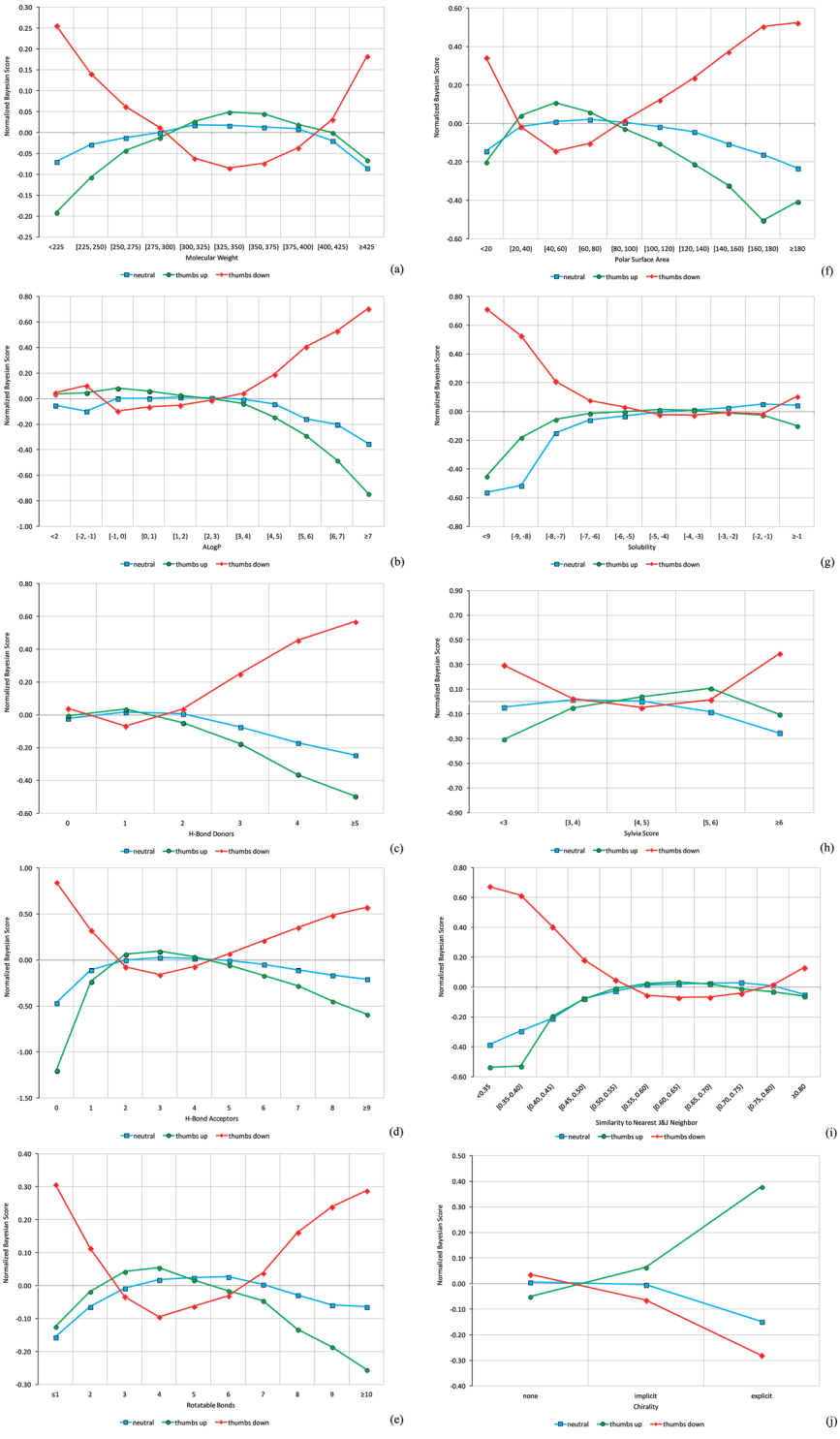


Figure 3. Normalized probabilities (Bayesian scores) for each property and vote type as a function of the property value. (a) Molecular weight, (b) logP, (c) hydrogen bond donors, (d) hydrogen bond acceptors, (e) rotatable bonds, (f) polar surface area, (g) solubility, (h) synthetic accessibility score computed by Sylvia, (i) similarity to closest J&J compound, (j) chirality.

through at least half the clusters, and six (4%) processed the entire collection of 22 015 clusters. However, it should be emphasized that because the clusters were presented to each chemist in a completely randomized order there should be no systematic bias in the *scores* of each cluster when averaged over all participants (*vide infra*).

Figure 2 also shows the average percentage of positive, negative, and neutral votes cast by the chemists every 500 clusters (values plotted on the primary *y* axis). In total, there were 181 492 neutral, 113 297 positive, and 110 623 negative votes. Interestingly, the relative fraction of neutral votes increases steadily for the first 4000 clusters and then starts to decrease

until 9500 clusters, at which point all three types of votes oscillate around the same levels. The increase in the fraction of neutral votes for the first 4000 clusters is consistent with voter fatigue, as this is also the phase with the fastest voter attrition. While there can be multiple interpretations, we believe that the instinctive tendency of humans under stress or weariness (or even boredom) is to be lenient and avoid making mistakes (a "mistake" in this case would be to exclude a potentially useful chemotype from our compound collection; including an irrelevant chemotype is not thought to be as detrimental). After 4000 clusters, the rate of attrition begins to drop, and the remaining voters are more dedicated and more deliberate in their choices.

To understand the preferences of our medicinal chemists from a chemical point of view, we calculated a number of molecular properties for each cluster representative exposed through the voting interface. These properties are widely used in drug-like profiling and include molecular weight, AlogP,[44] number of hydrogen bond donors, number of hydrogen bond acceptors, number of rotatable bonds, polar surface area, and solubility.[45] Three additional properties were included: the implicit or explicit chirality of the molecule, its synthetic accessibility as calculated by the program Sylvia,[46] and its Tanimoto similarity to the nearest J&JPRD neighbor.

The results are illustrated in Figure 3. Each of these plots shows the modified naïve Bayesian score (normalized probability) for each type of vote at various property intervals/values. A higher positive value for a given bin and type of vote indicates a higher probability that molecules whose property falls within that particular bin will receive that particular vote. The plots were constructed as follows. For every cluster prototype, the properties were calculated and subsequently divided into a number of bins. A multiple category naïve Bayesian model as implemented in Pipeline Pilot[47,48] was constructed with the type of vote (thumbs up, thumbs down, neutral) as the category property and the property bins as descriptors, distinguishing a given category versus the complete set. Thus, three Bayesian scores were generated for each property bin: one for the positive votes, one for the negative, and one for the neutral. For each binned property and voting category, we recorded the normalized probabilities (normalized Bayesian scores) as computed in Pipeline Pilot and plotted them as a function of the property bin in Figure 3. (Note that these scores incorporate a Laplacian correction that stabilizes the estimator when the subset to be learned consists of very few samples.)

The plots in Figure 3 reveal some very interesting and consistent trends. Consider, for example, molecular weight. At the two extremes of the range (<250 and ≥ 425), the normalized probabilities for the negative (thumbs-down) votes are the highest, while the ones for the positive (thumbs-up) and neutral votes are the lowest. This means that molecules with a molecular weight less than 250 or greater than 425 are much more likely to be voted down. As molecular weight increases, the relative probability for a negative vote decreases but remains higher than positive and neutral votes until about 300, at which point the positive and neutral probabilities become higher. Molecules with molecular weight between 300 and 400 are more likely to receive a positive vote, slightly less likely to received a neutral vote, and significantly less likely to be receive a negative vote. At a molecular weight of about 400, preferences are once again reversed, and the probability of a molecule receiving a negative vote is higher than positive or neutral. Thus, the molecular weight range preferred by our medicinal chemists is between 300 and 400, with the optimum around 350. Very small and very large molecules are clearly undesirable.

All of the other parameters follow similar trends to those observed for molecular weight. Molecules with less than three or more than six rotatable bonds had a higher probability of receiving a negative vote, with the optimum being four (where the probability of a negative vote is at its lowest value). Chemists clearly disliked molecules that were either too rigid or too flexible. Likewise, the favorable range for hydrogen bond donors was between 0 and 2 (optimum at 1), for hydrogen bond acceptors between 2 and 4 (optimum at 3), and for AlogP below 4. In all cases, the probability of a compound being rejected becomes progressively higher as the property value deviates more and more from the ideal range.

Three parameters that are harder to estimate when evaluating thousands of chemical structures through visual inspection are polar surface area, solubility, and synthetic accessibility. Nonetheless, these properties exhibit the same characteristic trends observed for the more familiar parameters discussed above. The ideal range for polar surface area was between 40 and 80 $Å^2$ (optimum at 40−60). Molecules with too little or too much polar surface area had an increased likelihood of being rejected, as did molecules with poor predicted solubility (less than −5). In this regard, it is reassuring that the chemists' perception of solubility was in line with the computational estimates. Synthetic accessibility as quantified by Sylvia also exhibited a clearly defined preferred range between 4 and 5. Sylvia estimates synthetic tractability using a combination of three types of criteria: (1) structure-based (molecular graph, ring, and stereochemical complexity), (2) starting-material-based (synthetic similarity to available chemicals), and (3) reaction-based (frequency analysis on the presence of strategic bonds that are extracted from reaction databases). These criteria are combined to produce a composite score that can be used to rank order a set of compounds based on their expected ease of synthesis. As seen in Figure 3, molecules that are too simple or too complex to make were clearly disfavored by our chemists.

Another common feature in all of the plots in Figure 3 is that the probability for a neutral vote typically lies between positive and negative and tracks closely the positive vote. The sharpest contrast is between negative and non-negative votes, which means that the decision to vote up or neutral is usually more blurry than the decision to vote down. Stated differently, the Bayesian analysis shows that in practice the neutral vote was perceived as a soft positive vote and that it was easier for chemists to agree on molecules they did not like than on molecules they did like.

The eight properties that we used in our analysis are not completely orthogonal but are not highly correlated either. As seen in Figure 4, the highest positive correlation is between polar surface area and hydrogen bond acceptors (0.66), molecular weight and rotatable bonds (0.5), and molecular weight and synthetic accessibility (0.45) and the highest negative correlation between ALogP and solubility (−0.79) and molecular weight and solubility (−0.53). The correlations between all other parameters are much weaker, both for the cluster prototype as well as all five cluster representatives (Figure 4 includes two rows/columns for each property; "property" denotes the property value of the cluster prototype and "property mean" the average property value of the five cluster representatives displayed in the voting plug-in). Molecular weight is clearly an important factor, as the larger the molecule, the greater the

	ALogP	ALogP Mean	H-Bond Acceptors	H-Bond Acceptors Mean	H-Bond Donors	H-Bond Donors Mean	Molecular Weight	Molecular Weight Mean	Polar Surface Area	Polar Surface Area Mean	Rotatable Bonds	Rotatable Bonds Mean	Solubility	Solubility Mean	Synthetic Accessibility	Synthetic Accessibility Mean
ALogP		0.87	-0.30	-0.23	-0.21	-0.17	0.29	0.22	-0.31	-0.24	0.08	0.03	-0.79	-0.67	0.06	0.03
ALogP Mean	0.87		-0.26	-0.27	-0.18	-0.18	0.20	0.25	-0.25	-0.26	0.05	0.06	-0.69	-0.79	0.03	0.02
H-Bond Acceptors	-0.30	-0.26		0.92	0.11	0.08	0.36	0.32	0.66	0.61	0.27	0.23	0.20	0.19	0.14	0.15
H-Bond Acceptors Mean	-0.23	-0.27	0.92		0.08	0.08	0.34	0.36	0.60	0.66	0.24	0.27	0.16	0.20	0.13	0.16
H-Bond Donors	-0.21	-0.18	0.11	0.08		0.91	-0.05	-0.05	0.32	0.26	0.13	0.13	0.09	0.08	-0.02	-0.02
H-Bond Donors Mean	-0.17	-0.18	0.08	0.08	0.91		-0.04	-0.05	0.27	0.29	0.13	0.14	0.07	0.07	-0.03	-0.03
Molecular Weight	0.29	0.20	0.36	0.34	-0.05	-0.04		0.89	0.31	0.32	0.50	0.42	-0.53	-0.41	0.45	0.47
Molecular Weight Mean	0.22	0.25	0.32	0.36	-0.05	-0.05	0.89		0.32	0.35	0.45	0.48	-0.44	-0.49	0.41	0.48
Polar Surface Area	-0.31	-0.25	0.66	0.60	0.32	0.27	0.31	0.32		0.93	0.22	0.21	0.03	-0.01	0.10	0.13
Polar Surface Area Mean	-0.24	-0.26	0.61	0.66	0.26	0.29	0.32	0.35	0.93		0.21	0.22	-0.01	-0.01	0.10	0.14
Rotatable Bonds	0.08	0.05	0.27	0.24	0.13	0.13	0.50	0.45	0.22	0.21		0.91	-0.29	-0.25	0.03	0.02
Rotatable Bonds Mean	0.03	0.06	0.23	0.27	0.13	0.14	0.42	0.48	0.21	0.22	0.91		-0.23	-0.27	0.00	0.02
Solubility	-0.79	-0.69	0.20	0.16	0.09	0.07	-0.53	-0.44	0.03	-0.01	-0.29	-0.23		0.87	-0.20	-0.17
Solubility Mean	-0.67	-0.79	0.19	0.20	0.08	0.07	-0.41	-0.49	-0.01	-0.01	-0.25	-0.27	0.87		-0.15	-0.16
Synthetic Accessibility	0.06	0.03	0.14	0.13	-0.02	-0.03	0.45	0.41	0.10	0.10	0.03	0.00	-0.20	-0.15		0.94
Synthetic Accessibility Mean	0.03	0.02	0.15	0.16	-0.02	-0.03	0.47	0.48	0.13	0.14	0.02	0.02	-0.17	-0.16	0.94	

Figure 4. Correlation coefficients between various cluster properties. "Property" refers to the property value of the cluster prototype and "property mean" to the average property value of the five cluster representatives displayed in the voting plug-in.

probability of finding more hydrogen bond acceptors and rotatable bonds, having greater polar surface area, being less soluble, etc. Further, as seen in the near-diagonal elements, the correlation between the properties of the cluster prototype and their respective means is very strong (0.87−0.94), suggesting that the favorable property ranges identified in Figure 3 would hold true for the five cluster representatives as well (which we have confirmed to be the case; the results are not shown in order to preserve space).

The property ranges favored by medicinal chemists are fully consistent with the Lipinski rules for orally active drugs (molecular weight ≤ 500, donors ≤ 5, acceptors ≤ 10, logP ≤ 5), if not a little more stringent.[49] It is important to note that, with the exception of molecular weight, these properties were not exposed to the chemists in the voting interface and that all of the molecules were evaluated primarily through visual inspection of their structures. In that respect, the consistency with the rule-of-five is very satisfying. In fact, our results suggest that the compounds preferred by our chemists are more lead-like than drug-like according to Oprea's analysis,[50] who showed that the median values for many of these properties were significantly lower for leads compared to drugs (leads had a median molecular weight 69 Da lower than drugs, one less hydrogen bond acceptor, two less rotatable bonds, and 0.43 lower ClogP units, suggesting that, on average, lead compounds tend to be less bulky, less flexible, and less hydrophobic). Whether this is coincidental with the stated purpose of this exercise to enhance the diversity of our corporate library (which is used more as a source of leads rather than drugs) is unclear. The point remains that the intuitive preferences of our medicinal chemists are fully consistent with general rules established through systematic analysis of large databases of historical data. It also suggests that enhancing the diversity of a chemical library need not come at the expense of lead- or drug-likeness. Indeed, as seen in the second-to-last plot in Figure 3 which shows the voting probability as a function of the Tanimoto similarity of the cluster representative to its nearest J&JPRD neighbor, molecules that were too similar (Tanimoto >0.8) or too dissimilar (Tanimoto <0.5) to existing compounds in our collection were more likely to be rejected. This plot also suggests that many of the chemists took into consideration the nearest J&JPRD neighbors displayed in the voting plug-in; otherwise they would have had no way of knowing that some of these compounds had closely related analogs in the existing J&JPRD library.

Another property that seems to have played a defining role in chemists' selection was chirality. The last plot in Figure 3 shows the normalized voting probabilities for molecules containing explicit, implicit, or no chiral centers. The first category refers to molecules that contained at least one chiral center with explicitly designated stereochemistry (dotted or wedged bonds), the second category to molecules that contained chiral centers but whose stereochemistry was not specified (no dotted or wedged bonds), and the third to molecules with no chiral centers at all. This plot shows a clear preference for compounds containing explicit chiral centers and a slight preference for compounds with implicit ones. For molecules with no chiral centers, the voting probabilities were roughly equal. It is interesting that this is the only property for which the neutral vote trends closer to the negative than the positive vote. This suggests that positive votes were more deliberate and intentional than negative ones. In other words, the presence of chiral centers was seen as a reason for accepting a compound rather than rejecting it, i.e., as an attractive feature that increased the likelihood of casting an explicit positive vote. All other properties were used as a reason for rejecting a compound, e.g., too large, too flexible, too polar, too insoluble, etc.

1. Controversial Clusters. The final part of our analysis concerns controversial clusters, i.e., clusters that elicited the most conflicting responses among chemists. These were identified by rank ordering the clusters on the basis of the standard deviation of the votes they received using the same numerical scale that was used for computing the score (−1 for negative, +1 for positive, 0 for neutral). The standard deviation is maximized

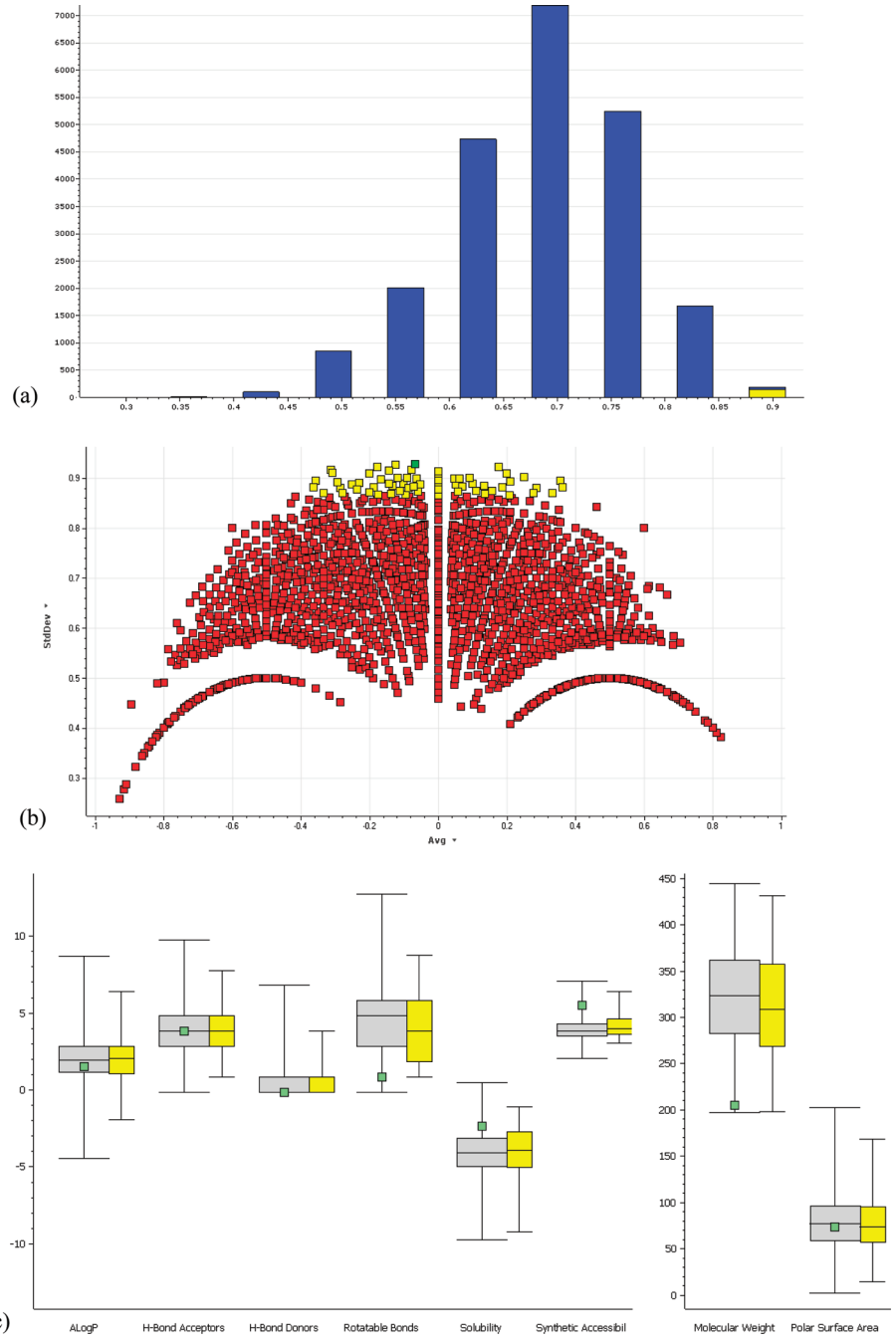


Figure 5. (a) Distribution of the standard deviation of the scores across all clusters. (b) Mean cluster scores plotted against their standard deviation. The 149 most controversial clusters are highlighted in yellow (and the most controversial one in green). They are characterized by a low average score and high standard deviation. (c) Box plot showing the distribution of eight molecular properties across all clusters (gray) and the 149 most controversial ones (yellow).

when there is a large number of positive votes, an equally large number of negative votes, and as few neutral votes as possible. Using 0.88 as an arbitrary cutoff, we identified the 149 most controversial clusters and visually examined them to identify any common themes.

As shown in Figure 5a, the standard deviations of the votes follow a typical normal distribution, with only a small fraction of clusters exhibiting strong contradictions among voters. The relationship between the scores and standard deviations is illustrated in Figure 5b. The 149 most controversial clusters highlighted in yellow have a high standard deviation and a low average score, because the negative and positive votes cancel each other out. The most controversial cluster is highlighted in green at the top center of the plot, while the most unpopular and most popular clusters are located on the bottom left and bottom right side of this plot, respectively. The structures of the prototypes of these clusters are illustrated in Figure 6.

As shown in the box plot in Figure 5c, the molecular properties of the controversial clusters are in line with the overall population. Therefore, we conjecture that the reason for the dispute is

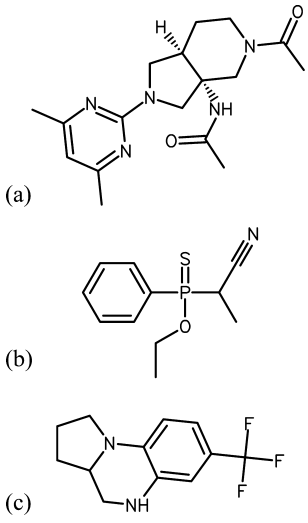


Figure 6. Prototypes of (a) the most popular, (b) the least popular, and (c) the most controversial clusters in the entire collection. Cluster a received 14 positive, 3 neutral, and 0 negative votes; cluster b received 0 negative, 1 neutral, and 13 negative votes; and cluster c received 6 positive, 2 neutral, and 7 negative votes.

the presence of certain structural motifs that some chemists perceived as highly undesirable, while others considered them desirable or at least easily replaceable if these molecules ever emerged as hits in a high throughput screening campaign. Through visual inspection and substructure analysis, we identified several patterns that occurred more frequently in the controversial clusters compared to the broader population. These include carbamates, carboxylic acids, and esters, which occurred in 22% of the controversial clusters compared to 16% in the broader population (1.4-fold enrichment), and thiophenes, pyroles, and furans, which occurred in 24% of the controversial clusters versus 18% in the broader population (1.3-fold enrichment). Whether this correlation implies causality is unclear. What is certain is that different experiences lead to different preferences and that these preferences are not always congruent. An alternative explanation for some of the controversial clusters is that voters disagreed on how close the available compounds were to existing compounds in the collection. If the compounds were desirable but very similar to existing compounds then they may have engendered an overall split vote.

CONCLUDING REMARKS

The library enhancement approach presented in this paper stemmed from a desire to tap into the collective wisdom and experience of our global medicinal chemistry community, and an even deeper organizational desire to decentralize decision making, empower our scientists, and give them ownership and accountability for the outcome. In designing this experiment, we took all necessary precautions to guard against factors that could lead to a poor collective decision, including (1) homogeneity, by allowing input from multiple demographic groups with diverse backgrounds and experience, (2) centralization, by eliminating organizational and/or intellectual authority as a deciding factor, (3) division, by synthesizing input in a fair and unbiased way, (4) imitation, by ensuring that each person's choices were invisible to others until all of the votes were collected, thus preventing intentional or unintentional information cascade, and (5) emotion, by allowing scientists to cast their votes in private, thereby eliminating any form of peer pressure.

Poor decisions arise when intellectual conformity replaces independence of thought, when the members of the community, for whatever reason, become too conscious of the opinions of others and begin to emulate them. Authority, organizational or intellectual, poses the greatest risk. Its influence can be subtle or direct, unintentional or deliberate. The subtle aspects were identified by Aristotle in his artistic proofs of persuasion: *ethos*, *pathos*, *logos*—the credibility of the speaker, his emotional or motivational appeal, and the logical grounding of his arguments. Although crowds can be swayed very easily by persuasive people, the main reason for intellectual conformity is that there is a systematic flaw in the system for making decisions.

The results of our experiment are fully consistent with prior literature on what confers drug- or lead-like characteristics to a chemical substance. Whether the strategy will yield the desired results in the long term with respect to quality, novelty, and number of hits/leads remains to be seen. It is also unclear whether this strategy can lead to sufficient differentiation from a competitive stand-point. In the meantime, the only undeniable benefit we can point to is that we harnessed our chemists' opinions to select lead-like molecules that are totally within reasonable property ranges, that fill diversity holes, and that have been purchased for screening, and we did so in a way that promoted greater transparency, greater awareness, greater collaboration, and a renewed sense of involvement and engagement of our employees.

Going forward, our plan is to use a similar voting system to evaluate proposed proprietary structures that we will prepare specifically for addition to our compound library. The experience of examining large sets of commercial compounds should serve well in guiding this effort since our chemists are now well-versed in what is already available commercially. We will ask that they generate ideas for novel compounds and chemotypes, partly on the basis of the structures they voted on and partly on the basis of using synthetic chemistry not employed to make existing compounds.

ASSOCIATED CONTENT

Supporting Information. The filters used to eliminate clearly nondrug-like matter are provided. This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*Tel: (215) 628-6814 (D.K.A.), (858) 320-3482 (T.M.). E-mail: dagrafio@its.jnj.com, tmirzade@its.jnj.com.

ACKNOWLEDGMENT

We thank our global medicinal chemistry community for participating in this effort and Dr. Roger Bone for providing executive sponsorship and funding for the library enhancement project. We also thank Joseph Ciervo who developed an earlier version of a voting application called Screen-or-Scrap, Dr. Pascal Bonnet for suggesting the Sylvia analysis, and Drs. Peter J. Connolly and John (Jake) J. M. Wiener for their critical review of this manuscript.

REFERENCES

(1) Editorial: *Nat. Rev. Drug Discovery* **2004**, *3*, 375.

(2) Dunbar, D. B. Compound acquisition strategies. *Pac. Symp. Biocomput.* **2000**, 555–565.

(3) Engels, M. F.; Gibbs, A. C.; Jaeger, E. P.; Verbinnen, D.; Lobanov, V. S.; Agrafiotis, D. K. A cluster-based strategy for assessing the overlap between large chemical libraries and its application to a recent acquisition. *J. Chem. Inf. Model.* **2006**, *46*, 2651–2660.

(4) Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702–712.

(5) Baurin, N.; Baker, R.; Richardson, C.; Chen, I.; Foloppe, N.; Potter, A.; Jordan, A.; Roughley, S.; Parratt, M.; Greaney, P.; Morley, D.; Hubbard, R. E. Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 643–651.

(6) Merlot, C.; Domine, D.; Cleva, C.; Church, D. J. Chemical substructures in drug discovery. *Drug Discovery Today* **2003**, *8*, 594–602.

(7) Shemetulskis, N. E.; Dunbar, J. B.; Dunbar, B. W.; Moreland, D. W.; Humblet, C. Enhancing the diversity of a corporate database using chemical database clustering and analysis. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 407–416.

(8) Agrafiotis, D. K. Stochastic algorithms for maximizing molecular diversity. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 841.

(9) (a) Agrafiotis, D. K. Multiobjective optimization of combinatorial libraries. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 335–356. (b) Agrafiotis, D. K. Multiobjective optimization of combinatorial libraries. *IBM J. Res. Dev.* **2001**, *45* (3/4), 545–566.

(10) Higgs, R. E.; Bermis, K. G.; Watson, I. A.; Wikel, J. H. Experimental designs for selecting molecules from large chemical databases. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 861–870.

(11) Rassokhin, D. N.; Agrafiotis, D. K. Kolmogorov-Smirnov statistic and its applications in library design. *J. Mol. Graphics Modell.* **2000**, *18* (4−5), 370–384.

(12) Turner, D. B.; Tyrrel, S. M.; Willett, P. Rapid quantification of molecular diversity for selective database acquisition. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 18–22.

(13) Rhodes, N.; Willett, P. Bit-string methods for selective compound acquisition. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 210–214.

(14) Cummins, D. J.; Andrews, C. W.; Bentley, J. A.; Cory, M. Molecular diversity in chemical databases: comparison of medicinal chemistry knowledge bases and databases of commercially available compounds. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 750–763.

(15) Nilakantan, R.; Bauman, N.; Haraki, K. S. Database diversity assessment: new ideas, concepts, and tools. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 447–452.

(16) Muegge, I. Selection criteria for drug-like compounds. *Med. Res. Rev.* **2000**, *23*, 302–321.

(17) Surowiecki, J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*; Little, Brown & Company: New York, 2004; ISBN 0-316-86173-1.

(18) Breiman, L. Bagging predictors. *Machine Learning* **1996**, *24*, 123–140.

(19) Feund, Y.; Schapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. In *Proceedings of the Second European Conference on Computational Learning Theory*; Springer-Verlag: Berlin, 1995; pp 23−37.

(20) Breiman, L. Stacked regressions. *Machine Learning* **1996**, *24*, 49–64.

(21) Agrafiotis, D. K.; Cedeno, W. On the use of neural network ensembles in QSAR and QSPR. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 903–911.

(22) Mattioni, B. E.; Kauffman, G. W.; Jurs, P. C.; Custer, L. L.; Durham, S. K.; Pearl, G. M. Predicting the genotoxicity of secondary and aromatic amines using data subsetting to generate a model ensemble. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 949–963.

(23) Seierstad, M.; Agrafiotis, D. K. A QSAR model of hERG binding using a large, diverse and internally consistent training set. *Chem. Biol. Drug. Des.* **2006**, *67* (4), 284–296.

(24) Ginn, C. M. R.; Willett, P.; Bradshaw, J. Combination of molecular similarity measures using data fusion. *Perspect. Drug Discovery Des.* **2000**, *20*, 1–16.

(25) Salim, N.; Holliday, J.; Willett, P. Combination of fingerprint-based similarity coefficients using data fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435–442.

(26) Baurin, N.; Mozziconacci, J. C.; Arnoult, E.; Chavatte, P.; Marot, C.; Morin-Allory, L. 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 276–285.

(27) Baber, J. C.; Shirley, W. A.; Feher, M. The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 277–288.

(28) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.

(29) Paul, N.; Rognan, D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins* **2002**, *47*, 521–533.

(30) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477–491.

(31) Lajiness, M. S.; Maggiora, G. M.; Shanmugasundaram, V. Assessment of the consistency of medicinal chemists in reviewing sets of compounds. *J. Med. Chem.* **2004**, *47*, 4891–4896.

(32) Oprea, T. I.; Bologa, C. G.; Boyer, S.; Curpan, R. F.; Glen, R. C.; Hopkins, A. L.; Lipinski, C. A.; Marshall, G. R.; Martin, Y. C.; Ostopovici-Halip, L.; Rishton, G.; Ursu, O.; Vaz, R. J.; Waller, C.; Waldmann, H.; Sklar, L. A. A crowdsourcing evaluation of the NIH chemical probes. *Nat. Chem. Biol.* **2009**, *5*, 441–447.

(33) *Pipeline Pilot*; Accelrys: San Diego, CA. http://accelrys.com/products/pipeline-pilot/ (accessed June 30, 2011).

(34) Hack, M. D.; Mirzadegan, T. The use of maximum common fingerprints in fast, high-quality substructure clustering. Submitted.

(35) Shemetulskis, N. E.; Dunbar, J. B.; Dunbar, B. W.; Moreland, D. W.; Humblet, C. Enhancing the diversity of a corporate database using chemical database clustering and analysis. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 407–416.

(36) Bayada, D. M.; Hamersma, H.; van Geerestein, V. J. Molecular diversity and representativity in chemical databases. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1–10.

(37) Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236.

(38) Agrafiotis, D. K.; Alex, S.; Dai, H.; Derkinderen, A.; Farnum, M.; Gates, P.; Izrailev, S.; Jaeger, E. P.; Konstant, P.; Leung, A.; Lobanov, V. S.; Marichal, P.; Martin, D.; Rassokhin, D. N.; Shemanarev, M.; Skalkin, A.; Stong, J.; Tabruyn, T.; Vermeiren, M.; Wan, J.; Xu, X. Y.; Yao, X. Advanced Biological and Chemical Discovery (ABCD): Centralizing discovery knowledge in an inherently decentralized world. *J. Chem. Inf. Model.* **2007**, *47* (6), 1999–2014.

(39) Agrafiotis, D. K.; Shemanarev, K.; Connolly, P. J.; Farnum, M.; Lobanov, V. S. SAR maps: a new SAR visualization technique for medicinal chemists. *J. Med. Chem.* **2007**, *50* (24), 5926–5937.

(40) Kolpak, J.; Connolly, P. J.; Lobanov, V. S.; Agrafiotis, D. K. Enhanced SAR maps: Expanding the data rendering capabilities of a popular medicinal chemistry tool. *J. Chem. Inf. Model.* **2009**, *49*, 2221–2230.

(41) Agrafiotis, D. K.; Wiener, J. J.M. Scaffold explorer: an interactive tool for organizing and mining SAR data spanning multiple chemotypes. *J. Med. Chem.* **2010**, *53* (13), 5002–5011.

(42) Agrafiotis, D. K.; Wiener, J. J. M.; Skalkin, A.; Kolpak, J. Single R-group polymorphisms (SRPs) and R-cliffs: An intuitive framework for

analyzing and visualizing activity cliffs in a single analog series. *J. Chem. Inf. Model.* **2011**, *51*, 1122–1131.

(43) Cepeda, M. S.; Lobanov, V. S.; Farnum, M.; Weinstein, R.; Gates, P.; Agrafiotis, D. K.; Stang, P.; Berlin, J. A. Broadening access to electronic health care databases. *Nat. Rev. Drug Discovery* **2010**, *9*, 84.

(44) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577.

(45) Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. Estimation of aqueous solubility of chemical compounds using E-state indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488–1493.

(46) Boda, K.; Seidel, T.; Gasteiger, J. Structure and reaction based evaluation of synthetic accessibility. *J. Comput.-Aided. Mol. Des.* **2007**, *21*, 311–325.

(47) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **2004**, *47*, 4463–4470.

(48) Nidhi; Glick, M.; Davies, J. W.; Jenkins, J. L. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **2006**, *46*, 1124–1133.

(49) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **2001**, *46*, 3–26.

(50) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315.