Do Structurally Similar Molecules Have Similar Biological Activity?

Yvonne C. Martin,*,† James L. Kofron,‡ and Linda M. Traphagen‡

Global Pharmaceutical Research and Development, Abbott Laboratories, Abbott Park, Illinois 60064-6100

Received April 12, 2002

To design diverse combinatorial libraries or to select diverse compounds to augment a screening collection, computational chemists frequently reject compounds that are ≥0.85 similar to one already chosen for the combinatorial library or in the screening set. Using Daylight fingerprints, this report shows that for IC_{50} values determined as a follow-up to 115 high-throughput screening assays, there is only a 30% chance that a compound that is ⩾ 0.85 (Tanimoto) similar to an active is itself active. Although this enrichment is greater than that found with random screening and docking to three-dimensional structures, this low fraction of actives within similar compounds occurs not only because of deficiencies in the Daylight fingerprints and Tanimoto similarity calculations but also because similar compounds do not necessarily interact with the target macromolecule in similar ways. The current study emphasizes the statistical or probabilistic nature of library design and that perfect results cannot be expected.

Introduction

It is a central premise of medicinal chemistry that structurally similar molecules have similar biological activities. This premise is validated by the long experience that suggests rules-of-thumb such as that β-lactams frequently possess antibacterial activity, that phenethylamines are likely to have activity in the central nervous system, and that aromatic nitro compounds are likely to be mutagenic. Computational chemists have exploited this premise in their analysis of molecular diversity of compound libraries and in the selection of compounds for high-throughput screening (HTS).[1−5] These uses raise two questions: (i) Do the computational calculations of molecular similarity[6] adequately reflect the factors that lead to biological similarity, and (ii) do the generalizations apply across many types of molecules or to only a subset of them?

Earlier, we examined the screening sets available at the time, containing 2000−3000 members, to address this question. We concluded that if a compound has a Tanimoto similarity,[6] based on Unity fingerprints,[7] of ≥0.85 to an active compound, then the compound has an 80% chance of itself being active in the same assay.[8,9] As a result, we do not purchase from an outside vendor any compound that is ≥0.85 similar to a compound in the Abbott screening collection. (This criterion typically eliminates approximately 75% of the compounds offered for purchase.) Others have reached a similar conclusion.[10−12] For example, Matter evaluated several different molecular descriptors to quantify molecular diversity relevant to biological activity.[11] Using Unity fingerprints, he concluded that "if two molecules have a Tanimoto coefficient larger than 0.85...the biological activity of the first molecule is similar to that of the second one...This concept now allows to reduce the redundancy of a database by rejecting structurally similar molecules based on this similarity radius."

In contrast, in an early study of the influence of the method of compound selection on the hit rate and diversity of hits obtained, Taylor used a range of 0.012−0.50 as the frequency with which a compound similar to an active is itself active.[13] That is, his empirical estimate was much lower than the one that was later accepted by computational chemists. A later study from the same company showed 0.40−0.6 as the proportion of 0.85 similars that are active.[14]

Kubinyi presents compelling examples of the lack of parallel between structural and biological similarity.[15] This evidence is supported by the behaviors of medicinal chemists. For example, we noticed that medicinal chemists more frequently follow up, by synthesizing more analogues, those hits for which several analogues had also been tested. This tendency appears to be independent of the source of the analogues: although many of the sets of analogues had been synthesized in-house, some sets of analogues had been purchased because the collection did not go through the usual computational screening or before we had demonstrated the superiority of Ward's clustering over Jarvis−Patrick clustering.[16] When questioned about this behavior, the medicinal chemists responded that the activity/inactivity of analogues suggests if a hit might have "flat structure−activity relationship (SAR)"—that is, that all analogues have the same potency and hence no improvement in potency might be expected if other analogues were synthesized. Hence, the medicinal chemists expected to see, and apparently did see, that similar compounds do not have similar activity, a contradiction to the hypothesis underlying diversity selection of compounds for screening.

Four factors led us to undertake the current study. First, we were turning away many compounds that otherwise were attractive to purchase. Second, we knew even from the original analysis that by not purchasing compounds similar to those in the screening set that we would miss active compounds: the question was how many. Third, the behavior of medicinal chemists suggested that analogues have value that we previously

* To whom correspondence should be addressed. Tel: 1-847-937-5362. Fax: 1-847-937-2625. E-mail: yvonne.c.martin@abbott.com.
† Computer-Assisted Molecular Design Department R47E.
‡ Department of Biomolecular Screening, R4PN.

10.1021/jm020155c CCC: $22.00
Published on Web 08/13/2002

Table 1. Statistics of the HTS Results in Different Assays

	mean	SD	median	min	max
no. of actives	93.4	145	48	3	880
total no. of ≥0.85 similars to all actives in the assay	208	279	109	4	1563
max no. of ≥0.85 similars to any active in the assay	23.3	16.7	19	2	76
actives with similar compds	44.3	71.8	18	1	467
fraction of ≥0.85 similars that are active	0.288	0.128	0.278	0.046	1.00
fraction of actives with no ≥0.85 similars	0.532	0.139	0.534	0.139	0.857

ignored. Last, we now had a large data set that could help us more adequately quantitate the similarity principle, particularly as it applies to selecting compounds for screening. The current investigation provides information that will allow us to improve our criteria for selecting compounds for purchase and for targeted libraries.

Materials and Methods

IC_{50} Data Set. We had collected a database of IC_{50} or null values for compounds tested as a follow-up to HTS. A compound is tested for IC_{50}, as a single compound, if either the HTS screen suggests that it may be active or it is similar to a hit in screening or a previous IC_{50} determination. If a compound is not active in the IC_{50} determination, only its Abbott ID is entered into the database. In this data set of 19 533 compounds tested in one or more of 115 assays, 6784, or 35%, are active in one or more assay; 5584 are active in only one assay, 690 are active in two, 424 are active between three and ten, and 86 are active in 10−19 assays. Hence, 82% of the compounds are active in only one assay.

We built a Daylight[17] database of the structures, results from each of the IC_{50} assays, a summary field for the number of IC_{50} values reported for a compound, and another summary field for which assays an IC_{50} was reported. Although the compounds tested were often salts, we built our database and based our analysis on the part of the structure that is not a common salt: these structures are available in our registration system. The reasons for not using the salt are that most often the common salt portion of the compound does not impart the biological activity but also that different salts of the same molecule do not have identical fingerprints and hence do not have a similarity of 1.0. All of the analyses were performed on normal smiles with the result that differences in stereochemistry were ignored.

Table 1 summarizes the character of the database: the number of actives in a particular assay ranges from three to 880, and the number of compounds similar to any active in an assay ranges from four to 1563.

Monoamine Oxidase (MAO) Data Set. This data set of 1645 compounds, which had been screened for inhibition of monoamine oxidase, was used in our previous evaluation of clustering algorithms and structural descriptors.[16,18,19] It will allow the new results to be placed in context of our previous work and allow others to compare their algorithms.[18] It contains 287 actives divided into three potency ranges. The compounds were hand-selected by a medicinal chemist and augmented with synthesized analogues that explore the SAR. This database has the advantages that every compound was tested in the assay and that three, approximately 10× different, levels of potency are reported. Because it contains some compounds synthesized for the project, the fraction of actives in this database, 17.4%, is larger than is typically found in screening databases.

Similarity Analyses. We performed automated similarity searches using the Daylight toolkit program merq.[20] Merq processes each smiles string in a file, all structures with a reported IC_{50} in a particular assay, for example. For each compound in the file, merq reports how many and which structures are similar to it using the Tanimoto measure of similarity. The analysis then involves using Excel to remove duplicate entries resulting from structures that are similar to more than one active, to remove active compounds for which no similars were found, and last to calculate the fraction of actives in the remainder. Note that our measure includes only those actives for which at least one similar compound was found. This eliminates the problem, discussed by Delaney,[14] of counting actives that are singletons and hence inflating the fraction of actives at high similarities.

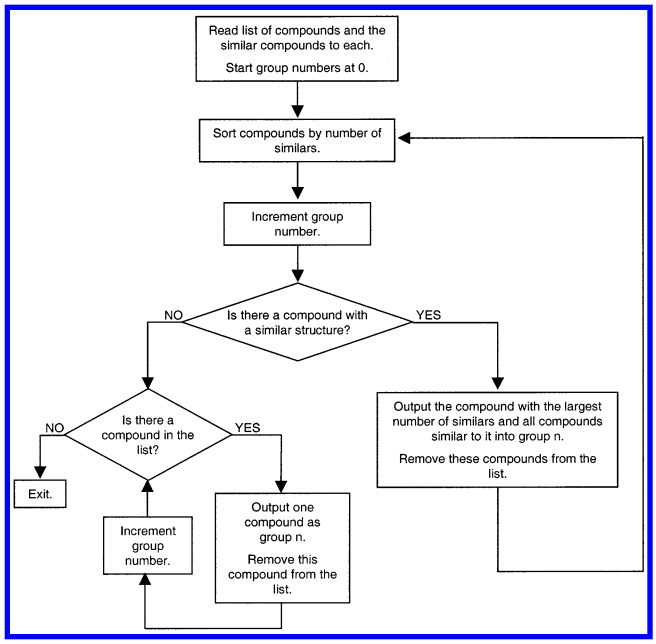


Figure 1. Flow diagram of the procedure to cluster compounds by similarity.

Cluster Identification. We also clustered the compounds using an unsupervised nonhierarchical clustering algorithm described by Taylor[13] and later by Butina.[21] It is summarized in Figure 1. For the IC_{50} data set, there are 11615 true singletons at ≥0.85 similarity. The clustering algorithm identified 14 113 clusters of similar compounds at ≥0.85 similarity, or an average of 1.38 compounds per cluster. Of these clusters, 11 897 are singletons; hence, only 2.4% of the singletons are artifacts generated by the clustering algorithm. The largest cluster contained 80 variously substituted 4-phenyl-6-amino-7-fluoro quinolones that had been synthesized for antibacterial activity.[22−25] Only four of these were active, all in the same assay: in this assay, another 177 compounds from different clusters were also active. The program found 1164 clusters in the monoamine oxidase data set, or an average of 1.42 compounds per cluster. In this data set, there are 943 true singletons: the clustering algorithm finds an additional nine, or 1%, false negatives. The largest cluster contains 38 compounds, one of which is moderately active and seven of which are slightly active. This cluster is structurally unrelated to known MAO inhibitors.

Results

Table 1 and Figure 2 provide an overview of the analysis of the HTS results. For only 3015 of the active compounds, less than half had a ≥0.85 similar compound that had been tested for IC_{50}. This set forms the basis for the further analysis of the HTS results. From Table 1, it can be seen that the mean fraction of ≥0.85

Table 2. Summary of Detailed Analysis of Selected HTS Screens at ≥0.85 Similarity

assay	1	2	3	4	5	6	7	8	9
Calculations That Are Based on Similarity Searching of All Actives									
total no. of actives with similars	105	46	15	266	390	25	309	467	37
total no. of actives plus similars	513	139	43	770	1114	92	938	1075	71
ratio actives/total	0.20	0.33	0.35	0.34	0.35	0.27	0.33	0.43	0.52
Calculations That Are Based on Similarity Searching of Only Potent and Selective Actives									
no. of actives at 10 μM that hit in only one assay and have similar structures	32	12	14	97	120	7	74	377	9
total no. of selected actives plus similars	260	40	42	359	496	17	258	962	19
total no. of active compds	48	13	14	133	185	7	118	476	14
fraction of similars that are active	0.18	0.33	0.34	0.37	0.37	0.44	0.46	0.49	0.74

Table 3. Diversity of the Data Sets Used for the Detailed Analysis

assay	1	2	3	4	5	6	7	8	9
no. of ≥ 0.85 clusters	27	15	13	92	106	11	69	260	7
compds per cluster	9.63	2.67	3.23	3.90	4.68	1.55	3.74	3.70	2.71
mean no. of ≥0.85 similar compds for each potent active with similars	18.4	4.17	3.07	5.42	5.49	2.42	7.12	7.48	2.89

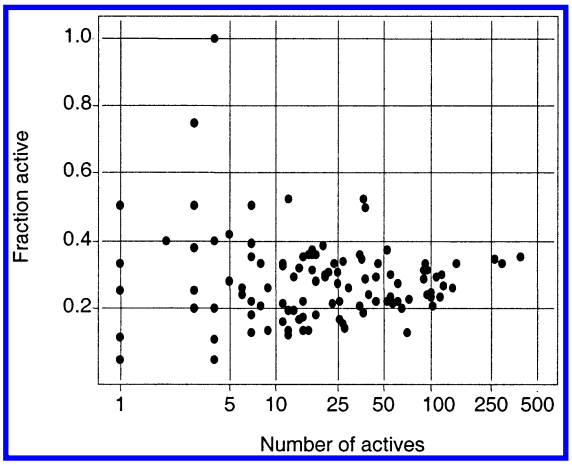


Figure 2. For 115 assays, the fraction of molecules that are similar to any active in the particular assay that are themselves active as a function of the number of actives with similars. Similarity was evaluated as a Tanimoto coefficient ≥0.85 using Daylight fingerprints of length 1024.

similars that are active is 0.29 and the median is 0.28. Figure 2 shows that for a large fraction of the assays, the fraction of compounds similar to an active that are themselves active in the same assay is between 0.20 and 0.40. Figure 2 also shows that the main conclusions are not unduly influenced by data sets in which there are only a few active compounds for which similars had been tested. In fact, such data sets are responsible for the proportions less than 0.2 and greater than 0.45. The fraction of similars that are active in data sets that contain more than 75 actives falls in the range of 0.20−0.35, and in data sets that contain more than four actives, the fraction is always less than 0.52.

Because the enrichment in actives was less than expected, we explored the data further. Two factors might skew the results: first, it is possible that for some assays not all similars in the database had actually been tested in the assay of interest and could be falsely counted as inactive. This would be especially likely to occur if the search for similar compounds was based on a compound that was active in more than one assay. Second, our initial analysis searched for similars of all actives regardless of the potency of the active compound. It is possible that an analogue of a weak compound would be even weaker, with potency undetectable in our assay. To investigate these factors, we identified nine assays for which we were certain that all compounds similar to an active had been tested. We then repeated the above analysis using as search targets only compounds with an IC_{50} < 10 μM, 10× lower than the highest concentration tested in the IC_{50} determination, but continued to count as active any compound with a reported IC_{50}. In addition, we confined the similarity search to use only those compounds that were active in only the target assay.

The results, summarized in Table 2, reveal a slightly higher fraction of actives in similars to a potent selective active. In three of the nine cases, the increase is substantial. However, the median fraction of actives increases only from 0.35 to 0.37. The results also support the use of these sets of actives to derive answers that apply to the larger data set, although for this set the fraction of similars that are active is somewhat higher than in the whole data set. Note also that the fraction actives for assay 9 jumps from 0.52 to 0.74; it might be that this is an artifact of the small number of compounds. Table 3 summarizes the diversity of the compounds used in this analysis.

The above analyses were performed using Daylight fingerprints 1024 bits long. The default Daylight fingerprints encode all of the substructures in a molecule that contain from zero to seven bonds.[17] To accommodate such a large number of possible substructures, a hashing algorithm is used such that each type of substructure sets a certain number of bits but no or few substructures set exactly the same bits. Because short fingerprints lose information by the hashing process,[6] we reexamined the latter data sets using fingerprints of lengths 2048 and 4096. The longer fingerprints have a reduced chance that two different substructures will set the same bits and hence artificially inflate the similarity between two compounds. Figure 3 shows that in general there is an increase in the fraction of similars to an active that are themselves active, but the increase is small. The lower panel shows that the longer fingerprints are more selective because fewer (inactive) com-

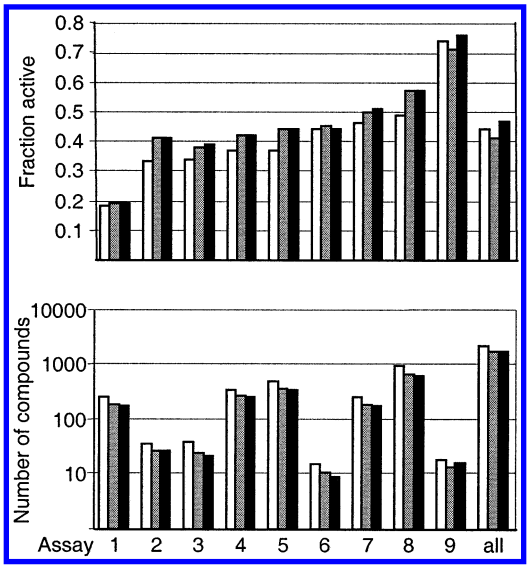


Figure 3. For nine selected assays, a comparison of the fraction of ≥0.85 similars to a potent selective active that are themselves active using Daylight fingerprints of lengths 1024 (white bars), 2048 (gray bars), and 4096 (black bars).

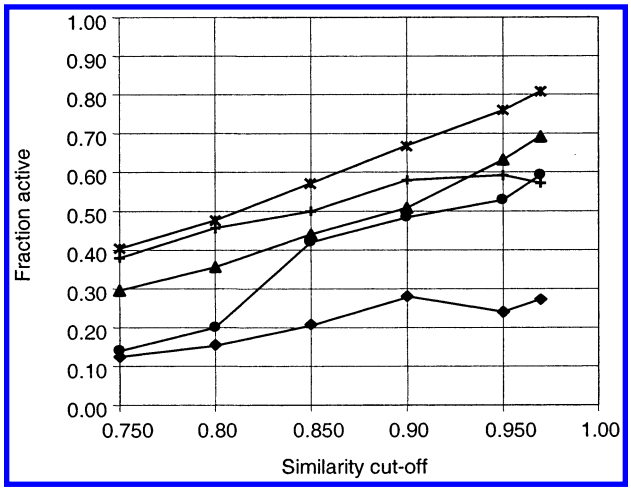


Figure 4. For five selected assays, a comparison of the fraction of similars to a potent active that are themselves active as a function of the similarity threshold used for the searching.

pounds are retrieved in the searches based on longer fingerprints.

In light of the above results, it is important to demonstrate that molecular similarity is indeed related to biological similarity. Figure 4 shows the effect of increasing the similarity threshold on the fractions of similars to an active that is active. Four assays were excluded because not enough actives or similars were in the database to give reliable results at higher similarities. The figure shows a clear trend toward a parallel between chemical and biological similarity.

To further expand this analysis, we examined the MAO data set. Figure 5 shows the fraction actives within those compounds similar to inactive, weakly active, moderately active, and potent compounds. Because we knew that all compounds had been tested in this assay, we expanded the similarity search range down to ≥0.5 similarity. Note that the fraction active within the set of compounds similar to an inactive is lower than the fraction of actives within the whole data

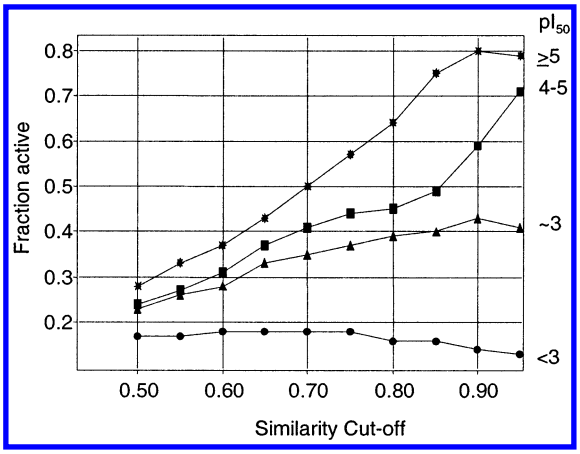


Figure 5. For the monoamine oxidase screening set, a comparison of the fraction of similars to an active that are themselves active as a function of the potency of the compounds used for the similarity search and of the similarity threshold used for the search.

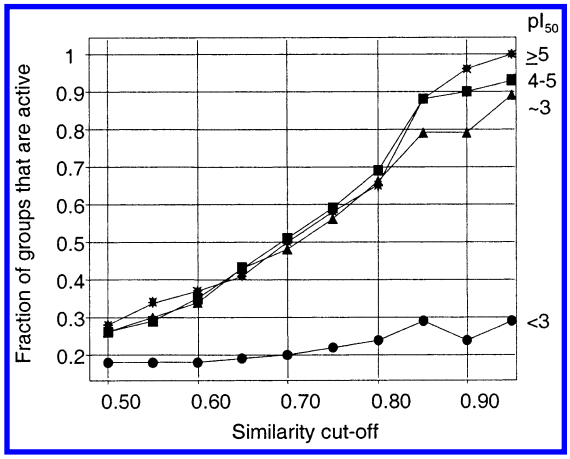


Figure 6. For the monoamine oxidase screening set, the fraction of clusters identified at ≥0.85 similarity that are active vs the similarity threshold used to identify the similar compounds.

set and that this fraction decreases as the similarity to an inactive is increased. Also note that the fraction active increases with the potency of the compounds used for the similarity search and with the similarity cutoff used in the search. Together, these observations validate the principle that two similar compounds are more likely to share the same biological properties than are two compounds selected at random.

It is possible that the MAO data set is biased because it contains compounds synthesized to follow up a lead; therefore, we also investigated how many unique clusters of compounds were found in the various similarity searches. For this, we catalogued the clusters, calculated based on ≥0.85 similarity, to which the similars belong and also the clusters to which the active subset of similars belong. Figure 6 shows that the fraction of clusters identified that contains an active molecule increases as the similarity threshold is tightened. Hence, we conclude that the results from the MAO data set are not misleading.

Discussion

How Frequently Are Molecules Similar in Structure to an Active Themselves Active? As expected,

for compounds that are similar to an active, the proportion of actives is increased as compared to "random" screening. The proportion of ≥0.85 similars to an active that are themselves active is approximately 0.30. This represents an enrichment of at least 30-fold as compared to random screening for which hit rates of less than 0.01 are observed. The enrichment is approximately 10-fold higher than we found with virtual screening using Dock and PMF scoring, which itself was 2-fold higher than Dock and Amber scoring.[26] Hence, two-dimensional (2D) similarity searching is a valuable way to find active compounds. Of course it will not, by definition, find new types of active compounds.

However, although pairs of ≥0.85 similar compounds are more likely to have similar biological activity than are two dissimilar compounds, this frequency is low enough that screening truly diverse data sets will likely miss finding clusters of active compounds. If only one compound of a similarity set is screened, there is a 70% chance that the activity within this cluster will not be discovered; if five similars are screened, the probability of missing the active compound drops to 17% $(0.70)^5$ and if ten, it drops to 3%. Of course, increasing the number of similars in a database also decreases the diversity of the database. Further studies will be needed to optimize the number of similars that should be included.

Both Table 2 and Figure 5 suggest that the fraction of actives within similar compounds increases with the potency of the active used for the search. If this observation holds true, it further supports the notion that it is a poor strategy to exclude similar compounds from a screening library.

These results with Daylight or Unity fingerprints are not unprecedented. In an early study, as noted above, Delaney found a range of fraction actives of 0.4−0.6.[14] Recall, however, that he included in the ratio actives for which no similar was identified and that hence his values were inflated as compared to those reported here.

A number of alternative descriptors tested by the Merck group again show less than ideal relationships between similarity in chemical descriptors and biological activity. They compared the ability of similarity searching using a number of descriptors to find actives in the highest ranked 300 compounds of a large data set. In the test of performance of physicochemical property topological descriptors in only one of the 10 test cases, comparing the ability of nine possible descriptors was the most favorable fraction active >0.5.[27] In three cases, it was less than 0.1. Slightly worse results were found using geometric atom pair descriptors with eight of the data sets.[28] Similar results were found with similarities calculated from the same two types of fingerprint by the LaSSI method, latent semantic structural indexing.[29]

In a report concentrating on a novel clustering method, Reynolds[30] reported 0.43−0.70 fraction actives in searches run at various similarity levels using either topological torsions or atom pair fingerprints. In their comparison of similarity searching to binary kernel discrimination, Harper et al. found some but not extraordinary enrichment of the fraction of actives in the set of similar compounds.[31] For example, using the MAO data set, if 100 similars to half of the actives are selected, 50 of these compounds are active. In a corresponding test for a HTS data set, the number of actives is only 10.

Dixon and Merz showed that in a SAR series the correlation between the Daylight similarity of a molecule to the tightest CBG binder, 11-deoxycortisol, and its binding affinity is 0.66.[32] While significant, this result indicates that only 43% of the variance in affinity is explained by the similarity or lack of it to the reference compound. From their plots, it appears that fully one-third of the pairs of compounds that are ≥0.85 similar (Daylight fingerprints) differ by more than one log unit in potency. Additionally, within series of compounds whose potency is predicted by an optimum number of near neighbors, the average correlation coefficient for such predictions is 0.48.

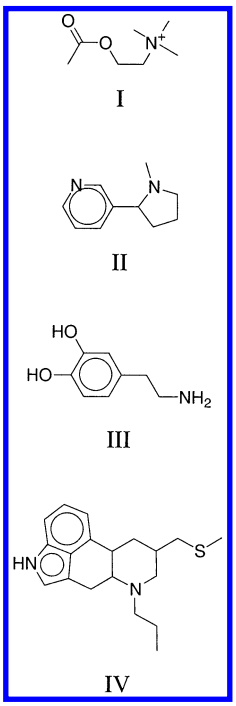
Dixon and Merz also investigated the biological information contained in a novel one-dimensional (1D) representation of a molecule.[32] In the case of 1D descriptors generated from three-dimensional (3D) structures, denoted 3D→1D, the five compounds that are ≥0.85 similar to 11-deoxycortisol vary in potency by approximately an order of magnitude and the 10 that are ≥0.80 similar vary in affinity by 2 orders of magnitude.[32] For 1D fingerprints generated from 2D structures, 2D→1D, the corresponding numbers are approximately 2 and 3 orders of magnitude. In the two examples shown, plots show that compounds that are ≥0.85 similar in 3D→1D fingerprints vary in potency by 3 orders of magnitude. Again, similar compounds by this measure do not have similar biological activity.

Although the results with alternative types of fingerprints discussed above emphasize their lack of ability to capture all of the features relevant for biological activity, because they are different from traditional substructure descriptors, they identify different molecules as similar to an active. Hence, they provide a complementary method of searching for potentially active compounds.

Do Biologically Similar Compounds Have Similar Structures? Considering such classic example pairs as the nicotinic agonists acetylcholine (I, Chart 1) and nicotine (II) or the dopaminergic agonists dopamine (III) and pergolide (IV), the expected answer is no. In fact, the highest Daylight Tanimoto similarity within this group of four compounds (0.32) is between nicotine and pergolide and the second highest (0.22) is between nicotine and dopamine. Nevertheless, other types of fingerprints might encode the features necessary for biological activity and hence be suitable alternatives for designing a screening collection. However, several groups have shown that Daylight or Unity fingerprints are more similar for compounds with the same biological property than to compounds with different biological activity. In the context of the present report, the interesting finding is that one must go to similarity levels much lower than 0.85 to find this relationship.

In a comparison of structurally different but bioisosteric compounds, Schuffenhauer, Gillett, and Willett showed that 2995 bioisosteric pairs of molecules had a mean Unity similarity of 0.54 as compared with 0.22 for random pairs.[33] Makara compared ToPD, total pharmacophore diversity, fingerprints with Unity fingerprints on 10 small series of molecules chosen from the Protein Data Bank.[34] Only 4% of the structural pairs

Chart 1



chosen from the PDB are ≥0.85 similar in Unity fingerprints; the median similarity is 0.5. Despite this, at a similarity level that includes only 5% of the negatives (similarity of approximately 0.4), Unity fingerprints recognize 94% of the similar pairs. For the ToPD descriptors, Makara found that a similarity of 0.7 is required to identify all pairs of biologically similar molecules—at ≥0.85 similarity, only 50% of the pairs of close analogues are identified.[34] At ≥0.70 similarity, there is a complete separation of the pairs of biologically similar molecules from the dissimilar pairs. A more difficult test compared molecules that bind to the same binding site in a protein with those that bind to a different protein. In this case, the best result was a 14% false negative rate, or a fraction active of 0.86. The corresponding number for Unity fingerprints was 0.69. Because this method was not tested in large diverse data sets, it is difficult to discern the ultimate value of the ToPD descriptors.

Dixon and Merz studied the ability of similarity measures to predict which of six possible therapeutic classes a compound belongs.[32] At the optimum similarity radius of 0.54, Daylight fingerprints were able to correctly predict the class only 66% of the time. ACE inhibitors, β-blockers, and H2 antihistamines, all series known to include extensive 2D similarity, dominate the correct classifications.

In a series of studies, workers from New Chemical Entities compared traditional 2D descriptors with short binary bit string representations that characterize molecular properties related to intermolecular interactions as well as including some traditional substructural keys.[35−38] These minifingerprints are designed to provide an alternative to 3D searching for identifying new structural classes with the same biological activity. Although they report high accuracy in predicting the biological class to which the compounds belong, in some cases the compositions of the fingerprints are optimized to provide this prediction, similar to a quantitative SAR investigation. Nonetheless, the good performance of short fingerprints suggests that they merit further investigation.

Are the Daylight Fingerprints an Acceptable Way to Quantify Similarity? Although we do not argue that the Daylight fingerprint is the best descriptor to use for similarity analysis, there is ample evidence that it does contain information with respect to the biological activity of compounds. Because they are complex descriptions of the 2D structures of the molecules, molecules that are similar in Daylight fingerprints in general appear similar to a synthetic chemist. For this reason this type of fingerprint is appropriate for diversity analysis aimed at populating a database with compounds that could appear in different patents. Other types of fingerprints may be appropriate for other uses.

Despite the fact that our early work showed slightly better performance of substructure-based descriptors, we did find good performance of both Unity and Daylight fingerprints in clustering biologically active compounds[16] and in predicting physical properties.[19] In our studies, Unity and Daylight fingerprints were essentially identical in performance. Hence, we assume that validations of Unity fingerprints also validate Daylight fingerprints.

Many groups have studied alternate molecular descriptors for similarity measurements related to biological activity. Often, but not always, Unity or Daylight fingerprints outperform more novel descriptors. Delaney reported that the Daylight fingerprints outperform Molconn topological indices and Unity fingerprints.[14] Matter showed the superiority of Unity fingerprints to various topological and 3D descriptors.[11] Additionally, Ginn, Turner, and Willett demonstrated that Unity descriptors perform equally well as the EVA descriptor in predicting octanol−water log *P* from similar compounds.[39] Also, Briem and Kuntz compared Dock-generated fingerprints with Daylight fingerprints for their ability to recognize the activity class of a molecule when seeded into a database that also contains molecules of the same class, molecules of four other classes, and twice as many compounds that do not belong to any of the five classes.[40] They found that using the most similar compound to predict the activity class of another was successful 62−86% of the time with Daylight fingerprints and less with Dock-derived fingerprints.

There are a number of recognized limitations of similarity calculations based on Daylight fingerprints.[6,41] Because Daylight and 2D Unity fingerprints consider the character of bond paths up to only a certain number of bonds, they do not discriminate between molecules that differ only in bond paths longer than the maximum. In addition, these fingerprints do not explicitly include shape information because molecules with different atom types in the same environment are not considered identical to the algorithms. The default fingerprints do not include information about the stereochemistry of the molecules, with the result that the fingerprints cannot distinguish stereoisomers although the biological target might recognize the difference. Using 2D Unity fingerprints, Flower demonstrated that the more bits set in the query molecule, the higher is the average Tanimoto coefficient of all other molecules in the database to it.[41]

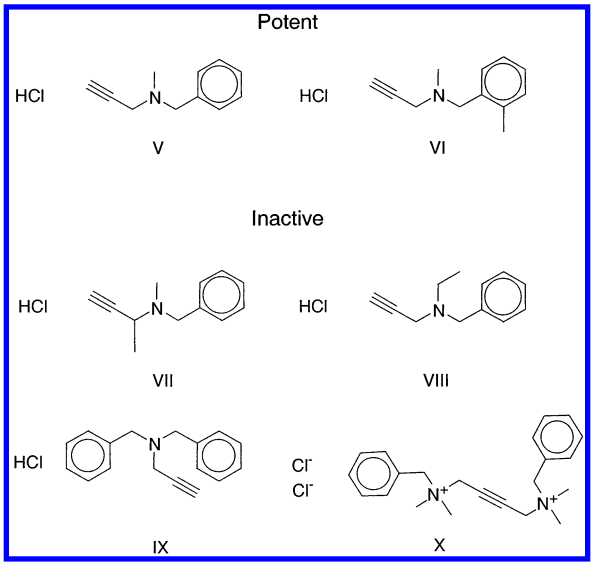


Figure 7. Pargyline (V) and five compounds ≥0.85 similar to it.

This bias, which artificially inflates the number of similars toward more complex molecules, complicates the interpretation of results such as reported here.

What Is the Error in the Analyses That Led to the 0.85/80% Rule? The previous analyses of activity of similar molecules were not based on screening results on truly diverse data sets. For example, our earlier conclusions used two data sets that were rich in compounds prepared for the particular targets.[8,9] The current results suggest that the medicinal chemists had already identified the pharmacophore required for biological activity and so prepared only compounds that they thought should be active, hence the high rate of activity in similar compounds. Others used SAR series[42] or databases that had been constructed of sets of molecules active in different assays.[11,12] Additionally, in general, these data sets contain potent compounds, which we show in Figure 5 are more likely to identify similar active compounds than are weakly active compounds. As noted above, the results, albeit on only three sets of compounds, reported by Delaney contradict the 0.85/80% rule and support the conclusions reported here.[14]

Why Does the Similarity Principle Fail? The similarity principle could fail for at least two reasons. A problem that could in principle be overcome would be if the similarity computation did not correctly quantitate the intuitive similarity between two chemical structures. A more serious challenge to molecular similarity/diversity analysis would be if in fact intuitively similar chemical structures do not have similar biological activity.

Figure 7 shows that Daylight fingerprints do not encapsulate the structural similarity and dissimilarity between compounds. It contains a sample of the compounds that are ≥0.85 similar to pargyline, V. The one active analogue, VI, shown would not surprise a medicinal chemist. However, compounds VII and VIII provide valuable SAR information because they are perceived to be similar but are inactive. However, compounds IX and X, although similar by the algorithm, would be perceived to be less similar to V; hence, it is not surprising or perhaps even interesting that they are inactive. Work to improve the match between empirical and computational perception of similarity would certainly be valuable. Many examples of studies of this issue are cited above.

The Tanimoto coefficient also has its limitations.[43] In particular, it is not an even distribution. However, the more serious challenge to the similarity principle comes from biology itself. Structural biology has taught us that protein structures are complex and flexible.[44] For example, compounds that look very similar to a chemist sometimes bind in very different orientations in the protein active site, bind to a different conformation of a protein, or bind to a different protein altogether.[15] In fact, such observations are why medicinal chemists need to make so many compounds to optimize the biological activity of a structural class, even when they are designing to a biological target of known structure.

Implications for Compound Selection and Library Design. The current work does not support the practice of rejecting compounds that are ≥0.85 similar to a compound already in the screening collection. Rather, the optimum library will have a small number of such similar compounds but not so many as to compromise diversity. For the computational chemist, a simple comparison of every potential added compound with every available compound no longer suffices. Instead, one would need to keep a tally, for each compound, of the number of similars it has in the existing database plus the number of similars to it that one has already proposed to add to the database. The target number for the size of a similarity cluster might contain factors that allow for errors in the structure of the compound or the probability that a vendor will not be able to supply the compound. The informatics aspect of such a strategy would involve keeping lists of both the available and the selected compounds, updating lists as new compounds are registered, updating compounds that are removed from the vendor's catalog, or updating the current sample when it is exhausted.

If a combinatorial library is to be designed without regard to similarity to existing compounds, then methods such as genetic algorithms could be used to optimize the size of clusters.[4,45−49] Although the implications of this work for similarity searching are beyond the scope of the current report, Figures 4−6 suggest that an enrichment of actives will be found at similarity thresholds well below the traditional ≥0.85 cutoff.

Conclusions

This work demonstrates that structurally similar compounds do have similar biological activity and that as the structural similarity is increased, so is the biological similarity. The enrichment in actives is higher than is found with current methods of docking to proteins of known 3D structure.

This work also shows that the biological similarity is not so strong as has previously been assumed. For example, at ≥0.85 Tanimoto similarity in Daylight fingerprints, only 30% of compounds similar to an active are themselves active. These results require a rethinking of strategies for compound acquisition and design of combinatorial libraries.

References

(1) Johnson, M.; Lajiness, M.; Maggiora, G. M. Molecular Similarity: A Basis for Designing Drug Screening Programs. In *QSAR*: *Quantitative Structure−Activity Relationships in Drug Design*; Fauchere, J. L., Eds.; Alan R. Liss: New York, 1989; pp 167−171.

(2) Lajiness, M. S.; Johnson, M. A.; Maggoria, G. M. Implementing Drug Screening Programs Using Molecular Similarity Methods. In *QSAR*: *Quantitative Structure−Activity Relationships in Drug Design*; Fauchere, J. L., Eds.; Alan R Liss: New York, 1989; pp 173−176.

(3) Lajiness, M. Molecular Similarity-Based Methods for Selecting Compounds for Screening. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Science Publishers: New York, 1990; pp 299−316.

(4) Lewis, R. A.; Pickett, S. D.; Clark, D. E. Computer-Aided Molecular Diversity Analysis and Combinatorial Library Design. *Rev. Comput. Chem.* **2000**, *16*, 1−51.

(5) Martin, Y. C.; Willett, P.; Lajiness, M.; Johnson, M.; Maggiora, G.; Martin, E.; Bures, M. G.; Gasteiger, J.; Cramer, R. D.; Pearlman, R. S.; Mason, J. S. Diverse Viewpoints on Computational Aspects of Molecular Diversity. *J. Comb. Chem.* **2001**, *3*, 231−250.

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

(7) *Unity Chemical Information Software*; Tripos Associates: St. Louis, MO.

(8) Brown, R. D.; Bures, M. G.; Martin, Y. C. Similarity and Cluster-Analysis Applied to Molecular Diversity. *Chem. Abstr.* **1995**, *209*, 3-COMP.

(9) Brown, R. D.; Martin, Y. C. An Evaluation of Structural Descriptors and Clustering Methods for Use in Diversity Selection. *SAR QSAR Environ. Res.* **1998**, 23−39.

(10) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior—a Useful Concept for Validation of Molecular Diversity Descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.

(11) Matter, H. Selecting Optimally Diverse Compounds from Structure Databases—a Validation Study of Two-Dimensional and Three-Dimensional Molecular Descriptors. *J. Med. Chem.* **1997**, *40*, 1219−1229.

(12) Potter, T.; Matter, H. Random or Rational Design—Evaluation of Diverse Compound Subsets from Chemical Structure Databases. *J. Med. Chem.* **1998**, *41*, 478−488.

(13) Taylor, R. Simulation Analysis of Experimental Design Strategies for Screening Random Compounds as Potential New Drugs and Agrochemicals. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 59−67.

(14) Delaney, J. S. Assessing the Ability of Chemical Similarity Measures to Discriminate between Active and Inactive Compounds. *Mol. Diversity* **1996**, *1*, 217−222.

(15) Kubinyi, H. Similarity and Dissimilarity − a Medicinal Chemists View. *Perspect. Drug Discovery Des.* **1998**, *11*, 225−252.

(16) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.

(17) *Thor and Merlin*; Version 4.62; Daylight Chemical Information Systems Inc.: Irvine, CA. Theory at www.daylight.com.

(18) Martin, Y. C. Information on Accessing the Monoamine Oxidase Dataset: yvonne.c.martin@abbott.com, 1996.

(19) Brown, R. D.; Martin, Y. C. The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand−Receptor Binding. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1−9.

(20) Yang, J. *Merq, a Daylight Contrib Program*; Daylight Chemical Information Systems Inc.: Irvine, CA. http://www.daylight.com/support/f_download.html

(21) Butina, D. Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way to Cluster Small and Large Data Sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747−750.

(22) Chu, D. T.; Fernandes, P. B.; Claiborne, A. K.; Pihuleac, E.; Nordeen, C. W.; Maleczka, R. E., Jr.; Pernet, A. G. Synthesis and Structure−Activity Relationships of Novel Arylfluoroquinolone Antibacterial Agents. *J. Med. Chem.* **1985**, *28*, 1558−1564.

(23) Chu, D. T.; Fernandes, P. B.; Claiborne, A. K.; Gracey, E. H.; Pernet, A. G. Synthesis and Structure−Activity Relationships of New Arylfluoronaphthyridine Antibacterial Agents. *J. Med. Chem.* **1986**, *29*, 2363−2369.

(24) Chu, D. T.; Fernandes, P. B.; Maleczka, R. E., Jr.; Nordeen, C. W.; Pernet, A. G. Synthesis and Structure−Activity Relationship of 1-Aryl-6,8-Difluoroquinolone Antibacterial Agents. *J. Med. Chem.* **1987**, *30*, 504−509.

(25) Chu, D. T.; Fernandes, P. B. Structure−Activity Relationships of the Fluoroquinolones. *Antimicrob. Agents Chemother.* **1989**, *33*, 131−135.

(26) Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF Scoring in Docking Weak Ligands to the Fk506 Binding Protein. *J. Med. Chem.* **1999**, *42*, 2498−2503.

(27) Kearsley, S. K.; Sallamack, S.; Fluder, E. M.; Andose, J. D.; Mosley, R. T.; Sheridan, R. P. Chemical Similarity Using Physicochemical Property Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 118−127.

(28) Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical Similarity Using Geometric Atom Pair Descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128−136.

(29) Hull, R. D.; Fluder, E. M.; Singh, S. B.; Nachbar, R. B.; Kearsley, S. K.; Sheridan, R. P. Chemical Similarity Searches Using Latent Semantic Structural Indexing (LASSI) and Comparison to TOPOSIM. *J. Med. Chem.* **2001**, *44*, 1185−1191.

(30) Reynolds, C. H.; Druker, R.; Pfahler, L. B. Lead Discovery Using Stochastic Cluster Analysis (SCA) − a New Method for Clustering Structurally Similar Compounds. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 305−312.

(31) Harper, G.; Bradshaw, J.; Gittins, J. C.; Green, D. V. S.; Leach, A. R. Prediction of Biological Activity for High-Throughput Screening Using Binary Kernel Discrimination. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1295−1300.

(32) Dixon, S. L.; Merz, K. M. One-Dimensional Molecular Representations and Similarity Calculations: Methodology and Validation. *J. Med. Chem.* **2001**, *44*, 3795−3809.

(33) Schuffenhauer, A.; Gillet, V. J.; Willett, P. Similarity Searching in Files of Three-Dimensional Chemical Structures: Analysis of the Bioster Database Using Two-Dimensional Fingerprints and Molecular Field Descriptors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 295−307.

(34) Makara, G. M. Measuring Molecular Similarity and Diversity: Total Pharmacophore Diversity. *J. Med. Chem.* **2001**, *44*, 3563−3571.

(35) Xue, L.; Godden, J. W.; Bajorath, J. Evaluation of Descriptors and Mini-Fingerprints for the Identification of Molecules with Similar Activity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1227−1234.

(36) Xue, L.; Stahura, F. L.; Godden, J. W.; Bajorath, J. Mini-Fingerprints Detect Similar Activity of Receptor Ligands Previously Recognized Only by Three-Dimensional Pharmacophore-Based Methods. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 394−401.

(37) Xue, L.; Godden, J. W.; Bajorath, J. Database Searching for Compounds with Similar Biological Activity Using Short Binary Bit String Representations of Molecules. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 881−886.

(38) Xue, L.; Bajorath, J. Molecular Descriptors for Effective Classification of Biologically Active Compounds Based on Principal Component Analysis Identified by a Genetic Algorithm. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 801−809.

(39) Ginn, C.; Turner, D. B.; Willett, P.; Ferguson, A. M.; Heritage, T. W. Similarity Searching in Files of 3-Dimensional Chemical Structures − Evaluation of the Eva Descriptor and Combination of Rankings Using Data Fusion. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 23−37.

(40) Briem, H.; Kuntz, I. D. Molecular Similarity Based on Dock-Generated Fingerprints. *J. Med. Chem.* **1996**, *39*, 3401−3408.

(41) Flower, D. R. On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 379−386.

(42) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.

(43) Godden, J. W.; Xue, L.; Bajorath, J. Combinatorial Preferences Affect Molecular Similarity/Diversity Calculations Using Binary Fingerprints and Tanimoto Coefficients. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 163−166.

(44) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; Bode, W.; Huber, R.; Poll, T.; Rudolph, R.; Vondersaal, W. Enzyme Flexibility, Solvent and Weak Interactions Characterize Thrombin-Ligand Interactions − Implications for Drug Design. *Structure* **1996**, *4*, 1353−1362.

(45) Gillet, V. J.; Willett, P.; Bradshaw, J. The Effectiveness of Reactant Pools for Generating Structurally-Diverse Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 731−740.

(46) Brown, R. D.; Martin, Y. C. Designing Combinatorial Library Mixtures Using a Genetic Algorithm. *J. Med. Chem.* **1997**, *40*, 2304−2313.

(47) Liu, D. X.; Jiang, H. L.; Chen, K. X.; Ji, R. Y. A New Approach to Design Virtual Combinatorial Library with Genetic Algorithm Based on 3d Grid Property. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 233−242.

(48) Agrafiotis, D. K. Multiobjective Optimization of Combinatorial Libraries. *IBM J. Res. Dev* **2001**, *45*, 545−566.
(49) Zheng, W. F.; Cho, S. J.; Waller, C. L.; Tropsha, A. Rational Combinatorial Library Design. 3. Simulated Annealing Guided Evaluation (Sage) of Molecular Diversity: A Novel Computational Tool for Universal Library Design and Database Mining. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 738−746.

JM020155C