Similarity Search Profiling Reveals Effects of Fingerprint Scaling in Virtual Screening

Ling Xue,[†] Florence L. Stahura,[†] and Jürgen Bajorath*[,†,‡]

Department of Computer-Aided Drug Discovery, Albany Molecular Research, Inc. (AMRI), AMRI Bothell Research Center (AMRI-BRC), 18804 North Creek Parkway, Bothell, Washington 98011-8012, and Department of Biological Structure, University of Washington, Seattle, Washington 98195

Received April 29, 2004

Fingerprint scaling is a method to increase the performance of similarity search calculations. It is based on the detection of bit patterns in keyed fingerprints that are signatures of specific compound classes. Application of scaling factors to consensus bits that are mostly set on emphasizes signature bit patterns during similarity searching and has been shown to improve search results for different fingerprints. Similarity search profiling has recently been introduced as a method to analyze similarity search calculations. Profiles separately monitor correctly identified hits and other detected database compounds as a function of similarity threshold values and make it possible to estimate whether virtual screening calculations can be successful or to evaluate why they fail. This similarity search profile technique has been applied here to study fingerprint scaling in detail and better understand effects that are responsible for its performance. In particular, we have focused on the qualitative and quantitative analysis of similarity search profiles under scaling conditions. Therefore, we have carried out systematic similarity search calculations for 23 biological activity classes under scaling conditions over a wide range of scaling factors in a compound database containing ∼1.3 million molecules and monitored these calculations in similarity search profiles. Analysis of these profiles confirmed increases in hit rates as a consequence of scaling and revealed that scaling influences similarity search calculations in different ways. Based on scaled similarity search profiles, compound sets could be divided into different categories. In a number of cases, increases in search performance under scaling conditions were due to a more significant relative increase in correctly identified hits than detected false-positives. This was also consistent with the finding that preferred similarity threshold values increased due to fingerprint scaling, which was well illustrated by similarity search profiling.

INTRODUCTION

Molecular fingerprints are widely used tools for similarity searching,[1] in addition to, for example, structural queries,[2] graph representations,[3] or pharmacophore models.[4] Fingerprints are designed as linear bit strings that encode molecular structure or properties by use of diverse types of descriptors.[1] Many different fingerprints have been introduced that in part substantially vary in terms of size, complexity, and descriptors that are incorporated.[1,5] A major distinguishing feature is whether bit strings are keyed or hashed.[5] In keyed representations, every bit is associated with a specific structural feature, property, or value range, whereas hashing (or folding) maps encoded properties to overlapping bit segments. Database searching using fingerprints transforms molecular similarity analysis into a comparison of bit strings and patterns of query and source compounds. It is assumed that similarity of bit patterns corresponds to molecular similarity and thus it is required to quantify fingerprint overlap, for which a variety of similarity coefficients and metrics are available.[6] For fingerprint-based virtual screening of compound databases, threshold values of chosen similarity coefficients must be predefined in order to identify potential hits, which is a nontrivial problem for many applications.[7] Appropriate calculation parameters and search performance typically depend not only on the search tools used but also on the characteristic features of active compounds.[5,8]

Methods have been introduced to modulate and improve fingerprint search calculations when multiple query compounds are available including the generation of consensus or modal fingerprints[9] and profile scaling.[10,11] These approaches essentially try to identify common characteristics within a series of template molecules that are relevant for their activity and capture them in fingerprint representations. This often increases the information content of search calculations based on single query compounds. Other approaches to increase search performance include consensus scoring[12] and data fusion.[13] These techniques can also be applied to search calculations based on single query molecules because they operate at the level of scoring, rather than fingerprinting, and combine information from multiple scoring schemes or similarity metrics.

The fingerprint scaling technique we have developed is applicable to any keyed fingerprint design and exploits the observation that compounds having similar activity often produce characteristic bit patterns in these fingerprints that are distinct from other activity classes.[10] Application of scaling factors to signature bits increased the probability of identifying compounds with similar activity in databases.[10]

* Corresponding author phone: (425)424-7297; fax: (425)424-7299; e-mail: jurgen.bajorath@albmolecular.com. Correspondence should be addressed at AMRI-BRC.

[†] AMRI Bothell Research Center.

[‡] University of Washington.

10.1021/ci0400819 CCC: $27.50
Published on Web 09/03/2004

Table 1. Compound Classes and Search Results[a]

				nonscaled				scaled				
abbrev	biological activity	Nt	Np	$[avTc]_I$	$[hit/dbc]_I$	hit/100dbc	avTc	sf	$[avTc]_I$	$[hit/dbc]_I$	hit/100dbc	avTc
BLC	β-lactamase inhibitors	7	7	0.66	4.9/162.7	4.9	0.67	50	0.89	6.0/109.9	6.0	0.89
PKC	protein kinase C inhibitors	7	8	0.71	2.9/64.6	2.9	0.70	20	0.92	2.9/61.4	2.9	0.92
ADR	adrenergic receptor ligands	8	8	0.69	6.3/191.6	5.5	0.72	20	0.93	7.0/95.6	7.0	0.93
GLU	glucocorticoid analogues	7	7	0.83	1.4/45.3	2.0	0.80	4	0.90	2.9/101.7	2.9	0.90
BEN	benzodiazepine receptor ligands	11	11	0.76	2.9/43.2	3.3	0.75	20	0.91	3.8/56.2	3.8	0.91
CAE	carbonic anhydrase II inhibitors	11	11	0.73	2.9/50.9	3.1	0.72	10	0.89	5.8/125.0	5.8	0.89
H3E	H3 antagonists	10	11	0.69	7.2/192.7	7.0	0.71	50	0.90	9.0/186.7	7.0	0.91
TKE	tyrosine kinase inhibitors	10	10	0.82	1.2/25.5	1.8	0.79	50	0.98	8.0/188.8	7.4	0.99
5HT	serotonin receptor ligands	10	11	0.81	1.2/28.4	1.6	0.77	50	0.98	2.2/29.0	5.0	0.97
HIV	HIV protease inhibitors	9	9	0.75	1.3/39.1	1.8	0.74	50	0.93	6.0/99.0	6.0	0.93
COX	cyclooxygenase-2 inhibitors	8	9	0.76	2.3/58.9	2.3	0.75	50	0.94	7.0/170.0	7.0	0.95
ANG	angiotensin AT1 antagonists	5	5	0.70	2.0/81.2	2.0	0.70	50	0.89	4.0/136.2	4.0	0.89
ARO	aromatase inhibitors	5	5	0.78	0.4/14.8	0.4	0.75	50	0.96	0.4/11.8	0.8	0.94
DIH	dihydrofolate reductase inhibitors	5	6	0.77	0.8/36.2	0.8	0.75	50	0.93	2.4/114.2	2.4	0.93
FAC	factor Xa inhibitors	7	7	0.76	1.1/38.4	1.1	0.74	50	0.94	2.9/46.6	2.9	0.94
MAT	matrix metalloproteinase inhibitors	6	6	0.69	3.3/141.7	3.3	0.69	2	0.75	3.7/140.0	3.3	0.76
VIT	vitamin D analogues	6	6	0.72	4.3/153.0	4.3	0.73	50	0.86	5.0/121.7	5.0	0.86
RTI	reverse transcriptase inhibitors	7	8	0.76	0.6/16.9	0.9	0.72	50	0.96	1.1/29.4	1.4	0.94
PPAR	PPARgamma agonists	8	8	0.69	3.8/88.4	3.8	0.70	4	0.80	4.5/109.8	4.5	0.80
DD2A	dopamine D2 antagonists	7	7	0.72	0.9/28.9	1.4	0.70	4	0.83	0.9/29.1	0.9	0.81
CRF1	CRF1 antagonists	6	6	0.75	1.0/41.5	1.3	0.73	2	0.79	1.3/51.0	1.7	0.78
CALA	calcium antagonists	9	9	0.86	0.2/5.0	0.9	0.77	50	0.93	2.0/57.9	2.2	0.92
ARI	aldose reductase inhibitors	6	6	0.80	0.3/11.0	1.0	0.73	50	0.94	1.0/34.3	1.3	0.93

[a] Nt and Np report the number of compounds in the test and profile set, respectively, and dbc means database compounds. In each search calculation, the number of potential hits is (Nt − 1) because one active compound at a time is used as the query. $[avTc]_I$ is the similarity threshold value at the intersection point between hit and dbc curves and $[hit/dbc]_I$ the ratio of correctly identified hits and dbc at this point; hits/100dbc reports the number of hits simultaneously recognized with ∼100 dbc, which represents the approximate hit rate, and avTc is the corresponding similarity threshold value; sf stands for scaling factor. Nonscaled calculations correspond to previous MP-MFP virtual screening trials.[14]

In extensive similarity search calculations on many activity classes, scaling improved overall search performance by up to 15% dependent on the fingerprint that was used.[11]

Recently, similarity search profiling has been introduced as a diagnostic method to evaluate fingerprint-based virtual screening.[14] This analysis is based on systematic similarity search calculations using multiple template compounds over the entire value range of a chosen similarity coefficient followed by side-by-side graphical representations of numbers of correctly identified hits and other database compounds. The profiling method monitors distributions of hits and false-positives during virtual screening calculations as a function of similarity threshold values. This makes it possible to estimate whether a virtual screening trial can in principle succeed for a given compound class, source database, search tool, and similarity metric. Therefore, this approach extends other measures of compound retrieval.[15]

Here we have used similarity search profiling to analyze the effects of fingerprint scaling on different biological activity classes. Consistent with previous findings,[11] scaling generally increased observed hit rates and analysis of search profiles revealed that scaling influenced similarity search calculations in different ways, dependent on the studied compound classes. Profiles often displayed distinct and systematic differences. Furthermore, although high scaling factors frequently produced best results, profile analysis also showed that improvements were often marginal compared to lower factors, thus indicating the potential of moderate fingerprint scaling.

METHODS

Test Compounds and Structural Diversity. In this study, 23 different classes of active compounds were used as test cases, as reported in Table 1. These classes were originally assembled for fingerprint scaling calculations.[11] For similarity searching, each activity class was added to a source database containing ∼1.3 million molecules collected from various vendor sources.[16] For each class, fingerprint scaling requires a separate set of active compounds (not included in the search calculations) to determine consensus bit positions[11] that is called the profile set (Table 1). From each activity class, compounds for the profile set were randomly selected. Previously, we have found that fingerprint profiles and thus the results of scaling calculations were not very sensitive to the composition of the profile set for different activity classes. Furthermore, for many activity classes, fewer than 10 compounds were sufficient to produce scalable MP-MFP profiles with significant predictive value.[11]

Many of the activity classes studied here are structurally diverse. We illustrate this point by analyzing the structural diversity of our randomly selected profile sets. For each set, compounds were compared in a pairwise manner using the MP-MFP fingerprint[17] described in the following section. Representative data are reported in Table 2. Based on these calculations, profiles sets of at least 14 of the 23 activity classes are structurally diverse.

Fingerprint. As a fingerprint for this analysis, MP-MFP was selected that consists of a total of 175 bit positions, 114 of which encode structural fragment descriptors and 61 molecular property descriptors.[17,18] MP-MFP was chosen because it performed overall best under scaling conditions among various keyed fingerprint designs.[11] As a special feature, it contains binary transformed property descriptors. Binary transformation is based on calculation of statistical medians of descriptor value distributions in large compound databases.[16,17] Following this scheme, a descriptor bit is set

Table 2. Structural Diversity of Compounds in Profile Sets[a]

ID	1	2	3	4	5	6	7
				ARI			
1	1.00	*0.39*	*0.48*	0.51	*0.33*	0.64	
2	*0.39*	1.00	*0.48*	0.54	0.60	*0.44*	
3	*0.48*	*0.48*	1.00	0.67	*0.40*	0.61	
4	0.51	0.54	0.67	1.00	*0.48*	0.64	
5	*0.33*	0.60	*0.40*	*0.48*	1.00	*0.34*	
6	0.64	*0.44*	0.61	0.64	*0.34*	1.00	
				DIH			
1	1.00	**0.80**	*0.47*	0.55	*0.49*	*0.46*	
2	**0.80**	1.00	*0.48*	0.52	*0.44*	*0.44*	
3	*0.47*	*0.48*	1.00	0.64	0.59	0.68	
4	0.55	0.52	0.64	1.00	0.61	0.73	
5	*0.49*	*0.44*	0.59	0.61	1.00	0.68	
6	*0.46*	*0.44*	0.68	0.73	0.68	1.00	
				GLU			
1	1.00	0.56	**0.86**	**0.92**	**0.86**	0.65	**0.86**
2	0.56	1.00	*0.50*	0.54	0.59	*0.46*	0.57
3	**0.86**	*0.50*	1.00	**0.78**	**0.76**	**0.75**	0.72
4	**0.92**	0.54	**0.78**	1.00	**0.86**	0.60	**0.88**
5	**0.86**	0.59	**0.76**	**0.86**	1.00	0.61	**0.76**
6	0.65	*0.46*	**0.75**	0.60	0.61	1.00	0.67
7	**0.86**	0.57	0.72	**0.88**	**0.76**	0.67	1.00
				VIT			
1	1.00	**0.80**	0.68	**0.91**	**0.76**	**0.82**	
2	**0.80**	1.00	0.69	**0.82**	**0.87**	**0.77**	
3	0.68	0.69	1.00	0.74	0.73	0.72	
4	**0.91**	**0.82**	0.74	1.00	**0.84**	**0.86**	
5	**0.76**	**0.87**	0.73	**0.84**	1.00	**0.78**	
6	**0.82**	**0.77**	0.72	**0.86**	**0.78**	1.00	

[a] Reported are avTc[17] (see Methods) matrices using MP-MFP for pairwise comparison of compounds in the profile sets of four activity classes. avTc values smaller or equal to 0.5 are italic and values equal to or greater than 0.75 are shown in bold face. From the top to the bottom, structural diversity within the profile set decreases. The top two activity classes are considered to contain compounds of significant structural diversity. For (nonscaled) similarity searching with MP-MFP, avTc values ~0.75 or greater often indicate similar biological activity.[17]

to one if its value for a compound is equal to or greater than the database median and to zero if it is smaller. This provides a mechanism to incorporate property descriptors as single bits in keyed fingerprints.

For fingerprint comparison, the presence of binary transformed descriptors requires the application of a variant of the conventional Tanimoto coefficient (Tc)[6] because in these cases bit positions set to zero capture as much information as those set to one and must be considered when calculating fingerprint overlap. This previously reported Tc variant is called average Tc (avTc).[17] Tc is defined as Tc = bc/(b1+b2-bc); with b1 being the number of bits set to one in molecule 1, b2 the number of bits set to one in molecule 2, and bc the number of bits set on in common to both molecules; avTc is defined as avTc = (Tc+Tc′)/2; with Tc′ being the Tc calculated for bit positions set to zero (rather than one). This averaging procedure produced a sound similarity metric for the application of MP-MFP and related fingerprints.[17]

Fingerprint Scaling. For a set of compounds and a given keyed fingerprint, a so-called fingerprint profile[19] is calculated (for profile set) to identify consensus bit positions. This is done by adding all bits at each position of the fingerprint and dividing the sum by the total number of compounds per class. Therefore, the fingerprint profile reports the relative frequency of bit settings between 0.0 (a bit position is never set on) and 1.0 (a bit is always set on). Then, for similarity search calculations, scaling factors are applied to consensus bits that are set on with a chosen minimum frequency. In this study, scaling factors of two, four, six, eight, 10, 20, 30, 40, and 50 were applied to bit positions that were always or never set on (i.e., frequency 0.0 or 1.0), consistent with the use of MP-MFP as search tool.

Similarity Searching and Profiling. Each active compound was searched against the source database after adding the remaining active compounds as potential hits under nonscaled and scaled conditions (applying different scaling factors in separate runs). For each condition and query compound, avTc similarity threshold values were systematically varied from zero to one in 0.01 increments. Any database compound other than potential hits that was detected at a given threshold level was considered to be inactive and a false-positive. Of course, considering the very large size of our source database, it is expected that a number of database compounds would be novel hits for some of the activity classes. Thus, the hit rates reported here represent a lower limit because some detected database compounds that were considered false-positives could be true hits. Routines for systematic similarity searching with MP-MFP were implemented in the Molecular Operating Environment.[20,21]

Similarity search profiles (SSPs) were generated by separately monitoring correctly identified hits and false-positives at each similarity threshold value and averaging over all compounds per activity class. Then hits and false-positives were plotted versus avTc using separate scales so that a window capturing the range of desired similarity search results was obtained, which in turn was determined by chosen compound selection criteria. Here we have focused on the final selection of ~100 database compounds (which would be a reasonable choice in many practical virtual screening applications) and adjusted the scales for profiling accordingly. The number of correctly identified hits within ~100 database compounds (hit/100dbc) provided the hit rate for each calculation. Compound retrieval rates are not explicitly discussed in the text but also are provided in Table 1 (hit/100dbc versus Nt). Concerning the practicability of virtual screening analysis, hit rates adjusted to a number of database compounds that provide a reasonably sized selection set (e.g., hit/100dbc) are much more relevant than retrieval rates. As discussed in the following, this information can easily be extracted from SSPs that were designed to evaluate the feasibility of virtual screening calculations.

RESULTS AND DISCUSSION

Fingerprint Scaling. The basic ideas of fingerprint scaling are illustrated in Figure 1. First a fingerprint profile for an activity class is generated to determine consensus bit positions. As demonstrated previously, these profiles encode at least some compound class-specific information that can be exploited to increase the sensitivity of similarity search calculations and the probability of identifying compounds having similar activity.[10,11] This is accomplished by applying scaling factors to consensus bit positions that emphasize these bits during similarity searching. Thus, two important variables for fingerprint scaling are the profile cutoff value that determines the minimum bit frequency level for consensus positions and the profile scaling factor that determines the

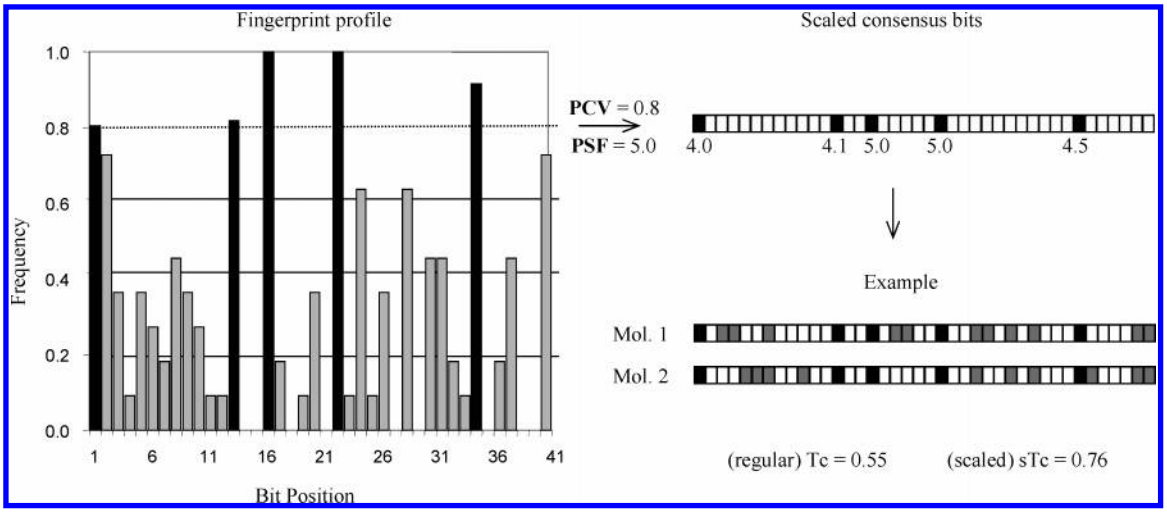

Figure 1. Steps involved in fingerprint scaling. A fingerprint profile for a hypothetical activity class and a short model fingerprint is shown. PCV stands for profile cutoff value and PSF for profile scaling factor. A PCV of 0.8 means that bit positions with a minimum frequency of 0.8 are considered consensus bits (shown in black). Within the consensus frequency interval [0.8,1.0], a PSF of 5.0 is linearly adjusted yielding bit frequency-dependent effective scaling factors. Model fingerprints are shown for two active molecules, the comparison of which results in a Tc value of 0.55. Under scaling conditions, fingerprint overlap effectively increases resulting in a higher Tc value of 0.76.

weights. As shown in Figure 1, scaling factors can be linearly adjusted within the bit frequency interval above the profile cutoff value. However, since in this study only bits that were either always or never set on were considered consensus positions, linear adjustments of scaling factors were not required. The net effect of scaling is that fingerprint overlap increases for molecules sharing consensus bits, which produces higher similarity coefficients, as demonstrated in Figure 1 for two short model fingerprints. This favors the detection of molecular similarity.

Similarity Search Profiles. Figure 2a shows a nonscaled SSP for activity class H3E covering ranges from zero to nine potential hits and zero to 200 detected database compounds (dbc) and outlines a preferred search window in accordance with our compound selection criteria (hits within ~100 database compounds). As the similarity threshold value and search stringency increase during these calculations, fewer compounds are detected as similar to the query molecules. Due to the different scales on the compound ordinate axes, all potential hits are monitored, but the dbc curve enters the profile window only at the preferred level of search stringency (i.e., if fewer than 200 dbc are recognized). The presence of an intersection point between the hit and dbc curves within this profile window away from its axes generally indicates a successful similarity search result, as this example illustrates. At the intersection point ~160 dbc and ~7 hits are simultaneously detected at an avTc threshold value of 0.69. From the intersection point we can extrapolate to the level of ~100 dbc and determine that these compounds contain on average four to five hits. This represents a high-quality search result, given the very large size of the source database and the small number of potential hits.

By contrast, Figure 2b shows the SSP for another test case (calcium antagonists), and we can recognize that this similarity search experiment has little, if any, chance to succeed. Here too many dbc are detected as similar to active query compounds, and the number of dbc is only reduced at high levels of search stringency where true similarity between hits is no longer detected. In this case, 100 selected dbc would contain on average less than one hit, and single search calculations would very likely fail. Clearly, if searches among known actives produce SSPs such as the one shown in Figure 2b, the probability of identifying novel hits in the same source database and using the same search tool is a priori low. Such comparisons nicely illustrate the potential value

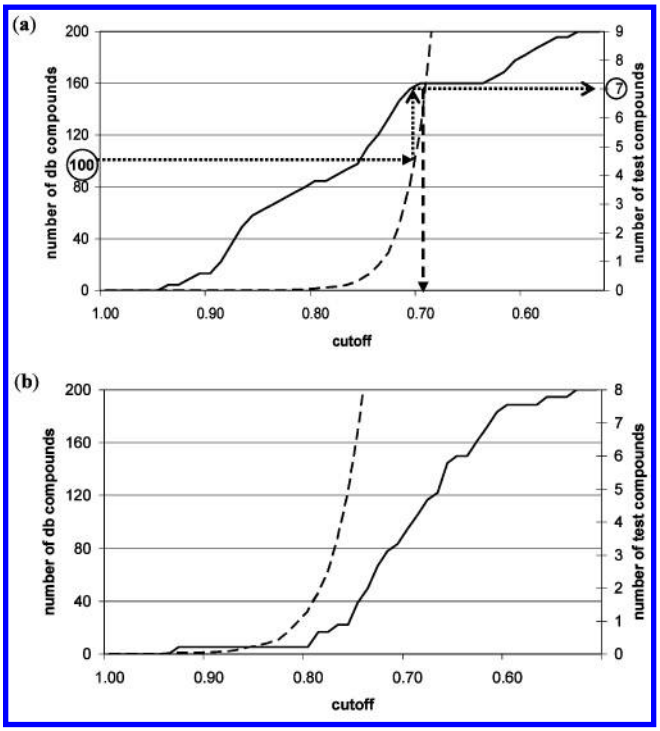

Figure 2. Similarity search profiles and their analysis. Profiles are shown for H3 antagonists (a) and calcium antagonists (b) covering ranges of zero to eight or nine potential hits (test compounds) and zero to 200 database (db) compounds. Curves representing potential hits and other database compounds are drawn as solid and dashes lines, respectively. On the horizontal axis, cutoff stands for the avTc similarity cutoff or threshold value. The SSP in (a) represent a successful similarity search and also illustrates how similarity threshold values and compound numbers reported in Table 1 were determined. (b) shows an SSP for a similarity search experiment that could not succeed for reasons discussed in the text.

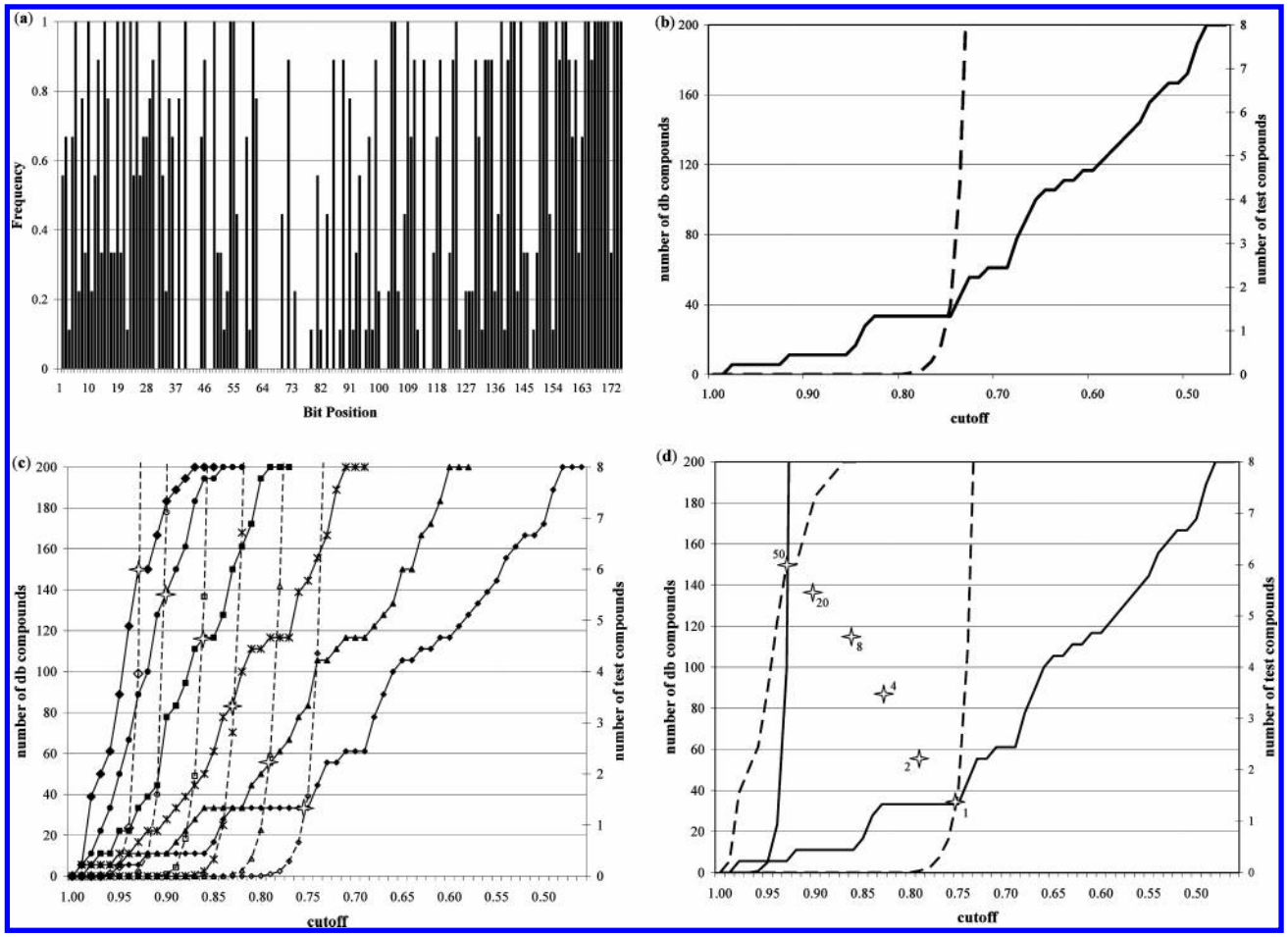

Figure 3. Representative profiles. Shown are the MP-MFP profile (a) for activity class HIV and nonscaled (b) as well as scaled SSPs (c and d). In (c), profiles with different symbols are produced by application of different scaling factors (from the right to the left: 1 (nonscaled), 2, 4, 8, 20, 50) and intersection points between hit (solid) and dbc (dashed) curves are marked with stars. In (d), a simplified scaled SSP is shown with curves representing the starting and end points of scaling and marked intersection points of curves for intermediate scaling factors.

of SSPs to better understand the characteristics of similarity search calculations.

Profiling under Scaling Conditions. The 23 compound classes studied here were selected because all of them were previously tested in nonscaled virtual screening calculations using MP-MFP.[14] In Figure 3, profiles obtained for activity class HIV are shown as a representative data set including its MP-MFP fingerprint profile (a) and nonscaled (b) and scaled SSPs (c and d). Corresponding data sets were collected for all 23 activity classes and analyzed as described in the following. The SSP for nonscaled similarity search calculations reveals that at the intersection point ~40 dbc are recognized together with on average 1.3 hits (yielding an extrapolated hit rate of 1.8% for selection of ~100 dbc; Table 1). Figure 3c shows SSPs for a subset of scaling calculations (for clarity, not all of them are shown within the same graph), which makes it possible to visualize the effects of fingerprint scaling. Under scaling conditions, the curves shift to the left toward higher similarity threshold values, and intersection points move up with increasing scaling factors toward higher dbc and hit numbers. This is well illustrated in the simplified scaled SSPs in Figure 3d where only curves representing the starting (nonscaled) and end point of scaling are displayed together with stars representing the intersection points for intermediate scaling factors. These simplified profiles are shown in the following for other activity classes. For HIV, the profile in Figure 3d reveals that scaling leads to a relative enrichment of active compounds while steadily increasing the optimum similarity threshold value and thus the stringency of similarity searching. The consequence of these effects is a net increase in hit rate from 1.8% to 6% (Table 1).

Comparison of Different Activity Classes. Table 1 reports the search results for all activity classes obtained by SSP analysis, as illustrated in Figure 2a. In general, these similarity search calculations were rather challenging because of the large number of database compounds (all of which were considered inactive) and very small numbers of potential hits (between five and 11 within 1.3 million dbc). In four test cases (ADR, H3E, MAT, DD2A), fingerprint scaling did not further increase similarity search performance. However, scaling improved hit rates of nonscaled MP-MFP calculations in 19 of 23 cases, sometimes significantly (e.g., TKE, COX), which is well in accord with previous observations.[11] Four of our 23 test cases (ARO, DIH, RTI, CALA) were previously considered unsuccessful because MP-MFP produced on average of less than one hit within ~100 database compounds[14] and in all but one case (ARO), scaling notably improved the results.

Preferred scaling factors varied, but overall best results were often obtained for a scaling factor of 50 (Table 1). In each case, scaling increased the avTc value at the intersection

point and the extrapolated threshold value for selection of ~100 dbc. The comparison also revealed systematic differences in preferred similarity threshold values. In general, while an avTc value ~0.75 would produce reasonable results over all classes in nonscaled calculations, scaling would generally require an avTc threshold value of ~0.9, more or less independent of the factors that were applied. Thus, analyses of this kind also make it possible to determine preferred fingerprint-dependent similarity threshold values, which is generally difficult to accomplish.[5,7] It is important to note that scaling effectively increases the length of compared fingerprints at the level of avTc calculations, since the settings of scaled bits are multiplied and counted many times (according to the scaling factor). This explains why higher similarity threshold values better distinguish the recognition of true hits from false-positives under scaling conditions. It also means that comparison of full-length fingerprints is required and that scaled bits positions alone would be insufficient to achieve these results.

Representative Scaled Similarity Search Profiles. Profiles of different activity classes are shown in Figure 4. Analysis of scaled SSPs for all compound classes revealed that the steady upward shift of intersection points due to scaling (and corresponding increase in hit rates) observed for HIV (Figure 3d) occurred frequently but not always. As illustrated in Figure 4, scaled SSPs displayed some significant compound-class specific differences. For example, the HIV profile might suggest that application of high scaling factors would generally be preferred, as also indicated in Table 1. A similar example was TKE (Figure 4) where scaling dramatically improved the results up to a scaling factor of 20. TKE also showed the most significant increase in hit rate under scaling conditions, from 1.8% to 7.4%. Profiles of ANG or ADR are representative of another group of activity classes with optimum search performance at high scaling factors. However, in these cases, most of the improvement in hit rates due to scaling was already achieved at much lower scaling factors and high factors only marginally improved the results. This group represented overall successful test cases because under moderate scaling conditions (e.g., a factor of 4 for ANG), all potential hits were always identified together with no more than ~200 dbc. Very high scaling factors essentially transform similarity search calculations into a fingerprint "identity check" at similarity threshold values approaching one, which represents an extreme scenario for similarity searching. However, the SSPs show that application of high scaling factors was often not required to achieve most of the improvements in hit rates. A similar example was COX where scaling significantly increased hit rates and where a scaling factor of 10 was already sufficient to recognize all potential hits. Thus, differences in average values between calculations at scaling factors of 10 or 50 become more or less "statistical" in such cases.

PKC, on the other hand, was an exemplary case where fingerprint scaling did not improve hit rates. Hit and dbc curves also shifted toward higher cutoff values, but their intersection points formed a flat line and, therefore, no relative enrichment in hits was observed. Profiles of GLU, PPAR, or CAE represent another distinctly different case. Here moderate scaling led to some improvement but all or most of the hits were never identified. An intermediate scaling factor (e.g., four for GLU) produced the best results, but search performance decreased when higher scaling factors were applied.

Thus, SSP analysis not only indicated that scaling influences similarity search calculations in different ways, dependent on the test cases, but also that search calculations on different activity classes can follow similar pathways under scaling conditions. These findings have implications for practical applications. For example, pilot calculations on known compound classes can be carried out in combination with SSP analysis to estimate how successful scaling calculations might be in the search for novel hits, given a source database and specific search tools. Detection of profiles such as those observed for HIV or TKE indicates that fingerprint scaling is a promising search strategy.

Scaling Performance and Related Methods. In addition to insights obtained from SSP analysis, it was encouraging to find that scaling detectably improved results for 19 of the 23 test cases that we studied. As mentioned above, due to the small number of potential hits and the large number of dbc, these calculations were quite challenging. However, on the basis of the results obtained herein and our previous analysis,[11] we can conclude that fingerprint profile scaling has significant predictive value. Are there similar approaches? Most closely related is the Stigmata algorithm[9] that originally made use of the 2048-bit version of Daylight fingerprints[22] to derive modal fingerprints characteristic for different compound classes. From Daylight fingerprints for a set of similar compounds, the modal fingerprint was generated by setting bits on that were consistently set on in a prespecified minimum number of single fingerprints (between 50% and 100%). At the 100% level, these bit positions would correspond to consensus bits in our formulation. However, different from the scaling approach, Stigmata produced a consensus fingerprint for each activity class. The similarity between a test compound and the modal or consensus fingerprint could then be compared using a modified Tc-like similarity metrics.[9] As mentioned before, scaling of consensus bits only indirectly affects fingerprint composition at the level of quantitative fingerprint comparison. Furthermore, Daylight fingerprints employed by Stigmata are hashed, whereas fingerprint profile scaling was exclusively designed for keyed fingerprints (where each bit position is firmly associated with a specific descriptor).

Since both Stigmata and fingerprint profile scaling essentially make use of relative frequencies of descriptor bit settings, another approach that should be mentioned in this context is the generation of so-called biological activity profiles,[23] which is related to the original implementation of substructure analysis.[24] In this case, the frequency of occurrence of substructures or incrementally encoded property descriptors (with binned value ranges) in two databases was compared and frequency differences were converted into weights for each substructure or descriptor value. When a database of synthetic compounds was used as a reference, differentially weighted descriptor bit strings (akin to fingerprints) could be produced as profiles, for example, for drugs in general or selected therapeutic compound classes.[23] Profile descriptors could then also be calculated for test molecules, and, using the sum of profile weights for their specific descriptor settings as a score, a ranking was obtained reflecting their similarity to the profile. Thus, although this

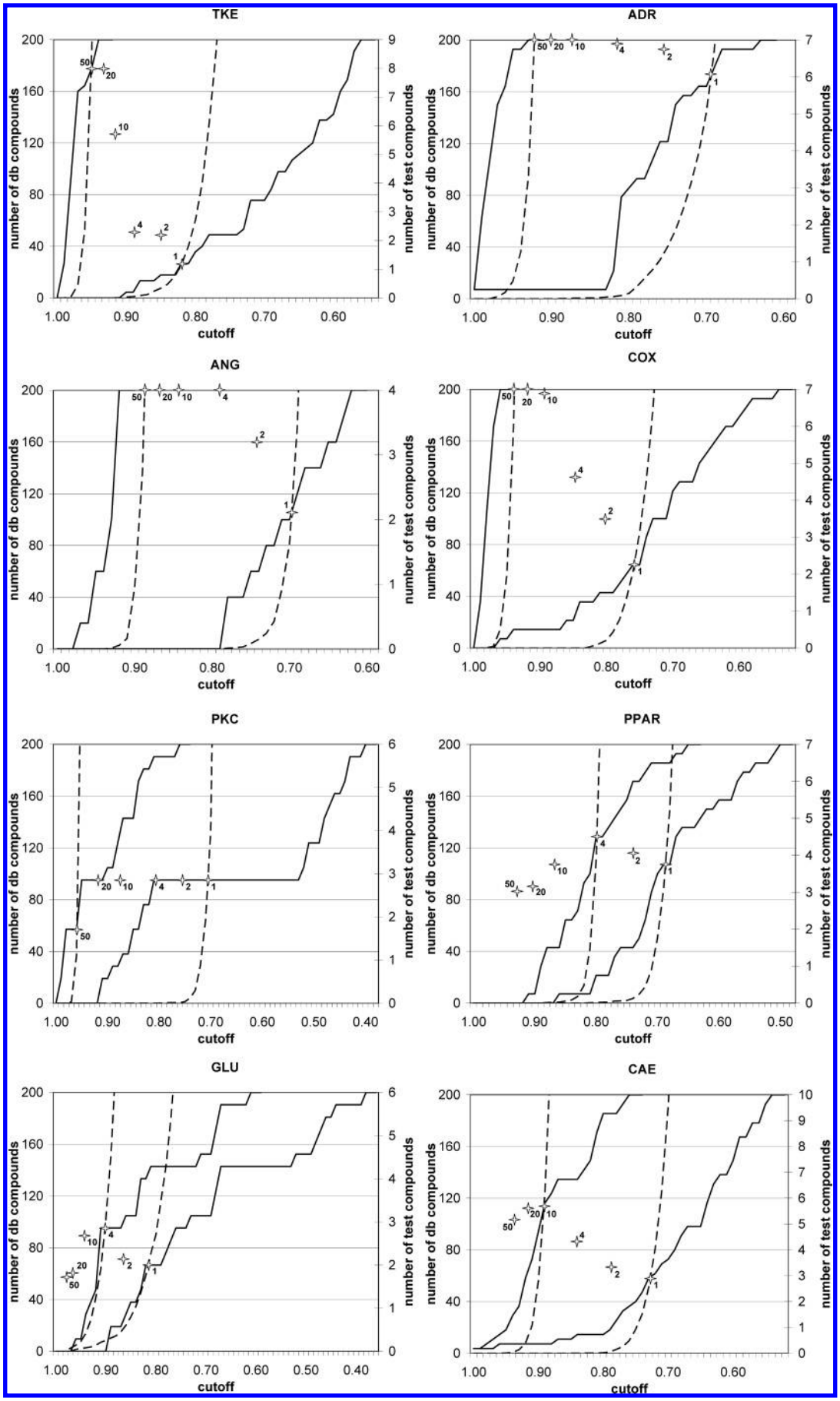

Figure 4. Scaled similarity search profiles. For different activity classes, scaled SSPs are shown in a simplified representation according to Figure 3d. Hit and dbc curves are shown for the nonscaled calculation and the scaling factor producing the best hit rate (except for PKC where scaling did not improve hit rates).

approach is not closely related to fingerprint scaling, it certainly contains a number of similar features for compound profiling.

CONCLUSIONS

Here we have applied a recently developed analysis tool termed similarity search profiling to better understand the influence of fingerprint scaling on similarity search calculations. Analysis of scaled SSPs has been the focal point of this study and has shown that scaling generally shifted hit and dbc curves toward higher optimum similarity threshold values, regardless of search performance. Importantly, it revealed that increases in search performance were often due to a more significant relative increase in hits than simultaneously recognized dbc. SSPs also showed that different search calculations followed similar pathways when scaling was applied, at least for MP-MFP. In this study, we have focused on evaluating the influence of scaling on different compound classes, but this type of analysis can also be applied to comparison of different fingerprints or evaluation of different compound libraries or databases available as sources for virtual screening.

REFERENCES AND NOTES

(1) Bajorath, J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 233−245.
(2) Merlot, C.; Domine, D.; Cleva, C.; Church, D. J. Chemical substructures in drug discovery. *Drug Discov. Today* **2003**, *8*, 594−602.
(3) Gillet, V. J.; Willett P.; Bradshaw J. Similarity searching using reduced graphs. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 338−345.
(4) Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, *35*, 2145−2154.
(5) Bajorath, J. Integration of virtual and high-throughput screening. *Nature Rev. Drug Discov.* **2002**, *1*, 882−894.
(6) Holliday, J. D.; Hu, C.-Y., Willett, P. Grouping of coefficients for the calculation of intermolecular similarity and dissimilarity using 2D fragment bit strings. *Combi. Chem. High Throughput Screen.* **2002**, *5*, 155−166.
(7) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(8) Sheridan, R. P.; Kearsley, S. K. Why do we need so many chemical similarity search methods? *Drug Discov. Today* **2002**, *7*, 903−911.
(9) Shemetulskis, N. E.; Weininger, D.; Blankley, C. J.; Yang, J. J.; Humblet, C. Stigmata: an algorithm to determine structural commonalities in diverse datasets. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 862−871.
(10) Xue, L.; Stahura, F. L.; Godden, J. W.; Bajorath, J. Fingerprint scaling increases the probability of identifying molecules with similar activity in virtual screening calculations. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 746−753.
(11) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Profile scaling increases the similarity search performance of molecular fingerprints containing numerical descriptors and structural keys. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1218−1225.
(12) Wang, R.; Wang; S. How does consensus scoring work for virtual library screening? An idealized computer experiments. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422−1426.
(13) Salim, N.; Holliday, J. D.; Willett, P. Combination of fingerprint-based similarity coefficients using data fusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435−442.
(14) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Similarity search profiles as a diagnostic tool for the analysis of virtual screening calculations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1275−1281.
(15) Edgar, S. J.; Holliday, J. D.; Willett, P. Effectiveness of retrieval in similarity searches of chemical databases: a review of performance measures. *J. Mol. Graph. Model.* **2000**, *18*, 343−357.
(16) Godden, J. W.; Furr, J. R.; Bajorath, J. Recursive median partitioning for virtual screening of large databases. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 182−188.
(17) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Design and evaluation of a molecular fingerprint involving the transformation of property descriptor values into a binary classification scheme. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1151−1157.
(18) SVL Exchange, URL: svl.chemcomp.com.
(19) Godden, J. W.; Stahura, F. L.; Xue, L.; Bajorath, J. Searching for molecules with similar biological activity: analysis by fingerprint profiling. *Pac. Symp. Biocomput.* **2000**, *5*, 566−575.
(20) Sanatvy, M.; Labute, P. SVL: the scientific vector language. *J. Chem. Comput. Group*, URL: www.chemcomp.com/feature/svl.htm.
(21) Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, Quebec, Canada, 2003, URL: www.chemcomp.com.
(22) James, C. A.; Weininger, D. Daylight theory manual. Daylight Chemical Information Systems, Inc., Irvine, CA, 1995, URL: www.daylight.com.
(23) Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165−179.
(24) Cramer, R. D.; Redl, G.; Berkoff, C. E. Substructural analysis. A novel approach to the problem of drug design. *J. Med. Chem.* **1974**, *17*, 533−535.

CI0400819