Neighborhood Behavior of in Silico Structural Spaces with Respect to In Vitro Activity Spaces–A Benchmark for Neighborhood Behavior Assessment of Different in Silico Similarity Metrics

Dragos Horvath, and Catherine Jeandenans

J. Chem. Inf. Comput. Sci., **2003**, 43 (2), 691-698• DOI: 10.1021/ci025635r • Publication Date (Web): 07 February 2003

Neighborhood Behavior of in Silico Structural Spaces with Respect to In Vitro Activity Spaces−A Benchmark for Neighborhood Behavior Assessment of Different in Silico Similarity Metrics

Dragos Horvath* and Catherine Jeandenans

Cerep, 128 rue Danton, 92506 Rueil-Malmaison, France

Received November 7, 2002

In a previous work, we have introduced Neighborhood Behavior (NB) criteria for calculated molecular similarity metrics, based on the analysis of in vitro activity spaces that simultaneously monitor the responses of a compound with respect to an entire panel of biologically relevant receptors and enzymes. Now, these novel NB criteria will be used as a benchmark for the comparison of different in silico molecular similarity metrics, addressing the following topics: (1) the relative performance of 2D vs 3D descriptors, (2) the importance of the similarity scoring function for a given descriptor set, and (3) binary or Fuzzy Pharmacophore Fingerprints—can they capture the similarity of the spatial distribution of pharmacophoric groups despite different molecular connectivity? It was found that fuzzy pharmacophore descriptors (FBPA) displayed an optimal NB and, unlike their binary counterparts, were successful in evidencing pharmacophore pattern similarity independently of topological similarity. Topological FBPA, identical to the former except for the use of topological instead of 3D atom pair distances, display a somehow weaker, but significant NB. Metrics based on "classical" global 2D and 3D molecular descriptors and a Dice scoring function also performed well. The choice of the similarity scoring function is therefore as important as the choice of the appropriate molecular descriptors.

INTRODUCTION

The computational evaluation of molecular similarity is nowadays widely used for library design and analysis. Molecules are represented by points of a Structural Space (SS) defined by molecular descriptors, in which the in silico metric (dissimilarity score) represents the pairwise distance between points. In silico focused library[1] design relies on the computational exploitation of the similarity hypothesis or Neighborhood Behavior (NB).[2]

In a previous work,[3] we have reported novel NB-assessment criteria, relying on the analysis of complete activity profiles. Obtained by means of robotized High Throughput Profiling[4] (HTP) experiments, these activity profiles can be seen as the basis of a multidimensional Activity Space (AS), monitoring on each axis the activity of compounds with respect to the associated biological test. In this context, neighborhood relationships in the activity space are characterized by activity dissimilarity metrics $\Lambda(m,M)$ monitoring how similar the activity profiles of compounds m and M are, unlike in previous works where the latter term only referred to activity differences with respect to a single biological target. Neighborhood Behavior can be accordingly understood as a relationship between the structural dissimilarity metric $\Sigma(m,M)$ of the calculated molecular descriptor space and the activity dissimilarity metric $\Lambda(m,M)$. To avoid confusion, we have reserved the term "similarity" to design structural similarity, while "activity-relatedness" has been chosen to design similarity of the biological activity profiles.

The present work uses the previously defined benchmark criteria in order to compare the NB of various in silico dissimilarity metrics, based on different categories of molecular descriptors. Previous attempts[5−7] to quantify NB were based on an empirical assignment of ligands into activity classes on the basis of their "main" activity with respect to the target they were designed for. Therefore, the comparison of the NB of various in silico similarity metrics in terms of the novel, activity profile-related criteria is expected to offer new insight with respect to topics such as the relative performance of 2D vs 3D descriptors[8] or the relative importance of connectivity vs pharmacophore similarity with respect to activity similarity. The descriptors and metrics have been chosen such as to sample the main types of structural spaces: 2D topological[9] and electrotopological[10] descriptors, 2D fingerprints,[11] 3D shape and polarity-weighed accessible surface indices,[12] and both fuzzy[13] two-point (FBPA) and bitwise three- and four-point 3D pharmacophore fingerprints.[14,15]

2. EXPERIMENTAL DATA—TARGET PANEL AND TESTED COMPOUNDS

The same panel of $N_{targets} = 42$ tests[16] (see Table 2 of the preceding paper[3]) defining the activity space, and the same set of $N_{mols} = 584$ commercially available drugs and reference ligands have been used here. The measured activity values $A_{m,t}$ represent the average percentages of inhibition, at a 10 μM concentration, of a compound m against every biological target t.

* Corresponding author phone: +33.1.55.94.84.49; fax: +33.1.55.94.84.10; e-mail: d.horvath@cerep.fr.

10.1021/ci025635r CCC: $25.00
Published on Web 02/07/2003

Table 1: In Silico Metrics Explored in This Work[a]

metric	description
Σ^1	A/V normalized FBPA metric with optimal parametrization
Σ^5	Dice metric with normalized FBPA
Σ^8	normalized TFBPA metric with parameters such as in Σ^1
Σ^9	Dice metric with normalized TFBPA
Σ^{10}	four-point pharmacophore fingerprint Dice metric
Σ^{11}	three-point pharmacophore fingerprint Dice metric
Σ^{12}	Dice metric in the "2D+3D" PC space—eq 2
Σ^{13}	Euclidean metric in the "2D+3D" PC space—eq 1
Σ^{14}	Dice metric in the "2D" PC space—eq 2
Σ^{15}	Euclidean metric in the "2D" PC space—eq 1

[a] Note: metric numbering is not contiguous, in order to avoid confusion with the previous paper[3] of this series in which mentions other scoring schemes (#2,3,4,6,7), that will no longer be discussed here. Metric numbering is therefore consistent throughout the two papers.

3. SPACES AND METRICS—DEFINITIONS AND DISCUSSION

3.1. Structural Spaces. Molecules are represented as points of a Structural Space (SS), where the distances between a pair of such points (*m*,*M*) is given by the in silico dissimilarity score $\Sigma(M,m)$. Several structural spaces have been explored in this work, and for each space one or more similarity metrics have been evaluated—an exhaustive list of which can be found in Table 1).

•The "2D" and "2D+3D" structural spaces were defined on the basis of "classical" molecular descriptors available in the Cerius[2] (MSI)[17] QSAR module. The 2D space includes 42 topological (Balaban,[9] Kier and Hall,[18] Wiener,[19] Information Content, ...) and structural (Molecular Weight, hydrogen bond donor/acceptor counts, rotatable bonds, ...) descriptors, reduced—using principal component analysis (PCA)[20] after average/variance normalization—to six orthogonal components that explained 90% of the initial variance. The 2D+3D space includes, in addition to the previous, the electrotopological state keys[10] as well as various 3D terms such as the Jurs[12] polar surface areas, shadow indices, principal moments of inertia, and others. This gave a total of 142 descriptors, reduced to 33 principal components at 90% of explained variance. Both the Euclidean distance—eq 1—and the Dice[21] correlation coefficient—eq 2—have been considered as similarity metrics of these PC spaces, in which the $i = 1..N_{Pc}$ components of the principal component position vector of a molecule *M* will be denoted as PC_M^i

$$\Sigma(m,M) = \sqrt{\sum_{i=1}^{N_{pc}} (PC_M^i - PC_m^i)^2} \tag{1}$$

$$\Sigma(m,M) = 1 - \frac{2\sum_{i=1}^{N_{pc}} PC_M^i PC_m^i}{\sum_{i=1}^{N_{pc}} PC_M^i PC_M^i + \sum_{i=1}^{Npc} PC_m^i PC_m^i} \tag{2}$$

•The Fuzzy Bipolar Pharmacophore Autocorrelogram[13] (FBPA) space has been thoroughly described elsewhere.[3,13]

•Topological Fuzzy Bipolar Pharmacophore Autocorrelograms (TFBPA) are identical in all respects with the FBPA, except for the use of *topological* (e.g. minimal numbers of bonds separating two atoms) instead of 3D distances. The two metrics using FBPA descriptors have been equally employed with TFBPA, as noted in Table 1.

•Bitwise three-point[14] (3PF) and four-point[15] (4PF) pharmacophore fingerprint spaces are characterized by a huge number of dimensions representing the many possible configurations—each corresponding to a bit in the fingerprint—that can be formed in 3D-space with three or four pharmacophoric centers of any considered type, at all the considered combinations of distance ranges separating them. They have been built starting from a Catalyst database with up to 100 conformers per molecule (compared to a maximum of 20 used with the FBPA). Dice[21] metrics have been associated to these spaces.

An overview of the characteristics of the metrics used in this work is given in Table 1.

3.2. Activity Space. The considered Activity Space (AS) and its associated dissimilarity scores, generically denoted by $\Lambda(M,m)$, have been introduced previously[3] and will be briefly revisited here:

•The $\Lambda^{(1)}$ metric (3) is related to the number of targets with respect to which two compounds *M* and *m* display significant potency differences, defined in terms of empirical high (H%) and low (L%) values of inhibition percentage differences:

$$\Lambda^{(1)}(m,M) = \sum_{t=1}^{N_{targets}} \lambda_t(m,M) \quad \text{where } \lambda_t(m,M) = \begin{cases} 1 & \text{if } |A_{M,t} - A_{m,t}| \geq H \\ 0 & \text{if } |A_{M,t} - A_{m,t}| \leq L \\ \dfrac{|A_{M,t} - A_{m,t}| - L\%}{H\% - L\%} & \text{otherwise} \end{cases} \tag{3}$$

•The $\Lambda^{(2)}$ metric (4) relies on the same principle but includes correction factors ω_t accounting for the interrelatedness[3] of target *t* and other targets from the activity panel:

$$\Lambda^{(2)}(m,M) = \sum_{t=1}^{N_{targets}} \omega_t \lambda_t(m,M) \tag{4}$$

In addition to these activity metrics, this paper also considers, for purposes of comparison, the standard and respectively target intercorrelation-corrected Euclidean distances (5) and (6) as potential activity similarity metrics.

$$\Lambda^{(3)}(M,m) = \sqrt{\sum_{t=1}^{N_{targets}} (A_{M,t} - A_{m,t})^2} \tag{5}$$

$$\Lambda^{(4)}(M,m) = \sqrt{\omega_t \sum_{t=1}^{N_{targets}} (A_{M,t} - A_{m,t})^2} \tag{6}$$

3.3. Which Is the Best Activity Metric? The validation of in vitro similarity metrics is not a trivial problem, since a NB study is typically considered to be a test for the descriptors and their associated in silico metrics, e.g. the mapping of neighboring points of SS onto remote points from AS counts as a failure of the in silico metric. However, the activity metrics, although based on experimental data, are

empirical expressions such as any other similarity score. In this context, a NB study becomes a bootstrapping approach, simultaneously validating both in silico and in vitro metrics against each other. If one of the activity metrics is *systematically* found to maximize NB criteria, irrespectively of the employed in silico counterparts, that definition of activity neighborhood may arguably have some intrinsic advantages.

4. NEIGHBORHOOD BEHAVIOR OF IN SILICO METRICS WITH RESPECT TO ACTIVITY METRICS

4.1. How Similar is Similar? Translating the Significance of Structural Similarity Scores in Terms of Expected Activity Profile Differences. What dissimilarity level s can still be tolerated in order to allow a confident extrapolation of the properties of one partner of the pair to the other? This can be answered by monitoring $\langle\Lambda\rangle_{P(\Sigma\approx s)}$, the expectation value of observed activity profile difference within subsets of compound pairs at given level of dissimilarity s.

$$\langle\Lambda\rangle_{P(\Sigma\approx s)} = 1/N_{P(\Sigma\approx s)} \sum_{\Sigma(m,M)\approx s} \Sigma(m,M) \qquad (7)$$

In eq 7 the summation is carried out over all the $N_{P(\Sigma\approx s)}$ pairs of structural dissimilarity falling within a small range around s, of width $\Delta s = 0.02 \times s_{60\%}$, where $s_{60\%}$ is an upper dissimilarity threshold chosen such that 60% of the compound pairs have $\Sigma(m,M) < s_{60\%}$ (typically, $\Delta s \approx 0.01...0.02$ for Dice-type metrics). The low end of the structural dissimilarity range, within which the profile dissimilarity averages remain conveniently low, represents the "similarity radius" or "validity range" of the metric.

4.2. The NB Quality Criteria. The reader is referred to the previous paper in this series[3] for a detailed discussion of the complementary NB quality scores—the consistency criterion $\chi(s)$ and respectively the optimality score $\Omega(s)$ at given dissimilarity threshold s as well as for the introduction to the "$\Omega-\chi$" plots that will be further one used for Neighborhood Behavior benchmarking.

5. RESULTS AND DISCUSSION

5.1. How Similar is Similar? The plots in Figure 1 evidence the nonlinear relationship between structural dissimilarity scores and the expected activity profile differences $\Lambda^{(1)}(m,M)$ of compound pairs at given structural dissimilarity $\Sigma(m,M) \approx s$. All metrics are *consistent* at the lower end of their value range—the least performant being Σ^{14}, a metric according to which many False Similar pairs score dissimilarity values close to 0. The average degree of activity-relatedness of pairs located at the upper ends of the dissimilarity ranges reaches (or even exceeds) the average degree of expected dissimilarity over the entire set ($\langle\Lambda^{(1)}\rangle_P = 4.6$). The mapping of dissimilarity scores with respect to the activity-relatedness criterion widely differ from metric to metric. Sets of pairs that are 90% dissimilar ($s = 0.9$) according to Σ^{10} (based on four-point bitwise fingerprints) are roughly equally consistent as the 30% dissimilar ($s = 0.6$) according to Σ^{1} (fuzzy two-point FBPA). Very few compound pairs—most of which are obviously similar "me too" analogues with slightly different substitution patterns—score a dissimilarity as low as 0.3 according to Σ^{10}. To match

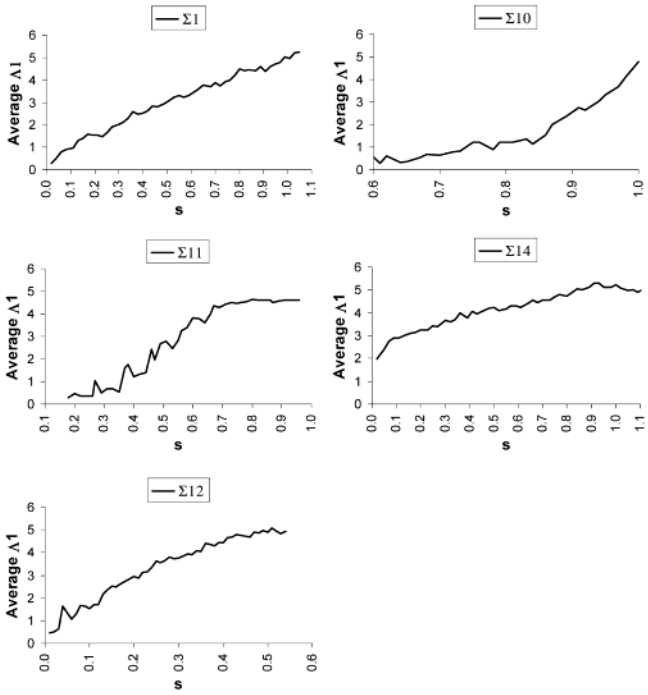

Figure 1. Expected average activity dissimilarity score $\langle\Lambda^{(1)}\rangle_{P(\Sigma\approx s)}$ taken over subsets of compound pairs scoring, according to various metrics, a dissimilarity value Σ roughly equal to the running variable s of the X axis.

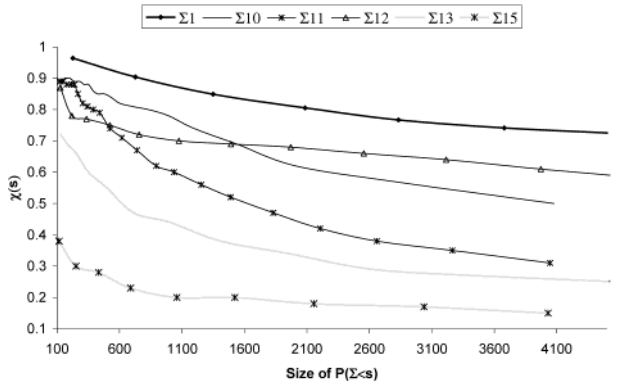

Figure 2. Consistency within the 4500 pairs ranked as the most similar by various in silico metrics, with respect to the $\Lambda^{(1)}$ activity-relatedness score.

the pattern attached to a given bit, six geometric degrees of freedom *simultaneously* need to match specific values. Therefore, similar compounds in most of the cases fail to highlight *exactly* the same bits in the fingerprint. They may instead highlight bits that are *related* (standing for quite similar four-point configurations), some information ignored by the bitwise metric.

5.2. Pair Ranking: The Relative Similarity Scale. The plots of $\chi(s)$ against $N_{P(\Sigma<s)}$ in Figure 2 illustrate the duality of the NB concept, in terms of two complementary aspects of (a) consistency—"Which metric offers the best guarantees that the few best-ranked pairs will be activity-related?" and (b) completeness—"Which metric is able to pick a maximum of activity-related pairs at given consistency?". The latter is illustrated by the *rate* of consistency loss with respect to the selected set size.

Most metrics present excellent guarantees of enrichment in activity-related pairs among the best-ranked 100...500 pairs, the worst performers with this respect being the 2D

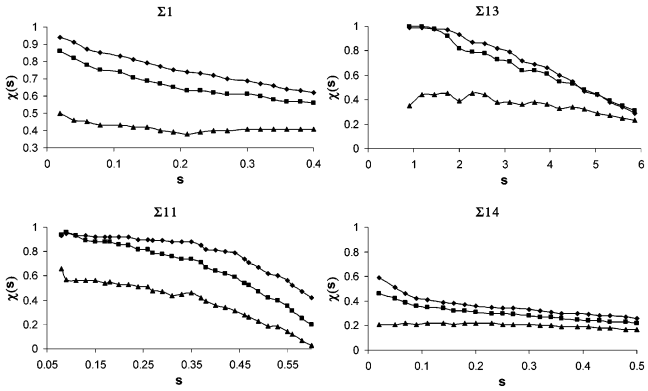

Figure 3. Comparison of consistency scores of structural metrics against several activity-relatedness metrics. The different rendered curves correspond to $\Lambda^{(1)}$ − with rhombs, $\Lambda^{(2)}$ − squares, and $\Lambda^{(4)}$ − triangles.

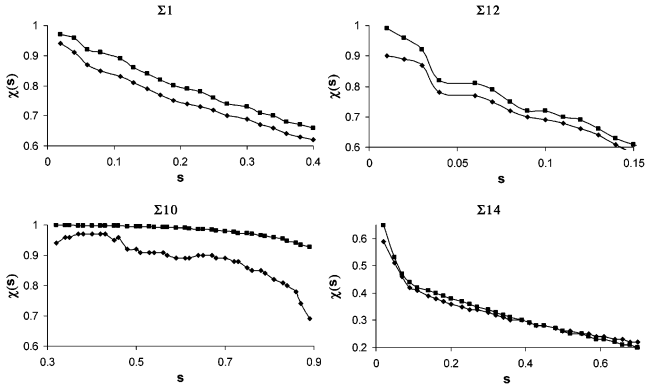

Figure 4. Consistency criteria χ(s) vs dissimilarity cutoffs *s* scored by various metrics with respect to the $\Lambda^{(1)}$ activity dissimilarity criteria in the (a) default activity space defined by the entire test panel, shown with rhombic dot marks and respectively the (b) subspace including only the GPCR binding assays, plotted with squares.

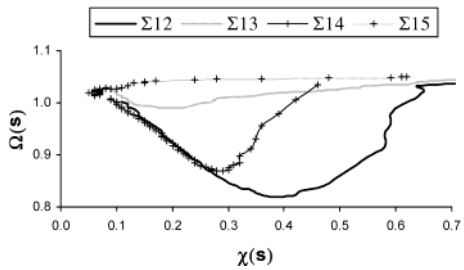

Figure 5. $\Omega-\chi$ plot of Euclidean and Dice metrics of the 2D and 2D+3D principal component descriptor spaces, featuring $\Lambda^{(2)}$ as activity space metric.

and 2D+3D descriptor-based Euclidean metrics Σ^{15} and Σ^{13} as well as the Dice metric in the 2D descriptor space Σ^{14}. Furthermore, the number of activity-related compound pairs from the current data set in which the activity-relatedness may actually be due to some underlying structural similarity may be estimated at a minimum of 5000. These are compound pairs that are perceived as structurally similar by at least one of the considered metrics (Σ^1), successfully selecting all of them without any significant loss of consistency. Σ^1 is at this point the most complete metric of the study, setting a challenge for other metrics to evidence "hidden" structural similarities among other pairs of compounds in addition to the 5000 ones picked by Σ^1, without co-opting any False Similars.

5.3. Relative Quality of the Activity Metrics. As discussed in paragraph 3.3, a consensual preference of various in silico metrics for a given activity metric, expressed for example in terms of increased consistency factors, may underscore the appropriateness of that activity metric as a monitoring tool of activity likeness. *All* structural dissimilarity scores (representatives of each type being shown in Figure 3) were found to display highest consistency values with respect to the $\Lambda^{(1)}$ metric, followed by $\Lambda^{(2)}$ and $\Lambda^{(4)}$. This is expected, the drawback of Euclidean metrics in terms of inhibition percentages being that many small differences due to experimental noise may spuriously sum up to yield dissimilarity scores of magnitude comparable to the ones of less noisy profiles, showing in exchange some significant activity differences against a few targets. Although the Euclidean metric is the most straightforward distance function—and might, furthermore, be used in connection with Principal Component Analysis (PCA) to eliminate intertarget correlation artifacts in a more elegant way than it has been done in $\Lambda^{(2)}$ and $\Lambda^{(4)}$—these results illustrate that inhibition percentage noise "suppression" mechanisms such as the ones used in $\Lambda^{(1)}$ and $\Lambda^{(2)}$ are quite important in the herein considered activity space.

The differences between $\Lambda^{(1)}$ and $\Lambda^{(2)}$ are due of the downscaling of the importance of *related* tests in the panel. As shown in the previous work, strong correlations are practically confined within the family of G-Protein Coupled Receptors (GPCR). Or, Figure 4 shows that the consistency of pharmacophore metrics with respect to a GPCR-only activity subspace (Alpha1, Alpha2, Beta1h, D1h, D2h, DaUpt, H1c, M1h, M3h, Muh, 5HT1Ah, 5HT1D, 5HT2ch, 5HT3h, 5HT6h, 5HTUpt, Sigma1) is *better* than the one relative the entire activity space. This is expectable, since the characteristic pharmacophoric signature of bioactive amines (aromatic ring − positive charge) is easily detected by pharmacophore metrics and the atom-type sensitive 2D+3D descriptors. A purely topological metric such as Σ^{14} does however not behave better with respect to the GPCR receptors. Therefore, the better performance of $\Lambda^{(1)}$ most likely stems from the repeated (and unjustified) counting of its high success rates against GPCR targets, while $\Lambda^{(2)}$ is the more realistic metric of activity space. Unless stated otherwise, this latter activity metric has been used to derive all of the following results.

5.4. Optimality Analysis of in Silico Metrics. Like in the previous work,[3] the NB optimality of in silico metrics will be discussed on hand of plots of the overall optimality criterion Ω(s) against the consistency criterion χ(s), where $\Omega = f(\chi)$ displays a local minimum, of depth 1-Ω(s*) and situated at χ(s*). An in silico metric showing *lower* Ω(s*) values at *higher* χ(s*) scores is unambiguously the one with a better NB. This insight will be used to compare the relative performance of the different metrics while trying to highlight the impact of specific aspects of the considered SS on the Neighborhood Behavior.

5.4.1. Euclidean or Dice? The comparison (Figure 5) of the performances of Euclidean (Σ^{15},Σ^{13}) vs Dice (Σ^{12},Σ^{14}) metrics in the 2D and 2D+3D PC descriptor spaces unambiguously prove the superiority of latter. Euclidean metrics hardly do better than random in retrieving similar pairs with similar activities. Accordingly, it can be concluded that the angular component of the Dice metric, showing how

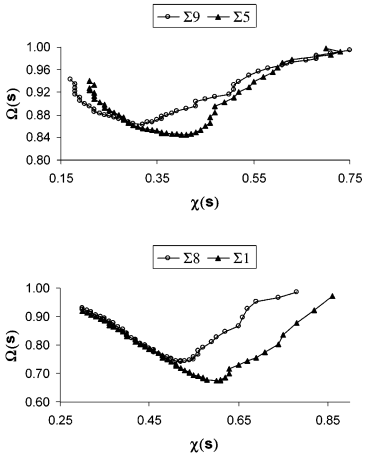

Figure 6. Impact of the substitution of geometric 3D distances (FBPA) by topological 2D distances (TFBPA) on the NB of bipolar pharmacophore metrics, all the other parameters being equal (Standard Dice metrics—Σ^5 vs Σ^9—in the upper graph, and metrics using fitted parameters—Σ^1 vs Σ^8—in the lower graph).

divergent two molecular vectors are, is essential for the definition of biologically relevant neighborhoods in structure space. The classical intuitive image of activity-related "globular clusters" in the structural "Universe" should therefore be more accurately described as "activity cones" (although the Dice score, unlike the cosine correlation coefficient,[21] also introduces a penalty related to the differences in vector norms). Unsurprisingly, the metric based on the information-richer 2D+3D descriptor set is the better one.

5.4.2. How Important Is the Knowledge of the Three-Dimensional Structure for Similarity Evaluation? The three-dimensional shape of molecules is of paramount importance for ligand recognition but not necessarily so for similarity scoring. Indeed, besides their computational expense, the sometimes important variance of 3D descriptors function of the considered conformation may introduce a noise level overriding the benefits of considering the 3D aspects. To minimize such artifacts, the FBPA[13] were specifically designed as averages over an ensemble of *conformational* fingerprints—e.g. they capture well the distance relationships that recur in many sampled conformers, in detriment of those that are specific to a given conformer.

The debate concerning the superiority of 2D vs 3D descriptors in QSAR and similarity evaluation is ongoing[22] and seemingly heading for the consensus conclusions that 3D descriptors are not outperforming 2D or 1D[23] approaches. Previous authors have however sometimes compared widely different classes of 2D and 3D descriptors, making it unclear whether the observed differences in performance were actually the consequence of the difference in dimensionality. Figure 6 compares the NB of two metrics that are identical in all respects, except for the (3D and respectively topological) nature of the distances used to build the fingerprints. With equal weights for pharmacophoric features (upper plot), the FBPA (Σ^5) and the TFBPA (Σ^9) appear to perform equally well in terms of overall optimality at consistency levels above $\chi > 0.6$. At intermediate consistency levels, the FBPA metric is however more performant. The comparison shifts in favor of the FBPA metric when optimal pharmacophore feature weights are used (lower plot). An attempt to specifically optimize the NB of the TFBPA-based metric (results not shown) failed to outperform the quality of Σ^1. Also, FBPA generation is extremely rapid[13] for combinatorial libraries and may therefore be a method of choice to characterize such compound collections, while noncombinatorial compounds may be more effectively described using 2D methodologies.

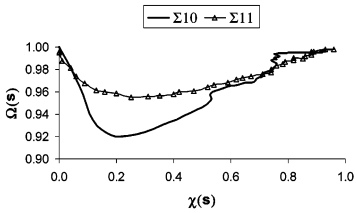

Figure 7. Neighborhood Behavior of three- vs four-point pharmacophore fingerprint metrics.

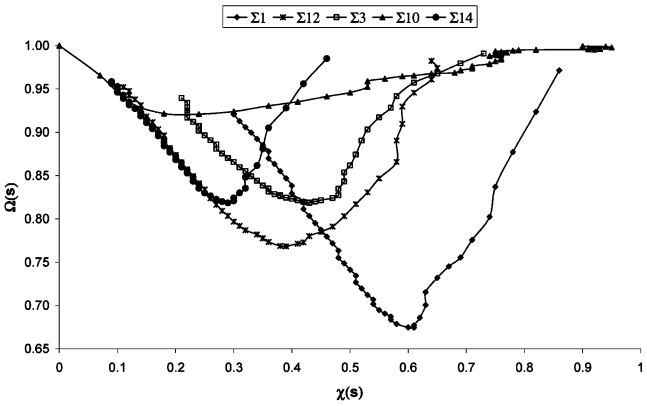

Figure 8. $\Omega-\chi$ plots comparing the best structural metrics representing each of the considered structural spaces.

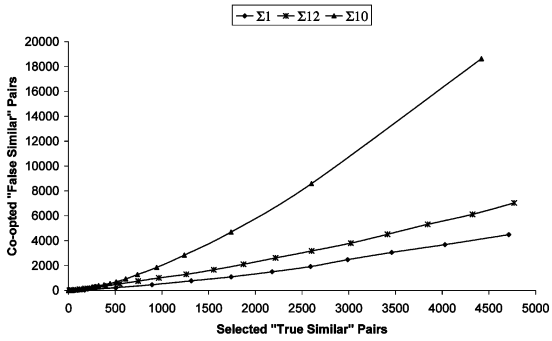

Figure 9. Numbers of co-opted "False Similar" N_{FS} pairs (on Y) vs the numbers of "True Similar" pairs N_{TS} (on X) found within an increasing subset of top-ranking structurally similar pairs.

5.4.3. Three- and Four-Point Pharmacophore Fingerprints. Three- (Σ^{11}) and four-point (Σ^{10}) pharmacophore metrics score excellent consistency with respect to the small subsets of <500 compound pairs ranked as most similar (Figure 2). However, a common characteristic of these two metrics is the relative paucity of compound pairs classified within their similarity radii—although these radii (0.8..0.9) almost cover their entire defined value range [0,1] (see paragraph 5.1 and Figure 1). Due to low completeness, overall optimality scores are deceivingly low, as shown in Figure 7.

5.4.4. Comparative Analysis of the NB of Different Categories of Molecular Descriptors. Figure 8 shows $\Omega-\chi$ plots of in silico metrics based on different categories of descriptors. Figure 9 offers some additional information with respect to some of the metrics in Figure 8, displaying the

numbers of false similars co-opted into the smallest subset of most similar pairs that includes the given number of true similar pairs N_{TS} (on the *X* axis).

Even in absence of weighing factor calibration, the FBPA-based metric (Σ^5) shows much better overall optimality than the ones based on bitwise four-point pharmacophore fingerprints, a trend that is further emphasized by weighing factor calibration. The 2D+3D Principal Component Space in conjunction with a Dice metric Σ^{12} also displays strong neighborhood behavior, outperforming Σ^5.

5.4.5. Nontrivial Similarity: Pharmacophorically Similar Analogues with Different Topology. The concept of molecular similarity, rooted in the medicinal chemist's know-how to recognize compounds with related biological properties, heavily relies on connectivity similarity apparent from the 2D structure sketches. Scaffold-centric (Markush)[24] definitions of compound families are a direct consequence of this fact. By contrast, the similarity of geometry-dependent pharmacophoric patterns is much less obvious to a human observer. Therefore, in silico metrics able to specifically recognize nontrivial near neighbors with different connectivity,[25,26] that would not have been perceived as similar by a chemist, are of special interest—provided that pharmacophore similarity *alone* is a reasonably good guarantee of activity similarity. Though it is the spatial distribution of functional groups that drives the molecular recognition phenomena, connectivity-related factors such as flexibility and electronic density effects play a nonnegligible role in binding.

To evaluate the specific ability of pharmacophore descriptor metrics to detect nontrivial pharmacophore similarity backed by an activity similarity, a NB analysis of the FBPA-based Σ^1 and four-point pharmacophore fingerprint-based Σ^{10} metrics has been run against subsets of *topologically dissimilar* pairs with Σ^{14} scores larger than 0.1 (165 961 pairs) and 0.25 (153 404 pairs), respectively, out of the total of 170 236. Figure 10 witnesses the expected general decrease of NB quality of both metrics, after the elimination of the pairs of "me too" (topologically and pharmacophorically similar) analogues. This only moderately affects the NB of the FBPA-based Σ^1 metric—the exclusion of the first 4275 topologically most related pairs having no sizable effect. However, fluctuations appear in the high consistency zone of the $\Omega-\chi$ plot of the Σ^{10} metric, signaling that the numbers of selected pairs within that consistency range suffered a significant backlash. According to Figure 11, the bitwise fingerprint-driven selection corresponding to a consistency level of 0.5 includes about 400 examples of true similars, out of which all but 5(!!) also belong to the subset of 4275 *topologically* most similar pairs. This proves that Σ^{10} achieved its excellent consistency scores by specifically ranking pairs of "me too" compounds at the top of its nearest neighbors list. It fails to evidence pharmacophore similarity not backed by topological similarity—probably due to the low degree of exact bit string matches (see discussion in paragraph 5.1). It might be concluded that the upper limit of 100 conformations/molecule used to build the Catalyst database is largely insufficient to ensure the continuous conformational space coverage required in order to render four-point pharmacophores independent to sampling artifacts. These descriptors were originally designed to work with a much denser conformational space coverage[15]—however, it has been reported that three-point fingerprints performed

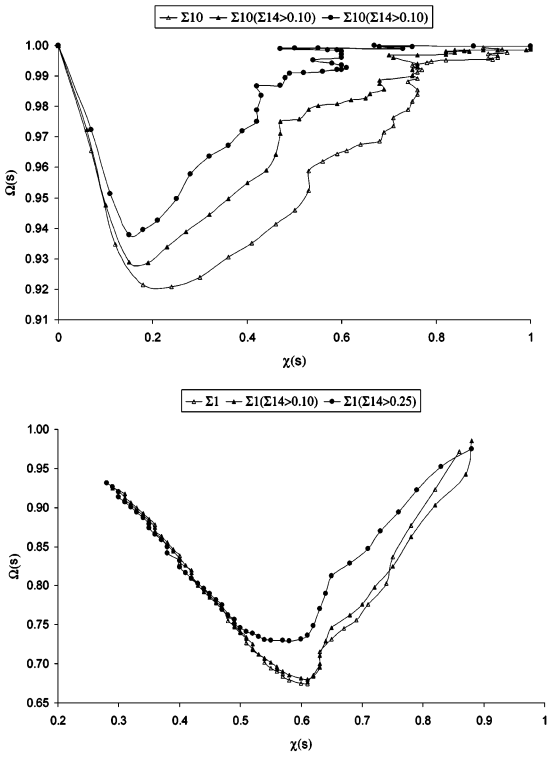

Figure 10. Changes in the NB of the Σ^{10} and Σ^1 metrics upon removal of atom pairs of high *connectivity* similarity ($\Sigma^{14} > 0.10$ and 0.25, respectively) from the set.

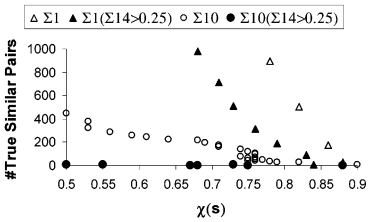

Figure 11. Number of "True Similar" pairs (on Y) found within an expanding the subset of top-ranked structurally similar pairs, plotted against the consistency criterion of the subset (on X), before and after elimination of the topologically most similar compound pairs ($\Sigma^{14} > 0.10$ and 0.25, respectively).

better when built on the basis of less conformers.[7] Our own results (not shown) tend to validate this counterintuitive finding, for three-point pharmacophore fingerprints built on hand of only 20 conformers slightly outperformed the NB of the 100-conformer metric used throughout this paper.

The fuzzy character of the FBPA ensures a better ability to detect examples of nontrivial molecular similarity. Figure 12 exemplifies three such cases of compound pairs with no obvious relationships at the molecular connectivity level, which all score Σ^1 scores well within the validity range of that metric (all lower than 0.16), and also show a significant degree of activity similarity ($\Lambda < 1$, while sharing a common pool of targets to which they bind). These pairs are "missed" by the bitwise pharmacophore metrics (with four-point fingerprints, all score worse than the similarity radius of 0.95).

6. CONCLUSIONS

The application of the activity profile-based NB assessment criteria to different types of molecular similarity metrics leads to a series of interesting insights, some in agreement

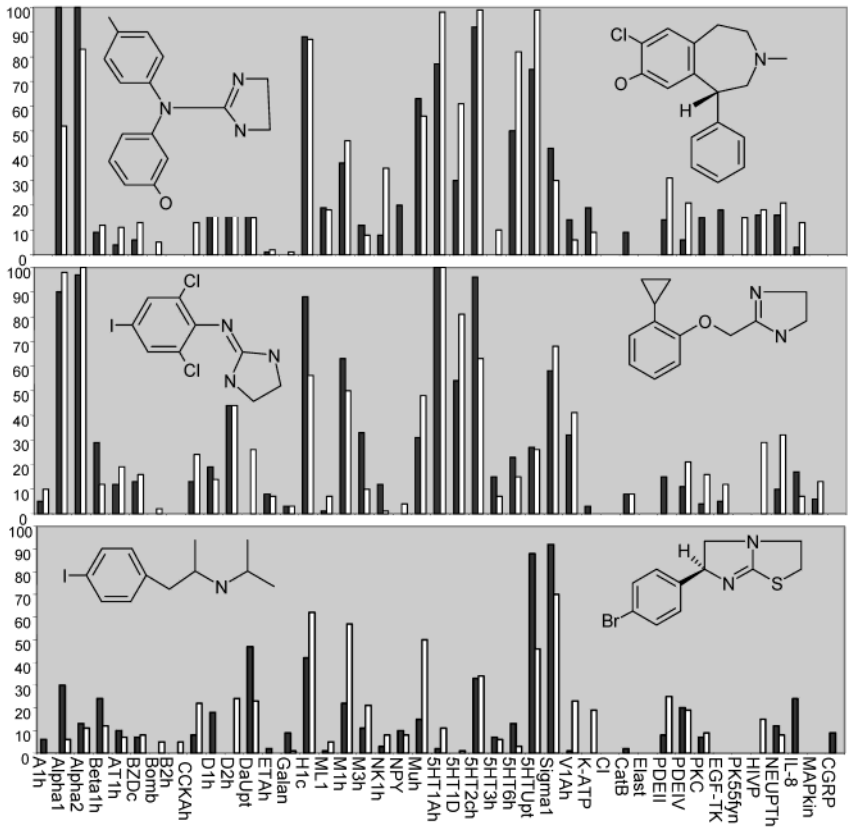

Figure 12. Examples of topologically different but *pharmacophorically similar* pairs of compounds with similar activity profiles. These pairs are perceived as similar by the Σ^1 metric, but *not* the Σ^{10} metric.

and others in contradiction to the observations reported in the literature, on hand of empirical biological activity classifications of molecules.

Notably, our work showed that 3D information, captured in terms of average interatomic distances over a set of up to 20 diverse conformations (FBPA), improved the NB of this metric with respect to its 2D counterpart (Topological FBPA). This is interesting, for literature[27] typically reported comparisons between completely unrelated 2D and 3D descriptors—the observed differences being therefore not strictly related to descriptor dimensionality. However, the net increase in NB with 3D fingerprints may not justify the supplementary effort required to generate 3D conformers of noncombinatorial compounds.

The very good NB of FBPA-based metrics is dependent on the parametrization scheme of pharmacophore feature weights. Unless the latter are optimized, the Dice metric in the 2D+3D principal component space performs better. Interestingly, Euclidean metrics preferentially used in PC spaces perform significantly worse than correlation coefficient-based measures of molecular similarity. Therefore, the functional form of the similarity metric is as important as the choice of molecular descriptors. Purely topological similarity metrics ignoring the atom types obviously fail to display any significant NB.

Three- and four-point pharmacophore fingerprints supposed to capture the 3D distribution of pharmacophore features have been shown to score an excellent consistency, however, limited to the few best-ranked pairs of compounds that happened to be also topologically similar. Compound pairs with significantly different chemotypes are extremely unlikely to set *exactly* the same bits in the pharmacophore string, even if their overall pharmacophore group distribution is fairly similar. In the latter case, compounds would highlight different bits corresponding to *related* triplets/quadruplets, but these "near misses" will not count in the similarity evaluation. These descriptors can be successfully used for pharmacophore queries—retrieving all molecules with a given pattern of set bits from a database—when the issue of undetected compounds is unimportant as far as all of the retrieved ones display the wanted features.

ACKNOWLEDGMENT

This work has been part of a proof-of-concept study assessing the importance of HTP data for in silico model validation, nowadays continued within the framework of Cerep's BioPrint chemoinformatics platform. The biological tests were set up and carried out by large teams of biologists and chemists in Poitiers, France (Valérie Hammond et al.) and Redmond, Washington (Marie-Christine Bodinier, Cheryl Wu, Christian Eckhof, Margaret Davis, and others). Many thanks to the BioPrint team 1999−2000 (Jacques Migeon, John Cargill, Xianqun Wang, Dave Porubek − Redmond) and to Frederic Revah, Rueil-Malmaison, France, for suggestions and helpful discussions.

REFERENCES AND NOTES

(1) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Chapman & Hall: London, 1995.

(2) 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.

(3) Horvath, D.; Jeandenans, C. Neighborhood Behavior of In Silico Structural Spaces with Respect to in Vitro Activity Spaces—A Novel Understanding of the Molecular Similarity Principle in the Context of Multiple Receptor Binding Profiles. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 680−690.

(4) Jean, T.; Chapelain, B. Method of identification of leads or active compounds, European Patent EP0905512, **1999**.

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

(6) 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.

(7) Matter, H.; Pötter, T. Comparing 3D Pharmacophore Triplets and 2D Fingerprints for Selecting Diverse Compound Subsets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1211−1225.

(8) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH Verlag GmbH: Weinheim, 2000.

(9) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.

(10) Hall, L. H.; Kier, L. B. Electrotopological state indices for atom types: a novel combination of electronic, topologic and valence state information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.

(11) The Daylight Theory Manual, http://www.daylight.com.

(12) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 726−735.

(13) Horvath, D. High Throughput Conformational Sampling & Fuzzy Similarity Metrics: A Novel Approach to Similarity Searching and Focused Combinatorial Library Design and its Role in the Drug Discovery Laboratory. In *Combinatorial Library Design and Evaluation: Principles, Software Tools and Applications*; Ghose, A., Viswanadhan, V., Eds.; Marcel Dekker: New York, 2001; pp 429−472.

(14) Pickett, S. D.; Mason, J. S.; McLay, I. M. Diversity Profiling and Design Using 3D Pharmacophores: Pharmacophore-Derived Queries. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 1214−1223.

(15) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem*. **1998**, *38*, 144−150.

(16) Cerep Online Catalog, http://www.cerep.fr/Cerep/Utilisateur/CatalogOnLine/frs_index_cat.asp.

(17) Cerius^2 v 4.0, Molecular Simulations Inc., San Diego, CA.

(18) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure−Property Modeling. In *Reviews in Computational Chemistry II*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; pp 367−422.

(19) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Chem. Phys.* **1947**, *69*, 17−20.

(20) Glen, W. G.; Dunn, W. J.; Scott, D. R. Principal Components Analysis and Partial Least Squares Regressions. *Tetrahedron Comput. Technol.* **1989**, *2*, 349−376.

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

(22) 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*. **1996**, *36*, 572−584.

(23) Dixon, S. L.; Merz, K. M. One-Dimensional Molecular Representation and Similarity Calculations: Similarity and Validation. *J. Med. Chem*. **2001**, Web publication 10.1021/jm010137f.

(24) Brown, R. D.; Downs, G. M.; Barnard, J. M. Use of Markush structureanalyzis techniques for rapid processing of large combinatorial libraries, Conference at the 218th National ACS Meeting, New Orleans, Louisiana, Aug. 22−26, 1999.

(25) Poulain, R.; Horvath, D.; Bonnet, B.; Eckoff, C.; Chapelain, B.; Bodinier, M.-C.; Déprez, B. From Hit to Lead. Combining Two Complementary Methods for Focused Library Design Application to μ Opiate Ligands. *J. Med. Chem.* **2001**, *41*, 3378−3390.

(26) Poulain, R.; Horvath, D.; Bonnet, B.; Eckoff, C.; Chapelain, B.; Bodinier, M.-C.; Déprez, B. From Hit to Lead. Analyzing Structure-Profile Relationships. *J. Med. Chem.* **2001**, *41*, 3391−3401.

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

CI025635R