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

Dragos Horvath* and Catherine Jeandenans

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

Received November 7, 2002

As a consequence of recent advances in the field of High Throughput Screening, the systematic testing ("in vitro profiling") of compounds against a panel of targets covering different therapeutic areas is nowadays used to generate relevant information with respect to the in vivo behavior of drug candidates. However, the development of chemoinformatics tools required for the exploitation of such data is yet in an incipient phase. In this paper, a formalism for the analysis of activity profile vectors (describing the experimental responses of compounds in each of the considered activity tests) is introduced and applied at the study of Neighborhood Behavior (NB; the hypothesis that structurally similar compounds display similar biological properties) of molecular similarity metrics. The experimental activity profiles define an Activity Space in which more than 500 drugs and reference compounds are positioned, their coordinates being inhibitory propensities in the included tests and unambiguously characterizing a molecule in terms of its receptor binding properties. While previous studies of Neighborhood Behavior had to rely on a loose classification of compounds in terms of the therapeutic areas they were designed for, here the NB of a calculated "in silico" similarity metric has been redefined as a relationships between intermolecular dissimilarity scores in the "structural" and "activity" spaces, respectively, and expressed in terms of two quantitative criteria: "consistency" (the propensity of the metric to selectively rank activity-related compound pairs among the structurally most similar pairs) and "completeness" (monitoring the retrieval rate of activity-related compound pairs among the best ranked pairs of structural neighbors). These criteria were used to calibrate and validate a similarity metric based on Fuzzy Bipolar Pharmacophore Fingerprints.

INTRODUCTION

Nowadays, the early assessment[1] of potential liabilities of tens to hundreds of lead compounds, milestones of the path towards the drug candidate, plays a leading role in drug discovery. It relies on robotized High Throughput Profiling (HTP), monitoring the activities $A_{m,t}$ of molecules m ($m = 1...N_{\text{mols}}$) against a series of $t = 1...N_{\text{target}}$ receptor and enzyme binding, functional (cell-based) and/or ADMET-related biological assays. The resulting experimental output—the "activity profiles" of compounds—provides an extensive characterization of in vitro biological properties (potency and selectivity), being therefore a decisional aid for lead prioritization and potentially useful for the understanding of in vivo drug properties. However, the analysis of these vast activity matrices[2] A_{mt} is not a straightforward data mining exercise, and the development of specific mathematical tools accounting for the peculiarities of the acquired activity data (percentages of inhibition/activation, IC_{50} values, ...) is still at its beginning. In this context, a compound m is represented by an activity profile vector denoting a point $\vec{A}_m = (A_{m,1}, A_{m,2}, ..., A_{m,t}, ..., A_{m,N\text{target}})$ in an Activity Space (AS). The goal of this paper is the development of a mathematical formalism describing the neighborhood properties of the AS, e.g. the definition of activity profile ("in vitro") similarity metrics subsequently used to optimize the Neighborhood Behavior of in silico similarity metrics.

The in silico assessment and exploitation of molecular similarity,[3−5] aimed at enhancing the success rate of drug discovery programs,[6,7] is one important trend in chemoinformatics and high throughput molecular modeling. The heuristic concept of similarity, used by medicinal chemists to classify active compounds on the basis of their structural features, became the underlying principle of computed molecular similarity scores. Molecular structures (further on denoted by M and m) can be represented as points of a structural space[8] (SS) defined by their calculated vectors of molecular descriptors, in which the in silico dissimilarity metric[9] $\Sigma(m,M)$ represents the distance between these points.

The computational cost of molecular descriptors[10] significantly decreased in the latter years, notably due to the development of modeling approaches for combinatorial products.[11−13] This has opened the way to a whole series of time-and-expense-saving computational applications based on the molecular similarity concept, including the following: (i) cluster[13] analysis of the SS, (ii) the selection of a most diverse[14] core sublibrary of minimal size, homogeneously covering the SS occupied by a given (virtual or real) compound set, and (iii) the search for nearest neighbors of an active compound within a collection of yet untested (or even unsynthesized) compounds.[15]

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

10.1021/ci025634z CCC: $25.00
Published on Web 02/07/2003

Table 1. Most Important Symbols and Abbreviations, in Alphabetical Order

symbol	meaning
A_{mt}	activity of a molecule *m* in a test *t* included in the HTP panel
A/V	average/variance normalization of a variable, setting its average value at the origin and its variance as unit length of the transformed coordinate
ADME-T	absorption, distribution, metabolism, excretion−toxicity
AS	activity space—each axis stands for a test in the activity profile
$\chi^{\Sigma}(s)$	consistency criterion at similarity threshold *s*, according to metric Σ—as defined in eq 10
FBPA	fuzzy bipolar pharmacophore autocorrelograms
HTP	High Throughput Profiling (testing of a compound against a target panel)
HTS	High Througput Screening
FS	False Similar compound pairs—*m* and *M* being *similar*: $\Sigma(m,M) < s$, although activity-*unrelated*: $\Lambda(m,M) > l$
$\Lambda^{(k)}(m,M)$	activity dissimilarity score of molecules *M* and *m*. If present, *k* relates to a specific scoring scheme.
l	activity dissimilarity threshold chosen within the validity range of Λ
NB	Neighborhood Behavior
N_{targets}	number of biological assays in the activity panel
N_{mols}	number of molecules in the available data set
P	The set of all $N_p = 170\ 236$ molecule pairs $(m,M) \in P$. The subset of *similar* pairs, $P(\Sigma<s)$ includes the $N_{P(\Sigma<s)}$ pairs with $\Sigma(m,M) < s$, while $P(\Lambda < l)$ of pairs with $\Lambda(m,M) < l$ is the subset of *activity-related* pairs
PFD	Potentially False Dissimilar pairs—*not* perceived as similar: $\Sigma(m,M) > s$, although activity-*related*: $\Lambda(m,M) < l$
$\Sigma^{k}(m,M)$	Structural dissimilarity score of molecules *M* and *m*. *k* denotes a specific score, with descriptors, equation, and weights given in Table 3
s	a structural dissimilarity threshold chosen within the validity range of Σ
SS	Structural Space—its axes being molecular descriptors
TS	True Similar compound pairs—*m* and *M* being *similar*: $\Sigma(m;M) < s$, and also activity-*related*: $\Lambda(m;M) < l$
TD	True Disimilar pairs−*m* and *M* being neither *similar*: $\Sigma(m;M) > s$, nor activity-*related*: $\Lambda(m;M) > l$
w_k, γ	weighing factors specifying the relative importance of each pharmacophore feature k. γ controls the "fuzziness" of the metric
$\Omega^{\Sigma}(s)$	overall optimality criterion at similarity threshold *s*, according to metric Σ—as defined in eq 11

These procedures implicitly assume that the structural space displays a significant Neighborhood Behavior,[16] e.g. that a subset of the structurally nearest neighbors of an active compound is likely to include more actives than expected to be found in any other, randomly chosen, set of the same size. Therefore, calculated similarity scores displaying maximal NB need to be discovered—and a quantitative NB estimator must be provided in order to serve as the objective function in the metric optimization procedure.

Up to date,[17−21] the NB of various metrics has been monitored with respect to a single molecular property at a time—the "main" activity of the compound in the therapeutic area it had been designed for. However, in the novel context of High Throughput Profiling, *NB can be regarded as the property of a Structural Space to map onto the Activity Space in a way ensuring that neighboring points in the former are likely to correspond to neighboring points in the latter*. This paper describes, unlike in previous works on this topic,[17−21] a novel perspective on the Neighborhood Behavior of in silico similarity metrics confronted to experimental activity *profiles*. In other words, the main question debated in this paper is whether and how the classical NB concept "Structural similarity implies a similar activity with respect to a given target" can be generalized to "Structural similarity implies similar *activity profiles*". Prerequisites to this quest are as follows: (i) a quantitative definition of what "similar activity profiles" are—e.g. the introduction of adequate AS metrics quantifying the similarity of activity profiles (in a context where the measured activities A_{mt} are inhibition percentages at 10 μM) and potentially accounting for the bias due to the nonorthogonality of the AS axes (due to the relatedness of some receptors in the test panel) and (ii) the definition of quantitative NB scores adapted to this context.

These NB criteria are then used as objective functions for the choice of the optimal functional form and fittable parameters of a 3D similarity metric based on Fuzzy Bipolar Pharmacophore Autocorrelograms[15] (FBPA). A validation study assessing the NB parameters of the optimal FBPA metrics against different compounds in a different activity space is undertaken in order to cross-check the extent to which the optimal parameters of this metric are biased by the actual choice of targets in the panel used for calibration. For a quick overview of the novel terminology used in this paper, see the glossary in Table 1.

EXPERIMENTAL DATA—TARGET PANEL AND TESTED COMPOUNDS

Although HTP experiments may include cellular and/or ADMET-related assays, in this work the test panel has been restricted to receptor binding and enzyme inhibition tests, best illustrating the recognition of ligands by macromolecular targets. The panel of $N_{\text{targets}} = 42$ tests defining the activity space considered in this work is given in Table 2. Additional information with respect to these tests is available.[22]

A total of $N_{\text{mols}} = 584$ drugs, reference ligands, and precursors used in pharmaceutical industry have been profiled, measuring their average percentages of inhibition at 10 μM (duplicate testing). These inhibition values are suited for use in drug discovery,[23,24] discriminating between the (micromolar or stronger) leads that may be starting points of drug candidate optimization and inactive molecules.

Table 2. Biological Test Panel Used in This Work[a]

label	description	label	description
A1h	human rec. adenosine	5HT6h	human rec. serotonin-6
Alpha1	nonselective rat brain alpha1	5HTUpt	human rec. serotonin transporter
Alpha2	nonselective rat brain alpha2	Sigma1	guinea pig brain sigma-1
Beta1h	human rec. beta1	V1Ah	human rec. vasopresin-1A
AT1h	human rec. angiotensin AT1	K-ATP	rat brain ATP-dep. K channel
BZDc	rat brain central benzodiazepine	Cl	chloride channel
Bomb	nonselective rat brain bombesin	CatB	human placenta cathepsin B
B2h	guinea pig ileum bradykinin 2	Elast	human leukocyte elastase
CCKAh	human rec. cholecystokinin A	PDEII	human phosphodiesterase-II
D1h	human rec. dopamine D1	PDEIV	human phosphodiesterase-IV
D2h	human rec. dopamine D2	PKC	rat brain protein kinase C
DaUpt	human rec. dopamine transporter	EGF-TK	human EGF tyrosine kinase
ETAh	human rec. endothelin A receptor	PK55fyn	bovine thymus p55fyn kinase
Galan	nonselective rat striatum galanin	TNFa	human TNF-α growth factor
H1c	guinea pig central histamine	HIVP	HIV protease
ML1	chicken brain melatonin	NEUpth	human norepinephrine transporter
M1h	human rec. muscarinic M1	MAPkin	rat recombinant MAP kinase
M3h	human rec. muscarinic M3	CGRP	human calcitonin gene-related peptide
NK1h	human glioblastoma neurokinin 1		
NPY	human rec. neuropeptide Y		
Muh	human μ opiate		
5HT1Ah	human rec. serotonin-1A		
5HT1D	bovine caudate serotonin-1D		
5HT2ch	human rec. serotonin-2C		
5HT3h	human rec. serotonin-3		

[a] rec. − recombinant; dep. − dependent; nonselective binding involves the different receptor subtypes.

Both structural and activity dissimilarity scores have been generated for each of the $N_P = N_{mols} \times (N_{mols}-1)/2 = 170\,236$ possible compound pairs (m,M). In the following, this set of all compound pairs will be denoted by P, its size being N_P.

SPACES AND METRICS—DEFINITIONS AND DISCUSSION

4.1. FBPA Structural Space. The 252-dimensional structural space of Fuzzy Bipolar Pharmacophore Autocorrelograms (FBPA) has been described in detail elsewhere.[15] The FBPA of a molecule M, denoted as $\Psi_M(a,b,\Delta)$, is a three-dimensional fingerprint monitoring the numbers of atom pairs within each of the $252 = 21 \times 12$ categories that can be defined in terms of the 21 combinations of $N_f = 6$ pharmacophoric features $a,b \in$ {Hydrophobicity, Aromaticity, Hydrogen Bond Acceptor & Donor, Cation, Anion} times $N_{bin} = 12$ considered distance ranges $\Delta \in$ {(3,4), (4,5), ..., (14,15)} [Å]. Several metrics are applicable to this space:

(i) The original[15] fuzzy similarity metric $\Sigma(m,M)$ relies on partial similarity scores $\delta_{ab}(M,m)$ calculated with respect to the atom pair distributions matching each pharmacophore feature pair a,b:

$$\delta_{a,b}(m,M) = 1 - \frac{2(\Psi_m \otimes \Psi_M)_{a,b}}{(\Psi_M \otimes \Psi_M)_{a,b} + (\Psi_m \otimes \Psi_m)_{a,b}} \quad (1)$$

where

$$(\Psi_m \otimes \Psi_M)_{a,b} = \sum_{\Delta_1=1}^{N_{bin}} \sum_{\Delta_2=1}^{N_{bin}} \Psi_m(a,b,\Delta_1)\Psi_M(a,b,\Delta_2) exp[-\gamma(\Delta_1 - \Delta_2)^2] \quad (2)$$

The global dissimilarity score is defined as a weighted average of the pharmacophore pair partial scores $\delta_{a,b}(m,M)$, with weighing factors w_k associated to each pharmacophore feature:

$$\Sigma(m,M) = \frac{\sum_{a=1}^{N_f}\sum_{b=a}^{N_f} w_a w_b \delta_{a,b}(m,M)|_{(\Psi_M \otimes \Psi_M)_{a,b} + (\Psi_m \otimes \Psi_m)_{a,b} > 0}}{\sum_{a=1}^{N_f}\sum_{b=a}^{N_f} w_a w_b|_{(\Psi_M \otimes \Psi_M)_{a,b} + (\Psi_m \otimes \Psi_m)_{a,b} > 0}} \quad (3)$$

(ii) A fuzzy and weighted Dice[9] metric, bypassing partial similarity score evaluation:

$$\Sigma(m,M) = 1 - \frac{2\sum_{a,b} w_a w_b (\Psi_m \otimes \Psi_M)_{a,b}}{\sum_{a,b} w_a w_b (\Psi_M \otimes \Psi_M)_{a,b} + \sum_{a,b} w_a w_b (\Psi_m \otimes \Psi_m)_{a,b}} \quad (4)$$

(iii) The usual Dice[9] metric—a particular case of eq 4 with weighing factors set to 1 and an infinitely large "fuzziness" parameter γ from eq 2:

$$\Sigma(m,M) = 1 - \frac{2\sum_{a,b}\sum_{\Delta} \Psi_m(a,b,\Delta)\Psi_M(a,b,\Delta)}{\sum_{a,b}\sum_{\Delta} \Psi_m(a,b,\Delta)\Psi_m(a,b,\Delta) + \sum_{a,b}\sum_{\Delta} \Psi_M(a,b,\Delta)\Psi_M(a,b,\Delta)} \quad (5)$$

Table 3. In Silico Metrics Explored in This Work

metric	eq	value range and remarks
Σ^1 Σ^2	(4)	[0,2] − average/variance normalized FBPA metrics using best 2 (sub)optimal parametrization schemes
Σ^3	(4)	[0,2] − normalized FBPA with equal weights $w_k = 1$ and fuzziness factor $\gamma = 0.32$
Σ^4	(3)	[0,2] − normalized FBPA with $w_k = 1$ and $\gamma = 0.32$
Σ^5	(5)	[0,2] − dice metric with normalized FBPA
Σ^6	(3)	[0,2] − normalized FBPA with parameters from ref 15
Σ^7	(3)	[0,1] − FBPA metric as calibrated in ref 15

An overview of the characteristics of the metrics used in this work is given in Table 3. The metric Σ^7 from Table 3 uses the *absolute* number of atom pairs $\Psi_M(a,b,\Delta)$. However, the newly introduced diversity scoring rules $\Sigma^{1...6}$ make use of Average/Variance (A/V) normalized FBPA (see Table 1). Note that $0 \le \Sigma^7 \le 1$, while $0 \le \Sigma^{1...6} \le 2$.

4.2. Activity Space. As already mentioned, each of the $N_{targets}$ biological test of the HTP panel is assimilated to one dimension of the Activity Space (AS), featuring inhibition percentages of 10 μM ligand solutions. Activity space metrics—further on denoted by $\Lambda(m,M)$—have been defined as the number of targets with respect to which two compounds M and m display "significant potency differences", using the following rules (6): if the degrees of inhibition of a target t by m and M differ by more than a high threshold value H% (typically 70%), then this difference is considered to be "significant" and $\Lambda(m,M)$ is incremented by one unit. If the difference is less than a low threshold L% (30%), this is most likely accounted by experimental noise and will be ignored. In intermediate situations, a subunitary linearly interpolated increment will be applied:

$$\Lambda^{(1)}(m,M) = \sum_{t=1}^{N_{targets}} \lambda_t(m,M) \quad \text{where } \lambda_t(m,M) = \begin{cases} 1 & \text{if } |A_{M,t} - A_{m,t}| \ge H \\ 0 & \text{if } |A_{M,t} - A_{m,t}| \le L \\ \dfrac{|A_{M,t} - A_{m,t}| - L\%}{H\% - L\%} & \text{otherwise} \end{cases} \quad (6)$$

If however T and T' are related targets, then the ability of the in silico metric to distinguish binders from nonbinders of T somehow *implies* its ability to do so with respect to T'. Its success with respect to the second target should not count, for it is the consequence of target interrelatedness and not a merit of the metric. Therefore, we have chosen to account for target intercorrelations by means of weighing the distance contribution $\lambda_t(m,M)$ of the target t by a subunitary factor, unless t is "orthogonal" to all the others in the test panel—e.g. if the activities of the molecules against j are not correlated with the ones against other targets. To do so, a measure of the target interrelatedness $R(T,t)$ has been adopted.

$$R(T,t) = \frac{2\sum_{m=1}^{N_{mols}} A_{m,T} A_{m,t}}{\sum_{m=1}^{N_{mols}} A_{m,T}^2 + \sum_{m=1}^{N_{mols}} A_{m,t}^2} \quad (7)$$

The weight ω_T of the increment $\lambda_T(m,M)$ contributed by target T to the activity distance between M and m becomes

$$\omega_T = 1/\sum_{t=1}^{N_{targets}} R^*(T,t) \quad \text{where } R^*(T,t) = \begin{cases} R(T,t) \text{ if } R(T,t) \ge R_{min} \\ 0 \text{ otherwise} \end{cases} \quad (8)$$

where $\omega_T = 1$ if $R(T,t) < R_{min}$ (typically 0.5) for any $T \neq t$. The introduction of a minimal relevant correlation coefficient R_{min} was necessary in order to focus on *meaningful* target correlations only, avoiding possible artifacts due to spurious summation of the numerous very small but positive R values of actually uncorrelated target pairs. It is now straightforward to define the correlation-corrected metric:

$$\Lambda^{(2)}(m,M) = \sum_{t=1}^{N_{targets}} \omega_t \lambda_t(m,M) \quad \text{where } \lambda_t(m,M) \text{ is defined in (6)} \quad (9)$$

MOLECULAR SIMILARITY IN STRUCTURAL AND ACTIVITY SPACES; NEIGHBORHOOD BEHAVIOR—DEFINITIONS AND TERMINOLOGY

Molecular similarity is a pairwise property, inversely related to the distance between the points representing the two compounds in the given structural or activity space. Classically, the term "molecular similarity" is used in the context of a structural descriptor space, implicitly assuming that the employed similarity criteria are of structural (topological, electron-density, or pharmacophoric) nature. Accordingly, in the present work, any reference to "similarity" without any further specifications will imply the structural similarity scores Σ. By "subset of similar compound pairs", we mean the set of all molecule pairs (m,M) of structural dissimilarity score $\Sigma(m,M)$ lower than a given structural dissimilarity threshold s: $P(\Sigma < s) = \{(m,M)|\Sigma(m,M) < s\}$. If $(m,M) \in P(\Sigma < s)$, then m and M are told to be similar compounds—a meaningful definition of what similar molecules are therefore implies (1) the design of an optimal similarity metric Σ *and* (2) the choice of an associated optimal value for the dissimilarity threshold (a.k.a. "dissimilarity radius") s.

The current paper, however, also introduced a different category of molecular similarity metrics $\Lambda(m,M)$, based not on structural criteria, but on a difference in response within a panel of biological tests. To avoid the possible confusion arising from the indiscriminate use of "molecular similarity", we will employ the term "activity-related" to denote a pair of compounds (m,M) having a level of activity panel dissimilarity below an activity dissimilarity threshold l. $P(\Lambda < l) = \{(m,M)|\ \Lambda(m,M) < l\}$ is the subset of "activity-related" molecule pairs—by contrast to $P(\Sigma<s)$, the subset of "similar" pairs. The choice of the appropriate activity dissimilarity threshold l is experiment-driven and would represent the tolerated amount of activity differences in a focused library design based on nearest neighbor retrieval.

The subset of similar pairs that are also activity-related—formally, $TS = P(\Sigma<s) \cap P(\Lambda<l)$—is the subset of "True Similar" pairs, of size N_{TS}. Similar pairs that are not activity-related will be denoted as "False Similar" pairs, members

of the subset FS = $P(\Sigma<s) - P(\Lambda<l)$ of size N_{FS}. Compound pairs that are neither similar nor activity-related will be termed as "True Dissimilar" (TD).

A NB analysis must rely on a clear understanding of the concept of molecular similarity. We consider the following aspects to be important guidelines for the development of NB scoring functions:

A. An important structural similarity *implies* (in a statistical sense) an *activity* similarity of the concerned compounds. More precisely, a subset $P(\Sigma<s)$ of similar pairs should include more activity-related pairs than might be encountered by pure chance in a random pair subset of the same size.

B. The reciprocal of the previous statement is not true, e.g. structural similarity is in no way *necessary* for activity similarity. This is obvious in the case of totally unrelated molecules that do not hit any of the targets and therefore display identical "blank" profiles. Structurally dissimilar compounds *m* and *M* may also share a common set of targets at which both bind, if (i) *m* and *M* adopt different binding modes with respect to each target, or (ii) they have a same binding pharmacophore, but are nevertheless *globally* dissimilar because of structural differences that are neutral with respect to the binding properties (different moieties that protrude out of the site, for example).

Therefore, it would not be correct to denote *dis*similar but activity-related compound pairs as "False Dissimilar" (FD)—herewith implying that their dissimilarity score was overestimated (similarity metric failed to evidence the common underlying structural features). This may but *must* not always be the case, since g*enuinely* dissimilar but activity-related pairs do exist. In the following, we will use the terminology "Potentially False Dissimilar" (PFD) for the category of activity-related but dissimilar pairs members of the subset PFD = $P(\Lambda<l) - P(\Sigma<s)$.

It might also be argued that not all of the structurally similar but nevertheless activity-unrelated compound pairs necessarily deserve to be branded as "False" Similars. The medicinal chemistry literature[25] abounds of examples of activity-unrelated analogues that are, according to any calculated or empirical criterion, as similar as two compounds can get (methyl-normethyl homologues, for example). However, such cases are quite rare compared to the number of methyl-normethyl homologues that show no significant activity differences. In other words, in the case of False Similar compound pairs, the hypothesis of an overestimated molecular similarity score is more likely to apply than the assumption of a genuine "violation" of the similarity principle. By contrast, it is quite difficult to decide which of the possible alternatives best explains the encountered Potentially False Dissimilar pairs.

5.1. Overview of Existing NB Monitoring Criteria. Several criteria[17,18] have been used to compare the NB of various dissimilarity metrics. Unfortunately, those works did not benefit from the knowledge of the activity profile, but relied on activity classes in terms of the supposedly "major" activity against the target each compound had been "designed for".

Cluster separation analysis monitors the distribution of actives from different classes within each cluster of compounds in structural space, following the principle (see paragraph 5A) that binders to different biological targets should *not* cluster together. This is in principle a sound approach, but it is dependent on the choice of the clustering procedure and its adjustable parameters. Borderlines between neighboring clusters may be ill-defined, and it is not true that a couple of members of a same cluster are always closer than they are with respect to members of different clusters.

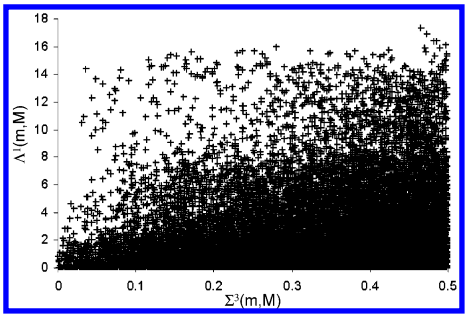


Figure 1. Plots of the activity dissimilarity score $\Lambda^{(1)}$ against the Σ^3 structural dissimilarity score intuitively illustrate the existence of a "forbidden", e.g. less densely populated zone in the upper left corner of the graph.

Monitoring of the number of activity classes sampled by a most diverse compound selection[26] relies on the assumption that picking of compounds that are as structurally different as possible should avoid redundant sampling of a same activity class. This has been also used as an underlying working hypothesis for diversity metric optimization.[15] However, it relies on the biased assumption that members of a same class *have* to be close in structure space (or else they would be simultaneously eligible for a most diverse subset selection). This is equivalent, in other words, to the statement that activity similarity implies structural similarity. It may be of limited usefulness for the calibration of similarity metrics to be specifically used for diversity analysis, but in our opinion (paragraph 5A), the requirement that compound pairs with similar activities should be the *only* ones scoring low structural dissimilarity is a better definition of NB.

Eventually, the plotting of the activity versus in silico dissimilarity scores for all the compound pairs[18] (Figure 1) illustrates the intuitive concept of a "forbidden area" in the upper left corner of the graph (at low structural/high activity dissimilarity) expected to be void of compound pairs as a direct consequence of the "axiom" in paragraph 5A. However, it is less straightforward to derive a quantitative measure of NB based on this idea, because (a) there is no unambiguous rule to define the "borderline" of the "forbidden zone" and (b) these plots may be misleading since they fail to illustrate the local density of pairs at given structural dissimilarity and activity dissimilarity.

5.2. Consistency Criterion. The herein defined *consistency* criterion captures the proficiency of a structural similarity metric to selectively perceive pairs of activity-related compounds as structurally similar. Accordingly, the average of the activity dissimilarity score $\langle\Lambda(m,M)\rangle_{P(\Sigma<s)}$ over the subset $P(\Sigma<s)$ of similar pairs is a measure of consistency. In the worst-case scenario of a metric introducing no bias in favor of activity-related pairs within $P(\Sigma<s)$, this value would equal the global average over the entire pair set $\langle\Lambda(m,M)\rangle_P$. At the other extreme, the best result that a similarity metric may achieve is to pick the $N_{P(\Sigma<s)}$ *most* activity-related pairs—the lower threshold for $\langle\Lambda(m,M)\rangle_{P(\Sigma<s)}$ therefore being $\langle\Lambda(m,M)\rangle_{opt}$, the average of the $N_{P(\Sigma<s)}$ lowest

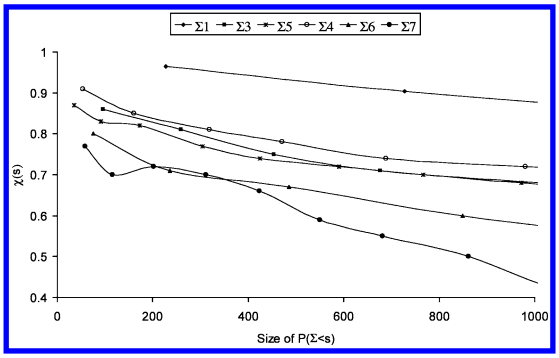


Figure 2. Consistency within the set of 1000 pairs ranked as the most similar by various FBPA metrics, with respect to the $\Lambda^{(1)}$ activity metric.

$\Lambda(m,M)$ scores observed within the N_p pairs. With these considerations, the consistency at a structural dissimilarity threshold s can be written as

$$\chi^{\Sigma}(s) = \frac{\langle \Lambda(M,m)\rangle_P - \langle \Lambda(M,m)\rangle_{P(\Sigma<s)}}{\langle \Lambda(M,m)\rangle_P - \langle \Lambda(M,m)\rangle_{\text{opt}}} \qquad (10)$$

and may take values comprised between 1 (ideal NB) and 0 (random pair picking)—or even below 0 if the in silico "similarity" actively translates into activity dissimilarity!

It can be seen that χ(s) is well suited for the role of an universal dissimilarity coordinate, e.g. a consistent way to map the similarity value ranges of radically different scoring functions onto each other. Indeed, the typical problem encountered when trying to perform similarity searches with different metrics, ex. "what is the appropriate cutoff $0 < s' < \infty$ for an open-ended, Euclidean metric Σ' roughly equivalent to the cutoff s of the Tanimoto score Σ", can be answered by imposing $\chi^{\Sigma}(s) = \chi^{\Sigma'}(s')$. A statement like "the similarity search has been performed with metric Σ, at $s = 0.2$" is difficult to interpret when unacquainted with the characteristics of Σ. By contrast, any similarity search configured such that $\chi^{\Sigma}(s) = 80\%$ corresponds to a roughly equivalent enrichment in activity-related pairs, irrespectively of the metrics in question.

5.3. Overall Optimality Criterion. The consistency criterion χ implicitly keeps track of False Similar pairs, their spurious selection by the structural similarity metric triggering an unjustified increase of $\langle \Lambda(m,M)\rangle_{P(\Sigma<s)}$. It does however not address the problem of the Potentially False Dissimilars: if activity-related pairs are perceived as similar according to *at least one* structural similarity scoring scheme, then the metrics failing to evidence this structural similarity backed by activity-relatedness need to be penalized. A scoring scheme according to which all but very few closely related compound pairs are *dissimilar* may score an excellent consistency but be of limited use because of its failure to pick most of the activity-related compound pairs found by a better similarity scoring scheme.

The "completeness" of the in silico metric can be defined as the fraction of activity-related pairs that actually rank among the structurally most similar ones—and are therefore selected by a computational similarity search. Plots of the consistency criterion χ(s) against the selected subset *size* $N_{P(\Sigma<s)}$ (Figure 2) offer an indirect illustration of completeness: a rapid decrease of χ with respect to the set size suggests that the concerned metric may only achieve a high consistency within a limited subset of few nearest neighbor pairs, e.g. at low completeness.

Table 4. Split of the Set of Considered Compound Pairs with Respect to Activity and Structural Similarity Thresholds

"True Similar": N_{TS}	*"False Similar"*: N_{FS}
$\Lambda(m,M) \le l$ and $\Sigma(m,M) \le s$	$\Lambda(m,M) > l$ and $\Sigma(m,M) \le s$
Potentially False Dissimilar: N_{PFD}	*"True Dissimilar"*: N_{TD}
$\Lambda(m,M) \le l$ & $\Sigma(m,M) > s$	$\Lambda(m,M) > l$ and $\Sigma(m,M) > s$

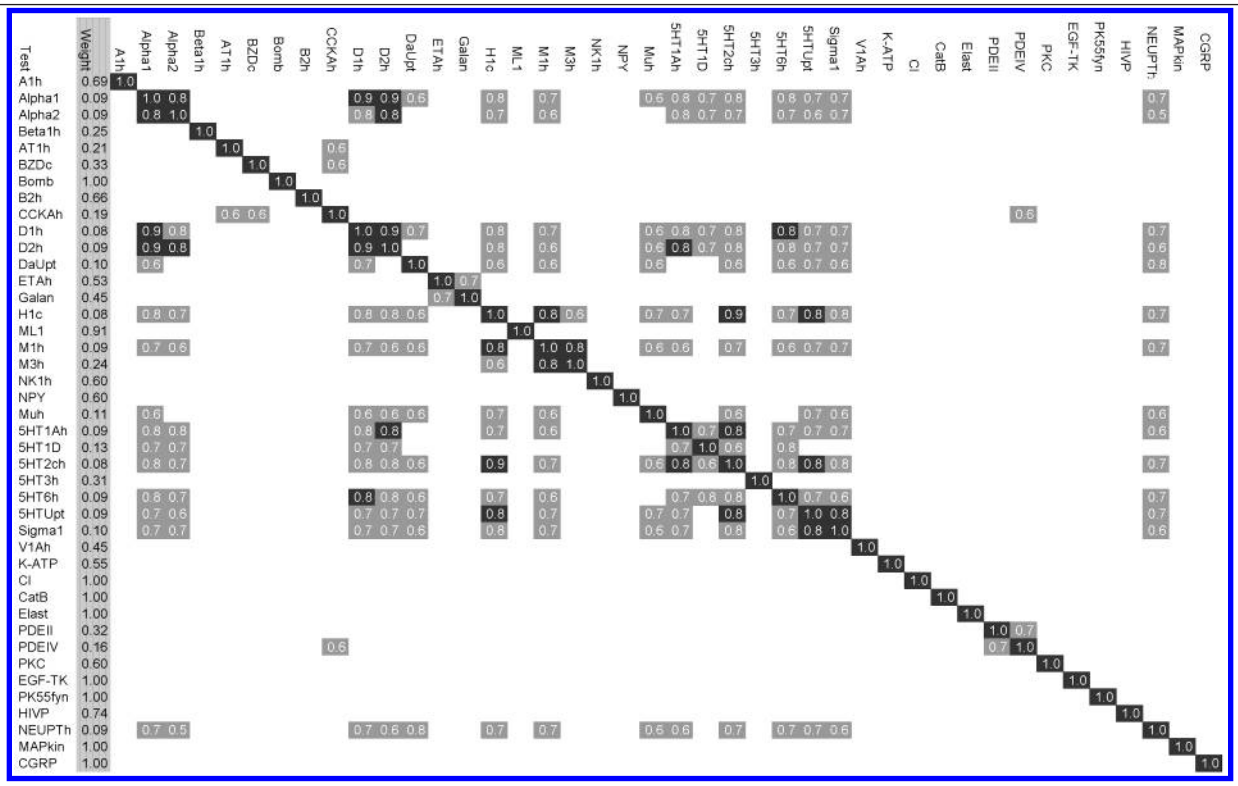
Thus, an overall optimality criterion should describe the dual nature of NB, defined by the two complementary aspects of "consistency" and "completeness". Obviously, an increase in the choice of the dissimilarity cutoff *s* always triggers decrease of consistency in favor of increasing completeness and the role of an overall optimality parameter is to settle for the experimentally most meaningful compromise.

The optimality score proposed in this work relies on the double splitting of the pair set according to both an activity and a structural dissimilarity threshold (*l* and *s*, respectively) shown in Table 4. In all the studies reported here, *l* has been chosen such that the set of activity-related pairs $P(\Lambda<l)$ will include the top $\alpha_l = 15\%$ compound pairs of *minimal* activity dissimilarity. The so defined $N_{P(\Lambda<l)} = \alpha_l N_p$ activity-related pairs include the N_{TS} "true similar" and the N_{PFD} "potentially false dissimilar" pairs, e.g. $N_{P(\Lambda<l)} = N_{TS} + N_{PFD}$. Similarly, the $N_{P(\Sigma<s)}$ pairs of similar compounds (according to metric Σ at cutoff *s*) include the N_{TS} "true similar" and N_{FS} "false similar" pairs and are assumed to represent a fraction β_s of the entire pair set: $N_{P(\Sigma<s)} = N_{TS} + N_{FS} = \beta_s N_P$. A metric with significant NB, operating at its optimal dissimilarity radius s*, is expected to minimize both the N_{FS} and N_{PFD} terms. Since in practice the retrieval of false positives engenders losses in terms of unnecessary synthesis and testing efforts, the penalty related to N_{FS} should be emphasized. Therefore, a scaled sum $\kappa N_{FS} + N_{PFD} = \kappa N_{FS} + \alpha_l N_P - N_{TS}$ may capture well the optimality requirements of a similarity metric (κ has been empirically set to 5 in the present work). This sum must however be compared to the equivalent score obtained by a *random* splitting of *P* into subsets of sizes $\beta_s N_P$ "alleged similar" and $(1-\beta_s)N_P$ "alleged dissimilar" pairs, respectively. In the random splitting scheme, the activity-related pairs will be proportionally distributed within the two subsets, so that $N_{FS}^{(rand)} = (1-\alpha_l) \times \beta_s N_p$ and $N_{PFD}^{(rand)} = \alpha_l \times (1-\beta_s)N_p$. Accordingly, the overall optimality factor Ω^{Σ}(s) can be defined as

$$\Omega^{\Sigma}(s) = \frac{\kappa N_{FS} + N_{PFD}}{\kappa N_{FS}^{(rand)} + N_{PFD}^{(rand)}} = \frac{\kappa N_{FS} + N_{PFD}}{[\kappa\beta_s + \alpha_l - (\kappa+1)\beta_s\alpha_l]N_p} \qquad (11)$$

Practically, Ω is evaluated according to the following steps: (1) at given *s*, the set of pairs *P* is split into the subsets of similar $P(\Sigma<s)$ and dissimilar $P(\Sigma>s)$ pairs; (2) N_{FS} is determined by counting the number of compound pairs from $P(\Sigma < s)$ with $\Lambda > l$; (3) N_{PFD} is provided by the count of pairs from $P(\Sigma>s)$ with $\Lambda < l$; (4) eventually, $\beta_s = N_{P(\Sigma<s)}/N_P$ is calculated and entered in eq 11. The variation of Ω in function of *s* can be easily inferred: at $s = s_{min}$, no similar molecule pairs are selected, therefore $N_{FS} = 0$, and all activity-related pairs are categorized as "Potentially" False

Table 5. Pairwise Correlations $R(t,T)$ between the Activities of the Compound Set against Two Targets T and t, as Defined in Eq 7[a]



[a] Levels of shading correspond to the following—white: $R(t,T) < 0.5$; grey: $0.5 \le R(t,T) < 0.8$; black: $0.8 \le R(t,T) < 1.0$.

Dissimilar: $N_{PFD} = N_{PFD}^{(rand)} = \alpha_l N_P$, which gives $\Omega(s_{min}) = 1$. At $s = s_{max}$ all the pairs are considered "similar"; therefore, $N_{PFD} = 0$, and all activity-unrelated ones count as "False" Similars: $N_{FS} = N_{FS}^{(rand)} = (1-\alpha_l)N_P$, so $\Omega(s_{max}) = 1$. Therefore, the optimality criterion of any metric with significant Neighborhood Behavior will match the random selection level at both extremes of the value range, while reaching a *local optimum (minimum)* at some optimal similarity cutoff s^*. However, the shapes of $\Omega^{\Sigma}(s)$—s plots are conditioned by the peculiar definitions of the value ranges of each metric, whereas plots of $\Omega^{\Sigma}(s)$ against $\chi^{\Sigma}(s)$, the "universal" dissimilarity coordinate—see discussion above—would be consistently comparable throughout any series of arbitrarily different similarity metrics Σ. *These plots will be the main tool for NB analysis, as detailed further on.*

5.4. FBPA Metric Optimization Strategy. An optimization of the seven fittable parameters of FBPA-based metrics, e.g. the six weighting factors w_k associated to the pharmacophoric features and the fuzziness parameter γ, has been undertaken by means of a random search in a space defined by the limits $0 < w_k < 1$ and $0.1 < \gamma < 1.5$. At every explored configuration of parameters, the distance matrix $\Sigma(m,M)$ was rebuilt according to eq 4 and used to determine the consistency with respect to the Λ^2 activity metric within the set of 2000 most similar in silico pairs. The Σ^1 and Σ^2 metrics in Table 3 are based on herein determined optimal parameter configurations.

$$\chi(s)|_{NP(\Sigma<s)=2000} = \mathrm{MAX} \qquad (12)$$

RESULTS AND DISCUSSION

6.1. Preliminary Analysis of the Activity Space. Table 5 displays the correlation coefficients of the ligand responses against the different targets. Unsurprisingly, the strongest correlations can be evidenced between biologically related members of the GPCR superfamily. Many of the found correlations merely reflect the phylogenetic relatedness (sequence homology) of those receptors. The strong bias in favor of GPCR targets in the activity profile and the quite strong correlations within this subset of targets make the use of the correlation-corrected Λ^2 metric mandatory.

Targets that appear to be completely uncorrelated to the other receptors of the panel are often found to display very low hit rates. For each GPCR receptor, the learning set includes at least 35 compounds scoring inhibition percentages above 50% (5HT3h is the less promiscuous GPCR receptor, with 35 hits in the above-defined sense, while 5HT1A, 5HT2ch, or Sigma1 have more than 160). By contrast, there are no significant hits for CGRP, Elast, Cat-B, and B2h and less than 5 hits for AT1h, Bomb, CCKAh, ETAh, Galan, NPY, NK1h, V1Ah, PKC, EGF-TK, and Mapkin. Clearly, the compound set under study, although representative of the ensemble of the currently marketed drugs, is as well biased in favor of GPCR binders—unsurprisingly, since GPCRs are both biologically important and yield high hit rates in HTS experiments, which makes them attractive therapeutic targets.

6.2. Optimality Analysis of in Silico Metrics. The following discussion will be centered around plots of the overall optimality criterion $\Omega(s)$, against the consistency criterion $\chi(s)$. The *lower* (that is "better") the overall optimality values at *high* consistency scores, the better the NB. Situations in which one metric displays a *deeper* overall optimality well at *lower* consistency are however less obvious to interpret. Ultimately, the quality of a metric must be

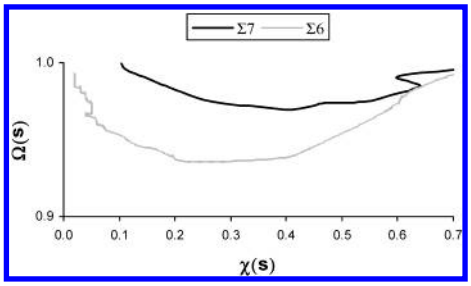


Figure 3. $\Omega-\chi$ plot of the Σ^7 vs Σ^6 metrics, illustrating the importance of A/V normalization.

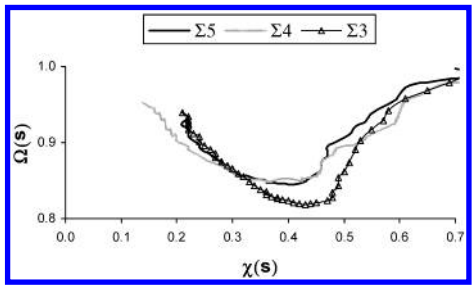


Figure 4. $\Omega-\chi$ plot illustrating the impact on FBPA metric formalism and fuzziness on the NB (at a same weighing scheme).

considered relative to its intended applications, with respect to which the importance of consistency vs completeness needs to be balanced. $\Lambda^{(2)}$ is the AS metric of choice used to build the $\Omega-\chi$ plots throughout this work.

6.2.1. Dice Metric: Choice of the Reference System. The Dice similarity scores are sensitive to both the translation of the origin of the reference system and the rescaling of vector components. The FBPA space is homogeneous in the sense that all its axes monitor atom pair numbers of given pharmacophoric properties at given separation distance ranges, but the pharmacophoric feature pairs are unequally represented within the subset of pharmacologically relevant compounds. In the context of this "privileged" subspace populated by biologically active molecules, A/V normalized FBPA show whether a compound contains more or less pharmacophore feature pairs at given distance than expected for an "average" drug molecule, in terms of "natural" units of observed pair number variances. Figure 3 shows that A/V normalization (metric Σ^6) allows a wider selection of "true similars" at comparable consistency $\chi(s^*) \approx 0.4$. This is consistent with the slower decrease of the Σ^6 consistency vs selected subset size in Figure 2.

6.2.2. NB Dependence on the Functional Form of FBPA Metrics. Figure 4 compares the relative performance of three different ways to estimate molecular dissimilarity scores on hand of FBPA descriptors, at fixed configuration of the fittable parameters $w_k = 1$, (when applicable, $\gamma = 0.32$—see Table 3). The net increase in performance of the fuzzy Σ^3 metric with respect to the equivalent nonfuzzy formalism Σ^5 evidences the benefits of blurring the artificial cutoffs between the distance bins defining the FBPA. By fine-tuning of the γ parameter, the dissimilarity penalty due to differences in 3D distance distributions can be conveniently controlled—from complete ignoring, (a FBPA metric with $\gamma = 0$ behaving like a pure 2D score) to full acknowledgment (at very high γ values, atom pairs of a same type categorized in different, although neighboring, distance bins will no longer contribute to a similarity increase). So, the FBPA

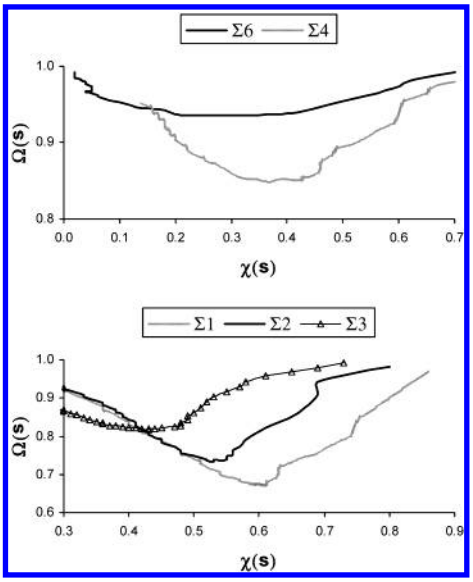


Figure 5. $\Omega-\chi$ plots illustrating the impact of the choice of the pharmacophore feature weighing factors w_k on the NB of FBPA similarity scores.

metric may be seen as a metric of fractional dimension, between 2D and 3D, where γ controls its actual "dimensionality" or sensitivity to 3D information. While the site—ligand interaction mechanisms are certainly controlled by the three-dimensional features of a potential ligand, similarity scoring may not, for practical reasons, be based on the "bioactive" conformers of compounds. The fact that 2D metrics fully ignore geometrical aspects does however not necessarily imply a bad NB, for (a) they are clearly able to evidence the dissimilarity of compounds that, no matter what geometries they may adopt, would *fail to display a similar pharmacophore pattern because they are constituted of pharmacophorically different functional groups*, and (b) they may correctly point out the similarity of pharmacophore patterns of compounds that adopt *similar* conformations—whichever those might be—as a consequence of similar molecular connectivity.

A 3D metric should be more accurate in as far as it explicitly checks the geometric similarity of pharmacophore patterns—and is accordingly expected to detect the pharmacophore similarity of topologically different compound pairs. On the other hand, the artifactual dependence of 3D similarity scores on the peculiar selection of the considered conformers might counterbalance these benefits. The role of the γ factor is therefore to find an optimal consensus between 2D and 3D scoring schemes in order to maximize the benefits in terms of NB.

In terms of the employed formalism for similarity scoring, all other parameters being equal, it can be seen that scoring formula Σ^4 based on feature pair subscores according to eq 3 appears to be a less good idea with respect to the "fuzzy Dice" eq 4 used by Σ^3.

6.2.3. NB Dependence on the Pharmacophore Feature Weights. Metric Optimization Results. Figure 5 illustrates the influence of the fittable parameters w_k and γ on NB. The upper graph shows a surprising *increase* in performance when replacing the previously[15] calibrated pharmacophore feature weights w_k (used in the Σ^6 and Σ^7 metrics) by equal weights $w_k = 1$ in Σ^4. The poor performance of the Σ^6

Table 6. Previously Reported "Reference" and Currently Optimized FBPA Metric Parameters

weighing factor	reference[15] value	optimal	suboptimal
aliphatic	0.41	0.67	0.83
aromatic	0.24	1.00	0.90
HB acceptor	1.00	0.16	0.42
HB donor	0.78	0.15	1.00
cation	0.68	0.61	0.53
anion	0.76	0.03	0.23
fuzziness factor γ	0.32	0.60	0.37

parametrization scheme can be attributed to the use of "maximal diversity" optimality criterion as objective function for its fitting. The difficulty of that calibration procedure did not actually consist in training the metric to achieve a differentiation between easily distinguishable families but rather in forcing the metric to be "tolerant" with respect to the quite high degree of structural diversity *within every family*. The Σ^6 and Σ^7 similarity metrics obtained on the basis of this reasoning do nevertheless display a significant consistency *within the subset of the few best-ranked neighbor pairs* but would not allow an in-depth retrieval of a significant fraction of the activity-related analogues.

The lower plot displays the outstanding improvement of NB obtained by calibration of the fittable parameters (w_k,γ). Although this calibration exercise consisted in optimizing the consistency of the metric with respect to the 2000 most similar in silico pairs, the resulting metrics spontaneously reached an excellent tradeoff between consistency and completeness. The best and second-best different optimums have been considered for analysis—the corresponding parameter configurations being given in Table 6.

The most striking differences between the optimal weighing factors and the previously[15] calibrated weights is the decrease of the relative importance of all the polar pharmacophore features except for positive charge. However, the weighing factor values cannot be directly compared, since the latter have been derived with respect to A/V normalized FBPA, unlike the former. The relative increase of the weighing factors associated to aliphatic and aromatic groups compensates for the downscaling of those FBPA terms at the A/V normalization step, given the high variances of atom pair numbers involving these ubiquitous pharmacophore elements.

6.2.4. NB Optimization of Metrics—General Improvement of Their Relevance with Respect to Biological Activity or Training-Set Dependent Artifact? The suspected dependence of the metric optimization outcome on the target selection of the activity profile prompted us to verify the relevance of the optimized metrics in a completely different environment than the one they were trained against. This NB validation test involved an activity space featuring a series of 12 novel targets (Table 7) that are not *biologically related* to *any* ones from the training set (Table 2) and a set 357 compounds, none of which was part of the initial molecule selection. The relative performances of three FBPA metrics submitted to this validation test (marked by *) are compared in Figure 6 to the ones they scored against the initial NB problem. Σ^3, not resulting from an optimization with respect to the initial data set, is seen to perform slightly better against the validation problem. The overall optimality of the "suboptimal" metric Σ^2 is, in the validation test,

Table 7. A Validation Panel of Targets that Are Completely Unrelated to the Ones Used in the Learning Set (Table 2) Has Been Used To Check the Relevance of NB-Optimal Metrics with Respect to a Different NB Problem than the One Used for Training

label	description
GABA-A	rat brain γ−aminobutiric acid receptor
Kainate	rat brain kaininc acid receptor
NMDA	rat brain *N*-methyl-D-asparagine receptor
P2X	rat urinary bladder purinergic receptor
COX2	human cyclooxygenase II
NOS	mouse macrophage inducible NO synthase
CYP1A2	human recombinant cytochrome P450−1A2
CYP2B6	human recombinant cytochrome P450−2B6
CYP2C9	human recombinant cytochrome P450−2C9
CYP2C19	human recombinant cytochrome P450−2C19
CYP2D6	human recombinant cytochrome P450−2D6
CYP2E1	human recombinant cytochrome P450−2E1

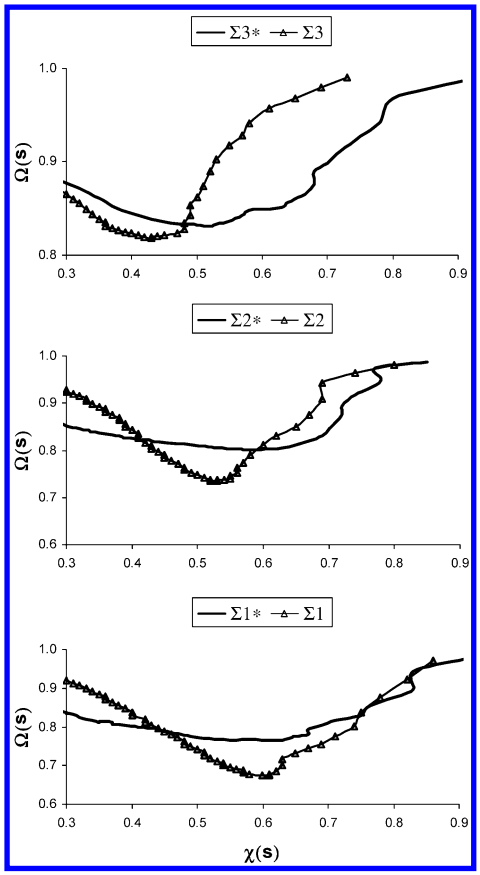


Figure 6. Comparative $\Omega-\chi$ plots of three metrics (not optimized, partially optimized, and thoroughly optimized), monitoring their performances with respect to the completely different training and validation (*) environments.

superior at very high consistency levels, while in the training environment it reaches a deeper minimum, but at slightly worse consistency. Σ^1 shows however a more effective NB with respect to its training environment, this metric being therefore a metric of choice in similarity screening searches for GPCR-binding active analogues. The comparison of the performance of the three metrics in the validation environment showed (Figure 7) the same (although less marked) trend as evidenced in the training environment (Figure 5). In other words, the optimization procedure did not generate an overtrained metric, meaningless outside the environment it was developed for. Unsurprisingly, the accentuation of the relative importance of aromatic and cationic features appears to be a *generally* favorable move toward improved NB, the

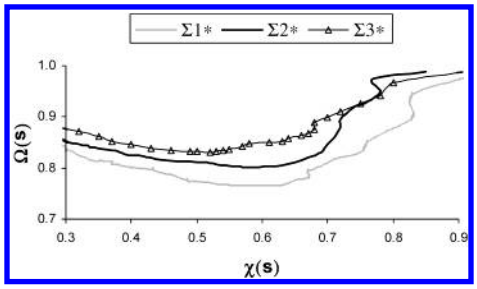


Figure 7. Comparative $\Omega-\chi$ plots of three metrics (not optimized, suboptimal, and optimal) within the validation environment.

relevance of these groups extending to other target families than GPCRs (cytochrome 2D6 being such an example). With a representative panel of relevant receptors of a living organism, the methodology developed here allows the optimization of a general dissimilarity metric capturing the "average" importance of the pharmacophoric features in the ligand recognition process.

CONCLUSIONS

The development of quantitative NB criteria of in silico metrics greatly benefits from the availability of High Throughput Profiling data. The somewhat ill-defined concept of "biological activity classes" used to group together ligands in terms of what is loosely considered to be their "principal" biological target can be successfully replaced by the representation of compounds as "vectors" of a multidimensional activity space. Despite the partial and noisy information conveyed by the inhibition percentages used in this work, meaningful activity metrics can be defined in order to estimate how dissimilar the biological responses of two compounds are.

This work attempted to formalize the definition of NB in terms of two quantitative criteria capturing on one hand the "consistency" of the metric (its ability to specifically avoid scoring high similarity values to False Similar pairs of activity-unrelated compounds) and on the other its "completeness", e.g. its ability to recognize the underlying structural relatedness of compound pairs with similar activity profiles. These criteria have proven to form a solid basis for the calibration of similarity metrics. Testing of a metric calibrated on hand of a diverse, although GPCR-dominated activity panel against a radically different activity space showed that, despite the initial bias in terms of targets, the fitting of pharmacophoric weighing factors lead to the definition of a metric maintaining its good NB properties outside the activity space it had been conceived for. However, the use of customized activity profiles can serve to define "privileged" activity spaces for the training of "biased" similarity metrics with a maximal discriminative power with respect to characteristic ligands for the classes of receptors entered in the profile.

ACKNOWLEDGMENT

This work is 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 platform. The tests were set up and carried out by 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 (Dave Porubek, Jacques Migeon, John Cargill, Xianqun Wang-Redmond) and to Frédéric Revah, Rueil-Malmaison, France, for suggestions and helpful discussions.

REFERENCES AND NOTES

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

(2) Weinstein, J. N.; Myers, T. G.; O'Connor, P. M.; Friend, S. H.; Fornace, A. J.; Kohn, K. W.; Fojo, T.; Bates, S. E.; Rubinstein, L. V.; Anderson, N. L.; Buolamwini, J. K.; van Osdol, W. W.; Monks, A. P.; Scudiero, D. A.; Sausville, E. A.; Zaharevitz, D. W.; Bunow, B.; Viswanadhan, V. N.; Johnson, G. S.; Wittes, R. E.; Paull, K. D. An information-intensive approach to the molecular pharmacology of cancer. *Science* **1997**, *275*, 343−349.

(3) Martin, E. J.; Blaney, J. M.; Siani, M. A.; Spellmeyer, D. C.; Wong, A. K.; Moos, W. H. Measuring, Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery. *J. Med. Chem.* **1995**, *38*, 1431−1436.

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

(5) Maggiora, G. M.; Johnson, M. A. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, 1990; pp 99−117.

(6) Gallop, M. A.; Barret, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M.; Applications of Combinatorial Technologies to Drug Discovery. 1. Background and Peptide Combinatorial Libraries. *J. Med. Chem.* **1994**, *37*, 1233−1251.

(7) Gordon, E. M.; Barret, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. Applications of Combinatorial Technologies to Drug Discovery. 2. Combinatorial Organic Synthesis, Library Screening Strategies and Future Directions. *J. Med. Chem.* **1994**, *37*, 1385−1401.

(8) Oprea, T. I.; Gottfires, J. Chemography: The Art of Navigating in Chemical Space. *J. Comb. Chem.* **2001**, *3*, 157−166.

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

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

(11) Cramer, R. D.; Patterson, D. E.; Clark, R. D.; Soltanshahi, F.; Lawless, M. S. Virtual Compound Libraries: A New Approach to Decision Making in Molecular Discovery Research. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1010−1023.

(12) Horvath, D.; Deprez, B.; Tartar, A. T. High Throughput Molecular Modeling using "Fast 3D" Descriptors. *Act. Chim. Ther.* **1997**, *23*, 55−69.

(13) Barnard, J. M.; Downs, G. M. Clustering of Chemical Structures on the Basis of Two-Dimensional Similarity Measures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 644−649.

(14) Dixon, S. L.; Villar, H. O. Bioactive diversity and screening library selection via affinity fingerprints. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 1192−1203.

(15) 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.

(16) 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.

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

(18) 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.

(19) 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.

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

(21) 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.
(22) Cerep Online Catalog, http://www.cerep.fr/Cerep/Utilisateur/CatalogOnLine/frs_index_cat.asp.
(23) 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.
(24) 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.
(25) Kubiny, H. *Structure-Based Design of Enzyme Inhibitors and Receptor Ligands*, oral presentation at the Second European Workshop in Drug Design; Certosa di Pontignano: May 17−24, 1998.
(26) Kahn, S. Molecular Simulations Inc.: personal communication.

CI025634Z