Virtual Screening with Generative Topographic Maps: How Many Maps Are Required?

Iuri Casciuc, Yuliana Zabolotna, Dragos Horvath, Gilles Marcou, Jürgen Bajorath, Alexandre Varnek

Virtual Screening with Generative Topographic Maps: How many maps are Required?

Iuri CASCIUC[1], Yuliana ZABOLOTNA[1], Dragos HORVATH[1], Gilles MARCOU[1], Jürgen BAJORATH[2], Alexandre VARNEK[1*]

1)Laboratory of Chemoinformatics, University of Strasbourg, France

2)B-IT, Limes, Unit Chem. Biol. & Med. Chem., University of Bonn, Germany

e-mail: varnek@unistra.fr

Abstract

Universal Generative Topographic Maps (GTM) provide 2D representations of chemical space selected for their "polypharmacological competence", *i.e.* the ability to simultaneously represent meaningful activity and property landscapes, associated with many distinct targets and properties. Several such GTMs can be generated – each based on a different initial descriptor vector, encoding distinct structural features. While their average polypharmacological competence may indeed be equivalent, they may nevertheless significantly diverge with respect to the quality of each property-specific landscape. In this work, we show that distinct universal maps represent complementary and strongly synergistic views of biologically relevant chemical space.

Eight universal GTMs were employed as support for predictive classification landscapes, using more than 600 active/inactive ligand series associated with as many targets from the ChEMBL database (v.23). For nine of these targets, it was possible to extract, from the Directory of Useful Decoys (DUD), truly external sets featuring sufficient "actives" and "decoys" not present in the landscape-defining ChEMBL ligand sets. For each such molecule, projected on every class landscape of particular universal map, a probability of activity was estimated, in analogy to a Virtual Screening (VS) experiment.

Calculation of Cross-Validated (CV) Balanced Accuracy (BA) on landscape-defining ChEMBL data was unable to predict the success of that landscape in VS. Thus, the universal map with best CV results for a given property should not be prioritized as the implicitly best predictor. For a given map, predictions for many DUD compounds are not trustworthy, according to Applicability Domain (AD) considerations. By contrast, simultaneous application of all universal maps, and rating of the likelihood of activity as the mean returned by all applicable

maps, significantly improved prediction results. Performance measures in consensus VS using multiple maps were always superior or similar to those of the best individual map.

Keywords: Generative Topographic Mapping, Virtual Screening, Classification Models

1 Introduction

Nowadays we are facing a growing problem with Big Data in many areas and chemistry is not an exception. Currently, ensemble of academic, commercial and propriety databases records more than 100 million of compounds[1]. An estimation of the drug-like chemical space size gives us around 10^{33} virtual compounds[1]. Hence, selection of potential drug molecules from vast collections of candidate compounds is a real challenge for medicinal chemists.

Chemical information is intrinsically multidimensional, as it may alternatively focus on, for example, connectivity, electronic cloud densities, shape, or pharmacophore patterns, and each aspect may prove to be very important for understanding chemical properties and biological activities. These various properties can be encoded by specific molecular descriptors, i.e. specific vectors of N numbers derived from chemical structure, thus representing a molecule as a point in N-dimensional descriptor space. In principle and at arbitrarily high N, this conceptual space may contain almost all known information about molecules, which, in theory, should allow researchers to predict any desired properties using already obtained experimental values as a training input. However, it is impossible to handle such amount of information without advanced data mining techniques. Even though a variety of methods exist[23], the main difficulty is striving for a balance between the accuracy of the results and the computational cost of the required calculations.

One of the techniques that is well suited to reach this balance is Generative Topographic Mapping[4](GTM) - a non-linear mapping method that is widely used as visualization tool for analysis of a multidimensional space. GTM landscapes have already been used as QSAR models[567], and their predictive performance in Virtual Screening (VS) tends to increase with the size and diversity of the dataset used to “color” the landscape. GTM was successfully used for structure-activity analysis of an anti-viral compound set[8] and also of an anti-malarial mode of action database[9]. Recently, it has also been successfully applied to visualize large public chemical databases such as PubChem, ChEMBL[10] and FDB-[11]. Sidorov et al.[12] applied GTM to create “universal” maps of chemical space, that easily distinguished active and inactive compounds for more than 400 ChEMBL targets, yielding an averaged Balanced Accuracy (BA) higher than 0.6 for all targets, indicating high potential of this method for such applications.

The advantage of universal GTM models over classical QSAR approaches is that the most relevant descriptor space that guarantees polypharmacological competence and preferred operational parameter settings defining the manifold are “learned” only once, at the map construction stage. At this stage, large random collections of relevant (drug-like) compounds are used to span biologically relevant chemical space, serving as a “frame set” for unsupervised GTM manifold fitting, while a large and diverse ensemble of structure-activity sets are employed

as “selection sets”. Their role is to score the quality of the current manifold for its ability to host predictive landscapes corresponding to each selection set activity. Top manifolds scoring well at this stage are selected as the final “universal” maps, with the expectation that they will also be able to support predictive landscapes for other, distinct properties, beyond those present in the selection set. This expectation was well met by more than 400 structure-activity sets consisting of novel compounds associated with completely unrelated targets and properties by Sidorov et al.[12]. Certainly, dedicated models that might be built for a given property could exceed the predictive power of universal GTM-based property landscapes – if sufficient training data are available. By contrast, universal GTM manifolds act like “default”, zero-parameter models that can even be employed to explore scarcely studied properties with little experimental data. Therefore, they are both the best strategy to use with incipient, small structure-activity series, and an economic, rapid, fitting-free approach to model building for large and diverse series.

In this work, we assess the predictive performance of eight newly constructed universal GTM models in VS of nine target-specific compound sets extracted from Directory of Useful Decoys (DUD)[13]. These GTMs have been constructed on basis of ChEMBL[14] (v.23) structure-activity data for the respective targets – each based on a different initial descriptor vector, encoding distinct structural features. Their average polypharmacological competence is (roughly) equivalent – they are all members of the top ranked population produced by the evolutionary map building process. Nevertheless, they significantly differ in the quality of each property-specific landscape. We show that distinct universal maps represent complementary and strongly synergistic view of chemical space. The predictive power of any classification landscape built for ChEMBL data can be internally assessed by the Cross-validated Balanced Accuracy (BA_{CV}) criterion in an “aggressive” three-fold cross-validation experiment repeated five times, with data scrambling. However, the BA_{CV} indices were shown unable to predict the success of that landscape in VS. Thus, it would be an error to prefer the universal map with best CV results for a given property as the implicitly best predictor. For a given map, predictions for many DUD compounds are not trustworthy, according to Applicability Domain (AD) considerations. By contrast, simultaneous application of all universal maps, and rating of the likelihood of activity as the mean returned by all applicable maps, significantly improved prediction results. On the basis of different measure, the performance of consensus maps in VS was consistently better than of individual maps.

2 Methods

2.1 Data

The target-specific compound series extraction protocol by Sidorov[12] has been applied to release 23 of the ChEMBL database. A total of 618 datasets containing ligands of different ChEMBL human targets have been extracted. The same structure standardization procedure (*vide infra*) has been applied to DUD database, followed by removal of molecules that were

present in ChEMBL to create orthogonal external datasets. For most of the targets shared by ChEMBL and DUD, this required elimination of all the actives from the DUD series. However, in nine cases the DUD target-specific series contained sufficiently numerous original actives and were used for VS. Table 1 summarizes the composition of selected compound datasets.

Table 1: Description of target-specific subsets used for model training (ChEMBL) and VS (DUD).

ChEMBL ID	Target Name	DUD dataset		ChEMBL dataset	
		Active	Inactive	Active	Inactive
1827	Phosphodiesterase 5A	170	25334	691	1515
1952	Thymidylate synthase	63	6113	124	455
251	Adenosine A2a receptor	79	28001	1303	3618
260	MAP kinase p38 alpha	100	32925	1453	2567
279	Vascular endothelial growth factor receptor 2	94	22595	2047	4663
301	Cyclin-dependent kinase 2	189	25675	638	2305
4282	Serine/threonine-protein kinase AKT	52	14228	725	2619q
4338	Purine nucleoside phosphorylase	102	6334	100	111
4439	TGF-beta receptor type I	82	8013	282	385

2.2 Workflow

The following workflow was applied:

1) Standardization of ChEMBL and DUD datasets followed by descriptor generation;
2) Coloring the manifolds of universal maps by each of nine target-specific class landscapes using ChEMBL subsets;
3) 5-fold CV of predictive landscapes using ChEMBL datasets
4) VS applying these landscapes to the DUD subsets;

For some of these steps a dedicated section is presented below.

2.3 Data preparation and descriptors generation

Structures from both databases ChEMBL (version 23) and DUD were standardized accordingly to the procedure implemented on virtual screening server of the Laboratory of Chemoinformatics in the University of Strasbourg (infochimie.u-strasbg.fr/webserv/VSEngine.html) using the ChemAxon Standardizer[15]:

- Dearomatization and final aromatization according to the “basic” setup of the ChemAxon procedure (heterocycles like pyridone are not aromatized)
- Dealkalization
- Conversion to canonical SMILES
- Salts and mixtures removal

- Neutralization of all species, except nitrogen (IV)
- Generation of the major tautomer according to ChemAxon

After the standardization, 1 540 615 compounds from ChEMBL and 914,379 compounds from DUD remained.

The descriptors used here were ISIDA descriptors computed by ISIDA Fragmentor[161718]. More than 100 different types of descriptors sets were generated. They include sequences, atom pairs, circular fragments and triplet counts of different length, colored by formal charges, pharmacophore features or force field types. These fragmentation schemes were selected for relatively low number of fragments they generate.

2.4 Generative Topographic Mapping

Generative Topographic Mapping (GTM) is a non-linear mapping method used for data visualization originally described by Bishop. In GTM 2D latent space (called manifold) is embedded into the descriptor space. The points which are close in the latent space remain neighbors in the data space. The manifold represents a grid of k x k nodes; each node is mapped in the initial descriptor space using the mapping function y(x, **W**). The mapping function is given as a grid of *m* x *m* radial basis functions (RBF). In order to build a GTM-based QSAR model, the weighted average of properties of all molecules associated with any particular node is used to "color" the manifold according to that property. Here, the projected property is activity class membership, resulting into a fuzzy activity landscape. Molecule "responsibilities" are used as weights. Red and blue zones are only populated by active and inactive compounds, respectively; all colors in between correspond to the regions occupied by compounds of both classes in different proportions. White zones represent unpopulated areas.

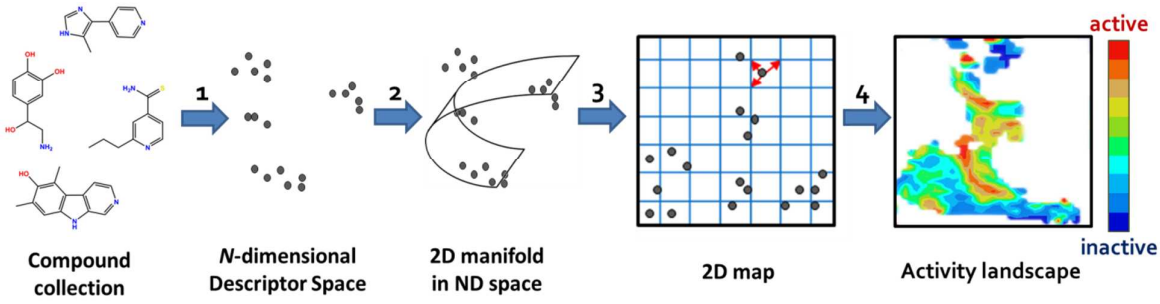

Figure 1: A frame set of compounds is represented in the N-dimensional descriptor space. A flexible 2D manifold, which is a square grid of nodes, is injected into that space and is fitted to the data. The molecules are non-linearly projected onto it, and when the manifold is unbent, a 2D map is obtained. Each node can be colored according to the activities of molecules residing there, producing "activity landscapes", where red zones are populated only by active molecules, blue – by inactive, all colors in between correspond to the regions occupied by compounds of both classes in different proportions. White zones are empty.

GTM supports several Applicability Domain (AD)[6] definitions, but only the density-based AD is applied here. Compounds projected onto a "white zone" of the map (accumulating no responsibilities of "training" compounds used to build the landscape) are out of the AD.

Note, however, that the AD considerations in VS may differ from those in predictive QSAR. In the latter case, compounds outside of the AD should be ignored – no prediction of their property should be attempted. In VS, however, the inability to obtain a trustworthy prediction for out-of-AD compounds practically implies that those compounds will be never selected for synthesis and testing because they were predicted to be inactive. Therefore, external compounds falling within the blank spots of the employed class landscapes were assigned zero probability to be active placing them at the bottom of rankings.

Global manifolds (Universal maps) were derived following the procedure in the cited work[12] but employing updated compound datasets. They are based on frame sets of maximal diversity (aimed at spanning the entire drug-like chemical space), employed 236 of the above-mentioned 618 compound series for map selection. Like in any global mapping approach, they are not meant to capture the detailed SAR of every target-specific set but allow analysis of several activities at the same time. Note that global activity landscapes are relying on a common manifold, itself derived from a selected descriptor space in order to maximize the mean predictive power of all these landscapes. It is obvious that global manifolds represent a best compromise to describe biological activity in general, based on some 'consensus' descriptor space. Interestingly, several such descriptor spaces were identified, each focusing on different aspects of chemical structures. Eight global (universal) maps based on eight distinct ISIDA fragment descriptor spaces were selected (Table 2). On the average, their mean predictive power over all the 618 considered activity sets is similar, while corresponding predictions for each activity series fluctuate.

Table 2: Description of eight universal maps, their descriptor types and the descriptor space dimensionality.

Map	Abbreviation	Definition	Descriptor space dimensionality
1	IA-FF-FC-AP-2-3	Sequences of atoms with a length of 2 to 3 atoms labeled by a Force Field and Formal charges using all paths.	5161
2	IIRAB-FF-1-2	Atom-centered fragments of restricted atom and bonds of a length 1 to 2 atoms labeled by a Force Field.	3172
3	IAB-PH-FC-AP-2-4	Sequences of atoms and bonds of a length 2 to 4 atoms labeled by	4245

		pharmacophoric atom types and formal charges using all paths.	
4	IA--2-7	Sequences of atoms of a length 2 to 7 atoms.	6520
5	IAB-FC-AP-FC-2-4	Sequences of atoms and bonds of a length 2 to 4 atoms labeled by formal charge using all paths	3437
6	IA-FF-P-2-6	Sequences of atoms pairs with a length of 2 to 6 atoms labeled by Force Field	2901
7	III-PH-3-6	Atom triplets labeled by pharmacophoric atom types with topological distance from 3 to 6 bonds	4846
8	III-FF-3-4	Atom triplets labeled by a Force Field with topological distance from 3 to 4 bonds	8953

2.5 Performance evaluation

Model performance was evaluated using BA in 3-fold CV and VS, Receiver Operating Characteristic Area Under Curve (ROC AUC) in VS and Enrichment Factor (EF) in VS. BA has been mainly used during cross-validation. BA serves to assess the ability of landscapes to predict the correct activity class of candidates not used for landscape construction, i.e. both in "internal" cross-validation and "external" VS. However, ROC AUC is a more natural VS evaluation criterion than BA, since the latter requires a formal prediction, active *vs* inactive, for each external compound. In VS, however, the key element is the relative ranking of candidates – a significant prioritization of the actives with respect to the inactives is sufficient to guarantee VS success. Ranking was performed according to the GTM landscape-predicted probability of each compound to be active. The compounds falling outside the applicability domain were assigned zero probability of activity thus they were placed at the bottom of the ranking list. To complement ROC AUC values, the EF of actives ranked within the 100 top compounds was also monitored. EF for the top 100 ranked molecules was calculated according to the equation below.

$$EF_{100} = \frac{Actives_{100}/100}{Actives_{total}/N_{total}}$$

where $Actives_{100}$ is the number of true positives in the top 100 compounds, $Actives_{total}$ is the total number of active compounds in the dataset, N_{total} is the total number of compounds in dataset.

However, selection of the top 100 compounds may be considered only if there is a significant gap between the probabilities to be active of the 100th selected compound and the

one of the 101st, not selected candidate. In practice, several candidate compounds will have the same predicted probability to be active (reported with a precision of 0.01) and therefore all those that are equiprobable to the 100th selected compound would be equally deserving to enter the selection. In order to force selection of a top 100 compounds, a random subset of these equiprobable must be picked in completion of the better ranked candidates. In this *a posteriori* study, three scenarios are considered to compute the EF:

1. Pessimistic: out of candidates that are equiprobable to the 100th selected compound, inactives are selected first, and then the remaining places in the pessimistic top 100 are completed by actives.
2. Optimistic: the opposite strategy (actives are filled in first, remaining places taken by inactives)
3. Stochastic pick out of candidates that are equiprobable to the 100th selected compound.

Scenarios 1 and 2 are deterministic. The values obtained are termed Pessimistic Enrichment Factor (PEF) and Optimistic Enrichment Factor (OEF), respectively. Scenario 3 is not deterministic and repeated random drawing/averaging would be required to converge to expectation values. Yet, it is possible to estimate an average value, termed Interpolated Enrichment Factor (IEF) using the following equation:

$$IEF = \lambda \times PEF + (1 - \lambda) \times OEF$$

$$\lambda = \frac{n}{N}$$

where IEF - interpolated enrichment factor; OEF - optimistic enrichment factor; PEF - pessimistic enrichment factor; λ - the ratio n/N, with N being the size of set including all the candidates that are equiprobable to the 100th selected compound and n the number of these latter candidates. For instance, if the set including all the four candidates that are equiprobable to the 100th selected compound contains 102 hits, then N=4and n=2 such that λ=0.5.

3 Results

3.1 Cross-Validation of ChEMBL activity class landscapes

Three-fold CV of the BA was repeated five times for each of the ChEMBL series. For the 236 “selection” series, this was part of GTM manifold scoring process, where the fitness score reflects the mean of each BA_{CV} value. For the eight selected manifolds, the same CV procedure was applied to the remaining 618-236 “external” series, thus obtaining the complete matrix of the predictive power of every map for each of the 618. Unsurprisingly, not every property is equally well predicted by each map, albeit the average BA_{CV} value may not differ much from map to map. Each map was examined in order to identify the number of targets for which it is able to solve the active/inactive classification problem at BA_{CV} above a given threshold.

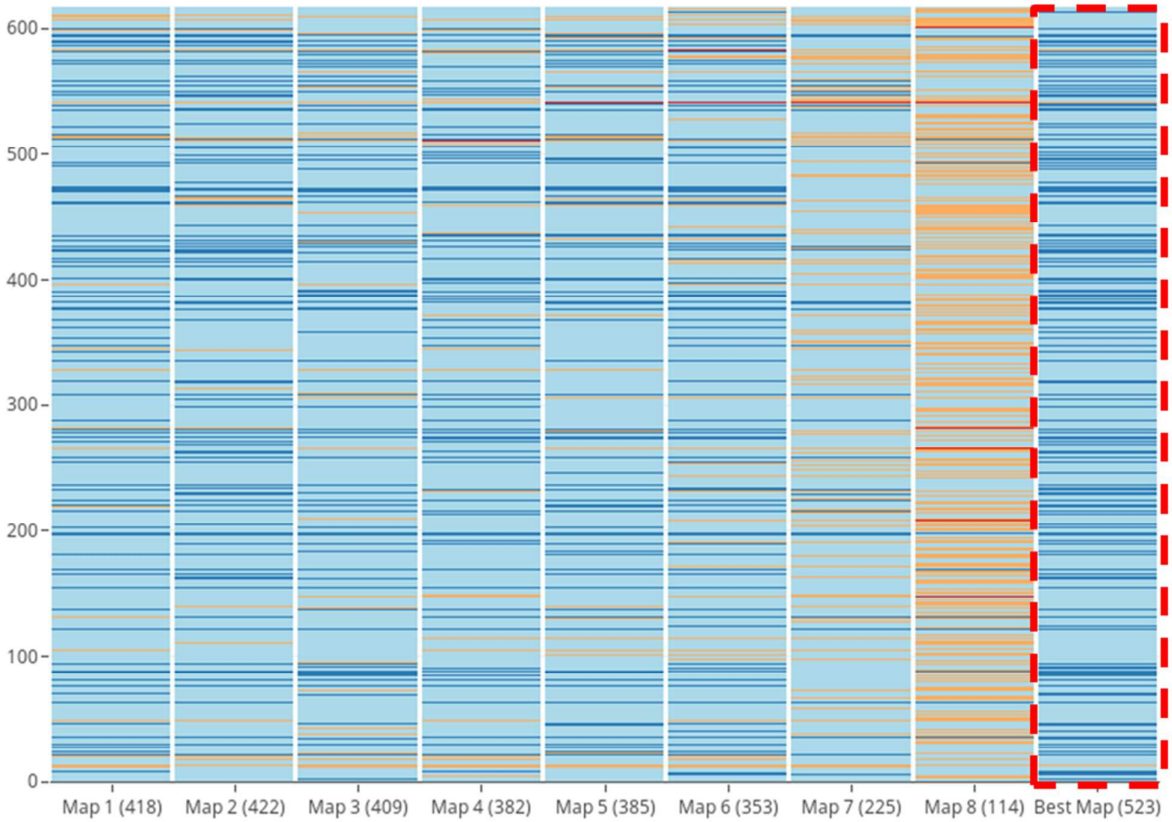

Figure 2: Heatmap showing the performance of universal maps on 618 selected series. Color-codes: dark blue – BA>0.85, light blue – 0.65 < BA ≤ 0.85, orange – 0.5 < BA ≤ 0.65 and red BA ≤ 0.5. Between parenthesis is shown the number of target-specific classification problems for which a map scores BA > 0.75.

Figure 3 shows that for 617 of 618 targets, BA_{CV} scores of 0.6 or better are achieved by at least one of the maps. The exception (ChEMBL5678) represents a set with too few compounds. Note that maps are ranked according to their original fitness score (mean BA_{CV} scores over the 236 selection SAR series) and it can be seen from Figure 3 that the first map is strongly predictive (BA_{CV} >0.75) for 418 distinct series. Note that part of these 418 are selection series but include a significant number of external series nevertheless. It is also noteworthy that every single map is able to provide significantly better-than-random separation of actives and inactives (BA_{CV} >0.6) for virtually all (609/618 – in case of map #1) SAR sets, which fully justifies the label of "universal" maps. However, one single map is not expected to flawlessly model all series – no single descriptor space (fragmentation scheme) on which a map is built could capture all the relevant chemical information that might impact so many different structure-activity relationships. The eight selected maps are highly complementary: series less well explained by one map will work better on another manifold, exploiting specific information from its distinct descriptor space to host a strongly predictive model. Cumulated prediction performance increases with the number of considered maps (Figure 3) which clearly demonstrates the maps complementarity: Seven universal maps based on as many distinct

descriptor spaces are sufficient to provide at least one satisfactory result for more than 85% of used targets even at the very stringent BA_{CV} >0.75. Thus, for further analysis, only seven universal maps were used.

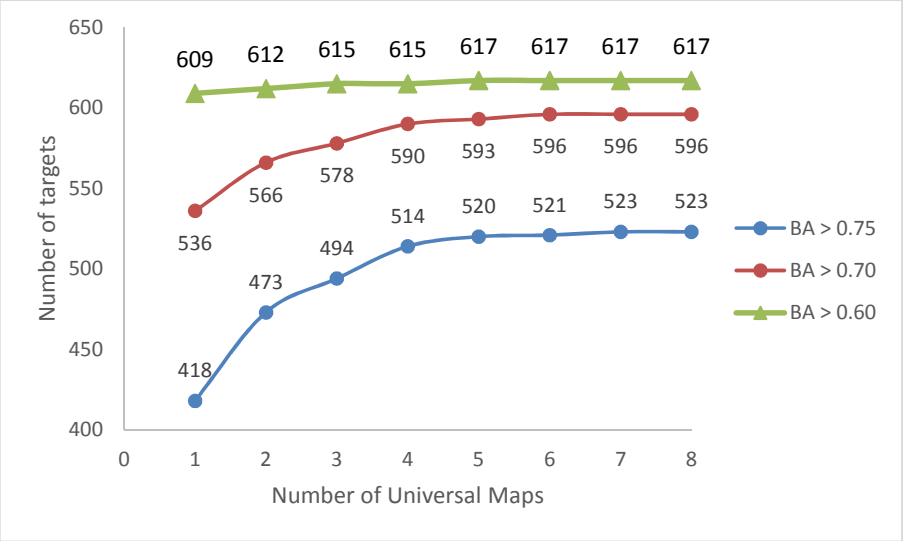

Figure 3: Cumulated performance of universal maps: number of predicted target-specific series vs number of used maps

3.2 Is BA_{CV} a reliable indicator of VS success?

Next, the question how to identify the best universal map for a particular activity was addressed. It may be expected that the model which shows highest predictive CV performance in target-specific ligand classification would be the best model in VS. In order to test this hypothesis, correlation between landscape performance in CV and VS was evaluated for each of the 63 QSAR models (activity landscapes for nine targets on seven universal maps). Figure 4 compares, for the specific activity landscapes of target CHEMBL260 hosted on each map, the "internal" estimation predictive power (BA_{CV}) on one hand, and, the observed predictive power in "external" VS of the DUD subset, on the other hand.

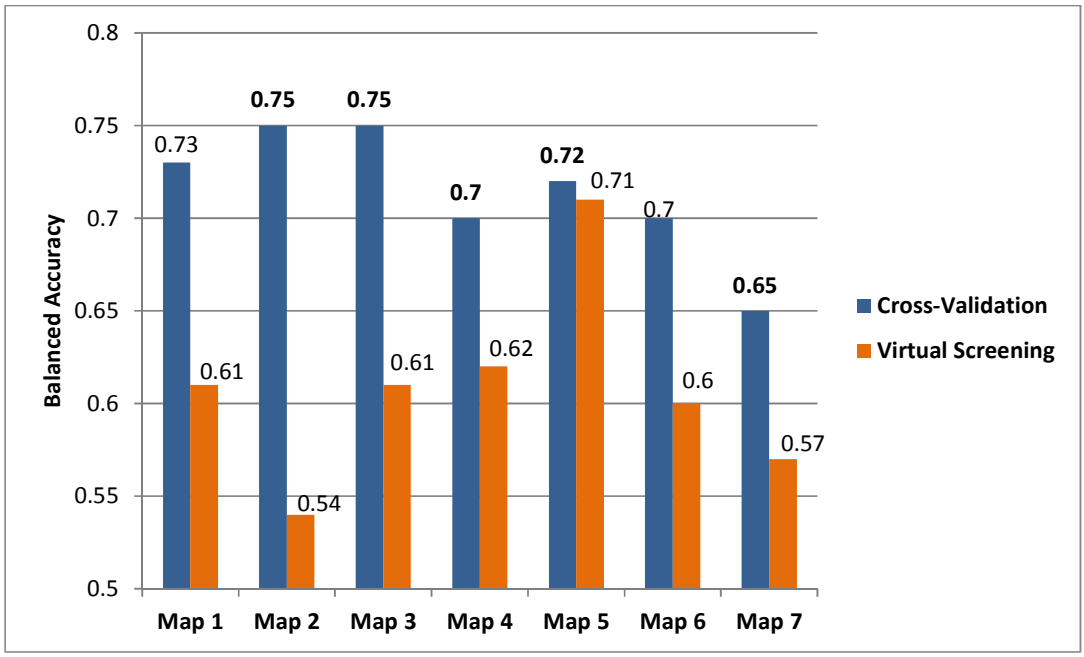

Figure 4: BA values obtained in CV and VS of the CHEMBL260 dataset.

The Pearson correlation coefficient of BA_{CV} versus BA_{VS} over the seven maps was calculated for all the nine sets; they vary in the range 0.02 – 0.63 which means that a map can hardly be chosen on the basis of its CV performance. Unfortunately – but not unexpectedly[19] – high BA_{CV} is a necessary, but not sufficient guarantee of model success in VS. The success in a predictive challenge depends on the peculiar composition of the test set.

3.3 Consensus of Universal Maps.

Given the genuine complementarity of the seven maps, consensus predictions by averaging results these complementary views of chemical space might be a promising strategy. Here, averaging was applied to the probability of activity from each of the seven landscapes, for each compound from the external test set, excluding, however, landscapes in which the compound was projected into an "empty" zone (Figure 5). In this study, the density-based AD criterion as implemented by default in ISIDA GTM)was applied [6]. Compounds that fell outside the AD in all existing maps were excluded from the consensus model.

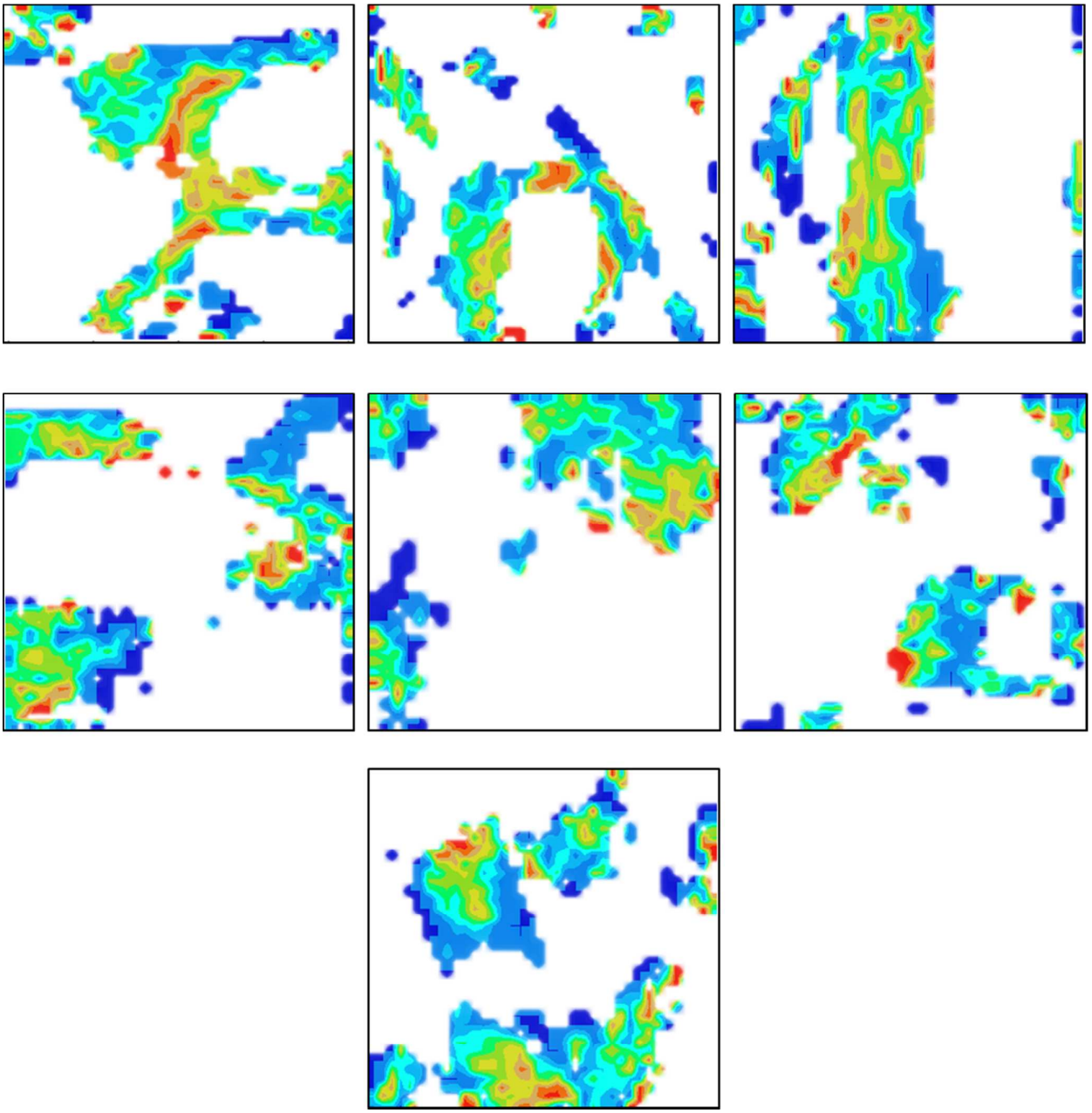

Figure 5: Activity landscapes of the CHEMBL260 dataset in seven universal maps**.**

Apart from the fact that consensus allows making predictions without choosing *a priori* one best map, it has another important advantage - data coverage increase (percentage of the compounds that are considered to be in AD). For example, none of the maps of the CHEMBL260 subset provided 100% data coverage achieved by the consensus. Similar observations were made for the remaining eight datasets. Only for two, coverage was less than 100% (ChEMBL4338 – 79,8%; ChEMBL4439 – 97,5%). Recall that in a VS context, compounds out of AD are not “discarded”, but given a probability of zero to be active, which implicitly ranks them at the bottom of the list. Thus, data coverage in this context does not impact on the size of the screened compound set (BA, EF and ROC AUC values are reported with respect to the full

DUD sets, respectively). Data coverage however impacts the reliability of results since increasing data coverage reduces compounds with zero probability of activity.

Figure 6 shows that consensus BA values generally exceed the majority of BA scores achieved by individual universal maps. Only universal the map of CHEMBL260 outperformed the consensus model

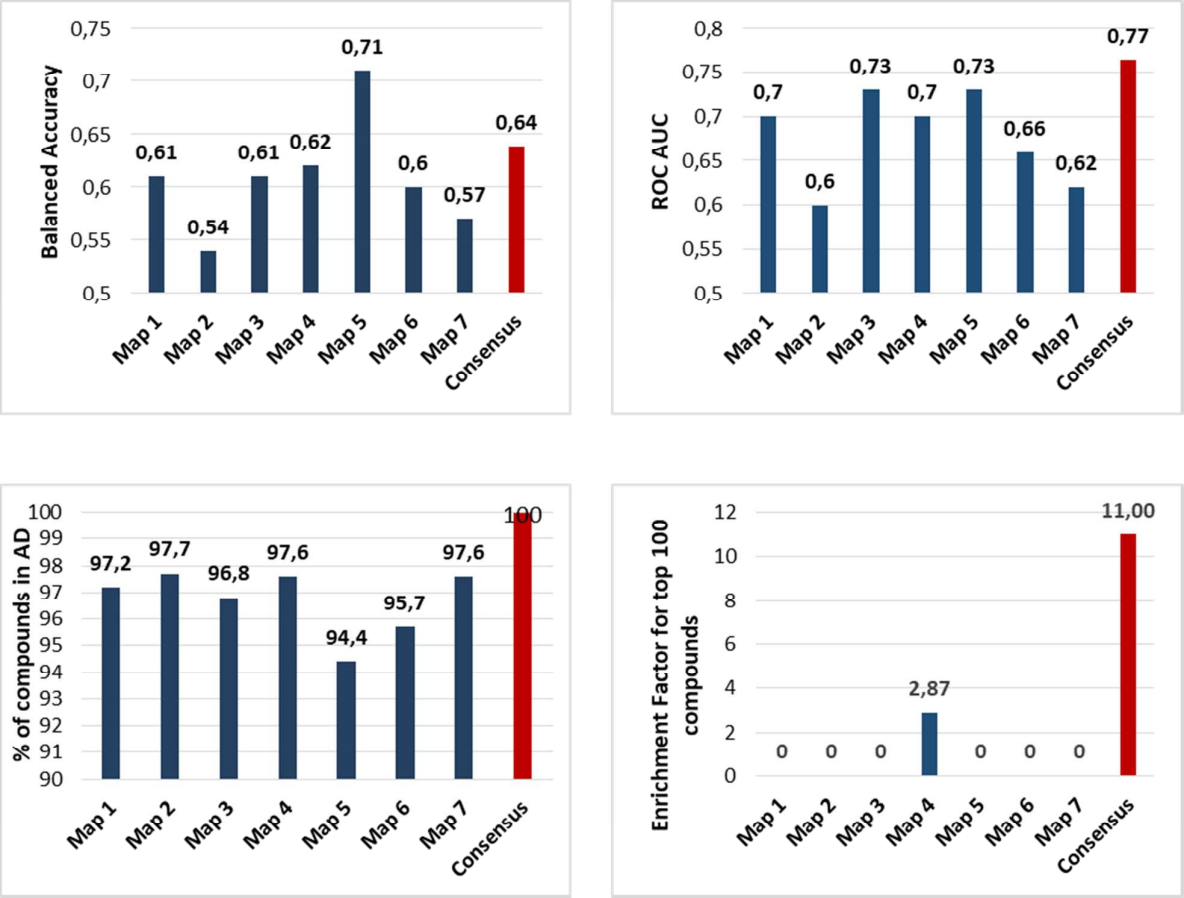

Figure 6: **Performance of VS on DUD with the models developed for the CHEMBL260 dataset assessed on the basis of BA (top left), ROC AUC (top right), data coverage (bottom left) and EF calculated for top 100 compounds (bottom right)**

In terms of EF, no individual model except universal map 4 was able to rank any of active compounds from DUD into the top 100. For the universal map 4, EF=2.87 corresponded to a single active compound in the top 100. However, the EF for the consensus model reached 11, which resulted from five true actives in the top 100.

In **Error! Reference source not found.** the results for all nine datasets are shown. The consensus model performed than any individual map on the basis of EF. We note that CHEMBL4338 represents an atypical dataset because about 60% of the compounds fell outside the applicability domain.

Table 3: Performance in CV and VS for individual universal maps compared to consensus models.

Target	Cross-Validation		Virtual Screening				Consensus model		
	Best Map number	BA	Best Map number	BA	ROC AUC	EF	BA	ROC AUC	EF
ChEMBL1827	4	0,82	7	0,70	0,73	0,00	0,67	0,74	1.5
CHEMBL1952	4	0,83	5	0,82	0,85	0,13	0,82	0,86	14.7
CHEMBL251	2	0,77	3	0,77	0,84	1,56	0,80	0,88	17.8
CHEMBL260	2	0,75	5	0,71	0,73	0,00	0,64	0,77	11,00
CHEMBL279	2	0,73	4	0,71	0,78	0,00	0,66	0,82	4.83
CHEMBL301	3	0,80	5	0,74	0,80	0,60	0,81	0,87	5.47
CHEMBL4282	5	0,81	3	0,81	0,87	17,39	0,83	0,92	52.18
CHEMBL4338	5	0,84	3	0,71	0,73	0,00	0,54	0,66	0,00
CHEMBL4439	5	0,81	5	0,75	0,88	1,97	0,67	0,88	4.94

Analysis of CHEMBL4338 revealed the presence of distinct structural subsets in DUD and ChEMBL, which provided a rationale for low performance on the basis of BA, ROC AUC and EF.. The ChEMBL series used to build the activity landscape mainly contained fused aromatic heterocycles such as hypoxantine, pyrrolopyrimidne, benzimidazole-4,7-quinone (Figure 7).

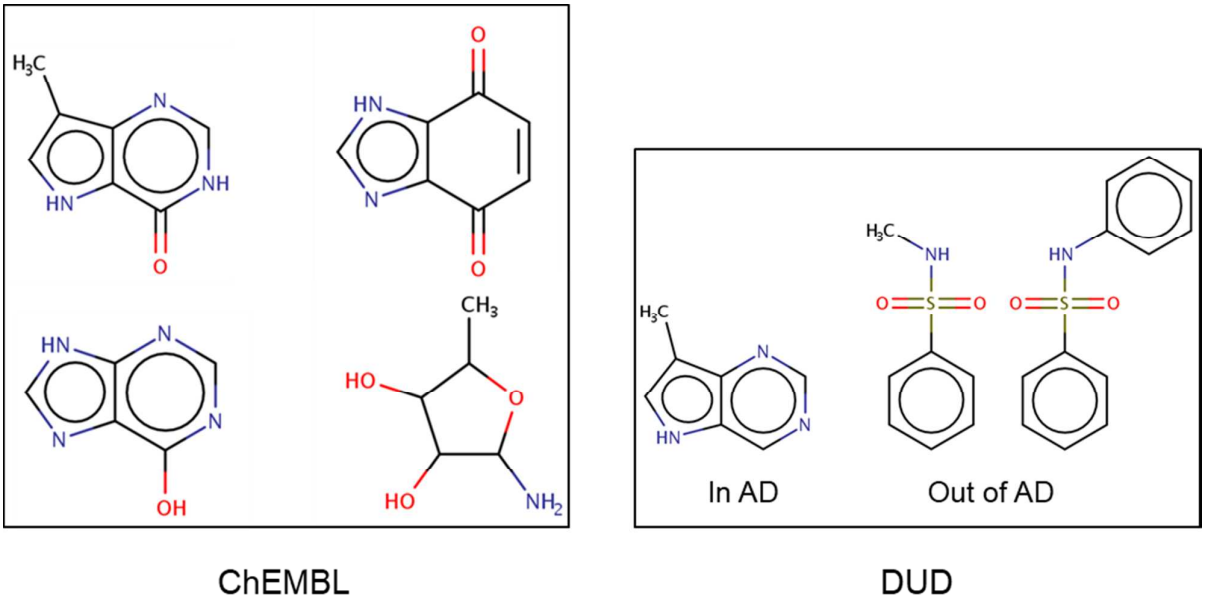

Figure 7: Maximum common substructures of compound subsets active against the purine nucleoside phosphorylase receptor in the CHEMBL4338 dataset and DUD.

In the DUD series, the majority of compounds that were correctly predicted contained a purine moiety similar to training set molecules. (Figure 7). However, compounds outside of AD were N-phenylsulfonamides (Figure 7) that were not present in the ChEMBL dataset.

4 Conclusion

A new series of "universal" chemical space maps from datasets in the ChEMBL23 database was built using the GTM dimensionality reduction algorithm and following a previously reported evolutionary procedure to select preferred descriptor spaces and GTM parameter stings. These maps were able to provide better than random separation (BA_{CV}>0.6) of actives and inactives in 609 of 618 ChEMBL sets, irrespective of whether series were used for map selection or not. However, consistently accurate predictions for each activity class could not be achieved be achieved by any individual map. However, these maps, which were each based on a different descriptor space, were highly complementary. For 617 of 618 activity classes, at least one out of the seven top universal maps represented a highly discriminatory activity landscape.

Since there is no correlation between performance in CV and external predictive power of individual activity landscapes, the one possible solution is to use a consensus approach. The, all landscapes with favorable density distributions of VS candidates make positive contributions to the consensus model. The most important advantages of a consensus map are: 1) 100% data coverage in most of the cases; 2) significant increase in EF for the 100 top ranked compounds; 3) high performance of the consensus model compared to individual models on the basis of ROC AUC. Thus, while any single universal map displays moderate predictive power f, the combination of complementary maps results in a strong consensus effect in VS. Seven universal maps were sufficient to generate complementary views of biologically relevant chemical space that resulted in further increased VS performance.

Supporting Information

Activity landscapes for all nine DUD subsets used in VS are provided.

Acknowledgment

IC thanks the Région Grand Est for a PhD fellowship.

Bibliography

(1) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the Size of Drug-like Chemical Space Based on GDB-17 Data. *J. Comput. Aided. Mol. Des.* **2013**, *27* (8), 675–679.

(2) Kohonen, T. The Self-Organizing Map. *Proc. IEEE* **1990**, *78* (9), 1464–1480.

(3) Singh, N.; Guha, R.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A.; Medina-Franco, J. L. Chemoinformatic Analysis of Combinatorial Libraries, Drugs, Natural Products, and Molecular Libraries Small Molecule Repository. *J. Chem. Inf. Model.* **2009**, *49* (4), 1010–1024.

(4) Bishop, C. M.; Svensén, M.; Williams, C. K. I. GTM: The Generative Topographic Mapping. *Neural Comput.* **1998**, *10* (1), 215–234.

(5) Kireeva, N.; Baskin, I. I.; Gaspar, H. A.; Horvath, D.; Marcou, G.; Varnek, A. Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison. *Mol. Inform.* **2012**, *31* (3–4), 301–312.

(6) Gaspar, H. A.; Baskin, I. I.; Marcou, G.; Horvath, D.; Varnek, A. GTM-Based QSAR Models and Their Applicability Domains. *Mol. Inform.* **2015**, *34* (6–7), 348–356.

(7) Kayastha, S.; Kunimoto, R.; Horvath, D.; Varnek, A.; Bajorath, J. From Bird$\blacksquare$s Eye Views to Molecular Communities: Two-Layered Visualization of Structure--Activity Relationships in Large Compound Data Sets. *J. Comput. Aided. Mol. Des.* **2017**, *31* (11), 961–977.

(8) Klimenko, K.; Marcou, G.; Horvath, D.; Varnek, A. Chemical Space Mapping and Structure--Activity Analysis of the ChEMBL Antiviral Compound Set. *J. Chem. Inf. Model.* **2016**, *56* (8), 1438–1454.

(9) Sidorov, P.; Davioud-Charvet, E.; Marcou, G.; Horvath, D.; Varnek, A. AntiMalarial Mode of Action (AMMA) Database: Data Selection, Verification and Chemical Space Analysis. *Mol. Inform.* **2018**.

(10) Kayastha, S.; Horvath, D.; Gilberg, E.; Gu?tschow, M.; Bajorath, J.; Varnek, A. Privileged Structural Motif Detection and Analysis Using Generative Topographic Maps. *J. Chem. Inf. Model.* **2017**, *57* (5), 1218–1232.

(11) Lin, A.; Horvath, D.; Afonina, V.; Marcou, G.; Reymond, J.-L.; Varnek, A. Mapping of the Available Chemical Space versus the Chemical Universe of Lead-Like Compounds. *ChemMedChem* **2018**, *13* (6), 540–554.

(12) Sidorov, P.; Gaspar, H.; Marcou, G.; Varnek, A.; Horvath, D. Mappability of Drug-like

Space: Towards a Polypharmacologically Competent Map of Drug-Relevant Compounds. *J. Comput. Aided. Mol. Des.* **2015**, *29* (12), 1087–1108.

(13) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.

(14) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2011**, *40* (D1), D1100--D1107.

(15) Standardizer, C Version 5.12. ChemAxon, Ltd: Budapest, Hungary 2012.

(16) Ruggiu, F.; Marcou, G.; Solovev, V.; Horvath, D.; Varnek, A. ISIDA Fragmentor2015-User Manual.

(17) Varnek, A.; Fourches, D.; Horvath, D.; Klimchuk, O.; Gaudin, C.; Vayer, P.; Solov'ev, V.; Hoonakker, F.; Tetko, I. V; Marcou, G. ISIDA-Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Curr. Comput. Aided. Drug Des.* **2008**, *4* (3), 191.

(18) Varnek, A.; Fourches, D.; Solov'Ev, V.; Klimchuk, O.; Ouadi, A.; Billard, I. Successful in Silico Design of New Efficient Uranyl Binders. *Solvent Extr. Ion Exch.* **2007**, *25* (4), 433–462.

(19) Golbraikh, A.; Tropsha, A. Beware of Q2! *J. Mol. Graph. Model.* **2002**, *20* (4), 269–276.