JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

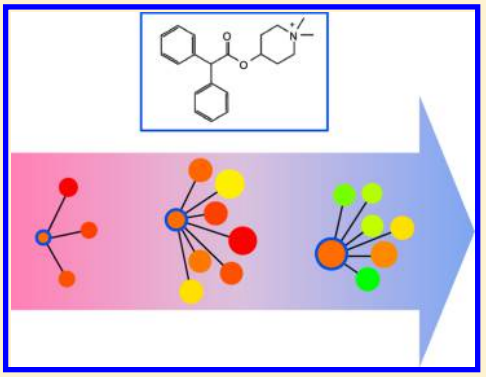
pubs.acs.org/jcim

Quantifying the Fingerprint Descriptor Dependence of Structure−Activity Relationship Information on a Large Scale

Dilyana Dimova,[†] Dagmar Stumpfe,[†] and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstr. 2, D-53113 Bonn, Germany

ABSTRACT: It is well-known that different molecular representations, e.g., graphs, numerical descriptors, fingerprints, or 3D models, change the numerical results of molecular similarity calculations. Because the assessment of structure−activity relationships (SARs) requires similarity and potency comparisons of active compounds, this representation dependence inevitably also affects SAR analysis. But to what extent? How exactly does SAR information change when alternative fingerprints are used as descriptors? What is the proportion of active compounds with substantial changes in SAR information induced by different fingerprints? To provide answers to these questions, we have quantified changes in SAR information across many different compound classes using six different fingerprints. SAR profiling was carried out on 128 target-based data sets comprising more than 60 000 compounds with high-confidence activity annotations. A numerical measure of SAR discontinuity was applied to assess SAR information on a per compound basis. For ∼70% of all test compounds, changes in SAR characteristics were detected when different fingerprints were used as molecular representations. Moreover, the SAR phenotype of ∼30% of the compounds changed, and distinct fingerprint-dependent local SAR environments were detected. The fingerprints we compared were found to generate SAR models that were essentially not comparable. Atom environment and pharmacophore fingerprints produced the largest differences in compound-associated SAR information. Taken together, the results of our systematic analysis reveal larger fingerprint-dependent changes in compound-associated SAR information than would have been anticipated.



INTRODUCTION

The assessment of structure−activity relationship (SAR) information associated with active compounds is of central relevance in medicinal chemistry.[1] A variety of computational methods are applied to aid in SAR analysis[2] including QSAR approaches,[3] visualization techniques,[4] and numerical analysis functions for SAR profiling.[5,6] SAR analysis generally requires the comparison of structures of active compounds and their potency values. Structure comparison focuses on the assessment of molecular similarity/dissimilarity, in qualitative and/or quantitative terms.[7,8] Both from a medicinal chemistry[1,7] and computational[8,9] perspective, molecular similarity is often difficult to assess.

A quantitative computational assessment of similarity requires the choice of a consistently applied molecular descriptors (representations) and a similarity or distance metric/measure.[10,11] Comparisons are typically carried out in a pairwise manner, either by quantifying the similarity of given molecular representations or by calculating the distance in chemical reference space. The molecular descriptor and, to a lesser extent, metric dependence of similarity calculations are well-appreciated caveats of molecular similarity analysis.[8−12] Similarity relationships display a tendency to change when alternative molecular representations are used, which inevitably affects similarity search calculations[12] and quantitative SAR analysis.[14] Although this dependence is widely appreciated, its potential magnitude remains unclear.

In addition to QSAR approaches focusing on individual compound series, numerical SAR analysis functions,[5,6] such as the SAR index (SARI),[5] are available for quantitative SAR exploration. SARI integrates similarity and potency comparisons of active compounds, thereby producing a numerical score that quantitatively characterizes SAR features. This characterization can be carried out both at the level of complete data sets (i.e., describing global SAR features) or compound subsets (local SARs). Different SAR feature categories can be distinguished on the basis of numerical SAR analysis. For example, SAR continuity refers to the presence of structurally diverse compounds having comparable potency. The presence of continuous SARs provides the basis of scaffold hopping[13] in similarity searching and virtual screening.[9−12] By contrast, SAR discontinuity refers to the presence of structurally similar compounds with a large potency differences.[5,14] The extreme form of SAR discontinuity is represented by activity cliffs[14−17] that are rationalized as pairs of similar compounds (structural analogs) with large potency differences. For the analysis of activity cliffs, the definition of similarity and potency difference threshold values is typically required.[16] For large-scale SAR analysis using numerical analysis functions, different types of fingerprints,[9−12] i.e., bit representations of molecular structure and properties, are

Received: July 11, 2013
Published: August 22, 2013

dx.doi.org/10.1021/ci4004078 | J. Chem. Inf. Model. 2013, 53, 2275−2281

generally preferred molecular representations and most widely used for similarity calculations.[5,6,14]

It is currently unknown to what extent active compounds across different targets might change their SAR phenotype, e.g., from a continuous to a discontinuous one and vice versa, when different fingerprints are used as molecular representations. Obtaining this information would provide a first quantification of fingerprint-dependent changes in SAR characteristics on a per compound basis and help to estimate the influence of these effects on the reliability of computational SAR models. Therefore, we have systematically analyzed 128 compound sets with high-confidence activity data using different fingerprints to quantitatively describe and compare changes in SAR information at the level of individual compounds.

MATERIALS AND METHODS

Compound Data Sets. Data sets were assembled from ChEMBL (version 15).[18] For our analysis, only compounds with precisely specified equilibrium constants (K_i values) below 1 μM for human targets at the highest confidence level (confidence score 9) were selected. A compound with multiple measurements for the same target was only considered if it had consistent potency measurements (i.e., if all values fell within 1 order of magnitude). In this case, the average potency was calculated as the final annotation. In addition, each data set was required to contain at least 100 compounds. On the basis of these selection criteria, 128 target-based data sets comprising a total of 60 248 compounds were obtained. The data sets are made freely available via the downloads section of following: http://www.lifescienceinformatics.uni-bonn.de.

Fingerprints. Six fingerprints of different design and complexity were used as molecular representations:

(a) Extended connectivity fingerprint with bond diameter 4 (ECFP4)[19] producing ~4×10^9 theoretically possible features. ECFP4 is a topological fingerprint that encodes layered atom environments with a maximum diameter of four bonds around each atom.

(b) Functional class fingerprint with bond diameter 4 (FCFP4) is a closely related derivative of ECFP4 that replaces atom types with pharmacophore features (such as hydrogen-bond donors and acceptors, charged, or aromatic atoms), which renders atom information less specific compared to ECFP4 and reduces the size of feature sets.

(c) Molecular access system (MACCS)[20] consists of 166 structural fragments with 1−10 nonhydrogen atoms.

(d) Typed graph distances (TGD)[21] is an atom pair-type fingerprint consisting of 420 bits. Shortest distances in the molecular graph between two atoms (represented as seven pharmacophore features) are calculated and assigned to 15 distance ranges.

(e) Typed graph triangles (TGT)[21] contains 1704 bits positions representing three-point pharmacophore patterns in molecular graphs. Each atom is assigned to one of four atom types (hydrogen-bond donor and acceptor, donor/acceptor, or hydrophobic), and six bond distance ranges are applied.

(f) Donor−acceptor polar-hydrophobe triangle (GpiDAPH3)[21] is a molecular graph-based three-point pharmacophore fingerprint generating 30 240 possible features. In this case, each atom is assigned to one of eight atom types derived from three atomic properties (π system, donor, or acceptor), and eight bond distance ranges are utilized.

All fingerprint representations were calculated using the Molecular Operating Environment (MOE).[21] The TGD, TGT, and GpiDAPH3 fingerprints are MOE-internal developments.

SARI Discontinuity Score. SARI is a numerical SAR analysis function to quantify SAR features of compound data sets and consists of separate terms accounting for SAR continuity and discontinuity, respectively.[5] We have used the discontinuity score component as a quantitative measure of per compound SAR discontinuity. The raw (non-normalized) discontinuity score is defined as[5]

$$\mathrm{raw}_{\mathrm{disc}} = \frac{\sum_{\{i,j|\mathrm{sim}(i,j)>x, i>j\}} |\mathrm{pot}(i) - \mathrm{pot}(j)| \cdot \mathrm{sim}(i,j)}{|\{i,j|\mathrm{sim}(i,j) > x,\ i > j\}|}$$

Here, pot(i) is the potency (pK_i) value of compound i and sim(i,j) the calculated fingerprint similarity for compounds i and j. The raw discontinuity score is derived from the average of potency differences between pairs of ligands multiplied by their similarity. Accordingly, the score emphasizes the presence of structurally similar compounds with large potency differences. Therefore, it is advisable to limit discontinuity scoring to compounds that have at least limited molecular similarity.[5] Hence, a similarity threshold value (x) is usually applied, as further discussed below.

Depending on the normalization procedure, raw discontinuity scores can be converted into scores that either quantify the degree of SAR discontinuity for a complete data set, compound subsets, or individual compounds.[5,22,23] For our analysis, we utilized per compound discontinuity scores to assess local SAR discontinuity. Therefore, for each fingerprint, raw discontinuity scores were systematically calculated for all 128 classes. The raw scores were then converted into Z-scores using the sample mean and standard deviation of the score distribution for each individual data set, i.e., the intraclass score distribution for each fingerprint. Finally, cumulative probabilities were calculated to map scores onto the value range [0, 1].[5] All similar compound pairs were taken into consideration, regardless of potency differences, to fully account for the structural neighborhood of each individual compound,[22,23] a prerequisite for assessing local SAR discontinuity. As shown herein, per compound discontinuity scores provide a meaningful measure of compound-associated SAR information and make it possible to characterize local SAR environments in compound data sets in detail.

Similarity Calculation and Threshold Correspondence. Pairwise fingerprint similarity was quantified using the Tanimoto coefficient (Tc).[11] As a reference point for the discontinuity score similarity threshold, a MACCS Tc of 0.70 was applied, which typically indicates remote similarity.[12] Compound pairs yielding further increasing MACCS Tc values become increasingly similar (as an activity cliff criterion, a MACCS Tc of 0.85 is often applied).[16] Across all 128 data sets, a MACCS Tc of 0.70 yielded 12% of all possible compound pairs meeting or exceeding this threshold.

In order to determine corresponding Tc threshold values for all fingerprints yielding the same number of similar compound pairs, a three-step procedure was applied: (a) For each of the 128 data sets, the number of compound pairs yielding a MACCS Tc $\geq$ 0.70 was determined; (b) for each of the remaining fingerprints, the Tc threshold was determined that yielded the same number of compound pairs for each set; and

(c) for each fingerprint, the median of its 128 individual threshold values was calculated and used as its global threshold.

On the basis of this analysis, the following corresponding Tc threshold values were derived: MACCS, 0.70; ECFP4, 0.31; FCFP4, 0.38; TGD, 0.69; TGT, 0.65; GpiDAPH3, 0.15. These thresholds were applied for the calculation of per-compound discontinuity scores.

RESULTS AND DISCUSSION

Study Concept and Goals. We have aimed to quantify the influence of different fingerprint descriptors on compound-associated SAR information on a large scale across many different targets. To our knowledge, this is the first study that systematically and quantitatively assesses fingerprint-dependent changes in SAR information content at the level of individual active compounds. To these ends, we have selected more than 60 000 compounds with high-confidence activity annotations for 128 targets and calculated pairwise similarities using six fingerprints of different design and complexity. For these

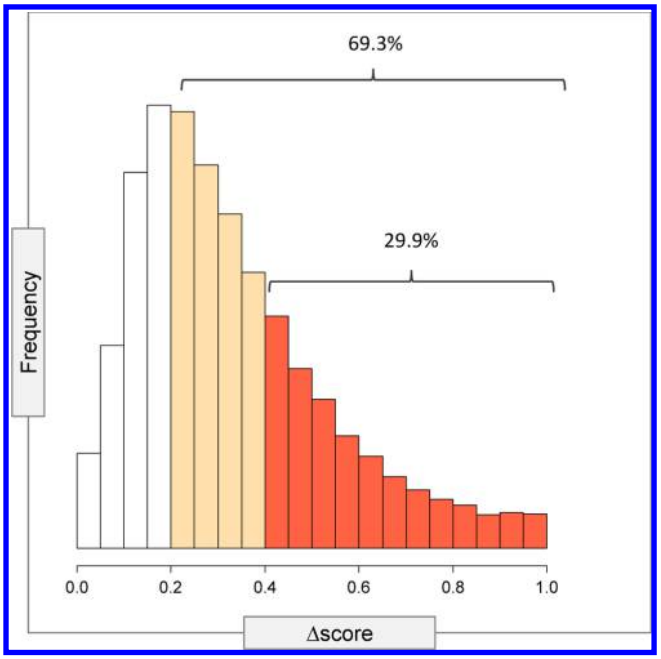


Figure 1. Distribution of pairwise score differences. For each compound, discontinuity score differences are calculated for all pairs of fingerprints, yielding a total of 15 score differences for an individual compound. The histogram reports the distribution of the maximal score differences for all 60 248 active compounds.

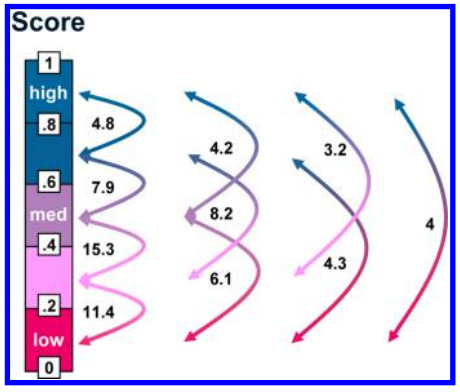


Figure 2. Discontinuity score shifts. Five score ranges ([0, 0.2], (0.2, 0.4], (0.4, 0.6], (0.6, 0.8], (0.8, 1]) are defined. Ranges of low, medium, and high discontinuity are marked. Arrows represent largest observed score differences and are labeled with the percentage of compounds displaying such fingerprint-dependent differences.

Table 1. Distribution of Score Differences for Fingerprint Combinations[a]

fingerprint combination	Δscore ≥ 0.2	Δscore ≥ 0.4
ECFP4/FCFP4	12.2	3.8
ECFP4/GpiDAPH3	20.5	6.9
ECFP4/MACCS	20.0	6.9
ECFP4/TGT	28.6	11.2
ECFP4/TGD	31.1	12.7
FCFP4/GpiDAPH3	20.1	7.0
FCFP4/MACCS	20.7	7.4
FCFP4/TGT	24.6	9.7
FCFP4/TGD	26.7	10.9
GpiDAPH3/MACCS	25.6	9.1
GpiDAPH3/TGT	25.8	9.9
GpiDAPH3/TGD	27.0	10.7
MACCS/TGT	28.9	11.7
MACCS/TGD	30.2	12.8
TGT/TGD	18.8	7.9

[a]For all possible combinations of fingerprints, the percentage of compounds with discontinuity score differences (Δscore) of at least 0.2 or 0.4 is reported as an average over all classes.

Table 2. Number of Fingerprint Combinations Producing Large Score Changes[a]

	Δscore ≥ 0.2			Δscore ≥ 0.4		
no. FP combinations	no. cpds	% cpds	avg. no. neigh	no. cpds	% cpds	avg. no. neigh
0	18 509	30.7	108.2	42 217	70.1	94.1
1	4155	6.9	96.6	3639	6.0	66.5
2	4079	6.8	88.5	3045	5.1	60.9
3	3655	6.1	82.3	2358	3.9	55.1
4	4077	6.8	78.9	2255	3.7	54.0
5	7126	11.8	64.3	2822	4.7	38.4
6	3459	5.7	70.1	1044	1.7	49.1
7	3249	5.4	64.9	795	1.3	41.7
8	4774	7.9	58.2	1445	2.4	30.0
9	3552	5.9	49.7	561	0.9	24.9
10	1671	2.8	48.6	52	0.1	19.4
11	1510	2.5	37.5	15	0.0	21.3
12	360	0.6	32.2	0	0	0
13	71	0.1	23.6	not possible		
14	1	0.0	3.2	not possible		
15	0	0	0	not possible		

[a]The number of fingerprint (FP) combinations yielding a score difference (Δscore) of at least 0.2 and 0.4 is determined for all compounds. For all combinations, the corresponding number and percentage of compounds (cpds) are reported. In addition, the average number of structural neighbors (no. neigh) calculated for all fingerprints (no. FP = combinations 0) and for only those fingerprints that participate in pairs with Δscore ≥ 0.2 or Δscore ≥ 0.4 (no. FP combinations ≥1) are reported. It should be noted that the maximal number of fingerprint combinations that can yield Δscore ≥ 0.2 and Δscore ≥ 0.4 is 15 and 12, respectively. Hence, it is not possible that all numbers of combinations of six fingerprints can meet the Δscore ≥ 0.4 condition. The three numbers of combinations for which Δscore ≥ 0.4 cannot be obtained (13, 14, and 15) are designated as "not possible".

dx.doi.org/10.1021/ci4004078 | J. Chem. Inf. Model. 2013, 53, 2275−2281

Table 3. Compound Neighborhood Statistics for Different Fingerprint Combinations[a]

	Δscore ≥ 0.2		Δscore ≥ 0.4	
no. FP combinations	category A	category B	category A	category B
0	2.9	97.1	2.3	97.6
1	1.1	98.9	4.4	95.5
2	1.6	98.4	6.1	93.9
3	2.1	97.9	8.2	91.8
4	3	97.0	10.4	89.6
5	9.9	90.2	29	70.9
6	4.2	95.8	15.8	84.2
7	4.6	95.4	23.2	76.8
8	11.5	88.5	39.2	60.7
9	14.3	85.7	46.5	53.4
10	13.5	86.5	62.4	37.7
11	25.3	74.7	69.1	30.9
12	30.7	69.3	0	0
13	42.4	57.6	not possible	
14	78.5	21.4	not possible	
15	0	0	not possible	

[a]The number of fingerprint combinations yielding a score difference (Δscore) of at least 0.2 and 0.4 is reported for all compounds. For each pair of fingerprints, the fingerprint-dependent compound neighborhoods are determined and classified as distinct (i.e., two fingerprints produce different sets of structural neighbors; category A) or overlapping (i.e., two fingerprints produce a subset of shared structural neighbors; category B). In addition, the average percentage of distinct and overlapping neighborhoods calculated for all fingerprints (no. FP combinations = 0) and for only those fingerprints that participate in pairs yielding Δscore ≥ 0.2 or Δscore ≥ 0.4 (no. FP combinations ≥1) are reported. Three numbers of fingerprint combinations for which Δscore ≥0.4 cannot be obtained (see legend of Table 2) are designated as "not possible".

fingerprints, corresponding Tc threshold values were determined to ensure that the same proportion of all possible compound pairs was classified as similar.

SAR discontinuity is an indicator of SAR information content.[14,15] In order to account for SAR discontinuity in a consistent manner, a SAR discontinuity scoring scheme was applied that quantifies the discontinuity contributions of individual compounds.[14] The latter criterion is of critical relevance for our current analysis. What does local SAR discontinuity mean? Local SAR discontinuity is high if a compound has a potency value that substantially deviates from the potency of its structural neighbors.

Through systematic assessment of compound-centric SAR discontinuity, fingerprint-dependent changes in SAR information were quantified for all test compounds and local SAR environments were analyzed.

Distribution of Discontinuity Score Differences. For each compound, discontinuity scores using all 6 fingerprints and score differences for all possible 15 pairwise fingerprint combinations were calculated. In Figure 1, the distribution of the maximal score differences per compound is reported. For ∼70% and ∼30% of all compounds, maximal score differences (Δscore) ≥0.2 and ≥0.4 were detected, respectively. Discontinuity score differences ≥0.2 are substantial, and score differences ≥0.4 indicate a major change in SAR information associated with a given compound. This is the case because a score change of magnitude 0.4 or larger transforms a compound with low SAR discontinuity into one with intermediate or high discontinuity and vice versa, as further discussed below. Thus, for ∼30% of all active compounds across the 128 target-based data sets, large-magnitude changes in SAR information were observed that altered their SAR phenotype, giving rise to fingerprint-dependent SAR phenotype switches, which were quantified for the first time. Overall, an unexpectedly high proportion of compounds active against 128 different targets was found to change their SAR phenotype when different fingerprints were used. On the basis of fingerprint similarity value distributions,[24] a much smaller proportion of compounds would have been predicted. This meant that the structural neighborhood of many compounds substantially changed when different fingerprints were used, thereby completely changing many local SAR environments, as further discussed below. In Figure 2, the range distribution of score differences is reported. Differences were not confined to certain score subranges. Rather, shifts of varying magnitudes across the entire discontinuity score range were observed.

Fingerprint Comparison. In Table 1, the proportion of compounds with discontinuity score differences ≥0.2 or ≥0.4 is reported for all 15 pairwise fingerprint combinations, ranging from12.2−31.1% (score difference ≥0.2) and 3.8−12.8% (≥0.4). All fingerprint combinations were found to induce substantial discontinuity score differences including closely related fingerprints, such as ECFP4/FCFP4, for which 12.2% and 3.8% of the compounds displayed score differences ≥0.2 and ≥0.4, respectively. The comparison made it possible to identify combinations of fingerprint types that generated inconsistent SAR models. A major finding has been that atom environment fingerprints and 2D pharmacophore fingerprints produced largely different SAR models. For example, the ECFP4/TGD combination yielded score differences ≥0.2 and ≥0.4 for 31.1% and 12.7% of active compounds, respectively. On the basis of the fingerprint pair-based score differences in Table 1, all 15 fingerprint combinations were considered for further analysis of discontinuity score distributions.

Score Differences and Fingerprint Combinations. Table 2 reports the number of fingerprint combinations that produced large score differences and the corresponding numbers of compounds. In addition, the average number of neighbors (meeting or exceeding the fingerprint similarity thresholds) of these compounds is reported. For 30.7% and 70.1% of all compounds, no fingerprint combination generated score differences ≥0.2 and ≥0.4, respectively. These data are consistent with the score difference distribution in Figure 1. However, 1−11 different fingerprint combinations yielded score differences ≥0.2 for thousands of compounds each and 1−8 combinations differences ≥0.4 for comparable numbers of compounds. Thus, multiple fingerprint combinations were generally responsible for substantial changes in SAR discontinuity, and large proportions of active compounds experienced large changes in their structural neighborhoods for multiple fingerprints.

Fingerprint Combinations and Compound Neighborhoods. Table 2 also reveals a steady decrease in the number of structural neighbors of compounds for which increasing numbers of fingerprint combinations generated large changes in SAR discontinuity. For compounds that did not yield score differences ≥0.2 and ≥0.4, on average more than 100 and 90 structural neighbors were detected, respectively. Thus, these compounds were very similar to many others. By contrast, for compounds for which increasing numbers of fingerprint combinations led to substantial score changes, decreasing numbers of structural neighbors were detected. For example,

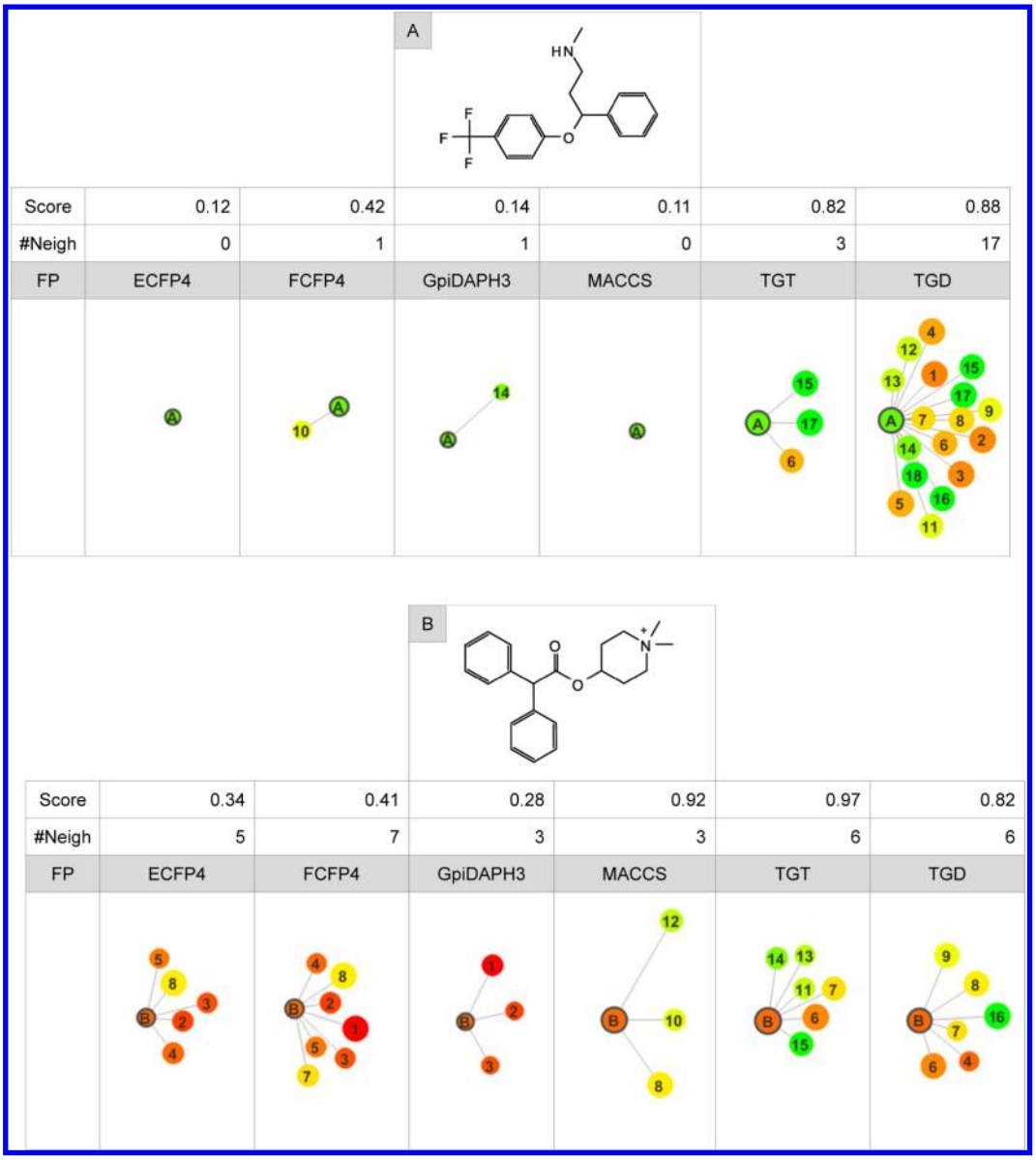


Figure 3. Local SAR environments. In (A) and (B), the chemical neighborhood of two exemplary compounds in their data sets is shown when similarity relationships are calculated using the six different fingerprints. Compounds A and B are muscarinic acetylcholine receptor M5 antagonists (ChEMBL ID 2035). Sections of network-like similarity graphs (NSGs)[23] are displayed representing the fingerprint-dependent neighborhoods of each compound. In NSGs, compounds are represented as nodes and edges similarity relationships. For each fingerprint, the corresponding Tc similarity threshold values reported in the Materials and Methods section are applied as a similarity (edge) criterion. Nodes are numbered, color-coded by compound potency using a color spectrum ranging from red (high) over yellow (intermediate) to green (low potency) and scaled in size by local discontinuity scores. For compounds A and B and each fingerprint, the discontinuity score and the number of structural neighbors (#Neigh) applying the fingerprint-dependent Tc threshold values are reported.

for compounds whose discontinuity scores were notably changed by eight different fingerprint combinations, on average only 58 (Δscore $\geq$0.2) and 30 (Δscore $\geq$0.4) neighbors were identified (Table 2). Thus, compounds whose SAR characteristics were increasingly fingerprint-dependent formed fewer similarity relationships than other compounds; another unexpected finding.

Compound neighborhoods are further analyzed in Table 3. Most of the compound neighborhoods generated with different fingerprints were overlapping (Category B). However, for compounds for which multiple fingerprint combinations produced substantial score changes, increasing numbers of distinct neighborhoods (Category A) were observed, i.e., neighborhoods that did not share any compounds. For example, for compounds whose scores were significantly altered by eight different fingerprint combinations, on average ~12% (Δscore $\geq$0.2) and ~39% (Δscore $\geq$0.4) of their neighborhoods generated with these fingerprints were distinct (Table 3). Thus, different fingerprints often produced nonoverlapping similarity relationships that led to large changes in local SAR discontinuity.

Local SAR Environments. In Figure 3, local SAR environments centered on exemplary compounds are compared for different fingerprints with the aid of similarity-based molecular networks,[23] which illustrate changes in local SAR information. In Figure 3A, an active compound is shown for which two fingerprints (ECFP4, MACCS) did not yield structural neighbors. Thus, in these cases, the compound was

dx.doi.org/10.1021/ci4004078 | J. Chem. Inf. Model. 2013, 53, 2275−2281

a singleton carrying essentially no SAR information. In addition, two other fingerprints (FCFP4, GpiDAPH3) identified only a single neighbor. By contrast, TGT and especially TGD detected multiple neighbors with varying potency. Thus, in these local environments, the test compound introduced a high degree of SAR discontinuity. Discontinuity scores for this compound varied from 0.11 (MACCS) to 0.88 (TGD). In the TGD-dependent neighborhood, the compound formed multiple activity cliffs (combinations of large red and green nodes in Figure 3A) and hence obtained a high discontinuity score. Thus, the SAR character of this compound and its associated SAR information content completely changed dependent on the fingerprint that was used.

Furthermore, in Figure 3B, another exemplary compound is shown for which different fingerprints consistently detected multiple neighbors. The fingerprint-dependent neighborhoods were partly overlapping and distinct and resulted in very different local SAR environments, reflected by either low or high discontinuity scores ranging from 0.28 (MACCS) to 0.97 (TGT). For MACCS, TGD, and TGT, the test compound formed increasing numbers of activity cliffs in its neighborhood and was thus highly discontinuous. Thus, in these cases, this compound would be assigned a key compound for SAR analysis. By contrast, in the ECFP4-, FCFP4-, and GpiDAPH3-dependent environments, most structural neighbors had potency values very similar to the test compound, which resulted in low/intermediate discontinuity scores. In these cases, the compound would not be considered as a focal point of SAR analysis. Thus, the SAR characteristic of the compound in Figure 3B also fundamentally changed with different fingerprint representations.

The exemplary compounds in Figure 3 had score differences $\geq$0.4 for eight (compound A) and nine (B) different fingerprint combinations, respectively. Score differences were in seven (A) and four (B) cases due to the presence of distinct neighbors and in one (A) and five (B) cases due to partly overlapping fingerprint-dependent neighborhoods. Thus, the local SAR environments of these exemplary compounds often fundamentally differed. These findings illustrate that a very different conclusion concerning local SAR features in compound data sets might be drawn when alternative fingerprints are used.

CONCLUSIONS

Herein we have reported a first quantitative and large-scale analysis of fingerprint-dependent changes in SAR information associated with ~60 000 compounds active against 128 different targets. Different types of 2D fingerprints were compared including structural fragment, atom environment, and pharmacophore fingerprints. The analysis was facilitated by a consistent quantitative assessment of per compound SAR discontinuity and revealed a number of unexpected findings, as summarized in the following: For ~70% of all compounds, substantial changes in SAR discontinuity were detected, and for ~30% of all compounds, different numbers of fingerprint combinations were found to change their SAR phenotype; a much larger proportion of compounds that would have been estimated on the basis of similarity score distributions. Furthermore, we have found that compounds forming below-average numbers of similarity relationships to others displayed a particularly strong fingerprint dependence of their local SAR environments. Moreover, we have determined that in nearly all of the combinations of fingerprints we analyzed (except the most closely related ones such as ECFP4 and FCFP4), we led to large differences in compound-associated SAR information, more so than we had anticipated. Atom environment and pharmacophore fingerprints produced largest differences in SAR information. A major conclusion from our study is that SAR phenotypes of many compounds change when alternative fingerprints are used. Thus, ensuing SAR differences are not gradual but often fundamental. This makes it essentially impossible to draw general conclusions concerning local SARs and individual compounds. Key compounds representing centers of SAR discontinuity in data sets identified with a given fingerprint might not be detected when other fingerprints are used. In SAR analysis, awareness should be raised concerning the unexpectedly large magnitude of these effects and the resulting inconsistency of many SAR models produced by widely used fingerprints. In the future, the current analysis might be further extended to alternative sets of numerical descriptors, which are typically preferred for QSAR modeling.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bajorath@bit.uni-bonn.de

Author Contributions

†These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

D.S. is supported by Sonderforschungsbereich 704 of the Deutsche Forschungsgemeinschaft.

REFERENCES

(1) Wermuth, C. G. *The Practice of Medicinal Chemistry*, 3rd ed.; Academic Press: Waltham, MA, 2008.

(2) Wawer, M.; Lounkine, E.; Wassermann, A. M.; Bajorath, J. Data Structures and Computational Tools for the Extraction of SAR Information from Large Compound Sets. *Drug Discovery Today* **2010**, *15*, 631−639.

(3) Esposito, E. X.; Hopfinger, A. J.; Madura, J. D. Methods for Applying the Quantitative Structure-Activity Relationship Paradigm. *Methods Mol. Biol.* **2004**, *275*, 131−214.

(4) Stumpfe, D.; Bajorath, J. Methods for SAR Visualization. *RSC Adv.* **2012**, *2*, 369−378.

(5) Peltason, L.; Bajorath, J. SAR Index: Quantifying the Nature of Structure-Activity Relationships. *J. Med. Chem.* **2007**, *50*, 5571−5578.

(6) Guha, R.; Van Drie, J. H. Structure-Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48*, 646−658.

(7) Kubinyi, H. Similarity and Dissimilarity: A Medicinal Chemist's View. *Perspect. Drug Discovery Des.* **1998**, *9*−*11*, 225−232.

(8) *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; John Wiley & Sons: New York, 1990.

(9) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225−233.

(10) Willett, P. Searching Techniques for Databases of Two- and Three-Dimensional Structures. *J. Med. Chem.* **2005**, *48*, 4183−4199.

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

(12) Stumpfe, D.; Bajorath, J. Similarity Searching. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 260−282.

(13) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. Scaffold-Hopping by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem., Int. Ed.* **1999**, *19*, 2894−2896.

(14) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure−Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209−8223.
(15) Maggiora, G. M. On Outliers and Activity Cliffs − Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535−1535.
(16) Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.
(17) Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of Activity Landscapes using 2D and 3D Similarity Methods: Consensus Activity Cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477−491.
(18) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2011**, *40*, D1100−D1107.
(19) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754.
(20) *MACCS Structural Keys*; Accelrys: San Diego, CA.
(21) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: Montreal, Quebec, Canada.
(22) Peltason, L.; Hu, Y.; Bajorath, J. From Structure-Activity to Structure-Selectivity Relationships: Quantitative Assessment, Selectivity Cliffs, and Key Compounds. *ChemMedChem.* **2009**, *4*, 1864−1873.
(23) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. Structure-Activity Relationship Anatomy by Network-like Similarity Graphs and Local Structure-Activity Relationship Indices. *J. Med. Chem.* **2008**, *51*, 6075−6084.
(24) Vogt, M.; Stumpfe, D.; Geppert, H.; Bajorath, J. Scaffold Hopping Using Two-dimensional Fingerprints: True Potential, Black Magic, or a Hopeless Endeavor? Guidelines for Virtual Screening. *J. Med. Chem.* **2010**, *53*, 5707−5715.