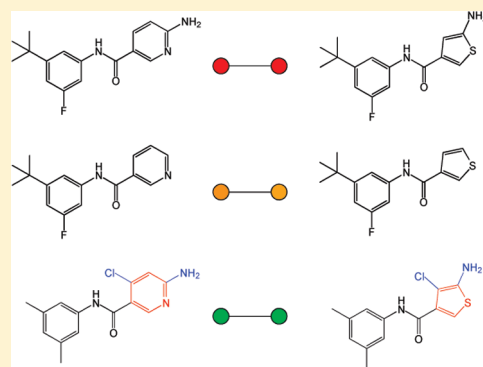


A Data Mining Method To Facilitate SAR Transfer

Anne Mai Wassermann and Jürgen Bajorath*

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

ABSTRACT: A challenging practical problem in medicinal chemistry is the transfer of SAR information from one chemical series to another. Currently, there are no computational methods available to rationalize or support this process. Herein, we present a data mining approach that enables the identification of alternative analog series with different core structures, corresponding substitution patterns, and comparable potency progression. Scaffolds can be exchanged between these series and new analogs suggested that incorporate preferred R-groups. The methodology can be applied to search for alternative analog series if one series is known or, alternatively, to systematically assess SAR transfer potential in compound databases.



INTRODUCTION

The chemical lead optimization process involves simultaneous improvement of multiple properties such as compound potency, selectivity, oral availability, or metabolic stability and is considered one of the most challenging tasks in medicinal chemistry.^{1,2} Often considered more of an art form than a well-defined scientific exercise,² lead optimization typically depends on chemical expertise, experience, and intuition and publications describing and rationalizing lead optimization processes are relatively rare.² During hit-to-lead and lead optimization efforts, SAR information is sampled by designing, generating, and testing series of analogs and alternative starting points are often considered. There are many possible complications along the path to developing viable candidate compounds, and it is not uncommon that a compound series displaying otherwise promising SAR behavior hits a roadblock, perhaps due to metabolic liability or unwanted side effects, which then prevents its further development.

In such situations, one would ideally like to build upon prior knowledge, utilize available SAR information, and evaluate the possibility of an “SAR transfer”, i.e., the exploration of an alternative chemotype that displays similar SAR characteristics and potency progression. Accordingly, one would search for alternative molecular core structures (scaffolds) where corresponding chemical substitutions yield comparable SAR trends (consistent with a conserved mechanism of action) but circumvent liabilities associated with the original chemotype. However, facilitating such SAR transfers goes far beyond the identification of alternative scaffolds displaying a specific biological activity,³ because it requires the identification of corresponding analog series with equivalent SAR characteristics. In addition to replacing one chemical series by another, SAR transfer also has other facets. For example, if parallel compound series with varying degrees of chemical exploration (e.g., different numbers of

analogs, different potency levels) would be available, one might learn more about SAR progression than on the basis of a single series. In addition, it might be possible to suggest potent analogs for one series based on another, whose other lead-relevant properties could then be compared.

Although SAR transfer is a common task in practical medicinal chemistry, there are currently no computational methods available to aid in this process. PubMed and Google Scholar searches currently reveal no publications that address the issue of SAR transfer from a computational point of view. As a first step in this direction, we report herein a data mining method that is designed to (i) identify chemical series with SAR transfer potential if one analog series is available as a starting point and (ii) detect all SAR transfer events that occur in compound databases.

MATERIALS AND METHODS

Scaffold Generation. Scaffolds were defined to consist of ring systems and linkers between them⁴ and obtained from compounds by removal of all R-groups from rings and linkers, utilizing a Pipeline Pilot implementation.⁵ Different from the original scaffold definition,⁴ terminal atoms forming exocyclic double bonds to ring atoms (mostly carbonyl oxygens) were also considered part of a scaffold and not removed.

Scaffold Exchanges. Pairs of scaffolds with changes in one ring structure were identified using an in-house-designed variant of the matched molecular pair (MMP)⁶ search algorithm of Hussain and Rea.⁷ The procedure is as follows: All contiguous ring systems in a scaffold are separately removed by deleting all connecting bonds between the ring system and the remaining

Received: June 8, 2011

Published: July 20, 2011

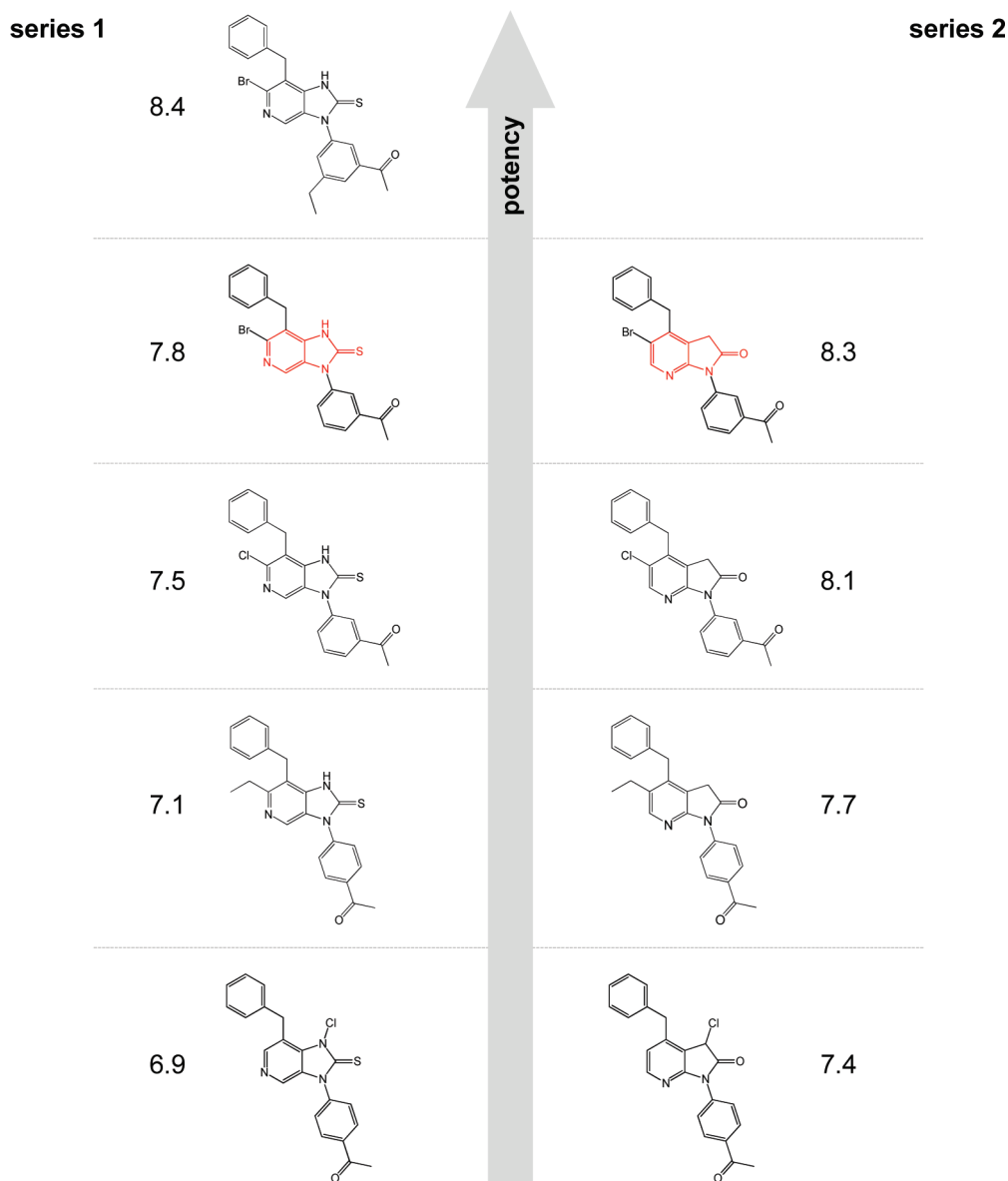


Figure 1. SAR transfer. The principal idea of SAR transfer is illustrated. Two structurally related model analog series are shown. The ring systems that distinguish the two corresponding scaffolds are highlighted in red in a compound of each series. Compounds in both series that carry the same substituents and only differ in the exchanged rings are aligned, i.e., they form pairs of corresponding analogs. The alignment organizes compounds (labeled with their pK_i values) in the order of increasing potency. For potency-based ordering, one series serves as a reference. In this case, the two ordered series display a steady potency progression and potency differences between analogs in pairs are comparable, although the absolute potency values differ. Hence, the alignment represents a prototypic SAR transfer between two analog series.

structure. Connectivity information for the resulting fragments is retained by marking the attachment points. According to Hussain and Rea, in each fragment combination the removed ring system is referred to as value fragment and the remaining part(s) of the scaffold as key fragment(s). Here, the value fragment is annotated with a scaffold identifier and key and value fragments are represented as canonical SMILES⁸ strings to assemble a key–value pair in an index table. For two scaffolds that only differ by a single ring, removal of the exchanged ring structures leads to identical key fragments. Hence, this scaffold pair can be identified from the index table. After identification of a scaffold pair, all compounds represented by the two scaffolds are also fragmented by removing the distinguishing ring and attached R-groups. The removed fragment is then further decomposed into the invariant

ring structure and the R-groups, which are marked with the numeric identifier of the ring atom to which they were attached. Hence, each molecule of a series can be unambiguously represented by the combination of the residual substructure after ring removal (key fragment) and the set of R-groups. Corresponding compounds of two analog series are required to have identical key fragments and R-groups. These R-groups are used to define corresponding substitution sites in the exchanged ring systems. To determine corresponding substitution sites all possible mappings between R-group positions are explored. All compound pairs yielding the same R-group mapping are combined into two “matching series”.

Potency Scaling. For all compounds represented by a given scaffold within a matching series, the mean potency is

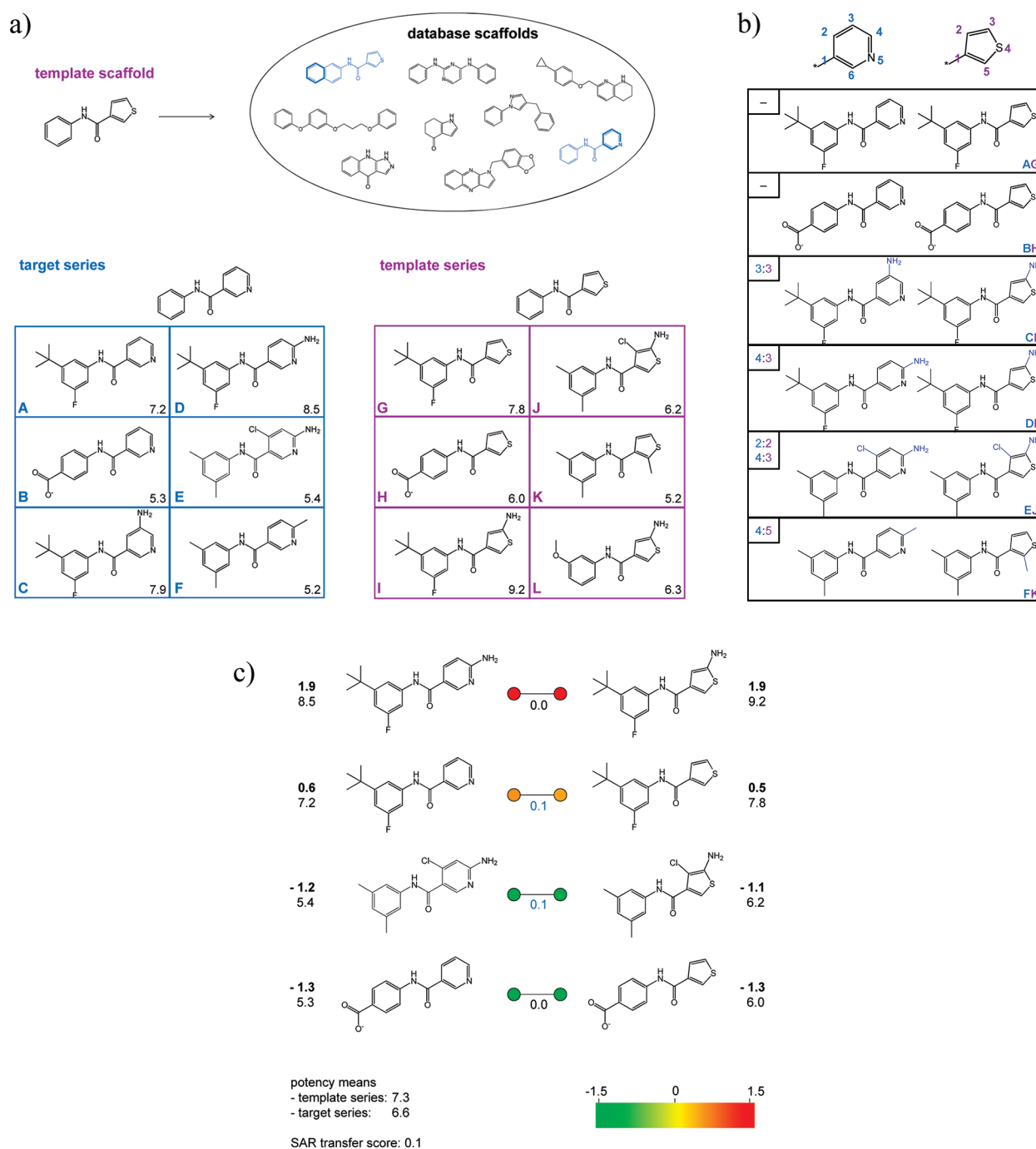


Figure 2. SAR transfer methodology. The figure outlines the approach to facilitate SAR transfer. (a) Scaffold detection. A scaffold representing a template series is used to search a database for scaffolds that differ by the replacement of a ring system. Two scaffolds meeting this structural criterion are highlighted in blue. For one of these scaffolds, all analogs are assembled representing a target series. In this example, both the template and the target series contain six analogs. These compounds are consecutively named A–L. The pK_i value of each analog is reported. (b) Pairwise analog assembly. Pairs of analogs belonging to the template and target series are identified that show corresponding R-group patterns. In order to establish substitution site correspondences for exchanged rings of different size and composition, different mappings of ring positions are systematically explored. Analog pairs are named according to a, and mappings of ring substituent positions are reported (e.g., 3:3). The hyphen (–) indicates the absence of substituents at the exchanged rings. In this example, position three of the thiophene ring is mapped to alternative positions of the pyridine ring in different analog pairs. The mapping to position four of the pyridine is prioritized according to the criteria reported in the text. (c) Analog pair alignment. For the preferred mapping, all pairs of analogs are selected. Corresponding compounds of the target (left) and template (right) series are vertically aligned and represented by nodes (connected by an edge). Analog pairs are ranked in the order of increasing potency of the target series. The mean potency for the four aligned compounds from each series is calculated, and for each compound, the potency difference from the mean (relative potency) is calculated. The compounds are annotated with their relative (bold) and absolute potency values. Nodes are color coded by relative potency using a continuous spectrum from green (via yellow) to red. Green indicates lowest and red highest relative potency in a series. Edges are labeled with differences between rescaled potencies of corresponding compounds (nodes), and maximal differences representing the SAR transfer score are highlighted in blue.

determined and potency values are rescaled with respect to the mean. Hence, zero-centered potency values for all analogs are calculated by subtracting the mean from actual compound potency (in pK_i units). For visual inspection of SAR transfer potential, corresponding compounds in two matching series are ordered by increasing potency of one analog series and displayed as colored nodes connected by an edge. A uniform color code is applied for all series by using a green-to-red color spectrum that represents potency differences from the mean within the range from -1.5 to 1.5 pK_i units around the mean. Potency differences falling below or above this range are represented in green and red, respectively.

SAR Transfer Score. For a quantitative assessment of SAR transfer potential, absolute differences between rescaled potencies of all compound pairs in matching series are calculated. The maximal difference between two compounds in a pair is utilized as an SAR transfer score because it describes the observed deviation of matching series from an ideal SAR transfer scenario. Accordingly, a score of 0 corresponds to perfect SAR transfer with identical potency progressions between ordered pairs of analogs in matching series, whereas high scores indicate a substantial discrepancy in potency progression.

Data Sets. As a source database for our analysis, BindingDB⁹ was used. All ring-containing compounds with available K_i values for human targets were extracted from BindingDB. For all molecules with multiple potency measurements against the same target, the arithmetic mean was calculated to yield a final potency value, unless reported K_i values spanned a potency range of more than 1 order of magnitude. In this case, the target activity was excluded from the analysis. A total of 53 760 different qualifying compounds were selected and organized into 708 target sets. Then target set compounds were grouped by scaffolds containing at least two ring systems.

Implementation. The MMP algorithm and its modifications were implemented in Java using the OpenEye chemistry toolkit.¹⁰ Determination of matching ring positions for compounds in two different analog series and the potency-based ordering of compound pairs as well as score calculations were carried out with in-house-generated Perl programs.

RESULTS AND DISCUSSION

SAR Transfer Events. Figure 1 illustrates an SAR transfer scenario that emphasizes both the alternative series and learning aspects. In general, to qualify as an SAR transfer event, matching series must consist of structurally corresponding analog pairs with comparable potency differences between ordered analogs in each set. Matching series can differ in their absolute potency values, but potency differences between ordered analogs in each series should be comparable. In the hypothetical example in Figure 1, compound series 1 is utilized as a template and a database search yielded series 2. Both series contain distinct core structures and consist of analogs with pairwise corresponding R-group patterns that show a comparable increase in potency. Thus, the SARs of series 1 and 2 are essentially interchangeable and the two series represent a prototypic SAR transfer model. Moreover, series 1 contains a potent analog that has no counterpart in series 2, and hence, the corresponding analog might be suggested for synthesis. This illustrates potential learning aspects associated with detected SAR transfer events. Although chemotype replacement might be considered the primary task of SAR transfer, the comparative learning aspect is also relevant for SAR exploration.

Methodological Concept. We approached SAR transfer from a data mining perspective. The methodology we devised consists of three different stages, as illustrated in Figure 2. Depending on the application of the approach, i.e., (i) identification of alternative (target) series with SAR transfer potential if one template series is available or (ii) general analysis of SAR transfer events in a database, only stage 1 needs to be modified.

Stage 1: Identification of Alternative Scaffolds. (i) From the template series, the common scaffold is extracted. Then a database search is carried out to find related but chemically distinct scaffolds (Figure 2a). We applied the criterion that candidate scaffolds were permitted to differ from the original scaffold by replacement of one contiguous ring system (i.e., consisting of one or more rings). There are no restrictions on ring sizes or composition. Hence, depending on the exchanged ring systems, corresponding scaffolds might display different degrees of (dis-)similarity. If an alternative scaffold is identified, all compounds represented by this scaffold are retrieved, representing a potential target series.

(ii) For the general assessment of SAR transfer potential, all scaffolds are retrieved from a database and scaffolds pairs meeting the ring exchange criterion are identified. The subsequent evaluation proceeds in analogy to (i). In this case, however, template and target series are not distinguished. In addition, a scoring scheme is applied to quantify SAR transfer potential, as described in the Methods section.

Stage 2: Identification of Corresponding analogs. All analogs are subjected to R-group decomposition of the exchanged ring structure and grouped by corresponding substitution sites (Figure 2b). For analogs in the template series, matching compounds in the target series are identified that carry the same substituents and thus only differ in the exchanged ring system. R-groups are often attached to exchanged ring systems that differ in size and composition. Therefore, all possible pairwise mappings of ring positions are explored and alternative mappings are compared (Figure 2b). For (i), if R-groups are found in rings of different size, mappings are prioritized by considering R-group positions relative to the attachment point of the exchanged ring to the remaining part of the scaffold and their placement relative to heteroatom positions in exchanged rings. Then all pairs of corresponding analogs are extracted from the compared series.

Stage 3: Potency-Based Compound Ordering. For each series, the mean compound potency of all selected analogs is calculated, potency (pK_i) values of individual compounds are centered on the mean, and corresponding compound pairs are ranked in the order of increasing potency of the target series (Figure 2c). A uniform color code is applied to account for the potency difference of a compound from the mean (Figure 2c). The color code represents a continuous spectrum from green (over yellow) to red to account for lowest (green) to highest (red) compound potency in a series. Individual analogs are labeled with color-coded nodes that indicate their relative potency within the series.

Interpretation of SAR Transfer. Assessment of SAR transfer potential generally requires identification of corresponding analog series (structural criterion) and evaluation of potency effects resulting from corresponding structural modifications (potency criterion). SAR transfer potential is high if corresponding structural modifications lead to equivalent relative potency progression within each series. The graphical representation of the pairwise compound alignment of two series exemplified in Figure 2c is straightforward to analyze. R-group correspondence

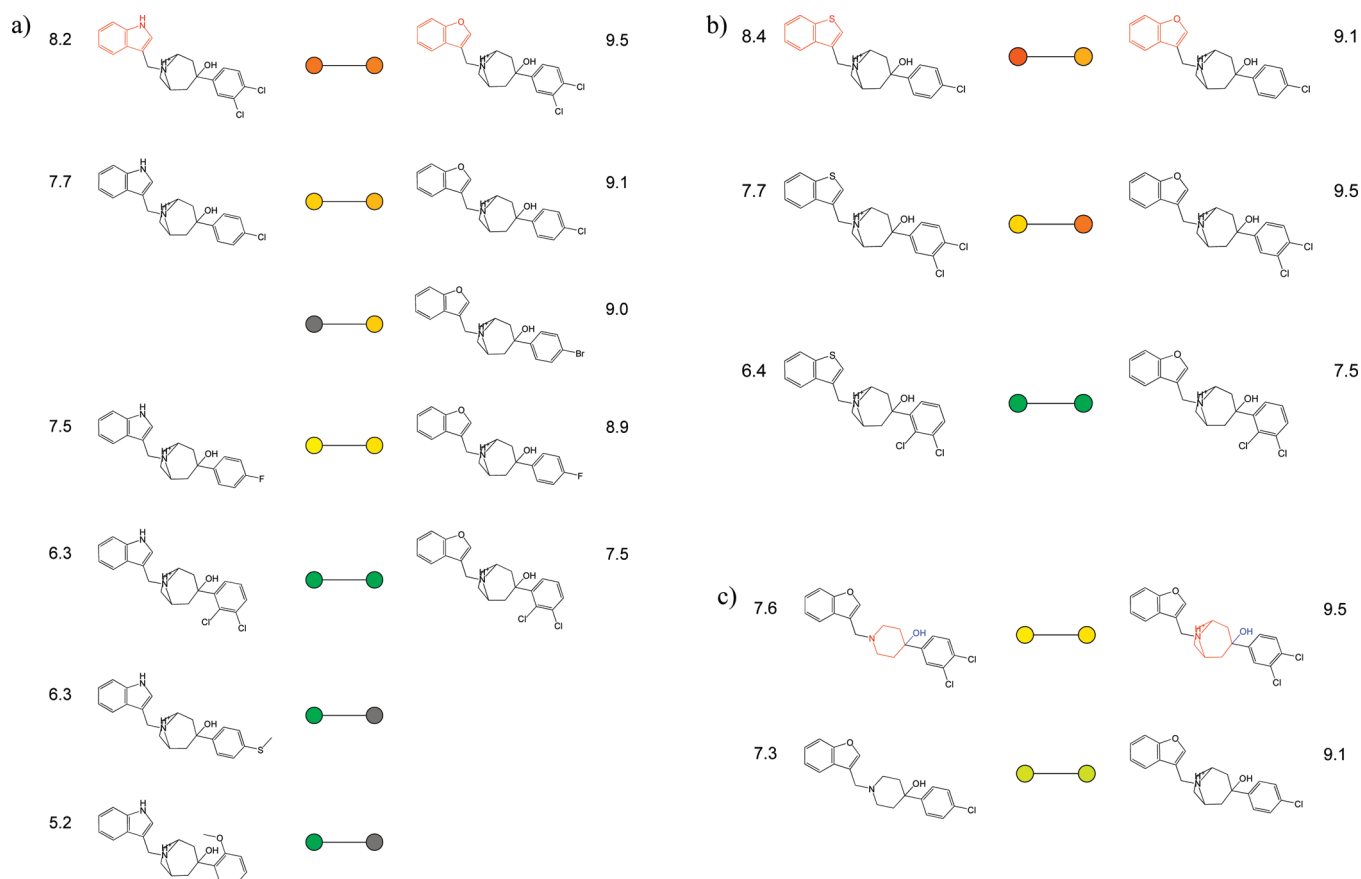


Figure 3. Template and target series for dopamine D3 receptor antagonists. (a–c) Alignments with three different target series are shown for a template series consisting of dopamine D3 receptor antagonists. In each case, compounds belonging to the template series are shown on the right and the target series is shown on the left. Corresponding compound pairs are ranked in the order of increasing potency of the target series. Node colors are determined on the basis of centered potency differences according to Figure 2c. Gray nodes indicate "missing" analogs. For each compound, its pK_i value is reported. In the compound pair at the top of each alignment, the exchanged ring structures are colored red. R-groups at the highlighted exchanged rings are colored blue.

can be examined, and color matches along the ranking indicate whether corresponding replacements have similar effects on potency progression within the analog series. Thus, if compounds forming a pair are always assigned the same color, their substitutions consistently cause the same relative potency effects, leading to complete SAR transfer along the alignment. By contrast, if corresponding analogs have differently colored nodes, substitutions have different relative potency effects. If this occurs for only a few pairs of analogs within larger series, SAR transfer is locally incomplete. If no corresponding color patterns are observed, two analog series meeting the structural criterion do not fulfill the potency criterion and hence there is no SAR transfer. Thus, SAR transfer potential can be easily assessed by comparing corresponding analogs on the basis of pairwise series alignments. Furthermore, if SAR transfer is observed and individual compounds exist in one analog series that do not have counterparts with corresponding R-groups in the other, new analogs can be readily suggested.

Identification of SAR Transfer Series. We first tested the methodology by searching for alternative analog series using a known series as a template. For different activity classes, compound series consisting of multiple analogs were selected as template series and the remaining active compounds were searched for potential target series that contained corresponding compounds

with comparable potency progression. In several instances, target series with SAR transfer potential were identified.

Dopamine D3 Receptor Antagonists. Figure 3 reports search results for a target series consisting of eight analogs of dopamine D3 receptor antagonists. Three target series with different SAR transfer potential were detected. In Figure 3a, the template (right) and target (left) series are related by the exchange of a benzofuran versus an indole moiety. The target series also consists of eight analogs. In the template and target series, five and six analogs, respectively, are unsubstituted at the exchanged ring and form four corresponding pairs in the alignment shown in Figure 3a. Among these pairs, clear SAR transfer is observed, indicated by very similar node colors for compounds forming each pair. In the second target series shown in Figure 3b, which consist of three analogs, the benzofuran moiety is replaced by a benzothiophene. Here, three corresponding pairs are also formed for unsubstituted exchanged rings. In contrast to the first target series, no SAR transfer is observed, due to inconsistent potency progression, although the core structures of these series are nearly identical. In both instances, the template series reaches a slightly higher potency level. Furthermore, Figure 3c shows a rudimentary yet structurally qualifying match with only two pairs of analogs. In this case, the central saturated ring moiety is exchanged and the rings carry a hydroxyl substituent. Analogs containing the original bridged aliphatic ring

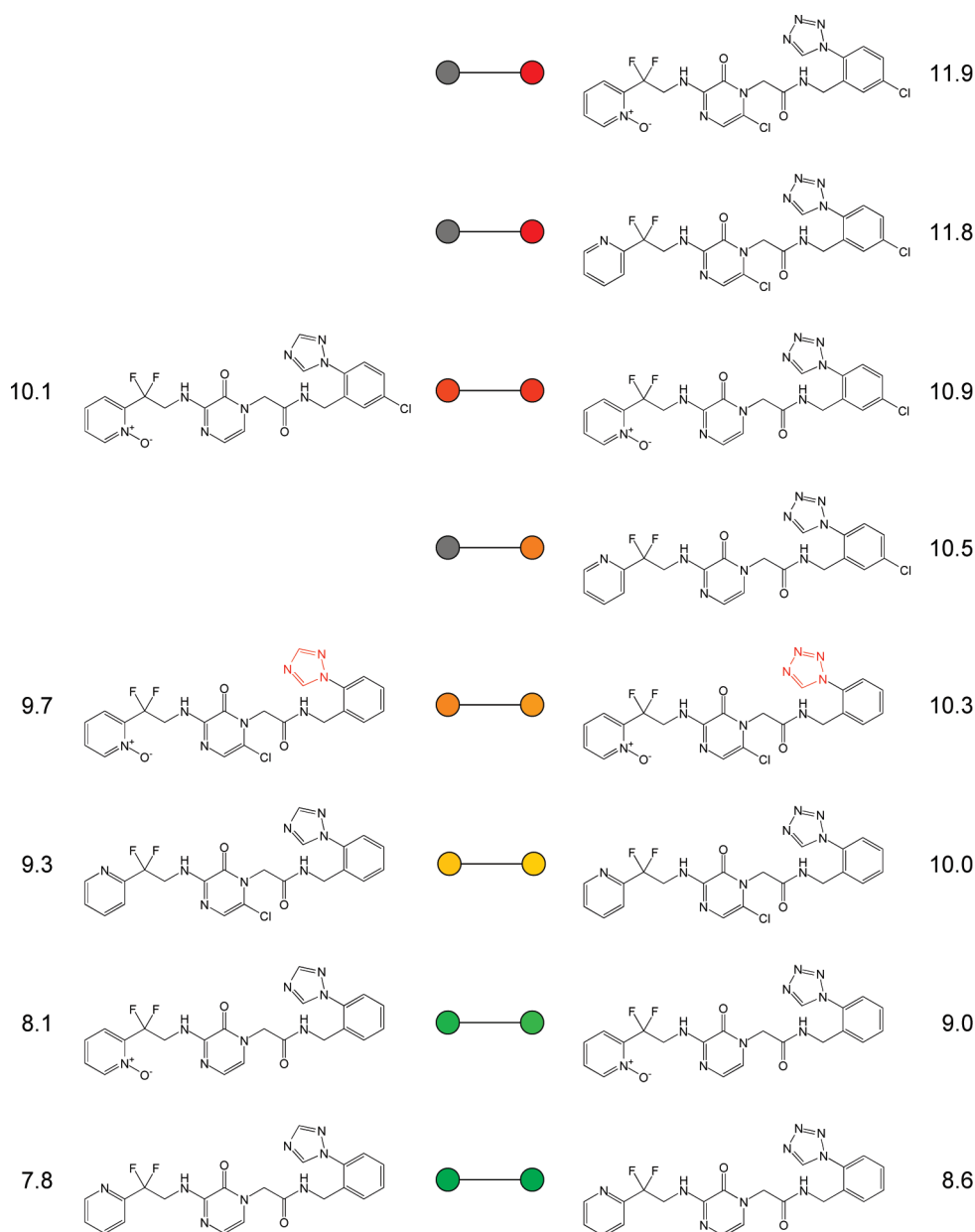


Figure 4. Thrombin inhibitors. Shown is a template series consisting of thrombin inhibitor analogs (right) and a single target series (left). The representation is according to Figure 3.

are nearly 100-fold more potent, but other SAR transfer conclusions cannot be drawn in this case.

Thrombin Inhibitors. Figure 4 shows a thrombin inhibitor template series consisting of eight analogs and the single target series we identified. The target series contains five analogs, all of which form pairs with template compounds. The template and target series are related by the exchange of a tetrazole versus a triazole ring. The alignment reveals clear SAR transfer character. In addition, the two most potent template compounds have no counterparts in the target series. In light of the observed potency progression, the corresponding analogs would be expected to have higher potency than the currently most potent triazole-containing compound, hence providing a prototypic example for alignment-based compound suggestions and the comparative learning aspect associated with SAR transfer analysis.

Factor Xa Inhibitors. We also searched for target series using a large template series of 88 factor Xa inhibitors. In this case, a small target series containing only six analogs was detected. These series were related by the exchange of phenyl and pyridine rings, and four pairs of corresponding analogs could be aligned (Figure 5). Although the resulting alignment conveys comparably little SAR transfer information, it shows that the potency of corresponding compounds is very similar and that the conserved R-group pattern at the exchanged rings in combination with substitutions at the two terminal phenyl rings leads to a large potency increase in both series. In the most and least potent analog pairs, the exchanged rings contain a fluorine–methyl–fluorine substituent pattern. In addition, in the most potent analog pair, the two terminal phenyl rings are meta substituted with carbamimidoyl groups, whereas in the least potent pair, one of these rings is

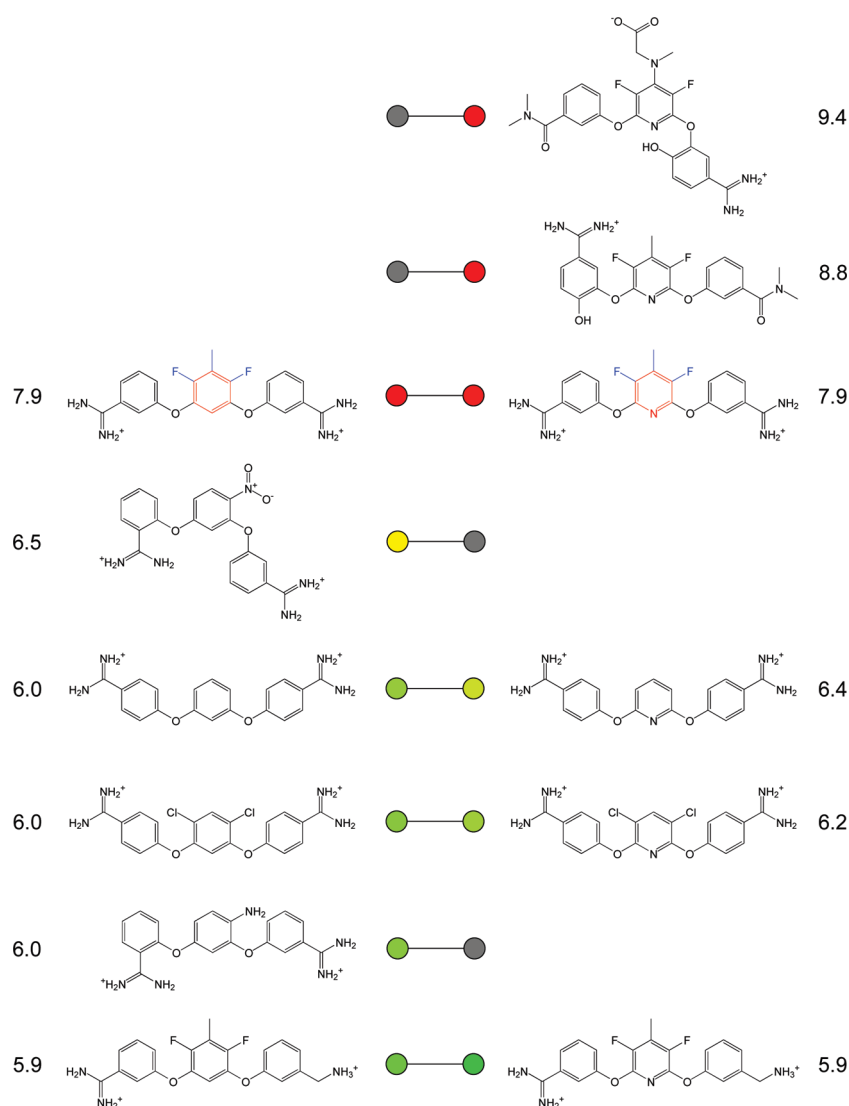


Figure 5. Factor Xa inhibitors. A template and target series are shown. The representation is according to Figure 3.

substituted in the meta position with an aminomethyl instead of the carbamimidoyl group. The other two weakly potent analog pairs contain both carbamimidoyl substituents but lack the fluorine–methyl–fluorine pattern at the exchanged ring. Other highly potent compounds of the template series (representative examples are shown in Figure 5) also contain this or similar R-group patterns.

Carbonic Anhydrase I Inhibitors. In Figure 6, search results are shown for another large template series of sulfonamide-containing carbonic anhydrase I inhibitors (51 analogs). Here, a target series with 39 analogs was identified. In this case, a phenyl and a thiadiazole ring were exchanged that carried the critical sulfonamide substituent. The best alignment consisting of eight analog pairs was obtained for sulfonamide substituents in the meta position of the phenyl ring. In this alignment, SAR transfer is locally incomplete because the relative potencies of the weakly potent analog pairs differ for the two series, as indicated by the color code. For the remaining analog pairs, potency progression is comparable. Thus, the alignment represents an example of partial SAR transfer. Importantly, SAR transfer was observed for the more potent compounds and several potent analogs in both

series had no counterparts in the alignment (representative examples are shown in Figure 6), thus providing another opportunity for comparative learning from two series and the design of other potent analogs.

Systematic Analysis of SAR Transfer Events. We then systematically searched for possible SAR transfer events in BindingDB. Therefore, all possible matching series in a target set were identified. Then all possible mappings of R-group positions in matching series were separately considered for score calculation, i.e., for each possible analog alignment the maximal difference of rescaled potency values for corresponding compounds was calculated. In order to focus the search on series that displayed a comparable potency progression over multiple compounds, we only retained alignments containing at least three compounds pairs. Furthermore, analogs in at least one of the matching series must span a potency range of at least 1 order of magnitude.

On the basis of these criteria, 306 matching scaffold pairs were identified in 93 target sets. Because some scaffold pairs occurred in multiple sets, a total of 405 different scaffold pair–target combinations were obtained. For the general assessment of SAR

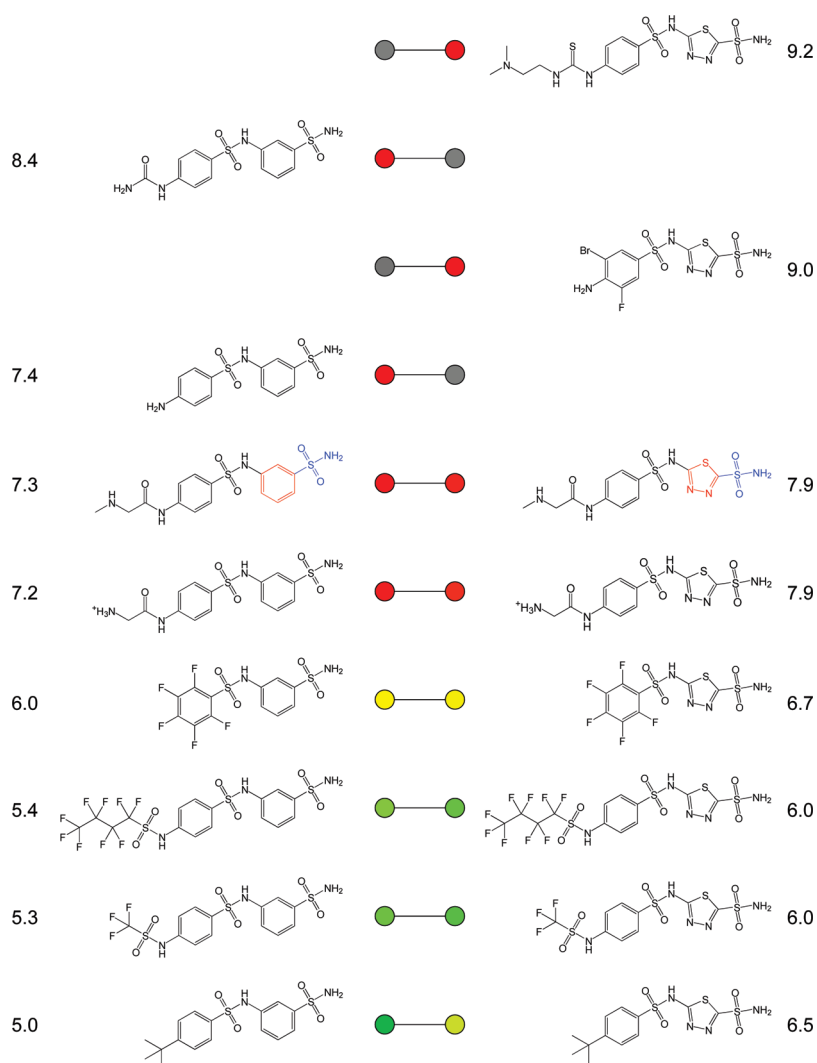


Figure 6. Carbonic anhydrase I inhibitors. A template and target series are shown. For clarity, only a part of the global series alignment is displayed. The representation is according to Figure 3.

transfer potential, the matching series yielding the lowest score for a scaffold pair–target combination were selected among alternative mappings.

We found that matching series consisted on average of 4.15 corresponding compound pairs. Figure 7 shows the score distribution observed for the 405 compound pair alignments. Scores between 0.2 and 1.0 were most frequently obtained, with a mean score of 0.69. However, the right tail of the distribution indicates that very high scores also occurred in some instances, i. e., for some matching series the exchange of a ring structure led to dramatic SAR discrepancies.

In order to investigate whether the number of transferred ring positions had an influence on the SAR transfer potential of a matching scaffold pair, statistics were separately generated for different numbers of mapped R-group sites. Table 1 reveals that matching series without R-groups at the exchanged rings were most frequently observed and that these series displayed the tendency to yield low scores, consistent with high SAR transfer potential. Furthermore, we observed the trend that with increasing numbers of R-groups at exchanged rings the scores also increased. Visual inspection of many analog alignments

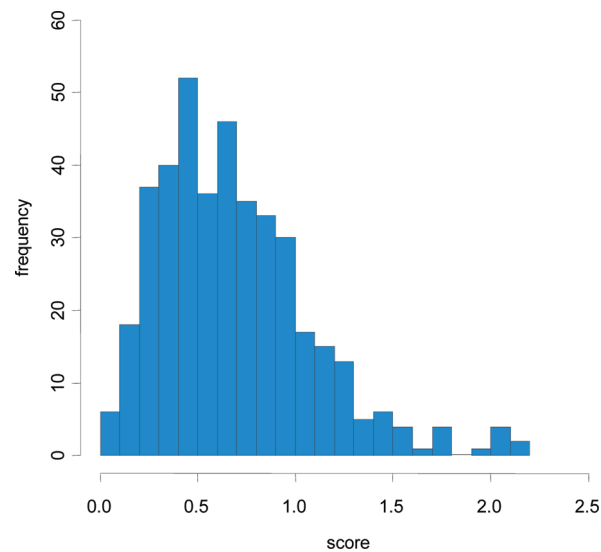


Figure 7. Score distribution. The distribution of SAR transfer scores is reported for 405 compound pair alignments extracted from BindingDB.

suggested that a score lower than or equal to 0.3 typically represented matching series representing SAR transfer events. For example, for the series shown in Figures 3a, 4, and 5 scores of 0.15, 0.12, and 0.24 were obtained, respectively. Hence, we applied a score threshold of 0.30 to search for SAR transfer series. A total of 61 SAR transfer series were identified in BindingDB that occurred in 39 different target sets and contained 59 different scaffold pairs. As also shown in Table 1, 70% of these SAR transfer series did not carry R-groups at the exchanged ring systems.

Table 1. Global Assessment of SAR Transfer Potential

all compound alignments			
no. of R-sites	no. of freq	score	no. of pairs
0	186	0.60	4.04
1	134	0.75	4.11
2	57	0.64	4.37
3	27	1.01	4.63
4	1	1.23	4.00

no. of freq		
no. of R-sites	SAR transfer series	nontransfer series
0	43	49
1	11	58
2	6	12
3	1	15
4	0	1

The 405 compound alignments extracted from BindingDB are grouped by the number of ring positions (R-sites) mapped for corresponding compounds from different analog series. For each group, its absolute frequency of occurrence (no. of freq), average SAR transfer score (score), and average number of aligned compound pairs (no. of pairs) are reported. The absolute frequency of occurrence is also reported for all groups of SAR transfer and nontransfer series.

Finally, we also identified matching series without SAR transfer potential. Therefore, a lower score cutoff of 0.80 was applied. In this case, 135 matching series were identified where the exchange of a ring system resulted in very different SAR behavior. A representative example is shown in Figure 8. Hence, matching series with limited or no SAR transfer potential were more frequently found than SAR transfer series. Interestingly, ~65% of matching series with scores larger than or equal to 0.80 comprised compounds with R-groups at the exchanged ring systems.

CONCLUDING REMARKS

Herein we reported a computational approach to search for individual SAR transfer series and systematically analyze SAR transfer events in databases. Partial core structure replacements were considered for the purpose of SAR transfer analysis, with no restrictions on the size, complexity, and composition of exchanged ring systems. In addition to designing a suitable computational search procedure, we put much emphasis on the chemical interpretability of the results. The analog pair alignments we introduce are straightforward to analyze and provide a basis for comparative SAR analysis. Several representative examples have been discussed, and a statistical analysis of SAR transfer events has been presented, which also included identification of structurally corresponding analog series with differing potency progression (i.e., nontransfer series). In addition, we have also shown how analog pair alignment information can be translated into compound design. The method reported herein can also be easily modified to account for scaffold replacements other than our preferred ring transformation.

AUTHOR INFORMATION

Corresponding Author

*Phone: +49-228-2699-306. Fax: +49-228-2699-341. E-mail: bajorath@bit.uni-bonn.de.

ACKNOWLEDGMENT

The authors would like to thank Martin Vogt and Mathias Wawer for helpful discussions.

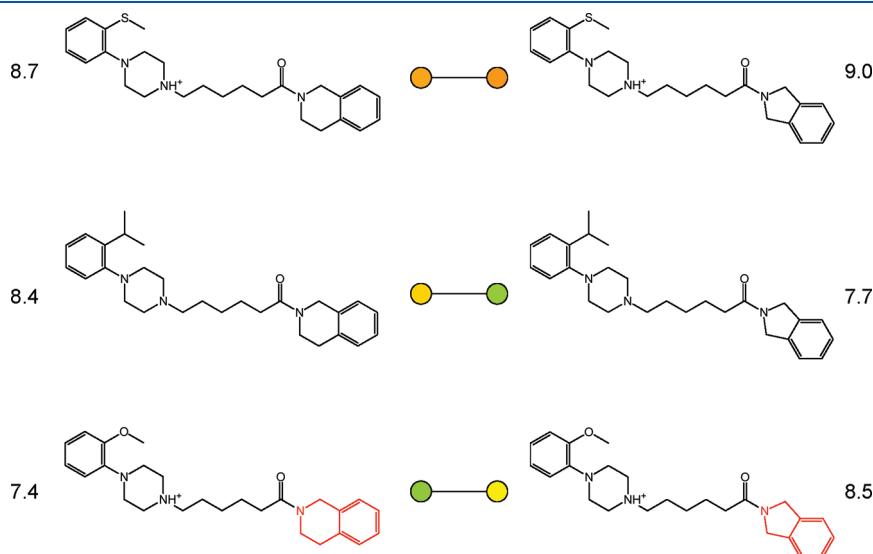


Figure 8. Serotonin receptor 1a antagonists. Matching series with no SAR transfer are shown. The representation is according to Figure 3.

■ REFERENCES

- (1) In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C. G., Ed.; Academic Press-Elsevier: Burlington, San Diego, USA; London, UK, 2008.
- (2) Wess, G.; Urmann, M.; Sickenberger, B. Medicinal Chemistry: Challenges and Opportunities. *Angew. Chem., Int. Ed.* **2001**, *40*, 3341–3350.
- (3) Brown, N.; Jacoby, E. On Scaffolds and Hopping in Medicinal Chemistry. *Mini Rev. Med. Chem.* **2006**, *6*, 1217–1229.
- (4) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
- (5) *Pipeline Pilot, Student ed.*, version 6.1; Accelrys, Inc.: San Diego, CA, 2007.
- (6) Kenny, P. W.; Sadowski, J. Structure Modification in Chemical Databases. In *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Wiley-VCH: Weinheim, Germany, 2005; pp 271–285.
- (7) Hussain, J.; Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model* **2010**, *50*, 339–348.
- (8) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (9) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201.
- (10) *OEChem TK*, version 1.7.4.3; OpenEye Scientific Software Inc.: Santa Fe, NM, 2010.