JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

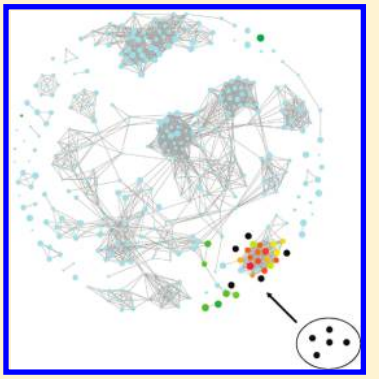
pubs.acs.org/jcim

Extraction of Discontinuous Structure−Activity Relationships from Compound Data Sets through Particle Swarm Optimization

Vigneshwaran Namasivayam, Preeti Iyer, 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: The characterization of structure−activity relationship (SAR) features of large compound data sets has been a hot topic in recent years, and different methods for large-scale SAR analysis have been introduced. The exploration of local SAR components and prioritization of compound subsets have thus far mostly relied on graphical analysis methods that capture similarity and potency relationships in a systematic manner. A currently unsolved problem in large-scale SAR analysis is how to automatically select those compound subsets from large data sets that carry most SAR information. For this purpose, we introduce a numerical optimization scheme that is based on particle swarm optimization guided by an SAR scoring function. The methodology is applied to four large compound sets. We demonstrate that compound subsets representing the most discontinuous local SARs are consistently selected through particle swarm optimization.



INTRODUCTION

The extraction of structure−activity relationship (SAR) information from compound data sets presents a major challenge in pharmaceutical research,[1,2] especially as increasingly large amounts of compound activity data become available. The evaluation of SAR information that might be present in large data sets is of critical importance for the selection of active compounds for hit-to-lead projects and for the prioritization of candidates for further optimization. For large-scale SAR analysis, the activity landscape concept is of high interest.[2] In general terms, an activity landscape is formed by the addition of a biological response hypersurface to a given chemical reference space representation.[2] It has been shown that the activity landscapes of many large data sets of specifically active compounds are characterized by the presence of SAR heterogeneity.[2,3] This means that different subsets of compounds are typically found in these data sets that display different types of similarity and potency relationships. For SAR analysis and the exploration of structural determinants of biological activity, regions of local SAR discontinuity are most relevant.[1,2] These regions are formed by structurally similar compounds having different potency.[2] The extreme form of SAR continuity is provided by activity cliffs,[2,4] i.e., pairs of structurally very similar (or analogous) compounds with large potency differences. Activity cliffs tend to appear in regions of strong local SAR discontinuity, and from these regions, most SAR information can usually be extracted.

The exploration of SAR heterogeneity in large compound sets requires employing computational methods that further extend the classical QSAR concept[5] and systematically search for available SAR information, without engaging in potency prediction.[1,2] For this purpose, SAR analysis functions have been introduced, such as the structure−activity relationship index (SARI)[6] or the structure−activity landscape index (SALI).[7] These numerical functions systematically compare similarity and potency relationships in compound data sets and calculate scores that account for global[6] and local[7,8] SAR continuity[6,8] or discontinuity.[6−8] These scoring schemes provide SAR diagnostics and are usually combined with graphical analysis methods such as SALI graphs[7] or network-like similarity graphs (NSGs)[8] to focus on SAR-informative compound subsets. In addition, a variety of other graphical SAR analysis tools have been introduced.[2]

However, it has thus far not been possible to automatically select compound subsets from large data sets that contain most SAR information, i.e., compounds forming local regions of high SAR discontinuity. For this purpose, we introduce a numerical optimization scheme for compound subset selection that combines particle swarm optimization (PSO)[9] with SAR discontinuity scoring. The methodology and exemplary applications are reported herein.

MATERIALS AND METHODS

Particle Swarm Optimization. PSO is an adaptive optimization method originally introduced by Kennedy and Ebehart.[9] It is designed to simulate coordinated behavior of individuals in a group (such as the flocking of birds and schooling of fish). The PSO algorithm is initialized with a population ("swarm") of solutions called "particles" at random positions in D-dimensional search space. The positional vector of each particle in the swarm is represented as $x_{id}(t)$. Each particle is assigned a randomized velocity $v_{id}(t)$ and moved through search space during each iteration (t). The particles attempt to improve their position in

Received: April 14, 2011
Published: June 06, 2011

```
begin
    Initialize the swarm S randomly with N particles
        with position x_id=x_i1,x_i2,.....x_iD and velocity v_id=v_i1,v_i2,.....v_iD
    Repeat
        Evaluate each particle in swarm by a fitness function
        For each particle ( i )
            Update individually best ( pbest ) and neighborhood best ( nbest ) positions
            For each dimension ( d )
                Update velocity v_id
                Compute sigmoidal transformation S(v_id)
                Update position x_id
            Next dimension
        Next particle
    Until termination criterion is met
end
```

Figure 1. PSO pseudocode. The steps involved in particle swarm optimization are presented as a pseudocode outline.

the search space on the basis of their best "experience" (pbest) and the best "experience" of the neighboring particles (nbest). Hence, PSO uses information from all particles. During the classical PSO optimization process, each particle updates its velocity $v_{id}(t)$ and position $x_{id}(t)$ in the swarm according to eqs 1 and 2, respectively.

$$v_{id}(t+1) = wv_{id}(t) + c_1\,\mathrm{rand}()[\mathrm{pbest}_{id}(t) - x_{id}(t)] + c_2\,\mathrm{rand}()[\mathrm{nbest}_{id}(t) - x_{id}(t)] \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

where $d = 1, 2, ..., D$ is the dimension of each particle and $i = 1, 2, ..., N$ is the number of particles in the swarm. $v_{id}(t)$ is the velocity at the current position of particle $x_{id}(t)$, and $x_{id}(t+1)$ is the updated particle position reached with velocity $v_{id}(t+1)$; w is an inertia factor. A large inertia weight facilitates a global exploration of search space and a small weight a more local exploration. c_1 and c_2 are confidence coefficients that control stochastic acceleration factors that pull each particle toward best individual and experience of neighborhood particles in the swarm. rand() is a random function uniformly distributed within the range (0,1). The optimization process is guided by experiences calculated using a fitness function. In the end, the best position found from all particles is returned as the final solution.

Given our objective to identify compound subsets representing highly discontinuous local SARs in a data set, we have used a standard structural fingerprint, MACCS keys,[10] to define the dimensionality for particles in the swarm. MACCS keys were chosen because of their robustness in large-scale SAR profiling[3] and because we intended to analyze the results of PSO in 2D activity landscape models[8] that utilize MACCS keys as a standard molecular representation. The choice of other fingerprints or numerical property descriptors would be expected to yield at least partly different compound similarity relationships and thus also alter local and global SAR features of compound data sets.[2,3]

Given our choice of a binary fingerprint descriptor, we have employed a binary version of PSO.[11] Therefore, the original PSO algorithm is adapted for discrete binary search space. However, in this case, the particle position is not a real value. Through application of a sigmoidal transformation to the velocity $v_{id}(t+1)$ of each particle in the swarm, values between 0 and 1 are obtained according to eq 3:

$$S(v_{id}(t+1)) = \frac{1}{1 + e^{-v_{id}(t+1)}} \quad (3)$$

Finally, the position of the particle is updated according to eq 4:

$$\text{if} \quad (\mathrm{rand}() < S(v_{id}^{t+1})) \quad \text{then} \quad x_{id}(t+1) = 1 \quad \text{else} \quad x_{id}(t+1) = 0 \quad (4)$$

A pseudocode outline of PSO is provided in Figure 1.

Fitness Function. A modified version of the discontinuity score component of the SARI score[6] according to eq 5 was used as a fitness function for particles.

$$\mathrm{raw}_{\mathrm{disc}} = \frac{\sum_{\{i,j|i\neq j\}} \mathrm{potdiff}(i,j) \times \mathrm{sim}(i,j)}{|\{i,j|i \neq j\}|} \quad (5)$$

Here pot(i), pot(j) are the potency (IC_{50} or K_i) values of compound i and j, respectively, potdiff(i,j) is the absolute potency difference between i and j, and sim(i,j) is the MACCS Tanimoto similarity[12] between compound i and j using MACCS fingerprints.

For the purpose of our analysis, the original discontinuity score was modified by omitting similarity and potency threshold values for score calculations[6] to enable an unrestricted search for SAR discontinuity in a compound data set. Raw scores were converted to Z-scores by utilizing the mean (μ) and standard deviation (σ) obtained from an external reference panel of 117 compound data sets assembled from ChEMBL.[13] Z-scores were then normalized by calculating the cumulative probability assuming a normal distribution, which yields discontinuity scores ($\mathrm{score}_{\mathrm{disc}}$) between 0 (no discontinuity) and 1 (maximal discontinuity).

PSO Calculations. PSO parameter settings are typically dependent on the specific application.[14] For our analysis, PSO parameters were adjusted on the basis of initial test calculations as follows: swarm size, 35; inertia weight, w = 0.721 348; confidence coefficients, c_1 and c_2, 1.193 147; maximum number of evaluations, 10 000. Inertia and confidence coefficients were set as suggested by Maurice.[15]

According to our scoring scheme, the objective for particles in the swarm is to search for compound subsets yielding high SARI discontinuous scores. Thus, during the optimization, particles should move toward a set of n compounds in the most discontinuous local SAR region of the data set. During each individual PSO iteration, the Euclidean distance was calculated between the particle position $x_{id}(t)$ and compounds in the data set. Compounds were ranked in the order of increasing Euclidean distance from particle positions, and the n top ranked compounds were selected to compute the discontinuity score. In subsequent optimization trials on each data set, n was set to

dx.doi.org/10.1021/ci2001692 |J. Chem. Inf. Model. 2011, 51, 1545–1551

10, 15, 20, 25, and 30 compounds. In each case, 10 independent trials were carried out. From these 10 trials, the best individual run was selected.

Compound Data Sets. For PSO analysis and compound subset selection, four sets of different enzyme or transporter inhibitors were selected from ChEMBL[15] that consisted of 217−392 active compounds spanning a large potency range, as summarized in Table 1. PSO can be applied to data sets of larger size. However, from the perspective of SAR analysis, the current data sets already are of fairly large size. On the basis of SARI profiling,[6] all four compound sets were characterized by the presence of notable global SAR heterogeneity and thus provided relevant test cases for our analysis.

Table 1. Compound Data Sets[a]

target	set	size	potency range (nM)
norepinephrine transporter	NOR	392	0.62−5454000
matrix metalloproteinase-2	MM2	334	0.09−685000
tyrosine-protein kinase LCK	LCK	369	0.01−500000
protein kinase C alpha	PKC	217	0.93−3590000

[a] For the four sets of inhibitors assembled from ChEMBL, the name of the target, the abbreviation used for the compound data set, its size (i.e. number of compounds), and potency range are reported.

Table 2. Discontinuity Scores of Selected Compound Subsets[a]

	discontinuity score			
subset	NOR	MM2	LCK	PKC
10	**0.97**	**0.94**	**0.98**	**0.82**
15	**0.87**	**0.85**	**0.90**	**0.72**
20	0.63	0.69	**0.79**	0.62
25	0.43	0.63	**0.73**	0.54
30	0.33	0.51	0.68	0.45

[a] For the four data sets, the best discontinuity scores obtained after PSO are reported for subsets containing 10−30 compounds. Scores above the threshold of 0.70 are given in bold.

Compound Analysis. In order to evaluate the numerical optimization of compound subsets representing discontinuous local SARs, PSO selections were mapped on NSG[8] representations of the source data sets. An NSG is an annotated similarity-based compound network where nodes represent compounds that are connected by edges if their calculated MACCS key Tanimoto similarity exceeds a predefined threshold value, set here to 0.80. Nodes are color-coded by the potency distribution in a data set using a continuous gradient from green (lowest potency in the data set) via yellow to red (highest potency). In addition, nodes are scaled in size according to the SARI per-compound discontinuity score,[8] which monitors the contribution of an individual compound to local SAR discontinuity. Accordingly, large nodes indicate that the potency of a compound significantly differs from the potency of its immediate structural neighbors. Combinations of large red and green nodes mark activity cliffs. The graphical layout of NSGs is calculated using the Fruchterman−Reingold algorithm[16] that positions multiple densely interconnected compounds closely together and separates weakly or unconnected groups of compounds from each other. Therefore, the distance between separated groups of compounds has no chemical meaning. NSGs were generated using an in-house Java implementation.

RESULTS AND DISCUSSION

PSO in Molecular Design. Different from evolutionary/genetic algorithms[17] that have for long been utilized to address different optimization problems in drug design,[18,19] PSO represents

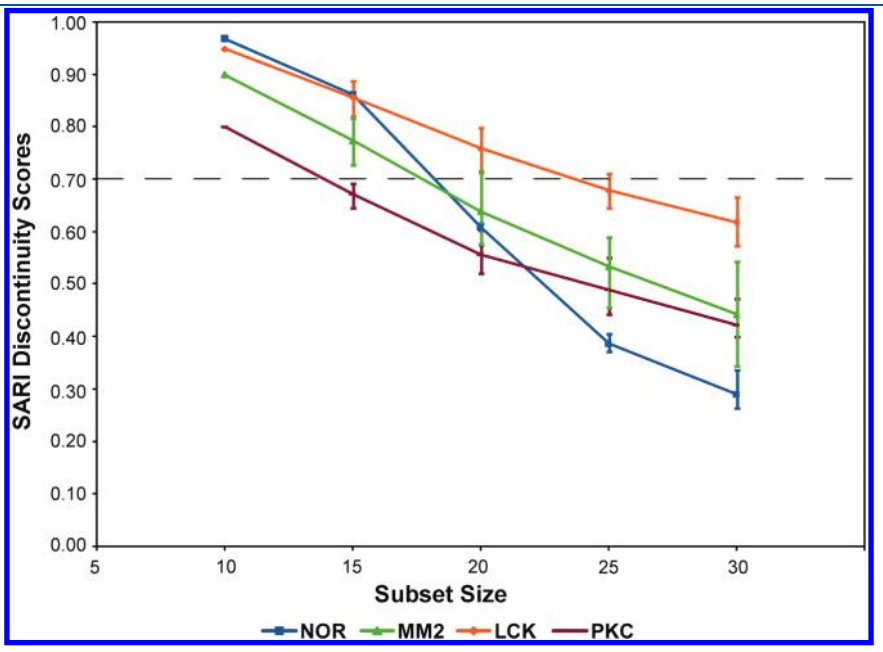


Figure 2. PSO optimization. Plotted are average discontinuity scores of 10 independent optimization trials for compound subsets of increasing size from all four data sets: NOR (blue), MM2 (green), LCK (orange), and PKC (red). Bars indicate the smallest and largest individual scores observed for each subset size. A discontinuity score of 0.70 (dashed line) was used as the threshold for compound selection.

dx.doi.org/10.1021/ci2001692 |J. Chem. Inf. Model. 2011, 51, 1545–1551

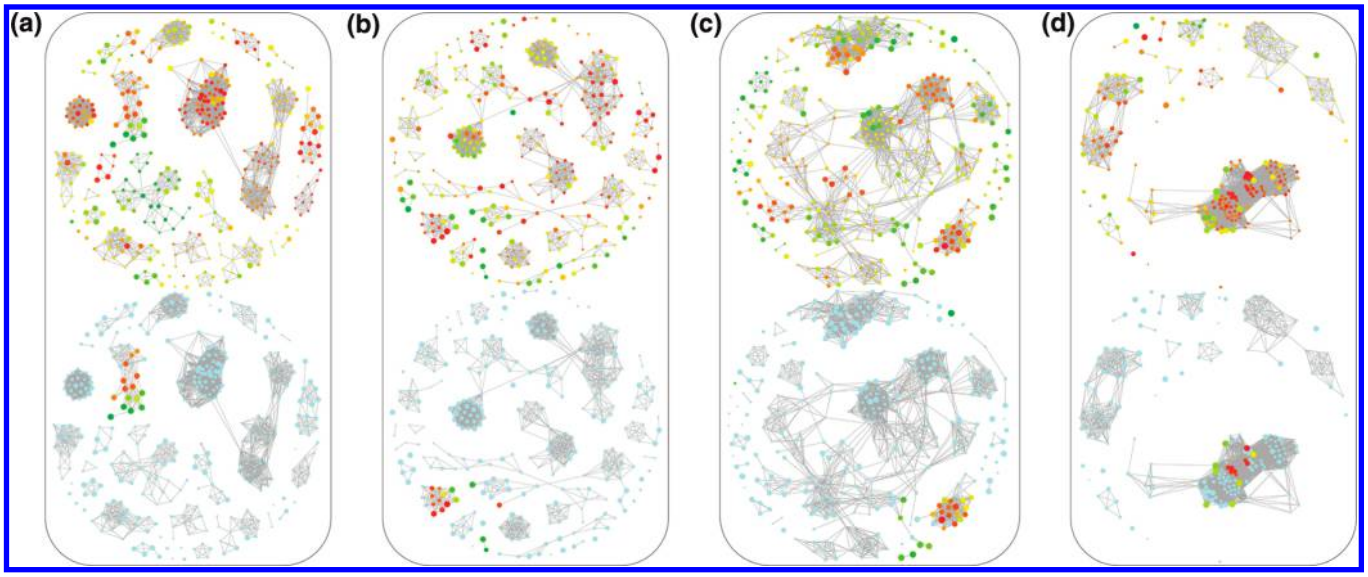


Figure 3. NSG Analysis. Selected compound subsets representing locally discontinuous SARs on the basis of PSO analysis were projected on NSG representations of the entire compound sets. In each case, the largest compound subset above the discontinuity threshold according to Table 2 is mapped. At the top of each panel, the original NSG of the data set is shown with potency-based node coloring (see Materials and Methods). At the bottom, only selected compound subsets are displayed with potency-based node coloring, while all remaining nodes are colored light blue. (a) NOR, (b) MM2, (c) LCK, and (d) PKC.

a more recent approach that has not yet been extensively investigated.[19] Only a few molecular design studies have thus far been reported that apply the PSO concept including QSAR modeling,[20,21] the optimization of docking algorithms,[22] or combinatorial de novo design.[23] Here we report a different type of PSO application, aiming at the selection of compound subsets with desired SAR properties. For this purpose, standard compound classification approaches such as clustering techniques cannot be applied in a meaningful manner. Advanced classification and ranking methods such as support vector machines are also difficult to apply here because the task at hand is in its essence an optimization problem. For our analysis, we selected PSO over evolutionary algorithms, for the following reason. In population-based heuristic methods such as genetic algorithms, sharing of information encoded in chromosomes principally differs from PSO. In the former case, selection, crossover, and mutation of chromosomes are essential to transfer information from one generation to another, which generally requires the exploration of a large number of generations. However, in PSO, particles share the information within the same and the following generation, which typically leads to efficient convergence, as reported previously.[22,24]

PSO and Local SARs. Discontinuous SARs are formed by structurally related or analogous compounds with significant differences in potency. Such compound sets are generally preferred sources of SAR information, as we further illustrate below. Thus, in addition to intuitive graphical SAR analysis, it would also be very useful to identify and extract the most discontinuous SAR regions from different data sets in a consistent manner, through automated data mining. In order to address this question, we have applied PSO in combination with the SARI discontinuity score component as a fitness function. This discontinuity score is calculated for a given compound set of varying size and can also be applied in modified form for per-compound scoring.[8] The more similar the compared compounds and the larger their potency differences are, the higher the discontinuity score will be. On the basis of SARI profiling,[6] strong discontinuity in a compound subset is typically observed at discontinuity scores greater than 0.70.

For the extraction of local SAR information from data sets, an important issue to take into account is the relationship between SAR discontinuity and subset size. The highest degree of discontinuity is represented by an activity cliff, i.e., a pair of very similar compounds with a large potency difference. However, discontinuous local SARs are not formed by isolated activity cliffs; they typically involve multiple cliffs and/or compounds in the immediate structural environment of prominent cliffs.[8] Thus, it is generally important to select multiple compounds in order to characterize local SARs.

PSO-Based Compound Selection. In the light of the above considerations, we have restricted the size of the smallest compound subset for PSO-based selection to 10 active compounds and subsequently carried out independent optimizations for subsets of up to 30 compounds, with an incremental increase of five compounds per subset. The threshold score for local discontinuity was set to 0.70.

The results of our optimization trials are summarized in Table 2 and Figure 2. All PSO trials converged within the chosen limit of 10 000 iterations. As reported in Table 2, the best discontinuity scores for subsets of 10 compounds were greater than 0.90 for three of our data sets (NOR, MM2, and LCK) and greater than 0.80 for one (PKC). In two cases (NOR and LCK), the best discontinuity scores were close to 1. Thus, in each case, PSO successfully extracted compound subsets that were characterized by strong local SAR discontinuity. Figure 2 shows that score deviations between independent PSO trials were generally small over all compound subsets. Figure 2 also shows that discontinuity scores constantly decreased with increasing subset sizes. This is exactly what one would expect in this case. As selected compound subsets increase in size, an increasing number of compounds peripheral to centers of strong SAR discontinuity are selected until, at some point, the subset extends beyond

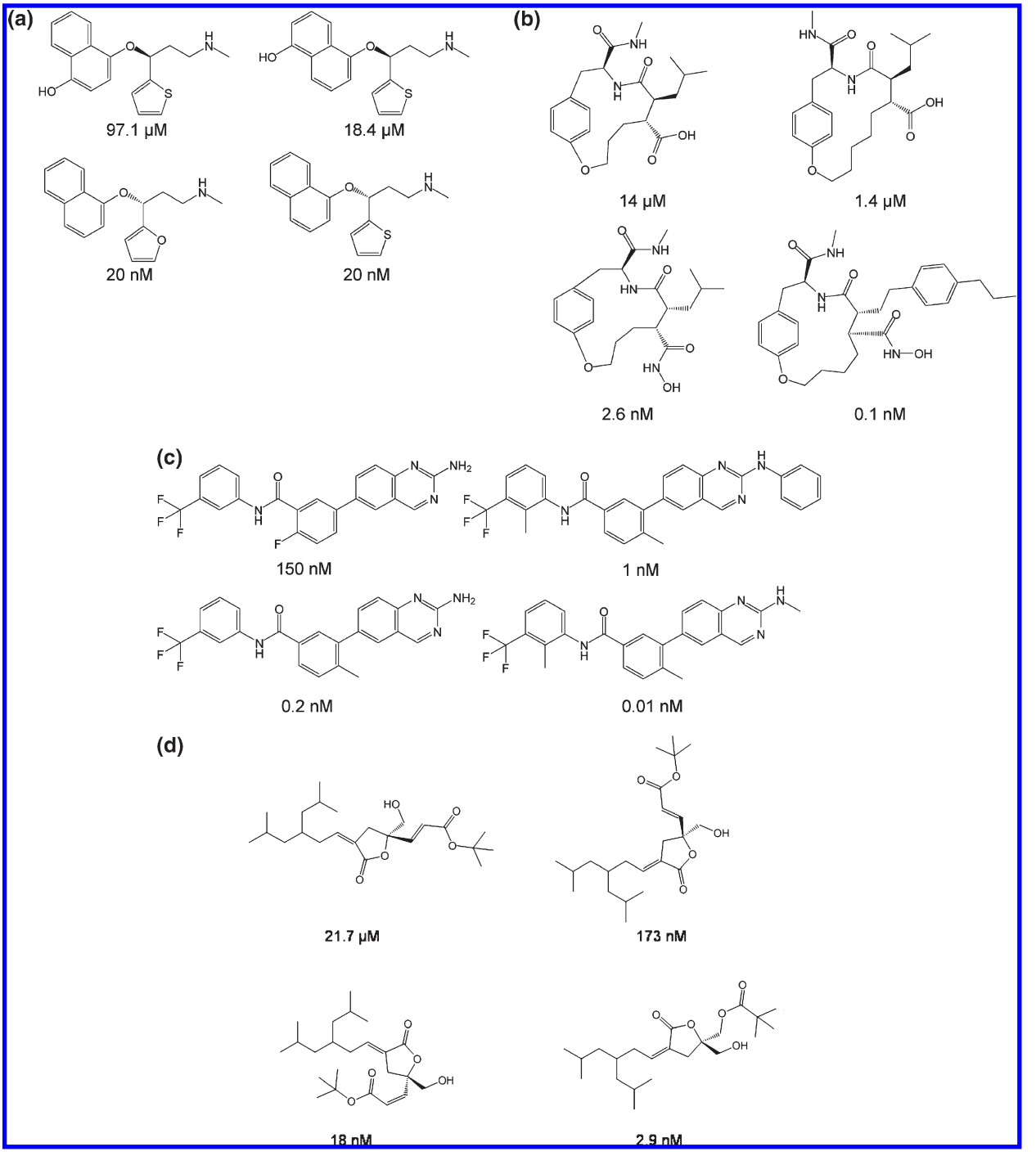


Figure 4. Compound structures. For each compound subset in Figure 3, structures of two compounds with high potency and two with low potency are shown as representative examples. (a) NOR, (b) MM2, (c) LCK, and (d) PKC.

the region of local discontinuity. This is evident in Table 2. For two data sets (NOR and MM2), subsets containing 10 and 15 compounds were strongly discontinuous in nature, but a notable decrease in discontinuity scores occurred when 20 compounds were selected. For LCK and PKC, a more gradual decrease in scores was observed. In three instances (NOR, MM2, and PKC), subsets containing 30 compounds yielded intermediate or low discontinuity scores that did no longer represent SAR discontinuity. For further analysis, we selected the top-scoring individual subsets of increasing size from each data set until the best subset score fell below the discontinuity threshold of 0.70. Hence, for NOR, MM2, and PKC, subsets containing 10 and 15 compounds were selected and, for LCK, subsets of 10, 15, 20, and 25 compounds (Table 2).

Compound Mapping. The selected subsets were then mapped on NSG representations of the entire data sets, shown in Figure 3. The NSGs of all four data sets reflected the presence of SAR heterogeneity. In each case, compound clusters were observed that contained combinations of large red and green nodes, indicating strong local discontinuity. In addition, other

dx.doi.org/10.1021/ci2001692 |J. Chem. Inf. Model. 2011, 51, 1545–1551

clusters mostly consisted of small and similarly colored nodes, i.e., structurally similar compounds having similar potency. These clusters represent a form of local SAR continuity and do not provide information about structural modifications that are critical for activity. Another form of SAR continuity is presented by structurally diverse compounds having similar potency. These compounds are not connected by edges in NSGs and thus not grouped together.

When mapping selected compound subsets of increasing size, we found, without an exception, that PSO subsets from each data set always covered the same region in the NSG (Figure 3). Thus, PSO produced stable compound selections. Importantly, on the basis of graphical analysis, the respective compound subsets would have been selected in all four cases as those representing the most prominent discontinuous SARs. Hence, the PSO selections were also fully consistent with graphical analysis.

In Figure 3, the largest compound subsets yielding a discontinuity score above 0.70 are mapped. In Figure 3a, the NSG of the NOR data reveals that the 15 selected compounds essentially extract all activity cliffs from the compound cluster representing the highest degree of discontinuity in the data set. In the case of MM2 (Figure 3b), the most prominent region of SAR discontinuity involves a small compound cluster at the lower left of the NSG containing three compounds with high and two with low potency that form activity cliffs. Thus, in this case, strongest SAR discontinuity is confined to a very small region that is, however, covered by the PSO selection. In addition, several weakly potent compounds are contained in the selected 15-compound subset that make large contributions to SAR discontinuity (i.e., they are represented by large green nodes), although they do not have highly potent neighbors in the graph. This is the case because their similarity to highly potent clustered compounds falls below the applied edge threshold. If the threshold is slightly lowered, these compounds join the discontinuous cluster (data not shown). Hence, in this case, numerical optimization yields a larger highly discontinuous subset than would have been possible to select on the basis of graphical analysis. Similar observations are made for LCK in Figure 3c. In this case, a 25-compound subset is mapped. This subset selects the most discontinuous cluster at the lower right of the NSG together with six peripheral weakly potent compounds that participate in the formation of additional activity cliffs but also fall below the edge similarity threshold, as discussed above. As shown in Figure 3d, the topology of the NSG for PKC differs from the others because there is a large and densely connected graph component present that includes compounds forming a strongly discontinuous SAR. Here the 15-compound PSO subset also selects the most prominent compounds from this region.

In Figure 4, representative highly and lowly potent compounds from all PSO subsets are shown whose comparison reveals SAR information. For example, for NOR inhibitors in Figure 4a, strong influence of isomerization at a single stereocenter on compound potency is observed. For MM2 inhibitors in Figure 4b, the conversion of a carboxylic acid function into an *N*-oxidized acetamide and changes in substituent stereochemistry increase compound potency. Furthermore, the exchange of simple fluoro and methyl substituents at the central phenyl moiety and the introduction of an additional peripheral phenyl ring modulate the potency of LCK inhibitors (Figure 4c). Moreover, a cis−trans isomerization and the introduction of an ether group lead to large-magnitude potency changes in a series of PKC inhibitors (Figure 4d). Thus, compound subsets selected by PSO displayed strong SAR continuity, mapped to the most prominent regions of local discontinuity in NSGs, and revealed readily interpretable SAR information.

CONCLUSIONS

We have reported a PSO-based approach for the selection of compound subsets from large data sets that represent strong local SAR discontinuity. On the basis of our analysis, PSO yielded stable compound selections and consistently extracted highly discontinuous subsets from our test data sets. In addition, the PSO selections were consistent with graphical analysis of local SAR discontinuity. Hence, the PSO approach introduced herein presents an automated numerical optimization scheme for the identification of compound subsets yielding SAR information. As such, it further extends the spectrum of currently available methods for large-scale SAR analysis.

AUTHOR INFORMATION

Corresponding Author

*Tel: +49-228-2699-306; fax: +49-228-2699-341; e-mail: bajorath@bit.uni-bonn.de.

REFERENCES

(1) Bajorath, J.; Peltason, L.; Wawer, M.; Guha, R.; Lajiness, M. S.; Van Drie, J. H. Navigating Structure−Activity Landscapes. *Drug Discovery Today* **2009**, *14*, 698–705.

(2) Wassermann, A. M.; Wawer, M.; Bajorath, J. Activity Landscape Representations for Structure−Activity Relationship Analysis. *J. Med. Chem.* **2010**, *53*, 8209–8223.

(3) Peltason, L.; Bajorath, J. Systematic Computational Analysis of Structure−Activity Relationships: Concepts, Challenges and Recent Advances. *Future Med. Chem.* **2009**, *1*, 451–466.

(4) Maggiora, G. M. On Outliers and Activity Cliffs—Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535–1535.

(5) 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.

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

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

(8) 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.

(9) Kennedy, J.; Eberhart, R. C. Particle Swarm Optimization. *Proceedings of the IEEE International Conference Neural Networks IV (ICN95)*; 1995; pp 1942−1948.

(10) *MACCS Structural Keys*; Symyx Software: San Ramon, CA, 2005.

(11) Kennedy, J.; Eberhart, R. C. A Discrete Binary Version of the Particle Swarm Algorithm. *Proceedings of the World Multiconference on Systemics, Cybernetics, and Informatics*; 1997; pp 4104−4109.

(12) Willett, P. Searching Techniques for Databases of Two- and Three-Dimensional Structures. *J. Med. Chem.* **2005**, *48*, 1–17.

(13) *ChEMBL*. http://www.ebi.ac.uk/chembl (accessed July 1, 2010).

(14) Shi, Y.; Eberhart, R. C. Parameter Selection in Particle Swarm Optimization. *Proceedings of the Seventh Annual Conference on Evolutionary Programming*; 1998; pp 591−600.

(15) Maurice, C. Stagnation Analysis in Particle Swarm Optimization or What Happens When Nothing Happens. Technical Report CSM-460, Department of Computer Science, University of Essex. 2006. ISSN:

(16) Fruchterman, T. M. J.; Reingold, E. M. Graph Drawing by Force-Directed Placement. *Software—Pract. Exper.* **1991**, *21*, 1129–1164.
(17) Judson, R. Genetic Algorithms and Their Use in Chemistry. *Rev. Comput. Chem.* **1997**, *10*, 1–73.
(18) Terfloth, L.; Gasteiger, J. Neural Networks and Genetic Algorithms in Drug Design. *Drug Discovery Today* **2001**, *6* (Suppl.), S102–S108.
(19) Schneider, G.; Hartenfeller, M.; Reutlinger, M.; Tanrikulu, Y.; Proschak, E.; Schneider, P. Voyages to the (Un)known: Adaptive Design of Bioactive Compounds. *Trends Biotechnol.* **2009**, *27*, 18–26.
(20) Agrafiotis, D. K.; Cedeno, W. Feature Selection for Structure—Activity Correlation Using Binary Particle Swarms. *J. Med. Chem.* **2002**, *45*, 1098–1107.
(21) Lü, J. X.; Shen, Q; Jiang, J. H.; Shen, G. L.; Yu, R. Q. QSAR Analysis of Cyclooxygenase Inhibitors Using Particle Swarm Optimization and Multiple Linear Regression. *J. Pharm. Biomed. Anal.* **2004**, *35*, 679–687.
(22) Namasivayam, V.; Günther, R. PSO@Autodock3: A Fast Flexible Molecular Docking Program Based on Swarm Intelligence. *Chem. Biol. Drug Des.* **2007**, *70*, 475–484.
(23) Hartenfeller, M.; Proschak, E.; Schüller, A.; Schneider, G. Concept of Combinatorial De Novo Design of Drug-like Molecules by Particle Swarm Optimization. *Chem. Biol. Drug Des.* **2008**, *72*, 16–26.
(24) Eberhart, R. C.; Shi, Y. Comparison of Genetic Algorithms and Particle Swarm Optimization. Evolutionary Programming VII. *Lecture Notes Comput. Sci.* **1998**, *1447*, 611–616.