JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

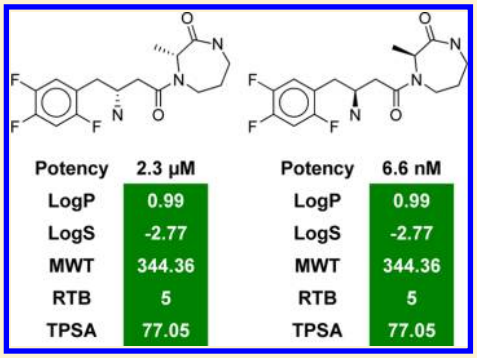
pubs.acs.org/jcim

Multiobjective Particle Swarm Optimization: Automated Identification of Structure−Activity Relationship-Informative Compounds with Favorable Physicochemical Property Distributions

Vigneshwaran Namasivayam 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: The selection of active compounds for chemical optimization efforts typically requires the consideration of multiple properties beyond potency. Herein we introduce a multiobjective particle swarm optimization approach to automatically extract compound subsets from large data sets that reveal structure−activity relationship (SAR) information and display physicochemical property distributions that are indicative of favorable absorption, distribution, metabolism, and excretion (ADME) characteristics. The approach is based on Pareto optimization of multiple objectives and does not require subjective intervention. It is automated and can be easily modified. We have applied the method to screen 10 compound data sets of different composition and global SAR phenotypes. In five of these data sets, between one and more than hundred compound subsets were identified that represented discontinuous local SARs and had desirable property distributions.



1. INTRODUCTION

The extraction of structure−activity relationship (SAR) information from large compound data sets is a major task in computational medicinal chemistry.[1] Large-scale SAR exploration aids in the characterization of activity landscapes of compound data sets, identification of under- or overexplored compound series, and prioritization of active compounds for further chemical optimization.[1] For SAR information extraction, analysis functions such as SARI[2] or SALI[3] have been introduced that characterize the information SAR content of compound data sets, describe global and local SAR features, and identify activity cliffs. SAR discontinuity arises from the presence of structurally similar or analogous compounds that are specifically active at different potency levels. For example, if a compound has a potency that is very different from its immediate structural neighbors, local SAR discontinuity is introduced. Accordingly, activity cliffs represent the extreme form of SAR discontinuity.[1] In addition, SAR continuity is due to the presence of structurally increasingly diverse compounds that maintain similar potency. These SAR phenomena are quantitatively accounted for by the discontinuity and continuity components of the SARI score.[2] The composite score balances these contributions.

SAR analysis functions such as SARI or SALI systematically calculate similarity and potency relationships between active compounds but do not take other physicochemical properties into account. With the aid of numerical molecular descriptors, preferred ranges of selected physicochemical properties have been delineated and used to estimate drug- or lead-likeness of active compounds.[4] Among others, popular property descriptors utilized for these purposes include LogP (the octanol/water partition coefficient; a measure of hydrophobicity), LogS (solubility), RTB (number of rotatable bonds; a measure of molecular flexibility), H-bond acceptors, H-bond donors, and TPSA (total polar surface area of a molecule; a measure of hydrophilicity and membrane permeability). For compound selection on the basis of SAR information, the inclusion of additional criteria accounting for drug- or lead-likeness is of course desirable. Compounds that are SAR-informative and display favorable physicochemical property ranges are preferred candidates for further chemical exploration. In compound design, desirable combinations of different molecular properties can be obtained through multiobjective property optimization.[5,6] Currently available optimization methods such as SELECT[5] or MOGA[6] utilize genetic algorithms (GAs) in combination with different property descriptors.

Previously, we have introduced computational approaches for the automated identification of compound subsets in large data sets that represented highly discontinuous SARs[7] or continuous local SARs in the vicinity of activity cliffs.[8] These approaches have combined particle swarm optimization (PSO),[9] a social behavior-mimicking evolutionary algorithm (EA), with SARI component scoring as a fitness function.

Herein, we report a further extension of SAR-focused PSO to automatically identify compound subsets with high SAR information content and, in addition, favorable physicochemical property distributions. Such compounds are expected to provide meaningful starting points for optimization and development efforts. For this purpose, a multiobjective PSO (MO-PSO) scheme was designed that combines SAR and molecular property

Received: August 24, 2012
Published: October 7, 2012

evaluation. In the following, we introduce our MO-PSO methodology and report its application to a variety of compound data sets.

2. PARTICLE SWARM OPTIMIZATION

PSO is an EA that simulates the coordinated behavior of animals in a swarm.[9] In PSO, patterns of social behavior are modeled as rules to guide so-called particles to move in a coordinated manner toward preferred regions of a search space. Similar to other EAs such as GAs or evolution strategy algorithms, PSO is initialized using a population of N random solutions constituting a swarm. Each solution is termed a particle (i) and represented by a coordinate vector x_{iD} in a D-dimensional search space. PSO uses the rate of change in particle positions to navigate the search space.[10] The rate of change is represented as a velocity vector v_{iD}. In addition to its individual feature information, each particle stores the information of its previous best position $(pbest)$ and the best positions of its neighboring particles $(nbest)$. The movement of a particle in search space is influenced by its interactions with neighboring particles. During each iteration (t), the particle adjusts its velocity and position according to eqs 1 and 2, respectively, and moves in the search space toward a favorable global solution:

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

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

Here, c_1 is termed the cognitive confidence coefficient and c_2 is the social confidence coefficient,[11] which modify the velocity of a particle toward pbest and nbest, respectively; rand() is a random function uniformly distributed between values 0 and 1; w is the inertia weight, a parameter that controls the influence of the velocity history on the current velocity of a particle in the swarm.

Equations 1 and 2 are applied to search spaces with real-valued dimensions. For search spaces with binary-valued (0 or 1) dimensions, a binary PSO variant is utilized and the particle position is updated by applying a sigmoidal transformation to the velocity according to eqs 3 and 4.[12]

$$x_{iD}(t+1) = x_{iD}(t) + v_{iD}(t+1) \quad (3)$$

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

The steps involved in the binary PSO are summarized in Figure 1.

3. SARI DISCONTINUITY SCORE

SARI and its discontinuity score component according to eq 5 can be calculated for entire data sets (global score) or compound subsets (local score):

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

Here, potdiff(i,j) is absolute potency difference between compound i and j and sim(i,j) is the calculated similarity of i and j. Scores of 0 and 1 indicate minimal and maximal SAR discontinuity, respectively. For both global and local score calculations, the raw scores are converted into Z-scores on the basis of the score distribution of an external reference panel of

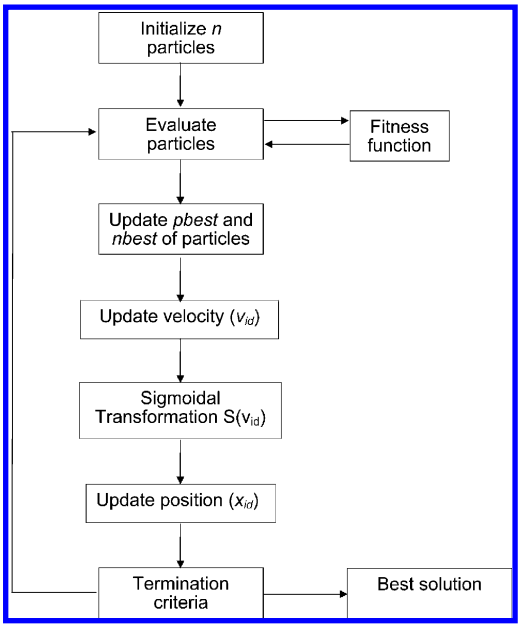


Figure 1. Binary PSO. Steps involved in binary PSO calculation are presented in a flowchart diagram.

compound data sets[2] and mapped to the range $[0, 1]$ via a cumulative distribution function.[2] Tanimoto similarity[13] was calculated using MACCS structural keys[14] as the molecular representation. Discontinuity scores of 0.7 or greater are indicative of compound subsets representing strongly discontinuous local SARs.[7] Compounds forming discontinuous SARs are characterized by limited chemical changes leading to significant potency alterations. Hence, SAR discontinuity is generally associated with high SAR information content.

4. CLASSIFICATION OF DESCRIPTOR VALUE RANGES

For all test compounds, the LogP, LogS, RTB, TPSA, and molecular weight (MW) descriptors were calculated using the molecular operating environment.[15] We utilized the compound ADME-related property descriptor classification scheme of Lobell et al.[16] Following this work, descriptor value range combinations were organized using traffic light-analogous classification to yield property ranges corresponding to favorable (green), intermediate (yellow), or unfavorable (red) ADME characteristics. The following value range combinations were utilized[16] (solubility values given in milligrams per liter were converted into molar LogS values). Favorable value range combinations were the following: LogP ≤ 3, MW ≤ 400, TPSA ≤ 120, RTB ≤ 7, LogS ≤ -3.5. Intermediate range combinations were the folloing : LogP = 3−5, MW = 400−500, TPSA = 120−140, RTB = 8−10, LogS = −3.5 to −4.0. Unfavorable range combinations were the following: LogP > 5, MW > 500, TPSA > 140, RTB > 10, LogS < −4.0.

5. COMPOUND DATA SETS

Ten compound data sets were assembled from ChEMBL[17] as described[8] to represent different SAR phenotypes, as reported in Table 1. These activity classes contained between 177 and 1553 compounds. On the basis of global SARI scoring, they included data sets that were mostly discontinuous in nature (low SARI scores; CAS, CFX, REN, and AKT), heterogeneous in their SAR character (intermediate scores; PKC, DPP, CA1, and MAP), or mostly continuous (high SARI scores; NOR and COX). If

dx.doi.org/10.1021/ci300402g | J. Chem. Inf. Model. 2012, 52, 2848−2855

Table 1. Compound Data Sets[a]

no.	activity class	set	no. cpds	global SARI
1	Caspase-1 inhibitors	CAS	177	0.19
2	coagulation factor X inhibitors	CFX	1553	0.26
3	Renin inhibitors	REN	557	0.28
4	Ser/Thr kinase AKT inhibitors	AKT	458	0.31
5	protein kinase C alpha inhibitors	PKC	274	0.43
6	dipeptidyl peptidase IV inhibitors	DPP	1232	0.44
7	carbonic anhydrase I inhibitors	CA1	931	0.55
8	MAP kinase p38 alpha inhibitors	MAP	807	0.58
9	norepinephrine transporter ligands	NOR	1378	0.74
10	cyclooxygenase-2 inhibitors	COX	562	0.76

[a]For each of 10 activity classes assembled from ChEMBL, the name of the target, set abbreviation, number of compounds (no. cpds), and global SARI score are reported. Both the global SARI and the discontinuity score (a SARI score component) range from 0 to 1. An increasing global SARI score (a composite score of continuity and discontinuity contributions) indicates increasing global SAR continuity of a data set, whereas increasing discontinuity scores, as calculated herein, indicate increasing local SAR discontinuity.[2]

multiple (K_i or IC_{50}) potency values were available for a compound, the geometric mean was calculated as the final potency annotation. In Figure 2, the value range distribution of each selected descriptor is reported for these activity classes.

6. MULTIOBJECTIVE PARTICLE SWARM OPTIMIZATION

6.1. Principles of Multiobjective Optimization. MO optimization schemes generally balance contributions of multiple fitness functions. The optimization problem is then reformulated as the minimization of eq 6:

$$F(x) = \sum f_i(x) \quad (6)$$

where $f_i(x)$ is a function capturing $i = 1$ to k different objectives. PSO can be modified to utilize more than one objective to find a preferred global solution.[18] For example, this might be achieved by combining all objectives into a single function or by primarily optimizing one parameter while constraining the remaining ones to predefined settings. These approaches requires prior knowledge of appropriate weights and/or property values. By contrast, Pareto-based optimization,[18,19] as illustrated in Figure 3, simultaneously optimizes multiple equivalent objective functions to search for nondominated solutions that cannot be further improved with respect to a given objective without compromising others. Hence, this optimization scheme eliminates the need to predefine weights or constraints.

6.2. MO-PSO Setup and Search Trials.

(a) For our MO-PSO analysis, the Pareto-based approach was implemented.

(b) In conventional PSO using single objective functions, nbest guides particles in the swarm and updates their position. By contrast, in MO-PSO, the global best (gbest), i.e., the best position of neighboring particles with respect to all objectives, was used to guide particles toward nondominated solutions in search space.

(c) In order to identify subsets of compounds that are SAR-informative and have favorable physicochemical property distributions, the local SARI discontinuity score and five numerical property descriptors were used as objectives including LogP, LogS, RTB, TPSA, and MW.

(d) For each descriptor, the average value for a compound subset and the standard deviation were calculated as independent objective functions. Thus, together with the discontinuity score, this resulted in a total of 1 + 5*2 = 11 objective functions for our analysis. The calculation of descriptor value averages was required to characterize compound subsets (instead of individual compounds). In addition, requiring low standard deviations as additional objectives favored ultimate solutions in which the majority of compounds populated preferred value ranges.

(e) As in our previous PSO analysis,[7] the inertia weight was set to $w = 0.721348$ and the confidence coefficients were set to $c_1 = c_2 = 1.193147$.[20]

(f) Swarms were initialized with 30 particles using MACCS structural keys as a representation. Hence, the search space was a 166-dimensional structural fragment/pattern space. A total of 2000 evaluation steps were carried out, and a maximum of 200 nondominated solutions were stored per optimization trial. If more than 200 qualified solutions were obtained, the best 200 solutions were retained. Each data set was searched for subsets consisting of 5 or 10 compounds. In each case, 10 independent MO-PSO trials were carried out and nondominated solutions were sampled.

(g) During each iteration of the optimization process, the Euclidean distance was calculated between particle position x_{iD} and compounds in a data set. Compounds were ranked in the order of increasing distance from particle positions. The n (5 or 10) top-ranked compounds were selected to compute the discontinuity score and the average and standard deviation of the physicochemical descriptors. Nondominated solutions were sampled with respect to the 11 objective functions. These solutions were used as "leaders" to update the position of the particles in the swarm. Each particle was randomly assigned to a leader from the top 10% of nondominated solutions (ranked according to the discontinuity score). At the end of each trial, nondominated solutions yielding a discontinuity score of at least 0.7 were retained. The solutions with highest discontinuity score were selected for which all average descriptor values fell into preferred (green) ranges.

All MO-PSO calculations were carried out with in-house generated programs.

7. MO-PSO RESULTS

7.1. Study Objective. The primary goal of our study has been the development of a methodology to search for small subsets of compounds in data sets that are SAR-informative and have favorable (ADME-friendly) physicochemical property ranges. For this purpose, we have implemented an MO-PSO approach. During multiobjective optimization, 11 objective functions were taken into account, 10 of which were descriptor-based, and nondominated solutions were sampled and sorted as described above. Given the requirement of SAR discontinuity among qualifying active compounds, the search protocol prioritized compound subsets representing discontinuous local SARs and selected best-scoring solutions with desired physicochemical property profiles.

7.2. Qualifying Solutions. The 10 compound data sets we searched were selected to cover the entire range of global SAR phenotypes, from mostly discontinuous over heterogeneous to

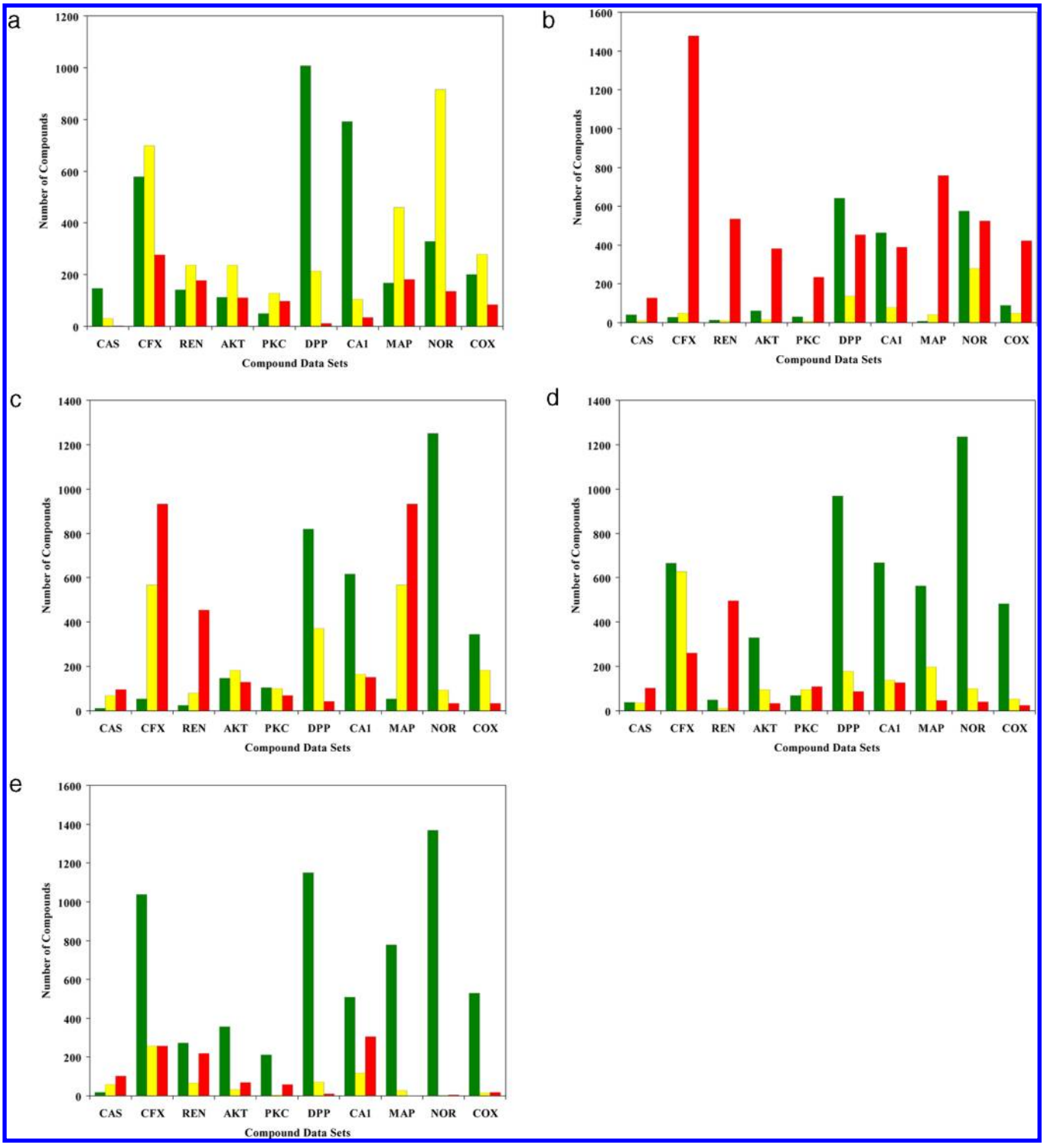


Figure 2. Descriptor value range distribution. For each descriptor, the value distribution in each activity class is assessed following the traffic light value range classification scheme specified in the text. For each activity class, the numbers of compounds falling into the three different value ranges are reported. (a) LogP, (b) LogS, (c) MW, (d) RTB, (e) TPSA.

mostly continuous global SARs. Regardless of the SAR phenotype, small compound subsets representing discontinuous local SARs (with a discontinuity score of at least 0.7) were detected in all data sets. However, only in 5 of 10 data sets, subsets containing 5 compounds were identified that met our SAR and descriptor value range criteria, as reported in Table 2. These activity classes included AKT, DPP, CA1, NOR, and COX. Thus, descriptor value ranges, rather than SAR characteristics, were a limiting factor in a number of cases, which can be rationalized on the basis of Figure 2. For example, for activity classes DPP and NOR, many compounds populated preferred value ranges of all five (DPP) or four of five (NOR) descriptors. By contrast, in CFX (1553 compounds; the largest set) or REN for which no qualifying solutions were obtained, many compounds displayed unfavorable values for two (CFX) or three (REN) descriptors. Thus, in these cases, it was very

dx.doi.org/10.1021/ci300402g | J. Chem. Inf. Model. 2012, 52, 2848−2855

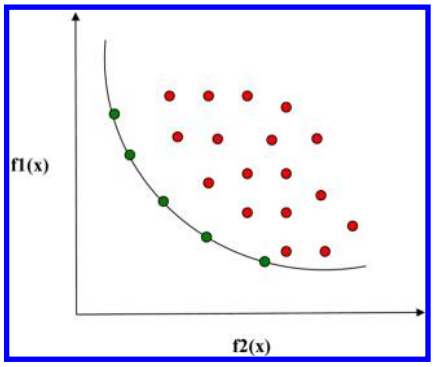


Figure 3. Pareto optimization. The concept of Pareto optimization is illustrated for two objective functions $f_1(x)$ and $f_2(x)$. Solutions colored in red are dominated by $f_1(x)$ or $f_2(x)$. By contrast, solutions colored in green are nondominated solutions.

Table 2. Nondominated Solutions for Compound Data Sets[a]

no.	set	no. compound subsets
1	AKT	1
2	DPP	113
3	CA1	14
4	NOR	69
5	COX	4

[a]MO-PSO successfully identified subsets of five compounds meeting our search criteria in 5 of 10 different data sets. For each of these five sets, the number of qualifying subsets (nondominated solutions) is reported.

difficult, if not impossible, to find compounds meeting all objectives.

For the five activity classes in Table 2, the number of qualifying compound subsets ranged from 1 (AKT) to 113 (DPP). For COX, CA1, and NOR, 4, 14, and 69 qualifying solutions were identified, respectively. Thus, most solutions were found for DPP, followed by NOR. As reported in Table 1, DPP was globally heterogeneous in its SAR character and thus expected to contain regions of local SAR discontinuity. By contrast, NOR displayed a largely continuous SAR phenotype. Nevertheless, it also contained a variety of compound subsets with favorable descriptor settings that represented discontinuous local SARs, which were successfully detected. DPP and NOR also were among the three largest data sets we investigated, comprising 1232 and 1378 compounds, respectively. In subsequent calculations, qualifying subsets consisting of 10 (instead of 5) compounds were only detected for these two data sets including 22 different subsets for DPP and 3 for NOR.

It should also be noted that failures to identify qualifying solutions in data sets might often be due to narrowly defined physicochemical property value ranges that are not well represented. Relaxing individual value ranges increases the probability of identifying qualifying solutions. The Pareto-based optimization selects subsets of compounds by simultaneous optimization of multiple objective functions and searching for nondominated solutions. Instead, one could also apply other approaches such as, for example, weighted summation, which would require prior knowledge to assign relative weights to objective functions. Changes in relative weights might substantially influence the calculations. Regardless, narrowly defined property value ranges constrain calculations using alternative methods and reduce the likelihood of identifying qualifying solutions.

In addition, one should consider how outliers might influence the calculations. Statistical analysis might be carried out to identify likely outliers with respect to individual properties and filter data sets prior to optimization trials to remove potential outliers. However, for PSO, outliers do not present a major limitation because particles share information about neighbors in the swarm. This typically leads to the exclusion of outliers from preferred solutions. In PSO, outliers will likely not be selected and their presence is not expected to negatively affect the selection of other subsets of compounds.

7.3. Preferred Compound Subsets. In Figure 4, compound subsets are shown for all five activity classes that yielded highest discontinuity scores with average descriptor values consistently falling into green ranges. Subset discontinuity scores are reported as well as individual and average property descriptor values. Figure 4a shows the only qualifying solution identified for AKT (458 compounds), yielding a discontinuity score of 0.81. These compounds were only moderately similar. Four of them were weakly potent in the micromolar range, but one had very high potency (0.59 nM). The least potent compound (23.9 μM) had consistently favorable descriptor values whereas the most potent one yielded an unfavorable LogS value. The remaining compounds also displayed one or two intermediate descriptor values and one of them had an unfavorable TPSA value. Similar observations were made for the best of four qualifying solutions for COX shown in Figure 4b. In this subset yielding a discontinuity score of 0.87, two compounds were remotely similar to three others and associated with descriptor liabilities. One of these compounds was highly potent (2.4 nM), whereas the other was only lowly potent (16.8 μM). Hence, this subset contained limited SAR information. Different from AKT and COX, the NOR compound subset in Figure 4c, with a discontinuity score of 0.89, contained structurally very similar compounds spanning a large potency range. Four of five compounds in this series had consistently favorable descriptor values. Only for one compound, LogP and LogS values fell into the intermediate range. Furthermore, the top-ranked DPP compound subset in Figure 4d was highly relevant. In this case, a discontinuity score of 1.00 was observed that resulted from the presence of very similar compounds only distinguished by stereochemistry or a single fluorine substituent. In addition, these compounds covered a large potency range from 6.6 nM to 67 μM. Hence, this compound series was associated with high SAR information content. Moreover, the descriptor values of all compounds consistency fell into favorable ranges, without an exception. Similar observations were made for the series of CA1 sulfonamide inhibitors shown in Figure 4e. Only one moderately potent compound (415 nM) had a single unfavorable LogS value; all other descriptor values of this series mapped to favorable ranges. Comparison of the four other compounds, one of which was highly potent (2.1 nM), whereas the others were only borderline active, immediately revealed detrimental effects of large substituents at the para-position of the benzene moiety or increasing volume of the fluorinated ring. Thus, in this case, SAR determinants were immediately apparent. Taken together, these examples illustrate that MO-PSO was capable of automatically extracting compound subset from diverse data sets that met multiple SAR- and ADME-related objectives, without the need to compare compounds individually. Of course, the calculations can only succeed if qualifying compounds are available in a data set, which is not guaranteed, even if a data set is large.

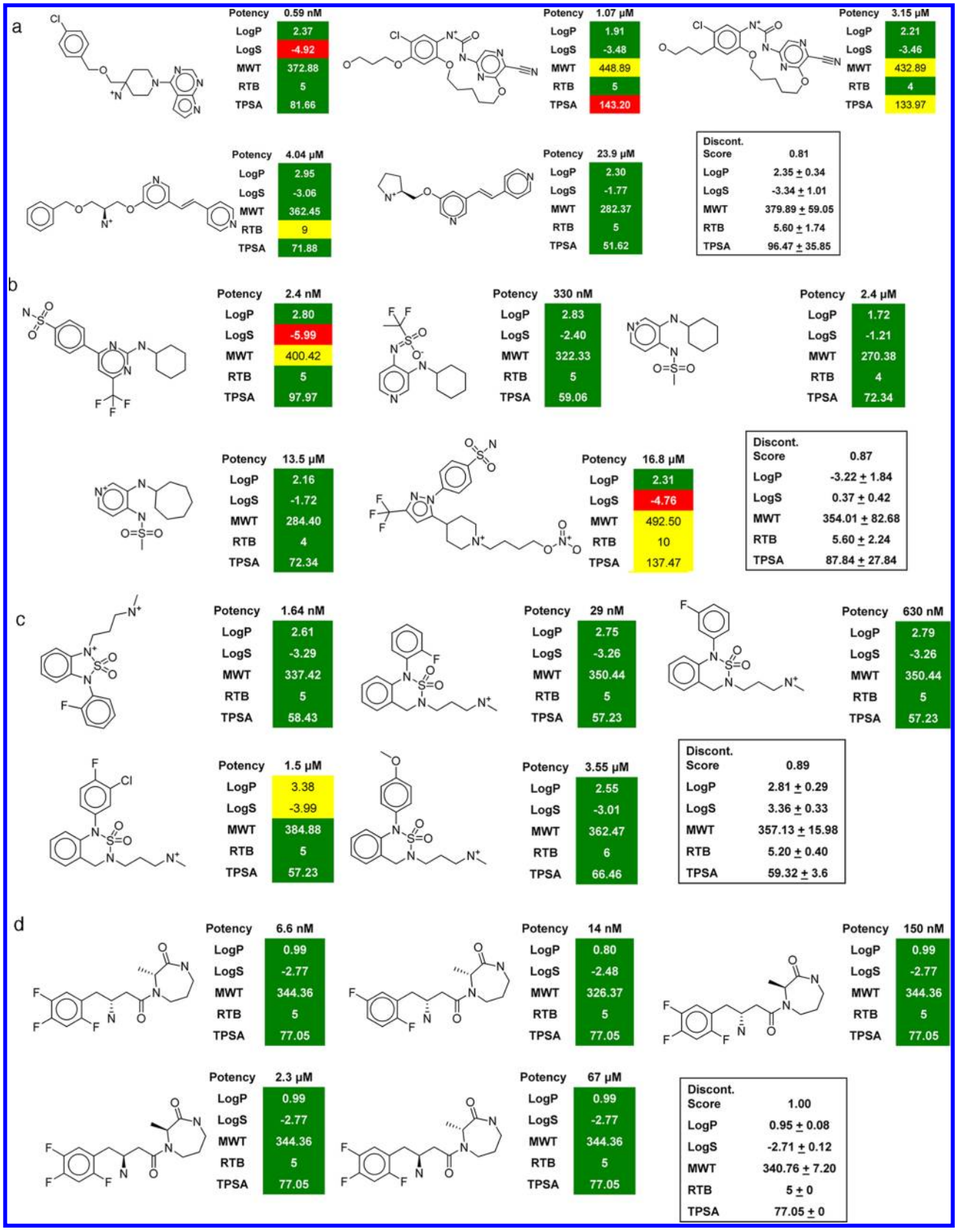


Figure 4. continued

dx.doi.org/10.1021/ci300402g | *J. Chem. Inf. Model.* 2012, 52, 2848−2855

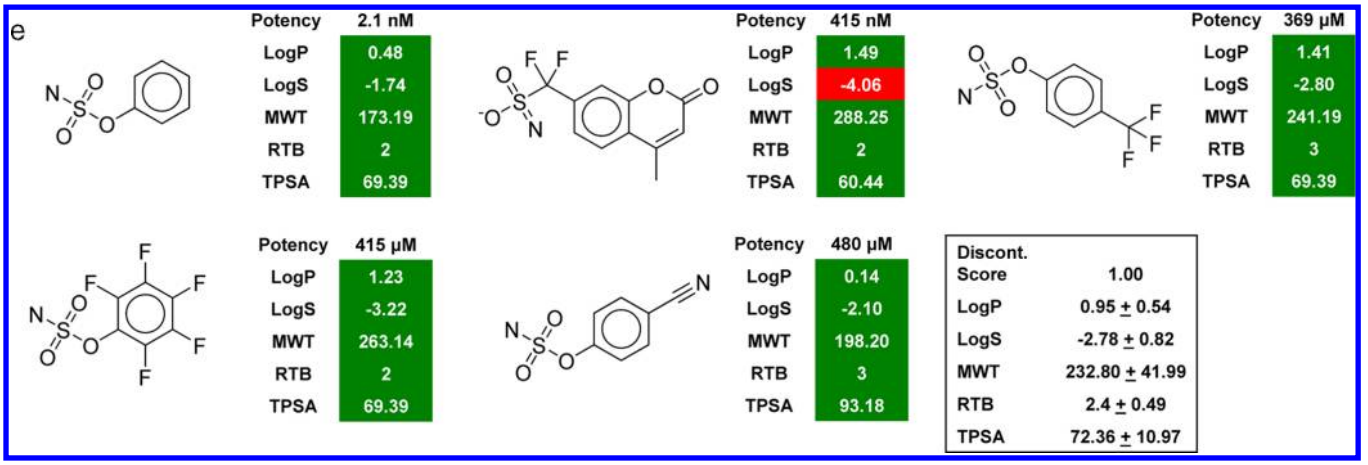


Figure 4. Preferred solutions. Top-ranked subsets containing five compounds are displayed. For each compound, its descriptor values are reported and classified. These compound subsets had the largest discontinuity score and average descriptor values consistently falling into favorable value ranges. The box on the lower right reports the discontinuity score for the subset and the average descriptor values and standard deviations. (a) AKT, (b) COX, (c) NOR, (d) DDP, (e) CA1.

8. CONCLUSIONS

Previously, we have successfully applied PSO to extract compounds forming discontinuous local SARs from large data sets. The selection of compounds for medicinal chemistry applications typically requires the consideration of additional parameters beyond compound potency. Therefore, we have been interested in developing a multiobjective search scheme to rapidly identify series of compounds in data sets that are not only SAR-informative but also display physicochemical properties associated with favorable ADME characteristics. To these ends, a multiobjective PSO approach was designed that employs Pareto optimization and combines the evaluation of SAR information with physicochemical property distribution analysis. In our proof-of-concept investigation, we have utilized the SARI discontinuity score in combination with five ADME-related numerical property descriptors to guide MO-PSO calculations and demonstrated that compound subsets with desired SAR and property distributions subsets could be readily extracted from different data sets. The search protocol reported herein and the chosen objectives can be easily modified given specific requirements of other applications. Automated extraction of SAR-informative compound subsets with additional desired properties should be helpful to provide candidates for further chemical exploration in an efficient manner.

AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

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

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

(4) Leeson, P. D.; Davis, A. M.; Steele, J. Drug-like Properties: Guiding Principles for Design − or Chemical Prejudice. *Drug Discovery Today* **2004**, *1*, 189−195.

(5) Gillet, V. J.; Willet, P.; Bradshaw, J. Selecting Combinatorial Libraries to Optimize Diversity and Physical Properties. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 167−177.

(6) Wright, T.; Gillet, V. J.; Green, D. S.; Pickett, S. D. Optimizing the Size and Configuration of Combinatorial Libraries. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 381−390.

(7) Namasivayam, V.; Iyer, P.; Bajorath, J. Extraction of Discontinuous Structure-Activity Relationships from Compound Data Sets through Particle Swarm Optimization. *J. Chem. Inf. Model.* **2011**, *51*, 1545−1551.

(8) Namasivayam, V.; Iyer, P.; Bajorath, J. Exploring SAR Continuity in the Vicinity of Activity Cliffs. *Chem. Biol. Drug Des.* **2012**, *79*, 22−29.

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

(10) Eberhart, R. C.; Shi, Y. Comparison between Genetic Algorithms and Particle Swarm Optimization. *Evol. Programming VII, Lect. Notes Comput. Sci.* **1998**, *1447*, 611−616.

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

(12) 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.

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

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

(15) *MOE (Molecular Operating Environment)*; Chemical Computing Group Inc.: Montreal, Quebec, Canada, 2011.

(16) Lobell, M.; Hendrix, M.; Hinzen, B.; Keldnich, J.; Meier, H.; Schmeck, C.; Schohe-Loop, R.; Wunberg, T.; Hillisch, A. In silico ADMET Traffic Lights as Tool for the Prioritization of HTS Hits. *Chem. Med. Chem.* **2006**, *1*, 1229−1236.

(17) 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.* **2012**, *40*, D1100−D1107.

(18) Reyes-Sierra, M.; Coello, C. A. C. Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art. *Int. J. Comput. Intell. Res.* **2006**, *2*, 287−309.

(19) Wickramasinghe, W.; Li, X. Choosing Leaders for Multi-objective PSO Algorithms Using Differential Evolution. *Proceedings of the Seventh International Conference on Simulated Evolution and Learning (SEAL'08)*; 2008; pp 249−258.

(20) Clerc, M. *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: 1744-8050.