JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

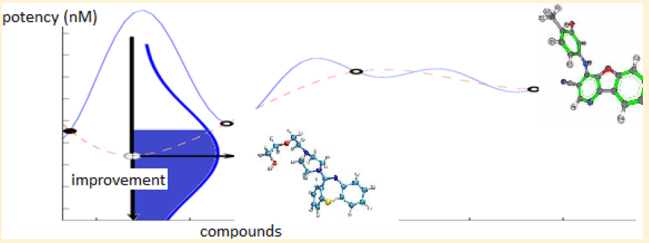
Predicting Potent Compounds via Model-Based Global Optimization

Mohsen Ahmadi,[†,‡] Martin Vogt,[†,§] Preeti Iyer,[†,§] Jürgen Bajorath,[†,§] and Holger Fröhlich*[,‡]

[†]Department of Life Science Informatics and [‡]Algorithmic Bioinformatics, Bonn-Aachen International Center for IT, and [§]LIMES, Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Dahlmannstrasse 2, D-53113 Bonn, Germany

Supporting Information

ABSTRACT: Finding potent compounds for a given target in silico can be viewed as a constraint global optimization problem. This requires the use of an optimization function for which evaluations might be costly. The major task is maximizing the function while minimizing the number of evaluation steps. To solve this problem, we propose a machine learning algorithm, which first builds a statistical QSAR-model of the SAR landscape and then uses the model to identify regions in compound space having a high probability to contain a highly potent compound. For this purpose, we devise the so-called *expected potency improvement* (EI) criterion to rank candidate compounds with respect to their likelihood to exhibit higher potency than the most active compound in the training data. Therefore, this approach significantly differs from a purely prediction-oriented classical QSAR model. The method is superior to a nearest neighbor approach as significantly fewer evaluation steps are needed to identify the most potent compound for the given target.



INTRODUCTION

The biologically relevant chemical space is vast. Hence, in silico approaches for virtual screening, compound selection, and activity prediction have experienced increasing interests during the past decade.[1,2] Among these are QSAR-type approaches, which have been utilized in many different ways. Generally, these techniques rely on the assumption that structurally similar compounds exhibit a similar biological activity, which is often, but not always true.[2] Regardless, a critical question for in silico compound assessment is how to define structural similarity in a sensible way.[3,4]

QSAR modeling is often associated with prediction inaccuracies, depending on the nature of the underlying SAR, which might limit practical utility. Typically, the ultimate goal of QSAR modeling is the identification or design of a highly potent compound. Given the uncertainties that are often associated with QSAR predictions, one would like to preferentially focus model building and application to regions in compound space that are SAR-informative and have a high probability to contain highly potent compounds. This challenge can be viewed as a global optimization problem, for which a chosen optimization function should be maximized while controlling the number of evaluation steps that are required. The optimization problem is constrained by the possibility to synthesize a candidate compound.

To address this problem, we propose a machine learning algorithm, which generates a QSAR-like-model and searches regions in compound space that are likely to contain highly potent compounds. A critically important aspect of the approach is the introduction of the so-called *expected potency improvement* (EI) criterion to rank candidate compounds with respect to their likelihood to exhibit a higher potency than the most active compounds in the training data. In extensive simulation studies, we compare the EI method with random compound selection, nearest neighbor selection, and a pure QSAR model to demonstrate that the EI criterion can significantly reduce the number of evaluation steps to identify the most active compound for a given target.

METHODS

Model Based Global Optimization. *Basic Idea.* The basic idea underlying our technique is that finding a compound with maximal potency can be viewed as an optimization process within a multidimensional SAR landscape, the dimensions of which are defined by chosen molecular descriptors as well as compound potency values. Accordingly, each molecule M corresponds to a pair $(\mathbf{x}, y)$, in which $\mathbf{x} \in \mathbb{R}^d$ is a d-dimensional vector of molecular descriptors and $y \in \mathbb{R}$ is the compound potency.

The key question we address is the following: Given a training set $D = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n, y_n)\}$ of n compounds with available potency values, in which region of an SAR landscape would it be most likely to find the most potent compound?

This task can be interpreted as finding the global maximum of a nonlinear response surface, where responses are compound potencies. Global optimization of nonlinear response surfaces has been studied in the engineering field and our methodology results from an adaptation of an optimization algorithm by

Received: October 3, 2012
Published: January 30, 2013

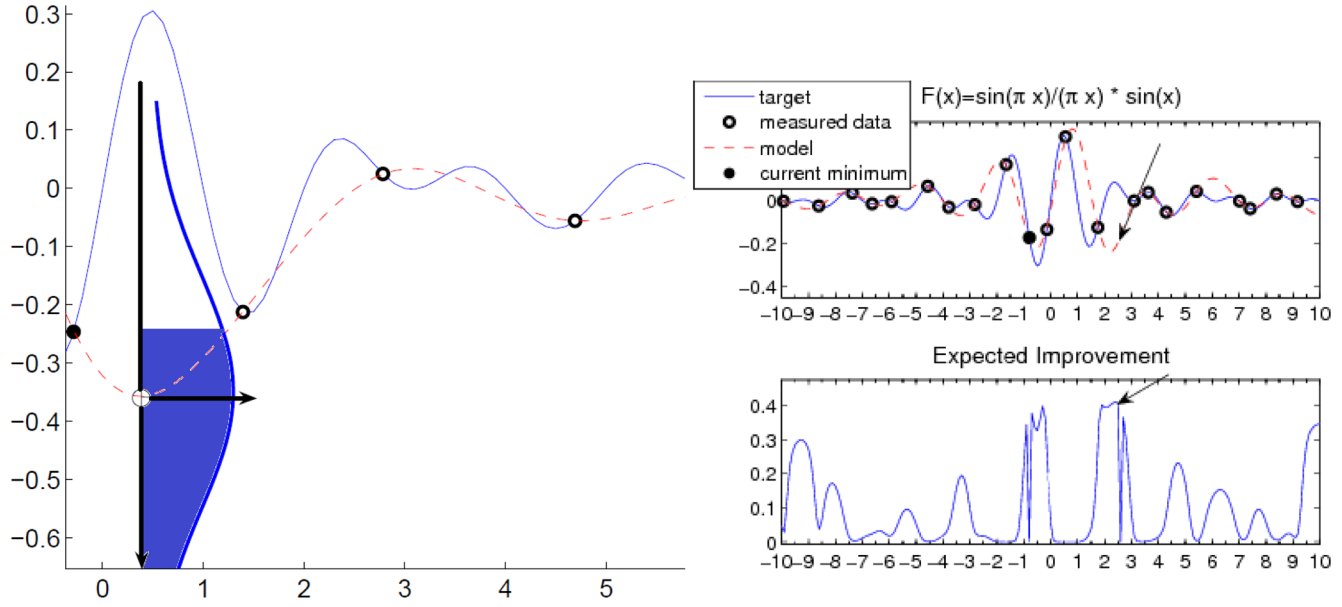


Figure 1. Potency and expected potency improvement. (left) Potency improvement illustrated for a candidate test compound (white circle) in a 1D toy example. The predictive distribution for the test compound is a normally distributed random variable. The potency improvement corresponds to the shaded part of this distribution. "Target" indicates the true SAR landscape, which is approximated by a response surface model that is constructed on the basis of measured data. The solid black dot indicates the most potent among all so far tested compounds. (right) *Expected* potency improvement viewed as a function over compound space (1D toy example with target function $F(x) = ([\sin(\pi x)]/\pi x)\cdot\sin(x)$). The arrow indicates the sample point (i.e., candidate compound) with highest expected improvement.

Jones et al.[5] Following this approach, the optimization process is guided by an iteratively refined statistical model of the response surface, which preferentially facilitates compound selection from regions of the SAR landscape where (i) model uncertainty is high (corresponding to a rough SAR landscape) and (ii) the model estimates close proximity to the global maximum.

Modeling the Response Surface via Gaussian Processes. We employ a Gaussian process regression[6,7] to model the response surface of the SAR landscape. The underlying idea is that any regression function f: $\mathbb{R}^d \rightarrow \mathbb{R}$ can be interpreted as a sample drawn from a Gaussian distribution over functions:

$$f(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

where $m(\mathbf{x})$ is the mean function (typically 0) and $k(\mathbf{x}, \mathbf{x}')$ is a covariance or kernel function measuring the similarity of descriptor vectors $\mathbf{x}, \mathbf{x}'$. The notation GP distinguishes between the conventional Gaussian distribution over data points and the one used here. We use a squared exponential covariance function:

$$k(\mathbf{x}, \mathbf{x}') = {\sigma_0}^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\gamma_i(x_i - x_i')^2\right) + {\sigma_n}^2\delta(\mathbf{x} = \mathbf{x}') \tag{2}$$

where $\theta := (\gamma_1, ..., \gamma_d, \sigma_0, \sigma_n)^{\mathrm{T}}$ are adaptable hyper-parameters, which we fit to the training data via maximum likelihood, with $\delta(\mathbf{x} = \mathbf{x}')$ being the indicator function.[7] More precisely, let $K = (k(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$ be the covariance or kernel matrix for training data D. The log marginal likelihood is then given as

$$\log p(\mathbf{y}|X, \theta) = -\mathbf{y}^{\mathrm{T}}{K_y}^{-1}\mathbf{y} - \log|K_y| - n\log 2\pi \tag{3}$$

where $K_y = K + {\sigma_n}^2 I$, $\mathbf{y}$ is the vector of compound potencies and X is the descriptor matrix for the training data. Maximum likelihood estimation corresponds to maximizing the log marginal likelihood with respect to hyper-parameters θ. This can be achieved via numerical procedures such as conjugate gradient descent. For this purpose we need to calculate the partial derivatives of the log marginal likelihood with respect to hyper-parameters θ_j, $j = 1, 2, ..., d + 2$:

$$\frac{\partial}{\partial\theta_j}\log p(\mathbf{y}|X, \theta) = \mathbf{y}^{\mathrm{T}}{K_y}^{-1}\frac{\partial K_y}{\partial\theta_j}{K_y}^{-1}\mathbf{y} - \mathrm{tr}\left({K_y}^{-1}\frac{\partial K_y}{\partial\theta_j}\right) \tag{4}$$

After fitting, the posterior predictive distribution of the potency for a candidate compound with descriptor vector $\mathbf{x}_*$ can be computed as

$$f(\mathbf{x}_*)|D \sim N(\mathbf{k}_*{}^{\mathrm{T}}{K_y}^{-1}\mathbf{y}, k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*{}^{\mathrm{T}}{K_y}^{-1}\mathbf{k}_*) \tag{5}$$

where $\mathbf{k}_* = (k(\mathbf{x}_*, \mathbf{x}_1), ..., k(\mathbf{x}_*, \mathbf{x}_n))^{\mathrm{T}}$.

Optimizing Expected Potency Improvement. The advantage of Gaussian process regression over many other regression techniques is that the procedure is well-suited for high dimensional data (thanks to the use of kernel functions) and enables the estimation of the variance (i.e., uncertainty) of each prediction. The optimzation of compound potencies corresponds to the minimization of potency values given as molar concentrations. Here, we consider log-potency values for each compound. For each candidate compound, we now define its *potency improvement* as

$$I(\mathbf{x}_*) = \max(0, y_{\min} - p(f(\mathbf{x}_*)|D)) \tag{6}$$

where $y_{\min}$ is the best potency value (i.e., lowest log-potency value) in the training data. It should be noted that $f(\mathbf{x}_*)|D$ is a random variable, and thus $I(\mathbf{x}_*)$ is also a random variable (Figure 1).

Let $\mu(\mathbf{x}_*) = E[f(\mathbf{x}_*)|D]$ denote the predicted potency for compound $\mathbf{x}_*$ and $\sigma^2(\mathbf{x}_*) = E[(f(\mathbf{x}_*)|D - \mu(\mathbf{x}_*))^2]$ the variance

dx.doi.org/10.1021/ci3004682 | J. Chem. Inf. Model. 2013, 53, 553−559

of this prediction. The expectation value of $I(\mathbf{x}_*)$, which we term *expected potency improvement* (EI), can be calculated as

$$E[I(\mathbf{x}_*)] = (y_{\min} - \mu(\mathbf{x}_*))\Phi\left(\frac{y_{\min} - \mu(\mathbf{x}_*)}{\sigma(\mathbf{x}_*)}\right) + \sigma(\mathbf{x}_*)\phi\left(\frac{y_{\min} - \mu(\mathbf{x}_*)}{\sigma(\mathbf{x}_*)}\right) \quad (7)$$

where Φ and ϕ are the standard normal distribution and density function, respectively.[5] The expected potency improvement can be viewed as a function $\mathbb{R}^d \rightarrow \mathbb{R}$ over the compound space (Figure 1). Given any candidate compound $\mathbf{x}_*$ in the SAR landscape, we can thus predict the expected potency improvement compared to the best compound in our training data. This can be done very quickly and efficiently thanks to eq 5. Once the candidate compound's potency has been obtained, the Gaussian process model can be retrained to further improve the initial response surface model.

Modified Variance Estimation. In our initial simulations, we observed that the prediction variance, $\sigma^2(\mathbf{x}_*)$, was often significantly underestimated, leading to a very low or zero expected potency improvement. To overcome this problem, we replaced

$$\sigma^2(\mathbf{x}_*) \leftarrow \max\{\sigma^2(\mathbf{x}_*), (y_{NN} - \mu(\mathbf{x}_*))^2\} \quad (8)$$

where NN is the index of the closest training compound to $\mathbf{x}_*$. The closest training compound was determined via the generalized Tanimoto coefficient for numerical descriptors:

$$T(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\|^2 \|\mathbf{b}\|^2 - \langle \mathbf{a}, \mathbf{b} \rangle} \quad (9)$$

The rationale behind this modification was that the model uncertainty (i.e., the prediction variance) should increase, the more the predicted potency of $\mathbf{x}_*$ deviates from the potency of its nearest neighbor. This deviation depends on the smoothness of the SAR landscape as well as the distance of $\mathbf{x}_*$ to $\mathbf{x}_{NN}$: The larger the distance, the higher the likelihood that y_{NN} significantly differs from $\mu(\mathbf{x}_*)$.

On the other hand, if the SAR landscape is very smooth, and $\mathbf{x}_*$ is proximal to $\mathbf{x}_{NN}$, the residual $\mu(\mathbf{x}_{NN}) - \mu(\mathbf{x}_*)$ will be small. Moreover, a smooth landscape usually allows for a good model fit, which makes $y_{NN} - \mu(\mathbf{x}_{NN})$ also small. Therefore, in this case, y_{NN} will not substantially differ from $\mu(\mathbf{x}_*)$.

Descriptor Selection. Since many descriptors are expected to be irrelevant or redundant for the response surface model, before each training step descriptors were selected that revealed a significant Spearman rank correlation with the logarithmic potency values (q-value < 5%) of the training compounds. Q-values were computed from original p-values via the multiple testing correction method by Benjamini and Hochberg.[8]

SAR Index (SARI). In order to characterize the correlation of compound similarities to potencies in our compound data sets, the SAR index (SARI) was applied.[9] The SARI is defined as

$$\text{SARI} = \frac{1}{2}(\text{score}_{\text{cont}} + (1 - \text{score}_{\text{disc}})) \quad (10)$$

where $\text{score}_{\text{cont}}$ and $\text{score}_{\text{disc}}$ range between 0 and 1 and describe the degree of SAR landscape continuity (similar and increasingly diverse structures have similar potencies) and discontinuity (similar structures have different potencies), respectively. The discontinuity score depends on a defined similarity cutoff, which was set for each data set to the median Tanimoto coefficient value (eq 9).

RESULTS

Data Sets and Descriptors. For 12 human targets, compound data sets containing at least 100 molecules were assembled from the ChEMBL database, version 13.[10] Defined end point measurements with the highest confidence score were selected. K_i values were utilized whenever available. Alternatively, IC50 values were used. For compounds with multiple measurements, the geometric mean was calculated to obtain the final potency values. The data sets are reported in Table 1.

Table 1. Data Set Identifiers, Corresponding Target Names, Number of Compounds, and Minimum/Maximum Compound Potencies Are Reported

			potency [nM]	
data set ID	target	size	minimum	maximum
3	phosphodiesterase 5A	274	0.007	98000
235	leukocyte elastase	308	0.022	1700000
242	aldose reductase	135	0.11	40000
10131	caspase-3	281	0.0001	60900
10185	adenosine kinase	100	0.2	8500
10561	oligopeptide transporter small intestine isoform	164	0.09	28000000
10702	methionine aminopeptidase 2	162	0.39	57500
10911	phenylethanolamine *N*-methyltransferase	148	2.8	890000
11407	c-Jun *N*-terminal kinase 2	109	11	404000
11624	caspase-1	177	0.009	200000
11910	phosphodiesterase 7A	129	3.5	28000
12261	c-Jun *N*-terminal kinase 1	250	2	385000

A total of 166 numerical 2D descriptors were calculated using the Molecular Operating Environment (MOE).[11]

Simulation Setup. Each data set was randomly divided to yield equally sized training and test sets. The training data was used to fit the Gaussian process model. The test set was then scanned for the compound with maximal expected potency improvement. This compound was subsequently added to the training data to further refine the SAR surface model (thereby simulating iterative compound generation and evaluation cycles). The process of adding compounds according to the expected potency improvement criterion was continued until all compounds in the original test data were selected. At the end, the number of simulation steps was determined, which the algorithm required to select the most potent compound in the original test set. To obtain statistically meaningful results, simulations were carried out using 25 independently selected training and test sets. The EI algorithm was compared to random compound selection (RA), a nearest neighbor (NN) approach, and a Gaussian process QSAR model (GP). In RA, a compound was randomly selected and added to the training data. In NN, the test compound was selected that was most similar on the basis the generalized Tanimoto similarity (eq 9) to the most potent training compound. In GP, the test compound, which was predicted to be most potent according to the fitted Gaussian process model, was added to the training data.

dx.doi.org/10.1021/ci3004682 | J. Chem. Inf. Model. 2013, 53, 553−559

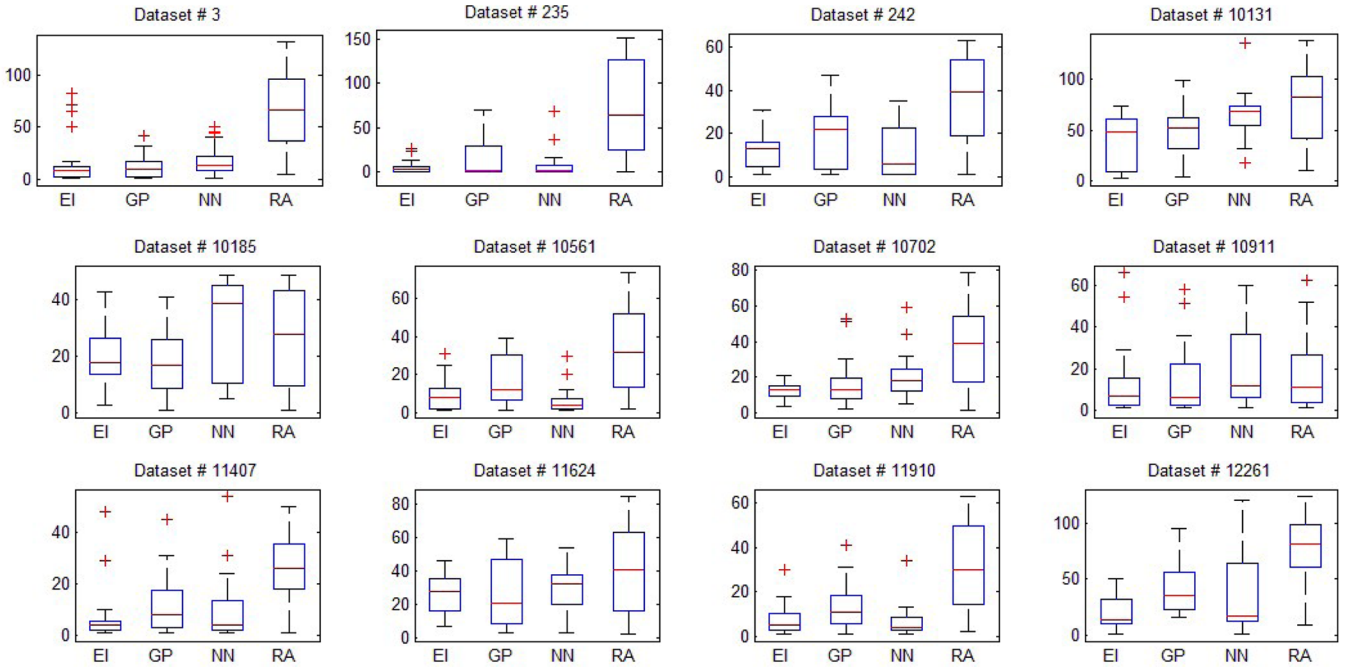


Figure 2. Evaluation steps. Box plot representations for each of the 12 data sets compare the average number of evaluation steps to select the most potent compound from the original test set using EI, RA, NN, and GP.

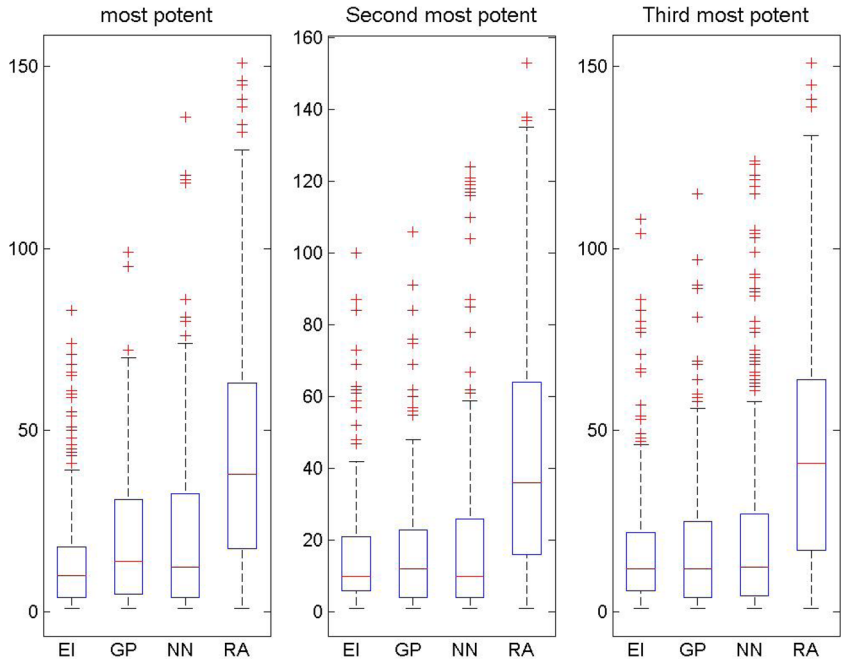


Figure 3. Average performance of different compound selection strategies. Reported is the median number of search steps to obtain the most potent, second most potent, and third most potent compound. Significances of EI vs NN according to a paired Wilcoxon signed rank test: $p = 7.7 \times 10^{-9}$ (most potent compound), $p = 0.09$ (second most potent compound), $p = 0.09$ (third most potent compound). Significances of EI vs GP: $p = 4.7 \times 10^{-7}$ (most potent compound), $p = 0.92$ (second most potent compound), $p = 0.82$ (third most potent compound). No significant differences between GP and NN were observed. All other differences are significant with $p < 1 \times 10^{-10}$.

EI vs NN and GP Based Compound Selection. On the basis of these simulations, the EI algorithm led to a significant reduction ($p < 5\%$ for the *paired Wilcoxon signed rank test*) of the number of evaluation steps compared to NN that were required to obtain the most potent compound from the original test sets for six data sets, including nos. 10131, 10185, 10702, 10911, 11407, and 12261 (Figure 2). Furthermore, EI outpeformed GP significantly for four data sets (nos. 10561, 11407, 11910, and 12261). For all but two data sets (nos. 10911 and 10185), EI needed significantly fewer search steps than random compound selection. NN was significantly better than RA in all but four data sets (nos. 10131, 10185, 10911, and 11624) and GP in all but three data sets (nos. 10185, 10911, and 11624).

When averaged over all data sets, the median number of evaluation steps was significantly lower for EI compared to NN

dx.doi.org/10.1021/ci3004682 | J. Chem. Inf. Model. 2013, 53, 553−559

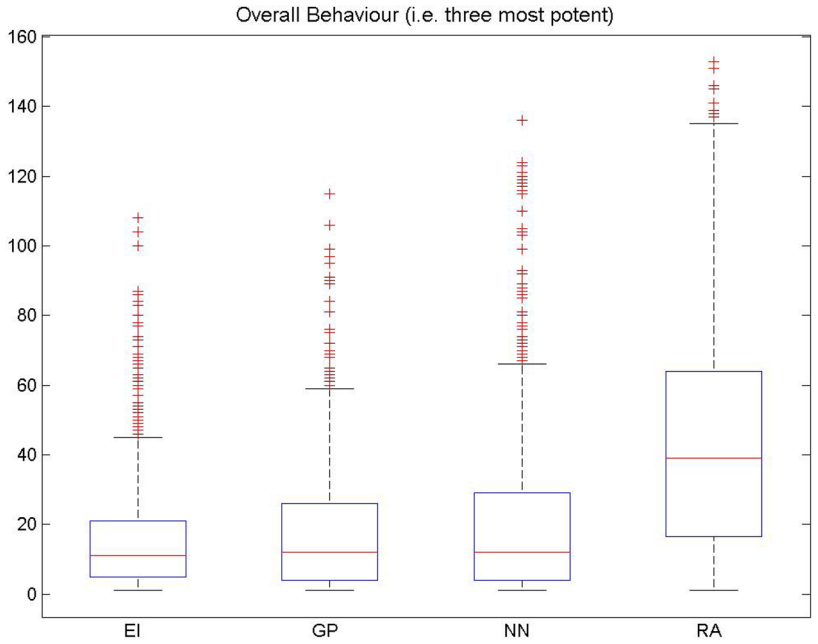


Figure 4. Average performance of different compound selection strategies. Depicted is the median number of search steps to find any of the three most potent compounds. Significances according to Wilcoxon's signed rank test: $p = 1.8 \times 10^{-7}$ (EI vs NN), $p = 0.004$ (EI vs GP), $p = 0.1$ (GP vs NN), $p = 7.2 \times 10^{-89}$ (EI vs random), $p = 3.5 \times 10^{-60}$ (NN vs random), $p = 7.1 \times 10^{-78}$ (GP vs random).

(paired Wilcoxon signed rank test: $p = 7.7 \times 10^{-9}$ for the most potent compound as well as compared to GP ($p = 4.7 \times 10^{-7}$, Figure 3). The same behavior was observed for the number of steps required to find any of the three most potent compounds (EI vs NN $p = 1.8 \times 10^{-7}$, EI vs GP $p = 0.004$, Figure 4).

In order to better understand the observed improvements by the EI algorithm in several, but not all cases, we computed for each data set the SAR index (SARI),[9] as a measure of the smoothness (high SARI score, high continuity score) or ruggedness (low SARI score, high discontinuity score) of the global SAR landscape. This analysis suggested that the EI method preferentially (four out of six cases for EI vs NN and three out of four cases for EI vs GP) performed well on data sets with a relatively smooth SAR landscapes (SARI > 0.4, Table 2). On the other hand, in cases of high discontinuity (data set nos. 3, 235, and 11624), no significant difference of EI to the other methods was observed.

Descriptor Selection Frequency. During our simulation, we recorded the frequency by which individual descriptors were selected (25 times at maximum). In general a large heterogenity of selected descriptors and their selection frequency was observed across data sets. Whereas, for example, on data set no. 3 a large proportion of descriptors was selected consistently (several volume surface area related ones), on data set no. 10911 only one descriptor (rsynth) was chosen 11 out of 25 times (see figures in the Supporting Information). In general we could neither see any correlation between the performance of the EI algorithm and descriptor selection stability nor observe any dependency of the EI method performance on the selection of certain descriptors.

Influence of Training Set Size. We also investigated to what extent the EI method was dependent on the initial size of the training set. Therefore, we randomly removed 25%, 50%, and 75% of the initially selected training compounds and carried out simulations as described above. Supporting Information Figures S1−S3 overall reveal decreasing differences in the number of evaluation steps between EI and NN and GP for training sets of decreasing size. However, even with just 25% of the original training data, the difference was still statistically significant according to Wilcoxon's signed rank test (Figure 5). These findings also illustrate the robustness of the EI approach.

Table 2. SAR Index (SARI) of Used Data Sets

data set	$score_{cont}$	$score_{disc}$	SARI	p (EI vs NN)	p (EI vs GP)
3	0.3102	0.635	0.3376	0.3807	0.3954
235	0.9966	0.5233	0.7366	0.9721	0.1532
242	0.384	0.232	0.576	0.7261	0.0671
10131	0.34	0.5651	0.3874	**1.77×10^{-5}**	0.367
10185	0.2458	0.5145	0.3657	**0.0061**	0.4426
10561	0.7926	0.996	0.3983	*0.0351*[a]	**0.0156**
10702	0.6671	0.1735	0.7468	**0.0012**	0.1652
10911	0.6346	0.3241	0.6553	**0.0449**	0.728
11407	0.3676	0.2054	0.5811	**0.0489**	**0.0215**
11624	0.1719	0.8102	0.1809	0.2875	0.968
11910	0.2176	0.3375	0.4401	0.6062	**0.0017**
12261	0.3664	0.2235	0.5715	**0.0257**	**0.0006**

[a]EI is outperformed by NN.

CONCLUSION

We have introduced a computational model based global optimization strategy to identify maximally potent compounds. In particular, the strategy made use of the expected potency improvement criterion, which balances the prediction variance of a QSAR-like model with the likelihood to identify a more potent compound. It is worth emphasizing that in contrast to classical QSAR models, the primary goal of the EI approach is not the prediction of potency values, but rather the selection of most potent compounds from an SAR landscape. Our simulation results indicated that the expected potency improvement criterion led to overall better search performance (i.e., fewer evaluation steps) than the nearest neighbor approach that

dx.doi.org/10.1021/ci3004682 | J. Chem. Inf. Model. 2013, 53, 553−559

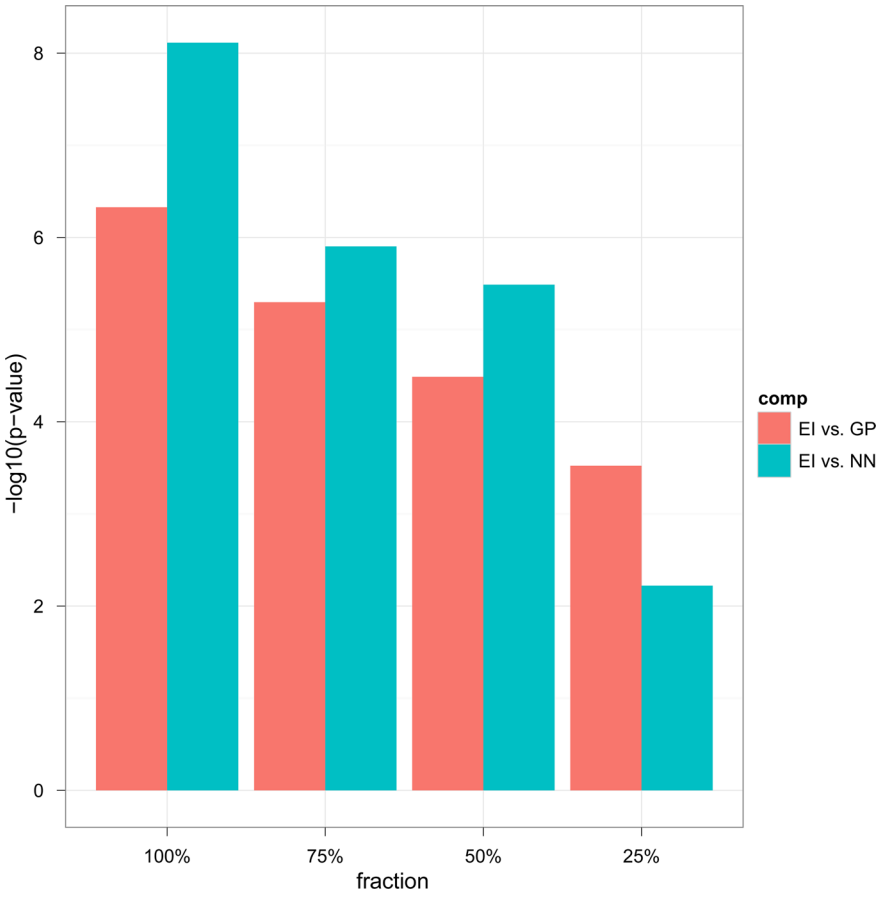


Figure 5. Dependency on training set size. Reported are differences in the median number of evaluation steps for finding the most potent compound between EI and NN and GP for differently sized training sets.

is often applied in virtual screening as well as a QSAR model. This was especially the case when the SAR landscapes of compound data sets were relatively smooth, thus indicating that the topology of SAR landscape influences simulation results, as one might expect. A priori characterization of data sets, for example via SAR indices, is therefore useful to estimate the likelihood to select potent compounds via the EI approach, especially in practical applications.

■ ASSOCIATED CONTENT

Supporting Information

File ci3004682_si_001.pdf: Descriptor selection frequencies for individual data sets. File ci3004682_si_002.pdf: Average performance of different compound selection strategies (75% of original training data). Reported is the median number of search steps to obtain the most potent, second most potent, and third most potent compound. Significances of EI vs NN according to Wilcoxon's signed rank test: $p = 3.25 \times 10^{-6}$ (most potent compound), $p = 4.79 \times 10^{-}5$ (second most potent compound), $p = 0.24$ (third most potent compound). Significances of EI vs GP: $p = 3.25 \times 10^{-5}$ (most potent compound), $p = 0.4$ (second most potent compound), $p = 0.58$ (third most potent compound). There are no significant differences between GP and NN. All other differences are significant with $p < 1 \times 10^{-10}$. File ci3004682_si_003.pdf: Average performance of different compound selection strategies (50% of original training set). Reported is the median number of search steps to obtain the most potent, second most potent, and third most potent compound. Significances of EI vs NN according to Wilcoxon's signed rank test: $p = 1.25 \times 10^{-6}$ (most potent compound), $p = 0.001$ (second most potent compound), $p = 0.001$ (third most potent compound). Significances of EI vs GP: $p = 5.04 \times 10^{-6}$ (most potent compound), $p = 0.001$ (second most potent compound), $p = 0.66$ (third most potent compound). There are no significant differences between GP and NN. All other differences are significant with $p < 1 \times 10^{-10}$. File ci3004682_si_004.pdf: Average performance of different compound selection strategies (25% of original training data). Reported is the median number of search steps to obtain the most potent, second most potent, and third most potent compound. Significances of EI vs NN according to Wilcoxon's signed rank test: $p = 0.006$ (most potent compound), $p = 0.14$ (second most potent compound), $p = 0.002$ (third most potent compound). Significances of EI vs GP: $p = 0.0002$ (most potent compound), $p = 0.47$ (second most potent compound), $p = 0.7$ (third most potent compound). There are no significant differences between GP and NN. All other differences are significant with $p < 1 \times 10^{-10}$. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Böhm, H.-J.; Schneider, G. *Virtual Screening for Bioactive Molecules*; Wiley-VCH: Weinheim, 2000.
(2) Kubinyi, H. Changing Paradigms in Drug Discovery. *Proceedings of the International Beilstein Workshop*, Berlin, June 27, 2004; pp 51−72.
(3) Bender, A.; Glen, R. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.
(4) Eckert, H.; Bajorath, J. *Drug Discovery Today* **2007**, *12*, 225−233.
(5) Jones, D.; Schonlau, M.; Welch, W. J. *Global Optim.* **1998**, *13*, 455−492.
(6) Williams, C. K. I.; Rasmussen, C. E. Gaussian Processes for Regression. In *Advances in Neural Information Processing Systems 8*; MIT Press: Cambridge, MA, 1996; pp 514−520
(7) Rasmussen, C.; Williams, C. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, 2006.
(8) Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc. B* **1995**, *57*, 289−300.
(9) Peltason, L.; Bajorath, J. *J. Med. Chem.* **2007**, *50*, 5571−5578.
(10) ChEMBL. *European Bioinformatics Institute (EBI)*. http://www.ebi.ac.uk/chembl/ (accessed Sep 18, 2012).
(11) Molecular Operating Environment (MOE), Chemical Computing Group: Montreal. http://www.chemcomp.com/ (accessed Sep 18, 2012).