SVM-Based Feature Selection for Characterization of Focused Compound Collections

Evgeny Byvatov and Gisbert Schneider*

Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universität, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany

Received December 9, 2003

Artificial neural networks, the support vector machine (SVM), and other machine learning methods for the classification of molecules are often considered as a "black box", since the molecular features that are most relevant for a given classifier are usually not presented in a human-interpretable form. We report on an SVM-based algorithm for the selection of relevant molecular features from a trained classifier that might be important for an understanding of ligand−receptor interactions. The original SVM approach was extended to allow for feature selection. The method was applied to characterize focused libraries of enzyme inhibitors. A comparison with classical Kolmogorov-Smirnov (KS)-based feature selection was performed. In most of the applications the SVM method showed sustained classification accuracy, thereby relying on a smaller number of molecular features than KS-based classifiers. In one case both methods produced comparable results. Limiting the calculation of descriptors to only the most relevant ones for a certain biological activity can also be used to speed up high-throughput virtual screening.

INTRODUCTION

Feature selection methods can help determine molecular descriptors that are important for the characterization of target-family specific classes of drugs and drug-like molecules by machine learning systems. Currently large numbers of descriptors are available for molecule characterization. Traditional feature selection methods such as forward and backward selection[1] or evolutionary algorithms[2] are computationally too expensive to be applied to very large descriptor sets directly. The most time-consuming step is retraining of the classifier after every modification of the set of selected features. This step needs to be reiterated sufficiently often before the process converges to the final set of features. Parallelizing computations is usually the only way to speed up the procedure.

An alternative approach is to select the important features prior to classifier training. In this case, the classifier needs to be trained only once for the selected features. Several techniques are known to implement this concept, e.g. correlation coefficients,[3,4] Fisher discriminant analysis,[1] and Kolmogorov-Smirnov (KS) statistics.[5] KS statistics was shown to be well-suited for feature selection in different fields of research.[6,7] Recently several model-dependent methods for feature selections were developed,[8] where the classifier is trained prior to feature selection, and features are selected based on a statistical model of the trained classifier. These methods have been predicted to outperform model-independent feature selection algorithms.[8]

For the present study we developed and applied a support vector machine (SVM)-based feature selection and compared it with a KS-based algorithm. An advantage of the SVM-based classification[9] in comparison to other methods, e.g. multilayered feed-forward neural networks,[1] is that the construction of the surface that separates classes of data depends only on the support vectors.[10] Support vectors are samples that are lying close to the border that separates two classes. Using only these samples can help increase the accuracy of the SVM prediction.[11] We extended the same principle to feature selection. Once an SVM classifier has been trained with all molecular descriptors, feature selection is based on the identified support vectors only, disregarding other samples.

The method was applied to feature selection from SVM classifiers for kinase inhibitors, factor Xa inhibitors, and thrombin inhibitors. The approach complements related work on "drug-likeness" prediction[12] and extends it to target- and target-family specific sets of inhibitors.

DATA AND METHODS

Data Sets. For SVM training and feature selection we used subsets of the COBRA database, version 2.1.[13] Three different splits of the COBRA collection were used for evaluation of the feature selection algorithms: 226 kinase inhibitors and 4479 noninhibitors; 227 factor Xa inhibitors and 4478 noninhibitors; and 227 factor Xa inhibitors and 195 thrombin inhibitors. The subset of kinase inhibitors represents a diverse set of molecules in that they are specific to a family of targets that differ significantly from each other. On the contrary, factor Xa and thrombin inhibitors are drug molecules which are specific for a single target. We expected that factor Xa and thrombin inhibitors should share a certain degree of similarity due to the similarity of the target binding sites.

Two sets of descriptors were calculated: 182 descriptors from MOE (Molecular Operating Environment)[14] and 225 topological pharmacophore (CATS) descriptors.[15] MOE descriptors include various 2D and 3D descriptors. 2D descriptors were physical properties, subdivided surface areas, atom and bonds counts, Kier-Hall connectivity and

* Corresponding author phone: +49-69 79829821; fax: +49-69 79829826; e-mail: gisbert.schneider@modlab.de.

10.1021/ci0342876 CCC: $27.50
Published on Web 03/03/2004

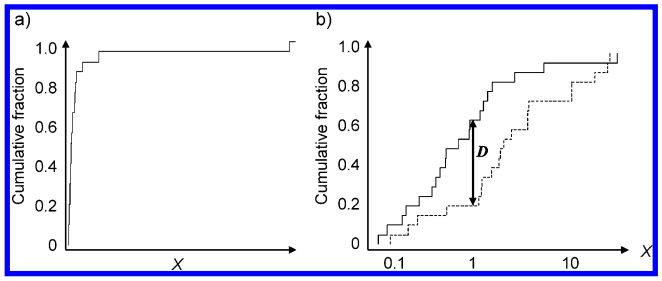

Figure 1. (a) Cumulative fraction plot. *X* denotes a molecular feature. (b) KS-test comparison. Cumulative fraction plots for two classes of data are shown by solid and dotted lines. *D* denotes the maximum difference of feature *X* values observed for the two classes.

Kappa Shape indices, adjacency and distance matrix descriptors, pharmacophore feature descriptors, and partial charges descriptors. 3D descriptors were potential energy descriptors; surface area, volume and shape descriptors; and conformational dependent charge descriptors. Before calculating MOE descriptors, single 3D conformers were generated by Corina.[16] CATS descriptors were calculated taking into consideration pairs of atom types separated by 0 up to 15 bonds. All descriptor columns were scaled to have zero mean and unit standard deviation.

Kolmogorov-Smirnov Statistics. KS-based statistics represent a model-independent method for feature selection. It is routinely used for feature selection from different data sets and features. Its main advantage over other methods is the independence from the particular statistical model that generates the data, in contrast to other methods, that perform well only if the data adopts certain statistics. For instance, "correlation coefficient"[3,4] based feature selection performs best if the data can be modeled by Gaussian mixtures,[1] and its accuracy drops otherwise. Very often it is impossible to correctly guess statistical models of the data a priori, which results in only approximately correct models. If the underlying statistics is not known or a Gaussian mixture model is not appropriate, KS statistics can be a method of choice.

In KS statistics each feature is first tested to have different statistics for class and nonclass samples. This is done by merging feature values for class and nonclass and building two separate cumulative fraction functions, one for class and one for nonclass. The cumulative fraction function represents the dependency of the percentage of samples whose feature values are below a certain threshold, on the position of the threshold value in the sorted list of feature values. An example of the cumulative function for the data set {0.08, 0.10, 0.15, 0.17, 0.24, 0.34, 0.38, 0.42, 0.49, 0.50, 0.70, 0.94, 0.95, 1.26, 1.37, 1.55, 1.75, 3.20, 6.98, 50.57} is given in Figure 1a. The maximum difference *D* of two cumulative functions for class and nonclass is then used as a measure for the significance of a distinguishing feature. An example of this measure is given in Figure 1b.

A KS statistics test is performed for all available features, which are then sorted with respect to the KS test results, and only the most relevant features are considered for further training.

SVM-Based Feature Selection. Usually feature selection algorithms are applied prior to the classifier training: A feature selection algorithm first selects a set of features and then a classifier is trained based on the features of this subset. Recently it was demonstrated that feature selection schemes,

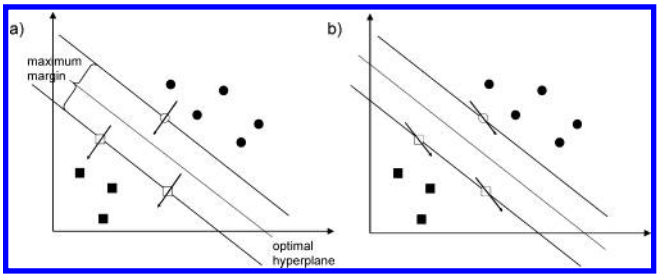

Figure 2. SVM-based feature selection. The optimal SVM hyperplane is shown with examples of class and nonclass samples (filled circles and squares). In the example support vectors are indicated by open symbols. For an estimation of the feature relevance the gradient (shown by arrows) of the feature change is calculated only for support vectors. (a) relevant features have a gradient perpendicular to the separating hyperplane; (b) irrelevant features.

where the feature selection algorithm relies on the model that is created during training, produce better results.[8] Accordingly an alternative scheme for feature selection was suggested: The classifier is first trained using all available features. Then, the least important features are deleted. The drawback of this approach is that the trained classifier usually assumes a certain statistical model for the data, which might be only approximately correct. Current algorithms for nonlinear classifier training like artificial neural networks or SVM estimate a statistical model for the data sufficiently well to make this approach an alternative to model-independent feature selection.

The separating surface generated by SVM is given by

$$f(\mathbf{x}) = \sum_i a_i \cdot K(\mathbf{x_i^{sv}}, \mathbf{x}) + b$$

Here a_i, b, and $\mathbf{x_i^{sv}}$ are parameters of the SVM, determined during training. $\mathbf{x_i^{sv}}$ are support vectors, which represent a subset of the training samples that determine the separating surface. This surface corresponds to the linear separation in a very high-dimensional space, where data points are mapped during SVM training.[17] This mapping is determined solely by the kernel function $K(x,x')$.[18] In this high-dimensional space the separating surface is given by

$$f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b$$

where

$$\mathbf{w} = \sum_{\mathbf{i}} a_i \mathbf{x_i^{sv}}$$

is a normal vector of the separating hyperplane. To estimate the importance R_f of a feature to the accuracy of the SVM prediction we calculated a projection of the feature change in the mapped space to the normal of the SVM plane (Figure 2):

$$R_f = \frac{(\mathbf{w} \cdot \Delta \mathbf{x}_f)}{\Delta \mathbf{x}_f} = \frac{(\mathbf{w} \cdot \mathbf{x}_f^{e}) - (\mathbf{w} \cdot \mathbf{x}_f^{b})}{\Delta \mathbf{x}_f} = \frac{\Delta f(\mathbf{x})}{\Delta \mathbf{x}_f} \rightarrow \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_f}$$

Calculating the derivative we obtain:

$$R_f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_f} = \sum_i a_i * \frac{\partial K(\mathbf{x_i^{sv}}, \mathbf{x})}{\partial \mathbf{x}_f} + b$$

For estimating the relevance of a feature to classification we should calculate R_f only in the vicinity of the separating hyperplane. To achieve it we will sum R_f only over support vectors, extending the principle of SVM that the position of the classifying hyperplane depends only on support vectors:

$$R_f = \sum_j R_j(\mathbf{x_j}^{sv}) = \sum_{i,j} a_i^* \frac{\partial K(\mathbf{x_i}^{sv}, \mathbf{x_j}^{sv})}{\partial \mathbf{x}_f} + b.$$

Empirically we observed that data normalization improved the performance in some cases; therefore, the final formula that we used to perform feature selection is

$$R_f = \sum_j R_j(\mathbf{x_j}^{sv}) = \sum_j \left(\sum_i a_i^* \frac{\partial K(\mathbf{x_i}^{sv}, \mathbf{x_j}^{sv})}{\partial \mathbf{x}_f} + b \right) / \left(\sum_{i,k} a_i^* \frac{\partial K(\mathbf{x_i}^{sv}, \mathbf{x_j}^{sv})}{\partial \mathbf{x_k}} + b \right)$$

Summarizing, R_f was calculated for all features, and those features with low R_f value were excluded from the features used for training. It is important to note that R_f depends only on the support vectors.

For constructing SVM models we used the SVM-light package.[19] A fifth-order polynomial kernel was used in SVM training: $K(\mathbf{x}',\mathbf{x}) = (s(\mathbf{x}'\cdot\mathbf{x}) + 1)^5$. Training parameters $\boldsymbol{s}$ and $\boldsymbol{C}$ were optimized using a gradient decent-like algorithm to achieve maximum accuracy of prediction for the validation set. Parameter $\boldsymbol{C}$ is an internal parameter that is set prior to SVM training. It defines the tradeoff between the separating margin and the penalty for incorrect predictions.[17]

Model Validation. Classification accuracy was evaluated based on prediction accuracy and the correlation coefficient according to Matthews[20]

$$\mathrm{cc} = \frac{NP - \mathrm{OU}}{\sqrt{(N+O)(N+U)(P+O)(P+U)}}$$

where P, N, O, and U are the numbers of true positive, true negative, false positive, and false negative predictions, respectively. Active molecules with specific activity were considered as the "positive set", and the other molecules formed the "negative set". The values for cc can range from −1 to 1. A perfect prediction gives a correlation coefficient of 1. Different training and test subset were selected, 80% of samples for the training and 20% for the test. Ten cross-validations were performed to estimate average and standard deviation of the accuracy. Prediction accuracy and average value of ⟨cc⟩ were calculated for the test subsets.

RESULTS AND DISCUSSION

We compared two methods for feature selection, KS-based and SVM-based. Both methods were able to effectively select sets of the most relevant features. Figure 3 shows the dependency of the classification accuracy and Matthews correlation coefficient on the number of selected features for each subset. In all three sample applications the SVM-based feature selection method outperformed the KS-based approach, i.e., the classification accuracy remained at a high level even for small numbers of remaining features. The prediction accuracy dropped when the number of features fell between 100 and 200 for the KS-based method. In contrast, using the SVM-based method for feature selection we were able to go down to about 40 features with only a slight reduction in classification accuracy. This indicates potential advantages of the SVM-based method. Considering the error margins in the thrombin vs factor Xa classification, KS-based feature selection may be regarded as comparable to the SVM approach. This might have a relatively simple explanation: A large portion of features might be relatively easily discarded as "irrelevant" for correct classification. In this case no significant advantage of an SVM-based versus a KS-based scheme is observed. Still, when the number of features was below 100 SVM-based feature selection performed better. We wish to stress that a general statement about the relative usefulness of the two methods is not possible based on this single study. Also, we cannot fully exclude that the difference seen in Figure 3 between SVM and KS might in part result from different levels of parameter optimization.

Table 1 contains a list of the features which were selected being the most relevant for subset classification. Table 2 contains average property values calculated for the sets of inhibitors used in this study.

Both factor Xa and thrombin inhibitors are relatively large molecules containing characteristic fragments that are specific for binding to the S1 pocket of the trypsin-like serine proteases.[21] Typically, these fragments are positively charged. Most of the known faxtor Xa inhibitors exploit the S4 pocket and S3 "cation recognition pocket" of factor Xa to gain binding affinity.[22] A difference between the two classes of the molecules might be noted by observing the most relevant features in more detail. The distance of a positive charge on the one side and lipophilic, hydrogen-bond donor and acceptor groups on the other side was suggested being a key property for a distinction between factor Xa and thrombin inhibitors by our SVM-based feature selection. This property is most easily observed by comparing CATS descriptors for large distances. As expected, these descriptors are found in the top listed of the ranked features (Table 1a). These features can be highlighted in the two-dimensional structures of selective factor Xa inhibitors (Figure 4). Compound **1**[23] and compound **2**[24,25] have an approximately 3300-fold selectivity for factor Xa over thrombin and contain the topological pharmacophores selected by SVM. Structure **2** is a representative member of several covalent, peptide-derived biscation factor Xa inhibitors which were used for SVM-training. It is not surprising, therefore, that the most "relevant" molecular features according to the SVM classifier are found in these molecular structures. Structure **1** was not part of the training data, but some of the high-ranking features are present in this molecule, too.

Our compilation of kinase inhibitors represents a compound collection containing much broader activities than the collection of factor Xa and thrombin inhibitors. Looking at their average molecular weight and lipophilicity (clogP) one can conclude that they are smaller and more lipophilic than factor Xa and thrombin inhibitors (Table 2). This might explain the observation that in the list of top-ranking SVM features the topological descriptors are less prominent, and various van der Waals based estimations of surface charges were selected as "relevant" (Table 1b).

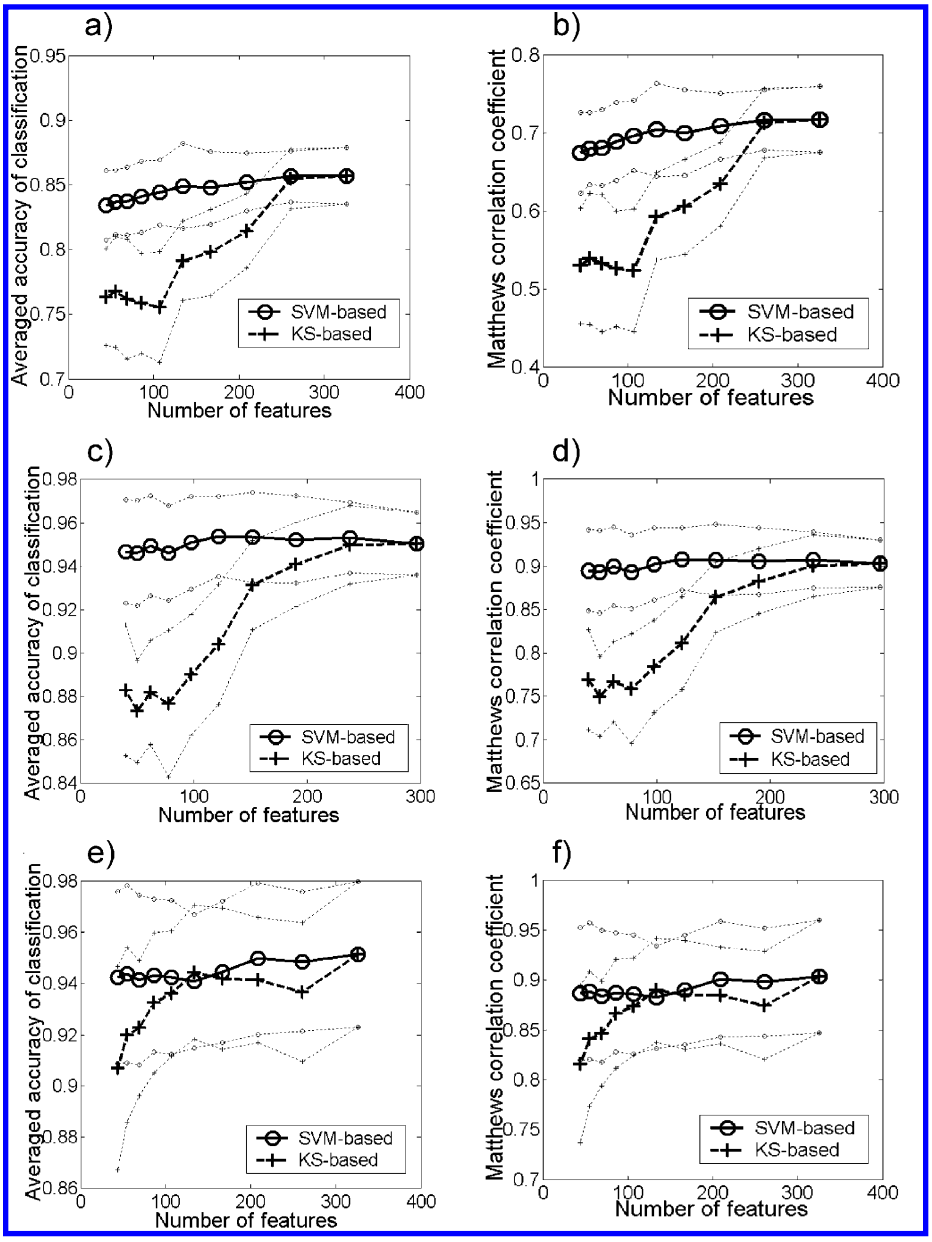

Figure 3. Results of feature selection by SVM- and KS-based algorithms. Matthews correlation coefficient and average classification accuracy are plotted as a function of the number of selected features. Standard deviations are shown as dotted lines. (a, b) Classification of kinase inhibitors versus the remainder of the COBRA data set. (c, d) Classification of factor Xa inhibitors versus the remainder of the COBRA data set. (e, f) Classification of factor Xa versus thrombin inhibitors.

Factor Xa inhibitors represent a relatively diverse set of molecules. Nonetheless, by examining their structures it is possible to assume that they have certain topological similarity. This could be a reason, why various topological descriptors are found within the first 20 most important descriptors (Table 1c). Surprisingly simple descriptors, like the number of aromatic atoms and aromatic bonds are also at the top of the list. Certainly, these simplistic descriptors cannot explain selectivity of factor Xa inhibitors, rather the whole list of "relevant" features must be taken into consideration if one tries to make sense out of a classifier system. This example demonstrates that feature selection does not necessarily deliver clear answers.

Although similar approaches were applied to perform SVM-based feature selection by Guyon and co-workers,[26] an advantage of our method is that feature selection was performed only based on the position of support vectors. It allows us to discard a large portion of data which is irrelevant for construction of the separating hyperplane. A potential additional advantage of our implementation is that classification of new molecules is quick and straightforward: computation time needed for a single molecule is approximately comparable to the time for reading its descriptors. Further information about computational efficiency of SVM can be found elsewhere.[18] Our results demonstrate that a central idea of SVM, namely the construction of a separating surface which is based only on support vectors, results in an efficient algorithm for feature selection when equipped with a feature selection scheme. We have successfully applied this algorithm to characterize groups of enzyme inhibitors. The algorithm was able to select crucial molecular features from a rather loosely defined compound class (kinase inhibitors) as well as features that might be relevant for inhibition of a particular target (factor Xa). It is important to mention that such feature selection methods do not explain why subsets can be classified or what the chemical explanation for an

Table 1. (a) Selected Features of Factor Xa Inhibitors versus Thrombin Inhibitors,[a] (b) Selected Features of Kinase Inhibitors,[b] and (c) Selected Features of Factor Xa Inhibitors[c]

feature	description
	(a)
SMR_VSA4	sum of υ_i such that R_i is in (0.39,0.44]
CATS_207	correlation for the distance of 13 bonds between positive and lipophilic atoms
CATS_171	correlation for the distance of 11 bonds between acceptor and acceptor atoms
CATS_153	correlation for the distance of 10 bonds between donor and positive atoms
CATS_120	correlation for the distance of 8 bonds between lipophilic and lipophilic atoms
a_nN	number of nitrogen atoms
CATS_91	correlation for the distance of 6 bonds between donor and donor atoms
CATS_63	correlation for the distance of 4 bonds between donor and positive atoms
CATS_57	correlation for the distance of 3 bonds between positive and lipophilic atoms
CATS_50	correlation for the distance of 3 bonds between acceptor and acceptor atoms
CATS_47	correlation for the distance of 3 bonds between donor and acceptor atoms
SMR_VSA5	sum of υ_i such that R_i is in (0.44,0.485]
PEOE_FPNEG	fractional negative polar van der Waals surface area. This is the sum of the υ_i such that q_i is less than -0.2 divided by the total surface area. The υ_i were calculated using a connection table approximation.
PEOE_VSA+3	sum of υ_i where q_i is in the range [0.15,0.20).
CATS_187	correlation for the distance of 12 bonds between acceptor and positive atoms
CATS_33	correlation for the distance of 2 bonds between donor and positive atoms
Dens	mass density: molecular weight divided by van der Waals volume.
PEOE_VSA_PNEG	total negative polar van der Waals surface area. This is the sum of the υ_i such that q_i is less than -0.2. The υ_i were calculated using a connection table approximation.
PEOE_VSA-1	sum of υ_i where q_i is in the range $[-0.10,-0.05)$.
	(b)
VDistEq	If m is the sum of the distance matrix entries, then VdistEq is defined to be the sum of $\log_2 m\text{-}pi \log_2 pi/m$ where p_i is the number of distance matrix entries equal to i.[28]
diameter	largest value in the distance matrix[28]
CATS_188	correlation for the distance of 12 bonds between acceptor and negative atoms
SMR_VSA4	sum of υ_i such that R_i is in (0.39,0.44].
VSA_other	approximation to the sum of VDW surface areas of atoms that are not a donor, acceptor, positive, negative, or hydrophobe
a_nCL	number of chlorine atoms
std_dim1	standard dimension 1: the square root of the largest eigenvalue of the covariance matrix of the atomic coordinates. A standard dimension is equivalent to the standard deviation along a principal component axis
FASA_H	fractional ASA_H calculated as ASA_H/ASA. Here, ASA_H is the water accessible surface area of all hydrophobic ($\lvert q_i \rvert < 0.2$) atoms and ASA is the water accessible surface area of all atoms.
Q_VSA_FPOS	fractional positive van der Waals surface area. This is the sum of the υ_i such that q_i is nonnegative divided by the total surface area. The υ_i were calculated using a connection table approximation.
Q_VSA_FHYD	fractional hydrophobic van der Waals surface area. This is the sum of the υ_i such that $\lvert q_i \rvert$ is less than or equal to 0.2 divided by the total surface area. The υ_i were calculated using a connection table approximation.
radius	If r_i is the largest matrix entry in row i of the distance matrix D, then the radius is defined as the smallest of the r_i[28]
CATS_192	correlation for the distance of 12 between positive and lipophilic atoms
b_ar	number of aromatic bonds
a_aro	number of aromatic atoms
CATS_147	correlation for the distance of 9 bonds between donor and lipophilic atoms
a_nF	number of fluorine atoms
petitjian	value of (diameter-radius)/diameter.[28] Here *diameter* is the largest value in the distance matrix; radius is defined as follows, if r_i is the largest matrix entry in row i of the distance matrix D, then the radius is defined as the smallest of the r_i
petitjianSC	Petitjean graph shape coefficient as defined in ref 28
CATS_200	correlation for the distance of 13 bonds between donor and lipophilic atoms
CATS_186	correlation for the distance of 12 bonds between acceptor and acceptor atoms
	(c)
PEOE_VSA+1	sum of υ_i where q_i is in the range [0.05,0.10).
balabanJ	Balaban's connectivity topological index[29]
b_ar	number of aromatic bonds
a_aro	number of aromatic atoms
SLogP_VSA1	sum of υ_i such that L_i is in $(-0.4,-0.2]$
wienerPol	Wiener polarity number: half the sum of all the distance matrix entries with a value of 3 as defined in ref 30
vsa_acid	approximation to the sum of VDW surface areas of acidic atoms
a_acc	number of hydrogen bond acceptor atoms (not counting acidic atoms but counting atoms that are both hydrogen bond donors and acceptors such as −OH).

Table 1 (Continued)

feature	description
reactive	indicator of the presence of reactive groups. A nonzero value indicates that the molecule contains a reactive group. The table of reactive groups was based on the Oprea set[31] and includes metals, phospho-, N/O/S−N/O/S single bonds, thiols, acyl halides, Michael acceptors, azides, esters, etc
b_rotR	fraction of rotatable bonds: b_rotN divided by b_count. Here b_count is the number of bonds including implicit hydrogens.
vsa_pol	approximation to the sum of VDW surface areas of atoms that are both hydrogen bond donors and acceptors, such as −OH
vsa_base	approximation to the sum of VDW surface areas of basic atoms.
Q_VSA_FPOS	fractional positive van der Waals surface area. This is the sum of the υ_i such that q_i is nonnegative divided by the total surface area. The υ_i were calculated using a connection table approximation.
Q_VSA_FHYD	fractional hydrophobic van der Waals surface area. This is the sum of the υ_i such that $\|q_i\|$ is less than or equal to 0.2 divided by the total surface area. The υ_i were calculated using a connection table approximation.
b_heavy	number of bonds between heavy atoms
CATS_91	correlation for the distance of 6 bonds between donor and donor atoms
SLogP_VSA2	sum of υ_i such that L_i is in (−0.2,0].
Pmi	principal moment of inertia.
Zagreb	Zagreb index[32]
Chi1_qC	carbon connectivity index (order 1)[32]

[a] Calculation of the subdivided surface areas descriptors, like SMR_VSA, PEOE_VSA, was based on an approximate accessible van der Waals surface area calculation for each atom, υ_i, along with some other atomic property p_i. The υ_i were calculated using a connection table approximation. Each descriptor in a series was defined to be the sum of the υ_i over all atoms i such that p_i is in a specified range (a,b]. For SMR_VSA p_i is R_i, which denotes the contribution to Molar Refractivity for atom i. For PEOE_VSA p_i is L_i, which denotes the contribution to logP(o/w) for atom i.[27] [b] Here, υ_i is the van der Waals surface area of atom i (as calculated by a connection table approximation). R_i denotes the contribution to the molar refractivity of atom i.[27] [c] For definition of υ_i, R_i, and L, see Table 1a.

Table 2. Average Property Values of the Three Sets of Inhibitors[a]

target	MW	PSA[b]/Å^2	clogP
factor Xa	490	132	2.9
thrombin	503	140	2.6
kinase	405	89	3.2

[a] Properties were calculated using MOE.[14−32] [b] PSA: polar surface area.

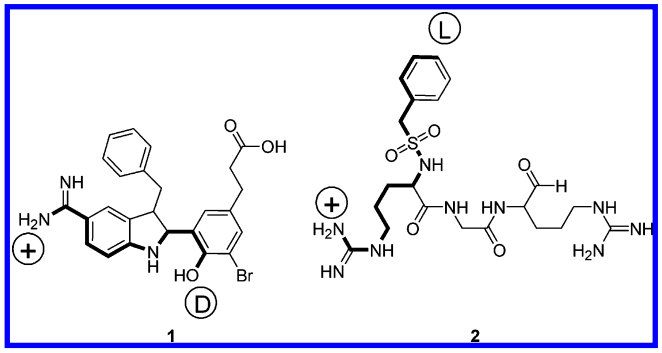

Figure 4. Examples of molecular features selected by SVM. Compounds 1 and 2 are selective factor Xa inhibitors. Two features are highlighted which were identified by an SVM classifier for discrimination between factor Xa and thrombin inhibitors. In structure **1** a positive charge (+) is separated by 10 bonds from a hydrogen-bond donor (D) site; in structure **2** a positive charge is separated by 13 bonds from a lipophilic point (L). These two-point pharmacophore features might be relevant for binding to the factor Xa active site pocket.

observed biological activity is. They might be suited for reducing the number of variables used in QSAR studies. It should be stressed that different feature selection algorithms tend to select different sets of "relevant" features. Therefore, the ranked list of features produced by the SVM-based method need not necessarily be more meaningful than a selection obtained by other methods, as one might conclude from the observation that the selected features resulted in a sustained high level of classification accuracy. It is possible that certain feature sets represent approximately the same chemical information, and as long as we only roughly describe a chemical agent using molecular descriptors, there will exist several almost equally suited partial solutions to the same classification task.

ACKNOWLEDGMENT

Uli Fechner is thanked for contributing his software for calculation of the CATS descriptors. This work was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt.

REFERENCES AND NOTES

(1) Richard O. Duda; Hart, P. E.; Stork, D. G. *Pattern Classification*; Wiley-Interscience, 2000.
(2) Mitchell, M. *An Introduction to Genetic Algorithms* (*Complex Adaptive Systems*); MIT Press: 1998.
(3) Golub, T. R.; Slonim, D. K.; Tamayo, P.; Huard, C.; Gaasenbeek, M.; Mesirov, J. P.; Coller, H.; Loh, M. L.; Downing, J. R.; Caligiuri, M. A.; Bloomfield, C. D.; Lander, E. S. *Science* **1999**, *286*, 531−537.
(4) Furey, T. S.; Cristianini, N.; Duffy, N.; Bednarski, D. W.; Schummer, M.; Haussler, D. *Bioinformatics* **2000**, *16*, 906−914.
(5) Rassokhin, D. N.; Agrafiotis, D. K. *J. Mol. Graphics Modell.* **2000**, *18*, 370−384.
(6) Harter, H. L.; Khamis, H. J.; Lamb, R. E. *Commun. Statistics, Simulat. Comput.* **1984**, *13*, 293−323.
(7) Khamis, H. J. *J. Statistical Planning Inference* **1990**, *24*, 317−335.
(8) Kohavi, R.; John, G. H. *Artif. Intelligence* **1997**, *97*, 273−324.
(9) Byvatov, E.; Schneider, G. *Appl. Bioinf.* **2003**, *2*, 67−77.
(10) Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: 1995.
(11) Cortes, C.; Vapnik, V. *Machine Learning* **1995**, *20*, 273−297.
(12) Byvatov, E.; Fechner, U.; Sadowski, J.; Schneider, G. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1882−1889.
(13) Schneider, P.; Schneider, G. *QSAR Comb. Sci.* **2003**, *22*, 713−718.
(14) MOE; Chemical Computing Group Inc.: Montreal, 2003.
(15) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2894−2896.

(16) Gasteiger, J.; Rudolph, C.; Sadowski, J. *Tetrahedron Comput. Method* **1990**, *3*, 537−547.
(17) Burges, C. J. C. *Data Min. Knowledge Discov.* **1998**, *2*, 121−167.
(18) Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*; Cambridge University Press: 2000.
(19) Joachims, T. In *Advances in Kernel Methods − Support Vector Learning*; Schölkopf, B., Burges, C., Smola, A., Eds.; MIT-Press: Cambridge, MA, 1999; pp 41−56.
(20) Matthews, B. W. *Biochim. Biophys. Acta* **1975**, *405*, 442−451.
(21) Banner, D. In *Protein−Ligand Interactions. From Molecular Recognition to Drug Design*; Böhm, H. J., Schneider, G., Eds.; Wiley-VCH: Weinheim, 2003; pp 163−185.
(22) Rai, R.; Sprengeler, P. A.; Elrod, K. C.; Young, W. B. *Curr. Med. Chem.* **2001**, *8*, 101−119.
(23) Rai, R.; Kolesnikov, A.; Li, Y.; Young, W. B.; Leahy, E.; Sprengeler, P. A.; Verner, E.; Shrader, W. D.; Burgess-Henry, J.; Sangalang, J. C.; Allen, D.; Chen, X.; Katz, B. A.; Luong, C.; Elrod, K.; Cregar, L. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1797−1800.
(24) Marlowe, C. K.; Sinha, U.; Gunn, A. C.; Scarborough, R. M. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 13−16.
(25) Tamura, S. Y.; Levy, O. E.; Uong, T. H.; Reiner, J. E.; Goldman, E. A.; Ho, J. Z.; Cohen, C. R.; Bergum, P. W.; Nutt, R. F.; Brunck, T. K.; Semple, J. E. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 745−749.
(26) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. *Machine Learning* **2002**, *46*, 389−422.
(27) Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1965**, *5*, 868−873.
(28) Petitjean, M. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 331−337.
(29) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(30) Balaban, A. T. *Theor. Chim. Acta* **1979**, *53*, 355−375.
(31) Oprea, T. I. *J. Comput.-Aided. Mol. Des.* **2000**, *14*, 251−264.
(32) Hall, L. H.; Kier, L. B. *Rev. Comput Chem.* **1991**, *2*, 367−422.

CI0342876