Development of a Structural Model for NF-κB Inhibition of Sesquiterpene Lactones Using Self-Organizing Neural Networks

Steffen Wagner,[†] Angelika Hofmann,[§] Bettina Siedle,[†] Lothar Terfloth,[§] Irmgard Merfort,*[,†] and Johann Gasteiger[§]

Institut für Pharmazeutische Wissenschaften, Lehrstuhl für Pharmazeutische Biologie und Biotechnologie, Universität Freiburg, 79104 Freiburg, Germany, and Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, 91052 Erlangen, Germany

Received November 8, 2005

A variety of sesquiterpene lactones (SLs) possess considerable anti-inflammatory activity. Several studies have shown that they exert this effect in part by inhibiting the activation of the transcription factor NF-κB. In the present study we elaborated on the investigation of a data set of 103 structurally diverse SLs for which we had previously developed several different QSAR equations dependent on the skeletal type. Use of 3D structure descriptors resulted in a single model for the entire data set. In particular, local radial distribution functions (L-RDF) were used that centered on the methylene−carbonyl substructure believed to be the site of attack of cysteine-38 of the p65/NF-κB subunit. The model was developed by using a counterpropagation neural network (CPGNN), attesting to the power of this method for establishing structure−activity-relationships. The investigations shed more light onto the influence of the chemical structure on NF-κB inhibitory activity.

Introduction

Sesquiterpene lactones (SLs) are a large group of secondary plant metabolites from which more than 4000 structures are known mostly from the Asteraceae family.[1] Numerous species of this family are used in traditional medicine for the treatment of inflammation, and SLs are described as their active constituents. The anti-inflammatory activity of SLs has been corroborated using various assays, and several studies have established that they exert their activity by inhibiting the transcription factor NF-κB.[2−5] Using helenalin and parthenolide as models, we could provide evidence that DNA binding of NF-κB is prevented by alkylation of cysteine-38 in the p65/NF-κB subunit.[2,6] There are strong indications that this is a general mechanism for SLs, which possess α,β-unsaturated carbonyl structures such as α-methylene-γ-lactones or α,β-unsaturated cyclopentenones. These functional groups are known to react with nucleophiles, especially with the sulfhydryl group of cysteine, in a Michael-type addition.[1]

NF-κB is a central mediator of the human immune system promoting the expression of over 400 target genes in response to inflammatory stimulators such as cytokines as well as genes encoding cell adhesion molecules and enzymes such as COX-II and iNOS.[7,8]

Because of its central role in regulating inflammatory responses, a pharmacological inhibition of NF-κB activation in vivo could be beneficial in the treatment of inflammation.[9−12] Therefore, searching for lead structures that can be optimized for the development of a pharmaceutically used anti-inflammatory cytokine suppressing remedy is an interesting task.

Previously, we presented a QSAR study on the NF-κB DNA binding activity of a great variety of structurally different SLs.[13] Multiple linear regression analysis revealed that a strong NF-κB inhibitory activity correlates with the number of alkylating centers, such as the methylene lactone and conjugated carbonyl functions, but that topological and structure-coding properties also contribute to the NF-κB inhibitory activity of SLs possessing a rigid skeleton. Good correlations were only obtained for individual subgroups of SLs and not for the entire data set.

Here, we report on the development of a counterpropagation neural network (CPGNN) model to predict the inhibitory potency of unknown SLs using the same data set of 103 SLs as in the previous QSAR study.[13] These SLs represent six different skeletal types with a wide structural diversity: 44 germacranolides, 16 furanoheliangolides, 22 guaianolides, 9 pseudoguaianolides, 2 hypocretenolides, and 10 eudesmanolides.

Artificial neural networks (ANN) are defined as computational models having structures derived from the simplified concept of the brain in which a number of nodes, called neurons, are interconnected in a network-like structure.[14] Kohonen introduced an ANN, which he called self-organizing network[15] and which projects objects from a multidimensional space into a space of lower-dimensionality, usually into a 2D plane. In this projection, the similarity relationship between objects is conserved. Thus, in principle, Kohonen networks can be used for clustering of objects. Whereas the training of these networks is unsupervised—i.e., the investigated property is not used during the training process—the Kohonen learning algorithm can also be utilized for supervised learning. This results in counterpropagation neuronal networks and can be used to model problems and to predict properties of new objects.

Here, we generated CPGNN to improve the prediction of the inhibitory potency of unknown SLs compared to the previous QSAR study. CPGNN is a suitable option because of the complex data set, the possibility of using high-dimensional data sets, and the existence of noncontinuous values. We have succeeded in modeling the entire data set as a whole in a single model. This has been achieved by using values of 3D structure representations as descriptors. This global model provides more detailed information on the structural influence of the biological activity than the previous QSAR studies. This structural model can contribute to the search and optimization of lead structures for the development of therapeutically used cytokine suppressing

* To whom correspondence should be addressed. Phone: +49-761-2038373. Fax: +49-761-2038383. E-mail: irmgard.merfort@pharmazie.uni-freiburg.de.
[†] Universität Freiburg.
[§] Universität Erlangen-Nürnberg.

10.1021/jm051125n CCC: $33.50
Published on Web 02/24/2006

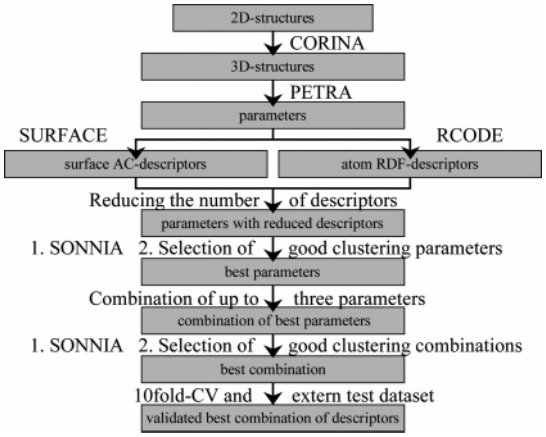

Figure 1. Strategy for creating the best counterpropagation neural network (CPGNN).

remedies valuable for the treatment of various inflammatory diseases.

Results and Discussion

Generation of a Counterpropagation Neural Network. To generate a CPGNN for the prediction of the NF-κB inhibitory potency of unknown SLs, the data set of 103 SLs and the resulting micromolar concentrations (5, 10, 12.5, 20, 25, 50, 100, 200, or 300 μM) that completely inhibit NF-κB DNA binding and that are described as IC_{100} values were used. We embarked on the following strategy that is summarized in Figure 1. The vectors were reduced from 128 to between 7 and 12 descriptors of the radial distribution function (RDF) or autocorrelation coefficient (AC) (Table 1) as described in the Experimental Section. These reduced vectors were used for input as descriptors of a molecule. Abbreviations of the atomic properties indicate which property had been encoded into the RDF vector. In a similar manner, the abbreviations HBP (hydrogen-bonding potential) and MEP (molecular electrostatic potential) indicate the surface properties that were encoded into the AC vector (Tables 1−3). Eight of the 15 RDF-coded atom properties show a clustering with less than 13 conflicts and an occupancy of more than 75%. From the surface properties, only the HBP surface exhibits any clustering. Different activities could be distinguished by a few entries of the local radial distribution function (L-RDF) vectors using π-electronegativity, χ_π, values.

To study different properties of the SLs that could contribute to the inhibitory activity, up to four different properties were combined. This was done in such a way that two of the best eight RDF-coded atom properties (Table 1) were combined with each other, resulting in 28 combinations from which three exhibited a good clustering (conflicts less than 9, occupancy more than 79%). These three combinations were effective polarizability (α_d) and π-electronegativity (χ_π), π-electronegativity (χ_π) and σ-electronegativity (χ_σ), and π-electronegativity (χ_π) and the number of neighbor atoms ($n_{neighbor}$). Subsequently, the best eight RDF-coded atom properties and all 28 combinations were combined with the AC-encoded surface of the hydrogen-bonding potential. Only three show good clustering (conflicts less than 9, occupancy more than 79%): $n_{el} + \chi_\pi +$ HBP, $\chi_\pi +$ HBP, $\chi_\pi + n_{neighbor} +$ HBP. Finally, the L-RDF encoded χ_π properties were combined with the six good clustered combinations mentioned above, resulting in one combination with good clustering: $\chi_\pi +$ HBP $+ \chi_\pi$(L-RDF).

The resulting seven best combinations were characterized by clustering, conflicts, and occupancy. The model $\chi_\pi +$ HBP $+ \chi_\pi$(L-RDF) is the only one that belongs to the three best ones in all quality categories (Table 2). The exact composition of this model is given in Table 3. An example of a CPGNN-output map is given in Figure 2. The SLs with the highest inhibitory activity (activity classes 1 and 2) build an almost quadratic cluster that is well separated into all other activities. A similar behavior could be observed for the middle active substances (activity classes 3 and 4), whereas low active SLs (activity classes 5 and 6) build a small tapelike cluster. Interestingly, empty neurons predominate between the cluster of high activity and low activity SLs, indicating a good separation of these activity classes despite their neighborhood in the output map.

Evaluation of the Best CPGNN. The internal validation by 10-fold cross-validation (CV) revealed that a correct prediction of the NF-κB inhibitory activity was possible to 80.6% (Table 4). However, overfitting is a main risk by any kind of model building, especially if a high number of descriptors is associated with a low number of molecules. Whereas the ratio is sufficient in our best model (19 descriptors to 103 molecules), it has to be considered that the 19 descriptors were chosen from a large number of possible descriptors. Therefore, validation with an external data set that was not used during selection and training was carried out. External validation with 14 new SLs resulted in a correct prediction of 78.6%. This means that the NF-κB inhibitory activity for 11 of the 14 SLs was correctly predicted (see Table 5).

Performing a 10-fold CV with all 117 SLs (103 and the 14 new SLs) gave a correct prediction of 77.8%. Altogether, similar values of correct prediction were obtained using these three data sets, indicating that the best created model is not an artifact and covers a wide area of SLs structures.

Considering the quality criteria "clustering" and "conflicts", our best model shows outliers. Both types of outliers that will be discussed in the following can be depicted from Figure 3.

Outliers of clustering represent SLs that form islands of another activity class in a cluster. Because a resulting Kohonen map depends on the initializing of the network, four different random-initialized CPGNN models were used.

Interestingly, no SL is isolated in all networks. Only one SL (**83**) is isolated in 3 of 4 CPGNNs; all the other ones are isolated only in 2 (six SLs) or 1 (seven SLs) of 4 CPGNNs (Table 5). Thus, the best model enables good clustering without excluding any SL. In addition, isolation results only in a wrong prediction for six SLs, whereas four of them are also involved in conflict neurons. Furthermore, outliers occur looking at the quality criteria "conflicts". As mentioned above, a conflict occurs if SLs with different activity classes are within the same neuron. One (or more) of these SLs can be regarded as outliers: one is an outlier if it has an activity class differing from those in the neighborhood (Figure 3). Considering this definition, seven SLs can be regarded as outliers (Table 5). By exclusion of these outliers,[16,17] validation of the remaining data set of 96 SLs showed an increased correct prediction of 88.5% by a 10-fold CV. These outliers may be SLs that have structural features that are not yet considered or that differ in their modes of action compared with the other ones.

Nevertheless, the original data set of 103 SLs was retained unchanged to cover a wide range of structures including structures that are not yet correctly predictable with this model. Unknown substances whose activity is predicted by being mapped into neurons containing outliers may be regarded as not predictable. Thus, the risk to predict incorrect inhibitory activities can be avoided.

Because the best model will be used in the future to search and develop lead structures with a strong inhibitory activity on

Table 1. List of the Atom Properties and Surfaces Used[a]

symbol	abbreviation used by PETRA	description	dimensionality after reduction
q_{formal}	FORMCH	formal charge of atoms	10[b]
$\boldsymbol{n_{el}}$	**NEL**	number of electrons	10[b]
$\boldsymbol{a_{peripher}}$	**PERIPH**	atom is peripheral or not	12[b]
$\boldsymbol{\alpha_d}$	**POLARIZ**	effective atom polarizability	10[b]
$n_{ringsize}$	RINGSIZ	size of the smallest ring the atom belongs to	9[b]
χ_{LP}	ENLP	lone pair electronegativity	10[b]
$\boldsymbol{\chi_\pi}$	**ENPI**	π-electronegativity	8[b]
$\boldsymbol{\chi_\sigma}$	**ENSIG**	δ-electronegativity	9[b]
$n_{free-el}$	FREEEL	number of free electrons	10[b]
a_{LPstab}	LPSTAB	mesomeric stabilization by lone pairs	10[b]
$\boldsymbol{n_{neighbor}}$	**NEIGHBOR**	number of neighbor atoms	7[b]
$\boldsymbol{n_{non-H}}$**-neighbor**	**NONHNEIG**	number of non-hydrogen neighbor atoms	7[b]
q_π	QPI	π-charge	11[b]
q_σ	QSIG	δ-charge	10[b]
q_{tot}	QTOT	total charge	10[b]
MEP	MEP	surface of the molecular electrostatic potential	9[c]
HBP	**HBP**	surface of the hydrogen binding potential	9[c]

[a] The atom properties were calculated by PETRA. The surfaces were generated using SURFACE. The bold-faced properties show a clustering with a low number of conflicts and a sufficient occupancy. [b] RDF encoded. [c] AC-encoded.

Table 2. Data That Characterize the Best Combinations Using the Kohonen Neural Network[a]

	clustering				
combination	correct	drawn	wrong	occupancy	conflicts
$\alpha_d + \chi_\pi$	72.75	11.50	**18.75**	74.25	**4.75**
$\chi_\pi + \chi_\sigma$	71.25	11.75	20.00	73.00	7.25
$\chi_\pi + n_{neighbor}$	68.75	12.75	21.50	**76.75**	5.50
$n_{el} + \chi_\pi$ + HBP	**73.75**	**9.25**	20.00	73.25	6.50
χ_π + HBP	**73.50**	**9.75**	**19.75**	**75.50**	5.25
$\chi_\pi + n_{neighbor}$ + HBP	72.25	10.75	20.00	73.75	**4.50**
χ_π + HBP + χ_π(L-RDF)	**77.25**	**9.00**	**16.75**	**74.50**	**5.00**

[a] The values are the average of four trainings. The three best values of each category are in bold.

Table 3. Descriptors of the Best Model Using CPGNN

property	encoded by	number of descriptors	distance (Å)
χ_π	RDF	8	1.4, 1.6, 2.6, 3.5, 4.6, 5.1, 5.9, 6.6
HBP	AC	9	4.4, 6.9, 7.1, 7.3, 7.7, 8.2, 8.6, 8.9, 10.3
χ_π	L-RDF	2	1.5, 4.5

NF-κB DNA binding, a good prediction is especially interesting for the activity classes 1 and 2 (Table 4). Our model allows the correct prediction of 31 from 37 SLs (84%) belonging to the activity class 1 or 2 containing the higher active components. Only 4 out of the 66 SLs (6%) from the lower activity classes (classes 3−6) are wrongly assigned to activity class 1 or 2. From these results it can be concluded that our model is a valuable tool for the screening of lead structures possessing potent NF-κB inhibitory activity.

Previously, a similar number of wrongly predicted SLs was obtained when carrying out classical QSAR studies with the same data set.[13] However, the data set had to be divided into five subgroups according to the carbon skeletons. Here, only one CPGNN describes the whole data set uniformly. This attests to the global validity of the model developed here, having discovered descriptors that can describe the structures of different carbon skeletons. In contrast to the former QSAR study, descriptors based on the RDF-encoded atomic properties and the AC-encoded surfaces have been used here. These descriptors are based on a 3D structure representation giving much more detailed information on the structural influence on biological activity, in our case of the NF-κB DNA binding inhibitory activity. In addition, this study also emphasizes the advantage of using counterpropagation neural network compared to classical QSAR studies. Thus, CPGNN is a very suitable tool for the prediction of the NF-κB DNA inhibitory activity and for the search for lead structures.

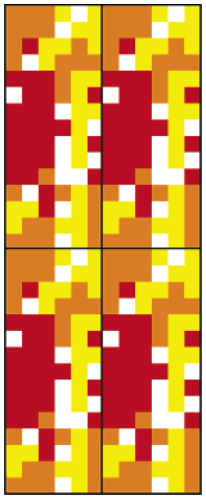

Figure 2. Output map of a toroidal CPGNN with a one-dimensional output layer of the best model χ_π + HBP + χ_π(L-RDF). High inhibitory activity for activity classes 1 and 2 is in red, that for activity classes 3 and 4 is in orange, and that for activity classes 5 and 6 is in yellow. The white areas indicate empty neurons. To illustrate the clustering of the different classes, four toroidal maps were arranged like tiles to indicate the closed nature of a toroidal surface.

Table 4. Confusion Matrix Based on a 10-Fold Cross-Validation of the Best Model χ_π + HBP + χ_π(L-RDF) Using CPGNN[a]

predicted activity class	exptl activity class						n_{false} of the predicted activity class
	1	2	3	4	5	6	
1	**5**	**8**	1	0	1	0	2
2	**7**	**9**	**7**	2	0	0	2
3	1	**2**	**4**	**4**	1	3	5
4	0	2	**5**	**12**	**1**	4	6
5	2	0	0	**1**	**3**	**4**	2
6	0	1	0	2	**6**	**5**	3
n_{false} of the exptl activity class	3	3	1	4	2	7	20

[a] n_{false} is the number of wrongly predicted SLs.

Structural Information of the Used Descriptors for the NF-κB Inhibitory Activity. The best model is obtained by a combination of χ_π + HBP + χ_π(L-RDF). The exact composition is given in Table 3. These descriptors can provide us with information about which structural features influence the

Table 5. Wrongly Predicted SLs and Outliers Regarding Quality Criterias "Conflicts" and "Clustering"[a]

no. of SLs	exptl activity class	no. of wrongly predicted SLs by validation: predicted activity class	no. of wrongly predicted SLs by validation: predicted by SL	conflicts: no. of SLs involved in conflict neurons	clustering: rate of wrong clustering using four CPGNNs (%)
5#	2	6	82	82	50
82*	6	–	–	**5**	–
92	4	6	**82**	–	–
19*	4	2	**20**	**20**	–
20	2	4	**19**	**19**	–
23	4	6	**79**	**79**	–
79	6	4	**23**	**23**	–
25*,#	3	–	–	**85/100**	25
85	6	3	**25**	**25**	–
36	4	2	**38**	**38**	–
38*	2	4	**36**	**36/100**	–
100	6	4	**36**	**25/38/102**	–
102*	4	–	–	**100**	–
33#	4	–	–	–	25
34#	3	–	–	–	25
37	4	–	–	**65**	–
65*,#	1	5	**103**	**37**	25
103#	5	1	**65**	–	50
39#	3	–	–	–	25
43	6	4	**37**	–	–
45#	4	6	**46**	**46**	25
46#	6	4	**45**	**45**	25
47#	6	3	**88**	–	50
48	1	–	–	**49/58**	–
49*	3	1	**58**	**48/58**	–
58	1	3	**49**	**49**	–
53	1	5	**91**	–	–
71#	1	–	–	–	50
72#	2	–	–	–	50
83#	2	–	–	–	75
87#	3	–	–	–	50
93	5	3	**76**	–	–
N2	1	3	**88**		
N4	3	6	**47**		
N8	2	6	**43**		

[a] SLs that are within the same neuron and/or predict each other are grouped together. SLs marked with "∗" are conflict outliers. Those marked with "#" are outliers of clustering. In column 3, the wrongly predicted activity class by 10-fold CV is shown. In column 4, the SLs that are in the same neuron are specified. Clustering outliers with the rate of occurrence regarding four CPGNNs are in column 5. A dash in a column indicates that the corresponding SL is not wrongly predicted and/or not an outlier.

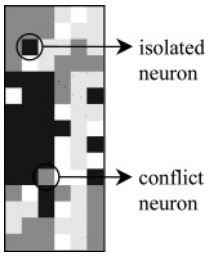

Figure 3. Output map of the best model demonstrating the two types of outliers. High inhibitory activity (activity classes 1 and 2) is in dark-gray, that for activity classes 3 and 4 is in gray, and that for activity classes 5 and 6 is in light-gray. The white areas indicate empty neurons. The isolated neuron is surrounded by other activities. The SL attached to this neuron is defined as an outlier of clustering. Two SLs are found within the conflict neuron: **19** (activity class 4) and **20** (activity class 2). Because the surrounded neurons are predominated by activity classes 1 and 2 (dark-gray), SL **19** is regarded as an outlier.

NF-κB inhibitory activity. Interestingly, our investigation strengthens the importance of the π-electronegativity; the atom property χ_π is included in the best model encoded by both RDF and L-RDF (Table 3). Only atoms with π-electrons, such as

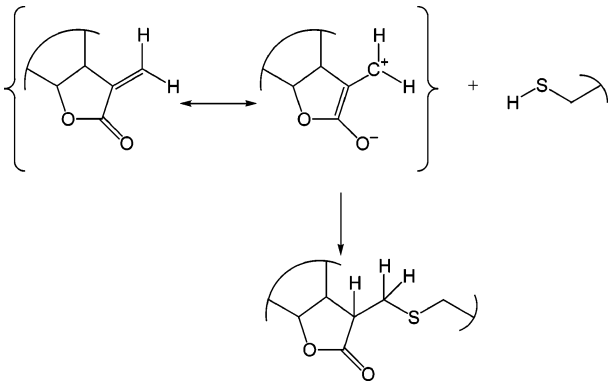

Figure 4. Reaction mechanism for the addition of the sulfhydryl group of cysteine to the α-methylene-γ-lactone groups of an SL.

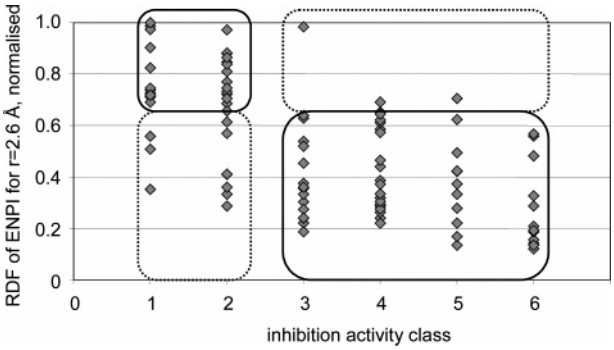

Figure 5. Distribution of the normalized RDF-encoded property ENPI (χ_π) at an atom distance of 2.6 Å within the inhibitory activity classes 1−6. Each rhomboid symbolizes one molecule. On the basis of an RDF value of 0.65, an almost complete differentiation between SLs of activity classes 1 + 2 and 3−6 is possible. Right separated SLs are framed by continuous lines, and false separated are framed by dashed lines. Only 9 of 37 SLs (24%) of activity classes 1 + 2 and only 3 of 66 SLs (5%) of activity classes 3−6 would be false-classified using only this single descriptor.

atoms in multiple bonds and atoms with free electron pairs that are in conjugation with multiple bonds, obtain χ_π-values different from zero. These atoms representing α,β-unsaturated carbonyl structures such as the α-methylene-γ-lactone moiety are mainly involved in the proposed reaction with cysteine-38 of the p65 subunit (Figure 4). Therefore, the exocyclic carbon atom was defined as the central atom by encoding with L-RDF. But not all distances of the RDF or L-RDF curves are important; therefore, only few were selected (Table 3) by a statistical *t*-test. Considering all selected distances, those with high values correlate with a high inhibitory activity.

The short distance of 1.5 Å of the L-RDF encoded χ_π exactly correlates with the atom distances between the exocyclic carbon atom and the neighboring carbon atom of the lactone ring. This emphasizes the importance of the α-methylene-γ-lactone structure for the reaction with cysteine. Including similar distances (1.4 and 1.6 Å) of the RDF-encoded χ_π revealed that other α,β-unsaturated carbonyl structures are also necessary for high reactivity. Analogously, distances of 2.6 Å (RDF-encoded) can be discussed. As mentioned above, high values of these descriptors correlate with a high inhibitory activity, as shown in Figure 5. Obviously, when only this single descriptor is used, an almost complete differentiation between SLs of activity classes 1 + 2 and 3−6 is possible. The atom distance of $r =$ 2.6 Å is of particular significance in α,β-unsaturated carbonyl structures (Figure 6a). Thus, the value of this descriptor correlates with the number of such structural elements. The more such α,β-unsaturated carbonyl units are present in a molecule,

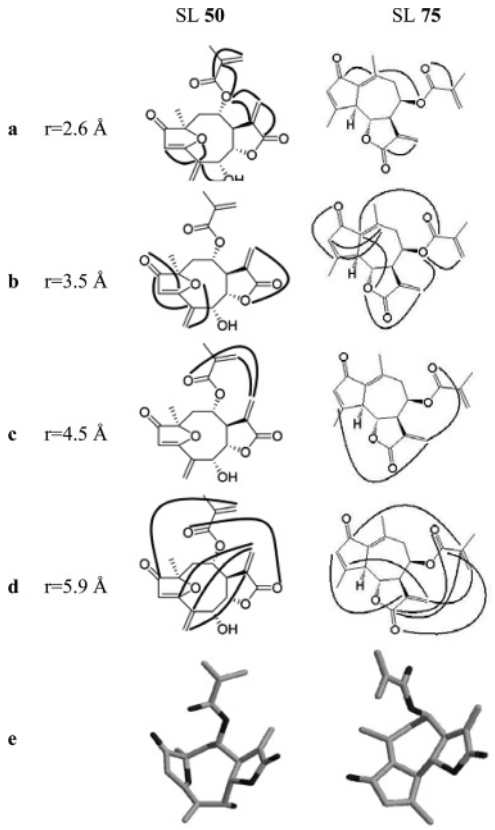

Figure 6. SL **50** (left) and SL **75** (right) with atom distances showing RDF or L-RDF values different from zero. Row a shows atom distances of $r = 2.6$ Å and row b of $r = 3.5$ Å of the RDF-coded property χ_π. Row c shows atom distances of $r = 4.5$ Å of L-RDF coded property χ_π with the exocyclic C atom of the lactone ring as the central atom. Row d shows the atom distances of $r = 5.9$ Å of the RDF-coded property χ_π. For better illustration, the 3D structures of the SLs **50** and **75** are shown in row e.

the higher is the activity. The significance of such elements was already shown.[13,18]

The distance of 3.5 Å is found within the α-methylene-γ-lactone as well as within α,β-unsaturated carbonyl structures that are part of a ring system (Figure 6b). Consequently, the importance of such structure elements that occur within the ring system and not within acyl moieties, such as methacryloyl moieties, is emphasized.[19] The intermediate distances of 4.6 Å (RDF-encoded) and 4.5 Å (L-RDF encoded) contain information on the neighborhood of α,β-unsaturated carbonyl structures. Thus, the L-RDF encoded χ_π at a distance of 4.5 Å gives information on the occurrence of hydroxy, carboxy, and ester groups adjacent to the exomethylene group (Figure 6c). The existence of these groups results in an increased inhibitory activity that agrees with previous QSAR investigations.[13,18]

The long distances of 5.1, 5.9, and 6.6 Å describe the positions of different α,β-unsaturated carbonyl structures relative to each other (Figure 6d).

Nine distances between 4.4 and 10.3 Å from the AC-encoded HBP were included in the best model. In the calculation of the hydrogen-bonding potential, the functional groups were divided into proton donors and proton acceptors and mapped onto a surface. A correlation between the atom distances of the atom property χ_π and the distances of the surface points of HBP could not be found. However, in contrast to the descriptors resulting from χ_π, high values of HBP descriptors do not always correlate with high inhibitory activity. HBP descriptors probably describe the possibility of docking at the p65 subunit. Interestingly, SLs of activity classes 5 and 6 show AC values of HBP of about zero at a surface point distance of $r = 10.3$ Å, whereas SLs of activity class 1 exhibit values between 0.2 and 0.5, indicating that active SLs surface areas with a hydrogen-bonding potential have to be on the opposite side of the molecule to reach a distance of 10.3 Å.

Conclusions

Altogether, we show that the best generated CPGNN model is a valuable tool for the prediction of the NF-κB inhibitory activity of new SLs. In contrast to our previous study, only a single model based on a CPGNN was needed to describe the entire data set. The 19 descriptors used in our best model confirm the importance of α,β-unsaturated carbonyl structures for the NF-κB inhibitory activity. Moreover, by use of the 3D structure representation, the descriptors provide us with further, more detailed information on the structural features necessary for high inhibitory activity. Thus, our best CPGNN model can be used for the screening of structure databases for lead structures. Furthermore, the clear-cut structure interpretation of our model allows the de novo construction of new sesquiterpene lead structures.

Experimental Section

Data Set. The 103 SLs are listed in Table 6, and the structures are shown in Figure 7. The biological activity of SLs is expressed as the concentration in μM that totally blocks DNA binding of the transcription factor NF-κB in the electrophoretic mobility shift assay (EMSA), here called IC_{100}. Because of the experimental procedure, IC_{100} values of 5, 10, 12.5, 20, 25, 50, 100, 200, 300 μM can be assigned. To get better clustering, it was necessary to reduce the number of activity classes and to combine compounds with similar activity. This division gave classes with nearly the same number of SLs. For the training of neural networks (NN) and the prediction of the activities, the IC_{100} values were expressed for classes 1−6 (class 1, $IC_{100} = 5$ μM; class 2, $IC_{100} = 10$ and 12.5 μM; class 3, $IC_{100} = 20$ and 25 μM; class 4, $IC_{100} = 50$ μM; class 5, $IC_{100} = 100$ μM; class 6, $IC_{100} = 200$ and 300 μM). The origin of the SLs and their inhibitory activity are summarized in Siedle et al.[13]

An external test set of 14 SLs was used for validation. This set represents results of recent investigations and is listed in Table 7. The corresponding structures are presented in Figure 8.

Study of the NF-κB DNA Inhibitory Activity in an Electrophoretic Mobility Shift Assay. Jurkat T-cells were incubated for 1 h and subsequently stimulated with TNF-α for 1 h. Total protein extracts were prepared and analyzed for NF-κB binding activity in an EMSA. All experiments were reproduced at least once. Further experimental information is given in Siedle et al.[13]

Structure Representations. ANN vectorial representations of the objects need to be studied with a fixed number of entries. This was achieved in the following way.

The 3D structure generator CORINA[24−26] was used to generate single low-energy 3D conformations. SLs only exhibit few low-energy conformations. Moreover, the most important structural feature—the lactone ring—is always rather rigid. This encouraged us to go ahead with a single conformation only for each compound. Next, the atoms of a molecule were characterized by various physicochemical properties. All atomic properties used in this work (see Table 1) were calculated by PETRA (parameter estimation for the treatment of reactivity applications).[26,27] This program package comprises various empirical methods for the calculation of a wide panel of atomic and other physicochemical properties in organic molecules. In particular, the following properties were used: the σ-charge q_σ,[28] the π-charge q_π,[29] the polarizability α_d,[30] and the π-electronegativity χ_π.[31] To obtain a vectorial representation with a fixed number of entries, irrespective of the number of atoms in a molecule, a mathematical transformation of the atomic properties is required. This was achieved by using the radial distribution

Table 6. List of the Investigated SLs and Their Inhibitory Activity (IC_{100}) in the NF-κB DNA Binding Assay[a]

compd	name	IC_{100} (μM)
	I. Germacranolides without Furanoheliangolides	
1	2α-acetoxy-15-isovaleroylmiguanin + 2α-acetoxy-15-(2-methylbutyryl)miguanin	12.5
2	15-isobutyrylmiguanin	12.5
3	15-isovaleroylmiguanin + 15-(2-methylbutyryl)miguanin	12.5
4	9α,14-dihydroxy-15-isobutyryloxycostunolide	100
5	14-hydroxy-15-isovaleroyloxy-9-oxomelampolide + 14-hydroxy-15-(2-methylbutyryloxy)-9-oxomelampolide	12.5
6	1α-methoxy-15-isobutyryloxy-9-oxogermacra-4*e*,10(14),11(13)-trien-12,6α-olide	25
7	1*β*-methoxy-15-isobutyryloxy-9-oxogermacra-4*e*,10(14),11(13)-trien-12,6α-olide	25
8	1*β*-methoxy-15-isovaleroyloxy-9-oxogermacra-4*e*,10(14),11(13)-trien-12,6α-olide + 1*β*-methoxy-15-(2-methylbutyryloxy)-9-oxogermacra-4*e*,10(14),11(13)-trien-12,6α-olide	12.5
9	8*β*-hydroxy-9α-methacryloyloxy-14-oxoacanthospermolide	5
10	9α-hydroxy-8*β*-methacryloyloxy-14-oxoacanthospermolide	10
11	15-acetoxy-9α-methacryloyloxy-8*β*-hydroxy-14-oxoacanthospermolide	5
12	15-acetoxy-9α-hydroxy-8*β*-methacryloyloxy-14-oxoacanthospermolide	5
13	9α-hydroxy-8*β*-methacryloyloxy-14-oxoacanthospermolid-4α,5*β*-epoxide	10
14	miller-9*e*-enolide	5
15	miller-9*z*-enolide	10
16	1*β*-methoxy-miller-9*z*-enolide	10
17	4*β*,15-epoxy-miller-9*e*-enolide	5
18	4*β*,15-epoxy-miller-9*z*-enolide	5
19	9α-methoxy-miller-1(10)*z*-enolide	50
20	9α-acetoxy-miller-1(10)*z*-enolide	10
21	9α-acetoxy-4*β*,15-epoxy-miller-1(10)*z*-enolide	5
22	tatridin A	200
23	1-epitatridin B	50
24	tamirin	50
25	parthenolide	20
26	15-(2′,3′-epoxy)isobutyryloxymicrantholide	50
27	15-isobutyryloxymicrantholide	50
28	15-(2′-methyl-3′-hydroxy)butyryloxymicrantholide	50
29	15-(2′-hydroxy)isobutyryloxymicrantholide	50
30	14-acetoxy-15-(3′-hydroxy)methacryloyloxymicrantholide	50
31	15-(2′-methyl)butyryloxymicrantholide	50
32	15-(3′-hydroxy)isobutyryloxymicrantholide	100
33	15-methacryloyloxymicrantholide	50
34	15-(4-hydroxy)methacryloyloxymicrantholide	20
35	11*β*,13-dihydro-14-oxo-15-hydroxygermacra-1(10)*e*,4*z*-dien-12,8α-olide	300
36	costunolide	50
37	3-acetoxycostunolide	50
38	7-hydroxycostunolide	10
39	eupatoriopikrin	20
40	enhydrine	10
41	molephantin	10
42	molephantinin	10
43	9*β*-acetoxycostunolide	200
44	scandenolide	10
	II. Furanoheliangolides	
45	diversifolin	50
46	diversifolin methyl ether	200
47	tirotundin	200
48	centratherin	5
49	goiazensolide	20
50	isogoiazensolide	10
51	1-oxo-5-chlor-3,10-epoxy-8-methacryloyloxy-germacra-2,4(15),11(13)-trien-12,6α-olide	5
52	15-hydroxy-eremantholide B	100
53	15-acetoxy-eremantholide B	5
54	15-deoxybudlein A	5
55	atripliciolidtiglate	5
56	3,10-hydroxy-2-methoxy-8-(2-methylpropanoyloxy)germacra-4,11(13)-dien-12,6α-olide	200
57	15-deoxygoiazensolide	10
58	budlein A	5
59	2*β*-methoxy-2-deethoxyphantomolin	10
60	2*β*-methoxy-2-deethoxy-8-*O*-deacylphantomolin-8-*O*-tiglinate	10
	III. Guaianolides	
61	cumambrin A	20
62	cumambrin B	100
63	dehydroleucodin	50
64	3-chlorodehydroleucodin	20
65	3,4-epoxydehydroleucodin	5
66	2-oxoguaia-1(5),11(13)-dien-12,8*β*-olide	20
67	2-oxoguaia-1(5),11(13)-dien-12,8α-olide	20
68	thieleanin	20
69	3-oxoguaia-4,11(13)-dien-12,8*β*-olide	50
70	2-oxoguaia-1(5),11(13)-dien-12,6*β*-olide	20

Table 6 (Continued)

compd	name	IC_{100} (μM)
	III. Guaianolides (Continued)	
71	2-oxoguaia-1,4(15), 11(13)-trien-12,8β-olide	5
72	2-oxoguaia-1,4,11(13)-trien-12,8α-olide	10
73	3-oxoguaia-1(2),11(13)-dien-12,8α-olide	50
74	2-oxoguaia-1(5),11(13)-dien-12,6α-olide	20
75	2-oxo-8β-methacryloyloxyguaia-1(10),3,11(13)-trien-12,6α-olide	10
76	2-oxo-8β-epoxyangelicoyloxyguaia-1(10),3,11(13)-trien-12,6α-olide	20
77	2-oxo-8β-methacryloyloxy-10β-hydroxyguaia-3,11 (13)-dien-12,6α-olide	50
78	2-oxo-8β-methacryloyloxy-10α-hydroxyguaia-3,11 (13)-dien-12,6α-olide	20
79	3β,10α-dihydoxyguaia-4(15),11(13)-dien-12,6α-olide	200
80	2-oxo-15-hydroxyguaia-1(10),3,11(13)-trien-12,6α-olide	50
81	3β-acetoxyguaia-4(15),10(14),11(13)-trien-12,6α-olide	100
82	3β-senecioyloxyguaia-4(15),10(14),11(13)-trien-12,6α-olide	200
	IV. Pseudoguaianolides	
83	helenalin	10
84	11α,13-dihydrohelenalin	200
85	chamissonolide	200
86	2,3-dihydroaromaticin	50
87	mexicanin I	20
88	helenalinisobutyrate	20
89	11α,13-dihydrohelenalinacetate	200
90	11α,13-dihydrohelenalintiglinate	100
91	11α,13-dihydrohelenalinmethacrylate	100
	V. Hypocretenolides	
92	14-hydroxycretenolide	50
93	14-acetoxycretenolide	100
	VI. Eudesmanolides	
94	douglanin	300
95	santamarin	100
96	ludovicin A	200
97	3α-hydroxyreynosin	200
98	ludovicin B	100
99	1β-hydroxy-4α-hydroxy-15-isobutyryloxyeudesma-11(13)-en-12,8β-olide	200
100	1β-hydroxy-4β-hydroxy-15-isobutyryloxyeudesma-11(13)-en-12,8β-olide	200
101	1β-acetoxy-4α-hydroxy-15-isobutyryloxyeudesma-11(13)-en-12,8β-olide	100
102	1β-hydroxy-15-isobutyryloxyeudesma-3,11(13)-dien-12,8β-olide	50
103	2-oxo-3-acetoxyeudesma-3,11(13)-dien-12,8β-olide	100

[a] Further information concerning the origin is given in Siedle et al.[13]

function (RDF) code (eq 1) as obtained with the program RCODE.[32] A slightly simplified interpretation of the radial distribution function for an ensemble of atoms is a kind of probability distribution of the individual interatomic distances r:

$$g(r) = \sum_{i=1}^{N-1} \sum_{j>i}^{N} p_i\, p_j\, e^{-B(r-r_{ij})^2} \quad (1)$$

where r_{ij} represents the distance between atoms i and j, N is the number of atoms in a molecule, p_i and p_j are atomic properties associated with the atoms i and j, respectively, and B is a smoothing factor.[33,34]

Furthermore, an RDF-like function was used for encoding atomic properties called local RDF (L-RDF, eq 2) calculated with the program ARDF:[35]

$$f(r) = p_c \sum_{i=1, i\neq c}^{N} p_i\, e^{-B(r-r_{i,c})^2} \quad (2)$$

Similar to the RDF code, the L-RDF can be interpreted as a probability distribution to find an atom in a spherical shell around one (the central) atom c. The exocyclic carbon atom of the lactone ring was selected as the central atom because it is involved in alkylating p65/NF-κB in a Michael-type addition.[18] In effect, we selected this exocyclic carbon atom as the reaction site and analyzed how its reactivity is influenced by other atoms whereby this influence is considered as distance-dependent. Furthermore, we investigated different atomic properties that are expected to influence this reactivity.

The SURFACE program[26,36] was used to generate vectors of autocorrelation coefficients[37] (AC, eq 3) of different molecular surface properties: the molecular electrostatic potential (MEP) and the hydrogen-bonding potential (HBP).

$$A(d_l,d_u) = \sum_{i=1}^{N} \sum_{j=i}^{N} \delta(d_{ij},d_l,d_u)\, p_j\, p_i \quad (3)$$

$$\delta(d_{ij},d_l,d_u) = \begin{cases} 1 \;\forall\; d_l < d_{ij} \le d_u \\ 0 \;\forall\; d_{ij} \le d_l \vee d_{ij} > d_u \end{cases}$$

In eq 3 between the boundaries d_l (lower) and d_u (upper) the products of property p for surface points i and j having a Euclidian distance d within this interval were summarized.

All representations (RDF, L-RDF, and AC) are uniform and invariant under translations and rotations of molecules. The functions are given as vectors calculated with a sampling rate of 0.1 Å. The dimension was set to 128, $g(r)$, $f(r)$, and $A(r)$ were defined in the interval 0.0−12.8 Å.

For a clear understanding, the following definitions are established. Each value within the vector is called descriptor, whereas the atomic property or molecular surface property that results in a vector is called property. The symbols and abbreviations of the atomic properties that have been encoded into RDF vectors are given in Table 1. In a similar fashion, the abbreviation HBP or MEP indicates which surface properties were encoded into the AC (see Tables 1−3).

Reduction of the Number of Descriptors and Preprocessing. Two data sets were obtained from each atom property using RDF and L-RDF codes, and one data set was obtained for each molecular surface property. From these initial data sets, descriptors with

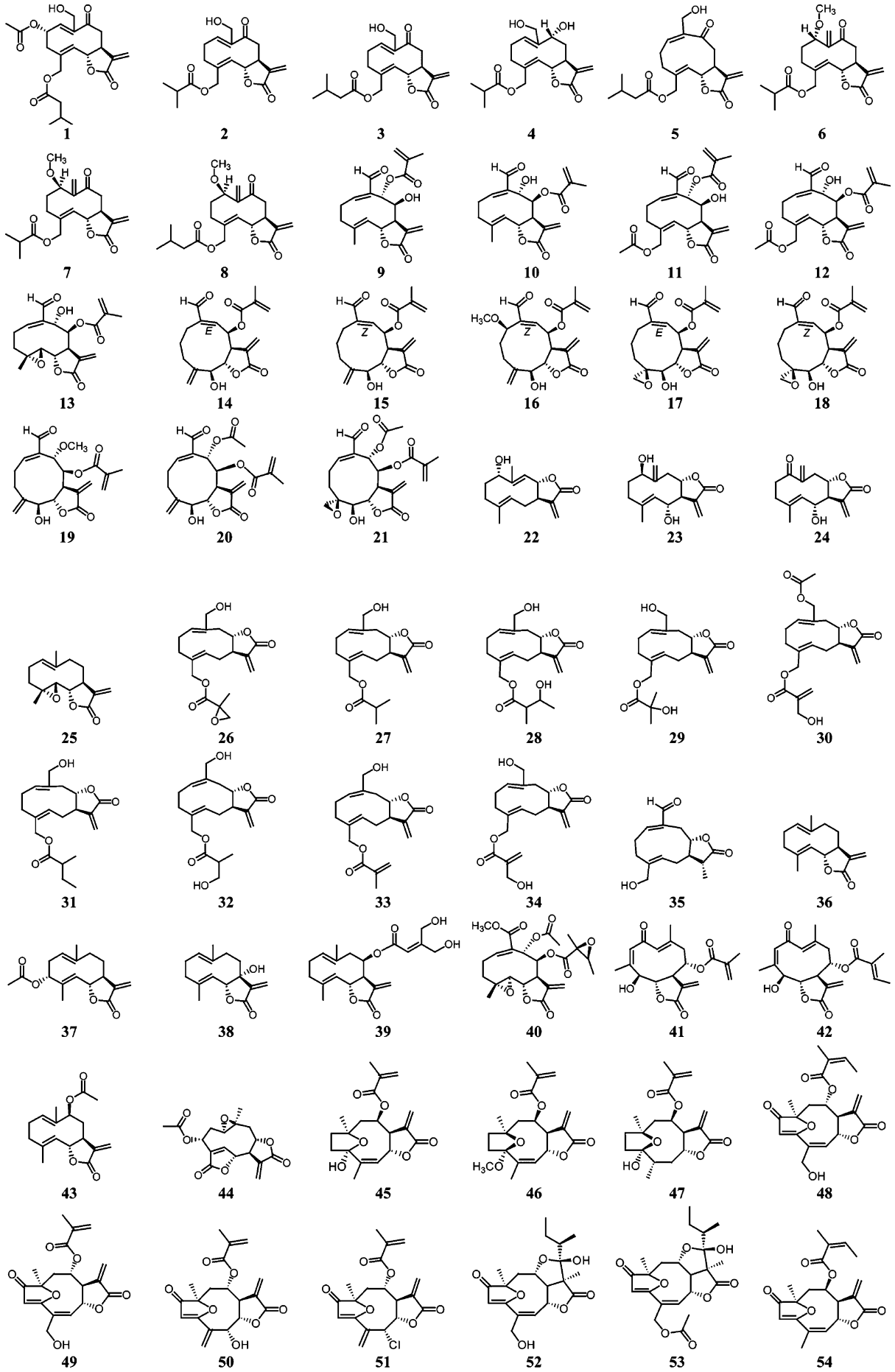

Figure 7. (Continued on next page)

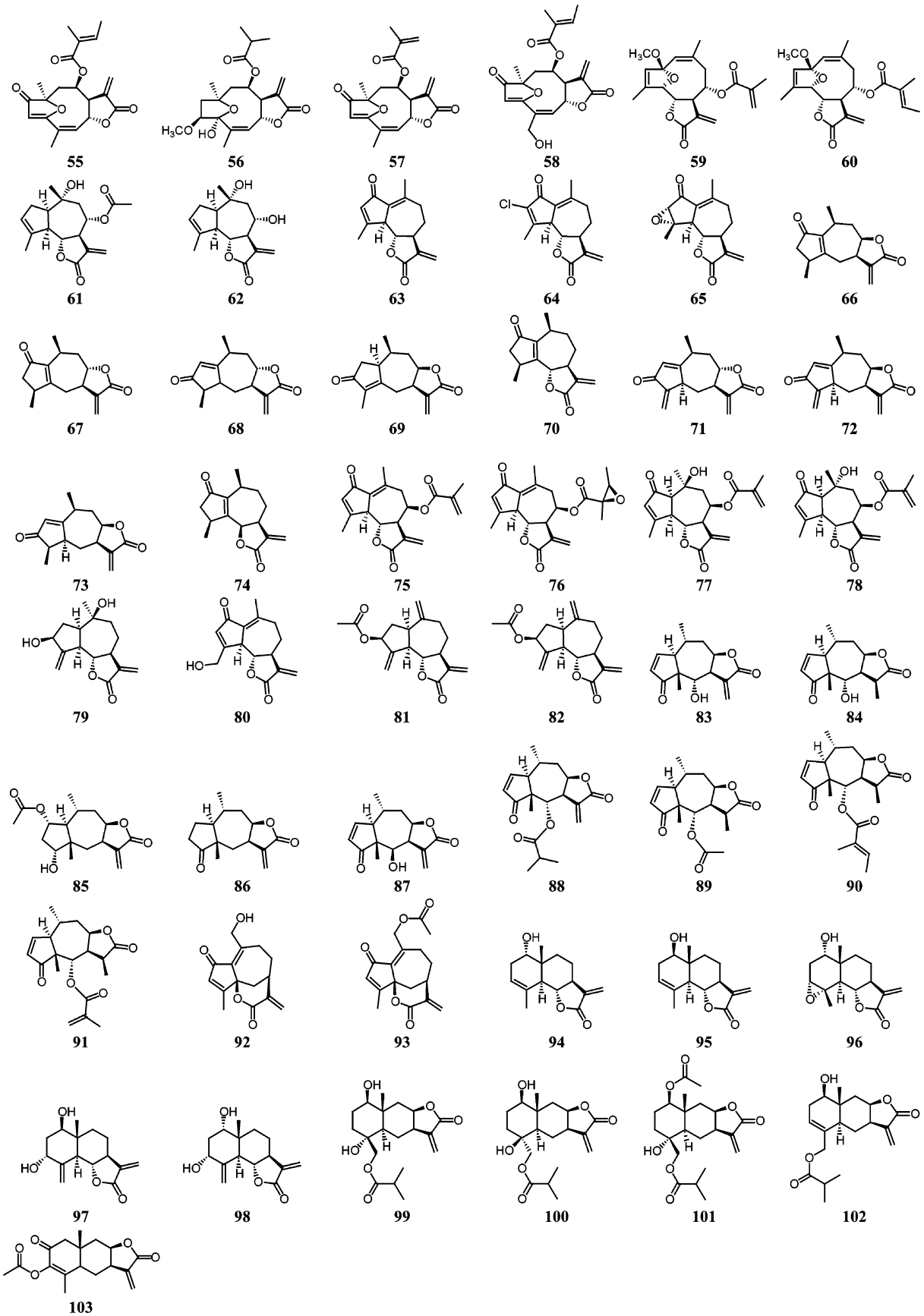

Figure 7. Structures of the investigated sesquiterpene lactones representing six structural classes: germacranolides (**1**−**44**), furanoheliangolides (**45**−**60**), guaianolides (**61**−**82**), pseudoguaianolides (**83**−**91**), hypocretenolides (**92** and **93**), and eudesmanolides (**94**−**103**).

constant values for all molecules were excluded. Subsequently, for each data set the mean values and standard deviations of the descriptor values of each activity class were calculated and compared using a statistical *t*-test. Only those descriptors were selected by which the individual classes of the NF-κB inhibitory activity could be partially differentiated. This was possible by a

Table 7. List of the SLs of the External Test Set, Their Inhibitory Activity (IC_{100}) in the NF-κB DNA Binding Assay, and Their Origin

compd	name	IC_{100} (μM)	origin
N1	2β-ethoxy-2,3-dihydrohelenalin-6-*O*-acetate	50	20
N2	helenalin-2-methylbutyrate	10	20
N3	helenalinmethacrylate	5	20
N4	niveusin A	20	21
N5	2β-hydroxy-1-desoxyniveusin A	50	21
N6	niveusin B	20	21
N7	4,5-isobudlein A	10	21
N8	8α-(2′*R*,3′*R*-epoxy-2′-methylbutyryloxy)-4α-hydroxy-9-oxo-5β*H*-eudesm-1*Z*,11(13),dien-6β,12-olide	10	22
N9	8α-(2′*R*,3′*R*-epoxy-2′-methylbutyryloxy)-9α-hydroxymontahibisciolide	200	22
N10	9β-(2′*S*,3′*S*-epoxy-2′-methylbutyryloxy)-8α-hydroxy-germacra-4*E*,1(10)*E*-dien-6β,12-olide	50	22
N11	8α-(2′*S*,3′*S*-epoxy-2′-methylbutyryloxy)-9-oxo-germacra-4*E*,1(10)*E*-dien-6β,12-olide	50	22
N12	8α-(2′*S*,3′*S*-epoxy-2′-methylbutyryloxy)-1α-methoxy-9-oxo-10α*H*-germacra-4*E*-en-6β,12-olide	20	22
N13	8α-hydroxy-9β-tigloyloxy-germacra-4*E*,1(10)*E*-dien-6β,12-olide	100	22
N14	dehydrosaussurealactone	100	23

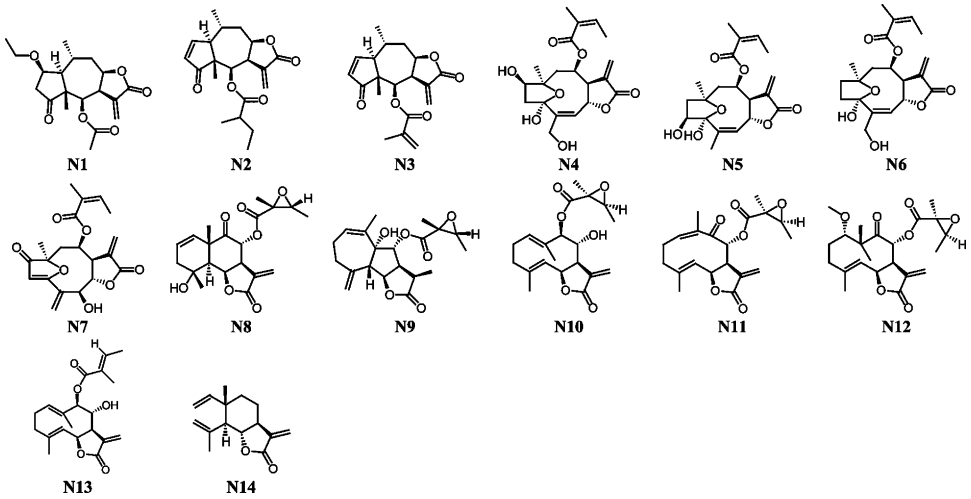

Figure 8. Structures of the SLs of the external test set representing six structural classes: pseudoguainolides (**N1**−**N3**), furanoheliangolides (**N4**−**N7**), eudesmanolides (**N8**), montahibisciolides (**N9**), germacranolides (**N10**−**N13**), and elemanolides (**N14**).

t-value of more than 2 as controlled by a graphical comparison of descriptors with different *t*-values. The selected descriptors of each data set were combined into a new data set. Thus, a reduction from 128 descriptors to between 7 and 12 was achieved for each data set (Table 1).

All descriptors of the reduced vectors were normalized between zero and 1 using a range scaling procedure:

$$x_{\mathrm{new},i} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \qquad (4)$$

The original distribution is thereby maintained.[16]

Generation of a Self-Organizing Network. The software SONNIA (self-organizing neural network for information analysis)[38] was used to generate the Kohonen networks and the CPGNN.

Both ANNs consist of the *n*-dimensional input layer, with *n* descriptors resulting from RDF, L-RDF, or AC encoding. In addition, CPGNNs also have an output layer with the NF-κB inhibitory activity expressed as classes between 1 and 6. Thereby, the activity is encoded by natural numbers from 1 to 6 in one layer or in a binary code using one output layer for each activity class (Figure 9).[14,15,39]

At first, Kohonen networks were used to search for atomic properties important for the inhibitory activity of the SLs. Training of the Kohonen network started with the presentation of a vector (a molecule) of input variables to all neurons. The neuron that has weights being closest to the input variables is selected as the winning neuron in the learning algorithm. To improve the response to the same input in the next epoch, the weights of the so-called winning neuron are adjusted to the input vector as well as the weights of neurons in the neighborhood. The degree of adaptation decreases with increasing distance to the winning neuron. This adaptation is repeated for each vector of an input molecule. After training, the response of the network is calculated for each vector of the data set. Subsequently, the projection of the data set into the two-dimensional space is performed by mapping the activity of each vector into the coordinates of its winning neuron. Additionally, the CPGNN were used for validation and prediction. During training of CPGNN the weights of both input and output layers are adapted, but only the input layers are used to determine the winning neuron. The results are visualized as two-dimensional maps by looking at the output layers and identifying each molecule by its class assignment.

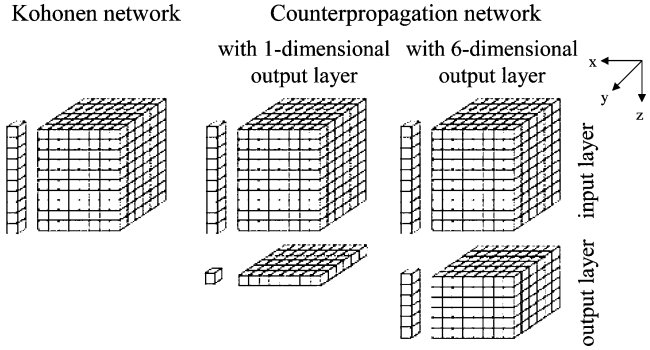

Figure 9. Architecture of the used neural networks and the associated vectors. Whereas the Kohonen network consists only of an input layer, counterpropagation networks also have an output layer. The output layer contains the information on the activity that is classified into six activity classes. By use of a one-dimensional output layer, the activity is encoded as a natural number between 1 (highly active) and 6 (low activity). By use of a six-dimensional output layer, the activity is binary-coded whereby each layer refers to one activity class.

Different sizes of the networks were tested using different data sets. The clustering ability was evaluated by the same quality criteria as described in the next paragraph. It turned out that a size of $6 \times 15 \times n$ was the best, where n is the dimension of the input vector. Therefore, this size was used throughout the study. The topology of both ANNs was toroidal. To allow a direct comparison of all experiments, the same training parameters[40] were used.

Search for Good Clustering Properties. Kohonen networks were trained with each data set. The resulting output maps were evaluated according to their clustering ability. At first, a visual evaluation of the output maps was performed. Subsequently, the clustering for each neuron was calculated. Clustering was defined as correct when the same inhibitory activity dominated in neighboring neurons. If other activities predominate the neighborhood, the neuron is isolated and the clustering of this neuron was wrong. If there were no preponderance of any activity, the clustering is defined as drawn. Occupancy and the number of conflicts were further quality criteria. Thus, a low occupancy, indicated by many empty neurons, implies missing data. Conflicts occur when molecules with different inhibitory activity occur within the same neuron. It must be pointed out that neurons including SLs with neighboring activity classes were not regarded as conflicts. Considering that all SLs with activities between >10 and 20 μM fall in the activity class of "20 μM" and between >5 and 10 μM in the activity class of "10 μM", the activity of SLs with neighboring activity classes (e.g., 10 and 20 μM) can have a more similar activity than two SLs within the same activity class. In an analogous manner it was decided whether the predicted class of an SL was correct or not in the validation process.

Validation. The best model obtained by evaluation, selection, and combination of RDF, L-RDF, and AC-descriptors was validated by using internal and external validation models. Both models use CPGNN with a six-dimensional output layer (Figure 9).

For the internal validation, 10-fold cross-validation (CV) was used.[41] The original data set was divided by random splitting into 10 subsets. A model was built based on nine subsets using the remaining set as test set. To eliminate possible initializing effects, 19 random seeds were used to initialize the CPGNN. The procedure was repeated 10 times until each subset had been used as a test set. Then a new splitting was performed and the whole procedure was repeated. Altogether, 19 random splittings were performed. This number is independent of the 19 random seeds. Consequently, 3610 ($10 \times 19 \times 19$) CPGNN simulations using 6×15 rectangular topology were the basis for one 10-fold CV.

Furthermore, we tested the best model using the external test set of 14 SLs, which was not used before, on 19 CPGNNs trained with the entire data set of 103 SLs. Each CPGNN was initialized by one random seed.

Acknowledgment. We appreciate the helpful discussions with S. Spycher and E. Pellegrini from the Computer-Chemie-Centrum, Erlangen, and are grateful to the BMBF (HepatoSys) for supporting this work within the systems biology scheme.

References

(1) Schmidt, T. J. Toxic activities of sesquiterpene lactones: structural and biochemical aspects. *Curr. Org. Chem.* **1999**, *3*, 577−608.
(2) Garcia-Pineres, A. J.; Castro, V.; Mora, G.; Schmidt, T. J.; Strunck, E.; Pahl, H. L.; Merfort, I. Cysteine 38 in p65/NF-κB plays a crucial role in DNA binding inhibition by sesquiterpene lactones. *J. Biol. Chem.* **2001**, *276*, 39713−39720.
(3) Mazor, R. L.; Menendez, I. Y.; Ryan, M. A.; Fiedler, M. A.; Wong, H. R. Sesquiterpene lactones are potent inhibitors of interleukin 8 gene expression in cultured human respiratory epithelium. *Cytokine* **2000**, *12*, 239−245.
(4) Hehner, S. P.; Heinrich, M.; Bork, P. M.; Vogt, M.; Ratter, F.; Lehmann, V.; Schulze-Osthoff, K.; Droge, W.; Schmitz, M. L. Sesquiterpene lactones specifically inhibit activation of NF-κB by preventing the degradation of IκBα and IκBβ. *J. Biol. Chem.* **1998**, *273*, 1288−1297.
(5) Lyss, G.; Knorre, A.; Schmidt, T. J.; Pahl, H. L.; Merfort, I. The anti-inflammatory sesquiterpene lactone helenalin inhibits the transcription factor NF-κB by directly targeting p65. *J. Biol. Chem.* **1998**, *273*, 33508−33516.
(6) Garcia-Pineres, A. J.; Lindenmeyer, M. T.; Merfort, I. Role of cysteine residues of p65/NF-κB on the inhibition by the sesquiterpene lactone parthenolide and *N*-ethyl maleimide, and on its transactivating potential. *Life Sci.* **2004**, *75*, 841−856.
(7) Bonizzi, G.; Karin, M. The two NF-κB activation pathways and their role in innate and adaptive immunity. *Trends Immunol.* **2004**, *25*, 280−288.
(8) Senftleben, U.; Karin, M. The IKK/NF-κB pathway. *Crit. Care Med.* **2002**, *30*, S18−S26.
(9) Bacher, S.; Schmitz, M. L. The NF-κB pathway as a potential target for autoimmune disease therapy. *Curr. Pharm. Des.* **2004**, *10*, 2827−2837.
(10) Chariot, A.; Meuwis, M. A.; Bonif, M.; Leonardi, A.; Merville, M. P.; Gielen, J.; Piette, J.; Siebenlist, U.; Bours, V. NF-κB activating scaffold proteins as signaling molecules and putative therapeutic targets. *Curr. Med. Chem.* **2003**, *10*, 593−602.
(11) Karin, M.; Yamamoto, Y.; Wang, Q. M. The IKK NF-κB system: a treasure trove for drug development. *Nat. Rev. Drug Discovery* **2004**, *3*, 17−26.
(12) Verma, I. M. Nuclear factor (NF)-κB proteins: therapeutic targets. *Ann. Rheum. Dis.* **2004**, *63* (Suppl. 2), ii57−ii61.
(13) Siedle, B.; Garcia-Pineres, A. J.; Murillo, R.; Schulte-Monting, J.; Castro, V.; Rüngeler, P.; Klaas, C. A.; Da Costa, F. B.; Kisiel, W.; Merfort, I. Quantitative structure−activity relationship of sesquiterpene lactones as inhibitors of the transcription factor NF-κB. *J. Med. Chem.* **2004**, *47*, 6042−6054.
(14) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 1999.
(15) Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59−69.
(16) Mazzatorta, P.; Vracko, M.; Jezierska, A.; Benfenati, E. Modeling toxicity by using supervised Kohonen neural networks. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 485−492.
(17) Vracko, M.; Gasteiger, J. A QSAR Study on a set of 105 flavonoid derivatives using descriptors derived from 3D structures. *Internet Electron. J. Mol. Des.* **1997**, 527−544.
(18) Schmidt, T. J. Helenanolide-type sesquiterpene lactones. III. Rates and stereochemistry in the reaction of helenalin and related helenanolides with sulfhydryl containing biomolecules. *Bioorg. Med. Chem.* **1997**, *5*, 645−653.
(19) Rungeler, P.; Castro, V.; Mora, G.; Goren, N.; Vichnewski, W.; Pahl, H. L.; Merfort, I.; Schmidt, T. J. Inhibition of transcription factor NF-κB by sesquiterpene lactones: a proposed molecular mechanism of action. *Bioorg. Med. Chem.* **1999**, *7*, 2343−2352.
(20) Kos, O.; LIndenmeyer, M. T.; Tubaro, A.; Sosa, S.; Merfort, I. New sesquiterpene lactones from *Arnica* tincture prepared from fresh flowerheads of Arnica montana. *Planta Med.* **2005**, *71*, 1044−1052.
(21) Lindenmeyer, M. T. Untersuchungen zum molekularem Wirkmechanismus der antiinflammatorischen Aktivität von Sesquiterpenlactonen (Understanding the molecular mechanism of antiinflammatory activity of sesquiterpene lactones). Ph.D. Thesis, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany, 2004.
(22) Müller, S.; Murillo, R.; Castro, V.; Brecht, V.; Merfort, I. Sesquiterpene lactones from *Montanoa hibiscifolia* inhibit the transcription factor NF-κB. *J. Nat. Prod.* **2004**, *67*, 622−630.
(23) Müller, S. Isolierung und Strukturaufklärung terpenoider Inhaltsstoffe aus *Montanoa hibiscifolia* und *Oyedaea verbesinoides* sowie Untersuchungen ihrer biologischen Aktivität (Isolation and structure determination of terpene materials from *Montanoa hibiscifolia* and *Oyedaea verbesinoides* as well as investigations of their biological activities). Ph.D. Thesis, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany, 2004.
(24) Computer-Chemie-Centrum Universität Erlangen-Nuremberg. http://www2.chemie.uni-erlangen.de/software/corina/index.html (accessed 2004).
(25) Sadowski, J.; Gasteiger, J. From atoms and bonds to 3-dimensional atomic coordinates. Automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(26) Molecular Networks GmbH, Erlangen, Germany. http://www.mol-net.com (accessed 2005).
(27) Computer-Chemie-Centrum, University of Erlangen-Nuremberg. http://www2.chemie.uni-erlangen.de/software/petra/index.html (accessed 2005).
(28) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(29) Gasteiger, J.; Saller, H. Calculation of the charge-distribution in conjugated systems by a quantification of the resonance concept. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687−689.
(30) Gasteiger, J.; Hutchings, M. G. Quantification of effective polarizability. Applications to studies of X-ray photoelectron-spectroscopy

and alkylamine protonation. *J. Chem. Soc., Perkin Trans. 2* **1984**, 559−564.
(31) Marsili, M.; Gasteiger, J. Pi-charge distributions from molecular topology and pi-orbital electronegativity. *Croat. Chem. Acta* **1980**, *53*, 601−614.
(32) *Computer Program RCODE*; Computer-Chemie-Centrum, Universität Erlangen-Nuremberg: Erlangen, Germany, 2005; unpublished results.
(33) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. Deriving the 3D structure of organic molecules from their infrared spectra. *Vib. Spectrosc.* **1999**, *19*, 151−164.
(34) Hemmer, M. C.; Gasteiger, J. Prediction of three-dimensional molecular structures using information from infrared spectra. *Anal. Chim. Acta* **2000**, *420*, 145−154.
(35) *Computer Program ARDF*; Computer-Chemie-Centrum, Universität Erlangen-Nuremberg: Erlangen, Germany, 2005; unpublished results.
(36) Computer-Chemie-Centrum, University of Erlangen-Nuremberg. http://www2.chemie.uni-erlangen.de/software/surface/index.html (accessed 2005).
(37) Gasteiger, J.; Engel, T. *Chemoinformatics: A Textbook*; Wiley-VCH: Weinheim, Germany, 2003.
(38) *SONNIA Software*, version 4.1; Molecular Networks GmbH: Erlangen, Germany, 2005.
(39) Anzali, S.; Gasteiger, J.; Holzgrabe, U.; Polanski, J.; Sadowski, J.; Teckentrup, A.; Wagener, M. The use of self-organizing neural networks in drug design. In *3D QSAR in Drug Design*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Kluwer/ESCOM: Dordrecht, The Netherlands, 1998; Vol. 2, pp 273−299.
(40) *SONNIA User Manual*; Molecular Networks GmbH: Erlangen, Germany, 2005; http://www.mol-net.de/software/sonnia/sonnia_manual.pdf.
(41) Spycher, S.; Pellegrini, E.; Gasteiger, J. Use of structure descriptors to discriminate between modes of toxic action of phenols. *J. Chem. Inf. Model.* **2005**, *45*, 200−208.

JM051125N