Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity

Gerhard Klebe,* Ute Abraham, and Thomas Mietzner

BASF AG, Main-Laboratory, Carl-Bosch Strasse, D-67056 Ludwigshafen, FRG

Received June 2, 1994[⊗]

An alternative approach is reported to compute property fields based on similarity indices of drug molecules that have been brought into a common alignment. The fields of different physicochemical properties use a Gaussian-type distance dependence, and no singularities occur at the atomic positions. Accordingly, no arbitrary definitions of cutoff limits and deficiencies due to different slopes of the fields are encountered. The fields are evaluated by a PLS analysis similar to the CoMFA formalism. Two data sets of steroids binding to the corticosteroid-binding-globulin and thermolysin inhibitors were analyzed in terms of the conventional CoMFA method (Lennard-Jones and Coulomb potential fields) and the new comparative molecular similarity indices analysis (CoMSIA). Models of comparative statistical significance were obtained. Field contribution maps were produced for the different models. Due to cutoff settings in the CoMFA fields and the steepness of the potentials close to the molecular surface, the CoMFA maps are often rather fragmentary and not contiguously connected. This makes their interpretation difficult. The maps obtained by the new CoMSIA approach are superior and easier to interpret. Whereas the CoMFA maps denote regions apart from the molecules where interactions with a putative environment are to be expected, the CoMSIA maps highlight those regions within the area occupied by the ligand skeletons that require a particular physicochemical property important for activity. This is a more significant guide to trace the features that really matter especially with respect to the design of novel compounds.

Introduction

Prerequisite for a successful binding of a low molecular weight ligand to a macromolecular receptor is a negative Gibbs free energy of binding. Experimentally, this value is available from the observed binding constant. In QSAR studies, one of the important goals is the quantitative correlation of molecular structure with the binding constant and subsequently the prediction of this property for novel compounds. Furthermore, these methods should help to characterize those spatial features that are responsible for activity changes in a series of drug molecules. The methods that can be applied to achieve this goal are rather different whether the 3D structure of the receptor is known or not. In the latter case, any quantitative correlation has to be focused on relative differences of appropriate molecular descriptors within a series of drug molecules. Some of these descriptors require a mutual structural alignment of the molecules in order to determine their gradual changes within the series. In the present paper, we will concentrate on this situation.

In the absence of structural information about the receptor, the derivation of a relevant structural alignment is not trivial. Recently, we reported on different approaches to resolve the alignment problem.[1] Convincing results have been obtained using a conformational analysis with MIMUMBA[2] combined with a subsequent superposition according to a modified version of the program SEAL.[3] The alignment condition in SEAL is based on mutual similarity indices pairwise calculated between all atoms of the molecules being compared. To consider the local environment of each atom in this pairwise comparison, these similarity indices take the properties of the neighboring atoms in a distance-dependent fashion into account.

With respect to a set of ligands for which the binding geometry has been experimentally observed by protein crystallography and which bind to common protein receptors, the combined MIMUMBA/SEAL approach[1] allows one to reproduce the experimentally observed alignment within a mean rms deviation of about 1-1.5 Å. Considering the accuracy of experimentally determined ligand/protein complexes, this deviation roughly approaches the accuracy limits defined by the experimental references (the positional errors of such alignments can be estimated to about 0.7 Å in space). In light of these results, it is supposed that alignments obtained by the combined MIMUMBA/SEAL approach are relevant for an attempted comparative analysis.

As mentioned above, the target property to be correlated and predicted in a comparative analysis is a free energy value. It can be imagined that enthalpic contributions to the binding constant are covered by molecular descriptors that explore the capabilities of molecules to perform intermolecular interactions with a putative receptor. In the CoMFA method[4] gradual changes of the interaction properties are mapped by evaluating the potential energy at regularly spaced grid points surrounding the mutually aligned molecules. The contributions due to dispersion forces between molecules are frequently described by Lennard-Jones-type potentials, and electrostatic properties are characterized by Coulomb-type potentials.

Entropic contributions to the binding affinity are more difficult to describe. A major factor arises from the solvent-to-protein transfer. As shown in several studies, this portion approximately correlates with the size of the hydrophobic surface area of the drug molecules.[5] Accordingly, descriptors are required that appropriately

[⊗] Abstract published in *Advance ACS Abstracts*, October 15, 1994.

0022-2623/94/1837-4130$04.50/0

quantify the relative differences of the hydrophobic surface areas. In addition, in a data set covering molecules with distinct conformational flexibility, differences in this property have to be considered since the immobilization at the binding site involves important entropy changes.

The CoMFA approach uses in its standard implementation only Lennard-Jones and Coulomb potentials. Evidence has been collected that these potentials solely describe the energetic contributions to the binding constants.[6] Entropic influences seem to be neglected or insufficiently covered. In order to include entropic contributions, some kind of field considering the differences in hydrophobic surface contributions is required. Hydrophobic fields have been described by Kellog and Abraham[7] and are implemented into the program HINT. Furthermore, using a water probe in Goodford's GRID program[8] allows one to map hydrophobic surface regions in terms of a field. This field and other potential fields with various functional forms have been applied in CoMFA analyses.[9]

The fields presently used in CoMFA imply some additional problems. For example, the Lennard-Jones potential is very steep close to the van der Waals surface. As a consequence, the potential energy expressed at grid points in the proximity of the surface changes dramatically. It is likely that values from this region display significant descriptors in a QSAR.[10,11] Accordingly, just some small mutual shifts of the molecules or minor conformational changes can result in strong variations of these descriptors. Nevertheless, these shifts can be so small that they are easily accepted as "nearly identical" by visual inspection.

Furthermore, the Lennard-Jones and Coulomb potentials show singularities at the atomic positions. To avoid unacceptably large values, the potential evaluations are normally restricted to the regions outside the molecules, and some arbitrarily fixed cutoff values are defined. Due to differences in the slope of the potentials (e.g., Lennard-Jones and Coulomb) these cutoff values are exceeded for the different terms at different distances from the molecules.[11] This requires further arbitrary settings to adjust the two fields in a simultaneous evaluation and can involve the loss of information about one of the fields. For the interpretation of CoMFA results, in particular with respect to the design of novel compounds, contour maps of the relative spatial contributions of the different fields are extremely useful tools.[12] However, due to the described cutoff settings and the steepness of the potentials close to the molecular surfaces, these maps are often not contiguously connected and accordingly are difficult to interpret.

To overcome the outlined problems, an alternative approach to derive molecular descriptors for a comparative analysis is described in the present paper. On the basis of the alignment function used in SEAL that revealed convincing results for a spatial comparison of molecules, similarity indices are calculated in space. Using a common probe, these similarity indices are enumerated for each of the aligned molecules in the data set at regularly spaced grid points. They do not exhibit a direct measure of similarity determined between all mutual pairs of molecules. Instead, they are indirectly evaluated via the similarity of each molecule in the data set with a common probe atom which is placed at the intersections of a surrounding lattice. In determining this similarity, the mutual distance between the probe atom and the atoms of the molecules of the data set is considered. A functional form has been selected (Gaussian-type functions, no singularities) for this distance dependence so that no arbitrary definition of cutoff limits is required and the indices can be calculated at all grid points. In principle, any relevant physicochemical property can be considered in this approach to calculate a "field" of similarity indices. However, in a first approach only steric, electrostatic, and hydrophobic properties have been used. According to the considerations above, it is supposed that the most important contributions responsible for binding affinity are covered by these properties. The distance dependence of the different properties is equivalently handled in all cases. The applied Gaussian-type functional form defines a significantly smoother distance dependence compared to, e.g., the Lennard-Jones potential. The obtained indices are evaluated in a PLS analysis[13] according to the usual CoMFA protocol. This comparative molecular similarity index analysis (CoMSIA) has been applied to data sets of steroids binding to the corticosteroid-binding globulin (CBG) and thermolysin inhibitors. The results are compared to those obtained from CoMFA studies.

The present approach implies moving from field descriptors based on well established and generally accepted potentials (Lennard-Jones and Coulomb) to some arbitrary descriptors considering the spatial similarity or dissimilarity of molecules. However, a statistical approach such as a 3D-QSAR analysis seeks to correlate *relative* differences of discriminating molecular descriptors with a dependent property, e.g., the binding affinity of the drug molecules. The descriptors need not necessarily display partitions of interaction energy terms. They only have to correlate with these contributions in a uniform manner. Recently, Good et al.[14,15] reported on the successful evaluation of similarity indices in correlation and predicting the activity of aligned molecules. Since the authors only used integral similarity indices of the entire molecules in the analysis, limited information about spatial features and characteristics is available that are responsible for the variation of the activity with the 3D structure. However, this spatial aspect is perhaps the most important guide to understand what really matters and a valuable tool to assist the design of novel compounds. With the present evaluation technique substantially improved contour maps are obtained that can easily be interpreted and used as visualization tool in designing novel compounds. Whereas the level-dependent contouring of usual CoMFA-field contributions highlights those regions in space where the aligned molecules would favorably or unfavorably interact with a possible environment, the CoMSIA-field contribution denote those areas within the region occupied by the ligands that "favor" or "dislike" the presence of a group with a particular physicochemical property. This association of required properties with a possible ligand shape is a more direct guide to check whether all features important for activity are present in the structures under consideration.

Computational Methods

CoMFA Analyses. All CoMFA analyses were performed using SYBYL,[16] version 6.2, running on a Silicon Graphics Indigo R4000. The steric and electrostatic potential fields (Lennard-Jones potential as defined in SYBYL[16] and Coulomb potential from AM1 point

Table 1. Summary of the Results of the Different CoMFA and CoMSIA Analyses of the Steroid Data Set

	alignment according to ref 4[a]		alignment present study[a]	
	CoMFA[a,b] (1)	CoMSIA (2)	CoMFA (3)	CoMSIA (4)
q^2	0.662	0.662	0.598	0.665
s_{press}	0.719	0.763	0.832	0.759
r^2	0.897	0.941	0.947	0.937
S	0.397	0.320	0.303	0.330
no. comp	2	4	4	4
fraction				
steric		0.086	0.543	0.073
electrostatic		0.535	0.457	0.513
hydrophobic		0.378		0.415
box				
stepsize (Å)		1	2	1
x		−10 to 9	−9.5 to 8.9	−10 to 9
y		−9 to 10	−8.6 to 9.1	−9 to 10
z		−10 to 7	−9.5 to 6.4	−10 to 7

[a] Numbers of analysis given in parentheses. [b] Results taken from ref 4.

charges[17] or Gasteiger-Marsili charges[18]) for CoMFA were calculated using a lattice with 2 and 1 Å grid spacing and a Csp^3-probe atom with a charge of 1.0. The lattice size (Table 1) for the steroid example was automatically generated with 720 points. For the thermolysin series, a box of the size given in Table 3 with 2352 points was used. Energy truncation values of 30 kcal/mol were set for the steric and electrostatic interactions. Columns in the data table with a standard deviation ≤ 2 were dropped. The analyses were performed with a scaling according to CoMFA standard deviations. The relative contributions of steric and electrostatic fields are given in Tables 1 and 3.

Six orthogonal latent variables were first extracted by the standard PLS algorithm[13] and subsequently subjected to a cross-validation in the order of their correlation with the dependent variable. If the analysis indicated that more latent variables were required for an optimal description of the variance in the data set, additional PLS runs were performed considering a higher number of components. The "best" model was accepted as that which showed the sum of the squared differences between predicted and actual dependent property values (−lg K_I, Tables 2 and 4) to be a minimum from a leave-one-out cross-validation method (number of cross-validation groups equal to the number of compounds considered). Comparing these values to the overall variance of the actual dependent property values yielded cross-validated q^2s (Tables 1 and 3). The number of components considered corresponds to the lowest s_{press} obtained in each analysis. It has been checked whether the finally added component improves the cross-validated q^2 by more than 5% and reduces the s_{press} value. Results from the leave-one-out technique are given, since they are the only reproducible ones. Several cross-validations were performed for the steroid data set (SEAL alignment, Figure 2) leaving more than one object (selected by random) out of the analyses. The obtained q^2 values all fall into a range of ±0.05 around the value given in Table 1.

In addition to the predictive q^2s, the explanatory r^2 values (no cross-validation) are reported for the different analyses. In these calculations, the optimal number of components was regarded as revealed from the cross-validated analyses. For all examples, the standard errors are given in Tables 1 and 3. The experimentally determined binding constants (expressed as −lg K_I or pK_I) are listed in Tables 2 and 4 together with the residuals obtained in the different cross-validated CoMFA and CoMSIA analyses.

Table 2. Data Set of 21 Steroids Used in the Training Set To Obtain the Different 3D-QSAR Models Referred to in Table 1 and 10 Examples Used to Predict Binding Affinities[a]

	pK_I	CoMSIA (2)[b,d]	CoMFA (3)[c,d]	CoMSIA (4)[c,d]
1, aldrosterone	6.28	0.06	0.25	0.07
2, deoxycorticosterone	7.65	0.04	−0.12	−0.04
3, deoxycortisol	7.88	−0.04	0.07	0.08
4, dihydrotestosterone	5.92	−0.46	−0.32	−0.46
5, estradiol	5.00	0.04	0.05	−0.04
6, estriol	5.00	0.08	−0.12	0.20
7, estrone	5.00	0.08	−0.07	0.11
8, etiocholanolone	5.26	−0.37	−0.14	−0.36
9, pregnenolone	5.26	−0.40	−0.05	−0.32
10, 17-OH-pregnenlone	5.00	−0.50	−0.64	−0.59
11, progesterone	7.38	0.10	0.34	0.04
12, androstanediol	5.00	−0.12	0.02	−0.08
13, hydroxyprog	7.74	0.25	0.35	0.35
14, testosterone	6.72	0.47	0.29	0.36
15, androstenediol	5.00	0.25	0.17	0.21
16, androstendione	5.76	−0.44	−0.52	−0.43
17, androsterone	5.61	0.40	0.27	0.41
18, corticosterone	7.88	−0.01	−0.18	0.14
19, cortisol	7.88	0.15	0.06	0.16
20, cortisone	6.89	0.07	−0.05	−0.19
21, dehydroepiandrostrone	5.00	0.33	0.32	0.38
MEAN	6.15	0.00	0.00	0.00
ST_DEV	1.17	0.29	0.27	0.29
HIGH	7.88	0.47	0.35	0.41
LOW	5.00	−0.50	−0.64	−0.59
Predictive Data Set				
22, steroid 1	7.51	−0.79	−0.02	−0.11
23, steroid 2	7.55	−0.80	−0.34	−0.13
24, steroid 3	6.78	0.19	0.10	0.26
25, steroid 4	7.20	−0.78	−0.57	−0.55
26, steroid 5	6.14	0.25	0.30	0.30
27, steroid 6	6.25	−1.12	−0.80	−0.93
28, steroid 7	7.12	−0.25	−0.11	−0.23
29, steroid 8	6.82	−0.17	0.08	−0.08
30, steroid 9	7.69	−0.64	−0.10	0.03
31, steroid 10	5.80	−2.60	−2.17	−2.02
MEAN	6.89	−0.67	−0.36	−0.35
ST_DEV	0.65	0.82	0.71	0.69
HIGH	7.69	0.25	0.30	0.30
LOW	5.80	−2.60	−2.17	−2.02

[a] Chemical formulae are given in refs 4 and 14. The measured pK_I values are given together with the residuals remained for the different models; for the predictive set the differences to the measured values are listed. [b] Alignment according to ref 4. [c] Alignment present study with SEAL. [d] Number of analysis given in parentheses.

CoMSIA Analyses. Similar to the usual CoMFA approach, a data table has been constructed from similarity indices[3] calculated via a common probe atom which is placed at the intersections of a regularly spaced lattice. A grid spacing of 1 Å has been used throughout this study. Similarity indices $A_{F,k}$ between the compounds of interest and a probe atom, systematically placed at the intersections of the lattice, have been calculated according to (e.g., at grid point q for molecule j of the data set):

$$A_{F,k}{}^{q}(j) = -\sum_{i=1}^{n} w_{probe,k} w_{ik} e^{-\alpha r_{iq}{}^2}$$

where i = summation index over all atoms of the molecule j under investigation; w_{ik} = actual value of the physicochemical property k of atom i; $w_{probe,k}$ = probe

Table 3. Summary of the Results of the Different CoMFA and CoMSIA Analyses of the Thermolysin Data Set

	alignment according to ref 22[a]		alignment present study[a]			
	CoMFA (1)	CoMSIA (2)	CoMFA (3)	CoMFA (4)	CoMSIA (5)[b]	CoMSIA (6)
q^2	0.522	0.580	0.641	0.513	0.512	0.587
s_{press}	1.493	1.413	1.317	1.535	1.508	1.414
r^2	0.842	0.908	0.942	0.937	0.785	0.896
S	0.859	0.661	0.532	0.552	1.001	0.710
no. comp	5	6	7	7	5	7
fraction						
steric	0.672	0.320	0.672	0.636	0.540	0.296
electrostatic	0.328	0.252	0.328	0.364	0.460	0.263
hydrophobic		0.428				0.441
box						
stepsize (Å)	2	1	2	1	1	1
x	−9.4 to 17.4	−10 to 17	−8.6 to 16.4	−9 to 17	−9 to 17	−9 to 17
y	−16.5 to 10.0	−17 to 10	−16.9 to 9.9	−17 to 10	−17 to 10	−17 to 10
z	−12.0 to 10.5	−12 to 11	−9.9 to 12.6	−10 to 13	−10 to 13	−10 to 13

[a] Number of analysis given in parentheses. [b] Considering steric and electrostatic properties only.

atom with charge +1, radius 1 Å, and hydrophobicity +1; α = attenuation factor; and r_{iq} = mutual distance between probe atom at grid point q and atom i of the test molecule. Large values of α will result in a strong attenuation of the distance-dependent consideration of molecular similarity. Accordingly, there is little averaging of local feature matches of the molecules being compared. The global molecular similarity becomes less important. With small values of α also remote parts of each molecule will be experienced by the probe and the global molecular features become more important. In the present study α has been set to 0.3. With this selection, at a given lattice point the property value of an atom of the molecule under investigation (e.g., the partial atomic charge) is experienced in 1 Å distance by 74.1%, in 2 Å by 30.1%, and in 3 Å by 6.7% of its total value. This permits a reasonable "local smearing" of the molecular similarity indices and should help to avoid extreme dependencies on small changes of the mutual alignments.

In the present study, three physicochemical properties have been evaluated. The steric contribution has been reflected by the third power of the atomic radii of the atoms. Electrostatic properties have been introduced as atomic charges determined either by the semiempirical AM1 method[17] or Gasteiger–Marsili charges.[18] An atom-based hydrophobicity has been assigned according to the parametrization developed by Viswanadhan et al.[19] The lattice dimensions were selected with a sufficiently large margin (>4 Å) to enclose all aligned molecules. A box similar to that generated by CoMFA has been used for the steroids considering 7200 points (Table 1) for the thermolysin example with 18144 points (Table 3).

In order to extract a QSAR relationship from the high dimensional data table, the partial least-squares (PLS) method in SYBYL has been used.[16] The data handling between SYBYL and our own Fortran routines has been performed via SYBYL contour files. Once imported, they were transferred into field files and subsequently into QSAR tables within SYBYL.[20] The externally produced data tables were used in SYBYL applying "CoMFA standard scaling". Alternatively, for the thermolysin example an evaluation technique recently reported by Bush and Nachbar[21] has been applied. The described SAMPLS algorithm has been programmed by us and produces the results of a cross-validation in dramatically shorter CPU timings. The above-mentioned scaling and σ-cutoff has also been applied in SAMPLS. The statistical evaluation for the CoMSIA analyses were performed in the same way as described for CoMFA. The obtained statistical parameters are reported in Tables 1 and 3. The remaining residuals revealed for the different cross-validated analyses (leave-one-out method) are listed in Tables 2 and 4. As in CoMFA, cross-validations leaving randomly selected groups of compounds out of the analysis result in q^2 values ranging approximately ±0.05 around the value given for the leave-one-out technique (steroid data, SEAL alignment). To check the statistical significance of the models, random permutations of the dependent property variable were generated for the steroid data set. Using the permuted affinities, only nonpredictive models were obtained (both for CoMFA and CoMSIA).

Seven analyses were carried out considering the three fields separately or in all possible combinations. The obtained q^2 values (leave-one-out technique) for both data sets (SEAL alignment, Figures 2 and 4) are listed in Table 5. For the steroid example, all three fields are of comparable predictive power (similar results were obtained for CoMFA using steric and electrostatic fields separately). In the thermolysin case, the combination of all three fields reveals the best predictive q^2. Considering the three properties separately, the steric and electrostatic field are of less predictive power than the hydrophobic one. Since only the combination of all fields provides full insight into the spatial features of the different field contributions, the combined models have been used for further analyses.

Structural Alignment. Alignments of the data sets (Tables 2 and 4) evaluated in this study were either taken as described in literature[4,22] or re-aligned using the alignment condition in SEAL simultaneous with an angle optimization in torsion space. This approach has been described recently in detail.[1] For the steroid example, 10 additional molecules (Table 2, for chemical formulas see refs 4 and 12) described by Cramer et al.[4] were built from the coordinates of related structures in the Cambridge Crystallographic Database[23] and matched onto the 21 examples of the training set either by a least-squares atom-by-atom fit (atoms: 3, 5, 6, 13, 14, 17 of the steroid numbering scheme) or the SEAL condition. Predictions of the CBG receptor affinity based on the model obtained for the 21 compounds have been calculated (predicted values see Table 2). The realignment of the thermolysin ligands (Table 4) has been performed with respect to those reference examples that were structurally determined by X-ray crystallography (Table 4, name_crys). Compared to the alignment derived by DePriest et al.,[22] rms deviations between 0.3

Table 4. Data Set of 61 Thermolysin Inhibitors Used in the Training Set To Obtain the Different 3D-QSAR Models Referred to in Table 3 and 15 Examples Used To Predict Binding Affinities[a]

	pK_I	CoMFA (1)[b,d]	CoMSIA (2)[b,d]	CoMFA (3)[c,d]	CoMFA (4)[c,d]	CoMSIA (5)[c,d,e]	CoMSIA (6)[c,d]
1, CLTZNCRYS_TS	7.47	0.24	0.21	−0.10	−0.04	0.89	−0.21
2, ZGPLLZNCRYS_	8.04	0.88	0.95	0.86	0.88	1.69	0.60
3, ZFPLAZNCRYS_TS	10.17	2.13	0.96	0.65	0.83	2.45	1.23
4, ZGPOLLZNCRYS_TS	5.05	−2.15	−0.74	−0.63	−1.21	−0.95	−0.88
5, ZGPCLLZNCRYS_TS	6.74	−0.49	0.28	0.13	−0.19	−0.20	0.31
6, ZGPLA_TS	7.78	1.80	0.69	0.26	0.90	1.69	0.54
7, DAH53_TS	6.66	−0.28	1.27	−0.29	−0.79	−1.95	−0.81
8, DAH50_TS	7.96	−0.02	−0.07	−0.51	−0.77	−1.15	−1.07
9, BZSAG_TS	6.12	0.96	1.16	0.56	0.31	0.96	0.94
10, DAH51_TS	6.22	0.27	0.29	−0.21	−0.01	0.04	0.31
11, DAH52_TS	5.55	1.15	1.00	0.38	−0.44	1.60	1.48
12, DAH54_TS	5.77	0.28	−0.10	0.10	−0.08	0.02	0.24
13, DAH55_TS	2.42	−1.36	−1.68	−0.86	−0.58	−2.26	−1.41
14, ZGLY_TS	6.39	0.61	0.47	0.30	0.32	−0.13	1.02
15, ZALA_TS	6.07	0.39	0.31	0.74	0.45	−0.92	0.49
16, C6PLTNME_TS	8.82	0.72	0.56	0.93	1.65	1.93	1.01
17, NHOHBZMAGNA_TS	6.37	0.67	−0.04	−0.16	0.12	−0.05	−0.41
18, ZGPOLA_TS	4.89	−0.86	−0.55	−1.13	−0.82	−0.82	−1.00
19, ZAPOLA_TS	5.74	−0.29	0.10	0.07	−0.01	−0.03	−0.06
20, NHOHLEU_TS	3.72	−0.11	−0.57	0.25	−0.47	1.37	−0.07
21, ZGGLNHOH_TS	4.41	0.17	0.11	0.21	0.18	−0.24	0.38
22, ZGG_D_LNHOH_TS	3.60	−0.50	−0.22	0.26	−0.14	−0.68	0.15
23, ZG_D_LNHOH_TS	4.32	0.68	0.02	−0.22	−0.37	−0.29	0.22
24, ZGLNHOH_TS	4.89	0.13	−0.06	0.35	0.20	0.28	0.79
25, NHOHBZMAGNH2_TS	6.18	0.70	0.90	0.61	0.60	0.70	0.57
26, NHOHBZMOET_TS	4.70	−0.72	−0.29	−0.53	−0.60	0.02	−0.25
27, PHOSPHORAMIDON_TS	7.55	0.11	−0.34	−0.10	−0.10	1.07	−0.49
28, CBZPHE_TS	3.29	1.15	0.72	0.38	0.15	0.74	0.69
29, C6PLCTNME_TS	7.28	−0.56	0.49	0.14	0.20	−0.30	0.11
30, C6POLTNME_TS	5.84	−1.77	−1.02	−0.50	−0.48	−0.77	−0.65
31, S_THIORPHAN_TS	5.74	−0.78	0.02	0.63	0.01	0.57	0.47
32, R_THIORPHAN_TS	5.64	−0.60	−0.37	0.14	−0.24	0.68	−0.04
33, ZLPOLA_TS	6.17	−0.48	−0.54	0.04	−0.19	−0.68	−0.38
34, Z_D_FPOLA_TS	4.52	−0.60	0.37	0.20	0.04	−0.40	0.31
35, ZFPOLA_TS	7.35	−0.65	−0.46	−0.48	−0.64	0.07	−0.65
36, Z_D_LPOLA	4.38	−0.35	0.23	0.28	0.46	−0.42	0.45
37, Z_D_APOLA_TS	4.62	−0.57	−0.45	−1.15	−0.23	−0.46	−0.46
38, Z_D_FPLA_TS	6.32	0.68	0.56	0.82	0.68	0.75	0.48
39, ZGLNH2_TS	1.68	−1.14	−0.21	−0.85	−0.62	0.49	−0.35
40, ZGLNMEOH_TS	2.65	−0.62	−1.11	−0.19	−0.03	0.29	−0.62
41, Z_NH_GLNHOH_TS	5.57	0.40	−0.70	0.37	0.04	0.47	0.56
42, ZGGNHOH_TS	3.03	0.06	0.72	0.37	0.32	0.37	0.24
43, Z_NH_GLNH2_TS	3.42	0.36	1.32	−0.62	−0.79	−0.99	−0.34
44, NHOHBZMAGOH_TS	6.18	−0.46	−0.47	0.03	−0.05	−0.42	−0.67
45, NHOHIBMAGNH2_TS	6.32	1.07	0.99	0.10	0.81	1.17	0.77
46, NHOHMALAGNH2_TS	2.96	−0.53	0.18	−0.22	−0.50	−0.38	−0.37
47, OHBZMAGNH2_TS	3.38	−1.17	−0.21	−0.43	−0.58	−1.43	−0.01
48, HOCH2CO_FAGNH2_TS	2.54	−0.31	−0.42	0.05	0.00	−0.94	−0.69
49, CH3COCH2CO_FAGNH2_TS	2.51	−0.06	0.50	0.37	0.49	0.03	0.87
50, PO3_FAGNH2_TS	5.59	0.83	−0.08	0.04	0.33	0.84	0.93
51, SO2P_FAGNH2_TS	5.16	0.58	−0.35	−0.59	0.01	0.30	−0.09
52, SO3_FAGNH2_TS	2.37	−1.10	−0.82	−1.08	−0.61	−2.05	−1.92
53, CH3O2S_FAGNH2_TS	0.52	−0.33	−0.13	−0.39	−0.01	−0.10	−0.37
54, PLEUNH2_TS	4.10	0.37	−0.46	0.38	0.29	−0.24	−0.64
55, PNHET_TS	0.52	−1.18	−0.52	0.53	0.38	−0.61	0.06
56, P_OPHE_OME_LEUNH2_TS	0.52	0.28	0.33	0.52	0.26	−1.27	−0.29
52, PAAOH_TS	4.06	0.11	0.03	0.09	0.46	0.30	0.25
58, PPHEOH_TS	4.14	0.13	0.50	−0.12	−0.42	−1.14	−0.28
59, P_ILE_AOH_TS	6.44	1.06	0.35	−0.62	0.02	0.37	−0.23
60, CHO_OHLEU_AGNH2_TS	2.47	0.17	−0.78	−0.33	0.05	−0.14	−0.68
61, ACE_OHLEU_AGNH2_TS	2.47	0.88	−0.31	0.22	0.59	0.26	−0.10
MEAN	4.97	0.00	0.00	0.00	0.00	0.00	0.00
STD_DEV	2.07	0.82	0.63	0.50	0.52	0.96	0.67
HIGH	10.17	2.13	1.32	0.93	1.65	2.45	1.48
LOW	0.52	−2.15	−1.68	−1.15	−1.21	−2.26	−1.92
Predictive Data Set							
1, ZGPLG_TS (17)	6.57			0.00	−0.58	0.99	−0.50
2, ZGPLF_TS (18)	7.12			−0.33	−0.74	0.21	−0.94
3, ZGPLNH2_TS (19)	6.12			2.28	1.53	2.12	1.18
4, ZGPOLG_TS (21)	3.64			−2.16	−2.80	−1.36	−1.99
5, ZGPOLF_TS (22)	4.27			−2.61	−2.96	−2.19	−2.49
6, ZGPOLNH2_TS (23)	3.18			−0.49	−1.36	−0.35	−0.73
7, ZGPCLA_TS (24)	7.73			0.29	0.10	0.98	0.98
8, ZGPCLG_TS (25)	6.52			−0.35	−0.69	0.18	0.05
9, ZGPCLF_TS (26)	7.18			−0.16	−0.84	−0.33	−0.19
10, ZGPCLNH2_TS (27)	5.85			1.87	0.78	1.41	1.48
11, PPPHE_TS	2.79			0.53	−1.42	−1.54	−0.66

Table 4 (Continued)

	pK_I	CoMFA (1)[b,d]	CoMSIA (2)[b,d]	CoMFA (3)[c,d]	CoMFA (4)[c,d]	CoMSIA (5)[c,d,e]	CoMSIA (6)[c,d]
12, PLFOH_TS	7.72			1.09	1.83	0.53	0.53
13, ZFGNH2_TS	3.46			1.99	1.76	1.27	1.36
14, ZLGNH2_TS	2.51			1.84	2.14	0.88	1.15
15, ZYGNH2_TS	3.66			2.21	1.80	1.51	1.64
MEAN	5.22			0.40	−0.10	0.29	0.06
STD_DEV	1.91			1.51	1.68	1.24	1.28
HIGH	7.73			2.28	2.14	2.12	1.64
LOW	2.51			−2.61	−2.96	−2.19	−2.49

[a] Amino acids in the usual one letter notation, additional abbreviations as given in ref 22 and 24; for those structures in the predictive data set already used in ref 6, the corresponding numbering scheme of the compounds is given. The measured pK_I values are given together with the residuals remained for the different models; for the predictive set the differences to the measured values are listed. [b] Alignment according to ref 22. [c] Alignment present study with SEAL. [d] Number of analysis given in parentheses. [e] Steric and electrostatic properties only.

Table 5. Results from Different PLS Analyses for the Steroid and Thermolysin Data Sets Based on the SEAL Alignment (Figures 2 and 4)[a]

	CoMSIA fields						
	electrostatic +steric +hydrophobic	steric	electrostatic	hydrophobic	electrostatic +steric	electrostatic +hydrophobic	steric +hydrophobic
				Steroids			
q^2	0.665	0.623	0.526	0.733	0.515	0.683	0.749
S_{press}	0.759	0.781	0.903	0.647	0.914	0.739	0.620
no. comp	4	3	4	3	4	4	4
				Thermolysin Inhibitors			
q^2	0.587	0.482	0.277	0.542	0.512	0.527	0.563
S_{press}	1.414	1.555	1.853	1.476	1.508	1.472	1.442
no. comp	7	4	6	6	5	4	6

[a] The similarity index fields have been considered separately or in all possible combinations. The obtained q^2 values (leave-one-out technique), s_{press}, and the number of considered components are given.

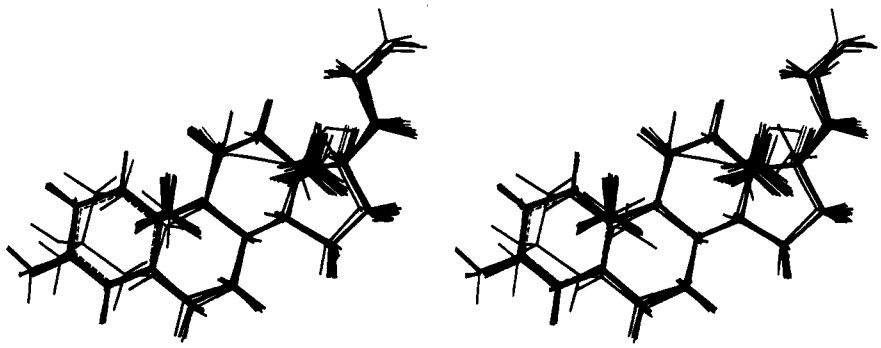

Figure 1. Stereodiagram of the structural superposition of 21 steroids, given by Cramer et al.[4] according to a least-squares atom-by-atom fit. In this alignment the structures were used as training set to perform CoMFA and CoMSIA analyses.

and 1.8 Å were obtained. The most pronounced deviations between both alignments are observed in those areas where the ligands extend into the solvent exposed area. In all cases, it has been checked whether the assumed binding geometry is in agreement with the pattern of the functional groups of protein residues oriented toward the binding site (Figure 10). Similar to the steroid example, 15 additional ligands (Table 4) have been aligned with the training set according to the SEAL alignment and predictions of their binding affinities have been calculated. These structures have been taken from a data set used in previous studies.[6,22,24] Chemical formulas are given therein. In the steroid examples charges were calculated by the AM1 method.[17] For the thermolysin inhibitors, Gasteiger–Marsili charges[18] have been used as produced by SYBYL.[16] As suggested by DePriest et al.,[22] the carboxylate oxygen atoms occasionally present in the C-termini of some inhibitors were assigned atom type O.CO2; if two oxygen atoms were present in the group next to Zn, the atom directly coordinating to Zn has been assigned O.3, the other O.2. A rough idea about the differences in the two alignments can be obtained from Figures 1–4.[25]

QSAR Coefficient Contour Maps. The visualization of the results of the comparative analyses in terms of field contributions has been performed by means of computer graphics enclosing the volumes above and below particular field values by isocontours (isopleths).[4,12] We have used the field type "STDEV * COEFF" to obtain contours from a CoMFA analysis that elucidate the relationship between differences in the fields and variations in the dependent variable. An appropriate selection of the contour levels has been performed in an iterative manner using histograms of field levels as indicator. Most information about the results is contained in the distribution of actual field values and the percentage of field contributions. The first evaluation indicates in which spatial areas significant variations among the molecules occur that correlate with the dependent property variable. The percentage of field contributions above and below particular levels allows one to find those lattice points where the product of the associated QSAR coefficient and the standard

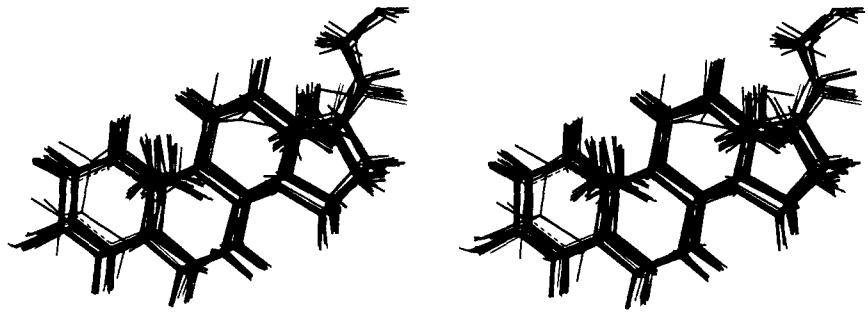

Figure 2. Stereodiagram of the structural superposition of 21 steroids, obtained by an alignment using the condition implemented into SEAL together with an optimization in torsion angle space.[1] In this alternative alignment the structures were used as training set to perform CoMFA and CoMSIA analyses.

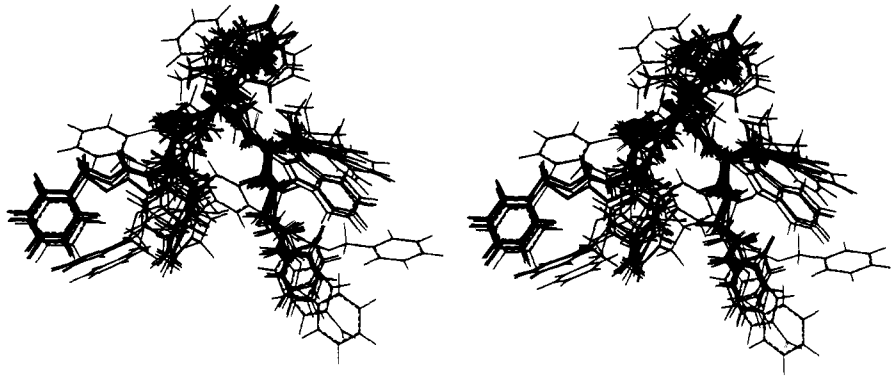

Figure 3. Stereodiagram of the structural superposition of 61 inhibitors of thermolysin, obtained according to the alignment rule defined by DePriest et al.[22] In this alignment the structures were used as training set to perform CoMFA and CoMSIA analyses.

Figure 4. Stereodiagram of the structural superposition of 61 inhibitors of thermolysin, obtained by an alignment using the condition implemented into SEAL together with an optimization in torsion angle space.[1] In this alignment the structures were used as training set to perform CoMFA and CoMSIA analyses.

deviation of all values in the corresponding column of the data table are above or below a particular value.[4] This indicates the areas where appropriate alterations of the field values will enhance or reduce the value of the dependent property variable. In favorable cases, a fairly small amount of points is responsible for the major part of these correlations. Accordingly, the corresponding maps are easy to interpret. The maps obtained in the various analyses are shown in Figures 7, 8, and 11–18. A gray scaling has been applied to separate "enhancing" and "reducing" contributions.

Differences between CoMFA and CoMSIA Fields. From a technical point of view, the fields used in a CoMFA or CoMSIA analysis are of equivalent form. As a consequence, the subsequent PLS analyses are identically performed. In both cases, numerical data tables are evaluated that contain quantities calculated at regularly spaced grid points using some distance-dependent functional form. However, the rationale behind the fields is different.

In CoMFA, for each molecule in the data set interaction energies with respect to a probe atom are calculated according to a particular potential function. Usually, Lennard-Jones and Coulomb potentials are used (also other potentials have been applied successfully[9]) that taper off hyperbolically in space, however with a quite different slope. In Figure 5 the Lennard-Jones and Coulomb potentials are shown as calculated at 171 points (steps of 0.1 Å) along an axis of a lattice encompassing the molecule under consideration (here: benzoic acid). The lattice expands along the x-axis from −4 to 13 Å, and the molecule falls into the range between 1 and 8 Å. The axis under consideration runs through one of the carboxylate oxygens and is parallel to the C2–C3 bond of the phenyl ring (Figure 5). Along the y-axis the potential energy is plotted in kilocalories per mole. Owing to the hyperbolic functional form, grid points close to or inside the molecule obtain very large values. To avoid these unacceptable large quantities, the potentials are set in these regions to arbitrarily fixed

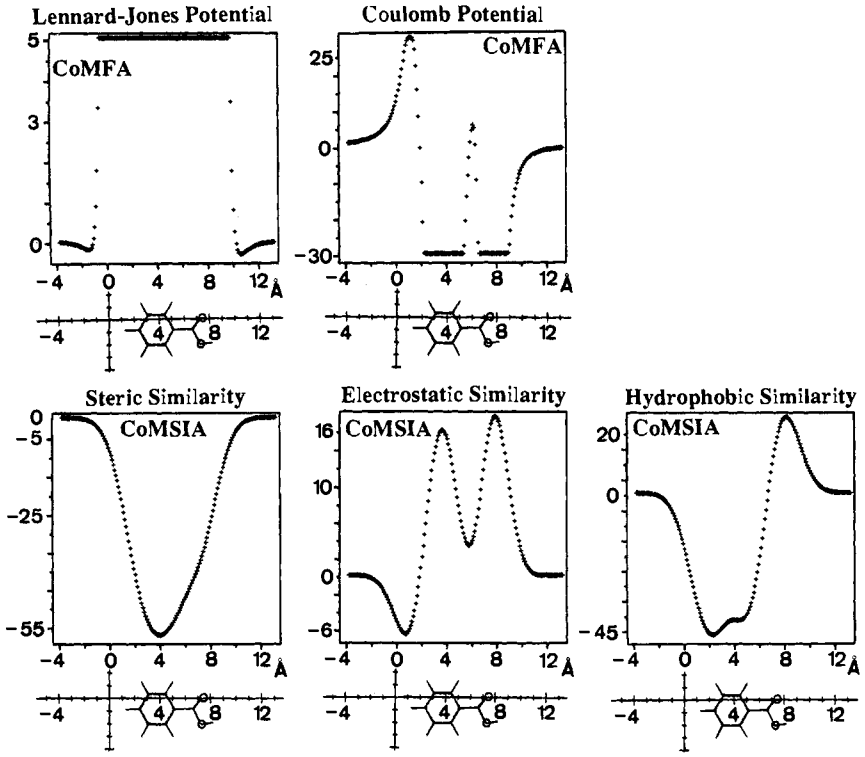

Figure 5. Change of the Lennard-Jones and Coulomb potential (upper row, kcal/mol) and the steric, electrostatic, and hydrophobic similarity indices (lower row, arbitrary units) calculated for benzoic acid along an axis passing through the molecule as indicated below each diagram. The molecule extends from about 1 to 8 Å, the grid values have been calculated between −4 and 12 Å in steps of 0.1 Å.

cutoff values (here: 5 kcal/mol for the Lennard-Jones and ±30 kcal/mol for the Coulomb potential). The cutoff limits are exceeded for the two potentials at different distances from the molecule (see Figure 5, upper row).

In CoMSIA, the similarity indices compiled for the different molecules in the data set at the intersections of a regularly spaced lattice obtain a different meaning. To compare a set of molecules in space, for each of the molecules its similarity with a common probe is determined in a distance-dependent fashion. The probe atom scans the entire lattice embedding each of the molecules. Thus, lattice points inside and outside the molecules are used, and no arbitrary cutoffs are required. In Figure 5 (lower row), the changes of the similarity indices for steric, electrostatic, and hydrophobic properties are plotted along the axis described above using the same grid spacing. The scaling along the y-axis is in arbitrary units measuring the similarity $A_{F,k}$ between the probe and the molecule according to the functional form given above. For steric contributions, A_F obtains only negative values by definition. Approximately, in the same range where the Lennard-Jones potential increases (at about −1.5 and 10 Å), the similarity indices change from zero toward increasingly negative values. The Lennard-Jones potential exceeds at −0.8 and 9.5 Å the cutoff value of 5 kcal/mol. Numerically, the same absolute value (−5 units) is exceeded for the steric similarity indices at −0.7 and 9.6 Å. The latter function reveals an extremum at 3.7 Å. This value falls close to the center of mass along this axis. Compared to the Lennard-Jones potential, the selected Gaussian-type functional form results in a smoother change of the descriptor of the steric properties. For electrostatic and hydrophobic features, positive or negative similarity indices are calculated. The electrostatic similarity obtains close to the meta and para hydrogens negative, toward the meta (partial charge −0.15 e) and ortho carbons (−0.07 e) and the carboxylate group positive values. Similarly, the values for hydrophobic indices change sign moving from the hydrophobic phenyl moiety toward the hydrophilic carboxylate group.

CoMFA and CoMSIA Analyses of Steroids. The two alignments shown in Figures 1 and 2 have been submitted to CoMFA and CoMSIA analyses. As a reference, the CoMFA results published by Cramer et al.[4] for the atom-by-atom alignment are used. All four analyses reveal comparable cross-validated q^2 values (Table 1). Only the CoMFA analysis based on the SEAL alignment (no. 3) reveals a slightly smaller q^2. The test data set of 10 compounds used by Cramer et al.[4] to predict activities of examples not included to the training set have been aligned as described above (Table 2). Predictions were performed using the CoMFA and CoMSIA models. The results are shown in Figure 6. For comparison purposes, the predictions obtained by Cramer et al.[4] are also given. The two analyses based on the SEAL alignment reveal more convincing predictions considering the reduced data scatter around the diagonal. The "predictive" r^2 as used by Cramer et al.[4] and Waller and Marshall[26] amounts for the analyses based on the SEAL alignment to 0.36 (Table 2, no. 3) and 0.40 (no. 4, Figure 6, center and right). Taking the values in Table III, part B (standard parameter setting), from the study of Cramer et al.[4] yields a "predictive" r^2 of 0.30 (Figure 6, left, omitting the values for compounds **1**, **9**, **10** we obtain in accordance with ref 4 a "predictive" $r^2 = 0.81$). The only pronounced outlier is a 9-fluoro-substituted steroid (steroid **10**). Since the data set did not contain a compound with a non-hydrogen atom in the 9-position, reliable predictions cannot be expected by the models. The present type of correlation analysis cannot extrapolate into areas where no data have initially been included in the training set.

Besides a quantitative estimate of the dependent

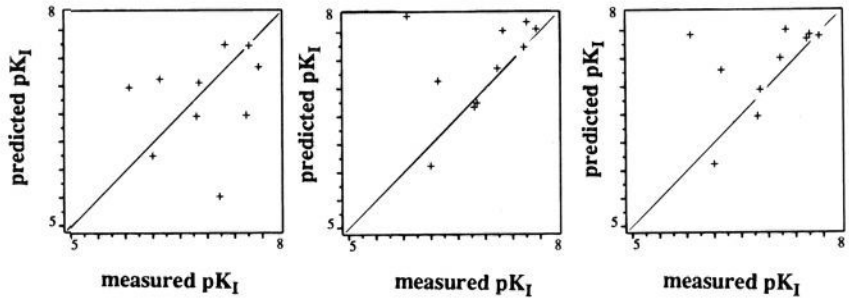

Figure 6. *Predicted vs measured pK_I values for the set of 10 different steroids as obtained from the three different analyses:* (1) CoMFA analysis based on the least-squares atom-by-atom alignment (values taken from ref 4 Table III, part B, standard parameter setting), "predictive" $r^2 = 0.30$, (2) CoMFA analysis based on the SEAL alignment (Table 2, no. 3), "predictive" $r^2 = 0.36$, (3) CoMSIA analysis based on the SEAL alignment (no. 4), "predictive" $r^2 = 0.40$.

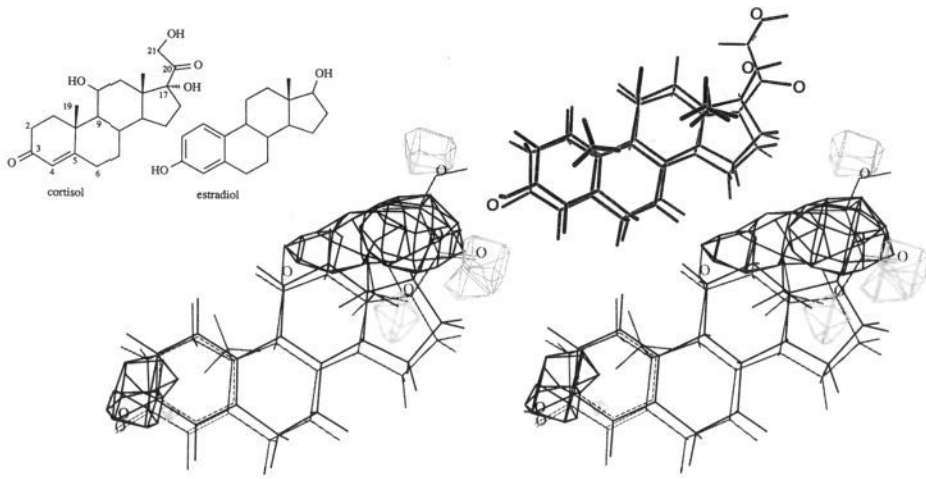

Figure 7. Stereoview of the contours at 0.009 and −0.008 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the electrostatic properties obtained for the set of steroids; the gray isopleths enclose areas where an increase of negative charge will enhance the affinity, and black isopleths encapsulate regions where a more positively charged group will improve the binding properties. Superimposed onto the map are cortisol and estradiol exhibiting high and low CBG receptor affinity, respectively. For clarity, in the center the mutual alignment of the two steroids is given (contour plots performed with SYBYL,[16] molecular superposition represented by SHADEMOL[36]).

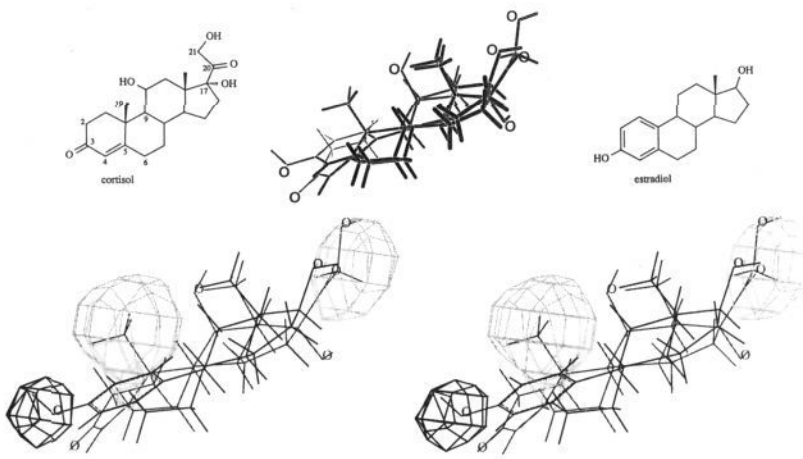

Figure 8. Stereoview of the contours at ±0.001 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the steric properties obtained for the set of steroids; the gray isopleths enclose areas where any occupation with sterically demanding groups will enhance the activity, and black isopleths encompass regions where steric bulk should be reduced. Superimposed onto the map are cortisol and estradiol exhibiting high and low CBG receptor affinity, respectively. For clarity, in the center the mutual alignment of the two steroids is given.

property variable, the analysis of the field contributions with respect to the different properties allows one to elucidate those features that are responsible for the differences in biological activity. Cramer et al.[4] discussed in their paper the steric and electrostatic CoMFA field contributions. For a direct comparison, the corresponding CoMSIA field contributions are shown in Figures 7 and 8. Superimposed onto the maps are cortisol and estradiol as examples for high and low CBG-receptor affinity in the alignment obtained by SEAL.

At first, the contributions of the electrostatic properties should be inspected. Figure 7 shows two regions (black isopleths) where additional positive charge will enhance activity: around the side chain at C17 and

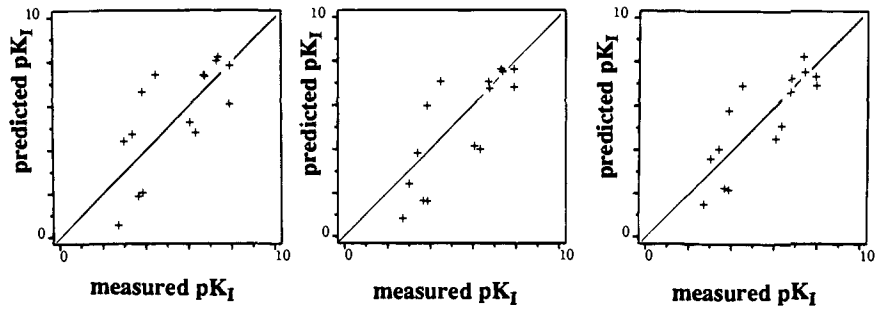

Figure 9. Predicted vs measured pK_I values for the set of 15 different thermolysin inhibitors as obtained from the three different analyses (Table 3) using the SEAL alignment: (1) CoMFA analysis with grid spacing 1 Å (no. 4), "predictive" $r^2 = 0.24$, (2) CoMFA analysis with grid spacing 2 Å (no. 3), "predictive" $r^2 = 0.34$, (3) CoMSIA analysis (no. 6), "predictive" $r^2 = 0.56$.

about the C2–C3 bond. Extension of the side chain, e.g., from 17-hydroxy or 17-carbonyl to 17-(C=O)CH_2-OH, increases activity. The additional positive charge in this area is mainly produced by the carbonyl carbon and the methylene group. The isopleth close to C2 and C3 indicates that hydrogenation at these positions produces a positive charge that is advantageous for activity. The three lobes (gray isopleth) where negative charge results in higher affinity coincides with the oxygen atoms at C17, C20, and C21. Compounds bearing oxygen at these positions achieve better binding to the CBG receptor.

The map of steric properties (Figure 8) shows an extended region spanning from the "19-methyl position" to the 5,6-bond in the steroid skeleton. In this region any occupancy with sterically demanding groups enhances activity (gray isopleth). Compounds with low affinity all lack the 19-methyl group or show a double bond between C5 and C6. Hydrogenation of the latter bond introduces additional steric bulk in this region. Furthermore, an area close to the 17-position is highlighted by a gray isopleth. Steric bulk introduced by a side chain at this position increases biological activity. A region to be avoided by sterically demanding groups is indicated close to the 3-position. Due to unsaturation of the A ring, the low-affinity estradiol-type steroids occupy a volume section slightly "above" the cortisol skeleton. The isopleth contoured in black indicates that steric bulk produced by, for example, the estradiol-type steroids will reduce affinity.

The map of hydrophobic properties (no figure) indicates significant correlations between changes in the hydrophobicity and biological activity only in the area around the A ring. Increasing hydrophobicity in this area produces improved receptor affinity. A closer inspection of the structures in the training set discovers that the A-ring replacement from a phenolic or cyclohexanolic to a cyclohexenonic moiety increases activity. Apparently, this coincides with the trend indicated by the maps of hydrophobic properties.

CoMFA and CoMSIA Analyses of Thermolysin Inhibitors. Several CoMFA and CoMSIA analyses using the two alignments in Figures 3 and 4 have been accomplished. In the original study, DePriest et al.[22] reported a cross-validated $q^2 = 0.70$ and a conventional $r^2 = 0.98$ using 11 PCs. Obviously, only 30 cross-validation groups for the data set of 61 compounds have been completed. In a second study reported by Waller and Marshall,[26] the same data set has been analyzed (using PM3 charges) with 61 cross-validation groups. This analysis revealed a $q^2 = 0.536$ and $r^2 = 0.851$. We reperformed the conventional CoMFA analysis and

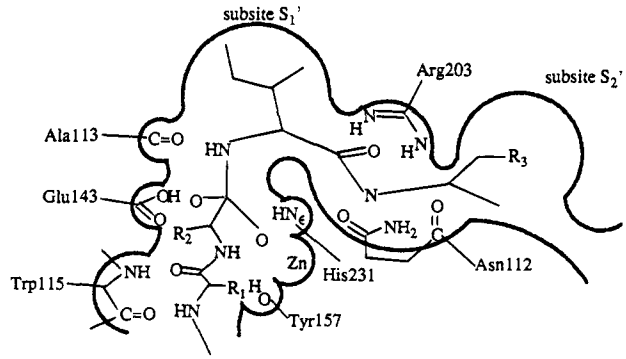

Figure 10. Schematic sketch of the assumed interactions between thermolysin and a peptide substrate. The formal orientation of ligand and protein binding site corresponds to those in Figures 11–18.

obtained a $q^2 = 0.522$ ($r^2 = 0.842$, Table 3, no. 1) using Gasteiger–Marsili charges and a box defined according to DePriest et al.[22] Referring to the same alignment, the CoMSIA approach yielded a better $q^2 = 0.580$ (non-cross-validated $r^2 = 0.908$, Table 3, no. 2). With respect to the second alignment, derived from the SEAL-alignment condition, higher cross-validated q^2 values have been obtained. A CoMFA analysis with 2 Å grid spacing converged to $q^2 = 0.641$ ($r^2 = 0.942$, no. 3), whereas the 1 Å grid spacing produced a lower $q^2 = 0.513$ ($r^2 = 0.937$, Table 3, no. 4). This grid-spacing dependency of q^2 has already been pointed out by Folkers et al.[11] Our CoMSIA result ($q^2 = 0.587$, $r^2 = 0.896$, no. 6) based on a 1 Å grid spacing reveals a q^2 between those of the CoMFA analyses with respect to a 1 and 2 Å lattice.

The ligand affinities to thermolysin (expressed as pK_I) are given in Table 4 together with the residuals obtained for the different models. A data set of 15 inhibitors not included in the training set of 61 compounds has been aligned according to the SEAL condition. The subsequent prediction reveals better agreements with the CoMSIA models. The results are shown in Figure 9. "Predictive" r^2 values of 0.24 (no. 4), 0.34 (no. 3), and 0.56 (no. 6) were obtained for the three models considered.

An impressive amount of crystal data has been collected on ligand/protein complexes of the endopeptidase thermolysin mainly due to studies of the group of Matthews.[27] As a consequence, detailed knowledge about the structural requirements for ligand binding is available. The following binding mode is assumed for the natural peptide cleaved by the proteinase (Figure 10). Four subsites S_2, S_1, S_1', and S_2' are characterized by a pattern of specific hydrogen bonds and hydrophobic pockets of limited size.[27] The S_2 and S_1 subsite provide

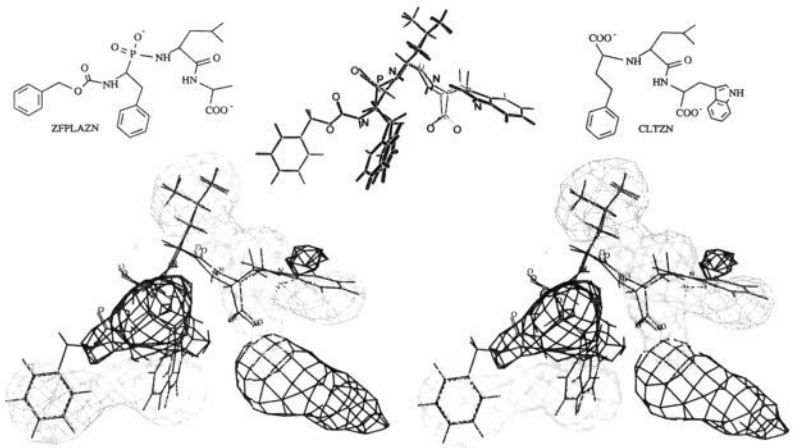

Figure 11. Stereoview of the contours at 0.003 and −0.004 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the steric properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where any occupation with sterically demanding groups will enhance the affinity. Black isopleths encapsulate regions where steric bulk should be reduced. Superimposed onto the map are the two tight-binding inhibitors CLTZN and ZFPLAZN. For clarity, the mutual alignment of the two inhibitors is given in the center. The two compounds differ in the occupancy with bulky groups in the "lower left" region of the diagram.

hydrogen bonding sites through the backbone C=O and NH groups of Trp115. At the cleavage site a zinc atom is found. In addition to this metal atom, the enzyme orients the OH group of Tyr157, the proton at N_ϵ of His231, and the carboxylate group of Glu143 toward this site. Next to this center, the backbone carbonyl oxygen of Ala113 is found that serves as a hydrogen bond acceptor. In addition, the side chain of Asn112 orients its carbonyl oxygen toward this site. The S_1' subsite consists of a hydrophobic pocket that can accommodate ligand or substrate side chains such as Val, Leu, or Phe.[28] Significantly larger groups (e.g., Trp) cannot be placed into this site. The substrate's peptide groups connecting the residues that occupy the S_1' and S_2' subsites find matching groups in Arg203 and Asn112. The S_2' subsite that readily accommodates larger groups such as an indole moiety of tryptophan can also accept a hydrogen bond of the inhibitor through the backbone carbonyl of Asn112.[29] The terminal NH_2 of the side chain of Asn112 is involved in H-bonding to the following peptide group of the substrate.

It is interesting to see how much of the described spatial requirements can be discovered in the results of the 3D-QSAR analysis. To perform this comparison, stereodiagrams of the steric, electrostatic, and hydrophobic field contributions expressed by the product of the QSAR coefficients and the standard deviation have been contoured. These diagrams are oriented similar to the sketch of the binding pocket shown in Figure 10. In their study, DePriest et al.[22] performed a similar analysis and found clear correlations of such features. In addition to features directly related to interactions in the binding pocket, regions along the aligned ligands exposed to the solvent are indicated that correlate significantly with the binding affinity.

In Figure 11, the contours of the steric properties are shown. Spatial occupation of regions that increase activity are contoured by gray isopleths, regions to be left vacant by black isopleths. Occupation of the S_1' subsite is known to be a prerequisite for tight binding.[28] Accordingly, the contour map of steric properties underlines this region. Furthermore, steric occupancy of the S_2' subsite is indicated as favorable for activity. For example, the two inhibitors DAH50 and DAH53 (Figure 12) differ by an indole and phenyl moiety oriented toward this site. The inhibitor with the smaller phenyl group (Figure 12, slightly rotated compared to Figure 11 and contours in front clipped) does not fully occupy this area. Hence, its affinity is reduced by 1.3 logarithmic units.

The orientation of sterically demanding substituents into an area "above" the group coordinated to zinc is unfavorable for binding (black isopleth). On the contrary, occupancy of the neighboring region by sterically crowded substituents increases activity. With respect to the enzyme structure, this region is located at the surface of the protein, thus interfering with the solvent/protein interface. The two tight-binding inhibitors CLTZN and ZFPLAZN both orient substituents into the outlined region (Figure 11). Both display high affinities (pK_I = 7.47/10.17). However, the latter inhibitor occupies this region by either a carbobenzoxyl and a benzyl group, whereas the former, weaker-binding compound only orients its benzyl moiety into this area. This latter "benzyl binding site" should be closer inspected. The three inhibitors ZLPOLA, ZFPOLA, and Z-(*d*)-FPOLA only differ by the occupancy of the "right" part of this subsite (Figure 13, only a section through this subsite is shown for clarity). ZFPOLA is the strongest inhibitor (pK_I = 7.35). It occupies this area most satisfactorily by its benzyl group. The sterically less demanding leucine side chain in ZLPOLA (pK_I = 6.17) fills less space of the sterically favorable region. Changing the stereochemistry at C_α from *l*- to *d*-phenylalanine (ZFPOLA → Z-(*d*)-FPOLA) reveals only partial occupancy of the sterically favorable region, in contrast, the sterically less favorable region (black isopleth) is now partially used. The latter inhibitor is characterized by a substantially reduced activity (pK_I = 4.52).

The contour maps of steric properties indicate reduced affinity for inhibitors that extend at the C-terminus far beyond the S_2' subsite (carbonyl group hydrogen bonded to Asn112, Figures 10 and 11). A comparison with the protein structure shows that the indicated sterically unfavorable region extends into the solvent exposed area. Most likely, inhibitors that occupy this region (e.g., several hydroxamic acid derivatives that all show

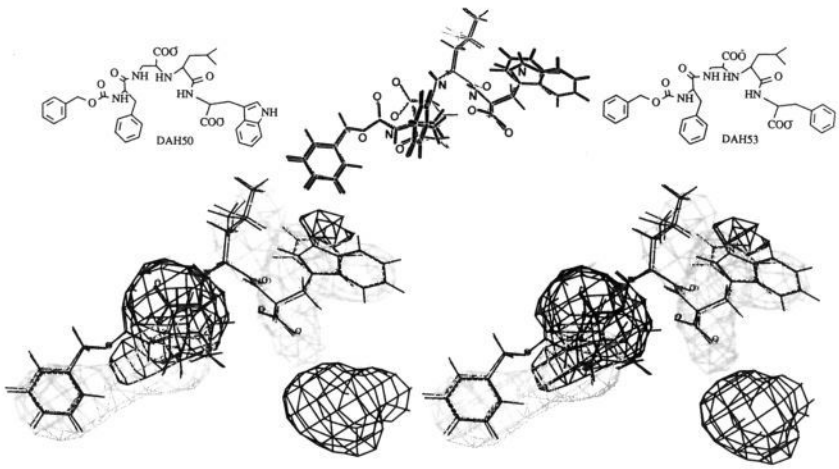

Figure 12. Stereoview of the contours at 0.003 and −0.004 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the steric properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where any occupation with sterically demanding groups will enhance the affinity, and black isopleths encapsulate regions which should be avoided. The view in this diagram is slightly rotated compared to that in Figure 11 in order to show the differences between the two inhibitors DAH50 and DAH53 that are superimposed onto the map. For clarity, their mutual alignment is given in the center. Both compounds differ in the "filling" of the sterically favorable region corresponding to the S_2' subsite (cf. Figure 10).

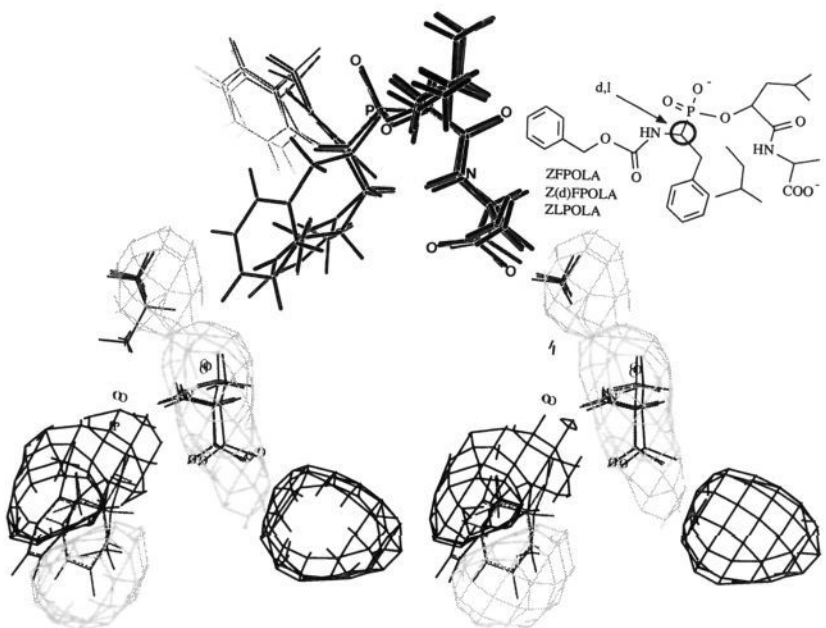

Figure 13. Stereoview of the contours at 0.003 and −0.004 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the steric properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where any occupation with sterically demanding groups will enhance the affinity, and black isopleths encapsulate regions which should be avoided. Superimposed onto the map are three inhibitors ZFPOLA, ZLPOLA, and Z-(*d*)-FPOLA that possess different R_2 substituents (see Figure 10). Their mutual alignment is also shown in the center of the diagram. For clarity only a section through the contours in this particular area is given. ZFPOLA and ZLPOLA orient their phenyl or isopropyl groups, respectively, into the area predicted as favorable (gray) for sterically demanding groups. Due to the inverted stereochemistry in Z-(*d*)-FPOLA, its phenyl substituent hardly dips into the favorable region from "above"; however, a substantial part of this phenyl ring remains in the area indicated as sterically unfavorable (black contoured area).

reduced activity) interfere unfavorably with water molecules at the solvent/protein interface.

Analyzing the contours of electrostatic properties reveals two regions where increasing negative charge (gray isopleths) enhances activity and one region where a more positive charge improves affinity (Figure 14). One of the regions favorable for increasing negative charge can be ascribed to the group coordinating to zinc. For example, in all hydroxamic acids less charge is found in this region. As obvious from the training set, in general, the latter inhibitors are less active than comparable phospho derivatives. The second region, indicating negative charge to be advantageous, is located around the functional group potentially involved in a hydrogen bond toward Asn112. In many inhibitors this group corresponds to a terminal COO^- group, apart from a slight size difference between a Leu and Ile side chain (Figure 14). The former compound leaves this region unoccupied, accordingly its affinity is reduced by more than two pK_I units (4.10/6.44). An interesting series of isostructural inhibitors are ZGPLLZN, ZGPOLLZN, and ZGPCLLZN (Figure 15). Crystallographic studies[30] revealed closely similar binding geometry for the phosphonamidate, phosphonate, and phosphinate. Phosphonamidate and phosphinate possess comparable affinities, the phosphonate is by a

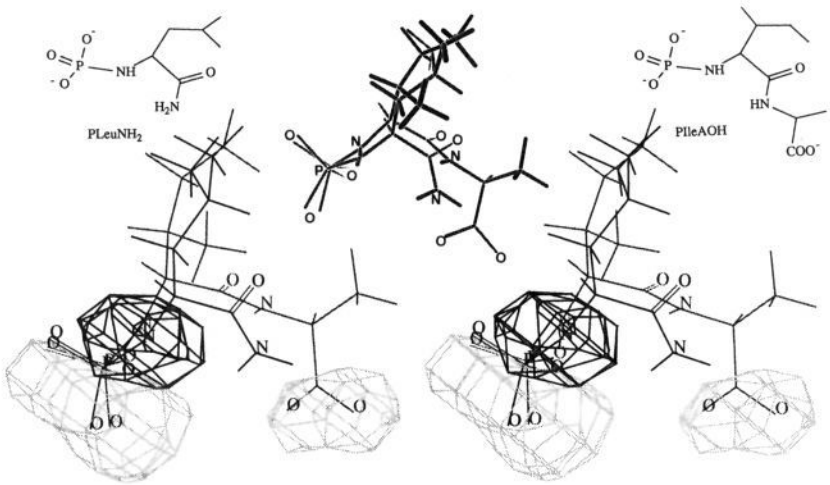

Figure 14. Stereoview of the contours at ±0.02 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the electrostatic properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where an increase of negative charge will enhance the affinity, and black isopleths encapsulate region where a more positively charged group will improve the binding properties. Superimposed onto the map are the two inhibitors $PLeuNH_2$ and PIle-AOH. For clarity, in the center the mutual alignment of the two inhibitors is given. The latter tighter binding inhibitor orients its carboxylate group into the area indicated by the analysis to be favorable for negatively charged groups.

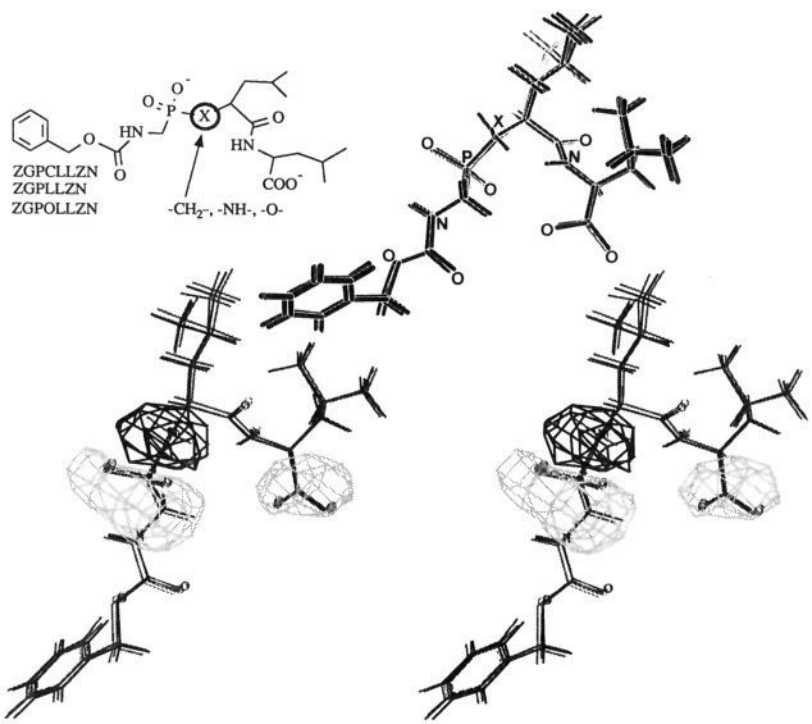

Figure 15. Stereoview of the contours at ±0.02 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the electrostatic properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where an increase of negative charge will enhance the affinity, and black isopleths encapsulate regions where a more positively charged group will improve the binding properties. Superimposed onto the map are the three isostructural inhibitors ZGPLLZN, ZGPOLLZN, and ZGPCLLZN. For clarity, the mutual alignment of the three inhibitors is given in the center. Structural differences (CH_2/NH/O) only occur next to the PO_2 group. This position falls into the center of a contoured volume in which an increasing positive charge should enhance affinity.

factor of 1000 less active[31,32] than the amidate. In the latter, the NH group forms a hydrogen bond to the carbonyl oxygen of Ala113 (Figure 10). Since a suitable hydrogen-bond donating group is missing in the analog phosphonate and phosphinate at this position, these two inhibitors are not able to form an equivalent hydrogen bond with the protein. In solution, most likely, all polar groups in the three inhibitors are involved in hydrogen bonding to solvent molecules, in particular the NH and O of the phosphonamidate and phosphonate. The methylene group of the phosphinate cannot perform a comparable interaction. A simple comparison of the hydrogen bonding inventory in solution and in the protein reveals a compensated situation for phosphonamidate and phosphinate, whereas the phosphonate loses its hydrogen-bonded environment about the O of the phosphate group. This uncompensated situation for the phosphonate leads to a lower binding constant corresponding to a 4.1 kcal/mol deficit in intrinsic binding constant.[31] Since the binding constants of phosphonamidate and phosphinate are rather similar, it can be concluded that the PO_2NH···O=C

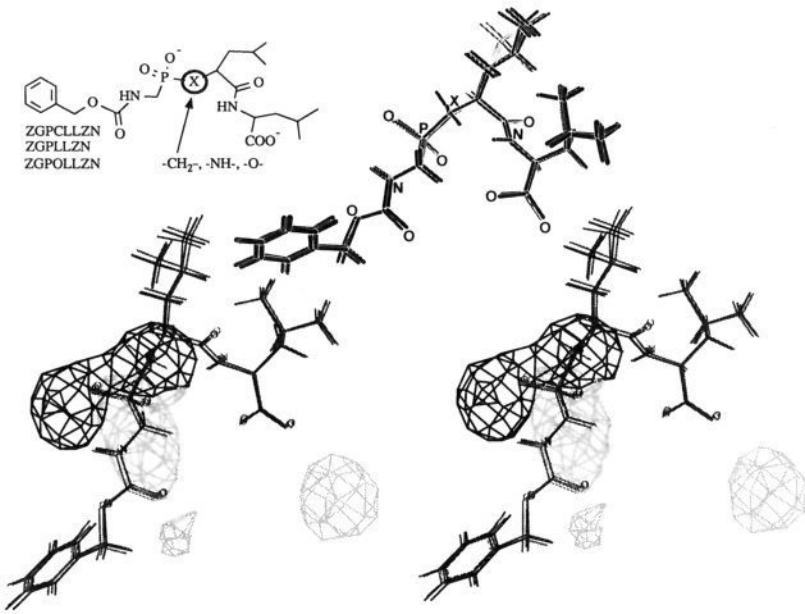

Figure 16. Stereoview of the contours at 0.017 and −0.010 "kcal/mol"[35] of the CoMSIA coefficient * standard deviation for the hydrophobic properties obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where a decrease of hydrophobicity will enhance the affinity. Black isopleths encapsulate regions where a more hydrophobic group will improve the binding properties. Superimposed onto the map are three isostructural inhibitors ZGPLLZN, ZGPOLLZN, and ZGPCLLZN (mutual alignment as in Figure 15). Structural differences (CH_2/NH/O) only occur next to the PO_2 group. This position falls into a subspace of a region (black contour) for which an increasing hydrophobicity should enhance affinity.

interaction between phosphonamidate and Ala113 is energetically of the same order as the interaction of the PO_2NH group with the surrounding water molecules in solution. A similar trend in binding affinity is observed between the two isostructural inhibitors C6PLCTNME and C6POLTNME that differ by a comparable CH_2/O exchange (pK_I = 7.28/5.84).

How is this O/NH/CH_2 replacement expressed in the contour maps? The replacement of O by NH and CH_2 clearly increases positive charge in this region. According to the contour plot of electrostatic properties, a more positive charge in this region should increase affinity (Figure 15). Considering the contour map of hydrophobic properties (Figure 16), further evidence for the observed trend in the series is found. Increasing hydrophobicity (black isopleth corresponding to growing hydrophobicity) is required in exactly the same region which has already been indicated by the electrostatic properties to favor the presence of a more positive charge (compare Figures 15 and 16). This trend also explains the higher affinity of the phosphinate especially over the phosphonate. Thus, a nonessential hydrogen bond to the receptor is expressed by this, on a first glance surprising correlation.

The neighboring area, indicated in the contour map of hydrophobic properties as less favorable for hydrophobic groups (Figure 16, gray isopleths), coincides with the atoms coordinating to zinc. Groups of a more hydrophilic character are important for tight ligand binding. An additional region favorable for hydrophilic groups extends toward the solvent-exposed area. This region is mainly occupied by those inhibitors in the training set that further extend at their C-termini such as the hydroxamic acid inhibitors. However, as already indicated by the steric properties, any extension of the inhibitors into this region reduces biological activity.

For comparison purposes, the corresponding maps derived from the CoMFA analysis based on the SEAL alignment are shown in Figures 17 and 18. In Figure 17, the steric (Lennard-Jones) field contributions are shown in the same orientation as in Figure 11. Superimposed onto the map are the two inhibitors CLTZN and ZFPLAZN. Regions where increasing steric-field contributions will enhance activity are contoured by gray isopleths, sterically disadvantageous areas are shown in black. A direct comparison with the maps obtained by CoMSIA for the steric properties vaguely allows one to detect similar features, e.g., the sterically favorable S_1' subsite and the region around the carbobenzoxyl and benzyl moiety in ZFPLAZN or the unfavorable area beyond the S_2' subsite.

The electrostatic (Coulombic) field contributions based on the CoMFA analysis are given in Figure 18. Superimposed onto this map are the two inhibitors $PLeuNH_2$ and PIle-AOH in the same orientation as in Figure 14. Similar to the previous case, corresponding features can be extrapolated from the map. For example, the gray isopleth adjacent to the terminal carboxylate group permits a rough guess that increasing charge in this area will enhance affinity. The contouring around phosphorus is rather fuzzy and accordingly difficult to interpret. However, the importance of this area has clearly been indicated by the corresponding CoMSIA maps.

From the field-contribution maps it is evident that the analysis based on spatial similarity indices reveals a much more significant guide to trace the features that really matter especially with respect to the design of novel compounds.

Discussion and Conclusions

In the present paper, an alternative approach is reported to compute property fields based on similarity indices of drug molecules that have been brought into a common alignment. The functional forms to calculate fields of different physicochemical properties all possess

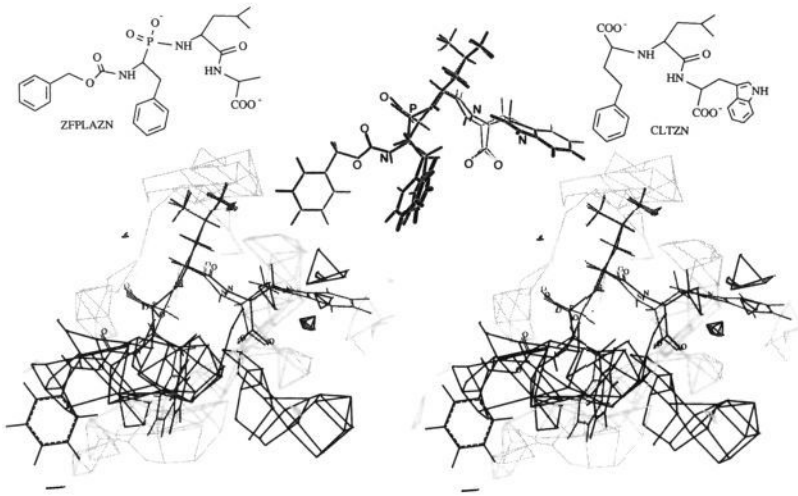

Figure 17. Stereoview of the contours at 0.05 and −0.03 kcal/mol of the CoMFA coefficient * standard deviation for the steric (Lennard-Jones) field obtained for the set of thermolysin inhibitors; the gray isopleths encompass areas where increasing steric interactions will enhance the affinity. Black isopleths indicate regions where contributions to the steric field should be reduced. The view in this figure is equivalent to that in Figure 11. Superimposed onto the map are the two tight-binding inhibitors CLTZN and ZFPLAZN.

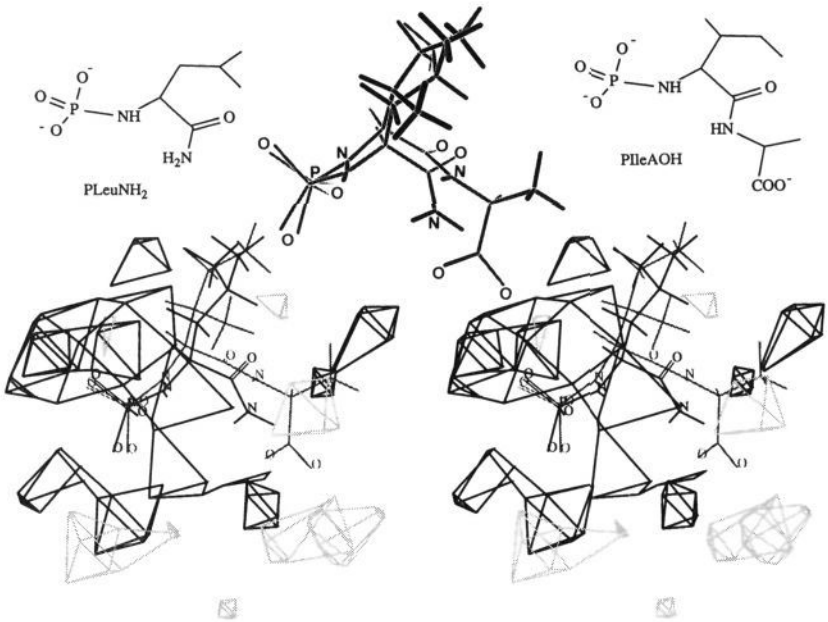

Figure 18. Stereoview of the contours at ±0.035 kcal/mol of the CoMFA coefficient * standard deviation for the electrostatic (Coulomb) field obtained for the set of thermolysin inhibitors; the gray isopleths enclose areas where an increasing coulombic interaction with a negative charge will enhance the affinity, and black isopleths encapsulate a region where interactions with a more positively charged group will improve the binding properties. The view in this figure is equivalent to that in Figure 14. Superimposed onto the map are the two inhibitors $PLeuNH_2$ and PIle-AOH.

the same distance dependence, and no singularities occur at the atomic positions. Accordingly, no arbitrary definitions of cutoff limits and deficiencies due to different slopes of the fields are encountered. Recently, we reported on an alignment method capable of approximately reproducing alignments that are crystallographically observed at the binding site of proteins.[1] The functional form used in this alignment condition and in the above-mentioned field calculations is equivalent. Two data sets of steroids and thermolysin ligands were analyzed in terms of the usual CoMFA method (Lennard-Jones and Coulomb potentials) and the alternative CoMSIA approach using two different alignments. One of them has been computed with the above-mentioned alignment condition. An alignment technique commonly applied in congeneric series of rather rigid compounds, such as steroids, tries to superimpose the molecules by a least-squares atom-by-atom match (Figure 1). This approach has been applied to the steroid data set. Using the alignment condition that seeks to superimpose molecular portions according to their similar physicochemical properties in space reveals an alignment with bonding skeletons less sharply superimposed (Figure 2).

Models with comparable statistical significance are derived for both alignments applying CoMFA and CoMSIA. However, the SEAL alignment is superior in the prediction of receptor affinities of novel compounds. This alignment with the more "fuzzy" superposition of the steroid skeletons performs slightly worse in CoMFA. This observation is in agreement with a study of Horwitz et al.[33] who superimposed pyrazoloacridines

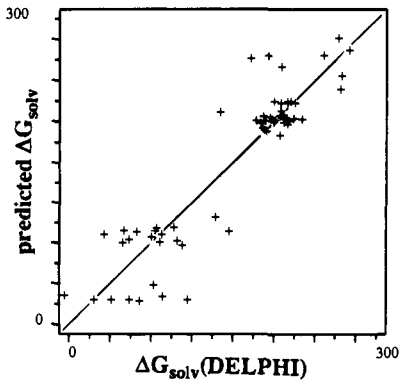

Figure 19. Fitted predictions from the PLS analysis using a field of hydrophobic similarity indices *vs* the ΔG_{solv} values computed by the DELPHI method[34] of 61 thermolysin inhibitors. The model was derived using 6 principle components having a $q^2 = 0.365$, $r^2 = 0.842$.

with SEAL[3] and subsequently subjected the aligned structures to CoMFA. The best results were obtained with an alignment using a strong distance-dependent attenuation (large α). This dictates a superposition where global similarity is of minor importance. It approaches a direct atom-by-atom matching.

The thermolysin data set is structurally much more diverse, and the alignment condition has to consider the intermolecular recognition features with the surrounding protein. The alignment based on the SEAL condition performs slightly better in the statistical analyses than one obtained from docking and minimizing the structures within the protein binding pocket. Differences between both alignments mainly occur in regions where the ligands extend beyond the protein into solvent-exposed areas. In these areas the SEAL approach imposes additional constraints that accomplish an alignment better suited for a molecular comparison (cf. Figures 3 and 4). Predictions of inhibitors not included into the training set perform better by CoMSIA than with the conventional CoMFA.

In the Introduction, the assumption has been put forward that hydrophobic fields are introduced into the present CoMSIA approach in order to cover, at least partially, the contribution of the free solvation enthalpy to the binding constant. A CoMSIA analysis omitting the hydrophobic properties reveals a slightly worse correlation (Table 3, analysis no. 5). Waller and Marshall[26] tried to include computed ΔG_{solv} values into a CoMFA. Slight improvements have been observed. To assess the above-mentioned assumption, the ΔG_{solv} values computed by Waller and Marshall for the 61 thermolysin ligands with DELPHI[34] have been subjected to a PLS evaluation using the field of hydrophobic similarity indices only. A $q^2 = 0.365$ (six components, $s_{press} = 65.10$) and a conventional $r^2 = 0.842$ ($s = 32.38$) have been obtained. The cross-validated q^2 indicates a correlation at the edge of significance. The predictions based on the derived model are shown in Figure 19. It has to be considered that the ΔG_{solv} values used in this correlation are not experimentally observed but computed quantities. Despite being at present the best method to determine solvation energies, the DELPHI method is approximate. The same holds for the parametrization applied for the calculation of hydrophobic properties in CoMSIA. Furthermore, it could well be that hydrophobic fields only comprise, besides additional effects, the ΔG_{solv} contributions to the binding constant. These different aspects might be responsible for the reduced performance of the correlation.

The contribution maps obtained by the CoMSIA approach are by far superior to the usual CoMFA maps. This is mainly so because no cutoffs are required and the fields are of equivalent and "smoother" distance-dependent functional form. Contribution maps derived from CoMFA only denote regions apart from the molecules (cutoff!) where interactions with a putative environment are to be expected, whereas the maps obtained by CoMSIA allow one to recognize those regions within the area occupied by the ligand skeletons that require a particular physicochemical property important for activity. This is a more direct guide toward new ideas about the design of novel compounds.

There is much room to improve the presently used fields of similarity indices. Three physicochemical properties have been considered so far. The approach can be easily extended to additional properties using a common distance dependence. The only prerequisite is a reliable assignment of atom-based parameters. Possibly additional similarity indices can consider further contributions to the binding constant that are still unreflected in the presently used fields. However, even if these property fields would not enhance the correlation they still can support the design process significantly. Contouring the important regions in space that matter with respect to a particular property suggests to the modeler where to modify the known structures in order to develop compounds with higher affinity. Including additional fields into the approach increases the computational requirements enormously. However, due to the dramatic speed-up of the SAMPLS method,[21] this expansion appears to be feasible.

Acknowledgment. The authors are grateful to Hugo Kubinyi (BASF, Ludwigshafen) for helpful discussions, especially during the implementation of the SAMPLS method.

References

(1) Klebe, G.; Mietzner, T.; Weber, F. Different Approaches Toward an Automatic Alignment of Drug Molecules: Applications to Sterol Mimics, Thrombin and Thermolysin Inhibitors. *J. Comput.-Aided Mol. Design*, in press.
(2) Klebe, G.; Mietzner, T. A Fast and Efficient Method to Generate Biologically Relevant Conformations. *J. Comput.-Aided Mol. Design* **1994**, *8*, 583–606.
(3) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Method.* **1990**, *3*, 615–633.
(4) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(5) Pace, C. N. Contribution of the Hydrophobic Effect to Globular Protein Stability. *J. Mol. Biol.* **1992**, *226*, 29–35. Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models. *Biochemistry* 1991, *30*, 9686–9697.
(6) Klebe, G.; Abraham, U. On the Prediction of Binding Properties of Drug Molecules by Comparative Molecular Field Analysis. *J. Med. Chem.* **1993**, *36*, 70–80.
(7) Kellogg, G. E.; Abraham, D. J. KEY, LOCK, and LOCKSMITH: Complementary Hydrophathic Map Predictions of Drug Structure from a Known Receptor-Receptor Structure from Known Drugs. *J. Mol. Graph.* **1992**, *10*, 212–217.
(8) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(9) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular Lipophilicity Potential, a tools in 3D QSAR: Method and Applications. *J. Comput.-Aided Mol. Design* **1994**, *8*, 83–96. Davis, A. M.; Gensmantel, N. P.; Johansson, E.; Marriott, D. P.

The Use of the GRID Program in the 3D QSAR Analysis of a Series of Calcium-Channel Agonists. *J. Med. Chem.* **1994**, *37*, 963–972. Cocchi, M.; Johansson, E. Amino Acids Characterization of GRID and Multivariate Data Analysis. *Quant. Struct.-Act. Relat.* **1993**, *12*, 1–8. Kim, K. H.; Martin, Y. C. Direct Prediction of Linear Free Energy Substituent Effects from 3D Structures using Comparative Molecular Field Analysis. I. Electronic Effects of Substituted Benzoid Acids. *J. Org. Chem.* **1991**, *56*, 2723–2729. Kim, K. H.; Martin, Y. C. Direct Prediction of Dissociation Constants (pK_a) of Clonidine-like Imidazoles, 2-substituted Imidazoles and 1-Methyl-2-substituted-imidazoles from 3D Structures using a Comparative Molecular Field Analysis (CoMFA). *J. Med. Chem.* **1991**, *34*, 2056–2060. Kim, K. H. 3D-Quantitative Structure Activity Relationships: Description of Electronic Effects directly from 3D Structures using a GRID-Comparative Molecular Field Analysis (CoMFA) Approach. *Quant. Struct.-Act. Relat.* **1992**, *11*, 127–134. Kim, K. H. 3D-Quantitative Structure Activity Relationships: Nonlinear Dependence Described directly from 3D Structures using a GRID-Comparative Molecular Field Analysis (CoMFA) Approach. *Quant. Struct.-Act. Relat.* **1992**, *11*, 453–460.

(10) Klebe, G. Structural Alignment of Molecules. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 173–199.

(11) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 583–618.

(12) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 443–485.

(13) Stahle, L.; Wold, S. Multivariate Data Analysis and Experimental Design in Biomedical Research. *Progr. Med. Chem.* **1988**, *25*, 292–334.

(14) Good, A. C.; So, S.-S.; Richards, W. G. Structure-Activity Relationships from Molecular Similarity Matrices. *J. Med. Chem.* **1993**, *36*, 433–438.

(15) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR's form Similarity Matrices. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. *J. Med. Chem.* **1993**, *36*, 2929–2937.

(16) SYBYL Molecular Modeling System (Version 5.40), Tripos Ass., St. Louis, MO.

(17) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(18) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(19) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.

(20) The following steps have been performed: (1) define a CoMFA region with a particular grid spacing (e.g., 1 Å), (2) produce a QSAR data table containing the molecules of the training set, (3) generate with an external FORTRAN program for each property the similarity indices at all grid points and store them in the SYBYL contour file format (FORTRAN interfaces XC-Create and XXClose, see Toolkit Manual), (4) import contour files into SYBYL and transfer them into field files "COMFA FIELD CONTOUR IMPORT region_name field_out contour_file", (5) produce QSAR columns for the field files (option "TAILOR COMFA FIELD TYPE" either to "ELECTROSTATIC_ONLY" or "STERIC_ONLY") and fill the columns by the following two commands: "TABLE ENTER cell row_expr yes field_file", "TABLE EVALUATE new_only row_expr col_expr", (6) save table and run PLS.

(21) Bush, B. L.; Nachbar, R. B., Jr. Sample-distance Partial Least-Squares: PLS optimized for many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Design* **1993**, *7*, 587–619.

(22) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372–5384.

(23) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink-Peters, T.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystallographic Data Centre: Computer-based Search, Retrieval, Analysis and Display of Information. *Acta Crystallogr. Sect. B* **1979**, *B35*, 2331–2339.

(24) Klopman, G.; Bendale, R. D. Computed Automated Structure Evaluation (CASE): A Study of Inhibition of the Thermolysin Enzyme. *J. Theor. Biol.* **1989**, *136*, 67–77.

(25) The coordinates of the steroid and thermolysin data sets revealed by the SEAL alignment can be obtained from the authors upon request in electronic form in SYBYL-Molfile format.

(26) Waller, C. L.; Marshall, G. R. Three-Dimensional Quantitative Structure-Activity Relationship of Angiotensin-Converting Enzyme and Thermolysin Inhibitors. II. A Comparison of CoMFA Models Incorporating Molecular Orbital Fields and Desolvation Free Energies Based on Active-Analog and Complementary-Receptor-Field Alignment Rules. *J. Med. Chem.* **1993**, *36*, 2390–2403.

(27) Matthews, B. W. Structural Basis of the Action of Thermolysin and Related Zinc Peptidases. *Acc. Chem. Res.* **1988**, *21*, 333–340.

(28) Tronrud, D. E.; Monzingo, A. F.; Matthews, B. W. Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin. *Eur. J. Biochem.* **1986**, *157*, 261–268.

(29) Holden, H. M.; Matthews, B. W. The Binding of 1-Valyl-1-tryptophan to Crystalline Thermolysin Illustrates the Mode of Interaction of a Product of Peptide Hydrolysis. *J. Biol. Chem.* **1988**, *263*, 3256–3260.

(30) Tronrud, D. E.; Holden, H. M.; Matthews, B. W. Structures of Two Thermolysin-Inhibitor Complexes That Differ by a Single Hydrogen Bond. *Science* **1987**, *235*, 571–574.

(31) Bartlett, P. A.; Marlowe, C. K. Evaluation of Intrinsic Binding Energy from a Hydrogen Bonding group in an Enzyme Inhibitor. *Science* **1987**, *235*, 569–571.

(32) Morgan, B. P.; Scholtz, J. M.; Ballinger, M. D.; Zipkin, I. D.; Bartlett, P. A. Differential Binding Energy: A Detailed Evaluation of the Influence of Hydrogen-Bonding and Hydrophobic Groups on the Inhibition of Thermolysin by Phosphorus-Containing Inhibitors. *J. Am. Chem. Soc.* **1991**, *113*, 297–307.

(33) Horwitz, J. P.; Massova, I.; Wiese, T. E.; Wozniak, A. J.; Corbett, T. H.; Sebolt-Leopold, J.; Capps, D. B.; Leopold, W. R. Comparative Molecular Field Analysis of in Vitro Growth Inhibition of L1210 and HCT-8 Cells by Some Pyrazoloacridines. *J. Med. Chem.* **1993**, *36*, 3511–3516.

(34) DELPHI program, available from Biosym Technologies, Inc., 9685 Scranton Road, San Diego, CA 92121-2777.

(35) As mentioned, the CoMSIA similarity index fields are defined in arbitrary units; nevertheless, they are formally treated within SYBYL equivalent to potential energy fields. Accordingly the contour levels are given in kilocalories per moles. However, the functional form and parametrization of these similarity index fields does not correspond to any potential describing different partitions of molecular interactions.

(36) Hahn, M.; Wipke, W. T. SHADEMOL: An Algorithm for Presentation of Three-dimensional Structures on a Laser Printer Using Depth-Shading. *Tetrahedron Comput. Method. 1* **1988**, 81–86.