Journal of Computer-Aided Molecular Design, **13:** 1–10, 1999.
KLUWER/ESCOM
© 1999 *Kluwer Academic Publishers. Printed in the Netherlands.*

Comparative Molecular Similarity Index Analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries

Gerhard Klebe[a,*] & Ute Abraham[b]
[a]*Institute of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, D-35032 Marburg, Germany;*
[b]*BASF AG, Main Laboratory, Carl-Bosch Strasse, D-67056 Ludwigshafen, Germany*

Received 15 November 1997; Accepted 3 April 1998

Key words: comparative molecular field analysis, contribution map interpretation, molecular similarity, thermolysin inhibition, 3D QSAR

Summary

Comparative molecular field analysis has been applied to a data set of thermolysin inhibitors. Fields expressed in terms of molecular similarity indices (CoMSIA) have been used instead of the usually applied Lennard-Jones- and Coulomb-type potentials (CoMFA). Five different properties, assumed to cover the major contributions responsible for ligand binding, have been considered: steric, electrostatic, hydrophobic, and hydrogen-bond donor or acceptor properties. The statistical evaluation of the field properties by PLS analysis reveals a similar predictive potential to CoMFA. However, significantly improved and easily interpretable contour maps are obtained. The features in these maps intuitively suggest where to modify a molecular structure in terms of physicochemical properties and functional groups in order to improve its binding affinity. They can also be interpreted with respect to the known structural protein environment of thermolysin. Most of the highlighted regions in the maps are mirrored by features in the surrounding environment required for binding. Using the derived correlation model, different members of a combinatorial library designed for thermolysin inhibition have been scored for affinity. The results obtained demonstrate the prediction power of the CoMSIA method.

Introduction

The drug discovery and lead optimization process is currently dominated by developments in two fields: a ‘rational design’ based on structural information and sophisticated computer methods to elucidate the structural prerequisites important for binding to a particular target, and a ‘random screening’ using high-throughput screening technologies to discover possible leads from large compound libraries provided increasingly by combinatorial chemistry [1]. The two approaches are complementary: Structural characteristics about a particular series of compounds, e.g. hits from random screening, can be used to establish a structure-activity relationship. The derived model helps to explain the important relative differences within a compound series, suggests how to improve their binding properties and assists in ranking and selecting novel candidates for synthesis. This latter step is also of great value in the specific design of compound libraries.

This strategy requires a powerful tool to analyze and compare molecules in terms of either their similarity or diversity. Over the last eight years the CoMFA method [2] has been established as such a tool in 3D QSAR [3]. It maps gradual changes of the interaction properties of molecules by evaluating the potential energy at regularly spaced grid points surrounding the mutually aligned molecules of a data set. Recently, we have extended this method to a comparative analysis of molecular similarity [4]. This approach avoids some of the inherent deficiencies arising from the functional form of the Lennard–Jones and Coulomb potentials used in the original version of CoMFA. Both potentials are very steep close to the van der Waals surface, and as a consequence, the potential energy

[*] To whom correspondence should be addressed.

expressed at grid points in the proximity of the surface changes dramatically. However, it is precisely this region that contains important information in a QSAR analysis [5]. Furthermore, they produce singularities at the atomic centers. To avoid unacceptably large values, the potential evaluations are normally restricted to regions outside the molecules, and some arbitrarily determined cut-off values are defined. Due to differences in the slope of the Lennard–Jones and Coulomb potentials, these cut-off values are exceeded at different distances from the molecules [5]. This requires further arbitrary scaling of the two fields in a simultaneous evaluation and can involve the loss of information about one of the fields.

To overcome such problems, we calculate molecular similarity indices 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. The values obtained in form of a field do not exhibit a direct measure of similarity 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 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. For this distance dependence, a Gaussian-type functional form is selected that avoids singularities at the atomic positions. No arbitrary definition of cut-off limits is required and the indices can be calculated at all grid points. The obtained ensemble of indices is evaluated in a PLS analysis [6] according to the standard CoMFA protocol.

Often enough comparative molecular field analyses are performed to derive a correlation model of predictive power. Novel compounds, designed for synthesis, are compared with the obtained model and their expected activity is predicted. This step evaluates a newly designed compound, however it does not directly support its design. To achieve the latter aspect the graphical interpretation of CoMFA results is extremely important. Contour maps of the relative spatial contributions (or better importance) of the different fields are usually used [7]. However, due to the described cut-off settings and the steepness of the potentials close to the molecular surfaces, these maps are often not continuously and smoothly connected and accordingly they are difficult to interpret. Using molecular similarity indices, substantially improved contour maps are obtained that are easy to interpret and very intuitive as a visualization tool in actively supporting the design of novel compounds. Whereas the level-dependent contouring of CoMFA-field contributions highlights those regions in space where the aligned molecules would favorably or unfavorably interact with a possible environment, the CoMSIA-field contributions 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.

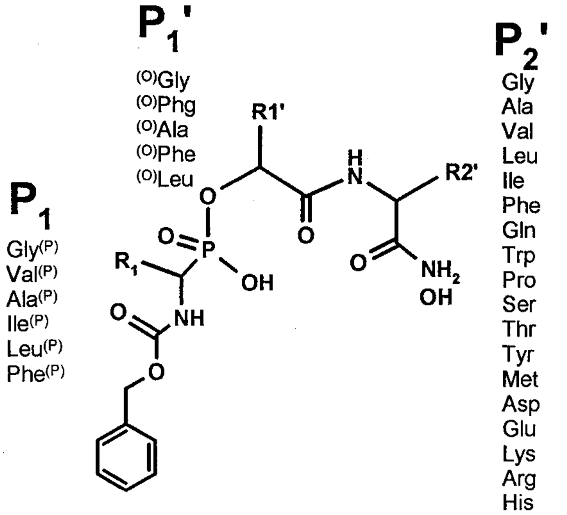

Figure 1. Combinatorial library of different peptidylphosphonates Cbz-X^{P}-^{O}Y-Z-resin published by Campbell et al. [9]. Different residues were considered at the three positions P$_1$, X^{P} = six α-aminophosphonic acids, P$'_1$ (^{O}Y = five α-hydroxy acids), and P$'_2$ (Z = 18 natural amino acids, excluding Cys and Asn).

In the present paper we describe the introduction of hydrogen-bonding properties into the similarity analysis. Putative hydrogen-bonding sites about functional groups of the molecules are generated. In the comparative analysis they are enumerated similarly to the other properties. If the presence of a particular hydrogen-bonding site about the molecules correlates with changes in the binding affinities, this site will occur as a particular feature in the final contour maps. This again is a very intuitive indicator of where to alter the molecules under consideration.

The newly introduced hydrogen-bond properties are evaluated for a data set of thermolysin inhibitors already used in previous studies [4, 8]. Since a crystal structure of the protein is available a relevant structural alignment of the molecules can be defined straight forwardly. It should be emphasized that knowledge of the protein structure as a reference is not a prerequisite to perform the present analysis. However, it provides an opportunity to interpret features indicated in the contour maps with respect to the protein environment. In order to estimate the predictive power of such 3D QSAR models for the design of combinatorial libraries, the binding properties of a published and tested peptide library [9] inhibiting thermolysin is evaluated.

Computational methods

Calculation of the hydrogen-bonding site

In order to consider hydrogen-bonding properties, a strategy is followed similar to that in the program DISCO [10, 11]. The various functional groups are classified, e.g. as carboxylate, amide, alcohol or amine. To define putative hydrogen-bonding sites about these groups, the results from mapping composite crystal-field environments in small molecule crystal structures found in the Cambridge Crystallographic Database [12] have been evaluated [13]. Representative spatial positions for H-bond donors and acceptors are generated at the centers of these distributions by placing pseudoatoms, e.g. at four such centers on a spherical cone about a carboxylate oxygen with two in and one above and below a plane through the functional group (distance 1.9 Å). For bifunctional groups capable of operating as donor and acceptor, e.g. an alcohol group, acceptor and donor pseudoatoms are generated at these centers. Any such position is discarded from the list that clashes with or falls closer than 1.8 Å to an atom of the probe molecules. In the subsequent step of this approach, each position of a H-bond donor or acceptor is represented by a Gaussian function. The width of these functions (see below) has been selected in a way that a local smearing is achieved which reasonably well represents the donor or acceptor distribution in the corresponding composite cystal-field environments.

Determination of the CoMSIA fields

Similar to CoMFA, a data table has been constructed from similarity indices [4] calculated at the intersections of a regularly spaced lattice (1.1 and 2 Å spacing). Similarity indices $A_{F,k}$ between the compounds of interest and a probe atom have been calculated according to (e.g. at grid point q for molecule j of the data set):

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

with i: summation index over all atoms of the molecule j; w_{ik}: actual value of the physicochemical property k of atom i, $w_{\mathrm{probe},k}$: probe atom with charge $+1$, radius 1 Å, hydrophobicity $+1$, H-bond donor and acceptor property $+1$; α: attenuation factor; r_{iq}: mutual distance between probe atom at grid point q and atom i of the test molecule.

Large values of α correspond to a strong distance-dependent attenuation of the similarity measure. Only if the probe closely approaches the surface of a test molecule, similarity is enumerated. The probe ‘looks’ just locally into the molecules and similarity is experienced by local feature matches. For small α’s a soft distance dependence is computed, accordingly a probe close to an aligned test compound will experience similarity with many more of its atoms. The global molecular features obtain higher importance. In the present study α has been set to 0.3 for all five properties. 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. Other values have been checked, however no better statistical results could be detected.

In the present study five physicochemical properties have been evaluated: steric contributions by the third power of the atomic radii, electrostatics by atomic AM1 charges [14], hydrophobicities by atom-based parameters [15] and hydrogen-bonding properties by suitably placed pseudoatoms. The dimensions of the surrounding lattice were selected with a sufficiently large margin ($>$ 4 Å) to enclose all aligned molecules.

As in previous studies [4], a lattice box similar to that generated by CoMFA was used for the data set of 61 thermolysin inhibitors. Due to the size of the training data set and our intention to use three physicochemical and two hydrogen-bonding properties simultaneously in PLS, the grid-spacing of the box has been fixed to 2 Å first. This was mainly done to keep the number of variables down and accordingly the problem in computationally tractable size. For better graphical interpretation of the PLS results a step size of 1.1 Å has also been chosen (see below).

Structural alignment

Model building and alignment of the original training set was performed as described in a previous study [4] using Sybyl [16] (MaxiMin) and the SEAL method [17, 18].

In a study of Campbell et al. [9] the synthesis of a combinatorial library of different peptidylphosphonates (Cbz-X^{p}-^{o}Y-Z-resin, Figure 1) that contains a number of potent thermolysin inhibitors is described. The library has been assayed for thermolysin inhibition while attached to the resin. The most interesting candidates from this library are the following [19, 20]: Cbz-X^{p}-^{o}Leu-Ala(OH) in which X varied, Cbz-Phe^{p}-^{o}Y-Ala(OH) where Y is varied, and Cbz-Phe^{p}-^{o}Leu-Z-resin where Z is varied.

Four members of these libraries (Cbz-Leu^{p}-Leu-Ala = ZLPOLA, Cbz-Phe^{p}-Leu-Ala = ZFPOLA, Cbz-Gly^{p}-Leu-Ala = ZGPOLA, Cbz-Ala^{p}-Leu-Ala = ZAPOLA) were already present in our original training set [4]. They were used as structural reference to construct additional candidates from the above mentioned libraries (Sybyl [16]). The alignment of these new tripeptide inhibitors with the crystallographically studied reference Cbz-Phe^{p}-Leu-Ala (ZFPLAZNCRYS) has been performed using the SEAL method as described previously [4]. Additional investigations were performed with respect to the carboxy terminus of Cbz-Phe^{p}-^{o}Leu-Ala-X with X = OH or NH_2 in order to compare our calculations with the experimental results obtained by Campbell et al. [9].

Results of the CoMSIA analyses

The previously performed CoMSIA PLS analyses [4] with the training set of 61 thermolysin inhibitors, fitted by SEAL, and considering three fields revealed a cross-validated q^2 of 0.587 (see Reference 4, Table 3 r^2 = 0.896, CoMSIA (6)) based on a grid spacing of 1.0 Å). Table 1 (present paper) contains the results of two additional CoMSIA PLS analyses using the same 61 thermolysin inhibitors but now considering all five fields. Analyses (2) and (3) show that there is essentially no difference related to the grid spacing. These analyses, combining the previously used steric, electrostatic and hydrophobic CoMSIA fields with the two additional hydrogen-bonding fields, reveal a slightly improved q^2 and r^2. Only an insignificant increase in the predictive power (q^2) of the five-field model compared to the former three-field model is observed. It is likely that the properties considered are intercorrelated however in a complicated way. The intercorrelation of these numerically intensive grid fields is difficult to detect. An obvious correlation caused by similar field contributions at similar lattice points is rather unlikely. As many examples in literature demonstrated, already a one- or two-field model (cf. CoMFA performed on the present data set [4]) can obtain convincing predictive power. Accordingly, this observation does not justify the introduction of two additional properties, especially if comparative molecular-field methods are primarily used for affinity prediction of novel compounds. However, this cannot be the main purpose of the application of such elaborate techniques. The advantage of using five or more different fields of well defined molecular properties has to be seen in the straight forward partitioning of these properties into spatial locations where they take a determining role on biological activity. This aspect is of utmost importance if a targeted optimization of molecules in a design program is anticipated and 3D-QSAR is supposed to support this step.

Table 1. Results of the different CoMSIA analyses of a data set of thermolysin inhibitors: (1) three fields with 2 Å grid spacing; (2) five fields with 2 Å grid spacing; (3) five fields with 1.1 Å grid spacing

	CoMSIA (1)	CoMSIA (2)	CoMSIA (3)
q^2	0.587	0.591	0.591
S_{press}	1.414	1.408	1.406
r^2	0.899	0.927	0.926
S	0.698	0.593	0.598
No. comp	7	7	7
Fraction			
steric	0.303	0.174	0.171
electrostatic	0.270	0.152	0.145
hydrophobic	0.427	0.223	0.232
H-acceptor		0.259	0.262
H-donor		0.191	0.190
Box			
Stepsize (Å)	2	2	1.1
x	−9 to 16	−9 to 16	−9 to 16
y	−17 to 11	−17 to 11	−17 to 11
z	−11 to 10	−11 to 10	−11 to 10
points	2145	2145	11960

At this point, the major advantage of CoMSIA compared to standard CoMFA becomes important: the better ability to visualize and interpret the obtained correlations in terms of field contributions. Strictly speaking, the plots represent isocontours of the ob-

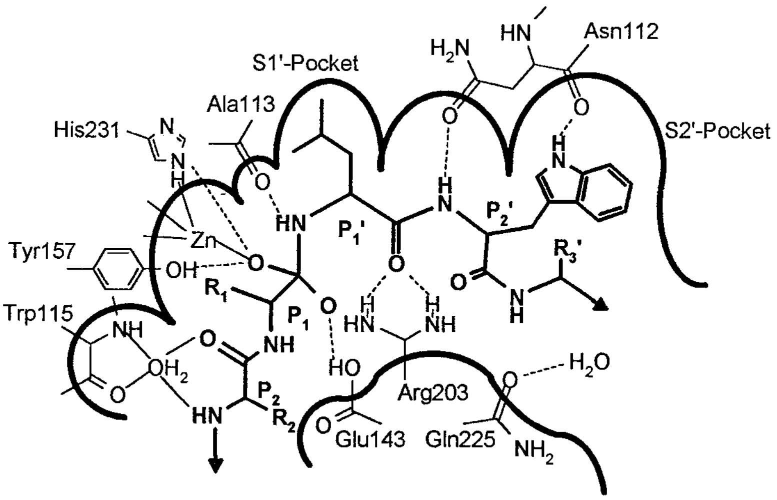

Figure 2. Schematic sketch of the binding site of thermolysin with a substrate-like ligand. The peptide cleavage proceeds via a tetrahedral transition state coordinating to the zinc. Typical interactions between the ligand and key active-site residues or structurally important water molecules are indicated.

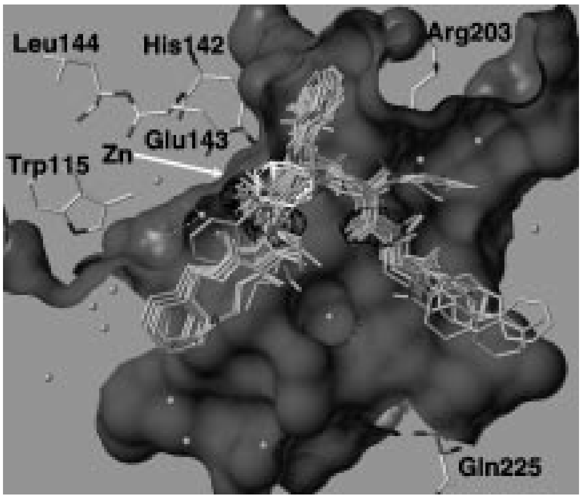

Figure 3. Diagram of the isocontour plot of field contributions of the electrostatic properties based on CoMSIA (contour level at ±0.02 kcal/mol of the CoMSIA coefficient *standard deviation). Superimposed are the aligned inhibitors (grey) and some key residues in the active site of thermolysin according to the crystal structure (atom coded), the solvent accessible surface of the protein is shown in solid, active site water molecules as balls. A z-clipping has been applied in all diagrams to focus on the important part of the active site. Areas contoured in black (chicken-wire) correspond to regions where negatively charged residues enhance binding affinity. In areas surrounded by white isopleths increasingly positive charge will enhance affinity.

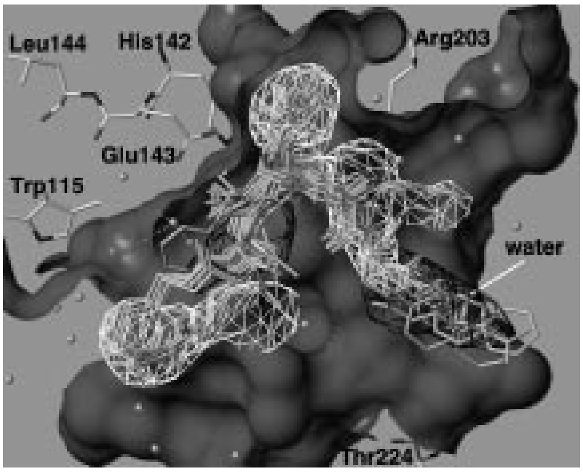

Figure 4. Diagram of the isocontour plot of field contributions of the steric properties based on CoMSIA (contour level at −0.004 and 0.003 kcal/mol), protein surface in solid. Areas contoured by a white chicken-wire indicate regions favorable for steric occupancy. Areas where steric bulk reduces binding affinity are contoured in black.

tained coefficients from PLS. They indicate those lattice points where a particular property significantly contributes and thus explains the variation in affinity data. They give an excellent insight into the relationship between structure and activity for the different physicochemical properties of the considered structures. To obtain smoother and more detailed contour plots for graphical interpretation, smaller grid spacings (e.g. of 1 Å) are desired. In the present study, we had to select a grid spacing of 1.1 Å (analysis (3)). Due to insufficient memory dimensioning in the program version used, an even higher resolution could not be handled considering five different fields and 61 structures at a time in the training set. However, the

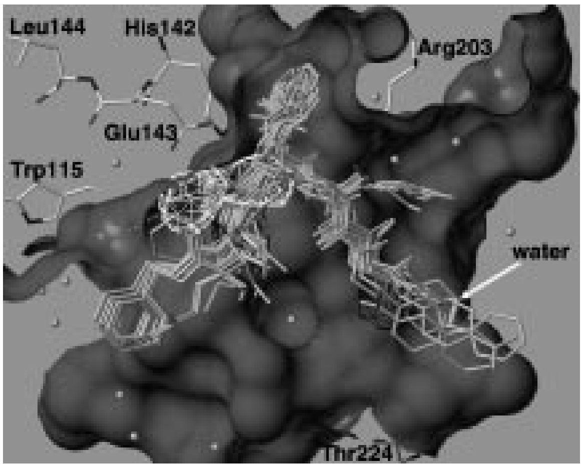

Figure 5. Diagram of the isocontour plot of field contributions of hydrophobic properties based on CoMSIA (contour level at −0.010 and 0.017 kcal/mol), protein surface in solid. Black isopleths encompass regions where hydrophilic groups improve binding, in the white contoured areas hydrophobic substituents are more favorable.

statistical results of analysis (3) are nearly identical to those of analysis (2) [21].

Graphical interpretation and discussion of the observed correlations

In the following the isocontour plots of the various properties will be interpreted. In this discussion, reference is taken to the protein environment of thermolysin to explain observed correlations with respect to protein-ligand interactions. Usually, 3D-QSAR studies are not performed if the 3D structure of the target protein is known, since more powerful tools for drug design are available. However, to judge and validate the physical meaning of an observed correlation, studies like the present one are very helpful.

A general overview of the binding site of thermolysin is given in Figure 2. The assumed binding of a peptidic substrate is sketched. In Figure 3, the electrostatic properties are summarized. The aligned ligands are shown together with some key residues in the active site. The solvent-accessible surface is given as grey solid surface. Areas where negatively charged groups enhance affinity are contoured by a black chicken-wire. Groups of increasing positive charge enhance affinity if they coincide with regions surrounded by white isopleths. A black contoured area falls close to the zinc binding site. The obviously required negative charge in this region corresponds to negatively charged functional groups that serve as potent coordinating groups for the metal ion. The second black contour coincides with the location of the substrate's amide bond following the P2′ position (Figure 2). Some of the potent ligands in the data set possess a charged carboxy terminus at this position. Apparently, the presence of this group improves affinity.

The steric contour map (Figure 4) indicates a preferred occupancy of the S1′ and S2′ pocket (cf. white isopleths in these areas). As for substrate recognition, occupancy of the specificity pockets is important for inhibitor binding. Another extended region requiring steric bulk falls close to the protein-solvent interface (beyond the P2 position). Groups of increasing steric bulk in this region enhance binding affinity. Three areas are unfavorable for steric occupancy (black isopleths). One above zinc at the P1 position, another at the rim of the S2′ pocket (not visible in Figure 4), and a third at the far end of the binding site where it opens to the solvent. This latter area is only occupied by ligands with substituents extending beyond the P2′ position. How can this observation be understood? The crystal structure with phosphoramidon, a potent inhibitor that does not extend beyond P2′, shows a water molecule in this region indicated as sterically unfavorable. Ligands binding to this area would have to replace this water molecule. It could well be that the replacement of the water molecule bound to Gln 225 is rather costly in terms of ΔG. Accordingly, any ligand substituting this water molecule upon binding might lose part of its affinity.

This effect is also partially indicated in the hydrophilicity plots (Figure 5). A black isopleth points to the requirement for hydrophilic groups. Close to the surface water molecule bound to Gln 225, a black contoured region denotes the necessity of polar groups in this area.

The position adjacent to the zinc chelating group (toward P1′) is contoured to be favorable for hydrophobic groups (white). The same region had been indicated in the electrostatic map (Figure 3) to favor a more positively charged group. This correlation can be explained by analysing the trends in biological data for some of the inhibitors. Next to this position, the carbonyl oxygen of Ala 113 provides a H-bond acceptor facility to the ligands. A closely related series of phosphonamidates, phosphonates and phosphinates all show similar binding geometries [21], however, whereas phosphonamidates and phosphinates possess comparable affinities, the phosphonates are less active by a factor of 1000 [22, 23]. The amidates

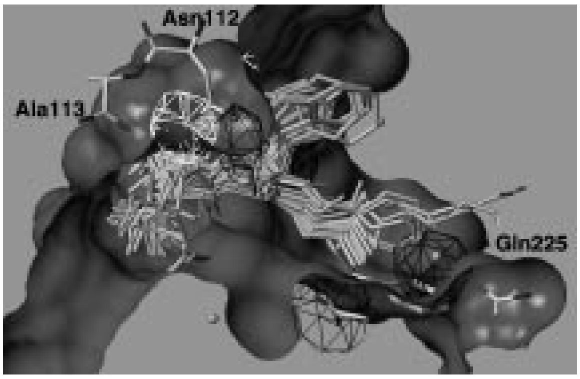

Figure 6. Diagram of the isocontour plot of field contributions of hydrogen-bond acceptor properties based on CoMSIA (contour level at −0.02 and 0.007 kcal/mol), protein surface in solid. White chicken-wire contours map areas beyond the ligands where an acceptor group in the receptor will be favorable for binding. Regions surrounded by black isopleths point toward hydrogen-bond acceptor capabilities that do not enhance receptor affinity.

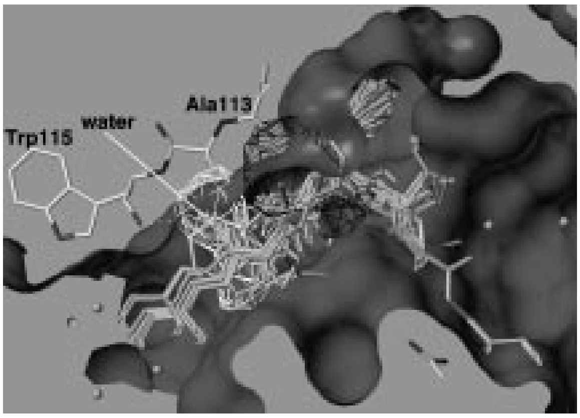

Figure 7. Diagram of the isocontour plot of field contributions of hydrogen-bond donor properties based on CoMSIA (contour level at −0.008 and 0.006 kcal/mol), protein surface in solid. Black isopleths indicate areas for which the presence of a donor group in the receptor is unlikely. White contoured regions point towards the required occurrence of a donor group in the protein.

form a hydrogen bond with their NH group to the carbonyl oxygen of Ala 113 (Figure 2). A comparable hydrogen-bond donor is missing in the analog phosphonates and phosphinates. Accordingly, these inhibitors are not able to form equivalent hydrogen bonds to the protein. In solution, all polar groups in the three inhibitor classes are probably involved in hydrogen bonding to solvent molecules, in particular the NH and O of the phosphonamidates and phosphonates. 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 phosphonamidates and phosphinates, whereas the phosphonates lose their hydrogen-bonding environment about O. They even expose the atom to an electrostatic repulsion with the carbonyl oxygen. This uncompensated situation results in lower affinity.

In both cases, an O → CH_2 or O → NH replacement reveals an enhancement of binding affinity. On a molecular level, these replacements increase hydrophobicity and introduce a more positive charge in this region. Both such changes are indicated as affinity-enhancing in the contour maps.

The graphical interpretation of the field contributions of the hydrogen-bonding properties are shown in Figure 6 (hydrogen-bond acceptor field) and Figure 7 (hydrogen-bond donor field). In principle, they should highlight areas beyond the ligands where putative hydrogen partners in the enzyme could form H-bonds that influence binding affinity significantly. White contouring in these maps indicate that groups possessing the analyzed property will be favorable for binding, whereas black contours show that this property should be absent in this area.

A white isopleth in the acceptor field (Figure 6) surrounds the carbonyl oxygen in the side chain of Asn 112, indicating this area to be favorable for hydrogen-bond acceptors. In fact, the carbonyl oxygen of the Asn 112 side chain is frequently involved as acceptor in hydrogen bonds toward potent thermolysin inhibitors. The three regions contoured in black should lack hydrogen-bond acceptor capabilities. This is in agreement for the amide group of the side chain of Asn 112. The two other black contours, next to Gln 225, encompassing a structural water, and next to Asp 226, are more difficult to interpret. They result from a series of low-affinity inhibitors in the data set that all orient a NH_2 group in this region. In consequence, next to these NH_2 groups, the statistical analysis generates positions for putative H-bond acceptors that are classified to be unfavorable for binding. One can only speculate why these compounds show low affinity. One reason might be that the adjacent Asp and the water molecule cannot adopt a geometry suited to accept an efficient H-bond.

In the donor field (Figure 7), the black isopleth around the backbone carbonyl oxygen of Ala 113 indicates an area where the presence of a hydrogen-bond donor would reduce affinity. This is in agreement with the fact that the protein orients a carbonyl oxygen into this area which accepts a hydrogen bond from potent

Table 2. Data set of peptidylphosphonate sequences used to predict binding affinities with the CoMSIA approach, pK$_i$'s experimentally determined (left) and predictions by CoMSIA

	pK$_i$	CoMSIA (3)
X^p-Position (= P$_1$-position)		
Cbz-Phe-Leu-Ala (ZFPOLA)	7.35	7.29
Cbz-Leu-Leu-Ala (ZLPOLA)	6.17	6.27
Cbz-Ile-Leu-Ala		5.25
Cbz-Ala-Leu-Ala (ZAPOLA)	5.74	5.57
Cbz-Val-Leu-Ala		5.38
Cbz-Gly-Leu-Ala	4.89	5.48
Y^p-Position (= P$'_1$-position)		
Cbz-Phe-Leu-Ala (ZFPOLA)	7.35	7.29
Cbz-Phe-Phg-Ala		6.60
Cbz-Phe-Phe-Ala	5.56	6.72
Cbz-Phe-Ala-Ala	3.93	4.22
Cbz-Phe-Gly-Ala	3.02	3.80
Z-Position (= P$'_2$-position)		
Cbz-Phe-Leu-lfis	7.24	7.34
Cbz-Phe-Leu-Arg	7.19	8.01
Cbz-Phe-Leu-Gln	6.90	7.53
	Hydrolysis rate mA/min	Calculated pK$_i$
Cbz-Phe-Leu-Ala (ZFPOLA)	16	7.29
Cbz-Phe-Leu-Asp	26	7.04
Cbz-Phe-Leu-Gly	28	6.44
Cbz-Phe-Leu-Met	30	6.16
Cbz-Phe-Leu-Tyr	32	6.90
Cbz-Phe-Leu-Trp	36	8.02
Cbz-Phe-Leu-Glu	44	6.75
Cbz-Phe-Leu-Ser	52	6.56
Cbz-Phe-Leu-Lys	56	6.51
Cbz-Phe-Leu-Phe	60	6.18
Cbz-Phe-Leu-lleu	68	6.33
Cbz-Phe-Leu-Leu	68	6.44
Cbz-Phe-Leu-Val	76	6.46
Cbz-Phe-Leu-Thr	76	6.08
Cbz-Phe-Leu-Pro	124	6.01

inhibitors. A white-contoured area pointing toward the favorable occurrence of hydrogen-bond donor groups in the protein encompasses a water molecule at the cleavage site. In this case a protein residue is not suggested as bonding partner, but a structurally important water that mediates a hydrogen bond between a ligand and Trp 115.

Affinity scoring of a combinatorial library

The previous discussion of the graphical results of the different field contributions has demonstrated that many of the features in these maps can be interpreted in terms of properties reflected in the surrounding protein environment. In order to examine the predictive power of the present comparative molecular field analysis, especially with respect to ranking and designing candidates for combinatorial libraries, results from a study by Campbell et al. [9] have been used. Our modeling study has been performed as a post-predictive investigation, however it demonstrates that the CoMSIA and CoMFA method can be used to rank possible candidates from a library in a predictive manner. Such information is important in the design and selection of components for a combinatorial library. Using the split bead method, Campbell et al. have constructed a library of tripeptidylphosphonate inhibitors of the sequence Cbz-X^p-^{o}Y-Z. The attempted variations at the P$_1$, P$'_1$ and P$'_2$ position are outlined in Figures 1 and 2. The library was assayed for thermolysin inhibition and a rank ordering has been observed for modifications at the three sites.

For our CoMSIA-predictions, the P1 derivatives Cbz-X^p-Leu-Ala(OH) with X^p = Gly, Val, Ala, Ile, Leu, Phe, have been constructed and aligned as described. Using the model derived from the above described training set of 61 inhibitors the affinities of the library candidates were computed. For the Phe, Leu, Ala, and Gly derivative K$_i$'s were published. CoMSIA predicts the affinities of the most potent Phe and Leu derivatives convincingly well. The less potent (factor 100 or 10 respectively) Ile, Ala, Val, and Gly compounds are predicted in the correct range, however an individual discrimination among them is hardly possible.

At the P$'_1$ site the library has been varied by five different α-hydroxy acids. In our study the most potent Phe residue has been placed at the P$_1$ position, at P$'_2$ Ala(OH) has been considered. Measured pK$_i$'s were taken from literature or the reported k$_{cat}$/K$_m$ values were used to estimate approximate pK$_i$'s (Table 2). CoMSIA predicts the trends correctly, only the phenylglycyl derivative is ranked too high. Finally, the P$'_2$ position has been inspected. Campbell et al. [9] checked 18 different amino acids at this position and the C-terminus has been studied as free acid and amide. At the two remaining positions, the residues Phe^p and ^{o}Leu, indicated as optimal, have been considered. First we tried to discriminate acid and amide.

The computed affinities confirm the amides to be more potent. For four members of the Cbz-Phe^{p}-^{o}Leu-Z-library detailed pK$_i$'s are given in literature. The prediction with CoMSIA meets the experimental values with a deviation of clearly less than one order of magnitude. For the remaining amino-acid variations no pK$_i$ values are given. However, the hydrolysis rate described in the study of Campbell et al. [9] can serve as a first estimate to derive an affinity rank order. CoMSIA reveals a trend among the different Z-derivatives. According to experiment, the more potent (low hydrolysis rate) peptidylphosphonates obtain the higher predicted pK$_i$'s.

The scoring of different members of a combinatorial library shows that 3D-QSAR techniques can be used to select and compose optimal components for such libraries. The example given demonstrates the predictive potential of the CoMSIA method. Since we considered the different substitutions at the three positions in a subsequent manner, alternating each site at a time, additivity of the different groups to binding affinity has been assumed. In a first approximation this hypothesis appears to be justified.

Conclusions

The present study shows a comparative molecular field analysis based on similarity indices using five different fields. Interestingly enough the newly introduced hydrogen-bonding fields do not improve the predictive power of the analysis. This fact suggests strong intercorrelations with the previously considered fields. Accordingly, if solely the development of a predictive QSAR model is anticipated, the consideration of such additional fields cannot be justified. However, the challenge to perform such elaborate analyses has to be more ambitioned. A 3D-QSAR tool is expected to translate structural variations within a set of biologically graded molecules into spatially resolved features that indicate where different physicochemical properties matter and how they create affinity changes. In that respect as many relevant property fields a possible should be evaluated. This further calls upon a powerful tool to assess the obtained correlations by graphical means.

The CoMSIA approach described in this paper uses at present five different property fields. The method projects and filters out where steric, electrostatic, hydrophobic and H-bonding properties account in space for affinity. Since for the present case study the thermolysin environment is known and can be consulted as a reference for a better understanding, the complexity of some of the indicated correlations can be interpreted in structural terms (e.g. exchange of $CH_2 \longrightarrow NH \longrightarrow O$).

Even though this information is hardly given in a real-life 3D-QSAR application the indicated property features help to modify and optimize given molecules in a design project. As an example the selection of components for a combinatorial library is described. This selection could be based on predictions from such a comparative field analysis.

Acknowledgements

The authors are grateful to Markus Böhm (University Marburg, Germany) for his help in preparing the figures of the different contribution maps.

References

1. Müller, K., Perspect. Drug Discov. Design, 3 (1995) 1.
2. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
3. Thibaut, U., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design, ESCOM, Leiden, 1993, pp. 661–696.
4. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
5. Folkers, G., Merz, A. and Rognan, D., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design, ESCOM, Leiden, 1993, pp. 583–618.
6. Stahle, L. and Wold, S., Prog. Med. Chem., 25 (1988) 292.
7. Cramer III, R.D., De Priest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.), 3D QSAR in Drug Design, ESCOM, Leiden, 1993, pp. 443–485.
8. DePriest, S.A., Mayer, D.C.B. and Marshall, G.R., J. Am. Chem. Soc., 115 (1993) 5372.
9. Campbell, D.A., Bermate, J.C., Burkoth, T.S. and Patel, D.V., J. Am. Chem. Soc., 117 (1995) 5381.
10. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P., J. Comput.-Aided Mol. Design, 7 (1993) 83.
11. Program DISCO is available from Tripos Ass., St. Louis, MO, USA.
12. Allen, F.H., Kennard, O. and Taylor, R., Acc. Chem. Res., 16 (1983) 146.
13. Klebe, G., J. Mol. Biol., 237 (1994) 212.
14. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
15. Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K., J. Chem. Inf. Comput. Sci., 29 (1989) 163.
16. Sybyl Molecular Modeling System (Version 6.1), Tripos Ass., St. Louis, MO, USA.
17. Kearsley, S.K. and Smith, G.M., Tetrahed. Comput. Meth., 3 (1990) 615.
18. Klebe, G., Mietzner, T. and Weber, F., J. Comput.-Aided Mol. Design, 8 (1994) 751.

19. Bartlett, P.A. and Marlowe, C.K., Biochemistry, 26 (1987) 8553.
20. Morihara, K. and Tsuzuki, H., Eur. J. Biochem., 15 (1970) 374.
21. We have to admit that is hardly possible to document by black-and-white drawings the graphical advantage of the CoMSIA maps that become obvious using an interactive computer terminal.
22. Morgan, B.P., Scholtz, J.M., Ballinger, M.D., Zipkin, I.D. and Bartlett, P.A., J. Am. Chem. Soc., 113 (1991) 297.
23. Tronrud, D.H., Holden, H.M. and Matthews, B.W., Science, 235 (1987) 571.
24. Bartlett, P.A. and Marlowe, C.K., Science, 235 (1987) 569.
25. Bash, U.C., Singh, P.A., Brown, F.K., Langridge, R. and Kollman, P.A., Science, 235 (1987) 574.