On the Prediction of Binding Properties of Drug Molecules by Comparative Molecular Field Analysis

Gerhard Klebe* and Ute Abraham

BASF AG, Main Laboratory, Carl-Bosch Strasse, D-6700 Ludwigshafen, FRG

Received May 21, 1992

Comparative molecular field analysis (CoMFA) has been applied to three different data sets of drug molecules binding to human rhinovirus 14 (HRV14), thermolysin and renin, respectively. Different structural alignments have been tested to predict binding properties. An alignment based on crystallographically determined coordinates of the inhibitors bound to the proteins has been compared with alignments obtained from multiple-fit and field-fit procedures. These methods are commonly used for systems where no reference to protein structural data is available. For HRV14, two different models, one based on experimental evidence and one based on a hypothetical alignment reveal moderate predictions of the binding constant of comparable quality. For thermolysin, hypothetical alignments allow a substantially better prediction than an alignment based on experimental evidence. The prediction of binding properties (expressed as ΔG, ΔH, and ΔS) of renin inhibitors, which were aligned on the basis of crystallographic data from related inhibitors bound to the aspartyl protease endothiapepsin, gives evidence that only enthalpies (ΔH) and not free enthalpies (ΔG) or binding constants can be properly predicted by comparative molecular field analysis.

Introduction

One of the main goals in drug design is the prediction of new biologically active compounds on the basis of previously synthesized and tested compounds. Many methods have been described to approach this challenging and difficult problem.[1] Comparative molecular field analysis (CoMFA) developed by Cramer et al.[2] has become a popular and valuable tool in this field.[3-5] The basic assumption in this method is that "a suitable sampling of the steric and electrostatic fields around a set of aligned ligand or drug molecules might provide all the information necessary for understanding their biological properties".[2]

The suitable sampling is achieved by calculating the steric and electrostatic interactions[6] between each ligand and an appropriate probe at regularly spaced grid points of a three-dimensional lattice. A stable structure/activity relationship is then extracted from the data table of interaction energies by a partial least-squares (PLS) analysis.[7] The sampling requires that all ligands have been preliminarily aligned in a way which—at best—is equivalent to the relative spatial arrangement of the various drug molecules at the protein binding site. Without knowing the geometry of the binding site, this appears at a first glance to be a crucial and rather unresolvable task. However, many techniques have been developed[1,8] to derive model receptors and achieve alignments which allow—at least—a reasonable (within the model) superposition of ligands in conformations corresponding to the biologically active ones. The biological properties of the ligands are usually expressed by their binding constants K. These quantities are related to the free enthalpy of binding through $\Delta G = -RT \ln K$.[9]

This present paper tries to shed some light on the importance of the mutual alignment of the ligands and on

(1) Marshall, G. R.; Motoc, I. Approaches to the conformation of the drug bound to the receptor. In *Molecular Graphics and Drug Design*; Burgen, A. S. V., Roberts, G. C. K., Tute, M. S., Eds.; Elsevier Sci. Publ.: Amsterdam, 1986; pp 116-156. Marshall, G. Computer-Aided Drug Design. *Ann. Rev. Pharmacol. Toxicol.* 1987, *27*, 193-213. Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. Molecular Modelling Software and Methods for Medicinal Chemistry. *J. Med. Chem.* 1990, *33*, 883-894. Dean, P. M. Molecular foundations of drug-receptor interaction; Cambridge Univ. Press: Cambridge, 1987. Franke, R. Theoretical Drug Design Methods; Elsevier: Amsterdam, 1984. Tintelnot, M.; Andrews, P. Geometries of Functional Group Interactions in Enzyme-Ligand Complexes: Guides for Receptor Modelling. *J. Comput.-Aided Mol. Des.* 1989, *3*, 67-84. Richards, W. G. *Computer-Aided Molecular Design*; IBC Technical Services Ltd.: London, 1989. Goodford, P. J. Receptors, Spare Receptors and other Binding Sites. *Trends Pharmacol. Sci.* 1984, *4*, 90. Goodford, P. J. Drug Design by the Method of Receptor Fit. *J. Med. Chem.* 1984, *27*, 557-564. Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* 1985, *28*, 849-857. Bedell, C. R. Designing Drugs to Fit a Macromolecular Receptor. *Chem. Soc. Rev.* 1984, *13*, 279-319. Fruehbeis, H.; Klein, R.; Wallmeier, H. Computer-assisted Molecular Design (CAMD)-an overview. *Angew. Chem., Int. Ed. Engl.* 1987, *26*, 403-418.

(2) 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.

(3) Greco, G.; Novellino, E.; Silipo, C.; Vittoria, A. Comparative Molecular Field Analysis on a Set of Muscarinic Agonists. *Quant. Struct.-Act. Relat.* 1991, *10*, 289-299.

(4) Kim, K. H.; Martin, Y. C. Direct Prediction of Dissociation Constants (pKa's) of Clonidine-like Imidazolines, 2-Substituted Imidazoles, and 1-Methyl-2-substituted-imidaziles from 3D Structures Using a Comparative Molecular Field Analysis (CoMFA) Approach. *J. Med. Chem.* 1991, *34*, 2056-2060.

(5) Allen, M. S.; Tan, Y. C.; Trudell, M. L.; Narayanan, K.; Schindler, L. R.; Martin, M. J.; Schultz, C.; Hagen, T. J.; Koehler, K. F.; Codding, P. W.; Skolnick, P.; Cook, J. M. Synthetic and Computer-Assisted Analyses of the Pharmacophore for the Benzodiazepine Receptor Inverse Agonist Site. *J. Med. Chem.* 1990, *33*, 2343-2357. Thomas, B. F.; Compton, D. R.; Martin, B. R.; Semus, S. F. Modeling the Cannabinoid Receptor—A 3-dimensional Quantitative Structure Activity Analysis. *Mol. Pharmacol.* 1991, *40*, 656-665. Carroll, F. I.; Gao, Y. G.; Rahman, M. A.; Abraham, P.; Parham, K.; Lewin, A. H.; Boja, J. W.; Kuhar, M. J. Synthesis, Ligand Binding, QSAR, and CoMFA Study of 3β-(p-Substituted phenyl)tropane-2β-carboxylic Acid Methyl Esters. *J. Med. Chem.* 1991, *34*, 2719-2725. Bjorkroth, J. P.; Pakkanen, T. A.; Lindroos, J.; Pohjala, E.; Hanhijarvi, H.; Lauren, L.; Hannuniemi, R.; Juhakoski, A.; Kippo, K.; Kleimola, T. Comparative Molecular Field Analysis of Some Clodronic Acid Esters. *J. Med. Chem.* 1991, *34*, 2338-2343. Kim, K. H.; Martin, Y. C. Direct Prediction of Linear Free Energy Substituent Effects from 3D Structures Using Comparative Molecular Field Analysis. 1. Electronic Effects of Substituted Benzoic Acids. *J. Org. Chem.* 1991, *56*, 2723-2729. Kellogg, G. E.; Semus, S. F.; Abraham, D. J. Hint: A new methods of empirical hydrophobic field calculations for CoMFA. *J. Comput.-Aided Mol. Des.* 1991, *5*, 545-552.

(6) The use of these two fields is justified by the experience that in a molecular mechanics approach these two account for those nonbonded interactions important for a drug/receptor interaction.

(7) Stahle, L.; Wold, S. Multivariate Data Analysis and Experimental Design in Biomedical Research. *Prog. Med. Chem.* 1988, *25*, 292-334 (Ellis, G. P., West, G. B., Eds.).

0022-2623/93/1836-0070$04.00/0

the predictability of binding constants using only steric and electrostatic field differences of the drug molecules.

To reduce the uncertainties about biologically active conformations and mutual alignments within a set of different ligands, only systems where structural data on the protein (enzymes, coating proteins) and ligand/protein complexes are available from X-ray crystallography were considered. Antiviral compounds of human rhinovirus 14 (HRV14) and inhibitors of thermolysin and renin have been selected. In all cases, crystallographic results (references in Tables I–III) indicate that the residues at the binding site show only minor displacements due to the influences of an induced fit. Thus, the assumption of "common" fields around the ligands in their adopted conformations and relative alignments appears to be justified.

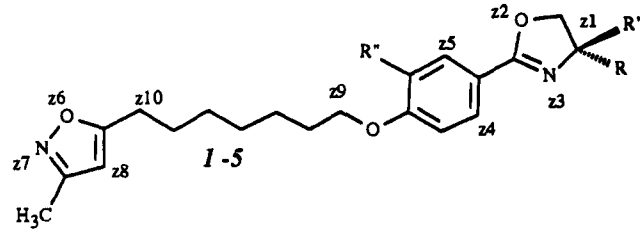
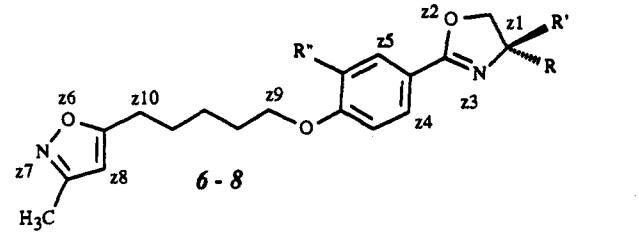
These alignments, based on experimental evidence, are used in CoMFA to describe binding affinities (expressed as ln K) in a quantitative way. The obtained results are placed against those derived from other alignments which would appear as a reasonable model without having reference to the experimentally determined binding geometries. In the case of renin inhibitors, CoMFA analyses have been performed to predict either the enthalpy, entropy, or free enthalpy of binding.

Molecular Conformation and Alignment

The most crucial variable in CoMFA is the positioning of the molecules within a fixed lattice. For two of the three case studies, deviating alignments were evaluated. All structural manipulations were performed with the molecular modeling package SYBYL.[10]

In a first attempt, a CoMFA analysis was performed on the basis of coordinates taken from ligand/protein complexes determined by X-ray crystallography (molecular constitution and references to PDB codes, see Tables I–III). Starting with the coordinates from the PDB file,[11] structures 1–8 and 9–15 were optimized for bond lengths and bond angles (but fixed torsion angles) with the molecular mechanics program MOMO.[12] This optimization was performed to ensure that all molecules (including those derived by homology building, see below) had consistent bond-length and bond-angle geometry. Furthermore, this aspect is important for a uniform determination of charges (see below). Root-mean-square deviation determined for a fit of all non-hydrogen atoms in the various molecules before and after the optimization shows that the binding geometries are not altered by this procedure (for 1–8, 0.1–0.3 Å; 9–15, 0.2–0.4 Å, which is in the same range as the experimental error for the atomic positions in protein X-ray structures[13]). Hydrogen atoms were added with standard geometries, assuming reasonable protonation states for the polar groups at physiological pH and the likelihood of forming hydrogen bonds with the adjacent active-site residues. The relative arrangement of 1–8 or 9–15 respectively was obtained by aligning the backbone atoms of HRV14 or thermolysin in the active-site region of the corresponding protein/ligand complexes. The obtained alignments are shown in Figures 1 and 2.[14] For thermolysin, the data set was expanded by 16–28. These

Table I. Chemical Structures of Eight Inhibitors (1–8) of HRV14 Used in CoMFA with Reference to the PDB Codes of the Corresponding Ligand/Protein Complexes[32]

no.	R	R′	R″	PDB
				
1	Me	H	H	2RS1
2	H	Me	H	2RR1
3	Me	H	Cl	2RM2
4	Et	H	H	2RS3
				
5	H	H	H	2R04
6	Me	H	H	2RS5
7	H	H	H	2R06
8	H	H	Cl	2R07

[a] The 10 centers used in the multiple-fit procedure for the different molecules in the data set are indicated (z1–z10).

(8) Labanowski, J.; Motoc, I.; Naylor, C. B.; Mayer, D.; Dammkoehler, R. A. Three-Dimensional Quantitative Structure-Activity Relationships. 2. Conformational Mimicry and Topographical Similarity of Flexible Molecules. *Quant. Struct.-Act. Relat.* **1986**, *5*, 138–152. Sit, S. Y.; Parker, R. A.; Motoc, I.; Han, W.; Balasubramanian, N.; Catt, J. D.; Brown, P. J.; Harte, W. E.; Thompson, M. D.; Wright, J. J. Synthesis, Biological Profile, and Quantitative Structure-Activity Relationship of a Series of Novel 3-Hydroxy-3-methyl glutaryl Coenzyme A Reductase Inhibitors. *J. Med. Chem.* **1990**, *33*, 2982–2999. Turconi, M.; Nicola, M.; Quintero, M. G.; Maiocchi, L.; Micheletti, R.; Giraldo, E.; Donetti, A. Synthesis and Biological Evaluation of New Antimuscarinic Compounds with Amidine Basic Centers. A Useful Bioisosteric Replacement of Classical Cationic Heads. *J. Med. Chem.* **1990**, *33*, 2101–2108. Madsen, U.; Brehm, L.; Schaumburg, K.; Jorgensen, F. S.; Krogsgaardlarsen, P. Relationship between Structure, Conformational Flexibility, and Biological Activity of Agonists and Antagonists at the N-Methyl-D-aspartic Acid Subtype of Excitatory Amino Acid Receptors. *J. Med. Chem.* **1990**, *33*, 374–380. Hibert, M. F.; McDermott, I.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. Radioligand binding study of a series of 5-HT_{1A} receptor agonists and definition of a steric model of this site. *Euro. J. Med. Chem.* **1989**, *24*, 31–37. Hibert, M. F.; Gittos, M. W.; Middlemiss, D. N.; Mir, A. K.; Fozard, J. R. Graphics Computer-Aided Receptor Mapping as a Predictive Tool for Drug Design: Development of Potent, Selective, and Stereospecific Ligands for the 5-HT(1A) Receptor. *J. Med. Chem.* **1988**, *31*, 1087–1093.

(9) Some comments on the use of the expressions "energy" (U), "enthalpy" (H), "free energy" (A), and "free enthalpy" (G) through out this paper should be given: The state function U is defined as internal energy of a system under constant-volume conditions. A change in internal energy, e.g., a negative ΔU means that the system loses energy and that this energy is dissipated in heat that is evolved and work that is done by the system. Since most chemical reactions are usually carried out at constant pressure, an equivalent state function, the enthalpy H is defined under constant-pressure conditions. Considering irreversible processes, entropic contributions (S) have to be taken into account. The Helmholtz free energy A refers to processes under constant-volume conditions whereas the Gibbs free energy or free enthalpy G refers to constant-pressure conditions.

(10) SYBYL Molecular Modeling System (Version 5.40 and 5.50), Tripos Ass., St. Louis, MO.

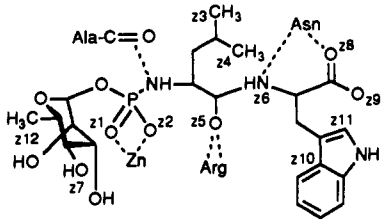
(11) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rogers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structure. *J. Mol. Biol.* **1977**, *112*, 535–542.

(12) Beck, H.; Egert, E. *Force-Field-Program MOMO*; Univ. Göttingen: Göttingen, 1988. Smith, A. E.; Lindner, H. J. π-SCF- Molecular Mechanics PIMM: Formulation, parameters, applications. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 235–262.

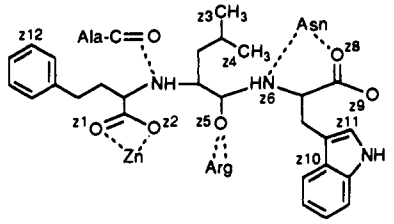
(13) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand Binding to Proteins: Escherichia coli Dihydrofolate Reductase-Trimethoprim, A Drug-Receptor System. *Proteins* **1988**, *4*, 31–42.

(14) Coordinates of the aligned molecules can be obtained in SYBYL-molfile format[10] from the authors on request.

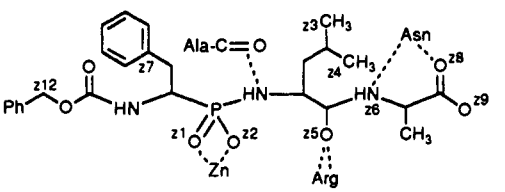
Table II. Chemical Structures of 20 Inhibitors of Thermolysin Used in CoMFA[a]



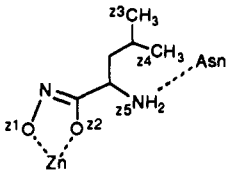
9. 1TLP



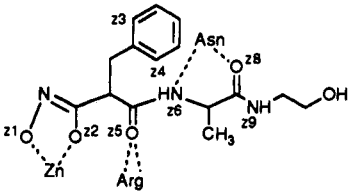
10. 1TMN



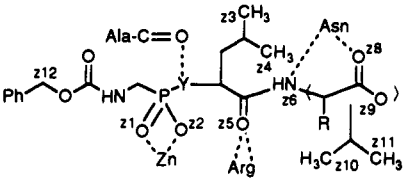
11. 4TMN



12. 4TLN



13. 5TLN



R	Y = NH	Y = O	Y = CH_2	R	Y = NH	Y = O	Y = CH_2
iBu (Leu)	14[b]	15[c]	28[d]	CH_2Ph (Phe)	18	22	26
CH_3 (Ala)	16	20	24	NH_2[e]	19	23	27
H (Gly)	17	21	25				

[a] For 9–15, references to the PDB codes of the corresponding ligand/protein complexes are given. The structures of 16–28 which belong to three congeneric series of inhibitors were constructed from the coordinates of 14 and 15. The fit centers (up to 12) used in the multiple-fit procedure for the different molecules in the data set are indicated (z1–z12). [b] 5TMN. [c] 6TMN. [d] References 15 and 43. [e] The last residue, in brackets, is missing.

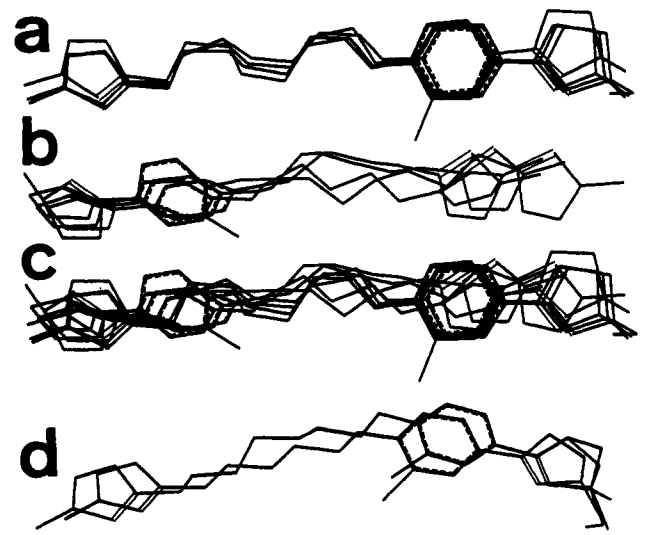


Figure 1. Superposition of inhibitors of HRV14 in the conformations and relative alignment used in CoMFA (hydrogen atoms removed for clarity). Alignments shown in a–c are based on coordinates obtained from X-ray crystal structure analysis:[32] (a) 1–4 with oxazole ring to the left; (b) 5–8 with oxazole ring to the right; (c) all eight inhibitors superimposed; (d) the alignment based on a multiple-fit with two basic conformations is shown.

additional inhibitors[15] are closely related to examples of similar constitution (14, 15, and 28) whose structures were experimentally determined. For these additional compounds, binding modes equivalent to the references were assumed. Each of the constructed ligands was structurally optimized by molecular mechanics.

The data set of renin inhibitors 29–41 was constructed on the basis of structural data taken from closely related inhibitor complexes with endothiapepsin.[16] This procedure was performed as described above for the additional inhibitors of thermolysin. Since all of the nine experimentally determined complexes with endothiapepsin, present in the data base, show a conserved binding mode, the prerequisite for a successful prediction of reasonable binding geometries appears to be given (see Figure 3).

In comparison to these alignments, directly taken or derived from experimental evidence, some additional alignments were tested which are based upon techniques frequently used in molecular modeling.[8] They are usually applied to systems where no information on the receptor binding site is present. In a flexible multiple-fit procedure, it is attempted to superimpose "equivalent" pharmacophoric groups by simultaneously minimizing the distances between these target atoms in the various molecules together with the intramolecular force-field of each individual molecule. Clearly the selection of "pharmacophoric" groups is biased by the conviction of the modeler, and thus many other, maybe even "better", definitions of a pharmacophor might be possible. This procedure, as implemented in the molecular mechanics program MOMO,[17] has been used to obtain complementary alignments.

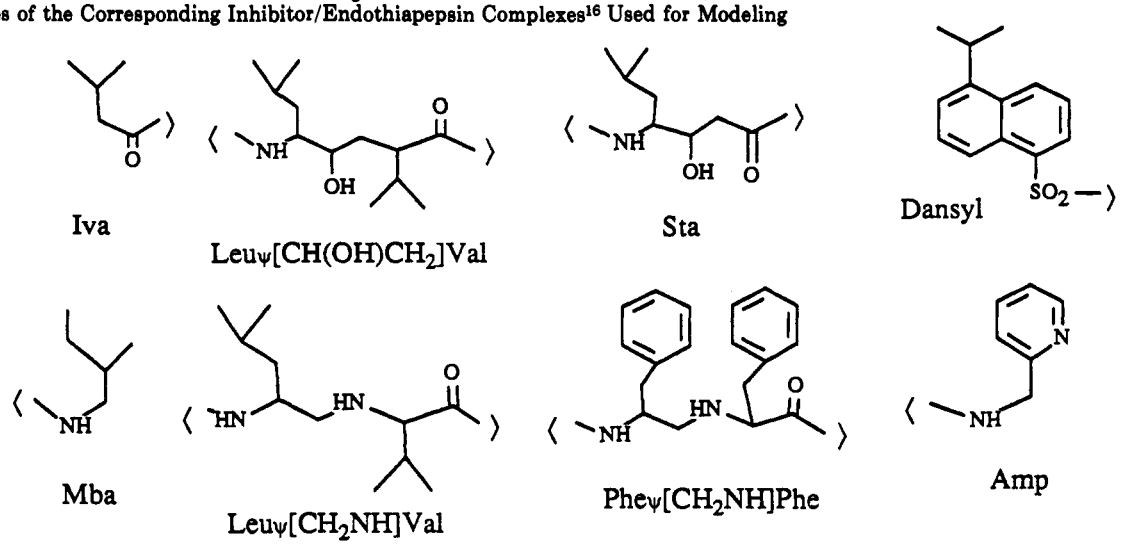
The results of such a procedure are clearly dependent upon the starting geometries. Since, for HRV14 antiviral compounds the reverse binding mode of a part of the inhibitors would hardly be predictable without crystallographic evidence, a uniform alignment was chosen leading to two basic conformations (Figure 1, see Discussion). For thermolysin, an alignment was determined originally starting with the coordinates taken from the protein/inhibitor complexes. Before applying the multiple-fit option in MOMO, each inhibitor was separately relaxed into the next local minimum in its isolated state. The obtained alignment is shown in Figure 2b.

(15) 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.

(16) Pearl, L.; Blundell, T. The active site of aspartic proteinases. *FEBS Lett.* **1984**, *174*, 96–101.

(17) Mietzner, T.; Klebe, G. Manuscript in preparation.

Table III. Chemical Structures and Original Codes[23] of 13 Inhibitors (29–41) of Renin Used in CoMFA with Reference to the PDB codes of the Corresponding Inhibitor/Endothiapepsin Complexes[16] Used for Modeling



No.	Reference	Peptide	PDB
29	U-80215E	Dansyl-His-Leuψ[CH(OH)CH2]Val-Ile-Amp	2ER7
30	U-80631E	Ac-Phe-His-Leuψ[CH(OH)CH2]Val-Ile-NH2	2ER7
31	U-77646E	Ac-Pro-Phe-His-Leuψ[CH(OH)CH2]Val-Ile-NH2	2ER7
32	U-77647E	Ac-D-Pro-Phe-His-Leuψ[CH(OH)CH2]Val-Ile-NH2	2ER7
33	U-76780E	H-Pro-His-Pro-Phe-His-Leuψ[CH2NH]Val-Ile-His-Lys-OH	2ER7
34	U-62168E	H-Pro-His-Pro-Phe-His - Phe - Phe -Ile-His-Lys-OH	2ER7
35	U-77451E	Ac-Pro-Phe-His - Pheψ[CH2NH]Phe-Mba	2ER7
36	U-71909E	Ac-Pro-Phe-His - Pheψ[CH2NH]Phe-NH2	2ER7
37	U-73777E	Ac-Phe-His - Pheψ[CH2NH]Phe-NH2	2ER7
38	U-72407E	Ac-Phe-His - Sta -Ile-NH2	2ER9
39	U-72408E	Ac-Pro-Phe-His - Sta -Ile-NH2	2ER9
40	U-72409E	Ac-His-Pro-Phe-His - Sta -Ile-NH2	2ER9
41	U-77455E	Iva-His-Pro-Phe-His - Sta -Ile-Phe-NH2	2ER0

In addition, for the thermolysin example, the field-fit option as implemented in SYBYL[10] was applied. This procedure is used to increase field similarity within a series of inhibitors.[2] The root-mean-square difference in the sum of the steric and electrostatic interaction energies, averaged across all lattice points, between an individual molecule in the data set and the average of all molecules in the set, is minimized with respect to the six rigid-body degrees of freedom (keeping the individual conformations rigid). According to this procedure, the field-fit alignment, shown in Figure 2c, is derived from the previously described multiple-fit alignment. Modulating the purely steric multiple fit with an additional field fit should better account for differences in the electrostatic properties of the considered molecules.

CoMFA Interaction Energy Calculations

The steric and electrostatic potential energy fields of each molecule in the three data sets with different alignments were calculated with the TRIPOS force field,[18] using a Csp^3 probe atom with a charge of +1.0. Partial atomic charges for the molecules in the various conformations were determined with the MNDO–AM1 method.[19]

Two different energy truncation values of 30/30 and 5/30 kcal/mol were set, respectively, for the steric and electrostatic interactions. In the second choice, similarly used by other authors,[3,4] the steric energy was truncated at lower values to avoid unjustified large parametric variance due to the steep increase of the steric field contribution at lattice points close to the molecules. The electrostatic interactions experienced at lattice points whose steric energy exceeded the truncation value were set to the mean of all the electrostatic values at the same locations. However, the truncation of the steric energy at a low cutoff value can imply that some lattice points where the electrostatic field displays a significant value are not appropriately considered in the analysis.[20] The relative contributions of steric and electrostatic fields are given in the Tables IV, XI, VIII, IX, and X.

The CoMFA lattices were used with 2-Å grid spacing. (Smaller grid spacing increases computing time enor-

(18) Vinter, J. G.; Davis, A.; Saunder, M. R. Strategic Approaches to Drug Design. I. An Integrated Software Framework for Molecular Modelling. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 31–35.

(19) 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.

(20) For the renin example, a cut-off ratio of 100/30 for the steric and electrostatic fields was tested (ΔG). Compared to the values given in Table VIII, the following results were obtained: r^2_{cross}, –0.381 (2 comp); $r^2_{nocross}$, 0.598. The fraction between steric and electrostatic field contributions in this case changes to 0.497/0.503.

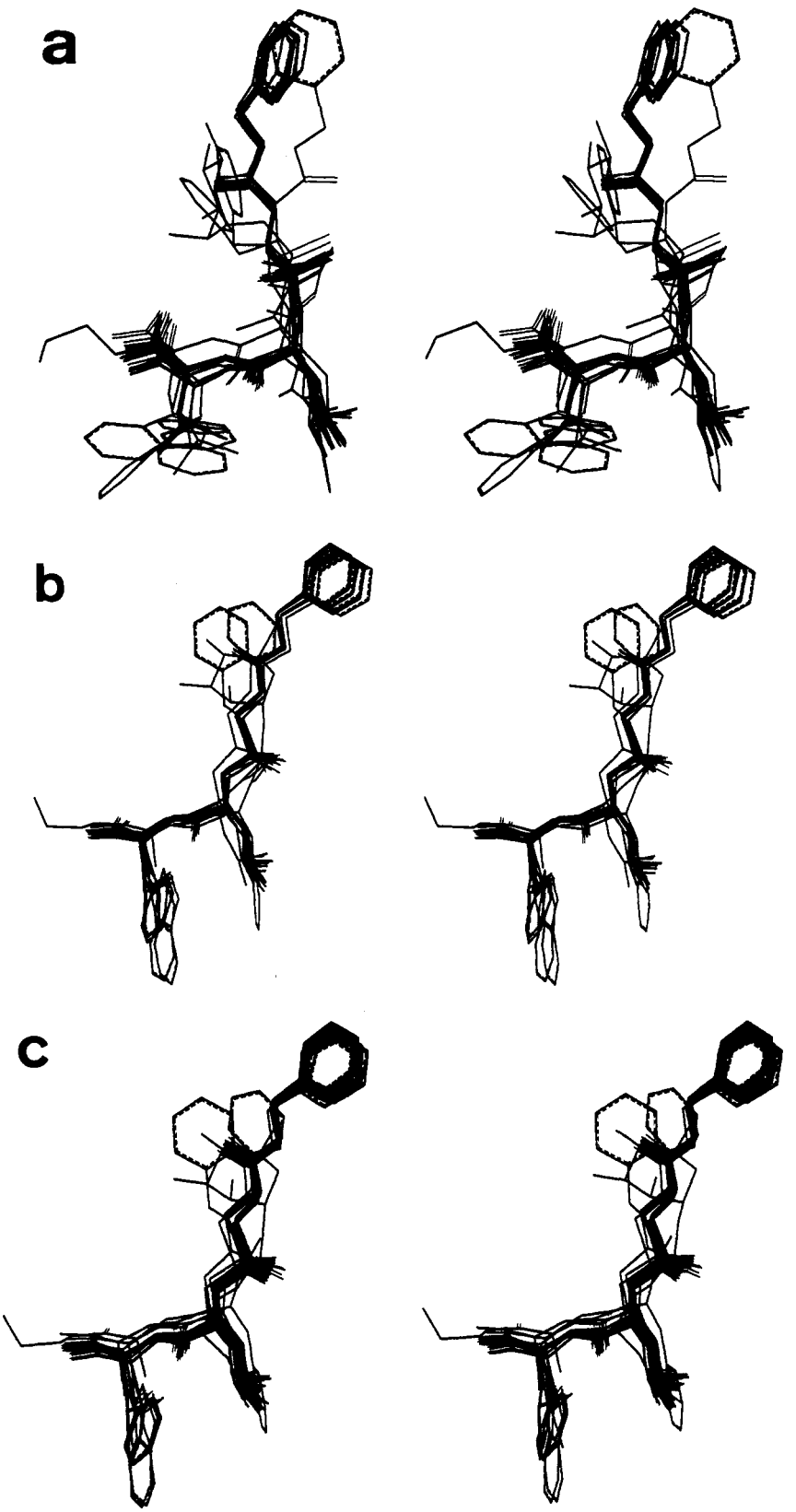


Figure 2. Stereodiagram of the superposition of 20 inhibitors of thermolysin in the conformations and relative alignment used in CoMFA (hydrogen atoms removed for clarity): (a) for **9–15** based on coordinates taken from X-ray crystal structure analysis[33-41,43] and for **16–28** constructed from the crystallographically obtained coordinates of **14** and **15**; (b) for all inhibitors based on a multiple fit; (c) based on a field fit.

mously but according to a study of Greco et al.[3] does not greatly improve the results; however, contrary experience has also been obtained.[21]) The analyses were performed either with no scaling of the field contributions or with a scaling according to CoMFA standard deviations, as suggested by the authors of the program.[22] Scaling influences the finally obtained r^2 values (see below); however, in most cases the better results are obtained by

(21) Folkers, G. Personal communication

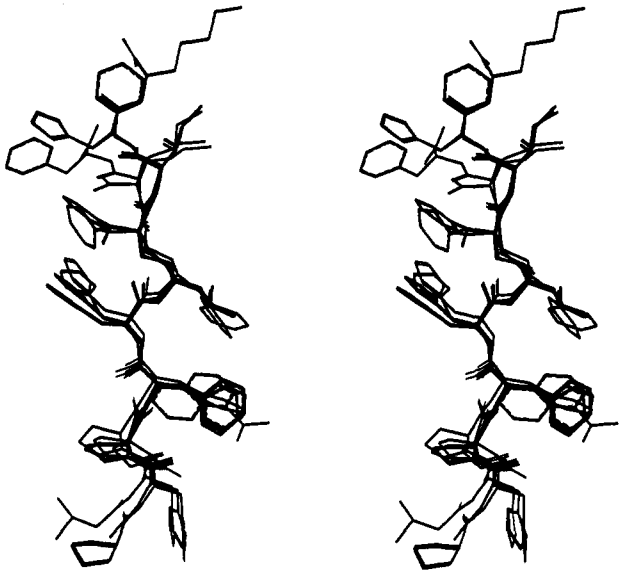


Figure 3. Stereodiagram of the superposition of 13 inhibitors of renin (**29–41**) in the conformations and relative alignment used in CoMFA (hydrogen atoms removed for clarity), based on coordinates taken from X-ray crystal structure analysis[16] of inhibitor/enzyme complexes of endothiapepsin.

Table IV. CoMFA-PLS Analyses of Eight Inhibitors of HRV14, Based on Two Alignments Crystallographically Confirmed (XS) and Calculated by Multiple Fit (MFIT)[a]

	rhinovirus ln (MIC)			
	XS		MFIT	
PLS analysis	30/30	5/30	30/30	5/30
r^2 cross	0.410	0.535	0.435	0.262
PRESS	2.324	1.459	1.609	1.644
r^2 nocross	1.000	0.983	0.931	0.799
S	0.044	0.280	0.561	0.859
no. comp	5	3	3	2
fraction				
steric	0.591	0.553	0.742	0.649
electrostatic	0.409	0.447	0.258	0.351
CoMFA				
init weights				
steric	0.007360	0.05388	0.01024	0.07827
electrostatic	0.003702	0.07353	0.10460	0.23380

[a] The cross-validated r^2_{cross} values and the explanatory $r^2_{nocross}$ values (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with regard to the number of components listed in the table (no. comp); the standard errors for both r^2 values (PRESS, S), the fraction of steric and electrostatic field contribution, and the initial weights in the analysis are listed.

the latter approach. Thus, the results listed in Tables IV, VI, and VIII refer to this scaling.

Partial Least-Squares Calculations

Six orthogonal latent variables were first extracted by the standard PLS algorithm[7] 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 (ln K, ΔG, ΔS, ΔH) to be a minimum from a leave-one-out cross-validation method. Comparing these values to the overall variance of the actual dependent property values yield a cross-validated r^2_{cross} (r^2_{PRESS}). The r^2_{cross} values listed for the different models are the maximal values which were obtained considering the number of components given in the tables. The number of components regarded corresponds to the highest cross-validated r^2 value obtained in each analysis. It has been checked to what extent the finally added component improved the cross-validated r^2 value. In all cases an improvement of more than 5% was achieved. In addition to the predictive r^2_{cross} values, the explanatory $r^2_{nocross}$ values (no cross-validation) are reported for the different analyses. In these determinations, the optimal number of components was regarded as revealed from the cross-validated analyses. For all examples, the standard errors are given (cross-validation PRESS, no cross-validation S, Tables IV, VI, and VIII–X).

Table V. Experimental Binding Constants for Eight Inhibitors of Human Rhinovirus 14, Expressed as Negative Logarithm of Minimal Inhibitory Concentration[32] and the Corresponding Values Predicted by CoMFA (5/30 truncation) for a Model Based on Crystallographic Evidence (XS) and on a Multiple-Fit Alignment (MFIT)[a]

	-ln (MIC)		
	exp	calcd (XS)	calcd (MFIT)
1	17.30	17.01	15.99
2	14.70	16.32	15.29
3	15.40	15.60	16.77
4	17.70	16.57	16.03
5	14.30	15.40	15.39
6	14.30	14.70	15.97
7	14.50	13.40	13.16
8	12.90	14.28	13.91
mean	15.14		
std dev	1.62		

[a] For the experimental data the mean value and the standard deviation are given.

Table VI. CoMFA-PLS Analyses of 20 Inhibitors of Thermolysin, Based on Three Different Alignments: Crystallographically Confirmed (XS); Calculated by Multiple Fit (MFIT); Calculated by Field Fit (FFIT)[a]

	thermolysin: ln K					
	XS		MFIT		FFIT	
PLS analysis	30/30	5/30	30/30	5/30	30/30	5/30
r^2_{cross}	0.254	0.434	0.505	0.725	0.513	0.775
PRESS	3.866	3.888	3.494	2.709	3.769	2.354
$r^2_{nocross}$	0.869	0.990	0.986	0.977	0.995	0.993
S	1.618	0.525	0.587	0.783	0.399	0.415
no. comp	3	7	6	7	8	6
fraction						
steric	0.468	0.576	0.498	0.549	0.401	0.660
electrostatic	0.532	0.424	0.502	0.451	0.599	0.340
CoMFA						
init weights						
steric	0.005120	0.047840	0.005482	0.05448	0.005710	0.05551
electrostatic	0.008614	0.009549	0.009218	0.01057	0.008927	0.01042

[a] The cross-validated r^2_{cross} values and the explanatory $r^2_{nocross}$ values (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with regard to the number of components listed in the table (no. comp); the standard errors for both r^2 values (PRESS, S), the fraction of steric and electrostatic field contribution, and the initial weights in the analysis are listed.

(22) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Toward its use with 3D-Structural Databases. *Tetrahedron Comput. Methods* **1990**, *3*, 47–59.

Table VII. Experimental Binding Constants for 20 Inhibitors of Thermolysin, Expressed by Their Negative Logarithms and the Corresponding Values Predicted by CoMFA (5/30 truncation) for a Model Based on Crystallographic Evidence (XS), on a Multiple-Fit Alignment (MFIT), and on a Field-Fit Alignment (FFIT)[a]

		-ln K				
		exp	XS	MFIT	FFIT	ref(s)
9	1TLP	17.39	15.86	16.17	17.53	33–35
10	1TMN	16.81	19.51	15.55	17.26	36,37
11	4TMN	23.41	13.81	18.18	17.73	38,39
12	4TLN	8.57	13.32	11.90	11.24	40
13	5TLN	14.66	14.39	12.47	12.71	40
14	5TMN-LEU	18.52	17.09	17.47	15.54	41, 42
15	6TMN-LEU	11.62	13.97	11.62	13.08	41, 42
16	5TMN-ALA	17.92	18.27	18.43	18.47	41, 42
17	5TMN-GLY	15.12	14.74	16.46	17.55	41, 42
18	5TMN-PHE	16.37	13.85	15.59	15.95	41, 42
19	5TMN-NH_2	14.10	10.30	13.54	14.02	41, 42
20	6TMN-ALA	11.25	10.24	10.16	12.45	41, 42
21	6TMN-GLY	8.38	9.62	9.55	9.21	41, 42
22	6TMN-PHE	9.85	10.59	10.20	11.48	41, 42
23	6TMN-NH_2	7.32	11.21	11.80	8.04	41, 42
24	PCH_2-ALA	17.81	17.76	17.99	17.30	15, 43
25	PCH_2-GLY	15.02	16.19	14.32	14.90	15, 43
26	PCH_2-PHE	16.53	18.46	16.71	14.70	15, 43
27	PCH_2-NH_2	13.48	11.25	9.66	12.73	15, 43
28	PCH_2-LEU	18.36	18.50	18.34	17.05	15, 43
	mean	14.62				
	std dev	4.11				

[a] For the experimental data the mean value and the standard deviation are given.

Binding Properties

In the case of HRV14 and thermolysin, the negative natural logarithms of the binding constant were used as dependent property values. The binding constants were taken from the literature (see Tables V and VII). For the renin example, thermodynamic data have been recorded by Epps et al.[23] These data were used as dependent variables in CoMFA (Table XI). The mean values of the dependent property value and the corresponding standard deviations (Tables V, VII, and XI) are given for each data set.

Description of the Case-Studies

Antiviral Compounds of Human Rhinovirus 14. Eight antiviral compounds (1–8) considered in this analysis are derived from two basic skeletons distinguished by the length of the aliphatic chain ($n = 5, 7$) and the substitution pattern at the oxazoline or phenyl ring, respectively (Table I). As indicated by X-ray crystallography, 1–4 bind with the oxazoline moiety toward the entrance of the binding pocket while 5–8 occur in reverse orientation with the isoxazole ring toward the entrance. The relative alignment of 1–8, as found crystallographically, was used in CoMFA. The obtained cross-validated r^2 values indicate a moderate prediction of the logarithm of the minimal inhibitory concentration (Tables IV and V). Very recently, a study on a similar set of antivirial compounds of HRV14, aligned according to crystallographic evidence, has been reported.[24] These results are difficult to compare with the present study, since some of the considered compounds have a different composition. An r^2 value of 0.728 is given. This r^2 value is not a predictive (cross-validated) r^2 but an explanatory one (no cross-validation). It has to be compared to the $r^2_{nocross}$ values of 1.000 and 0.983 in the present analysis (Table IV, XS alignment). Furthermore, the grid dimensions used by Diana et al.[24] are different from those used in the present study.[25]

In order to test whether CoMFA is able to render the experimentally confirmed alignment prominent to other hypothetically derived alignments, 1–8 were mapped upon each other according to a multiple-fit procedure. Since reverse binding of some of the inhibitors would be hard to predict without crystallographic evidence, all molecules were fitted with the oxazoline ring to one side and the isoxazole ring to the opposite side. Ten fit centers were distributed over each of the molecules (three in each heterocycle, two in the phenyl moiety, and at the two termini of the chain; the various centers assigned in this calculation are indicated in Table I). The alignment obtained, which only comprises two basic conformations (Figure 1b), was subjected to CoMFA. The conformations used in this alignment differ from those based on the protein data by root-mean-square deviations (for all non-hydrogen atoms) of 0.9–1.5 Å, which shows that not only the relative alignment but also the individual conformations of the molecules are different in both alignments.

Inhibitors of Thermolysin. Compared to the previous example, the structural variance of the inhibitors in this data set is appreciably larger. For seven inhibitors (9–15), the active site conformation and the relative alignment were taken from crystal structure analysis, 13 additional inhibitors (16–28) belonging to three congeneric series were constructed according to structural data of 14 and 15 (Table II). The relative alignment which is assumed to be as close as possible to the actual binding geometry is shown in Figure 2a. Applying CoMFA to this set of inhibitors reveals r^2 values which show that the prediction error is still smaller than the variance of the actual dependent property values (ln K, Tables VI and VII).

Similar to the HRV14 case, additional alignments were tested. A multiple fit was calculated with MOMO. Since the implemented algorithm allows for a varying number of "fit centers" in the inhibitors of different size, the following atoms were specified, if present, for the fit. In all molecules, the group coordinating to Zn and the H-bond partners for ARG203 and ASN112 of thermolysin were matched (Table II, see indicated fit centers z1, z2, z4, z5, z8, and z9). For the hydrophobic portions, in position P1′ (*i*-Pr, Ph, z3, z4), P1 ($PhCH_2CH_2$, $PhCH_2O$, sugar moiety, z7, and z12), and P2′ (-CH_2(indole), CH_3, *i*-Bu, CH_2Ph, z10, and z11) appropriate atoms were selected to allow for a superposition of these moieties. The achieved alignment, shown in Figure 2b, is *one* possible solution out of many such matching procedures. However, it follows alignment rules currently used in rational drug design for systems where no information on the geometry of the protein binding site is known. The conformations of the individual inhibitors obtained by the present multiple fit deviate substantially from those considered in the previous alignment based on experimental evidence (root-mean-

(23) Epps, D. E.; Cheney, J.; Schostarez, H.; Sawyer, T. K.; Prairie, M.; Krueger, W. C.; Mandel, F. Thermodynamics of the Interaction of Inhibitors with the Binding Site of Recombinant Human Renin. *J. Med. Chem.* **1990**, *33*, 2080–2086.

(24) Diana, G. D.; Kowalczyk, P.; Treasurywala, A. M.; Oglesby, R. C.; Pevear, D. C.; Dutko, F. J. CoMFA Analysis of the Interactions of Antipicornavirus Compounds in the Binding Pocket of Human Rhinovirus-14. *J. Med. Chem.* **1992**, *35*, 1002–1008.

(25) The grid dimensions in the present study are $x_{min}/x_{max} = -6.8, 17.0$; $y_{min}/y_{max} = -17.3, 8.3$; $z_{min}/z_{max} = -13.6, 2.9$ Å; those in ref 24 are −8, 8; −8, 2; −6, 5 Å. To our experience, the grid dimensions take influence on the finally obtained r^2 values.

Table VIII. CoMFA-PLS Analyses of ΔG for 13, 10 (without **36**, **38**, and **39**), and 12 (without **36** and data for **38**/**39** Interchanged) Inhibitors of Renin, Based on an Alignment Derived from Crystallographically Confirmed Complexes of Related Inhibitor/Enzyme Complexes of Endothiapepsin[a]

	renin: ΔG					
	13 compounds		10 compounds		12 compounds	
PLS analysis	30/30	5/30	30/30	5/30	30/30	5/30
r^2_{cross}	−0.329	−0.198	−0.479	−0.273	−0.565	−0.498
PRESS	1.962	1.863	2.530	2.347	2.276	2.227
$r^2_{nocross}$	0.848	0.859	0.899	0.916	0.826	0.898
S	0.663	0.638	0.660	0.604	0.758	0.582
no. comp	4	4	4	4	4	4
fraction						
steric	0.700	0.731	0.727	0.776	0.681	0.773
electrostatic	0.300	0.269	0.273	0.224	0.319	0.227
CoMFA						
init weights						
steric	0.004114	0.03292	0.004065	0.03008	0.004085	0.03240
electrostatic	0.064140	0.08376	0.052270	0.06972	0.059870	0.07834

[a] The cross-validated r^2_{cross} values and the explanatory $r^2_{nocross}$ values (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with regard to the number of components listed in the table (no comp); the standard errors for both r^2 values (PRESS, S), the fraction of steric and electrostatic field contribution, and the initial weights in the analysis are listed.

Table IX. CoMFA-PLS Analyses of ΔH for 13, 10 (without **36**, **38**, and **39**), and 12 (without **36** and data for **38**/**39** Interchanged) Inhibitors of Renin, Based on an Alignment Derived from Crystallographically Confirmed Complexes of Related Inhibitor/Enzyme Complexes of Endothiapepsin[a]

	renin: ΔH					
	13 compounds		10 compounds		12 compounds	
PLS analysis	30/30	5/30	30/30	5/30	30/30	5/30
r^2_{cross}	0.261	0.263	0.687	0.795	0.779	0.835
PRESS	6.203	6.195	6.722	4.217	4.727	3.731
$r^2_{nocross}$	0.476	0.489	0.995	0.987	0.993	0.991
S	5.222	5.157	0.862	1.078	0.818	0.868
no. comp	1	1	6	4	6	5
fraction						
steric	0.302	0.360	0.749	0.803	0.751	0.781
electrostatic	0.698	0.640	0.251	0.197	0.249	0.219
CoMFA						
init weights						
steric	0.004114	0.03292	0.004065	0.03008	0.004085	0.03240
electrostatic	0.064140	0.08376	0.052270	0.06972	0.059870	0.07834

[a] The cross-validated r^2_{cross} values and the explanatory $r^2_{nocross}$ values (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with regard to the number of components listed in the table (no. comp); the standard errors for both r^2 values (PRESS, S), the fraction of steric and electrostatic field contribution, and the initial weights in the analysis are listed.

square deviations between 0.89 and 2.8 Å). In this geometry the inhibitors cannot be accommodated to the binding site of thermolysin, at least if the structure of the binding site is considered as found in the crystal.

This hypothetical alignment allows for a substantially improved prediction (Tables VI and VII), compared with the previous described alignment based on structural data, which should be the superior model for the geometry actually adopted at the enzyme binding site. Applying the field-fit option as implemented in SYBYL on the data set aligned by the multiple-fit procedure reveals a superposition which enables an even better prediction (Tables VI and VII).

Inhibitors of Renin. A data set of renin inhibitors (Table III[23]) has been analyzed for which ΔG, ΔH, and ΔS were experimentally determined. Since for these compounds no crystallographic studies were performed in parallel, the active site conformations and the relative alignment had to be constructed on the basis of the binding geometries of closely related inhibitors of endothiapepsin, a closely related aspartyl protease. (Reference to the corresponding inhibitor/endothiapepsin complexes[16] used for modeling are given in Table III by their PDB codes; each constructed inhibitor has been optimized by molecular mechanics regarding the protein environment during the minimization as rigid aggregate.) From the design of renin inhibitors, it is known that binding geometry is extensively conserved in the active site.[16,26] The obtained binding geometry and the alignment of 13 renin inhibitors (Table III) is shown in Figure 3.

On the basis of this alignment, CoMFA allows no prediction of ΔG, the negative r^2 indicates that the prediction errors are greater than the overall variance of ΔG (Table VIII, first row, and Table XI). This clearly shows that for the present alignment no CoMFA model for ΔG can be established. Nevertheless, for ΔH (and also for ΔS, since ΔH and ΔS change with opposite sign, see Discussion) a cross-validated r^2 is obtained with this alignment which allows a prediction better than the total variance (Tables IX–XI). However, it has to be remembered that the total variance of ΔH within the data set is

(26) Sali, A.; Veerapandian, B.; Cooper, J. B.; Moss, D. S.; Hofmann, T.; Blundell, T. L. Domain Flexibility in Aspartic Proteinases. *Proteins* **1992**, *12*, 158–170.

Table X. CoMFA-PLS Analyses of ΔS for 13, 10 (without 36, 38, and 39), and 12 (without 36 and data for 38/39 Interchanged) Inhibitors of Renin, Based on an Alignment Derived from Crystallographically Confirmed Complexes of Related Inhibitor/Enzyme Complexes of Endothiapepsin[a]

	renin: ΔS					
	13 compounds		10 compounds		12 compounds	
PLS analysis	30/30	5/30	30/30	5/30	30/30	5/30
r^2_{cross}	0.178	0.196	0.640	0.807	0.733	0.844
PRESS	23.231	22.975	26.190	14.871	18.672	13.043
$r^2_{nocross}$	0.434	0.460	0.995	0.984	0.989	0.988
S	19.282	18.841	3.017	4.277	3.723	3.657
no. comp	1	1	6	4	6	5
fraction						
steric	0.312	0.370	0.780	0.835	0.771	0.831
electrostatic	0.668	0.630	0.220	0.165	0.229	0.169
CoMFA						
init weights						
steric	0.004114	0.03292	0.004065	0.03008	0.004085	0.03240
electrostatic	0.064140	0.08376	0.052270	0.06972	0.059870	0.07834

[a] The cross-validated r^2_{cross} values and the explanatory $r^2_{nocross}$ values (no cross-validation) are given for a 30/30 and 5/30 truncation of the steric and electrostatic field contributions. These values were obtained with regard to the number of components listed in the table (no. comp); the standard errors for both r^2 values (PRESS, S), the fraction of steric and electrostatic field contribution, and the initial weights in the analysis are listed.

Table XI. Experimental Thermodynamic Data (ΔG, ΔH, and ΔS)[23] for 12 (without 36, data for 38/39 interchanged) and 10 (without 36, 38, 39, values in Parentheses) Inhibitors of Renin and the Corresponding Values Predicted by CoMFA (5/30 Truncation) for a Model Derived from the Crystallographically Confirmed Binding Geometry of Related Inhibitors of Endothiapepsin[a]

	ΔG		ΔH		ΔS	
	exp	pre	exp	pre	exp	pre
29	8.60	9.63 (9.83)	14.45	13.16 (13.23)	74.20	75.10 (75.51)
30	9.20	11.36 (10.91)	14.28	16.78 (18.05)	75.70	89.51 (93.65)
31	11.50	12.20 (12.66)	28.75	25.50 (24.48)	131.10	118.77 (119.08)
32	12.40	10.75 (10.59)	20.33	21.68 (21.89)	105.50	103.80 (105.34)
33	11.20	8.63 (8.76)	33.56	29.13 (29.16)	144.30	126.96 (127.45)
34	8.60	11.30 (11.24)	29.35	33.22 (33.45)	127.40	143.51 (143.76)
35	12.20	10.34 (10.85)	26.70	27.09 (27.13)	125.30	123.40 (123.61)
37	9.40	9.56 (9.44)	14.20	13.05 (13.47)	76.30	74.80 (71.47)
38	9.90	8.00	14.69	15.70	79.60	77.12
39	9.50	11.43	26.10	24.13	114.80	115.38
40	10.80	10.07 (11.98)	22.63	24.70 (21.85)	108.00	114.04 (109.00)
41	12.40	11.65 (10.86)	21.36	25.60 (25.23)	108.90	117.00 (116.54)
mean	10.48		21.55		103.81	
std dev	1.39		6.91		24.54	

[a] For the experimental data the mean values and the standard deviations are given.

about five times larger than that of ΔG, while the remaining residuals for both properties are in the same range.

A more detailed analysis of the residuals reveals that three compounds in the set are poorly predicted (36, 38, and 39). Leaving out these three compounds produces a substantial improvement of the model For ΔH and ΔS (Tables IX and X) an excellent prediction is suggested, whereas ΔG is still not predictable (negative r^2).

Dropping compounds from an analysis as "outliers" must be justified. From the rational design of renin inhibitors, it is known that a residue at P4 (usually a PRO), capable of functioning as a hydrogen-bond acceptor, helps to increase binding.[27] In a congeneric series, inhibitors missing this residue normally show reduced binding affinity. The pair 30/31 shows this expected tendency and the increase of ΔH might be attributed to the formation of an essential hydrogen bond.[28] Surprisingly, the comparable pair 38/39 shows an opposite trend. CoMFA predictions reveal very high residuals for these two compounds. The possibility exists that data for these two compounds might have been exchanged by accident.[29] A CoMFA with the thermodynamic data of 38/39 interchanged reveals a model even better than that based on only 10 inhibitors of the original data set (Table VIII, third row). However, 36 is still unsatisfactorily described in this model. A prediction of the thermodynamic data of this compound (experimental values: ΔG = 10.60 kJ/mol, ΔH = 13.70 kJ/mol, ΔS = 78.40 J/K mol) based on the no cross-validated analysis with 12 and 10 compounds (in parentheses) reveals a ΔG = 11.45 (11.70) kJ/mol, ΔH = 26.60 (26.12) kJ/mol, and ΔS = 122.76 (121.68) J/K mol.

Independent of this possible exchange, the present study gives evidence that the analysis of steric and electrostatic field contributions seems to be capable of predicting the enthalpic contribution to the free enthalpy of binding, whereas the free enthalpy itself is predicted unsatisfactorily.

Discussion and Conclusions

Three data sets of protein inhibitors were evaluated by CoMFA to describe binding properties in a quantitative

(27) Haber, E.; Hui, K. Y.; Carlson, W. D.; Bernatowicz, M. S. Renin Inhibitors: A Search for Principles of Design. *J. Cardiovas. Pharm.* 1987, *10*, S54–S58.

(28) Klebe, G. Structure Correlation and Ligand/Receptor Interactions. In *Structure Correlation*; ed. Bürgi, H. B., Dunitz, J. D., Eds., VCH Weinheim: Weinheim, in press.

(29) Epps, D. E. Personal communication.

way. In most of the analyses, the models with a truncation of the steric field at 5 kcal/mol revealed better predictions. Alignments based on different approaches and assumptions were tested for the three data sets, among them for each case one, which should resemble—to the best knowledge—the actual binding geometry at the protein binding site. In the HRV14 example, the alignment based on crystallographic evidence and a hypothetical alignment with all molecules oriented similarly (which is clearly false, when judged on the basis of the crystallographic results) revealed comparable predictions. However, the set of compounds is quite homogeneous and since the differences in binding affinity are mainly attributed to substituent effects, these differences appear to be regarded comparably in both alignments.

For thermolysin, the alignment, evident from experimental data, permits a substantially less satisfactory prediction than two hypothetical alignments which differ positionally and conformationally from the former one. Binding to the active site appears unlikely in these geometries, at least accomodation to the active site in its crystallographically determined geometry can be excluded. The thermolysin example underlines the view that the alignment is the most important input variable in CoMFA. However, it is surprising that hypothetically derived alignments based on reasonable but clearly arbitrary assumptions allow a better prediction.

Different reasons might be invoked to explain these strange and totally unexpected observations. One aspect might be, that the data set as used, the data preparation, and the parameters selected for the analysis are inappropriate. This question can only be answered if additional studies on other systems are communicated which help to provide additional experience for a correct selection.

Another, perhaps more fundamental, aspect has to be considered. The CoMFA approach calculates interaction *energies* of aligned molecules with common surrounding steric and electrostatic fields. This should allow one to determine a total interaction *energy*. In a favorable case, this energy correlates directly with the interaction energy of the ligand and receptor. However, in the present analyses, a prediction of a dependent property (ln K) is attempted which comprises the *free enthalpy of binding* ΔG.[9] Thus, the interaction energy (or enthalpy) should solely correlate with the enthalpic contribution to ΔG. A prediction of ΔG (or ln K) would only be possible if the entropic contribution to ΔG is negligible or constant over the entire data set.

Several experimental studies have shown (for reviews see ref 30) that this assumption is by no means justified for drug/receptor binding. No clear-cut correlation exists between ΔG and ΔH, even in congeneric series of drugs (cf. renin case). However, ΔH and ΔS change with opposite sign, not due to a physical relationship but simply because ΔG varies over a rather small range and ΔH or $-T\Delta S$ cover a much broader range.[28]

In order to find some evidence for this assumption, a data set of renin-inhibitors has been analyzed for which thermodynamic data are available and information on the possible binding geometries at the enzyme active site can be estimated from crystallographic data of related inhibitor complexes with endothiapepsin. The CoMFA analyses indicate that only enthalpies (ΔH) and not free enthalpies (ΔG) or binding constants can be properly predicted.

Some conclusions can be drawn. CoMFA is quite sensitive to changes in the alignment. It might well be that binding geometries, as they are known from the limited resolution of protein crystal structure determinations, are still not precise enough to allow a successful comparative field analysis. Perhaps, the conformations of a binding site and of a bound inhibitor obtained from crystal-structure analysis are not sufficient to describe satisfactorily "molecular similarity" required for a CoMFA analysis. However, this fact would be rather discouraging, since protein crystallography is the only reliable source to learn about the binding geometry of drug molecules. The charges, as used in the analyses, might be inadequate to describe sufficiently the electrostatic field properties. However, experience[3] shows that this aspect appears not to be of dominant influence on the finally obtained CoMFA results. Assumptions about the protonation states of the various functional groups of the ligands (and the active site residues) might be false due to changes in the dielectric conditions induced through ligand binding.[28] These influences are very difficult to predict at present. The renin example gives some evidence that the CoMFA approach is only appropriate to predict enthalpic contributions to the free enthalpy of binding. In this context, it has to be regarded that the total variance of ΔH is about 5 times larger than that of ΔG which might facilitate its predictability. However, this observation might also indicate that hypothetical alignments (cf. thermolysin) through multiple fit (minimizing the steric differences of molecules) or field fit (in addition to the latter minimizing the electrostatic field differences) which clearly deviate from the crystallographically evident alignment, reveal a better CoMFA prediction of the binding constant (and thus ΔG) simply because they compensate in a complex and nontransparent way for an inadequate description of the entropic contributions.[31] Molecules with different degrees of conformational flexibility (exhibiting a different

(30) Gilli, G., Borea, A. Structural Chemistry and Drug Design. In *The Application of Charge Density Research to Chemistry and Drug Design*; Jeffrey, G. A., Piniella, J. F., Eds.; Nato Adv. Study Inst.; Plenum Pub. Co.: New York, 1991; pp 241–286. Hitzemann, R. Thermodynamic aspects of drug-receptor interactions. *Trends Pharmacol. Sci.* 1988, *9*, 408–411.

(31) For example, a linear correlation between the free enthalpy of transfer from aqueous solution to an hydrophobic medium (which covers an important contribution to the entropic portion of ligand binding[28] and the (hydrophobic) molecular surface has been established. (For a recent discussion, see: Sharp, K. A.; Nicholis, A.; Fine, R. F.; Honig, B. Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* 1991, *252*, 106–109.) Any alignment which reduces the relative surface differences of the molecules in a data set thus minimizes—at least partially—this contribution to ΔG.

(32) Badger, J.; Minor, I.; Kremer, M. J.; Oliveira, M. A.; Smith, T. J.; Griffith, J. P.; Guerin, D. M. A.; Krishaswamy, S.; Luo, M.; Rossmann, M. G.; McKinlay, M. A.; Diana, G. D.; Dutko, F. J.; Fancher, M.; Rueckert, R. R.; Heinz, B. A. Structural Analysis of a Series of Antivirial Agents Complexed with Human Rhinovirus 14. *Proc. Natl. Acad. Sci. U.S.A.* 1988, *85*, 3304–3308.

(33) 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.

(34) Weaver, L. H.; Kester, W. R.; Matthews, B. W. A Crystallographic Study of the Complex of Phosphoramidon with Thermolysin. A Model for the Presumed Catalytic Transition State and for Binding of Extended Substrates. *J. Mol. Biol.* 1977, *114*, 119–132.

(35) Kester, W. R.; Matthews, B. W. Comparison of the Structures of Carbopeptidase A and Thermolysin. *J. Biol. Chem.* 1977, *252*, 7704–7710.

(36) Monzingo, A. F.; Matthews, B. W. Binding of N-Carboxymethyl Dipeptide Inhibitors to Thermolysin Determined by X-ray Crystallography: A Novel Class of Transition-State Analogues for Zinc Peptidases. *Biochemistry* 1984, *23*, 5724–5729.

(37) Hangauer, D. G.; Monzingo, A. F.; Matthews, B. W. An Interactive Computer Graphics Study of Thermolysin-Catalyzed Cleavage and Inhibition by N-Carboxymethyl Dipeptides. *Biochemistry* 1984, *23*, 5730–5741.

number of accessible conformations in the unbound state) will show different entropic contributions to ΔG once they are conformationally immobilized at the protein binding site. If this supposition is correct, the CoMFA approach has to be extended by some other method to cover the entropic contribution to the free enthalpy of binding. Furthermore, alignments where no reference to experimental data is possible have to be regarded with some caution, because deviations from the "correct" biologically active conformation and alignment could occur in order to compensate for an inadequate consideration of entropic effects in a comparative field analysis.

Acknowledgment. The authors are grateful to their collegues Hugo Kubinyi and Hans-Joachim Böhm at BASF for helpful discussions. Constructive comments of Gerd Folkers (Zürich) on the present manuscript are gratefully acknowledged.

(38) Holden, H. M.; Tronrud, D. E.; Monzingo, A. F.; Weaver, L. H.; Matthews, B. W. Slow and Fast-Binding Inhibitors of Thermolysin Display Different Modes of Binding: Crystallographic Analysis of Extended Phosphonamidate Transition-State Analogues. *Biochemistry* 1987, *26*, 8542–8553.

(39) Kester, W. R.; Matthews, B. W. Crystallographic Study of the Binding of Dipeptide Inhibitors to Thermolysin: Implications for the Mechanism of Catalysis. *Biochemistry* 1977, *16*, 2506–2511.

(40) Holmes, M. A.; Matthews, B. W. Binding of Hydroxamic Acid Inhibitors to Crystalline Thermolysin Suggests a Pentacoordinate Zinc Intermediate in Catalysis. *Biochemistry* 1981, *20*, 6912–6920.

(41) 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.

(42) 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.

(43) Matthews, B. W. Unpublished results.