3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors: A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active Site Geometries

Scott A. DePriest,† Dorica Mayer,‡ Christopher B. Naylor,§ and Garland R. Marshall*

Contribution from the Center for Molecular Design, Washington University, St. Louis, Missouri 63130-4899

Received March 23, 1992

Abstract: The ability of comparative molecular field analysis (CoMFA), a three-dimensional, quantitative structure–activity relationship (3-D QSAR) paradigm, to predict the activity of inhibitors of angiotensin-converting enzyme (ACE) and thermolysin was examined. Correlations derived from computationally and experimentally determined alignment rules were compared. The correlations derived for the ACE series using alignment rules determined from a systematic conformational search (Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. *J. Comput.-Aided Molec. Des.* **1987**, *1*, 3–16) were comparable to those derived for the thermolysin inhibitors using alignment rules defined by crystallographic data. Models derived from potential fields alone, however, were insufficient for accurately quantifying and predicting the nature of enzyme–inhibitor interactions. The predictive ability of the ACE model for a series of molecules not included in the training set was improved by the addition of a zinc indicator variable which explicitly defined the nature of the zinc–ligand interaction, an effect not observed within the thermolysin series. The effects of additional parameters, such as torsional degrees of freedom and the change in conformational enthalpy, $\Delta H_{conform} = H_{aligned} - H_{min}$, were also examined. Experimentally derived alignment rules based on known structures of three-dimensional complexes produced predictive correlations for thermolysin inhibitors comparable, but not superior, to the correlations for ACE inhibitors based on alignment rules which were computationally deduced. The use of the active analog approach to determine active site geometries in the absence of structural data on the receptor is strongly supported by these results. Additionally, the correlations indicate that 3-D QSARs based on alignment rules derived from structure–activity data alone can produce statistically significant predictive correlations for quite diverse, noncongeneric compounds.

Introduction

Traditional QSAR applications, based on concepts from physical organic chemistry, have long attempted to correlate biological activity with measurable physicochemical parameters such as Hansch's log P, Hammett's s, Taft's E_s, and MR.[1] The literature is filled with accounts of successful applications of these parameters to the development of robust QSARs for congeneric series.[2-6] Advancements in computer technology and data analysis have allowed us to extend QSAR parameters to the level of 3-D properties of the molecules of interest. Current 3-D QSAR methods include molecular-shape analysis,[7] the hypothetical active site lattice (HASL),[8] the RECEPS programs,[9] Crippen's distance geometry[4] and Voronoi binding sites,[10] and comparative molecular field analysis (CoMFA).[11,12]

Angiotensin-converting enzyme (ACE) is a zinc-containing metallopeptidase which catalyses the hydrolysis of the C-terminal dipeptide His-Leu from the decapeptide angiotensin I to produce the octapeptide angiotensin II, a potent vasoconstrictor. Inhibitors of ACE, such as captopril, enalapril, and lisinopril, are widely prescribed to control essential hypertension.[13] Although the primary amino acid sequence of ACE is known,[14-16] its 3-D structure is still undetermined. What is known about the structural requirements for ACE inhibition has been derived from a plethora of SAR studies.[17-19] These studies, combined with crystallographic data from the analogous enzyme thermolysin and its inhibitors, indicate that the requirements for binding to ACE are (1) a C-terminal carboxyl group for ionic binding to a positively charged group on the enzyme; (2) a carbonyl oxygen which hydrogen bonds to some active site residue X–H; and (3) some zinc-binding functional group such as a carboxylate, hydroxamate, phosphonate, or thiolate.[19] This structural information defines the minimal set of active site groups necessary for ACE inhibition and has been used to analyze databases of di-

* To whom correspondence should be addressed.
† Current address: TRIPOS Associates, Inc., 1699 S. Hanley Road, Suite 303, St. Louis, MO 63144.
‡ Current address: Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, U.K.
§ Current address: SmithKline Beecham Research Centre, Cold Harbour Road, The Pinnacles, Harlow, Essex CM 19 5AD, U.K.

(1) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1979.
(2) Hansch, C.; Klein, T. E. *Acc. Chem. Res.* **1986**, *19*, 392–400.
(3) Clare, B. W. *J. Med. Chem.* **1990**, *33*, 687–702.
(4) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333–346.
(5) Gould, K. J.; Manners, C. N.; Payling, D. W.; Suschintzky, J. L.; Wells, E. J. *J. Med. Chem.* **1988**, *31*, 1445–1453.
(6) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *Mathl Comput. Modelling* **1990**, *14*, 505–510.
(7) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
(8) Doweyko, A. M. *J. Med. Chem.* **1988**, *31*, 1396–1406.
(9) Kato, Y.; Itai, A.; Iitake, Y. *Tetrahedron* **1987**, *43*, 5229–5236.
(10) Boulu, L. G.; Crippen, G. M. *J. Comput. Chem.* **1989**, *10*, 673–682.
(11) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. *Tetrahedron Comput. Method* **1990**, *3*, 47–59.
(12) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(13) Cushman, D. W.; Ondetti, M. A. *Hypertension* **1991**, *17*, 589–592.
(14) Bunning, P.; Kleemann, S. G.; Riordan, J. F. *Biochemistry* **1990**, *29*, 10488–10492.
(15) Chen, Y.-N. P.; Riordan, J. F. *Biochemistry* **1990**, *29*, 10493–10498.
(16) Soubrier, F.; Alhenc-Gelas, F.; Hubert, C.; Allegrini, J.; John, M.; Tregear, G.; Corvol, P. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9386–9390.
(17) Petrillo, W. W.; Trippodo, N. C.; DeForrest, J. M. In *Annual Reports in Medicinal Chemistry*; Robertson, D. W., Ed.; Academic: New York, 1989; Vol. 25, pp 51–60.
(18) Hangauer, d. G. In *Computer-Aided Drug Design: Methods and Applications*; Perun, T. J., Propst, C. L., Eds.; Marcel Dekker: New York, 1989; pp 253–295.
(19) Wyvratt, M. J.; Patchett, A. A. *Med. Res. Rev.* **1985**, *5*, 483–531.

0002-7863/93/1515-5372$04.00/0

verse structural classes of ACE inhibitors to determine a common three-dimensional geometry for the active site consistent with their activity.[20-22]

The CoMFA methodology of 3D-QSAR is based on the assumption that the interactions between a ligand (inhibitor) and its receptor (enzyme) are primarily noncovalent in nature and shape-dependent. Therefore, a QSAR can be derived by sampling the electrostatic and steric fields surrounding a set of inhibitors and correlating the differences in those fields to biological activity. This methodology has been successfully applied to studies of inverse agonism of the benzodiazepine receptor,[23] prediction of electronic effects of substituted benzoic acids,[24] pK_a's of substituted imidazoles,[25] various biological activities of clodronic acid esters,[26] and muscarinic agonists[27] to name a few. We have applied the CoMFA methodology to a series of 68 ACE inhibitors representing 28 different chemical classes. Using the active site geometry determined by Mayer et al.,[21] we derived a CoMFA model with a statistically significant crossvalidated R^2 and considerable predictive ability for inhibitors outside of the training set.

CoMFA is a shape-dependent parameter;[12] therefore, the calculated field values are highly dependent on the conformation of the considered molecules and their relative orientations (the "alignment rule" [11,28]). Since the geometry of the ACE inhibitors was determined computationally rather than experimentally, we decided to calibrate the results of the ACE series against a series of molecules for which there is crystallographic data to explicitly define the active site geometry and the resulting alignment rules. Thermolysin inhibitors were chosen for several reasons: (1) the structures of native thermolysin and a variety of bound inhibitors are available; (2) there are a number of inhibitors of thermolysin available from the literature; and (3) thermolysin is also a zinc-containing metallopeptidase and numerous similarities between thermolysin and ACE have been proposed.[18]

Methods

A. General Methods. All molecular modeling and CoMFA analyses were done using SYBYL[29] versions 5.32 and 5.41 running on Silicon Graphics Iris 4D/80 and 4D/380, respectively. Molecular coordinates for the alignment rule were determined by a least-squares fitting procedure described below. The steric and electrostatic interactions for the CoMFA analyses were calculated using a volume dependent lattice with a 2-Å step size, an sp^3 hybridized carbon probe atom carrying a charge of +1.0, and a distance-dependent dielectric constant ($1/r$). The CoMFA lattice for the ACE series was 25 × 24 × 19 Å (X = −10 to 15, Y = −15 to 9, Z = −8 to 11) with 1560 points. The lattice for the thermolysin series was 26 × 26.5 × 22.5 Å (X = −9.4 to 17.4, Y = −16.5 to 10, Z = −12 to 10.5) with 2352 points. The cutoff value for both the steric and electrostatic interactions was set to +30 kcal/mol. Charges were calculated using the Gasteiger and Marsili method.[30] In all cases, the zinc-binding functionalities were considered to be ionized,[31] so the partial charges were calculated using formal charges of −0.5 on each of the C-terminal carboxylate oxygens and −1.0 on the zinc-ligand sulfur or oxygen. Columns in the data table for which the standard deviation was less than 0.05 were dropped. Correlations were derived using the method of partial least squares (PLS)[32] with principal component analysis (PCA)[33] and crossvalidated to reduce the probability of obtaining chance correlations.

As used in this paper, the cross-validated R^2 refers to the squared correlation coefficient of the equation derived from the cross-validation of the training set to determine the optimum number of principal components. The conventional R^2 is the fitted correlation of the training set using the optimum number of principal components with no crossvalidation. The predictive R^2 refers only to test molecules not included in the training set and was calculated analogous to the definition for the conventional R^2 by Cramer et al.,[12] as

$$\text{predictive } R^2 = (\text{SD} - \text{press})/\text{SD}$$

where SD is the sum of squared deviations of each biological property value (the pIC_{50}) for each molecule in the test set from the mean of the training set, and press is the sum, over all molecules in the test set, of the squared differences between their actual and predicted biological property values. This is not simply an rms fit to the line corresponding to $R^2 = 1$, but rather a measure of the predictive ability of the QSAR model relative to the "best guess" case of using the mean biological property of the training set, where R^2 would be equal to 0. An $R^2 < 0$ indicates that the model predicts activities with larger residuals than one would obtain by using the mean activity of the training set as the prediction for each member of the test set.

B. CoMFA of ACE Inhibitors. A training set of 68 molecules (Table I; 2D structures and coordinates are available as supplementary material) was selected from the literature,[19,31,34-48] representing the diversity of structures and activities in the original

(20) DePriest, S. A.; Shands, R. F. B.; Dammkeohler, R. A.; Marshall, G. R. In *QSAR: Rational Approaches to the Design of Bioactive Compounds*; Silipo, C., Vittoria, A., Eds.; Elsevier Science: Amsterdam, 1991; Vol. 16, pp 405–414.
(21) Mayer, D.; Naylor, C. B.; Motoc, I.; Marshall, G. R. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 3–16.
(22) Andrews, P. R.; Carson, J. M.; Caselli, A.; Spark, M. J.; Woods, R. *J. Med. Chem.* **1985**, *28*, 393–399.
(23) Allen, M. S.; Tan, Y.-C.; Trudell, M. L.; Narayanan, K.; Schindler, L. R.; Martin, M. J.; Schultz, C.; Hagen, T. J.; Koehler, K.; Codding, P. W.; Skolnick, P.; Cook, J. M. *J. Med. Chem.* **1990**, *33*, 2343–2357.
(24) Kim, K. H.; Martin, Y. C. *J. Org. Chem.* **1991**, *56*, 2736–2729.
(25) Kim, K. H.; Martin, Y. C. *J. Med. Chem.* **1991**, *34*, 2056–2060.
(26) Bjorkroth, J.-P.; Pakkanen, T. A.; Lindroos, J.; Pohjala, E.; Hanhijarvi, H.; Lauren, L.; Hannuniemi, R.; Juhakoski, A.; Kippo, K.; Kleimola, T. *J. Med. Chem.* **1991**, *34*, 2338–2343.
(27) Greco, G.; Novellino, E.; Silipo, C.; Vittoria, A. *Quant. Struct.-Act. Relat.* **1991**, *10*, 289–299.
(28) Marshall, G. R.; Cramer, R. D., III *Trends Pharmacol. Sci.* **1988**, *9*, 285–289.
(29) Tripos Associates, Inc.: St. Louis, Missouri, USA.
(30) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219–3228.
(31) Saunders, M. R.; Tute, M. S.; Webb, G. A. *J. Comput.-Aided Molec. Des.* **1987**, *1*, 133–142.
(32) Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1–17.
(33) Wold, S.; Esbensen, K.; Geladi, P. *Chemometr. Intell. Lab. Sys.* **1987**, *2*, 37–52.
(34) Hassal, C. H.; Krohn, A.; Moody, C. J.; Thomas, W. A. *FEBS Lett.* **1982**, *147*, 175–179.
(35) Hassal, C. H.; Krohn, A.; Moody, C. J.; Thomas, W. A. *J. Chem. Soc., Perkin Trans.* **1984**, *1*, 155–164.
(36) Kim, D. H.; Guinosso, C. J.; Buxby, G. C.; Herbst, D. R.; McCaully, R. J.; Wicks, T. C.; Wendt, R. L. *J. Med. Chem.* **1983**, *26*, 394–403.
(37) Galardy, R. E.; Kontoyiannidou-Ostrem, V.; Kartylewicz, Z. P. *Biochemistry* **1983**, *22*, 1990–1995.
(38) Cushman, D. W.; Cheung, H. S.; Sabo, E. F.; Ondetti, M. A. *Biochemistry* **1977**, *16*, 5484–5491.
(39) Condon, M. E.; Petrillo, E. W.; Ryondo, D. E.; Reid, J. A.; Neubeck, R.; Puar, M.; Heikes, J. E.; Sabo, E. F.; Losee, K. A.; Cushman, D. W.; Ondetti, M. A. *J. Med. Chem.* **1982**, *25*, 250–258.
(40) Ondetti, M. A.; Cushman, D. W. *CRC Crit. Rev. Biochem.* **1984**, *16*, 381–411.
(41) Patchett, A. A.; Harris, E.; Tristram, E. W.; Wu, M. T.; Taub, D.; Peterson, E. R.; Ikeler, T. J.; ten Broeke, J.; Payne, L. G.; Ondeyka, D. L.; Thorsett, E. D.; Greenlee, W. J.; Lohr, N. S.; Hoffsommer, R. D.; Joshua, H.; Ruyle, W. V.; Rothrock, J. W.; Aster, S. D.; Maycock, A. L.; Robinson, F. M.; Hirschmann, R.; Sweet, C. S.; Ulm, E. H.; Gross, D. M.; Vassil, T. C.; Stone, C. A. *Nature* **1980**, *288*, 280–283.
(42) Harris, R. B.; Ohlsson, J. T.; Wilson, I. B. *Arch. Biochem. Biophys.* **1981**, *206*, 105–112.
(43) Ryan, J. W.; Chung, A. *Adv. Exp. Med. Biol.* **1983**, *156B*, 1133–1139.
(44) Thorsett, E. D.; Harris, E. E.; Aster, S.; Patchett, A. A.; E. H., U.; Vassil, T. C. *Biochem. Biophys. Res. Commun.* **1983**, *111*, 166–171.
(45) Watthey, J. W. H.; Gavin, T.; Desai, M. *J. Med. Chem.* **1984**, *27*, 816–818.
(46) Attwood, M. R.; Francis, R. J.; Hassall, C. H.; Krohn, A.; Lawton, G.; Natoff, I. L.; Nixon, J. S.; Redshaw, S.; Thomas, W. A. *FEBS Lett.* **1984**, *165*, 201–206.
(47) Weller, H. N.; Gordon, E. M.; Rom, M. B.; Pluscec, J. *Biochem. Biophys. Res. Commun.* **1984**, *125*, 82–89.
(48) Natarajan, S.; Gordon, E. M.; Sabo, J. D.; Godfrey, J. D.; Weller, H. N.; Pluscec, J. *Biochem. Biophys. Res. Commun.* **1984**, *124*, 141–147.

Table I. Activities for the ACE Training Set

molecule	pIC_{50}	ref	molecule	pIC_{50}	ref
mol 01	6.15	34	mol 35	6.19	31
mol 02	7.42	35	mol 36	2.70	31
mol 03	7.00	35	mol 37	2.70	31
mol 04	8.43	36	mol 38	2.87	31
mol 05	8.22	19	mol 39	4.51	31
mol 06	6.34	19	mol 40	5.52	31
mol 07	6.11	37	mol 41	4.96	31
mol 08	9.00	19	mol 42	2.74	31
mol 09	7.64	38	mol 43	2.96	31
mol 10	8.05	39	mol 44	3.21	31
mol 11	7.19	19	mol 45	2.98	31
mol 12	7.31	19	mol 46	3.26	31
mol 13	8.77[a]	19	mol 47	3.35	31
mol 14	7.30[a]	40	mol 48	3.64[a]	31
mol 15	8.54[a]	19	mol 49	3.38[a]	31
mol 16	8.52[a]	19	mol 50	3.89[a]	31
mol 17	9.64[a]	19	mol 51	3.22	31
mol 18	8.92[a]	41	mol 52	3.72	31
mol 19	8.92[a]	41	mol 53	4.28	31
mol 20	8.96[a]	19	mol 54	3.03	31
mol 21	8.55[a]	19	mol 55	3.62	31
mol 22	9.22[a]	19	mol 56	4.77	31
mol 23	8.40	19	mol 57	2.96	31
mol 24	8.00[a]	19	mol 58	3.62	31
mol 25	8.11[a]	36	mol 59	3.19	31
mol 26	7.92	19	mol 60	5.62	31
mol 27	8.52	19	mol 61	4.41	31
mol 28	8.54	19	mol 62	6.15	31
mol 29	8.15	19	mol 63	3.48	19
mol 30	5.55	31	mol 64	4.99	19
mol 31	6.07	31	mol 65	5.31	19
mol 32	5.80	31	mol 66	5.62	19
mol 33	6.37	31	mol 67	5.08	19
mol 34	6.70	31	mol 68	4.32	19

[a] Activity is for a diastereomeric or racemic mixture. Ambiguous chiral centers were modeled in the *S* configuration.

set of 28 classes of inhibitors used to define the active site geometry of Mayer et al.[21] The molecules were built from the SYBYL fragment database, using the Mayer molecules as templates when possible, and minimized using the TRIPOS force field.[49] Those molecules for which the activities were reported for diastereomeric mixtures were built in the *S* configuration at the ambiguous chiral centers in accordance with previous structure–activity studies.[35,36,41,44] Using Constrained Search,[50] all molecules were submitted to a conformational search using a 10° scan of all rotatable bands to verify that they could adopt the active site geometry defined by Mayer. Using MULTIFIT within SYBYL, the molecules were aligned by superimposing the C-terminal carboxylate, the amide carbonyl, and the zinc ligand of each molecule and minimized to fit the Mayer geometry (Figure 1) of either captopril or enalapril. The molecules were then input as rows of a QSAR table along with their respective IC_{50} values (input as pIC_{50}). CoMFA steric and electrostatic fields were calculated as described above and entered as columns in the QSAR table. Initially an analysis with 68 cross-validation groups was performed to determine the optimum number of principal components (PCs) using only the CoMFA steric and electrostatic fields as explanatory variables. This was followed by a non-crossvalidated analysis using the optimum number of PCs to derive a predictive model. Twenty additional inhibitors[19,31,34–48] (Table II), reflecting the diversity of the training set, were selected as a predictive set to test the robustness of the resulting model. To test the CoMFA model's ability to distinguish between different zinc–ligand classes, predictive sets for 10 carboxylate[51] (Table

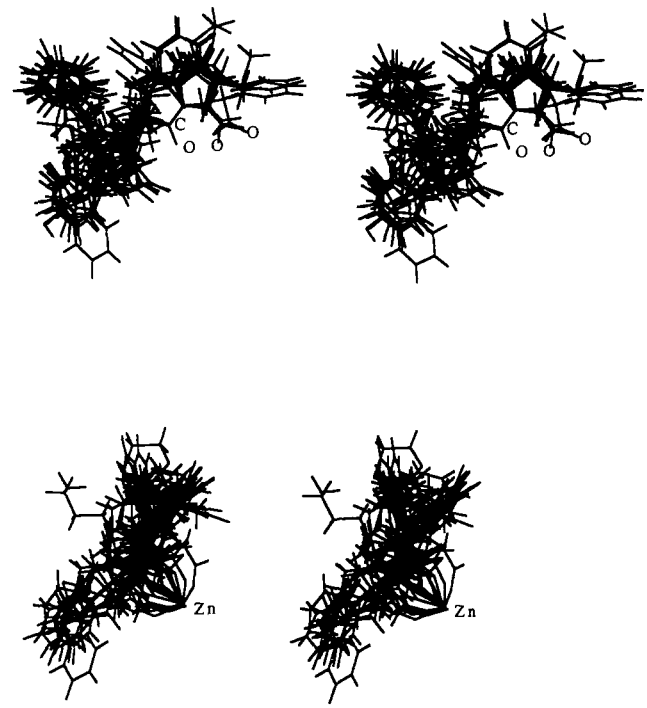
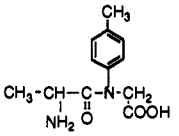
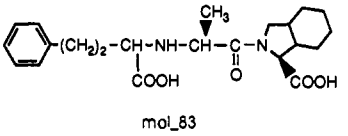
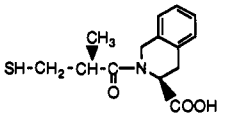
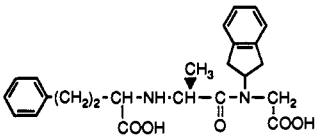
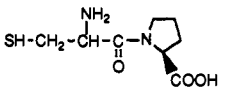
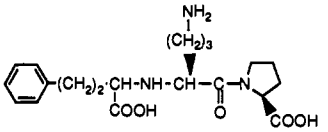
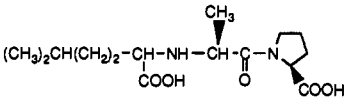
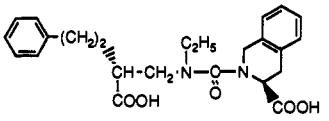
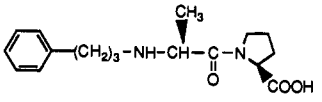
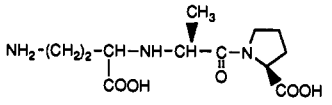
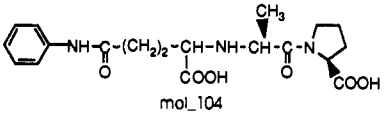
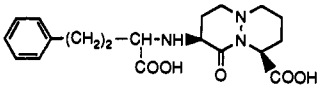
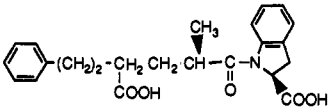
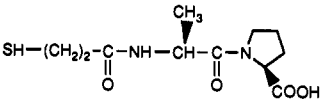
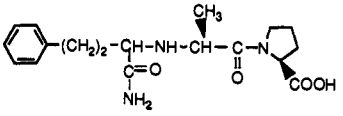
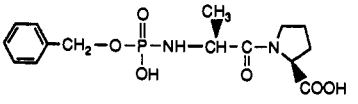
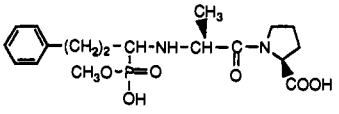
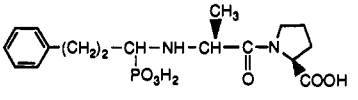
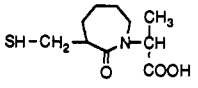
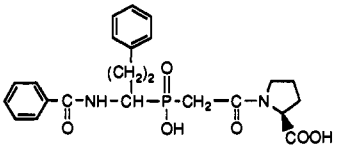


Figure 1. Orthogonal stereoviews of the 68 ACE inhibitors in the training set superimposed according to the alignment rule defined by the geometry of Mayer et al.[21]

III), 19 phosphate[52] (Table IV), and 17 sulfhydryl[53] (Table V) zinc–ligands were selected from the literature and MULTIFIT to the active site geometry as previously described, and their activities predicted. Analysis of the resulting correlation indicated a failure of the CoMFA to adequately distinguish between the different types of zinc–ligands. To explicitly define the type of zinc–ligand in the training set, additional columns were added to the QSAR table as indicator variables (carboxylate, sulfhydryl, phosphonate, phosphoramidate, or amine). This was done in a binary fashion. For example, if the inhibitor contained a sulfhydryl zinc–ligand, a factor of 10 was entered into the sulfhydryl column and a zero into all other zinc–ligand columns. These descriptor constants were varied by factors of 10 in an attempt to increase the signal from the zinc–ligand parameter in the PCA. [In SYBYL 5.3, the user is provided with two choices for weighting columns in the data table: either no scaling or autoscaling with equal weights assigned to all columns. In SYBYL 5.4, the user is provided with a choice of (1) no scaling, (2) autoscaling, (3) user defined scaling, or (4) CoMFA standard weights. The analyses using SYBYL 5.3 used no scaling of the data with weighting controlled by the magnitude of the zinc descriptor. These same options were applied for the analyses using SYBYL 5.4 so that results from SYBYL 5.3 could be reproduced in SYBYL 5.4. Attempts to produce better correlations in SYBYL 5.4 using user defined scaling were unsuccessful.] The resulting models were tested for their predictive ability against the same four test data sets. Finally, since it is quite possible that difference zinc–ligand classes bind to the active site with different geometries about the zinc atom as well as different binding strengths, separate CoMFA models were derived for the 12 phosphorus-based inhibitors, the 30 thiolates and the 23 carboxylates within the original training set of 68 molecules (the remaining three molecules were not included since they have amino groups at the zinc–ligand position which probably do not interact with the active site zinc atom).

(49) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982–1012.
(50) Dammkoehler, R. A.; Karasek, S. F.; Shands, E. F. B.; Marshall, G. R. *J. Comput.-Aided Molec. Des.* **1989**, *3*, 3–21.
(51) Sawayama, T.; Tsukamoto, M.; Sasagawa, T.; Nishimura, K.; Yamamoto, R.; Deguchi, T.; Takeyama, K.; Hosoki, K. *Chem. Pharm. Bull.* **1989**, *37*, 2417–2422.
(52) Karanewsky, D. S.; Badia, C.; Cushman, D. W.; DeForrest, J. M.; Dejneka, T.; Lee, V. G.; Loots, M. J.; Petrillo, E. W. *J. Med. Chem.* **1990**, *33*, 1459–1469.
(53) Oya, M.; Matusmoto, J.; Takasinea, H.; Watanabe, T.; Iwao, J.-I. *Chem. Pharm. Bull.* **1981**, *29*, 940–947.

Table II. Structures and Activities of the Molecules in the Test Set of 20 Diverse ACE Inhibitors

molecule	$p(IC_{50})$	ref	molecule	$p(IC_{50})$	ref
 mol_69	5.59[a]	19	 mol_83	8.66[a]	19
 mol_71	7.70	19	 mol_85	7.40[a]	19
 mol_75	7.20	19	 mol_91	8.66[a]	19
 mol_80	8.59[a]	19	 mol_99	8.19	19
 mol_82	5.24	19	 mol_101	6.49[a]	19
 mol_104	8.28[a]	19	 mol_131	7.25[a]	46
 mol_112	8.32	19	 mol_134	5.59	19
 mol_118	4.59	19	 mol_143	7.39	44
 mol_121	6.36[a]	19	 mol_147	7.29[a]	19
 mol_129	7.28[a]	44	 mol_154	7.14[a]	19

[a] Activity is for a diastereomeric or racemic mixture. Ambiguous chiral centers were modeled in the *S* configuration.

C. CoMFA of Thermolysin Inhibitors. The crystal structures of 10 inhibitors (phosphoramidon,[54] CLT,[55] PLN,[54] VW,[56] ZFPLA,[57] ZGPLL,[57] ZGPOLL,[58] L-Leu-NHOH,[59] HONH-BZMALONYL-Ala-Gly-NITROANILINE,[59] and CH_2CO-(NOH)-Leu-OCH_3[60]) bound to the active site of thermolysin

(54) Tronrud, D. E.; Monzingo, A. F.; Matthews, B. W. *Eur. J. Biochem.* **1986**, *157*, 261–268.
(55) Monzingo, A. F.; Matthews, B. W. *Biochemistry* **1984**, *23*, 5724–5729.
(56) Holden, H. M.; Matthews, B. W. *J. Biol. Chem.* **1988**, *263*, 3256–3260.
(57) Holden, H. M.; Tronrud, D. E.; Monzingo, A. F.; Weaver, L. H.; Matthews, B. W. *Biochemistry* **1987**, *26*, 8542–8553.
(58) Tronrud, D. E.; Holden, H. M.; Matthews, B. W. *Science* **1987**, *235*, 571–574.
(59) Holmes, M. A.; Matthews, B. W. *Biochemistry* **1981**, *20*, 6912–6920.

Table III. ACE Carboxylate Predictive Set[51]

molecule	pIC_{50}	molecule	pIC_{50}	molecule	pIC_{50}
coo_23A	7.92	coo_24C	9.60	coo_26A	7.92
coo_23J	6.41	coo_25A	8.46	coo_26C	7.00
coo_23E	5.60	coo_25E	7.82	coo_26H	6.66
coo_24A	5.60				

Table IV. ACE Phosphate Predictive Set[52]

molecule	pIC_{50}	molecule	pIC_{50}
sq29852	7.44	sq29852_2W	8.59
sq29852_2A	8.49	sq29852_2Y	9.06
sq29852_2B	9.31	sq29852_2Z	7.80
sq29852_2P	6.47	sq29852_2X	9.20
sq29852_2Q	9.88	sq29852_2E	8.97
sq29852_2R	9.05	sq29852_2G	9.54
sq29852_2S	9.11	sq29852_2H	9.94
sq29852_2T	9.28	sq29852_2K	9.53
sq29852_2U	9.28	sq29852_2I	9.53
sq29852_2V	8.72		

Table V. ACE Thiol Predictive Set[53]

molecule	pIC_{50}	molecule	pIC_{50}
thiol_2	2.46	thiol_18	6.12
thiol_4	4.11	thiol_20A	7.19
thiol_5	4.72	thiol_20B	6.44
thiol_7A	5.10	thiol_22	4.42
thiol_7B	3.27	thiol_27	5.34
thiol_9	2.14	thiol_28	<3.00
thiol_10	4.17	thiol_30A	7.46
thiol_12	3.59	thiol_30B	4.48
thiol_14	3.64		

were extracted from the Brookhaven Protein Data Bank[61,62] and minimized using the TRIPOS force field within SYBYL so that the bond lengths and bond angles would be consistent with the other inhibitors built within SYBYL. Since the TRIPOS force field is not parameterized for zinc, the zinc atom in the crystal structure was defined as "uninteresting" and zinc–ligand bond(s) were defined as an aggregate so as to be ignored by the minimizer. An additional 51 inhibitors (Table VI) with K_i values ranging from 10^{-2} to 10^{-10} M were taken from the literature[18,54,55,57,58,63–73] and modeled as previously described using the crystal structures as templates. After minimizing, each inhibitor was aligned with the crystal structure of the inhibitor most similar to it using MULTIFIT and then minimized within the active site of the same crystal structure as previously described. The 61 molecules were entered as rows of a QSAR table along with their respective pK_i values. CoMFA steric and electrostatic fields were calculated as previously described and entered as columns into the QSAR table. An initial PLS analysis with 30 cross-validation groups was performed to determine the optimum number of PCs using only the CoMFA steric and electrostatic fields as explanatory variables. This was followed by a non-crossvalidated run using the optimum number of PCs to derive a predictive model. An additional 11 inhibitors (Table VI) reflecting the diversity of structures in the training set were selected from the literature,[65,69,73] aligned with the crystal structure geometry and their activities predicted. As was done for the ACE series, additional analyses were performed using descriptors for the zinc–ligand type: carboxylate, phosphate (including phosphoramidates, phosphonates, and phosphinates), hydroxamates, and sulfur (includes thiols and sulfates). A constant of 10, 100, or 1000 was used to indicate the presence of the respective ligand type. The resulting noncrossvalidated models were then used to predict the inhibitory potencies of the 11 molecules in the test data set.

Table VI. Thermolysin Training Set

molecule	pK_i ($-\log M$)	ref	molecule	pK_i ($-\log M$)	ref
Training Set					
ace_ohleu_agnh2	2.47	73	ppheoh	4.14	73
bzsag	6.12	63	p_ile_aoh	6.44	73
c6pcltnme	7.28	64	(R)thiorphan	5.64	66, 67
c6pltnme	8.82	64	so2p_fagnh2	5.16	73
c6poltnme	5.84	64	so3_fagnh2	2.37	73
cbzphe	3.29	65	(s)thiorphan	5.74	66, 67
ch3coch2co_fagnh2	2.51	73	z-D-apola	4.62	68
ch3o2s_fagnh2	0.52	73	z-D-fpla	6.32	68
cho_ohleu_agnh2	2.47	73	z-D-fpola	4.52	68
cltzncrys	7.47	55	z-D-lpola	4.38	68
dah50	7.96	18	z(nh)glnh2	3.42	73
dah51	6.22	18	z(nh)glnhoh	5.57	73
dah52	5.55	18	zala	6.07	70
dah53	6.66	18	zapola	5.74	68
dah54	5.77	18	zfplazncrys	10.17	57
dah55	2.42	18	zfpola	7.35	68
hoch2co_fagnh2	2.54	73	zb-d-lnhoh	4.32	71
nhohbzmagna	6.37	73	zgg-d-lnhoh	3.60	71
nhohbzmagnh2	6.18	73	zgglnhoh	4.41	71
nhohbzmagoh	6.18	73	zggnhoh	3.03	71
nhohbzmoet	4.70	73	zglnh2	1.68	71
nhohibmagnh2	6.32	73	zglnhoh	4.89	71
nhohleu	3.72	73	zglnmeoh	2.65	71
nhohmalagnh2	2.96	73	zgly	6.39	70
ohbzmagnh2	3.38	73	zgpcllzncrys	6.74	72
p(ophe)(ome)leunh2	0.52	73	zgpla	7.78	68
paaoh	4.06	73	zgpllzncrys	8.04	57, 69
phosphoramidon	7.55	54	zgpola	4.89	68
pleunh2	4.10	73	zgpollzncrys	5.05	58, 69
pnhet	0.52	73	zlpola	6.17	68
po3_fagnh2	5.59	73			
Predictive Set					
zgpolf	4.27	69	zlgnh2	2.51	73
zgpolg	3.64	69	zfgnh2	3.46	73
zgplf	7.12	69	plfoh	7.72	73
zgplg	6.57	69	zygnh2	3.66	73
zgplnh2	6.12	69	β_ppphe	2.79	65
zgpolnh2	3.18	69			

Results

A. CoMFA of ACE Inhibitors. The results of the CoMFA analyses of the ACE inhibitors are summarized in Table VII. Detailed listings of the predictions for each molecule are available as supplementary material. The initial analysis which considered only CoMFA steric and electrostatic fields produced a correlation with a crossvalidated $R^2 = 0.66$ and a conventional $R^2 = 0.89$ for the predictive model using 8 PCs (Figure 2). When the model was used to predict the activities of 20 inhibitors representing all zinc–ligand types, the predictive R^2 was 0.53 (Figure 3). The

(60) Holmes, M. A.; Tronrud, D. E.; Matthews, B. W. *Biochemistry* **1983**, *22*, 236–240.
(61) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
(62) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. In *Crystallographic Databases-Information Content, Software Systems, Scientific Applications*; Allen, G. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107–132.
(63) Monzingo, A. F.; Matthews, B. W. *Biochemistry* **1982**, *21*, 3390–3394.
(64) Grobelny, D.; Goli, U. B.; Galardy, R. E. *Biochemistry* **1989**, *28*, 4948–4951.
(65) Kester, W. R.; Matthews, B. W. *Biochemistry* **1977**, *16*, 2506–2516.
(66) Benchetrit, T.; Fournie-Zaluski, M. C.; Roques, B. P. *Biochem. Biophys. Res. Comm.* **1987**, *147*, 1034–1040.
(67) Roderick, S. L.; Zaluski-Fournie, M. C.; Roques, B. P.; Matthews, B. W. *Biochemistry* **1989**, *28*, 1493–1497.
(68) Bartlett, P. A.; Marlowe, C. K. *Biochemistry* **1987**, *26*, 8553–8561.
(69) Bartlett, P. A.; Marlowe, C. K. *Science* **1987**, *235*, 569–571.
(70) Feder, J.; Brougham, L. R.; Wildi, B. S. *Biochemistry* **1974**, *13*, 1186–1189.
(71) Nishino, N.; Powers, J. C. *Biochemistry* **1978**, *17*, 2846–2850.
(72) Morgan, B. P.; Scholtz, J. M.; Ballinger, M. D.; Zipkin, I. D.; Bartlett, P. A. *J. Amer. Chem. Soc.* **1991**, *113*, 297–307.
(73) Klopman, G.; Bendale, R. D. *J. Theor. Biol.* **1989**, *136*, 67–77.

Table VII. Summary of CoMFA Analyses of the ACE Series

QSAR	obs	SE[a]	R^2 cross[b]	F value	p value	R^2 [c]	SD
1. CoMFA only	68	1.47	0.66(8)[d]	70.77	0.00	0.89	0.71
2. CoMFA + Zn (10×)	68	1.27	0.64(4)	88.78	0.00	0.84	0.88
3. CoMFA + Zn (100×)	68	1.26	0.68(4)	75.71	0.00	0.83	0.90
4. CoMFA + Zn (1000×)	68	1.43	0.67(7)	47.60	0.00	0.81	1.00
5. CoMFA + Zn (10 000×)	68	1.43	0.66(7)	47.03	0.00	0.80	1.02
6. carboxylates	23	1.55	0.81(7)	1147.90	0.00	0.99	0.14
7. phosphates	12	3.56	0.46(2)	22.85	0.00	0.83	0.63
8. thiols	30	1.94	0.26(4)	22.84	0.00	0.79	0.90

[a] Cross-validated standard error of estimate. [b] Cross-validated R^2: R^2 value for analysis using designated number of cross-validation groups. [c] Conventional R^2: fitted R^2 value for predictive model derived using no cross-validation and the optimum number of principal components. [d] Numbers in parentheses are optimal numbers of components.

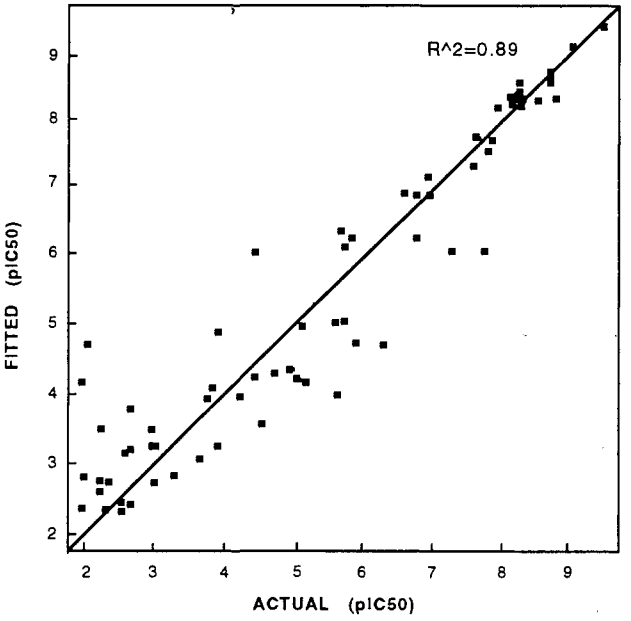


Figure 2. Fitted predictions vs actual pIC_{50} for the CoMFA analysis of 68 ACE inhibitors incorporating only CoMFA parameters (QSAR1). The model was derived using eight principal components having a cross-validated $R^2 = 0.66$.

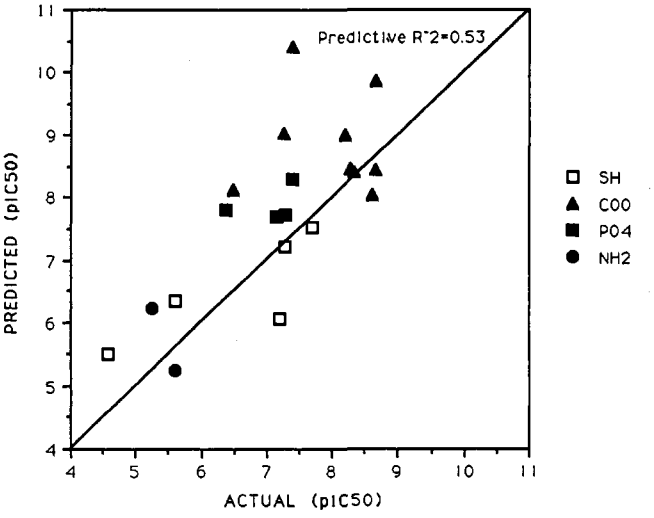


Figure 3. Predicted vs actual pIC_{50} for the test data set of 20 diverse ACE inhibitors incorporating only CoMFA parameters (QSAR1). The predictive model was derived using eight principal components having a cross-validated $R^2 = 0.66$ and a conventional $R^2 = 0.89$.

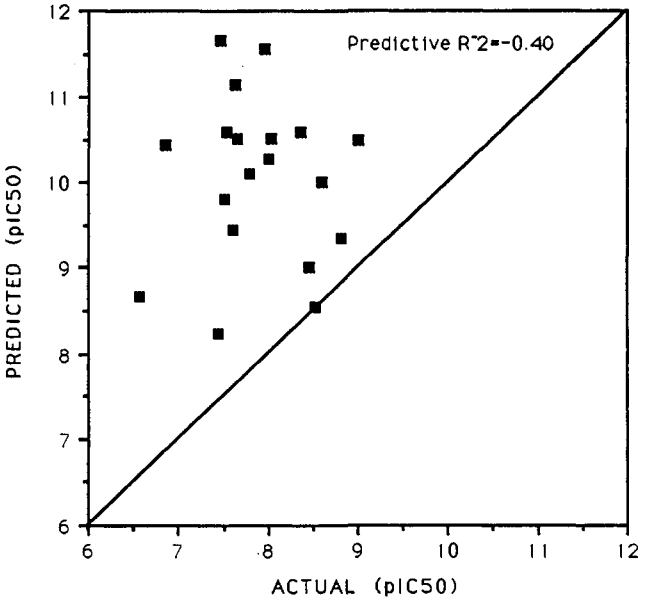


Figure 4. Predicted vs actual pIC_{50} for the test data set of 19 phosphate based ACE inhibitors incorporating only CoMFA parameters (QSAR1). The predictive model was derived using eight principal components having a cross-validated $R^2 = 0.66$ and a conventional $R^2 = 0.89$.

phosphorous-based inhibitors tend to be overpredicted, the sulfhydryl inhibitors underpredicted, and the carboxylate inhibitors equally distributed above and below the line representing a correlation of $R^2 = 1.0$. These results could indicate the presence of a systematic error in the analysis. In order to confirm this impression, the activities for a set of 19 phosphorous-based inhibitors were predicted. As Figure 4 shows, all were overpredicted, supporting the systematic error hypothesis. As the structure–activity data indicate, there is a correlation between ACE inhibitory activity and zinc–ligand type. These results indicate that the zinc–ligand interaction was inadequately represented in the CoMFA fields. This may be due to the fact that the steric field is more heavily weighted than the electrostatic field (3.4:1) in the PLS analysis. From the perspective of the active site zinc atom, the differences between the zinc ligands are both electrostatic in nature and represent ligand bonds (1–2 interactions) which are not evaluated in the nonbonded interaction fields used in CoMFA. Further predictions for the test data sets of the other individual classes of zinc ligands support this hypothesis, with R^2 values of 0.60 and 0.20 for the carboxylates and thiolates, respectively.

The analyses which included the explicitly defined zinc descriptors generally gave improved correlations. Among the four zinc–ligand indicators, 10, 100, 1000, and 10 000, the crossvalidated and conventional R^2 values are comparable to those of the "CoMFA only" model above, but the predictive ability of these models is significantly improved. Overall, no single model gives accurate predictions for all classes of zinc–ligands. The

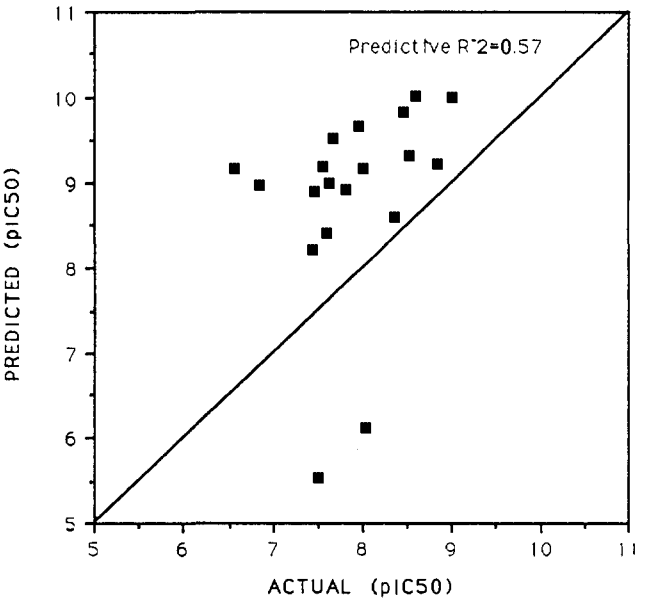


Figure 5. Predicted vs actual pIC_{50} for the test data set of 19 phosphate-based ACE inhibitors using CoMFA parameters plus a zinc indicator equal to 10 (QSAR2). The predictive model was derived using four principal components having a cross-validated R^2 = 0.64 and a conventional R^2 = 0.84. The inclusion of the zinc indicator improves the predictions for the phosphates by 50%.

Table VIII. X-Loadings from ACE CoMFA Analysis using Zn Descriptor of 10

	SH	COO	PO	NH_2
component 1	-0.635	1.111	0.209	-1.059
component 2	-1.257	-0.970	-0.621	2.844
component 3	4.786	-1.592	-0.526	-2.243
component 4	-0.667	0.351	0.722	0.296

model using a zinc descriptor of 10 (QSAR 2), however, gives the best overall performance (low crossvalidated standard error of estimate and highest F values and conventional R^2), especially in its ability to predict the phosphate inhibitors relative to the other models (Figures 4 and 5). It is worth noting that the model using a zinc descriptor of 10 required only four principle components to explain 84% of the variance in inhibitory potency compared to the eight components required for the analysis using only the CoMFA parameters. Analysis of the X-loadings (Table VIII) for this analysis indicates that the major contributors to each component are the zinc–ligand types, namely carboxylate, amino, thiol and phosphate for components 1, 2, 3, and 4, respectively. Additionally, the X-scores shows that for the well-predicted molecules in the training set, the model does recognize the difference in zinc–ligand type, as the principle component with the highest score for a given molecule is also the component representing the appropriate ligand type for that molecule.

The best correlations were obtained when CoMFA models were derived for each individual class of zinc–ligands (Figure 6). The conventional R^2 for the phosphorous-based inhibitors (Figure 6A) was 0.83, producing a predictive R^2 value of 0.74 for the phosphate test data set. For the carboxylate class (Figure 6C), the conventional R^2 was 0.99 and the predictive R^2 for the test data set was 0.66. Although the thiol compounds were much better correlated with a conventional R^2 = 0.79 (Figure 6C), the model was unable to predict the activities of the test data set.

B. CoMFA of Thermolysin Inhibitors. The results of the CoMFA analyses for the thermolysin inhibitors are summarized in Table IX. Analysis of 61 thermolysin inhibitors using only CoMFA parameters produced a correlation having a crossvalidated R^2 = 0.70 and a conventional R^2 = 0.98 using 11 PCs (Figure 7). The predictive R^2 for the test data set of 11 diverse

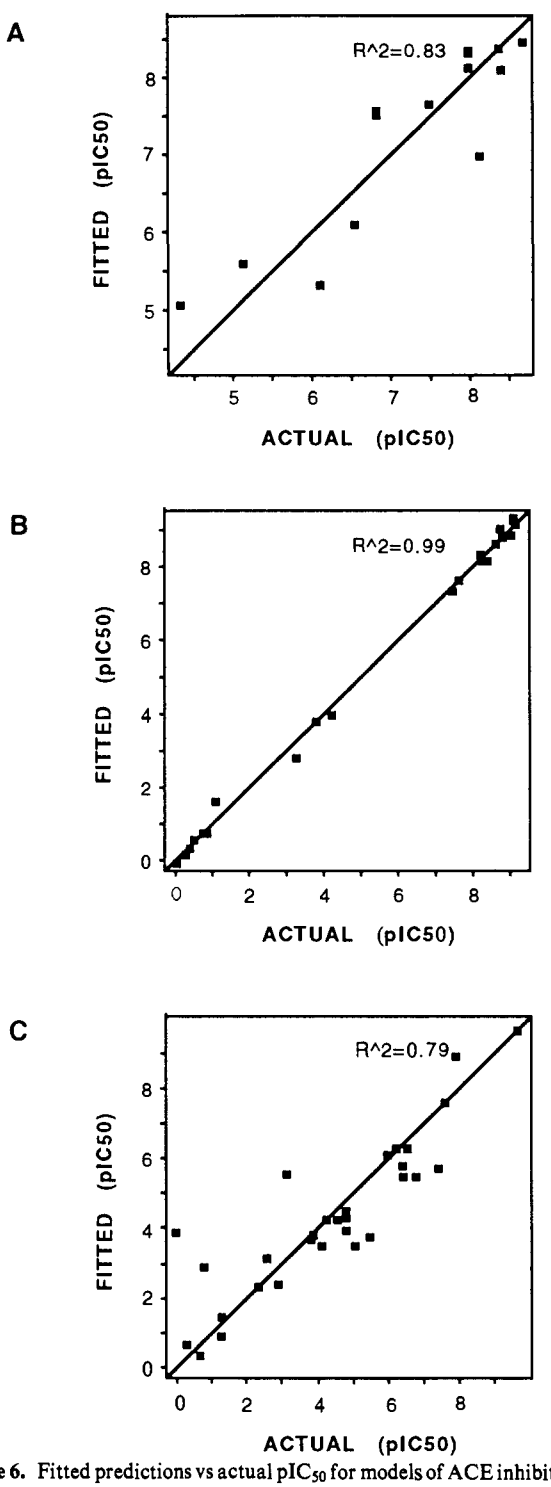


Figure 6. Fitted predictions vs actual pIC_{50} for models of ACE inhibition derived from training sets composed of the individual zinc-ligand classes using only CoMFA parameters: (a) 12 phosphates (QSAR7); (b) 23 carboxylates (QSAR6); (c) 30 thiols (QSAR8).

inhibitors, however, was only 0.29 (Figure 8), which is significantly poorer than the predictions obtained from the ACE "CoMFA only" model (conventional R^2 = 0.89, predictive R^2 = 0.53). While there were only minor differences in the conventional and crossvalidated R^2 values, the addition of the zinc descriptors had no effect on the predictive ability of the models. Although the number of principle components is similar for each model, the relative amount of the variance explained by each component differs. The model incorporating the zinc indicator equal to 1000 requires six principle components to explain 70% of the variance in the biological data (pK_i), whereas the models derived with the

Table IX. Summary of CoMFA Analyses of the Thermolysin Series

QSAR	obs	SE[a]	R^2 cross[b]	F value	p value	R^2 [c]	SD
9. CoMFA only	61	1.30	0.70(11)[d]	205.41	0.00	0.98	0.31
10. CoMFA + Zn (10×)	61	1.32	0.67(11)	205.41	0.00	0.98	0.31
11. CoMFA + Zn (100×)	61	1.35	0.64(10)	205.42	0.00	0.98	0.35
12. CoMFA + Zn (1000×)	61	1.40	0.63(12)	205.42	0.00	0.98	0.36

[a] Cross-validated standard error of estimate. [b] Cross-validated R^2: R^2 value for analysis using designated number of cross-validation groups. [c] Conventional R^2: fitted R^2 value for predictive model derived using no cross-validation and the optimum number of principal components. [d] Numbers in parentheses are optimal numbers of components.

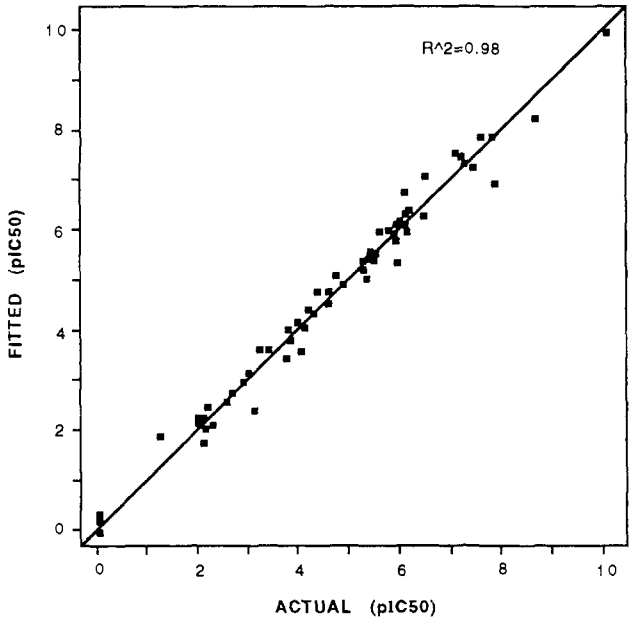


Figure 7. Fitted preditions vs actual pIC_{50} for the CoMFA analysis of 61 thermolysin inhibitors incorporating only CoMFA parameters (QSAR9). The model was derived using 11 principal components having a cross-validated $R^2 = 0.70$.

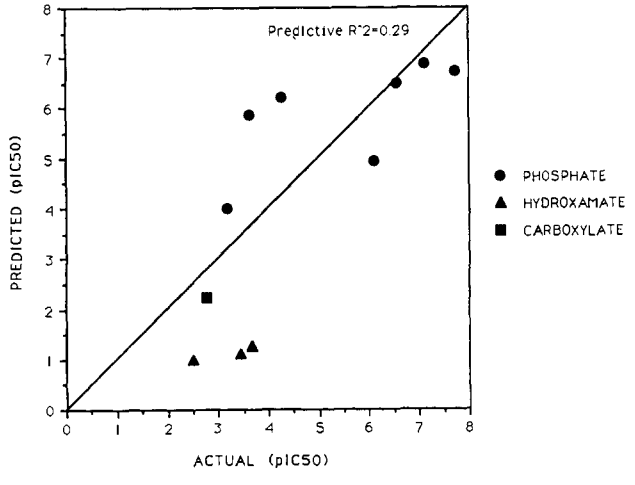


Figure 8. Predicted vs actual pIC_{50} for the test data set of 11 diverse thermolysin inhibitors incorporating only CoMFA parameters (QSAR9). The predictive model was derived using eight principal components having a cross-validated $R^2 = 0.70$ and a conventional $R^2 = 0.98$.

zinc descriptors equal to 10 and 100 require only three components to explain the same amount of variance.

C. CoMFA Fields. 1. ACE. The CoMFA coefficient contour maps for the steric and electrostatic potentials, which are calculated as the scalar product of the relevant QSAR coefficient and the standard deviation of all values in the corresponding columns of the data table, are depicted in Figure 9, parts a and b, respectively. These contours represent the lattice points where *differences in field values are associated with differences in inhibitory potency*. In the absence of an enzyme structure, the interpretation of these contours is primarily intuitive and highly subjective. From an electrostatic perspective, one would expect to see strong electrostatic interactions around the C-terminal carboxyl as well as the amide bond between the P_1' and P_2' residues, since these sites are proposed to exhibit strong hydrogen-bonding and charge–charge interactions, respectively. Examination of the contours of the CoMFA electrostatic field supports this binding hypothesis. Figure 9a shows the standard deviation times the CoMFA coefficients (STDEV*COEFF) for the electrostatic field. The blue contours represent regions where the STDEV*COEFF is less than −0.01, while the yellow contours represent where the STDEV*COEFF is greater than +0.01. A predominant feature of the electrostatic contour is the yellow contour about the amide proton of the P_1' residue (which is expected to function as a hydrogen bond donor), indicating that an increase in positive charge in this region will increase activity. Additionally, the blue contours about the C-terminal carboxyl group, which is proposed to interact with a cationic side chain such as the guanido group of an arginine residue, indicate that an increase in negative charge in this region with increase activity. The area of negative potential about the P_2' residue corresponds to the increased electron density about the aromatic ring of inhibitors such as the bicyclic lactams[74] as well as some of the proline surrogate analogs such as benzofused Pro[75] which is 3–7 times more potent than captopril or the 2-hydroxyphenyl analog, fentiapril,[76] which is 3–4 times more potent than captopril.

An unexpected feature of the CoMFA electrostatic potential is the blue contour about the P_1 residue, i.e., the pendant phenethyl group of enalapril indicating that negative charge in this region is associated with an increase in activity. According to the current model of ACE inhibitor binding,[19] the P_1 site interacts with a hydrophobic pocket which recognizes its aromatic ring. In support of this, the CoMFA steric field (Figure 9b) indicates that this is a region where additional steric bulk is associated with increased activity. Therefore, the enhancement in binding due to the the P_1 residue may not be merely a hydrophobic interaction but also a cation-π interaction[77] with a cationic residue in the active site. This type of stabilizing interaction between a quaternary ammonium group and the electrons of aromatic systems has been observed for aromatic residues in other protein structures.[78] This mode of interaction is further supported by the fact that branching at the P_1 site is not well tolerated nor are aminoalkyl side chains unless they are acylated.[19] The remaining yellow contours further define the steric boundaries of the S_1' and S_2' subsites, i.e., regions beyond which additional steric bulk will overlap with the atoms of the active site.

(74) Stanton, J. L.; Watthey, J. W. H.; Desai, M. N.; Finn, B. M.; Babiarz, J. E.; Tomaselli, H. C. *J. Med. Chem.* **1985**, *28*, 1603–1606.
(75) Stanton, J. L.; Gruenfeld, N.; Babiarz, J. E.; Ackerman, M. H.; Freidmann, R. C.; Yuan, A. M.; Macchhia, W. *J. Med. Chem.* **1983**, *26*, 1267–1277.
(76) Oya, M.; Kato, E.; Iwao, J.; Yasuoka, N. *Chem. Pharm. Bull.* **1982**, *30*, 484.
(77) Burley, S. K. *FEBS Lett.* **1986**, *203*, 139–203.
(78) Dougherty, D. A.; Stauffer, D. A. *Science* **1990**, *250*, 1558–1560.

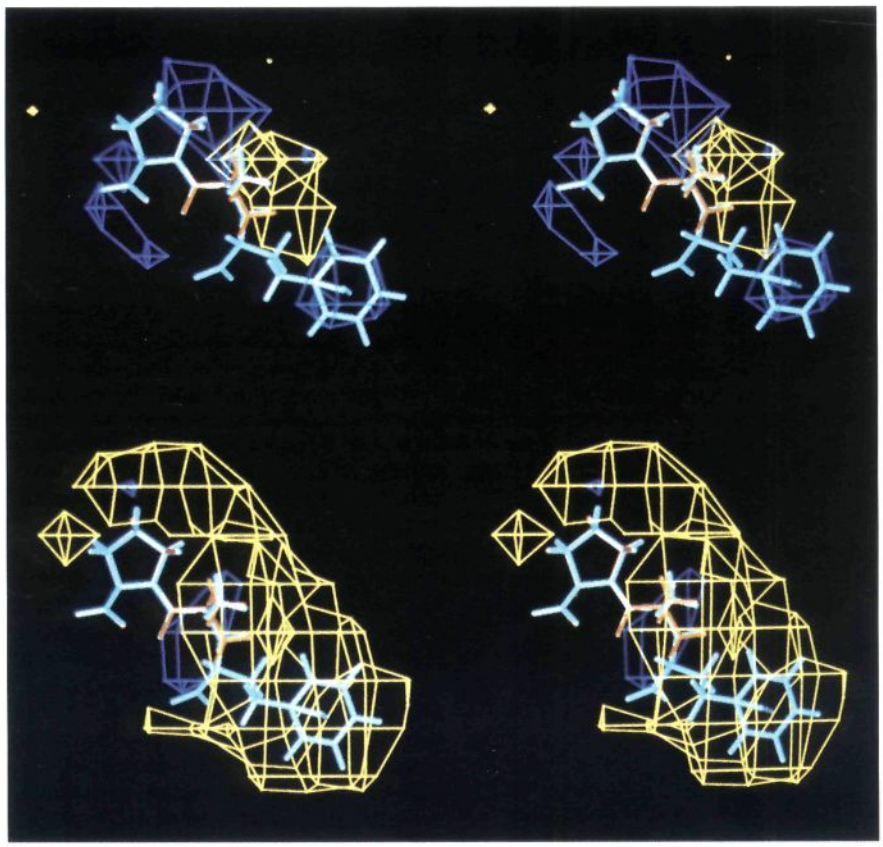


Figure 9. Stereoviews of the contours of the CoMFA STDEV*COEFF for the ACE series surrounding the ACE inhibitors captopril (red) and enalapril (cyan). (a, top) Electrostatic contours at the ±0.02 kcal/mol level: the blue contours (STDEV*COEFF = –0.02) indicate regions where the addition of negative charge will increase activity and the yellow contours (STDEV*COEFF = +0.02) indicate regions where the addition of positive charge will increase activity. (b; bottom) Steric contours at the ±0.01 kcal/mol level: the blue contours (STDEV*COEFF = –0.01) indicate regions where the additon of steric bulk is associated with decreased activity. The yellow contours (STDEV*COEFF = +0.01) indicate regions where the addition of steric bulk is allowed and also define the steric boundary of the active site.

2. Thermolysin. The principle electrostatic interactions between an inhibitor and the active site of thermolysin are well defined by the crystal structure of the phosphonamidate ZF^{P}-LA.[57] These interactions are thought to closely represent the interactions of a peptide substrate in the transition state. Tyr-157, His-231, and Glu-143 all serve as hydrogen bond donors to the phosphonamide oxygens. The amide oxygen of Ala-113 and oxygen OD1 of Asn-112 serve as H-bond acceptors of the phosphonamide proton. The carbonyl oxygen of the inhibitor's carbobenzoxyl group forms a strong hydrogen bond with the amide nitrogen of Trp-115. The contours of the CoMFA electrostatic field correlate well with these interactions (Figure 10a). These contours show regions where an increase in negative charge (top figure) and high positive potential (bottom figure) will increase activity. These regions correspond to the inhibitor's hydrogen bond acceptor and hydrogen bond donor sites, respectively. Additionally, there are extended areas of electrostatic potential in the solvent exposed region of the active site. Although there is no correlation for these contours with the phosphonamide inhibitor shown, they most likely indicate solvent/inhibitor interactions with other inhibitors in the training set that are further extended at their C-terminus such as the hydroxamic acid inhibitors.[71]

The binding sites (S_2, S_1, S_1', and S_2') of thermolysin correspond to the P_2 (carbobenzoxyl), P_1 (phenylalanyl), P_1' (leucyl), and P_2' (alanyl) residues of ZF^{P}LA. The bottom figure of Figure 10b indicates the region where the STDEV*COEFF for the steric potential is positive. These are regions of allowed steric interactions which coincide with the S_1' and S_2' binding sites around the Leu and Ala residues of the inhibitor. The Phe residue of the inhibitor is solvent exposed and does not interact with the enzyme, so that the contour around its phenyl ring may be representative of a favorable steric interaction with enzyme-bound water molecules. This is supported by the fact that the crystal structure of native thermolysin shows water molecules in the P_1 region that have been displaced in the inhibitor-bound crystal structure. This contour may also be a manifestation of the entropy gain in displacing solvent upon binding of the inhibitor, which would be expected to produce overpredicted binding affinities, such as those observed. The top figure of Figure 10b shows a contour region which fills the gap between the P_1 residue and the active site indicating that there are some sterically unfavorable orientations of the P_1 and P_2 residues with the training set. This is reflective of a number of inhibitors such as ZG^{P}LL,[57] which has no phenyl ring to occupy the S_1 site. In these inhibitors the carbobenzoxy carbonyl points away from the amide of Trp-115 so that the Z-group phenyl ring actually occupies a region between the S_1 and S_2 binding sites.

Discussion

In determining the active site geometry of the ACE inhibitors, Mayer et al.[21] assumed that all classes of ligands bind to the active-site zinc in the same mode. This assumption may be overly restrictive since most of the sulfhydryl inhibitors have at least one less rotatable bond between the zinc–ligand and the amide carbonyl than do the phosphorous- or carboxylate-based inhibitors, and may have a much different orientation at the zinc atom. Additionally, crystal structures of inhibitors of thermolysin indicate that some phosphorous-based inhibitors bind to the active site zinc in a bidentate manner. In the Mayer model, the

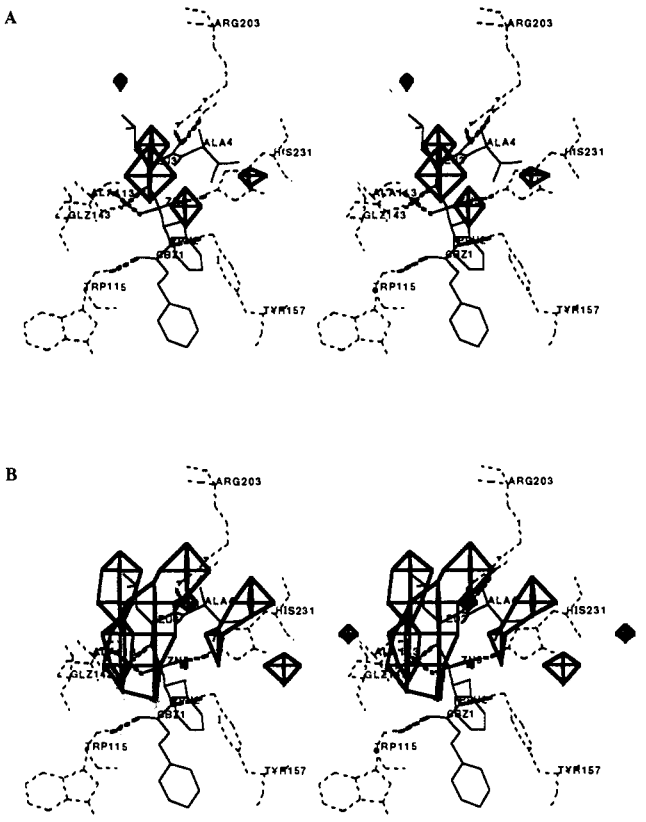


Figure 10. Stereoviews of the electrostatic contours at the ±0.04 kcal/mol level of the CoMFA STDEV*COEFF for the thermolysin series shown surrounding the crystal structure of the potent inhibitor ZF^{P}LA. The essential interactions of the inhibitor (solid lines) with the active site (dashed lines) are shown. (a) Regions where the STDEV*COEFF is negative show where the addition of negative charge will increase activity. (b) Regions where the STDEV*COEFF is positive show where the additon of positive charge will increase activity. These regions correlate well with the inhibitor's hydrogen bond acceptor and donor sites, respectively.

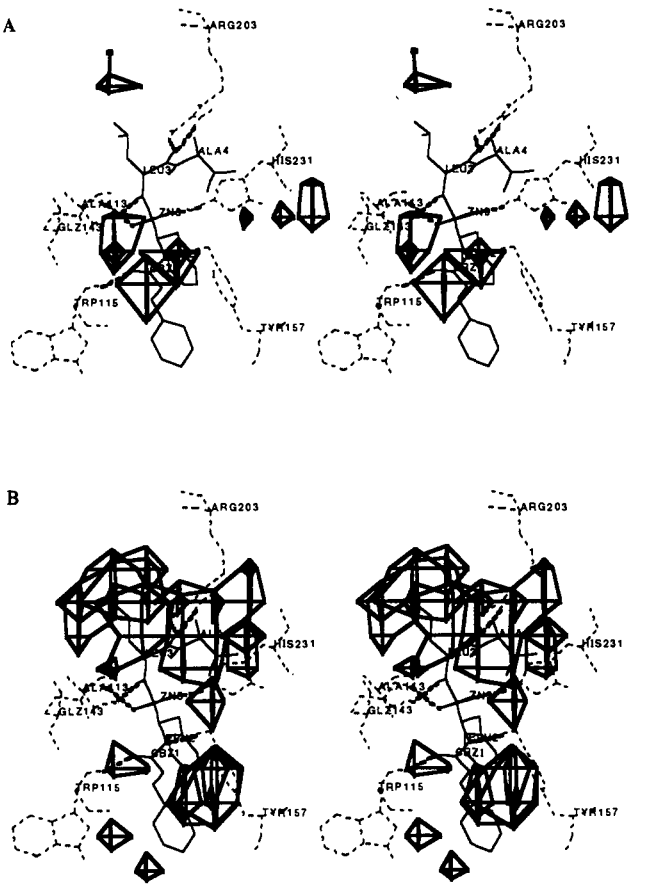


Figure 11. Stereoviews of the steric contours at the ±0.08 kcal/mol level of the CoMFA STDEV*COEFF for the thermolysin series shown surrounding the crystal structure of the potent inhibitor ZF^{P}LA. The essential interactions of the inhibitor (solid lines) with the active site (dashed lines) are shown. (a) Regions where STDEV*COEFF is negative show where additional steric bulk will decrease activity. (b) Regions where STEDEV*COEFF is positive show regions of allowed steric interactions and also define the boundaries of the active site.

phosphorous-based inhibitors were modeled as binding to the zinc in a monodentate manner which may explain the variance in the phosphorous predictions. Hausin and Codding[79] have studied a number of crystal structures of zinc-bound ligands and have suggested that the zinc coordination geometry of Andrews[22] that was used by Mayer to define the zinc–sulfhydryl and the zinc–carboxyl geometry for the systematic search is incorrect and can lead to a variance in zinc position by as much as 1 Å. An additional consideration, which we have not explicitly taken into account, is the strength of the zinc–ligand interaction. SAR data indicate a strong correlation of inhibitor activity to zinc–ligand type, i.e., phosphate > carboxylate > thiol. Although the zinc descriptor is an approximation of these different interactions, it is an overly simplistic one and does not explicitly consider the energetics of the zinc–ligand bond. This may be an essential parameter for explaining the differences in activity.

In the case of the thermolysin analysis, we have removed the variable of the active site geometry from the model, but we are presented with a new problem—optimization of the geometry about the zinc atom. As discussed previously, the TRIPOS force field is not parameterized for zinc. Optimized structures could only be obtained by ignoring the zinc atom and defining the zinc–ligand bond(s) as an aggregate. The active site geometries about the zinc atom were based entirely on those of the crystal structures of similar inhibitors. An alternative solution is to optimize the structures using a force field that has zinc parameters. Such a force field is Vedani's YETI force field[80] which is a modification of the Kollman force field[81] incorporating a charge-transfer model for metal–ligand interactions which yields atomic charges that are in agreement with the semiempirical calculations of Giessner-Prettre and Jacob[82] for Zn^{2+} in the thermolysin active site. We have optimized several thermolysin inhibitors using YETI v5.0 on a VAXStation 3520 and have obtained structures consistent with X-ray data. YETI, however, requires some *a priori* knowledge, or assumptions, about the binding of the ligand to the active site zinc atom. Crystal data indicate that the zinc atom may be tetracoordinated with the ligand bound to the zinc in a monodentate arrangement, as is the case with inhibitors such as ZG^{P}LL[57] and phosphoramidon,[83] or in a bidentate manner with the zinc in a pentacoordinated transition state as found with the inhibitor ZF^{P}LA.[57] YETI requires the user to define the coordination number of the metal center prior to optimization, and there are no rules consistent with all of the crystal data to guide such a prediction. The most effective use of YETI would be to optimize each structure in both coordination states and select the one with the lowest energy. Another question arises concerning the orientation of the ligand with respect to the S_1, S_2, S_1', and S_2' subsites. In most cases the orientation would seem to be straightforward, but how does one explain or predict the orientation seen with the inhibitor L-Leu-NHOH, which binds "backward" in the active site with respect to its N-terminal to

(79) Hausin, R. J.; Codding, P. W. *J. Med. Chem.* **1990**, *33*, 1940–1947.
(80) Vedani, A.; Huhta, D. W. *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767.
(81) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. J. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
(82) Giessner-Prettre, C.; Jacob, O. J. *J. Comput.-Aided Molec. Des.* **1989**, *3*, 23–37.
(83) Weaver, L. H.; Kester, W. R.; Matthews, B. W. *J. Mol. Biol.* **1977**, *114*, 119–132.

Table X. Summary of Results Incorporating Approximations of Thermodynamic Parameters

QSAR	obs	SE[a]	R^2 cross[b]	F test	p value	R^2 [c]	SD	predictive R^2
			ACE					
13. CoMFA + DOF[d] (Zn = 10×)	68	1.29	0.68(6)[e]	43.07	0.00	0.81	0.99	0.55[g]
14. CoMFA + ΔH[f] (Zn = 10×)	68	1.31	0.66(5)	41.34	0.00	0.77	1.07	0.62[g]
15. CoMFA + DOF + ΔH (Zn = 10×)	68	1.33	0.65(6)	34.86	0.00	0.77	1.07	0.63[g]
			Thermolysin					
16. CoMFA + DOF (Zn = 10×)	61	1.32	0.67(11)	243.11	0.00	0.98	0.31	0.30[h]

[a] Cross-validated standard error of estimate. [b] Cross-validated R^2: R^2 value for analysis using designated number of cross-validation groups. [c] Conventional R^2: Fitted R^2 value for predictive model derived using no cross-validation and the optimum number of principal components. [d] DOF is defined as the number of torsional degrees of freedom in each inhibitor. [e] Numbers in parentheses are optimal numbers of components. [f] ΔH is the change in conformational enthalpy, $\Delta H_{conform} = H_{aligned} - H_{min}$, where $H_{aligned}$ is the energy of the molecule in its aligned conformation, and H_{min} is the energy of the conformation at the nearest local minima. [g] 20 inhibitors. [h] 11 inhibitors.

C-terminal sense, having its leucyl side chain in the S_1' site rather than in the S_1 site as the "normal mode" of binding for extended substrates would dictate?[59] In aligning the structures for which there are no crystal data, we have had to make assumptions about both the zinc geometry and the orientation in the active site which may not be correct.

An important question to ask at this point is what factors are sufficient to parameterize ligand-receptor binding, and does CoMFA adequately represent those parameters. Williams et al.[84] have developed an expression for estimating the binding constants for bimolecular associations in solution in which they partition the free energy of binding into five terms: (i) $\Delta G_{(trans+rot)}$, the change in translational and rotational free energy upon ligand-receptor association, (ii) $\Delta G_{(rotor)}$, the free energy change due to the freezing out of internal rotations upon association, (iii) $\sum \Delta G_i$, the free energies of interactions between polar functional groups, (iv) ΔG_{vdw}, the free energy of hydrocarbon-to-hydrocarbon interactions (molecular packing of the complex relative to the dissociated form in the solvent), and (v) ΔG_H, the transfer free energy for nonpolar groups from water to the active site as well as the associated release of ordered water (the hydrophobic effect). Thus, the free energy of binding is controlled by a balance between the unfavorable terms, (i) and (ii), and the favorable terms, (iii), (iv), and (v), such that the $\Delta G_{binding} < 0$ for complex formation to occur. CoMFA electrostatic and steric potentials are shape-dependent, molecular parameters which may approximate only terms (iii) and (iv), respectively. CoMFA alone only considers interactions *at the active site of the enzyme* with the implicit assumption that the entropic contributions do not differ significantly. There is no explicit information describing the entropy loss due to the decrease in internal rotational degrees of freedom upon complex formation or the entropy gain due to the hydrophobic effect. Thus, these thermodynamic considerations which can strongly influence $\Delta G_{binding}$ are ignored when using only the CoMFA steric and electrostatic interactions. A molecule, therefore, may be predicted to be a potent inhibitor based on its CoMFA fields, but its activity may be much different due to solvation and hydrophobic effects, which are not explicitly described by the normal CoMFA parameters.

In an effort to address the hydrophobicity issue, some of our early analyses incorporated CLOGP[85] and dipole moment values with the CoMFA data. The resulting correlations were no better than the models derived from CoMFA parameters alone (R^2 = 0.66). Analyses using only CLOGP and dipole moment produced an R^2 = 0.50. Sharp et al.[86] have suggested that solvent partition coefficients underestimate the strength of the hydrophobic effect and that a better estimation is derived from transfer free energies which have been corrected for differences in volume entropy between solvents. A better companion to CoMFA for considering hydrophobic effects may be Kellogg and Abraham's HINT (Hydrophobic INTeraction) program[87] which provides an interface to SYBYL's QSAR module. Studies which incorporate HINT parameters into our CoMFA analyses are forthcoming.

Additionally, we attempted to approximate binding energies and entropy effects using the overly simplified parameters of the number of torsional degrees of freedom in the inhibitor (DOF) and the change in conformational enthalpy, $\Delta H_{conform} = H_{aligned} - H_{min}$, where $H_{aligned}$ is the energy of the molecule in its aligned conformation, and H_{min} is the energy of the conformation at the nearest local minimum. These parameters were added in addition to the zinc indicator (Table X). For the ACE series, these additional parameters clearly influenced the model, since the number of principal components increased from 4 to either 5 or 6. There was, however, no significant change in either the conventional or crossvalidated R^2 values with the addition of these parameters. The predictive ability of these models was comparable to those derived from the CoMFA parameters and zinc indicators alone but were no better than the best model derived using a zinc indicator of 10. Only the DOF parameter was added to the thermolysin analysis, and it had no effect on the results. The number of principal components, conventional R^2, crossvalidated R^2, and predictive R^2 were unchanged from the analysis using a zinc descriptor of 10. In both analyses, the first four principal components explain 84% of the variance in the biological data. Analysis of the X loadings indicates that the DOF parameter is influencing the model, as the second component is heavily loaded by the DOF variable.

The utility of a QSAR method such as CoMFA is ultimately as a pharmacological prescreen for biological activity prior to undertaking a potentially costly and time consuming synthesis project. The medicinal chemist, in selecting a training set and test sets, would not employ methods of statistical design. Instead, the chemist would construct a training set composed of molecules with known activities from an in-house database. The test molecules would be hypothetical structures derived from the chemist's knowledge of the training set combined with his/her chemical intuition. This is the same strategy that we have employed in the random selection of our test sets from the literature.

The composition of the training sets and test sets, however, can have a substantial effect on the resulting correlations.[88] One would not expect a molecule that is significantly different from those represented in the training set to be well predicted. The

(84) Williams, D. H.; J. P. L., C.; Doig, A. J.; Gardner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. *J. Am. Chem. Soc.* **1991**, *113*, 7020–7030.

(85) Daylight Chemical Information Systems, Inc.: Irvine, CA, USA.

(86) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686–9697.

(87) Kellogg, G. E.; Abraham, D. J. In Virginia Commonwealth University: Richmond, VA, USA, 1991.

Table XI. Mean Activities of the Data Sets Used in the ACE Analyses

data set	no. obs	mean p(IC_{50}) (log M)	std dev
training set	68	5.91	2.15
diverse inhibitors	20	7.11	1.27
phosphates	19	7.88	0.63
carboxylates	10	7.41	0.93
thiols	17	4.63	1.57

members of the test data sets reflect chronological appearance in the literature rather than rational design. This may explain, to some degree, the poor predictions obtained for the thiol-based inhibitors. Because of their generally marginal activities as well as their side effects, sulfhydryl-based ACE inhibitors have not received much attention in the recent literature. The set of thiols chosen as a test set was published in 1981[53] and includes molecules with marginal activities (pIC_{50} = 2.5–7.0) which are not well-represented in the training set. This is supported by a comparison of the mean activities of each data set (Table XI), which shows that the mean activity of the thiol test set is at least one order of magnitude lower than that of the training set and 2–3 orders of magnitude lower than those of the other test data sets. The 68-membered ACE training set, while being diverse in structure, does not necessarily reflect the diversity among all of the structures represented in the test data sets. Future approaches will include reselection of the training and test data sets from the entire set of 137 molecules using hierarchical cluster analysis[89,90] which factors data sets into reasonably homogeneous subgroups with respect to their various physicochemical parameters. This will insure that the training set is representative of the entire range of structures currently in our database.

An additional variable to consider it the method of charge calculation. For this study Gasteiger–Marsili charges were used for all molecules. There have been no systematic studies comparing the method of charge calculation and the quality of the correlations from CoMFA. Kim and Martin made a preliminary investigation in this area and concluded that AM1 charges gave the best results for correlating the electronic effects of substituted benzoic acids.[24] While they compared correlations using AM1 and STO-3G charges, they did not derive models for charges calculated using the Gasteiger method. Recent publications have reported successful CoMFA analyses using both semiemperical and Gasteiger charges.[91,92] Our choice of Gasteiger charges was based primarily on speed and convenience of calculation as well as an assumed independence of charge calculation method on CoMFA results. Since more recent data indicates that the preferred method of charge calculation may be dependent on the molecules being studied as well as the target property being correlated, further CoMFA studies using semiemperically derived charges will be performed for purposes of comparison.

The results of our studies give strong support to both the active analog approach[93] used to define the alignment rule for the ACE series and the CoMFA methodology itself. In the absence of an experimentally known active site geometry, we have derived correlations which explain as much as 84% of the variance in activities among a set of 68 diverse ACE inhibitors using CoMFA steric and electrostatic potentials plus a zinc indicator variable. If the set of 68 ACE inhibitors is divided into three classes and

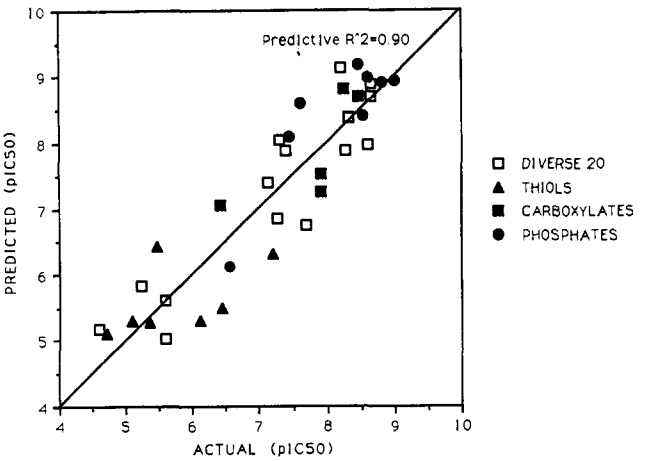


Figure 12. Predicted vs actual pIC_{50} for the collective ACE data set of test molecules having residuals <1.0, using CoMFA parameters plus a zinc indicator equal to 10 (QSAR2). The predictive model was derived using four principal components having a cross-validated R^2 = 0.64 and a conventional R^2 = 0.84.

correlations are derived for each class, CoMFA parameters alone explain 79% to 99% of the variance in activities. It is notable that we can derive statistically significant correlations in spite of the fact that CoMFA does not consider hydrophobicity or solvation. This demonstrates the intrinsic power and utility of the method. In further support of the active analog approach, the crossvalidated results of the ACE series are equivalent to those of the thermolysin series (crossvalidated R^2 = 0.65–0.70), for which the alignment rule was defined by crystallographic data.

Among the existing CoMFA analyses in the literature, only Greco et al.[27] have analyzed a large data set of noncongeneric structures. Their analyses of 39 diverse muscarinic agonists produced correlations with conventional R^2 values in the range of 0.87–0.90, also using an alignment rule derived from the active analog approach. Their objective, however, was to determine the optimal conditions for a CoMFA run by systematically varying the method of charge calculation, lattice point spacing, and standard deviation cut-off. Their results indicated that correlations were convergent to a narrow range of R^2 values, independent of the conditions chosen. Their analyses, however, did not include any attempts to predict activities of molecules outside the training set, i.e., the use of a test set. The predictions for the molecules outside the training sets are a valid test of the predictive ability of the model, rather than just a confirmation of self-consistency of the derived model. The predictive correlations presented represent a total of 66 diverse inhibitors that were not chosen as analogs of compounds present in the training set but by selecting recently published papers and testing all compounds in the paper. Using the ACE model with a zinc indicator of 10 as an example, 35 out of the 66 predicted molecules had residuals less than one log value. The predictive R^2 values for these 35 molecules are 0.91, 0.93, 0.95, and 0.24 for the set of 20 diverse inhibitors, the carboxylates, the phosphates, and the thiolates, respectively. The predictive R^2 for the collective set of these 35 test molecules is 0.90 (Figure 12). Of the 31 inhibitors with residuals greater than 1.0, 8 are carboxylates, 12 are phosphates, and 11 are thiols. Clearly, no single class of inhibitors dominates the distribution of residuals. Considering both the composition and the method of selection of the test data sets, the fact that more than 50% of the molecules are predicted with correlations greater than R^2 = 0.90 lends strong support to the use of CoMFA as a tool for QSAR development.

Conclusions

CoMFA steric and electrostatic potentials are useful parameters for deriving 3-D QSAR models both in the presence and absence

(88) Wold, S.; Sjostrom, M.; Carlson, R.; Ludstedt, T.; Hellberg, S.; Skagerberg, B.; Wikstrom, C.; Ohman, J. *Anal. Chim. Acta* **1986**, *191*, 17–32.
(89) Hansch, C.; Unger, S. H.; Forsythe, A. B. *J. Med. Chem.* **1973**, *16*, 1217–1222.
(90) Bratchell, N. *Chemometrics Intell. Lab. Systems* **1989**, *6*, 105–125.
(91) Diana, G. D.; Kowalcsyk, P.; Treasurywala, A. M.; Oglesby, R. C.; Pevear, D. C.; Dutko, F. J. *J. Med. Chem.* **1992**, *35*, 1002–1008.
(92) McFarland, J. W. *J. Med. Chem.* **1993**, *35*, 2543–2550.
(93) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. D.; Dunn, D. A. Olson, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; Vol. 112, pp 205–226.

of experimentally known active site geometries. Correlations derived using only CoMFA steric and electrostatic potentials are equivalent, if not superior, to those derived using the conventional QSAR parameters such as log P, Hammett's s, Taft's E_s, and MR. The models derived from the thermolysin training set, for which the alignment rule was defined by crystal structures of enzyme–inhibitor complexes, were somewhat better, based on conventional R^2s, than those for the ACE series with its computationally derived active site geometry. This indicates a clear dependence on the alignment rule for the development of robust QSARs. However, cross-validated R^2 values for the two were comparable, which supports the use of the active analog approach as a reliable method for the derivation of active site geometries in the absence of experimentally derived structural data. Predictions for the ACE series were improved by the addition of a zinc-descriptor to explicitly define the type of zinc–ligand interaction, although no single model could consistently predict the activities of all classes of ACE inhibitors. More parameters than just CoMFA parameters are apparently necessary to describe the interactions which dictate ligand binding. Specifically, the CoMFA methodology ignores the thermodynamics of binding, namely the entropy loss due to complex formation and the hydrophobic effect. While CoMFA's greatest utility lies in its ability to provide insight into the topographical properties necessary for binding, additional parameters which consider all the factors determining the free energy of binding will be necessary if one is to derive QSARs that not only explain the variance in the training set but also can predict activities of new structures as well.

Acknowledgment. The authors gratefully acknowledge support for this research from the National Institutes of Health Grants GM24483 and HL07275.

Supplementary Material Available: Tables of structures and activities for the ACE training set; the ACE carboxylate, the ACE phosphate, and the ACE thiol predictive sets; predictions for the set of diverse ACE inhibitors and for the carboxylate-, phosphate-, and thiol-based ACE predictive set; X-scores from CoMFA analysis; and predictions for the thermolysin predictive set and coordinate data for molecules in the ACE (Appendix A) and thermolysin (Appendix B) series (95 pages). Ordering information is given on any current masthead page.