

Three-Dimensional Quantitative Structure–Activity Relationships of Sulfonamide Endothelin Inhibitors

Stanley R. Krystek, Jr.,[†] John T. Hunt,[‡] Philip D. Stein,[‡] and Terry R. Stouch*,[†]

Departments of Macromolecular Modeling and Chemistry, Cardiovascular Agents, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543-4000

Received September 16, 1994[⊗]

A three-dimensional quantitative structure–activity relationship (QSAR) using steric and electrostatic fields (comparative molecular field analysis, CoMFA) applied to 36 aryl sulfonamides assayed for endothelin receptor subtype-A (ET_A) antagonism provided high cross-validation correlations (0.7) and showed promising predictive ability. The results were validated through trials using scrambled activities as well as trials using scrambled orientation of molecules. CoMFA was used to discriminate between alternate hypothetical biologically active conformations. CoMFA was also used to discriminate between two different molecular superpositions representing possible positioning within the receptor binding site. The preferred superposition supports hypotheses that suggest Tyr¹²⁹ in the ET_A receptor as a key residue for antagonist binding. Significant CoMFA results were obtained when crudely optimized geometries and simple charge schemes were used. The results improved on refinement, most substantially with refinement of the atomic charges.

Endothelin was originally isolated from porcine endothelial cells¹ and is the most potent vasoconstrictor known. There exist three endothelin isopeptides (ET-1, ET-2, and ET-3),² each containing 21 amino acids and two disulfide bonds. The actions exerted by endothelins include vasoconstriction, vasodilation, and pressor and depressor effects as well as mitogenicity.^{3–7} These diverse actions are attributed to the existence of multiple endothelin receptor subtypes with discrete cellular distributions and functions. Two receptor subtypes have been identified on the basis of molecular and pharmacological evidence.^{8–10} Receptor subtype-A (ET_A) binds ET-3 with much lower affinity than ET-1 or ET-2, while subtype-B (ET_B) binds the isopeptides with equivalent affinities. Due largely to the potent and long-acting vasoconstrictor effects of endothelin isopeptides, endothelins have been proposed as targets for therapeutic intervention in numerous diseases.¹¹

The discovery of selective endothelin receptor antagonists will facilitate investigations of the pathophysiology for endothelin isopeptides. Recent reports have described a large number of endothelin antagonists which differ in their receptor subtype selectivity. BQ-123, a cyclic pentapeptide discovered by random screening of fermentation products from *Streptomyces misakiensis*, has become established as a highly selective ET_A receptor antagonist.^{12,13} Clozel et al. provided the first report of a nonpeptide, orally active, nonselective endothelin receptor antagonist, Ro462005.¹⁴ We have recently reported the discovery of a class of benzenesulfonamide ET_A antagonists and the development of the naphthalenesulfonamide BMS-182874, a potent, orally active, and highly ET_A selective antagonist.¹⁵ In order to understand the structural and conformational requirements for its endothelin receptor affinity, a molecular-modeling-based three-dimensional quantitative struc-

ture–activity relationship (QSAR) was used to study the naphthalenesulfonamide ET_A receptor antagonists.

QSAR can be used to develop a better understanding of the biochemical properties of a set of molecules and to develop predictions of biological activities based upon a set of molecular properties. As data sets become large and diverse, SAR can become complicated and require computerized techniques. It has been proposed that the most significant interactions between a ligand and its receptor are of a nonbonded nature. Upon the basis of this premise, differences in binding and/or activity are often ascribed to steric and electrostatic variations between structures. The method of comparative molecular field analysis (CoMFA) can be used to develop a 3-D QSAR model (pharmacophore) describing the three-dimensional structure–activity relationships for a series of compounds.¹⁶ One advantage of such an approach is the graphical representation of the results of the analyses as three-dimensional grids which represent regions where steric bulk detracts or contributes to activity as well as contours displaying regions where positive or negative charge favorably contributes to activity. These data and representations can be used to guide future syntheses and develop hypotheses for ligand–receptor interactions.

Here we present the first in a series of studies that apply CoMFA methodology to rationalize the relationship between ET_A selective aryl sulfonamide antagonist structures and their activities. Our overall goal is to better understand the specificity and selectivity of various ligands for the ET_A receptor subtype. In addition, since CoMFA is a new and evolving technique, we use this set of compounds to probe aspects of CoMFA methodology. For example, we present several approaches to validate the CoMFA results, and we probe the level of precision required of the molecular models and atomic charges. Finally, the CoMFA results are also used to identify the putative bioactive conformation of the sulfonamide from possible alternatives.

Methods

Determination of Bioactive Conformation. A total of 36 compounds were included in this study. All structures

* Communicating author: Terry R. Stouch, Department of Macromolecular Modeling, Bristol-Myers Squibb Research Institute, P.O. Box 4000, Princeton, NJ 08543-4000. Fax: (609) 252-6030.

[†] Department of Macromolecular Modeling.

[‡] Department of Chemistry.

[⊗] Abstract published in *Advance ACS Abstracts*, January 1, 1995.

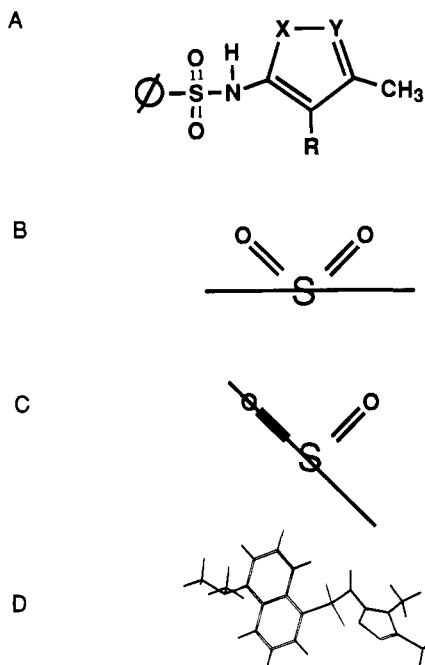


Figure 1. (A) The phenyl group attached to the sulfonamide sulfur can be a substituted phenyl or a 1- or 2-substituted naphthyl group. The sulfonamide moiety is attached via the nitrogen to an isoxazole analog. X = O and Y = N for the 5-isoxazolyl and X = N and Y = O for the 3-isoxazolyl group. R can be either H or CH₃. See Table 1 for a complete description of all compounds used. (B and C) Views down the C_{aryl}-S bond showing positions of the aryl ring. (D) Overhead view showing the working active conformation and the relationship between the two rings.

contained the aryl (substituted phenyl or 1- or 2-naphthyl bonded to sulfur) sulfonamide moiety with an isoxazole analog bonded to the amide nitrogen (Figure 1). Tertiary sulfonamides were always of low activity, and our study does not include them. For the core of these molecules then, this left three conformational degrees of freedom: the torsion about the sulfur-aryl bond, that about the sulfur-nitrogen bond, and that about the nitrogen-isoxazolyl bond. All of these structures contained an isoxazole bound to the sulfonamide nitrogen. Most of the compounds contained a 3,4-dimethyl-5-isoxazole. In order to increase the diversity of the data set and provide a uniform sampling over the range of activities, three compounds were included that contained 4,5-dimethyl-3-isoxazole, one compound contained 3-methyl-4-bromo-5-isoxazole, and two compounds contained 5-methyl-3-isoxazole. These variations resulted in only modest variations in activity (<16-fold for otherwise identical congeners) relative to the 10 000-fold range in the entire data set. The same conformation was used for the 4-unsubstituted-3-isoxazole analogs as for the 4,5-dimethyl analogs since lack of the 4-methyl group would easily allow this conformation. The small variation in activity caused by these changes in the isoxazole ring resulted in only small contributions to the CoMFA fields which are not visible at the significance levels of the plots presented later.

The Cambridge Crystallographic Database (CSD)¹⁷ contains a large number of aryl sulfonamides (over 450), and we used that information to help suggest likely conformations for these molecules. (See the Appendix for summary information on the internal geometry of the sulfonamide group.)

Amide Planarity. The distance from the nitrogen to the plane defined by its three attached atoms was calculated. By far, this distance tended toward zero, indicating a preference toward an sp²-hybridized nitrogen. Since these compounds were assayed at a pH above the pK_a of the sulfonamide, the sp² hybridization was chosen for this study.

Aromatic Ring Position. The distribution of the torsion angles C-C-S-N bond indicated a clear preference for the ring to be at 90° to the S-N bond, hence for the plane of the ring to be equidistant from the sulfonyl oxygens (see Figure

1B). This would be the expected minimum energy conformation for a phenyl ring symmetrically substituted at positions 2 and 6. However, there was a considerable range to this value, and no positions appeared to be prohibited. In fact, a number of structures had values close to 0 for this torsion, indicating that the aryl ring bisected the O-S-O angle. Some had values of the O-S-C_{aryl}-C_{aryl} of 0, indicating that a C-C bond of the aryl ring eclipsed the S-O bond (see Figure 1C). Two related structures were found with 1-naphthyl rings linked to the sulfur. Both of these structures showed a severe tilt of this ring with the C₁-C₂ phenyl bond essentially eclipsing as S=O bond. Presumably, the interaction between the other S=O and the C8 substituent on the fused ring was sufficiently unfavored to cause this arrangement. The wide range of possible conformations about the S-aryl linkage, noted above, suggested that this eclipsed state might not necessarily be highly strained. This ring position would be the most likely one for the 2-substituted phenyl sulfamides, also. This is, perhaps, understandable considering the 1.76 Å S-C_{aryl} bond length and the 120° aryl C-C-C bond angle.

Heteroring Position. The O-S-N-C torsion angle appeared much more restricted than the C-S torsion and had a clear preference for angles of about 60–70°. This presumably is due to the combined effects of the planarity of the nitrogen and whatever repulsion might occur between the isoxazole ring and the S=O. These angles, and the apparent preference of the nitrogen to be planar, suggested that the heteroring could be placed in only two general locations, on either side of the sulfone.

Working Active Conformation. These results suggested an active conformation for the sulfonamides. First, the low-energy conformations of the more active naphthyl and 2-substituted phenyl compounds would exist with a considerable tilt of the ring relative to the O-S=O group (where nontilted would place the plane of the ring equidistant from the oxygens). Second, the isoxazole ring would exist in a relatively restricted position trans (relative to the plane defined by C-S-N) to the second fused ring of the naphthyl compounds or to the side of the ring with the 2-substituent; a cis arrangement would put the two rings in close proximity (see Figure 1D). The N-isoxazole torsion was determined by conformational analysis which showed a limited number of possible positions.

Molecular Superposition. Three different superpositions were evaluated. Considering their similarity, the first and most obvious superposition of these molecules was a direct maximal overlap of the atomic positions within the restrictions of the low-energy conformations, described above. In this first overlap scheme, the sulfonamide and C1 of the aryl group were directly superimposed. The isoxazole moieties were maximally superimposed. The aryl rings were left in their presumed low-energy conformations wherein the planes of those phenyl and 1-naphthyl rings, which were symmetrically substituted at positions 2 and 6, were equidistant from the sulfone oxygens and those that were unsymmetrically substituted and the 2-naphthyl rings were placed with the C1-C2 bond eclipsing one S-O bond (Figure 2A,B).

The most active compounds contained the 1-naphthyl¹⁸ and the ortho-substituted phenyl rings (unpublished data) and would have an aromatic C-C bond eclipsing an S=O bond, as mentioned above. The symmetrically substituted (2, 6) phenyls and naphthyls could achieve this conformation also, although it would not be a low-energy conformer for these structures (rigid body rotation showed less than a 0.5 kcal/mol difference using the Tripos force field). The second overlap modified the conformation of these other rings so that their planes coincided with those of the most active compounds. This required only a small (approximately 30°) rotation about the S-C_{aryl} bond (Figure 2C).

The third superposition emphasized the role of the sulfonamide group and the aromatic ring substituents in a binding scheme. The conformations were maintained as in the first overlap scheme, but rather than maximally overlapping all atoms in the structures, a smaller subset of atoms was maximally overlapped including the sulfonamide sulfur and nitrogen and the amine nitrogen substituent on the aryl rings. Additionally, the isoxazole moieties were roughly overlaid by including the two methyl groups on this ring in the maximal

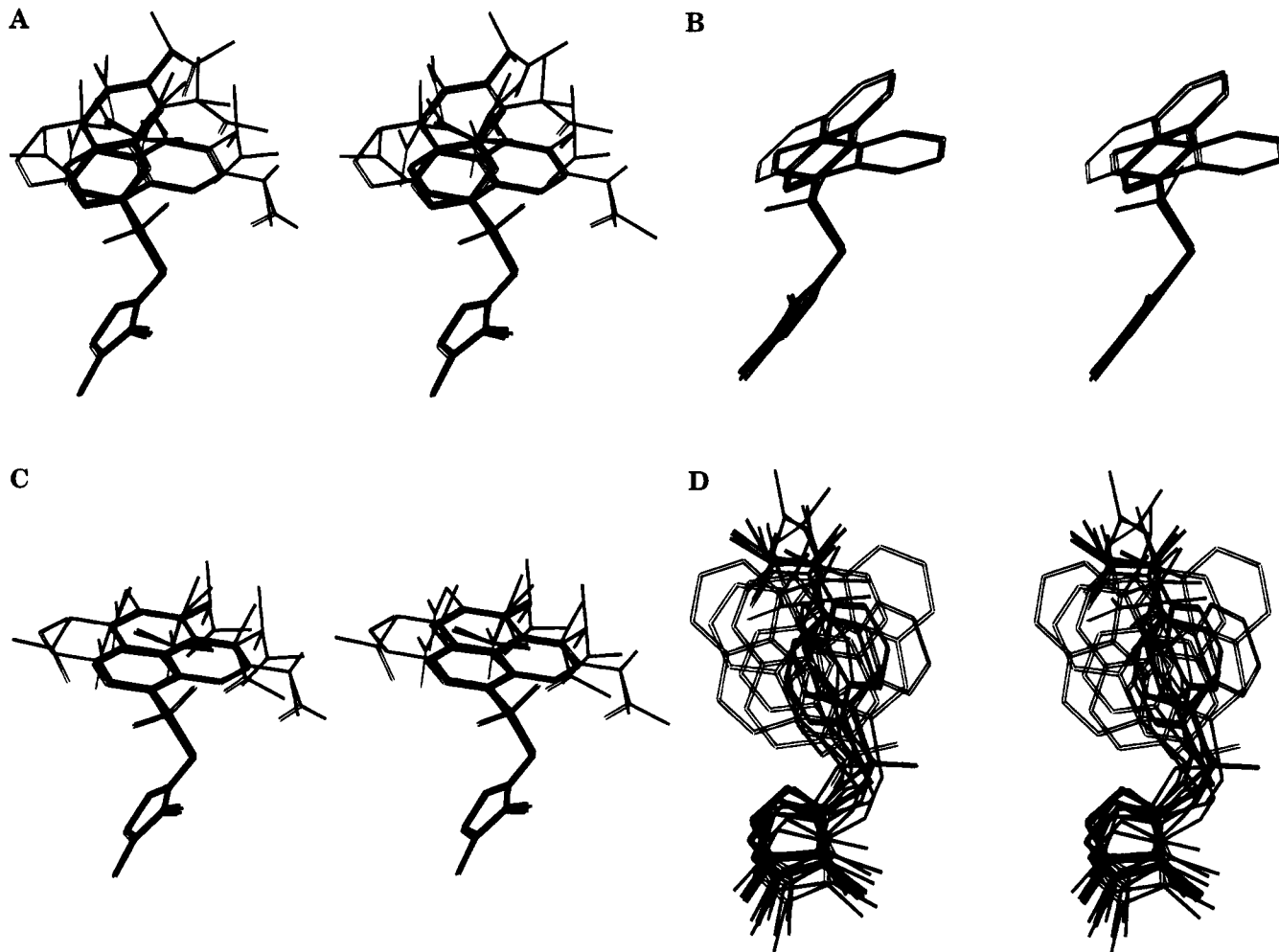


Figure 2. Stereoviews of the molecular superpositions. (A) Superposition of all 36 compounds based upon maximal atomic overlap. (B) As in A but with all hydrogens removed as the differences between 1-naphthyl and 2-naphthyl aryl rings can be seen. (C) Superposition with the 2-naphthyl and substituted phenyl rings eclipsing the S=O bond similar to the 1-naphthyl compounds. (D) Superposition emphasizing the sulfonamide group and the substituents attached to the aryl ring.

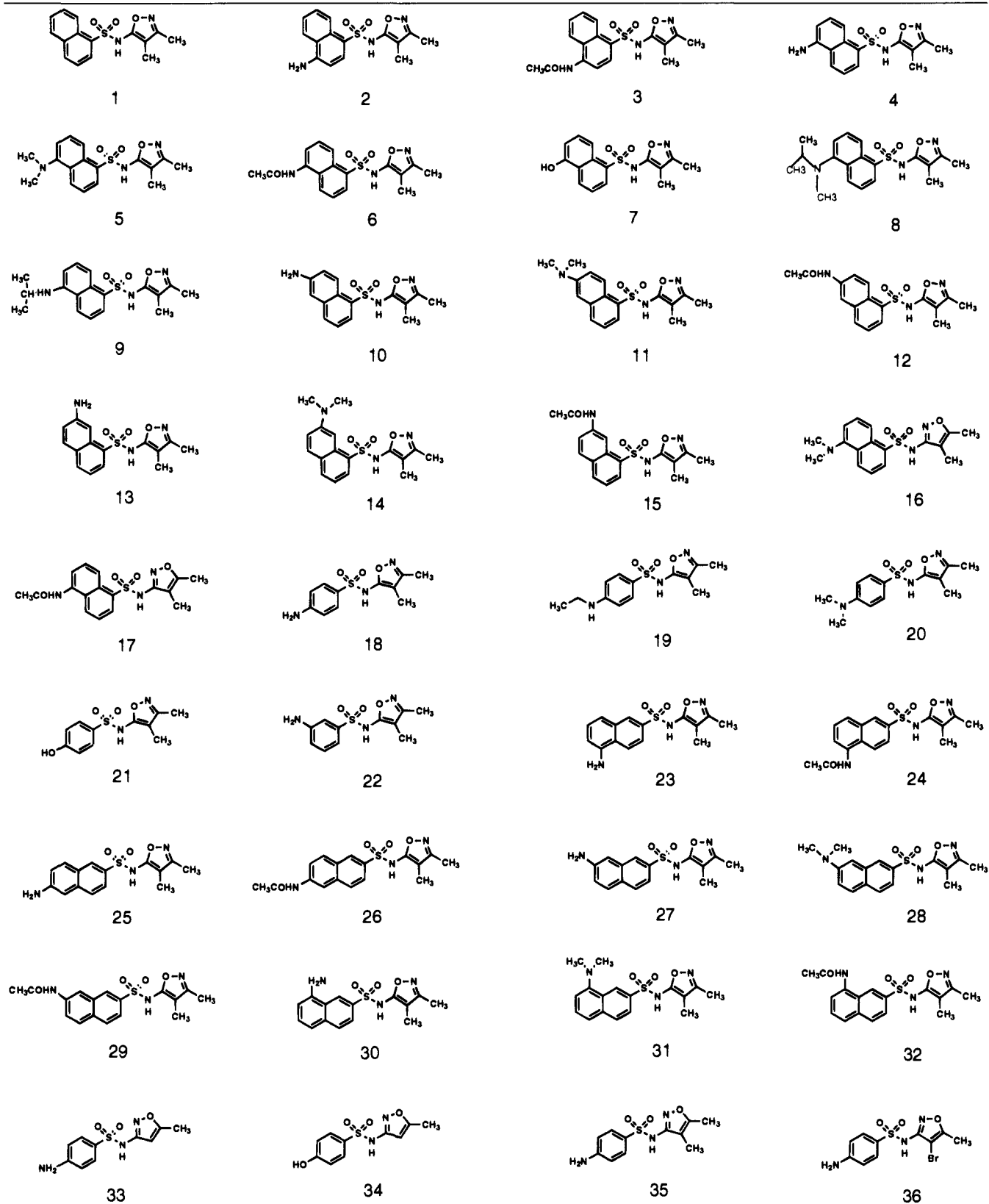
overlap (Figure 2D). Another minor variation included the S-linked carbon in the aryl ring in the superposition in order to bring the aryl rings of all compounds somewhat in line.

Optimization of Molecular Structures. The required degree of accuracy of the three-dimensional geometry of molecular models used in CoMFA analysis has not yet been studied. However, no doubt it largely depends on the relative variations within the series of molecules undergoing analysis. Obtaining optimized structures can be problematical and time-consuming depending on the desired degree of accuracy, availability of force field parameters, size, number of conformational degrees of freedom, and elemental composition. Because of this, we first established the existence of an SAR within the data set through the use of quickly generated "crude" models. Once an SAR was established, we then took the time to refine the models and redo the CoMFA. We wanted to establish if such crude models could provide information before taking the time to optimize the structures. The "crude" structures were based on the results of the Cambridge Structural Database searches with aromatic rings and heterorings positioned as previously described. Initially the aromatic ring and heterorings were obtained from the Tripos fragment library which was derived from average geometries from the Cambridge Structural Database.¹⁷ The Tripos fragment library was also used to add the substituents to the rings. In some cases, this led to nonoptimum conformations including planar geometries for the amino groups. Once the SAR was established, the optimized geometries for the substituents were determined using Mopac¹⁹ (PM3). The sulfonamide group was maintained at the crystallographically determined bond lengths and angles which were relatively invariant between crystal structures.

Atomic Partial Charges. Even more so than for the molecular geometries, calculation of "realistic" partial point

charges can be time-consuming and problematical.^{20,21} To our knowledge, for the charges as for the structures, little has been done to compare the relative advantages to a CoMFA of different ways of calculating the charges. Kim and Martin, prior to initiating a CoMFA study, evaluated several charge calculation methods and choose AM1 over several others.²² As for the molecular geometries, we sought to establish the existence of a useful SAR using approximate charges prior to more detailed computations. In this crude charge scheme, only the sulfonamide and those portions of the molecules that varied structurally were provided charges that were based on previous experience with potential derived charges determined for a large number of compounds. (See the Appendix for a detailed list of the charge scheme.) Once the SAR was established using this crude set of charges, "refined" potential derived charges were calculated using Mopac 6.0¹⁹ (PM3 and MNDO Hamiltonians) and the geometry optimized structures.

Biological Activities and Assay. Compounds were tested for their ability to inhibit [¹²⁵I]-endothelin-1 binding to A10 rat thoracic aorta smooth muscle cells as previously described.^{18,23} Briefly, A10 cells were cultured, detached with trypsin, collected, and stored at -70 °C. Cells were thawed, homogenized, and centrifuged in order to obtain a pellet containing cell membranes. The receptor binding assay had aliquots of the membranes incubated with radiolabeled endothelin-1 in either the absence or presence of competing ligand. The samples were filtered through glass fiber filters and washed, and the radioactivity retained in the filters was measured. Specific binding was calculated as the difference in binding of [¹²⁵I]-endothelin-1 attained in the absence and presence of excess "cold" endothelin-1. Analysis on the data was performed as described and reported as an IC₅₀ value.^{15,18,23}

Table 1. Structures of Compounds used in QSAR Analysis

Computational Methods. The comparative molecular field analysis (CoMFA) was performed using Sybyl 6.04 molecular modeling program.²⁴ The IC_{50} values were taken from previous work.^{15,18} Electrostatic (Coulombic) and steric (Lennard-Jones) potentials were sampled on a three-dimensional grid which was defined to extend at least 4 Å beyond the van der Waals surfaces of all molecules. The grid spacing was set at 1 Å (rather than the default of 2 Å) in all three dimensions. The finer grid density was used in order to capture more of the molecular details.²⁵ The interaction

energy between an sp^3 carbon probe and a point charge of 1.0 and each molecule in the set was calculated at each grid point as described previously.¹⁶

The QSAR table was constructed with rows containing the molecule names with structures and columns containing the dependent data (IC_{50}) as well as the individual steric and electrostatic field potential values at each grid point for each molecule. A separate QSAR table was created for each manipulation of the structure or change in calculation parameter.

Partial least squares (PLS) methodology was used to develop the relationship between the independent variables (steric and electrostatic properties) and the IC_{50} values. The optimum number of components in the final PLS analysis was determined by full cross-validation (leave-one-out) yielding 36 groups using a maximum number of seven components. The optimum number of components was determined to be that which yielded the highest cross-validated r^2 and lowest standard error of predictions (SE) as per established methodology.^{16,24} Following the cross-validated analysis a non-cross-validated analysis was performed using the optimum number of components previously identified. The non-cross-validated analyses were used to make predictions of activities and to analyze the CoMFA results. For both the cross-validated and non-cross-validated analyses, the σ used was 2.0 as we found that $\sigma = 0.0$ did not alter the calculated r^2 or SE significantly (see Table 3, trials A and B).

Discussion

This work addresses several questions concerning the specific structure-activity relationships of these molecules as well as some investigation into the CoMFA methodology itself. First, we present the most refined study demonstrating the "best" SAR that we obtained. We then present some verification studies of this study wherein either the activities or molecular orientations were randomized and the CoMFA recalculated. A study employing an alternative plausible overlay for these molecules is also presented. Also, CoMFA studies to differentiate between two distinct conformations for the phenyl and 2-naphthyl compounds are described. Finally, a time-consuming part of any study such as this can be generation of realistic molecular models and partial charges. We ask the question of how precise these structures and charges need to be in order to uncover useful SAR and to determine whether further effort is warranted.

Using the set of 36 molecules which were selected to represent diversity of structure (at positions in the aryl ring, Table 1) and activity (IC_{50} values 0.015–53 μ M, Table 2), several CoMFA analyses were initiated. The best results were obtained using the first overlay discussed in the Methods section, optimized structures, and PM3 charges (Table 3, trial A). Standard CoMFA procedure dictates that the optimum number of components be chosen at the point at which the PLS results degrade. Here, the cross-validated r^2 levels off at three and four components, and the SE increases only slightly at four components. However, the results at six components show the r^2 to be significantly greater than at four components, and the standard error is substantially less. The optimum number of components was selected by identification of the point at which the r^2 or SE values decreased significantly or were significantly greater than previous values. On the basis of this criteria, we selected six components for trial A ($r^2 = 0.70$, SE = 0.69). Table 4 contains the corresponding non-cross-validated analysis for the six components ($r^2 = 0.94$, SE = 0.30). The ratio of steric and electrostatic contributions to the final model is 69:31. These results indicate good agreement between the experimentally determined IC_{50} values and the predicted values derived from the CoMFA model. Although other CoMFA studies required other properties in order to explain activity,^{26–28} these data suggest that the steric and electrostatic properties of these molecules alone may be useful for providing information about these ligand-receptor interactions.

Figure 3 shows the contribution of the steric fields to the CoMFA. For both the steric and electrostatic fields,

Table 2. Sulfonamide Binding Affinity Data

compd ^a	IC_{50} ^b	BMS no.
1	3.1	183085
2	3.5	182499
3	3.4	182293
4	0.15	182542
5	0.015	182874
6	0.087	182401
7	0.068	184157
8	0.004	184289
9	0.005	183348
10	2.4	183820
11	0.84	184411
12	1.3	183545
13	1.45	184031
14	5.6	184212
15	0.87	183611
16	0.015	183240
17	0.016	183346
18	0.8	SQ7474
19	5.0	181737
20	84.0	181734
21	9.0	181977
22	1.3	182475
23	24.0	182228
24	2.7	182227
25	0.6	182489
26	3.8	182302
27	10.0	183984
28	1.4	184213
29	42.0	183640
30	9.8	183985
31	31.5	184214
32	27.5	183765
33	4.0	SQ28583
34	53.0	181241
35	0.26	182220
36	0.51	182400

^a Structures for these compounds can be found in Table 1.

^b Binding data from Stein et al.^{15,18}

the STDEV*COEFF of the QSAR equation was contoured. The contour map displays a region (colored green) where addition of steric bulk should be favorable. This region corresponds to the location of the groups attached to the 5 position of the 1-naphthyl compounds. Eight of the 36 compounds had increased activities when bulky substituents were attached to this position. There are also two regions where steric bulk is contraindicated (colored yellow). They are located on each side of the region where steric interactions are favorable. The region (to the right) represents a series of compounds with low activity. The 1-naphthyl compounds with groups at position 6 of the ring and the 2-naphthyl compounds with groups at the 8 position in the ring have bulk in this region of the space. The other (larger) region where steric interaction detracts from receptor binding was caused by groups located in the para position of the 2-substituted phenyl compounds, the 6 position of the 2-substituted naphthyl compounds, and the 4 position of the 1-substituted naphthyl compounds. These results suggest that the receptor can accommodate steric bulk in a narrow range of space, namely the 5 position of the 1-substituted naphthyl compounds.

The analysis of electrostatics is more complicated than sterics due, no doubt, to the long-range nature of the interactions. Figure 4 shows the electrostatic contour map from this CoMFA. A region (colored blue) where addition of positive charge would increase activity can be identified below the functional groups at either the 4 or 5 positions of the 1-substituted naphthyl compounds. There is also a region (colored red) where addition of a negatively charged group would be sug-

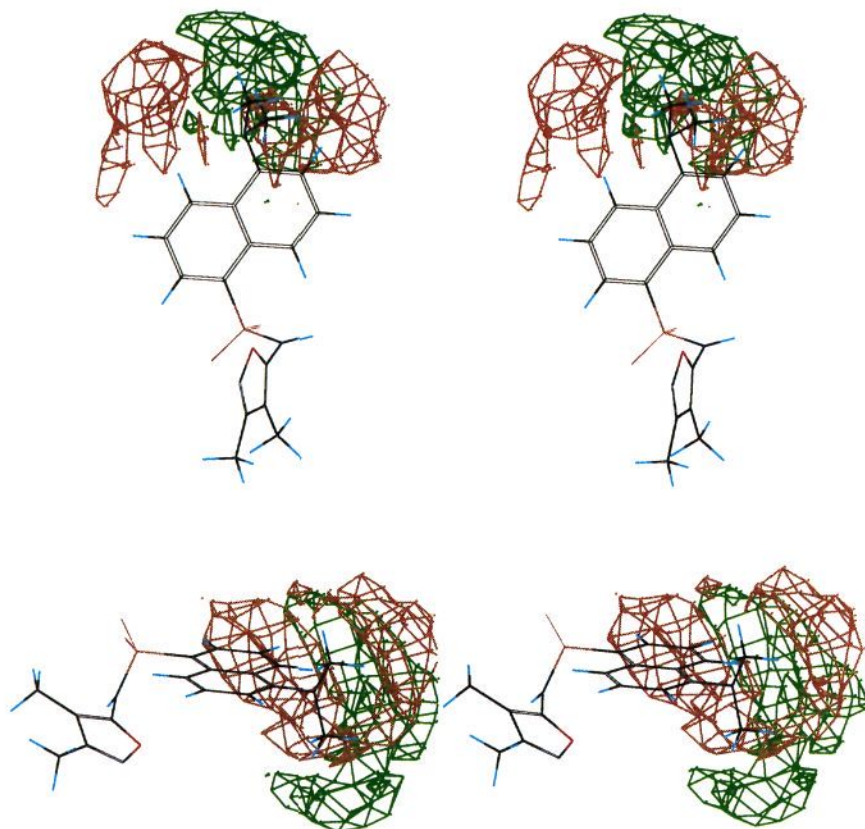


Figure 3. Stereoviews of the steric contours for the "best" (trial A) CoMFA STDEV*COEFF. One of the most potent molecules (number 5) is displayed for reference. Regions which favor steric interactions are colored green, while regions where steric interactions would detract from the biological activity are colored orange.

gested to increase activity. This region is probably caused by the carbonyl oxygen from the acetyl groups attached to the aryl ring.

Validation. Multivariate analyses such as these can often be lead to spurious results.²⁹⁻³³ In order to validate the CoMFA results, two approaches were used. First, the biological activities (IC_{50}) were randomized and reassigned to the 36 compounds, and the CoMFA analysis was redone (Table 3, trial D). The r^2 values for all number of components were essentially zero, and the SE was always very high. Even a non-cross-validated run (Table 4, trial D) using these data showed a low r^2 as well as a very large SE. Although we have no explanation, a curious observation is that the individual contributions of steric and electrostatic interactions to the final equations were similar to those for the trial using optimized structures with PM3 charges, whereas for the trial using randomized orientations, described below, the contributions are equal. Alternatively, a reviewer suggested one possibility is that the randomization of dependent variables or alignment simply averages out the fields in three-dimensional space, thus resulting in equal contributions of each field to the QSAR.

The second validation of the CoMFA results addressed the importance of the overlay orientation for the set of compounds. The orientations of all 36 compounds were randomly rotated in all three dimensions so there was essentially no overlay of any like functional groups. CoMFA analysis was rerun, and the results are presented in Table 3 (trial E). As would be expected for *three-dimensional scrambling of orientations*, the cross-validated r^2 was near zero for all components and the SE was very large. For the non-cross-validated run

(Table 4, trial E), using these data the r^2 was 0.99 and the SE was 0. The steric and electrostatic contributions to the final equation describing the data were 50:50, respectively. This demonstrates the importance of molecule orientation and overlay for the CoMFA methodology. It also showed that a specific overlay was required for significant CoMFA results.

A plot of the actual versus predicted values for compounds should be scattered in a straight line with a slope near 1.0 if the predicted values are similar to the actual data. Figure 5A displays the plot for the optimized structures with PM3 charges of actual IC_{50} versus calculated IC_{50} values. As demonstrated by the high-cross-validated r^2 , there is a good correlation between the calculated and actual data. In contrast, in Figure 5B,C are the corresponding plots for the CoMFA analyses with scrambled data (trial D) or randomized orientations (trial E), respectively. In both cases, the data appears randomly scattered.

Alternative Overlays and Conformations. We also carried out CoMFA analyses on several plausible alternative orientations of the compounds. The second orientation described in the Methods section involved rotation of the sulfur-aryl bond of the phenyl and 2-naphthyl rings to make the aromatic rings coplanar with those of the highly active 1-naphthyl analogs. Conformational analysis indicated that the 1-naphthyl compounds have a clear preference for orientation about the S-C bond which is attainable by the phenyl and 1-naphthyl compounds but which is not their lowest energy conformation. The cross-validated results from this overlay (Table 3, trial F) show a slight decrease in r^2 from 0.70 (trial A) to 0.66 for six components and an

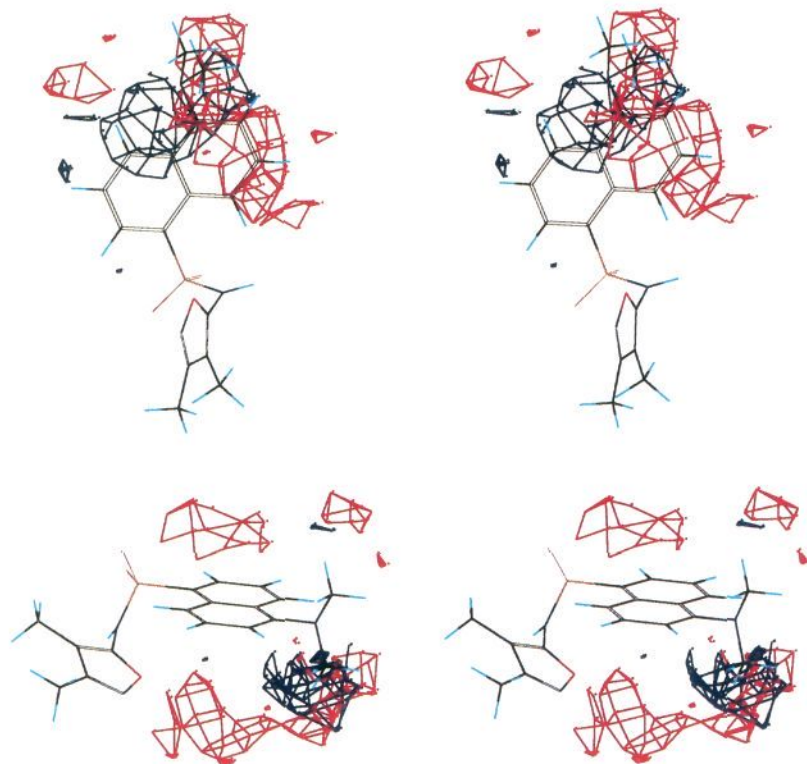


Figure 4. Stereoviews of the electrostatic contours for the "best" (trial A) CoMFA STDEV*COEFF. One of the most potent molecules (number 5) is displayed for reference. Regions colored blue show where addition of positive charge would increase activity, and regions colored red show where addition of negative charge would increase activity.

Table 3. Summary of Results for the Cross-Validated PLS Analyses

trial ^a	statistic	number of components						
		1	2	3	4	5	6	7
A	r^2	0.38	0.55	0.61	0.61	0.65	0.70	0.71
	SE	0.91	0.79	0.74	0.75	0.73	0.69	0.69
B	r^2	0.41	0.58	0.60	0.62	0.65	0.68	0.70
	SE	0.89	0.76	0.75	0.75	0.73	0.71	0.70
C	r^2	0.40	0.54	0.59	0.56	0.61	0.66	0.67
	SE	0.90	0.80	0.77	0.81	0.77	0.73	0.73
D	r^2	0.43	-0.16	-0.44	-0.50	-0.79	-0.76	-0.78
	SE	1.37	1.26	1.42	1.47	1.63	1.65	1.68
E	r^2	-0.18	-0.21	-0.02	-0.01	0.01	-0.02	-0.02
	SE	1.26	1.29	1.20	1.21	1.22	1.26	1.28
F	r^2	0.40	0.54	0.56	0.57	0.60	0.66	0.67
	SE	0.90	0.80	0.80	0.80	0.78	0.73	0.73
G	r^2	0.40	0.49	0.46	0.37	0.28	0.24	0.29
	SE	0.89	0.84	0.88	0.96	1.04	1.09	1.08
H	r^2	0.36	0.44	0.55	0.65	0.61	0.64	0.69
	SE	0.93	0.88	0.80	0.71	0.77	0.75	0.71
I	r^2	0.31	0.48	0.50	0.47	0.45	ND	ND
	SE	0.94	0.83	0.83	0.86	0.89	ND	ND

^a Trial A utilized optimized structures, PM3 charges, and $\sigma = 2.0$. Trial B utilized optimized structures, PM3 charges, and $\sigma = 0$. Trial C utilized optimized structures, MNDO charges, and $\sigma = 2.0$. Trial D utilized optimized structures, PM3 charges, $\sigma = 2.0$, and scrambled data. Trial E utilized optimized structures, PM3 charges, $\sigma = 2.0$, and scrambled overlay. Trial F utilized optimized structures, PM3 charges, $\sigma = 2.0$, and alternative conformation for 2-substituted compounds. Trial G utilized optimized structures, "simple" charges, and $\sigma = 2.0$. Trial H utilized "crude" structures, PM3 charges, and $\sigma = 2.0$. Trial I utilized "crude" structures, "simple" charges, and $\sigma = 2.0$.

increase in the SE from 0.69 (trial A) to 0.73 for this orientation. The results suggest that the previous orientation (using minimum energy conformations) is preferable. However, the CoMFA results differ only slightly, and this might indicate some conformational flexibility for the aryl rings of these compounds.

The third orientation emphasized the superposition of the sulfonamide sulfur and nitrogen and the aromatic ring substituents (Table 5, trial J). The results ($r^2 = 0.49$, SE = 0.84 at two components) improved with the enforcement of the isoxazole (methyl group) superposition (trial K, $r^2 = 0.58$, SE = 0.79 at four components) but were little changed on some reinforcement of the aryl ring superposition by including the S-linked aryl carbon in the superposition penalty function (trial L, $r^2 = 0.59$, SE = 0.82 at seven components). Although the results are significant, none approached the "best" results of trial A in Table 3. The major differences between this latest superposition and the "best" one were that all the functional groups were positioned close in space and the steric bulk was less well overlapped. That the results degraded suggests that close proximity of the functional groups might not be required.

These superposition results are especially interesting in that receptor-modeling and mutagenesis studies of the ET_A receptor have shown that Tyr¹²⁹ plays a key role in ligand binding.³⁴ The CoMFA results for the different superpositions qualitatively agree with hypotheses which suggest that Tyr¹²⁹ of the ET_A receptor plays a key role in binding the aromatic ring substituents of the antagonists. The tyrosine's several degrees of conformational freedom would allow it to reposition to compensate for slight variations in the location of the functional groups, and so enforcement of the superposition of these groups would not be necessary to explain activity. The enforcement causes a wider distribution of the steric bulk in space and creates a larger difference between the structures than is found in the other (first) overlay, suggesting a specific fit of steric bulk in the receptor binding site. Incorporation of these compounds and corresponding superpositions into the receptor models are underway using a variety of docking methods.

Table 4. Summary of the Results for the Non-cross-validated PLS Analyses

trial ^a	statistic	
A	r^2	0.94
	SE	0.30
	components ^c	6
B	contribution ^b	69:31
	r^2	0.94
	SE	0.31
C	components ^c	6
	contribution ^b	80:20
	r^2	0.94
D	SE	0.28
	components ^c	5
	contribution ^b	72:28
E	r^2	0.74
	SE	0.62
	components ^c	1
F	contribution ^b	63:37
	r^2	0.99
	SE	0.079
G	components ^c	5
	contribution ^b	50:50
	r^2	0.93
H	SE	0.32
	components ^c	5
	contribution ^b	67:33
I	r^2	0.91
	SE	0.37
	components ^c	2
J	contribution ^b	74:26
	r^2	0.94
	SE	0.30
K	components ^c	3
	contribution ^b	66:34
	r^2	0.91
L	SE	0.35
	components ^c	3
	contribution ^b	62:38

^a See Table 3. ^b Contribution of steric and electrostatic properties, respectively. ^c Number of components was determined from full cross-validated trials as described.

Precision of Models and Charges. In order to save time during our preliminary studies using these compounds, we examined the utility of using "crude", quickly determined, geometries and atomic charges. The quickest and simplest scheme had charges only on "important" or varying atoms (see Appendix) and used the crude geometries for all compounds (Table 3, trial I, $r^2 = 0.50$, SE = 0.83). The non-cross-validated analysis (Table 4, trial I) showed $r^2 = 0.91$, SE = 0.35, and a ratio of 74:26 for steric to electrostatic contributions. Given the simplistic nature of the model used to describe the structures ("simple" charges and sterics), it is encouraging that the r^2 produced such a significant value. As described previously, a cross-validated r^2 of 0.3 corresponds to a probability of chance correlation with activity of <0.05 ,^{24,35} thus, the current results are significant.

When the geometries were optimized, there was essentially no change in the CoMFA results (Table 3, trial G, $r^2 = 0.49$, SE = 0.84). However, when the charges were refined (PM3), the CoMFA results improved substantially ($r^2 = 0.65$, SE = 0.71) even though the crude geometry for molecules was used (Table 3, trial H). These CoMFA results are similar to the "best" results ($r^2 = 0.70$, SE = 0.69) which had refined charges (PM3) and optimized geometries, suggesting that utilization of refined charges is important for these CoMFA.

These data indicate that an approximate location of the steric bulk may be a good enough description for this set of molecules. The data also demonstrate that use of "simplified" charge schemes and crude structures may be used to evaluate QSAR procedures in order to determine early on in the study if a more comprehensive

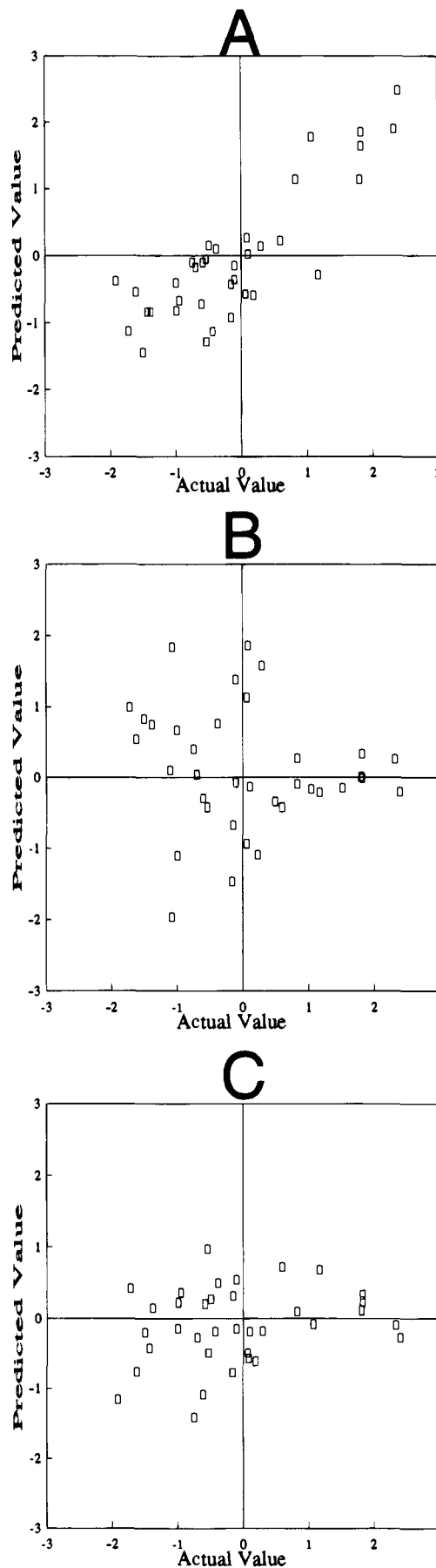


Figure 5. Predicted versus actual values plotted as $\log 1/IC_{50}$. (A) Plot for trial A showing high correlation of predicted versus actual values. (B) Plot for trial D (scrambled activities) showing low correlation of predicted versus actual values. (C) Plot for trial E (randomized orientations) which shows low correlation of the predicted values with the actual values.

Table 5. Summary of Cross-Validated PLS Analyses for Alternative Overlays

trial	statistic	number of components						
		1	2	3	4	5	6	7
J ^a	r ²	0.47	0.49	0.49	0.38	0.33	0.25	0.32
	SE	0.84	0.84	0.85	0.96	1.00	1.08	1.05
K ^b	r ²	0.44	0.48	0.55	0.58	0.57	0.58	0.56
	SE	0.86	0.85	0.80	0.79	0.81	0.81	0.85
L ^c	r ²	0.35	0.47	0.51	0.53	0.53	0.57	0.59
	SE	0.93	0.86	0.83	0.84	0.84	0.82	0.82

^a Trial J utilized optimized structures, PM3 charges, and $\sigma = 2.0$. The atoms overlaid were sulfonamide N and S as well as N in the ground attached to the aryl ring. ^b Trial K utilized optimized structures, PM3 charges, and $\sigma = 2.0$. The atoms overlaid were sulfonamide N and S as well as N in the group attached to the aryl ring and the two C's of the isoxazolyl ring to which the methyl groups are attached. ^c Trial L utilized optimized structures, PM3 charges, and $\sigma = 2.0$. The atoms overlaid were sulfonamide N and S as well as N in the group attached to the aryl ring and the two C's of the isoxazolyl ring to which the methyl groups are attached. In addition the S-aryl ring bond was rotated so as to make the 1-substituted compound aryl rings similar to the 2-substituted aryl rings.

Table 6. Summary of Data for Predictions using CoMFA

trial	statistic	number of components				
		1	2	3	4	5
Cross-Validated PLS						
M ^a	r ²	0.33	0.48	0.57	0.52	0.58
	SE	0.95	0.85	0.79	0.85	0.80
Non-cross-validated PLS						
N ^a	r ²	0.94				
	SE	0.30				
Calculated versus Experimental Values (log 1/IC ₅₀)						
molecule		predicted				actual
27		-0.41				-1.00
12		-0.11				-0.11
4		0.89				0.82
16		1.84				1.82

^a Trials M and N utilized optimized structures, PM3 charges, and $\sigma = 2.0$ for 32 selected compounds.

analysis is warranted. Additionally, it indicates that it is more important to have refined charge sets than refined molecular models.

Prediction. Four molecules were removed from the original 36 data set used in the previous CoMFA trials. The molecules were chosen on the basis of activity (one with high, one with low, and two of moderate activity). CoMFA was redone for the remaining 32 compounds using the overlay which provided the "best" previous CoMFA results (maximum overlay of all atoms), optimized geometries, and PM3 charges.

The CoMFA results for the 32 compounds are presented in Table 6. The $r^2 = 0.58$ for five components was close to 0.57 for three components, but five components were chosen due to SE being lower, 0.80 compared to 0.85. The non-cross-validated results for five components are also found in Table 6 and show $r^2 = 0.94$ and SE = 0.30. This CoMFA was applied to the four omitted compounds. The actual versus predicted values for the four compounds tested show good agreement overall (Table 6). The differences for the active and moderately active compounds were essentially zero, and the predicted value for the inactive compounds showed a significant but not extreme difference.

Conclusions

For this series of 36 endothelin receptor subtype-A antagonists, CoMFA provided significant correlation of

steric and electrostatic fields with biological activity and also provided predictions which agreed well with experimental values. CoMFA was used to assist in discriminating between possible bioactive conformations of the compounds, and the best correlations were obtained for compounds in their lowest energy conformations. CoMFA was also used to discriminate between different molecular overlay schemes. The best results were obtained for maximal atomic overlap of the structures. That an alternate overlay which superimposed functional groups gave slightly worse results might support a hypothesis suggesting a binding role for the orientationally variable Tyr¹²⁹ in the ET_A receptor.

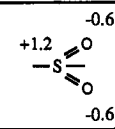
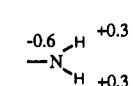
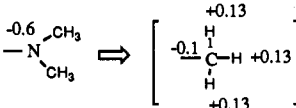
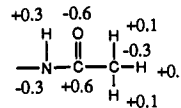
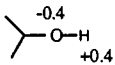
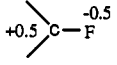
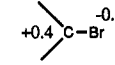
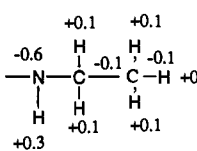
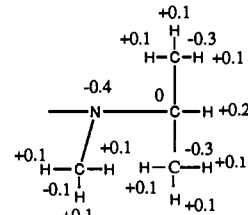
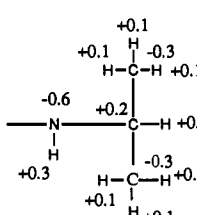
The significance of the CoMFA results were validated through randomization trials of both biological activities and the molecular superpositions, both of which yielded insignificant cross-validation results. Finally, we tested the importance of "simple" charge schemes versus refined charge sets and crude molecular geometries versus optimized geometries. Significant CoMFA results were obtained even with crude geometries and "simple" partial charge schemes. That the results improved substantially when the charges and geometries were refined further validates the physical significance of the results. Refining the charges improved results more than did refinement of geometries.

Appendix

Table A-1. C-SO₂-N Internal Structure

	S-C	S-N	S=O	C-S-N	O=S=O
mean	1.756	1.633	1.433	106.800	119.249
SD sample	0.014	0.020	0.012	1.829	1.284
SD mean	0.001	0.002	0.001	0.166	0.115
minimum	1.727	1.563	1.405	96.445	115.342
maximum	1.794	1.686	1.485	110.405	121.757
nobs	124	124	124	122	124

Table A-2. "Simple" Charge Scheme

References

- (1) Yanagiawa, M.; Kurihara, H.; Kimura, S.; Tomobe, Y.; Kobayashi, M.; Mitsui, Y.; Yazaki, Y.; Goto, K.; Masaki, T. A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature (London)* **1988**, *332*, 441–415.
- (2) Inoue, A.; Yanagisawa, M.; Kimura, S.; Kasuya, Y.; Miyuchi, T.; Goto, K.; Masaki, T. The human endothelin family: Three structurally and pharmacologically distinct isopeptides predicted by three separate genes. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2863–2867.
- (3) Simonson, M. S.; Dunn, M. J. Endothelin-1 stimulates contraction of rat glomerular mesangial cells and potentiates β -adrenergic-mediated cyclic adenosine monophosphate accumulation. *J. Clin. Invest.* **1990**, *85*, 1365–1372.
- (4) Secrest, R. J.; Cohen, M. L. Endothelin: Differential effects in vascular and nonvascular smooth muscle. *Life Sci.* **1989**, *45*, 1365–1372.
- (5) Warner, T. D.; de Nucci, G.; Vane, J. R. Rat endothelin is a vasodilator in the isolated perfused mesentery of the rat. *Eur. J. Pharmacol.* **1989**, *159*, 325–326.
- (6) Martin, E. R.; Marsden, P. A.; Brenner, B. M.; Ballerman, B. J. Identification and characterization of endothelin binding sites in rat renal papillary and glomerular membranes. *Biochem. Biophys. Res. Commun.* **1989**, *162*, 130–137.
- (7) Weber, H.; Webb, M. L.; Serafino, R.; Taylor, D. S.; Moreland, S.; Norman, J.; Molloy, C. J. Endothelin-1 and angiotensin-II stimulated delayed mitogenesis in cultured rat aortic vascular smooth muscle cells: evidence for common signaling mechanisms. *Mol. Endocrinol.* **1994**, in press.
- (8) Arai, H.; Hori, S.; Aramori, I.; Ohukubo, H.; Nakanishi, S. Cloning and expression of a cDNA encoding an endothelin receptor. *Nature (London)* **1990**, *348*, 730–732.
- (9) Sakurai, T.; Yanagisawa, M.; Takuwa, Y.; Miyazaki, H.; Kimura, S.; Goto, K.; Masaki, T. Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. *Nature (London)* **1990**, *348*, 732–735.
- (10) Rubanyi, G. M.; Parker Botelho, L. H. Endothelins. *FASEB J.* **1991**, *5*, 2713–2720.
- (11) Nayler, W. Endothelin: isoforms, binding sites and possible implications in pathology. *Trends Pharmacol. Sci.* **1990**, *11*, 96–99.
- (12) Ihara, M.; Fukuroda, T.; Saeki, T.; Masaru, N.; Kojiri, K.; Suda, H.; Yano, M. An endothelin receptor antagonist isolated from *Streptomyces Misakiensis*. *Biochem. Biophys. Res. Commun.* **1991**, *178*, 132–137.
- (13) Ihara, M.; Noguchi, K.; Saeki, T.; Fukuroda, T.; Tsuchida, S.; Kimura, S.; Takehiro, F.; Ishikawa, K.; Nishikibe, M.; Yano, M. Biological profiles of highly potent novel endothelin antagonists selective for the ET_A receptor. *Life Sci.* **1992**, *50*, 247–255.
- (14) Clozel, M.; Breu, V.; Burri, K.; Cassal, J.-M.; Fischli, W.; Gray, G. A.; Hirth, G.; Loffler, B.-M.; Muller, M.; Neidhart, W.; Ramuz, H. Pathophysiological role of endothelin revealed by the first orally active endothelin receptor antagonist. *Nature (London)* **1993**, *365*, 759–761.
- (15) Stein, P. D.; Hunt, J. T.; Floyd, D. M.; Moreland, S.; Dickenson, K. E. J.; Mitchell, C.; Liu, E. C.-K.; Webb, M. L.; Murugesan, N.; Dickey, J.; McMullen, D.; Zhang, R.; Lee, V. G.; Serafino, R.; Delaney, C.; Schaeffer, T. R.; Kozlowski, M. The discovery of sulfonamide endothelin antagonists and the development of the orally active ET_A antagonist BMS-182874. *J. Med. Chem.* **1994**, *37*, 329–331.
- (16) 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.
- (17) Cambridge Crystallographic Data Centre, 12 Union Rd., Cambridge CB2 1EZ, U.K.
- (18) Stein, P. D.; Floyd, D. M.; Kozlowski, M.; Bisaha, S.; Dickey, J.; Girota, R. N.; Lee, V. G.; Liu, E. C.-K.; McMullen, D.; Mitchell, C.; Moreland, S.; Murugesan, N.; Serafino, R.; Webb, M.; Zhang, R.; Hunt, J. T. The discovery and structure-activity relationships of sulfonamide ET_A-selective antagonists. Manuscript to be submitted.
- (19) The program Mopac is available from Quantum Chemistry Program exchange no. 455.
- (20) Stouch, T. R.; Williams, D. E. Conformation dependence of electrostatic potential derived charges of a lipid headgroup: Glycerylphosphorylcholine. *J. Comput. Chem.* **1992**, *13*, 622–632.
- (21) Stouch, T. R.; Williams, D. E. On the conformational dependence of electrostatic potential derived charges: Studies of the fitting procedure. *J. Comput. Chem.* **1993**, *14*, 858–866.
- (22) 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.
- (23) Hunt, J. T.; Lee, V. G.; Stein, P. D.; Hedberg, A.; Liu, E. C.-K.; McMullen, D.; Moreland, S. Structure-activity relationships of monocyclic endothelin analogs. *Bio-Org. Med. Chem. Lett.* **1991**, *1*, 33–38.
- (24) The program Sybyl 6.0 is available from Tripos Associates, St. Louis, MO.
- (25) Stouch, T. R.; Jurs, P. C. A simple method for the representation, quantification, and comparison of the volumes and shapes of chemical compounds. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4–12.
- (26) Kellog, G. E.; Semus, S. F.; Abraham, D. J. HINT: A new method of empirical field calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545–552.
- (27) DePriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: A comparison of CoMFA models based on deduced and experimentally determined active site geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372–5384.
- (28) Waller, C. L.; Marshall, G. R. Three-dimensional quantitative structure-activity relationship of angiotensin-converting enzyme and thermolysin inhibitors. II. A comparison of CoMFA models incorporating molecular orbital fields and desolvation free energies based on active-analog and complementary-receptor-field alignment rules. *J. Med. Chem.* **1993**, *36*, 2390–2403.
- (29) Stouch, T. R.; Jurs, P. C. Monte Carlo Studies of the Classifications Made by Nonparametric Linear Discriminant Functions 2. Effects of Nonideal Data. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 92–98.
- (30) Stouch, T. R.; Jurs, P. C.; Monte Carlo Studies of the Classifications Made by Nonparametric Linear Discriminant Functions. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 45–50.
- (31) Stouch, T. R.; Jurs, P. C. Chance Factors in Pattern Recognition Analysis Using Nonparametric Linear Discriminant Functions. *Quant. Struct.-Act. Rel. Pharmacol. Chem., Biol.* **1986**, *5*, 57–61.
- (32) Stouch, T. R.; Jurs, P. C. Computer-Assisted Studies of Molecular Structure and Genotoxic Activity Using Pattern Recognition Techniques. *Environ. Health Perspect.* **1985**, *61*, 329–343.
- (33) Topliss, J. G.; Edwards, R. P. Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* **1979**, *22*, 1238–1244.
- (34) Krystek, S. R., Jr.; Patel, P. S.; Rose, P. M.; Fisher, S. M.; Kienzle, B. K.; Lach, D. A.; Liu, E. C.-K.; Lynch, J. S.; Novotny, J.; Webb, M. L. Mutation of a peptide binding site in transmembrane region of a G protein-coupled receptor accounts for endothelin receptor subtype selectivity. *J. Biol. Chem.* **1994**, *269*, 12383–12386.
- (35) 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. Methodol.* **1990**, *3*, 47–59.
- (36) Urade, Y.; Fujitani, Y.; Oda, K.; Watakabe, T.; Umemura, I.; Takai, M.; Okada, T.; Sakata, K.; Karaki, H. An endothelin B receptor-selective antagonist: IRL 1038, [Cys¹¹-Cys¹⁵]-endothelin-1 (11-21). *FEBS Lett.* **1992**, *311*, 12–16.

JM940618H