

Prasanna Datar · Prashant Desai · Evans Coutinho
Krishna Iyer

CoMFA and CoMSIA studies of angiotensin (AT1) receptor antagonists

Received: 17 May 2002 / Accepted: 16 July 2002 / Published online: 2 October 2002
© Springer-Verlag 2002

Abstract Two 3D-QSAR methods – CoMFA and CoMSIA – were applied to a set of 38 angiotensin receptor (AT1) antagonists. The conformation and alignment of molecules were obtained by a novel method – consensus dynamics. The representation of biological activity, partial charge formalism, absolute orientation of the molecules in the grid, and grid spacing were also studied for their effect on the CoMFA models. The models were thoroughly validated through trials using scrambled activities and bootstrapping. The best CoMFA model had a cross-validated correlation coefficient (q^2) of 0.632, which improved with “region focusing” to 0.680. This model had a “predictive” r^2 of 0.436 on a test series that was unique and with little representation in the training set. Although the “predictive” r^2 of the best CoMSIA model, which included steric, electrostatic, and hydrogen bond acceptor fields was higher than that of the best CoMFA model, the other statistical parameters like q^2 , r^2 , F value, and s were unsatisfactory. The contour maps generated using the best CoMFA model were used to identify the structural features important for biological activity in these compounds.

Keywords AT1 receptor antagonists · CoMFA · CoMSIA · Consensus dynamics

Introduction

The renin–angiotensin system (RAS) is an important element in regulation of blood pressure and maintenance of

electrolyte balance. [1] The RAS system elaborates several points for intervention and development of therapeutics for controlling hypertension. The angiotensin-converting enzyme (ACE) and rennin inhibitors have an established place in the management of hypertension. [2, 3] Because the active component in the RAS is angiotensin II (AII), antagonism of AII at the receptor level (AT1) represents the most effective and specific way of RAS modulation. The first non-peptide AT1 receptor antagonist was Losartan developed by Dupont Merck Pharm. Co. [4]

There are some reports on the development of a 3D-pharmacophore and a 3D-QSAR (CoMFA) model for non-peptide AT1-receptor antagonists, which appeared in the mid-1990s. [5, 6, 7] Also, an alignment strategy for just six AT1 receptor antagonists by combined use of a simulated annealing method and cluster analysis has appeared. [8] However, since then, several new AT1-receptor antagonists have been reported and their inclusion in a model would definitely have a profound effect on any new 3D-QSAR models.

In this paper, two 3D-QSAR methods – CoMFA and CoMSIA – have been used to develop 3D-QSAR models for AT1-receptor antagonists. The latter method, to the best of our knowledge, has never been applied to AT1-receptor antagonists. In the CoMFA methodology, steric and electrostatic properties are calculated by the Lennard-Jones and Coulomb potentials. [9] In the more recent CoMSIA approach (comparative molecular similarity indices analysis), similarity indices in the space surrounding the aligned molecules are computed. [10, 11] In CoMFA, the hyperbolic functional form of the Lennard-Jones potential yields interaction energies with unrealistic values near the van der Waals surface of the molecules. This is avoided in the analysis by truncating the potentials (cut-off) to 30 kcal mol⁻¹. In CoMSIA, instead of interaction energies, similarity indices are calculated, with a probe of radius 1 Å having charge, hydrophobicity, and hydrogen bond properties of +1. The functional form of the interaction is Gaussian with an attenuation factor α (generally 0.3), which overcomes the limitations

P. Datar · P. Desai · E. Coutinho (✉) · K. Iyer
Department of Pharmaceutical Chemistry,
Bombay College of Pharmacy, Kalina, Santacruz (E),
Mumbai 400 098, India
e-mail: evans-im@eth.net
Tel.: +91-22-6126284, Fax: +91-22-6100935

Present address:

P. Desai, Department of Medicinal Chemistry,
School of Pharmacy, University of Mississippi, MI 38677-1848,
USA

of CoMFA. The introduction of hydrophobicity and hydrogen bonding brings a much-needed dimension to CoMFA interactions. Another aspect of CoMSIA is that, unlike CoMFA, the method is insensitive to the absolute orientation of the molecules in the grid.

Materials and methods

Biological data

The molecules composing the training set were selected based on homogeneity and consistency of biological data. Compounds were selected whose binding affinity was expressed as an IC_{50} value which is the concentration of the antagonist that displaces 50% of specifically bound [^{125}I][Sar¹, Ile⁸]AII in a rabbit aorta preparation. In spite of this strict selection criterion, there were difficulties in collating data originating from different laboratories. One such problem was the use of different reference compounds and the IC_{50} values of the reference. For example, in reports where Losartan (DuP-753) was used as a reference, the IC_{50} values reported by the researchers ranged from 19 to 50 nM. To offset these errors, we followed two scales for expressing the binding values. In the first scale, the negative logarithm of the IC_{50} (in M) value (pIC_{50}) was proportionated to the value of Losartan with IC_{50} value as 50 nM. This value was chosen because it was common among the various reported values for Losartan. Activity so modified has been named as proportionate data (Y1). The second was the negative logarithm of the original IC_{50} values (in M) reported by the various researchers without considering the differences observed in the value of the reference compound Losartan. This we refer to as non-proportionate data (Y2).

The two pIC_{50} values (Y1 and Y2) for the 38 molecules (Table 1, consisting of parts a, b, c, d, e, f, g, h, i, j, and k) that constituted the training set are listed in Table 2. It is interesting to note that there is a large diversity of structures belonging to the imidazole, [12] pyrazole, [13] imidazopyridine, [14, 15, 16, 17] triazole, [18, 19, 20, 21] and imidazotriazole [22] series in the training set. Furthermore, the range of biological activity of molecules in the training set spans about five log orders or more. All molecules have been considered in their neutral form.

Molecular modeling and alignment

All molecular modeling was carried out on a Silicon Graphics O2 workstation with Sybyl v 6.7 (Tripos Inc., U.S.A.), Cerius2 v 4.0, and Discover v 98 (MSI, U.S.A.) molecular modeling software.

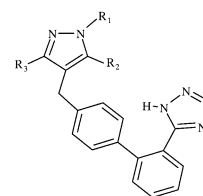
The molecular conformation and alignment of molecules are two sensitive input areas affecting CoMFA models. The “active” conformation has generally been extracted from the X-ray analysis of an inhibitor–

Table 1a Structure of molecules that form the training and test sets for CoMFA and CoMSIA studies

ID	R ₁	R ₂	R ₃
1	Cl	CH ₂ OH	
36	H	COOCH ₃	
37	H	COOH	
38	H	COOH	
40	C ₂ H ₅	COOCH ₃	
41	H	COOCH ₃	
51	Cl	COOCH ₃	
52	H	COOH	
53	H	COOCH ₃	

Table 1b

Pyrazole derivatives:

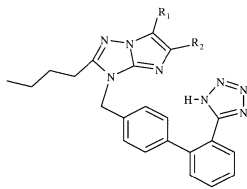


ID	R ₁	R ₂	R ₃
15	H	COOC ₂ H ₅	n-Pr
16		COOC ₂ H ₅	n-Bu
17		COOH	n-Bu
18		COOH	n-Bu
42		COOH	n-Bu
54		COOC ₂ H ₅	n-Bu
55	H	COOC ₂ H ₅	n-Bu

enzyme complex, [23] while alignments have been based on “atom-fit” or “field-fit” methods. [24, 25] We have followed a completely different approach – consensus dynamics (CD) – to address the twin problems of conformation and alignment. [26, 27] In this technique, a dynamics simulation of an ensemble of molecules (in this case, the 38 molecules constituting the training set) is carried out, in which the common elements in the set are

Table 1c

Imidazo[1,2-b][1,2,4]triazole derivatives:





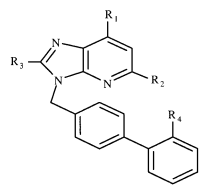



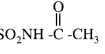
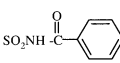

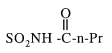
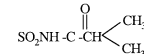
ID	R ₁	R ₂
27	CH ₃	CH ₃
28		H
50	H	CH ₃
62	CH ₃	

Table 1d

Imidazo[4,5-b]pyridine derivatives:



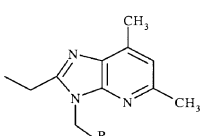
ID	R ₁	R ₂	R ₃	R ₄
5	H	H	n-Pr	
6	H	CH ₃	n-Pr	
7	CH ₃	CH ₃	C ₂ H ₅	
19	CH ₃	CH ₃	C ₂ H ₅	
20	CH ₃	CH ₃	C ₂ H ₅	
47	CH ₃	H	n-Pr	
58	CH ₃	CH ₃	C ₂ H ₅	
59	CH ₃	CH ₃	C ₂ H ₅	

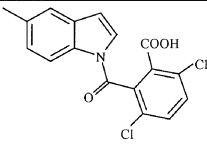
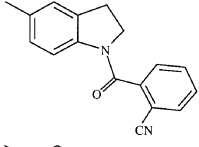
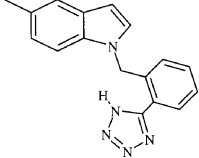
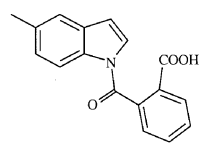
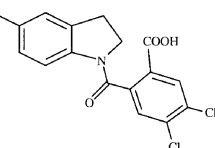
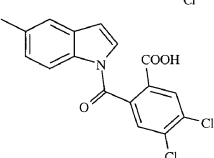
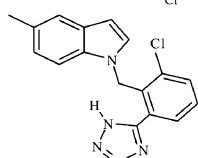
tethered together by a constraining function to the force-field. The intermolecular interactions are switched off. The procedure samples the conformational manifold of the ensemble while constraining corresponding groups within these molecules to occupy similar locations in space. The procedure guarantees that the conformations generated are accessible to all molecules in the ensemble. This makes CD an ideal technique for generating conformations and alignments critical to CoMFA.

Previously derived structure–activity relationships have identified minimum pharmacophoric points, which are critical for AT1 antagonism. [5, 6] These are a het-

Table 1e

Imidazo[4,5-b]pyridine derivatives:

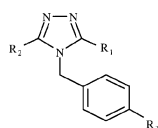


ID	R
8	
9	
10	
48	
49	
60	
61	

erocyclic ring, a hydrogen bond acceptor function, an acidic moiety, and an aromatic spacer that links the acidic group to the heterocyclic ring. These features have been highlighted in Fig. 1 for Losartan as an example. Related elements can be easily identified in all 38 molecules of the training set. The corresponding points in the different molecules were tied to each other by a restraining force with a constant of 100 kcal mol⁻¹ Å⁻². The maximum permitted constraining force was 10 kcal mol⁻¹. After tying up the molecules, the ensemble was energy-minimized beginning with 15,000 steps of steepest descents, which were succeeded by 17,000

Table 1f

1,2,4-Triazole derivatives:



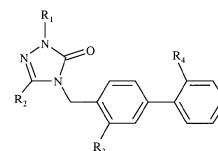
ID	R ₁	R ₂	R ₃
21	CH ₂ OCH ₃	S-C ₂ H ₅	
22		n-Bu	
23		n-Bu	
24	S-C(CH ₃) ₃	n-Bu	
25		n-Bu	
26		n-Bu	
43	CF ₃	SCH ₃	
44	n-Bu		
56		n-Bu	
57		n-Bu	

steps of conjugate gradients. This was followed by a simulated annealing protocol, where the temperature of the system was slowly raised to 900 K, in steps of 100 K, with a 9 ps dynamics run at each new temperature. On reaching 900 K, productive dynamics was carried out for 100 ps. The trajectory of the ensemble was sampled every 1 ps, resulting in 100 ensemble structures. Each of the 100 ensemble structures was cooled to 300 K, lowering the temperature in steps of 100 K by a 1 ps dynamics. Finally, the structures were energy minimized employing 12,000 steps of steepest descents and 15,000 steps of conjugate gradients. At the end of this minimization, all structures had a gradient of 0.001 kcal mol⁻¹ Å⁻¹ or lower.

In the MD simulations, the energy of the system was calculated with the CVFF forcefield. [28] The dielectric constant was 1.0; Newton's equations of motion were integrated with the Verlet algorithm [29] with a step size of 1 fs. Temperature control was achieved either by direct

Table 1g

1,2,4-Triazole-3-one derivatives:



ID	R ₁	R ₂	R ₃	R ₄
2		n-Pr	F	
3		C ₂ H ₅	F	
4		C ₂ H ₅	F	
11		n-Pr	H	
12		n-Bu	H	
13	CH ₂ CH ₃	n-Bu	F	
14		n-Bu	H	
29		n-Bu	H	SO ₂ NH-C[CH ₃] ₃
30		n-Bu	H	SO ₂ NH ₂
31		n-Bu	H	
32		n-Pr	H	
33		n-Bu	H	
34		n-Bu	H	SO ₂ NH-C-[CH ₃] ₃ COOH
35		n-Bu	F	
39		n-Bu	H	
45		n-Bu	H	
46		n-Bu	H	

scaling of atom velocities or by a weak coupling to a temperature bath. [30] Finally, the 100 ensemble structures were clustered into families by hierarchical clustering [31] using rmsd as the descriptor.

Table 1h

Quinazolinone derivatives:

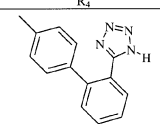
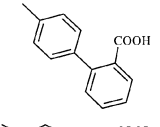

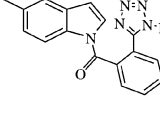

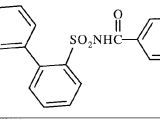
ID	R ₁	R ₂	R ₃	R ₄
63	NH ₂	H	n-Bu	
64	H	CH ₃	n-Bu	
65		H	n-Bu	
66		H	n-Pr	

Table 1i

Benzimidazole derivatives:

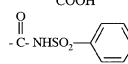


ID	R ₁	R ₂	R ₃	R ₄	R ₅
67	H	CH ₃	CH ₃	n-Bu	COOH
68	H	H	H	n-Bu	
69	H	COOCH ₃	H	n-Bu	
70	H	CH ₂ OH	H	n-Bu	
71	CH ₃	H	H	n-Pr	COOH

Table 1j

Purine derivatives:


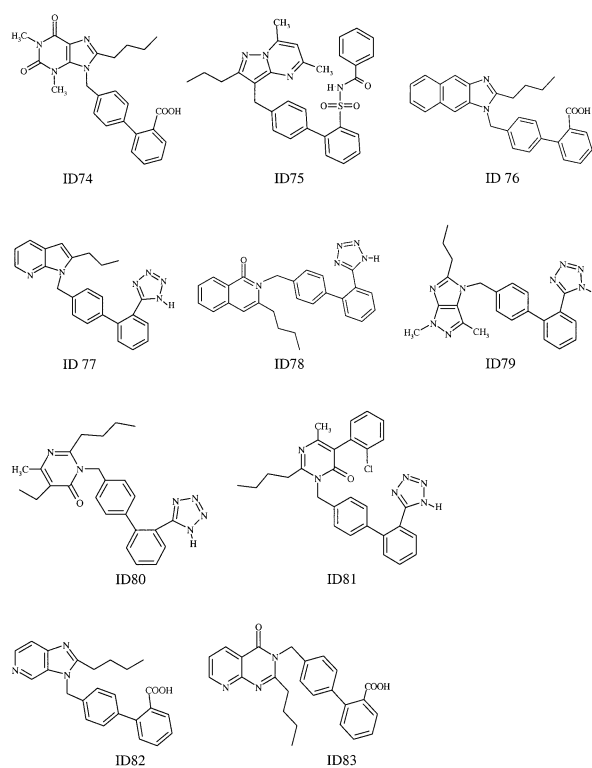
ID	R ₁	R ₂	R ₃
72	n-Pr	n-Bu	
73	n-Bu	H	H

Table 1k

Unique structures:



Alignment by field-fit

Besides the CD generated alignment, a second type of alignment employing field-fit strategy as followed in Cerius2 was also carried out. The conformation of the molecules for field-fit was chosen from the ensemble conformation with the lowest energy which was obtained through consensus dynamics. Losartan was used as the reference or target molecule for the field-fit process. The field-fit method aligns two molecules by first calculating both the steric and electrostatic fields around the molecules. The two molecules are then oriented (in a rigid fashion) to attain a maximum similarity overlap of these fields. A pre-alignment based on moments of inertia is done, to get the two molecules in a decent orientation and position, before performing field similarity calculations. The steric and electrostatic fields were calculated by the universal force field [32] using a neutral carbon atom for the estimation of the steric field component.

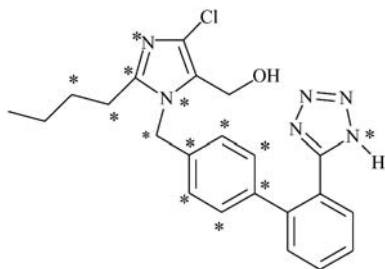
Assignment of partial charges

There are numerous formalisms for calculating the partial charges of atoms, and each method has a decisive effect on the overall CoMFA model. To gauge the effect of

Table 2 Experimental binding affinity of molecules in the training (1–38) and test (39–83) sets

Molecule ID	pIC50 (M) Y1 ^a	pIC50 (M) Y2 ^a	Reference	Molecule ID	pIC50 (M) Y1 ^a	pIC50 (M) Y2 ^a	Reference
1	7.30	7.30	12	43	5.32	5.31	22
2	9.92	10.14	21	44	8.48	8.48	22
3	8.20	8.43	21	45	9.55	9.76	21
4	9.02	9.25	21	46	6.92	6.92	20
5	8.09	8.09	15	47	9.0	9.0	15
6	8.69	8.69	15	48	7.73	7.95	17
7	9.52	9.52	15	49	6.80	7.76	17
8	9.13	9.09	17	50	7.26	7.25	22
9	6.20	6.17	17	51	8.06	8.06	12
10	8.09	8.06	17	52	8.22	8.22	12
11	8.11	8.11	20	53	8.72	8.72	12
12	9.13	9.13	20	54	7.60	7.69	13
13	7.11	7.11	20	55	6.79	6.88	13
14	6.16	6.16	20	56	7.62	7.61	22
15	6.69	6.79	13	57	6.92	6.92	22
16	7.64	7.74	13	58	9.57	9.56	14
17	9.64	9.74	13	59	9.68	9.67	14
18	8.67	8.76	13	60	7.86	7.82	17
19	8.65	8.65	14	61	9.03	9.0	17
20	9.69	9.69	14	62	7.89	7.88	22
21	5.28	5.28	22	63	7.98	8.30	41
22	8.52	8.52	22	64	6.11	6.43	7
23	7.52	7.52	22	65	5.91	6.23	42
24	6.05	6.05	22	66	8.60	8.92	43
25	7.76	7.76	22	67	5.70	6.09	44
26	8.85	8.85	22	68	6.44	6.76	45
27	8.10	8.10	22	69	7.03	7.35	7
28	6.65	6.65	22	70	6.56	6.88	44
29	5.78	5.88	19	71	6.75	7.07	44
30	5.98	6.08	19	72	8.20	8.52	46
31	9.26	9.36	19	73	5.88	6.2	44
32	7.89	7.88	19	74	6.91	7.22	47
33	9.67	9.76	19	75	8.03	9.15	48
34	8.16	8.25	19	76	6.0	6.31	44
35	9.72	9.82	19	77	6.94	7.26	49
36	7.58	7.67	12	78	7.26	7.58	7
37	9.69	9.69	12	79	6.60	6.92	50
38	8.44	8.53	12	80	7.26	7.85	51
39	8.20	8.30	19	81	8.38	8.69	51
40	9.30	9.30	12	82	5.70	6.02	52
41	7.70	7.69	12	83	6.35	6.67	47
42	9.20	9.30	13				

^a Y1 and Y2 denote proportionate and non-proportionate activity data. For details see text

**Fig. 1** The labeled atoms (*) shown in Losartan as an example were tethered to corresponding atoms in the ensemble of 38 molecules of the training set for consensus dynamics simulations

various assignment schemes on the CoMFA results, we have calculated partial charges using the CVFF force-field, [28] Gasteiger–Marsili method, [33] Gasteiger–Hückel method, [34] and AM1 molecular electrostatic

potential (AM1-ESP) fitted partial charge method using the ESP routine in MOPAC 6.0. [35]

Absolute orientation

Besides parameters of conformation, alignment, and partial charges, the absolute orientation of the molecules in the lattice also influences the CoMFA results. [36] The effect of orientation relative to grid position was investigated by systematically rotating the assembly of molecules in the x , y , and z directions in a step size of 30° , using the STATIC ROTATE command in Sybyl v 6.7. The q^2 for each orientation in the grid was determined and was seen to vary from 0.260 to 0.457, suggesting a dependence on absolute orientation. The orientation which gave the highest cross-validated correlation coefficient (q^2) was selected for further analysis.

CoMFA settings

The CoMFA grid was automatically generated with the default 4 Å, beyond the van der Waals radii of the assembly of aligned molecules along each of the three principal co-ordinate axes. Analysis was done for each of the three grid spacings of 0.5 Å, 1.0 Å, and 2.0 Å. The last spacing of 2.0 Å yielded the best CoMFA results. An sp^3 carbon atom with a formal charge of +1 served as the probe. A distance-dependent dielectric ($1/r$) was used for evaluation of the electrostatic energy. Cut-off was set to 30 kcal mol⁻¹ for both steric and electrostatic energies. The standard CoMFA scaling was applied to both steric and electrostatic energies. Columns with a variance lower than 1.0 kcal were “filtered out” from the analysis.

PLS regression and validation

To determine the optimal number of components corresponding to the lowest standard error of prediction (PRESS) and the highest q^2 , SAMPLS [37] (sample-distance partial least squares) with leave-one-out (LOO), no column filtering, and a maximum of ten variables was carried out. This was followed by cross-validation by LOO with the optimal number of components (determined in SAMPLS) and a column filtering of 1.0 to yield the final cross-validated correlation coefficient q^2 (also known as r^2_{cv}). To cross check for any chance correlations in the PLS analysis, [38] the PLS was repeated with complete randomization of the biological data and the q^2 recalculated. This was repeated ten times.

The non-cross-validated models were then assessed by the variance (r^2), standard error of estimate (s), the F value and standard deviation of error of calculation (SDEC).

A bootstrap analysis [39] was carried out in 100 runs for the various models and the mean r^2 is reported as $r^2_{bootstrap}$.

A guiding factor for the choice between two models was a higher q^2 and a lower standard deviation of error of prediction (SDEP).

q^2 -guided region selection (q^2 -GRS)

The q^2 -GRS process has been described in detail elsewhere [36] and in recent reviews. [40] In CoMFA, a significant number of lattice points are excluded from the analysis because CoMFA standard scaling applies equal weight to data from each lattice point in a given field. The q^2 -GRS technique refines a model by increasing the weight for those lattice points, which are most pertinent to the model. Thus, the resultant q^2 exhibits better statistics compared to the conventional q^2 .

q^2 -GRS was executed with the REGION FOCUS command in CoMFA.

The test set

The predictive ability of the CoMFA models was tested against a set of 45 diverse compounds (39–83) in Table 2, belonging to quinazolinones, [41, 42, 43] benzimidazoles, [44, 45] purines, [46, 47] pyrazolopyridines, [48] and other unique compounds (Table 1, consisting of parts a, b, c, d, e, f, g, h, i, j, and k). [15, 49, 50, 51] It is important to note that nearly 21 of these molecules are unique in the sense that they have no representation in the training set. This puts a stringent test on the predictive abilities of our CoMFA models. For the test set, the “predictive” r^2 , s , and F ratio were calculated.

Conformation and alignment of the test set

The molecules in the test set were aligned to a template in the CD trajectory that most closely matches their structure. This was achieved with TEMPLATE FORCE in Discover v 98 (MSI, U.S.A.). Matching points between the template and the test molecule were template forced with a constant of 100 kcal mol⁻¹ Å⁻².

Extrapolation values for the test set

The extrapolation values were obtained with the PREDICT command in the ANALYSIS menu of Sybyl QSAR. Extrapolations were found to be below 0.3 log unit predicted activity for each test set molecule, expressing confidence in the predictive values of the models.

CoMSIA settings

The five CoMSIA similarity index fields – steric, electrostatic, lipophilic, and hydrogen bond donor and acceptor – were calculated at grid points using a common probe atom of 1 Å radius with charge, hydrophobicity, and hydrogen bond donor and acceptor properties of +1. The attenuation factor α was set to 0.3.

QSAR contour maps

Contour maps were generated as scalar products of coefficients and standard deviation (stdev*coeff), associated with each CoMFA column. Favored and disfavored levels were fixed at 80% and 20%, respectively, for both the steric and electrostatic fields.

Results

The 100 ensemble conformations sampled by consensus dynamics clustered into three groups. From each group the lowest energy ensemble conformation was subjected to CoMFA analysis. The three conformations have been labeled as A, B, and C. As a representation of these geometries, the Losartan molecule in each of the three con-

formations is shown in Fig. 2. Conformation C differs from A in the sign of the torsion angle τ_1 , while B differs from A in τ_2 . Conformation A corresponds to the X-ray determined structure for Losartan and labeled “x” by Belvisi et al. [6] The “g-type” conformation of Belvisi et al. identifies with our conformation B. Conformation C is unique to our study. As an example the alignment of all the molecules in conformation A generated with consensus dynamics is shown in Fig. 3.

A trial PLS run on the training set in each of the three conformations A, B, and C is given in Table 3. The best statistics (q^2 , r^2 , F , and s) were seen for conformation A.

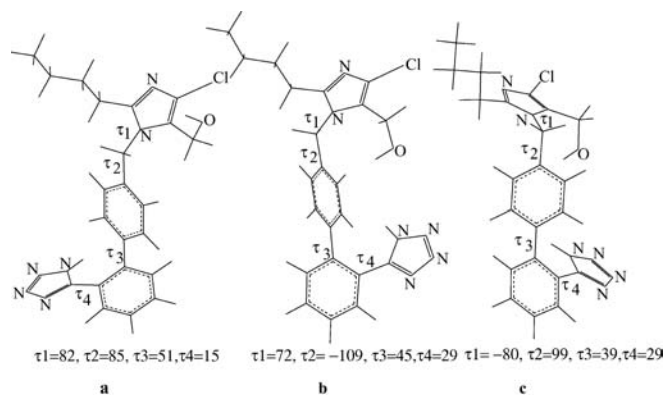


Fig. 2 The three major conformations A, B, and C being represented by Losartan. The torsions τ_1 , τ_2 , τ_3 , and τ_4 uniquely identify these conformations

Fig. 3 Alignment of the 38 molecules (in stereo view) in the training set in conformation A as generated by consensus dynamics

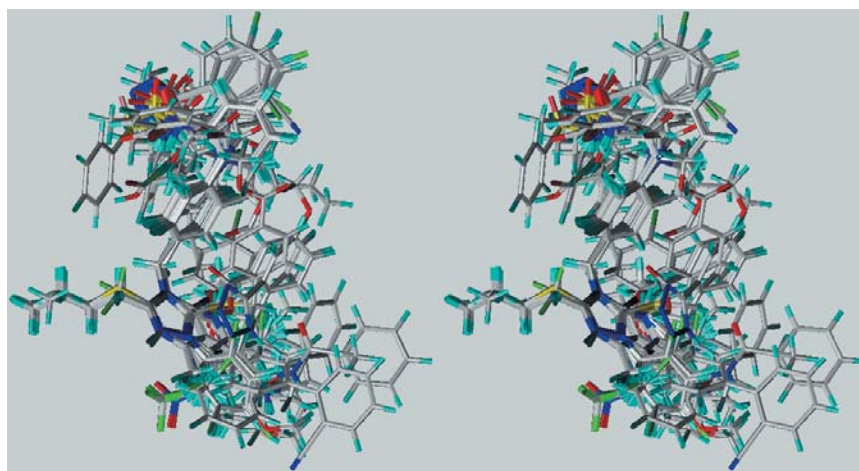


Table 3 Differentiation of conformations by CoMFA

Conformation	Analysis	No. of molecules	q^2	No. of components	r^2	F	s
A	Std. CoMFA	32	0.546	3	0.926	116.0	0.386
	Region Focusing		0.710	5	0.982	291.0	0.195
B	Std. CoMFA	31	0.497	4	0.957	145.9	0.275
	Region Focusing		0.595	4	0.962	163.0	0.261
C	Std. CoMFA	34	0.516	4	0.944	121.0	0.331
	Region Focusing		0.700	6	0.976	185.1	0.223

This result did not vary significantly with the method of representation of biological activity (Y1 or Y2 in Table 2) for the molecules. Conformation A was then selected for further studies.

Having decided the geometry to be used for the generation of the CoMFA models, the effect of atom partial charges on the CoMFA analysis is given in Table 4. In a conventional CoMFA setup, the AM1-ESP charges gave a model with q^2 and r^2 nearly comparable to other charge calculation methods, but fared poorly as regards the stan-

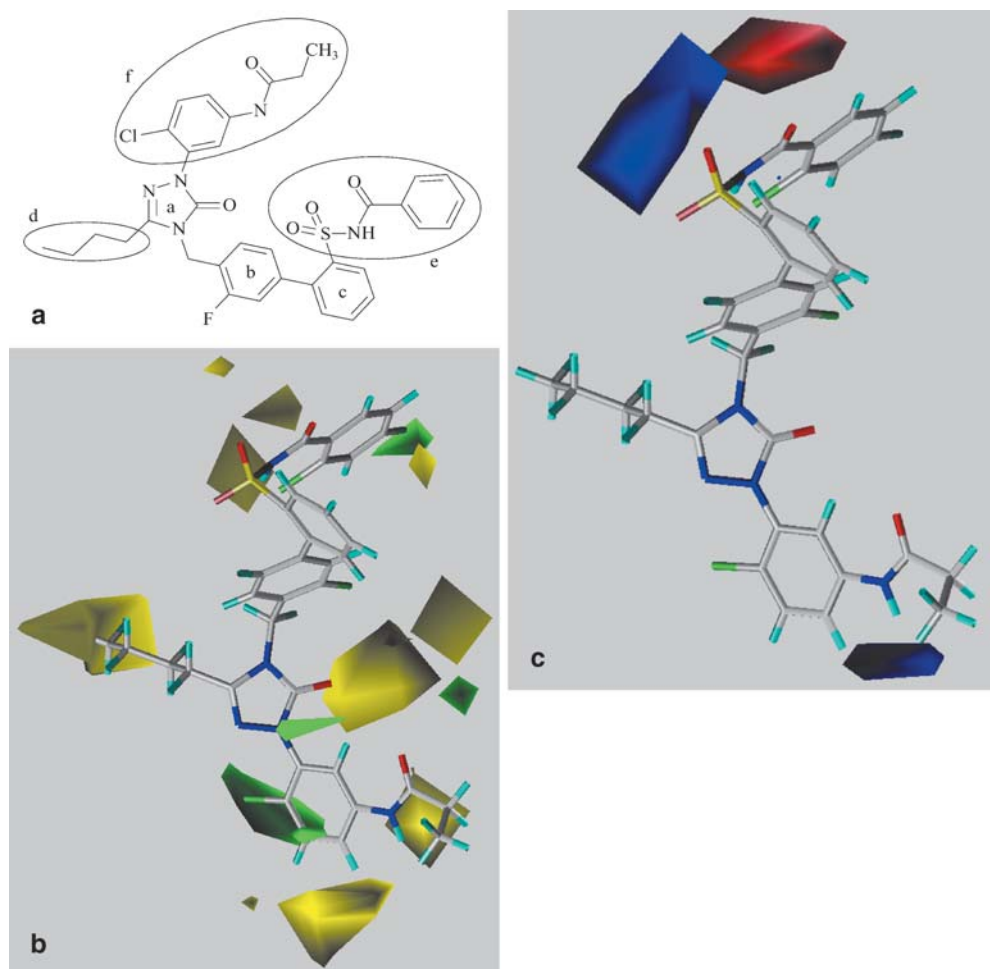
Table 4 Effect of partial charge on q^2 for the training set in conformation A

CoMFA data	Charge calculation method			
	CVFF	Gasteiger–Huckel	Gasteiger–Marseili	AM1-ESP
Conventional				
PCs	6	7	6	3
q^2	0.525	0.512	0.474	0.546
r^2	0.982	0.990	0.981	0.926
s	0.201	0.156	0.209	0.386
F	228.8	325.9	209.8	116.8
Region focusing				
PCs	6	6	6	5
q^2	0.626	0.559	0.586	0.710
r^2	0.980	0.970	0.980	0.982
s	0.213	0.262	0.212	0.195
F	202.5	133.1	194.5	291.6

Table 5 Summary of CoMFA and CoMSIA results for various models

	CoMFA Model 1		CoMFA Model 2		CoMFA Model 3		CoMSIA Model 4 (EHA)		Model 5 (SEA)	
	Std	Region focusing	Std	Region focusing	Std	Region focusing	Std	Region focusing	Std	Region focusing
No. of molecules	32	32	34	34	31	31	32	32	34	34
No. of components	3	5	5	5	4	4	6	6	6	4
q^2	0.546	0.710	0.632	0.680	0.539	0.596	0.524	0.601	0.452	0.468
SDEP	0.907	0.909	0.860	0.791	0.862	0.808	0.905	0.837	0.998	0.986
PRESS	0.956	0.792	0.893	0.834	0.944	0.886	1.028	0.948	1.111	1.056
r^2	0.926	0.982	0.981	0.976	0.959	0.950	0.969	0.941	0.978	0.892
s	0.386	0.195	0.201	0.226	0.282	0.312	0.266	0.364	0.222	0.476
SDEC	0.361	0.175	0.182	0.204	0.258	0.285	0.235	0.322	0.198	0.439
F value	116.7	291.5	295.5	232.3	151.9	123.6	128.6	66.6	200.6	59.9
r^2_{pred}	0.364	0.368	0.358	0.436	0.276	0.224	0.263	0.281	0.358	0.483
Steric	0.424	0.487	0.453	0.495	0.538	0.582	–	–	0.214	0.335
Electrostatic	0.576	0.513	0.547	0.505	0.462	0.418	0.252	0.117	0.415	0.243
Hydrophobic	–	–	–	–	–	–	0.432	0.513	–	–
H-acceptor	–	–	–	–	–	–	0.316	0.370	0.371	0.422

Fig. 4 **a** Demarcation of the steric and electrostatic regions of interest in the CoMFA contour maps for Model 2 associated with molecule 2. **b** Steric fields generated with CoMFA using region focusing around molecule 2. *Yellow* shows unfavorable steric areas, while *green* indicates regions where bulk is predicted to improve affinity. **c** Electrostatics fields generated with CoMFA in region focusing around molecule 2. *Blue* indicates regions where positively charged substituents are predicted to improve affinity, while *red* identifies regions where negatively charged substituents are likely to improve affinity



standard error of estimate (s) and F ratio. However, region focusing (q^2 -GRS) was seen to markedly improve the statistics for the AM1-ESP charges over other methods of charge calculation. The AM1-ESP charges were therefore selected for subsequent studies and development of the final CoMFA model.

With the training set in conformation A, and the partial charges for molecules assigned according to the AM1-ESP formalism, the orientation of the assembly in the grid was then optimized.

Table 5 summarizes the results of the final CoMFA models with the optimized parameters. Model 1 was de-

veloped with proportionated activity data (Y1 in Table 2). In this model, region focusing greatly improved the q^2 and lowered both PRESS, standard error of estimate (s) and standard deviation of error of calculation (SDEC). The conventional CoMFA model and the model resulting from region focusing were comparable in their predictive capabilities on the test compounds, i.e. their “predictive” r^2 were nearly the same.

When binding affinities were expressed in a non-proportionate fashion (Y2 in Table 2), optimization of the CoMFA yielded Model 2. Although the standard q^2 of Model 2 was slightly higher than the corresponding value of Model 1, its q^2 -GRS was a little lower than Model 1. However, it fared better as regards prediction of activity of molecules in the test set, with a “predictive” r^2 of 0.436. (Predicted activities with residuals are given in Table 6.) The standard deviation in error of prediction (SDEP) was also the smallest for Model 2 as compared to other models generated for this series (Table 5) and overall Model 1 and Model 2 are fairly close to each other. They have nearly the same steric and electrostatic contributions.

The CoMFA parameters for Model 3 generated with the same conformations as Model 1 and aligned based on field-fit are also listed in Table 5. This model was inferior in all respects to the earlier Models 1 and 2; suggesting that alignment based on pharmacophoric points is a better picture of the binding conformation of these molecules.

The CoMSIA studies were carried out with the same geometry, alignment, and partial charges used to generate Models 1 and 2. With the activity represented in a proportionate manner (Y1), of the 31 possible combinations of the five property fields, the best CoMSIA model (Model 4) had a q^2 of 0.524, which improved with region focusing to 0.601. It included electrostatic, hydrophobic, and hydrogen bond acceptor fields (Table 5). Further, inclusion of the steric field was seen to reduce the correlation marginally. With Y2 activity, the CoMSIA model (Model 5) is best described by steric, electrostatic, and hydrogen bond acceptor fields. This model has a better “predictive” r^2 than the previous CoMSIA Model 4; however, other statistical parameters leave a lot to be desired. In summation, the inclusion of hydrophobic and hydrogen bonding acceptor fields did not improve the quality of the models over those with the standard CoMFA fields. This suggested that standard CoMFA fields were adequate descriptors of receptor interaction for this group of AT1 antagonists.

Analysis of contour maps

The CoMFA contour maps for Model 2 (model with the best “predictive” r^2) are discussed below. The mapped regions are discussed in relation to molecule 2 (Fig. 4a), which was the most potent molecule in the training set.

Analysis showed that the important steric fields (Fig. 4b) were mainly concentrated around the side chain

Table 6 Predicted activities and residuals of test compounds (39–83) with Y2 activity by CoMFA Model 2

ID	Activity (Y2)	Predicted activity	Residual
39	8.30	8.99	-0.69
40	9.30	8.91	0.39
41	7.69	8.89	-1.20
42	9.30	8.82	0.48
43	5.31	6.74	-1.43
44	8.48	7.34	1.14
45	9.76	9.73	0.03
46	6.92	7.24	-0.32
47	9.00	8.40	0.60
48	7.95	7.39	0.56
49	7.76	6.93	0.83
50	7.25	7.68	-0.43
51	8.05	7.95	0.10
52	8.22	8.82	-0.60
53	8.72	7.29	1.43
54	7.69	8.56	-0.87
55	6.88	6.64	0.24
56	7.61	7.14	0.47
57	6.92	8.29	-1.37
58	9.56	8.10	1.46
59	9.67	8.20	1.47
60	7.82	7.42	0.40
61	9.00	7.94	1.06
62	7.88	8.13	-0.25
63	8.30	8.05	0.25
64	6.43	6.20	0.23
65	6.23	7.01	-0.78
66	8.92	9.05	-0.13
67	6.09	6.96	-0.87
68	6.76	8.10	-1.34
69	7.35	7.34	0.01
70	6.88	7.55	-0.67
71	7.07	7.03	0.04
72	8.52	8.07	0.45
73	6.20	7.68	-1.48
74	7.22	6.11	1.11
75	9.15	10.0	-0.85
76	6.31	6.89	-0.58
77	7.26	8.26	-1.00
78	7.58	7.68	-0.10
79	6.92	8.08	-1.16
80	7.85	7.75	0.10
81	8.69	7.97	0.72
82	6.02	6.75	-0.73
83	6.67	6.31	0.36

(f) attached to the heterocyclic ring (a). The sterically favored region was juxtaposed with a sterically forbidden region, indicating a strict conformational requirement of the group (f). This is in accordance with structure–activity relationships [52] of e.g. 2-substituted triazolones, which indicates an optimal length and orientation of the side chain at this position. Thus, the orientation and bulkiness of the group (f) plays an important role in the activity. The lower activity of molecules 14, 28, 65, and 74 could be attributed to the fact that phenyl or methyl groups (the “f” related groups) extend into disfavored steric regions. Correspondingly, the higher activity of molecules 2, 4, 12, 17, 31, 33, 42, 47, 63, and 66 was commensurate with the fact that the aromatic or the methyl groups (again the “f” related groups) lay in sterically favored regions. Another sterically disfavored area

denoted as “g” in Fig. 4a, (more apparent in those molecules that have a large group in this position) was also observed. This accounts for the moderate activity of molecules 50, 54, 55, 56, 57, and 62, which had subgroups extending into this disfavored region. This correlates well with SAR studies [22]. There was a small disfavored steric region around the C5 and C6 atoms of the phenyl ring (c) and also at the terminal carbon atom of the butyl chain (d). Consequently, propyl or ethyl groups (“d” related groups) would seem appropriate for activity in these molecules, falling in line with SAR studies of imidazole-5-carboxylic acids. [53] The high activity of molecules 40 and 51 results from the fact that the phenyl rings of the benzoylsulfonamide moiety (“e” related groups) fell in a sterically favored site. This nicely relates to SAR studies seen for molecules in which the tetrazole ring has been replaced by the acylsulfonamide group. [12] When the acylsulfonamide group is changed to acylsulfamide (e.g. molecules 41 and 52), SAR studies show a fall in the activity; [12] this is because the phenyl group (“e” related group) of such molecules extends into an unfavorable steric region. However, when the relatively “rigid” phenyl group is changed to a more “flexible” heptyl chain (molecule 53), the activity improves slightly on account of the fact that a portion of the aliphatic chain reached into a favorable steric site.

An electrostatically favorable region (Fig. 4c) surrounds the biphenyl moiety, suggesting that affinity might be improved by placing an electronegative function on this moiety. This was apparent in molecules 2, 3, and 4, which have a fluorine atom on the biphenyl ring and have high activities. For the “e” related groups a blue contour region was seen juxtaposed to a red contour region. This suggested critical conformational placements of the electropositive and electronegative groups. These suggestions from the model comply with SAR studies [54] where e.g. placement of a 5-methyl group on the tetrazole moiety was seen to decrease the activity. Another electrostatically favorable region was associated with the side chain (f) of the heterocyclic ring (a). An electropositive region was also seen near the tail portion (e) of those molecules which had an $\text{SO}_2\text{-NH}$ group attached to the biphenyl system. This implies that the activity of such molecules may be improved by introduction of an electropositive group on the phenyl ring of the $\text{SO}_2\text{-NH-X-Ph}$ moiety as present in molecules 20, 35, 40, 51, and such others.

Finally, the poor activity of molecule 24 could be partly attributed to the location of the *t*-butyl group in both sterically and electrostatically disfavored regions.

Discussion

Two 3D-QSAR methods – CoMFA and CoMSIA – were applied on a training set of 38 AT1 antagonists to understand structure–activity relationships. Significant CoMFA models were obtained which discriminated between various geometries, alignment schemes, orientations, and

partial charge formalisms. The conformation and alignment generated by consensus dynamics yielded a CoMFA model superior to that derived using field-fit. Likewise, partial charges calculated with the AM1-ESP formalism gave a model possessing a better q^2 value than other methods of charge calculation. Further analysis indicated that the CoMFA models possessed better predictive power and greater robustness compared to the CoMSIA models. This suggested that the steric and electrostatic fields of these molecules are adequate descriptors for structure–activity relationships in this class of compounds. The poor activity of some of these molecules could be rationalized based on the CoMFA-generated contour maps; these maps also portray how activity could be improved by varying the bulk and electrostatics for these molecules.

The earlier CoMFA studies by Belvisi et al. [6] and Prendergast et al. [7] were carried out on training sets of 28 and 50 molecules, respectively. In this aspect, the number of molecules in this work and the diversity of the test set straddle the two limits. One of the reasons for our inability to include more compounds and increase diversity was limited by the strict adherence to our selection criteria. We made all efforts to include only data obtained with respect to the same tissue and the same reference/standards. The conformations A and B generated by consensus dynamics are related to the “x” and “g” geometries of Belvisi et al. In their study, the geometry was not seen to influence the outcome of the CoMFA results. This contrasts with our study, where the geometry does seem to have an influence on the CoMFA models, but not the manner of representation of activity.

Belvisi et al. [6] and Prendergast et al. [7] found MNDO-ESP calculated partial charges and MNDO derived charges to be best suited for their CoMFA models, respectively. This work also showed that partial charges fitted to electrostatic potentials calculated at the AM1 level of theory are the most ideal. Further, Belvisi et al. [6] had carried out a variable selection optimization using D-optimal design and fractional factorial design in the GOLPE program to yield models with a high q^2 of 0.80. This was higher than our best q^2 of 0.71. However, the reported q^2 and r^2 values in the Prendergast et al. work [7] were only 0.64 and 0.76, respectively, possibly because the model was not optimized by removal of outliers. More importantly, Belvisi et al.’s CoMFA model was not rigorously tested on an independent test set, unlike Prendergast et al. who used the model to make predictions of molecules outside the training set. Although Prendergast et al. have not reported a “predictive” r^2 value, our calculations using data reported in their paper yield a “predictive” r^2 of 0.05. It is possible that the diversity of their test set and the lack of optimization of the CoMFA model may be probable factors for this low value. The “predictive” r^2 of our CoMFA model is 0.436. Overall our model for AT1 receptor antagonist is a refinement of the existing CoMFA models and exhibits better predictive capabilities and helps to better understand structure–activity relationships of this class of compounds.

Acknowledgements This work was made possible by a grant from the All India Council of Technical Education (AICTE), New Delhi to E Coutinho (F. No. 8020/RID/R&D-60/2001-02). P.A. Datar thanks the Amrut Mody Research Foundation for support.

References

- Vallotton MB (1987) *Trends Pharmacol Sci* 8:69–74
- McAreevey DW, Robertson JIS (1990) *Drugs* 40:326–345
- Greenlee WJ (1990) *Med Res Rev* 10:173–236
- Duncia JV, Carini DJ, Chiu AT, Johnson AL, Price WA, Timmermans PBMWM, Wong PC (1992) *Med Res Rev* 12:149–191
- Belvisi L, Bravi G, Scolastico C, Vulpetti A, Salimbeni A, Todeschini R (1994) *J Comput-Aided Mol Des* 8:211–220
- Belvisi L, Bravi G, Catalano G, Mabilia M, Salimbeni A, Scolastico C (1996) *J Comput-Aided Mol Des* 10:567–582
- Prendergast K, Adams K, Greenlee WJ, Nachbar RB, Patchett AA, Underwood DJ (1994) *J Comput-Aided Mol Des* 8:491–512
- Perkins TDJ, Dean PM (1993) *J Comput-Aided Mol Des* 7:155–172
- Cramer RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110:5959–5967
- Bohm M, Sturzebecher J, Klebe G (1999) *J Med Chem* 42:458–477
- Klebe G, Abraham U, Mietzner T (1994) *J Med Chem* 37:4130–4146
- Naylor EM, Chakravarty PK, Costello AC, Chang RS, Chen T, Faust KA, Lotti VJ, Kivlighn SD, Zingaro G J, Siegl PKS, Pancras CW, Carini DJ, Wexler AA Greenlee WJ (1994) *Bioorg Med Chem Lett* 4:69–74
- Wallace TA, Steven MH, Greenlee WJ, Doss GA, Chang RSL, Lotti VJ, Faust KA, Chen T, Zingaro GJ, Kivlighn SD, Siegl PKS (1993) *J Med Chem* 36:3595–3605
- Chakravarty PK, Naylor EM, Chang RSL, Chen A, Chen T, Faust KA, Lotti VJ, Kivlighn SD, Zingaro GJ, Schorn TW, Schaffer LW, Siegl PKS, Patchett AA Greenlee WJ (1994) *J Med Chem* 37:4068–4072
- Mantlo NB, Chakravarty PK, Ondeyka DL, Siegl PKS, Chang RSL, Lotti VJ, Faust KA, Chen T, Schorn TW, Sweet CS, Emmert SE, Patchett AA, Greenlee WJ (1991) *J Med Chem* 34:2919–2922
- Dhanoa DS, Bagley SW, Chang RSL, Lotti VJ, Chen T, Siegl PKS, Chakravarty PK, Patchett AA, Greenlee WJ (1993) *J Med Chem* 36:3738–3742
- Dhanoa DS, Bagley SW, Chang RSL, Lotti VJ, Chen T, Siegl PKS, Zingaro GJ, Patchett AA, Greenlee WJ (1993) *J Med Chem* 36:4230–4238
- Linda LC, Wallace TA, Flanagan KL, Chen T, Zingaro GJ, Siegl PKS, Kivlighn SD, Lotti VJ, Chang RSL, Greenlee WJ (1994) *J Med Chem* 37:4464–4478
- Linda LC, Flanagan KL, Wallace TA, Naylor EM, Chakravarty PK, Patchett AA, Greenlee WJ, Chen T, Faust KA, Chang RSL, Lotti VJ, Zingaro GJ, Siegl PKS, Kivlighn SD (1994) *J Med Chem* 37:2808–2824
- Linda LC, Wallace TA, Flanagan KL, Greenlee WJ, Chang RSL, Lotti VJ, Faust KA, Chen T, Zingaro GJ, Kivlighn SD, Siegl PKS, Bunting P, MacCoss M, Strelitz RA (1993) *J Med Chem* 36:2558–2568
- Linda LC, Wallace TA, Flanagan KL, Chen T, Zingaro GJ, Siegl PKS, Lotti VJ, Chang RSL, Greenlee WJ, O'Malley SS (1995) *J Med Chem* 38:3741–3758
- Linda LC, Wallace TA, Cantone CL, MacCoss SS, Chang RSL, Lotti VJ, Faust KA, Chen T, Strelitz RA, Bunting P, Kivlighn SD, Siegl PKS (1993) *J Med Chem* 36:591–609
- Matter H, Schwab W, Barbier D, Billen G, Hasse B, Neises B, Schudok M, Thorwort W, Schreuder H, Lonze P (1999) *J Med Chem* 42:1908–1920
- Marshall GR, Barry CD, Bosshard HE, Dammkoehler RA, Dunn DA (1979) The conformational parameter in drug design: active analog approach. In Olson EC, Christoffersen RE (eds) *Computer-assisted drug design*. American Chemical Society, Washington, D.C. pp 205–226
- Cramer RD, Clark M, Simeroth P, Patterson DE (1991) Recent developments in comparative molecular field analysis (CoMFA). In Silipo C, Vittoria A (eds) *QSAR: rational approaches to the design of bioactive compounds*. Elsevier, Amsterdam, pp 239–242
- Kamath S, Coutinho E (1997) *J Biosci* 22:315–324
- McDowell RS, Gadek TR, Barker PL, Burdick DL, Chan KS, Quan CL (1994) *J Am Chem Soc* 116:5069–5076
- Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT (1988) *Proteins: Struct Funct Gen* 4:31–47
- Verlet L (1967) *Phys Rev* 159:98–103
- Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) *J Chem Phys* 81:3684–3690
- Murtaugh F (1983) *Comput J* 26:354–359
- Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM, (1992) *J Am Chem Soc* 114:10024–10035
- Gasteiger J, Marsili M (1980) *Tetrahedron* 36:3219–3228
- Tripos (2001) SYBYL 6.7 Force field manual. Tripos, St Louis, Mo. 290 pp
- Stewart JJP (1990) *J Comput-Aided Mol Des* 4:1–103
- Cho SJ, Tropsha A (1995) *J Med Chem* 38:1060–1066
- Bush BL, Nachbar RB (1993) *J Comput-Aided Mol Des* 7:587–619
- Wold S, Johansson E, Cocchi M (1993) PLS – partial least-squares projections to latent structures. In Kubinyi H (ed) *3D-QSAR in drug design: theory methods and applications*. ESCOM, Leiden, The Netherlands, pp 523–550
- Leger C, Politis DN, Romano JP (1992) *Technometrics* 34:378–399
- Tropsha A, Cho SJ (1998) Cross-validated r^2 guided region selection for CoMFA studies. In Kubinyi H, Folkers G, Martin YC (eds) *3D QSAR methodology CoMFA and related approaches*, vol III. Kluwer, Dordrecht, The Netherlands, pp 57–69
- Allen EE, deLaszlo SE, Huang SX, Quagliato CS, Greenlee WJ, Chen RT, Faust KA, Lotti VJ (1993) *Bioorg Med Chem Lett* 3:1293–1296
- Bagley S, Greenlee WJ, Dhanoa DS, Patchett AA (1992) *US Patent* 5,175,164
- Glinka TW, deLaszlo SE, Siegl PKS, Chang RSL, Kivlighn SD, Schorn TS, Faust KA, Chen TB, Zingaro GJ, Lotti VJ, Greenlee WJ (1991) *Bioorg Med Chem Lett* 4:81–84
- Chakravarty PK, Camara VJ, Chen A, Marcin LM, Greenlee WJ, Patchett AA, Chang RSL, Lotti VJ, Siegl, PKS (1990) 200th National Meeting of the American Chemical Society. American Chemical Society, Washington, D.C. pp 26–31
- Chakravarty PK, Patchett AA, Camara VJ, Walsh TF, Greenlee WJ (1990) *European Patent Application* 400,835
- Ondeyka DL, Mantlo NB, Chakravarty PK, Chen A, Camara VJ, Chang RSL, Lotti VJ, Siegl PKS, Patchett AA, Greenlee WJ (1991) Joint Central–Great Lakes Regional Meeting of the American Chemical Society, Indianapolis, 31 May. American Chemical Society, Washington D.C.
- Chakravarty PK, Greenlee WJ, Mantlo NB, Patchett AA, Walsh TF (1990) *European Patent Application* 400,974
- Allen EE, Huang SX, Chang RSL, Lotti VJ, Siegl PKS, Patchett AA, Greenlee WJ (1991) 202nd National Meeting of the American Chemical Society, New York, August 25–30
- Patchett AA, Mantlo NB, Greenlee WJ (1992) *US Patent* 5,124,335
- Greenlee WJ, Johnston DBR, MacCoss M, Mantlo NB, Patchett AA, Chakravarty PK, Walsh TF (1992) *US Patent* 5,164,407
- Allen EE, Greenlee WJ, Chakravarty PK, Patchett AA, Walsh TF (1992) *US Patent* 5,100,897
- Huang, HC, Reitz PB, Chamberlain TS, Olins GM, Corpus VM, McMohan EG, Palomo MA, Koepae JP, McGraw DE (1993) *J Med Chem* 36:2174–2181
- Yanagisawa H, Amemiya Y, Kanazakai T, Shimoji Y, Fujimoto K, Kitahara Y, Sada T, Mizuno M, Ikeda M, Koike H (1996) *J Med Chem* 39:323–338
- Kubo K, Inada Y, Kohara Y, Sugiura Y, Ojima M, Itoh K, Furukawa Y, Nishikawa K, Naka T (1993) *J Med Chem* 36:1772–1784