A 3D-QSAR of Angiotensin II (AT1) Receptor Antagonists Based on Receptor Surface Analysis

Prasanna A. Datar, Prashant V. Desai,[†] and Evans C. Coutinho*

Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (E),
Mumbai 400 098, India

A hypothetical receptor surface model has been constructed for a set of 38 AT1 antagonists using activity data of each molecule as a weight in the building of the receptor surface. The best model was derived by optimizing various parameters such as atomic partial charges, surface fit, and the manner of representation of electrostatics on the surface. Descriptors such as van der Waals energy, electrostatic energy, and total nonbonded energy were used individually or in combination to derive a family of quantitative structure−activity relationship equations using G/PLS as the statistical method.

INTRODUCTION

The peptide hormone Angiotensin II (AII) is known to play an important role in the regulation of blood pressure and salt-water homeostasis.[1] Two distinct subtypes of the AII receptor labeled as AT1 and AT2 are known. Numerous studies in the past few years have shown that the major biochemical and functional responses of AII are mediated by activation of the AT1 receptor. The various antihypertensive regimens such as angiotensin converting enzyme inhibitors and renin inhibitors have several side effects and limitations. The AII antagonists on the other hand do not have these disadvantages and have become a globally accepted therapy for hypertension.

The AT1 receptor has been identified as a G-protein coupled receptor, and a three-dimensional model of a complex of AII with the transmembrane region has been solved by homology modeling with rhodopsin.[2] Recently, with the aid of molecular modeling tools such as docking and homology modeling, several AT1 receptor residues involved in the binding of ligands and activation of the receptor by agonists have been identified. A plausible mechanism for AT1 receptor activation has also been proposed.[3] Although some information regarding the receptor binding site is known by site-directed mutagenesis studies, there is a lot more yet to be thoroughly understood. Therefore indirect approaches such as quantitative structure−acitvity relationships (QSARs) have been used to understand the ligand−receptor interactions.

In the past 2 decades various pharmacophore and 3D-QSAR methodologies have been used to study AT1 antagonists;[4−7] however, these need to be updated with the current knowledge and newer reported analogues. The main goal of this present work was to develop a receptor site model for the AT1 receptor, which could lead us to design more potent analogues. Receptor surface analysis (RSA) uses

explicit surfaces from compounds of varying activity to characterize the shape of the active site, providing clues on how ligands may be modified to improve activity. We report here a 3D-QSAR study using receptor surface analysis as implemented in the software Cerius2 (Accelrys Inc., San Diego, CA). We used the RSA methodology to develop a receptor surface model (RSM) with hypothetical binding site features for AT1 antagonists.

Receptor Surface Analysis (RSA).[8] RSA is a useful tool in situations when the 3D structure of the receptor is unknown, since one can build a hypothetical model of the receptor site. RSM differs from pharmacophore models in that the former tries to capture essential information about the receptor, while the latter captures information about the commonality of compounds that bind to a receptor. A receptor surface model embodies essential information about the hypothetical receptor site as a three-dimensional surface with associated properties such as hydrophobicity, partial charge, electrostatic (ELE) potential, van der Waals (VDW) potential, and hydrogen bonding propensity. The surface points that organize as triangle meshes in the construction of the RSM store these properties as associated scalar values. Receptor surface models provide compact, quantitative descriptors, which capture three-dimensional information of interaction energies in terms of steric and electrostatic fields at each surface point, which in other techniques such as CoMFA are calculated using probe interactions[9] at various grid points. These descriptors can be used for 3D QSAR studies. Variable selection techniques such as genetic function approximation[10] (GFA) can then be used to develop a family of equations. GFA was initially conceived by drawing inspiration from two algorithms, Holland's genetic algorithm[11] and Friedman's[12] multivariate adaptive regression splines (MARS) algorithm. The GFA algorithm develops an initial population of individuals, and a fitness function (lack of fit, LOF)[12] is used to estimate the quality of each individual. Individuals with the best fitness scores are allowed to mate and propagate their genetic material to offsprings through the crossover operation. After repeatedly performing these steps, the average fitness of the individuals in the

* Corresponding author phone: +91-22-26126284; fax: +91-22-26100935; e-mail: evans-im@eth.net.
† Present address: Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, University, MS 38677-1848.

3D-QSAR OF AT1 RECEPTOR ANTAGONISTS

J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004 **211**

population increases, as good combination of "genes" are discovered and spread through the population. This can be observed as the frequency at which each term (descriptor) appears in all equations in the final population. The entire population of equations can then be searched for information on features, patterns, and regions in which different equations predict well.

The most preferred equations in QSARs are those that contain only linear terms. but if there is a possibility that some subrange of descriptors may have a nonlinear relationship, then GFA can uncover such nonlinear relations using spline ($\langle x - a \rangle$) and quadratic (x^2) terms. We have used G/PLS in which both GFA and partial least squares (PLS) algorithms are applied to extract the best relationships that exist for the given data set. GFA selects the appropriate basis functions by using combination of linear (x) and nonlinear such as spline ($\langle x - a \rangle$), quadratic (x^2), offset quadratic (($x - a$)^2), and quadratic spline ($\langle x - a \rangle^2$) terms, while PLS regression as a fitness function weighs the basis function's relative contributions in the final model.

MATERIALS AND METHODS

Training Set. The training set comprised 38 molecules (molecule ID's 1−38 in Table 1) selected from several different structural classes such as imidazole,[13] pyrazole,[14] imidazopyridine,[15] triazole,[16] and imidazotriazole[17] as shown in Table 1. All molecules in the training set were with activity measured on the same tissue, i.e., rabbit aorta. This strict selection criterion entailed the loss of many diverse analogues that were reported active on other tissue models. 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][Sar^1,Ile^8]AII in a rabbit aorta preparation. In the majority of the research papers, the IC_{50} value for the reference molecule, losartan, is not the same but covers a large range, with a majority value of 50 nM. Therefore activity values, i.e., the parameter Y for the QSAR, for the different compounds, were converted with respect to 50 nM for losartan. The range of activity of molecules included in both the training and test sets (vide infra) spanned 5 log orders.

Test Set. The test set included 35 diverse compounds (molecule ID's 39−73 in Table 1), belonging to quinazolinone,[18] benzimidazole,[19] purine,[20] and pyrazolopyridine[21] classes and other unique compounds.[15b,22−24] It is important to note that nearly 15 of these molecules were unique; i.e., they had no representation in the training set. Thus the predictive ability of our RSA models was subjected to a stringent test.

All molecular modeling work was carried out on a Silicon Graphics O2 workstation with the QSAR module of Cerius2 v 4.6[25] and Discover v 98[26] (Accelrys) molecular modeling software.

Conformation and Alignment. The conformation and alignment of molecules was generated with consensus dynamics. The detailed methodology and theory of this technique may be found in our earlier paper.[7] In brief, a dynamics simulation of an *ensemble* of molecules (the 38 molecules constituting the training set) was carried out, in which the common elements (pharmacophoric points)[4,5] in the set were tethered together by a constraining function to

the force field. The intermolecular interactions were switched off. Thus a set of conformations aligned on the basis of the pharmacophoric features[4,5] were obtained. The lowest energy conformation among the ensemble of 100 conformations was selected for further studies.

Generation of Receptor Surface Model (RSM).[8] In RSA, a volumetric field, characterizing molecular shape is constructed for each aligned structure, which is termed as a shape field. The shape field is calculated around a set of field sources such as a set of atoms, at different points in space (typically on a three-dimensional grid). These field sources give rise to a local field defined by some distance-dependent function. The shape fields are then combined to give the isosurface of the field to create an explicit object with well-defined shape.

We have generated the shape field using a van der Waals field function. The van der Waals field function allows a surface to be created inside or outside the van der Waals surface of a set of atoms at a specified distance from the van der Waals surface. The function "$V(r) = r - VDWr$" is used to create a shape field at each surface point around the molecules, at which the atomic coordinates of the contributing atoms of the aligned molecules are used to compute field values at each point on a 3D grid. In this function, r is the distance from the point to the atom and VDWr is the van der Waals radius of the atom. This function allows a surface to be constructed at any arbitrary isovalue of density on the grid. Thus at the van der Waals surface $V(r)$ is zero, inside the surface $V(r)$ is negative, and outside the volume $V(r)$ is positive. The isovalue chosen is called the surface fit and the Marching Cubes algorithm[27] is used to create a triangulated surface fit with an average surface density of about six points per square angstrom. A cutoff distance is employed, as the function is unbounded.

Mapping Properties on the Surface. (a) Partial Charge. In RSA each surface point has an associated charge. The assumption made is that the charge at any point on the surface is complementary to the partial atomic charge of any atom in contact with the surface. In the case of a set of aligned molecules each surface point is given a charge, which is equal but opposite in sign to the average partial charge of the closest atoms in each molecule. We have critically evaluated the effect of atomic partial charges assigned using Gasteiger Huckel,[28] Gasteiger Marsili,[29] CVFF,[30] ESP,[31] and charge equilibration method[32] formalisms on the RSA analysis.

In addition to the charge, surface points also store values of the electrostatic potential. The electrostatic potential at the surface to be constructed over a bundle of molecules is the complementary potential calculated as the average of the potentials exerted by each molecule in the bundle. This is calculated by summing Q_aQ_p/r for all atoms (where r is the distance between point and atom, Q_a is the charge on the atom, and Q_p is the charge on the surface point).

A hydrogen bond property is also stored with each surface point and corresponds to the tendency of the point to be involved in a hydrogen bond. For hydrogen bonding a value of $+1.0$ is assigned to volume vertex points where a hydrogen bond acceptor would be desirable in the receptor. A value of -1.0 is assigned to volume vertex points where a hydrogen bond donor hydrogen might be desirable. These

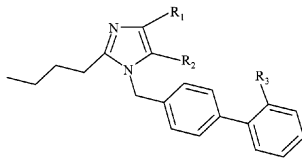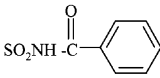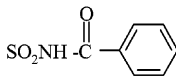**Table 1.** Structure of Molecules that Form the Training and Test Sets for the RSA Study
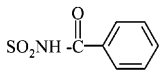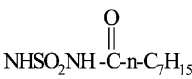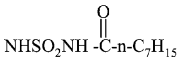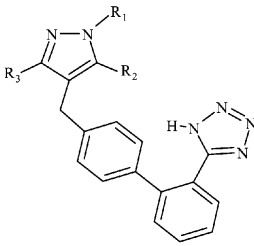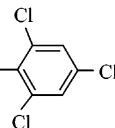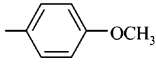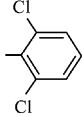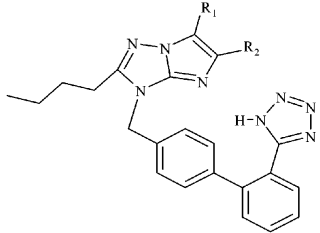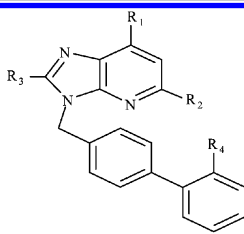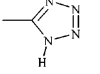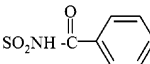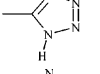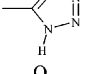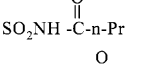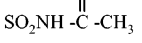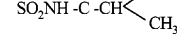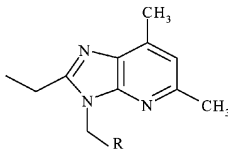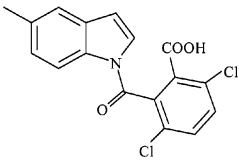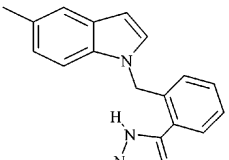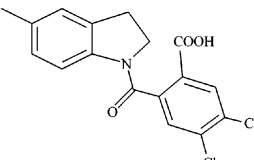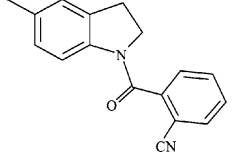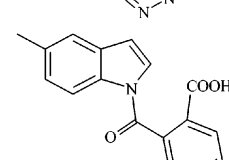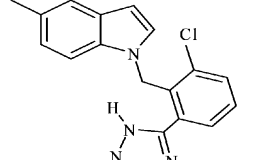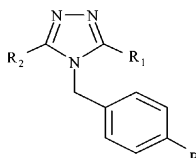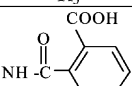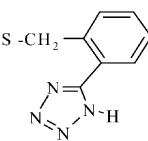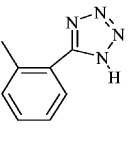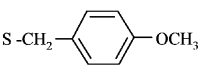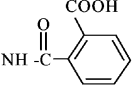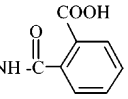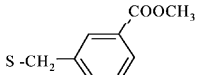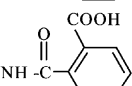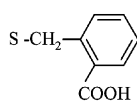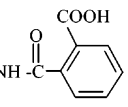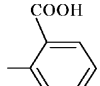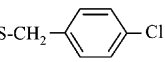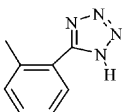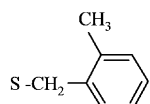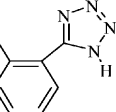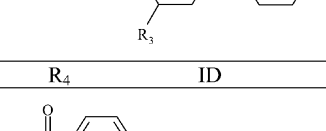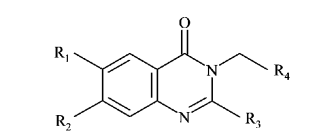
Imidazole derivatives:

ID	Activity	R_1	R_2	R_3	ID	Activity	R_1	R_2	R_3
1		Cl	CH_2OH		40		C_2H_5	$COOCH_3$	
36		H	$COOCH_3$		50		Cl	$COOCH_3$	
37		H	COOH						
38		H	COOH		51		H	$COOCH_3$	

Pyrazole derivatives:

ID	R_1	R_2	R_3	ID	R_1	R_2	R_3
15	H	$COOC_2H_5$	n-Pr	41		COOH	n-Bu
16		$COOC_2H_5$	n-Bu				
17		COOH	n-Bu	52		$COOC_2H_5$	n-Bu
18		COOH	n-Bu	53	H	$COOC_2H_5$	n-Bu

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

ID	R_1	R_2	ID	R_1	R_2
27	CH_3	CH_3	49	H	CH_3
28		H	58	CH_3	

3D-QSAR OF AT1 RECEPTOR ANTAGONISTS

J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004 **213**

Table 1. (Continued)

Imidazo[4,5-b]pyridine derivatives:

ID	R_1	R_2	R_3	R_4	ID	R_1	R_2	R_3	R_4
5	H	H	n-Pr	(methyl-tetrazole)	20	CH_3	CH_3	C_2H_5	$SO_2NH-C(=O)-$phenyl
6	H	CH_3	n-Pr	(methyl-tetrazole)	46	CH_3	H	n-Pr	(methyl-tetrazole)
7	CH_3	CH_3	C_2H_5	(methyl-tetrazole)	55	CH_3	CH_3	C_2H_5	$SO_2NH-C(=O)-$n-Pr
19	CH_3	CH_3	C_2H_5	$SO_2NH-C(=O)-CH_3$	56	CH_3	CH_3	C_2H_5	$SO_2NH-C(=O)-CH(CH_3)_2$

Imidazo[4,5-b]pyridine derivatives:

ID	R	ID	R	ID	R
8	(5-methylindoline, COOH, 2,6-diCl-benzoyl)	10	(5-methylindole, tetrazole-biphenyl)	48	(5-methylindoline, COOH, 3,4-diCl-benzoyl)
9	(5-methylindoline, 2-CN-benzoyl)	47	(5-methylindole, COOH-benzoyl)	57	(5-methylindole, Cl, tetrazole-phenyl)

1,2,4-Triazole derivatives:

ID	R_1	R_2	R_3	ID	R_1	R_2	R_3
21	CH_2OCH_3	$S-C_2H_5$	(NH-C(=O)-, COOH benzene)	26	$S-CH_2$(2-tetrazolyl)phenyl	n-Bu	(methyl, tetrazole-phenyl)
22	$S-CH_2$(4-OCH_3)phenyl	n-Bu	(NH-C(=O)-, COOH benzene)	42	CF_3	SCH_3	(NH-C-, COOH benzene)
23	$S-CH_2$(3-$COOCH_3$)phenyl	n-Bu	(NH-C(=O)-, COOH benzene)	43	n-Bu	$S-CH_2$(2-COOH)phenyl	(NH-C-, COOH benzene)
24	$S-C(CH_3)_3$	n-Bu	(COOH benzene)				
25	$S-CH_2$(2-CH_3)phenyl	n-Bu	(methyl, tetrazole-phenyl)	54	$S-CH_2$(4-Cl)phenyl	n-Bu	(methyl, tetrazole-phenyl)

Table 1. (Continued)

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

ID	R_1	R_2	R_3	R_4	ID	R_1	R_2	R_3	R_4
2	4-Cl-3-CH₃-C₆H₃-NH-C(O)-C₂H₅	n-Pr	F	SO₂NH-C(O)-C₆H₄(2-Cl)	31	2-CF₃-C₆H₄-CH₂	n-Bu	H	SO₂NH-C(O)-C₆H₅
3	4-Cl-3-CH₃-C₆H₃-NH-C(O)-C₆H₅	C₂H₅	F	SO₂NH-C(O)-O-C[CH₃]₃	32	2-CF₃-C₆H₄-CH₂	n-Pr	H	SO₂NH-C(O)-C₆H₅
4	4-Br-3-CH₃-C₆H₃-NH-C(O)-C₂H₅	C₂H₅	F	SO₂NH-C(O)-O-C[CH₃]₃	33	2-CF₃-C₆H₄-CH₂	n-Bu	H	SO₂NH-C(O)-benzofuran-2-yl
11	2-F-C₆H₄-CH₂	n-Pr	H	tetrazol-5-yl	34	2-CF₃-C₆H₄-CH₂	n-Bu	H	SO₂NH-C(O)-[CH₂]₃COOH
12	5-OCH₃-2-NO₂-4-CH₃-C₆H₂	n-Bu	H	tetrazol-5-yl	35	2-CF₃-C₆H₄-CH₂	n-Bu	F	SO₂NH-C(O)-cyclohexyl
13	CH₂CH₃	n-Bu	F	tetrazol-5-yl	39	2-CF₃-C₆H₄-CH₂	n-Bu	H	SO₂NH-C(O)-thiophen-3-yl
14	CH₂-C₆H₄-C₆H₄(2-CN)	n-Bu	H	1,2,3-triazol-4-yl	44	Cl-CH₃-C₆H₃-NH-C(O)-C₂H₅	n-Bu	H	SO₂NH-C(O)-C₆H₄(2-Cl)
29	2-Cl-C₆H₄-CH₂	n-Bu	H	SO₂NH-C[CH₃]₃	45	3-Cl-C₆H₄-CH₂	n-Bu	H	tetrazol-5-yl
30	2-Cl-C₆H₄-CH₂	n-Bu	H	SO₂NH₂					

Quinazolinone derivatives:

ID	R_1	R_2	R_3	R_4	ID	R_1	R_2	R_3	R_4
59	NH₂	H	n-Bu	4′-CH₃-biphenyl-2-(tetrazol-5-yl)	61	NH-C(O)-C₆H₅	H	n-Pr	4′-CH₃-biphenyl-2-(SO₂NH-C(O)-C₆H₅)
60	H	CH₃	n-Bu	4′-CH₃-biphenyl-2-COOH					
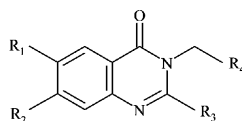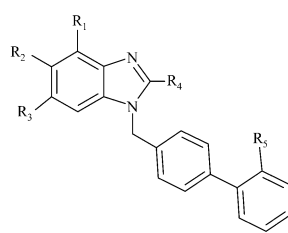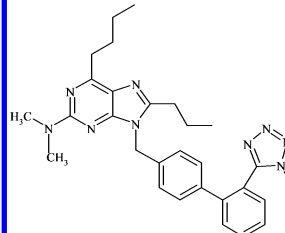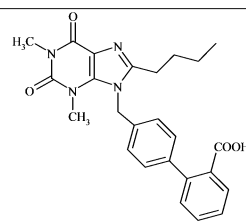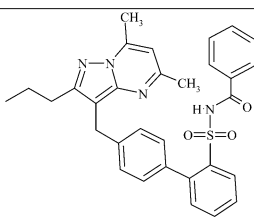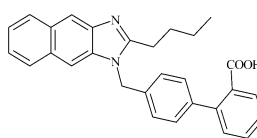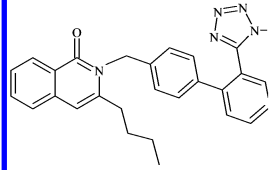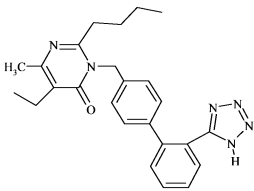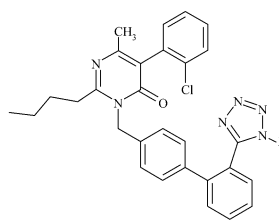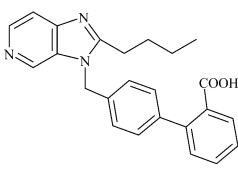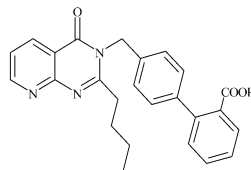
Table 1. (Continued)

Benzimidazole derivatives:

ID	R_1	R_2	R_3	R_4	R_5	ID	R_1	R_2	R_3	R_4	R_5
62	H	CH_3	CH_3	n-Bu	COOH	64	CH_3	H	H	n-Pr	COOH
63	H	$COOCH_3$	H	n-Bu							

Unique structures:

ID65   ID66   ID67   ID68

ID69   ID70   ID72   ID73

ID71

vertex points are found by projecting a cone away from each hydrogen bond donor or acceptor atom in the *ensemble* of molecules.

Each point can also be assigned a hydrophobic or hydrophilic attribute. This is distinguished by giving a value of 1.0 to a hydrophobic point and zero otherwise. A hydrophobic point is a point with low partial charge (absolute value less than 0.15), a low electrostatic potential (absolute value less than 0.01), and a low hydrogen bond-donating or -accepting propensity (absolute value less than 0.1).

The RSM can be constructed as open for the consideration of solvent accessible regions or can be generated as a closed surface enclosing some region of space. We have used the closed model and added the solvation correction factor, which penalizes the polar groups placed in a hydrophobic region by 0.3 kcal/(mol $Å^2$).

Nonbonded interaction energies are calculated between atoms of each molecule in the aligned set and the points on the receptor surface by means of van der Waals and electrostatic terms[33] as defined by the following two relationships:

$$E_{VDW} = K[(RA/r)^{12} - 2D(RA/r)^6]$$

where RA = $(VDWr)C_h$, with

RA = hybridization-corrected van der Waals radius of the atom,

VDWr = van der Waals radius,

C_h = hybridization factor,

r = distance between atom and surface point,

K = well-depth constant (set to 0.1 for all atom pair interaction), and

D = empirically derived point density scaling factor. D scales VDW energy and forces so that ideal atom/surface interactions yield a typical value of 0.0125 kcal/$Å^2$ of surface contact.

The electrostatic term is a monopole−monopole Coulombic function.

$$E_{ele} = (322.1Q_aQ_p/r)DS(r)$$

where r, Q_a, Q_p, and D are as stated above and $S(r)$ is an atom-based switching function with a cutoff of 8 Å.

(b) Receptor Surface Model Development. The receptor surface model is normally generated from the most active structures in the data set. The rationale is that the most active molecules tend to explore the best spatial and electronic interactions with the receptor, while the least active do not and tend to have unfavorable steric and electrostatic interactions. We initiated our study with receptor models generated using in the following order—most active, four most active, six most active, twelve most active, and finally all molecules in the training set as templates. All models generated using these strategies resulted in a prediction,

which overestimated the activity of the poorly active molecules; therefore a second strategy was implemented where models were generated using "QSAR activity data". In this methodology, RSA uses activity data from all the training set molecules that are used as templates to generate the receptor model. The activity data are used to weight the relative importance of each template structure in generating the receptor model.

As mentioned previously, the receptor surface model can be generated at a distance from the van der Waals surface of the set of atoms of a molecule with varying "surface fit value". We have studied the effect of three different surface fit values in the generation of the RSM on the QSAR equations. A tight fitting of -0.5 Å, the default value of 0.1 Å, and the loose-fitting value of 1.0 Å were all evaluated.

(c) Evaluation of Energies of Molecules within RSM. After the receptor surface model has been generated, all the structures in the training and test sets can be evaluated against the model. The model can be used to calculate the energy associated with the binding of a molecule in the model. It can also be used to minimize a molecule by adjusting the geometry of the structure into a "best-fit configuration" based on the constraints imposed by the receptor model. Three energy values can be calculated for each molecule by evaluating its interactions with the RSM. These are - energy (internal strain energy of the molecule as it sits in and is constrained by the receptor), $E_{interact}$ (interaction energy of the molecule with the receptor), and E_{strain} (internal strain energy of a molecule as it sits in the receptor without being subject to receptor model constraints). These energy terms help to understand the quality of fit of each structure with respect to RSM.

For each RSM generated either from a tight-, default-, and loose-surface-fit value, E_{strain} and $E_{interact}$ were examined. It was seen that all the molecules had an unfavorable (positive) E_{strain} and $E_{interact}$ value when evaluated against the RSM generated with a tight surface fit. For both loose- and default-surface-fit options, the corresponding energy values for all molecules were negative, suggesting a favorable condition.

We have generated models where molecules have been energy minimized or left without minimization within the receptor surface. It was observed that energy minimization of molecules within the receptor surface lowered the E_{strain} values for all of them.

The descriptors generally used to construct a QSAR[34] of RSM models are the electrostatic, steric, and total potential energy at each point on the receptor surface. To these we added the energy terms calculated during evaluation of the molecule within the receptor model. These were internal energy of the molecule inside the receptor (IntraEnergy), the electrostatic energy between molecule and the receptor (InterELEEnergy), the steric energy between the molecule and the receptor (InterVDWEnergy), the total energy between the molecule and the receptor (InterEnergy), the internal energy of the molecule without the receptor (MinIntra-Energy), and the strain energy of the molecule within the receptor (StrainEnergy). We refer to these as adjunct energy terms.

We have constructed a large number of receptor models by varying the surface fit (tight or default or loose), using different formalisms for assigning partial charges to the atoms of molecules, and mapping the electrostatic energies at points

on the surface either by electrostatic potential or partial charge complementarities. In all these models, the number of surface points were as large as 8000 or even more. Therefore data filtration techniques based on selecting every Nth surface point or selecting surface points with the highest variance or correlation were employed. The data filtration techniques also helped to make the calculations computationally less intensive.

The statistical method G/PLS was then used to derive meaningful relationships between the activity values and the RSA descriptors such as ELE, VDW, TOT, and the adjunct energy terms discussed earlier. The following parameters were set for deriving the QSAR equation. The descriptor values were scaled to a variance of 1.0, the optimal number of components was fixed as 4, the length of the equation was set to 5 terms with a smoothness of 1.0 (smoothness function controls the bias in the scoring factor between equations with different numbers of terms), and genetic crossovers were limited to 5000 generations.

The best models generated using the receptor surface descriptors were selected on the basis of equations having higher values for the square of the correlation coefficient (r^2) and cross-validated r^2 (q^2). The predictive power of the equations was estimated by cross-validation performed with the leave-one-out (LOO) technique on the training set and by calculating the predictive r^2 for the test set. The best models obtained were checked for chance correlation by carrying out randomization[35] and bootstrapping tests.[36]

RESULTS AND DISCUSSION

To derive the best predictive models various possible combinations of adjustable parameters such as atomic partial charge, surface fit, charge complementarities, and evaluation with or without minimization of molecules within the receptor surface were exploited. Some of the most significant equations that were obtained are described below.

The best model (eq 1) was obtained with nonlinear variables using Gasteiger Hückel atomic charges, the receptor surface constructed with a surface fit of 0.1, electrostatic charge complementarity for surface point potential, and descriptors of VDW, ELE, TOT, and the adjunct energy terms described earlier.

$$\text{activity} = 10.46 - 7.55(\langle \text{TOT}/914 - 0.0 \rangle) -$$
$$6.40(\langle \text{ELE}/2750 + 0.02 \rangle) + 2.49\,(\text{VDW}/4918) -$$
$$6.58\,(\langle \text{TOT}/1840 + 0.23 \rangle)\ (1)$$

$$n = 26, \quad r^2 = 0.928, \quad q^2 = 0.733, \quad \text{SDEP} = 0.840,$$
$$\text{LSE} = 0.116, \quad \text{BS}r^2 = 0.925, \quad \text{PRESS} = 4.155,$$
$$\text{SDEC} = 0.548, \quad \text{and}$$
$$\text{residual sum squared error} = 0.301$$

where n is the number of molecules, LSE is the least-squares error, $\text{BS}r^2$ is bootstrap r^2, PRESS is the predicted sum of squares, SDEP is the standard deviation of error of prediction, and SDEC is the standard deviation of error of calculation.

In eq 1, the term $\langle \text{TOT}/914 - 0.0 \rangle$ is a spline term with a "knot at 0.0" and TOT/914 is the sum of steric and electrostatic potentials at surface point 914. The descriptor ELE/2750 is the electrostatic interaction energy of the molecule with the receptor at point 2750, while descriptor
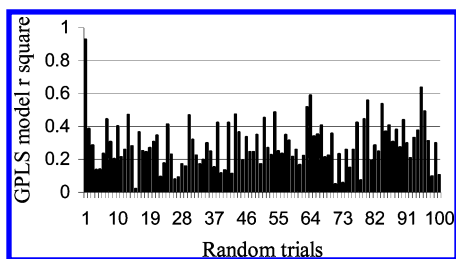
3D-QSAR OF AT1 RECEPTOR ANTAGONISTS

J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004 **217**

Figure 1. G/PLS randomization test at 99% confidence. The first bar is the r^2 value of the model based on actual activity data, and the remaining 99 bars are for r^2 derived from permuted data.

Table 2. Predicted Activities and Residuals of Test Compounds (39−73) for RSM Models As Described by Equations 1 and 2

molecule ID	activity ($-\log IC_{50}$)	RSA model (eq 1)		RSA model (eq 2)	
		pred activity	residual	pred activity	residual
39	8.20	8.66	−0.46	7.79	−0.41
40	9.30	9.98	−0.68	9.21	−0.09
41	9.20	8.18	1.02	10.39	1.19
42	5.32	5.13	0.19	5.93	0.61
43	8.48	8.02	0.46	9.60	1.12
44	9.55	9.75	−0.20	9.55	0.00
45	6.92	8.18	−1.26	5.81	−1.11
46	9.00	8.60	0.40	9.87	0.87
47	7.73	8.80	−1.07	9.08	1.35
48	6.80	6.34	0.46	8.15	1.35
49	7.26	8.10	−0.84	5.93	−1.33
50	8.06	9.28	−1.22	6.81	−1.25
51	8.72	9.63	−0.91	7.59	−1.13
52	7.60	8.19	−0.59	7.06	−0.54
53	6.79	8.20	−1.41	5.55	−1.24
54	7.62	8.11	−0.49	6.75	−0.87
55	9.57	9.39	0.18	10.08	0.51
56	9.68	8.84	0.84	10.68	1.00
57	9.03	8.08	0.95	10.09	1.06
58	7.89	7.94	−0.05	8.16	0.27
59	7.98	8.25	−0.27	7.94	−0.04
60	6.11	5.30	0.81	6.16	0.05
61	8.60	7.22	1.38	7.84	−0.76
62	5.77	5.28	0.49	5.47	−0.30
63	7.04	8.09	−1.05	6.27	−0.77
64	6.75	5.76	0.99	7.44	0.69
65	8.20	8.58	−0.38	8.32	0.12
66	6.91	5.69	1.22	7.67	0.76
67	8.84	9.66	−0.82	8.43	−0.41
68	6.00	5.38	0.62	5.91	−0.09
69	7.27	8.18	−0.91	6.43	−0.84
70	7.54	8.19	−0.65	6.71	−0.83
71	8.38	8.19	0.19	8.71	0.33
72	5.71	5.27	0.44	5.32	−0.39
73	6.36	5.15	1.21	6.68	0.32

VDW/4918 is the van der Waals interaction energy of the molecule with the receptor at point 4918.

The randomization test at 99% confidence was carried out by repeatedly permuting the activity values of the data set and evaluating the r^2 values. The r^2 values for 99 trials are shown in Figure 1. It is evident that the r^2 value of the original model is much higher than that of the trials with permuted data, suggesting the robustness of the model.[36] The activity predictions for molecules in the test set are listed in Table 2.

In another significant QSAR equation using ESP atomic charges, a nonlinear relationship was obtained with solely the TOT (sum of steric and electrostatic) descriptor. For this model (eq 2) the receptor surface was constructed with a
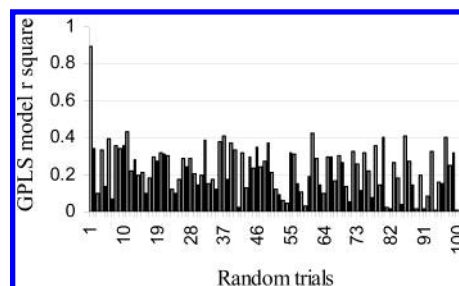

Figure 2. r^2 values for 100 runs of the randomization test. For details see Figure 1.

surface-fit value of 0.1 and the potential on the receptor surface points expressed by electrostatic charge complementarity.

$$\text{activity} = 7.70 - 4.01(\text{TOT}/1682) - 77.98(\langle \text{TOT}/695 + 0.01 \rangle) - 1.01(\text{TOT}/2357) - 2.39(\langle \text{TOT}/3471 + 0.09 \rangle) \quad (2)$$

$$n = 28, \quad r^2 = 0.893, \quad q^2 = 0.639, \quad \text{SDEP} = 0.775,$$
$$\text{LSE} = 0.172, \quad \text{BS}r^2 = 0.884, \quad \text{PRESS} = 7.06,$$
$$\text{SDEC} = 0.356, \quad \text{and}$$
$$\text{residual sum squared error} = 0.127$$

The randomization test, as shown in Figure 2, assured us that the model was not a result of chance correlation.

For the two best receptor surface models which are described by eq 1 and eq 2 the predictive r^2 (r^2_{pred}), for a set of test molecules that were quite diverse from the training set, were 0.535 and 0.538, respectively.

Mapping of RSM. In RSA, the information about the receptor site is displayed graphically by mapping properties onto the surface. Regions of the surface are color-coded to indicate particular chemical properties. The intensity of the color on the surface corresponds to the magnitude of the property. The location and magnitude of a descriptor can be used as a guideline to improve the activity of molecules. For each molecule, the descriptor that appears in a RSM equation can be associated with a structural feature of a molecule located closest to the position of that variable. Parts a and b of Figure 3 show a graphical representation of the receptor constructed over the training set of molecules, which have been aligned by consensus dynamics. The steric (Figure 3 a) and electrostatic fields (Figure 3b) are depicted on the receptor surface as colored regions. The magenta color represents negative energy values as favorable interaction sites, while the green-colored regions represent positive energy values that are unfavorable sites for binding of a molecule on the receptor surface. The locations of some important descriptors that have emerged from the model of the receptor surface analysis as described by eq 1 are displayed in Figure 3a,b. For the ease of discussion of the effect of nonbonded interactions at these surface points on the activity we have demarcated regions a−g around the highest active molecule, and this is shown in Figure 4.

A significant number of RSM equations were generated using Gasteiger Hückel atomic charges (e.g., eq 1) which shows a linear relationship with VDW at point 4918 (near the "d" region) on the receptor surface. The VDW interaction at point 4918 is more favorable for the ethyl than the butyl
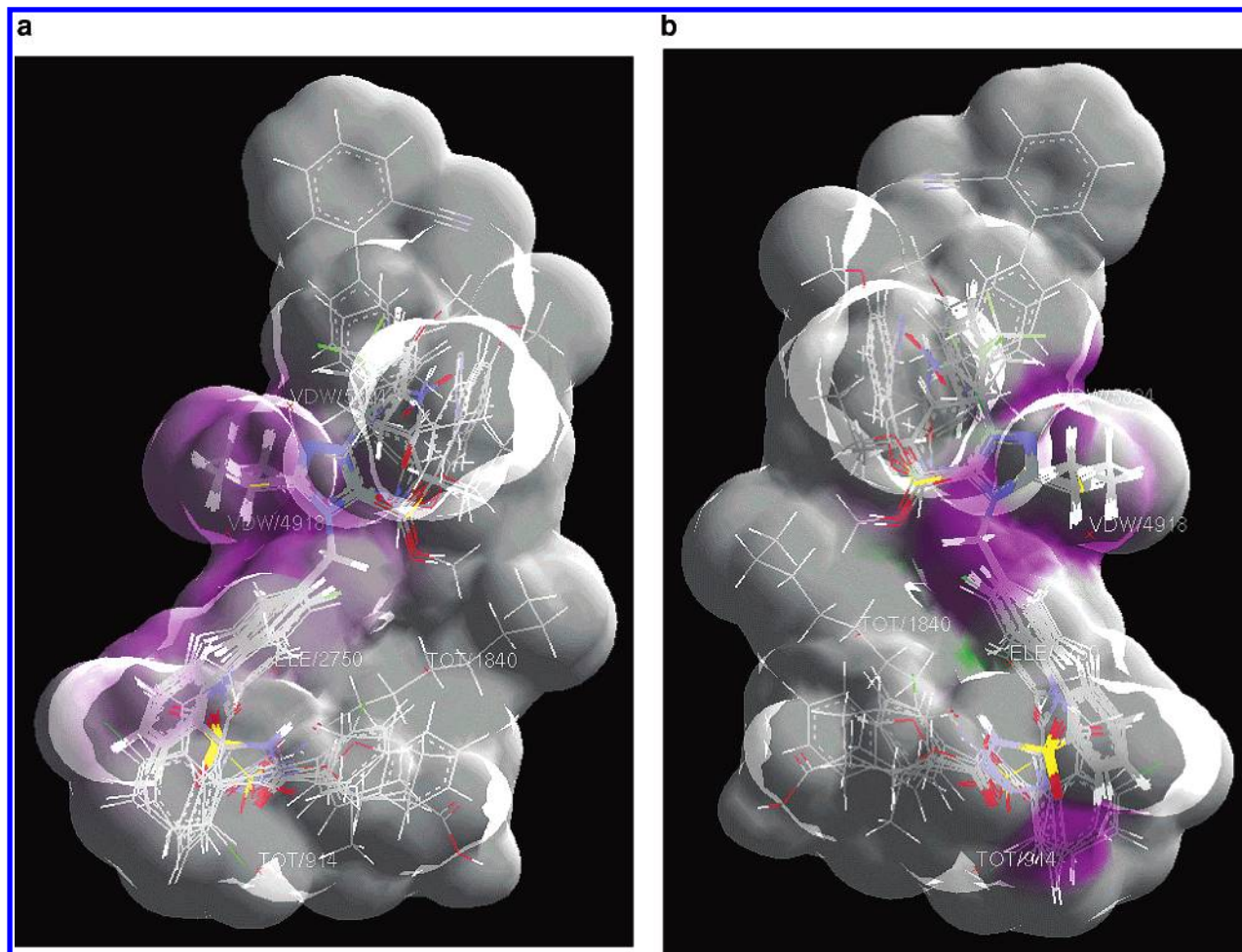
Figure 3. Graphical representation of the receptor surface generated around the training set of molecules showing (a) steric fields and (b) electrostatics fields.
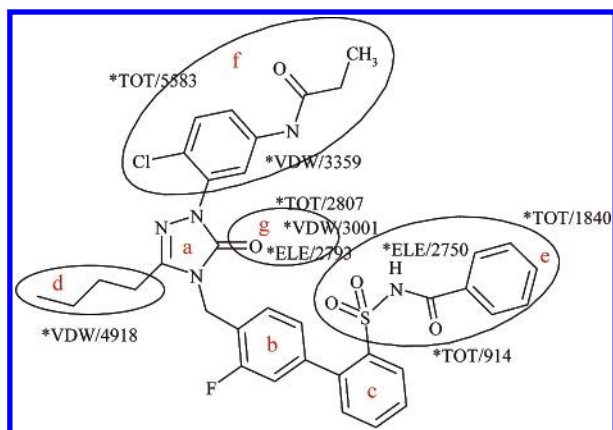

Figure 4. RSA descriptors with their locations in the regions demarcated as a−g around the most active molecule (ID2 in Table 1).

group. This implies that at the site marked d in Figure 4, a two-carbon chain is optimal. The remaining terms in eq 1 are nonlinear, i.e., ELE/2750, TOT/914, and TOT/1840 (all near the "e" region), and these descriptors describe an unfavorable effect for molecules containing -COOH, -CN, -SO$_2$NH$_2$, -SO$_2$NHC(CH$_3$)$_3$, -NHSO$_2$NHCOPh, and -CONHSO$_2$Ph functional groups, while the -SO$_2$NHCOPh group seems to be the most ideal substituent at this site. In addition to these, some of variables that frequently appear

in the family of RSM equations are VDW/3001 and ELE/2793 for the "g"-related region, ELE/4159 for ring "a", and VDW/3359 for the region indicated by "f".

Equation 2 has four TOT descriptors that are correlated with the activity. All four variables TOT/695, TOT/1682, TOT/2357, and TOT/3471 describe the same locations and properties as explained by TOT/914, TOT/1840, and ELE/2750 terms in eq 1. Thus it is interesting to note that both the equations lay a large emphasis on the e-related region indicating a major contribution for groups in this site toward a specific AT1 receptor blockade.

An analysis of the family of equations generated using ESP atomic charges indicates some descriptors occur frequently, such as TOT/2807 and TOT/5583. Out of these TOT/2807 is located near the g-marked region (Figure 4), where substituents such as -COOCH$_3$, -OCH$_3$, and -CH$_2$OH seem to contribute unfavorably to the activity, while -COOH and -SCH$_2$Ph−R moieties impart a favorable effect on the activity. The term TOT/5583 is located near the region marked f in Figure 4, where aromatic groups such as phenyl, trifluromethylphenyl, and 2,4,6-trichlorophenyl and aliphatic groups such as methyl and ethyl contribute positively toward the activity while unsubstituted phenyl rings in f which lack the electrostatic contribution make a poor effect toward the activity.

These observations are in line with some of the structure−activity relationships[13−17] reports. Our RSA models have

3D-QSAR OF AT1 RECEPTOR ANTAGONISTS

J. Chem. Inf. Comput. Sci., Vol. 44, No. 1, 2004 **219**

identified two new points near the biphenyl moiety that can be explored for further improvements of AT1 antagonists.

CONCLUSIONS

Receptor surface models do not contain atoms but give a representation of the essential features of an active site by assuming complementarity between the shape and properties of the receptor site with respect to the set of binding molecules.

Out of the various receptor surface models generated in the present work, the best models eq 1 and eq 2 gave a predictive r^2 of 0.535 and 0.538 respectively, for a set of test molecules that were quite diverse from the training set. The graphical representation of nonbonded interactions, hydrophobicity, and H-bonding on the receptor surface provided a clear picture of regions on a molecule that can be modified to improve its activity. The structure−activity relationships that have emerged from this study can be used advantageously in the development of new and specific AT1 antagonists.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Dzau, V. J.; Pratt, R. E. Renin−Angiotensin System: Biology, Physiology and Pharmacology. In *The Heart and Cardiovascular System*; Fozzard, H. A., Haber, E., Jennings, R. B., Katz, A. M., Morgan, H. E., Eds.; Raven Press: New York, 1986; pp 1631−1662.

(2) Joseph, M. P.; Maigret, B.; Bonnafous, J. C.; Marie, J.; Scheraga, H. A. A Computer Modeling Postulated Mechanism for Angiotensin II Receptor Activation. *J. Protein Chem.* **1995**, *14*, 381−398.

(3) Nikiforovich, G. V.; Marshall, G. R. 3D Model for TM Region of the AT-1 Receptor in Complex with Angiotensin II Independently Validated by Site-Directed Mutagenesis Data. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 1204−1211.

(4) Laura, B.; Bravi, G.; Scolastico, C.; Vulpetti, A.; Salimbeni, A.; Todeschini, R. A 3D QSAR Approach to the Search for Geometrical Similarity in a Series of Nonpeptide Angiotensin II Receptor Antagonists. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 211−220.

(5) Laura, B.; Bravi, G.; Catalano, G.; Mabilia, M.; Salimbeni, A.; Scolastico, C. A 3D QSAR CoMFA Study of Non-peptide Angiotensin II Receptor Antagonists. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 567−582.

(6) Prendergast, K.; Adams, K.; Greenlee, W. J.; Nachbar, B.; Patchett, A. A.; Underwood, D. J. Derivation of a 3D Pharmacophore Model for the Angiotensin-II Site One Receptor. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 491−512.

(7) Datar, P.; Desai, P.; Coutinho, E.; Iyer, K. CoMFA and CoMSIA Studies of Angiotensin (AT1) Receptor Antagonists. *J. Mol. Model.* **2002**, *8*, 290−301.

(8) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem.* **1995**, *38*, 2080−2090.

(9) Cramer, R. D.; 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.

(10) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure−Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.

(11) Holland, J. *Adaptation in Artificial and Natural Systems*; University of Michigan Press: Ann Arbor, MI, 1975.

(12) Friedman, J. *Multivariate Adaptive Regression Splines*; Technical Report 102, Laboratory for Computational Statistics, Department of Statistics, Stanford University: Stanford, CA, 1988; revised 1990. Gasteiger, J.; Marsali, M. *Tetrahedron Lett.* **1978**, *34*, 3181.

(13) Naylor, E. M.; Chakravarty, P. K.; Costello, A. C.; Chang, R. S.; Chen, T.; Faust, K. A.; Lotti, V. J.; Kivlighn, S. D.; Zingaro, G. J.; Siegl, P.

K. S.; Pancras, C. W.; Carini, D. J.; Wexler, A. A.; Greenlee, W. J. Potent Imidazole Angiotensin II Antagonists: Acyl Sufonamides and Acyl Sulfamides as Tetrazole Replacements. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 69−74.

(14) Wallace, T. A.; Steven, M. H.; Greenlee, W. J.; Doss, G. A.; Chang, R. S. L.; Lotti, V. J.; Faust, K. A.; Chen, T.; Zingaro, G. J.; Kivlighn, S. D.; Siegl, P. K. S. Nonpeptide Angiotensin II Antagonists Derived from 1*H*-Pyrazole-5-carboxylates and 4-Aryl-1*H*-imidazole-5-carboxylates. *J. Med. Chem.* **1993**, *36*, 3595−3605.

(15) (a) Chakravarty, P. K.; Naylor, E. M.; Chang, R. S. L.; Chen, A.; Chen, T.; Faust, K. A.; Lotti, V. J.; Kivlighn, S. D.; Zingaro. G. J.; Schorn, T. W.; Schaffer, L. W.; Siegl, P. K. S.; Patchett, A. A.; Greenlee, W. J. A Highly Potent, Orally Active Imidazo[4,5-*b*]pyridine Biphenylacylsulfonamide [MK-996; L-159-282]: A New AT$_1$ selective Angiotensin II Receptor Antagonist. *J. Med. Chem.* **1994**, *37*, 4068−4072. (b) Mantlo, N. B.; Chakravarty, P. K.; Ondeyka, D. L.; Siegl, P. K. S.; Chang, R. S. L.; Lotti, V. J.; Faust, K. A.; Chen, T.; Schorn, T. W.; Sweet, C. S.; Emmert, S. E.; Patchett, A. A.; Greenlee, W. J. Potent, Orally Active Imidazo[4,5-*b*]pyridine Based Angiotensin II Receptor Antagonists. *J. Med. Chem.* **1991**, *34*, 2919−2922. (c) Dhanoa, D. S.; Bagley, S. W.; Chang, R. S. L.; Lotti, V. J.; Chen, T.; Siegl, P. K. S. Chakravarty, P. K.; Patchett, A. A.; Greenlee, W. J. (Dipropylphenoxy)phenylacetic Acids: A New Generation of Non-peptide Angiotensin II Receptor Antagonists. *J. Med. Chem.* **1993**, *36*, 3738−3742. (d) Dhanoa, D. S.; Bagley, S. W.; Chang, R. S. L.; Lotti, V. J.; Chen, T.; Siegl, P. K. S.; Zingaro, G. J.; Patchett, A. A.; Greenlee, W. J. Non-Peptide Angiotensin II Receptor Antagonists. 1. Design, Synthesis, and Biological Activity of N-Substituted Indoles and Dihydroindoles. *J. Med. Chem.* **1993**, *36*, 4230−4238.

(16) (a) Linda, L. C.; Wallace, T. A.; Flanagan, K. L.; Chen, T.; Zingaro, G. J.; Siegl, P. K. S.; Kivlighn, S. D.; Lotti, V. J.; Chang, R. S. L.; Greenlee, W. J. Triazolinone Biphenylsulfonamides as Angiotensin II Receptor Antagonist with High Affinity for Both the AT$_1$ and AT$_2$ Subtypes. *J. Med. Chem.* **1994**, *37*, 4464−4478. (b) Linda, L. C.; Flanagan, K. L.; Wallace, T. A.; Naylor, E. M.; Chakravarty, P. K.; Patchett, A. A.; Greenlee, W. J.; Chen, T.; Faust, K. A.; Chang, R. S. L.; Lotti, V. J.; Zingaro, G. J.; Siegl, P. K. S.; Kivlighn, S. D. Triazolinone Biphenylsulfonamide Derivatives as Orally Active Angiotensin II Antagonists with Potent AT$_1$ Receptor Affinity and Enhanced AT$_2$ Affinity. *J. Med. Chem.* **1994**, *37*, 2808−2824. (c) Linda, L. C.; Wallace, T. A.; Flanagan, K. L.; Greenlee, W. J.; Chang, R. S. L.; Lotti, V. J.; Faust, K. A.; Chen, T.; Zingaro, G. J.; Kivlighn, S. D.; Siegl, P. K. S.; Bunting, P.; MacCoss, M.; Strelitz, R. A. Triazolinones as Nonpeptide Angiotensin II Antagonists. 1. Synthesis and Evaluation of Potent 2,4,5-Trisubstituted Triazolinones *J. Med. Chem.* **1993**, *36*, 2558−2568. (d) Linda, L. C.; Wallace, T. A.; Flanagan, K. L.; Chen, T.; Zingaro, G. J.; Siegl, P. K S.; Lotti, V. J.; Chang, S. L.; Greenlee, W. J.; O'Malley, S. S. Potent and Orally Active Angiotensin II Receptor Antagonists with Equal Affinity for Human AT$_1$ and AT$_2$ Subtypes. *J. Med. Chem.* **1995**, *38*, 3741−3758.

(17) Linda, L. C.; Wallace, T. A.; Cantone, C. L.; MacCoss, S. S.; Chang, R. S. L.; Lotti, V. J.; Faust, K. A.; Chen, T.; Strelitz, R. A.; Bunting, P.; Kivlighn, S. D.; Siegl, P. K. S. Nonpeptide Angiotensin II Antagonists Derived from 4*H*-1,2,4-Triazoles and 3*H*-Imidazo[1,2-*b*]-[1,2,4]triazoles. *J. Med. Chem.* **1993**, *36*, 591−609.

(18) (a) Allen, E. E.; de Laszlo, S. E.; Huang, S. X.; Quagliato, C. S.; Greenlee, W. J.; Chang, R. S.; Chen, T. B.; Faust, K. A.; Lotti, V. J. Quinazolinones 1: Design and Synthesis of Potent Quinazolinone-Containing AT$_1$-Selective Angiotensin II Receptor Antagonists. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1293−1298. (b) Bagley, S.; Greenlee, W. J.; Dhanoa, D. S.; Patchett, A. A. U.S. Pat. No. 5,175,164, 1992. (c) Glinka, T. W.; de Laszlo, S. E.; Siegl, P. K. S.; Chang, R. S. L.; Kivlighn, S. D.; Schorn, T. S.; Faust, K. A.; Chen, T. B.; Zingaro, G. J.; Lotti, V. J.; Greenlee, W. J. A New Class of Balanced AT$_1$/AT$_2$ Angiotensin II Antagonists: Quinazolinone AII Antagonists with Acylsulfonamide and Sulfonylcarbamate Acidic Functionalities. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 81−86.

(19) (a) Chakravarty, P. K.; Camara, V. J.; Chen, A.; Marcin, L. M.; Greenlee, W. J.; Patchett, A. A.; Chang, R. S. L.; Lotti, V. J.; Siegl, P. K. S. 200th National Meeting of the American Chemical Society; American Chemical Society,: Washington, DC, August 1990; pp 26−31. (b) Chakravarty, P. K.; Patchett, A. A.; Camara, V. J.; Walsh, T. F.; Greenlee, W. J. Eur. Pat. Appl. 400,835, 1990.

(20) (a) Ondeyka, D. L.; Mantlo, N. B.; Chakravarty, P. K.; Chen, A.; Camara, V. J.; Chang, R. S. L.; Lotti, V. J.; Siegl, P. K. S.; Patchett, A. A.; Greenlee, W. J. Joint Central-Great Lakes Regional Meeting of the American Chemical Society, Indianapolis, May 31, 1991. (b) Chakravarty, P. K.; Greenlee, W. J.; Mantlo, N. B.; Patchett, A. A.; Walsh, T. F. Eur. Pat. Appl. 400,974, 1990.

(21) Allen, E. E.; Huang, S. X.; Chang, R. S. L.; Lotti, V. J.; Siegl, P. K. S.; Patchett, A. A.; Greenlee, W. J. 202nd National Meeting of the American Chemical Society, New York, Aug. 25−30, 1991.

(22) Patchett, A. A.; Mantlo, N. B.; Greenlee, W. J. U.S. Pat. No. 5,124,335, 1992.

(23) Greenlee, W. J.; Johnston, D. B. R.; MacCoss, M.; Mantlo, N. B.; Patchett, A. A.; Chakravarty, P. K.; Walsh, T. F. U.S. Pat. No. 5,164,407, 1992.

(24) Allen, E. E.; Greenlee, W. J.; Chakravarty, P. K.; Patchett, A. A.; Walsh, T. F. U.S. Pat. No. 5,100,897, 1992.

(25) *Cerius^2 Modeling Environment*, Release 4.0; Accelrys Inc.: San Diego, CA, 1999.

(26) *Discover*, v.98, Cerius2 Modeling Environment; Accelrys Inc.: San Diego, CA, 1999.

(27) Lorensen, W. E.; Cline, H. E. Marching Cubes: A High Resolution 3-D Surface Construction Algorithm. *Comput. Graphics (Proc. SIGGRAPH, 21)* **1987**, *4*, 163−169.

(28) *SYBYL 6.7*, Force Field Manual; Tripos Inc.: St Louis, MO; p 290 (Oct 2000).

(29) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(30) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. Structure and Energetics of Ligand Binding to Proteins: E. Coli Dihydrofolate Reductase-Trimethoprim, A Drug-Receptor System. *Proteins: Struct., Funct. Genet.* **1988**, *4*, 31−37.

(31) Stewart, J. J. P. MOPAC: A Semiempirical Molecular Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−105.

(32) Rappé, A. K.; Goddard, W. A., III. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95*, 3358.

(33) *Cerius^2 User Guide*, July 1998; Accelrys Inc.: San Diego, CA, 1998.

(34) Hahn, M.; Rogers, D. Receptor Surface Models. 2. Application to Quantitative Structure−Activity Relationship Studies. *J. Med. Chem.* **1995**, *38*, 2091−2102.

(35) Leger, C.; Politis, D. N.; Romano, J. P. Bootstrap Technology and Applications. *Technometrics* **1992**, *34*, 378−399.

(36) Shi, L. M.; Fan, Y.; Myers, T. G.; O'Connor, P. M.; Paull, K. D.; Friend, S. H.; Weinstein, J. N. Mining the NCI Anticancer Drug Discovery Databases: Genetic Function Approximation for the QSAR Study of Anticancer Ellipticine Analogues. *J. Chem. Inf. Comput. Sci.* **1998**, 38, 189−199.