Quantitative Antidiabetic Activity Prediction for the Class of Guanidino- and Aminoguanidinopropionic Acid Analogs Based on Electron-Conformational Studies

Aleksandr V. Marenich, Pei-Han Yong, Isaac B. Bersuker, and James E. Boggs*

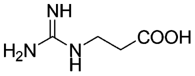
Institute for Theoretical Chemistry, Department of Chemistry and Biochemistry,
The University of Texas at Austin, 1 University Station A5300, Austin, Texas 78712

Received November 1, 2007

The electron-conformational method has been employed to reveal the pharmacophore (Pha) and to predict antidiabetic activity, studying 154 compounds in the class of guanidino- and aminoguanidinopropionic acid analogs. The derived Pha consists of four sites with certain electronic and topological characteristics which are represented by two oxygen atoms of the carboxyl group and two nitrogens of the guanidine group but may be substituted with any other atoms that have the same electronic and geometric features. The Pha flexibility and the influence of out-of-Pha features are described by only three model descriptors that predict the experimental activities quantitatively within experimental uncertainty for a training set of 120 compounds. The quality of the derived Pha and the corresponding quantitative model of activity has been validated (and deemed acceptable) by cross-validation including many-fold cross-validations within the training set and against an independent test set of 34 additional analogs with known experimental activities out of the training set. At last, several dozen compounds never tested experimentally have been screened theoretically using this model, and statistically significant hypoglycemic activities for a few of them are predicted.

INTRODUCTION

Non-insulin-dependent diabetes mellitus (type 2) is a long-term metabolic disorder of complex etiology characterized by insulin resistance of the peripheral target tissues, resulting in elevated blood glucose levels (hyperglycemia).[1] Various chemicals can antagonize hyperglycemia through improving the insulin sensitivity, reducing the glucose production (glycogenolysis and gluconeogenesis) from various substrates in the liver or elsewhere, or slowing the intestinal digestion of carbohydrates.[1] Improvement in insulin sensitivity has been proposed in recent studies on the antihyperglycemic activity of guanidinopropionic acid **1** and its derivatives as potential antidiabetic agents in the rodent model of non-insulin-dependent diabetes mellitus.[2,3]



1

Guanidino- and aminoguanidinopropionic acid analogs with strong hydrophilicity may be favored over lipophilic guanidine antidiabetic agents like biguanides[1] associated with lactic acidosis.[1,2] Although the biochemical mechanism for antidiabetic effect of **1** and its analogs has been deemed obscure,[2,3] a structure−activity relationship (SAR) analysis[2,3] shows that the biological activity of these compounds strongly depends on their molecular structure features: the biological activity of **1** tolerates only minor modifications such as isomerization and substitutions by an amino group.[2,3] One can assume that these molecules interact with a specific biological target (bioreceptor) and that there is an ensemble of some steric and electronic features that can facilitate the binding of the molecules with the receptor's active sites. The experimental data on guanidino- and aminoguanidinopropionic acid derivatives[2,3] provide a sufficient basis for a computer-aided study of quantitative structure−activity relationship (QSAR) in this class of compounds with the ultimate purpose of screening and prediction of other potentially even more active drugs for non-insulin-dependent diabetes mellitus.

During the last decades, computer-based biological activity prediction has been recognized as a cost-efficient tool for modern drug discovery in view of the high, and increasing, costs of experimental testing.[4] A few SAR and QSAR studies have been performed recently for screening new antidiabetic agents (see refs 5−7 and references therein) including the structure-based design and evaluation of ligand−receptor binding thermodynamics applied to a set of glucose analog inhibitors of glycogen phosphorylase,[5] a QSAR model for the thiazolidinedione- and tyrosine-based ligands of peroxisome proliferator-activated receptors,[6] and a molecular topology study of various active and inactive compounds with regards to hypoglycemic activity.[7]

We report here the results of pharmacophore (Pha) identification and quantitative prediction of hypoglycemic activity for the class of guanidino- and aminoguanidinopropionic acid analogs[2,3] using the electron-conformational (EC) method the modern (quantitative) version of which was first introduced in refs 8 and 9 and then reformulated in more detail in the review paper.[10] An older less accurate qualitative version of the electron-conformational method called the electron-topological method was worked out earlier[11] (see also Further Discussion). Distinguished from traditional QSAR approaches, the EC method describes the properties of a molecule in its interaction with the bioreceptor by means

* Corresponding author phone: (512)466-9145; fax: (512)471-8696; e-mail: james.boggs@mail.utexas.edu.

10.1021/ci700401p CCC: $40.75
Published on Web 02/16/2008

of electronic and geometric features obtained from direct quantum-chemical calculations without any arbitrary descriptors. The computational results are conveniently presented in terms of interaction indices,[9,10] bond orders, and interatomic distances arranged in the form of elements of the electron-conformational matrix of congruity (ECMC).[8−10] Multiple comparisons of these matrices in relation to the activities of the corresponding compounds by means of special computer programs allow one to separate a smaller electron-conformational submatrix of activity (ECSA) which is common for all active compounds and may not be present in the inactive ones. This set of matrix elements (electronic and geometry characteristics) is a rather full description of the given activity feature called the pharmacophore.[4] Identification of the Pha as the necessary condition of activity allows one to qualitatively predict activities (active/inactive) by calculating the ECMC of nontested molecules and revealing whether or not they have the ECSA.

The quantitative activity prediction by the EC method involves an appropriate handling of the multiconformation problem as well as a rather sophisticated parametrization of the pharmacophore flexibilities and the influence of out-of-Pha functional groups shielding the Pha.[8−10] The EC method has been recently applied to a variety of problems such as screening rice blast activity inhibitors,[8] angiotensin converting enzyme inhibitors,[9] group I metabotropic glutamate receptor agonists,[12] and inhibitors of human breast carcinoma[13] (see also refs 14−17 and references therein for applications of other electron-topological approaches).

In this paper, the EC method[8−10] is employed to study 154 guanidino- and aminoguanidinopropionic acid analogs[2,3] as potential antidiabetic agents with a full exploration of its two main parts: Pha identification and quantitative activity prediction. In the first part, ECMCs for the molecules of interest were calculated using semiempirical and ab initio quantum-chemistry methods of appropriate quality and processed by means of special programs to reveal the Pha. The quantitative prediction of activity based on the regression model is discussed in detail. The quality of the derived Pha and the corresponding quantitative model of activity was validated by cross-validation (including many-fold cross-validations) within the training set and against a test set of 34 additional analogs with known experimental activities. The test set was derived from the set of 154 analogs. This test set was not used for training the model. The remaining compounds (120 as total) served as a training set. The results obtained in this way allowed us to perform computer-based screening of several dozens of new compounds (never tested before) and to predict for a few of them statistically significant hypoglycemic activities.

METHODS AND MATERIALS

The electron-conformational method consists of three consecutive steps:[10] (1) construction of the ECMCs for the active and inactive compounds of a selected training set, their processing in relation to the activity by multiple comparison, and identification of the Pha (quantum chemical evaluation); (2) estimation of the influence of Pha flexibility, anti-Pha shielding groups, and other factors beyond the Pha by means of a proper parametrization and least-squares regression analysis; (3) and use of the obtained Pha description and formulas modeling the influence of Pha flexibility, anti-Pha shielding groups, and other factors beyond the Pha for screening new potentially active compounds and prediction of their activity.

Computation of Electron-Conformational Matrices of Congruity. To reveal the conformers heavily populated at room temperature for each molecule of the training set, a conformational analysis was performed using the Monte Carlo randomized search method with the molecular mechanics Merck force field.[18] Geometries for conformers located in this way were optimized with the semiempirical PM3 method[19] using analytical gradients. Then, single-point calculations of total energies and electronic properties of conformers (molecular orbital population analysis, partial atomic charges, and bond orders) were carried out with second-order Møller-Plesset (MP2) perturbation theory[20] using 6-31G(d) electronic basis sets.[21] The frozen core approximation was used in the MP2 calculations. Conformational analysis and electronic structure calculations were performed using the *Spartan* package.[22,23] The methods chosen here provide for an optimal balance of computational time-saving and the quality of models proven by variation of theoretical levels in the calculation on a few selected molecules.

Calculated electronic and molecular structure data are used for construction of the electron-conformational matrix of congruity. Off-diagonal matrix elements of the ECMC represent bond orders for chemically bonded atoms and interatomic distances for nonbonded pairs. In this work we use bond orders from Mulliken population analysis,[24] but one can use any other reasonable type of bond orders within the EC method. Diagonal matrix elements of the ECMC reproduce atomic interaction indices (II),[9,10] which are measures of electron-donor properties of the corresponding atoms in the molecule. In this paper, we derive the following form for the interaction indices (in Hartree units)

$$\mathrm{II} = g_{nl} \int_{r_o}^{\infty} \{R_{nl}(r)\}^2 r^2 \, \mathrm{d}r \qquad (1)$$

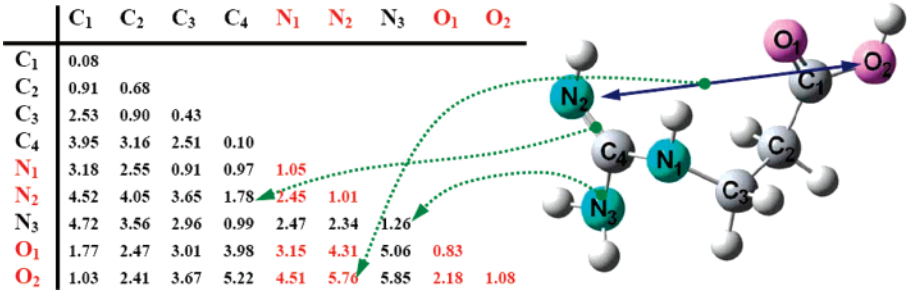
where r is a variable radius, g_{nl} (for instance, g_{1s}, g_{2s}, g_{2p}, etc.) is the electron population of the outermost orbital on a given atom (g_{np} for np-elements is equal to a third of the total occupancy of valence p-orbitals: p_x, p_y, and p_z). The function of $R_{nl}(r)$ is a radial component of the corresponding hydrogen-like atomic orbital[25]

$$R_{nl}(r) = -\left[(2\sqrt{2\epsilon})^3 \frac{(n-l-1)!}{2n\{(n+l)!\}^3}\right]^{1/2} e^{-r\sqrt{(2\epsilon)}} (2r\sqrt{2\epsilon})^l L_{n+l}^{2l+1}(2r\sqrt{2\epsilon}) \qquad (2)$$

where L_{n+l}^{2l+1} is the associated Laguerre polynomials (see for instance, ref 25). The quantity ϵ is the valence orbital ionization potential (VOIP) of the atom-in-molecule orbital calculated as a function of the atomic charge q and the electronic configuration of the atom

$$\epsilon(q) = Aq^2 + Bq + C \qquad (3)$$

where A, B, and C are reference VOIP data.[26] In this work we use partial atomic charges from the Mulliken population analysis,[24] but the formula for the interaction index can be



	C_1	C_2	C_3	C_4	N_1	N_2	N_3	O_1	O_2
C_1	0.08								
C_2	0.91	0.68							
C_3	2.53	0.90	0.43						
C_4	3.95	3.16	2.51	0.10					
N_1	3.18	2.55	0.91	0.97	1.05				
N_2	4.52	4.05	3.65	1.78	2.45	1.01			
N_3	4.72	3.56	2.96	0.99	2.47	2.34	1.26		
O_1	1.77	2.47	3.01	3.98	3.15	4.31	5.06	0.83	
O_2	1.03	2.41	3.67	5.22	4.51	5.76	5.85	2.18	1.08

Figure 1. The electron-conformational matrix of congruity for molecule **1**. Hydrogen atoms are omitted for simplicity from the ECMC. The diagonal elements refer to the atomic interaction indices calculated by eq 1, while the off-diagonal elements reproduce Mulliken bond orders for chemically bonded pairs of atoms and interatomic distances for nonbonded pairs. The electron-conformational submatrix of activity, the diagonal elements of which correspond to the atoms N_1, N_2, O_1, and O_2, is highlighted.

readily employed with any type of partial charges. Numerical integration in eq 1 is performed using the finite domain defined by the van der Waals radius of an atom (r_o) as a lower limit. We use Bondi's van der Waals radii.[27] The upper limit can be set to any large value of r so that any further increase of r should not affect the value of II for any atom (it has been set to 100 Å in the present paper). Thus the quantity of II equals the fraction of the electronic population on a given atom that can participate in nonbonding (noncovalent) interactions (for instance, can be donated for hydrogen bonding to facilitate a docking of a ligand with its biological receptor). In other words, the larger value of II corresponds to the higher electron-donor ability of an atom (for instance, in the case of nitrogen and oxygen), whereas the lower value can indicate the higher electron-acceptor ability (for instance, in the case of hydrogen).

Figure 1 illustrates an example of the ECMC calculated for the ground conformation of compound **1**. For simplicity, we exclude the hydrogen atoms in the ECMC from consideration hereafter.

Comparison of Electron-Conformational Matrices. The comparison of EC matrices is not a trivial procedure in general,[8−10] but it has been formalized now in a special FORTRAN algorithm called ELCOME (this is an abbreviation for the electron-conformational method). The procedure of comparison of the ECMCs begins with identification of the threshold of experimental activity that can separate active and inactive compounds in the training set. Then we make a comparison of the ECMCs for a limited number of 6−8 molecules that possess the most prominent experimental activities to identify an initial approximation for the electron-conformational submatrix of activity within some initial tolerances for its elements. For simplicity, at this step we compare only the ECMCs related to the lowest energy (ground) conformations. Choosing one of these ECMCs as the reference, we compare it to the ECMCs of the other molecules. At first, we consider only diagonal elements in the reference ECMC to exclude those that are not found in all other ECMCs within the given tolerances. Then we compare the off-diagonal elements (at the intersection of the rows and columns of the remaining diagonal ones) to identify the ECSA that is present in all ECMCs under consideration. Subsequently, choosing different ECMCs as the reference, we obtain the $N \times N$ submatrix that is optimal in terms of the least relative deviations (in %) of its elements from the corresponding ones in the ECMCs. By varying initial tolerances, one can identify several different pharmacophores even of different sizes that should be continued to the next step of comparison.

Next, we examine the ECMCs of all other active molecules to identify a submatrix of the ECMC in each active molecule as the best match to the approximate ECSA (the Pha) obtained in the previous step with a limited number of active molecules. We do not impose any tolerances to constrain the Pha at this step. The only criterion in selection of the optimal submatrix for each active molecule over multiple possibilities is the least relative deviations (in %) of the elements of such a submatrix from the corresponding ones in the approximate ECSA. As soon as the submatrices of activity are revealed in all of the active molecules the new values for the elements of the approximate ECSA are calculated by averaging the corresponding matrix elements of the ECMCs of all the active molecules. To consider a possibility for binding the biological target by higher-energy conformations of the ligand molecule we incorporate the ECMCs of the excited conformers into comparison. In this case the resulting ECSA is additionally averaged over all of the conformers having the Pha for each molecule with conformers. The procedure of comparison is iterative and must be repeated using the renovated ECSA several times to have the values of ECSA elements converged. When the final ECSA is obtained, the ECSA tolerances are calculated as the maximum deviations (in %) of the averaged values of ECSA elements from the corresponding values in the ECMCs of the molecules with the Pha including excited molecular conformers. Note that we do not use inactive molecules (as defined by the threshold of activity) in the evaluation of ECSA elements. The Pha derived in this way does not need any further optimization by means of alteration of the ECSA elements and further minimization of their tolerances. The final ECSA elements and their tolerances are not arbitrarily selected descriptors but functions of only electronic structure and geometry which are inherent and the most general characteristics of a molecule (within the accuracy of theoretical approximations and experimental data used).

Using the above procedure, one can identify several pharmacophores corresponding to different ECSAs with different tolerances but only one should be selected. The selection can be done by screening all inactive molecules in the training set to check whether or not the ECMCs of inactive molecules have a pharmacophore (or pharmacophores). One should prefer the Pha that provides the best possible separation of active and inactive molecules in the training set. In other words, active molecules should have the Pha, inactive molecules should not. However, the presence of Pha is a necessary but not sufficient condition of activity. Some compounds with low activities or even inactive compounds may still have the Pha, their activity being diminished by out-of-Pha functional groups and other factors which may obstruct their biological potency, as discussed in refs 8−10. It is important to continue the obtained Pha to the next step that is a quantitative modeling of the influence of the out-of-Pha environment affecting the Pha activity in a molecule. Thus selection of the right Pha should be based not only on good separation of active and inactive compounds in the training set but also on the predictive abilities of quantitative models corresponding to different versions of the Pha. The derived Pha should tolerate any reasonable cross-validation when removal of one or a few compounds from the Pha identification does not lead to a qualitatively different pharmacophore with a new qualitatively different set of quantitative descriptors.

Another possibility is to obtain different ECSAs for different levels of activity. If only very active compounds are sought for, one can simplify the search of Pha by trying only a limited number of the most active ones in the training set and to get the ECSA with minimum tolerances corresponding to such activities. The compounds with lower levels of activity will have a similar Pha but with larger tolerances. By following the changes in the tolerances when moving from more active to less active compounds one can reveal the role of the flexibilities of specific atoms (sites) in the activity quantitatively and the dependence of the latter on these flexibilities. However, we do not use this approach in the present paper.

Quantitative Model. According to refs 8−10, a general formula relating the biological activity of a ligand to the ligand−receptor binding affinity is expressed in terms of equilibrium for the ligand−receptor binding and with use of the Boltzmann distribution over the population of ligand conformations

$$a_i = a_{\mathrm{ref}} \exp\left(-\frac{E_i^{\mathrm{PHA}} - E_{\mathrm{ref}}^{\mathrm{PHA}}}{k_{\mathrm{B}}T}\right) \exp(-S_i(R)) \qquad (4)$$

$$S_i(R) = \sum_{j=1}^{P} k_j (R_j^{(i)} - R_j^{\mathrm{ref}}) \qquad (5)$$

where a_i and a_{ref} are activities (or functions of activities) of the ith compound and the reference compound, respectively, E^{PHA} is the relative energy of the lowest energy conformer that contains the Pha, and $S_i(R)$ is a function of the electronic and geometric parameters of the ligand molecule, where parameters R stand for the pharmacophore flexibility and the influence of anti-Pha shielding groups in the ligand molecule and other factors affecting activity. The value of $S(R)$ in eq 4 corresponds to a positive (if $S < 0$) or a negative ($S > 0$) addition to the value of a due to the pharmacophore flexibility and the influence of out-of-Pha environment for the compound under consideration as compared to the reference molecule. It seems reasonable to use the simplest compound with significant activity (guanidinopropionic acid **1**) as the reference system. Since the value of a in eq 4 decreases rapidly with the growth of E^{PHA} we consider only low-lying (below 1 kcal/mol) conformations of each molecule for quantitative activity prediction as well as for Pha identification (higher energy conformations are not significant in our case).

Model coefficients k in eq 5 are calculated by means of a least-squares fit of the values of a_i predicted by eq 4 for all the compounds in the training set containing the Pha to corresponding experimental values a_i^{exp}. The quality of the quantitative model can be evaluated by analysis of variance using the standard error (σ) and the F-statistic value[28]

$$\sigma^2 = 1/N \sum_{i=1}^{N} (x_i - \bar{x})^2 = \langle x^2 \rangle - \bar{x}^2,$$
$$\bar{x} = 1/N \sum_{i=1}^{N} x_i, \langle x^2 \rangle = 1/N \sum_{i=1}^{N} x_i^{\,2} \qquad (6)$$

$$F\left(\frac{\upsilon_1}{\upsilon_2}, \alpha\right) = \frac{r^2}{1 - r^2}\frac{\upsilon_2}{\upsilon_1} \qquad (7)$$

where x_i is the discrepancy of experimental and theoretical values for the activity of the ith compound, and r is the correlation coefficient. Calculated values of F can be compared to tabulated ones[28] with the confidence α and the degrees of freedom $\upsilon_1 = P$, $\upsilon_2 = N - P - 1$, where P is the number of descriptors R in eq 5, and N is the number of molecules that have the pharmacophore. However, F-statistics alone cannot be a sufficient tool in evaluation of the predictive ability of a model. Indeed, the model can work well for the training set of compounds (in other words, it may have high F-statistic and correlation coefficient values) but may have no predictive power for any independent test set. Moreover, a set of descriptors R in eq 5 may strongly depend on the choice of the ECSA and its tolerances.

After considerable trial and error we have elaborated two rules of thumb one should use in training a quantitative model to avoid chance correlations: (1) one pharmacophore should be selected from (multiple) possibilities based on (at least) leave-one-out cross-validation at the step of Pha identification to prove that removal of one compound at a time in the Pha identification does not lead to a qualitatively different pharmacophore (for instance, of different size); and (2) only those descriptors should be selected, the choice of which does not depend on the results of the Pha cross-validations.

In our work, we compare predictive abilities of alternative models (including different pharmacophores and different sets of quantitative descriptors) by comparing the values of the correlation coefficient corresponding to the correlation between theoretical predictions obtained with a cross-validated model and those obtained without such a cross-validation.

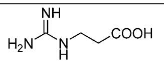
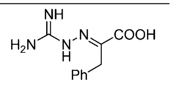
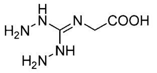
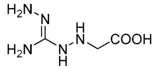
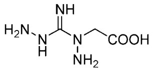
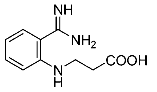
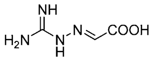
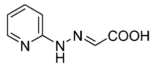
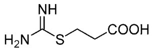
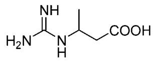
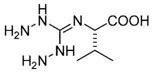
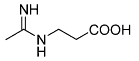
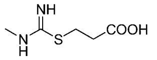
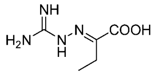
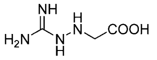
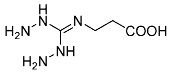
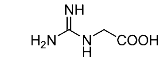
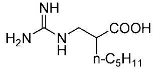
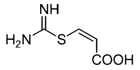
Description of Training and Test Sets. Previous experimental studies[2,3] on a series of guanidino- and aminoguanidinopropionic acid analogs contain the data on mouse insulin sensitizing screen (MISS) blood-glucose levels measured in the test (T) and control (C) groups of the KKA^{y} mice as a ratio of MISS T/C for 154 unique compounds.[2,3] These data serve as a reference database in the present theoretical study. The lower values of MISS T/C indicate the higher hypoglycemic activities. Compounds with MISS T/C ≈ 1 lack activity. After sorting these compounds in the order of increasing MISS T/C values (from more active to less active compounds), we selected every fifth compound for a test set (totally 30 compounds). We added four additional compounds to the test set. First of all, we added the two compounds numbered as 29 in ref 2 and 55 in ref 3, respectively, because their optical antipodes were already in the test set. We also added the compound with the highest activity (the aminopyridine derivative numbered 47 in ref 2) because it was suggested[2] to be acting via a mechanism of action that might be different from the one associated with guanidinopropionic acid **1** (it seems to be interesting to make an independent testing of the hypothesis). We also included its closest pyrimidine analog (numbered as 48 in ref 2) in the test set for the same reason. Thus the final test set contains 34 quite randomly selected compounds with various functionalities (22% of all data), and the training set contains the remaining 120 compounds that are sufficient to derive a model. The molecular structures of 154 compounds along with their order numbers introduced in the original references[2,3] and values of MISS T/C are presented in the Supporting Information. The Supporting Information also presents the Cartesian coordinates calculated in this work for all studied compounds including the conformers lying below 1 kcal/mol (the number of such conformers is indicated for each molecule). Note that many compounds have only one (main) conformation below 1 kcal/mol; however, the total number of all calculated conformers varies from a few conformations to one hundred.

RESULTS AND DISCUSSION

Pharmacophore Identification. Table 1 contains the molecular structures of the first 19 compounds (**1**−**19**) in the training set that possess statistically significant hypoglycemic activities (MISS T/C < 0.8) as defined in refs 2 and 3. According to the experimental results,[2,3] we define the threshold of activity as 0.8, and we use the 19 compounds for elucidation of the pharmacophore of activity in a series of 120 guanidino- and aminoguanidinopropionic acid analogs. The compounds with low activity are shown in Table 2.

Now we present the results of Pha identification. After screening the 19 active compounds (Table 1) within the initial ECSA tolerances not exceeding 25% for diagonal and 50% for off-diagonal elements we have found that the ECMC submatrix that is common for all of the active molecules contains five atoms corresponding to O_1, O_2, C_1, N_1, and N_2 in the ground conformer of compound **1** (see Figure 1). These five atoms characterize the guanidine and carboxyl functional groups. Since most of the compounds in our training set contain the carboxyl functionality and the electronic properties of the carboxylic carbon atom does not undergo any significant changes from molecule to molecule, we removed the carbon atom from the Pha. Table 3 presents the final Pha in terms of the ECSA and its tolerances. The Pha contains only four atoms that correspond to two oxygen atoms of the carboxyl group and two nitrogens of the guanidine group but may be substituted with any other atoms that have the same electronic and geometric features. A list of the atoms, which correspond to the Pha sites in the molecules that have the Pha, is given in the Supporting Information. For instance, the atoms labeled O_1, O_2, N_1, and N_2 in compound **1** are indeed replaced by S_1, N_1, O_1, and O_2, respectively, in the sulfur-containing compound **5**.

Table 1. Training Set of Guanidino- and Aminoguanidinopropionic Acid Analogs **1**−**19**[a]

N	MISS T/C	molecular structure	*N*	MISS T/C	molecular structure
1	0.52±0.24		11	0.68±0.34	
2	0.23± 0.05		12	0.69±0.22	
3	0.43±0.20		13	0.73	
4	0.52±0.24		14	0.73±0.21	
5	0.53±0.34		15	0.77	
6	0.56±0.21		16	0.77	
7	0.59±0.33		17	0.77±0.31	
8	0.61±0.29		18	0.78±0.14	
9	0.62±0.24		19	0.79	
10	0.68±0.33				

[a] Experimental data on mouse insulin sensitizing screen test/control (MISS T/C) blood glucose levels are quoted from refs 2 and 3 along with the corresponding experimental uncertainties when available. The compounds possess statistically significant antidiabetic activities (MISS T/C < 0.8) and comprise a subset of the training set of 120 compounds that was used for elucidation of the pharmacophore of activity.

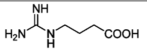
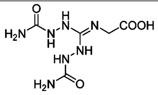
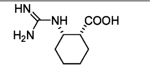
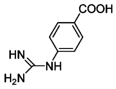
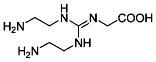
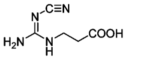
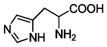
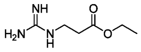
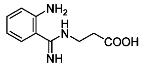
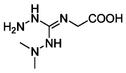
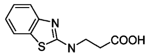
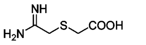
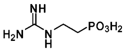
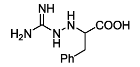
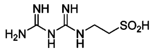
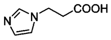
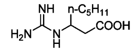
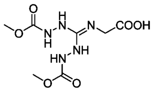
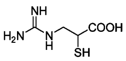
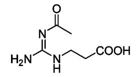
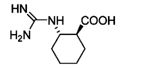
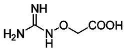
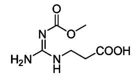
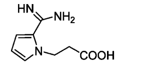
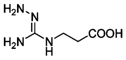
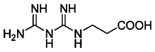
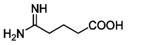
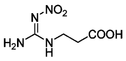
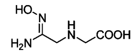
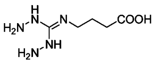
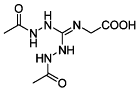
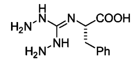
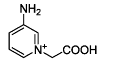
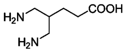
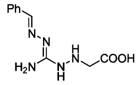
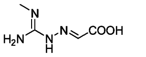
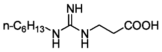
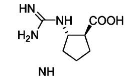
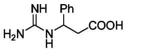
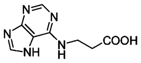
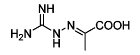
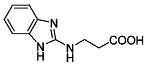
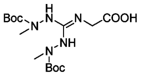
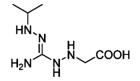
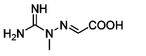
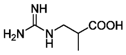
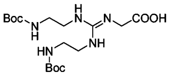
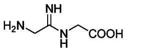
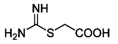
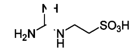
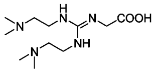
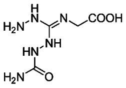
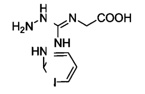
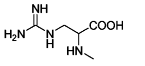
Quantitative Prediction of Activity: Pha Flexibility and Out-of-Pha Influence. There are 101 (out of 120) compounds in the training set with the Pha of activity in any possible conformer lying below 1 kcal/mol. For the compounds that have the Pha we predict values of hypoglycemic activity a using eqs 4 and 5 that can be expressed via MISS T/C as follows:

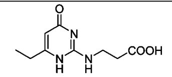
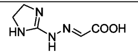
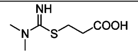
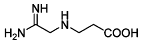
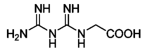
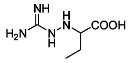
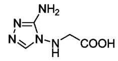
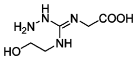
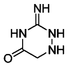
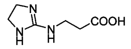
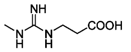
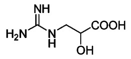
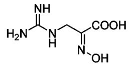
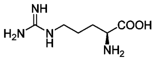
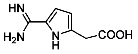
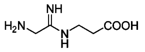
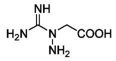
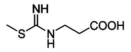
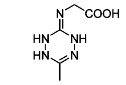
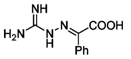
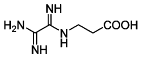
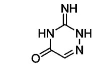
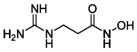
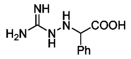
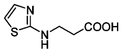
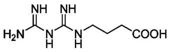
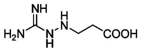
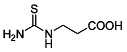
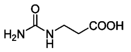
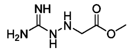
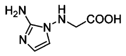
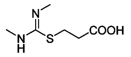
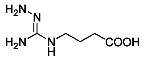
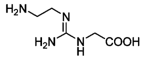
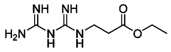
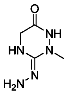
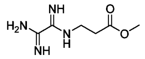
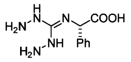
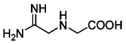
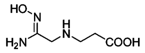
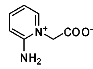
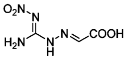
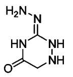
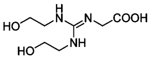
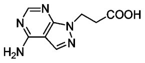
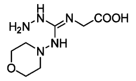
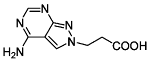
$$a = -\log \text{MISS T/C} \tag{8}$$

Values of a are always positive for potentially active compounds that can reduce blood glucose levels (MISS T/C < 1), whereas zero activity refers to compounds without the Pha (MISS $= 1$). To obtain the best possible fit of the calculated MISS T/C quantities to the corresponding experimental data, we minimize the squares of theory-experiment deviations in the MISS T/C values summed over 101 data points with the use of a nonlinear solver implemented in our Fortran program ELCOME. The 19 remaining compounds marked by an asterisk in Table 2 have no Pha (in the conformers heavily populated at room temperature), so they do not need further parametrization for quantitative predictions of activity. Their predicted activity is zero (or MISS T/C $= 1$).

By means of computer-aided alterations of possible combinations of various descriptors R in eq 4, we constructed a model composed of as few as three descriptors to describe the Pha flexibility and the influence of out-of-Pha (anti-Pha shielding) groups and other factors affecting the Pha activity. To describe the influence of out-of-Pha groups shielding the pharmacophore of activity we introduce the descriptor R_c

Table 2. Training Set of Guanidino- and Aminoguanidinopropionic Acid Analogs **20**−**120**[a]

N	molecular structure	*N*	molecular structure	*N*	molecular structure
20*		38		56	
21*		39		57	
22		40		58	
23		41		59	
24		42		60	
25*		43		61	
26		44		62	
27		45		63	
28		46		64	
29		47*		65	
30		48		66*	
31*		49		67	
32		50		68	
33		51		69	
34		52		70	
35		53		71	
36		54		72	
37		55		73	

N	molecular structure	*N*	molecular structure	*N*	molecular structure
74		90*		106	
75		91		107	
76*		92		108*	
77		93		109	
78		94*		110*	
79		95		111	
80		96		112	
81*		97*		113	
82		98		114	
83		99		115	
84*		100		116	
85		101		117*	
86		102		118	
87*		103*		119	
88*		104		120	
89		105*			

[a] Experimental data on mouse insulin sensitizing screen test/control (MISS T/C) blood glucose levels (refs 2 and 3) for compounds **20**−**120** are given in Part II of the Supporting Information. These compounds possess statistically low antidiabetic activity (MISS T/C ≥ 0.8). The asterisk refers to those compounds that do not have the pharmacophore of antidiabetic activity revealed in the present study for any conformation below 1 kcal/mol. The tertiary-butoxycarbonyl $(CH_3)_3C-O-CO$ and phenyl C_6H_5 groups were abbreviated to *Boc* and *Ph*, respectively.

that is equal to the distance from the geometric center of the pharmacophore to the most distant non-hydrogen out-of-Pha atom in a molecule. Large values of R_c correspond to relatively large (mainly hydrophobic) groups (for instance, phenylalkyl groups in **11**, **42**, **48**, **49**, **68**, **96**, **102**, and **113** or long alkyl chains with five or more carbon atoms like in **19**, **32**, and **43**) that obstruct the activity of guanidino- and aminoguanidinopropionic acid analogs.[2,3] Due to steric effects, relatively small (hydrophilic) anti-Pha shielding groups can also impede the ligand−receptor interaction: see, for instance, NH(CO)CH_3 in **30** or CH_2NH_2 in **71** and **79**. To describe the Pha flexibility we introduce the parameter a_{24} corresponding to the distance between the Pha sites noted A_2 and A_4 which is the distance between one of the nitrogen atoms in the guanidine group and one of the oxygen atom in the carboxyl group in a molecule. The a_{24} element is the largest element of the ECSA with the largest absolute tolerance. Figure 2 shows an example of the R_c and a_{24} descriptors for the ground conformer of compound **19**. The third descriptor is the solvation free energy in aqueous solution ($\Delta G_S°$) calculated by the semiempirical SM5.4P model[29] as implemented in the *Spartan* package.[22,23] The

Table 3. Pharmacophore of Antidiabetic Activity Revealed for the Series of Guanidino- and Aminoguanidinopropionic Acid Analogs **1−19**[a]

Pha site	A_1	A_2	A_3	A_4
	ECSA and Its Tolerances			
A_1	0.936 ± 0.160			
A_2	2.387 ± 0.413	1.132 ± 0.360		
A_3	3.376 ± 1.326	4.785 ± 0.996	0.988 ± 0.180	
A_4	4.492 ± 1.109	5.838 ± 1.885	2.257 ± 0.249	0.905 ± 0.195
	RMSE of the ECSA Elements			
A_1	0.0051			
A_2	0.0114	0.0067		
A_3	0.0329	0.0194	0.0055	
A_4	0.0241	0.0382	0.0080	0.0054
	RMSE of the ECSA Tolerances			
A_1	0.0048			
A_2	0.0110	0.0434		
A_3	0.0327	0.1127	0.0057	
A_4	0.0876	0.1550	0.0076	0.0059
	ECSA and Its Tolerances after 3-Fold Cross-Validation[b]			
A_1	0.935 ± 0.161			
A_2	2.397 ± 0.403	1.128 ± 0.355		
A_3	3.359 ± 1.343	4.785 ± 0.995	0.977 ± 0.168	
A_4	4.459 ± 1.077	5.795 ± 1.842	2.254 ± 0.252	0.904 ± 0.195

[a] The pharmacophore of activity, presented in terms of the electron-conformational submatrix of activity (ECSA), was revealed in the series of 19 compounds (out of the training set of 120 compounds) that possess statistically significant biological activities (MISS T/C < 0.8) as measured in refs 2 and 3. The root-mean-square errors (RMSE) of the ECSA elements and the tolerances of the elements evaluated using leave-one-out cross-validation show the variations in the ECSA and its tolerances upon removal of at least one compound at a time (out of the 19 compounds) from the Pha identification. [b] The ECSA was found after removal of compounds **1**, **2**, and **19** from the stage of Pha identification. The predicted cross-validated activities are 0.49, 0.32, and 1.00 for **1**, **2**, and **19**, respectively. They are close to the non-cross-validated activities that are respectively 0.52, 0.39, and 1.00.

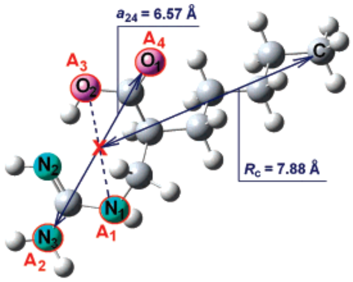


Figure 2. Examples of the anti-Pha shielding (R_c) and pharmacophore flexibility (a_{24}) descriptors in molecule **19**. The Pha sites (A_1, A_2, A_3, A_4) and the Pha center are highlighted.

solvation calculation was performed with the use of gas-phase geometries. The parameters R_c, a_{24}, and $\Delta G_S°$ are found to be linearly independent (orthogonal) by analysis of the matrix of collinearity[28] calculated in terms of correlation coefficients R_{ij}^2 with the maximum element (0.07) corresponding to the correlation between R_c and a_{24}. Numerical values of the three descriptors are listed for all molecules with the Pha in the Supporting Information.

Table 4 presents the optimal values of model coefficients k in eq 5 obtained by minimization of the squares of theory-experiment deviations in the MISS T/C values summed over 101 data points. The positive values of k_{R_c} and $k_{a_{24}}$ indicate the decrease of biological activity with the growth of R_c and a_{24}. Since the values of $\Delta G_S°$ are negative for all studied compounds and the sign of $k_{\Delta G_S°}$ is negative, the more negative values of $\Delta G_S°$ correlate with the lower activities.

Table 4. Model Coefficients k_i and Cross-Validation Statistics[a]

property of descriptor	R_c	a_{24}	$\Delta G_S°$
k_i	0.400	1.372	−0.0515
$r_{k_i}^2$ [b]	0.17	0.32	0.21
σ_{k_i} (3-fold CV)[c]	0.014	0.066	0.0019
σ_{k_i} (4-fold CV)[c]	0.017	0.077	0.0023
σ_{k_i} (5-fold CV)[c]	0.019	0.087	0.0026

[a] The quantitative model was derived using 101 compounds out of 120 compounds of the training set (Tables 1 and 2) that possess the pharmacophore of activity (Table 3) in at least one of the conformers lying below 1 kcal/mol. Dimensions of the regression coefficients k_i are reciprocal to the dimensions of corresponding descriptors: $\Delta G_S°$ is given in kcal/mol; R_c and a_{24} are in Å (see the text and Figure 3). The ground conformer of compound **1** with $R_c = 0$ Å, $a_{24} = 5.755$ Å, and $\Delta G_S° = -13.79$ kcal/mol was used as the reference compound (see eq 5). [b] The correlation coefficient obtained by removal of one descriptor out of three at a time should be compared to $R^2 = 0.41$ obtained by inclusion of all the descriptors in the model. [c] The standard errors (σ_{k_i}) of regression coefficients k_i (see eq 6) were evaluated by many-fold cross-validations (e.g., by removal of three, four, and five molecules from the set of 101 compounds). The mean values of k_i after such validations remain equal to the values of k_i obtained over the whole set of 101 compounds.

This is understandable because high affinity to water can indicate low affinity to lipophilic media that can impede the transportation of a compound across the lipophilic parts of cell membranes. Our investigation of structure−activity relationships in the series of guanidino- and aminoguanidinopropionic acid analogs indicates that $\Delta G_S°$ can be replaced (but less successfully in view of predictive ability of the model) by dipole moment. It is not surprising because the higher dipole moments correlate with more negative aqueous solvation free energies (water is one of the most polarizable media).

We found no significant dependence of hypoglycemic activity[2,3] on partition coefficients log P,[30] molecular weights, areas or volumes, electronegativity, or hardness.[23]

Statistical Analysis and Cross-Validation Using Training Set. In this section we evaluate the variability of both the pharmacophore and the quantitative model based on the Pha.

First of all, we performed a leave-one-out (LOO) cross-validation of the derived Pha by means of removal of one compound at a time from the subset of 19 compounds with statistically significant hypoglycemic activities, which were used in identification of the Pha. Every time the procedure of comparison of the EC matrices of remaining compounds was repeated from the very beginning to identify a new ECSA with a new set of tolerances. Then the new ECSA was used to construct a new quantitative model to describe the Pha flexibilities and the influence of out-of-Pha groups and other factors affecting the activity. At last, the new quantitative model was used to predict the activity of the eliminated compound. The LOO cross-validated values of the ECSA elements and their tolerances upon elimination of any of the 19 compounds are listed in the Supporting Information. Table 3 in this paper shows an excerpt of these data. The root-mean square errors of the LOO cross-validated ECSA elements and their tolerances with respect to those obtained using all of the 19 compounds do not exceed a few percent on average. Thus upon the LOO cross-validation the Pha remains qualitatively identical to that obtained without the cross-validation. We have found also that the choice of

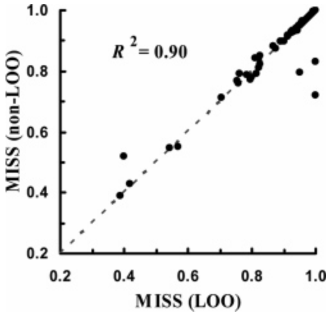


Figure 3. Theoretical values of MISS T/C (test/control blood glucose levels) for a training set of 101 compounds in the class of guanidino- and aminoguanidinopropionic acid analogs calculated with the use of leave-one-out cross-validation (LOO) and without the cross-validation (non-LOO). The LOO cross-validation includes removal of one compound at a time from the step of Pha identification as well as from the quantitative model. Only compounds with the Pha are included in the correlation.

model descriptors (R_c, a_{24}, $\Delta G_S°$) described in the previous section is not affected by these minor (quantitative) changes of the ECSA and its tolerance upon cross-validation. The LOO cross-validated values of MISS T/C calculated for each of the 19 active compounds are in good correlation with the corresponding values calculated without cross-validation with a resulting value of $R^2 = 0.85$ for correlation coefficient. Figure 3 illustrates the corresponding correlation for all 101 compounds having the Pha that yields $R^2 = 0.90$.

Table 3 also shows that there are only minor changes in the ECSA matrix elements and their tolerances upon removal of three compounds (out of 19) which are compounds with the highest activity (**2**), with the lowest activity (**19**), and with the medium activity (**1**). Of course, further decrease of a training set used in the Pha identification (especially, by removing compounds of the same class) can lead to a more significant drop of accuracy, but this is a rather trivial conclusion because the whole concept of QSAR modeling is based on the assumption that one has accurate and diversified experimental information on the activities.

Now we evaluate the quality of the model based on the derived Pha (Table 3) and the three selected descriptors (R_c, a_{24}, $\Delta G_S°$) with respect to prediction of experimental MISS T/C values for the training set of 120 guanidino- and aminoguanidinopropionic acid analogs. Theoretical values of activity versus experimental data are shown in the Supporting Information and in Figure 4. The analysis of variance performed by comparison of 120 theoretical and experimental values of MISS T/C yields $R^2 = 0.41$ (correlation coefficient), $\sigma = 0.13$ (standard error), and a value of $F(3/116,\alpha) = 27$ for F-statistics with nearly 100% of confidence α (eq 7). The root-mean-square error (RMSE) is approximately equal to the standard deviation σ because the mean error is nearly zero (see eq 6). In other words, there is no systematic overestimate or underestimate in the prediction.

The low value of R^2 is rather disappointing, but we have failed to identify any additional parameters to improve the statistics and at the same time to avoid an overfitting of the model (see ref 31 for general recommendations on building QSAR models). Note that the experimental data are in vivo, and they are far from being highly accurate. The experimental errors[2,3] for compounds with statistically significant hypoglycemic activities in the training set of 120 compounds vary from 0.05 to 0.34 with 0.25 on average. Although the errors for compounds with low activities were not evaluated in the experimental studies,[2,3] one can assume they are not lower than 0.25. Thus we still have only 8 outliers out of 120 compounds (that is less than 7%) for which the theoretical errors exceed (mostly slightly) the experimental uncertainties (see Figure 4 and the full set of data in the Supporting Information). The average experimental error (0.25) at 95% confidence corresponds to a standard error of ~0.13 if one assumes a normal distribution of the error. We have already reached this standard error by using only three descriptors of activity. Note also that compounds **86**−**120** in the training set have the MISS T/C activities that vary from 1.00 to 1.27.[2,3] These (hyperglycemic) activities cannot be predicted theoretically by the formula in eq 8 assuming that MISS T/C < 1. Excluding all compounds with experimental values of MISS T/C ≥ 1 that the model cannot predict from the analysis of variance we obtain the following statistics for the 85 remaining compounds with MISS T/C < 1: $R^2 = 0.59$, $\sigma = 0.12$, and $F(3/81, \alpha) = 39$.

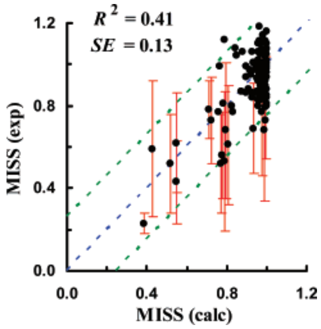


Figure 4. Theoretical (calc) versus experimental (exp) values of MISS T/C (test/control blood glucose levels) for a training set of 120 compounds in the class of guanidino- and aminoguanidinopropionic acid analogs. The experimental values of MISS T/C are quoted from refs 2 and 3 along with available experimental errors (red bars). The green lines indicate possible theory-experiment deviations corresponding to the average experimental error (±0.25). The standard error (SE) of theoretical predictions against experiment and the corresponding correlation coefficient are depicted.

To validate the quality of the quantitative prediction in this study we have also performed an M-fold cross-validation ($M = 3-5$). We have considered L new training sets obtained by exclusion of all possible combinations of M molecules out of the initial training subset of 101 compounds with the Pha ($N = 101$), where L is defined as follows:

$$L = \frac{N!}{M!(N-M)!} \qquad (9)$$

For each of the L new sets we recalculated the values of coefficients k_i corresponding to the model descriptors (R_c, a_{24}, $\Delta G_S°$) introduced in Table 4. Then we predicted activities (MISS T/C) of the M excluded compounds based on new values of the parameters. In other words, the M excluded compounds comprise a test subset. The quality of the predictions with any of the test sets is analyzed by evaluation of the root-mean-square of theory-experiment error (RMSE). For any M, the value of RMSE varies from zero to ~0.4 units of MISS T/C. The value of RMSE averaged over L M-fold cross-validations, where L is defined by eq 9, slowly and smoothly increases with M: 0.126 ($M = 3$), 0.128 ($M = 4$), and 0.130 ($M = 5$). The standard error of RMSE equals 0.04.

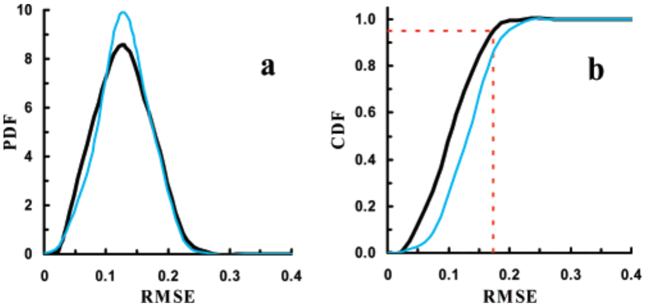


Figure 5. Probability density function (a, PDF) and cumulative distribution function (b, CDF) of the root-mean-square error (RMSE) of theoretical MISS T/C values calculated with 5-fold cross-validation (the black lines). The value of RMSE ≤ 0.17 corresponds to 95% confidence (the red dotted lines). The blue lines show the normal distribution corresponding to the mean value of RMSE equal to 0.13 and the standard deviation of RMSE equal to 0.04.

Figure 5a shows the probability density function of RMSE approximated from the corresponding discrete distribution over the ensemble of 79 208 745 5-fold cross-validations in the set of 101 compounds. This function is close to the normal distribution corresponding to the mean value of RMSE equal to 0.13 and the standard error of RMSE equal to 0.04. Figure 5b illustrates the cumulative distribution function of RMSE. Considering the case of 5-fold cross-validation, we note that activities of any five eliminated compounds are predicted with the value of RMSE equal to 0.17 or less at 95% confidence.

The mean values of regression coefficients k_i obtained over three-, four-, and 5-fold cross-validations are equal to those listed in Table 4. Standard deviations σ_{k_i} are quite acceptable; they are slightly larger for a_{24}, but this parameter is the least significant descriptor (see the values of $r^2_{k_i}$ in Table 4).

Statistical Analysis Using Test Set. Table 5 shows the molecular structures of 34 guanidino- and aminoguanidinopropionic acid analogs used as a test set along with the experimental and theoretically predicted MISS T/C values.

The test set contains 8 molecules with statistically significant experimental activities (MISS T/C < 0.8) that is 24% of the test set. For comparison, the training set used for training the model contains 19 active molecules that is 16%. Seven compounds with statistically significant experimental activities have the Pha. Therefore the EC method predicts their activity qualitatively correct. However, for three of them theory quantitatively predicts MISS T/C ≥ 0.8, but these quantitative predictions are still within the experimental error except for **122**. Compound **126** should be active with MISS T/C (exp) = 0.68 ± 0.35, but it does not have the Pha. However, the experimental error for this compound is rather large. For the 26 tested inactive compounds with MISS T/C (exp) ≥ 0.8 we have found the Pha in 22 of them, but theory still predicts their activity quantitatively correct (MISS T/C ≥ 0.8) with the exception of the following compounds: **134** (MISS T/C = 0.72 vs 0.87 in experiment), **141** (MISS T/C = 0.79 vs 0.95), **147** (MISS T/C = 0.42 vs 1.01), and **154** (MISS T/C = 0.72 vs 1.16).

The available experimental uncertainties for compounds in the test set of 34 compounds vary from 0.17 to 0.38 with 0.27 on average. Combining these errors with the errors of theoretical prediction estimated in the previous section as 0.17 with 95% confidence, we identify only three outliers, the activities of which cannot be correctly predicted: **122**, **147**, and **154**. Removing these three compounds (that is 9%) from the statistics we obtain quite acceptable statistical indices for the 31 remaining compounds in view of rather large experimental errors: $R^2 = 0.45$ and $\sigma = 0.16$ instead of $R^2 = 0.18$ and $\sigma = 0.22$ as in the case of 34 compounds. However, evaluation of the uncertainty for theoretical prediction (see the next section) using the formulas of activity derived in the present study should be done with a less optimistic value of $\sigma = 0.22$ corresponding to the 95% confidence interval equal to 0.43 if one assumes a normal distribution of the error.

The hypoglycemic activity of compound **122** and its optical antipode **138** strongly depends on chirality. We neglected the chirality in the present study (see ref 10 for treatment of the chirality influence by the EC method). However, theory predicts a rather correct value of MISS T/C = 0.96 for compound **138**. The theoretical model cannot distinguish the activities of compounds **121** and **147** (the latter structure differs only by an N-substitution in the ring), but compound **121** is correctly predicted active. The aminopyridine analog **121** demonstrates a strong hypoglycemic activity in experiment.[2] However, the authors[2] suggested that the mechanism of activity of compound **121** might be different from that of **1** because of the gross toxicity of **121**. Since the EC method gives a rather correct prediction of the experimental value of MISS T/C (experiment: 0.20 ± 0.17;[2] theory: 0.43), the mechanisms for **1** and **121** are expected to be the same.

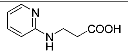
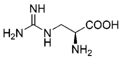
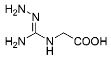
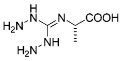
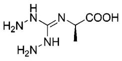
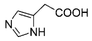
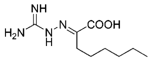
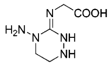
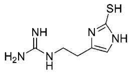
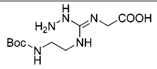
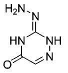
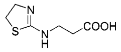
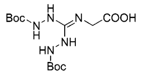
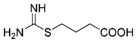
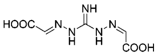
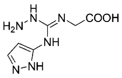
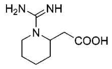
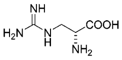
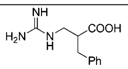
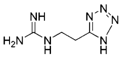
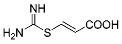
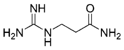
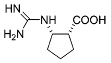
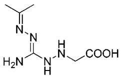
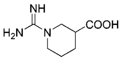
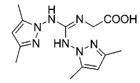
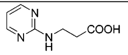
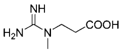
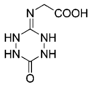
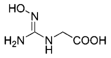
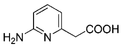
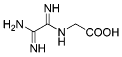
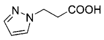
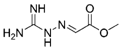
Screening for New Compounds. In possession of a theoretical formula for the prediction of hypoglycemic activity (Tables 3 and 4, eq 5), we are able to find new potentially active compounds in the class of guanidinopropionic acid analogs or in any other class of molecules that can be assumed to act via the same mechanism.

The analysis of the Pha shows that the presence of the carboxyl group is a key component of activity within the specifics of biochemical mechanism for derivatives of **1**. Any substitutions on the carboxyl functionality (which is apparently significant also for the solubility and extended hydrophilicity) obstruct the antidiabetic potency. Compounds with lipophilic substitutions are predicted to be devoid of activity. Loss of basicity of the compounds substituted with electron-withdrawing groups is correlated to loss of activity. On the other hand, modification on the guanidine functionality does not necessarily result in loss of activity (see for instance, compound **121**).

Any modification in **1** and its derivatives may lead to disappearance of the Pha due to the redistribution of the electronic density and changes in interaction indices, bond orders, and interatomic distances in the ECSA. In general, only quantum-chemical calculations of the electronic and geometric structure of the new molecule allow one to predict the activity using the EC method. Nevertheless, the analysis of the absolute values and the sign of model coefficients k (Table 4) can be useful for some preliminary predictions.

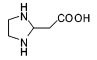
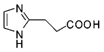
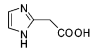
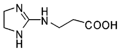
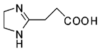
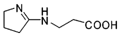
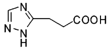
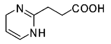
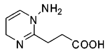
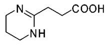
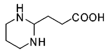
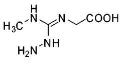
We screened a few dozen new species related to compound **1**. The electron-conformational matrices for these compounds were calculated and compared to the electron-conformational submatrix of activity obtained above (Table 3). Table 6 presents the molecular structures of compounds **155**−**166** that have the Pha (their ECMCs contain the ECSA) and may

Table 5. Test Set of Guanidino- and Aminoguanidinopropionic Acid Analogs **121−154**[a]

N	MISS T/C exp	MISS T/C calc	molecular structure
121	0.20±0.17	0.43	
122	0.35±0.20	0.96	
123	0.51±0.24	0.54	
124	0.57±0.28	0.30	
125	0.66±0.29	0.30	
126*	0.68±0.35	1.00	
127	0.73±0.27	0.99	
128	0.77±0.38	0.89	
129	0.80	0.93	
130	0.81	0.96	
131*	0.84	1.00	
132	0.85	0.99	
133	0.86	0.99	
134	0.87	0.72	
135	0.88	1.00	
136*	0.89	1.00	
137	0.91	0.99	
138	0.92	0.96	
139	0.93	1.00	
140*	0.94	1.00	
141	0.95	0.79	
142	0.96	0.82	
143	0.97	0.99	
144	0.98	0.98	
145	1.00	0.97	
146	1.00	0.98	
147	1.01	0.42	
148	1.02	0.89	
149	1.04	0.92	
150	1.06	0.81	
151	1.08	0.94	
152	1.08	0.97	
153*	1.11	1.00	
154	1.16	0.72	

[a] These 34 compounds comprise the test set used for testing the quantitative model of activity derived with the use of 120 compounds of the training set (Tables 1 and 2). Experimental data on mouse insulin sensitizing screen test/control MISS T/C (exp) blood glucose levels are quoted from refs 2 and 3 and given along with the corresponding theoretical values MISS T/C (calc) calculated with the use of data in Tables 3 and 4. The asterisk refers to those compounds that do not have the pharmacophore of antidiabetic activity revealed in the present study for any conformation below 1 kcal/mol. The tertiary-butoxycarbonyl $(CH_3)_3C-O-CO$ and phenyl C_6H_5 groups were abbreviated to *Boc* and *Ph*, respectively.

Table 6. Molecular Structures of Potential Antidiabetic Agents **155−166** Predicted by the Electron-Conformational Method[a]

N	molecular structure	MISS T/C
155		0.04
156		0.30
157		0.03
158		0.44
159		0.15
160		0.44
161		0.35
162		0.45
163		0.02
164		0.12
165		0.06
166		0.51

[a] Mouse insulin sensitizing screen test/control (MISS T/C) blood glucose levels are calculated with the use of data in Tables 3 and 4.

be potential antidiabetic agents. According to the EC method, these compounds are expected to demonstrate statistically significant hypoglycemic activity under experimental conditions and within the experimental error in the studies.[2,3] The 95% confidence interval for these predictions is equal to ±0.43. It was estimated from the standard error of $\sigma = 0.22$ obtained from the analysis of theoretical predictions over the test set of 34 compounds with known experimental activities using the derived pharmacophore of activity and three-parameter quantitative model (Table 5).

Compound **166** in Table 6 is an acyclic analog of **1** obtained with an amino group substitution in the α-position, whereas compounds **155−161** belong to the class of imidazole derivatives. Compounds **162−165** are pyrimidine derivatives. The analysis of pharmacokinetic properties of **155−166** and aspects of toxicity in vivo as well as ways of preparation of these compounds is beyond the scope of the present study. We refer to synthesis procedures for various analogs of **1** described in the literature.[2,3] Compound **156** is available commercially.

We performed screening of other groups of potential antidiabetic agents[1,32−36] out of the class of **1** including oxyiminoalkanoic acid derivatives[33] and thiazolidinediones analogs,[34] both supposedly targeting the peroxisome proliferator-activated receptors in insulin sensitive tissues such as adipocytes (fat-storing cells),[34] fluorinated pyrazol and pyrazolone analogs regulating renal glucose reabsorption,[35] imidazoline derivatives of piperazine improving insulin secretion,[36] and biguanides (metformine, for instance) that act mainly by reducing gluconeogenesis and stimulating peripheral glycolysis.[1] None of these compounds has the Pha revealed for the class of guanidino- and aminoguanidinopropionic acid compounds **1−154** and predicted for a new series of potential hypoglycemic agents **155−166**. Therefore, the mechanism of biochemical action of analogs of **1** in the rodent model of non-insulin-dependent diabetes mellitus[2,3] apparently differs from those suggested for other classes of compounds with significant antidiabetic potency.[1,32−36]

FURTHER DISCUSSION

The EC method was used for in silico screening of new compounds never tested experimentally and for prediction of statistically significant hypoglycemic activities for a few of them (**155−166**). However, the estimated error of these predictions (±0.43 at 95% confidence) is rather large. One of the reasons of that is the large uncertainties in the in vivo data (**1−154**) used for training and testing the present model. Indeed, the experimental errors[2,3] for compounds with statistically significant hypoglycemic activities vary from 0.05 for compound **1** to 0.38 for **128** with 0.26 on average.

The remaining discrepancy can be related to a variety of possible theoretical inaccuracies due to approximations used in the calculations as well as lack of explicit account for toxicity, solubility, membrane permeability, stability in the biological fluids, pharmacokinetic properties, etc. However, we included some of these factors in the model implicitly to some extent using solvation free energies and parameters describing the Pha-flexibility and the influence of out-of-Pha groups in a molecule. On the other hand, the semiempirical SM5.4P model[29] used in solvation calculations is not as accurate as the newest (ab initio) model of the same series.[37] Interaction indices are calculated by eq 1 with the use of Mulliken partial atomic charges. There can be some discrepancy related to the sensitivity of the charges from population analysis to basis set size.[38] Table S4 in the Supporting Information (Part II) lists Milliken atomic charges calculated for the ground molecular conformer of compound **82** as an example. Indeed, depending on the choice of basis set or electronic structure method the charges vary within up to 47% on average. Interestingly, the corresponding deviations in interaction indices that use the same Mulliken charges are twice smaller (only 23%).

The EC method described in refs 8−10 and used in this work is related to the class of ligand-based approaches in drug design. Considering properties of ligand rather than its biological target (biological receptor), such methods are especially useful in prediction of activity when the structure of the bioreceptor is unknown. The main disadvantage of these methods is the need for large arrays of experimental data on activities of the substrate (ligand) compounds, which should have diverse molecular structures but interact with the bioreceptor via the same mechanism. Besides, in the older electron-topological version of the method,[11] still in use by some authors,[15−17] there is no quantitative evaluation of the activity based on the out-of-Pha influence, the multiconformation problem is ignored, and Mulliken charges are used as diagonal elements of ECMCs instead of interaction indices. As part of the present study, we tested Mulliken charges as diagonal elements in the ECMC, and we failed to describe the experimental activities of sulfur-containing compounds qualitatively well. For instance, compounds **5**, **7**, and **10** do not have the Pha obtained in other active compounds if Mulliken charges are used probably because the charges do not properly describe the electron-donor properties of an atom in a molecule (for instance, sulfur atoms). We do not recommend the partial atomic charges for diagonal elements.

In an attempt to improve the older version[11] the authors[17] try to take into account the antipharmacophore shielding (first introduced in ref 8) by introducing rather arbitrary *multiple* pharmacophores and antipharmacophores that appear in active and inactive compounds with different probabilities correlated statistically to experimentally observed activities. The EC method[8−10] used in the present study describes the properties of a molecule in its interaction with the bioreceptor by means of the *single,* uniquely defined ECSA (pharmacophore) that provides the description of activity without employing statistical descriptors. In other words, in order to identify the pharmacophore of activity we calculate only electronic structure and geometry which are inherent and most general characteristics of the molecule (within the accuracy of theoretical approximations and experimental data used).

The active "skeletons" of those molecules that bind to the same bioreceptor are expected to be similar to each other with respect to their electronic structure and geometry (described in the EC method by ECSA). Such similar "contact skeletons" may belong to rather different classes of chemical compounds (in the case of musk odorants there are tens of different classes with the same active skeleton),[39] but it is obvious that even minor substitutions in the functionality of a given ligand (from the same class of compounds) may change its activity. All the compounds studied in the present work belong to the same class of guanidino- and aminoguanidinopropionic acid analogs, but they are diverse enough to be used for appropriate training of both qualitative and quantitative models of activity, taking into account a variety of substituted positions and substituents like amino, imino, alkyl, phenyl, acetyl, and tertiary-butoxycarbonyl groups and various heterocyclic species. The strong dependence of the activity on the Pha surroundings and other factors, confirmed also in this work, shows that the qualitative model of Pha identification employed in many QSAR approaches as well as in the older version of the EC method[11,15−17] may not be sufficient for any full description of the biological activity; the parametrized description in the second part of the EC method provides for the necessary quantitative analysis and prediction of the activity.

CONCLUSIONS

The electron-conformational method applied to describe the antidiabetic activity in the training set of guanidinopropionic acid **1** and its analogs (**1**−**120**) reveals that this activity is controlled by a pharmacophore that consists of four sites with certain electronic and topological characteristics which in these compounds are occupied by two oxygen atoms from the carboxyl group and by the two nitrogen atoms of the guanidine functional group but may be substituted with any other atoms that have approximately the same electronic structure parameters and geometric positions within the derived tolerances. The quantitative model of activity based on the derived pharmacophore contains only three descriptors. One of these parameters is a distance from the geometric center of the pharmacophore to the most distant non-hydrogen out-of-Pha atom in a molecule. It describes anti-Pha shielding functional groups. Another parameter describes the Pha flexibility as a distance between one of the nitrogen atoms in the guanidine group and one of the oxygen atoms in the carboxyl group. The third parameter is the solvation free energy of a ligand in aqueous solution calculated by the SM5.4P continuum solvation model. It accounts for the influence of other factors (outside the Pha) affecting the activity.

The pharmacophore selected from multiple possibilities has been validated by removing one or a few compounds at a time from the step of Pha identification. The choice of the three model descriptors tolerates these cross-validations. The derived model has been tested using the test set of 34 analogs (**121**−**154**) with known experimental activities containing various chemical functionalities. Flexibility of the Pha with respect to the guanidine group allowed us to design (never tested experimentally) modified analogs of guanidinopropionic acid **1** and to predict theoretically the novel molecular structures **155−166** with statistically significant antidiabetic potency as perspective candidates for future experimental studies.

ACKNOWLEDGMENT

Acknowledgments are made to the Welch Foundation (grant F−100) and to the Dreyfus Foundation (grant SI−05-011) for support of this research. We also appreciate valuable technical assistance from Wenli Zou (the University of Texas at Austin).

Supporting Information Available: Low-energy (<1 kcal/mol) conformers of guanidino- and aminoguanidinopropionic acid analogs **1−166**, PM3 optimized Cartesian coordinates and corresponding relative and absolute total energies calculated at the MP2(frozen core)/6-31G(d) level of theory (Part I), molecular structure formulas, experimental and theoretical values of MISS T/C for the studied compounds, values of molecular descriptors for prediction of antidiabetic activities, electron-conformational submatrices of activity and antidiabetic activities predicted by leave-one-out cross-validation, and interaction indices versus Mulliken partial atomic charges in compound **82** (Part II). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Karam, J. H. In *Basic & Clinical Pharmacology*; Katzung, B. G., Ed.; Appleton & Lange: Stamford, CT, 1998; pp 684−705.

(2) Larsen, S. D.; Connell, M. A.; Cudahy, M. M.; Evans, B. R.; May, P. D.; Meglasson, M. D.; O'Sullivan, T. J.; Schostarez, H. J.; Sih, J. C.; Stevens, F. C.; Tanis, S. P.; Tegley, C. M.; Tucker, J. A.; Vaillancourt, V. A.; Vidmar, T. J.; Watt, W.; Yu, J. H. Synthesis and Biological Activity of Analogues of the Antidiabetic/Antiobesity Agent 3-Guanidinopropionic Acid: Discovery of a Novel Aminoguanidinoacetic Acid Antidiabetic Agent. *J. Med. Chem.* **2001**, *44*, 1217−1230.

(3) Vaillancourt, V. A.; Larsen, S. D.; Tanis, S. P.; Burr, J. E.; Connell, M. A.; Cudahy, M. M.; Evans, B. R.; Fisher, P. V.; May, P. D.; Meglasson, M. D.; Robinson, D. D.; Stevens, F. C.; Tucker, J. A.; Vidmar, T. J.; Yu, J. H. Synthesis and Biological Activity of Aminoguanidine and Diaminoguanidine Analogues of the Antidiabetic/Antiobesity Agent 3-Guanidinopropionic Acid. *J. Med. Chem.* **2001**, *44*, 1231−1248.

(4) *Pharmacophore Perception, Development and Use in Drug Design*; Güner, O. F., Ed.; International University Line: La Jolla, CA, 2000.

(5) Venkatarangan, P.; Hopfinger, A. J. Prediction of Ligand−Receptor Binding Thermodynamics by Free Energy Force Field Three-Dimensional Quantitative Structure−Activity Relationship Analysis: Applications to a Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase. *J. Med. Chem.* **1999**, *42*, 2169−2179.

(6) Liao, C.; Xie, A.; Shi, L.; Zhou, J.; Lu, X. Eigenvalue Analysis of Peroxisome Proliferator-Activated Receptor γ Agonists. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 230−238.

(7) Calabuig, C.; Antón-Fos, G. M.; Gálvez, J.; García-Doménech, R. New Hypoglycaemic Agents Selected by Molecular Topology. *Int. J. Pharm.* **2004**, *278*, 111−118.

(8) Bersuker, I. B.; Bahceci, S.; Boggs, J. E.; Pearlman, R. S. An Electron-Conformational Method of Identification of Pharmacophore and Anti-Pharmacophore Shielding: Application to Rice Blast Activity. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 419−434.

(9) Bersuker, I. B.; Bahceci, S.; Boggs, J. E. Improved Electron-Conformational Method of Pharmacophore Identification and Bioactivity Prediction. Application to Angiotensin Converting Enzyme Inhibitors. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1363−1376.

(10) Bersuker, I. B. Pharmacophore Identification and Quantitative Bioactivity Prediction Using the Electron-Conformational Method. *Curr. Pharm. Des.* **2003**, *9*, 1575−1606.

(11) Bersuker, I. B.; Dimoglo, A. S. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, pp 423−460.

(12) Rosines, E.; Bersuker, I. B.; Boggs, J. E. Pharmacophore Identification and Bioactivity Prediction for Group I Metabotropic Glutamate Receptor Agonists by the Electron-Conformational QSAR Method. *Quant. Struct.-Act. Relat.* **2001**, *20*, 327−334.

(13) Makkouk, A. H.; Bersuker, I. B.; Boggs, J. E. Quantitative Drug Activity Prediction for Inhibitors of Human Breast Carcinoma. *Int. J. Pharm. Med.* **2004**, *18*, 81−89.

(14) Liu, S.-S.; Yin, C.-S.; Li, Z.-L.; Cai, S.-X. QSAR Study of Steroid Benchmark and Dipeptides Based on MEDV-13. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 321−329.

(15) Guzel, Y.; Ozturk, E. Study of the Binding Affinity for Corticosteroid-Binding Globulin (CBG) Using the Electron Topological Method (ETM) as Three-Dimensional Quantitative Structure−Activity Relationship (3D QSAR). *Bioorg. Med. Chem.* **2003**, *11*, 4383−4388.

(16) Altun, A.; Golcuk, K.; Kumru, M.; Jalbout, A. F. Electron-Conformational Study for the Structure−Hallucinogenic Activity Relationships of Phenylalkylamines. *Bioorg. Med. Chem.* **2003**, *11*, 3861−3868.

(17) Oruc, E. E.; Rollas, S.; Kandemirli, F.; Shvets, N.; Dimoglo, A. S. 1,3,4-Thiadiazole Derivatives. Synthesis, Structure Elucidation, and Structure−Antituberculosis Activity Relationship Investigation. *J. Med. Chem.* **2004**, *47*, 6760−6767.

(18) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(19) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods. I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.

(20) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.

(21) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(22) *Spartan'02* is available from Wavefunction, Inc., Irvine, CA, U.S.A. Except for molecular mechanics and semiempirical models, the calculation methods used in Spartan'02 have been documented in the following: Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Voorhis, T. V.; Oumi, M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. Q-Chem 2.0: a High-Performance Ab Initio Electronic Structure Program Package. *J. Comput. Chem.* **2000**, *21*, 1532−1548.

(23) Hehre, W. J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*; Wavefunction, Inc.: Irvine, CA, 2003.

(24) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. *J. Chem. Phys.* **1955**, *23*, 1833−1840.

(25) Pauling, L. *The Nature of The Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, New York, 1960.

(26) Basch, H.; Viste, A.; Gray, H. B. Valence Orbital Ionization Potentials from Atomic Spectral Data. *Theor. Chim. Acta* **1965**, *3*, 458−464.

(27) Bondi, A. Van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(28) Livingstone, D. *Data Analysis for Chemists: Applications to QSAR and Chemical Product Design*; Oxford University Press: New York, 1995.

(29) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for Aqueous Solvation Based on Class IV Atomic Charges and First Solvation Shell Effects. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(30) Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships III: Modeling Hydrophobic Interactions. *J. Comput. Chem.* **1988**, *9*, 80−90.

(31) Livingstone, D. J. In *Predicting Chemical Toxicity and Fate*; Cronin, M. T. D., Livingstone, D. J., Eds.; CRC Press: Boca Raton, FL, 2004; Chapter 7, pp 151−170.

(32) Lohray, B. B.; Bhushan, V. Advances in Insulin Sensitizers. *Curr. Med. Chem.* **2004**, *11*, 2467−2503.

(33) Imoto, H.; Sugiyama, Y.; Kimura, H.; Momose, Y. Studies on Non-Thiazolidinedione Antidiabetic Agents. 2. Novel Oxyiminoalkanoic Acid Derivatives as Potent Glucose and Lipid Lowering Agents. *Chem. Pharm. Bull.* **2003**, *51*, 138−151.

(34) Desai, R. C.; Han, W.; Metzger, E. J.; Bergman, J. P.; Gratale, D. F.; MacNaul, K. L.; Berger, J. P.; Doebber, T. W.; Leung, K.; Moller, D. E.; Heck, J. V.; Sahoo, S. P. 5-Aryl Thiazolidine-2,4-diones: Discovery of PPAR Dual α/γ Agonists as Antidiabetic Agents. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2795−2798.

(35) Kees, K. L.; Fitzgerald, J. J., Jr.; Steiner, K. E.; Mattes, J. F.; Mihan, B.; Tosi, T.; Mondoro, D.; McCaleb, M. L. New Potent Antihyperglycemic Agents in db/db Mice: Synthesis and Structure−Activity Relationship Studies of (4-Substituted benzyl)(trifluoromethyl)pyrazoles and -pyrazolones. *J. Med. Chem.* **1996**, *39*, 3920−3928.

(36) Bihan, G. L.; Rondu, F.; Pelé-Tounian, A.; Wang, X.; Lidy, S.; Touboul, E.; Lamouri, A.; Dive, G.; Huet, J.; Pfeiffer, B.; Renard, P.; Guardiola-Lemaître, B.; Manéchez, D.; Pénicaud, L.; Ktorza, A.; Godfroid, J.-J. Design and Synthesis of Imidazoline Derivatives Active

on Glucose Homeostasis in a Rat Model of Type II Diabetes. 2. Syntheses and Biological Activities of 1,4-Dialkyl-, 1,4-Dibenzyl, and 1-Benzyl-4-alkyl-2-(4′,5′-dihydro-1′*H*-imidazole-2′-yl)piperazines and Isosteric Analogues of Imidazoline. *J. Med. Chem.* **1999**, *42*, 1587−1603.

(37) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Non-aqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011−2033.

(38) Thompson, J. D.; Xidos, J. D.; Sonbuchner, T. M.; Cramer, C. J.; Truhlar, D. G. More Reliable Partial Atomic Charges When Using Diffuse Basis Sets. *PhysChemComm* **2002**, *5*, 117−134.

(39) Bersuker, I. B.; Dimoglo, A. S.; Gorbachov, M. Yu.; Vlad, P. F.; Pesaro, M. Origin of Musk Fragrance Activity: the Electron-Topologic Approach. *New J. Chem.* **1991**, *15*, 307−320.

CI700401P