

Quantitative structure and aldose reductase inhibitory activity relationship of 1,2,3,4-tetrahydropyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone derivatives

Kwangseok Ko and Youngdo Won*

Department of Chemistry, Hanyang University, Seoul 133-792, Korea

Received 18 November 2004; revised 20 December 2004; accepted 20 December 2004

Abstract—We investigate the quantitative structure–activity relationship of spirosuccinimide-fused tetrahydropyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives acting as aldose reductase inhibitors, which contain a chiral center. The published assay data of 30 training compounds are not for optically pure enantiomer preparations but for racemic mixtures. As the physicochemical descriptors for the QSAR analysis must be evaluated for either (*R*)-enantiomer or (*S*)-enantiomer, we devise a new ‘racemic’ descriptor as the arithmetic mean of the (*R*)-enantiomer descriptor and the (*S*)-enantiomer descriptor. The resultant QSAR model derived from the racemic descriptors outperforms the original QSAR models. The racemic QSAR model shows that the hydrophobic character of the benzyl moiety is the major contributing factor to the aldose reductase inhibitory activity and the polar surface area descriptors modulate the inhibitory activity.

© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

Aldose reductase is the first enzyme of the polyol pathway and catalyzes the conversion of D-glucose to D-sorbitol with NADPH.^{1,2} While the majority of glucose is phosphorylated by hexokinase under normal glycemic conditions, glucose flux through the polyol pathway is markedly increased under chronic hyperglycemia, which increases the intracellular sorbitol concentration. The high accumulation of sorbitol causes diabetic complications such as retinopathy, neuropathy, nephropathy, and cataracts.^{3,4} Therefore, the catalyst of the first step in the polyol pathway would be the effective therapeutic target for treatment of long-term diabetic complications.^{1–4} A great deal of effort has focused on the discovery of compounds that selectively inhibit the aldose reductase enzyme.

While a large number of structurally diverse aldose reductase inhibitors were synthesized,^{5–15} only a small part of these compounds were tested in clinical trials

on diabetic patients. Tolrestat, zopolrestat, and zenarestat were withdrawn from trials due to either low efficacy or possible toxic side effects.^{16–18} There is a strong need to develop aldose reductase inhibitors effective enough to meet the standard of the FDA. The quantitative structure activity relationship (QSAR) of a series of compounds with known activity should aid the lead optimization to enhance potency and deliverability. We investigated QSAR of 1,2,3,4-tetrahydropyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone derivatives and aldose reductase inhibitory activity.

While the spirosuccinimide congeners contain the chiral center, racemic mixtures are normally used in the aldose reductase inhibition assay. It is reported that the aldose reductase has high enantioselectivity for spirosuccinimides^{19,20} and the (*R*)-enantiomer is about 10 times more potent than the (*S*)-enantiomer in vitro.²¹ It is a challenging question, which molecular structure to employ in evaluating the QSAR descriptors. Here, we consider a simple arithmetic mean of enantiomeric descriptors as the QSAR basis and demonstrate its applicability to the QSAR analysis of spirosuccinimide aldose reductase inhibitors.

We derive QSAR models using the (*R*)-enantiomer structure and the (*S*)-enantiomer structure, separately,

Keywords: QSAR; Aldose reductase inhibitor; Genetic function approximation; Physicochemical descriptor.

* Corresponding author. Tel.: +82 2 2290 0944; fax: +82 2 2290 0762; e-mail: won@hanyang.ac.kr

and use the arithmetic mean of descriptors derived from each enantiomer structure to obtain a new racemic QSAR model. The behavior of QSAR models is examined with a variety of statistical measures and the contribution of each substituent group is separately analyzed. The racemic QSAR equation predicts the inhibitory activity of both (*R*)-enantiomer and (*S*)-enantiomer as well as the racemate better than the original QSAR models.

2. Methods

2.1. The biological activity data

Negoro et al. synthesized pyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives and performed the aldose reductase inhibition assay.²¹ The activity data of the series of compounds containing spirosuccinimide are collected in Table 1. The IC₅₀ value is the concentration of the compound required for 50% inhibition of Porcine lens aldose reductase activity. The negative logarithm of IC₅₀ (pIC₅₀) is used as the dependent variable of the QSAR equation.

All 30 spirosuccinimide congeners have the chiral center marked by '*' in the molecular formula of Table 1. However, the compounds were not prepared in optically pure enantiomeric form. The assay data were obtained for racemates.

2.2. Molecular model building

We utilized the Cerius² program package²² and built the (*R*)-enantiomer molecular structures of pyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone derivatives. The molecular structure of each compound was optimized by thorough conformational search and energy minimization. A set of conformers of a given compound was generated and the energy was minimized until the root mean square of the energy gradient reached 0.001 kcal/mol Å. We evaluated the molecular energy and forces acting on each atom using the Merck Molecular Force Field. The adopted basis Newton–Raphson minimization method was employed in the geometry optimization procedure. The molecular structure with the lowest energy of compound **24** was close to that of AS-3201 docked in the active site of AR.²³ The lowest energy structure of (*R*)-enantiomers was the reasonable choice for the QSAR analyses.

The flexible docking experiment also showed that the pyrrolo[1,2-*a*]pyrazine and the benzyl moieties of the (*S*)-enantiomer of compound **24** (SX-3202) took the similar conformation with that of the (*R*)-enantiomers while placing the succinimide ring in the opposite orientation.²³ The lowest energy (*R*)-enantiomer structure was used to build the molecular structure of the (*S*)-enantiomer by rotating the succinimide ring by 180°. The (*S*)-enantiomer structure was optimized through the same minimization protocol. The (*S*)-enantiomer compound **24** was optimized to the structure with energy of −92.942 kcal/mol. The global minimum was located at −93.046, 0.104 kcal/mol lower in energy. Other

compounds were optimized to the structure with energy no more than 1 kcal/mol (less than 1%) higher than that of the global minimum. The small energy can be easily overcome to fit into the active site of AR.

We selected the pyrrolo[1,2-*a*]pyrazine-1,3-dione moiety of the most active compound **22** to be the reference frame and set the Cartesian coordinate system as shown in Figure 1. The pyrrolo[1,2-*a*]pyrazine ring moiety was placed in the *yz*-plane and the bridging CN bond of the five-member aromatic ring was aligned to the *z*-axis. We superimposed the optimized molecular structure of each compound to the reference structure of compound **22**. The rigid fit method of the superposition procedure minimized the root mean square deviations of common atoms with respect to the reference structure. Figure 1 shows aligned molecular structures of 30 compounds, (*R*)-enantiomers at the left and (*S*)-enantiomers at the right.

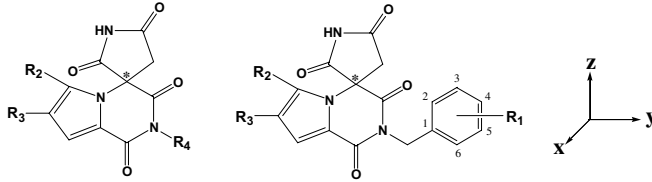
2.3. Molecular descriptors

We utilized the QSAR⁺ module of the Cerius² package to generate molecular descriptors for each enantiomer of the compounds in Table 1. The QSAR⁺ program offers 78 physicochemical descriptors, which include conformational, electronic, spatial, structural, thermodynamic, quantum mechanical, and molecular shape analysis descriptors. We obtained two sets of 78 molecular descriptors, one for (*S*)-enantiomer compounds and the other for (*R*)-enantiomer compounds called *S*-descriptors and *R*-descriptors, respectively. We took the respective arithmetic mean values of *R*-descriptors and *S*-descriptors to generate another set of descriptors, called *RS*-descriptors.

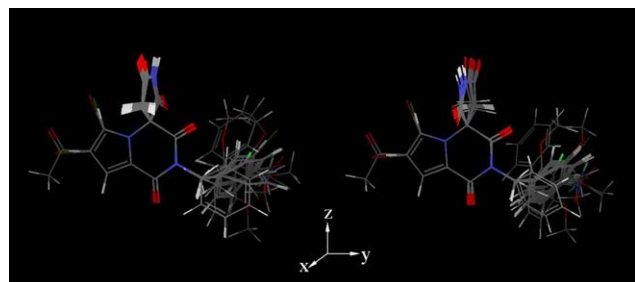
A spectrum of molecular variety is needed in deriving a useful QSAR equation, which requires descriptors evenly distributed in value. Descriptors with the standard deviation less than that of the activity, pIC₅₀, were removed from the set. The filtering procedure left 48 descriptors.

The descriptors of the resultant QSAR model are as follows:

- Alog P* the logarithm of the partition coefficient;
- V*_{NCOS} the non-common overlap steric volume;
- μ*_{*x*}, *μ*_{*y*}, *μ*_{*z*} the *x*, *y*, *z* components of the dipole moment;
- α*_{pol} the sum of the atomic polarizabilities;²⁴
- TASA the total hydrophobic surface area: sum of solvent-accessible surface areas of atoms with absolute value of partial charges less than 0.2;²⁵
- TPSA the total polar surface area: sum of solvent-accessible surface areas of atoms with absolute value of partial charges greater or equal than 0.2;
- PNSA₁ the partial negative surface area: sum of the solvent-accessible surface areas of all negatively charged atoms;
- PPSA₁ the partial positive surface area: sum of the solvent-accessible surface areas of all positively charged atoms;

Table 1. Predicted and observed aldose reductase inhibitory activity data of 2,6,7-substituted-1,2,3,4-tetrahydropyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone^a


Compd	R ₁	R ₂	R ₃	R ₄	pIC ₅₀ ^b			
					Obsd	R-Eq.	S-Eq.	RS-Eq.
1	–	H	H	H	0.201	0.430	0.343	0.330
2	–	H	H	CH ₃ –	0.553	0.489	0.426	0.365
3	–	H	H	C ₆ H ₅ –	0.699	0.581	0.531	0.611
4	–	H	H	C ₆ H ₅ –(CH ₂) ₂ –	0.569	0.591	0.801	0.584
5	H	H	H	C ₆ H ₄ –CH ₂ –	1.004	1.018	1.267	1.203
6	2-F	H	H	C ₆ H ₄ –CH ₂ –	1.215	1.230	1.170	1.259
7	4-F	H	H	C ₆ H ₄ –CH ₂ –	0.921	1.050	1.090	1.209
8	2-Cl	H	H	C ₆ H ₄ –CH ₂ –	1.432	1.140	1.348	1.188
9	3-Cl	H	H	C ₆ H ₄ –CH ₂ –	1.456	1.472	1.481	1.519
10	4-Cl	H	H	C ₆ H ₄ –CH ₂ –	1.420	1.216	1.373	1.308
11	2-Br	H	H	C ₆ H ₄ –CH ₂ –	1.284	1.134	1.184	1.193
12	3-Br	H	H	C ₆ H ₄ –CH ₂ –	1.432	1.540	1.345	1.475
13	4-Br	H	H	C ₆ H ₄ –CH ₂ –	1.328	1.319	1.245	1.376
14	4-CH ₃	H	H	C ₆ H ₄ –CH ₂ –	1.284	1.278	1.317	1.149
15	4-OCH ₃	H	H	C ₆ H ₄ –CH ₂ –	1.337	1.248	1.107	1.159
16	4-CF ₃	H	H	C ₆ H ₄ –CH ₂ –	1.097	0.996	1.104	1.116
17	4-NO ₂	H	H	C ₆ H ₄ –CH ₂ –	0.959	0.897	1.143	0.994
18	4-NH ₂	H	H	C ₆ H ₄ –CH ₂ –	0.854	0.790	0.961	0.786
19	2,4-(OCH ₃) ₂	H	H	C ₆ H ₃ –CH ₂ –	0.456	0.492	0.492	0.578
20	3,4-(OCH ₃) ₂	H	H	C ₆ H ₃ –CH ₂ –	0.959	0.915	0.900	1.000
21	2,4-F ₂	H	H	C ₆ H ₃ –CH ₂ –	1.143	1.252	1.032	1.253
22	3,4-Cl ₂	H	H	C ₆ H ₃ –CH ₂ –	1.638	1.538	1.689	1.559
23	2-F, 4-Cl	H	H	C ₆ H ₃ –CH ₂ –	1.387	1.377	1.285	1.291
24	2-F, 4-Br	H	H	C ₆ H ₃ –CH ₂ –	1.347	1.491	1.189	1.337
25	2-F, 4-Br	H	Cl	C ₆ H ₃ –CH ₂ –	1.301	1.262	1.355	1.168
26	2-F, 4-Br	H	Br	C ₆ H ₃ –CH ₂ –	1.000	1.338	1.364	1.239
27	2-F, 4-Br	H	CH ₃ CO	C ₆ H ₃ –CH ₂ –	0.678	1.041	0.776	0.835
28	2-F, 4-Br	Cl	H	C ₆ H ₃ –CH ₂ –	1.456	1.304	1.338	1.372
29	2-F, 4-Br	Br	H	C ₆ H ₃ –CH ₂ –	1.398	1.317	1.170	1.458
30	2-F, 4-Br	Br	Br	C ₆ H ₃ –CH ₂ –	1.284	1.342	1.264	1.174

^a The observed in vitro activity values are obtained from Ref. 21.^b pIC₅₀ = –log(IC₅₀), where the IC₅₀ value is in μM.**Figure 1.** The aligned molecular structures of 30 compounds of the training set: (*R*)-enantiomers (left) and (*S*)-enantiomers (right). Atoms are colored with the CPK convention: oxygen in red, nitrogen in blue, carbon in gray, and hydrogen in white. The pyrrolo[1,2-*a*]pyrazine ring moiety is placed in the *yz*-plane.

PNSA₂ total charge weighted negative surface area: partial negative solvent-accessible surface area multiplied by the total negative charge;

PPSA₂ total charge weighted positive surface area: partial positive solvent-accessible surface area multiplied by the total positive charge;

PPSA₃ atomic charge weighted positive surface area: sum of the product of solvent-accessible surface area and partial charge for each positively charged atom;

DPSA₁ difference in charged partial surface areas: PPSA₁ – PNSA₁;

DPSA₂ difference in total charge weighted surface areas: PPSA₂ – PNSA₂;

S_y the length of molecules in the *y*-direction;²⁶

I_x, I_y the *x*, *y* components of principle moment of inertia.

Alog *P* represents the hydrophobic character of the molecule.²⁷ V_{NCOS} , I_x , I_y , and S_y are the molecular shape

analysis descriptors. Atomic charges are assigned to the lowest energy molecular structure by the charge equilibration method,²⁸ which are used to calculate the atomic polarizability, the dipole moment, and charged surface area descriptors.

2.4. The QSAR analysis

The QSAR analysis with 48 descriptors for 30 compounds in the training set is a typical over-determined problem. The genetic function approximation (GFA) algorithm is a suitable method for such over-determined QSAR analyses.²⁹

In the GFA approach, the random selection of descriptors generates the initial models. Both linear and quadratic terms of descriptors are allowed in the selection. The genetic algorithm drives the evolution of the initial models through the genetic crossover operation. Recombining terms in the better-performing models yields improved models of the next generation. Genetic operation of 20,000 iterations is performed to draw the best QSAR model.

In the GFA procedure,²⁹ the constructed models are evaluated with the Friedman's 'lack of fit' (LOF) score. The model with the lowest LOF score is selected as the best one of the generation. The mutation of equation is performed with the user adjustable smoothing parameter set to the default value of 1.0. A new term is added with the probability of 50%. The resultant QSAR equation of each model is evaluated with the regression coefficient and the LOF score. The sequential *F* test is performed as the number of descriptors of the QSAR equation is increased. As a validation measure, the cross-validated r^2 (q^2) is computed with each molecule is left out from the training set.

3. Results and discussion

We derived three sets of descriptors *R*-descriptors, *S*-descriptors, and *RS*-descriptors for 30 pyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives. Each set contained 48 descriptors with the standard deviation greater than that of the activity data. The QSAR analysis was separately performed for each descriptor set, which yielded three sets of QSAR equations *R*-equations for *R*-descriptors, *S*-equations for *S*-descriptors, and *RS*-equations for *RS*-descriptors.

We first determined the number of descriptors necessary and sufficient for the QSAR equation representative for the training set. Taking a brute force approach, we increased the number of terms in the QSAR equation one by one and evaluated the statistical contribution of each addition. The genetic function algorithm was applied to the descriptor set with limiting the number of terms of the QSAR equation.

The QSAR equation improves as the number of terms increases as shown in Figure 2. The regression coefficient directly indicates how well the QSAR equation repre-

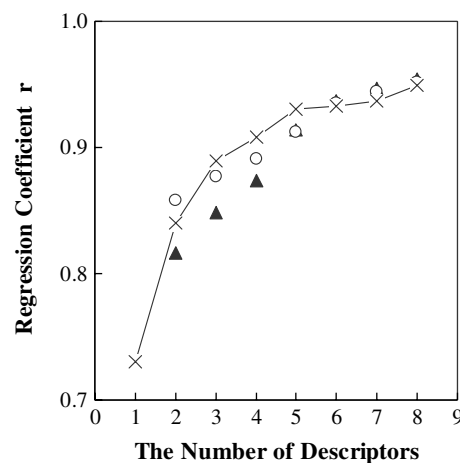


Figure 2. The regression coefficient as a function of the number of descriptors for *R*-equations (▲), *S*-equations (○), and *RS*-equations (×).

sents the activity data of the training set. Figure 2 shows that the *RS*-equations converge faster and reach larger regression coefficient with more than three terms than both *R*- and *S*-equations. However, all three series of QSAR equations perform with the similar level of confidence and reach the regression coefficient over 0.9 with five terms. We therefore concentrate on the five term QSAR equations.

The best QSAR equations derived from *R*-descriptors and their statistical values are given in Table 2. As the descriptor values in the QSAR equations are normalized, the coefficient is the direct measure on the contribution of the descriptor. *r* is the regression coefficient and LOF is the Friedman's 'lack of fit' score, which evaluates the QSAR model by considering the number of descriptors as well as the quality of fitness. The regression coefficient is 0.914 and the LOF score is 0.358 for the five term equation. Although it is little improvement, each additional term beyond the five term QSAR equation yields the regression coefficient increased by 0.01 approximately. q^2 is the cross-validated r^2 . *s* is the standard deviation and represents the absolute measure of the quality of fit. The standard deviation of all models is lower than that of the activity data (0.355). *F* is the *F* value, a measure of the level of statistical significance of the QSAR model. The *F* values given in parentheses are of 95% significance level. All *F* values of *R*-equations are greater than the 95% significance limits, indicating that *R*-equations are statistically significant at the 95% level. F_s is the sequential *F* value, which is used to compare two regression models with different number of descriptors. If the sequential *F* value is greater than the 95% significance limit, the use of the model containing the larger number of descriptors is justified. The F_s value of the best six term equation is 8.024, which means that the additional term improves the five term equation.

In the five term *R*-equation, μ_x^2 , TASA, and TPSA² descriptors positively contribute to the inhibitory activity. While the correlation coefficient of μ_x^2 to pIC₅₀ is

Table 2. Statistical assessments of the *R*-equations

Terms	Descriptors ^a								Statistics ^b					
	<i>AlogP</i>	V_{NCOS}^2	μ_x^2	TASA	PNSA ₁	TPSA ²	DPSA ₂ ²	S_y^2	<i>r</i>	q^2	LOF	<i>s</i>	<i>F</i>	F_s^c
1	0.730	–	–	–	–	–	–	–	0.730	0.467	0.518	0.247	31.994 (4.196)	
2	0.820	–0.375	–	–	–	–	–	–	0.816	0.593	0.430	0.213	26.918 (3.354)	10.743
3	0.801	–0.391	0.231	–	–	–	–	–	0.848	0.626	0.425	0.199	22.115 (2.975)	4.929
4	–	–0.572	0.280	0.696	0.281	–	–	–	0.874	0.638	0.424	0.186	20.237 (2.759)	4.740
5	–	–0.642	0.428	2.961	–	2.108	–1.700	–	0.914	0.688	0.358	0.158	24.346 (2.621)	10.428
6	–	–0.569	0.456	3.034	–	2.141	–1.674	–0.244	0.937	0.698	0.329	0.140	27.427 (2.528)	8.024

^a The normalized descriptors were used.

^b *r* is the regression coefficient; q^2 is the square of the cross-validated regression coefficient; LOF is the Friedman's 'lack of fit'; *s* is the standard deviation; *F* is the *F* value; F_s is the sequential *F* value.

^c $F_s = \frac{(r_2^2 - r_1^2) \cdot (n - k_2 - 1)}{(k_2 - k_1) \cdot (1 - r_2^2)}$; *k* is the number of terms ($k_1 < k_2$); *r* is the regression coefficient; *n* is the number of molecules.

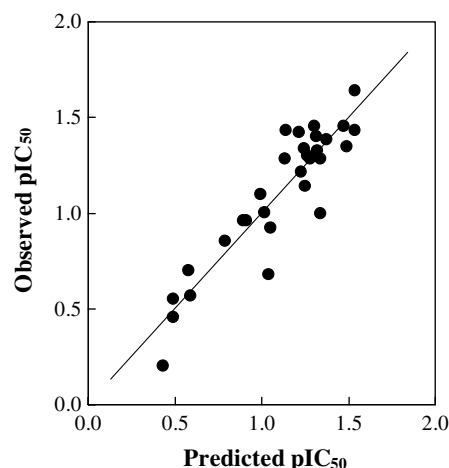
0.277, that of μ_x to pIC₅₀ is –0.204. A compound with the larger dipole moment in the negative *x*-direction is expected to have greater inhibitory activity against the aldose reductase. TASA, the descriptor for the total hydrophobic surface area, strongly correlates with *AlogP*, the descriptor for hydrophobic character of the compound. The TASA–*AlogP* correlation coefficient is 0.848. *AlogP* positively contributes to the activity in the QSAR equations with one, two, and three terms. TASA replaces *AlogP* in the QSAR equations with more than three terms and also positively contributes to the activity. The positive contribution of *AlogP* and TASA suggests an important role of the hydrophobicity in the inhibition of aldose reductase. In contrast, the positive coefficient of TPSA² indicates the importance of molecular interactions among polar groups. The negative coefficient of V_{NCOS}^2 indicates that the molecular shape needs to be similar to that of the reference compound in order to be favorable for the inhibitory activity. The QSAR equation also includes the negative term of the difference in total charge weighted surface area (DPSA₂²).

The correlation matrix of the descriptors of the five term QSAR model for *R*-descriptors is given in Table 3. Although TASA and TPSA² correlate moderately (–0.632), all the other correlation coefficients among descriptors are less than 0.5. Five descriptors construct reasonably independent terms of the QSAR model. The activities obtained from the five term QSAR equation are listed in Table 1 and compared with the biological activity data in Figure 3. The standard deviation is 0.158 and the cross-validated r^2 value is 0.688.

The best QSAR equations derived from *S*-descriptors are summarized in Table 4. The five term *S*-equation yields the regression coefficient 0.912 with the LOF score

Table 3. Correlation matrix for the biological activity and the descriptors of the five term *R*-equation

	V_{NCOS}^2	μ_x^2	TASA	TPSA ²	DPSA ₂ ²	pIC ₅₀
V_{NCOS}^2	1.000					
μ_x^2	0.087	1.000				
TASA	0.440	0.053	1.000			
TPSA ²	–0.178	–0.214	–0.632	1.000		
DPSA ₂ ²	0.294	–0.117	0.443	0.396	1.000	
pIC ₅₀	–0.179	0.277	0.617	–0.413	0.207	1.000

**Figure 3.** Observed and predicted activities of pyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives using the five term *R*-equation.

0.368. The *S*-equations show the similar statistical characteristics as the *R*-equations. The regression coefficient increases in small increments beyond five terms. The F_s value of the sixth term is 7.768 and justifies incorporation of the sixth term. However, the five term QSAR equation is analyzed in the same format as the *R*-equation.

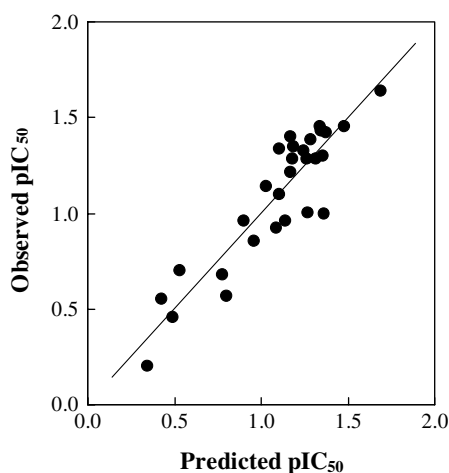
The five term *S*-equation includes *AlogP*, V_{NCOS}^2 , α_{pol}^2 , μ_y , and PPSA₃. *AlogP*, α_{pol}^2 , and μ_y positively contribute to the inhibitory activity while V_{NCOS}^2 and PPSA₃ negatively contribute. As in the *R*-equation, the hydrophobic character *AlogP* is the most important factor for the inhibitory activity. The molecular shape should be close to that of the reference compound in order to ensure efficient inhibition of the aldose reductase activity. The atomic polarizability α_{pol}^2 and μ_y take the role of μ_x^2 and TPSA² of the five term *R*-equation. PPSA₃ negatively contributes to the activity as DPSA₂² does in the *R*-equation. Table 5 is the correlation matrix of the descriptors in the five term *S*-equation. While α_{pol}^2 shows moderate correlation with both *AlogP* (the correlation coefficient 0.581) and PPSA₃ (0.525), the five term QSAR equation reproduces the observed activity at the similar level as the *R*-equation as shown in Figure 4. The predicted activities of the five term *S*-equation are also listed in Table 1. The standard deviation is 0.161 for the 30 compounds of the training set.

Table 4. Statistical assessments of the *S*-equations

Terms	Descriptors ^a							Statistics ^b					
	<i>Alog P</i>	V_{NCOS}^2	α_{pol}^2	μ_y	PPSA ₃	DPSA ₁	TPSA ²	<i>r</i>	q^2	LOF	<i>s</i>	<i>F</i>	F_s^c
1	0.730	–	–	–	–	–	–	0.730	0.467	0.518	0.247	31.994 (4.196)	
2	0.796	–0.454	–	–	–	–	–	0.858	0.668	0.340	0.189	37.530 (3.354)	20.801
3	0.669	–0.477	0.225	–	–	–	–	0.877	0.682	0.350	0.180	28.776 (2.975)	3.712
4	0.707	–0.563	0.237	0.182	–	–	–	0.891	0.698	0.371	0.174	24.008 (2.759)	3.002
5	0.603	–0.573	0.453	0.276	–0.276	–	–	0.912	0.739	0.368	0.161	23.604 (2.621)	5.401
6	0.541	–0.509	0.448	0.376	–	–0.991	–0.787	0.935	0.794	0.340	0.142	26.478 (2.528)	7.768

^{a–c}See Table 2.**Table 5.** Correlation matrix for the biological activity and the descriptors of the five term *S*-equation

	<i>Alog P</i>	V_{NCOS}^2	α_{pol}^2	μ_y	PPSA ₃	pIC ₅₀
<i>Alog P</i>	1.000					
V_{NCOS}^2	0.146	1.000				
α_{pol}^2	0.581	0.182	1.000			
μ_y	–0.175	0.433	–0.099	1.000		
PPSA ₃	0.014	0.202	0.525	0.315	1.000	
pIC ₅₀	0.730	–0.338	0.527	–0.209	–0.058	1.000

**Figure 4.** Observed and predicted activities of pyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives using the five term *S*-equation.

The best *RS*-equations are listed in Table 6 with the statistical measures. The five term *RS*-equation is the best among the five term QSAR equations with the regression coefficient 0.93 and the LOF score 0.295. It also has the most predictive power with the cross-validated r^2 value of 0.746. Additional degrees of freedom to the five term equation provide very little improvement of

statistics. The regression coefficient of the six term equation is 0.933 and that of the seven term equation is 0.937. Furthermore, the F_s value is 0.993 for the sixth term addition and 1.349 for the seventh term addition. The five term *RS*-equation contains necessary and sufficient set of descriptors representing the activity data of the training compounds.

In the five term *RS*-equation, TASA also positively contributes to the inhibitory activity, which emphasizes hydrophobic interactions in binding to the aldose reductase. The hydrophobic interaction is negatively modulated by the polar surface area descriptor, DPSA₁. The negative contribution of V_{NCOS}^2 is common to all QSAR models. The reference compound has well optimized shape for tight binding interactions. As discussed in the five term *R*-equation, μ_x negatively contributes to pIC₅₀ with the correlation coefficient of –0.405 (Table 7). The last entry into the five term *RS*-equation is the molecular shape descriptor, S_y^2 , negatively contributing to the inhibitory activity. The elongation along the *y*-axis would reduce the activity. As the R₃ substitution is aligned to the *y*-direction, a bulky substitution does not favor the inhibitory activity.

The correlation matrix of descriptors in the *RS*-equation is given in Table 7. The maximum value of the intercorrelation coefficients is 0.427 between V_{NCOS}^2 and S_y^2 . All other intercorrelation coefficients are small, indicating that the GFA algorithm has chosen a robust set of descriptors for the *RS*-equation. The activities obtained from the five term *RS*-equation are listed in Table 1 and compared with the observed activities in Figure 5. The standard deviation is 0.144, which is 0.158 for the *R*-equation and 0.161 for the *S*-equation.

The arithmetic mean *RS*-descriptors yield the better QSAR model for the 30 racemic compounds listed in

Table 6. Statistical assessments of the *RS*-equations

Terms	Descriptors ^a							Statistics ^b					
	<i>Alog P</i>	V_{NCOS}^2	TASA	DPSA ₁	μ_x	S_y^2	α_{pol}^2	<i>r</i>	q^2	LOF	<i>s</i>	<i>F</i>	F_s^c
1	0.730	–	–	–	–	–	–	0.730	0.467	0.518	0.247	31.994 (4.196)	
2	0.814	–0.424	–	–	–	–	–	0.840	0.639	0.379	0.200	32.381 (3.354)	15.839
3	–	–0.592	0.857	–0.712	–	–	–	0.889	0.722	0.315	0.171	32.831 (2.975)	10.505
4	–	–0.545	0.819	–0.672	–0.188	–	–	0.908	0.741	0.317	0.161	29.217 (2.759)	4.863
5	–	–0.442	0.869	–0.712	–0.274	–0.252	–	0.930	0.746	0.295	0.144	30.595 (2.621)	7.183
6	–	–0.425	0.778	–0.683	–0.260	–0.278	0.122	0.933	0.780	0.349	0.144	25.679 (2.528)	0.993

^{a–c} See Table 2.

Table 7. Correlation matrix for the biological activity and the descriptors of the five term *RS*-equation

	V_{NCOS}^2	TASA	DPSA ₁	μ_x	S_y^2	pIC ₅₀
V_{NCOS}^2	1.000					
TASA	0.367	1.000				
DPSA ₁	-0.020	0.337	1.000			
μ_x	0.170	-0.039	0.137	1.000		
S_y^2	0.427	0.306	-0.147	-0.299	1.000	
pIC ₅₀	-0.264	0.400	-0.411	-0.405	0.011	1.000

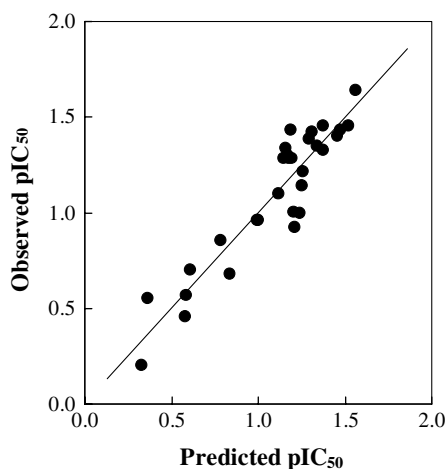
**Figure 5.** Observed and predicted activities of pyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives using the five term *RS*-equation.

Table 1 than the *R*-descriptors or *S*-descriptors do by themselves. We employ the *RS*-descriptor set to derive QSAR equations for each substitution position around the 1,2,3,4-tetrahydropyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone core. The training compounds of Table 1 are sorted into three groups according to the substituted moiety. Group I contains compounds with a simple substitution only at the R₄ position, compounds 1–5. Twenty compounds (compounds 5–24) with substituted benzyl group at the R₄ position belong to Group II. Group III includes 2-benzyl compounds with substituents at R₂ and/or R₃ positions, compounds 24–30.

Table 8 lists the most succinct *RS*-equation with the regression coefficient greater than 0.8 for each group. For Group I compounds, the inhibitory activity depends linearly on the TASA value. The optimal choice at the R₄ position is the benzyl moiety, which corresponds to

the reasoning of Negoro et al.²¹ Substitution on the benzyl group modulates the inhibitory activity and is optimized into compound 22 (the 2,4-dichloro substitution). The negative coefficient of S_y^2 in the Group III QSAR equation indicates that a large substituent at the R₂ or R₃ position is unfavorable to the inhibitory activity. There would be little space to hold a bulky R₂ or R₃ substituent in the binding site.

It is necessary to demonstrate the superior predictability of the *RS*-equation against its *R*- and *S*-counterparts. Negoro et al. prepared the enantiomers of the compound 24 (SX-3030) and evaluated the aldose reductase activity of the (*S*)-(+)-enantiomer (SX-3202) and (*R*)-(–)-enantiomer (AS-3201), 0.19 and 0.015 μM, respectively. The racemic mixture SX-3030 has the aldose reductase inhibitory activity 0.045 μM. As reported previously,^{19–21,23} the aldose reductase has high enantioselectivity for pyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone derivatives. Table 9 lists the predicted activity of five term QSAR equations.

The five term *R*-equation predicts the observed pIC₅₀ values with the standard deviation of 1.452 and the *S*-equation does with the standard deviation of 0.396. The *RS*-equation predicts those activity values with the standard deviation of 0.185. The *R*-equation poorly predicts the activity of both the (*S*)-enantiomer and the racemate and results in the large deviation. The *RS*-equation is the most robust and stable QSAR equation and closely reproduces the observed activity of either optically pure enantiomers or racemic mixtures.

4. Conclusion

The racemic descriptor set is derived as the simple arithmetic average of the corresponding *R*- and *S*-descriptors and yields the QSAR equation of better quality than the original descriptor set used separately. The *RS*-equation stands with the larger regression coefficient with relatively small number of less intercorrelating terms than *R*- and *S*-equations. The five term *RS*-equation with the regression coefficient 0.930 and the LOF score 0.295 shows that the hydrophobic character of the benzyl moiety is the major contributing factor to the aldose reductase inhibitory activity. The contribution is modulated by the polar descriptors, DPSA₁ and μ_x . The *RS*-equation also indicates that the bulky substitution at the R₂ or R₃ position is unfavorable for the inhibitory activity.

Table 8. The subgroup QSAR models with *RS*-descriptors

Group	TASA	V_{NCOS}^2	μ_x	S_y^2	Intercept	<i>n</i>	<i>r</i>	<i>F</i>
I ^a	0.331	–	–	–	–1.079	5	0.805	5.511
II ^b	0.400	–0.592	–0.557	–	–0.158	20	0.822	11.130
III ^c	–	–	–	–0.431	0.613	7	0.815	9.920

^a Group I contains molecules with substituent at R₄, compounds 1–5.

^b Group II contains molecules with substituent at R₁, compounds 5–24.

^c Group III contains molecules with substituent at R₂ and R₃, compounds 24–30.

Table 9. Predicted activity of five term QSAR equations for AS-3201, SX-3202, and SX-3030^a

QSAR model	AS-3201	SX-3202	SX-3030
R-Equation	1.491	2.670	1.900
S-Equation	1.517	1.189	1.353
RS-Equation	1.715	0.958	1.337
Observed activity	1.824	0.721	1.347

^a The inhibitory activity is the pIC₅₀ (μM) value.

The RS-equation is superior to its R- and S-counterparts in predicting the inhibitory activity of optically pure enantiomers as well as racemic mixtures. The RS-descriptors are suitable to probe QSAR of the activity data assayed with racemates. It would be possible to improve the racemic descriptors using the Boltzmann factor weighted average of R- and S-descriptors. A logical extension of this work is in progress to drive the RS-QSAR model for the in vivo activity data.²¹

References and notes

- Kador, P. F. *Med. Res. Rev.* **1988**, *8*, 325.
- Tomlinson, D. R.; Stevens, E. J.; Diemel, L. T. *Trends Pharmacol. Sci.* **1994**, *15*, 293.
- Lee, A. Y. W.; Chung, S. S. M. *FASEB* **1999**, *13*, 23.
- Nishimura, C. Y. *Pharmacol. Rev.* **1998**, *50*, 21.
- Yamagishi, M.; Yamada, Y.; Ozaki, K.; Asao, M.; Shimizu, R.; Suzuki, M.; Matsumoto, M.; Matsuoka, Y.; Matsumoto, K. *J. Med. Chem.* **1992**, *35*, 2085.
- Costantino, L.; Rastelli, G.; Vescovini, K.; Cignarella, G.; Vianello, P.; Del Corso, A.; Cappiello, M.; Mura, U.; Barlocco, D. *J. Med. Chem.* **1996**, *39*, 4396.
- Kotani, T.; Nagaki, Y.; Ishii, A.; Konishi, Y.; Yago, H.; Suehiro, S.; Okukado, N.; Okamoto, K. *J. Med. Chem.* **1997**, *40*, 684.
- Costantino, L.; Rastelli, G.; Gamberini, M. C.; Vinson, J. A.; Bose, P.; Iannone, A.; Staffieri, M.; Antolini, L.; Del Corso, A.; Mura, U.; Albasini, A. *J. Med. Chem.* **1999**, *42*, 1881.
- Costantino, L.; Rastelli, G.; Gamberini, M. C.; Giovannoni, M. P.; Piaz, V. D.; Vianello, P.; Barlocco, D. *J. Med. Chem.* **1999**, *42*, 1894.
- Fresneau, P.; Cussac, M.; Morand, J. M.; Szymonski, B.; Tranqui, D.; Leclerc, G. *J. Med. Chem.* **1998**, *41*, 4706.
- Nicolaou, I.; Zika, C.; Demopoulos, V. J. *J. Med. Chem.* **2004**, *47*, 2706.
- Settimo, F. D.; Primofiore, G.; Settimo, A. D.; Motta, C. L.; Simorini, F.; Novellino, E.; Greco, G.; Lavecchia, A.; Boldrini, E. *J. Med. Chem.* **2003**, *46*, 1419.
- Pau, A.; Asproni, B.; Boatto, G.; Grella, G. E.; Caprariis, P. D.; Costantino, L.; Pinna, G. A. *Eur. J. Pharm. Sci.* **2004**, *21*, 545.
- Bruno, G.; Costantino, L.; Curinga, C.; Maccari, R.; Monforte, F.; Nicolò, F.; Ottanà, R.; Vigorita, M. G. *Bioorg. Med. Chem.* **2002**, *10*, 1077.
- Costantino, L.; Corso, A. D.; Rastelli, G.; Petrash, J. M.; Mura, U. *Eur. J. Med. Chem.* **2001**, *36*, 697.
- Pfeifer, M. A.; Schumer, M. P.; Gelber, D. A. *Diabetes* **1997**, *46*, S82.
- Costantino, L.; Rastelli, G.; Vianello, P.; Cignarella, G.; Barlocco, D. *Med. Res. Rev.* **1999**, *19*, 3.
- Yasuda, H.; Terada, M.; Maeda, K.; Kogawa, S.; Sanada, M.; Haneda, M.; Kashiwagi, A.; Kikkawa, R. *Prog. Neurobiol.* **2003**, *69*, 229.
- Sarges, R.; Schnur, R. C.; Belletire, J. L.; Peterson, M. J. *J. Med. Chem.* **1988**, *31*, 230.
- Sarges, R.; Goldstein, S. W.; Welch, W. M.; Swindell, A. C.; Siegel, T. W.; Beyer, T. A. *J. Med. Chem.* **1990**, *33*, 1859.
- Negoro, T.; Murata, M.; Ueda, S.; Fujitani, B.; Ono, Y.; Kuromiya, A.; Komiya, M.; Suzuki, K.; Matsumoto, J. I. *J. Med. Chem.* **1998**, *41*, 4118.
- Cerius2, version 4.6; Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.
- Kurono, M.; Fujiwara, I.; Yoshida, K. *Biochemistry* **2001**, *40*, 8216.
- Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
- Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
- Rohrbaugh, R. H.; Jurs, P. C. *Anal. Chim. Acta* **1987**, *199*, 99.
- Ghose, A. K.; Crippen, G. M. *J. Comput. Chem.* **1986**, *7*, 565.
- Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.
- Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.