Racemic Descriptors for Quantitative Structure Activity Relationship of Spirosuccinimide Type Aldose Reductase Inhibitors

Jeong Rim Kim* **and Youngdo Won**†

Department of Applied Chemistry and Chemistry, Hanyang University, Ansan 425-791, Korea
†*Department of Chemistry, Hanyang University, Seoul 133-792, Korea*
Received September 22, 2004

Quantitative structure activity relationship has been probed for spirosuccinimide-fused tetrahydropyrrolo[1,2-*a*]pyrazine-1,3-dione derivatives acting as aldose reductase inhibitors. While the spirosuccinimide compounds contain a chiral center, the aldose reductase inhibition assay was performed with racemic mixtures in the published work. As the physicochemical descriptors of the QSAR analysis must be evaluated for a definite molecular structure, 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, closely reproducing the observed activity of optically pure enantiomers as well as racemic mixtures.

Key Words : QSAR, Racemic descriptor, Aldose reductase inhibitor

Modern drug discovery research relies on massive synthesis, assay and rational design of compounds to generate *de novo* lead molecular constructs.[1,2] The lead compound is further optimized to enhance potency and deliverability. Quantitative structure activity relationship (QSAR) investigation aids lead optimization by analyzing the assay data in terms of molecular descriptors. While drug candidate molecules often contain chiral centers, the activity assay is rarely performed with optically pure enantiomers. In such cases, it is a challenging question which molecular structure to employ in evaluating the QSAR descriptors. 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.

Aldose reductase (AR) is the first enzyme of the sorbitol pathway in which it converts glucose to sorbitol.[3] As high intracellular sorbitol accumulation causes chronic diabetic complications such as retinopathy, neuropathy, nephropathy, and cataracts,[4] inhibition of AR is an effective treatment for long-term diabetic malignity.[5] Negoro *et al.* synthesized tetrahydropyrrolo[1,2-*a*]pyrazine derivatives and performed the assay.[6] Although the spirosuccinimide cogeners contain the chiral center indicated by * in the figure of Table 1, the AR inhibition assay is performed with racemic mixtures.

We obtain IC_{50} values of 30 spirosuccinimide compounds from the published data of Negoro *et al.* The AR inhibitory activity is defined as $pIC_{50} = -\log(IC_{50})$. The normalized pIC_{50} values of the training set are listed in Table 1. We utilize the Cerius[2] program package[7] and separately build the (*S*)-enantiomer and (*R*)-enantiomer molecular structures of the pyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone derivatives. For each enantiomer, thorough conformational search and energy minimization result in the optimized molecular geometry. The adopted basis Newton Raphson minimization method is used with the Merck Molecular Force Field. The optimized geometry is aligned to that of the most potent compound, 2-(4-bromo-2-fluorobenzyl)-1,2,3,4-tetrahydropyrrolo[1,2-*a*]pyrazine-4-spiro-3'-pyrrolidine-1,2',3,5'-tetrone (SX-3030). We used atomic charges assigned by the charge equilibration method for descriptor evaluation.[8]

The QSAR[+] module of the Cerius[2] package generates 78 physicochemical descriptors for each enantiomer. *R*-descriptors, as we define, are based on the optimized structure of the (*R*)-enantiomers and *S*-descriptors are derived for the (*S*)-enantiomers. We take the respective arithmetic mean values of *R*-descriptors and *S*-descriptors to generate the new set of descriptors, *RS*-descriptors. In order to warrant evenly distributed descriptor basis, descriptors with the standard deviation smaller than that of the pIC_{50} values are removed to leave 48 descriptors in each set.

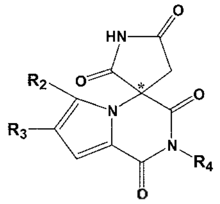
The genetic function approximation method is applied to the over-determined QSAR problem of 48 descriptors for 30 compounds.[9] The random selection of both linear and quadratic descriptors generates the initial models, which evolve through genetic crossover operations. 20,000 genetic operations are performed to derive the best QSAR equation.

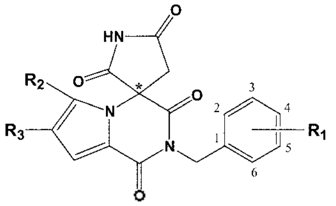
We drive three sets of QSAR equations: *R*-equations with *R*-descriptors, *S*-equations with *S*-descriptors, and *RS*-equations with *RS*-descriptors. As detailed QSAR analyses will appear somewhere else, here we concentrate on the performance comparison among the descriptor sets. The genetic function algorithm is applied to each descriptor set with successively increasing the number of terms in the QSAR equation. This procedure determines the necessary and sufficient QSAR equation with parsimony of terms. The statistics of the best QSAR equations with one through eight terms is summarized in Table 2.

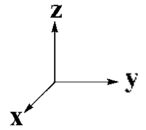
The QSAR equations fit the training data set more closely

*Corresponding Author. Tel: +82-31-400-5491; Fax: +82-31-407-3863; e-mail: jrkim@hanyang.ac.kr

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'-tetrones[a,b]







Compd	R_1	R_2	R_3	R_4	pIC_{50}[c]			
					observed	*R*-eqn	*S*-eqn	*RS*-eqn
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_6H_5-	0.699	0.581	0.531	0.611
4	–	H	H	C_6H_5-$(CH_2)_2$-	0.569	0.591	0.801	0.584
5	H	H	H	C_6H_4-CH_2-	1.004	1.018	1.267	1.203
6	2-F	H	H	C_6H_4-CH_2-	1.215	1.230	1.170	1.259
7	4-F	H	H	C_6H_4-CH_2-	0.921	1.050	1.090	1.209
8	2-Cl	H	H	C_6H_4-CH_2-	1.432	1.140	1.348	1.188
9	3-Cl	H	H	C_6H_4-CH_2-	1.456	1.472	1.481	1.519
10	4-Cl	H	H	C_6H_4-CH_2-	1.420	1.216	1.373	1.308
11	2-Br	H	H	C_6H_4-CH_2-	1.284	1.134	1.184	1.193
12	3-Br	H	H	C_6H_4-CH_2-	1.432	1.540	1.345	1.475
13	4-Br	H	H	C_6H_4-CH_2-	1.328	1.319	1.245	1.376
14	4-CH_3	H	H	C_6H_4-CH_2-	1.284	1.278	1.317	1.149
15	4-OCH_3	H	H	C_6H_4-CH_2-	1.337	1.248	1.107	1.159
16	4-CF_3	H	H	C_6H_4-CH_2-	1.097	0.996	1.104	1.116
17	4-NO_2	H	H	C_6H_4-CH_2-	0.959	0.897	1.143	0.994
18	4-NH_2	H	H	C_6H_4-CH_2-	0.854	0.790	0.961	0.786
19	2,4-$(OCH_3)_2$	H	H	C_6H_3-CH_2-	0.456	0.492	0.492	0.578
20	3,4-$(OCH_3)_2$	H	H	C_6H_3-CH_2-	0.959	0.915	0.900	1.000
21	2,4-F_2	H	H	C_6H_3-CH_2-	1.143	1.252	1.032	1.253
22	3,4-Cl_2	H	H	C_6H_3-CH_2-	1.638	1.538	1.689	1.559
23	2-F, 4-Cl	H	H	C_6H_3-CH_2-	1.387	1.377	1.285	1.291
24	2-F, 4-Br	H	H	C_6H_3-CH_2-	1.347	1.491	1.189	1.337
25	2-F, 4-Br	H	Cl	C_6H_3-CH_2-	1.301	1.262	1.355	1.168
26	2-F, 4-Br	H	Br	C_6H_3-CH_2-	1.000	1.338	1.364	1.239
27	2-F, 4-Br	H	CH_3CO	C_6H_3-CH_2-	0.678	1.041	0.776	0.835
28	2-F, 4-Br	Cl	H	C_6H_3-CH_2-	1.456	1.304	1.338	1.372
29	2-F, 4-Br	Br	H	C_6H_3-CH_2-	1.398	1.317	1.170	1.458
30	2-F, 4-Br	Br	Br	C_6H_3-CH_2-	1.284	1.342	1.264	1.174

[a]The observed *in vitro* activity values are obtained from Negoro *et al.*[6] [b]The chiral center is denoted by '*'. The pyrollo[1,2-*a*]pyrazine ring is placed in the yz-plane and the CN bond of the five membered aromatic ring is aligned to the z-axis. [c]$pIC_{50} = -\log(IC_{50})$ is normalized for the QSAR analyses. IC_{50} value in μM.

Table 2. Statistical Evaluation of QSAR Models*

Number of Terms	*R*-Equations			*S*-Equations			*RS*-Equations		
	r	s	F_s	r	s	F_s	r	s	F_s
1	0.730	0.247		0.730	0.247		0.730	0.247	
2	0.816	0.213	10.743	0.858	0.189	20.801	0.840	0.200	15.839
3	0.848	0.199	4.929	0.877	0.180	3.712	0.889	0.171	10.505
4	0.874	0.186	4.740	0.891	0.174	3.022	0.908	0.161	4.863
5	0.914	0.158	10.428	0.912	0.161	5,401	0.930	0.144	7.183
6	0.937	0.140	8.024	0.935	0.142	7.768	0.933	0.144	0.993
7	0.947	0.130	4.017	0.944	0134	3.417	0.937	0.142	1.349
8	0.954	0.125	3.109	0.952	0.128	3.400	0.949	0.132	4.781

*r is the regression coefficient. *s* is the standard deviation. F_s is the sequential F value. $Fs = (r_2^2 - r_1^2)\cdot(n - k_2 - 1)/(k_2 - k_1)\cdot(1 - r_2^2)$ where *k* is the number of terms ($k_1 < k_2$) and n is the number of compounds.

as the number of terms increases. While all three series of QSAR equations perform at the similar level of confidence and reach the regression coefficient over 0.9 with five terms, the *RS*-equations converge more rapidly. The outstanding convergence of the *RS*-equations is indicated by the F_s value, which compares two successive regression models. The large F_s value strongly justifies the addition of the new term. While the F_s values in Table 2 justify the sixth and the seventh terms of *R*- and *S*-equations, they manifest no requirement for those terms beyond the fifth of the *RS*-equations. Among the five term QSAR equations, the *RS*-equation best fits the inhibitory data as indicated by the smallest standard deviation.

In terms of the normalized descriptors, the five term QSAR equations are as follows.

$$pIC_{50} = 2.961\ TASA - 0.642\ V_{NCOS}^2 + 0.428\ \mu_x^2 + 2.108\ TPSA^2 - 1.700\ DPSA_2^2 \qquad (R\text{-Equation})$$

$$pIC_{50} = 0.603\ AlogP - 0.573\ V_{NCOS}^2 + 0.276\ \mu_y + 0.453\ \alpha_{pol}^2 - 0.276\ PPSA_3 \qquad (S\text{-Equation})$$

$$pIC_{50} = 0.869\ TASA - 0.442\ V_{NCOS}^2 - 0.274\ \mu_x - 0.252\ S_y^2 - 0.712\ DPSA_1 \qquad (RS\text{-Equation})$$

The predicted pIC_{50} values of are listed in Table 1 for each QSAR equation. AlogP is the logarithm of the partition coefficient. V_{NCOS} is the non-common overlap volume, μ is the dipole moment, α_{pol} is the sum of atomic polarizability, S_y is the length in the y-direction, and the other descriptors are surface areas incorporating atomic charges. TASA and AlogP strongly correlate each other with the correlation coefficient over 0.8. Both represent the hydrophobicity and positively contribute to the inhibitory activity. The negative coefficient of V_{NCOS}^2 indicates that the molecular shape needs to be similar to that of the reference compound for strong interaction with AR. The above QSAR equations suggest that the hydrophobic interaction plays important role in tight binding to AR and the interaction is modulated by atomic charge distribution. The correlation matrix of the descriptors in the *RS*-equation has the absolute value average of 0.225 with the maximum 0.427. The *RS*-descriptors in the QSAR equation substantially less correlate with each other than those in the *R*-equation (the corresponding values are 0.285 and 0.443, respectively) and the *S*-equation (0.266 and 0.585, respectively). The correlation matrix is another indicator showing that the *RS*-descriptors yield the QSAR equation in better quality.

Table 3. pIC_{50} values predicted by five term QSAR Equations

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

We further demonstrate superiority of the *RS*-equation against its *R*- and *S*-counterparts by predicting the inhibitory activity of enantiomers. While the racemic mixture SX-3030 has the AR inhibitory activity 0.045 μM, the (*S*)-(+)-enantiomer (SX-3202) and (*R*)-(−)-enantiomer (AS-3201) do 0.19 and 0.015 μM, respectively. Table 3 presents the predicted activity of each five term QSAR equation. The five term *R*-equation predicts the observed pIC_{50} values of the enantiomers and the racemate with the standard deviation of 1.452 and the *S*-equation does with the standard deviation of 0.369. The *RS*-equation predicts those activity values with the standard deviation of 0.185.

The *RS*-equation is the most robust and stable QSAR equation. It stands with the largest regression coefficient with relatively small number of terms that intercorrelate less than those of *R*- and *S*-equations. It closely reproduces the observed activity of either optically pure enantiomers or racemic mixtures. The arithmetic mean *RS*-descriptors are suitable to probe QSAR of the activity data assayed with racemates.

References

1. Suh, M. E.; Park, S. Y.; Lee, H. J. *Bull. Korean Chem. Soc.* **2003**, *23*, 417.
2. Cho, D. H.; Lee, S. K.; Kim, B. T.; No, K. T. *Bull. Korean Chem. Soc.* **2001**, *22*, 388.
3. Petrash, J. M. *Cell. Mol. Life Sci.* **2004**, *61*, 737.
4. Altan, V. M. *Curr. Med. Chem.* **2003**, *10*, 1317.
5. Suzen, S.; Buyukbingol, E. *Curr. Med. Chem.* **2003**, *10*, 1329.
6. 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.
7. *Cerius²*, version 4.6, Accelrys Inc., San Diego, U.S.A.
8. Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
9. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.