Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 244–253
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

A neural networks-based drug discovery approach and its application for designing aldose reductase inhibitors

LiHong Hu [a], GuanHua Chen [a,*], Raymond Ming-Wah Chau [b]

[a] Department of Chemistry, The University of Hong Kong, Pokfulam, HKSAR, Hong Kong, China
[b] Department of Anatomy, The University of Hong Kong, Pokfulam, HKSAR, Hong Kong, China

Received 8 June 2005; received in revised form 9 September 2005; accepted 9 September 2005
Available online 13 October 2005

Abstract

A novel approach that combines neural networks, computer docking and quantum mechanical method is developed to design potent aldose reductase inhibitors (ARIs). Neural networks is employed to determine the quantitative structure–activity relationship (QSAR) among the known ARIs. The physical descriptors of the neural networks, such as electronegativity and molar volume, are evaluated with first-principles quantum mechanical method. Based on the QSAR, new candidates for ARI are predicted, and subsequently screened via computer docking technique. The surviving candidates are further tested via quantum mechanical calculation for their bindings to aldose reductase. We find that the best 49 predicted ARI candidates have better calculated binding energies than those of experimentally known drug candidates.

Keywords: Neural networks; ARIs; QSAR

1. Introduction

Insulin-dependent diabetes mellitus results from metabolic disorders of glucose, which is due to the deficiency of insulin caused by the defects in insulin secretion. For the diabetes mellitus patients, the mortality is actually not caused by diabetes mellitus itself but by diabetic complications, such as neuropathy, nephropathy, retinopathy and cardiovascular disease [1]. These diabetic complications are induced by an enzyme, aldose reductase (AR) [2], which is a nicotinamide adenine dinucleotide phosphate (NADPH) dependent enzyme expressed in various human tissues such as in lens, nerve, retina, muscle, placenta and liver. Physiologically, AR participates in the polyol pathway and reduces a wide range of substrates, such as sugar, aldehyde, aldose and corticosteroid, to their corresponding alcohols [3]. This polyol pathway involves two enzymes, namely AR and sorbitol dehydrogenase. These two enzymes convert glucose to fructose via a two-step reaction: (1) glucose is reduced to sorbitol by AR and (2) the sorbitol is further oxidized to fructose by sorbitol dehydrogenase [4]. The diabetic complications are ascribed to the abnormity of this glucose metabolic process. With the glucose level increasing in blood, the activity of AR increases. Therefore, the rate of sorbitol being produced is much faster than that of sorbitol being oxidized. This results in the excessive accumulation of sorbitol. As a consequence, the excessive sorbitol seriously blocks the cell membrane pervasion, and ultimately leads to diabetic complications [1–4]. AR inhibition or avoiding the sorbitol accumulation has been thus proposed as a way to prevent or delay diabetic complications [5–8].

Designing potent ARIs for inhibiting the AR activity and reducing the concentration of sorbitol has been the target for many researchers [8–12]. Both experimental studies and computer simulation have been carried out to find potent ARIs. However, compounds that have been identified as potent ARIs in both vitro and vivo are few [13–16]. All these found ARIs can be classified into several types such as flavonoids, spirohydantoins, substituted acetic acid and phenylsulfonylnitromethane derivations [5,17,18]. In our study, we develop a novel strategy that combines neural networks, computer docking, quantum mechanical calculation to design potent ARIs. Neural networks is applied to determine the QSAR among the known ARIs, and the determined QSAR is subsequently used to predict new potent ARIs. The predicted ARIs are then subjected to computer docking simulation for screening. Those that passed the docking test are further subjected to the quantum mechanical calculation for confirmation.

* Corresponding author. Tel.: +852 2859 2164; fax: +852 2857 1586.
E-mail address: ghc@everest.hku.hk (G. Chen).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.09.002

The article is presented as follows. After an introduction, we present our neural networks method for determining the QSAR in Section 2. Computer docking simulations and quantum mechanical calculations for confirming the binding between the drug candidates and AR are presented in Section 3. Finally, discussion is given in Section 4.

2. Quantitative structure–activity relationship (QSAR)

A series of spirohydantoin derivatives, which showed their affinities to AR active site, were reported by Yamagishi et al. [19]. They synthesized several series of spirohydantoin derivatives with same lead structure (shown in Fig. 1a), and claimed one of these derivatives, (4*R*)-6′-chloro-3′-methyl-spiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione, could inhibit polyol accumulation in the sciatic nerve in vivo and had low toxicity for mice. The lead structure (shown in Fig. 1a) of these inhibitors is similar to that of fidarestat (see Fig. 1b) [20]. A series of molecules based on the lead structure shown in Fig. 1a are investigated for this study.

A QSAR can be constructed by analyzing the molecular structure/properties and the corresponding bioactivities. The resulting QSAR can thus be used to design new drug leads. Hansch et al. [21] developed a method to determine the QSAR from a free energy related constant (π-comparative substituent constant) and electronic parameter (σ-Hammett function) [21]. The QSAR equation was expressed as

$$\log \frac{1}{c} = -k\pi^2 + k'\pi + \rho\sigma + k'', \tag{1}$$

where c is the concentration that induces certain biological response, ρ the reaction constant, and k, k', k'' are regression coefficients [21]. Later, a three-dimensional QSAR calculation method, comparative molecular field analysis (CoMFA), was developed by Cramer et al. [22], and it describes the relationship between the molecular 3D-steric and electrostatic characteristics and the biological activity [23]. This method has been broadly employed in virtual screening [24].

We opt for an alternative approach. The QSAR of ARIs is determined via a novel neural networks-based approach. Two molecular descriptors, electronegativity and molar volume, are selected, and calculated by ab initio Hartree–Fock (HF) method. It is known that neural networks has been widely applied in engineering, finance and medicine [25,26]. Neural networks is employed here to determine the QSAR of ARI candidates, i.e., the relationship between their bioactivities and their electronegativities and molar volumes.

The electronegativity of a chemical group is related to both its hydrophobic property and its ability to form hydrogen bonds with surrounding molecules, and it is usually considered a very important factor for describing biological system properties [21,27,28]. Volume is a direct measure of the steric hindrance effect of a functional group. Because the size of active site is limited, the size of ligand is of course an important factor in an enzymatic reaction. The two descriptors are calculated by ab initio HF method with 6-311G(d, p) basis set, and the energies are calculated with zero point energy correction. These calculations are carried out by Gaussian 98 program [29]. The data for various functional groups are listed in Table 1.

In this study, the error-back propagation neural networks (BPNN) [25] is used to calculate the QSAR for ARIs. The neural network depicted in Fig. 2 is characterized by equations as follows:

$$Z = \sum_{j=1,2} Wy_j \mathrm{Sig}\left(\sum_{i=1,n} Wx_{ij}x_i\right) \tag{2}$$

$$\mathrm{Sig}(\upsilon) = \frac{1}{1+\exp(-a\upsilon)} \tag{3}$$

where Z is the output, Wy_j the weight connected hidden neuron with output, and Wx_{ij} is the weight connected input with hidden neurons. The neural networks has the $n-2-1$ structure. n is the number of input neurons that includes descriptors and a bias, and n is equal to 9 in this work, since substituents may be introduced at four sites, 5′, 6′, 7′, and 8′ in Fig. 1a. For each site there are two descriptors taken as input neuron, besides that there is one more bias. Therefore, there are nine input neurons in total. Two is the number of hidden neurons; and one is the output for the bioactivity.

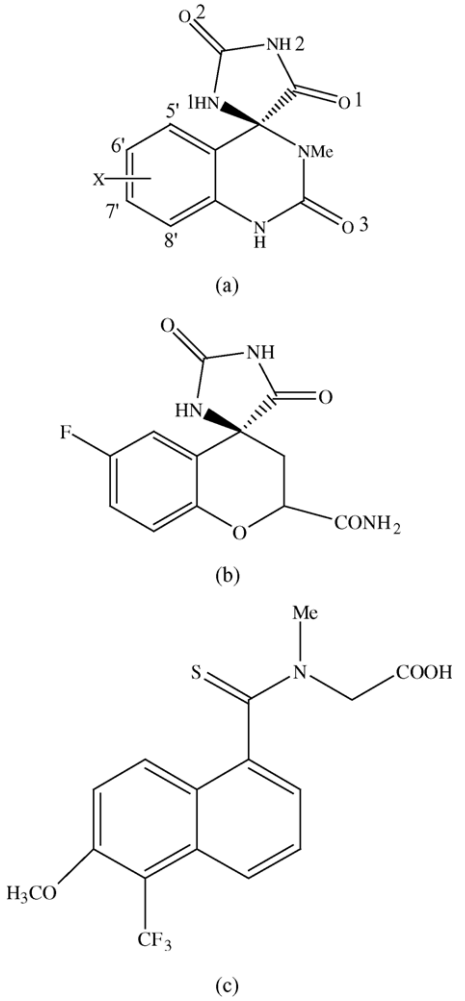

Fig. 1. The structure of (a) spiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-triones [19], (b) fidarestat, and (c) tolrestat.

There are total 30 spirohydantoin derivatives with measured IC_{50} (the concentration of the inhibitor required to produce 50% inhibition of the enzymatic activity in vitro) and the mean inhibition percentage of polyol accumulation in sciatic nerve in vivo [19]. In this series of experimental data, there are 16 mono-substituted, 11 di-substituted, and 2 tri-substituted compounds. Three of these compounds are with 5′-substitute, 21 of them are with 6′-substitute, 12 of them are with 7′-substitute, and 7 of

Table 1
Parameters of functional groups

Numbers	Functional groups	Electron-negativity[a] (eV)	Volume[b] (cm^3/mol)
1	H	5.748	8.55
2	2-Thienyl	4.662	64.90
3	3-4-Methylenedioxo	4.265	41.87
4	Morpholino	3.472	71.49
5	CH_3	3.266	22.00
6	C_2H_3	3.337	25.56
7	C_2H_5	2.631	43.41
8	CH_2–COOH	4.304	51.20
9	C_2H_5–OC═O	3.015	46.38
10	C_3H_7	2.209	57.04
11	C_4H_9	2.187	67.52
12	C_5H_{11}	2.190	73.61
13	C_6H_5	5.758	54.57
14	4-$C_6H_4NO_2$	4.504	74.03
15	4-C_6H_4OH	3.593	73.38
16	4-C_6H_4Br	3.978	86.44
17	C_6H_5COO	6.159	113.71
18	CH_2–C_6H_4-4-OH	2.336	85.72
19	CH_2–C_6H_4-4-OMe	2.227	112.20
20	CH_2–C_6H_3-3,4-OH	2.523	91.78
21	$(CH_2)_2$–C_6H_4-4-OH	1.627	93.32
22	C_6H_5CH═CHCOO	5.347	109.56
23	$(CH_2)_3$–O–Ph	2.394	105.19
24	$(CH_2)_3$–C_6H_4-4-OH	2.106	123.06
25	2-CH_2–$C_{10}H_7$	2.696	104.46
26	$(C_6H_5)_2$–CH	2.557	148.55
27	CN	9.158	16.03
28	COOH	3.636	31.11
29	COO-*t*-Bu	6.322	92.16
30	CF_3	4.194	25.98
31	NH_2	4.751	12.90
32	NHAC	3.558	56.14
33	N–$(CH_3)_2$	3.482	45.53
34	NO_2	5.308	23.86
35	OH	6.732	10.63
36	OCH_3	3.814	35.97
37	OCH_2CHCH_2	3.831	48.38
38	$OCH_2CH_2CH_3$	3.025	50.44
39	OCO-n-Hexane	3.915	118.88
40	O–$(CH_2)_2$–OH	3.227	47.88
41	O–$(CH_2)_3$–OH	3.140	67.17
42	O–$(CH_2)_3$–OCH_2Ph	3.179	137.15
43	OCO–$(CH_2)_3$–$CHMe_2$	3.895	113.36
44	OCO–$(CH_2)_2$–COCHMePh	4.142	154.37
45	OCF_3	4.999	33.15
46	F	9.767	10.31
47	Cl	8.325	21.33
48	Br	7.591	25.60
49	SCH_3	5.005	31.79
50	SO_2CH_3	4.989	50.89

[a] Electronegativity = 1/2(ionization potential + electron affinity).
[b] Molar volume.

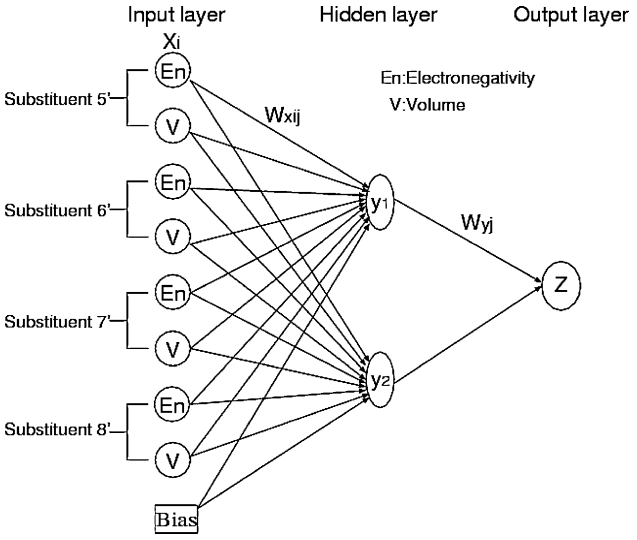

Fig. 2. Neural networks structure.

Table 2
The IC_{50} and inhibition percent of polyol accumulation of the training and testing data

Numbers	Compounds	Exp.[a]	NN calc.[b]	Exp.[c]	NN calc.[d]
1	H	0.325	0.266	2.447	2.007
2	5′-Cl	0.175	0.129	2.934	2.803
3	5′-Me	0.275	0.210	2.602	2.602
4	6′-Cl	0.680	0.682	1.748	1.814
5	6′-Br	0.747	0.670	1.716	1.751
6	6′-NO_2	0.155	0.200	1.863	2.166
7	6′-Ome	0.136	0.191	3.653	2.528
8	6′-OCOPh	0.123	0.186	3.301	3.443
9	6′-OCOCH═CHPh	0.293	0.185	3.792	3.676
10	6′-CO_2Et	0.222	0.191	2.380	2.591
11	6′-Me	0.309	0.192	2.903	2.576
12	7′-Cl	0.223	0.143	1.973	1.969
13	7′-Me	0.375	0.358	2.342	2.422
14	7′-OMe	0.210	0.187	2.556	2.509
15	8′-F	0.299	0.277	1.623	1.701
16	5′-Cl,6′-NO_2	0.000	0.001	2.672	3.002
17	6′,7′-Cl_2	0.529	0.546	1.869	1.912
18	6′-Cl,7′-Me	0.580	0.478	1.477	1.617
19	6′-NO_2,7′-Cl	0.077	0.068	2.079	2.132
20	6′-NO_2,7′-morpholino	0.103	0.002	2.633	2.616
21	6′,7′-OCH_2O-	0.284	0.253	3.255	3.176
22	6′,8′-Cl_2	0.631	0.615	1.568	1.518
23	6′-Cl,8′-NO_2	0.292	0.230	1.663	1.724
24	6′-Cl,8′-NH_2	0.446	0.432	2.114	2.103
25	6′,8′-NO_2	0.141	0.129	2.415	2.489
26	6′,8′-Cl_2,7′-OMe	0.375	0.374	1.875	1.785
27	6′,7′,8′-F_3	0.299	0.309	2.146	2.172
28	6′-F	0.587	0.682[e]	2.000	2.017
29	7′-NO_2	0.115	0.257[e]	2.681	2.298[e]
30	6′-Cl,7′-OMe	0.246	0.222	1.771	1.720[e]

[a] Experimental values of the inhibition proportion of polyol accumulation in sciatic nerve.
[b] Neural networks predicted values of the inhibition proportion of polyol accumulation in sciatic nerve.
[c] The experimental values of log(IC_{50}).
[d] Neural networks predicted values of log(IC_{50}).
[e] The data in the testing set.

them are with 8′-substitute. Usually in vivo the characteristics of bioactivity are different from that in vitro, and in this experiment the vivo data is more closely related to lipophilicity required to penetrate into the nerve. In neural networks calculation, we determine the QSAR from these 30 measured data (shown in Table 2), among which 28 data are used for the training set and two data for the testing set. To ensure validity of QSAR calculation, we perform cross-validation for the neural network. Cross-validation was done by picking one of the 28 molecules for validation at a time. The training results of both IC_{50} and the mean inhibition of polyol accumulation are plotted in Fig. 3(a) and (b), and are shown to agree well with the experimental data. Because there are only three compounds with 5′-substitute, we also perform neural network calculation for data with three sites 6′, 7′, and 8′. The neural network calculation results for these 27 data are shown in Fig. 3(c) and (d). This study focuses on the results of neural network calculation for compounds with four sites 5′, 6′, 7′ and 8′. As shown in Fig. 3(a) and (b) and Table 2, the training results agree well with the experimental data for both IC_{50} and mean decreasing polyol level, and the test data has similar RMSE as that of the training set. For training data, the range of experimental IC_{50} data is from 1.5 to 3.8, the root mean square error (RMSE) of calculated IC_{50} is 0.26, which is 11% of the range; and the range of experimental inhibition of polyol level is from 0 to 0.75, and its RMSE of calculated inhibition of polyol levels in sciatic nerve is 0.06, which is 8% of the range. The calculated data demonstrate the validity of the determined neural network for the QSAR.

We then utilize the QSAR to predict the values of IC_{50} and the inhibition percent of polyol level for new compounds by substituting various functional groups at different sites of the spiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-triones [19]. Potent inhibitors are selected from millions of new predicted compounds according to the following criteria: (1) 30 nM $\leq IC_{50} \leq$ 40 nM and (2) polyol level of sciatic nerve is decreased by more than 65% (all the bioactivity values are constrained within the range of the training set), and (3) each

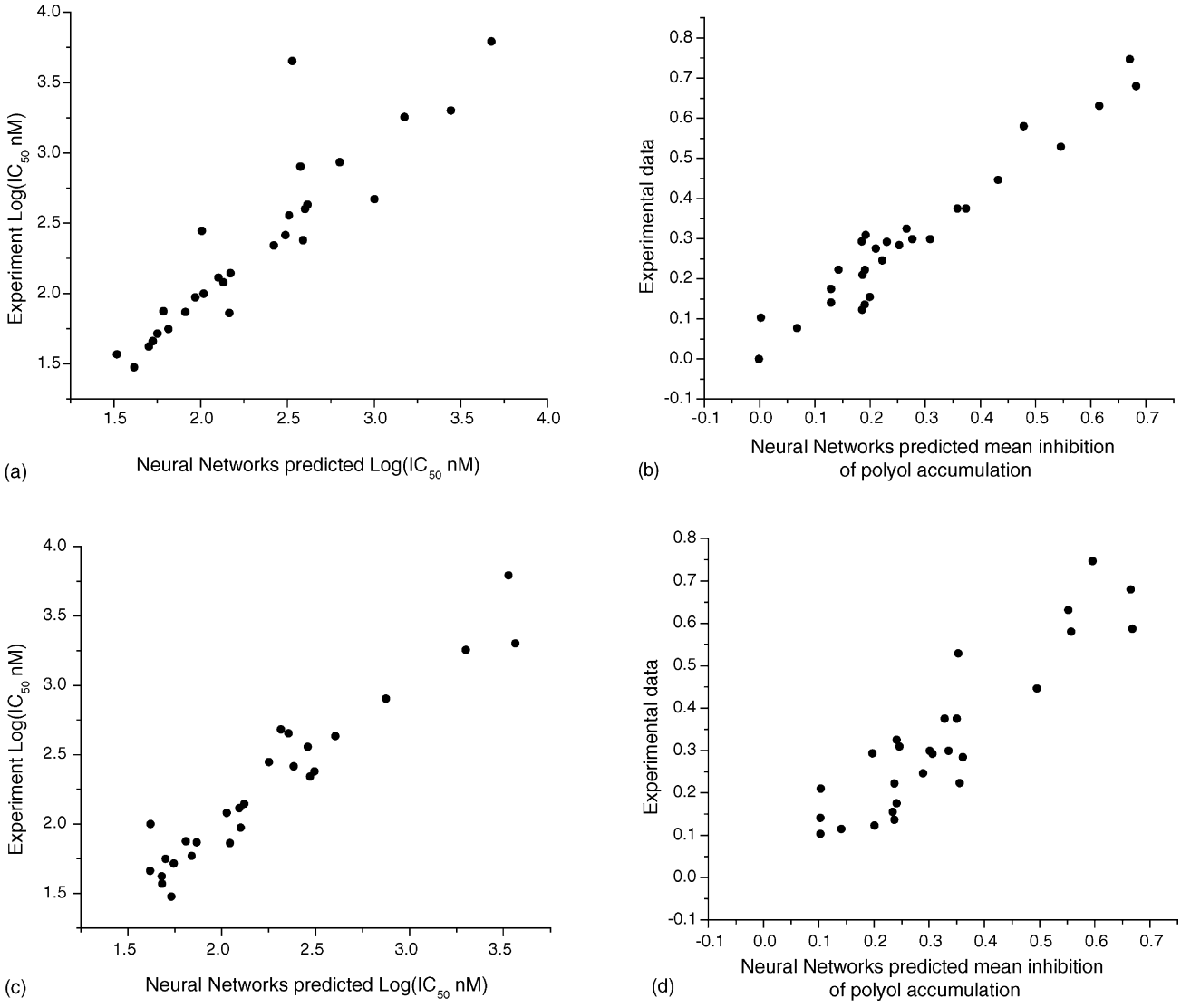

Fig. 3. Comparison between the experimental bioactivities and neural networks prediction: (a) log(IC_{50}) and (b) mean inhibition of polyol accumulation are obtained by drug lead with 5′-, 6′-, 7′- and 8′-sites; (c) log(IC_{50}) and (d) mean inhibition of polyol accumulation are obtained by drug lead with 6′-, 7′- and 8′-sites.

Table 3
Structure, calculated bioactivity values and scores of predicted ARIs

Numbers	Structure	IC_{50}[a] (nM)	Percentage[b] (%)	$E_{binding}$, Autodock (kcal/mol)	Ligand-score	Dock-score	$E_{binding}$, QM/MM (kcal/mol)	Total score
1	6′-Cl,8′-OH	38	71.6	−8.88	4.90	42.22	−30.90	2.72
2	6′-Cl,7′-NH_2,8′-Cl	35	65.6	−9.12	4.87	47.19	−43.87	3.16
3	6′-Cl,7′-NH_2,8′-OH	34	69.7	−8.80	5.03	43.34	−32.58	2.81
4	6′-CN,7′-C_2H_3,8′-CN	33	71.0	−8.89	4.97	45.10	−39.58	2.98
5	6′-CN,7′-C_2H_5,8′-F	35	66.9	−8.79	5.22	45.73	−31.34	2.94
6	6′-CN,7′-CF_3,8′-CN	32	70.3	−8.60	5.59	45.40	−35.14	3.07
7	6′-CN,7′-CH_3,8′-CN	32	71.7	−8.92	4.86	46.76	−28.67	2.84
8	6′-CN,7′-Cl,8′-Cl	31	65.1	−9.23	5.12	47.00	−30.74	3.09
9	6′-CN,8′-Cl	31	72.5	−9.01	4.80	45.47	−24.64	2.74
10	6′-CN,7′-NH_2,8′-Cl	32	70.5	−8.99	4.94	45.81	−27.12	2.84
11	6′-CN,7′-NH_2,8′-OH	32	70.1	−8.49	5.01	43.11	−31.47	2.67
12	6′-CN,7′-NO_2,8′-CN	31	68.8	−7.97	5.51	46.43	−37.06	2.87
13	6′-CN,7′-OCF_3,8′-CN	32	67.0	−8.67	5.49	48.97	−31.05	3.11
14	6′-CN,7′-OH,8′-Cl	32	71.7	−8.79	4.73	46.13	−35.07	2.82
15	6′-CN,7′-N-2CH_3,8′-F	34	65.1	−8.80	5.35	49.34	−36.52	3.19
16	6′-CN,7′-COOH,8′-CN	33	68.5	−8.27	5.55	47.36	−36.40	3.02
17	6′-CN,7′-COOEt,8′-F	35	65.6	−9.20	5.97	53.48	−40.25	3.77
18	6′-CN,7′-SCH_3,8′-CN	32	67.7	−8.84	5.16	47.22	−30.57	2.97
19	6′-F,7′-C_2H_3,8′-CN	31	72.5	−8.92	4.91	44.66	−29.69	2.81
20	6′-F,7′-C_2H_3,8′-F	31	71.9	−9.09	5.11	48.56	−42.23	3.26
21	6′-F,7′-C_2H_5,8′-F	32	70.6	−9.02	5.14	47.96	−43.67	3.25
22	6′-F,7′-CF_3,8′-CN	31	72.4	−8.45	5.08	47.91	−41.64	3.00
23	6′-F,7′-CF_3,8′-F	30	70.0	−8.68	5.38	42.65	−34.47	2.91
24	6′-F,7′-CH_3,8′-CN	31	73.8	−8.74	5.07	47.67	−41.42	3.08
25	6′-F,7′-CH_3,8′-F	30	70.4	−8.96	5.02	46.48	−43.28	3.13
26	6′-F,7′-Cl,8′-Cl	31	65.6	−8.90	5.05	47.90	−42.09	3.14
27	6′-F,7′-Cl,8′-CN	31	71.9	−8.97	5.24	43.01	−32.74	2.94
28	6′-F,8′-Cl	31	73.2	−8.76	4.91	46.06	−40.63	2.96
29	6′-F,8′-CN	30	73.0	−8.57	4.95	45.94	−39.46	2.89
30	6′-F,7′-3-4-methylenedioxo,8′-F	31	71.2	−8.68	5.23	49.67	−36.29	3.11
31	6′-F,7′-Br,8′-CN	30	72.3	−8.77	5.09	48.35	−43.15	3.15
32	6′-F,7′-CN,8′-CN	38	68.1	−8.60	5.08	47.17	−38.85	2.98
33	6′-F,7′-NH_2,8′-Cl	31	70.7	−8.87	5.01	46.78	−45.20	3.15
34	6′-F,7′-NH_2,8′-CN	30	74.8	−8.65	5.05	46.86	−43.90	3.06
35	6′-F,7′-NH_2,8′-OH	31	70.1	−8.48	4.94	41.59	−33.78	2.62
36	6′-F,7′-NO_2,8′-CN	30	73.3	−8.56	5.41	45.51	−38.97	3.04
37	6′-F,7′-OCF_3,8′-CN	31	68.4	−8.35	5.43	47.86	−32.84	2.96
38	6′-F,7′-OCF_3,8′-F	30	71.0	−8.37	5.31	45.79	−35.59	2.90
39	6′-F,7′-$OCH_2CH_2CH_3$,8′-F	33	65.4	−8.77	5.33	50.52	−33.85	3.17
40	6′-F,7′-OCH_2CHCH_2,8′-F	32	67.6	−8.73	5.27	52.81	−48.46	3.44
41	6′-F,7′-OCH_3,8′-CN	32	65.2	−8.47	5.13	50.32	−45.95	3.17
42	6′-F,7′-OCH_3,8′-F	31	72.7	−8.69	5.08	48.68	−30.42	2.94
43	6′-F,7′-OH,8′-Cl	32	72.5	−8.72	5.04	42.11	−27.14	2.66
44	6′-F,7′-OH,8′-CN	31	71.9	−8.49	5.01	41.97	−33.01	2.66
45	6′-F,7′-SO_2CH_3,8′-F	31	65.8	−8.71	5.41	49.47	−32.09	3.12
46	6′-F,7′-N (CH_3)$_2$,8′-F	32	69.5	−9.22	5.13	49.62	−45.17	3.40
47	6′-F,7′-CH_2COOH,8′-F	32	65.3	−9.24	5.20	51.67	−50.72	3.58
48	6′-F,7′-COOH,8′-CN	31	69.4	−8.35	5.55	46.23	−39.94	3.06
49	6′-F,7′-COOH,8′-F	31	72.7	−8.96	5.29	48.67	−31.10	3.12
50	6′-F,7′-O2(CH_2)OH,8′-F	32	67.8	−8.86	5.37	51.19	−44.80	3.41
51	6′-F,7′-EtOCO,8′-F	32	68.8	−8.73	5.44	51.95	−35.56	3.28
52	6′-F,7′-SCH_3,8′-CN	31	69.4	−8.76	5.29	50.52	−41.77	3.28
53	6′-F,7′-SCH_3,8′-F	30	70.6	−9.02	5.56	47.51	−34.89	3.27
54	5′-F,6′-CN,8′-Cl	32	69.6	−9.31	4.84	44.85	−20.65	2.78
55	5′-F,6′-F,8′-Cl	33	70.2	−9.17	4.97	40.44	−30.56	2.79
56	5′-F,6′-F,8′-CN	32	72.8	−8.90	4.76	45.57	−39.26	2.92
57	5′-NH_2,6′-CN,8′-Cl	33	71.0	−9.29	4.69	38.40	−26.14	2.58
58	5′-NH_2,6′-F,8′-Cl	34	71.5	−9.17	4.86	39.74	−26.71	2.66
59	5′-NH_2,6′-F,8′-CN	32	73.6	−8.93	4.83	38.85	−31.22	2.61
60	5′-OH,6′-CN,8′-Cl	32	71.2	−9.10	5.07	42.86	−17.90	2.69
61	5′-OH,6′-F,8′-Cl	32	71.7	−9.04	4.94	40.90	−25.58	2.67
62	5′-OH,6′-F,8′-CN	31	73.2	−8.79	4.96	40.64	−32.23	2.68

[a] Concentration that causes 50% inhibition in experiment.
[b] The inhibition percent of polyol level in sciatic nerve.

molecule should have no more than three non-hydrogen functional groups at 5′-, 6′-, 7′- and 8′-sites. According to these criteria, 62 possible ARIs are determined. These molecules will be subjected to further screening by computer docking experiment and quantum mechanical calculation to confirm their bindings to AR.

The QSAR is analyzed by examining the predicted ARIs' structure and compared with QSAR study by Yamagishi et al. In Table 3, among the 62 predicted ARIs, all identified functional groups at this 6′-site are hydrogen acceptors with small volume and large electronegativity. It is shown that the 6′-substituent of spiro[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-triones [19] is favorable. In their study, Yamagishi et al. found that (1) hydrophobic and electron-withdrawing substituents are favorable for high activity and (2) a smaller 6′-substituent is more active. They also indicated the bioactivity of compounds with 6′-amine and hydroxyl substituted is not favorable. Our prediction shows that there are absolutely no such functional groups at the site among the molecules with good bioactivities (see Table 2) [19]. In the functional group parameter database (Table 1), apparently F, CN, and Cl have larger electronegativity and smaller volume comparing with other functional groups. Hence, the predicted potent ARIs only contain these functional groups at the 6′-site. The 7′-site can accommodate larger functional groups containing hydrophobic functional groups (CH_3, C_2H_3, C_2H_5, etc.) as compared to the 6′- and 8′-sites. These results are consistent with the experimental results of Yamagishi et al. [19] and the QSAR obtained by the CoMFA method [30]. Because there are few data in the 5′- and 8′-positions, we need to analyze closely the predicted structures. The 8′-substituent is a preferable electronegative position, because these potent ARI candidates only have the functional groups such as F, OH, CN and Cl at this site. The 5′-substituent is very close to the spirohydantoin, due to steric hindrance, only small functional groups are suitable. Moreover, H of OH can form intra-molecular hydrogen bond with O of spirohydantoin, which stabilizes the inhibitor. It might be the reason that 5′-site is preferable to have OH functional group. In original experiments, Yamagishi et al. could not find anything important either for deciding the type of functional groups at 5′- and 8′-site. They concluded that 5′- and 8′-sites contain the functional groups with varying hydrophobic and electronic properties, and thus do not play important roles in binding. As analyzed above, the QSAR obtained by neural networks agree well with experimental results. The reason is that the hydrogen bonding and hydrophobic interactions are the most important interactions between ARIs and AR.

3. Virtual screening via docking of ARIs into active site and QM/MM calculation of binding energy

We report a new neural networks-based method for designing aldose reductase inhibitors. As many exists methods, our strategy has its limitations. For instance, the backbone of the protein is rigid during the docking. The computer docking is carried out by employing two softwares, Cerius2 ligand-fit (Accelrys Inc.) [31] and Autodock3.0.5 [32], and the binding energies of ARI candidates are calculated by using CHARMm QM/MM method [33–40].

The conformations of inhibitors within the AR active site are obtained by docking the ligands into the active site. The protein structure, human AR crystal structure (1PWM) together with a known inhibitor fidarestat (see Fig. 1b), is downloaded from Brookhaven Protein Data Bank, and its resolution is 0.92 Å. Fidarestat has similar structure as our predicted ARIs. The polypeptide chain of AR contains 315 residues with a molecular mass of 35.8 kDa [2]. The charges and potentials are assigned by CHARMm force field in InsightII 2000 [33]. Because the water molecules at the active site prevent the inhibitors to enter into the site, they are removed by using Cerius2 Structure Based Design module [31]. The active site of AR is quite hydrophobic and consists of residues Tyr48, Trp20, Lys77, His110, Phe122, Trp219, Cys298, Leu300, Val47 and Trp111 [2–6,9]. As shown in Fig. 4, the active site (shown by green cylinder) is located at the center of holoenzyme, and most residues surrounding the active site have hydrophobic rings, which results in the hydrophobic active site. Those hydrophobic rings are parallel with each other due to the hydrophobic interaction.

We find that all the predicted potent compounds dock well into the active site. The best docking conformation of each of these compounds, i.e., with the highest value of ligand-score in Cerius2 docking calculation (for each molecule 10 best docking conformations are saved in the docking calculation), are subjected further to QM/MM calculation of binding energy. All of scores including ligand-score, Docking score calculated by ligand-fit, the estimated free binding energy calculated by Autodock3.0.5 and interaction energy by QM/MM are listed in Table 3. Two predicted inhibitors in Table 3, nos. 17 and 47 are

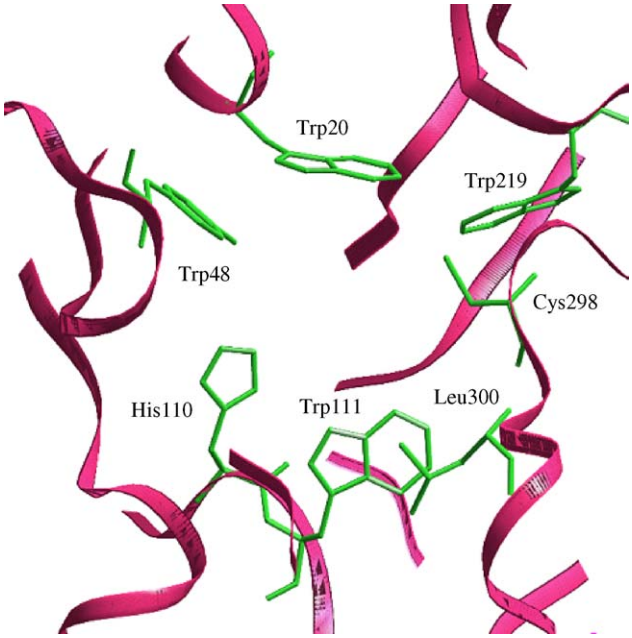

Fig. 4. Stereo-view of aldose reductase active site. Protein backbone is shown in pink ribbon. The active residues are shown in green cylinders.

shown in ball and stick in Fig. 5. Inhibitor no. 17 (see Fig. 5a) has hydrogen bonds (black dashed line) with four protein residues shown in cylinder, Trp20, Leu300, Leu301 and Ser302. It shows that O2 of spirohydantoin is in contact with Hε1 of Trp20 (2.09 Å), F of 8′-substituent with HN of Leu300 (2.14 Å), and two oxygen atoms of 7′-COOEt with HN of Leu301 (2.38 Å) and HN of Ser302 (2.441 Å), respectively. The hydrophobic ring is parallel to the inhibitor site residues, Trp20 and Trp219. In Fig. 5b, no. 47 has different orientation from no. 17 because of their different lengths. To have the most stable conformation, the hydrophilic head O1, O2 and H2 of no. 47 become in contact with Trp111 (Hε1: 2.25 Å), Tyr48 (Hh: 2.751 Å) and His110 (Nε2: 2.14 Å) through the hydrogen bonds. The hydrophobic ring is also parallel to the rings of Trp20 and Trp219. It can also be observed in Fig. 5b that the space around catalytic residues His110 and Tyr48 [6,9] is small, and the 6′-substituent points to the residues His110 and Tyr48. Therefore, there is no more room for a large substituent, and this is the reason why the potent inhibitors do not contain large 6′-substituents. These two orientations are the most favorite for most of the predicted inhibitor candidates. With Autodock3.0.5, the free binding energies are calculated and listed in Table 3 and the docked conformations obtained are consistent with those obtained using Cerius2. The comparison of inhibitor docking conformations by two docking softwares is shown in Fig. 6. The docking results show that if the steric hindrance of its functional groups on phenol ring is not too large, an inhibitor would have the same orientation as that of no. 47 and otherwise the other orientation is preferred. It was reported by Urzhumtsev et al. [6] that two contact zones are involved in the binding of ARIs. Some ARIs such as tolrestat (see Fig. 1c) may bind with a ‘specificity’ pocket, which is created by a conformational change of enzyme [6]. This specificity pocket indicates the adaptability of the AR active site and only opens for certain inhibitors, then we analyzed all the docking conformations of ligands, and found that the functional groups in the phenol ring contributes much conformational variation by adjusting the size of functional groups. Most of binding researches suggested the N and O on head of spirohydantoins contact with catalytic residues His110 and Tyr48. However, in this research we observed that inhibitor 17 adopts an unusual conformation although its binding energy is not low. This series of compounds have a methyl group on pyran ring instead of H or F single atom functional group on sorbinil or fidarestat in the corresponding position. Because the

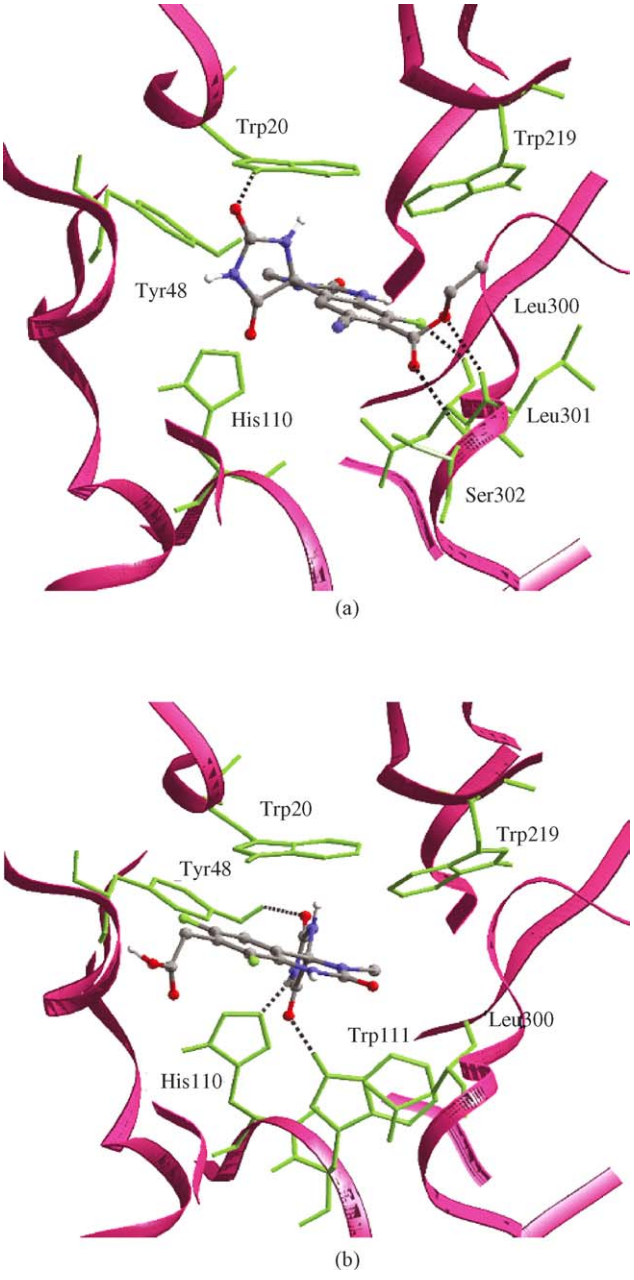

Fig. 5. The docking conformations for inhibitors no. 17 (shown in (a)) and no. 47 (shown in (b)). Protein backbone is shown in pink ribbon. Active site residues are shown in green cylinder, inhibitors are shown in cylinders, hydrogen bonds are shown in black dash line (carbon: grey, oxygen: red, nitrogen: blue, hydrogen: white, halogen: green).

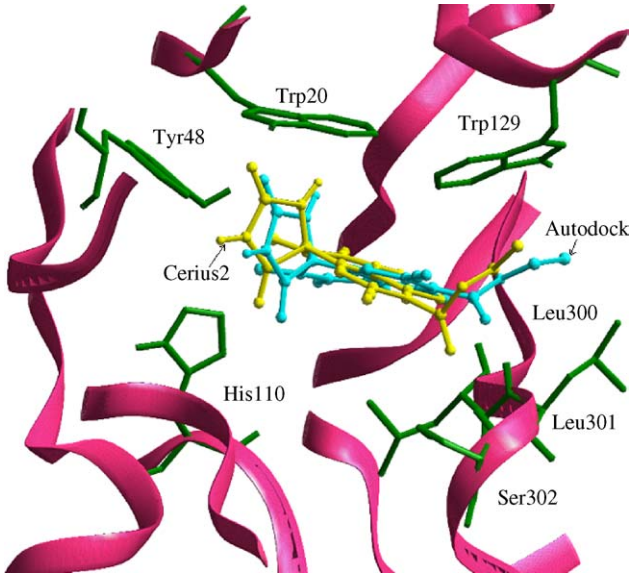

Fig. 6. Comparison of the docking conformations of inhibitor 17 by Cerius2 (yellow cylinder) and Autodock (cyan cylinder). Protein color scheme is same as Fig. 5.

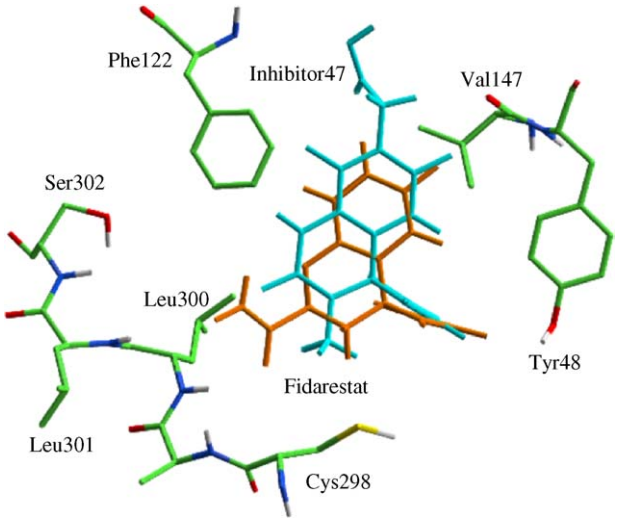

Fig. 7. Side view of inhibitor 47 and fidarestat docking conformations (protein residues shown in atomic color—carbon: light green, oxygen: red, nitrogen: blue, hydrogen: white; inhibitor 47: cyan; Fidarestat: orange).

space around residues His110 and Tyr48 is limited and methyl group is bulky group comparing with hydrogen, the position of the compound (cyan cylinder shown in Fig. 7) is tilted up by introducing methyl group on the pyran ring. This makes 6′-site closer to the residues VAL47. It is shown in Fig. 7 the distances between O of Val47 and 6′-Cl of inhibitor 47 and F of fidarestat are 2.79 and 3.21 Å, respectively; on the other side, the distances between S of Cys298 and a H of methyl group and the H of fidarestat in corresponding position are 2.81 and 2.95. The inhibitor 17 contains CN at the 6′-site, which is a bigger functional group than F or Cl, if the conformation is same as fidarestat (orange cylinder shown in Fig. 7), the N atom would be in the distance with large repulsion energy. Therefore, these molecules choose another conformation. Their binding energies are depending on hydrogen bonds; the more the hydrogen bonds, the larger the binding energies. Since this kind of molecules occupy the active site and have good binding with protein, they also can prevent the substrate from interacting with protein.

The inhibitors' binding energies are calculated by CHARMm QM/MM method [33–40]. The coordinates are taken from Cerius2 docking results. Single point energies are calculated for the protein with docked ligands (the conformations are ligand-fit docking results), protein and ligands, respectively. AM1 is used for quantum mechanical method. The quantum part refers all of ligand atoms and the molecular mechanical calculation includes all of the protein atoms. The binding energies are finally calculated by equation $E_{\text{binding}} = E_{\text{qmmm}} - (E_{\text{mm}} + E_{\text{qm}})$ and the values of QM/MM binding energies are listed in Table 3 as well.

4. Discussion and conclusion

So far although molecular mechanics plays a very important role in studying biological system, no molecular mechanical method can give comparably accurate values of energies. And different molecular mechanical method has its own advantages. By combining multiple energy calculation methods, we might consider the interaction energy more comprehensively to improve the reliability of the calculation results. We scale all types of scores including ligand-score, dock-score of ligand-fit, estimated binding free energy of Autodock and binding energies of QM/MM to the values within 0.5–1.0, and sum the scaled values of the scores up to obtain a total score for each molecule (listed in Table 3). These scaled values are simply determined by linear scaling method. Data in different range was mapped to same range. Because we consider the grads of original data is small, the range was only set to 0.5–1.0. We set the minimum value as 0.5 and the maximum value as 1.0 for different data range, and the scaled value of other data calculated by the equation

$$x_{\text{in}} = \frac{(a_{\text{mx}} - a_{\text{mn}})x_{\text{inp}} + a_{\text{mn}}x_{\text{max}} - a_{\text{mx}}x_{\text{min}}}{x_{\text{fac}}}$$

where x_{in} is the scaled value, x_{inp} the original input value, a_{mx} the maximum of the scaled value, a_{mn} the minimum of the scaled value, x_{max} the maximum of the input value, x_{min} the minimum of the input value, and $x_{\text{fac}} = x_{\text{max}} - x_{\text{min}}$. The relationships between the total score and some individual scores are shown in Fig. 8: (a) ligand-score, (b) dock-score and (c) QM/MM binding energy. It is shown in Fig. 8, three individual

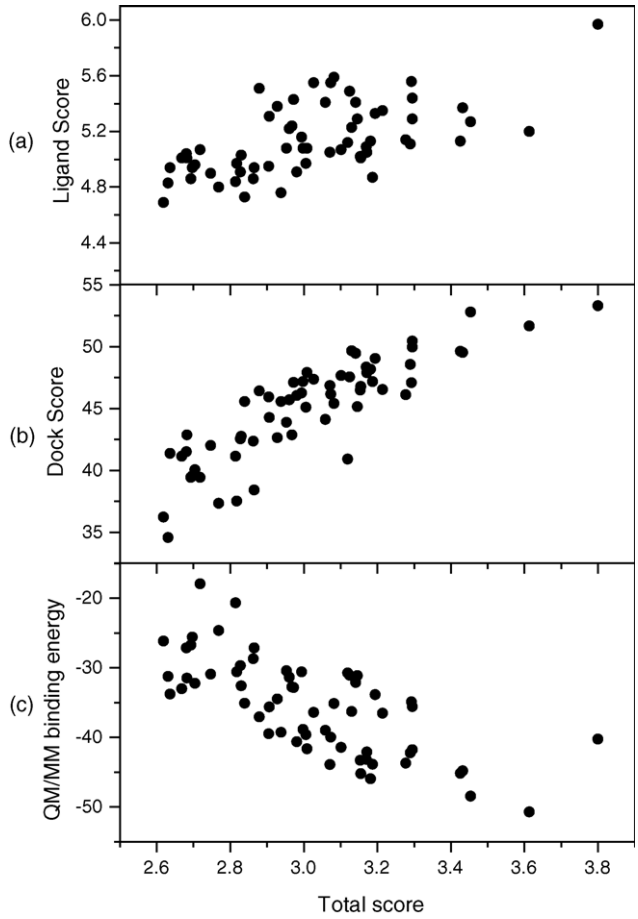

Fig. 8. The relationship between the total score and individual scores: (a) ligand-score, (b) dock-score, and (c) binding energy calculated via QM/MM method (in unit of kcal mol^{-1}).

scores, ligand-score, dock-score and QM/MM binding energy have the same trend as the total score. To compare calculated binding activity with experimental data, we calculated the various binding scores of the best experimental binding ARI (IC_{50} = 30 nM), (4*R*)-6′-chloro-7′-methyl-3′-methyl-spiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione (no. 18 in Table 2), and found that its total score is 2.78. In Table 3, inhibitors no. 17, (4*R*)-6′-cyano-7′-ethoxyl-carbonyl-8′-fluoro-3′-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione, and no. 47, (4*R*)-6′,8′-fluoro-7′-methylene-carboxylic-acid-3′-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione, have the highest total scores (3.77 and 3.58, respectively). There are 49 predicted molecules with better total scores than that of (4*R*)-6′-chloro-7′-methyl-3′-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione (see Fig. 1a and Table 3). It is found in Table 3 that the average binding energies of compounds with the 5′-substituent are comparably lower than those of other compounds. On the other hand, the compounds with the 6′- or 8′-substituents such as F, Cl and CN functional groups usually have higher binding energies than other compounds. The docking conformations are analyzed. We observe that the 6′- and 8′-sites always interact with the polar residues of protein in both two orientations (shown in Fig. 5). In Fig. 5a the 8′-site of no. 17 contacts with the polar residue Leu300. In Fig. 5b the 6′-site of no. 47 is near to polar site which consists of several polar hydrogen atoms such as Val47(HN), Gln49(HN) and Tyr48(HN). The corresponding electrostatic interactions between the 6′-site of ligand and the active site may strengthen the interaction between ligand and protein. Because it is near to the 6′- and 8′-sites, the 7′-site points to the hydrogen donor site (Leu300 and Leu301) and forms hydrogen bonds with those residues in Fig. 5a, therefore, the electron donor functional groups at this site stabilize this conformation. But in Fig. 5b the 7′-site points to the entrance of the active site, where it is far away from the active site residues and thus cannot form the hydrogen bond with active site residues, so this site is less effective than that in Fig. 5a. We thus conclude that: (1) the 6′and 8′-sites are the preferable binding sites, (2) the large functional groups in 5′-site may weaken the binding affinity, and (3) the 7′-substituent is less effective than the 6′ and 8′-substituents. The promising potent inhibitors should have functional groups with large electronegativity and small volume and the substituents at the 6′ and 8′-sites are preferred. To obtain potent inhibitors in both vitro and vivo, the rank of the total scores is further compared with that of the inhibition percent of polyol level. It indicates that the compounds with better binding activity in vitro may however have lower activity in vivo. In Table 3, inhibitors no. 20, (4*R*)-6′,8′-fluoro-7′-vinyl-3′-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione, no. 21, (4*R*)-6′,8′-fluoro-7′-ethyl-3v-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione and no. 46, (4*R*)-6′,8′-fluoro-7′-*N*,*N*-dimethyl-amine-3′-methylspiro-[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-trione, have both high total scores 3.26, 3.25 and 3.40 and predicted high inhibition percent of polyol level in vivo 71.9, 70.6 and 69.5%, respectively. These compounds are the most promising candidates for ARIs.

Acknowledgements

We thank Professor Stephen Chung and Dr. Pauline Chiu for very helpful discussions. This work was supported by the Generic Drug Research Program of the University of Hong Kong.

References

[1] N. Ruderman, J. Williamson, M. Brownlee, Hyperglycemia, Diabetes, and Vascular Disease, Oxford University Press, New York, 1992.
[2] D.K. Wilson, K.M. Bohre, K.H. Gabbay, F.A. Quiocho, An unlikely sugar substrate site in the 1.65 A structure of the human aldose reductase holoenzyme implicated in diabetic complications, Science 257 (1992) 81–84.
[3] K.M. Shaw, Diabetic Complications, John Wiley and Sons Ltd., Baffins Lane, Chichester, England, 1996.
[4] B. Draznin, S. Melmed, D. LeRoith, Molecular and Cellular Biology of Diabetes Mellitus, Vol. III: Complication of Diabetes Mellitus, Alan R. Liss Inc., New York, 1989.
[5] P.F. Kador, J.H. Kinoshita, N.E. Sharpless, Aldose reductase inhibitors: a potential new class of agents for the pharmacological control of certain diabetic complications, J. Med. Chem. 28 (1985) 841–849.
[6] A. Urzhumtsev, F. Tete-Favier, A. Mitschler, J. Barbanton, P. Barth, L. Urzhumtseva, J.-F. Biellmann, A.D. Podjarny, D. Moras, A 'specificity' pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors torestat and sorbinil, Structure 5 (1997) 601–612.
[7] P.F. Kador, N.E. Sharpless, Pharmacophor requirements of the aldose reductase inhibitor site, Mol. Pharmacol. 24 (1983) 521–531.
[8] J.G. Cannon, Pharmacology for Chemists, Oxford University Press, New York, 1999.
[9] P. Varnai, A. Warshel, Computer simulation studies of the catalytic mechanism of human aldose reductase, J. Am. Chem. Soc. 122 (2000) 3849–3860.
[10] Y.S. Lee, Z. Chen, P.F. Kador, Molecular modeling studies of the binding modes of aldose reductase inhibitors at the active site of human aldose reductase, Bioorg. Med. Chem. 6 (1998) 1811–1819.
[11] D.K. Wilson, I. Tarble, J.M. Petrash, F.A. Quiocho, A refined 1.8 Å structure of human aldose reducatse complexed with the potent inhibitor zopolrestat, Proc. Natl. Acad. Sci. 90 (1993) 9847–9851.
[12] K.M. Bohren, C.E. Grimshaw, The sorbinil trap: a predicted dead-end complex confirms the mechanism of aldose reductase inhibition, Biochemistry 39 (2000) 9967–9974.
[13] D. Amic, D. Davidovic-Amic, D. Beslo, The use of the ordered orthogonalized multivariate linear regression in a structure–activity study of coumarin and flavonoid derivatives as inhibitors of aldose reductase, J. Chem. Inf. Comput. Sci. 37 (1997) 581–586.
[14] Y. Iwata, M. Arisawa, R. Hamada, Y. Kita, M.Y. Mizutani, M. Tomioka, A. Itai, S. Miyamota, Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis, J. Med. Chem. 44 (2001) 1718–1728.
[15] M. Schlitzer, L. Rodriguez, P.F. Kador, Synthesis of potential aldose reductase inhibitors based on minimal pharmacophore requirements, J. Pharm. Pharmacol. 53 (2001) 831–839.
[16] L. Costantino, G. Rastelli, M.C. Gamberini, M.C. Gamberini, M.P. Giovannoni, V.D. Piaz, P. Vvianello, D. Barlocco, Isoxazolo-[3,4-*d*]-pyridazin-7-(6H)-one as a potential sbstrate for new aldose reductase inhibitors, J. Med. Chem. 42 (1999) 1894–1900.
[17] J. Angel de la Fuente, S. Manzanaro, Aldose reductase inhibitors for natural sources, Nat. Prod. Rep. 20 (2003) 243–251.
[18] K. Nakao, M. Asao, H. Shirai, K. Saito, T. Moriya, H. Iwata, M. Matsumota, Y. Matsuoka, R. Shimizu, Benzoxazoline and benzimdazoline derivatives as novel aldose reductase inhibitors. Part 2. Lead optimization, Med. Chem. Res. 9 (1999) 631–642.
[19] M. Yamagishi, Y. Yamada, K. Ozaki, M. Asao, R. Shimizu, M. Suzuki, M. Matsumoto, Y. Matsuoka, K. Matsumoto, Biological activities and quan-

titative structure–activity relationship of spiro[imidazolidine-4,4′(1′*H*)-quinazoline]-2,2′,5(3′*H*)-triones as aldose reductase inhibitors, J. Med. Chem. 35 (1992) 2085–2094.

[20] M. Oka, Y. Matsumoto, S. Sugiyama, N. Tsuruta, M. Matsushima, A potent aldose reductase inhibitors, (2*S*,4*S*)-6-fluoro-2′,5′-dioxospiro[chroman-4,4′-imidazolidine]-2-carboxamide (Fidarestat): its absolute configuration and interactions with aldose reductase by X-ray crystallography, J. Med. Chem. 43 (2000) 2479–2483;
R. Sarges, J. Bordner, B.W. Dominy, M.J. Peterson, E.B. Whipple, Synthesis, absolute configuration, and conformation of the aldose reductase inhibitor sorbinil, J. Med. Chem. 28 (1985) 1176–1720.

[21] C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani, E.J. Lien, "Aromatic" substituent constants for structure–activity correlations, J. Med. Chem. 16 (1973) 1207;
C. Hansch, T. Fujita, ρ–σ–π analysis. A method for the correlation of biological activity and chemical structure, J. Am. Chem. Soc. 86 (1964) 1616–1626.

[22] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[23] B. Skagerberg, D. Bonelli, S. Clementi, G. Cruciani, C. Ebert, Principal properties for aromatic substituents a multivariate approach for design in QSAR, Quant. Struct.–Act. Relat. 8 (1989) 32–38.

[24] J.J. Böhm, G. Schneider, Virtual Screening for Bioactive Molecules: Methods and Principles in Medicinal Chemistry, Wiley–VCH Verlag GmbH, Weinheim, 2000.

[25] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating correlations, Nature 333 (1986) 533–536;
S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, Upper Saddle River, 1999.

[26] G. Schneider, P. Wrede, Artificial neural networks for computer-based molecular design, Prog. Biophys. Mol. Biol. 70 (1998) 175–222.

[27] G.A. Jeffrey, An Introduction to Hydrogen Bonding, Oxford University Press, Oxford, 1997.

[28] K. Hasegawa, T. Deushi, H. Yoshda, Y. Miyashita, S. Sasaki, Chemometric QSAR studies of antifungal azoxy compounds, J. Comput. Aid. Mol. Des. 8 (1994) 449–456.

[29] J.B. Foresman, A. Frisch, Exploring Chemistry with Electronic Structure Methods, Gaussian Inc., 1996.

[30] K. Nakao, M. Asao, H. Shirai, R. Shimizu, 3D-pharmacophore analyses of aldose reductase inhibitory spiroquinazolinones, Drug Des. Disc. 16 (1999) 155–163.

[31] Cerius2 Ligandfit, Accelrys Inc., San Diego, CA, 2003.

[32] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian Genetic Algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.

[33] InsightII 2000, CHARMm Principles, Accelrys Inc., San Diego, CA, 2002.

[34] Y.S. Lee, M. Hodoscek, B.R. Brooks, P.F. Kador, Catalytic mechanism of aldose reductase studied by the combined potentials of quantum mechanics and molecular mechanics, Biophys. Chem. 70 (1998) 203–216.

[35] M.J. Field, P.A. Bash, M. Karplus, A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations, J. Comput. Chem. 11 (1990) 700–733.

[36] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[37] D.H. Harrison, K.M. bohren, D. Ringe, G.A. Petsko, D.H. Gabbay, An anion binding site in human aldose reductase: mechanistic implications for the binding of citrate, cacodylate, and glucose 6-phosphate, Biochemistry 33 (1995) 2011–2020.

[38] P.A. Bash, M.J. Field, M. Karplus, Free energy perturbation method for chemical reactions in the condensed phase: a dynamical approach based on a combined quantum and molecular mechanics potential, J. Am. Chem. Soc. 109 (1987) 8092–8094.

[39] P.J. Goodford, A computational procedure for determining energetically favorable binding sites on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.

[40] A.D. MacKerell Jr., B. Brooks, C.L. Brooks III, L. Nilsson, B. Roux, Y. Won, M. Karplus, CHARMM: the energy function and its parameterization, with an overview of the program, in: P.v.R. Schleyer, et al. (Eds.), The Encyclopedia of Computational Chemistry, vol. 1, John Wiley and Sons, Chichester, 1998, pp. 271–277.