Supporting Information

Scaffold hopping of potential anti-tumor agents by WEGA: a shape-based approach

Hu Ge[1], Yu Wang[1], Wenxia Zhao[2], Wei Lin[1], Xin Yan[1], and Jun Xu[1,*]

[1] School of Pharmaceutical Sciences, Sun Yat-Sen University, 132 East Circle Road at University City, Guangzhou, 510006, China.

[2] Department of Pharmacy, The Sun Yat-sen Memorial Hospital of Sun Yat-sen University, Guangzhou 510120, China

HepG2 Cell Viability Assays

Cell viability was determined using the Cell Counting Kit-8 (CCK-8) assay based on water-soluble tetrazolium salt (WST)-8. Like thiazolyl blue tetrazolium bromide or MTT, WST-8 can be catalyzed by mitochondrial dehydrogenases to yield a yellow, soluble Formazan dye. This dye is then used to estimate the number of viable cells. Briefly, cells were seeded in 96-well culture plates at a density of 2,000 per well, and were allowed to attach overnight. After serum starvation for 24 hours, the cells were treated in triplicate with graded concentrations of compounds for 48 hours. The cells were then incubated with 10% CCK-8 prepared in DMEM without phenol red for 1 hour at 37 °C. The optical density of each well was read using a plate reader (model VERSA Max, Molecular Devices) at a wavelength of 450 nm. Wells with drug-containing media without cells or untreated cells were used as negative and positive-controls, respectively. The inhibitory activity was expressed as the compound concentration required for 50% growth inhibition of cancer cells (IC_{50}). IC_{50} was calculated by the Logit method. The mean IC_{50} was determined from the results of the three independent tests.

WEGA

The Weighted Gaussian Algorithm (WEGA) for molecular shape similarity computation is described as follows: the shape density of a molecule is expressed as a linear combination of weighted atomic Gaussian functions as

$$G(r) = \sum_i w_i g_i(r) = \sum_i w_i p e^{-(\frac{3p\pi^{1/2}}{4\sigma_i^3})^{2/3}(r-r_i)^2}$$

where w_i is a weighting factor that can be determined by a simple formula. A good measurement for the crowdedness is the total overlap of an atom with all others. Therefore we introduce a simple empirical formula for calculating atomic Gaussian weights:

$$w_i = \frac{v_i^g}{v_i^g + k\sum_{j \neq i} v_{ij}^g}$$

where k is a universal constant determined by fitting the self-overlapping volumes to the hard-sphere volumes for a set of diverse molecules. In this new method, the overlap volume of two molecules becomes

$$V_{AB}^g = \sum_{i \in A,\, j \in B} w_i w_j v_{ij}^g$$

When the two molecules are identical, the above equation becomes the expression for the self-overlap volume of the molecule, and we want to make the value computed by this equation match the molecule's hard-sphere volume.

Like the ROCS-color[1] method, WEGA, as applied for molecular overlaying purposes, can be easily extended to include feature contributions so that one can put weights on both shape and pharmacophore features, so as to account for these features in alignments. Compared to some recent shape-feature based alignment methods, such as ShaEP[2], Phase-Shape[3], and SHAFTS[4], a major advantage of using Gaussian methods for shape-feature based alignment is that all contribution terms are treated consistently and the analytical first and second derivatives can be obtained easily so that the very robust Newton-Raphson method can be used for the optimization of the alignment[5].

Table S2 The activity data of the 36 compounds of the CoMFA model.

ID	pIC50
compound 01	4.6126
compound 02	4.4841
compound 03	4.8013
compound 04	3.8697
compound 05	4.433
compound 06	4.5436
compound 07	3.6123
compound 08	2.45
compound 09	3.9905
compound 10	4.8416
compound 11	5.3737
compound 12	0.4858
compound 13	4.2581
compound 14	4.1811
compound 15	3.3172
compound 16	0.6186
sysu-20064S	1.0877
sysu-20069S	0.6882
sysu-20152S	4.1475
sysu-20215S	3.7825
sysu-20218S	4.3219
sysu-20229S	3.1226
sysu-20254S	3.3076
sysu-20308S	0.6412
sysu-20309S	3.4428
sysu-20385S	4.7133
sysu-20529S	3.9066
sysu-20530S	4.7721
sysu-20532S	4.7595
sysu-20611S	4.7905
sysu-20727S	0.7659
sysu-20784S	4.466
sysu-20785S	3.1264
sysu-20913S	5.3565
sysu-22128S	3.6861
sysu-22977S	4.2062

Note: For the compounds whose precise IC50 values could not be determined, multiple sets of random pIC50 values, ranging from 0 to 3.976 were generated. Multiple CoMFA models were built and the one with the best cross-validation correlation coefficient value was chosen as the final model.

Table S2 The names and scaffolds of the 12 active hits.

Compound ID	Name	Scaffold
sysu-20152S	10-methoxyiminostilbene	Dibenzoazepine
sysu-20215S	2-(6-chloro-9*H*-carbazol-2-yl)propanoic acid	Carbazole
sysu-20218S	8-chloro-11-(4-methylpiperazin-1-yl)-5*H*-dibenzo[b,e][1,4]diazepine	Dibenzoazepine
sysu-20385S	10-oxo-10,11-dihydro-5*H*-dibenzo[b,f]azepine-5-carboxamide	Dibenzoazepine
sysu-20529S	4-hydroxycarbazole	Carbazole
sysu-20530S	4-glycidyloxycarbazole	Carbazole
sysu-20532S	1-((9*H*-carbazol-4-yl)oxy)-3-((2-(2-methoxyphenoxy)ethyl)amino)propan-2-ol	Carbazole
sysu-20611S	2-(4-(3-(2-chloro-10*H*-phenothiazin-10-yl)propyl)piperazin-1-yl)ethan-1-ol	Phenothiazine
sysu-20784S	iminostilbene carbonyl chloride	Dibenzoazepine
sysu-20913S	2-(4-(3-(2-(trifluoromethyl)-10*H*-phenothiazin-10-yl)propyl)piperazin-1-yl)ethan-1-ol	Phenothiazine
sysu-22128S	10*H*-phenoxazine	Phenoxazine
sysu-22977S	N,N,2-trimethyl-3-(10*H*-phenothiazin-10-yl)propan-1-amine	Phenothiazine

Table S3 The cytotoxic effect on Lo2 cells of the top five hits.

Compounds	IC50 (μM)
sysu-20913S	>160
sysu-20611S	>160
sysu-20530S	>160
sysu-20532S	85±8.2
sysu-20385S	>160

Table S4 The anti-proliferative activity of 20 random control compounds.

Compound ID	IC50 (μM)
sysu-00092	>160
sysu-00621	>160
sysu-10142N	>160
sysu-10278N	>160
sysu-10422N	>160
sysu-10442N	>160
sysu-20122S	>160
sysu-20268S	>160
sysu-20323S	>160
sysu-21757S	>160
sysu-21830S	130 ± 8.2
sysu-22252S	62 ± 5.7
sysu-22306S	>160
sysu-22382S	>160
sysu-22397S	>160
sysu-22857S	>160
sysu-22920S	>160
sysu-24926S	>160
sysu-24957S	>160
sysu-25112S	>160

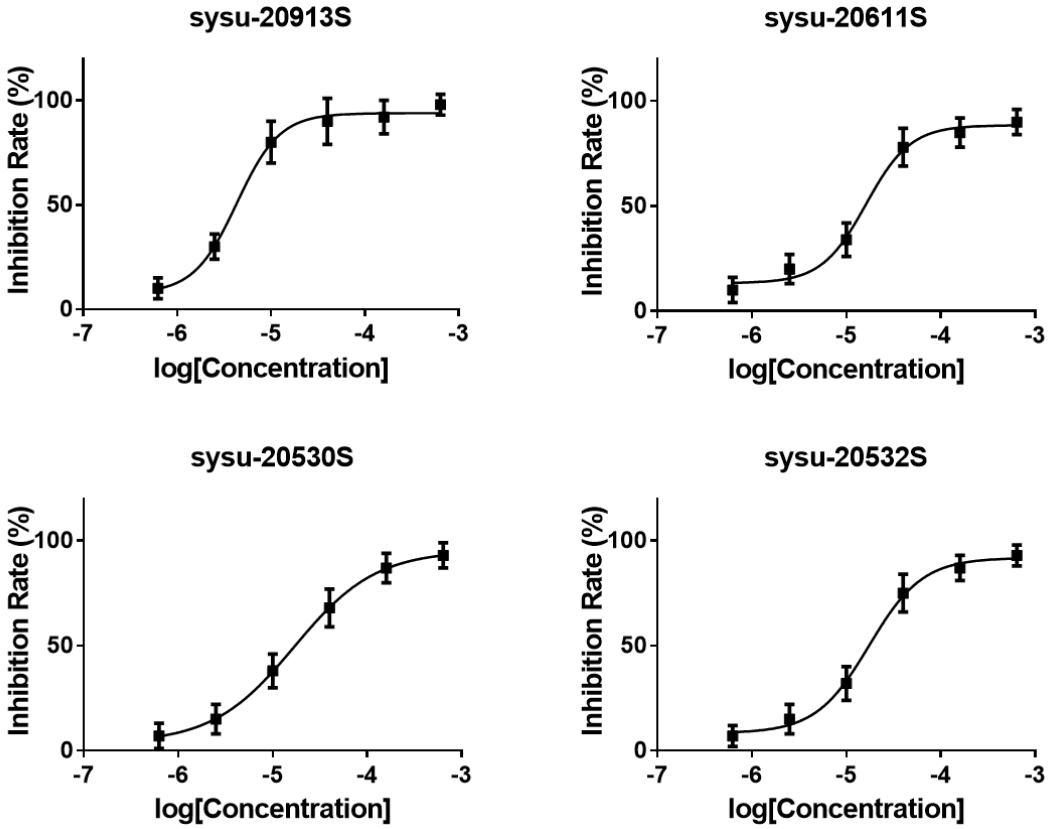

Fig. S1 The response curves of the four identified compounds with highest anti-proliferative activity on HepG2 cells.

References

(1) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *Journal of medicinal chemistry* **2007**, *50*, 2007.

(2) Vainio, M. J.; Puranen, J. S.; Johnson, M. S. ShaEP: molecular overlay based on shape and electrostatic potential. *Journal of chemical information and modeling* **2009**, *49*, 2009.

(3) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *Journal of chemical information and modeling* **2011**, *51*, 2011.

(4) Liu, X.; Jiang, H.; Li, H. SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. *Journal of chemical information and modeling* **2011**, *51*, 2011.

(5) Yan, X.; Li, J.; Liu, Z.; Zheng, M.; Ge, H.; Xu, J. Enhancing molecular shape comparison by weighted Gaussian functions. *Journal of chemical information and modeling* **2013**, *53*, 2013.