MedChemComm

CONCISE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Med. Chem. Commun.*, 2014, **5**, 737

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

Hu Ge,‡[a] Yu Wang,‡[a] Wenxia Zhao,[b] Wei Lin,[a] Xin Yan[a] and Jun Xu*[a]

Received 27th December 2013
Accepted 18th March 2014

DOI: 10.1039/c3md00397c

www.rsc.org/medchemcomm

In this paper, we describe the first prospective application of the shape-comparison program, WEGA (weighted Gaussian algorithm), to find new scaffolds for anti-tumor agents. A series of sixteen carbazole alkaloids extracted from *Clausena vestita* D. D. Tao, which have anti-tumor activities at the cellular level, were used as query molecules. A compound library was screened by ranking molecules based upon their 3D shape and pharmacophore similarities to known inhibitors. The relationship between the structures and activities was also studied through comparative molecular field analysis (CoMFA). Twelve hits show comparable growth inhibition activity against HepG2 cells (a hit rate of 60%); eight of the hits have new scaffolds (in comparison with known inhibitors). These results indicate that a shape-based screening approach, such as WEGA, can be efficiently used for scaffold hopping in a lead identification process.

Introduction

In lead optimization, systematic decoration of a common scaffold and bioisosteric replacement are the main techniques of structural variation. Scaffold hopping is an approach to generate novel chemical entities from known chemical frames. Since finding new druggable scaffolds is a bottle-neck in the pharmaceutical development process, a new scaffold hopping technology is demanded.[1] In this paper, we propose a new scaffold hopping technique that utilizes WEGA (weighted Gaussian algorithm),[2] a shape comparison program.

WEGA aligns chemical structures based upon three-dimensional shape and pharmacophore features. It is suitable for large-scale virtual screening with single or multiple bioactive compounds as the query "templates," regardless of whether corresponding experimentally determined conformations are available.

In our previous study, sixteen anti-cancer compounds were isolated from *Clausena vestita*, and their anti-tumor activities were evaluated *via* cell growth inhibition assays in our labs.[3] Those carbazole alkaloids share almost the same scaffolds. In this study, to identify potential anti-tumor agents based on the known inhibitors, and to discover new scaffolds, we employed WEGA for ligand-based scaffold hopping by screening the Guangdong small molecule tangible library (GSMTL).[4] The results were confirmed through exactly the same *in vitro* experiments. A high hit rate and various new scaffolds demonstrate the applicability of scaffold hopping *via* WEGA. Furthermore, to understand the three-dimensional quantitative structure–activity relationship (3D-QSAR),[5] a comparative molecular field analysis (CoMFA) study was performed based on our results; CoMFA analysis is widely used for lead optimization when ligand-based information is available.[6–13] We hope that this study provides another alternative approach for scaffold hopping and will inspire the development of new anti-tumor agents.

We reported a dataset of sixteen carbazole alkaloids[3] that are derived from *Clausena vestita D. D. Tao*. Their *in vitro* biological activity data were reported as IC_{50} values; these data are used in the current study (Table 1). The alkaloids' chemical structures are depicted in Fig. 1. The growth inhibition activity of the sixteen carbazole alkaloids on HepG2 cells were converted to the corresponding pIC_{50} values using the formula ($pIC_{50} = -\log IC_{50}$). Their activity was significantly affected by structural modifications within the carbazole skeleton (as revealed by the IC_{50} values of carbazole alkaloids).

Results and discussion

The virtual screening protocol by WEGA is depicted in Fig. 2. First, 3D conformations were generated for both the query template (known inhibitors) and the target compound database. Following shape-based alignment, the similarities between the query and target molecules were calculated by WEGA. Highly scored compounds exhibiting similar shape/pharmacophore features to the known inhibitors were identified. The hits were further confirmed through *in vitro* bioassays detecting anti-tumor activity.

The eleven anti-tumor compounds (no. **1–6, 9–11, 13, 14**) with exact IC_{50} values are used as query molecules in this study.

[a]*School of Pharmaceutical Sciences, Sun Yat-Sen University, 132 East Circle Road at University City, Guangzhou, 510006, China. E-mail: junxu@biochemomes.com*

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

† Electronic supplementary information (ESI) available: Details of the HepG2 cell viability assays, WEGA, and the CoMFA model. See DOI: 10.1039/c3md00397c

‡ These authors contributed equally.

Table 1 Anti-proliferative activity of the known compounds and the WEGA hits

Known compounds		Screened hits	
ID	pIC_{50}[b]	ID	pIC_{50}
1	4.61	**SYSU-20064S**	<3.80[a]
2	4.48	**SYSU-20069S**	<3.80[a]
3	4.80	**SYSU-20152S**	4.15
4	3.87	**SYSU-20215S**	3.78
5	4.43	**SYSU-20218S**	4.32
6	4.54	**SYSU-20229S**	<3.80[a]
7	<3.80[a]	**SYSU-20254S**	<3.80[a]
8	<3.80[a]	**SYSU-20308S**	<3.80[a]
9	3.99	**SYSU-20309S**	<3.80[a]
10	4.84	**SYSU-20385S**	4.71
11	5.37	**SYSU-20529S**	3.91
12	<3.80[a]	**SYSU-20530S**	4.77
13	4.26	**SYSU-20532S**	4.76
14	4.18	**SYSU-20611S**	4.79
15	<3.80[a]	**SYSU-20727S**	<3.80[a]
16	<3.80[a]	**SYSU-20784S**	4.47
		SYSU-20785S	<3.80[a]
		SYSU-20913S	5.36
		SYSU-22128S	3.69
		SYSU-22977S	4.21

[a] The mean IC_{50} value could not be determined as one or more of the corresponding data points were higher than the threshold value (160 μM). [b] $pIC_{50} = \log_{10}(1/IC_{50})$.

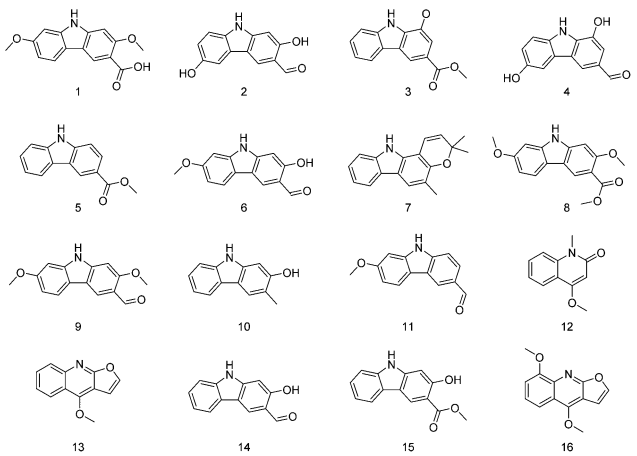

Fig. 1 Chemical structures of the sixteen alkaloids isolated from *Clausena vestita D. D. Tao*.

Their chemical structures were drawn using ChemBioDraw Ultra 13.0, and converted into 3D conformations by using the ligand preparation protocol in Discovery Studio 3.5. Twenty 3D conformations were generated by means of the CAESAR[14] algorithm in Discovery Studio 3.5 for each query molecule.

The target database used in this study is GSMTL, which contains more than 7200 annotated chemical compounds with purities >95%. Most of the compounds are natural products; the rest are synthesized. The 2D chemical structures of all compounds in the library were drawn with ISIS/Draw and stored

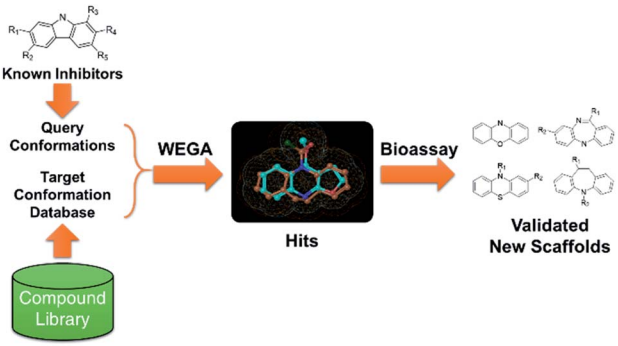

Fig. 2 Schematic representation of WEGA's workflow for database virtual screening.

in an ISIS/Base database. Conformational ensembles (maximum size of 250) were also generated for each compound in the database through the CAESAR algorithm.

GSMTL was virtually screened with WEGA based upon the 3D conformations of the query molecules. To achieve the best shape alignment for each pair of molecules, four initial alignments were considered for the superposition optimization, and the best one was selected. Both heavy atoms and hydrogens were considered for representing the molecular shape. For each compound in the target database, all its conformations were calculated for similarity with all query molecules. The highest similarity score was kept as the final score for each compound in the target library. The top 20 compounds were then selected for further bioactivity validation.

To validate whether the WEGA approach could distinguish actives from random compounds, the self-similarity between actives (known inhibitors) was also calculated (the scores between the same molecules were excluded). The distributions of the relative frequencies of the WEGA scores of both actives and random compounds (GSMTL in this case) are depicted in Fig. 3. The random compounds' scores have a normal

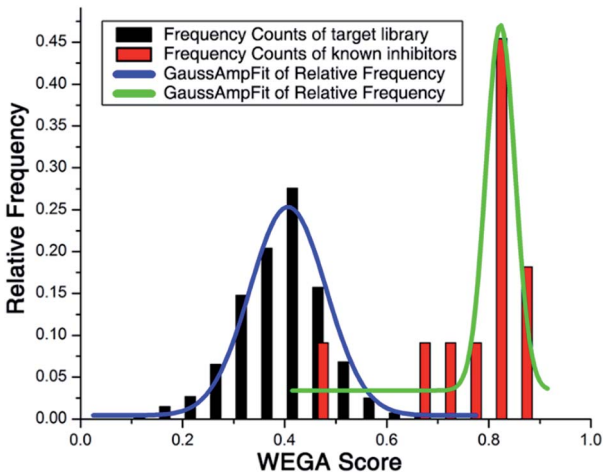

Fig. 3 The distribution of the relative frequency for the WEGA scores between known inhibitors themselves as well as scores between random compounds and known inhibitors. An amplitude version of the Gaussian peak function was used for curve fitting.

distribution; most scores range from 0.2 to 0.6, while the actives' scores distribute mainly from 0.7 to 0.9. This indicates that the WEGA similarity calculation can discriminate actives from random compounds and that the WEGA score is suitable for selecting probable inhibitors.

With WEGA, the top 20 compounds from the target library with the most similar shapes and pharmacophore features (relative to the query molecules) were retrieved (Fig. 4). Among these compounds, 4 hits have a carbazole scaffold (highlighted in Fig. 4 at the top). This scaffold was already known from the training set. Sixteen hits have new scaffolds belonging to four compound classes, which were not reported. Fig. 5 demonstrates two representative WEGA hits and their superimposed conformations.

Cell viability was determined using the Cell Counting Kit-8 (CCK-8) assay based on the water-soluble tetrazolium salt (WST)-8. The top 20 compounds with new scaffolds were screened for potential anti-tumor activities on HepG2; 12 of them were confirmed through CCK-8 experiments, with IC_{50} values ranging from 4 μM to 200 μM. Therefore, the confirmed hit rate was 60% (12/20). The experimental pIC_{50} values are also listed in Table 1. Besides the carbazole scaffold, a phenothiazine-like compound, **SYSU-20913S**, exhibits a comparably high potency (4.4 μM). The other scaffolds were also confirmed for their anti-proliferative activities. Details of the names and scaffolds of the 12 active hits are supplied in Table S1.†

The structural and activity information of all 36 (16 + 20) compounds were used to build a 3D-QSAR model *via* CoMFA. Their structures were generated and optimized through the energy minimization module in the Molecular Operating Environment (MOE) v2012.10 (Chemical Computing Group) with the MMFF94 force field. The energy minimized conformation of the most active compound **11** was chosen as the putative bioactive conformation. Then, all the other compounds were aligned, through WEGA, based on the template compound **11**. The aligned poses were further refined through the flexible alignment module in MOE and Schrodinger 2013.1. The aligned conformations are shown in Fig. 6.

The CoMFA model was built as described in ref. 15 and 16 with minor modifications. The steric (Lennard-Jones potential) and electrostatic (Coulombic potential) field energies were calculated from a standard Tripos force field. An atom having the van der Waals radius of an sp^3-hybridized carbon with one formal positive charge was used as a probe. A lattice with 0.5 Å grid spacing, and which extended at least 1 Å in each direction beyond the aligned molecules, was generated. The truncation for both steric and electrostatic energies was set at 30.00 kcal mol^{-1} and the electrostatic contributions were ignored at the lattice intersections with maximum steric interactions.

Partial least squares (PLS)[17] regression analysis was used to explore a quantitative relationship between molecular descriptors and biological activities. All regression analyses were done

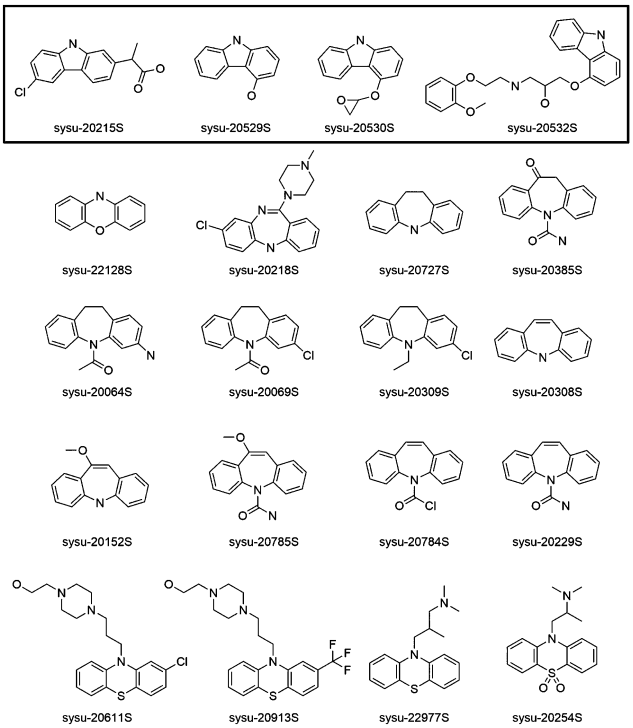

Fig. 4 Chemical structures of the compounds identified by WEGA. Compounds with known carbazole scaffolds are highlighted.

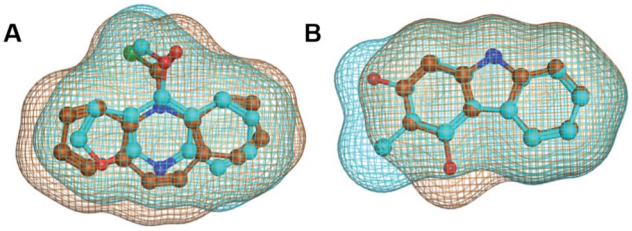

Fig. 5 Two compounds retrieved by WEGA screening and their superposed conformations. (A) **SYSU-20784** and **13** (B) **SYSU-20530** and **10**. The retrieved compounds are depicted in brown and the template molecules in cyan. The mesh surfaces indicate the molecular volumes.

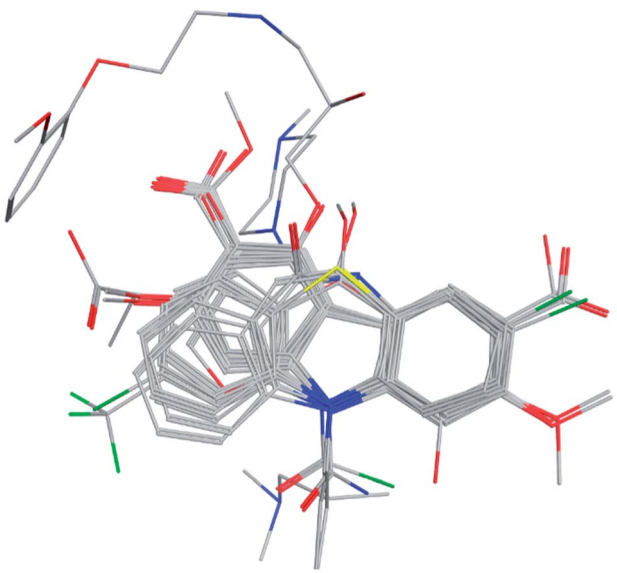

Fig. 6 The superposition of the 36 compounds for CoMFA.

in SYBYL-X 2.0. The leave-one-out (LOO) cross-validation, with a column filtering of a certain value, was performed to determine the optimum number of principal components. The cross-validated regression coefficient R_{cv}^2 suggests the robustness and predictive ability of the derived models.[18] According to a commonly recognized statistical standard, a reliable QSAR model should have an $R_{cv}^2 \geq 0.5$.[19] Then, non-cross-validation was carried out to derive the final PLS regression models with the conventional correlation coefficient r^2 and the standard error of estimate. The region focusing method[20] was performed to enhance the resolution and predictive ability of the derived model; this refined the model by increasing the weight for those lattice points pertinent to the model.

The final CoMFA model achieved a cross-validated correlation coefficient (q^2) of 0.609 with 7 principal components. The non-cross-validated PLS analysis generated a high conventional correlation coefficient (r^2) of 0.978, with a standard estimated error of 0.239. The Fischer's F value for the test of significance is 178.746. The relative contributions to this CoMFA model were 48% for the steric field and 52% for the electrostatic field. Detailed activity data are described in Table S2.†

The effects of all field descriptors contributing to bioactivity can be partitioned and viewed as CoMFA 3D coefficient contour plots as shown in Fig. 7 with compound **11** (7-methoxy-9*H*-carbazole-3-carbaldehyde) as an example. In Fig. 7A, plots in blue and red represent the positions that can be substituted with electropositive and negative groups to improve the activity. In Fig. 7B, plots in green and yellow represent the positions that can be substituted with bulky and slim groups to improve the activity.

In the case of compound **11**, if the site 1-, 2-, or 8- was substituted with an electropositive group, the activity would increase; if the site 3-, 4-, or 9- was substituted with an electronegative group, the activity would increase as well. Bulkier groups are favorable on positions 1-, 3-, and 9-, slimmer groups are good for positions 2- and 4- (Fig. 7).

To evaluate the selectivity of these agents, we estimated the cytotoxic effect of five active hits on the normal liver cell line LO2 (Table S3†). Their IC_{50} values on LO2 cells were either undetectable or much higher than those on HepG2 tumor cells, indicating a much lower cytotoxic effect on the normal cell line.

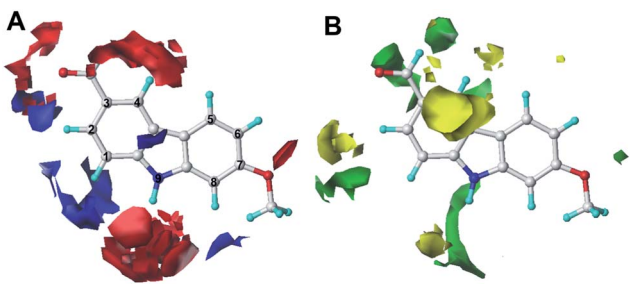

Fig. 7 The CoMFA contour maps, with compound **11** as the reference. Green and yellow refer to the sterically favored and disfavored areas, respectively, whereas blue and red contours refer to regions where electropositive substituents are favorable and unfavorable, respectively.

To determine whether this computational approach can truly enrich the yield of actives, we have conducted a control experiment. We randomly selected 20 compounds from the negative hits concluded by WEGA. Then, we tested the 20 compounds using an anti-proliferative assay. Only 2 compounds showed detectable activities. The data are listed in Table S4 (ESI†).

Conclusion

In summary, this study reported the first application of WEGA for identifying potential anti-tumor agents based on the known inhibitors; the results led to the discovery of even new scaffolds. A ligand-based scaffold hopping protocol was proved successful by the virtual screening of the GSMTL and by *in vitro* anti-tumor experiments. The high hit rate and various new scaffolds demonstrate the applicability of scaffold hopping by WEGA. Furthermore, the CoMFA study we performed provides additional information for future lead-optimization possibilities with respect to this series of anti-tumor compounds, and may inspire the development of new anti-tumor agents.

Abbreviations

WEGA	Weighted Gaussian algorithm
3D-QSAR	Three-dimensional quantitative structure–activity relationship
CoMFA	Comparative molecular field analysis
GSMTL	Guangdong small molecule tangible library
CCK-8	Cell Counting Kit-8
MOE	Molecular Operating Environment
PLS	Partial least squares
LOO	Leave-one-out

Acknowledgements

This work was funded in part by the National High-Tech R&D Program of China (863 Program) (2012AA020307), the Introduction of Innovative R&D Team Program of Guangdong Province (no. 2009010058), and the National Natural Science Foundation of China (no. 81001372, 81173470). This research was also supported in part by the Special Funding Program for the National Supercomputer Center in Guangzhou (2012Y2-00048/2013Y2-00045, 201200000037) and the Guangdong Innovative Research Team Program.

Notes and references

1 H.-J. Böhm, A. Flohr and M. Stahl, *Drug Discovery Today: Technol.*, 2004, **1**, 217–224.
2 X. Yan, J. Li, Z. Liu, M. Zheng, H. Ge and J. Xu, *J. Chem. Inf. Model.*, 2013, **53**, 1967–1978.
3 W. Lin, Y. Wang, S. Lin, C. Li, C. Zhou, S. Wang, H. Huang, P. Liu, G. Ye and X. Shen, *Eur. J. Med. Chem.*, 2012, **47**, 214–220.

4 Q. Gu, J. Xu and L. Gu, *Molecules*, 2010, **15**, 5031–5044.
5 J. Verma, V. M. Khedkar and E. C. Coutinho, *Curr. Top. Med. Chem.*, 2010, **10**, 95–115.
6 R. D. Cramer and B. Wendt, *J. Comput.-Aided Mol. Des.*, 2007, **21**, 23–32.
7 R. D. Cramer, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 197–201.
8 J. Fang, D. Huang, W. Zhao, H. Ge, H. B. Luo and J. Xu, *J. Chem. Inf. Model.*, 2011, **51**, 1431–1438.
9 V. D. Mouchlis, T. M. Mavromoustakos and G. Kokotos, *J. Chem. Inf. Model.*, 2010, **50**, 1589–1601.
10 A. Taladriz, A. Healy, E. J. Flores Perez, V. Herrero Garcia, C. Rios Martinez, A. A. Alkhaldi, A. A. Eze, M. Kaiser, H. P. de Koning, A. Chana and C. Dardonville, *J. Med. Chem.*, 2012, **55**, 2606–2622.
11 Y. Z. He, Y. X. Li, X. L. Zhu, Z. Xi, C. W. Niu, J. Wan, L. Zhang and G. F. Yang, *J. Chem. Inf. Model.*, 2007, **47**, 2335–2344.
12 H. Tang, X. S. Wang, X. P. Huang, B. L. Roth, K. V. Butler, A. P. Kozikowski, M. Jung and A. Tropsha, *J. Chem. Inf. Model.*, 2009, **49**, 461–476.
13 A. Gangjee and X. Lin, *J. Med. Chem.*, 2005, **48**, 1448–1469.
14 J. Li, T. Ehlers, J. Sutter, S. Varma-O'brien and J. Kirchmair, *J. Chem. Inf. Model.*, 2007, **47**, 1923–1932.
15 M. Liu, L. He, X. Hu, P. Liu and H. B. Luo, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 7004–7010.
16 Y. P. Li, X. Weng, F. X. Ning, J. B. Ou, J. Q. Hou, H. B. Luo, D. Li, Z. S. Huang, S. L. Huang and L. Q. Gu, *J. Mol. Graphics Modell.*, 2013, **41**, 61–67.
17 L. Stahle and S. Wold, *Prog. Med. Chem.*, 1988, **25**, 291–338.
18 B. L. Bush and R. B. Nachbar, Jr, *J. Comput.-Aided Mol. Des.*, 1993, **7**, 587–619.
19 M. Clark and R. D. Cramer, *Quant. Struct.-Act. Relat.*, 1993, **12**, 137–145.
20 S. J. Cho and A. Tropsha, *J. Med. Chem.*, 1995, **38**, 1060–1066.