Medicinal Chemistry Research (2019) 28:498–504
https://doi.org/10.1007/s00044-019-02303-x

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

4D-QSAR studies of CB_2 cannabinoid receptor inverse agonists: a comparison to 3D-QSAR

Houpan Zhang[1] · Qiaoli Lv[2] · Weidong Xu[1] · Xiaoping Lai[1] · Ya Liu[1] · Guogang Tu[1]

Received: 14 October 2018 / Accepted: 31 January 2019 / Published online: 11 February 2019

Abstract

Over the years QSAR methods have developed from 2D-QSAR to more complex 4D-QSAR which features freedom of alignment and conformational flexibility of individual ligands. This approach takes advantage of conformational ensemble profile (CEP) generated for individual compounds by molecular dynamics simulations. In present study, the 4D-QSAR methods called LQTAgrid-QSAR has been performed on a series of potent CB_2 cannabinoid receptor inverse agonists. Stepwise method was used to select the most informative variables. Partial least squares (PLS) and multiple linear regression (MLR) methods were used for constructing the regression models. Y-randomization and leave-N-out cross-validation (LNO) were carried out to verify the robustness of the model and to analysis of the independent test set. Best 4D-QSAR model provided the following statistics: $R^2 = 0.862$, $q^2_{LOO} = 0.737$, $q^2_{LNO} = 0.719$, $R^2_{Pred} = 0.884$ (PLS) and $R^2 = 0.863$, $q^2_{LOO} = 0.771$, $q^2_{LNO} = 0.761$, $R^2_{Pred} = 0.877$ (MLR). The comparison of the 4D-QSAR to 3D-QSAR was performed.

Keywords 4D-QSAR · 3D-QSAR · Molecular dynamic simulation · CB2 inverse agonists

Introduction

The endocannabinoid system is involved in a variety of pathological and physiological conditions. The main endocannabinoids are small molecules derived from 2-arachidonoyl-glycerol, arachidonic acid, and N-arachidonoyl ethanolamine (Di Marzo et al. 2004). The cannabinoid receptor belongs to the G-protein-coupled receptor superfamily (GPCRs) and includes two subtypes, namely CB_1 and CB_2, which were characterized pharmacologically in the 1990s (Matsuda et al. 1990; Munro et al. 1993). The two receptors are negatively coupled to adenylyl cyclase which inhibit cyclic AMP (cAMP) production and regulate voltage-gated N-, L- and Q- or P- type Ca^{2+} channels (Howlett et al. 2002). The CB_1 receptor is mainly located at the presynaptic nerve terminals and the central nervous system (CNS) and is also expressed in the periphery such as testis, eye, bladder, and gut. The CB_2 receptor is almost exclusively expressed in the immune cells such as NK cells, B-cells, and monocytes. The selective and high affinity agonists for CB_2 receptors without CB_1 central effects show a therapeutic potency to treat different pathologies such as ischemic stroke, pain transduction and perception, neurodegenerative diseases, severe inflammation, osteoporosis, autoimmune diseases, and cancers (Malfitano et al. 2014; Picone and Kendall 2015).

Now three-dimensional quantitative structure–activity relationship (3D-QSAR) is an important part for modern drug design which is helpful in suggesting to design new, more potent drug candidates. But there are some constraints about 3D-QSAR. Firstly it is dependent and sensitive to alignments and conformations of the compounds. Also only one conformation is considered for each compound, not a conformational ensemble profile (Andrade et al. 2010; Shim and MacKerell 2011). Secondly, the compounds must be properly aligned which may introduce user bias and is time consuming (Ghasemi et al. 2011). In order to overcome inherent problems of 3D-QSAR, the 4D-QSAR was

These authors contributed equally: Houpan Zhang, Qiaoli Lv

Supplementary information The online version of this article (https://doi.org/10.1007/s00044-019-02303-x) contains supplementary material, which is available to authorized users.

✉ Guogang Tu
tugg199@yahoo.com

[1] Department of Medicinal Chemistry, School of Pharmaceutical Science, NanChang University, 330006 Nanchang, China

[2] Department of Science and Education, JiangXi Key Laboratory of Translational Cancer Research, JiangXi Cancer Hospital, 330029 Nanchang, China

originally proposed which includes the freedom of alignment and the conformational flexibility to develop 3D-QSAR models by performing molecular state ensemble averaging, i.e., the fourth "dimension" (Hopfinger et al. 1997).

LQTA-QSAR (Martins et al. 2009), a new 4D-QSAR approach, generate the conformational ensemble profile (CEP) for each compound by molecular dynamics simulations, then calculate the 3D descriptors for a set of molecules. The major advantages of 3D-QSAR and 4D-QSAR was combined by this methodology. In this work, a 4D-QSAR model by LQTA method has been constructed on triazolo[1,5-a]pyrimidine compounds. A comparison to 3D-QSAR analysis was performed. This study is the first to report the 4D-QSAR and 3D-QSAR modeling of these compounds as CB_2 cannabinoid receptor inverse agonists.

Methods

Data set

All CB_2 cannabinoid receptor inverse agonists and the biological activities were taken from literature (Tabrizi et al. 2016). All of the compounds were divided into the training set (24 compounds) and the test set (5 compounds) randomly from original data set taking into account both the structural diversity and distribution of biological data. The chemical structures and biological activity values (IC_{50}) of the data set are presented in Fig. 1 and Table 1. The IC_{50} values in units of molarity were transformed to pIC_{50} ($-\log IC_{50}$) to provide numerically larger data values. Compounds in the training set were used to build the QSAR models, compounds in the test set were used to evaluate the predictive quality.

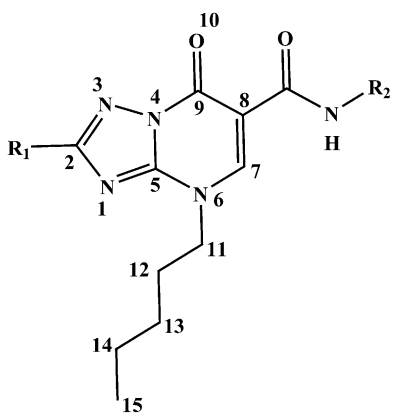

Fig. 1 The structure of data set and atoms chosen for alignment

4D-QSAR study

The 4D-QSAR was built by LQTAgrid (Ghasemi et al. 2012; Patil and Sawant 2015). In Ghemical program (Hassinen and Peräkylä 2001), the 3D structures of the data set were built. Energy minimizations were performed with the ffG43a1 force field. Then using UCSF Chimera (Pettersen et al. 2004), partial atomic charges were calculated with AMBER ff03 atom types by the AM1-BCC method. The topology files of compounds which were energy-optimized were generated by the topobuild program. By the GROMACS software (Van der Spoel et al. 2005), the MD simulations of all the compounds were performed in order to obtain conformational ensemble profile (CEP). Using the script of LQTAgrid software, MD simulation was run which included heating the system at 50, 200, and 350 K with 1 fs

Table 1 Structures of training and test set compounds

Compound	R_1	R_2	pIC_{50}
9[a]	H	Adamantan-1-yl	7.6990
10	Ph	Adamantan-1-yl	8.4437
15	H	Cycloheptyl	6.2890
16	H	3,5-dimethyladamantan-1-yl	7.6021
17	CH_3	Cyclohexyl	6.1713
18	CH_3	Adamantan-1-yl	6.4168
19[a]	Ph	Cyclohexyl	7.3665
20	Ph	Cycloheptyl	7.3188
21	Ph	3,5-dimethyladamantan-1-yl	8.0757
22	4-Cl-Ph	3,5-dimethyladamantan-1-yl	6.9318
23	4-Cl-Ph	Cyclohexyl	7.9469
24	4-OCH_3-Ph	Cyclohexyl	7.0177
25	4-OCH_3-Ph	Adamantan-1-yl	7.3279
26[a]	4-CH_3-Ph	Cyclohexyl	7.5086
27	4-CH_3-Ph	Cycloheptyl	7.6576
28	4-CH_3-Ph	Adamantan-1-yl	8.1675
29	Furan-2-yl	Adamantan-1-yl	7.3468
30	Pyridin-3-yl	Adamantan-1-yl	6.6421
31	S-CH_3	Cyclohexyl	6.1451
32[a]	S-CH_3	Adamantan-1-yl	7.3279
33	Morpholine	Cyclohexyl	6.0000
34	Morpholine	Cycloheptyl	6.2471
35	Morpholine	Adamantan-1-yl	7.2757
36	N-CH_3-piperazine	Cyclohexyl	5.0000
37	N-CH_3-piperazine	Cycloheptyl	6.0079
38	N-benzyl-N-methyl	Cyclohexyl	5.8755
39	N-benzyl-N-methyl	Adamantan-1-yl	7.6778
40[a]	Diallylamine	Cyclohexyl	5.8690
41	Diallylamine	Adamantan-1-yl	6.8013

[a]The test set compounds

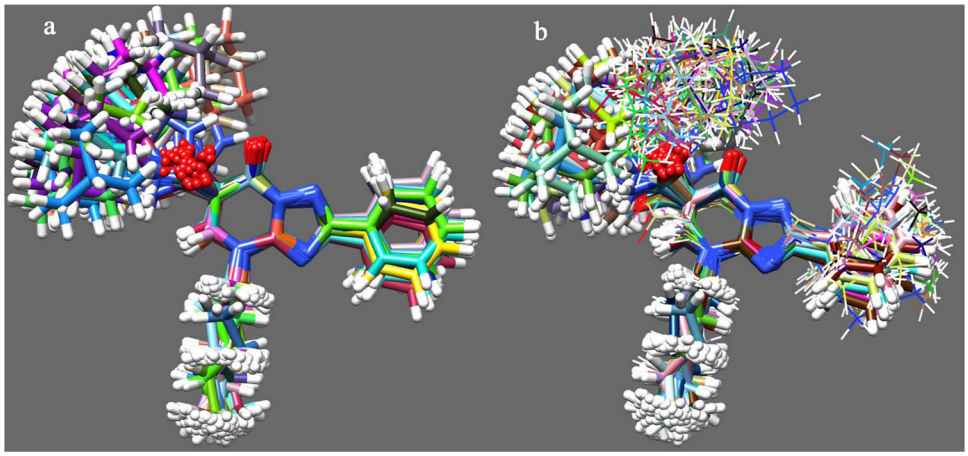

Fig. 2 The aligned CEPs resulting from MD simulations. **a** alignment of compound 10 (represented by licorice), **b** alignment of reference with the least active compound 36 (represented by line)

step size for 20 ps. Long-range electrostatics were calculated with Particle Mesh Ewald method (Darden et al. 1993). Van der Waals interaction energies were computed with a cut off radius of 1 Å. The temperature and pressure of the system was controlled by Berendsen thermostat and Parrinello-Rahman coupling respectively (Berendsen et al. 1984). The system was then cooled down to 300 K. The trajectory of the system was recorded every 2 ps simulation time, which is 2000 simulation steps.

All of the conformations that were generated in MD simulations at 300 K were aligned. Compound 10 was chosen as a reference compound because of the most active compound among all compounds. During the alignment, the intial conformer was chosen using the trajectory generated at 20 ps, then other trajectories generated up to 100 ps times with 2 ps increment each time were subjected to alignment. The atoms chosen for alignment are shown in Fig. 1. The alignments of conformers for the most active compound 10 (reference) and alignment with conformer for least active compound 36 are shown in Fig. 2.

The grid box of size 19 × 21 × 20 Å with a grid spacing of 1 Å, which was large enough to accommodate the conformers, was chosen. These aligned molecules were calculated as the energy descriptors of intermolecular interaction by LQTAgrid program. In order to compare results of the 4D-QSAR and the 3D-QSAR models, the CH_3^+ probe was selected which explores every grid point of a 1 Å grid cell lattice and computes 3D-energy interaction descriptors at each intersection. The electrostatic (Coulombic) and steric (van der Waals interactions) fields are calculated. The dimension of the descriptor matrix generated by LQTAgrid was 29 × 64,676. Descriptor matrix was refined by applying the filter of correlation coefficient. The correlation coefficient between the descriptors of interaction energy and pIC50 was calculated. In this step the descriptors with correlation coefficient lower than 0.5 were eliminated from the descriptor matrix. Furthermore, the descriptors having poor distribution when compared with pIC50 were also eliminated. With these steps, five descriptors were selected by step-wise method and were subjected to multiple linear regression (MLR) and partial least squares (PLS) analyses subsequently. Y-randomization and Leave-N-out (LNO) cross-validation were applied on the best model to examine the robustness (Kiralj and Ferreira 2009).

3D-QSAR study

As a comparison, a 3D-QSAR was constructed using the Open3DQSAR (Tosco and Balle 2011) running on UBUNTU Linux. All compounds were superimposed based on the reference compound using the atom-based alignment by the open3DALIGN tools (Tosco et al. 2011; Patel and Rajak 2018). The same training set and test set employed in 4D-QSAR was used to construct the model.

The electrostatic (Coulombic) and the steric (van der Waals) interaction energies were computed at each grid point for each compound using an alkyl carbon probe (default) with automatically assigned charges. Energies >30.0 kcal/mol and <−30.0 kcal/mol were cutoff, because a few high values may severely bias the model in the dataset. The Molecular Interaction Fields (MIFs) analyses were performed using partial least square (PLS) technique through the NIPALS algorithm methodology by open3-DQSAR tools (Wold et al. 2001). The smart region definition (SRD) and iterative variable elimination PLS (IVE-PLS) methods were used to further improve models. To determine the optimal number of components and to check the predictive ability of the 3D-QSARs models, cross-validation in PLS was carried out using the LNO method (Clark 2003). Because LNO-cross-validation is much more robust than LOO-cross-validation (Golbraikh and Tropsha 2002). We performed 20 LNO-cross-validation runs leaving 1 of 5 randomly composed groups of compounds out of the model (i.e., 20%) at each run, and predicted their activities

via the reduced model. The quality of the model is expressed with the cross-validated correlation coefficient q^2, and simultaneously the optimal number of components is decided by the largest value of q^2.

Results and discussion

4D-QSAR model

The 4D-QSAR model with MLR and PLS regressions were constructed to find a simple model between the molecular structures and observed bioactivities. By step-wise method the 5 variables were selected and imported into PLS_Toolbox and QSARINS (Gramatica et al. 2013). By the leave-one-out (LOO) cross-validation, the PLS model with three latent variables (LV) was indicated as the best model which showed $R^2 = 0.862$ and $q^2_{LOO} = 0.737$ and $R^2_{Pred} = 0.884$. The MLR model resulted in $R^2 = 0.863$, $q^2_{LOO} = 0.771$, $R^2_{Pred} = 0.877$ (Table 2). Figure 3a, b show the relationship between the predicted and experimental pIC_{50} based on PLS and MLR models, respectively. LNO-cross-validation was carried out to evaluate the robustness and reliability of the models. The average value for q^2_{LNO} was 0.719 (PLS) and 0.761 (MLR) being close to model q^2_{LOO}. In addition, Chance correlation and overfitting between the descriptors and the dependent variable were

Table 2 Summary of the statistical results for the constructed models

	4D-QSAR (PLS)	4D-QSAR (MLR)	3D-QSAR)
No. of latent variables	3	---	3
R^2_{ncv}	0.862	0.863	0.907
q^2_{lOO}	0.737	0.771	0.638
q^2_{lNO}	0.719	0.761	0.633
R^2_{Pred}	0.884	0.877	0.614

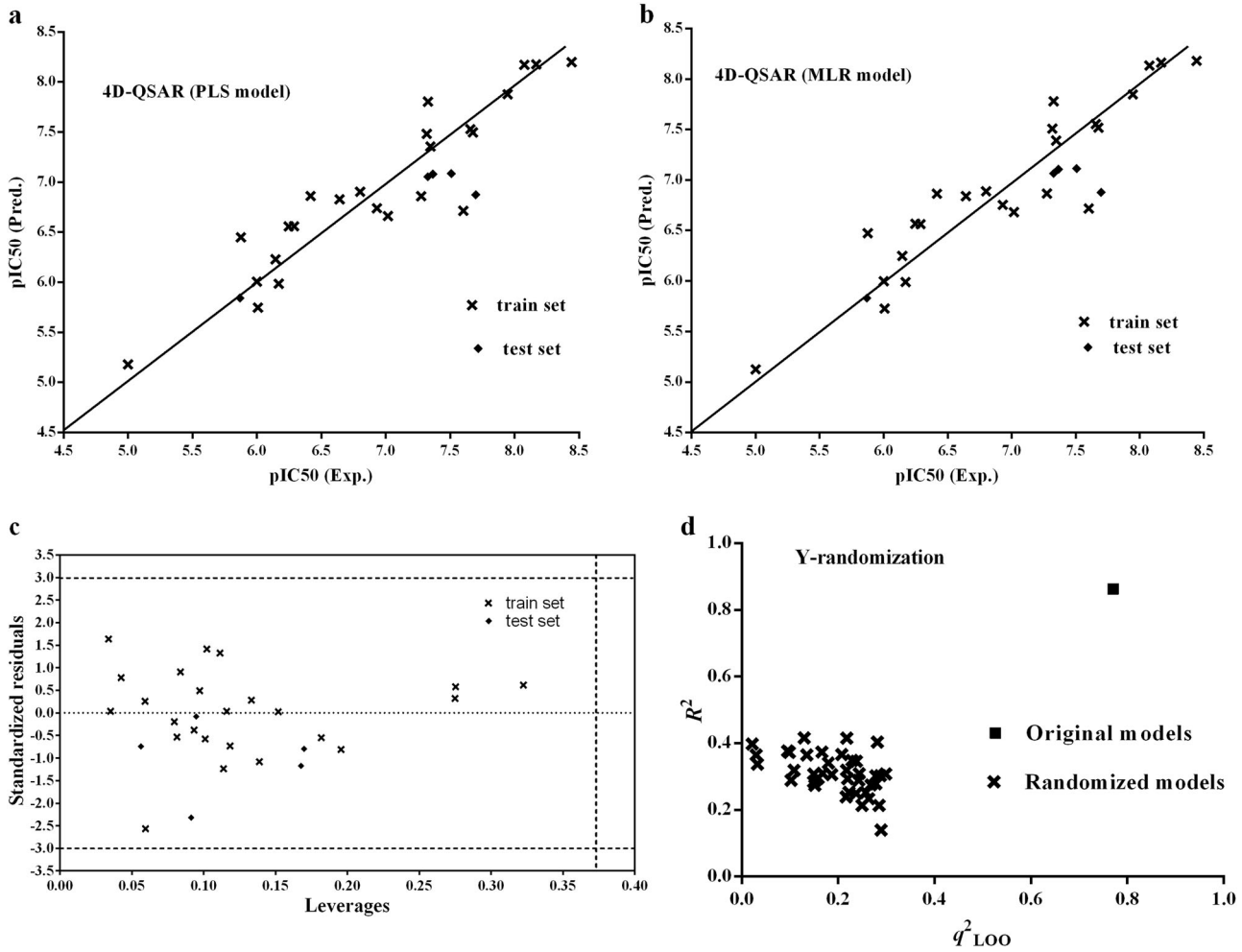

Fig. 3 Experimental versus predicted pIC_{50} values based on the 4D-QSAR, PLS **a** and MLR models **b**, plot of standardized residuals versus leverages. Dotted line represents warning leverage ($h^* = 0.375$) and dashed lines represent ±3 standardized residual **c** and scatter plot for R^2 versus q^2 for the randomized models (star) and original model (plain square) **d**

Table 3 Model statistical parameters for random series of train/test

No. model		Test set	R^2	q^2_{lOO}	R^2_{Pred}
	1	10, 20, 27, 33, 41	0.810	0.604	0.953
	2	15, 21, 28, 34, 39	0.813	0.681	0.994
4D-QSAR	3	16, 22, 29, 35, 37	0.880	0.808	0.673
	4	17, 23, 30, 36, 38	0.795	0.685	0.890
	5	18, 24, 25, 33, 40	0.851	0.718	0.762
	1	10, 20, 27, 33, 41	0.876	0.560	0.935
	2	15, 21, 28, 34, 39	0.926	0.618	0.677
3D-QSAR	3	16, 22, 29, 35, 37	0.906	0.606	0.402
	4	17, 23, 30, 36, 38	0.900	0.753	0.462
	5	18, 24, 25, 33, 40	0.938	0.752	0.436

checked with the Y-randomization validation method. The dependent variable-vector was shuffled many times randomly in the Y-randomization and a new model was constructed with the original independent variable matrix. The results of Y-randomization method are presented in Fig. 3c. The average R^2 and q^2_{LOO} values resulting from Y-randomization for the MLR model were 0.309 and 0.194. So the 4D-QSAR models presented robust results. The series of train and test set were randomly produced in order to further check the reliability and robustness of the models. The results were presented in Table 3. So the sensitivity of the obtained 4D-QSAR models with the different training and test set is very negligible.

Model applicability domain

The chemical applicability domain (AD) and the robustness of the obtained model are verified by the leverage method which calculates the leverage, h^*, for each molecule. The warning leverage is generally fixed at $3LV/m$, where m is the number of training set compounds. The results of the AD analysis are presented in Fig. 3d which reveals no outliers in all of the compounds. It is also important to note that the validation compounds which were not used to develop model are predicted with similar accuracy of the training compounds.

Graphical representations of the 4D-QSAR model are shown in Fig. 4a. Yellow regions denote steric regions corresponding to negative regression coefficients, while green regions indicate steric interactions related to positive MLR regression coefficients. Likewise, blue and red regions represent electrostatic descriptors with negative and positive regression coefficients, respectively. The green LJ + descriptor regions near the R_2 position (adamantan-1-yl, cyclohexyl, or cycloheptyl) indicate that a sterically group is favored in this region. It can show the fact that activities of compounds with bulkier adamantan-1-yl group (such as compounds 10, 23, and 35) are more than their analogs

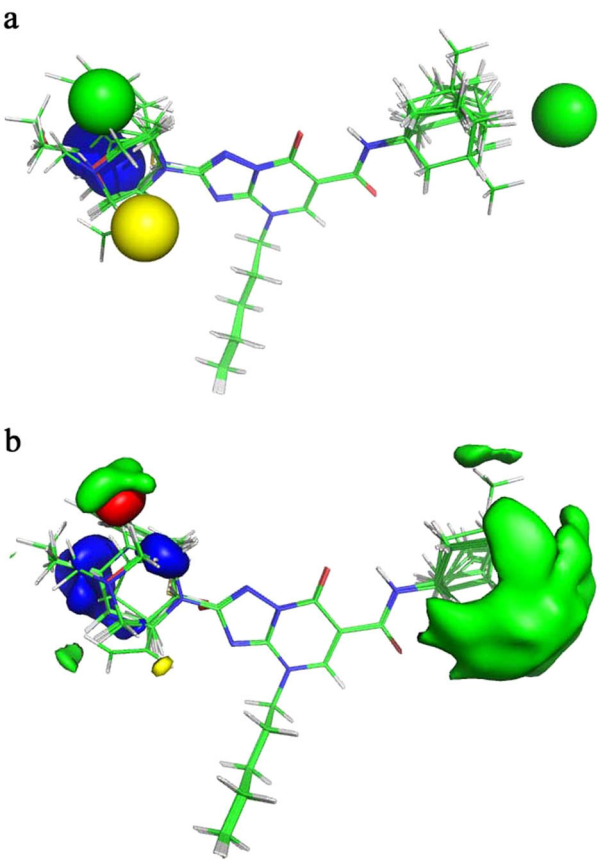

Fig. 4 Steric and electrostatic contour maps of 4D-QSAR model **a** and 3D-QSAR model **b**

(such as compounds 19, 22, and 34) at this position. The additional green LJ + descriptor region near *N*-benzyl-*N*-methyl showed that biological activity will be increased by this group. The yellow LJ- descriptor region near diallylamine indicate that bulky group is unfavorable for biological activity. So the biological activity of compound 40 is less than the analogs with small group at this position. The blue C- descriptor near aromatic ring displays an electrostatic unfavorable region and describes positively charged tolerance area would increase activity.

3D-QSAR model

The LNO-cross-validation was carried out to develop an effective 3D-QSARs model. The optimal number of components identified by selecting highest *q2* value in the LNO-cross-validation process was used in the final non-cross-validated PLS run. The best model yielded values $R^2 = 0.907$, a standard deviation of the error of calculation (SDEC) of 0.2593, $q^2_{\text{LNO}} = 0.633$ for 3 components, a SD on standard deviation of the error of prediction (SDEP) of 0.0378 and $R^2_{\text{Pred}} = 0.614$ using both electrostatic and steric fields with 1.0 grid spacing.

The correlation between the experimental activities and the predicted activities are shown in Fig. S1. The predicted and experimental activities for training and test set compounds with 4D-QSAR (PLS and MLR) and 3D-QSAR are depicted in Table. S1 of the Supplementary Material. The 3D-QSAR's steric and electrostatic field is presented as 3D colored contour maps in Fig. 4b. The steric interactions are represented by yellow and green colored contours, while electrostatic interactions are represented by blue and red colored contours, same as 4D-QSAR color contours (Uesawa and Mohri 2010). The 3D-QSAR map includes six contours, four of them were in the same position as 4D-QSAR model and their interpretation exactly were verified by the key descriptors in the 4D-QSAR. As mentioned in Table 3, the statistics of 3D-QSAR models were much more sensitive than 4D-QSAR models to separate train and test based on another train and test set of compounds were examined, and 3D-QSAR results are highly dependent on partitioning of train and test set of compounds.

Conclusion

A simple 4D-QSAR model was developed for the triazolo [1,5-a]pyrimidine compounds as CB_2 cannabinoid receptor inverse agonists by making use of MD simulation to obtain a CEP for each compound. Then the interaction energy descriptors were calculated and analyzed by PLS and MLR. Though the 4D-QSAR model obtained has a small number of variables, comparison to 3D-QSAR, the suitable predictive ability and the excellent statistical parameters of the results indicate that this model can help to rational design novel CB_2 cannabinoid receptor inverse agonists with preferred activities.

Acknowledgements The project was supported by the Jiangxi Province Science Foundation (20171BAB205104), the graduate innovation fund of Jiangxi Province (CX2016298).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ (2010) 4D-QSAR: perspectives in drug design. Molecules 15:3281–3294

Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Clark RD (2003) Boosted leave-many-out cross-validation: the effect of training and test set diversity on pls statistics. J Comput Aided Mol Des 17:265–275

Darden T, York D, Pedersen L (1993) Particle mesh ewald - an N-Log (N) method for ewald sums in large systems. J Chem Phys 98:10089–10092

Di Marzo V, Bifulco M, De Petrocellis L (2004) The endocannabinoid system and its therapeutic exploitation. Nat Rev Drug Discov 3:771–784

Ghasemi JB, Safavi-Sohi R, Barbosa EG (2012) 4D-LQTA-QSAR and docking study on potent gram-negative specific lpxc inhibitors: a comparison to comfa modeling. Mol Divers 16:203–213

Ghasemi JB, Salahinejad M, Rofouei MK (2011) Review of the quantitative structure-activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules. Supramol Chem 23:615–631

Golbraikh A, Tropsha A (2002) Beware of q(2)! J Mol Graph Model 20:269–276

Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S (2013) Qsarins: a new software for the development, analysis, and validation of qsar mlr models. J Comput Chem 34:2121–2132

Hassinen T, Peräkylä M (2001) New energy terms for reduced protein models implemented in an off-lattice force field. J Comput Chem 22:1229–1242

Hopfinger AJ, Wang S, Tokarski JS, Jin BQ, Albuquerque M, Madhav PJ, Duraiswami C (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J Am Chem Soc 119:10509–10524

Howlett AC, Barth F, Bonner TI, Cabral G, Casellas P, Devane WA, Felder CC, Herkenham M, Mackie K, Martin BR, Mechoulam R, Pertwee RG (2002) International union of pharmacology. Xxvii. Classif cannabinoid Recept Pharmacol Rev 54:161–202

Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J Braz Chem Soc 20:770–787

Malfitano AM, Basu S, Maresz K, Bifulco M, Dittel BN (2014) What we know and do not know about the cannabinoid receptor 2 (CB_2). Semin Immunol 26:369–379

Martins JPA, Barbosa EG, Pasqualoto KFM, Ferreira MMC (2009) LQTA-QSAR: a new 4D-QSAR methodology. J Chem Inf Model 49:1428–1436

Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI (1990) Structure of a cannabinoid receptor and functional expression of the cloned cDNA. Nature 346:561–564

Munro S, Thomas KL, Abushaar M (1993) Molecular characterization of a peripheral receptor for cannabinoids. Nature 365:61–65

Patel P, Rajak H (2018) Development of hydroxamic acid derivatives as anticancer agent with the application of 3D-QSAR, docking and molecular dynamics simulations studies. Med Chem Res 27:2100–2115

Patil R, Sawant S (2015) Molecular dynamics guided receptor independent 4D QSAR studies of substituted coumarins as anticancer agents. Curr Comput Aided Drug Des 11:39–50

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera - a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Picone RP, Kendall DA (2015) From the bench, toward the clinic: therapeutic opportunities for cannabinoid receptor modulation. Mol Endocrinol 29:801–813

Shim J, Mackerell AD (2011) Computational ligand-based rational design: role of conformational sampling and force fields in model development. Medchemcomm 2:356–370

Tabrizi MA, Baraldi PG, Ruggiero E, Saponaro G, Baraldi S, Poli G, Tuccinardi T, Ravani A, Vincenzi F, Borea PA, Varani K (2016) Synthesis and structure activity relationship investigation of

triazolo 1,5-a pyrimidines as CB_2 cannabinoid receptor inverse agonists. Eur J Med Chem 113:11–27

Tosco P, Balle T (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. J Mol Model 17:201–208

Tosco P, Balle T, Shiri F (2011) Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. J Comput Aided Mol Des 25:777–783

Uesawa Y, Mohri K (2010) Quantitative structure-activity relationship (QSAR) analysis of the inhibitory effects of furanocoumarin derivatives on cytochrome p450 3A activities. Pharmazie 65:41–46

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) Gromacs: fast, flexible, and free. J Comput Chem 26:1701–1718

Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130