Retrieving novel C5aR antagonists using a hybrid ligand-based virtual screening protocol based on SVM classification and pharmacophore models

Xiao-Yu Qing\textsuperscript{a}, Chun-Hui Zhang\textsuperscript{a}, Lin-Li Li\textsuperscript{b}, Pan Ji\textsuperscript{a}, Shuang Ma\textsuperscript{a}, Hua-Lin Wan\textsuperscript{a}, Ze-Rong Wang\textsuperscript{b}, Jun Zou\textsuperscript{a} & Sheng-Yong Yang\textsuperscript{a}

\textsuperscript{a}State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, No. 1, Keyuan Road 4, Gaopeng Avenue, Chengdu, 610041, Sichuan, China

\textsuperscript{b}West China School of Pharmacy, Sichuan University, 610041, Sichuan, China

Published online: 25 Jul 2012.

To cite this article: Xiao-Yu Qing, Chun-Hui Zhang, Lin-Li Li, Pan Ji, Shuang Ma, Hua-Lin Wan, Ze-Rong Wang, Jun Zou & Sheng-Yong Yang (2013) Retrieving novel C5aR antagonists using a hybrid ligand-based virtual screening protocol based on SVM classification and pharmacophore models, Journal of Biomolecular Structure and Dynamics, 31:2, 215-223, DOI: 10.1080/07391102.2012.698245

To link to this article: http://dx.doi.org/10.1080/07391102.2012.698245
Retrieving novel C5aR antagonists using a hybrid ligand-based virtual screening protocol based on SVM classification and pharmacophore models

Xiao-Yu Qinga, Chun-Hui Zhanga, Lin-Li Lib*, Pan Ji, Shuang Maa, Hua-Lin Wan, Ze-Rong Wangb, Jun Zoa and Sheng-Yong Yanganga*

aState Key Laboratory of Biotherapy and Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, No. 1, Keyuan Road 4, Gaopeng Avenue, Chengdu 610041, Sichuan, China; bWest China School of Pharmacy, Sichuan University, 610041 Sichuan, China

Communicated by Ramaswamy H. Sarma

(Received 24 January 2012; final version received 2 April 2012)

C5aR antagonists have been thought as potential immune mediators in various inflammatory and autoimmune diseases, and discovery of C5aR antagonists has attracted much attention in recent years. The discovery of C5aR antagonists was usually achieved through high-throughput screening, which usually suffered a high cost and a low success rate. Currently, the fast developing computer-aided virtual screening (VS) methods provide economic and rapid approaches to the lead discovery. In this account, we proposed a hybrid ligand-based VS protocol that is based on support vector machine (SVM) classification and pharmacophore models for retrieving novel C5aR antagonists. Performance evaluation of this hybrid VS protocol in virtual screening against a large independent test set, T-CHEM, showed that the hybrid VS approach significantly increased the hit rate and enrichment factor compared with the individual SVM classification model-based VS and pharmacophore model-based VS, as well as molecular docking-based VS in that the receptor structure was created by homology modeling. The hybrid VS approach was then used to screen several large chemical libraries including PubChem, Specs, and Enamine. Finally, a total of 20 compounds were selected from the top ranking hits, and shifted to the subsequent in vitro and in vivo studies, which results will be reported in the near future.

Keywords: C5aR antagonist; pharmacophore; support vector machine; hybrid virtual screening approach

1. Introduction

The complement component C5a is a 74-residue pro-inflammatory polypeptide generated during activation of the complement cascade by proteolysis of C5 (Huber-Lang et al., 2003; Monk, Scola, Madala, & Fairlie, 2007). C5a plays critical roles in a variety of biological processes, in particular the inflammatory and immune response (Stahel & Barnum, 2006). For example, C5a causes induction of several cytokines, such as IL-1, IL-6, IL-8, and TNF-α, from leukocytes, and augmentation of the humoral and cell-mediated immune response (Dunkelberger & Song, 2010; Ember, Sanderson, Hugh, & Morgan, 1994). Excessive complement activation leading to elevated plasma levels of C5a is associated with many clinical conditions, including sepsis, adult respiratory distress syndrome, rheumatoid arthritis, and ischemic heart disease (Guo & Ward, 2005). C5a exerts its bioactivities by binding to G-protein-coupled C5a receptor (C5aR), which is expressed widely on inflammatory and noninflammatory cells (Flierl, Schreiber, & Huber-Lang, 2006). Therefore, C5aR has been thought as an important therapeutic target for various inflammatory and autoimmune diseases, and discovery of C5aR antagonists has attracted much attention in recent years (Kurihara et al., 2010; Lee, Whitfeld, & Mackay, 2008; Qu, Ricklin, & Lambris, 2009).

Over the past decade, a number of C5aR antagonists have been discovered (Brodebeck et al., 2008; Fonseca et al., 2009; Hu, Xie, Chen, & Fang, 2011). Even so, there is only one small molecular C5aR antagonist, CCX-168, in clinical trials currently. Thus, discovering more potent C5aR antagonists is still needed, which could provide more lead candidates for drug development. Formerly, the discovery of C5aR antagonists was achieved through high-throughput screening, which usually suf-
federal a high cost and a low success rate (Chen et al., 2010; da Silva, Carvalho, & Taft, 2007; Yang, 2010). Currently, the fast developing computer-aided virtual screening (VS) methods provide economic and rapid approaches to the lead discovery (Atri et al., 2011; Chen et al., 2009, 2011, 2010; Hage-Melim, da Silva, Semighini, Taft, & Sampaio, 2010; Huang et al., 2010; Schwartd, Kolb, & Ernst, 2003). Commonly used VS methods include structure-based methods, such as molecular docking and ligand-based approaches, for example pharmacophore-based VS (Andrianov, 2008; Andrianov & Anishchenko, 2009; Kahlon, Roy, & Sharma, 2010). Nevertheless, structure-based methods are not suitable for the case of C5aR since for which there is no crystal structure reported at present; although the structure of C5aR could be created by homology modeling (Higginbottom et al., 2005; Lee et al., 2008), the reliability of structure generated together with the existing deficiencies of molecular docking including receptor flexibility, scoring function, and solvent problems has significant impact on the performance of DB-VS (Higginbottom et al., 2005; Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005). Thus, the ligand-based methods are the primary option for the VS of C5aR antagonists. The most widely used ligand-based method is the pharmacophore-based VS (PB-VS). Although PB-VS is one of the best VS methods, it still bears some inherent limitations including low speed, high false-positive rate, and low enrichment factor, especially in screening large chemical databases (Kirchmayr, Markt, Distinto, Wolber, & Langer, 2008; Yang, Chen, Shen, Kristal, & Hsu, 2005). Lately emerging support vector machine (SVM)-based VS (SB-VS) seems superior to the PB-VS in some aspects, in particular the screening speed (Han et al., 2008; Liew, Ma, Liu, & Yap, 2009; Liu et al., 2009). However, the SB-VS still suffers a low hit rate and a high false-positive rate, which might originate from the lack of consideration of information regarding the three dimensional (3D) pharmacophore features of small molecules (Mohan et al., 2005). A combination of these methods is expected to mutually compensate for these limitations and capitalize on their mutual strengths. The goal of this investigation is to establish a hybrid ligand-based VS approach that combines the SB-VS and PB-VS for discovering new potent antagonists of C5aR. The performance of this hybrid VS approach will be evaluated through comparing with that of individual SB-VS and PB-VS, as well as molecular docking-based VS (DB-VS) in that the structure of C5aR is created by homology modeling. Then, this hybrid VS approach will be applied to screen several large chemical libraries, including PubChem (18,831,686 compounds), Specs (202,408 compounds), and Enamine (980,000 compounds) for identifying novel C5aR antagonists. Finally, some of the retrieved hit compounds will be selected and shifted to the subsequent in vitro and in vivo studies.

2. Materials and methods

2.1. SVM modeling

2.1.1. Support vector machine

In this investigation, an optimal SVM modeling method, namely SVM method combined with genetic algorithm (GA) for feature selection and conjugate gradient method (CG) for parameter optimization (GA-CG-SVM), which was proposed by our group recently, was used (Yang et al., 2009). A brief description regarding SVM and GA-CG-SVM has been presented in Supplementary material.

2.1.2. Data-sets

We collected 456 compounds from different literature resources, which include 401 positives (C5aR antagonists, IC$_{50} < 50\mu$M) and 55 negatives (C5aR nonantagonists, IC$_{50} > 50\mu$M) (Barbay et al., 2008; Blagg et al., 2008a, 2008b; Brodbbeck et al., 2008; Chen et al., 2010; Schnatbam et al., 1999; Pellis & Wennogle, 1999; Fan et al., 2010; Qu et al., 2009; Sangee et al., 2009; Sumichika et al., 2002; Waters et al., 2005; Zenon Kontetatis & Springer, 1997). Using the protocol of “Generate Training and Test Data” of Discovery Studio 3.1 (DS 3.1), we grouped the 401 positive compounds into a training set, a test set, and an independent validation set based on the fingerprints FCFP_6, ECFP_6, FPFP_6, and EPFP_6. The finally formed training set contains 282 positive compounds, which was used for SVM modeling. The test set for validation of the SVM model includes 60 positive compounds and 55 collected negative compounds (all the training set and test set compounds are given in Supplementary material). The remaining 59 positive compounds and 79,534 decoys from ChEMBL (GPCR SARfari) database comprise the independent test set, termed as T-CHEM, which was used for evaluating the performance of the proposed hybrid VS approach.

As we have known, in the SVM modeling, an adequate amount of negatives in the training set would be very important for the quality of SVM model (Mohan et al., 2005). However, the collection of negatives is difficult since very few of literature definitely indicates that a compound is inactive. Therefore, we adopted a method suggested by Chen et al. to generate putative negatives without the knowledge of actual negatives (Han et al., 2008; Liu et al., 2009). In the generation of the putative negatives, we clustered the 18.83 M compounds from the PubChem into 8913 compound families by the K-means clustering method based on their molecular descriptors calculated by the DS 3.1 software. Finally, we randomly selected 17,853 putative negatives from those families that do not contain any of the known C5aR antagonists. These compounds together with 282 known C5aR antagonists constitute the training set for the SVM modeling.
2.1.3. Details for the SVM modeling

All the SVM modeling calculations were carried out by using GA-CG-SVM program (Yang et al., 2009; Zhang et al., 2009). In this investigation, the initial descriptors were calculated by the DS 3.1 program package. The termination criterion is that the generation number reaches 200 or that the fitness value does not improve during the last 10 generations. The crossover rate was set to 0.5. The mutation rate was 0.1. The starting values of the parameters (C, γ) were set to (8192, 0.38) in the program running.

2.2. Pharmacophore modeling

The HipHop algorithm implemented in DS 3.1 program package was employed for the pharmacophore modeling. Five C5aR antagonists (compounds 1–5, see Figure 1) were used as the training set compounds; these compounds were chosen since they are the most active compounds and bear structural diversity. Compound 1 (W-54,011, IC50 = 2.2 nM) (Sumichika et al., 2002), compound 3 (IC50=1.5nM) (Blagg et al., 2008a, 2008b), and compound 5 (IC50 = 7 nM) (Gong et al., 2008) (see Figure 1) were selected as “Reference Compound” specifying a “Principal” value of 2 and a “MaxOmitFeat” value of 0. The “Principal” and “MaxOmitFeat” values of the remaining two compounds were set to 1 and 0. The “Minimum Interfeature Distance” value was set to 1.0 Å. All the remaining HipHop parameters were kept at their default values.

2.3. Homology modeling of the structure of C5aR

The homology modeling was carried out using Modeler program within DS 3.1. The sequence of C5aR was obtained by NCBI database (http://www.ncbi.nlm.nih.gov/) showing in Table SI (see Supplementary material). The structure of bovine rhodopsin A (PDB Code: 1F88) was used as template for the homology modeling (Higginbottom et al., 2005); the amino acid identity between C5aR and the bovine rhodopsin A is 21.4%. The model with the lowest PDF (Probability Density Functions) total energy was selected and subjected to loop refinement and side chain refinement. The energy minimization was carried out using the conjugated gradient method with iterations of 50,000 steps and a RMS gradient of 0.001.

2.4. Molecular docking

All the molecular docking studies were carried out by GOLD (Genetic Optimization of Ligand Docking) 4.0. GOLD adopted the genetic algorithm to dock flexible ligands into protein binding sites. The C5aR structure generated by homology modeling was used as the reference receptor for the docking studies. Hydrogen atoms were added to the protein by using DS 3.1. The Charmm force field was assigned. A pocket surrounded by Ile116 (3.32), Tyr121 (3.37), Glu199 (5.35), Arg206 (5.42), Phe211 (5.47), Leu215 (5.51), Phe251 (6.44), Trp255 (6.48), Asp282 (7.35), Val286 (7.39), and Tyr290 (7.43) was defined as the binding site; this definition was based on studies by Gerber et al. (2001) and Higginbottom et al. (2005).
3. Results and discussion

3.1. Establishment of the SVM classification model

Firstly, we constructed a training set for generating a classification model of C5aR antagonists and nonantagonists. The training set contains 18,135 compounds, including 282 known C5aR antagonists and 17,853 putative nonantagonists (a detailed description about the generation of the training set is given in Section 2). Initially, 252 molecular descriptors, which cover diverse molecular properties such as geometrical, topological, and electronic properties, were selected. These descriptors were preprocessed to remove obvious bad ones. Here, the following descriptors were removed: (1) descriptors with too many zero values, (2) descriptors that are highly correlated with others (correlation coefficients > 95%), and (3) descriptors with very small standard deviation values (<0.5%). Seventy-nine molecular descriptors passed through this process. The descriptor values were then scaled to a range of 1 to +1, which was necessary since the use of descriptors with their values distributed in considerable different ranges would impact the quality of SVM models generated. The scaled descriptors were subjected to the GA–CG algorithm for a further descriptor selection. Finally, 35 molecular descriptors including constitutional descriptors (8), estate keys (8), lipophilicity descriptors (1), hydrogen-bonding descriptors (2), surface area and volume descriptors (5), and topological descriptors (12), were selected to build the SVM classification model (see Table 1).

3.2. Validation of the SVM classification model

The SVM classification model just developed was first validated by a fivefold cross-validation method (Hamilton, Pantelic, Hanson, & Teasdale, 2007). In the fivefold cross-validation, the training set was randomly partitioned into five subsets. Of the five subsets, one single subset was retained as the validation set, and the remaining four subsets were used as the training set. The cross-validation process was then repeated five times, with each of the five subsets used exactly once as a validation set. The five results obtained from these folds were then averaged to produce a single estimation. The fivefold cross-validation results for the training set are given in Table 2. From Table 2, we can see that, of the 282 C5aR antagonists, 245 were correctly predicted (TP, Table 2). The prediction accuracy of C5aR antagonists (SE,}

Table 1. Molecular descriptors selected by the GA–CG algorithm for the generation of SVM classification model of C5aR antagonists and nonantagonists.

| Descriptor class                     | Number of descriptors | Descriptors                                                                 |
|--------------------------------------|-----------------------|-----------------------------------------------------------------------------|
| Constitutional descriptors           | 8                     | N_Count, NPlusO_Count, Num_Atoms, Num_Bonds, Num_RingBonds, Num_Rings6, Num_ChainAssemblies, Num_AtomClass |
| Estate keys                          | 8                     | ES_Sum_aaCH, ES_Sum_ssCH2, ES_Sum_ssNH, ES_Sum_sssCH, ES_Sum_sssN, ES_Count_aasC, ES_Count_sCH3, ES_Count_ssCH2 |
| Lipophilicity descriptors            | 1                     | ALogP                                                                       |
| Hydrogen-bonding descriptors         | 1                     | Num_H_Acceptors                                                             |
| Surface area and volume descriptors  | 5                     | Molecular_Solubility, Molecular_SurfaceArea, Molecular_PolarSurfaceArea, Molecular_FractionalPolarSurfaceArea, Molecular_Polar_SASA |
| Topological descriptors              | 12                    | BIC, IC, SIC, CHI_2, CHI_V_0, CHI_V_1, CHI_V_2, CHI_V_3_P, Kappa_1_AM, Kappa_2_AM, Kappa_3, PHI |

Table 2. Results of the SVM model validation by fivefold cross-validation and an independent validation set.

| Method                        | Number of compounds | Positive | Negative | Total | Negative | Positive | Negative |
|-------------------------------|---------------------|----------|----------|-------|----------|----------|----------|
|                              |                     | TP       | FN       | SE%   | TN       | TP       | FN       |
| Fivefold cross-validation     | 18,135              | 245      | 37       | 86.88 | 17,830   | 23       | 99.87    | 99.66   |
| Independent validation set    | 115                 | 44       | 11       | 80    | 53       | 2        | 96.67    | 88.70   |

TP: true positive; TN: true negative; FP: false positive; FN: false negative; SE (%): sensitivity, SE = TP/(TP + FN); SP(%) : specificity, SP = TN/(TN + FP); and Q (%): overall accuracy, Q = (TP + TN)/(TP + TN + FP + FN).
Table 2) was 86.88%. For the 17,853 C5aR nonantagonists, 17,830 were correctly predicted (TN, Table 2). The accuracy for the prediction of C5aR nonantagonists (SP, Table 2) was 99.87%. The total prediction accuracy (Q) was 99.66%. These demonstrated that the generated SVM classification model was quite good for the prediction of training set agents.

A good SVM model needed the capability not only to classify the training set agents correctly into positives and negatives, but also to predict external agents which are outside of the training set accurately. Here, an independent validation set was employed to evaluate the established SVM model. The independent validation set comprises 60 antagonists and 55 nonantagonists. Of the 60 antagonists, 44 (TP, Table 2) were correctly predicted, indicating a prediction accuracy of 88.70%, which was comparable with that for the training set.

3.3. Pharmacophore modeling

A pharmacophore model was an ensemble of essential chemical features necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger its biological response (Daisy et al., 2011). The purpose of pharmacophore modeling was to extract these essential chemical features from a set of known ligands, called training set. In this investigation, five C5aR antagonists (compounds 1–5, see Figure 1) were selected to form the training set; these compounds were chosen since they are the most active compounds and have different scaffolds. The HipHop algorithm implemented in DS 3.1 program package was used to generate common feature based pharmacophore models. In the HipHop run, 10 pharmacophore hypotheses were finally generated. The top ranking hypothesis Hypo1, which will be used for the subsequent virtual screening, contains one hydrogen-bond acceptors (A), one ring-aromatic feature (Z), three hydrophobic features (H), and one hydrophobic aromatic feature (R). The 3D spatial relationship and geometric parameters of Hypo1 are shown in Figure 2A. Figure 2B depicts the alignments of Hypo1 with one active compound (compound 1).

Table 2) was 99.66%. These demonstrated that the generated SVM classification model was quite good for the prediction of training set agents.

3.4. Validation of the pharmacophore model Hypo1

To examine whether the pharmacophore model Hypo1 has the capacity to differentiate C5aR antagonists and nonantagonists, we constructed a test set, named T-DB, which comprised of 330 known C5aR antagonists and 4580 decoys (the selection method for the decoys see Supplementary material) from Drug Bank library. Each of the T-DB compounds was mapped onto Hypo1 by using the protocol of “Ligand Pharmacophore Mapping” within the DS 3.1. The “Maximum Omitted Feature” was set to 2, and the “Fitting Method” adopted was “Flexible”. Of the 330 C5aR antagonists, 311 were found to map very well with Hypo1, indicating a yield of 94.24%. And 1053 of the 4580 (22.99%) decoys were also mapped very well with Hypo1. These results showed, on the one hand, that Hypo1 was capable of discriminating to some extent the C5aR antagonists and nonantagonists. On the other hand, sole use of Hypo1 in VS also suffered a high false-positive rate, implied the necessity of employing a hybrid VS strategy.

3.5. Performance evaluation of the hybrid SB/PB-VS and the individual SB-VS and PB-VS in virtual screening

A large independent test set, called T-CHEM, was constructed to evaluate the overall performance of the hybrid SB/PB-VS and the individual SB-VS and PB-VS in virtual screening. The T-CHEM comprises 59 known C5aR antagonists (these compounds have never been used in previous modeling and model validations), and 79,534 decoys obtained from ChEMBL (GPCR SARfari) library. The yield (percentage of predicted compounds in known antagonists), hit rate (percentage of known antagonists in predicted compounds), and enrichment factor (ratio of hit rate to the percentage of known antagonists in T-CHEM), which showed the magnitude of hit rate improvement over random selection, were evaluated for the performance of virtual screening.

First, SB-VS and PB-VS were solely employed to screen the T-CHEM database. The screening results are...
From Table 3, we can see that for the individual use of SB-VS, the number of predicted hits is 836, in which 54 are positives, indicating a yield of 91.53% (54 out of the 59 antagonists, see Table 3). The hit rate and enrichment factor were 6.46 and 87.08%, respectively. The time used was just around 5 min (0.083 h) on a desktop PC equipped with an Intel E5420 (2500 MHz) processor if the molecular descriptors have been calculated (the time used for molecular descriptor calculations for all the T-CHEM compounds is about 0.22 h on the same computer). For PB-VS, the number of predicted positives was 15,216, and the total hits was 43, showed a yield of 72.88% (43 out of the 59 antagonists, see Table 3). The hit rate and enrichment factor were 0.283 and 3.81%, respectively. The time cost for the screening of T-CHEM by PB-VS was about 3.8 h.

Then, the hybrid VS approach was adopted to screen the T-CHEM library, in which the faster SB-VS was used as the first filter, followed by the slower PB-VS. The final number of positive compounds predicted was 330; in which the hits was 42 with a yield of 71.19% (42 out of the 59 antagonists, see Table 3). The hit rate and enrichment factor were 12.73 and 171.57%, respectively, which were significantly higher than the corresponding values of individual VS methods. The total time used was about 0.67 h.

### 3.6. A comparison of the hybrid SB/PB-VS and molecular docking-based VS (DB-VS)

As mentioned before, ligand-based VS methods were reasonably the primary choice for retrieving C5aR antagonists from virtual chemical databases because of no crystal structure reported for C5aR at present. Nevertheless, DB-VS could be another option since the structure of C5aR can be generated by homology modeling. Here, we made a comparison between the performances of the hybrid SB/PB-VS and DB-VS. In order to carry out molecular docking, we established a 3D structure of C5aR using homology modeling in advance using Mod-eler program within DS 3.1. The established model was evaluated by Profiles-3D. The calculated total "verify score" for the C5aR homology model was 146.06, with the "verify expected high score" and "verify expected low score" being 156.25 and 70.31, respectively. The calculated “verify score” for each of residues in C5aR model is shown in Figure 3A. Apparently, most of the residues that locate in favored region are shown in green, while residues locating in the region of unfavorable torsion angles are shown in red.

| Method     | Predicted positive | Hits  | Yield (%) | Hit rate (%) | Enrichment factor | Time cost (h) |
|------------|--------------------|-------|-----------|--------------|-------------------|---------------|
| SB-VS      | 836                | 54    | 91.53     | 6.46         | 87.08             | 0.083         |
| PB-VS      | 15,216             | 43    | 72.88     | 0.283        | 3.81              | 3.8           |
| SB/PB-VS   | 836/330            | 42    | 71.19     | 12.73        | 171.57            | 0.67          |
| DB-VS      | 10,054             | 16    | 27.12     | 0.159        | 2.15              | 648           |

Table 3. Comparisons of the performances of SB-VS, PB-VS, DB-VS, and SB/PB-VS in virtual screening against an independent test set, T-CHEM (59 antagonists + 79,534 decoys).
residues have a score larger than zero \(324/343\), implied rational structures for these residues. The established C5aR model was further evaluated using Ramachandran plot, which was generally used to check the correctness of predicted torsion angles of residues in proteins. Figure 3B shows the Ramachandran plot of the homology model of C5aR. We can see that only few of residues (1.17%) locate in the region of unfavorable torsion angles. Overall, the validation results demonstrated that the established homology model of C5aR would be a reasonable structure, which will be used in the following docking studies.

A regular DB-VS process was then performed to screen T-CHEM, in which GOLD was used for docking and the C5aR structure generated by homology modeling was taken as the reference receptor. The DB-VS gave 10,054 positives; here a compound was defined as a positive if its value of scoring function ChemPLP was greater than 52 (the threshold value of 52 was determined based on an initial test study made by us, in which we found that a compound has large possibility to be a positive if the value of scoring function ChemPLP is greater than 52). Among these positives, 16 were the known hits that indicated a yield of 27.12% (16 known antagonists out of 59, see Table 3). The hit rate and enrichment factor were 0.159 and 2.15%, respectively. The time cost was about 648 h on the same computer as before. Obviously, compared with the hybrid SB/PB-VS, the DB-VS was much worse in terms of the hit rate and enrichment factor as well as the screening speed. Even compared with the individual SB-VS and PB-VS, the DB-VS was still in the lower position. This situation was understandable because of the less reliability of receptor structure obtained by homology modeling, in addition to the existing deficiencies of molecular docking including receptor flexibility, scoring function, and solvent problems.

### 3.6. Application of the hybrid VS approach for retrieving novel C5aR antagonists

The optimal VS method, named the hybrid ligand-based VS approach, was used to screen several large chemical libraries including PubChem (18,831,686 compounds), Specs (202,408 compounds), and Enamine (980,000 compounds). All the compounds (20,014, 094) were first filtered by SB-VS. And 1031 compounds passed through this filter. These compounds were then subjected to screen by PB-VS. We selected 200 compounds as hits from top ranking ones based on the fitness values, which have been presented in Supplementary material. And a total of 20 compounds were visually chosen from the hit list and shifted to the subsequent experimental assays, which results will be reported in the near future.

### 4. Conclusion

In this investigation, we proposed a hybrid ligand-based virtual screening approach, namely SB/PB-VS, based on SVM classification and pharmacophore models for the identification of potential C5aR antagonists. The performance of SB/PB-VS approach in VS was evaluated, which results showed that it significantly increased the hit rate and enrichment factor compared with the individual use of SB-VS and PB-VS. This hybrid SB/PB-VS was also compared with DB-VS, in which the receptor structure was constructed by homology modeling. The results show that the hybrid SB/PB-VS is significantly more superior to DB-VS in terms of hit rate, enrichment factor, and screening speed. The hybrid VS approach was then used to screen several large chemical libraries including PubChem, Specs, and Enamine. Finally, a total of 20 compounds were selected from the top ranking compounds, and shifted to the subsequent \textit{in vitro} and \textit{in vivo} studies, which results will be reported in the near future. Overall, this study offers a strategy for the discovery of novel C5aR antagonists; this strategy has a potential to apply to other GPCRs whose structures are unknown.

### Supplementary material

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2012.698245.

### References

Andrianov, A.M. (2008). Determining the invariant structure elements of the HIV-1 variable V3 loops: Insight into the HIV-MN and HIV-Haiti isolates. Journal of Biomolecular Structure & Dynamics, 26, 247–254.

Andrianov, A.M., & Anishchenko, I.V. (2009). Computational model of the HIV-1 subtype A V3 loop: Study on the conformational mobility for structure-based anti-AIDS drug design. Journal of Biomolecular Structure & Dynamics, 27, 179–193.

Atri, M.S., Saboury, A.A., Moosavi-Movahedi, A.A., Goliae, B., Sefidbakht, Y., Alijanvand, H.H., … Niasari-Naslaji, A. (2011). Structure and stability analysis of cytotoxic complex of camel alpha-lactalbumin and unsaturated fatty acids produced at high temperature. Journal of Biomolecular Structure & Dynamics, 28, 919–928.

Barbay, J.K., Gong, Y., Bantin, M., Li, J., Claes, C., Hornby, P.J., … He, W. (2008). Synthesis and characterization of 5,6,7,8-tetrahydroquinoline C5a receptor antagonists. Bioorganic & Medicinal Chemistry Letters, 18, 2544–2548.

Blagg, J., Mowbray, C., Pryde, D., Salmon, G., Fairman, D., Schmid, E., & Beaumont, K. (2008). Small, non-peptide C5a receptor antagonists: Part 2. Bioorganic & Medicinal Chemistry Letters, 18, 5605–5608.

Blagg, J., Mowbray, C., Pryde, D.C., Salmon, G., Schmid, E., Fairman, D., & Beaumont, K. (2008). Small, non-peptide C5a receptor antagonists: Part 1. Bioorganic & Medicinal Chemistry Letters, 18, 5601–5604.

Brodbeck, R.M., Cortright, D.N., Kielyk, A.P., Yu, J., Baltazar, C.O., Buck, M.E., … Krause, J.E. (2008). Identification and characterization of NDT 9513727 [N,N-bis
(1,3-benzodioxol-5-ylmethyl)-1-butyl-2,4-diphenyl-1H-imidazole-5-m ethanamine], a novel, orally bioavailable C5a receptor inverse agonist. Journal of Pharmacology and Experimental Therapeutics, 327, 898–909.

Chen, C.Y., Chang, Y.H., Bau, D.T., Huang, H.J., Tsai, F.J., Tsai, C.H., & Chen, C.Y.C. (2009). Ligand-Based Dual Target Drug Design for H1N1: Swine Flu- A Preliminary First Study. Journal of Biomolecular Structure & Dynamics, 29, 171–178.

Chen, C.Y.C. (2010). Virtual screening and theoretical activity prediction of idenopyrazole derivatives of CDK2 inhibitors A QPLD and MM-GBSA approach. Journal of Biomolecular Structure & Dynamics, 27, 627–640.

Chen, C.Y. (2011). TCM Database@Taiwan: The world’s largest traditional Chinese medicine database for drug screening in silico. PloS ONE, 6, e15939.

Chen, J.J., Cole, D.C., Ciszewski, G., Crouse, K., Ellingboe, W., Nowak, P., … Li, W. (2010). Identification of a new class of small molecule C5a receptor antagonists. Bioorganic & Medicinal Chemistry Letters, 20, 662–664.

Daisy, P., Singh, S.K., Vijayalakshmi, P., Selvaraj, C., Rajalakshmi, M., & Suveena, S. (2011). A database for the predicted pharmacophoric features of medicinal compounds. Bioinformation, 6, 167–168.

Dunkelberger, J.R., & Song, W.C. (2010). Complement and its role in innate and adaptive immune responses. Cell Research, 20, 34–50.

Ember, J.A., Sanderson, S.D., Hugli, T.E., & Morgan, E.L. (1994). Induction of interleukin-8 synthesis from monocyes by human C5a anaphylatoxin. American Journal of Pathology, 144, 393–402.

Fan, P.C., Greenman, K.L., Leleti, M.R., Li, Y.L., Powers, J., Tanaka, H., Yang, J., & Zeng, Y.B. (2010). C5AR ANTAGONISTS. United States Patent Application Publication.

Fliert, M.A., Schreiber, H., & Huber-Lang, M.S. (2006). The role of complement, C5a and its receptors in sepsis and multiorgan dysfunction syndrome. Journal of Investigative Surgery, 19, 255–265.

Fonseca, M.I., Ager, R.R., Chu, S.H., Yazan, O., Sanderson, S. D., LaFerla, F.M., … Tenner, A.J. (2009). Treatment with a C5aR antagonist decreases pathology and enhances behavioral performance in murine models of Alzheimer’s disease. Journal of Immunology, 183, 1375–1383.

Gerber, B.O., Meng, E.C., Dotsch, V., Baranski, T.J., & Bourne, H.R. (2001). An activation switch in the ligand binding pocket of the C5a receptor. Journal of Biological Chemistry, 276, 3394–3400.

Gong, Y., Barbay, J.K., Buntinx, M., Li, J., Wauwe, J.V., Claes, C., … He, W. (2008). Design and optimization of aniline-substituted tetrahydroquinoline C5a receptor antagonists. Bioorganic & Medicinal Chemistry Letters, 18, 3852–3855.

Guo, R.F., & Ward, P.A. (2005). Role of C5a in inflammatory responses. Annual Review of Immunology, 23, 821–852.

Hage-Melim, L.I., da Silva, C.H., Semighini, E.P., Taft, C.A., & Sampaio, S.V. (2009). Computer-aided drug design of novel PLA2 inhibitor candidates for treatment of snakebite. Journal of Biomolecular Structure & Dynamics, 27, 27–36.

Hamilton, N.A., Pantele, R.S., Hanson, K., & Teasdale, R.D. (2007). Fast automated cell phenotype image classification. BMC Bioinformatics, 8, 110.

Han, L.Y., Ma, X.H., Lin, H.H., Jia, J., Zhu, F., Xue, Y., … Chen, Y.Z. (2008). A support vector machines approach for virtual screening of active compounds of single and multiplemechanisms from large libraries at an improved hit-rate and enrichment factor. Journal of Molecular Graphics and Modelling, 26, 1276–1286.

Higginbottom, A., Cain, S.A., Woodruff, T.M., Proctor, L.M., Madala, P.K., Tyndall, J.D., … Monk, P.N. (2005). Comparative agonist/antagonist responses in mutant human C5a receptors define the ligand binding site. Journal of Biological Chemistry, 280, 17831–17840.

Hu, Y., Xie, G.H., Chen, Q.X., & Fang, X.M. (2011). Small molecules in treatment of sepsis. Current Drug Targets, 12, 256–262.

Huang, H.J., Lee, K.J., Yu, H.W., Chen, C.Y., Hsu, C.H., Chen, H.Y., & Tsai, F.J. (2010). Structure-based and ligand-based drug design for HER 2 receptor. Journal of Biomolecular Structure & Dynamics, 28, 23–37.

Huber-Lang, M.S., Sarma, J.V., McGuire, S.R., Lu, K.T., Padgaonkar, V.A., Younkin, E.M., … Ward, P.A. (2003). Structure–function relationships of human C5a and C5aR. Journal of Immunology, 170, 6115–6124.

Kahlon, A.K., Roy, S., & Sharma, A. (2010). Molecular docking studies to map the binding site of squalene synthase inhibitors on dehydroperoxide synthase of Staphylococcus aureus. Journal of Biomolecular Structure & Dynamics, 28, 201–210.

Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection – what can we learn from earlier mistakes? Journal of Computer-Aided Molecular Design, 22, 213–228.

Kurihara, R., Yamaoka, K., Sawamukai, N., Shimajiri, S., Oshita, K., Yukawa, S., … Tanaka, Y. (2010). C5a promotes migration, proliferation, and vessel formation in endothelial cells. Inflammation Research, 59, 659–666.

Lee, H., Whitfeld, P.L., & Mackay, C.R. (2008). Receptors for complement C5a. The importance of C5aR and the enigmatic role of C5L2. Immunology and Cell Biology, 86, 153–160.

Liew, C.Y., Ma, X.H., Liu, X., & Yap, C.W. (2009). SVM model for virtual screening of Lck inhibitors. Journal of Chemical Information and Modeling, 49, 877–885.

Liu, X.H., Ma, X.H., Tan, C.Y., Jiang, Y.Y., Go, M.L., Low, B.C., & Chen, Y.Z. (2009). Virtual screening of Abl inhibitors from large compound libraries by support vector machines. Journal of Chemical Information and Modeling, 49, 2101–2110.

Mohan, V., Gibbs, A.C., Cummings, M.D., Jaeger, E.P., & Desjarlais, R.L. (2005). Docking: Successes and challenges. Current Pharmaceutical Design, 11, 323–333.

Monk, P.N., Scola, A.M., Madala, P., & Fairlie, D.P. (2007). Function, structure and therapeutic potential of complement C5a receptors. British Journal of Pharmacology, 152, 429–448.

Pellas, T.C., & Wennogle, L.P. (1999). C5a receptor antagonists. Current Pharmaceutical Design, 5, 737–755.

Qu, H., Ricklin, D., & Lambris, J.D. (2009). Recent developments in low molecular weight complement inhibitors. Molecular Immunology, 47, 185–195.

Sanganee, H.J., Baxter, A., Barber, S., Brown, A.J., Grice, D., Hunt, F., … Unitt, J. (2009). Discovery of small molecule human C5a receptor antagonists. Bioorganic & Medicinal Chemistry Letters, 19, 1143–1147.
Schnatbaum, K., Scharn, D., Loscardi, E., Polakowski, T., Richter, U., Reineke, U., & Hummel, G. (2009). United States Patent Application Publication US 2008/022003 A1. 1–119.

Schwardt, O., Kolb, H., & Ernst, B. (2003). Drug discovery today. Current Topics in Medicinal Chemistry, 3, 1–9.

da Silva, C.H., Carvalho, I., & Taft, C.A. (2007). Virtual screening, molecular interaction field, molecular dynamics, docking, density functional, and ADMET properties of novel AChE inhibitors in Alzheimer’s disease. Journal of Biomolecular Structure & Dynamics, 24, 515–524.

Stahel, P.F., & Barnum, S.R. (2006). The role of the complement system in CNS inflammatory diseases. Expert Reviews of Clinical Immunology, 2, 445–456.

Sumichika, H., Sakata, K., Sato, N., Takeshita, S., Ishibuchi, S., Nakamura, M., … Naka, Y. (2002). Identification of a potent and orally active non-peptide C5a receptor antagonist. Journal of Biological Chemistry, 277, 49403–49407.

Waters, S.M., Brodbeck, R.M., Steflik, J., Yu, J., Baltazar, C., Peck, A.E., … Krause, J.E. (2005). Molecular characterization of the gerbil C5a receptor and identification of a transmembrane domain V amino acid that is crucial for small molecule antagonist interaction. Journal of Biological Chemistry, 280, 40617–40623.

Yang, S.Y. (2010). Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. Drug Discovery Today, 15, 444–450.

Yang, J.M., Chen, Y.F., Shen, T.W., Kristal, B.S., & Hsu, D.F. (2005). Consensus scoring criteria for improving enrichment in virtual screening. Journal of Chemical Information and Modeling, 45, 1134–1146.

Yang, S.Y., Huang, Q., Li, L.L., Ma, C.Y., Zhang, H., Bai, R., … Wei, Y.Q. (2009). An integrated scheme for feature selection and parameter setting in the support vector machine modeling and its application to the prediction of pharmacokinetic properties of drugs. Artificial Intelligence in Medicine, 46, 155–163.

Zenon Konteatis, S.J.S., & Springer, Martin S. (1997). Assay to identify human C5a antagonists and agonists. United States Patent.

Zhang, H., Xiang, M.L., Ma, C.Y., Huang, Q., Li, W., Xie, Y., … Yang, S.Y. (2009). Three-class classification models of logS and logP derived by using GA-CG-SVM approach. Molecular Diversity, 13, 261–268.