Journal of Molecular Graphics and Modelling 32 (2012) 82–88

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Enrichment of virtual hits by progressive shape-matching and docking

Jiwon Choi, Ningning He, Nayoung Kim, Sukjoon Yoon*

Sookmyung Women's University, Department of Biological Sciences, Research Center for Women's Diseases, Hyochangwongil 52, Yongsan-gu, Seoul 140-742, Republic of Korea

ARTICLE INFO

Article history:
Received 15 July 2011
Received in revised form 1 October 2011
Accepted 3 October 2011
Available online 12 October 2011

Keywords:
Virtual screening
Lead hopping
Docking
Ligand-centric methods
Receptor-centric methods
Principal component analysis

ABSTRACT

The main applications of virtual chemical screening include the selection of a minimal receptor-relevant subset of a chemical library with a maximal chemical diversity. We have previously reported that the combination of ligand-centric and receptor-centric virtual screening methods may provide a compromise between computational time and accuracy during the hit enrichment process. In the present work, we propose a "progressive distributed docking" method that improves the virtual screening process using an iterative combination of shape-matching and docking steps. Known ligands with low docking scores were used as initial 3D templates for the shape comparisons with the chemical library. Next, new compounds with good template shape matches and low receptor docking scores were selected for the next round of shape searching and docking. The present iterative virtual screening process was tested for enriching Peroxisome proliferator-activated receptor and Phosphoinositide 3-kinase relevant compounds from a selected subset of the chemical libraries. It was demonstrated that the iterative combination improved the lead-hopping practice by improving the chemical diversity in the selected list of virtual hits.

1. Introduction

Fast, high-throughput virtual screening (HTVS) of chemical libraries has become a popular technique for the selection of a minimal number of compounds for experimental screening. HTVS also serves as an alternative to experimental high-throughput screening for the fast generation of lead compounds during drug discovery. The importance of HTVS is increasing simultaneously with the rapidly growing number of small molecules that are available in corporate and public compound libraries [1]. Despite the recent theoretical and technical improvements in virtual screening methods, there is still demand for the methodological development of more efficient and accurate virtual screening [2,3].

Virtual screening can be roughly divided into two categories: ligand-centric screening and receptor-centric screening. Ligand-centric methods essentially focus on the comparative analysis of the structural shape and chemical complementarity between compounds and known ligands. Therefore, knowledge of experimentally selected active compounds is a prerequisite for applying ligand-centric methods [4]. In contrast, receptor-centric methods predict the interaction of given compounds with a target receptor, which does not necessarily require experimental data on the active compounds. Molecular docking, which is a key method in receptor-centric virtual screening, is a technique that uses computers to predict the best binding mode of a given compound to a target receptor, resulting in a theoretically predicted binding affinity between the two molecules. Therefore, the docking step has become a primary component in many lead discovery practices [5–7].

It is likely that ligand-centric and receptor-centric virtual screening methods are complementary to one another. Thus, a well-tuned combination of these methods is required to maximize the effectiveness of virtual screening. In one of our previous studies, we presented a novel docking method that integrated ligand-centric and receptor-centric virtual screening methods to incorporate receptor flexibility with the time-efficiency of single conformation docking [8]. We elaborated on a distributed docking approach that took advantage of shape matching and multiple conformation docking methods, aiming to improve the activity of the hit enrichment in the selected list of virtual hits. The database compounds were classified in advance based on the shape similarities to one of the ligands that were complexed with the target protein in the available crystal structures. This classification enabled us to choose the appropriate receptor conformation for the single-receptor conformation docking of a given compound, thereby avoiding the time-consuming multiple docking approach. In the present study, we have improved the distributed docking by iterative combination of shape-search and docking methods.

Another important aspect of the application of virtual screening is lead hopping, which can be defined as the identification of isofunctional molecular structures with significantly different molecular backbones [9]. Lead hopping has become an important practice in the preparation of screening libraries and in maximizing the diversity of the chemical scaffolds in the potential hit list.

* Corresponding author. Tel.: +82 2 710 9415; fax: +82 2 2077 7322.
E-mail address: yoonsj@sookmyung.ac.kr (S. Yoon).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.10.002

However, it has been recognized that 3D docking can be more effective than 2D topology search or 3D shape matching methods in lead hopping for the selection of a diverse subset of the chemical library [10–13]. Molecular docking is more efficient than shape-matching algorithms for enriching chemicals with diverse scaffolds in the list of virtual hits, although the computational speed is a critical issue in this case. Thus, one of our main goals has been to optimize virtual high-throughput screening approaches to improve the lead-hopping activity while searching for virtual hits. To this end, we designed a "progressive distributed docking" method that iteratively combined a shape-matching method with a multiple-receptor conformation docking method. This method was based on the idea that docking provides new receptor-relevant queries with diverse chemical structures for the next round of the shape search. The ultimate benefit of this method in virtual screening depends not only on the enriching virtual hits but also on the maximizing chemical diversity in selected compounds. In the present study, two targets, PPARγ (peroxisome proliferator-activated receptor γ) and PI3Kγ (phosphoinositide 3-kinase γ), were used for the evaluation of the proposed "progressive distributed docking" method. We comparatively analyzed the receptor relevance (docking score) and the chemical diversity of the selected virtual hits for the iterative and non-iterative methods.

2. Computational methods

2.1. Database construction

We prepared a set of diverse, drug-like compounds from commercially available chemical libraries retrieved from Zinc web site (http://zinc.docking.org). First, the molecules that contained any atoms other than H, C, N, O, F, S, Cl and Br were removed using FILTER (OpenEye, Inc.). Remaining compounds were further filtered to satisfy Lipinski's rules that excluded compounds with poor properties in druglikeness, such as AlogP, molecular weight (MW) and H-bond donor–acceptors (HBDA). The physicochemical properties were calculated using Pipeline Pilot (SciTegic Inc., San Diego, CA, USA). We used the default criteria defined in Pipeline Pilot for the filtering. The final size of compound library for the present analysis was ~1,500,000 non-redundant compounds. The compounds were converted from 2D to 3D SDF format using Ligprep, and the OMEGA program (OpenEye, Inc., v2.1.0) was used to generate multiple conformers for the library compounds. A maximum of 200 conformers was allowed for each compound based on a default RMSD cutoff (0.8 Å) and energy window cutoff (25 kcal/mol).

2.2. Ligand-centric virtual screening (shape comparison)

To make diverse compound subsets that were not biased to a specific region of the chemical space, we eliminated the neighboring compounds with a Tanimoto similarity of >0.5 at each step. The Tanimoto similarity was calculated using a Daylight fingerprint descriptor. Pairwise 3D shape similarities comparisons of the generated conformers against the crystal structure ligands were quantitatively calculated using the ROCS program (OpenEye, Inc., v2.2). This program performs shape-based overlays of the conformers of two molecules and compares a Gaussian-based overlap that is parameterized to reproduce hard-sphere volumes between a query molecule and a single conformation of a database molecule. The two scoring functions that are implemented in ROCS are the Shape Tanimoto score (the shape similarity) and the ColorScore (chemical pattern similarity). All of our experiments were performed with ROCS in color optimization mode. In this mode, ROCS is able to optimize the molecular overlay to maximize the molecular shape overlay and the chemical functionality overlap. ROCS provides two color force fields: the ImplicitMillsDean force field and the ExplicitMillsDean force field. To quantify the matching of the chemical functionalities, we used the ImplicitMillsDean color force field, which is a simple pK_a model that defines cations, anions, donors and acceptors. We then ranked the subset of compounds based on their combo scores (the sum of the ShapeTanimoto score for the overlay and the ScaledColor score), which ranged from 0 to 2, for which a value of 2 represented an exact match of the shape and functional groups to query the molecule and a value of 0 represented no 3D similarity. As a result, the top-scored 10,000, 5000 and 1000 molecules based on ROCS combo score from each three shape-based searches were prepared as each new subsets to obtain the overall recovery of the iterative distributed docking for PPARγ and PI3Kγ.

2.3. Receptor-centric virtual screening (molecular docking)

GLIDE (version 4.0) and ICM (version 3.3) with distinctive features in conformation search and scoring, were used for the docking studies of the receptors. For the molecular docking and scoring, PDB structures of PPARγ and PI3Kγ were prepared using the Schrödinger software package. All of the water molecules were removed, and the multimeric complexes were simplified from the PDB structures. If a PDB structure was missing side-chain atoms, Prime was used to predict their locations. Prior to the molecular docking, the receptor structures were preprocessed using protein preparation and refinement components in the GLIDE docking package. Hydrogen atoms were added using the all-atom force field. Side chains that were not close to the ligand binding site and that did not participate in salt bridges were neutralized. A restrained minimization was performed for the refinement of the complex structure using the OPLS-AA force field. This procedure reoriented the side-chain hydroxyl groups and alleviated potential steric clashes. This minimization continued until the average RMS deviation of the non-hydrogen atoms reached the specified limit of 0.3 Å. The GLIDE docking algorithm performed a series of hierarchical searches for the possible locations of the ligand in the binding site region of the receptor. The details of the GLIDE docking and scoring methods are described elsewhere. All of the compounds in the test sets were energy minimized using LigPrep and were then docked to the selected initial receptor structures using the standard mode of GLIDE docking (GLIDE SP 4.0). For the ICM docking, the initial structures were converted to ICM objects, and the grid maps were calculated with a grid spacing of 0.5 Å. The docking was performed with the default docking parameters. During the docking, either one of the torsional angles of the ligand was randomly changed or a pseudo-Brownian move was performed. Each random change was followed by 100 steps of local conjugate-gradient minimization. The new conformation was either accepted or rejected according to the Metropolis rule, using a temperature of 600 K. The number of Monte Carlo steps in the docking run and the length of the local minimization were determined automatically using an adaptive algorithm, depending on the size and number of the flexible torsions in a ligand.

2.4. Data analysis

To calculate the Tanimoto similarity, we chose a Laplacian-Modified Bayesian Classifier with Functional Connectivity Fingerprints (FCFP) that was implemented in Pipeline Pilot. The query templates for PPARγ and PI3Kγ were selected from this set using the SciTegic fingerprint-based clustering [14]. Cluster 3.0 software (developed based on Eisen Lab's Cluster and Tree View software) was used to generate the 3D queries templates for the initial library screening. The clusters were constructed with an intra-cluster radius of 0.5 (in SciTegic fingerprint tanimoto units), which

Table 1
Identification of the best-scoring receptor conformation for each crystal structure ligand using preliminary cross-docking. The ICM docking score was calculated for all of the crystal structure ligands against all of the crystallized conformations of the receptor. A total of (A) 16 structures of PPARγ and (B) 10 structures of PI3Kγ were cross-docked with their crystal structure ligands. The grey box indicates the best-scoring receptor conformation for the given ligand. The bold numbers represent the best average docking score among the 16 receptor conformations.

	PPARγ PDB structures															
	1FM9	1K74	1FM6	1KNU	1NYX	1I7I	2GTK	1ZEO	2ATH	2F4B	2G0H	2G0G	2HFP	1WM0	2FVJ	4PRG
(A) PPARγ ligands																
1	−63	−19	−5	−9	−18	−22	−11	−43	−17	−23	−15	−24	−19	−4	−23	−19
2	−52	−43	−10	−37	−18	−22	−49	−51	−26	−22	−21	−30	−26	−16	−19	−19
3	−24	−22	−37	−23	−15	−27	−34	−25	−30	−24	−11	−21	−16	−31	−27	−20
4	−52	−28	−28	−41	−16	−47	−53	−43	−40	−32	−19	−26	−17	−27	−22	−11
5	−50	−43	−30	−44	−12	−49	−52	−41	−36	−36	−17	−26	−14	−5	−20	−17
6	−35	−29	−22	−32	−12	−46	−51	−26	−41	−30	−7	−22	−13	−18	−23	−18
7	−51	−22	−26	−44	−13	−16	−58	−39	−33	−28	−19	−21	−21	−16	−24	−20
8	−9	−8	0	−1	−4	−9	−9	−44	−5	−9	−10	4	−14	−9	−15	−14
9	−24	−27	−26	−18	−21	−30	−42	−32	−35	−33	−16	−25	−16	−10	−23	−18
10	−34	−17	−10	−24	−11	−35	−40	−21	−45	−42	−9	−14	−25	−10	−20	−18
11	−25	−17	−13	−25	−23	−26	−20	−28	−24	−18	−24	−33	−14	−18	−30	−22
12	−32	−21	−21	−25	−24	−27	−20	−26	−19	−21	−26	−38	−13	−21	−29	−25
13	−21	−17	−12	−9	−16	−25	−17	−30	−19	−20	−5	−9	−31	−19	−21	−20
14	−28	−22	−14	−23	−20	−25	−14	−26	−13	−19	−14	−15	−28	−41	−31	−23
15	−14	−13	−18	−10	−8	−19	−28	−23	−16	−25	−10	−21	−13	−5	−34	−14
16	−4	−3	−2	−4	−8	−18	−5	−32	−6	−12	7	−17	−5	1	−12	−23
Avg	**−32**	−22	−17	−23	−15	−30	**−31**	**−33**	−25	−25	−14	−21	−18	−16	−23	−18

	PI3Kγ PDB structures									
	2CHZ	2CHX	1E7V	2CHW	1E8Z	2A5U	2A4Z	1E90	1E8W	1E8X
(B) PI3Kγ ligands										
1	−16	−15	−14	−19	−12	−17	−22	−10	−8	−10
2	−25	−31	−15	−7	−20	−16	−2	−20	−16	−13
3	−19	−14	−25	−15	−11	−6	−12	−14	−16	−21
4	−14	−18	−21	−26	−7	8	6	−1	−8	−13
5	−6	−4	−3	−6	−30	21	69	3	6	−5
6	−17	−22	−24	−21	−26	−21	−21	−25	−19	−24
7	−17	−21	−21	−15	−9	−21	−18	−25	−17	−23
8	−19	−25	−18	−22	−5	−41	−32	−29	−22	−8
9	−22	−30	−19	−25	−24	−33	−26	−30	−35	−13
10	−14	−3	−16	−14	−22	11	44	28	36	−16
Avg	−17	**−18**	**−18**	−17	**−17**	−13	−1	−12	−10	−15

meant that each member of a cluster was guaranteed to be within a similarity of 0.5 to its cluster centroid. The hierarchical trees were constructed using the Tree View software, which graphically described the results of the clustering and other analyses from Cluster. The 2D chemical feature calculations were also conducted using various descriptors that are available in SciTegic's Pipeline Pilot. To measure the molecular diversity of the hits for each subset step, we used the Pipeline Pilot FCFP_4 (functional-class fingerprints) topology fingerprints of the molecules as the descriptors in the present study. It has been previously shown that the tanimoto distance between the two bit vector fingerprints is a valid metric; therefore, we used this value as our distance function. The principal component analysis involved a mathematical procedure that transformed a number of possibly correlated variables into a smaller number of uncorrelated variables that were called principal components. We used SciTegic's Pipeline Pilot (version 6.1 for Windows) for the analysis. PCA plots were created using MATLAB.

3. Results and discussion

3.1. Selection of the crystal structures for docking and the shape search

For the purpose of selecting appropriate receptor structures for the iterative docking method, we carried out the preliminary cross docking using a total of 16 and 10 crystal structures of PPARγ and PI3Kγ, respectively, that were co-crystallized with diverse ligands (Table 1). During the preliminary cross docking process, each crystal structure ligand was docked to all of the receptor structures, and the best-scoring receptor structure for each crystal structure ligand was identified (Table 1). In the case of PPARγ, three crystal structures (1FM9, 2GTK and 1ZEO) generated low docking scores for the majority of the diverse crystal ligands; therefore, these structures had a lower average docking score than the other structures. We used these three structures for the iterative dockings of the library compounds. For PI3Kγ, three crystal structures, 1E7V, 1E8Z and 2CHX, exhibited good docking performance with various crystal structure ligands; therefore, these structures were selected for further multiple conformation dockings. In a practical sense, a crystal conformation that yields low docking scores with known ligands is more effective in enriching the true hits from the library compounds than the crystal conformations with high docking scores with known ligands [8,15]. Thus, the presented preliminary cross-docking results will lead to the optimal selection of the receptor conformation for dockings.

Next, we selected a diverse subset of known ligands for the shape search of the chemical library. A total of 16 and 9 crystal ligands that were complexed with PPARγ and PI3Kγ, respectively, were cross-compared by performing shape-matching with the ROCS software. We built a hierarchical tree for each group of the PPARγ and PI3K ligands using the pairwise ROCS similarity scores and then selected the three most distinct ligands from each tree to be the 3D templates (i.e., the query ligands) for the initial library screening. The chemical structures of the selected ligands are displayed in Fig. 1.

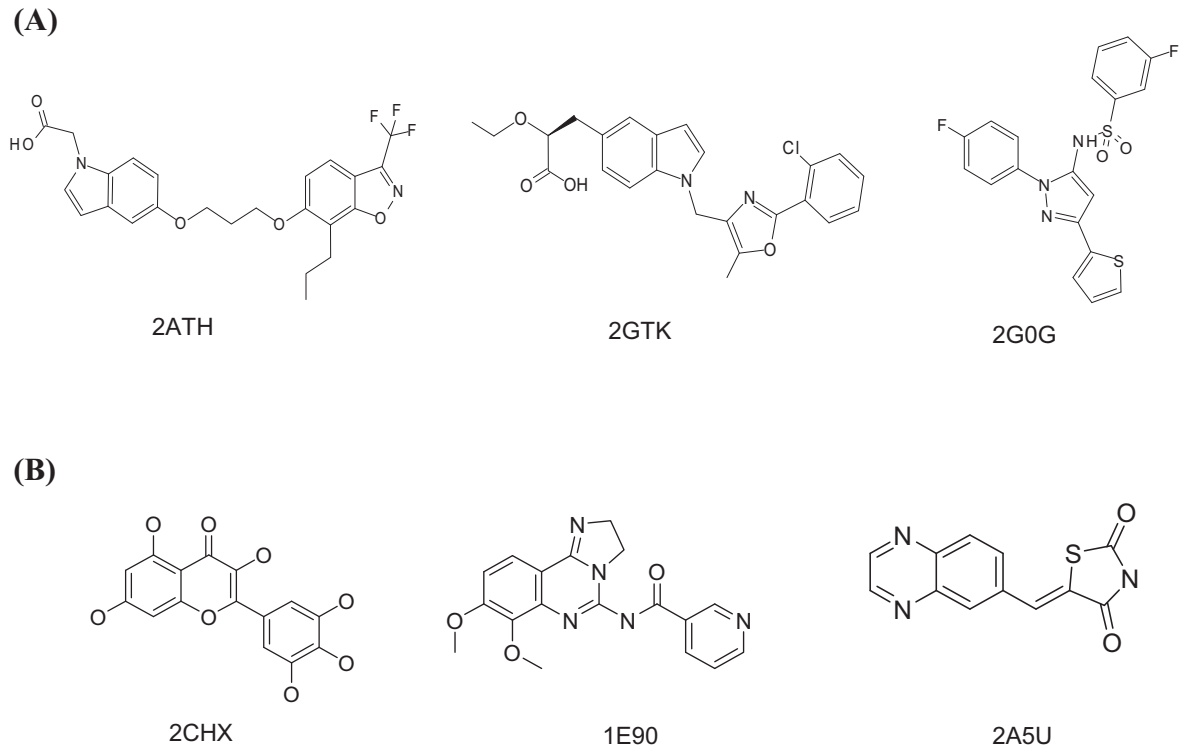

Fig. 1. Chemical structure of the 3D templates for the initial shape-matching. The templates were selected from the crystal structure ligands complexed with (A) PPARγ and (B) PI3Kγ.

3.2. Iterative shape search and docking for the enrichment of virtual hits

The procedure of the proposed progressive shape-matching and docking method is illustrated in Fig. 2. The ligand-centric 3D shape comparison method was conducted. Three selected crystal structure ligands for each protein (PPARγ or PI3Kγ) were used as the initial 3D queries templates for our library compound screening. The ROCS similarity search was conducted for each of the selected queries (i.e., crystal structure ligands) against all of the compounds in the library. The best score among three searches was determined for each of library compounds. To exclude the compounds that were topologically similar to the query molecules from the search, those with a Tanimoto similarity score of >0.5 to any of the initial crystal queries were excluded. As a result, three subsets of virtual hits (sampling size: 10,000, 5000 and 1000 molecules) with a high 3D similarity to one of the query ligands were selected based on the ROCS combo score (see Section 2 for details).

The next step of the structure-based virtual screening (docking) was conducted using the distributed docking method, which is a compromise between the accuracy of the multiple conformation docking and the speed of single conformation docking. Each of the selected virtual hits from the preceding step was docked to the selected PPARγ and PI3Kγ crystal structures using ICM and GLIDE. From the results of the docking, we calculated the hit rates using the ICM and GLIDE docking scores. The cutoffs for defining the hits were −35 for the ICM docking and −10.5 for the GLIDE docking for PPARγ and −25 for the ICM docking and −8.5 for the GLIDE docking

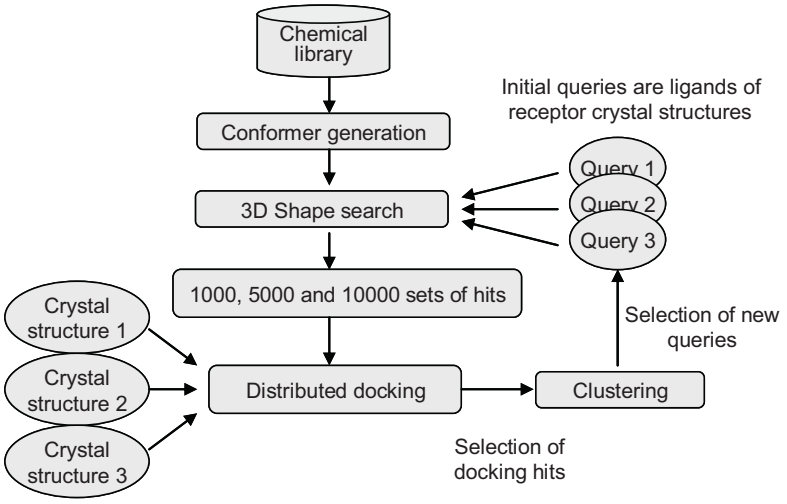

Fig. 2. Schematic representation of the progressive iteration of shape-matching and docking methods.

Table 2
Number of virtual hits obtained by the proposed progressive shape search and docking method. The hits were selected based on the docking score.

	ICM docking			GLIDE docking		
Sampling size	1000	5000	10,000	1000	5000	10,000
(A) PPARγ						
Control*	10	88	213	15	132	320
Iteration						
1st	91	439	890	110	509	1030
2nd	119	526	2127	113	833	2077
3rd	142	721	1953	346	778	2784
4th	165	486	–	290	1437	–
5th	127	758	–	266	1139	–
Sum	644	2930	4970	1125	4696	5892
(B) PI3Kγ						
Control*	6	47	102	14	109	238
Iteration						
1st	54	212	375	133	482	831
2nd	120	567	851	248	438	586
3rd	77	513	820	137	333	633
4th	55	553	–	137	349	–
5th	33	691	–	97	214	–
Sum	339	2536	2046	752	1816	2050

* A control set for the given sample size was selected randomly from the chemical library without the shape search.

for PI3Kγ. The cutoffs were established based on the docking score of known ligands to a crystal receptor conformation (Table 1) [8].

The docking hits from each sampling size were used as a 3D template for the next round of the shape search of the library, thereby generating a new subset for the given sampling size. This new subset was used for the next round of docking. We carried out a total of five iterations of the shape search and docking for the 1000 and 5000 sampling sizes. For the sampling size of 10,000 compounds, three iterations were conducted.

The number of docking hits was greater in all of the subsets that were selected from the shape search than the control sets that were randomly selected from the library (Table 2). The shape search, which is based on the 3D similarity to the crystal structure ligands, provides an improved enrichment of the receptor-relevant compounds with good docking scores. In addition, the number of hits increased in most cases as the iteration continued (Table 2 and Fig. 3). From each of the iterative processes, the docking hits were clustered based on the 2D similarity (FCFP descriptor in Pipeline Pilot) of the compounds. A compound with the best docking score that was within the cluster was selected as the cluster-representing compound. Next, the 3D query compounds for the next-step shape search were selected from the cluster-representing compounds of the three distinctive clusters.

We analyzed the efficiency of the progressive docking by counting the number of accumulative docking hits in the course of the iterations. First, the number of docking hits from the 5000-compound selection was compared with that of the docking hits from five sets of the 1000-compound selection in the five iterations (Fig. 3). In both of the targets (PPARγ and PI3Kγ), the progressive selection of the compounds provided more docking hits than the one-step selection of the compounds. Likewise, two iterations of the 5000-compound selection provided a better docking efficiency than the one-step selection of the 10,000-compound group. The ICM and GLIDE docking procedures had consistent results, confirming the efficiency of the iterative shape search and docking. This finding indicates that the iterative selection of compounds using a shape search increases the number of docking hits more efficiently than the one-step selection of compounds.

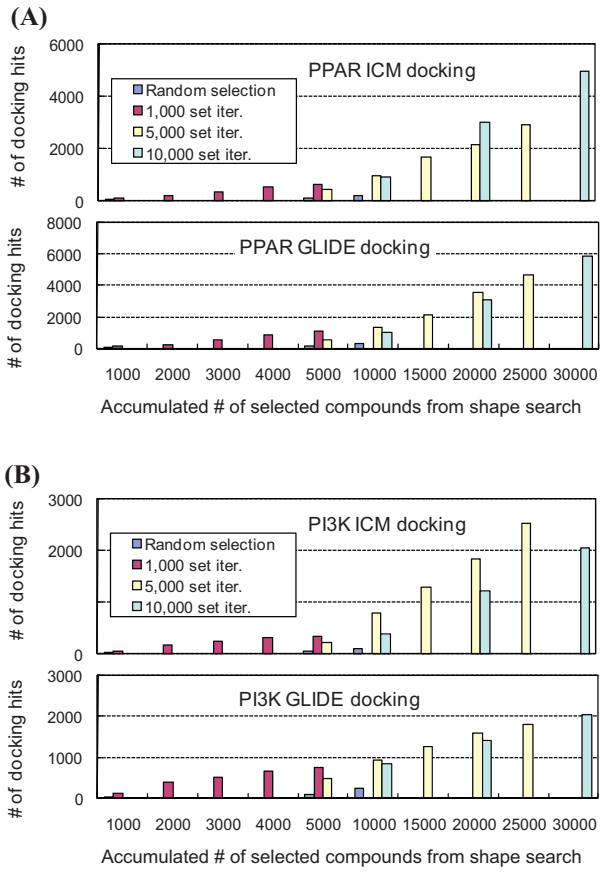

Fig. 3. Accumulation of the virtual hits by iterative shape search and docking. ICM and GLIDE dockings against (A) PPARγ and (B) PI3Kγ were conducted on each set of compounds with different size that were selected by the preceding shape search.

3.3. Chemical diversity in the virtual hits

We elaborated on the progressive shape search and docking to obtain an optimal list of virtual hits for a given target protein. As described above, the iterative combination of the shape search and docking exhibited a better enrichment of the virtual hits than the random selection or the non-iterative combination of the shape search and docking. Lead hopping is another important aspect of virtual screening that is used to maximize the chemical diversity of the candidate compounds for experimental screening [16–19]. In the present work, we also comparatively analyzed the chemical diversity of the virtual hits using a simple approach. To quantitatively compare the chemical diversity of the docking hits among the different methods, a principal component analysis (PCA) was conducted using the 2D fingerprint of the compounds that were generated using SciTegic Pipeline Pilot [20–22]. The docking hits from each set of the random selection, iterative and non-iterative procedures were separately used for the PCA. The principal components (PCs) identify orthogonal directions of variation within the data. The first principal components account for as much as possible of the variance in the data. Therefore, the first three components of the PCA were used to comparatively display the occupied chemical space of the docking hits in the 3D space (Figs. 4 and 5).

Table 3 summarizes the coverage of the chemical space that was occupied by the docking hits from the three methods (random selection, non-iterative shape search and iterative shape search). As a result, it is clear that the iterative shape search method provides greater coverage of the chemical space in the docking hits than the random selection or non-iterative shape search methods. For the

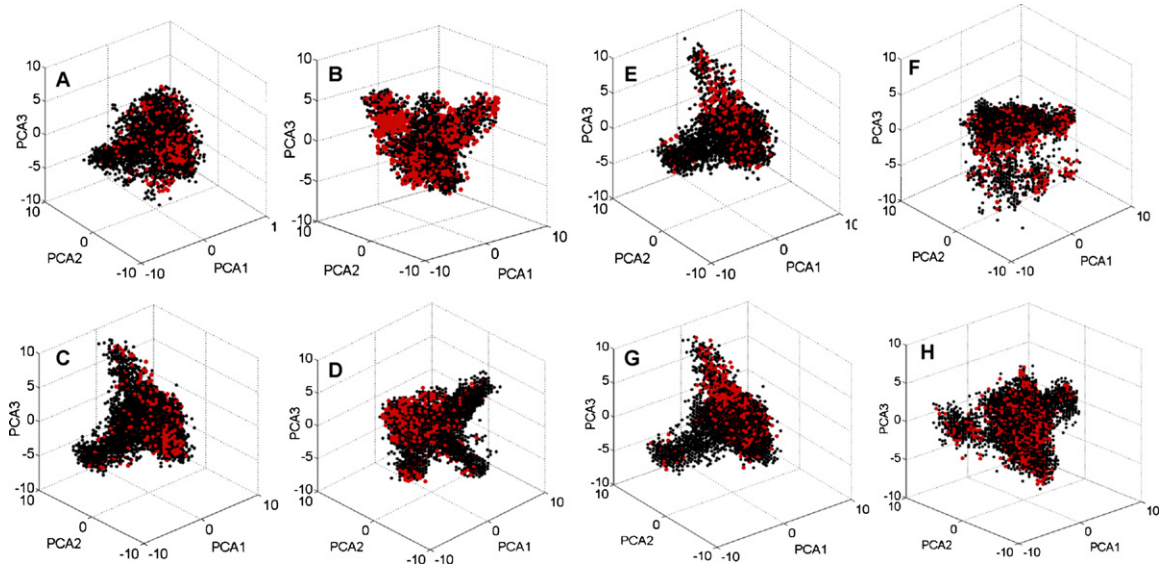

Fig. 4. Comparative analysis of the chemical diversity among the virtual hits that were prepared using the different methods. 3D plots of the PCA are displayed for the 2D chemical descriptor of the selected docking hits (red) and the total compounds from the shape-based search (black). GLIDE and ICM docking were conducted against PPARγ. The bin size of the PCA plot was 0.5. The plot was based on the first three Principle Components (PCs) that were calculated using the "FCFP_4" descriptors of the SciTegic Pipeline Pilot. Each panel represents a different selection of compounds (black dots): GLIDE (A–D) and ICM (E–H): (A and E) no iterations, 5000 compounds, (B and F) five iterations, 1000 compounds per iteration, (C and G) no iterations, 10,000 compounds, (D and H) two iterations, 5000 compounds per iteration.

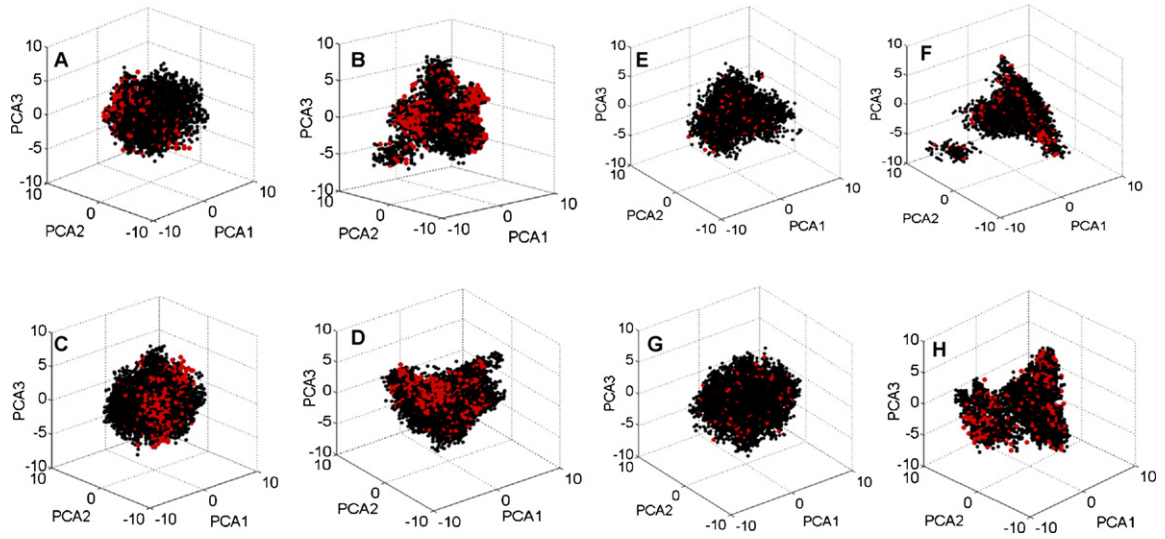

Fig. 5. Comparative analysis of the chemical diversity among the virtual hits that were prepared using the different methods. 3D plots of the PCA are displayed for the 2D chemical descriptor of the selected docking hits (red) and the total compounds from the shape-based search (black). GLIDE and ICM docking were conducted against PI3Kγ. The bin size of the PCA plot was 0.5. The plot was based on the first three PCs that were calculated using the "FCFP_4" descriptors of the SciTegic Pipeline Pilot. Each panel represents a different selection of compounds (black dots): GLIDE (A–D) and ICM (E–H): (A and E) no iterations, 5000 compounds, (B and F) five iterations, 1000 compounds per iteration, (C and G) no iterations, 10,000 compounds, (D and H) two iterations, 5000 compounds per iteration.

Table 3
Comparison of the chemical space occupied by the docking hits for PPARγ (A) and PI3Kγ (B). For each sampling size (5000 and 10,000), we compared the chemical space of the docking hits from the three sampling methods (random selection, non-iterative shape search, five iterations of shape search). The coverage represents the number of PCA bins in Figs. 4 and 5. The ratio is the relative coverage of the docking hits across all of the library compounds.

		5000 selection			10,000 selection		
		Random	No iteration	Five iterations	Random	No iteration	Three iterations
(A)							
ICM docking	Coverage	88	384	510	198	731	761
	Ratio	0.033	0.154	0.215	0.054	0.198	0.218
GLIDE docking	Coverage	130	447	795	287	819	874
	Ratio	0.053	0.178	0.347	0.088	0.221	0.270
(B)							
ICM	Coverage	47	95	286	99	338	609
	Ratio	0.017	0.067	0.116	0.027	0.080	0.174
GLIDE	Coverage	105	414	581	328	666	716
	Ratio	0.043	0.143	0.222	0.088	0.158	0.199

ICM and GLIDE docking procedures, the coverage of chemical space and its ratio over the total chemical space were increased after the iterative selection of compounds against PI3Kγ and PPARγ. Five iterations exhibited a greater improvement in the chemical coverage than three iterations for both of the docking methods. This observation confirms that a greater number of iterations that use a smaller sampling size is more effective than a fewer number of iterations that use a larger sampling size at obtaining a diversity of chemical scaffolds in the docking hits.

4. Conclusions

In the present study, we focused on the development and validation of the iterative combination of the shape search and docking methods, which aimed to maximize the enrichment of target-relevant virtual hits and their chemical diversity. The results indicate that an optimal combination of shape comparison and docking methods provide a better selection of the target-oriented virtual hits. Specifically, the iterative application of the combination of the shape search and docking methods was found to improve the hit enrichment and the efficacy of the lead hopping. The purpose of lead hopping is to find chemically novel molecules, starting from known active compounds. This technique requires a combination of several methods, which are described in the present study. We thoroughly validated this efficient method with the selection of PPARγ and PI3Kγ-oriented compounds. These progressive docking approaches that were based on the distributed docking strategy also meet the requirements for high-throughput virtual screening by shortening the amount of time that is needed for multiple conformation docking. This method has the potential for broad applications in various ligand-centric and receptor-centric virtual screening software programs for innovative drug discovery projects.

Acknowledgement

This research was supported by Sookmyung Women's University Research Grants 1-0903-0229.

References

[1] S. Yoon, A. Smellie, D. Hartsough, A. Filikov, Surrogate docking: structure-based virtual screening at high throughput speed, J. Comput. Aided Mol. Des. 19 (2005) 483–497.
[2] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: current status and future challenges, Proteins 65 (2006) 15–26.
[3] G. Klebe, Virtual ligand screening: strategies, perspectives and limitations, Drug Discov. Today 11 (2006) 580–594.
[4] F.L. Stahura, J. Bajorath, Virtual screening methods that complement HTS, Comb. Chem. High Throughput Screen. 7 (2004) 259–269.
[5] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (2004) 935–949.
[6] T. Tuccinardi, Docking-based virtual screening: recent developments, Comb. Chem. High Throughput Screen. 12 (2009) 303–314.
[7] A.N. Jain, Scoring functions for protein–ligand docking, Curr. Protein Pept. Sci. 7 (2006) 407–420.
[8] H.S. Lee, J. Choi, I. Kufareva, R. Abagyan, A. Filikov, Y. Yang, et al., Optimization of high throughput virtual screening by combining shape-matching and docking methods, J. Chem. Inf. Model. 48 (2008) 489–497.
[9] H.A. Carlson, J.A. McCammon, Accommodating protein flexibility in computational drug design, Mol. Pharmacol. 57 (2000) 213–218.
[10] B.B. Yan, M.Z. Xue, B. Xiong, K. Liu, D.Y. Hu, J.K. Shen, ScafBank: a public comprehensive Scaffold database to support molecular hopping, Acta Pharmacol. Sin. 30 (2009) 251–258.
[11] F. Fontaine, S. Cross, G. Plasencia, M. Pastor, I. Zamora, SHOP: a method for structure-based fragment and scaffold hopping, ChemMedChem 4 (2009) 427–439.
[12] B.A. Krueger, A. Dietrich, K.H. Baringhaus, G. Schneider, Scaffold-hopping potential of fragment-based de novo design: the chances and limits of variation, Comb. Chem. High Throughput Screen. 12 (2009) 383–396.
[13] F. Quintus, O. Sperandio, J. Grynberg, M. Petitjean, P. Tuffery, Ligand scaffold hopping combining 3D maximal substructure search and molecular similarity, BMC Bioinformatics 10 (2009) 245.
[14] P. Prathipati, N.L. Ma, T.H. Keller, Global Bayesian models for the prioritization of antitubercular agents, J. Chem. Inf. Model. 48 (2008) 2362–2370.
[15] S. Yoon, W.J. Welsh, Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring, J. Chem. Inf. Comput. Sci. 44 (2004) 88–96.
[16] S.W. Muchmore, D.A. Debe, J.T. Metz, S.P. Brown, Y.C. Martin, P.J. Hajduk, Application of belief theory to similarity data fusion for use in analog searching and lead hopping, J. Chem. Inf. Model. 48 (2008) 941–948.
[17] J.C. Saeh, P.D. Lyne, B.K. Takasaki, D.A. Cosgrove, Lead hopping using SVM and 3D pharmacophore fingerprints, J. Chem. Inf. Model. 45 (2005) 1122–1133.
[18] R.S. Vijayan, I. Bera, M. Prabu, S. Saha, N. Ghoshal, Combinatorial library enumeration and lead hopping using comparative interaction fingerprint analysis and classical 2D QSAR methods for seeking novel GABA(A) alpha(3) modulators, J. Chem. Inf. Model. 49 (2009) 2498–2511.
[19] R.D. Cramer, R.J. Jilek, S. Guessregen, S.J. Clark, B. Wendt, R.D. Clark, "Lead hopping". Validation of topomer similarity as a superior predictor of similar biological activities, J. Med. Chem. 47 (2004) 6777–6791.
[20] Z. Wang, Y. Lu, W. Seibel, D.D. Miller, W. Li, Identifying novel molecular structures for advanced melanoma by ligand-based virtual screening, J. Chem. Inf. Model. 49 (2009) 1420–1427.
[21] R. Mannhold, G. Cruciani, K. Dross, R. Rekker, Multivariate analysis of experimental and computational descriptors of molecular lipophilicity, J. Comput. Aided Mol. Des. 12 (1998) 573–581.
[22] D. Audisio, S. Messaoudi, I. Ijjaali, E. Dubus, F. Petitet, J.F. Peyrat, et al., Assessing the chemical diversity of an hsp90 database, Eur J. Med. Chem. 45 (2010) 2000–2009.