



Pharmacophore modeling and applications in drug discovery: challenges and recent advances

Sheng-Yong Yang

State Key Laboratory of Biotherapy and Cancer Center, West China Hospital, West China Medical School, Sichuan University, Chengdu, Sichuan 610041, China

Pharmacophore approaches have become one of the major tools in drug discovery after the past century's development. Various ligand-based and structure-based methods have been developed for improved pharmacophore modeling and have been successfully and extensively applied in virtual screening, *de novo* design and lead optimization. Despite these successes, pharmacophore approaches have not reached their expected full capacity, particularly in facing the demand for reducing the current expensive overall cost associated with drug discovery and development. Here, the challenges of pharmacophore modeling and applications in drug discovery are discussed and recent advances and latest developments are described, which provide useful clues to the further development and application of pharmacophore approaches.

Introduction

The concept of pharmacophore was first introduced in 1909 by Ehrlich [1], who defined the pharmacophore as 'a molecular framework that carries (*phoros*) the essential features responsible for a drug's (*pharmakon*) biological activity'. After a century's development, the basic pharmacophore concept still remains unchanged, but its intentional meaning and application range have been expanded considerably. According to the very recent definition by IUPAC [2], a pharmacophore model is 'an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response'. Apart from this official definition, some other similar definitions, as well as remarks, have been described in the literature [3–5]. The overall development and history of the pharmacophore concept through the past century has been reviewed by Günd [3] and Wermuth [4].

A pharmacophore model can be established either in a ligand-based manner, by superposing a set of active molecules and extracting common chemical features that are essential for their bioactivity, or in a structure-based manner, by probing possible interaction points between the macromolecular target and ligands. Pharmacophore approaches have been used extensively in virtual screening, *de novo* design and other applications such as

lead optimization and multitarget drug design (Fig. 1). A variety of automated tools for pharmacophore modeling and applications appeared constantly after the advances in computational chemistry in the past 20 years; these pharmacophore modeling tools, together with their inventor(s) and typical characteristics, are summarized in Supplementary Table S1. Many successful stories of pharmacophore approaches in facilitating drug discovery have been reported in recent years [6,7]. The pharmacophore approach, however, still faces many challenges that limit its capability to reach its expected potential, particularly with the demand for reducing the current high cost associated with the discovery and development of a new drug. This article discusses the challenges of pharmacophore modeling and applications in drug discovery and reviews the most recent advances in dealing with these challenges.

Ligand-based pharmacophore modeling

Ligand-based pharmacophore modeling has become a key computational strategy for facilitating drug discovery in the absence of a macromolecular target structure. It is usually carried out by extracting common chemical features from 3D structures of a set of known ligands representative of essential interactions between the ligands and a specific macromolecular target. In general, pharmacophore generation from multiple ligands (usually called training set compounds) involves two main steps:

E-mail address: yangsy@scu.edu.cn.

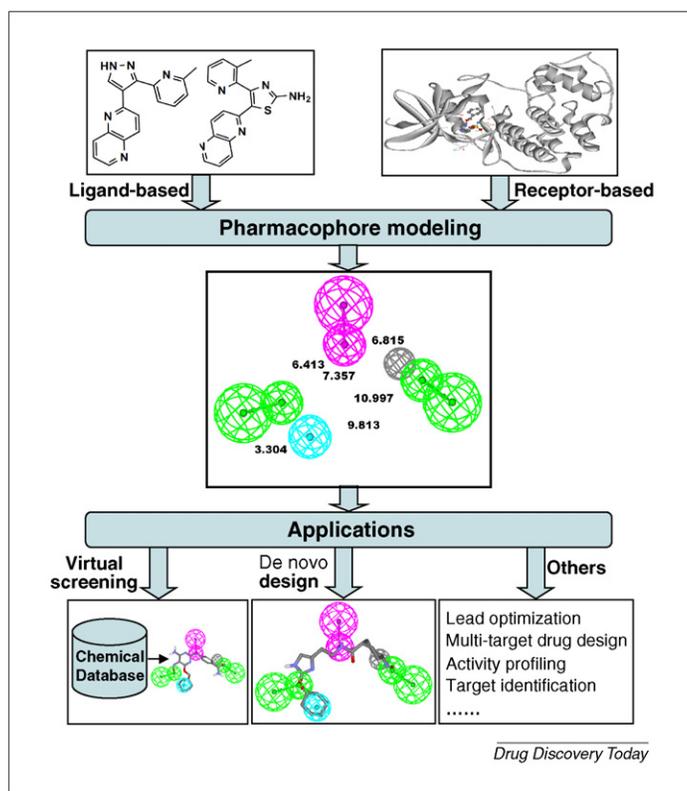


FIGURE 1

The full framework of pharmacophore architecture.

creating the conformational space for each ligand in the training set to represent conformational flexibility of ligands, and aligning the multiple ligands in the training set and determining the essential common chemical features to construct pharmacophore models. Handling conformational flexibility of ligands and conducting molecular alignment represent the key techniques and also the main difficulties in ligand-based pharmacophore modeling. Currently, various automated pharmacophore generators have been developed, including commercially available software – such as HipHop [8], HypoGen [9] (Accelrys Inc., <http://www.accelrys.com>), DISCO [10], GASP [11], GALAHAD (Tripos Inc., <http://www.tripos.com>), PHASE [12] (Schrödinger Inc., <http://www.schrodinger.com>) and MOE (Chemical Computing Group, <http://www.chemcomp.com>) – and several academic programs. These programs differ mainly in the algorithms used for handling the flexibility of ligands and for the alignment of molecules, which are outlined in Supplementary Table S1. There are some references in literature, such as Refs. [5,13,14], showing the differences, advantages and disadvantages of these programs; however, describing and analyzing the different programs is not our goal here.

Despite the great advances, several key challenges in ligand-based pharmacophore modeling still exist. The first challenging problem is the modeling of ligand flexibility. Currently, two strategies have been used to deal with this problem: the first is the pre-enumerating method, in which multiple conformations for each molecule are precomputed and saved in a database [13]. The second is the on-the-fly method, in which the conformation analysis is carried out in the pharmacophore modeling process

[13]. The first approach has the advantage of lower computing cost for conducting molecular alignment at the expense of a possible need for a mass storage capacity. The second approach does not need mass storage but might need higher CPU time for conducting rigorous optimization. It has been demonstrated that the pre-enumerating method outperforms the on-the-fly calculation approach [15]. Currently, a substantial number of advanced algorithms have been established to sample the conformational spaces of small molecules, which are listed in Supplementary Table S2. Some of these algorithms, such as poling restraints [16], systematic torsional grids [17], directed tweak [18], genetic algorithms [19] and Monte Carlo [20], have been implemented in various commercial and academic pharmacophore modeling programs. Nevertheless, a good conformation generator should satisfy the following conditions: (i) efficiently generating all the putative bound conformations that small molecules adopt when they interact with macromolecules, (ii) keeping the list of low-energy conformations as short as possible to avoid the combinational explosion problem and (iii) being less time-consuming for the conformational calculations. Several new or modified tools developed recently for conformational generation seem to outperform the previous algorithms in some aspects. MED-3DMC, developed by Sperandio *et al.* [21], uses a combination of the Metropolis Monte Carlo algorithm, based on a SMARTS mapping of the rotational bond, and the MMFF94 van der Waals energy term. MED-3DMC has been reported to outperform Omega when applied on certain molecules with a low to medium number of rotatable bonds [21]. Liu *et al.* [22] developed a conformation sampling method named 'Cyndi', which is based on a multiobjective evolution algorithm. Cyndi was validated to be markedly superior to other conformation generators in reproducing the bioactive conformations against a set of 329 testing structures [22]. CAESAR [23] is another conformer generator, which is based on a divide-and-conquer and recursive conformer buildup approach. This approach also takes into consideration local rotational symmetry to enable the elimination of conformer duplicates owing to topological symmetry in the systematic search. CAESAR has been demonstrated to be consistently 5–20 times faster than Catalyst/FAST.¹ The speedup is even more notable for molecules with high topological symmetry or for molecules that require a large number of conformational samplings.

Molecular alignment is the second challenging issue in ligand-based pharmacophore modeling. The alignment methods can be classified into two categories in terms of their fundamental nature: point-based and property-based approaches [15]. The points (in the point-based method) can be further differentiated as atoms, fragments or chemical features [5]. In point-based algorithms, pairs of atoms, fragments or chemical feature points are usually superimposed using a least-squares fitting. The biggest limitation of these approaches is the need for predefined anchor points because the generation of these points can become problematic in the case of dissimilar ligands. The property-based algorithms make use of molecular field descriptors, usually represented by sets of Gaussian functions, to generate alignments. The alignment optimization is carried out with some variant of similarity measure

¹ Catalyst is now incorporated into Discovery Studio, available from Accelrys Inc., San Diego, CA, USA.

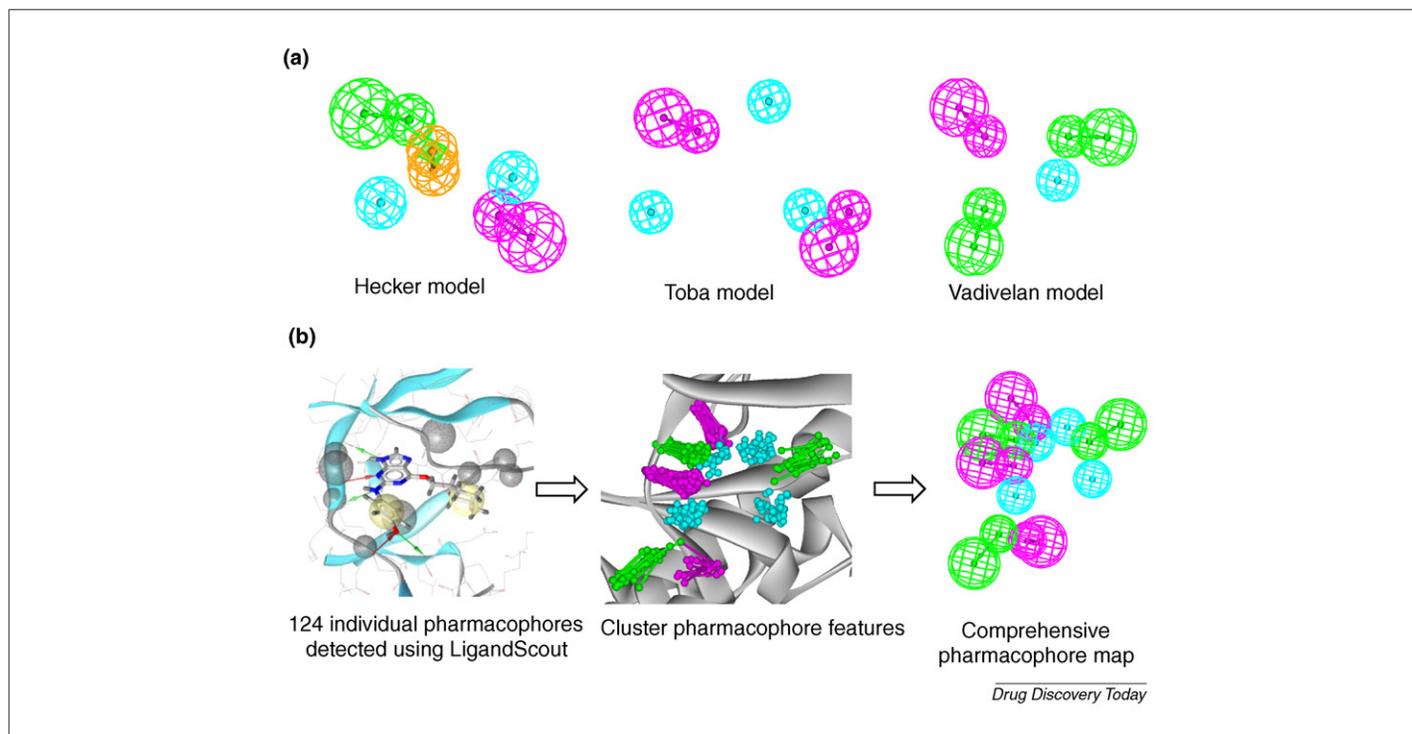


FIGURE 2

Pharmacophore models of cyclin-dependent kinase 2 (CDK2) inhibitors. **(a)** Pharmacophore models of CDK2 inhibitors developed using Catalyst by Hecker *et al.* [28], Toba *et al.* [29] and Vadivelan *et al.* [30]. **(b)** The basic process for the generation of multicompound-based comprehensive pharmacophore map of CDK2 inhibitors. The chemical features are color coded: green, hydrogen-bond acceptor; magenta, hydrogen-bond donor; light blue, hydrophobic feature; orange, aromatic ring.

of the intermolecular overlap of the Gaussians as the objective function. Conventional molecular alignment algorithms have been extensively reviewed elsewhere [15]. New alignment methods continue to be actively developed. Recently developed methods include stochastic proximity embedding [24], atomic property fields [25], fuzzy pattern recognition [26] and grid-based interaction energies [27].

Another challenging problem lies in the practical task of proper selection of training set compounds. This problem, apparently being simple and non-technical, often confuses users, even experienced ones. It has been demonstrated that the type of ligand molecules, the size of the dataset and its chemical diversity affect the final generated pharmacophore model considerably [13]. In some cases, completely different pharmacophore models of ligands interacting with the same macromolecular target could be generated from the same algorithm and program that uses different training sets. For example, Hecker *et al.* [28], Toba *et al.* [29] and Vadivelan *et al.* [30] have independently generated three pharmacophore models of cyclin-dependent kinase 2 (CDK2) inhibitors. They used the same program, Catalyst, but different training sets. The three pharmacophore models are found to be totally different from one another in terms of the feature categories, as well as the location constraints of features (Fig. 2a), for which a further discussion is presented in a subsequent section of this review.

Structure-based pharmacophore modeling

Structure-based pharmacophore modeling works directly with the 3D structure of a macromolecular target or a macromolecule–

ligand complex. The protocol of structure-based pharmacophore modeling involves an analysis of the complementary chemical features of the active site and their spatial relationships, and a subsequent pharmacophore model assembly with selected features. The structure-based pharmacophore modeling methods can be further classified into two subcategories: macromolecule–ligand-complex based and macromolecule (without ligand)-based. The macromolecule–ligand-complex-based approach is convenient in locating the ligand-binding site of the macromolecular target and determining the key interaction points between ligands and macromolecule. LigandScout [31] is an excellent representation that incorporates the macromolecule–ligand-complex-based scheme. Other macromolecule–ligand-complex-based pharmacophore modeling programs include Pocket v.2 [32] and GBPM [33]. The limitation of this approach is the need for the 3D structure of macromolecule–ligand complex, implying that it cannot be applied to cases when no compounds targeting the binding site of interest are known. This can be overcome by the macromolecule-based approach. The structure-based pharmacophore (SBP) method² implemented in Discovery Studio³ is a typical example of a macromolecule-based approach. SBP converts LUDI [34] interaction maps within the protein-binding site into Catalyst pharmacophoric features: H-bond acceptor, H-bond donor and hydrophobe. The main limitation of the SBP flowchart is that the derived interaction maps generally consist of a large number of

² SBP is now incorporated into Discovery Studio, available from Accelrys Inc., San Diego, CA, USA.

³ Discovery Studio available from Accelrys Inc., San Diego, CA, USA.

unprioritized Catalyst features, which complicates its application in such tasks as 3D database searches. To overcome this problem, Barillari *et al.* [35] recently proposed a fast knowledge-based approach, hot-spots-guided receptor-based pharmacophores (HS-Pharm). This approach enables the prioritization of cavity atoms that should be targeted for ligand binding, by training machine learning algorithms with atom-based fingerprints of known ligand-binding pockets. Tintori *et al.* [36] have also reported another apoprotein-based approach. With this approach, the GRID [37] molecular interaction fields (MIFs) are first calculated by using different probes for the binding site of interest, followed by the selection and subsequent conversion of the points of minimum of MIFs into pharmacophoric features.

A frequently encountered problem for structure-based pharmacophore modeling, not only the macromolecule-based approach, is that too many chemical features (generally not prioritized) can be identified for a specific binding site of the macromolecular target. However, a pharmacophore model composed of too many chemical features (for example, >7 chemical features) is not suitable for practical applications, such as 3D database screening. Thus, it is necessary to select a limited number of chemical features (typically three to seven features) to construct a practical pharmacophore hypothesis, although this is not an easy task in many cases. Another problem is that the obtained pharmacophore hypothesis cannot reflect the quantitative structure–activity relationship (QSAR) because the model is derived just based on a single macromolecule–ligand complex or a single macromolecule. In an attempt to overcome these problems, we, in a recent study, have suggested using a multicomplex-based comprehensive map and most-frequent pharmacophore model [38]. In that study, a multicomplex-based method was used to generate a comprehensive pharmacophore map of CDK2 based on a collection of 124 crystal structures of human CDK2 inhibitor complex. The chemical features for each complex were first identified by LigandScout, followed by clustering all the features to form a comprehensive pharmacophore map. The established pharmacophore map contains almost all the chemical features important for CDK2–inhibitor interactions (Fig. 2b). We found that, with the exception of a feature of aromatic ring (orange) in Hecker model, every pharmacophore feature in the reported ligand-based models (Hecker model, Toba model and Vadivelan model; Fig. 2a) can be matched to a feature in our comprehensive map, suggesting that these ligand-based models are subgraphs of our comprehensive map. The only exception (the aromatic ring feature in Hecker model) seems to occur in ligand scaffolds. A chemical feature occurred in small molecular scaffolds, which should be a pseudo-pharmacophore feature because it does not represent a ligand–macromolecule interaction, cannot be detected by a structure-based pharmacophore modeling approach. Because the comprehensive pharmacophore map is too restrictive and not suitable for the virtual screening, a reduced model is needed for a real application. A feasible solution is to select the most-frequent features that were recognized as the features important to the activity of the CDK2 inhibitors. Thus, the top-ranked seven features, which are present in the 124 complexes with more than 25% probability, have been selected and combined to form a most-frequent-feature pharmacophore model. Validation studies of the most-frequent-feature model have shown not only that it can discriminate successfully

between known CDK2 inhibitors and the molecules of focused inactive dataset but also that it is capable of correctly predicting the activities of a wide variety of CDK2 inhibitors in an external active dataset [38].

Pharmacophore-model-based virtual screening

Once a pharmacophore model is generated by either the ligand-based or the structure-based approach, it can be used for querying the 3D chemical database to search for potential ligands, which is so-called ‘pharmacophore-based virtual screening’ (VS). Pharmacophore-based VS and docking-based VS represent the mainstream of VS tools at the present time. In contrast to its counterpart, the docking-based VS method, pharmacophore-based VS reduces the problems arising from inadequate consideration of protein flexibility or the use of insufficiently designed or optimized scoring functions by introducing a tolerance radius for each pharmacophoric feature.

In the pharmacophore-based VS approach, a pharmacophore hypothesis is taken as a template. The purpose of screening is actually to find such molecules (hits) that have chemical features similar to those of the template. Some of these hits might be similar to known active compounds, but some others might be entirely novel in scaffold. The searching for compounds with different scaffolds, while sharing a biological activity is usually called ‘scaffold hopping’ [39]. The screening process involves two key techniques and difficulties: handling the conformational flexibility of small molecules and pharmacophore pattern identification. The strategies for handling the flexibility of small molecules in pharmacophore-based VS are very similar to those used in pharmacophore modeling. Again, the flexibility of small molecules is handled by either pre-enumerating multiple conformations for each molecule in the database or conformational sampling at search time. Pharmacophore pattern identification, usually called ‘substructure searching’, is actually to check whether a query pharmacophore is present in a given conformer of a molecule. The frequently used approaches for substructure searching are based on graph theory, which include Ullmann [40], the backtracking algorithm [41], and the GMA algorithm [42].

Pharmacophore-based VS can be very time-consuming, especially in cases of screening large chemical databases with flexible molecules, which is currently a key challenge in pharmacophore-based VS. A commonly used method to speed up the screening process is the multilevel searching approach [5]. In this approach, a series of screening filters are applied to the molecules in an increasing order of complexity so that the first filters are fast and simple, whereas successive ones are more time-consuming but are applied only to a small subset of the entire database.

However, the most challenging problem for pharmacophore-based VS is that in many cases, few percentages of the virtual hits are really bioactive; in other words, the screening results bear a higher ‘false positive’ rate and/or a higher ‘false negative’ rate. Many factors can contribute to this problem, including the quality and composition of the pharmacophore model and whether and how much the macromolecular target information is involved. First, the most apparent factor is associated with the deficiency of a pharmacophore hypothesis. To address this problem requires a comprehensive validation and optimization to the pharmacophore model. Various validation methods such as cross-validation

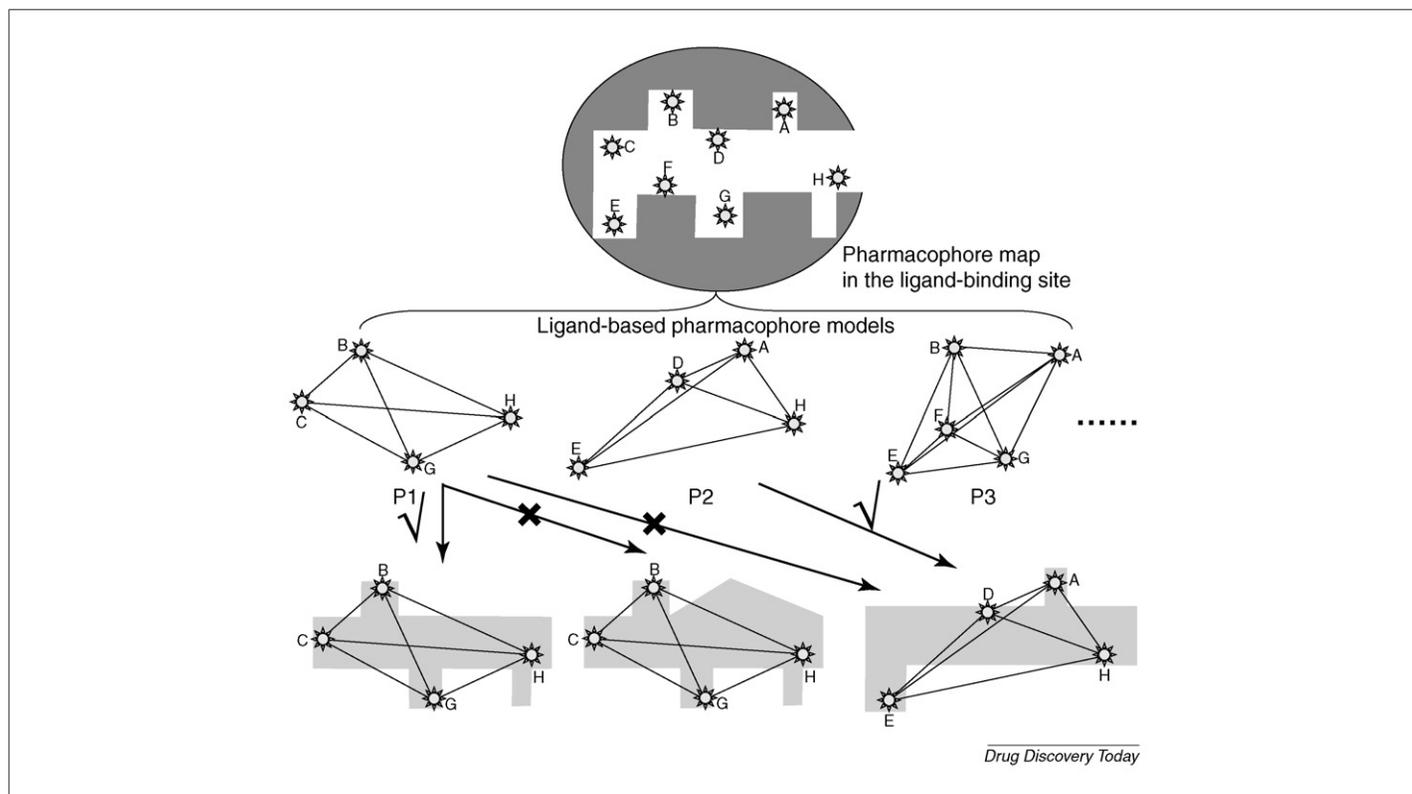


FIGURE 3

Schematic representation of the relationship between the full pharmacophore map identified by the structure-based approach and the pharmacophore models established by the ligand-based method. The upper schematically shows the chemical features in the ligand-binding site of a macromolecular target. The middle indicates the various pharmacophore models established by a ligand-based pharmacophore modeling method. The lower schematically depicts several possible cases in pharmacophore-based virtual screening. A tick (✓) means that the selected molecule conforms to the requirements of pharmacophore and the shape of ligand-binding site. A cross (×) indicates that the selected molecule does not satisfy the requirements of either pharmacophore or the shape of ligand-binding site (because of the atomic bumping).

and test set method have been suggested, which were reviewed recently by Triballeau *et al.* [43]. The validation process is usually associated with model optimization. Lately, Sun *et al.* [44] have developed a genetic algorithm-guided pharmacophore query optimization program, in which the optimization is carried out by automatically adjusting the position and tolerance radius of each pharmacophoric feature. The final query has been validated by using a test set method, which shows a considerably improved hit rate. Second, because the pharmacophore model used for 3D query is generally one of the subgraphs of the full pharmacophore map, screening with this pharmacophore query might not retrieve molecules that match other subgraphs except for the selected one, which is probably an important reason for the higher false negative rate in some studies (Fig. 3). Third, the flexibility of target macromolecule in pharmacophore approaches is handled by introducing a tolerance radius for each pharmacophoric feature, which is unlikely to fully account for macromolecular flexibility in some cases. Some recent attempts [45,46] to incorporate molecular dynamics simulations in pharmacophore modeling have suggested that the dynamics pharmacophore models generated from MD simulation trajectories show considerably better representation of the flexibility of pharmacophore.

Another factor that might lead to the high false positive rate is that the steric restriction by the macromolecular target is not sufficiently considered in pharmacophore models, although it is partly counted for by the consideration of excluded volumes. In

addition, most of the interactions between ligand and protein are distance sensitive – particularly the short-range interactions, such as the electrostatic interaction, for which a pharmacophore model is difficult to account for. An efficient approach is the synergistic combination of pharmacophore-based VS and docking-based VS. Because inherent limitations of each of these screening techniques are not easily resolved, their combination in a hybrid protocol can help to mutually compensate for these limitations and capitalize on their mutual strengths. Various combined virtual screening strategies and their validity have been well reviewed by Talevi *et al.* [47], Kirchmair *et al.* [48] and Muegge [49]. This approach has also been routinely used in our group, with which we have successfully obtained several real hits validated experimentally for inhibition against protein kinases Aurora-A [50], Syk [51] and ALK5 [52].

Pharmacophore-based *de novo* design

Besides the pharmacophore-based VS described above, another application of pharmacophore is *de novo* design of ligands. The compounds obtained from pharmacophore-based VS are usually existing chemicals, which might be patent protected. In contrast to pharmacophore-based VS, the *de novo* design approach can be used to create completely novel candidate structures that conform to the requirements of a given pharmacophore. The first pharmacophore-based *de novo* design program is NEWLEAD [53], which uses as input a set of disconnected molecular fragments that are consistent with a pharmacophore model, and the selected sets of

disconnected pharmacophore fragments are subsequently connected by using linkers (such as atoms, chains or ring moieties). Actually, NEWLEAD can only handle the cases in which the pharmacophore features are concrete functional groups (not abstract chemical features). Other shortcomings of the NEWLEAD program include that the sterically forbidden region of the binding site is not considered and that, as in traditional *de novo* design programs, the compounds created by the NEWLEAD program might be difficult to chemically synthesize. Other programs such as LUDI⁴ and BUILDER [54] can also be used to combine identification of structure-based pharmacophore with *de novo* design. They, however, need the knowledge of 3D structures of the macromolecular targets.

To overcome drawbacks of the current pharmacophore-based *de novo* design software, we have developed a new program, PhDD (a pharmacophore-based *de novo* design method of drug-like molecules) [55]. PhDD can automatically generate drug-like molecules that satisfy the requirements of an input pharmacophore hypothesis. The pharmacophore used in PhDD can be composed of a set of abstract chemical features and excluded volumes that are the sterically forbidden region of the binding site. PhDD first generates a set of new molecules that completely conform to the requirements of the given pharmacophore model. Then a series of assessments to the generated molecules are carried out, including assessments of drug-likeness, bioactivity and synthetic accessibility. PhDD was tested on three typical examples: pharmacophore hypotheses of histone deacetylase, CDK2 and HIV-1 integrase inhibitors. The test results showed that PhDD was able to generate molecules with completely novel scaffolds. A similarity analysis with the use of Tanimoto coefficients demonstrated that the generated molecules should have similar biological functions to the existing inhibitors, although they are structurally different.

Concluding remarks

Pharmacophore approaches have evolved to be one of the most successful concepts in medicinal chemistry through the collective efforts of many researchers in the past century. In particular, considerable progress of pharmacophore technology in the past two decades has made pharmacophore approaches one of the main tools in drug discovery. Despite the advances in key techniques of pharmacophore modeling, there is still room for further improvement to derive more accurate and optimal pharmacophore models, which include better handling of ligand flexibility, more efficient molecular alignment algorithms and more accurate model optimization. Lower efficiency (computational time cost) and poor effect (lower hit rate) of pharmacophore-based VS seriously obstructs the applications of pharmacophore in drug discovery. The former, however, will be further reduced and diminished by the increasing capacity and reducing cost of computer hardware. 'Synergistic' combination of pharmacophore method and other molecular modeling approaches such as docking is a good strategy to further improve the effect. Compared with pharmacophore-based VS, pharmacophore-based *de novo* design shows a unique advantage in building completely novel hit compounds. In addition to virtual screening and *de novo* design, the applications of pharmacophore have also been extended to lead optimization [56], multitarget drug design [57], activity profiling [58] and target identification [59]. The increasing application ranges of pharmacophore, together with success stories in drug discovery, enable further enrichment of the pharmacophore concept and promote the development and application of pharmacophore approaches.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.drudis.2010.03.013.

References

- Ehrlich, P. (1909) Ueber den jetzigen Stand der Chemotherapie. *Ber. Dtsch. Chem. Ges.* 42, 17–47
- Wermuth, C.G. *et al.* (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). *Annu. Rep. Med. Chem.* 33, 385–395
- Günd, P. (2000) Evolution of the pharmacophore concept in pharmaceutical research. In *Pharmacophore Perception, Development, and Use in Drug Design* (Güner, O.F., ed.), pp. 171–189, International University Line (IUL)
- Wermuth, C.G. (2006) Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In *Pharmacophores and Pharmacophore Searches* (Langer, T. and Hoffmann, R.D., eds), pp. 3–13, Wiley-VCH
- Dror, O. *et al.* (2006) Predicting molecular interactions *in silico*. I. An updated guide to pharmacophore identification and its applications to drug design. *Front. Med. Chem.* 3, 551–584
- Kubinyi, H. (2006) Success stories of computer-aided design. In *Computer Applications in Pharmaceutical Research and Development* (Ekins, S., ed.), pp. 377–424, Wiley-Interscience
- Mustata, G. *et al.* (2009) Discovery of novel Myc-Max heterodimer disruptors with a three-dimensional pharmacophore model. *J. Med. Chem.* 52, 1247–1250
- Barnum, D. *et al.* (1996) Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* 36, 563–571
- Li, H. *et al.* (2000) HypoGen: an automated system for generating 3D predictive pharmacophore models. In *Pharmacophore Perception, Development, and Use in Drug Design* (Güner, O.F., ed.), pp. 171–189, International University Line
- Martin, Y.C. (2000) DISCO: what we did right and what we missed. In *Pharmacophore Perception, Development, and Use in Drug Design* (Güner, O.F., ed.), pp. 49–68, International University Line
- Jones, G. and Willet, P. (2000) GASP: genetic algorithm superimposition program. In *Pharmacophore Perception, Development, and Use in Drug Design* (Güner, O.F., ed.), pp. 85–106, International University Line
- Dixon, S.L. *et al.* (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening. 1. Methodology and preliminary results. *J. Comput. Aid. Mol. Des.* 20, 647–671
- Poptodorov, K. *et al.* (2006) Pharmacophore model generation software tools. In *Pharmacophores and Pharmacophore Searches* (Langer, T. and Hoffmann, R.D., eds), pp. 17–47, Wiley-VCH
- Van Drie, J.H. (2004) Pharmacophore discovery: a critical review. In *Computational Medicinal Chemistry for Drug Discovery* (Bultinck, P., ed.), pp. 437–460, Marcel Dekker
- Wolber, G. *et al.* (2008) Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* 13, 23–29
- Smellie, A. *et al.* (1994) Poling: promoting conformational variation. *J. Comput. Chem.* 16, 171–187
- Gippert, G.P. *et al.* (1998) Distributed torsion angle grid search in high dimensions: a systematic approach to NMR structure determination. *J. Biomol. NMR* 11, 241–263
- Hurst, T. (1994) Flexible 3D searching: the directed tweak technique. *J. Chem. Inf. Comput. Sci.* 34, 190–196
- Jones, G. *et al.* (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aid. Mol. Des.* 9, 532–549
- Li, Z. and Scheraga, H.A. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 84, 6611–6615

⁴LUDI is now incorporated into Discovery Studio, available from Accelrys Inc., San Diego, CA, USA.

- 21 Sperandio, O. *et al.* (2009) MED-3DMC: a new tool to generate 3D conformation ensembles of small molecules with a Monte Carlo sampling of the conformational space. *Eur. J. Med. Chem.* 44, 1405–1409
- 22 Liu, X. *et al.* (2009) Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC Bioinformatics* 10, 1–14
- 23 Li, J. *et al.* (2007) CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J. Chem. Inf. Model.* 47, 1923–1932
- 24 Bandyopadhyay, D. and Agrafiotis, D.K. (2008) A self-organizing algorithm for molecular alignment and pharmacophore development. *J. Comput. Chem.* 29, 965–982
- 25 Totrov, M. (2008) Atomic property fields: generalized 3D pharmacophoric potential for automated ligand superposition, pharmacophore elucidation and 3D QSAR. *Chem. Biol. Drug Des.* 71, 15–27
- 26 Nettles, J.H. *et al.* (2007) Flexible 3D pharmacophores as descriptors of dynamic biological space. *J. Mol. Graph. Model.* 26, 622–633
- 27 Baroni, M. *et al.* (2007) A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* 47, 279–294
- 28 Hecker, E.A. *et al.* (2002) Use of Catalyst pharmacophore models for screening of large combinatorial libraries. *J. Chem. Inf. Comput. Sci.* 42, 1204–1211
- 29 Toba, S. *et al.* (2006) Using pharmacophore models to gain insight into structural binding and virtual screening: an application study with CDK2 and human DHFR. *J. Chem. Inf. Model.* 46, 728–735
- 30 Vadivelan, S. *et al.* (2007) Virtual screening studies to design potent CDK2-Cyclin A inhibitors. *J. Chem. Inf. Model.* 47, 1526–1535
- 31 Wolber, G. *et al.* (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169
- 32 Chen, J. and Lai, L.H. (2006) Pocket v.2: further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.* 46, 2684–2691
- 33 Ortuso, F. *et al.* (2006) GBPM: GRID based pharmacophore model. Concept and application studies to protein–protein recognition. *Bioinformatics* 22, 1449–1455
- 34 Böhm, H.J. (1992) The computer program LUDI: a new method for the *de novo* design of enzyme inhibitors. *J. Comput. Aid. Mol. Des.* 6, 61–78
- 35 Barillari, C. *et al.* (2008) Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J. Chem. Inf. Model.* 48, 1396–1410
- 36 Tintori, C. (2008) Targets looking for drugs: a multistep computational protocol for the development of structure-based pharmacophores and their applications for hit discovery. *J. Chem. Inf. Model.* 48, 2166–2179
- 37 Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857
- 38 Zou, J. *et al.* (2008) Towards more accurate pharmacophore modeling: multicomplex-based comprehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2. *J. Mol. Graph. Model.* 27, 430–438
- 39 Zhao, H. (2007) Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective. *Drug Discov. Today* 12, 149–155
- 40 Ullmann, J.R. (1976) An algorithm for subgraph isomorphism. *J. ACM* 23, 31–42
- 41 Barnard, J.M. (1993) Substructure searching methods: old and new. *J. Chem. Inf. Comput. Sci.* 33, 532–538
- 42 Xu, J. (1996) GMA: a generic match algorithm for structural homorphism, isomorphism, maximal common substructure match and its applications. *J. Chem. Inf. Comput. Sci.* 36, 25–34
- 43 Triballeau, N. *et al.* (2006) Are you sure you have a good model? Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In *Pharmacophores and Pharmacophore Searches* (Langer, T. and Hoffmann, R.D., eds), pp. 325–364, Wiley–VCH
- 44 Sun, H. (2008) Pharmacophore-based virtual screening. *Curr. Med. Chem.* 15, 1018–1024
- 45 Deng, J. *et al.* (2006) Dynamic pharmacophore model optimization: identification of novel HIV-1 integrase inhibitors. *J. Med. Chem.* 49, 1684–1692
- 46 Giménez-Oya, V. *et al.* (2009) Mimicking direct protein–protein and solvent-mediated interactions in the CDP-methylethritol kinase homodimer: a pharmacophore-directed virtual screening approach. *J. Mol. Model.* 15, 997–1007
- 47 Talevi, A. *et al.* (2009) Combined virtual screening strategies. *Curr. Comput. Aid. Drug. Des.* 5, 23–37
- 48 Kirchmair, J. *et al.* (2008) Enhancing drug discovery through *in silico* screening: strategies to increase true positives retrieval rates. *Curr. Med. Chem.* 15, 2040–2053
- 49 Muegge, I. (2008) Synergies of virtual screening approaches. *Mini Rev. Med. Chem.* 8, 927–933
- 50 Deng, X.-Q. *et al.* (2008) Pharmacophore modeling and docking study for identification of new Aurora-A kinase inhibitors. *Chem. Biol. Drug Des.* 71, 533–539
- 51 Xie, H.-Z. *et al.* (2009) Pharmacophore modeling study based on known spleen tyrosine kinase inhibitors together with virtual screening for identifying novel inhibitors. *Bioorg. Med. Chem. Lett.* 19, 1944–1949
- 52 Ren, J.-X. *et al.* (2009) Pharmacophore modeling and virtual screening for the discovery of new transforming growth factor- β type I receptor (ALK5) inhibitors. *Eur. J. Med. Chem.* 44, 4259–4265
- 53 Tschinke, V. and Cohen, N. (1993) The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.* 36, 3863–3870
- 54 Roe, D.C. and Kuntz, I. (1995) BUILDER v.2: improving the chemistry of a *de novo* design strategy. *J. Comput. Aid. Mol. Des.* 9, 269–282
- 55 Huang, Q. *et al.* (2010) PhDD: a new pharmacophore-based *de novo* design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model.* 10.1016/j.jmgs.2010.02.002
- 56 Brenk, R. and Klebe, G. (2006) ‘Hot spot’ analysis of protein-binding sites as a prerequisite for structure-based virtual screening and lead optimization. In *Pharmacophores and Pharmacophore Searches* (Langer, T. and Hoffmann, R.D., eds), pp. 171–192, Wiley–VCH
- 57 Wei, D. *et al.* (2008) Discovery of multi-target inhibitors by combining molecular docking with common pharmacophore matching. *J. Med. Chem.* 51, 7882–7888
- 58 Steindl, T.M. *et al.* (2006) Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* 46, 2146–2157
- 59 Rollinger, J.M. *et al.* (2004) Acetylcholinesterase inhibitory activity of scopolin and scopoletin discovered by virtual screening of natural products. *J. Med. Chem.* 47, 6248–6254