ELSEVIER

Molecule-pharmacophore superpositioning and pattern matching in computational drug design

Gerhard Wolber[1], Thomas Seidel[1], Fabian Bendix[1] and Thierry Langer[2]

[1] Inte:Ligand GmbH, Mariahilferstrasse 74B/11, 1070 Vienna, Austria
[2] Computer Aided Molecular Design Group, Institute of Pharmacy, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria

Three-dimensional (3D) pharmacophore modeling is a technique for describing the interaction of a small molecule ligand with a macromolecular target. Since chemical features in a pharmacophore model are well known and highly transparent for medicinal chemists, these models are intuitively understandable and have been increasingly successful in computational drug discovery in the past few years. The performance and applicability of pharmacophore modeling depends on two main factors: the definition and placement of pharmacophoric features and the alignment techniques used to overlay 3D pharmacophore models and small molecules. An overview of key technologies and latest developments in the area of 3D pharmacophores is given and provides insight into different approaches as implemented by the 3D pharmacophore modeling packages like Catalyst, MOE, Phase and LigandScout.

Introduction

The concept of pharmacophore modeling is simple and, in many respects, produces results that would be intuitive to an experienced medicinal chemist: the technique rigidly models the interaction between a ligand and its binding site in a specific binding situation. The result is a three-dimensional (3D) spatial arrangement of chemical features, which are derived using algorithms that take rules derived from chemical knowledge into account. Technologies for deriving 3D pharmacophores have been described in two books covering pharmacophore perception and modeling [1,2], and its importance as a drug discovery method has recently been described by Kubinyi [3]. These feature maps, called 3D pharmacophores, can be used to search for similarities between binding situations or even for similarities between molecules. This leads to the advantages and disadvantages of pharmacophore modeling: on the one hand, the rule-based creation of chemical features is the ideal basis for the definition of an interface between medicinal chemistry and computer science, providing the means to add intentional and necessary bias from the medicinal or computational chemist to the still imperfect representation of molecules in computers. On the other hand, heuristic modeling is not a systematic approach: important interactions may not be well-represented in a specific chemical feature model, increasing the likelihood of important information loss in the resulting 3D pharmacophore, and, as a result, estimating binding energy contributions of particular chemical features is practically impossible. Therefore, feature definitions are a crucial part of pharmacophore modeling techniques and will be an important part of this review article.

A pharmacophore can be derived either in a structure-based manner by determining complementarities between a ligand and its binding site, or in a ligand-based manner, by flexibly overlaying a set of active molecules and determining those conformations that are able to be overlaid in such a way that a maximum number of important chemical features geometrically overlap. The ligand-based approach inherently involves the flexible alignment of molecules, which, on the one hand, can be done only by taking into account atom contributions or by other methods that are unrelated to 3D pharmacophore representations. On the other hand, geometric information about all possible chemical features of a molecule can be used as input data for the flexible alignment. The alignment turns out to be the most computationally expensive and algorithmically most challenging part. Some pharmacophore alignment approaches incorporate algorithms where the computing time needed grows exponentially with the number of involved chemical features [4]. This limits their scalability and

Corresponding author: Wolber, G. (wolber@inteligand.com)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2007.09.007

applicability—also for small molecules as soon as the describing chemical feature set involves a larger number of chemical features. Practically, on current hardware, these approaches are limited to small molecules and simple chemical feature descriptions. Other approaches use algorithms that do not cover the whole search space but deliver one single optimal solution, and thus can be solved in polynomial time. This class of approaches is more flexible, since it can use more feature definitions and even place multiple features on the same atom group, and also more scalable, which allows their application to larger, more feature-rich molecules like peptides [5–7].

The challenge of molecular superpositioning (3D alignment) incorporates the problem of conformational flexibility that can be addressed by pre-enumerating a general-purpose conformational model or by changing molecule coordinates as needed by the alignment algorithm. Both show advantages and disadvantages: by pre-enumerating conformations, less computational time is spent during the alignment process at the detriment of not being able to get a 'tailored' conformational model for the specific problem. Current conformational model generators, however, seem to perform sufficiently well with respect to this problem [8,9]. If a conformational model is pre-generated, pattern-matching techniques can be applied to geometric pharmacophore patterns, which bear significant performance and applicability advantages in the alignment step mentioned above. In general, this also applies to different 3D substructure searching and virtual screening techniques, and therefore, pre-generating conformational models is commonly used to address the need for faster search times. This review will concentrate on methods that produce a 3D pharmacophore model that describes ligand-side chemical features, and therefore, refers to the ligand rather than to the macromolecule.

The term '3D pharmacophore'

The term 'pharmacophore' has become increasingly used in medicinal chemistry in recent years and has had different meanings attributed to it. 'Pharmacophores' are often regarded as structural fragments or functional groups being related to a chemical compound. However, the official IUPAC definition [10] from 1998 is more precise: 'A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response' [11]. This definition clearly emphasizes the abstraction of common steric and electronic interactions of bio-active compounds exhibiting comparable biological effects within the same binding site in a comparable situation. This abstract model, containing chemical functionalities (such as 'positive ionizable' instead of 'primary amine') can serve as an effective search filter for virtual screening. This concept is not new in medicinal chemistry and has already been successfully applied before computers were used in chemistry [12,13].

To be a useful tool for drug design, a pharmacophore model [14] has to provide valid information for medicinal chemists investigating structure–activity relationships. The pharmacophore model has to describe the nature of the functional groups involved in ligand–target interactions, as well as the type of the non-covalent bonding and intercharge distances. The model also has to show predictive power that, at its best, enables the design of novel chemical structures that are not evidently derived by the translation of structural features from one active series to the other, or even allows effective scaffold hopping [15].

Feature definitions and pharmacophore representation

Selecting the right chemical feature types is a first crucial step for the development of a high quality pharmacophore model. In early pharmacophore modeling techniques, such as the active analog approach described by Marshall *et al*. [16], features constituting a pharmacophore could contain any fragment or atom type. More recent techniques, as applied in the software package Catalyst [17] use a more general way for building pharmacophore queries, for example, a single geometric entity for all negative ionizable groups. The discussion below will show that several arguments exist for continuing with this trend and even to further extend the generalization of chemical functionalities. In real-life applications, however, built-in features are often tailored to achieve a desired filtering restrictivity level [18], that is, the ability to restrict a model to identify a specified set of compounds that the model was created for.

General definitions may result in models that are universal, at the cost of restrictivity. However, restrictivity is an important issue in pharmacophore searching and, therefore, feature descriptions that are too general need to be changed from reflecting universal chemical functionality to representing distinct functional groups. A common approach is to derive a model from distinct ligands in order to represent the specific mode of interaction as a chain of functional groups or exclusions thereof [18]. By restricting general chemical feature definitions in the way described above, the number of standard, well-known features increases at the cost of comparability. However, only comparable pharmacophores are sufficiently universal and can represent a mode of action and bioisosteric fragments of the ligand, instead of being restrictive to a set of already existing ligands. In order to describe the levels of universality and specificity of chemical features, a simple layer model is used in the following discussion to allow referral to these properties more easily. Table 1 shows a proposed classification of abstraction layers of the most important chemical features [19]. A lower level corresponds to higher specificity and, therefore, lower universality.

The most frequent reason for creating features on the low universality levels 1 and 2 is that the definitions of the higher levels are not sufficient to describe the features occurring in the training set (see reference [18] for an example). Even if customization results in a layer 1 or layer 2 feature, there should be a possibility of including layer 3 or 4 information in order to categorize and, thus, increase comparability (for example, a carboxylic acid as a layer 2 feature is a subcategory of 'negative ionizable', which is a layer 4 feature).

Current software packages for pharmacophore modeling like Catalyst [20], Phase [21], MOE [22], and LigandScout [7] always have to face a trade-off in the design of a generally applicable feature set that is universal and, at the same time, still selective enough to reflect all relevant types of ligand–receptor interactions. The most relevant interactions and their geometric representation in the applications mentioned before are described in the following section with Table 2 providing a summary.

TABLE 1

Abstraction layers of pharmacophoric feature constraints

Layer	Classification	Universality	Specificity
4	Chemical functionality without geometric constraint, for example, an H-Bond acceptor without a projected point or a lipophilic group	+++	–
3	Chemical functionality (H-bond acceptor, H-bond donor, positive ionizable, negative ionizable, hydrophobic) with geometric constraint, for example, an H-bond acceptor vector including an acceptor point as well as a projected donor point; aromatic ring including a ring plane	++	+
2	Molecular graph descriptor (atom, bond) without geometric constraint, for example, a geometrically unconstrained phenol group	–	++
1	Molecular graph descriptor (atom, bond) with geometric constraint, for example, a phenol group facing a parallel benzenoid system within a distance of 2–4 Å	– –	+++

Hydrogen bonding interactions

Hydrogen bonding occurs when covalently bound hydrogen atoms with a positive partial charge interact with other atoms with a negative partial charge. To capture the characteristics of hydrogen bonding, Catalyst and LigandScout model H-bond donor and acceptor features as a position for the heavy atom and a projected point representing the position from which the participating hydrogen will extend. These two positions form a vector that indicates the direction from the heavy atom to the projected point of the hydrogen bond. However, in Catalyst only a single hydrogen bonding feature is permitted per heavy atom, whereas LigandScout allows an acceptor or donor atom to be involved in more than one H-bonding interaction. In the situation shown in Figure 1 (PDB code 2GDE) either all hydrogen bonds or two hydrogen bonds and the ionizable feature must be omitted in Catalyst, or several different models must be created to reflect all interactions.

In a similar manner, Phase positions H-bond acceptor features on heavy atoms that carry one or more lone pairs, and, depending on the hybridization of the acceptor atoms, assigns vector attributes to each idealized hydrogen bond axis. Hydrogen bond donor features are centered on donor hydrogen atoms with a single vector constraint directed along the hydrogen bond axis. An alternative to this ligand-centric convention is to represent acceptors and donors as pure projected points located at complementary positions on theoretical binding sites. The projected point approach does not incorporate vector character into the site definition and permits situations where two ligands form

TABLE 2

Summary of pharmacophoric feature abstraction, representation and customization in different pharmacophore modeling applications

Feature	Phase	MOE	Catalyst	LigandScout
Hydrogen bonding interaction	H-donors located at hydrogen atom, acceptors at heavy atom. Modeled as layer 3 feature with direction constraint and position tolerance.	Geometric constraints depend on selected pharmacophore scheme. Features located at heavy atoms with tolerance sphere.	Acceptor and donor features positioned on heavy atom with tolerance sphere (layer 3 feature). Max. one donor or acceptor feature per atom.	Acceptor and donor features positioned on heavy atom with tolerance sphere. Represented as layer 3 features with feature position and projected point.
Lipophilic area	Represented as tolerance spheres. Aromatic rings not recognized as hydrophobic areas.	Represented as tolerance spheres. Aromatic rings not recognized as hydrophobic areas.	Represented as tolerance spheres.	Represented as tolerance spheres.
Aromatic interaction	Represented as position with tolerance and ring plane orientation.	Modeling of aromatic features depends on the selected pharmacophore scheme.	Represented as position with tolerance and ring plane orientation.	Represented as position with tolerance and ring plane orientation.
Charge–transfer interaction	Represented as tolerance spheres, no explicit charges necessary.	Represented as tolerance spheres, require charges (cationic and anionic).	Represented as tolerance spheres, no explicit charges necessary.	Represented as tolerance spheres, no explicit charges necessary.
Definition and customization	Definition and customization through SMARTS patterns with associated geometry attributes.	Implementation of new pharmacophore schemes possible with scripting language (SVL).	Definition of new features and customization via graphical interface.	Feature definitions as SMARTS or boxed algorithms, adjustment of feature specific parameters and geometric constraints.

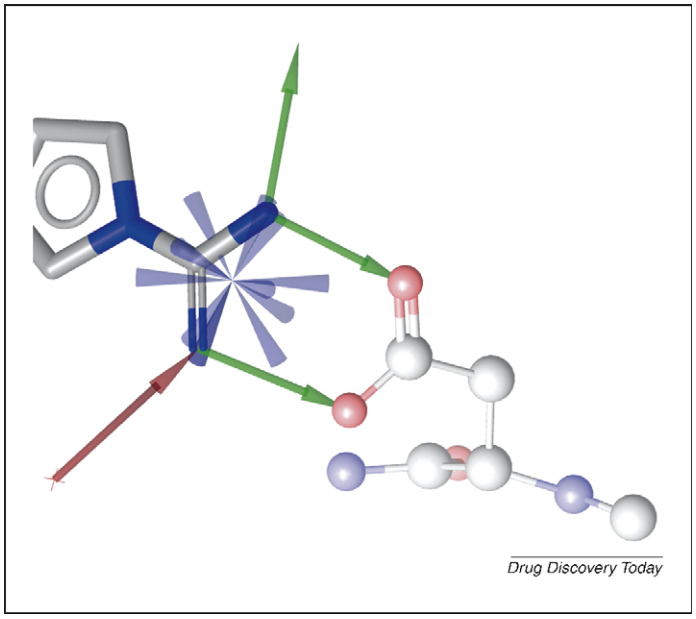


FIGURE 1

Thrombin inhibitor SN3401 in complex with thrombin (PDB entry 2GDE): three hydrogen bond donors (green), one acceptor (red), and one charge transfer feature (blue) are recognized by LigandScout. In order to show more detail of the interaction, only a portion of the structure of SN3401 has been shown.

hydrogen bonds to the same receptor atom, but from different locations and directions.

In MOE the modeling of H-donor and acceptor features depends on the selected pharmacophore scheme. There are six schemes supplied with MOE: PCH, PCHD, PCH_All, PPCH, PPCH_All, and CHD. The PCH scheme (which is the default) defines H-bond acceptors and donors as layer 4 point features. By contrast, the PCHD scheme also includes putative points from hydrogen bond donors and acceptors that are projected in the approximate direction of the hydrogen bond.

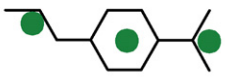
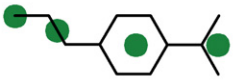
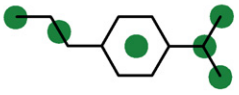
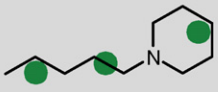
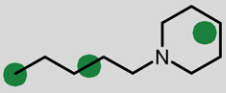
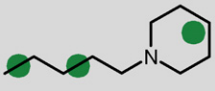
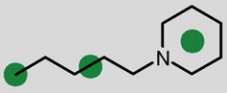
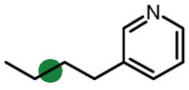
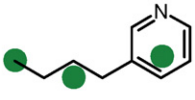
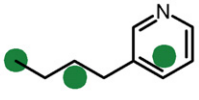
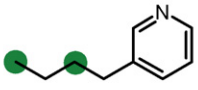
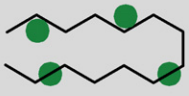
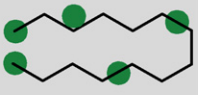
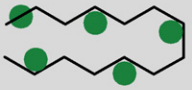
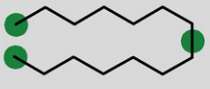
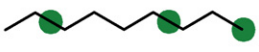
Lipophilic areas

Lipophilic contacts represent layer 4 features with no geometric constraints and are generally represented as tolerance spheres located in the center of hydrophobic atom chains, branches or groups. Although the perception of hydrophobic areas in Catalyst, Phase and LigandScout is based on the same algorithm described during Catalyst development by Green *et al.* [17], subtle deviations seem to exist, and the results differ considerably, which makes an otherwise possible program interoperability hard to accomplish.

Table 3 provides a schematic illustration of the differences in number and location of hydrophobic features generated by Phase, Catalyst, LigandScout and MOE for six representative molecules. In contrast to LigandScout and Catalyst, which place a hydrophobic feature shifted away from the ring center toward lipophilic atoms, Phase and MOE do not recognize a heteroaromatic ring as a hydrophobic area. For hydrocarbon chains, the number of hydrophobic features recognized by Catalyst, LigandScout and Phase depends on the length of the chain. MOE generally generates three annotation points per chain with one point in the middle and one on each end. This rough representation of lipophilic information in MOE makes it difficult to correctly describe lipophilic features for virtual screening – the dynamic placement as described by Green *et al.* [17] allows for a smoother overlay of lipophilic areas, such as aromatic rings, isopropyl moieties, and aliphatic chains. In MOE, currently there is no such possibility using the built-in

TABLE 3

Schematic illustration of hydrophobic feature placement (green dots) in different pharmacophore modeling applications

	Phase	Catalyst	LigandScout	MOE
1				
2				
3				
4				
5				
6				

feature definitions, but there is a possibility to define one's own algorithms, using the scripting language, SVL, provided with the software package. Scripting chemical feature placement, however, negatively influences virtual screening performance.

Aromatic interactions

Aromatic features can be modeled as layer 4 point features or as layer 3 features. In Catalyst, Phase and LigandScout these features are also attributed with a ring plane normal defining the spatial orientation (layer 3 feature). In MOE, the selected pharmacophore scheme determines whether a ring plane orientation constraint is included or if aromaticity is modeled as a pure point feature with a tolerance sphere.

Charge–transfer interactions

Positive or negative ionizable areas are single atoms or groups of atoms that are likely to be protonated or deprotonated at physiological pH. Ionizable features are commonly implemented as spheres with a certain tolerance radius for pharmacophore matching. While Catalyst, LigandScout and Phase are insensitive to the protonation state from the input molecules, MOE requires preprocessing of the molecules and the assignment of explicit charges. Additionally, in MOE, positive and negative ionizable areas are limited to single atoms (including resonance anions and cations) carrying a corresponding charge, causing limits for groups like carboxylic acids or guanidines.

Customization and definition of new features

Although feature definitions should be general and describe all possible interactions that are observed in ligand binding, some models still do not fulfill all requirements for restrictivity. For this reason, most applications provide means for customization or extension of the predefined feature set. Catalyst, for example, allows customization and extension of the built-in features via the graphical user interface allowing to specify one or more chemical groups that satisfy a particular feature (OR logic). Similarly, Phase allows specification of matching chemical groups and exclusions for a particular feature as a list of SMARTS [23] patterns. Each pattern can be associated with a geometric representation (point, group or vector) and additional flags for hydrogen bond acceptors and donors. Although the Phase approach is very flexible and user friendly, only a maximum of three additional custom feature definitions can be added. In MOE, the user has the possibility of implementing a new, or modifying an existing, pharmacophore scheme, using the scripting language SVL.

LigandScout defines all chemical features as SMARTS patterns in a single configuration file and additionally provides a means for defining geometric constraints, for example for hydrogen bonding or stacking of aromatic ring systems. No graphical user interface is provided for modifying feature definitions, but angle constraints and distance ranges for feature recognition in a macromolecule ligand complex can be fine-tuned in a graphical user interface.

The placement of pharmacophore points onto a 3D structure of a molecule heavily influences the performance of the following alignment algorithms used for detection of common pharmacophoric patterns among molecules (ligand-based design) or virtual screening. Software packages that do not place a hydrophobic feature on an aromatic ring, for example, will never be able to detect that an aromatic group or an aliphatic group may be positioned in a lipophilic area of a binding pocket. An algorithm that does not allow placing a charge point in the middle between the two oxygens of a carboxylic acid and allows a charge on only one of the two oxygens, will suffer from an insufficient cheminformatics representation.

Once chemical feature points are detected, they can be used as input for the computationally more challenging part: aligning a molecule to a pharmacophore. The algorithms described below can all be used to superposition two molecules according to their pharmacophoric annotation points, or to overlay a molecule to a pharmacophore, or to overlay two 3D pharmacophores. These algorithms form the basis for the elucidation of common pharmacophore patterns among different ligands with similar known biological activity, or for virtual screening. Both are tasks that are time consuming, even on modern hardware and, therefore, efficiency remains important.

Current superpositioning techniques for aligning 3D pharmacophores and molecules

In the broad field of possible pharmacophoric alignment techniques, one can distinguish between either point-based or property-based approaches [2]. With point-based approaches, atoms or chemical feature point distances are minimized, while property-based approaches use molecular field descriptors to generate alignments. Considering recent trends and examining currently available commercial software packages, the majority of the programs use point-based alignment algorithms, superposing pairs of points by minimizing distances. As already described, an important issue is the positioning of pharmacophoric anchor points, since this is the only chemical representation of the molecule for the algorithms. In order to discover the relevant chemical feature points, Dror *et al*. differentiate between points being either atoms, fragments, or chemical features [24]. In comparison with property-based techniques, this abstraction is one of the greatest limitations of all point-based methods, because aligning dissimilar ligands can become problematic in either case. Nevertheless, the feature-based approach – in the sense of pharmacophoric features – has become widely accepted and is used in nearly all drug discovery toolkits nowadays. If advanced feature point algorithms are used, the points represent bioisosterically comparable molecule parts, and the geometric sensitivity becomes an advantage of these algorithms.

Three-dimensional alignment incorporates the problem of conformational flexibility. One possibility to address the problem is to pre-generate conformations (like in DISCO [25] or Catalyst [17]), which makes the actual alignment less time consuming, but the user has to ensure that all relevant conformations are included. On the contrary, there are in-process approaches that perform the pattern identification and conformational search simultaneously.

A well-known pharmacophore elucidation program using such an approach is GASP [26,27] that was developed by Jones and coworkers in the mid-1990s and is marketed by Tripos. The program is based on a genetic algorithm, that is, a non-deterministic method that simulates evolution by randomly mutating chromosomes of a certain population. In terms of pharmacophore pattern matching, each chromosome represents a potential flexible pharmacophore by encoding all bond angles and by listing all feature

mappings to a manually selected reference compound. In each run, chromosomes that score best are selected, according to some crucial fitness function, and those are then mutated by applying random torsional rotations to cover conformational space on the fly. GALAHAD [28–30], developed at the University of Sheffield, Novo Nordisk and Biovitrum, and also marketed by the company Tripos, uses a modified genetic algorithm reducing bias toward a single template (base) molecule, introduces partial matching and an improved multi-objective scoring function. Searching is faster than with GASP, since GALAHAD allows the use of pre-generated conformations. For pharmacophore elucidation GALAHAD uses a very efficient atom-based alignment technique [RICHMOND] but lacks important feature definitions like the flexible placement of hydrophobic features, which is possible with the other programs.

In contrast to the in-process techniques providing fully flexible models, the other very different class of algorithms relies on rigid-body techniques for aligning molecular structures [31]. Either these methods are completely structure-based approaches, or consider conformational flexibility in terms of handling pre-generated conformations sequentially. The advantage of these techniques is that the time consuming process of generating conformations is out-sourced, and the conformations can be stored persistently because of the fact that multi-conformational generators actually have the ability to provide generally applicable conformational ensembles that sufficiently sample most small organic drug-like molecules [9].

Nearly all the commercial software using rigid-body alignment techniques is based on maximum common substructure search. One of the first programs that go along with this and that had a considerable influence on modern techniques [2] is DISCO [25]. It is based on distance geometry [32], and the alignment is implemented using the Bron–Kerbosh clique-detection algorithm [33] in terms of inter-distance comparisons. Because of exploring the complete conformational space, the technique is limited to a small number of input compounds of preferably limited flexibility—the main drawback when looking for an optimal solution.

A further development of the exhaustive search is Catalyst's HipHop [34] algorithm, just relaxing the GASP requirement that each feature in the pharmacophore must be present in each of the input compounds. Furthermore, HipHop starts by finding all two-feature models and expands the model until no more configurations can be found. All of the numerous results are listed and ranked according to their rarity-based score.

Another example for an exhaustive search is employed by Schrödinger's program Phase [35]. The algorithm enhances performance by narrowing the search space. The conformational search is dedicated to find a pharmacophore containing a user-defined number of features that are shared by a user-defined number of input molecules regarding a user-defined tolerance. All possible pharmacophores are grouped in a tree according to their inter-site distance, that is, a vector containing distances of all feature pairs. The tree is traversed and if a node fails to contain pharmacophores from the minimum number of actives, the complete subtree is eliminated from further investigation. Phase places high emphasis on user interaction, providing user intervention possibilities at each step of the pharmacophore elucidation process.

The most recent development regarding pharmacophore alignment technique is LigandScout's pattern matching approach [6]. In a first step, feature pairs are formed on the basis of feature types and distance characteristics, encoding the whole pharmacophore for each feature in a rotationally and translationally independent manner. For each feature type and feature, a distance shell contains a number of bins counting neighboring features, with each bin representing space at a certain distance interval. The core part of the algorithm for identifying pairs is a fast maximum weighted bipartite matching algorithm that scales polynomially with the number of features involved and thus allows its application to larger molecules like peptides. Through its performance, this algorithm has the possibility to perform rigid alignments within less than 100 ms on a modern single CPU allowing for interactive usage within the graphical user interface of LigandScout.

Conclusions

The pharmacophore concept is a successful and well-known approach for drug design (both ligand and structure based) as well as for virtual screening. Several methods for describing pharmacophores have been established showing significant differences and capabilities in their way of how to describe chemical features as building blocks for pharmacophores. It is important that the chemical feature representation used reflects the interactions that are important for the target being represented, and some chemical feature representations are more universal than others. The computational part of pharmacophore modeling, different alignment techniques used for pharmacophore elucidation and virtual screening, has significantly improved with the availability of new software packages. The use of new algorithms has led to performance optimizations over the past few years, leading to modern pattern recognition approaches that are capable of performing the superpositioning of pharmacophores and molecules in a fraction of the time that was needed by earlier approaches.

References

1 Güner, O.F. (2000) *Pharmacophore Perception Development and Use in Drug Design,* International University Line

2 Langer, T. and Hoffmann, R.D. (2006) *Pharmacophores and Pharmacophore Searches,* Wiley-VCH

3 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, New York

4 Patel, Y. *et al.* (2002) A comparison of the pharmacophore identification programs: catalyst, DISCO and GASP. *J. Comput. Aid. Mol. Des.* 16, 653–681

5 Wolber, G. and Langer, T. (2005) LigandScout: 3D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169

6 Wolber, G. and Dornhofer, A.A. (2006) Efficient overlay of small organic molecules using 3D pharmacophores. *J. Comput. Aid. Mol. Des.* 20, 773–788

7 Wolber, G. and Kosara, R. (2006) Pharmacophores from macromolecular complexes with LigandScout. In *Pharmacophores and Pharmacophore Searches*, (vol. 32) (Langer, T. and Hoffmann, R.D., eds) pp. 131–150, Wiley-VCH

8 Kirchmair, J. *et al.* (2005) Comparative analysis of protein-bound ligand conformations with respect to catalyst's conformational space subsampling algorithms. *J. Chem. Inf. Model.* 45, 422–430

9 Kirchmair, J. *et al.* (2006) Comparative performance assessment of the conformational model generators Omega and Catalyst: a large scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* 46, 422–430

10 Wermuth, C.G. *et al.* (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). *Annu. Rep. Med. Chem.* 33, 385–395
11 Wermuth, C.G. (2006) Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In *Pharmacophores and Pharmacophore Searches*, (vol. 32) (Langer, T. and Hoffmann, R.D., eds) pp. 3–13, Wiley-VCH
12 Dodds, E.C. and Lawson, W. (1938) Molecular structure in relation to oestrogenic activity. Compounds without phenanthrene nucleus. *Proc. R. Soc. Lond. Ser. B* 125, 122–132
13 Schueler, F.W. (1946) Sex hormone action and chemical constitution. *Science* 103, 221–223
14 Wermuth, C.G. and Langer, T. (1993) Pharmacophore identification. In *3D QSAR in Drug Design. Theory, Methods and Applications* (Kubinyi, H., ed.), pp. 117–136, ESCOM
15 Schneider, G. *et al.* (1999) ''Scaffold-hopping'' by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.* 38, 2894–2896
16 Marshall, G.R. *et al.* (1979) The conformational parameter in drug design: the active analog approach. In *Computer-Assisted Drug Design*, (vol. 112) (Olson, E.C. and Christoffersen, R.E., eds) pp. 205–225, American Chemical Society
17 Green, J. *et al.* (1994) Chemical function queries for 3D database search. *J. Chem. Inf. Comput. Sci.* 34, 1297–1308
18 Krovat, E.M. and Langer, T. (2003) Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. *J. Med. Chem.* 46, 716–726
19 Wolber, G. and Langer, T. (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model* 45, 160–169
20 *Catalyst*. Accelrys, Accelrys Inc. www.accelrys.com.
21 *Phase*. Schrödinger Inc. www.schrodinger.com., retrieved August 2007
22 *Molecular Operating Environment* (*MOE*). Chemical Computing Group. www.chemcomp.com., retrieved August 2007
23 SMARTS – A Language for Describing Molecular Patterns, http://www.daylight.com/dayhtml_tutorials/languages/smarts/index.html., retrieved August 2007
24 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, 551–581
25 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. La Jolla, CA
26 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, La Jolla, CA
27 Jones, G. *et al.* (1995) A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J. Comput. Aid. Mol. Des.* 9, 532–549
28 GALAHAD. Tripos. www.tripos.com.
29 Cottrell, S.J. *et al.* (2004) Generation of multiple pharmacophore hypotheses using multiobjective optimisation techniques. *J. Comput. Aid. Mol. Des.* 18, 665–682
30 Richmond, N.J. *et al.* (2004) Alignment of three-dimensional molecules using an image recognition algorithm. *J. Mol. Graph Model* 23, 199–209
31 Lemmen, C. and Lengauer, T. (2000) Computational methods for the structural alignment of molecules. *J. Comput. Aid. Mol. Des.* 14, 215–232
32 Blaney, J.M. *et al.* (1990) DGEOM: Distance Geometry. In QCPE 590 Quantum Chemistry Program Exchange
33 Bron, C. and Kerbosch, J. (1973) Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM* 16, 575–577
34 Barnum, D. *et al.* (1995) Identification of common functional configurations among molecules. *J. Chem. Inf. Comput. Sci.* 36, 563–571
35 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.* 0, 647–671