Part 6

3D Pharmacophore Modeling
Introduction

LigandScout enables chemists and molecular modelers to automatically derive 3D-interaction feature models starting from 1) a macromolecular ligand complex, 2) a set of ligands without active-site information, and 3) an active site where there are no ligands present. Within the platform the user can utilize these models to rapidly search very large libraries (millions of compounds) virtually to find new hit compounds, prioritize them for synthesis and biological testing, and make various 3D pharmacophore based alignment experiments to support lead optimization projects. The first version of LigandScout released in 2005, featured capabilities to automatically derive pharmacophore models from protein-ligand complexes.[1–4] Its evolution during the last ten years has expanded to include expert capabilities, such as, pharmacophore derivations from ligands, methods to explore active sites without ligands, docking, advanced filtering tools, fragment based approaches, advanced virtual screening, and alignment capacities available to users within an intuitive and friendly graphical user interface (Figure 20.1). LigandScout has been used widely by researchers in industry and academic research for early hit and lead discovery activities on a diverse number of clinically relevant targets such as, enzymes, G protein-coupled receptors (GPCRs), kinases, ion-channel, cytokine, integrin, and nuclear receptors to name a few. Recent examples in the literature cite the use of LigandScout for rapidly visualizing and deciphering key interaction features between proteins and ligands,[5,6] using 3D-pharmacophore models for virtual screening to successfully find biologically active compounds,[7,8] target fishing,[9,10] drug repurposing,[11] exploring protein-protein interfaces,[12–14] and profiling drug targets for side effects.[15–17] Furthermore, LigandScout pharmacophore models and virtual screening have been reported to outperform other virtual screening methods in a benchmarking study performed using C-X-C motif chemokine receptor type 4 (CXCR-4) antagonists.[18]

This chapter will review theory related to pharmacophore models and guide the user through six essential workflows using LigandScout: 1) structure-based pharmacophore modeling, 2) ligand-based pharmacophore modeling, 3) creating shared-feature pharmacophore models, 4) accurate virtual screening and pharmacophore editing in the
Figure 20.1 The graphical user interface of LigandScout, a molecular design platform that specializes in pharmacophore modeling and virtual screening technology to support hit finding and lead optimization activities in early stage drug discovery research.
active site, 5) hit analysis, and 6) parallel virtual screening. Tutorials presented in this chapter, as well as other advanced workflows for design not covered herein, and data sets are available upon request.[19]

Theory: 3D Pharmacophores

The tutorials in this chapter involve the creation of 3D-chemical feature pharmacophore models and the use of these models to find biologically active molecules using virtual screening methods. To better understand the academic concepts underlying these models we present in this section the theory behind the concept of a pharmacophore and the feature definitions assigned to a pharmacophore model. The fundamental function of the alignment algorithms and virtual screening modes implemented in LigandScout find foundation in the chemical feature definitions.

Pharmacophore modeling together with virtual screening have become increasingly popular in the last decades and matured to a valuable and efficient basis for a wide variety of computer-aided drug design projects.[20] Due to their simplistic nature, pharmacophores are easy to understand and illustrate, which renders them rather useful as a means to describe and explain ligand-target binding interactions. However, there still exists some confusion about the term pharmacophore which, depending on the background and context, is often attributed with different meanings. Historically, medicinal chemists used (and still use) the term pharmacophore to denote common structural or functional elements that are essential for the activity of a set of compounds toward a particular biological target. However, in 1998, Camille Wermuth, a well-known medicinal chemist, submitted a refined definition of a pharmacophore to the IUPAC that is now officially recognized.[21] It states: “A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supra-molecular interactions with a specific biological target structure and to trigger (or to block) its biological response.” According to this definition pharmacophores do not represent specific structural motifs of molecules (β-lactams, dihydropyridine) or associations of functional groups (e.g., primary amines, sulfonamides), but are an abstract description of essential steric and electronic properties that are required for an energetically favorable interaction between a ligand and receptor of the macromolecular target.[22] Pharmacophore models can thus be considered the largest common denominator of molecules that show a similar biological profile and are recognized by the same binding site of the target.[23]

Representation of Pharmacophore Models

To be useful for drug design, pharmacophore models must in some way uniformly represent the physico-chemical properties and location of functional groups involved in ligand-target interactions, as well as the different types of non-covalent bonding and their characteristics in a manner that is easy for humans to comprehend. The most common representation of pharmacophores is a spatial arrangement of so-called chemical (or pharmacophoric) features that describe essential structural elements and/or observed ligand-receptor interactions by means of geometric entities. Although this
representation is quite simple, it sufficiently fulfills the above requirements and has
found general acceptance among medicinal chemists because they already think of
molecules in terms of their pharmacophore space when modifying compounds in lead
optimization projects.[24]

However, from the perspective of developing pharmacophore features for computer-aided design purposes, special attention must be made regarding the right level
of abstraction of the chemical feature types used in the construction of the pharma-
cophore models that will be used for virtual screening. Rather general definitions
result in models that are universal, at the cost of selectivity. Selectivity, however, is
also an important issue for the quality of pharmacophore models and features that are
too general may need to be refined to additionally include characteristics of the
underlying functional groups instead of merely reflecting universal chemical func-
tionality. Being too restrictive, on the other hand, will increase the number of different feature types at the cost of comparability and the ability to identify novel, structurally unrelated, chemical compounds. Therefore, the design of pharmacophore modeling software faces a trade-off between a generally applicable feature set that is universal and, at the same time, still selective enough to reflect all relevant types of observed ligand-receptor interactions. LigandScout supports the derivation of fourteen feature types with defined geometries and tolerances including hydrogen bond (H-bond) donors, H-bond acceptors, hydrophobic, aromatic interactions, positive and negative ionizable features, coordination to metal ions as well as excluded volumes. The graphical representation of these features is shown in Figure 20.2. The

![Pharmacophore depictions in LigandScout. Hydrogen-bond donors and acceptors can be depicted as vectors or spheres depending on the model requirements.](image-url)
following sections give a brief overview of the most important types of ligand-receptor interactions and their corresponding geometric representations in pharmacophore models in LigandScout.

**Hydrogen-Bonding Interactions**

Hydrogen bonding is an attractive interaction of electropositive hydrogen atoms with an electronegative atom (H-acceptor) like oxygen, fluorine, or nitrogen. The participating hydrogen must be covalently bound to another electronegative atom (H-donor) to create the hydrogen bond. Hydrogen bonding is a relatively strong interaction (~2–167 kJ/mol) and one of the most important for the formation of strong ligand-receptor complexes.[25] To capture the characteristics of hydrogen-bonding interactions (Figure 20.3), they are usually modeled as a position with a certain tolerance for the acceptor (donor) atom and a projected point (also with a certain tolerance) for the position of the corresponding donor (acceptor) atom. Together these two positions form a vector that constrains the direction of the H-bonding axis and also the location of the interacting atom in the target receptor. When the direction constraint is omitted, they become less specific and will match any acceptor/donor atom irrespective of whether essential geometric preconditions for the formation of a hydrogen-bond are actually fulfilled.

**Hydrophobic Interactions**

Hydrophobic (lipophilic) interactions occur when non-polar amino acid side-chains in the protein come into close contact with lipophilic groups of the ligand. Lipophilic groups include, for example, aromatic and aliphatic hydrocarbons, halogen substituents like chlorine and fluorine, and many heterocycles like thiophene, and furan. Since lipophilic areas on the protein and ligand surface are not capable of participating in any polar interactions, attractive forces are negligible for the effect of hydrophobic interactions. They are driven instead by the displacement of water molecules from non-polar areas in the binding pocket to the outside of the protein. This leads to a higher entropy of the system due to a gain in mobility and allows the now unconstrained water molecules to form energetically favorable hydrogen bonds. According to the Gibbs free energy equation, \( \Delta G = \Delta H - T\Delta S \), both contributions will lower the change in free energy, \( \Delta G \), for the interaction and thus increase the ligand's overall binding affinity. Since hydrophobic interactions are undirected, they can be represented as tolerance spheres (Figure 20.2), which are located in the center of hydrophobic atom chains, branches or groups of the ligand.
Electron-rich π-systems like aromatic rings are capable of forming strong attractive interactions with other π-systems (π-stacking) and adjacent cationic groups (e.g., metal ions, ammonium cations in protein side-chains).[26] The interaction energies are the same order of magnitude as hydrogen bonds and thus play an important role in various aspects of molecular biology (e.g., stabilization of DNA and protein structures, enzymatic catalysis, and molecular recognition). Since cation-π and π-π interactions require a certain relative geometric configuration of the interacting counterparts (Figure 20.4), they belong to the class of directed interactions. In pharmacophore models, aromatic features are therefore at least represented by tolerance spheres located in the center of the aromatic ring-system (Figure 20.2). To account for the directional aspects of aromatic interactions, they are often attributed with additional information about the spatial orientation of the aromatic ring-system in the form of a ring-plane normal or two points that define this vector.

Ionic Interactions

Ionic interactions are strong attractive interactions (energies > 400 kJ/mol) that occur between oppositely charged groups of the ligand and the protein environment. Positive or negative ionizable areas can be single atoms (e.g., metal cations, ammonium ions) or groups of atoms (e.g., carboxylic acids, guanidines, aromatic heterocycles) that are likely to be protonated or deprotonated at physiological pH. Ionic interactions are of electrostatic nature and thus undirected which allows the corresponding pharmacophoric features to be represented by simple tolerance spheres.

Metal Complexation

Some proteins contain metal ions as co-factors. A prominent example involves metalloproteases that contain Zn^{2+} ions that are coordinated to the protein via three amino acids as shown in Figure 20.5.[27] In such proteins, a coordination complex of the metal ion with suitable electron donating atoms or functional groups of the ligand is often the
most important contribution to the overall binding affinity and essential for the ligand’s mode of action. Functional groups and structural elements that exert a strong affinity for metal ions are, for example, thiols R‐SH, hydroxamates R‐CONHOH, or sulfur and nitrogen containing heterocycles. In pharmacophore models, metal binding interactions are represented by tolerance spheres located on single atoms or in the center of groups that are capable of interacting with the metal ions. To additionally define or constrain the location of the coordinated metal ion or accommodate a particular coordination geometry, a vector representation is incorporated as well (Figure 20.2).

**Ligand Shape Constraints**

The chemical features in a pharmacophore model represent necessary, but not all of the, characteristics active molecules must possess to achieve specific, high affinity binding to a given target receptor. A molecule may be retrieved by a pharmacophore model in a virtual screening exercise because it is capable of matching a set of features that is entirely consistent with the pharmacophore model, but when tested it fails to show activity. A plausible reason for this that can be explained by modeling (rather than biological pathways) is that some part of the molecule experiences a steric clash with the receptor side-chains if it were to bind in the mode described by the pharmacophore model. A common way to avoid this situation is to add exclusion volumes to the model. Excluded volumes are represented by variably sized spheres (Figure 20.2) corresponding to variably assigned tolerances, which indicate regions of “forbidden” space, thereby defining a restricted area that a hit structure should not occupy when it is aligned to the pharmacophore. This means that molecules retrieved in a virtual screening exercise must not only fit the interaction features defined by the model but they must also fit within the region defined by the excluded volumes as well. Obviously,
the correct placement of excluded volumes is important in order to avoid the possibility of missing active compounds. The most reliable source of information for a proper placement of excluded volumes is the crystallographic structure of the ligand bound receptor. Such receptor-based excluded volumes are centered on appropriate atoms of the binding-site surface with sizes dictated by the corresponding atomic van der Waals radii. A clash of the aligned molecule with one of the excluded volume spheres directly corresponds to a steric overlap with an atom of the receptor surface and indicates a presumably poor fit of the molecule. When the three-dimensional receptor structure is not available (which is often the case), and ligand-based pharmacophore derivation methods are used, then the placement of exclusion volumes is less straightforward. In this case LigandScout will automatically distribute the location and size of the exclusion volumes spheres based on the union of the molecular shapes of a set of aligned known actives. The user can edit these excluded volumes manually to adjust the selectivity of the model as needed.

Pharmacophore Modeling

Pharmacophore models can be created by a variety of methods including manual construction, automated perception from the structure of one or more ligands, and receptor-based deduction from a crystallographic structure. The particular method or workflow that is best suited for a given modeling problem depends on a number of factors like the goals of the study, nature and quality of available data, computational resources, and the aim and further use of the created pharmacophore model. The following sections will give an overview of the methodological details and applicability of the most commonly used approaches for creating pharmacophore models.

Manual Pharmacophore Construction

The simplest way (in terms of algorithmic complexity) to create a pharmacophore model is manual construction based on information about the structure and/or special characteristics of a series of known active ligands. A manually constructed pharmacophore can be quite advantageous, particularly if it is derived from the x-ray structure of a ligand in its binding conformation or from a ligand with low conformational flexibility. In either case, the locations of the pharmacophoric features are essentially pinned down, so that conformational flexibility, one of the biggest uncertainties, is eliminated. However, there is still the question of which features should be incorporated into the model, which is not always easy to infer without additional information such as the structure of a ligand-receptor complex. With the advent of powerful computer-aided methods for pharmacophore modeling, the importance of a manual pharmacophore construction from scratch has largely diminished. Nevertheless, manual pharmacophore generation is still possible using LigandScout. Capabilities, such as, creating new features, modifying a feature type, removing and/or disabling a feature, and changing feature tolerances, will influence the selectivity and the performance of a pharmacophore model at the users discretion.
**Structure-Based Pharmacophore Models**

The availability of information about the three-dimensional structure (e.g., from NMR experiments or X-ray crystallography) of a ligand/receptor complex is a tremendous advantage when it comes to the development of high quality pharmacophore models. Knowledge of the 3D structure of the bound ligand and the surrounding receptor surface allows for the analysis of essential interactions and the correct placement of corresponding pharmacophoric features. Furthermore, detailed information about regions that are restricted can be incorporated into the final pharmacophore model in the form of excluded volumes and thus constrain the shape of ligands that are retrieved from virtual screening exercises.

A fundamental step in the development of a receptor-based (often also called structure-based) pharmacophore model is the analysis of the binding site and its associated ligand to identify potential interaction points. A number of methods can in principle be used to identify such regions. LigandScout, for example, takes a direct approach and derives a pharmacophore model from a single ligand/receptor complex as follows: 1) After the user loaded the protein-ligand complex, LigandScout analyzes and corrects hybridization states of unsaturated bonds and aromatic rings. 2) Following this step, both the ligand and binding pocket amino acids are analyzed for the presence of atoms and groups that can take part in hydrogen-bonding, hydrophobic, aromatic, ionic, and metal binding interactions. If there are complimentary interaction partners between the ligand and binding site functionalities, LigandScout will automatically add a feature to the model. Pharmacophoric feature detection can be customized with respect to interaction specific geometric characteristics like allowed distances and angle ranges. Whether a feature is incorporated into the final pharmacophore model depends on its location relative to a complementary feature in the binding-site. For example, a hydrogen-bond acceptor feature located on an acceptor atom of the ligand is only included if there is an opposing hydrogen-donor feature on the receptor side within a certain distance and angle range. 3) After all complementary feature pairs of the complex have been analyzed and the corresponding ligand-side features have been put into the derived pharmacophore model, exclusion volume spheres are added to mimic the shape of the binding pocket. Figure 20.6 illustrates a structure-based pharmacophore model for a cyclin dependent kinase 2 (CDK2) inhibitor from the PDB ID: 1KE8 that was created automatically using LigandScout.

**Ligand-Based Pharmacophore Models**

When information about the three-dimensional structure of the receptor is limited or not available, but a sufficient number of actives are available, then ligand-based methods provide an alternative way to leverage the available information and develop pharmacophore models that can be used for virtual screening or lead optimization workflows. An important precondition for ligand-based methods to work and deliver good models is that the ligands used for model generation should bind to the same receptor at the same location in the same way. Otherwise, the resulting pharmacophore models will not represent the correct mode of action and are essentially useless.
The derivation of pharmacophore models from a set of ligands involves many different algorithms.[30,31] For example, after the import and preparation of the input structures (SMI, SDF, or MOL files), one will need to generate a sufficiently large and diverse set of low energy ligand conformations. This is done because the bioactive conformations of the input ligands may not be known but it can be assumed that one conformation in each set of generated conformers is at least a good approximation thereof. This multi-conformational set of active ligands is designated the training set by the user and can be computed in an automated manner using the conformer generator iCon in LigandScout from the menu pull-down options or from a shortcut button above the table of ligands. The next step is at the heart of the overall procedure and aims to identify a chemical feature pattern[32] that is common to all training-set ligands and can be superimposed with at least one conformation of each ligand. Since often more than one such pharmacophoric pattern can be found due to multiple conformations of the ligands, LigandScout will produce multiple pharmacophore model solutions and rank them with a fitness function. From these ranked solutions the user can select the best model to fit the project needs. If the best model is not clear, then the user could opt to take the pharmacophore model with the best fit score or test each of the models against a set of active and inactive molecules in a virtual screening validation and ROC curve generation procedure.[31] Methods for pharmacophore validation can be divided into three categories:[30]

1) Statistical significance analysis and randomization tests.
2) Enrichment-based methods involve assuring the ability to recover active molecules from a test database in which a small number of known actives have been hidden among randomly selected compounds. Database mining and the utilization of receiver operating characteristic (ROC) curves[32] fall into this category.
3) Biological testing of matching molecules.

Figure 20.6 Receptor-based or structure-based pharmacophore generated by LigandScout for the CDK2/inhibitor complex with PDB ID: 1KE8. Gray spheres represent exclusion volumes that model the shape of the receptor surface. Yellow spheres represent hydrophobic interactions, green and red arrows represent hydrogen-bond donor and acceptor features, respectively. Vectors define the direction of the hydrogen bonds.
If pharmacophore validation results indicate a generally unsatisfying quality of the generated models, they may be refined manually (e.g., deletion/addition of features, modifying tolerances or features (vector to sphere or interpolation) applicable if only small changes are required) or the whole modeling procedure must be repeated with different set-up (e.g., changes to the composition of the training- and/or test-sets, refinement of ligand conformers, and editing pharmacophore model generation parameters) until acceptable results are obtained. The high number of influential variables and the disregard of the receptor structure makes pure ligand-based modeling relatively error prone and leaves much room for interpretation. The algorithmic power of the employed software, a high expertise of the user in terms of knowledge about the biological target, and a thorough validation of the obtained results are therefore critical for the successful application of this modeling approach.

### 3D Pharmacophore-Based Virtual Screening

Because of their abstract nature and simplicity, 3D pharmacophore models represent efficient filters for the virtual screening of large compound libraries.[33,34] The computational complexity of the hit identification process in virtual screening is greatly reduced by the sparse pharmacophoric representation of ligand–target interactions which results in rapid overall search times. Furthermore, since pharmacophore-based queries are based on pharmacophore feature alignments rather than compound scaffold alignments, they are able to find hit molecules with diverse scaffolds when compared to the original ligands used for the generating the pharmacophore models.[35] This is of special interest for researchers who need to find novel molecules that are patentable or lead candidates with better ADMET properties, and/or higher activity, and/or selectivity toward the target.[36,37]

### 3D Pharmacophore Creation

The first step in a typical pharmacophore-based virtual screening campaign is to create a query pharmacophore model that specifies the type and geometric constraints of the chemical features that have to be matched by the database molecules in a virtual screening experiment. Ligand-based and structure-based pharmacophore models can be created (see Tables 20.1, 20.3 & Fig. 20.9, 20.11) and used separately or in combination via parallel virtual screening workflows in LigandScout if desired. The strategy adopted depends on the goals of the project.

### Annotated Database Creation

One of the biggest challenges in pharmacophore modeling and virtual screening workflows is designing data sets that can be used for pharmacophore model validation.[36] In particular the design of a database of active molecules and inactive molecules that can be used for virtual screening in order to understand how well the pharmacophore models can distinguish between true active and inactive molecules is important before screening a library of commercially available untested compounds. The idea is to edit the pharmacophore model to maximize the retrieval of true active compounds and minimize the retrieval of inactive molecules. Data for creating such data sets may come from in house
sources of experimentally tested compounds or they could be extracted from public sources such as ChEMBL or PubMed.[38,39] LigandScout Expert/Knime provides a Knime node[40] for extraction of compounds from ChEMBL. In addition, databases of untested compounds from in house or commercial sources must also be computed for virtual screening campaigns using LigandScout in order to find new hit molecules.[36] An important aspect that needs to be considered when screening molecule libraries against 3D pharmacophores is conformational flexibility. Most major software applications including LigandScout deal with this problem by creating dedicated screening databases that store pre-computed conformations for each of the molecules. Another approach is to tweak the conformation of the molecules on the fly in the pharmacophore fitting process.[41] The advantage of the latter approach is the lower storage requirements. However, it also has the major disadvantage that the screening process is considerably slower and a dramatic reduction of the conformational search space while aligning bears the danger of falling into a local minimum.[29,42] Nowadays enough hard disk storage is available and screening databases with pre-generated conformations are clearly preferred. These databases are usually generated once or can be easily updated without having to recompute the entire library using LigandScout tools and reused whenever needed.

Virtual Screening-Database Searching

In LigandScout the 3D pharmacophore search of a library of compounds is implemented as a multistep filtering process. First, the user must open and designate the pharmacophore models and databases in the virtual screening perspective of LigandScout. The first step is a fast pre-filtering that aims to quickly identify and eliminate all molecules that cannot be fitted to the query pharmacophore model in 3D space. Only molecules that pass this pre-filtering step need to be processed in the final accurate, but computationally expensive, 3D alignment step. In the alignment step, LigandScout will examine closely the conformations of the remaining molecules that might match the query to see if they are able to match the spatial arrangement of the query features. Special care must be taken in this step because an ultimate decision has to be made whether to reject a database compound or to put it in the hit list and, thus, directly influencing the quality of the obtained screening results. To correctly identify a match of a molecule to the query pharmacophore within the defined feature tolerances an overlay in 3D space is required. This overlay is also necessary to check and score additional constraints imposed by vector features like hydrogen-bond acceptors/donors, plane features like aromatic rings, and exclusion/inclusion volume spheres. Commercial software for pharmacophore modeling that incorporate state-of-the-art screening functionality all perform some sort of geometric alignment in the 3D pharmacophore matching step, which is usually done by minimizing the root-mean-square deviation (RMSD) between associated feature pairs.[43]

Hit-List Analysis

Hits retrieved from the 3D database search are a good and recommended starting point for the validation and refinement of the query pharmacophore model. There are several useful measures to characterize the obtained hit list like Sensitivity, Specificity, Yield of Actives, and many others, which are described elsewhere.[44,45] A modern tool for the assessment of screening results against datasets of active and inactive compounds are Receiver Operating Characteristic curves (ROC).[32,46] A ROC curve displays the rate for retrieving true positives (actives) plotted against the rate of retrieving false positives
The Y-coordinate of the ROC curve represents the true-positive rate (rate of retrieving actives), whereas the X-coordinate denotes the appropriate false-positive rate (rate of retrieving inactives or decoys). An ideal curve would rise vertically along the Y-axis until it reaches the maximum true positive rate, which is one (1), and then continues horizontally to the right, which means that the hit list contains all active compounds in the database and that not one of the hits is a false positive. The ROC curve of a random database search is represented by the median.

On the basis of a hit-list analysis with the above measures and tools, the pharmacophore model is often refined to achieve more satisfying results. Adaption of feature definitions, modification of feature tolerances, addition or removal of features and exclusion volumes are some of the adjustments that can help to tune a pharmacophore model. Another possibility is to modify the database by readjusting the number of pre-generated conformations to address molecular flexibility more adequately. Because pharmacophore modeling and database screening are very complex tasks, several iterations of screening, analysis, and refinement are usually necessary to achieve good results. After the model has been suitably refined based on ROC curve analysis to the desired true active over false positive retrieval rate, the model is ready for virtual screening of set of compounds that have not been tested on the target. The hits retrieved will be ranked using a pharmacophore fit score. Higher pharmacophore fit scores represent a better fit to the model.
Tutorial: Creating 3D-Pharmacophore Models Using LigandScout

The next two sections detail stepwise tutorials for pharmacophore modeling with LigandScout. Figure 20.8 outlines a recommended and general pharmacophore modeling and virtual screening workflow.

Figure 20.8 A general and recommended pharmacophore modeling and virtual screening workflow using LigandScout.

Creating Structure-Based Pharmacophores From a Ligand-Protein Complex

Table 20.1 Description of actions for structure-based pharmacophore creation using LigandScout.[1]
Figure 20.9  Snapshots of the areas of the LigandScout GUI used for generating a structure-based pharmacophore model. The numbers circled in blue correspond to the steps in the description. Source: LigandScout.[1]
Description: Create a Structure-Based Pharmacophore Model

(Numbering below refers to numbers circled in blue in Figure 20.9).

1) Type the PDB ID: “1ke6” in the upper right area of the screen and press the button “Download 1ke6.” The protein will be downloaded from the protein data bank (PDB)[47,48] and displayed. You must be connected to the Internet to access the PDB. If you are not connected to the Internet you can use File → Open from the LigandScout pull down menu. In this case your PDB file should be stored locally on your computer.

2) Click on the yellow box within the protein to zoom into the active site (this is the “macromolecule view”).

3) Once inside the active site you will see the ligand. Since structures in the PDB contain incomplete information, you should always check whether the ligand is correctly depicted, for example, whether bond orders are correct.

4) If you find an error, you can edit the bond order by selecting a bond (by clicking on it either in the 2D or 3D view using the left mouse button) and use the bond order icon pull-down to select the correct bond order. Similarly you can select the bond you wish to edit using your left mouse button and use the keyboard numbers such as, “1” to create a single bond, “2” to create a double bond, or “3” to create a triple bond.

5) Once the ligand is chemically correct, create a pharmacophore by pressing Ctrl-F9 on Windows/Linux machines or Cmd-F9 on Apple machines.

Where to go from here:

● Move the pharmacophore model to the virtual screening perspective for virtual screening.

● Move the pharmacophore model to the alignment perspective for creating shared pharmacophore models.

● Edit the pharmacophore to make it more or less restrictive based on your project requirements (Table 20.2).

Create a Shared Feature Pharmacophore Model From Multiple Ligand-Protein Complexes

Table 20.2 Description of actions for creating a shared feature pharmacophore model using LigandScout.[1]

<table>
<thead>
<tr>
<th>Experience level: intermediate</th>
<th>Time needed: 15 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Views</td>
<td>Sequence</td>
</tr>
<tr>
<td>Macromolecule view</td>
<td>Download PDB File “1ke6”</td>
</tr>
<tr>
<td>Active site view</td>
<td>Check &amp; correct ligand</td>
</tr>
<tr>
<td>Alignment view</td>
<td>Create pharmacophore</td>
</tr>
<tr>
<td></td>
<td>Add ligand and pharmacophore to alignment view</td>
</tr>
</tbody>
</table>
### Table 20.2 (Continued)

<table>
<thead>
<tr>
<th>Views</th>
<th>Sequence</th>
<th>User Controls</th>
<th>Advanced Controls</th>
</tr>
</thead>
</table>
|       | ● Repeat these steps for “1ke7” and “1ke8”  
● Create a “shared feature pharmacophore” from the three pharmacophores  
● Align the three ligands to the new pharmacophore | ● Move pharmacophore model to alignment  
● Generate shared feature pharmacophore | |

Figure 20.10   Snapshots of the areas of the LigandScout GUI used for generating a shared featured pharmacophore model. The numbers circled in blue correspond to the steps in the description.  
Source: LigandScout.[1]

**Description: Create a Shared Feature Pharmacophore and Align it to Ligands**

(Numbering below corresponds to the numbers circled in blue in Figure 20.10).

First follow the steps described in Table 20.1 & Fig. 20.9 to generate a structure based pharmacophore model for PDB ID: 1KE6.

1) Use the data exchange widget and select “Copy current Ligand to Alignment Perspective” and the molecule “(1ke6) LS2201” will be added to the alignment view. Use the data exchange widget—select “Copy current Pharmacophore(s) to Alignment Perspective” to move pharmacophore model to the alignment perspective. Repeat the steps above using the PDB files with PDB ID: “1ke7,” and “1ke8.” After generating pharmacophore models for the three different PDB entries you should have six elements (three ligands and three pharmacophore models) in the alignment perspective. Click on the “Alignment” tab to move to the alignment perspective. There you will see your ligands and pharmacophore models in a list.
2) Select 1ke6 pharmacophore and use the first button above the table to set it as the reference (Set Reference). The color of the text will become red. Select the three pharmacophores 1ke6 (red text), 1ke7 and 1ke8 using Control key (Command on Apple) and left mouse button. You will see them appear in the 3D window.

3) Press the align button (icon with two squares) to align the pharmacophore models. In the 3D window you will see that the three models are now aligned.

4) Press the icon next to the align molecules icon to “Generate shared feature pharmacophore” button. A new pharmacophore model called “Shared [LS2201, LS3201, LS4299]” is appended to the list. Deselect all items in the list by clicking the empty space below the table with your left mouse. Now select the shared feature pharmacophore model you just created and mark it as the reference element using the “Set reference” button (the text will turn red). Use the control key (command on Apple) to select the ligands, 1ke6, 1ke7, and 1ke8 identified by the blue icon to the left of each entry and the reference shared pharmacophore model (now red). Hide the three pharmacophores derived from 1ke6, 1ke7, and 1ke8 by clicking on the eye symbols on the right next to these alignment entries. Now press the “Align selected elements” (icon with two squares) button to align the three ligands to the shared feature pharmacophore model.

5) You will see the three ligands from the x-ray structures aligned to the shared feature pharmacophore derived from the three PDB complexes. Move your shared feature model to the screening perspective for virtual screening (Table 20.3).

Create Ligand-Based Pharmacophore Models

Table 20.3 Description of actions for creating a ligand-based pharmacophore model using LigandScout.[1]

<table>
<thead>
<tr>
<th>Experience level: intermediate</th>
<th>Time needed: 15 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views</strong></td>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>● Ligand-based Modeling</td>
<td>● Import ligands</td>
</tr>
<tr>
<td></td>
<td>● Generate conformations</td>
</tr>
<tr>
<td></td>
<td>● Cluster ligand-set</td>
</tr>
<tr>
<td></td>
<td>● Generate ligand-based pharmacophore model</td>
</tr>
</tbody>
</table>
Figure 20.11  Snapshots of the areas of the LigandScout GUI used for generating a ligand-based pharmacophore model. The numbers circled in blue correspond to the steps in the description. Source: LigandScout.[1]
Description: Ligand-Based Pharmacophore Model Creation

(Numbers below correspond to the numbers in blue in Figure 20.11).

1) Open the SMILES file (“cdk2.smi”) by using the pull-down menu the “File → Open.” The ligands will appear in a table below the 3D view and are set to “training set” by default.

2) a) Generate conformations (3D structures) by clicking on the button (two blue arrows). LigandScout will prompt you with options. Click on the button “Apply FAST Settings.” This will generate a maximum of 25 conformations per ligand in your training set.

b) Cluster the ligand set according to the geometry of the 3D pharmacophore features by clicking on the “cluster” button. Use the default settings when prompted with settings choices and start the clustering process by pressing the “OK” button. LigandScout will create a new column in the table called “Cluster ID.”

c) Sort the table using this column by clicking on the column header “Cluster ID.”

3) Select all compounds in the table by selecting the first molecule in the list using your left mouse button then scroll to the end of the list and press the shift key and select the last molecule in list using your left mouse button. All molecules should be selected. Use the pull-down menu “Ligand-Set” → “Flag selected molecules as Ignored” to mark all molecules as ignored. Next select only the molecules with cluster id “1” using your left mouse and shift key. Use the pull-down menu “Ligand-Set” → “Flag selected molecules as Training-Set” to indicate that you want to compute a pharmacophore based on the ligands grouped in cluster 1. Only the ligands with cluster ID “1” should be marked as training the rest will be marked as ignored under the column header “Type.”

4) Click on the button “Create Ligand-Based Pharmacophore.” Give your pharmacophore model a name, for example,—“Cluster-1” and then press “Ok” to accept the default parameters. Since there will be more than one solution to this problem due to multiple conformations, LigandScout will generate 10 solutions (ligand-based pharmacophore models) with fit scores.

5) In the library view you will see a new column appear called “Feature pattern.” These patterns of colored squares correspond to the pharmacophore model features. Clicking on a colored square indicates which feature it corresponds to in the pharmacophore model displayed in the 3D-view.

6) The list of pharmacophore solutions will appear to the right of the ligand table. By selecting different pharmacophore models you will see them appear in the 3D view thereby allowing you to view how the ligands fit to each pharmacophore model solution. You can use the data exchange controls on the upper right corner of the 3D view (see Figure 20.10, number 1) to move the currently shown pharmacophore to other perspectives (e.g., the screening perspective) or save it using the “File -> Save as File” menu and selecting the appropriate file type (PML or PMZ).
Where to go from here:

- Follow the same procedure to generate a pharmacophore model for each cluster.
- Align structure-based pharmacophore models with ligand-based models.
- Virtual screening in LigandScout using several ligand-based pharmacophore models simultaneously.

**Tutorial: Pharmacophore-Based Virtual Screening Using LigandScout**

This section will provide a step-by-step procedure for virtual screening in the structure-based perspective, hit list analysis, and virtual screening in the virtual screening perspective. Using virtual screening methods, the user aims to retrieve the maximum enrichment of active compounds in a hit list by developing selective pharmacophore models. Therefore, before performing a virtual screening experiment on a commercial library of untested molecules, we recommend that the pharmacophore models be tested using virtual screening experiments on a set of known active compounds and a set of inactive compounds for the target of interest (see Table 20.4 & Fig. 20.12). If there are no known inactive compounds the user can create a data set of decoys using various methods described in the literature.[49] Such workflows enable the user to understand how well the pharmacophore models created are able to discriminate between actives and decoys (Table 20.4). A good pharmacophore model will be able to identify a significant portion of known active compounds and as few decoys as possible (Figure 20.7 & 20.8).

**Virtual Screening, Model Editing, and Viewing Hits in the Target Active Site**

<table>
<thead>
<tr>
<th>Experience level: advanced</th>
<th>Time needed: 15 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views</strong></td>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>Macromolecule view</td>
<td>Create pharmacophore for virtual screening</td>
</tr>
<tr>
<td>Active site view</td>
<td>Initiate virtual screening</td>
</tr>
<tr>
<td></td>
<td>Refine pharmacophore model</td>
</tr>
<tr>
<td></td>
<td>Repeat virtual screening</td>
</tr>
</tbody>
</table>
Description: Virtual Screening and Pharmacophore Model Editing

(Numbers below correspond to the numbers circled in blue in Figure 20.12).

In the structure-based perspective, create a structure-based pharmacophore as outlined in Table 20.1, Fig. 20.9 but instead of PDB ID: “1ke6” use PDB ID: “1ke7.”

1) Create a pharmacophore based on the 1ke7 complex by pressing the create pharmacophore icon in the tool bar (Figure 20.12). See Table 20.1, Fig. 20.9 if this is not clear.

2) Using the pull-down menu go to “Pharmacophore” → “Screen Pharmacophore Against External Library.”

Figure 20.12  Snapshots of the areas of the LigandScout GUI used for virtual screening and model editing. The numbers circled in blue correspond to the steps in the description. Source: LigandScout.[1]
3) Designate the screening settings in the dialog box. Designate the location and name of the multi-conformational database with known active compounds (file name: “cdk2-ligands.ldb”) in the box next to “Screened Database.”

4) In the 3D viewer edit the features of the pharmacophore model to test whether or not the edited model can retrieve more hits. Try designating a feature as optional (pull-down menu “Pharmacophore”→”Mark feature as optional”) and repeat the screening. You can also designate features as disabled and they will not be considered during screening but should you change your mind you change them back to enabled. The feature tolerances can be modified using the pull-down menu as well (Table 20.4).

5) View the screening results after modifying the pharmacophore model. Select hits in the hit list with your left mouse button and view them in the 1ke7 active site aligned to the pharmacophore model.

6) Use the widget to move the model or hits to other perspectives.

**Where to go from here:**

- Virtual screening of a library of actives and a library of inactives and create a ROC curve
- Docking of hits in the active site
- Analysis of hit list, compute standard properties, and filter based on desired properties.

### Analyzing Screening Results with Respect to the Binding Site

**Table 20.5** Description of actions for analyzing virtual screening results using LigandScout.[1]

<table>
<thead>
<tr>
<th>Experience level: intermediate</th>
<th>Time needed: 15 minutes</th>
<th>Prerequisites: Virtual Screening &amp; Pharmacophore Modeling Editing Tutorial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views</strong></td>
<td><strong>Sequence</strong></td>
<td><strong>User Controls</strong></td>
</tr>
<tr>
<td>● Library view (Table)</td>
<td>● Create or import screening results</td>
<td>● Library view (Table)</td>
</tr>
<tr>
<td>● Hierarchy view</td>
<td>● Sort virtual screening hit list</td>
<td>● Viewer controls</td>
</tr>
<tr>
<td></td>
<td>● Inspect and compare hits</td>
<td></td>
</tr>
<tr>
<td></td>
<td>● Discover controls of the library view</td>
<td></td>
</tr>
</tbody>
</table>
Figure 20.13  Snapshots of the areas of the LigandScout GUI used for analyzing hit lists from virtual screening. The numbers circled in blue correspond to the steps in the description. Source: LigandScout.[1]
Description: Analyzing Hits in the Active Site Using LigandScout

(Numbers below correspond to the numbers circled in blue in Figure 20.13).

Generate a structure based pharmacophore using the PDB ID: 1ke7 and perform virtual screening of the CDK2-ligands.ldb database as described in Table 20.4, Fig 20.12.

1) View the hit list in the library view.
2) Sort the hit list by the “Pharmacophore-Fit Score” by left clicking on the “Pharmacophore-Fit Score” column header (higher fit scores indicate a better match to the model). You may invert the sorting order by an additional click.
3) Select a compound of interest by single click on the corresponding table row. The hit ligand will be depicted in the active site view together with the original ligand from the x-ray structure and the pharmacophore model. Multiple compounds can be selected when holding shift or control (command on Apple) while clicking.
4) Bring compounds into view or out of view by clicking on the eye symbol to the left of the molecule in the table and to the right of the molecules in the hierarchy view.
5) The visibility of the core molecule (i.e., the original ligand) and other items in the active site view is toggled with the eye symbol located in the hierarchy view. Select a custom color for the core molecule using the square icon next to the eye icon in the hierarchy view. Calculation of the Gaussian Shape Similarity Score offers further functionality for scoring. Use the pull down menu “Library” → “Calculate Gaussian Shape Similarity Score” to compute a Gaussian Shape similarity score. A new property column “Gaussian Shape Similarity Score” will be added in the library view. Similarly you can compute standard properties using the pull-down menu “Library” → “Compute Standard Properties.” Several new columns will appear in the table. You can filter the table based on information (text or value) in any column (Table 20.6).

Where to go from here:

- Pharmacophore modeling: Creating shared and merged feature pharmacophores.
- Customize pharmacophore creation preferences.
- Filter you hit lists and export them to Excel for additional analysis.

Parallel Virtual Screening of Multiple Databases Using LigandScout

**Table 20.6** Description of actions for virtual screening of multiple databases simultaneously using LigandScout.[1]

<table>
<thead>
<tr>
<th>Experience level: intermediate</th>
<th>Time needed: 5 minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Views</strong></td>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>Screening view</td>
<td>Add pharmacophore model to screening view</td>
</tr>
<tr>
<td></td>
<td>Load virtual database</td>
</tr>
<tr>
<td></td>
<td>Perform virtual screening</td>
</tr>
</tbody>
</table>
Virtual Screening in the Screening Perspective of LigandScout

In contrast to the virtual screening procedure described in the Virtual Screening, Model Editing and Viewing Hits in the Target Active Site section, which featured a workflow for pharmacophore model refinement and hit viewing within an active site, the Screening Perspective in LigandScout provides additional options for managing virtual screening workflows. Those include such activities as screening multiple pharmacophore models against multiple libraries simultaneously and/or using the Boolean operator to create advanced screening workflows for fast library filtering using a set of pharmacophore models or model validation with automatic generation of ROC curves.

Description: Virtual Screening Using LigandScout

(Numbers correspond to the numbers circled in blue in Figure 20.14).

In the structure-based perspective, create a structure-based pharmacophore as outlined in Table 20.1, Fig. 20.9 but instead of PDB ID: “1ke6” use “1ke7.”

1) Use the widget to copy the pharmacophore model for 1ke7 to the virtual screening perspective.
2) Click on the “Load Screening Database” button to load the screening database “cdk2-ligands.ldb.”
3) Next to the database name will be an empty circle. Click on this empty circle to create a green box. By marking this library with a green box you have indicated to

Figure 20.14 Snapshots of the areas of the LigandScout GUI used for virtual screening. The numbers circled in blue correspond to the steps in the description. Source: LigandScout.[1]
LigandScout that this is a database of potentially active compounds that you wish to screen. If there is an empty circle next to the loaded database, it will not be screened. Select the pharmacophore model from the “Pharmacophores” list to make sure it is shown in the 3D view (1ke7 pharmacophore model should be there from step 1). Click on the “Perform Screening” button to initiate the virtual screening.

4) View the results in the table at the bottom of the screen. As you select hits from the table of you will see them in the 3D-view aligned to the pharmacophore model you used for screening.

Where to go from here:

- Analyze screening results using filtering functions
- Advanced screening: ROC curves and model combination
- Inject hits into the active site for viewing or for docking experiments.

Conclusions

The workflows presented herein are certainly not exhaustive by any means when considering the existing advanced functionalities of the molecular design platform LigandScout. However, they do cover the most fundamental aspects of pharmacophore modeling and virtual screening that serve as important starting points for computer-aided molecular design approaches for hit finding in early lead discovery research. The theory covered herein should also provide a deeper understanding of the representations and definitions of pharmacophore models and processes behind these in silico approaches that are used widely across various life science industries, including pharmaceutical drug discovery and academic research.

Acknowledgments

The authors are grateful to P Adaktylos, F Bendix, M Biely, A Dornhofer, R Kosara, M Langer, A Fuchs, J Hirschl, T Kainrad, L Moshuber, G Wolber, S Bachleitner, Y Aristei, E Bonelli, M Böhler, S Distinto, O Funk, T Huynh Buu, J Kirchmair, E M Krovat, C Lagner, P Markt, D Schuster, G Spitzer, O Wieder, R Martini and T Steindl for their valuable contributions to the development of LigandScout and the Inte:Ligand Scientific Advisory Board for their esteemed scientific guidance.

References


19 Tutorial materials, datasets and evaluation versions of LigandScout are available by request (support@inteligand.com).


38 ChEMBL. https://www.ebi.ac.uk/chembl/
44 O. F. Güner, Pharmacophore perception, development, and use in drug design, 2000, Int. Univ. Line
47 RCSB Protein Data Bank (PDB). http://www.rcsb.org/