Phase 3.4

Quick Start Guide

Schrödinger Press

Revision A, September 2012

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, command input and output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

Links to other locations in the current document or to other PDF documents are colored like this: Document Conventions.

In descriptions of command syntax, the following UNIX conventions are used: braces { } enclose a choice of required items, square brackets [ ] enclose optional items, and the bar symbol | separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

File name, path, and environment variable syntax is generally given with the UNIX conventions. To obtain the Windows conventions, replace the forward slash / with the backslash \ in path or directory names, and replace the $ at the beginning of an environment variable with a % at each end. For example, `$SCHRODINGER/maestro` becomes `%SCHRODINGER%\maestro`.

Keyboard references are given in the Windows convention by default, with Mac equivalents in parentheses, for example CTRL+H (⌘H). Where Mac equivalents are not given, COMMAND should be read in place of CTRL. The convention CTRL-H is not used.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Introduction

1.1 About Phase

Phase is a versatile product for pharmacophore perception, structure alignment, activity prediction, and 3D database searching. Given a set of molecules with high affinity for a particular protein target, Phase utilizes fine-grained conformational sampling and a range of scoring techniques to identify common pharmacophore hypotheses, which convey characteristics of 3D chemical structures that are purported to be critical for binding. Each hypothesis is accompanied by a set of aligned conformations that suggest the relative manner in which the molecules are likely to bind.

A given hypothesis may be combined with known activity data to create 3D QSAR models that identify overall aspects of molecular structure that govern activity. These models may be used in conjunction with the hypothesis to mine a 3D database for molecules that are most likely to exhibit strong activity toward the target.

Phase provides support for lead discovery, SAR development, lead optimization and lead expansion. Phase may also be used as a source of molecular alignments for third-party 3D QSAR programs.

Phase is integrated into Maestro, the graphical user interface (GUI) for all Schrödinger products. For more detailed information on Maestro, see the Maestro online help or the *Maestro User Manual*. For detailed information on Phase, see the *Phase User Manual*.

1.2 About this Document

This document provides tutorial instruction in the three main Phase workflows, both from the Maestro GUI and from the command line. The first three chapters give instruction in using the GUI; the following two chapters give instruction in using the command-line tools.

- Chapter 2 contains exercises on developing a pharmacophore model and building QSAR models.
- Chapter 3 contains exercises on preparing a database for searching. The database will contain both single structures, without conformers and sites, and structures for which the conformers and sites are generated and stored in the database.

- Chapter 4 contains exercises on searching the database for matches to a hypothesis, using a database and a hypothesis supplied with the distribution. You do not need to complete the exercises in Chapter 2 or Chapter 3 to work through this chapter.

To do the exercises, you must have access to an installed version of Maestro 9.3 and Phase 3.4. For installation instructions, see the *Installation Guide*. The exercises are designed for use with the Maestro profile (the default), not with the BioLuminate profile or the Elements profile.

A set of command-line tutorials is available. Each tutorial consists of a PDF file containing the instructions (in the documentation set) and a gzipped archive (`.tar`) file containing the necessary files (in the product distribution). The available tutorials are listed in Table 1.1.

Table 1.1. List of command-line tutorials

Tutorial	Archive file
Phase Command Line Tutorial: Developing Pharmacophore Models	`pharm_tutorial.tar.gz`
Phase Command Line Tutorial: Creating and Searching 3D Databases	`db_tutorial.tar.gz`

1.3 Preparing for the Exercises

To run the exercises, you need a working directory in which to store the input and output, and you need to copy the input files from the installation into your working directory. This is done automatically in the Tutorials panel, as described below. To copy the input files manually, just unzip the `phase` zip file from the `tutorials` directory of your installation into your working directory.

On Linux, you should first set the `SCHRODINGER` environment variable to the Schrödinger software installation directory, if it is not already set:

csh/tcsh: `setenv SCHRODINGER` *installation-path*

sh/bash/ksh: `export SCHRODINGER=`*installation-path*

If Maestro is not running, start it as follows:

- **Linux:** Enter the following command:

  ```
  $SCHRODINGER/maestro -profile Maestro &
  ```

- **Windows:** Double-click the Maestro icon on the desktop.

  You can also use Start → All Programs → Schrodinger-2012 → Maestro.

- **Mac:** Click the Maestro icon on the dock.

  If it is not on the dock, drag it there from the SchrodingerSuite2012 folder in your Applications folder, or start Maestro from that folder.

Now that Maestro is running, you can start the setup.

1. Choose Help → Tutorials.

   The Tutorials panel opens.

2. Ensure that the Show tutorials by option menu is set to Product, and the option menu below is labeled Product and set to All.

3. Select Phase Quick Start Guide in the table.

4. Enter the directory that you want to use for the tutorial in the Copy to text box, or click Browse and navigate to the directory.

   If the directory does not exist, it will be created for you, on confirmation. The default is your current working directory.

5. Click Copy.

   The tutorial files are copied to the specified directory, and a progress dialog box is displayed briefly.

If you used the default directory, the files are now in your current working directory, and you can skip the next two steps. Otherwise, you should set the working directory to the place that your tutorial files were copied to.

6. Choose Project → Change Directory.

7. Navigate to the directory you specified for the tutorial files, and click OK.

You can close the Tutorials panel now, and proceed with the exercises.

Building Pharmacophore Hypotheses

This chapter is designed to help you become familiar with the Develop Common Pharmacophore Hypotheses workflow of Phase. This workflow involves the identification of common pharmacophore hypotheses from a set of active ligands. A common pharmacophore hypothesis is a spatial arrangement of chemical features common to two or more active ligands, which is proposed to explain the key interactions involved in ligand binding. Each hypothesis identified by Phase is scored according to how well the active ligands superimpose when they are aligned on the features associated with that hypothesis. A high-scoring hypothesis might be used to search a 3D database for new potentially active molecules, or it might be used to align a series of ligands in order to create a 3D QSAR model.

The structures and data used in this portion of the tutorial were taken from *J. Med. Chem.* **2003**, *46*, 716-726. The data set consists of 50 angiotensin AT_1 antagonists, divided into training and test sets of 25 ligands apiece. For convenience and clarity, ligand names have been assigned to indicate membership in the training set (train01, train02, ...) or the test set (test01, test02, ...). You will be developing pharmacophore and QSAR models from the training set, and applying them to the test set.

2.1 Starting the Exercises

To begin the tutorial, first save the scratch project to preserve the results:

1. Click the Save As button on the Project toolbar.

   

   The Save Project As project selector opens.

2. In the File name text box, type `at1_tutorial`.

3. Click Save.

Maestro provides a wizard to guide you through the steps of the Develop Pharmacophore Model workflow, in the appropriate order. This workflow allows you to identify common pharmacophore hypotheses and create 3D QSAR models.

4. Choose one of the following in the main window:

   - Applications → Phase → Develop Common Pharmacophore Hypotheses
   - Tasks → Pharmacophore Modeling → Develop Common Hypotheses

   The Develop Common Pharmacophore Hypotheses panel opens.

At the bottom of the panel is a series of buttons labeled Prepare Ligands, Create Sites, and so on. These are the various steps in the workflow, and clicking on any enabled button takes you directly to that step. Currently, all buttons except Prepare Ligands are disabled. A button is enabled only if the prerequisite steps in the workflow have been successfully completed. The Step menu at the top of the panel can also be used to move around the workflow.

2.2 Adding Ligands

No structures have yet been added to the Phase workflow, so there are no entries in the Ligands table. The structures that you copied from the distribution must now be added.

1. Click From File, which is located near the top of the panel.

   The Add From File file selector is displayed.

2. Select the file `at1.maegz`, and click Open.

   This file contains 50 ligands. These ligands have been prepared and conformers generated for each. If you are starting with 2D single structures, you would have to clean the structures and then generate conformers.

   When you click add, the Choose Activity Property dialog box opens. If you intend to build QSAR models (which you will do in this tutorial), this is where you should select the property to use as the experimental activity variable.

3. Choose All properties from the Subset option menu.
4. Select the pIC50-Exp property.
5. Ensure that Convert property values is *not* selected.
6. Click OK.

   The Choose Activity Property dialog box closes and the ligands are imported. This process takes a few minutes, and a progress dialog box is displayed during the import.

The Ligands table is now filled with *copies* of the structures from the file. The Ligands table contains an In column, which allows you to view one or more ligands in the Maestro Workspace. The Name column holds the same information as the Title column of the Project Table, and the Activity column contains the pIC50-Exp data.

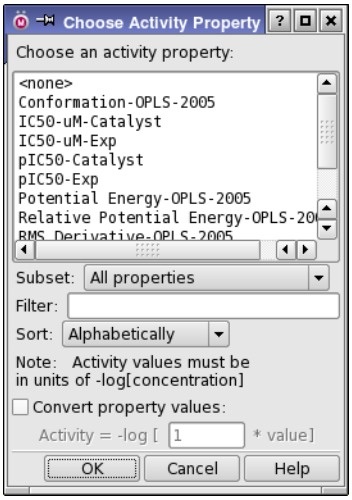

Figure 2.1. The Choose Activity Property dialog box.

The Pharm Set column indicates whether a molecule is in the set of actives used to identify common pharmacophore hypotheses, or in the set of inactives used to eliminate nondiscriminatory hypotheses, or in neither set. If ligands of widely varying activity are present, you would normally want to use only the most active ones in the set of actives. The most active ligands are assumed to contain the strongest binding, most important, or greatest number of pharmacophore features that are involved in binding to the protein target. The set of actives should contain as much structural diversity as possible, so that the resulting pharmacophore models are applicable across different chemical families.

The # Conformations column indicates how many conformations are present for each ligand. In this case, you imported multiple conformers for each ligand. When adding from the Project Table, you would normally see only a single conformation for each ligand unless you were selecting conformer sets.

The original set of 50 ligands contained a pair of duplicates, `test17` and `test24`. The latter is not in the Ligands table. These two ligands were found to be chemically identical and were therefore merged into a single entry when the conformers were generated. It turns out that these two structures are also identical in the publication from which they were taken, so you will proceed with a reduced data set of 49 ligands and the knowledge that the experimental activity value for `test17` might not be reliable.

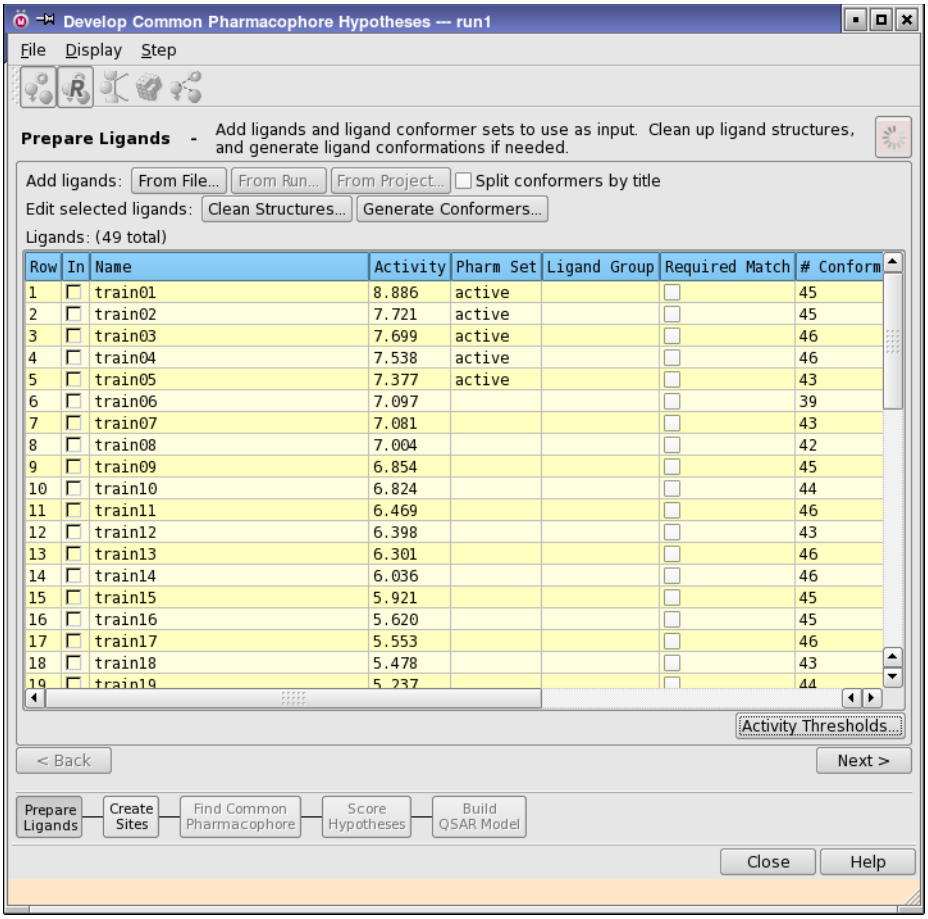

Figure 2.2. Ligands table in the Prepare Ligands step.

2.3 Choosing the Active and Inactive Sets

Only the most active compounds are normally considered when developing common pharmacophore hypotheses. Inactives can be used to eliminate hypotheses that do not provide a good explanation of activity on the basis of the pharmacophores alone. The active set determines the pool of pharmacophore models that are generated, and the initial scores that are assigned to them. The inactive set may be used subsequently to assign adjusted scores that reflect the degree to which the models distinguish actives from inactives. This is particularly useful if everything in the active set is built on a common scaffold, which can give rise to a number of spurious pharmacophore models that have nothing to do with ligand binding.

In this section, you will set a threshold for actives of $IC_{50} \leq 50$ nM, which translates to $pIC_{50} \geq 7.3$, and a threshold for inactives of $pIC_{50} \leq 5.0$.

1. Click Activity Thresholds.

   The Activity Thresholds dialog box appears, which allows you to specify the activity thresholds.

2. In the Active if activity above text box, type `7.3`.

3. In the Inactive if activity below text box, type `5.0`.

4. Click OK.

   The Pharm Set column now has active for each ligand whose pIC_{50} value is greater than 7.3, and inactive for each ligand whose pIC_{50} value is less than 5.0. The column is blank for ligands whose activity falls between these values.

5. Clear the Pharm Set column for any test set ligands (`test1` through `test25`) that are assigned to the active or the inactive set.

   You can do this by selecting the assigned ligands and control-clicking in the column for one of ligands until the selected rows are blank in this column. The values cycle through active, inactive and blank when you click.

   The test set ligands should not be part of the pharm set because the pharm set is used to develop the pharmacophore model, and the test set is used to validate the model.

   Five ligands should now be in the active set: `train01` through `train05`, and six ligands in the inactive set: `train20` through `train25`.

6. Click Next to proceed to the next step, Create Sites.

If you wish, you can do the optional exercises in Section 2.16 on page 31 at this point.

2.4 Examining and Modifying Feature Definitions

The pharmacophore features are mapped to the structures using a set of topological feature definitions, which you will examine here.

1. Place the first ligand (`train01`) in the Workspace by clicking the In check box in the first row of the Ligands table.

2. Click Edit Features.

   The Edit Features dialog box is displayed. This dialog box contains SMARTS patterns and other information that control the application of pharmacophore feature definitions.

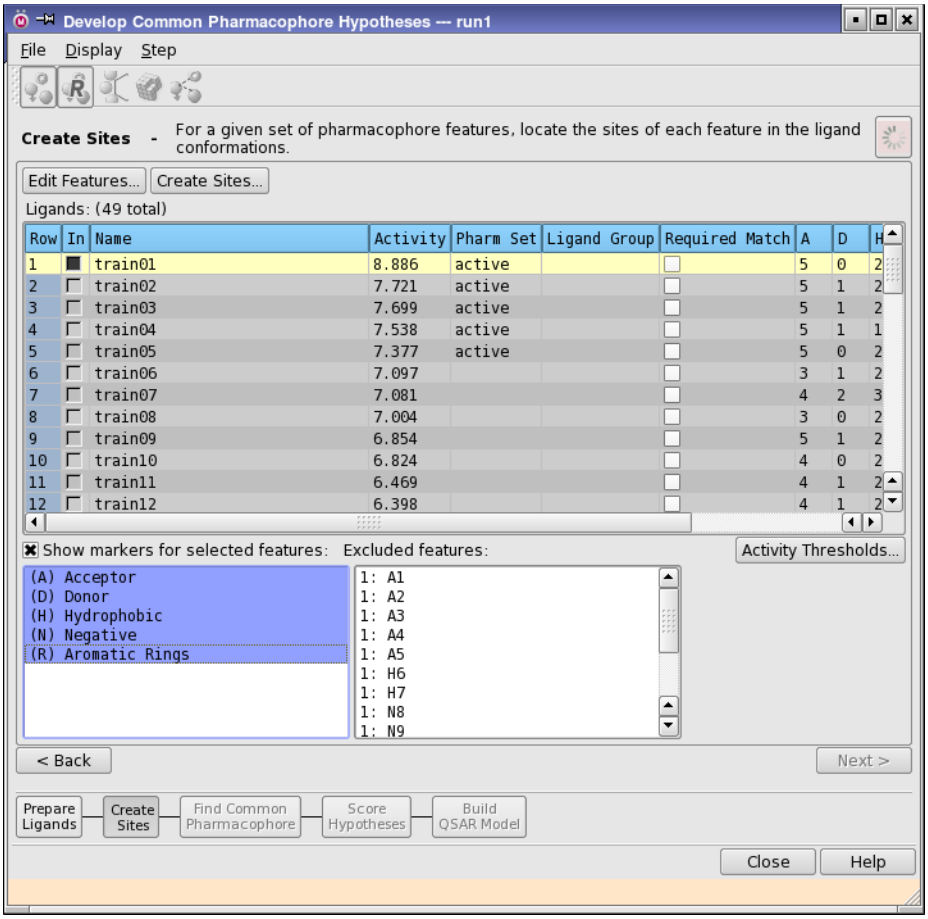

Figure 2.3. The Create Sites step.

By default, definitions for the Acceptor (A) feature type are displayed; other feature types may be viewed by selecting from the Feature option menu. These are the built-in definitions provided with every installation of Phase, and they cannot be modified or deleted. They can, however, be *ignored*, as you will see.

Definitions near the top of the table have higher precedence than those closer to the bottom of the table. So, for example, if the first pattern maps a particular nitrogen in the ligand as an acceptor, that same nitrogen will not be mapped as an acceptor by any subsequent pattern. The vertical positions of the built-in features may not be changed, but user-defined patterns (discussed in the next section) may be moved using the arrowhead buttons below the Pattern list table.

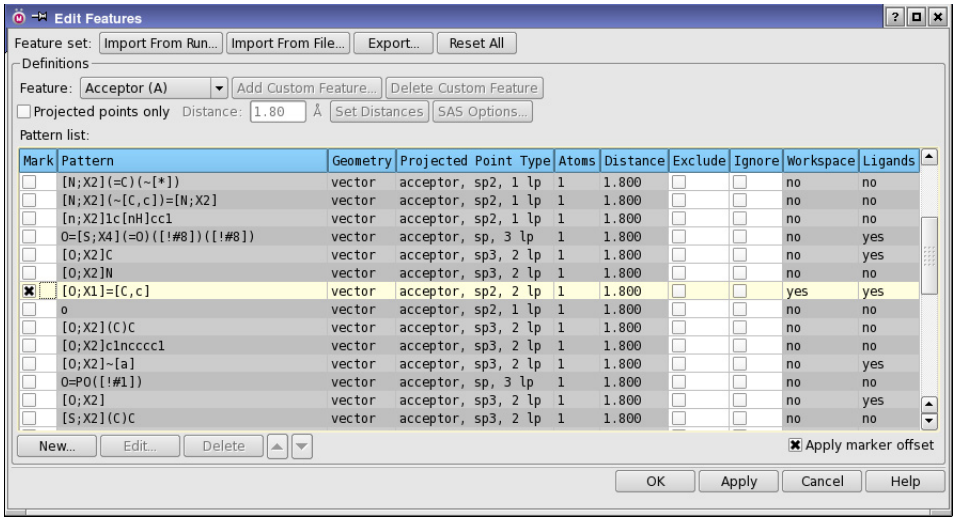

Figure 2.4. The Edit Features dialog box.

The order of precedence does not apply to *excluded* patterns, i.e., patterns for which a check mark appears in the Exclude column. Excluded patterns, which will be discussed shortly, are processed before all other patterns.

3. Click the Mark box next to the SMARTS pattern `[O;X1]=[C,c]`, which is about halfway down the Pattern list table.

   The atoms and bonds in the matching substructures are now marked in pink on the Workspace ligand. The red acceptor spheres, which should also be visible, clarify that the acceptor patterns are matched three times in the tetrazole ring and once in the imidazole ring.

   The Geometry for this feature definition is vector, meaning that the feature is located on a single atom and has one or more directions associated with it—the directions of the possible hydrogen bonds. Hydrogen bond acceptors and donors are vector features, and aromatic rings are also vector features because the orientation of the ring is important.

   However, there is no red sphere on the oxygen atoms of the carboxylate, so it is evidently not being perceived as an acceptor feature. This is because the built-in definitions assume that the user will *not* want to treat this as an acceptor due to the fact that the –COOH group is likely to be ionized. Accordingly, there is an *excluded* pattern further down in the table that matches this type of oxygen and excludes it from being flagged as an acceptor.

4. Clear the Mark box next to the SMARTS pattern `[O;X1]=[C,c]`.

   The C=O bond in `train01` is no longer marked.

5. Scroll down the table until you find the SMARTS pattern `O=C[O-,OH]`.

   The Exclude box is checked, which means this is an excluded pattern. Matches to this pattern are excluded from consideration for this feature type.

6. Click the Mark box next to this pattern.

   The O=C–O moiety is now marked on `train01`.

   Since the Geometry of this pattern is point (that is, the feature is located on a single atom), it excludes only features that are also located on a single atom, that is, features with point and vector geometries. There is no entry in the Atoms column, so the location of the feature defaults to the first atom in the pattern, which is the double bonded oxygen. So this exclude pattern matches all O=C–O moieties, and prevents the double bonded oxygen in -COOH from being flagged as an acceptor by any other pattern. As excluded patterns are processed first, the double bonded oxygen is flagged for exclusion before any of the regular acceptor patterns are applied.

   For the AT_1 data set, almost all of the ligands have an acceptor feature coming off the imidazole ring at the same position as the –COOH group in `train01`, indicating that there is probably an important hydrogen-bonding interaction. As `train01` is the most active ligand in the data set, and it differs from the second most active ligand (`train02`) only by replacement of a $–CH_2OH$ group with –COOH, it is likely that –COOH strengthens this hydrogen bonding interaction. Therefore, when looking for common pharmacophore hypotheses, at least one of the oxygens in this –COOH should be treated as an acceptor. Accordingly, you will *ignore* this exclude definition, so that it is not applied.

7. In the Ignore column, click the box for the `O=C[O-,OH]` pattern.

   The ignore operation is equivalent to removing the definition from the table, although you can always reinstate it by clearing the Ignore check box.

8. Click OK to update the definitions.

   A delay occurs as the new feature definitions are applied to the first conformation of each ligand. When this process is complete, a pink acceptor sphere appears on the oxygen in the C=O moiety. The excluded pattern definition `O=C[O-,OH]` is now being ignored, so that –COOH groups will be perceived as containing an acceptor feature.

You can add new SMARTS-based definitions to any of the feature types, and new features with their own definitions, as illustrated in the next two optional exercises. If you do not want to do these exercises, skip to Section 2.5 on page 13.

2.5 Creating Pharmacophore Sites

All of the operations done thus far applied the feature definitions only to the first conformation of each ligand. Before you can proceed to the next step in the workflow, you must apply the feature definitions to all conformations. Pharmacophore sites are created for all entries in the Ligands table, regardless of whether any rows are selected.

1. Click the Create Sites button near the top of the panel (not the one in the Guide).
2. Click Start in the Create Sites - Start dialog box.

   This job requires less than a minute on a 2 GHz Pentium 4 processor. Incorporation of results does not add any new information to the Ligands table, but the Find Common Pharmacophores button at the bottom of the panel is now active, and you can proceed to the next step in the workflow.
3. Click Next.

   The Find Common Pharmacophores step is displayed.

2.6 Changing the Number of Sites Allowed in the Common Pharmacophores

The possible common pharmacophores that Phase can identify depend on the number of features of each type that are available in the ligand set (listed in the Available column of the Feature frequencies table). They can also be restricted by setting the number of sites that a common pharmacophore may have, the number of ligands that must match, and the minimum and maximum number of each feature type permitted. The present set of five actives is not very diverse, so the value in the Must Match at least text box will be left at 5. The other parameters will be varied in this and the next exercise.

When the step is displayed, a list of 19 *variants* appears on the panel, reflecting the 19 possible combinations of features that could give rise to common pharmacophores.

This exercise examines how the features frequencies and variants change with the number of sites in the common pharmacophores, and sets the number that will be used in the search.

1. Choose 4 from the Maximum number of sites option menu.

   The only change in the Feature frequencies table is that the Maximum allowed frequency of acceptors has been reduced from 5 to 4. This merely reflects the fact that the common pharmacophores we are looking for cannot contain more than 4 sites. The number of variants has dropped from 19 to 16, and of course there are only 4 features in each variant.

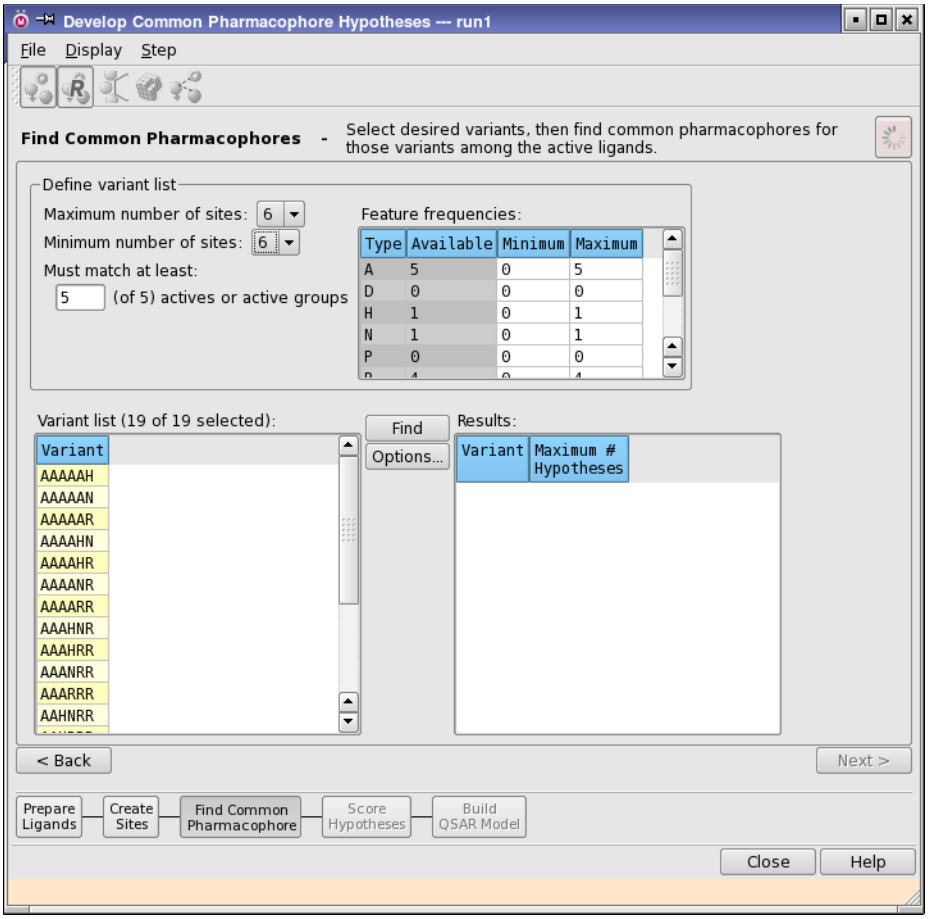

Figure 2.5. The Find Common Pharmacophores step.

2. Change the maximum number of sites to 6.

   The Feature frequencies table is identical to that observed when the maximum number of sites was 5, and there are once again 19 variants.

3. Change the maximum number of sites to 7.

   The Feature frequencies table is unchanged, but the number of variants drops to 16. Combinatorics dictates that there will be a peak in the number of variants as the number of sites is varied, and we see that the peak occurs at 5 and 6 sites. You will be searching for common 6-point pharmacophores.

4. Change the maximum and minimum number of sites to 6.

2.7 Changing the Allowed Feature Frequencies

The Available column of the Feature frequencies table a indicates that we could search for common pharmacophores with up to five acceptors. However, it is very unlikely that a ligand would bind to the receptor through five, or even four, hydrogen-bond acceptor interactions. Here, some chemical intuition is needed to filter out results that are scientifically unsound or unlikely. For example, we may decide that common pharmacophores should contain at least one, but no more than three acceptors (A).

The Minimum and Maximum columns in the Feature frequencies table contain user-settable limits on the number of times a feature is *allowed* to appear in a common pharmacophore. You will use these columns to set limits on the feature frequencies in this exercise.

1. In the Feature frequencies table, change the minimum and maximum number of acceptors (A) to 1 and 3, respectively.

   The number of variants drops from 19 to 11 and each variant contains between 1 and 3 acceptor features.

We know that each active ligand contains a tetrazole ring and therefore at least one negative feature. The tetrazole has no doubt been put there for a good reason, so it is pretty safe to assume that the hypothesis should contain a negative feature.

2. In the Feature frequencies table, change the minimum number of negative features (N) from 0 to 1.

   The number of variants drops from 11 to 6, and each variant contains a negative feature.

The Feature frequencies table also indicates that there could be as many as 4 aromatic rings. However, one of these rings is the tetrazole, which we presume will be acting as a negative feature, so it is not necessary to consider so many aromatic rings.

3. Change the maximum number of aromatic rings (R) from 4 to 3.

   The number of variants drops from 6 to 5, and no variant contains more than 3 aromatic rings.

2.8 Finding Common Pharmacophores

A total of five variants are examined to identify 6-point pharmacophores that are common to all five of the active ligands. The search is done by a succession of binary partitions of each intersite distance; each pharmacophore is placed in a “box” defined by the particular partitioning of each intersite distance for the pharmacophore. The default options for the search are used. For more information on the available options, see Section 5.3 of the *Phase User Manual*.

1. Ensure that all five variants are selected, then click Find.
2. Click Start in the dialog box.

This job requires about two minutes on a 2 GHz Pentium 4 processor. When the job finishes, the Results table shows that only two of the five variants yielded common pharmacophores. Maximum # Hypotheses is the number of boxes that provided distinct common pharmacophores for each variant. A given box contains one or more pharmacophores from each of the minimum required number of active ligands. Exactly one pharmacophore from each box is selected as a potential hypothesis, and this selection takes place in the Score Hypotheses step.

2.9 Scoring Hypotheses

In the scoring step, a scoring procedure is applied to identify the pharmacophore from each box that yields the best alignment of the active ligands. The scoring procedure also provides a ranking of the different hypotheses, allowing you to make informed choices about which hypotheses are most appropriate for further investigation.

In this exercise, you will examine settings that control how hypotheses are selected from each surviving box and how they are ranked with respect to one another, and then run scoring jobs to examine the effect of including various terms in the score.

1. Click Next to enter the Score Hypotheses step.

   The Score Hypotheses step is displayed. The Hypotheses table is empty because scoring has not yet been done.

2. Click Score Actives.
3. The Score Actives dialog box is displayed. For background information on the scoring process, see Section 6.1 of the *Phase User Manual*.

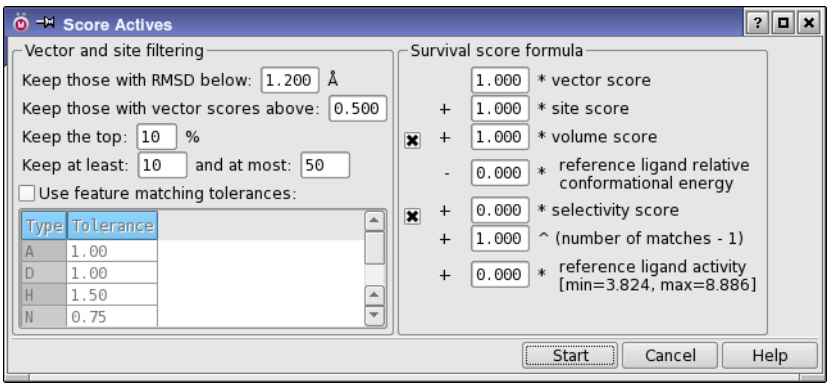

Figure 2.6. The Score Actives dialog box.

Several of the scores in the survival score formula have zero weights. By default, none of these optional scores contribute to the ranking of the hypotheses. However, all individual components of the score are reported, so you can always consider selectivity, if you so desire, when making choices among hypotheses that have very similar overall scores.

4. Check that the weights of the selectivity score and the number of matches are 0.0 and 1.0.
5. Click Start, and click Start again in the Start dialog box.

   This job requires less than one minute on a 2 GHz Pentium 4 processor. Once the job is incorporated, the Hypotheses table contains information about the highest scoring hypotheses from each variant.

   Note that the range of scores among the reported hypotheses is only 0.5, so none of the hypotheses is particularly poor.

Next, you will check whether any of these hypotheses can be eliminated based on its match to the inactives.

6. Click Score Inactives.

   The Score Inactives dialog box opens. Leave the inactives weight at its default value of 1.0.
7. Click Start, and click Start again in the Start dialog box.

   The job takes less than a minute. When it finishes, the Survival-inactive column of the Hypotheses table is populated with values.

   The survival score is reduced fairly evenly for all the hypotheses. A reduction of one unit is expected, because the scoring function for the inactives is the same as for the actives, and it includes a contribution of 1.0 for the number of matches. The remainder of the reduction is fairly uniform and fairly small, indicating that the inactives do not match the hypotheses well enough to eliminate any of them. On this basis, we can be fairly confident that the hypotheses are reasonable explanations of the activity.

Finally, you will adjust the scoring to take account of the activity. Of the five actives used for model development, one has a significantly higher activity than the others. Adding an activity reward could change the ranking of the hypotheses that have this ligand as the reference.

8. Click Rescore.

   The Rescore Hypotheses dialog box opens.
9. Enter `0.3` in the reference ligand activity text box.

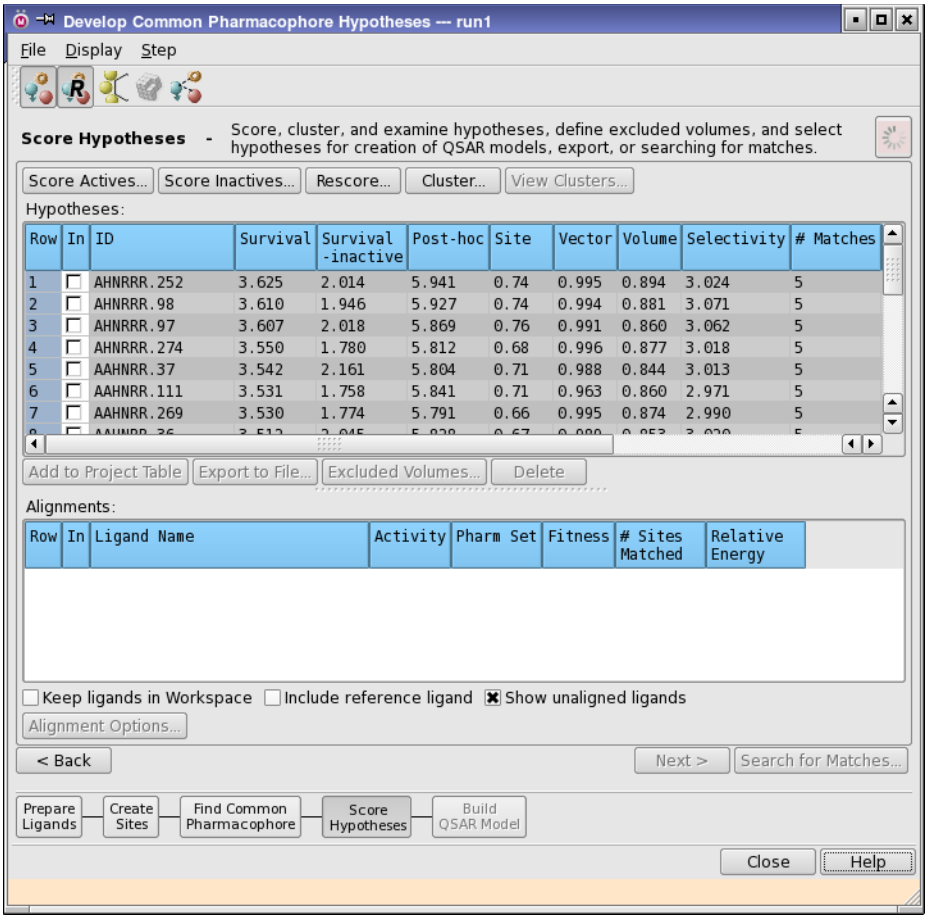

Figure 2.7. The Score Hypotheses step after scoring.

10. Click OK.

    The results are returned almost immediately, in the Post-hoc column. Scoring the actives and the inactives requires alignment of the ligands, which takes a little time. For rescoring the alignment is already done.

11. Click twice on the Post-hoc column heading, to sort the hypotheses in descending order by this score.

    Now, two other hypotheses are at the top, followed by the set of hypotheses with the top survival scores. In the next exercise, you will examine some of these hypotheses and the ligand alignments.

2.10 Viewing Hypotheses and Ligand Alignments

In this exercise you will examine the nature of the top-scoring hypotheses and the quality of the associated ligand alignments.

1. Sort the Hypotheses table by survival score, by clicking twice on the Survival column heading.

   The top scoring hypotheses come from an `AHNRRR` variant, `AHNRRR.252` and `AHNRRR.98`. (The numeric suffix is merely the index of the box from which the hypothesis came.) The scores of these two hypotheses are very close, due to the fact that both hypotheses align the active ligands in very similar ways.

2. Click the In column for `AHNRRR.98` in the Hypotheses table.

   The features of the hypothesis are displayed in the Workspace. If the hypothesis is *not* visible in the Workspace, click the Display Hypothesis toolbar button at the top of the Develop Pharmacophore Model panel.

   

   The Alignments table is filled with a record for each ligand. The records for the ligands not used for the model (the *non-model* ligands) are dark gray, indicating that no alignment was done for these ligands. The records provide information on the conformation whose pharmacophores yielded the best multi-ligand alignment to the hypothesis when it was selected as the reference from its box. Fitness measures the quality of each alignment using a weighted sum of alignment and volume scores, just as in the total scoring function. Note that the second row of the Alignments table is blue, indicating that `train02` is the reference ligand, i.e., the ligand from which the hypothesis came. Its fitness score is a perfect 3.0—the maximum possible score with the scoring options that were selected—because it corresponds to the alignment of `train02` onto itself.

3. Place the reference ligand in the workspace by clicking its In box in the Alignments table.

   `train02` is overlaid onto the hypothesis. The features of the hypothesis are perfectly positioned on the matching sites of this ligand, because the hypothesis comes from the reference ligand, so its features must coincide with those of the reference ligand.

4. Include `train03` in the Workspace by control-clicking its In box.

   `train03` is overlaid onto the hypothesis by aligning its 6-point pharmacophore with the six features in the hypothesis. Observe that the `train03` and `train02` ligands are only slightly out of alignment with each other.

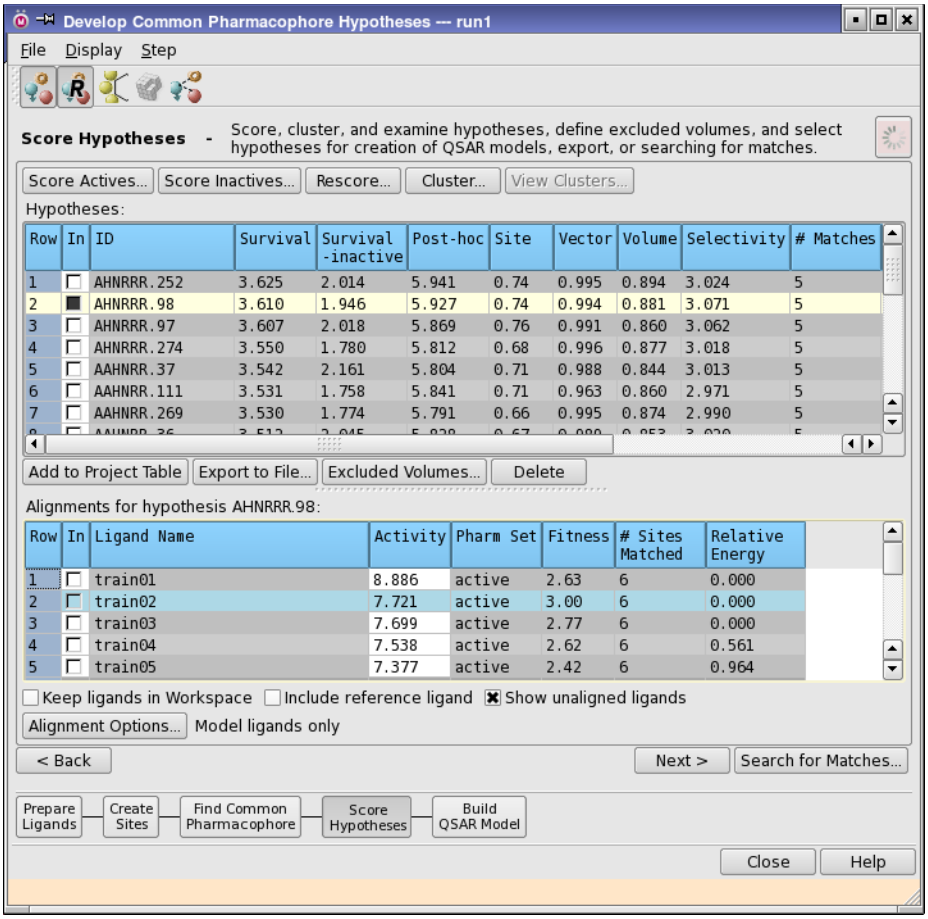

Figure 2.8. The Score Hypotheses step with alignments for AHNRRR.98.

5. Include the remaining actives in the Workspace.

   Most ligands superimpose on each other with only minimal offset. This should come as no surprise because most of the Fitness scores are close to 3.0. In fact, ligands `train02`, `train03`, and `train04` only differ by the substitution of Cl or I for H. The exception is `train01`, whose fitness score is about 2.6. If you place this ligand in the Workspace with the reference ligand, you can see the deviation in the alignment.

6. Examine the hypothesis `AHNRRR.252` by selecting its row in the Hypotheses table.

   The Alignments table is updated to reflect the best alignments for `AHNRRR.252`. For this hypothesis, `train02` is also the reference ligand. None of the values in the Relative Energy column are zero, indicating that there is an energy cost for this hypothesis.

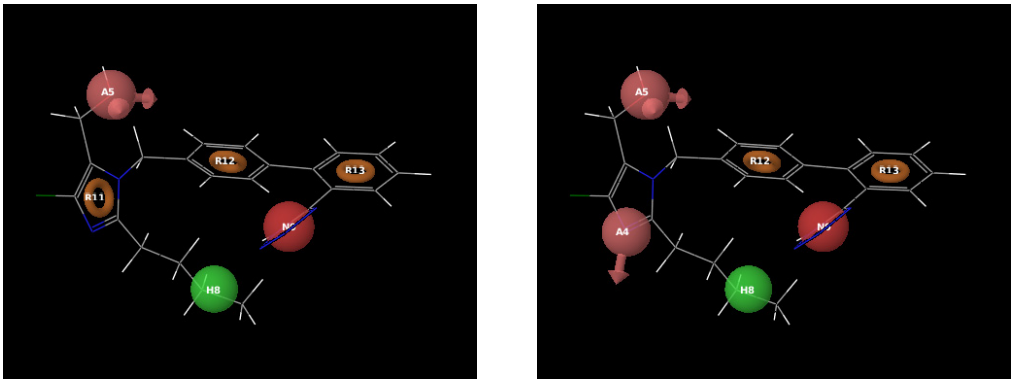

Figure 2.9. Hypotheses AHNRRR.98 (L) and AAHNRR.36 (R) with reference ligand train02.

7. Examine the hypothesis `AAHNRR.36` by selecting its row in the Hypotheses table.

   As for `AHNRRR.98`, `train02` is the reference ligand. The side-by-side view of `train02` aligned to `AHNRRR.98` and `AAHNRR.36` is shown in Figure 2.9.

   It should be evident that the same reference ligand conformation was selected in both cases, and that the two hypotheses differ only in whether the imidazole acts as an aromatic feature or as an acceptor. In situations like this, it is difficult to assess which hypothesis is most reasonable, so one often pursues both for further investigation. For this tutorial, we shall examine both `AHNRRR.98` and `AAHNRR.36`.

In the next exercise you will build a 3D QSAR model, which requires alignments for ligands outside the active set. The 49 training and test set ligands cannot all be expected to match all six sites in the hypothesis. Some of the weak binders will be missing one or more features contained in the hypothesis. To deal with this possibility, Phase uses *partial matching* to obtain alignments for these ligands. If at least three sites in the hypothesis can be matched, an unambiguous alignment is obtained. However, it is not always possible to obtain an alignment for every ligand. For each ligand outside the active set, then, Phase searches for matches involving the largest possible number of sites, and identifies the match that yields the highest fitness score. Alignments can give information about which features are important and which are not, especially for actives that are not in the training set.

8. Click Alignment Options below the Alignments table.

   The Alignment Options dialog box opens. In addition to selecting the option to align the "non-model" ligands (those that were not in the pharm set or did not match all sites in the hypothesis), you can require certain features to match when the alignment is performed.

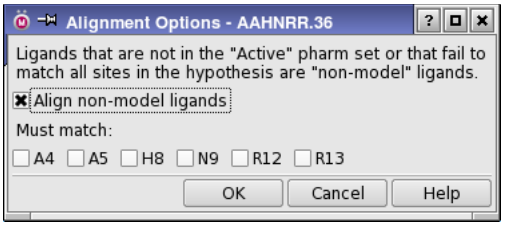

Figure 2.10. The Alignment Options dialog box.

If the features do not match for a particular ligand, the ligand is not aligned. For this tutorial we will not require any matches.

9. Select Align non-model ligands, and click OK.

   After a short delay, the Alignments table contains an entry for all ligands. Scrolling down the table, you will find matches involving 4 to 6 sites and a wide range of fitness values.

Now you will examine the alignments for several ligands: `train12`, which matches all 6 points but has a poor fitness score, `train25`, which matches 4 points and is inactive, and `test02`, which only matches 5 points but is highly active.

10. Place `train12` in the Workspace.

    The features in `train12` do not align well with the corresponding features of the hypothesis. This does not necessarily mean that `train12` could never achieve better alignment, but rather that the conformations we are working with contain no such alignment. However, an examination of the structure shows that a better alignment is unlikely, and there is some internal strain in achieving the current alignment: the energy of this conformer relative to the lowest is more than 3 kcal/mol. When building a QSAR model, it is a good idea to take into account the quality of the superimpositions, especially for members of the training set. Cleaner superimpositions usually yield models with greater statistical significance and greater predictive ability.

11. Place `train25` in the Workspace.

    It is immediately obvious that, although the four features that do match align well, there are two features missing. It is reasonable to suppose that at least one of these features is critical to activity.

12. Place `test02` in the Workspace.

    This is a highly active compound in the test set. It only matches on five features, implying that the sixth feature is not really necessary for activity.

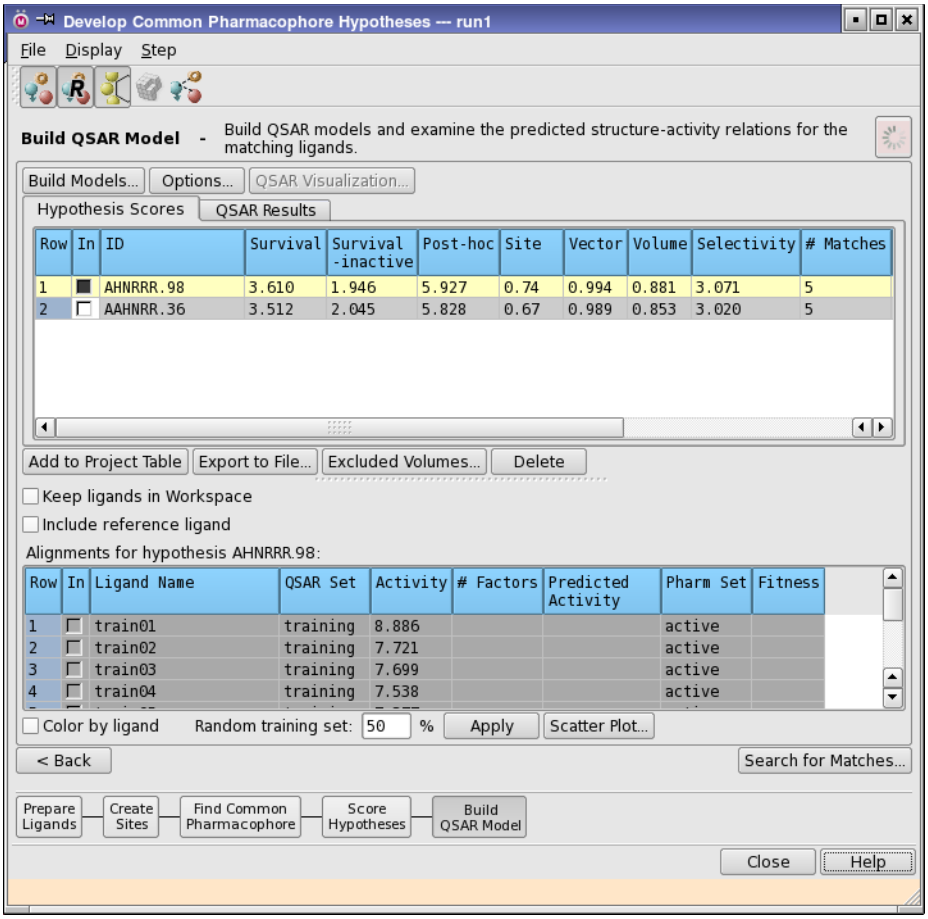

Figure 2.11. The initial view of the Build QSAR Model step.

2.11 Proceeding to Build QSAR Model

Phase QSAR models are based on partial least-squares (PLS) regression, applied to a large set of binary-valued variables that encode whether or not ligand atoms or ligand features occupy various cube-shaped elements of space. Using the hypotheses `AHNRRR.98` and `AAHNRR.36`, you will develop QSAR models to explain the activity data (`pIC50-Exp`), then apply these models to make activity predictions for the test set ligands. Although these hypotheses are the top-scoring hypotheses, there is no necessary connection between the score and the quality of the QSAR model. When you build QSAR models, you should try several hypotheses to ensure that you have a good model. You can build models for multiple hypotheses simultaneously.

1. Select AHNRRR.98 and AAHNRR.36 in the Hypotheses table.
2. Click Next.

   The Build QSAR Model step is displayed.

2.12 Assigning Training and Test Set Memberships

By default, all ligands are placed in the training set, so you must separate them into the proper training and test sets. You need only do this for one hypothesis, because the set membership is the same for all hypotheses in this step.

1. In the Hypothesis Scores table, ensure that AHNRRR.98 is included in the Workspace (the In column button is selected).

   The hypothesis is displayed in the Workspace, and the Alignments table is populated with data for this hypothesis.
2. Select all the test set ligands (test01, test02, etc.) in the Alignments table.

   You can do this by clicking test01 and shift-clicking test25. Make sure you do not click in the QSAR Set column. If the ligands are not sorted by name, click the Ligand Name column heading to sort them.
3. Control-click the QSAR Set column for one of the selected ligands, to change the value from training to test.

This results in a training set of 25 ligands and a test set of 24 ligands. If you wanted to eliminate a particular training set (or test set) ligand from consideration, possibly because of a poor superimposition, you could click the corresponding QSAR Set cell until its contents were blank.

Note that the hypothesis is used only to obtain ligand alignments: it does not contribute in any way to the QSAR model itself. But the hypothesis has an *association* with the model, because it defines how ligands should be pre-aligned before applying the model.

2.13 Setting QSAR Model Options

Phase can generate QSAR models that are atom-based or pharmacophore-based. The independent variables in the QSAR model are derived from a regular grid of cubic volume elements that span the space occupied by the training set ligands. Each ligand is represented by a set of bit values (0 or 1) that indicate which volume elements are occupied by either a van der Waals model of the atoms in that ligand or by pharmacophore features of that ligand. In the atom-based model, to distinguish different atom types that occupy the same regions of space, a given

cube in the grid may be allocated as many as six bits, accounting for six different classes of atoms. Likewise, the pharmacophore features in the hypotheses are represented by bits for each feature type.

In this exercise, you will be developing an atom-based model, and will set parameters to control the sizes of the cubes and the maximum number of PLS factors to include in the model.

1. Click Options.

   The Build QSAR Model options dialog box is displayed. Since we have selected the training set and the test set, there is no need to change settings in the Training set section.

2. Set Grid spacing to 1.0.

3. Set Maximum PLS Factors to 3.

4. Under Model type, ensure that Atom-based is selected.

   Note that for the pharmacophore-based models, you can adjust the feature radii.

5. Click OK.

   The cubes that define the independent variables will be 1 Å on each side, and atom-based linear regression models will be built containing one, two and three PLS factors.

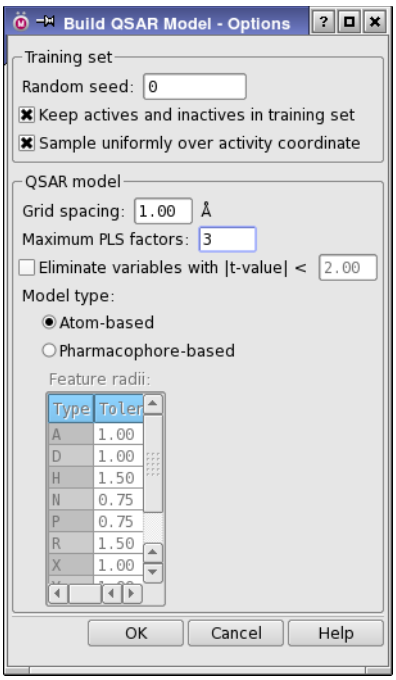

Figure 2.12. The Build QSAR Model - Options dialog box.

2.14 Building the QSAR Models

In this exercise, one-factor, two-factor and three-factor PLS regression models are created using the training set ligands and then applied to the test set ligands.

1. Click Build Models.

   The Start dialog box opens.

2. Select a host, and click Start.

   The job status icon at the top right turns green and spins. When it stops, data for three PLS regression models fills the QSAR results table. Each regression is a QSAR model. The columns in the table are described in Table 7.2, in Section 7.4 of the *Phase User*

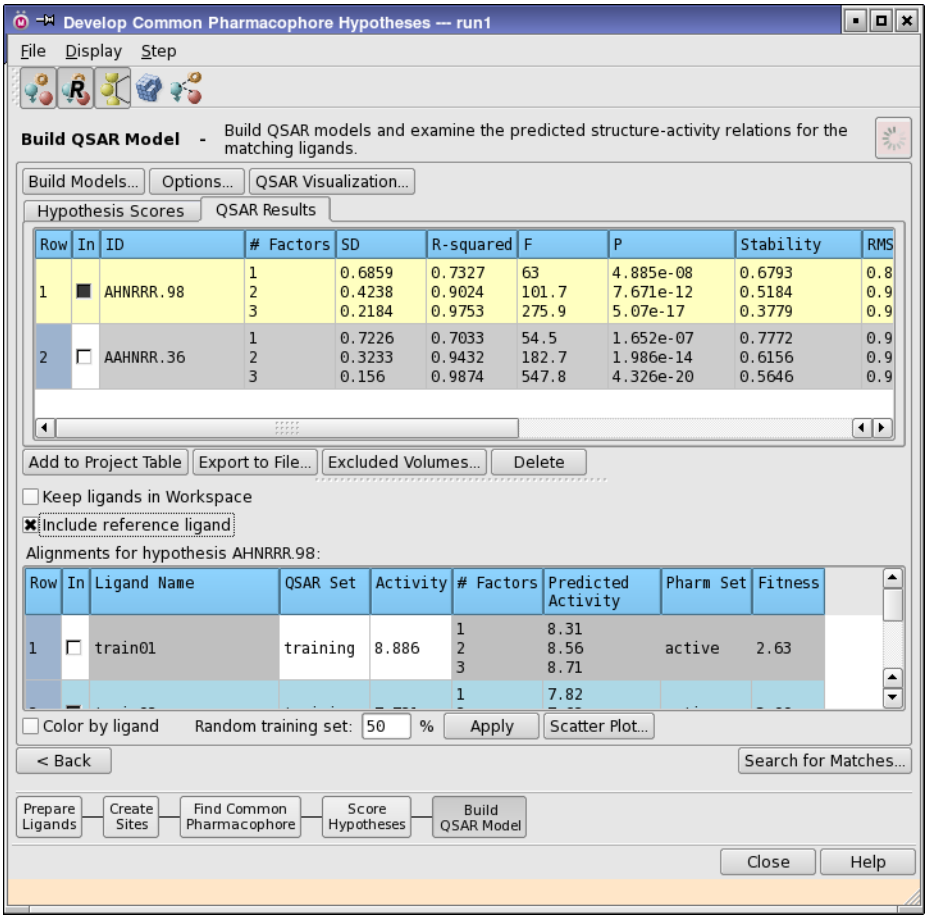

Figure 2.13. The Build QSAR Model step after building models.

Manual. The R-squared value always increases as the number of PLS factors increases, but the same is not necessarily true of Q-squared.

3. In the QSAR Results table, click the In column for `AAHNRR.36`.

   The hypothesis is displayed in the Workspace, and the Alignments table is populated with data for this hypothesis, including activity predictions for each ligand using one, two and three PLS factors.

   Scrolling down to the test set, we see that some predictions are good and some are bad. As might be expected, the prediction for `test02`, the active ligand that matched only five sites in the hypothesis, is significantly lower than experiment.

   Of particular interest is `test17`, whose experimental activity value may be in question since its structure was identical to that of `test24`, which was eliminated when the conformer sets were generated for this tutorial. The 2-factor prediction for this ligand is 4.37, while the experimental activities originally reported for test17 and test24 were 5.770 and 4.553, respectively. The first experimental value, 5.770, is the value that was used, so the error in the prediction is quite large. To see how much this observation is skewing the statistics for the test set, you can remove it and rebuild the model, and you can change the activity to use the smaller value. First, note the Q-squared values in the QSAR Results tab for this hypothesis.

4. In the Alignments table, click the QSAR Set cell for `test17` to change its contents from `test` to an empty cell.

   A dialog box appears warning you that you will invalidate data in forward steps.

5. Choose the Proceed option and click OK.

   The ligand `test17` is no longer a member of the test set.

6. Click Build Models, and click Start in the Start dialog box.

   The new models are identical to the old models because the training set is the same. However, the Q-squared value for the test set has increased because the poor prediction for `test17` is not being considered.

7. In the Alignments table, click the QSAR Set cell for `test17` twice to change its contents back to `test`.

   A dialog box appears warning you that you will invalidate data in forward steps.

8. Choose the Proceed option and click OK.

9. Edit the Activity cell for `test17` to change it to 4.553, the lower of the two reported values.

10. Click Build Models, and click Start in the Start dialog box.

    Again, the new models are identical to the old models because the training set is the same. However, the Q-squared value for the test set has increased significantly because the activity for `test17` is lower, in line with the QSAR model predictions. In this case the model gives us some grounds for accepting the lower of the two experimental values as the correct value.

2.15 Visualizing the QSAR Model

Three-dimensional aspects of the QSAR model are examined to help gain an understanding of how the structures of the ligands contribute to the computed activity.

1. Ensure that the View QSAR model toolbar button is selected.

   

2. Place `train01`, the most active ligand, in the Workspace.
3. Click QSAR Visualization.

   The QSAR Visualization Settings panel is displayed. This panel has various options for displaying characteristics of the QSAR model.

   You can view the cubic volume elements occupied by the ligands in the Workspace, or view all the cubes in the QSAR model (i.e., the union of cubes occupied by the 25 training set ligands). You can also view the cubes associated with all atom classes or a specific atom class. The number of PLS factors determines which of the three regression models you are viewing, and the positive and negative coefficient thresholds allow you to see only the cubes whose PLS regression coefficients exceed a particular tolerance.

4. Ensure that all items in the View effects from list are selected, and that Combine effects is selected
5. Ensure that Workspace ligands is selected under View volume occupied by.
6. Select 2 PLS factors.
7. Move the positive and negative coefficient threshold sliders to an intermediate value, such as 0.015 and –0.015.

   In the Workspace, you will see many blue cubes and a small number of red cubes. The blue cubes indicate regions that are favorable for activity and the red cubes indicate regions that are unfavorable. Note that you are viewing only those cubes whose regression coefficients exceed the intermediate thresholds we set.

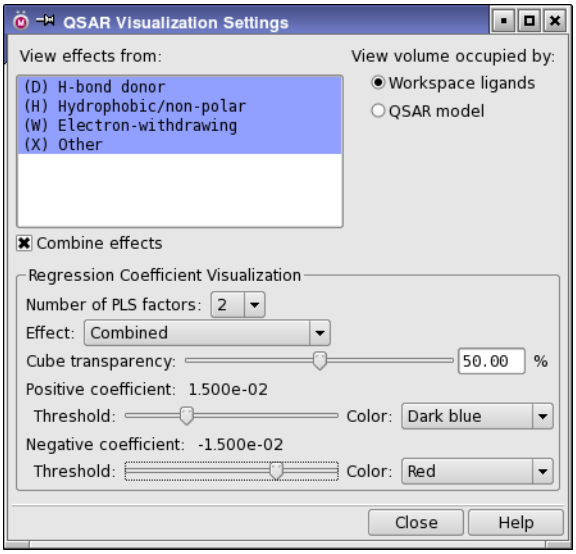

Figure 2.14. The QSAR Visualization Settings panel.

8. Change the positive and negative coefficient thresholds to more extreme values, such as 0.03 and –0.03.

   The number of cubes in the Workspace drops significantly because now you are viewing only very significant terms in the model.

9. Change the thresholds back to 0.015 and –0.015.

10. Replace `train01` in the Workspace with `train25`, the least active ligand.

    The Workspace now contains many more red cubes than blue cubes, indicating a preponderance of unfavorable interactions.

11. Add `train01` back into the Workspace, while keeping `train25` there as well (control-click the In column for `train01`).

    You are now viewing the union of the cubes occupied by the two ligands. It should be evident that `train25` fails to occupy much of the favorable blue volume of `train01`. It should also be evident that this volume is associated primarily with the aromatic and negative features that are not matched by `train25`. So while the hypothesis does not contribute in an explicit sense to the QSAR model, it is reflected implicitly in the regression coefficients and volume occupation patterns. Note that a Phase QSAR model may involve regions of space that extend beyond the physical bounds of the hypothesis because the QSAR considers the volume occupied by all atoms in the ligands.

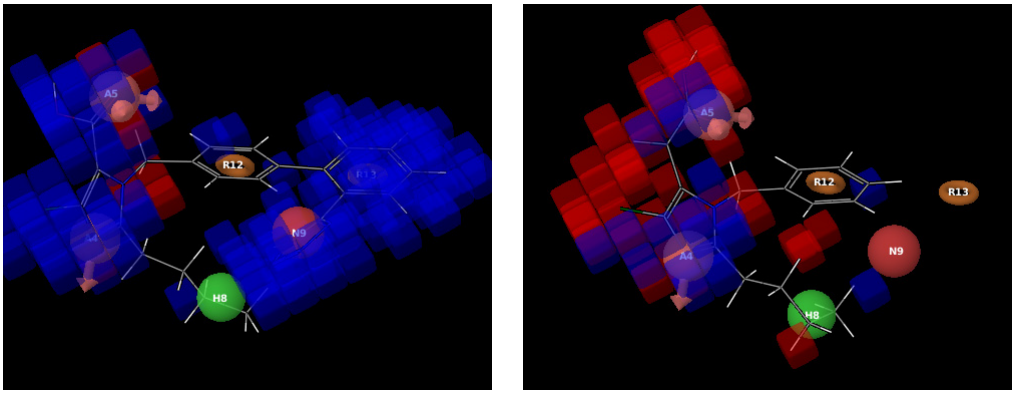

Figure 2.15. QSAR model for ligands train01 and train25.

12. Remove `train25` from the Workspace (click `train01`).

The current view of the model illustrates effects from all atom classes simultaneously, but it is possible to separate out the contributions of individual atom classes.

13. Under View effects from, deselect Combine effects, and select (D) H-bond donor.

    This category includes polar hydrogens bonded to nitrogen, oxygen and sulfur.

14. Choose (D) H-bond donor from the Effect option menu.

15. Change the coefficient thresholds to lower values, such as 0.004 and –0.004.

    The view in the workspace indicates that the tetrazole and carboxylic acid groups of `train01` occupy D-type volume, and that this occupation contributes favorably to activity. While we know that both of these groups are likely to be ionized, the QSAR model takes each ligand structure *as is* and assigns each atom to a particular class. So these groups would give rise to negative ionic atoms (type "N") only if the ligand structures were explicitly ionized. As it is, there are no type N features in the list.

16. Choose (H) Hydrophobic/non-polar in the View effects from list and the Effect option menu.

17. Change the positive and negative coefficient thresholds to 0.015 and –0.015.

    Green cubes are now visible throughout much of the structure of `train01`. These are favorable regions occupied by carbons, halogens, and nonpolar hydrogens. Note that the QSAR model does not distinguish between aromatic and nonaromatic carbons.

18. Add `train25` back into the Workspace.

    Many purple cubes appear, indicating that `train25` occupies a fair amount of unfavorable volume of type H.

19. Choose (W) Electron-withdrawing in the View effects from list and the Effect option menu.

20. Change the positive and negative coefficient thresholds to lower values, such as 0.005 and –0.005.

    The Workspace indicates favorable and unfavorable regions occupied by electron-withdrawing nitrogen and oxygen atoms. W-type volume includes hydrogen bond acceptor atoms, but it does not distinguish cases where a lone pair may not be available due to conjugation, e.g., amide nitrogens.

Experiment with different atom classes and different thresholds to view the various ways in which the QSAR model distinguishes ligands with high and low activities.

2.16 Additional Exercises

The exercises in this section are optional, but illustrate features that are not included in the main workflow.

2.16.1 Examining Feature Mappings

To fully understand how pharmacophore models are developed, it is important to understand how the various pharmacophore features are mapped onto the ligand structures. The Ligands table contain columns (A, D, N, etc.) showing the counts of each type of pharmacophore feature in each ligand. Below the table is a list of feature types that can be displayed.

Note: This exercise can be performed after the exercise in Section 2.3 on page 8.

1. If the Workspace is not empty, click the Clear button on the Workspace toolbar.

   

2. Select (A) Acceptor in the features list under Show markers for selected features.

3. Place the third ligand (`train03`) in the Workspace by clicking the In check box in the third row of the Ligands table.

   You should see various red transparent spheres with arrows protruding from them. The spheres are centered on the hydrogen bond acceptor features of the ligand, and the arrows indicate the axes along which ideal hydrogen bonds would be formed.

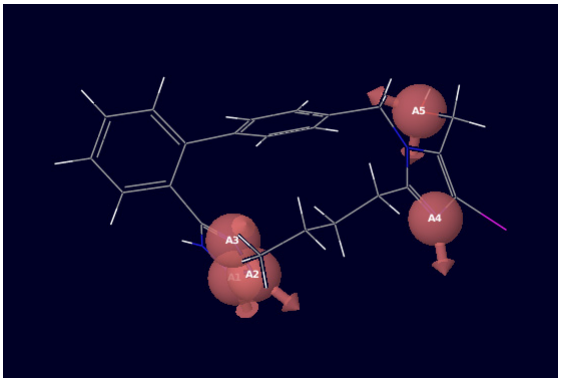

Figure 2.16. Acceptor features for ligand train03.

4. Select (D) Donor in the features list to display hydrogen-bond donors.

   A blue sphere appears on the hydroxyl hydrogen of this ligand. Once again, the arrow indicates the direction of the ideal hydrogen bond.

5. Select (H) Hydrophobic in the features list to display hydrophobic features.

   A green sphere appears in the middle of the n-butyl chain attached to the imidazole ring, and another appears on the iodine atom. Since there is no directionality to the hydrophobic feature, it is represented as a sphere without an arrow.

6. Select (N) Negative in the features list to display negatively-charged features.

   A red sphere appears in the center of the tetrazole ring. While there is no ionic charge in this ligand structure, tetrazole is known to be acidic, and so it is perceived as a negative ionizable feature, much like a carboxylic acid.

7. Select (R) Aromatic Rings in the features list.

   Orange toroids illustrate the locations of aromatic rings.

8. Select (A) Acceptor in the features list again.

2.16.2 Adding New Feature Definitions

The built-in feature definitions are reasonable, but there may be times when you need to add your own definitions to identify features that aren't accounted for by the built-in set. In this exercise you will define a new type of acceptor feature. The feature you will add is an aromatic ring, which can function as a weak acceptor.

Note: This exercise can be performed after the exercise in Section 2.4 on page 9.

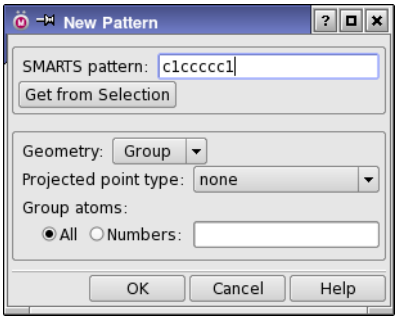

Figure 2.17. The New Pattern dialog box.

1. In the Create Sites step, select (A) Acceptor in the features list.

   The acceptor features are displayed in the Workspace.

2. Click Edit Features.

3. Select Acceptor (A) from the Feature option menu.

4. Click the first row in the Pattern list table.

5. Click New.

   The New Pattern dialog box is displayed.

6. In the SMARTS pattern text box, type `c1ccccc1`.

7. Choose Group from the Geometry option menu.

   The geometry in this case is defined by a group of atoms that together contribute an acceptor rather than a single atom.

8. Choose aromatic ring from the Projected point type option menu.

   Most acceptors have a directionality that can be precisely defined by considering the orientation of the acceptor lone pairs with respect to some plane in which the atom lies. In this case the direction is perpendicular to the ring.

9. Click OK.

   A new acceptor pattern appears at the top of the Pattern list table. In this position in the table, this pattern is matched before all others. You could move this new pattern vertically in the table using the arrowhead buttons, but doing so ultimately has no effect on the feature perception since there is no other similar functional group.

10. Click the Mark box for this new definition.

    The two occurrences of aromatic rings are marked in the Workspace.

11. Click Apply at the bottom of the panel to update the definitions.

    After a short delay, a pink sphere appears on each of the rings, indicating that they will be treated as an acceptor. This was for demonstration purposes only, so you should remove the definition before finishing this exercise.

12. Select the newly created pattern in the Pattern list table.

13. Click the Delete button below the table to remove the definition permanently.

14. Click Apply at the bottom of the panel.

    The rings are no longer perceived as acceptors. You could have accomplished the same thing by simply *ignoring* the new definition, without permanently deleting it.

15. Click Cancel to close the panel.

2.16.3 Adding New Feature Types

It is possible to create an entirely new type of feature (X, Y or Z) beyond the set of built-in feature types (A, D, H, N, P, R). In this exercise, you will create a feature that identifies aromatic rings as weak H-bond acceptors. In practice, you might want to create feature types for weak acceptors and strong acceptors so that the two can be separated in a pharmacophore model. Here, to define the SMARTS pattern, you will use Workspace selection.

Note: This exercise can be performed after the exercise in Section 2.4 on page 9.

1. Click Edit Features.

2. In the Edit Features panel, choose Custom (X) from the Feature option menu.

   The Pattern list table displays three default features that are not defined in terms of SMARTS patterns. The Ignore column is checked for each of these features, so they will not be used unless you clear the checks this column.

3. In the Workspace, select the six carbon atoms in an aromatic ring.

   You can do this using shift-click or by dragging over the atoms.

4. Click the New button below the Pattern list table.

   The New Pattern dialog box is displayed.

5. Click Get From Selection.

   A SMARTS pattern for the selected atoms appears in the SMARTS pattern text box.

6. Ensure that Geometry is Group, Projected point type is aromatic ring and Group atoms is All.
7. Click OK.

   The new pattern appears in the Pattern list table.
8. Check the Mark box next to the new pattern.

   The aromatic rings are marked in the Workspace.
9. Click Apply at the bottom of the Edit Features dialog box.
10. In the Develop Pharmacophore Model panel, ensure that (X) Custom is selected in the list under Show markers for selected features to display occurrences of this new feature.

    After a brief delay, a turquoise sphere with arrows appears at the centroid of each ring, indicating that the new X feature type has been perceived.

This example was for demonstration purposes only, so you should remove the custom feature X before finishing this exercise.

11. In the Edit Features dialog box, ensure that the Custom (X) definitions are displayed.
12. Click Delete Custom Feature.

    The Custom (X) definitions are removed and the Acceptor (A) definitions are displayed.
13. Click OK.

    The Edit Features dialog box is closed and the definitions are stored with the custom feature removed.
14. Verify that the custom X feature is no longer visible.

Creating a Phase Database

Once a pharmacophore model has been developed, it may be used to search a database, with the goal of identifying additional active molecules. This chapter describes the process of creating Phase 3D databases. You can search a set of 3D structures in a structure file, but if you plan to search the same database with the same feature set multiple times, it is quicker to create a Phase database that includes conformer sets and sites.

If you plan to search the database on a host that is different from the one you will use in this chapter, you must create the database on a file system that is accessible to the other host.

3.1 Creating a New 3D Database

In this exercise you will create a small database from a Maestro file that contains single-conformer models of 100 druglike molecules. The conformers and sites will be generated and stored in the database.

1. Choose one of the following in the main window.

   - Applications → Phase → Generate Phase Database
   - Tasks → Pharmacophore Modeling → Generate Database

   The Generate Phase Database panel opens.

You can add structures to a database from multiple sources. These sources can contain single molecules, or conformer sets, and both can be stored in the database. When you add structures from a file, you can choose to clean up the structures if they are 2D or are not well optimized.

2. Click Reset at the bottom of the panel to ensure that the panel is cleared and the default settings are selected.
3. Click Browse to the right of the Input Structure file text box.

   A file selector opens.

4. Select `dbMolecules.mae` and click Open.

   The file selector closes, and the file name is displayed in the Input structure file text box.

Each structure added to the database needs a unique title. In this exercise the default, title (Maestro) is used. However, you may also define structure titles based on selected properties.

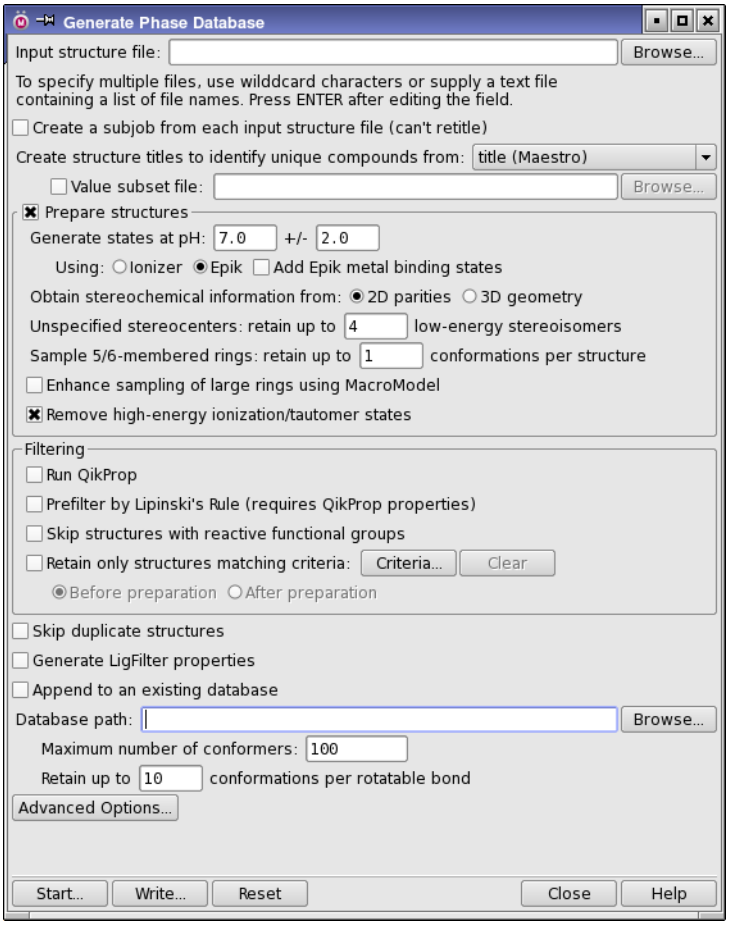

Figure 3.1. The Generate Phase Database panel.

5. Ensure that title (Maestro) is chosen from the Create structure titles to identify unique compounds from option menu.

Databases are often created from chemical files that contain fairly crude structures, so it is important to clean the structures unless you are certain that you have good 3D models with the proper numbers of hydrogens attached. If you are certain that you have good structures, you can deselect Prepare structures. In this exercise you will prepare the structures using the default settings.

6. Ensure that Prepare structures is selected.

   The structural preparation options should be left at the default settings. You will not be applying any filters so leave the filter options unselected (default).

7. Click Browse to the right of Database path.
8. Select the databases directory.
9. In Database name text box type tutorialDB.
10. Click Start to begin the database generation.

This job takes 15 minutes on a 2 GHz Pentium 4 processor. While the job is running, you can monitor the log files in the Monitor panel. Upon completion, the database contains 324 molecules. The number of molecules has increased from 100 to 324 due to the creation of stereoisomers and conformers.

3.2 Adding Structures to the Database

In this exercise, you will add the same structures to the database as in the previous exercise, but this time without generating conformers. This allows you to search the database with the conformers in the database and with conformer generation.

1. Choose one of the following in the main window.

   - Applications → Phase → Manage Phase Database
   - Tasks → Pharmacophore Modeling → Manage Database

   The Generate Phase Database panel opens.

2. Open the database you generated in the previous exercise with one of the following:

   - Type `tutorialDB.phdb` in the Open database text box and press Enter.
   - Click Browse, select `tutorialDB.phdb` and click OK.

   The database opens and the structures are listed in the Structures table.

3. Click Add.
4. In the file selector that opens, select `dbMolecules.mae` and click Open.

   A dialog box opens.

5. Ensure that Conformers present is selected

   The file does not actually have conformers, so selecting this option ensures that the structures are imported into the database without generating conformers. When creating a new database, conformers are automatically generated and added to the database.

6. Deselect Generate sites.

   Since conformers have to be generated for these molecules when a search is done, sites will automatically be generated as well.

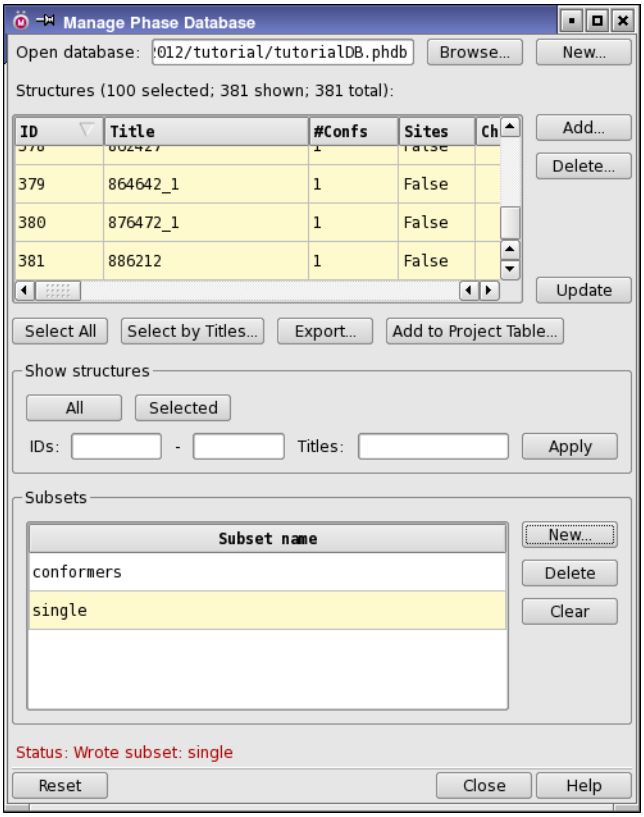

Figure 3.2. The Manage Phase Database panel.

7. Click OK.

   The structures are imported into the database.

3.3 Creating Database Subsets

In this exercises you will create database subsets for the two sets of compounds you imported. This will allow you to choose the subsets for searches, so you can search only the structures that have conformers, or only those that don't.

1. Select structures 1 to 281 in the table.

   You can use shift-click to select the range of structures. These structures should all have True in the Sites column.

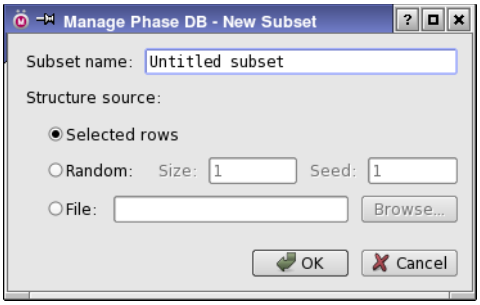

Figure 3.3. The Manage Phase DB - New Subset dialog box.

2. In the Subsets section, click New.

   The Manage Phase DB - New Subset dialog box opens. By default the Structure source is Selected rows.

3. Change the subset name to `conformers` and click OK.

   A new row is added to the Subsets table, named conformers.

4. Select structures 282 to 381 in the table.

   These structures should all have False in the Sites column.

5. In the Subsets section, click New.

6. In the Manage Phase DB - New Subset dialog box, change the subset name to `single` and click OK.

   A new row is added to the Subsets table, named single.

Finding Matches to a Hypothesis

One of the primary reasons for developing a pharmacophore model is to accelerate the identification of new active compounds. This is most commonly done by searching a 3D database for matches to a pharmacophore hypothesis. In this chapter, we demonstrate how an existing pharmacophore hypothesis is used to search the 3D database created in Chapter 3. If you do not want to create a database, you can use the database supplied with the tutorial.

4.1 Preparing for the Exercises

1. If you need to change to your tutorial directory, choose Project → Change Directory and navigate to the directory.
2. Choose one of the following in the main window:
   - Applications → Phase → Advanced Pharmacophore Screening
   - Tasks → Pharmacophore Modeling → Advanced Screening

   The Advanced Pharmacophore Screening panel opens.

4.2 Choosing the Database and Hypothesis

Before performing a search, you must select a database and a hypothesis.

1. Click Browse.

   A directory selector opens, with Files of type set to Phase 3D DB (*.phdb). Phase databases are directories, in the same way as Maestro projects.
2. Select `tutorialDB.phdb` and click Choose.

   If you did not generate a database, select `testDB.phdb` instead. The path to the database is now listed in the File name text box.

For the first database search, you will use the subset that contains conformers and sites.

3. Click the Browse button for Subset.
4. In the Subset File file selector, select `conformers_phase.inp` and click OK.

   The subset file is listed in the Subset text box.

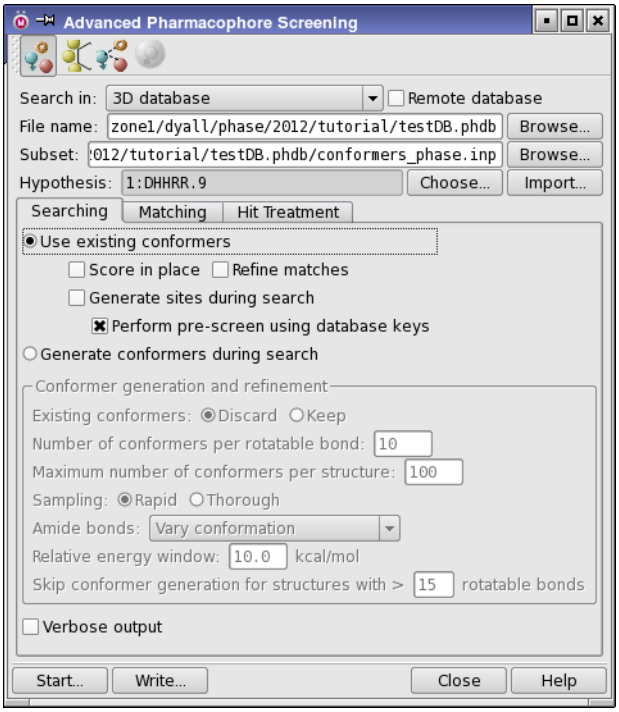

Figure 4.1. The Advanced Pharmacophore Screening panel, Searching tab.

Next, you will import the hypothesis.

5. Click Import.

   A file selector labeled Import Hypothesis opens.

6. Select `DHHRR_9.xyz` and click Open.

   The hypothesis title `DHHRR_9` is in the Hypotheses text box, prefixed by the entry ID.

If you click Search for Matches in the Develop Common Pharmacophore Hypotheses panel, the selected hypotheses are added to the Project Table and are available for database searching. Likewise, if you add hypotheses in the Manage Hypotheses panel, these are also available for database searching.

4.3 Searching the Existing Conformers

In this exercise, you will perform a search using default options that searches the subset with conformers already generated.

1. In the Searching tab, ensure that Use existing conformers is selected.
2. In the Matching tab, ensure that Excluded volumes is selected.
3. In the Hit Treatment tab, ensure that Use QSAR model is selected.

   This ensures that the QSAR models and excluded volumes are used. The remaining options can be left at the defaults.
4. Click Start.

   The Start dialog box opens. In this dialog box you can enter a job name, select the host to run the job and the number of processors to use. For the exercises in this chapter, you will run the job on the local host.
5. Enter `conformers` in the Name text box.
6. Click Start in the Start dialog box.

   The entire search should take less than a minute, after which a total of 12 hits are placed in the Project Table.
7. Click the Open/Close project table button on the main toolbar.

   

   The Project Table panel is displayed. The hits are in an entry group that is named after the job, `conformers-hits1`. You may view the hits in the Workspace and examine the variety of properties that are written to the Project Table, including the matching sites indices, the fitness and its various components, and the predicted activities from the QSAR model.

4.4 Searching with Site and Conformer Creation

In this exercise you will search the database subset for which conformers and sites were not generated. The conformers and sites will be generated during the search. This kind of search is known as “flexible searching”.

1. In the Find Matches to Hypothesis panel, click the Browse button to the right of the Subset text box.

2. Select the subset file `single_phase.inp` and click OK.
3. In the Searching folder, select Generate conformers during search.
4. Click Start.
5. In the Name text box of the Start dialog box, enter `single`.
6. Click Start.

The output to the Monitor panel now includes information about the conformation generation. This job should take several minutes, because the conformers are being generated during the search. When the job is finished, a total of 12 hits appear in the Project Table, in an entry group named `single-hits1`. The results should be identical to those obtained in the first search.

Getting Help

Information about Schrödinger software is available in two main places:

- The `docs` folder (directory) of your software installation, which contains HTML and PDF documentation. Index pages are available in this folder.
- The Schrödinger web site, http://www.schrodinger.com/, particularly the Support Center, http://www.schrodinger.com/supportcenter, and the Knowledge Base, http://www.schrodinger.com/kb.

Finding Information in Maestro

Maestro provides access to nearly all the information available on Schrödinger software.

To get information:

- Pause the pointer over a GUI feature (button, menu item, menu, ...). In the main window, information is displayed in the Auto-Help text box, which is located at the foot of the main window, or in a tooltip. In other panels, information is displayed in a tooltip.

  If the tooltip does not appear within a second, check that Show tooltips is selected under General → Appearance in the Preferences panel, which you can open with CTRL+, (⌘,). Not all features have tooltips.
- Click the Help button in a panel or press F1 for information about a panel or the tab that is displayed in a panel. The help topic is displayed in your browser.
- Choose Help → Online Help or press CTRL+H (⌘H) to open the default help topic in your browser.
- When help is displayed in your browser, use the navigation links or search the help in the side bar.
- Choose Help → Manuals Index, to open a PDF file that has links to all the PDF documents. Click a link to open the document.
- Choose Help → Search Manuals to search the manuals. The search tab in Adobe Reader opens, and you can search across all the PDF documents. You must have Adobe Reader installed to use this feature.

For information on:

- Problems and solutions: choose Help → Knowledge Base or Help → Known Issues → *product*.
- Software updates: choose Maestro → Check for Updates.
- New software features: choose Help → New Features.
- Scripts available for download: choose Scripts → Update.
- Python scripting: choose Help → Python Module Overview.
- Utility programs: choose Help → About Utilities.
- Keyboard shortcuts: choose Help → Keyboard Shortcuts.
- Installation and licensing: see the *Installation Guide*.
- Running and managing jobs: see the *Job Control Guide*.
- Using Maestro: see the *Maestro User Manual*.
- Maestro commands: see the *Maestro Command Reference Manual*.

Contacting Technical Support

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail: help@schrodinger.com
USPS: Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone: (503) 299-1150
Fax: (503) 299-4532
WWW: http://www.schrodinger.com
FTP: ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information:

- All relevant user input and machine output
- Phase purchaser (company, research institution, or individual)
- Primary Phase user
- Installation, licensing, and machine information as described below.

Gathering Information for Technical Support

This section describes how to gather the required machine, licensing, and installation information, and any other job-related or failure-related information, to send to technical support.

For general enquiries or problems:

1. Open the Diagnostics panel.
   - **Maestro:** Help → Diagnostics
   - **Windows:** Start → All Programs → Schrodinger-2012 → Diagnostics
   - **Mac:** Applications → Schrodinger2012 → Diagnostics
   - **Command line:** `$SCHRODINGER/diagnostics`
2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.
3. Attach the file specified in the dialog box to your e-mail message.

If your job failed:

1. Open the Monitor panel in Maestro.

   Use Applications → Monitor Jobs or Tasks → Monitor Jobs.
2. Select the failed job in the table, and click Postmortem.

   The Postmortem panel opens.
3. If your data is not sensitive and you can send it, select Include structures and deselect Automatically obfuscate path names.
4. Click Create.

   An archive file is created in your working directory, and an information dialog box with the name of the file opens. You can highlight and copy the name of the file.
5. Attach the file specified in the dialog box to your e-mail message.
6. Copy and paste any log messages from the window used to start Maestro (or the job) into the email message,or attach them as a file.
   - **Windows:** Right-click in the window and choose Select All, then press ENTER to copy the text.
   - **Mac:** Start the Console application (Applications → Utilities), filter on the application that you used to start the job (Maestro, BioLuminate, Elements), copy the text.

If Maestro failed:

1. Open the Diagnostics panel.

   - **Windows:** Start → All Programs → Schrodinger-2012 → Diagnostics
   - **Mac:** Applications → Schrodinger2012 → Diagnostics
   - **Linux/command line:** `$SCHRODINGER/diagnostics`

2. When the diagnostics have run, click Technical Support.

   A dialog box opens, with instructions. You can highlight and copy the name of the file.

3. Attach the file specified in the dialog box to your e-mail message.

4. Attach the file `maestro_error.txt` to your e-mail message.

   This file should be in the following location:

   - **Windows:** `%LOCALAPPDATA%\Schrodinger\appcrash`
     (Choose Start → Run and paste this location into the Open text box.)
   - **Mac:** `Documents/Schrodinger`
   - **Linux:** Maestro's working directory specified in the dialog box (the location is given in the terminal window).

5. On Windows, also attach the file `maestro.EXE.dmp`, which is in the same location as `maestro_error.txt`.

Glossary

Active compound—A compound that shows high affinity for the biological target. Synonymous with the term *ligand.*

Active set—The set of active compounds that is used to develop a pharmacophore model. This set does not necessarily include all active compounds.

Excluded volume—A region of space in a pharmacophore hypothesis that should not be occupied by any atom of an active compound.

Feature—see **Pharmacophore feature**

Hit—A structure in a 3D database that is found to contain an arrangement of site points that can be mapped to the pharmacophore hypothesis. A hit is not necessarily active, but it is presumed to have a greater than average probability of being active if it was retrieved using a valid hypothesis.

Hypothesis—see ***n*-Point pharmacophore hypothesis**

Inactive compound—A compound that shows little or no affinity for the biological target.

Intersite distance—The distance between any two site points in a pharmacophore.

Ligand—see **Active compound**

Negative compound—A compound that is inactive, yet highly similar in structure to one or more known actives. Some compounds are negative because they lack certain key pharmacophore features found in true actives. Other negatives may actually satisfy exactly the same pharmacophore hypotheses as the actives, but possess extraneous structural characteristics that prevent binding.

Pharmacophore feature—A characteristic of chemical structure that may facilitate a noncovalent interaction between a ligand and a biological target. Examples are hydrogen-bond acceptor (“A”), hydrogen-bond donor (“D”), hydrophobe (“H”), positive ionic center (“P”), negative ionic center (“N”).

Pharmacophore site—The labeling and location of a particular pharmacophore feature within a molecule. For example, a hydrogen bond acceptor site could simply be a nitrogen atom that carries an available lone pair. A hydrophobic site might be a methyl carbon or the centroid of a phenyl ring. The term *site point* is often used interchangeably with pharmacophore site.

Pharm set—The set of all compounds, active and inactive, used to develop a pharmacophore model.

***n*-Point pharmacophore**—Any 3D arrangement of *n* pharmacophore features.

***n*-Point pharmacophore hypothesis**—A specific 3D arrangement of *n* pharmacophore features, with associated uncertainties in the feature positions. High affinity ligands in their active conformations are expected to contain pharmacophore sites that can be mapped (within the limits of uncertainty) to any valid hypothesis. A given hypothesis may contain features that are associated with a single mode of binding, or it may contain features that are common to two or more modes of binding.

Reference ligand—The ligand that provides the pharmacophore that defines a hypothesis. In pharmacophore model development, this pharmacophore yields the highest multi-ligand alignment score for the active-set ligands. The reference ligand matches the hypothesis exactly, and has a perfect fitness score.

Site point—see **Pharmacophore site**

3D Database—A set of molecules, each of which is represented by one or more 3D conformational models, augmented with a pharmacophore-based representation of the molecules. A 3D database includes feature types and site point coordinates for each conformation.

Variant—The set of feature types in a pharmacophore. For example, the variant AHH indicates a 3-point pharmacophore containing one hydrogen bond acceptor and two hydrophobic sites.

Vector feature—A pharmacophore feature that contains directionality, such as a hydrogen bond acceptor, hydrogen bond donor, or aromatic ring. A vector feature does not necessarily have vector geometry.

Vector geometry—the geometric characteristics of hydrogen bond acceptors and donors. Refers to the direction of lone pairs in a hydrogen-bond acceptor or the direction of the heavy-atom–hydrogen-atom bond in hydrogen-bond donors. Features with vector geometry must be vector features.

120 West 45th Street
17th Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

245 First Street
Riverview II, 18th Floor
Cambridge, MA 02142

8910 University Center Lane
Suite 270
San Diego, CA 92122

155 Gibbs St
Suite 430
Rockville, MD 20850-0353

Dynamostraße 13
D-68165 Mannheim
Germany

Zeppelinstraße 73
D-81669 München
Germany

Potsdamer Platz 11
D-10785 Berlin
Germany

Quatro House
Frimley Road
Camberley GU16 7ER
United Kingdom

8F Pacific Century Place
1-11-1 Marunouchi
Chiyoda-ku, Tokyo 100-6208
Japan

No. 102, 4th Block
3rd Main Road, 3rd Stage
Sharada Colony
Basaveshwaranagar
Bangalore 560079, India

SCHRÖDINGER®