# SiteMap 2.6

## User Manual

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 to SiteMap

## 1.1 SiteMap Overview

The location of the primary binding site on a receptor such as a protein is often known from the structure of a co-crystallized complex. Efforts to design better ligands for these receptors can profit from an understanding of how well the known ligands complement the receptor, and how extension of the ligands into adjacent regions could promote binding. Determining whether there are nearby sites that might be useful for allosteric binding can also be important.

In some cases, however, the location of a binding site for protein-ligand or protein-protein interactions is not known in advance, even though the protein structures are available. Here, computational studies can help to suggest likely binding sites, and even to predict whether a given protein is likely to bind ligands tightly. Many such approaches have been explored; for references, see the recent paper by Nayal and Honig [1].

SiteMap [2] generates information on the character of binding sites using novel search and analysis facilities, and provides information to Maestro for visualization of the sites. A SiteMap calculation begins with an initial search stage that determines one or more regions on or near the protein surface, called *sites,* that may be suitable for binding of a ligand to the receptor. The search uses a grid of points, called *site points*, to locate the sites. In the second stage, contour maps (*site maps*) are generated, producing hydrophobic and hydrophilic maps. The hydrophilic maps are further divided into donor, acceptor, and metal-binding regions. The evaluation stage, which concludes the calculation, assesses each site by calculating various properties, which are added to the Maestro project.

Site maps can aid in the design of better ligands, by revealing “targets of opportunity”—for example, hydrophobic regions that have room to accommodate a larger hydrophobic group. Site maps can also be used to select the target for ligand docking with Glide and to evaluate docking hits, by showing how well the poses display proper complementarity to the receptor. The regions that are neither hydrophobic nor hydrophilic are important because they show places in which it may be possible to improve the physical properties of the ligand—for example, by changing the solubility—with minimal effect on the binding affinity.

In contrast to techniques that color-code the receptor surface to represent hydrophilicity or hydrophobicity, site maps depend on the site as a whole, not just the character of the nearest receptor atom. Moreover, site maps explicitly show the shape and extent of philic and phobic regions, something a surface-based display cannot do.

The most important property generated by SiteMap is an overall SiteScore, which has proven to be effective at identifying known binding sites in co-crystallized complexes. Other properties characterize the binding site in terms of:

- the size of the site,
- the degrees of enclosure by the protein and exposure to solvent,
- the tightness with which the site points interact with the receptor,
- the hydrophobic and hydrophilic character of the site and the balance between them,
- the degree to which a ligand might donate or accept hydrogen bonds.

SiteMap can be run either from Maestro or from the command line. A SiteMap calculation typically takes a few minutes for proteins having up to 5000 atoms. Version 2.6 of SiteMap contains many improvements over the original SiteMap, which is still available from the Surface submenu of the Workspace menu in Maestro, as Hydrophobic/philic.

## 1.2 Running Schrödinger Software

Schrödinger applications can be started from a graphical interface or from the command line. The software writes input and output files to a directory (folder) which is termed the *working directory*. If you run applications from the command line, the directory from which you run the application is the working directory for the job.

**Linux:**

To run any Schrödinger program on a Linux platform, or start a Schrödinger job on a remote host from a Linux platform, you must first set the `SCHRODINGER` environment variable to the installation directory for your Schrödinger software. To set this variable, enter the following command at a shell prompt:

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

**bash/ksh:** `export SCHRODINGER=`*installation-directory*

Once you have set the `SCHRODINGER` environment variable, you can run programs and utilities with the following commands:

```
$SCHRODINGER/program &
$SCHRODINGER/utilities/utility &
```

You can start the Maestro interface with the following command:

```
$SCHRODINGER/maestro &
```

It is usually a good idea to change to the desired working directory before starting Maestro. This directory then becomes Maestro's working directory.

**Windows:**

The primary way of running Schrödinger applications on a Windows platform is from a graphical interface. To start the Maestro interface, double-click on the Maestro icon, on a Maestro project, or on a structure file; or choose Start → All Programs → Schrodinger-2012 > Maestro. You do not need to make any settings before starting Maestro or running programs. The default working directory is the Schrodinger folder in your documents folder (Documents on Windows 7/Vista, My Documents on XP).

If you want to run applications from the command line, you can do so in one of the shells that are provided with the installation and that have the Schrödinger environment set up:

- Schrödinger Command Prompt—DOS shell.
- Schrödinger Power Shell—Windows Power Shell (if available).

You can open these shells from Start → All Programs → Schrodinger-2012. You do not need to include the path to a program or utility when you type the command to run it. If you want access to Unix-style utilities (such as `awk`, `grep`, and `sed`), preface the commands with `sh`, or type `sh` in either of these shells to start a Unix-style shell.

**Mac:**

The primary way of running Schrödinger software on a Mac is from a graphical interface. To start the Maestro interface, click its icon on the dock. If there is no Maestro icon on the dock, you can put one there by dragging it from the SchrodingerSuite2012 folder in your Applications folder. This folder contains icons for all the available interfaces. The default working directory is the Schrodinger folder in your Documents folder (`$HOME/Documents/Schrodinger`).

Running software from the command line is similar to Linux—open a terminal window and run the program. You can also start Maestro from the command line in the same way as on Linux. The default working directory is then the directory from which you start Maestro. You do not need to set the `SCHRODINGER` environment variable, as this is set in your default environment on installation. If you need to set any other variables, use the command

`defaults write ~/.MacOSX/environment` *variable* `"`*value*`"`

## 1.3 Citing SiteMap in Publications

The use of this product should be acknowledged in publications as:

SiteMap, version 2.6, Schrödinger, LLC, New York, NY, 2012.

# How SiteMap Works

SiteMap 2.6 represents an expansion of the original SiteMap facility in Maestro. As in the original procedure [4], site mapping operates in a manner analogous to Goodford's GRID algorithm [5]. A SiteMap calculation has three stages. First, a grid is set up, and the points are grouped into sets according to various criteria to define the sites. Second, the sites are mapped on another grid to produce files for visualization of the maps. Finally, properties are evaluated and sites are written in a Maestro-readable form. Each stage is accomplished by running an Impact job. The three stages are described in the sections below.

## 2.1 Finding Sites

The first stage of a SiteMap calculation is to locate the sites. A *site* is defined by a set of site points on a grid that are either contiguous or bridged by short gaps in solvent-exposed regions. The site-finding algorithm begins by placing a 1-Å grid of possible site points around the entire protein or around a placeholder species, such as a ligand. Identifying the sites involves several steps.

The first step is to classify the grid points as being either "inside" or "outside" the protein. The distance from each grid point to nearby protein atoms is compared to the van der Waals radius of each protein atom. If the ratio of the squares of these distances is larger than a given threshold, the point is considered to be outside the protein.

In the next step, the "outside" points are examined to determine which ones are in sufficiently good van der Waals contact with the receptor and sufficiently enclosed by the receptor to serve as site points. Enclosure is defined by sampling all possible directions from the grid point and determining the fraction of these directions ("rays") that strike the surface within a given distance. If the fraction is larger than a given threshold, the point is sufficiently enclosed and is therefore a candidate site point. The contact with the receptor is determined by a cutoff on the van der Waals interaction energy at the site point: if the interaction energy is too small in magnitude, the point is rejected. Points that meet all criteria are added to the list of site points.

The third step combines site points into distinct site-point groups. For a site point to be considered for membership in a group, it must have a minimum number of candidate site points within a given distance. Site points that do not have this minimum number are discarded. The process starts by assigning a site point to a group, then adding all candidate site points within a prescribed minimum distance. The addition process is repeated for each new site point. When

no further site points can be added, the group is considered complete and another site-point group is initiated. The process continues until all site points have been examined.

The final step merges site-point groups when the gap between them is relatively small and occurs in a solvent-exposed region. The merge is controlled by user-adjustable thresholds that determine how close two site-point groups must be for them to be considered for merging and whether the gap between them could plausibly be bridged by ligand atoms. The final groups constitute the sites.

The sites are written in order of the number of site points they contain to a Maestro file. Each site point is represented by a dummy atom, and zero-order bonds are used to join the site points for each site into a "structure" that Maestro recognizes as a single molecule.

## 2.2 Mapping the Sites

The second stage of a SiteMap calculation generates the various "maps" that define the sites. A map is defined by a set of values of a property on a given 3D grid. First, SiteMap uses the site points from the preceding stage to position a mapping box for each site. This box defines a grid with a given spacing, and extends beyond the site by a given amount.

Van der Waals and distance-dependent electrostatic-interactions of a probe placed at each of the grid points are then used to generate van der Waals and electric-field grids. The probe simulates a water molecule, and is represented by a van der Waals sphere of radius 1.6 Å, a well depth of 0.13 kcal/mol, and a point dipole moment of 2.4 Debye. To form the electrostatic-field grid, the probe's point dipole is oriented along the electric field and is offset by 0.15 Å from the van der Waals sphere toward the center of an optimally oriented O–H bond.

To more accurately represent the expected contact positions and interaction energies of donor and acceptor atoms, the force field is first modified by adjusting van der Waals radii and by reducing formal-charge contributions to the partial atomic charges by 50%. The reduction in formal charges, like the one employed in Glide [7], is used to keep regions around formally charged groups from inappropriately dominating the maps, which are meant to reflect interactions in solvent, not in the gas phase.

The resultant van der Waals and electric-field grids are then used to generate the phobic and philic potentials. Using these potentials, SiteMap partitions the accessible space in each site into the following three basic types of regions:

- Hydrophobic—regions that are favorable for occupancy by hydrophobic ligand groups
- Hydrophilic—regions that are favorable for occupancy by hydrophilic ligand groups
- Neither hydrophobic nor hydrophilic—regions that are of mixed character or are far enough from the receptor surface to be similar to bulk water

The hydrophilic regions are further subdivided into hydrogen-bond donor, hydrogen-bond acceptor, and metal-binding regions. The hydrophobic and hydrophilic regions (or maps) are obtained by contouring the computed phobic and philic potentials at specified threshold values. The "neither" regions are implicit: these are regions that lie outside the protein but are not marked as being either hydrophobic or hydrophilic. The methods for obtaining the maps are described in more detail in the sections below.

For each site, the five maps—hydrophilic, hydrophobic, donor, acceptor, and surface—are written to files (`.grd` files and `.vis` files) that can be used by Maestro to display the surfaces. If there is a metal, the metal-binding map is also written out.

### 2.2.1 Hydrophilic Map

SiteMap constructs a measure of hydrophilicity by adding an "electric-field reward" term to the van der Waals energy:

Grid_philic = vdW_energy + oriented-dipole_energy

where the oriented-dipole energy is necessarily negative. Hydrophilic regions are those within which the sum of the two terms is more negative than a given threshold, which by default is –8 kcal/mol.

### 2.2.2 Hydrophobic Map

The quantity representing hydrophobicity is constructed by adding an "electric-field penalty" (positive) term to the van der Waals term:

Grid_phobic = vdW_energy – 0.30 * oriented-dipole_energy

Hydrophobic regions thus are regions where something would like to be, but water would not. The starting point for defining the hydrophobic regions is to consider the regions within which the sum of the two terms is more negative than a given threshold.

SiteMap offers several alternatives for the definition of hydrophobic regions. The least restrictive definition uses a threshold of –0.75 kcal/mol, and includes all site points that lie within this region. This is not the default behavior, but represents the original SiteMap definition. The more restrictive definitions involve two possible modifications to the hydrophobic region.

In the first, grid points that border on too many assigned philic points, that have too few phobic-point or "inside-protein" neighbors, or that border on too many free-space "outside" points are reclassified as non-phobic.

In the second, for each phobic grid point the fraction of radial rays that intersect the protein surface within a given distance (default 6 Å) is calculated, and the phobic potential is then

multiplied by this fraction. By scaling down the phobic potential, exposed regions are less favored than regions that are sheltered from the solvent, like the Glide XP detection of "phobic enclosure" [7]. The threshold for defining the phobic region is also reduced, to –0.50 kcal/mol. This modification considers the nature of the site beyond the immediate vicinity of the grid point, in a way that the first modification cannot.

The default behavior is to include both of these modifications.

### 2.2.3 Donor, Acceptor, and Metal-Binding Regions

The hydrophilic map is further partitioned into separate hydrogen-bond donor and acceptor maps. When there is interaction with a metal center other than $Ca^{2+}$, which (as in Glide XP) is not considered to be a metal-binding center, a separate metal-binding map is also formed. Metal-binding grid points are philic grid points that lie within 3 Å of a qualifying metal center. Classification of the remaining philic points as donor or acceptor points is made by displacing them in the direction of the local electrostatic field and recomputing the value of the field. Donor and acceptor points are assigned depending on whether this displacement increases or decreases the magnitude of the field.

### 2.2.4 Surface Map

The surface map is obtained by removing attractive regions of the van der Waals grid and then contouring the repulsive part of this grid at a positive threshold value, which is set to +1 kcal/mol by default.

## 2.3 Evaluating the Sites

This stage uses the site-point groups produced in the site-finding stage and the grids produced in the mapping stage to evaluate the sites in terms of a number of properties. The same modifications to van der Waals radii and formal-charge contributions and the same definition of hydrophobicity are used as in the mapping stage. The properties for each site are added to the Maestro file for the site and recorded in the log file.

To minimize grid errors, the contact, phil, and don/acc SiteMap properties are calculated explicitly as average values computed at the site-point positions (including extension points), but the more complicated phob property is obtained by interpolation from the phobic grid file produced in the site-visualization step.

To make it easy to recognize sites that appear to be unusually favorable or deficient, key properties are expressed relative to the average value found for a large number of tight-binding (≤ 1 μM) sites. The procedure by which this average was obtained is described in Chapter 5. The properties and their use are described below.

**SiteScore.** The SiteScore is based on a weighted sum of several of the properties that are discussed below:

SiteScore = 0.0733 sqrt($n$) + 0.6688 $e$ - 0.20 $p$

where $n$ is the number of site points (capped at 100), $e$ is the enclosure score, and $p$ is the hydrophilic score, and is capped at 1.0 to limit the impact of hydrophilicity in charged and highly polar sites. This score is constructed and calibrated so that the average SiteScore for 157 investigated submicromolar sites is 1.0. Thus, a score of greater than 1 suggests a site of particular promise. A SiteScore of 0.80 has been found to accurately distinguish between drug-binding and non-drug-binding sites (see Chapter 5).

**Druggability Score, Dscore.** Dscore uses the same properties as SiteScore but different coefficients:

Dscore = 0.094 sqrt($n$) + 0.60 $e$ - 0.324 $p$

For Dscore, the hydrophilic score is not capped. This one of the keys for distinguishing "difficult" and "undruggable" targets from "druggable" ones [8]. The use of different functions for binding-site identification and for classifying druggability is justified because these are different, and sometimes conflicting, tasks. For example, ligands that bind to the PTP1B phosphate pocket with nanomolar, and even subnanomolar, affinity are known [9]. But these highly active ligands have charge structures like those of the natural phosphate substrate and are not drug-like. SiteMap should recognize that such a site can bind ligands tightly but should not rate it as druggable.

**Number of Site Points.** The number of site points that make up the site is a measure of the size of the site. As a rough rule of thumb, 2 to 3 site points typically correspond to each atom of the bound ligand, including hydrogens. The size of the site is often a good indicator of the preferred binding site.

**Exposure and Enclosure.** These two properties provide different measures of how open the site is to solvent.

To evaluate the exposure property, "extension" site points are added on the 1-Å grid. These points must lie within a given distance in $x$, $y$, or $z$ from an original site point (by default 3 Å), and must make good contact with the receptor or lie at least 4 Å from the nearest protein atom. The value of the property is the ratio of the number of extension points to the number of original plus extension points. A shallow, open site would allow many more site points to be added, giving a high exposure score. The lower the score, the better; the average for the tight-binding sites investigated is 0.49.

To evaluate the enclosure property, radial rays are drawn from the site points to sample all possible directions. The enclosure score is the fraction of rays that strike the receptor surface

within a distance of 10 Å, averaged over the original and the extension site points used in the exposure evaluation. The receptor surface is the same surface that was used to classify grid points as outside or inside the protein in the site-finding step. Here, higher scores are better, with the average enclosure score for a tight-binding site being 0.78.

**Contact.** The contact property measures how strongly the average site point interacts with the surrounding receptor via van der Waals nonbonded interactions, when the site point is given nominal van der Waals parameters. The contact score has been calibrated so that the average score for a tight-binding site is 1.0.

**Hydrophobic and hydrophilic character, and Balance.** These properties, labeled phob and phil, measure the relative hydrophobic and hydrophilic character of the site. The balance property expresses the ratio of the two. The phobic and philic scores have been calibrated so that the average score for a tight-binding site is 1.0. The average balance score for the investigated tight-binding sites, on the other hand is 1.6, not 1.0, because sites that have high phobic and low philic scores make large contributions to the average.

**Donor/Acceptor character.** This property, labeled don/acc, indicates the degree to which a well-structured ligand might be expected to donate, rather than accept, hydrogen bonds, as inferred from the sizes and intensities of donor and acceptor SiteMap regions.

**Reference distance properties.** When a supplied ligand or other species is used to define the region of the receptor to be mapped, refdist, refmin, refavg, and sitemin properties are also computed. The first of these specifies the distance between the centroid of the site points and the centroid of the reference ligand. The second specifies the closest approach of a site point to a ligand atom. Both are given in angstroms.

The sitemin property is the smallest distance between an atom of the reference species used to define the site and the site-point centroid. If that reference species is the co-crystallized ligand, 4 Å and less is normally taken as a "hit", by analogy to other practice in the literature. Larger values sometimes occur for cases in which the minimum distance of a reference atom to an individual site point (refmin) is small, showing that the site-point set does at least partly cover the reference ligand, because the site-point set is large and extends asymmetrically from the region occupied by the reference species.

In some cases, a small refmin value (typically < 1 Å) is accompanied by a moderately large refdist of 5 – 10 Å. These are cases in which the site extends asymmetrically beyond the reference ligand in one or more directions. In an endoprotease, such extensions may well map the channels that bind the N-terminal and C-terminal strands of the peptide undergoing cleavage, and hence are to be expected. These extensions are of interest because they may represent regions that a tight-binding ligand might usefully probe.

**Site Volume.** The volume of a protein site is well defined when the site is fully enclosed by the protein. More commonly, however, the site is open to solution on one or more sides. To assign the volume in such a case, what needs to be decided, as one proceeds outward from the protein surface, is where to stop counting. Understandably, different criteria will yield different site volumes.

SiteMap's approach approximates the "shrink-wrap" volume of the site by excluding regions that protrude too far into the solvent. This is accomplished by first identifying all points on the cubic mapping grid that lie within 4 Å of any site point and are outside the protein surface. By default, the grid spacing is 0.7 Å. A large number of radial rays are then drawn from each candidate volume point, and those for which fewer than 60% of the rays strike the protein surface within 8 Å are removed. The volume of the site is then computed from the number of remaining volume points and the grid-box volume, which is $(0.7\ \text{Å})^3$ in the default case.

If you run calculations from the command line, the mapping points considered to lie within the shrink-wrap can be saved by including the option `-keepvolpts`. This option returns a pdb-format file for each site that contains the calculated volume points. You can then visualize these points by importing the file.

# SiteMap Tutorial

This chapter provides an exercise on setting up a job to search a protein for possible active sites. The protein is 1ke8, which has a known site and ligand. This information will be used to assess how well the site is located. The results can be used to prepare a grid for Glide docking—see Chapter 5 of the *Glide Quick Start Guide* for grid generation and docking exercises using this site map.

## 3.1 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 `sitemap` 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 SiteMap Tutorial 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.

## 3.2 Running the SiteMap Calculation

In this exercise, you will set up and run the SiteMap calculation to locate sites on 1ke8.

1. Click the Import button on the Project toolbar.

   

   The Import panel opens.
2. From the Files of type option menu, ensure that Maestro is chosen.
3. Select the file `1ke8_protein.maegz` and click Open.

   The protein is displayed in the Workspace.
4. Choose Applications → SiteMap.

   The SiteMap panel opens.
5. In the Task section, ensure that Identify top-ranked potential receptor binding sites is selected.

You can leave the settings at their defaults.

6. Click Start.
7. Ensure that Append new entries is selected from the option menu in the Output section.
8. Name the job `1ke8_sitemap_find`, and click Start.

The job takes only a few minutes. When it finishes, the first (top-ranked) site found is included in the Workspace, along with the surfaces.

## 3.3 Examining the SiteMap Results

The main results of the calculation are displayed in the Workspace when the job finishes. In this exercise, you will see how well the results fit the native ligand, and assess the other sites on the protein.

1. Click the Import button on the Project toolbar.

   

   The Import panel opens.
2. From the Files of type option menu, ensure that Maestro is chosen.
3. Select the file `1ke8_ligand.maegz`.
4. If the options are not displayed, click Options.
5. Deselect Replace Workspace.
6. Click Open.

   The ligand is displayed in the Workspace, in wire representation. To view it more easily, it will be changed to tube representation.
7. If the Representation toolbar is not displayed, click Representation on the Manager toolbar, or choose Window > Toolbars > Representation.
8. From the Tube button menu on the Representation toolbar, choose Molecule.
9. Pick an atom in the ligand molecule.

   The molecule is now displayed in tube representation, which makes it easier to see through the surfaces. The hydrophobic part of the ligand fits well into the hydrophobic region (yellow surface). To see the complementarity more clearly, you can display the surfaces separately.

10. If the Project Table panel is not open, click the Table button on the Project toolbar.
11. In the Project Table panel, choose Entry > Fix.

    The ligand entry is now "fixed" in the Workspace, and this is indicated by a padlock icon in the In column.
12. From the Tile button menu on the Workspace toolbar in the main window, choose Tile by Surface.

    
13. Each of the surfaces is displayed in a separate square region (tile), along with the site points, and the ligand. The title of the surface and the entry are displayed at the top of each tile.
14. Enlarge the main window so you can see each tile clearly.
15. On the Tile button menu, deselect Transform all Tiles.

    

    You can now rotate or translate each tile separately.
16. Rotate each tile to see how the ligand occupies or interacts with each of the three regions: acceptor, donor, and hydrophobic.

    The oxygens from the ring carbonyl and the sulfonamide are all in (or near) acceptor regions, two of the NH hydrogens are in donor regions, and three of the rings are in the main hydrophobic region.
17. Click the Tile button again to revert to the composite view.

For more information on tiling, see Section 4.4 of the *Maestro User Manual*.

Another way of viewing the surfaces separately is to display or undisplay them. To do this, click the S button in the Title column for this project entry (or choose Workspace > Surface > Manage Surfaces). The Manage Surfaces panel opens, and allows you to control the display of the surfaces, including their visibility, color, style, transparency, and so on. For more information on surfaces, see Chapter 12 of the *Maestro User Manual*.

18. In the Project Table panel, scroll to the SiteScore column.

    The first site has a score greater than 1.0, and includes 149 site points. The other sites have only 30–45 points and have values around 0.6–0.75, which indicate that they are not likely to bind drugs. The Dscore shows similar trends.

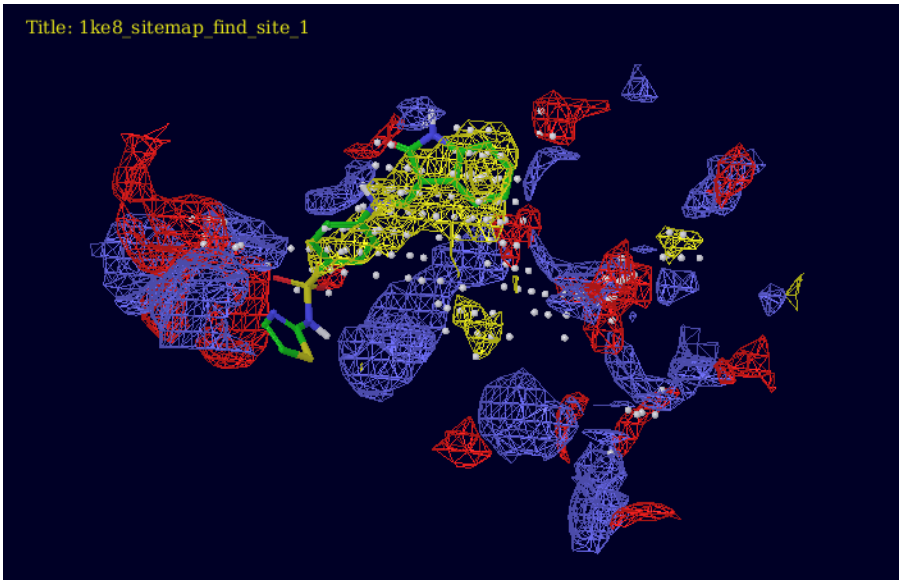


*Figure 3.1. The top-ranked site for 1ke8 with the ligand.*

# Running SiteMap

SiteMap can be run from within an existing Maestro session or from the command line. The original SiteMap facility is now accessible from the Surface submenu of the Workspace menu as Hydrophobic/philic.

## 4.1 Running SiteMap from Maestro

SiteMap calculations can be set up and run from the SiteMap panel, which is shown in Figure 4.1. To open this panel, choose SiteMap from the Applications menu. The panel is divided into two sections, Specify task and Settings, which are described below. When you have finished making settings, click Start to start the job.

### 4.1.1 Specifying the Task

In the Specify task section, you can choose between two tasks:

- Identify top-ranked potential receptor binding sites
- Evaluate a single binding site region

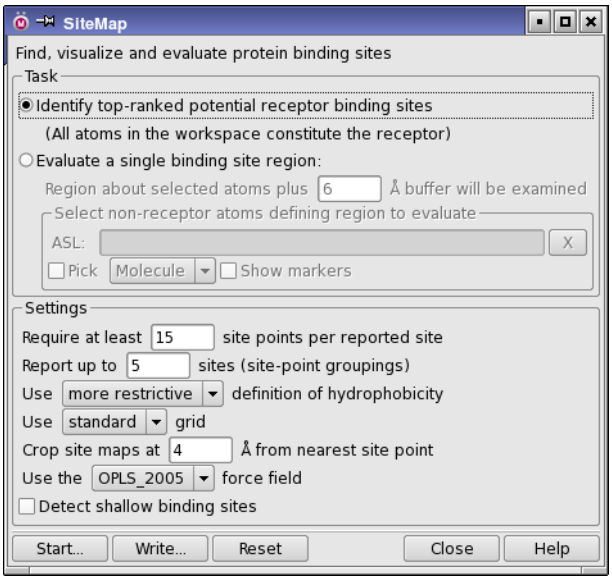


***Figure 4.1. SiteMap panel showing options for mapping an entire protein.***

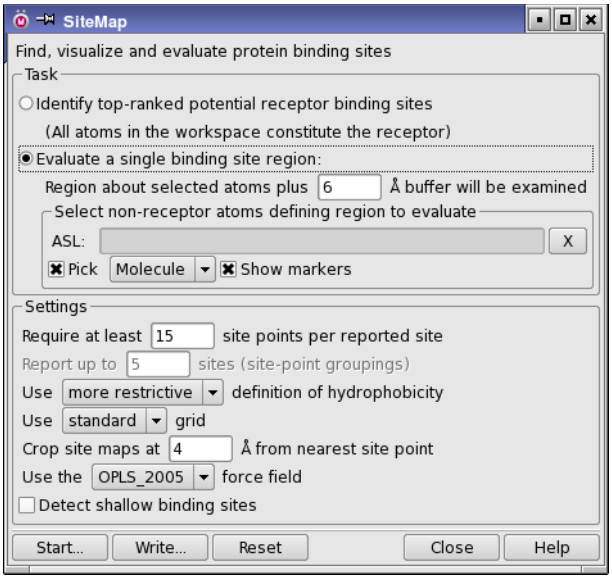


***Figure 4.2. SiteMap panel showing options for mapping a region defined by a set of atoms.***

For each task, the structure to be used must be displayed in the Workspace.

If you choose the first task, Identify top-ranked potential receptor binding sites, SiteMap looks for binding sites on the entire protein structure. The protein structure must be displayed in the Workspace, and should consist only of the protein—no ligand, waters, or cofactors. Figure 4.1 shows the SiteMap panel when this task is chosen.

If you choose the second task, Evaluate a single binding site region, SiteMap restricts the search for binding sites on the protein to a region around a specified structure. When you select this option, the controls below it become available. Figure 4.2 shows the SiteMap panel when this task is chosen.

For this task, you must pick a single molecule or entry in the Workspace to define the binding site, by choosing Molecule or Entry from the Pick menu in the Select non-receptor atoms defining region to evaluate section, and picking an atom in the Workspace. The ASL text box above the Pick menu displays the ASL expression for the selected structure. You can clear the selection by clicking the X button. The structure that you pick must not be the protein, and the contents of the Workspace must include only the protein and the structure you select to define the binding site. To mark the picked atoms, select Show markers.

If you want to include more than one molecule in the region for mapping (for example, a ligand and a cofactor), you should create an entry that includes only the molecules you want to use,

and another entry for the receptor. You can then display both entries, and pick the entry that contains the two molecules to define the region.

The region to be added around the selected reference molecule or entry extends out from the reference by 6 Å, by default. If you want to change the extent of the region, you can enter a value in the Region about selected atoms plus *n* Å buffer will be examined text box.

### 4.1.2 Setting Other Options

In the Settings section, you can make the following choices to control the calculations:

- Require at least *n* site points per reported site—Enter a value to set the minimum number of site points required in the initial site-finding stage to define a site. The default value of 15 site points should allow even relatively small sites to be detected. The average number of site points found for the tight-binding sites investigated in the calibration studies (see Chapter 5) is about 150.

  SiteMap always reports at least one site as long as that site contains at least 3 site points, which is the minimum number required to recognize a site. Thus, only the second and subsequent sites, ranked in order of the number of site points, must satisfy this threshold. Normally, this setting is not relevant when a molecule or entry is used to restrict the search region. However, SiteMap could find more than a single site even in this case if the region is large or contains subsites that SiteMap is unable to combine into a single site.

- Report up to *n* sites—Enter the maximum number of sites to report, ranked in order of decreasing size (number of site points). The default for this setting is to report up to the 5 largest sites found in the initial site-finding stage. This setting is not active when a ligand or other species is used to restrict the size and location of the mapping region.

- Use *type* definition of hydrophobicity—Choose the definition of hydrophobicity to use in the calculation. The choices are labeled more restrictive and less restrictive. The less restrictive definition corresponds closely to the definition used in the original version of SiteMap. The more restrictive definition eliminates points that are adjacent to defined hydrophilic regions and assigns reduced hydrophobicity to solvent-exposed regions.[1] See Section 2.2.2 on page 7 for more information. The more restrictive definition is the default choice.

- Use *type* grid—Choose the size of the grid to use in computing the displayed site maps. The choices are coarse, standard, and fine, corresponding to grid spacings of 1.0 Å, 0.7 Å, and 0.35 Å. The default is standard.

1. This definition favors enclosed phobic regions, like the Glide XP definition of hydrophobic enclosure. However, SiteMap's definition is not as restrictive as is the Glide XP definition, which has geometric elements that need not be met by the SiteMap definition.

It should be emphasized that the choice of grid increment has no effect on the site-finding or site-evaluation stages of the algorithm, which always position the site points on a 1-Å grid.

- Crop SiteMaps at *n* Å from nearest site point—Enter the distance from the nearest site point at which to crop the individual site maps for display in Maestro. The default is 4 Å. No data is lost when the map is cropped. This option merely affects the truncation of the displayed surface, and the distance can be altered during the Maestro session.
- Use the *type* force field—Choose the variant of the OPLS-AA force field to be used. The default is OPLS_2005; the alternative is OPLS_2001.
- Detect shallow binding sites—This option allows you to detect shallow binding sites that are suitable for protein-protein interactions, by modifying the `-enclosure` and `-maxvdw` settings (to 0.4 and 0.55).

### 4.1.3 Running the Job

To start the SiteMap job, click Start, make settings in the Start dialog box, and click Start in the dialog box. You can set the job name, the host, the user name, and the number of processors to use, and you can specify how to incorporate the results.

When a SiteMap job finishes, the site points and site maps are automatically incorporated into the current Maestro project, and associated with the receptor. The sites are returned in order of SiteScore, and each site is incorporated as a separate entry in the Project Table. By default, files that are not needed for incorporation into Maestro are removed. However, the file cleanup (and the SiteMap job) is aborted if a problem is found with an Impact job step, so that the log file is available for inspection.

If you choose not to incorporate a job, or you run a job from the command line, you can import the results using the Import panel. When you do so, ensure that you select the option Import associated data files.

If you want to run the job from the command line, or to modify any of the options, click Write. A dialog box opens, in which you can specify a job name. The command input file, *jobname* `.in`, is written to the current working directory. See Section 4.3 on page 26 for details.

### 4.1.4 Viewing the Results

When the results are incorporated, the Manage Surfaces panel is opened. You can use the controls in this panel to change the display attributes of the various maps. The accptr, donor, and phob maps are displayed by default; the phil and surf maps are not displayed by default.

The default appearance of the six map types is as follows:

- Hydrophobic map—yellow mesh
- Hydrophilic map—green mesh
- Hydrogen-bond donor map—blue mesh
- Hydrogen-bond acceptor map—red mesh
- Metal-binding map—pink mesh
- Surface map—gray surface, 50% transparency

Thus, a red ligand oxygen atom that accepts a hydrogen bond from the receptor or coordinates with a metal center should appear in a red acceptor or pink metal-binding region, and a polar hydrogen on a blue amide nitrogen should appear in a blue donor region. (Red and blue are the default colors for oxygen and nitrogen.)

You can change the cropping of any of the maps with the following procedure:

1. Select the map in the table.
2. Click Limit.
3. In the Limit dialog box, select Molecules from the Pick menu.
4. Pick a site point.
5. Enter the new distance from the site that the map will be cropped in the Distance text box.
6. Click OK.

**Note:** Do not change the molecular representation. Changing representations affects the display of the site points, and cannot be reversed without restarting Maestro.

You can change the value at which the maps are contoured (the isovalue), as follows:

1. Display the desired map, and undisplay all the others.
2. Adjust the Isovalue slider, or enter a value in the adjacent text box.

The other display properties of these surfaces, such as color or representation can also be changed. For more information, see Section 12.4 of the *Maestro User Manual*.

## 4.2 Sample Site Maps

To illustrate a typical application, Figure 4.3 and Figure 4.4 show the co-crystallized ligand for the thrombin 1ett receptor and the generated site points (white) in the context of the receptor structure and of the gray, translucent SiteMap surface. Figure 4.3 focuses on relatively exposed regions of the site, while Figure 4.4 profiles the buried specificity pocket.

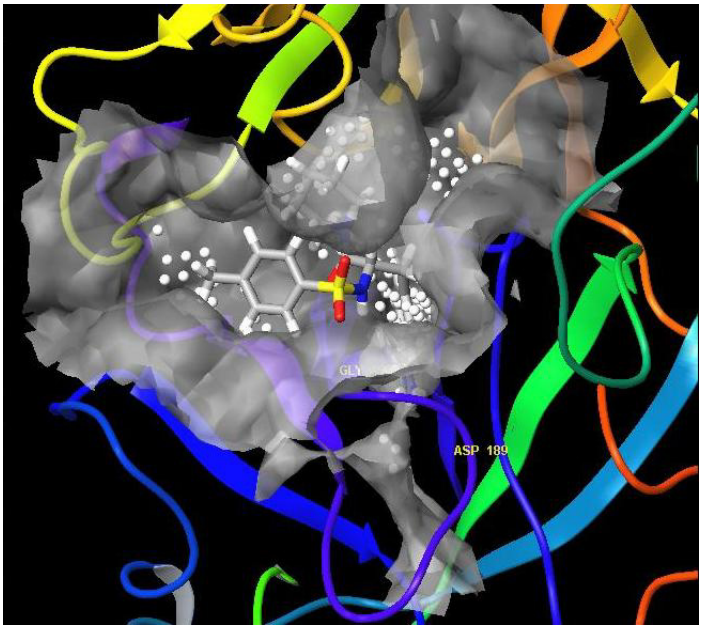


*Figure 4.3. SiteMap surface and site points for 1ett, exterior of pocket*

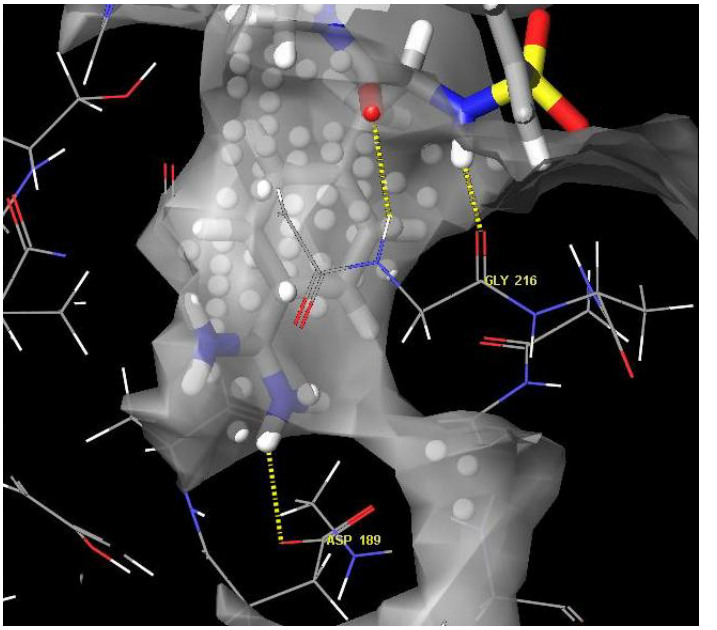


*Figure 4.4. SiteMap surface and site points for 1ett, inside pocket*

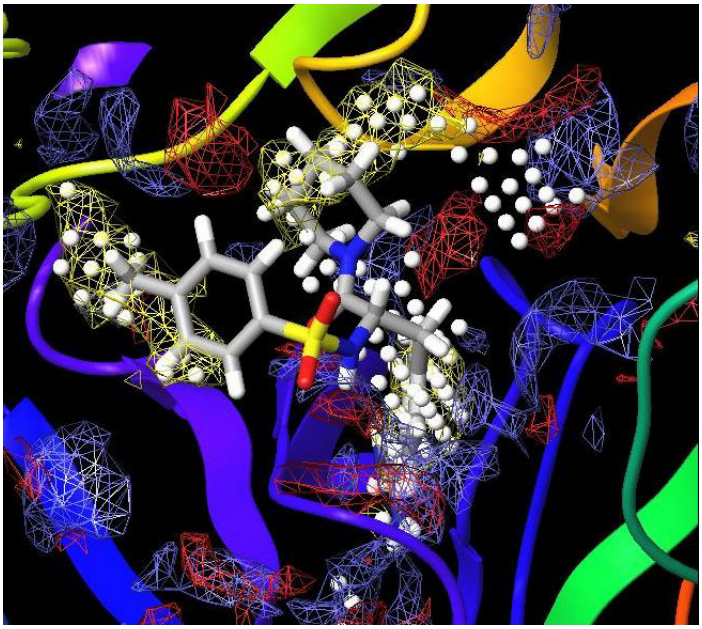

*Figure 4.5. Hydrophobic, donor, and acceptor maps for 1ett, exterior of pocket*

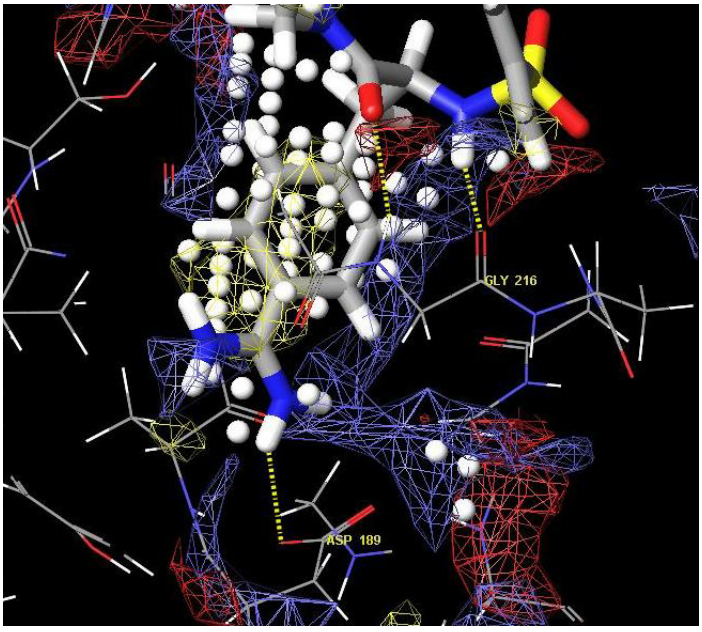


*Figure 4.6. Hydrophobic, donor, and acceptor maps for 1ett, inside pocket*

Figure 4.5 and Figure 4.6, taken from the same viewpoints, display the hydrophobic (yellow) and the hydrogen-bond donor (blue) and acceptor (red) maps, but for clarity suppress the receptor surface. The hydrophobic groups on the ligand can clearly be seen occupying hydrophobic regions, and the donors and acceptors of the ligand are located in or very close to the appropriate donor and acceptor regions.

## 4.3 Running SiteMap from the Command Line

SiteMap can be run from the command line with the `sitemap` command. The syntax of the `sitemap` command is as follows:

`$SCHRODINGER/sitemap [`*options*`] [-prot` *file*`.mae] [-i` *jobname*`.in]`

The principal arguments are described below.

| | |
|---|---|
| `-j[ob]` *jobname* | Specify the job name. *jobname* is used to make some file names unique. |
| `-prot` *file*`.mae` | Specify protein file in Maestro format, compressed or uncompressed. Required, either on the command line or in a command input file. |
| `-i` *jobname*`.in` | Specify the command options as keywords in a command input file. |

On Windows, you can run this command in a Schrödinger Command Prompt window, which you can open from the Start menu.

The job name is the determined by first of the following sources:

- the `-j[ob]` option
- the stem of the input file name as specified with `-i`
- the stem of the protein file name as specified with `-prot`

### 4.3.1 Command Options

The options for the `sitemap` command are given in Table 4.1. These options are classified into the common options, additional options, and advanced options. The common options include the ones that are set from the SiteMap panel in Maestro. The `sitemap` command also accepts the standard Job Control options, which are given in Section 2.3 of the *Job Control Guide*. You can use the `-HOST` option to distribute the job over multiple processors, for example.

The command options can be placed in a command input file as keyword-value pairs. The keywords are the command options without the `-` sign, and include the protein specification (`prot`). You can specify command options on the command line in addition to the command input file, and these command options override any settings in the command input file.

*Table 4.1. Options for the sitemap command.*

| Option | Description |
|---|---|
| `-compress {yes\|no}` | Compress the output Maestro files. Default: `yes`. |
| `-j[ob[name]]` *jobname* | Specify the job name. Default: stem of command input file, otherwise stem of protein input file. |
| `-ligmae` *filename* \| `-ligsdf` *filename* | Include optional ligand file in Maestro or SD format. |
| `-sitebox` *dist* | If an optional ligand file is specified, restrict the site-finding step to a box placed around the ligand plus a margin of *dist*. |
| `-reportsize` *size* | Minimum number of site points per reported site group in the first stage. The final number of site points may include extension points if `-extend` is used with a positive value or `-addsp yes` is given. Default: 15. |
| `-maxsites` *n* | Number of sites to report. Default: 5 |
| `-modphobic` *n* | Definition to be used for hydrophobic regions. Default: 3. Allowed values are:<br>0: Use the least restrictive definition, which is the original definition<br>1: Exclude site points that are exposed, close to hydrophilic regions, or too close to the protein<br>2: Scale phobic potential by exposure fraction<br>3: Exclude points and scale potential. |
| `-grid` *gridsize* | Grid size (Å) to be used in the mapping calculations. Default: 0.7, corresponding to `-resolution standard`. |
| `-resolution` *res* | Alternative means for specifying the grid size. Allowed values of *res* are `low` for 1.0 Å, `standard` for 0.7 Å or `high` for 0.35 Å. |
| `-mapdist` *dist* | Restrict displayed SiteMaps to *dist* from the nearest site point. Default: 4 Å. Can be overridden, up to the *margin* distance, by using the Limit facility in the Surface Table panel. |
| `-ffield` *version* | Version of OPLS force field to use. Allowed values are `OPLS_2001` and `OPLS_2005`. Default: `OPLS_2005`. |
| `-extend` *intdist* | When evaluating exposure in the site-evaluation stage, try adding "extension" site points at grid points up to ±*intdist* from the existing site points. Default: 3 Å. |
| `-addsp { yes\|no }` | Add "extension" site points to the site-point set if they fall in regions that satisfy the threshold criteria for phobicity or philicity. Default: `yes`. |
| `-cleanup { yes\|no }` | Remove all files not needed for the Maestro session. The file cleanup (and the SiteMap job) is aborted if a problem is found with an Impact job step, so that the log file is available for inspection. Grid files and potential files are not cleaned up. Default: `yes`. |

*Table 4.1. Options for the sitemap command. (Continued)*

| Option | Description |
|---|---|
| `-keepeval { yes \| no }` | Keep the evaluation-step log files even if `-cleanup yes` is given. This option can be used to obtain just the SiteScore and the SiteMap properties, which are listed at the end of the log files, while removing other files. Default: `no`. |
| `-keeplogs { yes \| no }` | Keep the log files from all stages even if `-cleanup yes` is given. Sets or resets `keepeval` to `yes`. This option is set to yes if the value of the `-verbosity` option is greater than 1. Default: `no`. |
| `-keepvolpts` | Keep the pdb-format files that contain the points that are summed to compute the volume property assigned for the site. |
| `-keepvdw` | Compute and keep grid-format (`.grd`) file of van der Waals potentials. |
| `-keepelec` | Compute and keep grid-format (`.grd`) file of electrostatic potentials. |
| `-writepot` | Write (x,y,z,potential) files that specify the phobic, philic, acceptor, donor, vdW, electrostatic, metal, and surface potentials computed by SiteMap. The file names are `phobic.smpot`, `philic.smpot`, `accptr.smpot`, `donor.smpot`, `vdw.smpot`, `elec.smpot`, `metal.smpot`, and `surface.smpot`. |
| *Additional options* | |
| `-enclosure` *fraction* | Fraction of ray directions from a grid point that must intersect the protein within a specified distance for that grid point to be considered as a potential site point. Maestro sets a value of 0.4 for shallow binding sites. Default: 0.5. |
| `-maxdist` *dist* | Distance within which a directional ray from a candidate grid point must intersect the protein surface. Default: 8 Å. |
| `-maxvdw` *vdw-energy* | Maximum van der Waals interaction energy (kcal/mol) for a grid point to be excluded as a potential site point. This quantity is the negative of the computed interaction energy, so the argument supplied must be positive. Maestro sets a value of 0.55 for shallow binding sites. Default: 1.1. |
| `-verbosity {0\|1\|2\|3}` | Control level of detail in output log files. Default: 0. |
| `-margin` *margin* | Grid-box margin (Å) to be used in the SiteMap calculations. Default: 6 Å. |
| `-addphob` *thresh* | Threshold for adding phobic points. Default: –0.50 if `-modphobic` option is 2 or 3, otherwise –0.75. |
| `-addphil` *thresh* | Threshold for adding philic points. Default: –8 kcal/mol |
| `-h` | Print brief usage summary. |
| `-help` | Print full usage summary. |

*Table 4.1. Options for the sitemap command. (Continued)*

| Option | Description |
|---|---|
| *Advanced options* | |
| `-dvscale` *ratio* | Squared distance ratio for determining whether a site point is inside or outside the protein. A site point is outside the protein if the ratio of the square of the protein-atom/site point distance to the protein-atom van der Waals radius is larger than this value for all protein atoms. Default: 2.5 Å$^2$. |
| `-nthresh` *n* | Minimum number of candidate site-point neighbors required to be within a given distance for a candidate site point to be eligible for inclusion in a site-point group. The square of the distance is specified by `-d2thresh`. Default: 3. |
| `-d2thresh` *value* | Squared distance used in `-nthresh` test. Default: 3.1 Å$^2$. |
| `-kmax` *n* | Maximum sum of differences in grid indices to nearest site point allowed to add a candidate site point to a site-point group. Default: 3. |
| `-kmax2` *n* | Maximum sum of squares of differences in grid indices to nearest site point allowed for a candidate site point to be added to a site-point group Default: 5. |
| `-mingroup` *n* | Minimum number of points in a site-point group required to constitute a site. Default: 3. |
| `-dthresh` *value* | Threshold for the minimum distance separating site points in two site-point groups for them to be considered for merging into a single group. If the minimum distance is larger than this value, the groups will not be merged. Default: 6.5 Å. |
| `-rthresh` *value* | Threshold for the ratio of the distance between the centroids of two site-point groups to the effective sizes of the groups for groups to be considered for merging. Default: 5. |
| `-r2thresh` *value* | Threshold for determining whether two site-point groups being considered for merging have successfully been interconnected by solvent-exposed bridging points. The squared distance from a bridging point between the groups to any site point must be less than *value* for the groups to be merged. Default: 4.1 Å$^2$. |
| `-cutoff` *cutdist* | Restrict the calculation of van der Waals and electrostatic potentials to protein atoms that lie within *cutdist* angstroms of a grid point. Default: 20 Å. |
| `-modvdw { yes \| no }` | Adjust van der Waals radii to improve the accuracy with which preferred locations of hydrogen donors and heavy-atom acceptors are represented in the site maps. Default: `yes`. |
| `-modcharges { yes \| no }` | Scale down formal-charge contributions to the protein partial atomic charges by 50%. Default: `yes`. |

### 4.3.2 SiteMap Output

The following output is generated by SiteMap:

- Site-point files (in compressed Maestro format) and map files (`.vis` and `.grd`). These files are read by Maestro to display the SiteMaps. For each site, a set of five or six map files is produced, containing the hydrophobic, hydrophilic, donor, acceptor, surface, and metal-binding maps. Sites are returned in order of SiteScore.
- Log files. These are removed by default when the job finishes successfully. The options `-keeplogs` and `-keepeval` can be used to control which log files are kept.

If you do not supply a project name with the `sitemap` command, the site maps are not automatically loaded into Maestro and displayed. To view the site maps in an existing Maestro session, you can import the results using the Import panel. When you do so, ensure that you select the option Import associated data files.

## 4.4 Adapting SiteMap

SiteMap provides control over many aspects of the calculations through command-line options, so that you can adapt SiteMap to the systems that you are interested in. The following subsections describe a number of different scenarios with the relevant command options.

### 4.4.1 Adjusting the Closeness of the Map to the Protein

The `-dvscale` option can be used to increase or decrease the number of grid points that lie outside the protein. The default value of 2.5 already allows fairly close approach to the protein. A value of 4 corresponds to the minimum-energy van der Waals distance for a homonuclear contact.

### 4.4.2 Tailoring the Definition of Hydrophobicity

Four `-modphobic` options are provided to tailor the definition of phobicity. Of these, Maestro employs the `-modphobic 3` ("more restrictive") and `-modphobic 0` ("less restrictive") options. It should be noted that the phobic-property score is calibrated such that the average phobic score and the average SiteScore for the previously discussed tight-binding complexes are 1.0, whichever phobic option is chosen. Thus, while the size and shape, and sometimes location, of displayed phobic regions will depend on the option selected, the average contribution of phobicity to the overall SiteScore for the tight-binding sites does not. What happens is that phobicity becomes more important for some binding sites and less important for others as the definition changes.

### 4.4.3 Finding Shallow Sites

If you are interested in finding relatively shallow sites, you can use the `-enclosure`, `-maxdist`, and `-maxvdw` options to make the site-finding step more receptive to finding relatively shallow sites. In particular, additional site-point positions will be recognized as valid candidates for inclusion in a site if some combination of the following is done:

- Make the threshold on the van der Waals energy for accepting points, set with `-maxvdw`, smaller than the default value of 1.1 kcal/mol.
- Make the maximum distance from a point to the protein, set with `-maxdist`, greater than the default value of 8 Å.
- Make the enclosure fraction, set with `-enclosure`, smaller than the default of 0.5.

You may need to explore these modifications of the default parameters to find sites appropriate for protein-protein interactions, for example.

### 4.4.4 Obtaining Only the SiteMap Properties

If you only want the values of the Maestro properties from the SiteMap calculation and not the maps, you can use the option `-keepeval yes`. The Maestro properties are listed in the final lines of the evaluation-step log files.

### 4.4.5 Specifying a Reference Ligand Without Restricting the Mapping Region

If a molecule or entry is used to restrict the mapping box, Maestro sets both the `-ligmae` and `-sitebox` options. From the command line, you can use `-ligmae` (or `-ligsdf`) without using `-sitebox` to specify a species that is to serve as a positional reference without restricting the region to be mapped. The refdist and refmin properties are generated in this case. This approach makes it easy to determine which receptor site, if any, corresponds to the reference site.

### 4.4.6 Controlling the Merging of Sites

You can control how sites are merged when they are separated by a short distance that could plausibly be bridged by a ligand in a solvent-exposed region. There are two measures of how close two groups must be to be considered for merging: the distance between the nearest site points in the two groups, and the ratio of the distance between the centroids of the groups to their effective size. The threshold for the first is set by `-dthresh`, and the threshold for the second is set by `-rthresh`. Both must be satisfied.

In addition to these two measures, there is a measure of whether two site-point groups being considered for merging have successfully been interconnected by solvent-exposed bridging points. The threshold for this measure is set by `-r2thresh`. When two groups are considered for merging, a bridging point is "grown" from one group. If it is sufficiently close to the nearest point in the other group, the groups are merged. The threshold is the minimum squared distance between the bridging point and the nearest point in the second group.

# SiteMap Results

To calibrate and characterize the SiteMap properties, SiteMap has been applied to an extensive set of 230 proteins, which were taken either from the Glide database-enrichment suite or from the PDBbind database [6]. These proteins bind ligands of molecular weight at least 150 with affinities of at least 100 µM. Of the 230 proteins, 155 have binding affinities of 1 µM or less. The proteins were prepared using standard Schrödinger techniques. To avoid prejudicing the search, all crystallographic water was removed.

The entire data set was used to optimize the contributions to the overall SiteScore of the SiteMap properties described in Section 2.3. The criterion for the optimization was that the site with the best SiteScore corresponded to the co-crystallized site as often as possible. The tight-binding set was further used to calibrate SiteScore and its contact, phobic, and philic components so that the average value for each of these quantities is 1.0. The most significant terms are the size of the site as measured by the number of site points, the relative openness of the site as measured by the exposure and enclosure properties, and the tightness of the site as measured by the contact property. The phobicity of the site plays a smaller role, and the site philicity plays a small enough role that it could have been excluded.

Table 5.1 summarizes SiteMap's accuracy in locating the primary (co-crystallized) binding site for the 230 proteins and for the 155 submicromolar binders. As Nayal and Honig [1] find for Screen and report for other methods, size is a fairly good predictor of the ligand-binding site. However, SiteScore is a better predictor, correctly locating the primary binding site in 96.5% of the proteins in the full set and 98.1% in the tight-binding set.

*Table 5.1. Performance in Locating the Primary Binding Site in Proteins*

| | 230 Proteins | | 155 Tight Binders | |
|---|---|---|---|---|
| **Comparison** | **Number** | **Percent** | **Number** | **Percent** |
| Primary site not found | 0 | 0.0 | 0 | 0.0 |
| Largest site scores best | 203 | 88.3 | 139 | 89.7 |
| Largest site is correct | 201 | 87.4 | 139 | 89.7 |
| Best-scoring site is correct | 222 | 96.5 | 152 | 98.1 |
| Largest or best-scoring site is correct | 224 | 97.4 | 153 | 98.7 |

SiteMap can also be employed as a "classifier" to discriminate sites that bind ligands from sites that don't. The objective is to determine whether a protein is likely to bind ligands tightly, not to decide which site in the protein to target. SiteMap can be used in this way by setting a threshold SiteScore value for recognition as a drug-binding site of 0.80 (80% of the average found for the 155 submicromolar sites). Used as a classifier, SiteMap performs as shown in Table 5.2. Similar results for the percentage of primary binding sites correctly classified (true positives) were reported for a different set of proteins by Nayal and Honig [1].

*Table 5.2. Performance of SiteScore Threshold in Classifying Primary Binding Sites in Proteins*

| | 230 Proteins | | 155 Tight Binders | |
|---|---|---|---|---|
| **Comparison** | **Number** | **Percent** | **Number** | **Percent** |
| Primary site not found | 0 | 0.0 | 0 | 0.0 |
| Primary site incorrectly classified | 24 | 10.4 | 15 | 9.7 |
| Primary site correctly classified | 206 | 89.6 | 140 | 90.3 |

For a more recent and more extensive set of tests, see Ref. 8.

# References


1. Nayal, M.; Honig, B. On the nature of cavities on protein surfaces: Application to the identification of drug-binding sites. *Proteins* **2006**, *63*, 892.

2. Halgren, T. A. New Method for Fast and Accurate Binding-site Identification and Analysis, *Chem. Biol. Drug Des.*, **2007**, *69*, 146.

3. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D. T. Extra Precision Glide: Docking and Scoring Based on a New Theory of Molecular Recognition. *J. Med. Chem.* **2006**, *49*, 6177.

4. Weber, A.; Halgren, T. A. et al. Design and Synthesis of P2-P1'-Linked Macrocyclic Human Renin Inhibitors. *J. Med. Chem.* **1991**, *34*, 2692.

5. Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849.

6. Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111.

7. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739.

8. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability, *J. Chem. Info. Model.*, **2009**, *49*, 377.

9. Therien, M.; Skorey, K.; Zamboni, R.; Li, C. S.; Lau, C. K.; LeRiche, T.; Truong, V. L.; Waddleton, D,; Ramachandran, C. Synthesis of a novel peptidic photoaffinity probe for the PTP-1B enzyme. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2319.

# 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
- SiteMap purchaser (company, research institution, or individual)
- Primary SiteMap 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`.

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®