

# 目 录

<b>一、 Discovery Studio 的基本操作.....</b>	<b>1</b>
<b>二、 Discovery Studio4.5 中的分子对接专题 .....</b>	<b>9</b>
蛋白活性位点的定义.....	10
Libdock——最快的分子对接技术 .....	13
LigandFit——基于几何形状匹配的快速分子对接技术 .....	19
CDOCKER——精准的分子对接技术 .....	27
Flexible Docking——全柔性受体-配体对接技术 .....	33
对接结果的分析: Analyze Ligand Poses.....	39
<b>三、 Discovery Studio4.5 中的片段药物设计专题.....</b>	<b>45</b>
MCSS.....	46
LUDI.....	50
Grow Scoffold.....	61
Replace Frament.....	68
<b>四、 Discovery Studio4.5 中的药效团专题.....</b>	<b>77</b>
构建基于分子共同特征的药效团模型 (HipHop) .....	78
构建具有活性预测能力的药效团 (Hypogen) .....	88
构建基于受体结构的药效团模型 (SBP) .....	100
构建及验证基于受体-配体复合物药效团 .....	110
<b>五、 Discovery Studio4.5 中的 QSAR/SAR 专题.....</b>	<b>119</b>

DS 中的定量构效关系分析: 2D-QSAR .....	120
DS 中的定量构效关系分析: 3D-QSAR .....	130
基于 MMP 分析 Activity Cliff.....	136
基于 MMP 分析构效关系.....	139
<b>六、 Discovery Studio4.5 中的虚拟组合库设计与分析专题 .....</b>	<b>143</b>
<b>七、 Discovery Studio4.5 中的化合物 ADMET 性质预测专题.....</b>	<b>151</b>
ADMET.....	152
TOPKAT .....	157
<b>八、 Discovery Studio4.5 中蛋白质序列的分析.....</b>	<b>163</b>
<b>九、 Discovery Studio4.5 中的蛋白质三维结构预测专题 .....</b>	<b>169</b>
球蛋白三维结构预测.....	170
<b>跨膜蛋白三维结构预测.....</b>	<b>185</b>
抗体三维结构预测.....	193
<b>十、 Discovery Studio4.5 中的蛋白质对接专题.....</b>	<b>201</b>
<b>十一、 Discovery Studio4.5 中的分子力学 / 分子动力学专题 .....</b>	<b>213</b>
<b>十二、 Discovery Studio4.5 中的分子力学/量子力学专题.....</b>	<b>225</b>
<b>十三、 Discovery Studio4.5 中的蛋白质设计专题.....</b>	<b>231</b>
虚拟氨基酸突变.....	232
蛋白质二硫键的预测.....	244
蛋白质自聚集预测.....	247

# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

**Discovery Studio 基本操作**

## Discovery Studio 基本操作

**目的：** 通过此教程，了解并掌握 Discovery Studio 中一些基本操作。

**所需功能和模块：** Discovery Studio Visualizer client

**所需数据文件：** pk-400.sd 和 1aq1.pdb。

**所需时间：** 30 分钟

### 介绍

在使用软件进行课题研究前，我们首先应该了解并掌握该软件使用的一些基本操作。为后续的体系处理做好准备工作。这个教程包括：

- 小分子配体准备
- 蛋白文件的处理

### 小分子配体准备

在 Discovery Studio (DS) 中，可以直接构建分子结构，也可以将在其它画图软件中画好的结构直接拷贝到 DS 中，本教程演示如何在 DS 中构建小分子结构。

#### 1. 调用 Sketching 功能

从 View 菜单下，打开 Toolbars，选择 Sketching。

Toolbars 中将显示各种 Sketching 的工具，这些工具可以用来构建化合物的初始结构。

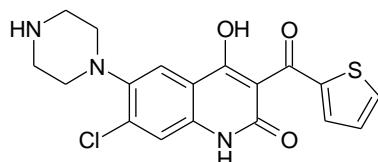


#### 2. 利用 Sketching 构建化合物的 3D 空间构象

打开一个分子显示窗口 (Molecule Window)，菜单栏 File|New|Molecule Window。

注：DS 中有四种窗口模式，包括 Molecule window (显示分子结构)，Protein Sequence Window (显示蛋白序列)，Nucleotide Sequence Window (显示核酸序列)，Script Window (显示脚本语言)，因此我们需要根据载入的文件类型选择窗口。

DS 中构建化合物的 3D 空间构象非常容易，也非常灵活。本教程以以下化合物为示例，以图示的方法演示如何构建化合物的结构。



选择 ，在窗口中画出结构 1 (图 1)。

点击 (可以通过菜单栏 View|Toolbars|View 调出) 将其选中，然后选择菜单栏 Chemistry|Bond|Aromatic 得到结构 2。

选择 ，鼠标指于芳环单键处并单击，构建稠环结构 3。

选择 ，构建连接单键，再选择 ，鼠标指于 C 原子处并单击构建环状结构，最后得到结构 4。

选择 和 构建单键和环状结构，选择 再次点击相应的键就可以构建双键结构，最终可得到结构 5。

更换元素类型， 选中某个碳原子，选择菜单栏 Chemistry|Element 更换相应元素即可，最后得结构 6。

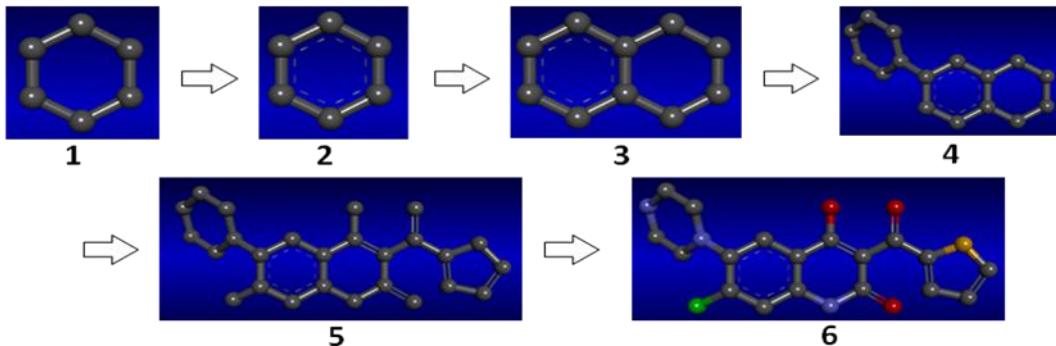


图 1 分子结构构建过程

选择菜单栏 Structure|Clean Geometry 可进一步优化此化合物的几何三维结构。

### 3. 使用 protocol 处理多个配体结构

如果配体数目较多，还要考虑对映异构体、pH 等因素时，可用流程浏览器（Protocol Explorer）中的 Prepare Ligands 处理该体系，通过此操作不仅可以产生三维结构，加氢，还可以产生异构体；此外也可以选择用 Linpiski's rules of five 作为筛选条件。

在文件浏览器（Files Explorer）中，展开 Samples|Tutorials|QSAR 文件夹，双击 pk-400.sdf 文件，默认以表格浏览器（Data Table View）形式打开。

选择前 10 行（即前 10 个分子），按下 **CTRL+C**。

点击 打开一个新的窗口，在新窗口中按下 **CTRL+V**，将前 10 个分子粘贴至新窗口中。

将相应分子的表格属性中 visible 前勾选上，并按下 **CTRL+G** 即可观看分子结构(图 2)。

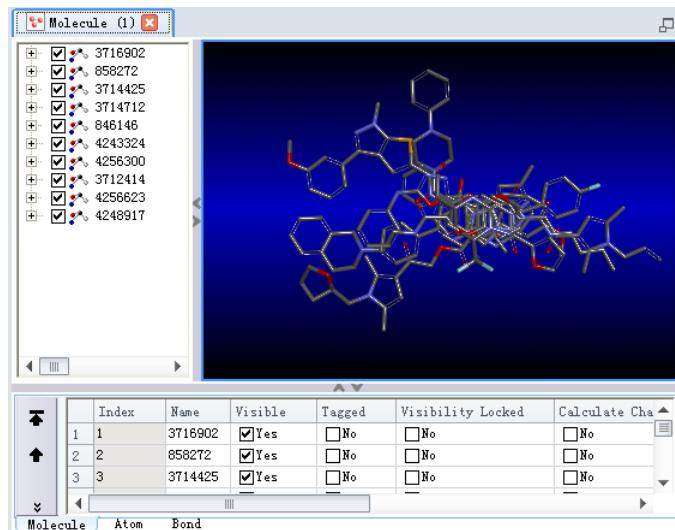


图 2 pk-10.sd 的输入分子结构

在工具浏览器（Tools Explorer）中，展开 Small Molecules|Prepare or Filter Ligands，点击 Prepare Ligands，打开相应的流程参数面板。  
设置 Input Ligands 参数为 Molecule:All。  
其余参数设为默认值。（图 3）

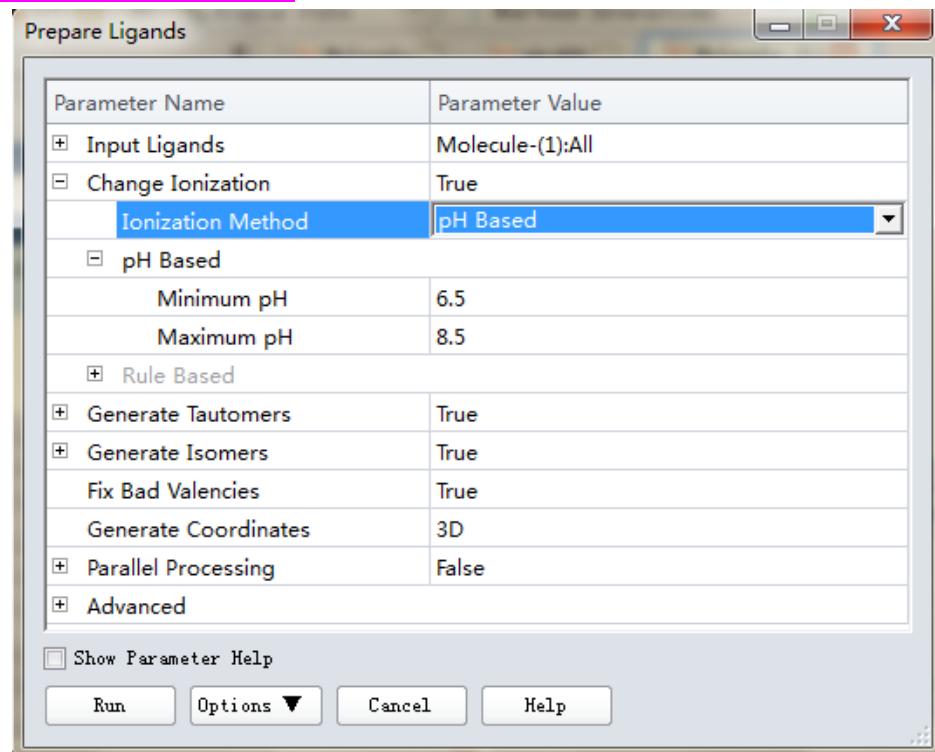


图 3 “Prepare Ligands”流程参数设置

点击 Run 运行作业，等待作业完成。

任务运行完毕，跳出如下提示框。

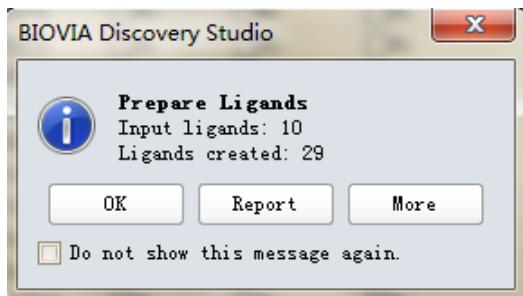


图 4 输出的结果提示框

输入的 10 个结构，经过处理后共产生 29 个配体分子。

点击 **Report**，打开 report 页面。

点击 **View Results** 观察处理后的分子结构。

从菜单中选择 **Window | Close All**，关闭所有窗口。

注：DS 中也可基于一定的力场对小分子结构进行优化，展开工具栏中 **Small Molecules|Minimize Ligands**，点击 **Full Minimization** 即可批量优化小分子结构。

## 蛋白文件的处理

### 1. 蛋白结构的载入

介绍三种蛋白结构的载入方法。

#### 1.1 可通过 DS 直接从 PDB 库中下载蛋白结构

点击菜单栏 **File|Open URL** 可出现图 5 对话框，直接输入蛋白 ID 号。

如果机器联网，即可直接下载到视窗中。

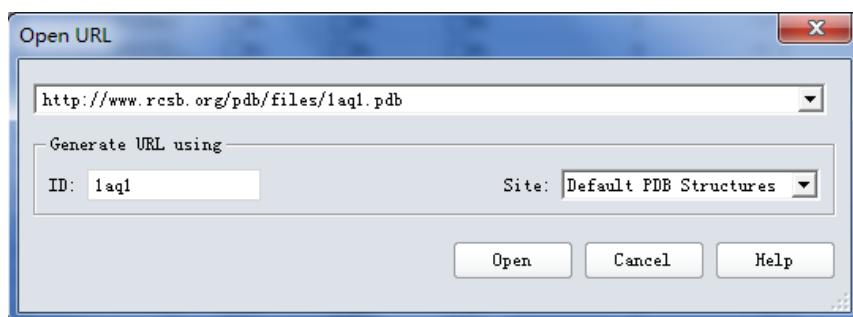


图 5 “Open URL”

#### 1.2 可通过流程浏览器中的 RCSB Structure Search 模糊查询所有符合条件的蛋白

在工具浏览器（Tools Explorer）中，展开 **Macromolecules|Query Online Databases|Search Databases**，点击 **RCSB Structure Search** 按钮，流程对应参数打开（图 6）。

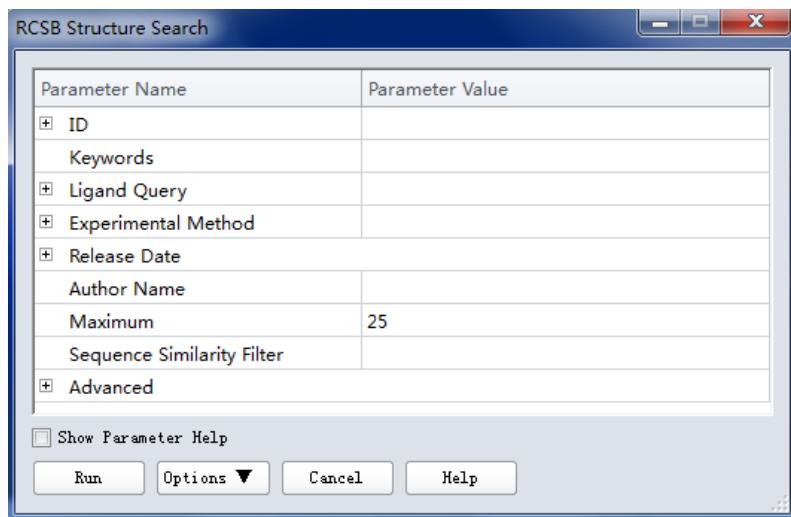


图 6 “RCSB Structure Search”流程参数设置

可通过 ID 号，配体结构信息，实验方法，作者名称等信息查询，可按需要进行检索。

### 1.3 从 File 中直接打开本地文件

#### 2. 蛋白文件处理

本例采用第三种载入蛋白的方式，在文件浏览器（Files Explorer）中，展开 Samples|Tutorials|Receptor-Ligand Interactions 文件夹，找到并双击打开 1aq1.pdb。

**Ctrl+H** 打开系统视图，**Ctrl+T** 打开表格视图，从左边的系统视图中选中水分子 Water（图 7），点击键盘 **Delete** 删除。

在工具浏览器（Tools Explorer）中，展开 Macromolecules|Prepare Protein，点击 Clean Protein。

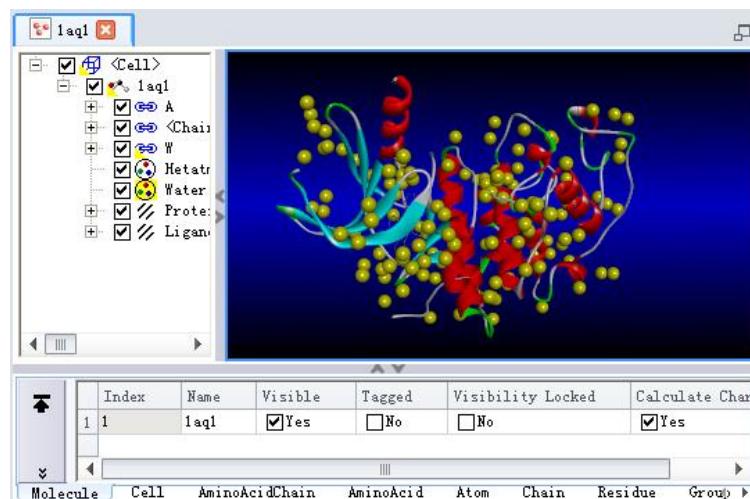


图 7 1aq1 结构

可去除蛋白多构象，补充非完整的氨基酸残基，为蛋白加氢等。

注：Clean protein 的具体参数设置可选择菜单栏 **Edit>Preference>Protein Utilities>Clean Protein** 进行设置（图 8）。

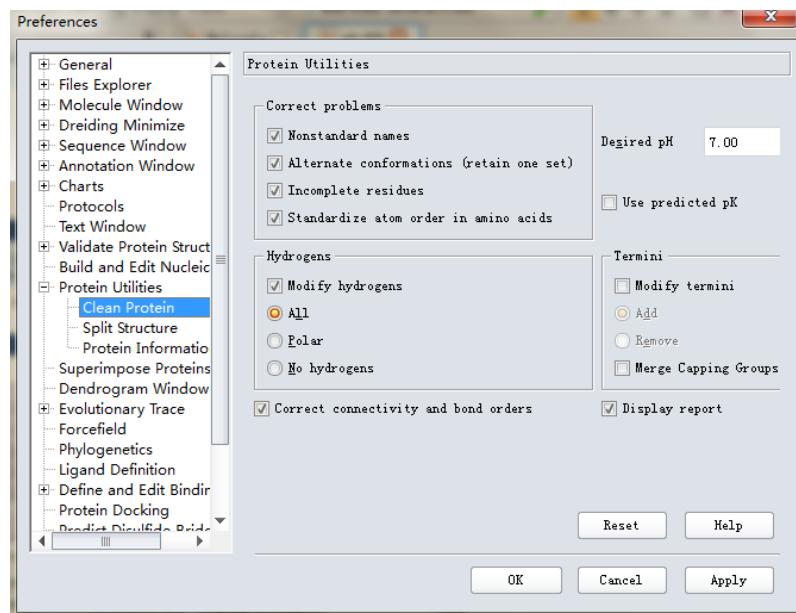


图 8 Preferences 对话框

经过处理的蛋白，可进行后续的对接等操作，也可对其显示方式作改变。

在蛋白分子窗口中点击鼠标右键选择 **Display Style**，**Atom** 一栏选择 **Line**，**Protein** 一栏选择 **Off**，其它参数默认。（图 9）

蛋白即可按照每个氨基酸残基以线状显示。（图 10）

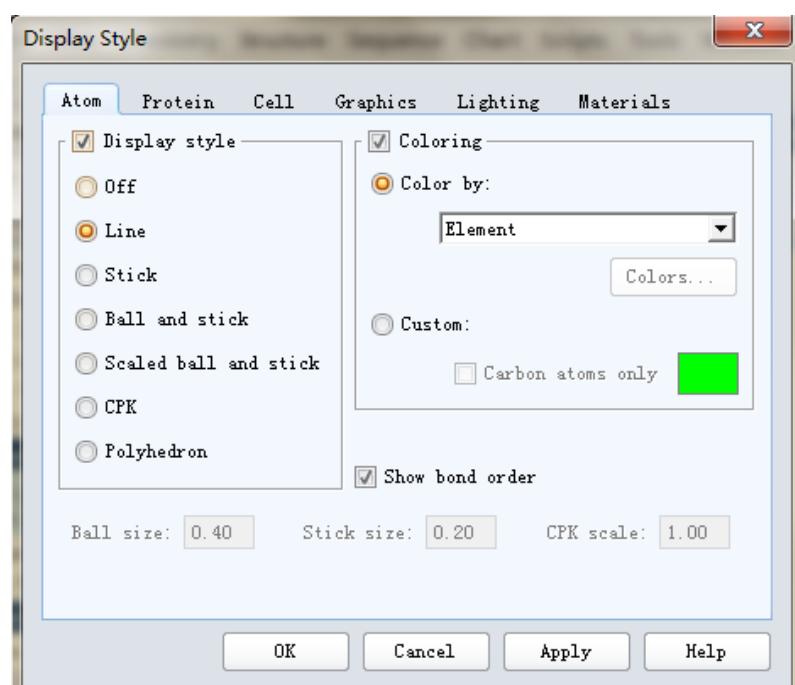


图 9 Display Style 对话框

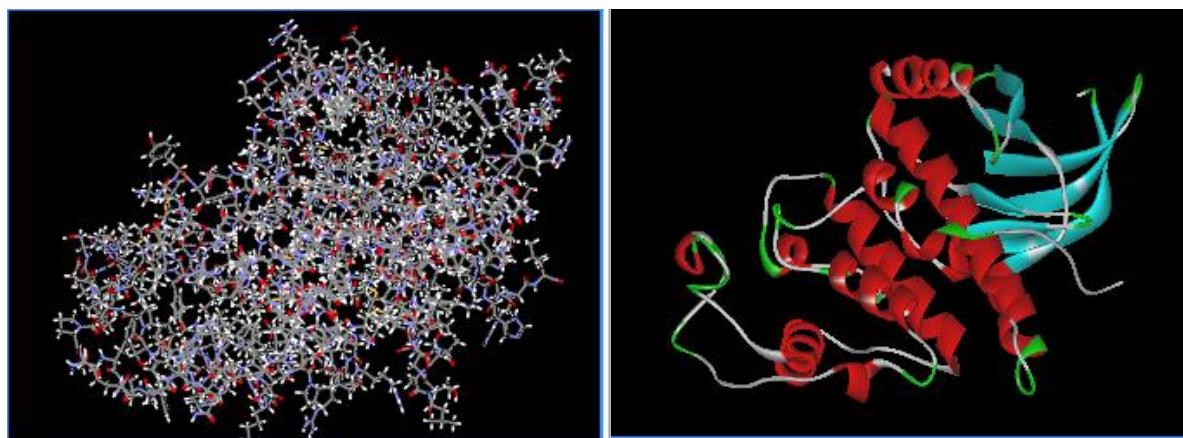


图 10 经过处理的蛋白

注：此外，也可以通过展开工具栏 Macromolecules|Prepare Protein，点击 Prepare Protein 对蛋白进行预处理，包括在不同 pH 条件下的质子化状态，缺失 loop 的补充等。（图 11）

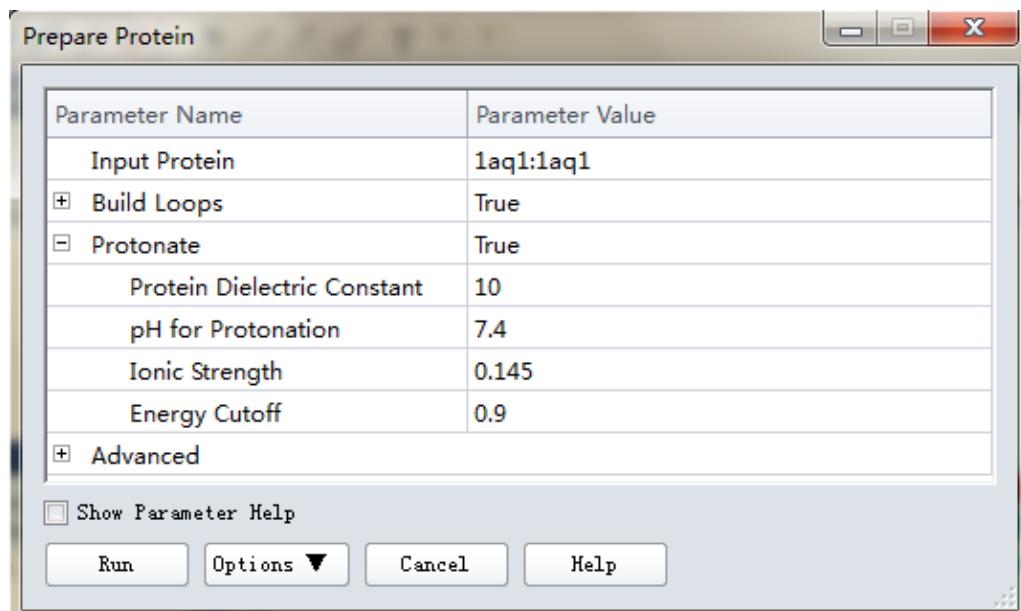


图 11 “Prepare Protein”流程参数设置

# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

分子对接专题:

**LibDock、LigandFit、CDOCKER、Flexible Docking**

## Discovery Studio Preparing Receptor Binding Site 教程

**目的:** 采用 Discovery Studio, 以一个蛋白为实例, 示范定义受体结合位点的操作过程。

**所需功能和模块:** Discovery Studio Visualizer client

**所需数据文件:** 1kim.pdb

**所需时间:** 10 分钟

### 介绍

受体结合位点的定义对于分子对接和打分都非常重要。寻找结合位点有两种方法: 在受体空腔中寻找结合位点和在指定位置寻找结合位点。

本教程中, 会介绍如何寻找并定义结合位点。具体包括:

- 基于蛋白空腔定义结合位点
- 基于指定位点定义结合位点

### 准备体系

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | 1kim.pdb。

在分子窗口中将打开一个蛋白质三维结构, 该蛋白已经过一定的预处理。

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 点击 Define Receptor, 将选择的蛋白分子 1kim 定义为受体分子。 (图 1)

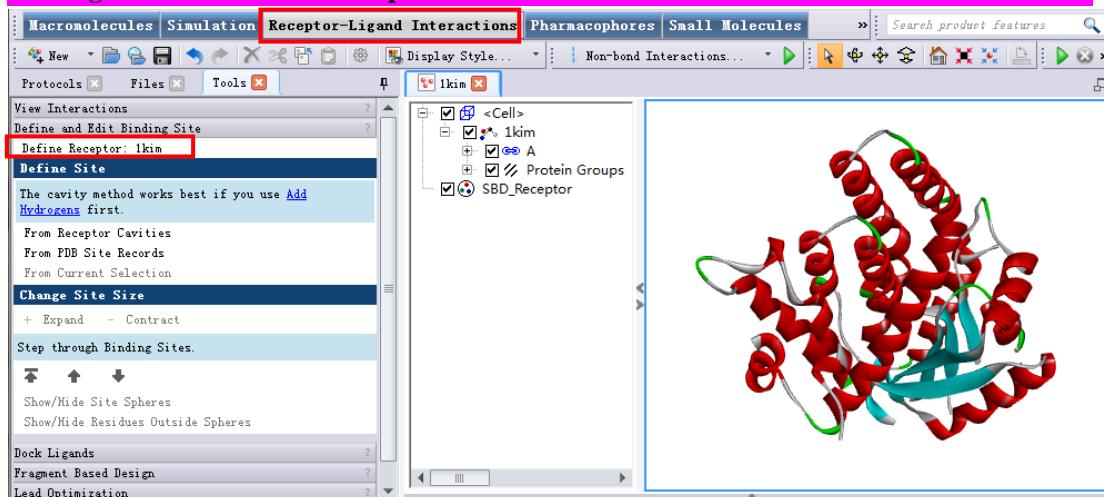


图 1 定义蛋白为受体

### 在受体空腔中寻找结合位点

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 点击 From Receptor Cavities。

通过寻找的空腔来定义受体中可能的结合部位。

在系统视图 (Ctrl+H) 中展开 1kim。

可以看到识别出 9 个可能的结合位点 (Site1-9), 最大的可能的结合位点被展示在图形视图中 (图 2)。

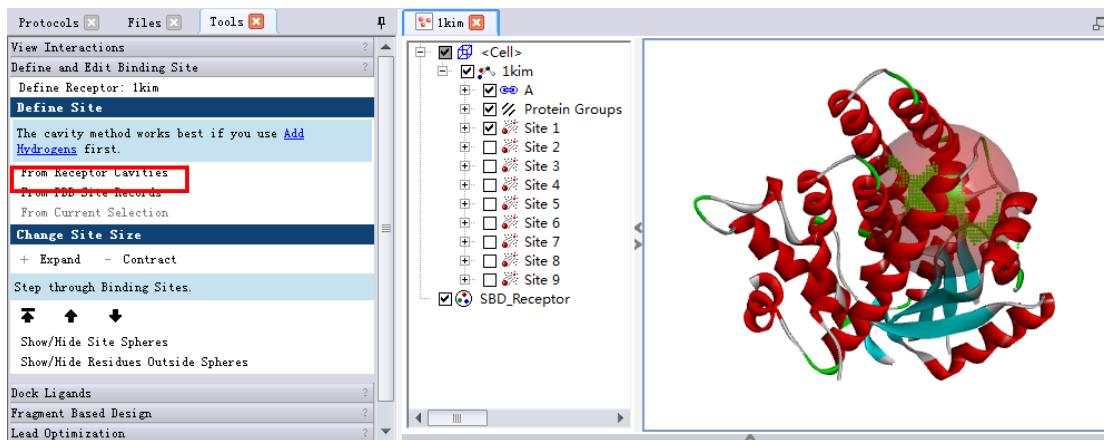


图 2 从受体中寻找结合位点

每一个找到的结合位点都由绿色的点 (Binding Site) 和一个红色的球 (SBD\_Site\_Sphere) 构成，绿色的点显示了空腔所在的位置和空腔形状，并被红色的球体所囊括。

红色的球体为 SBD\_Site\_Sphere，可通过 **Receptor-Ligand Interactions |Define and Edit Binding Site** 工具面板下的 **Show/Hide Site Spheres** 选择显示与否。

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interactions|Define and Edit**

**Step through Binding Sites.**

**Binding Site**，点击 **↑ ↓** 按钮可以观察不同的 Site。

此外，可以通过 **Receptor-Ligand Interactions |Define and Edit Binding Site** 工具面板下，**Expand** 或 **Contract** 按钮来调大或缩小 binding site 和 SBD Site Sphere 的大小。

### 在指定位置寻找结合位点

在系统视图 (Ctrl+H) 中展开 1kim，将上述步骤产生的 9 个位点全部选中，然后点击键盘上的 **Delete** 予以删除。

在系统视图 (Ctrl+H) 中展开 1kim 的链 A，选择残基 HIS58 和 VAL70。

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interactions|Define and Edit Binding Site**，点击 **From Current Selection**，在 HIS58 和 VAL70 处寻找可能的结合部位 (图 3)。

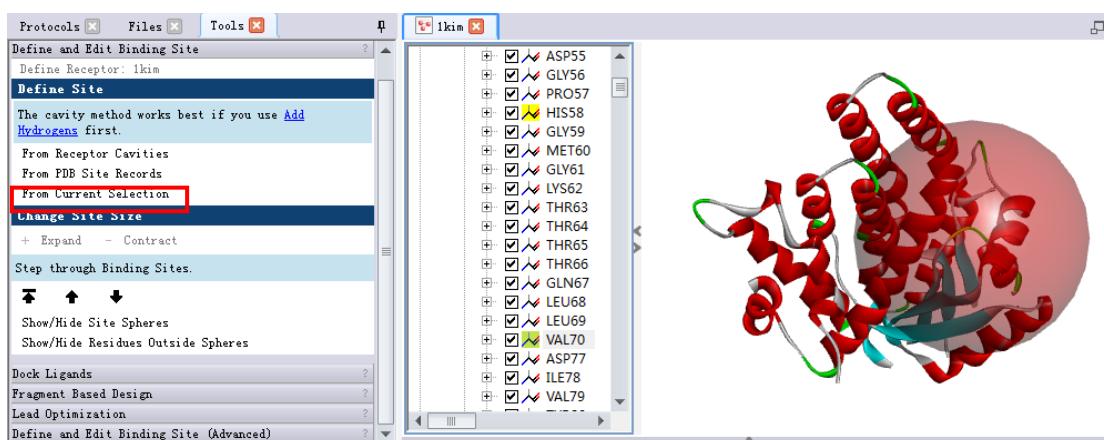


图 3 在指定位置寻找结合位点

注：该方法也可以选择任意原子或体系来进行结合位点的定义，如复合物晶体结构中自带的小分子配体等。

### 修改活性部位球体半径

单击选中 SBD\_Site\_Sphere 球体，点击鼠标右键选择 **Attributes of SBD\_Site\_Sphere...**，打开 **SphereObject Attributes** 对话框，在半径（Radius）选项输入设定的数值，点击 OK 按钮即可修改球体半径（图 4）。

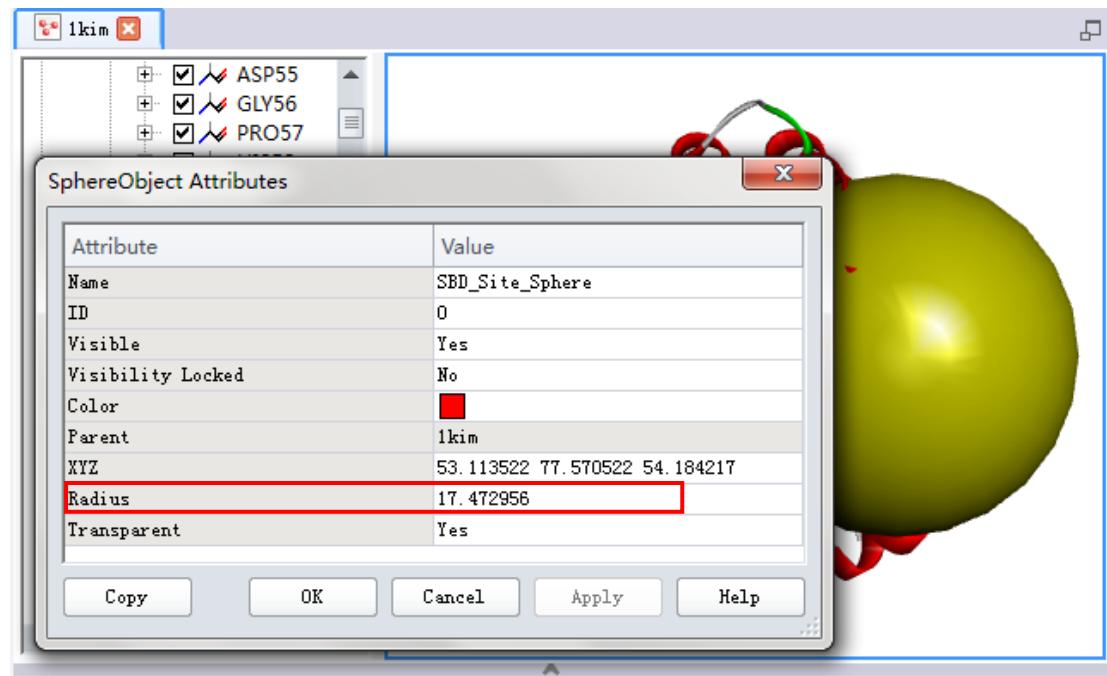


图 4 打开 Sphere Object Attributes 对话框，更改半径大小

## Discovery Studio LibDock 教程

### Libdock – 最快的分子对接技术

**目的:** 采用 Libdock, 以一组配体及一个明确活性位点的蛋白质为实例, 示范分子对接及结果分析的操作过程。

**所需功能和模块:** Discovery Studio Visualizer client, DS LibDock, DS Catalyst Conformation。

**所需数据文件:** 1kim\_prot.dsv 和 kinase\_ligands.sd。

**所需时间:** 15 分钟

#### 介绍

基于结构的药物设计技术在药物研发中起着非常重要的作用。在药物分子产生药效反应的过程中, 药物分子要与靶标相互结合, 首先需要两个分子充分接近, 采取合适的取向, 使两者在必要的部位相互契合, 发生相互作用, 继而通过适当的构象调整, 才能得到一个稳定的复合物构象。

基于结构的药物设计主要采用的是分子对接技术。分子对接就是把配体分子放在受体活性位点的位置, 然后按照几何互补、能量互补以及化学环境互补的原则来实时评价配体与受体相互作用的好坏, 并找到两个分子之间最佳的结合模式。分子对接是从整体上考虑配体与受体结合的效果, 能够较好避免其他方法中容易出现的局部作用较好而整体结合欠佳的情况。在药物设计中, 分子对接方法主要用来从小分子数据库中搜寻与受体生物大分子有较好亲和力的小分子, 进行药理测试, 从中发现新的先导化合物; 或者用于解释药靶之间的作用机制, 并在得到作用模式的基础上指导化合物结构改造。

LibDock 是 Discovery Studio 中的其中一种对接方法。该对接方法首先会针对受体活性位点计算得到热区图, 该热区图包含极性和非极性部分; 接着不同构象的配体分子分别刚性地叠合至热区图以形成比较合适的相互作用; 然后进行能量优化; 最后保留打分较高的对接构象。

本教程将采用 LibDock 将一组配体分子对接到胸苷激酶 (thymidine kinase) 中, 包括:

- 准备分子对接体系
- 执行分子对接计算
- 分析配体对接结果

#### 准备分子对接体系

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| 1kim\_prot.dsv 文件。

该蛋白将在一个新的分子窗口中出现, 该蛋白已经预处理过, 且活性位点也已定义好。

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 依次点击 Show/Hide Residues Outside Sphere 和 Show/Hide Sphere。

展开菜单栏 View|Transform, 点击 Fit To Screen 将蛋白结合位点居中显示。 (图 1)

以上操作可以将结合位点外的残基以及球体隐藏, 以便观察对接结果时更加便捷。

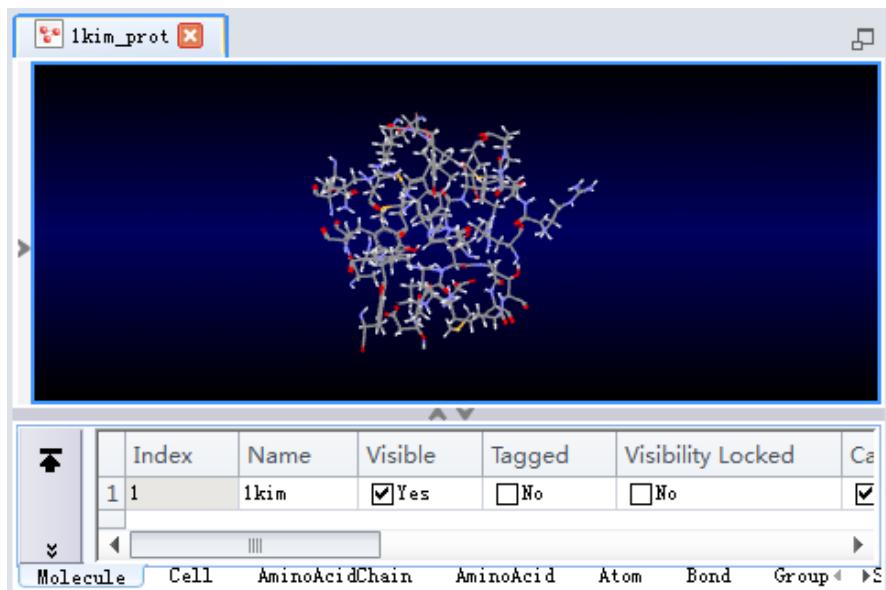


图 1 蛋白质三维结构示意图

在文件浏览器（Files Explorer）中，找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| kinase\_ligands.sd 文件（图 2）。

在新的分子窗口中打开一张共有 9 行的表格，每一行代表一个分子，共有 9 个配体分子。

	Index	Name	Visible	Tagged	Visibility Locked	Ligand Name
1	1	1e2k	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	1-[4-HYDROXY-5----]
2	2	1ki2	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	9- (1, 3-DIHYDROX...
3	3	1ki3	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	9- (4-HYDROXY-3----)
4	4	1ki6	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	1', 5' -ANHYDRO-2...
5	5	1ki7	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	5-IODODEOXYURID...
6	6	1kim	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	THYMIDINE
7	7	2ki5	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	9-HYDROXYETHOXYM...
8	8	1e2m	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	6-HYDROXYPROPYL...
9	9	1e2n	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	6-{ [4- OHYDROXYM...}

Molecule Atom Bond

图 2 以表格形式浏览待对接配体

### 执行分子对接计算

在工具浏览器（Tools Explorer）中，展开 Receptor-Ligand Interaction | Dock Ligands，点击 Dock Ligands (LibDock) 执行分子对接计算。

该模块相应参数在参数浏览器中打开。

在参数浏览器中，点击 Input Receptor 参数，从下拉列表中选择 1kim\_prot:1kim 设置受体蛋白。

点击 Input Ligands 参数，从下拉列表中选择 kinase\_ligands:All，指定对接配体。

点击 Input Site Sphere 参数，确保该参数设置为 48.6808, 82.5963, 54.0744, 9，如果不是该参数，则从下拉列表中选择该 sphere 的坐标及半径。

点击 Docking Preferences 参数，从下拉列表中选择 User Specified，根据需要改变对接计算的特定参数。

展开 Docking Preferences 参数，点击 Max Hits to Save 参数，输入值 10。

本教程中，修改默认设置以减少最终保留的对接构象，减少所用时间。（图 3）

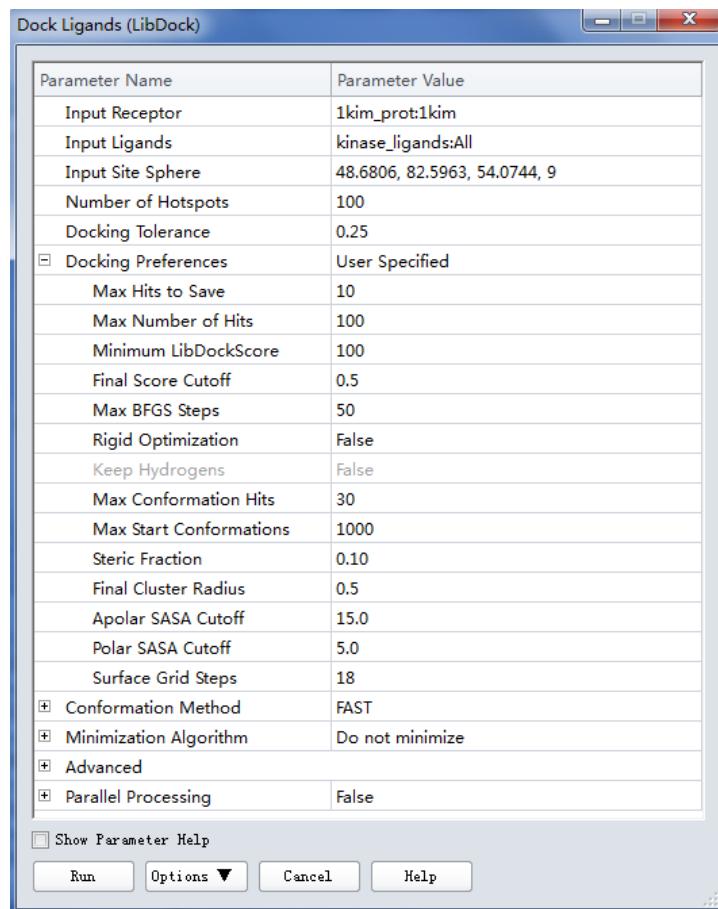


图 3 打开 Libdock 对接流程并设置参数

参数设置完毕后，点击 Run 运行作业。

## 分析分子对接结果

待作业完成后，对接结果会自动在一个新的窗口中打开，包含蛋白（只显示配体结合位点处残基）和所有对接构象。其中显示的蛋白已被锁在窗口中，当依次查看所有对接构象时该蛋白都可视。

（或者在作业浏览器（Jobs Explorer）中双击刚完成的分子对接作业，打开 Report 窗口，点击 View Results，同样可以打开对接结果。）

在打开的新的分子窗口 1kim 中确保系统视图（CTRL+H）和表格视图（CTRL+T）打开，在系统视图中可以观察到该窗口中包含了所有 9 个配体分子的对接结果，共 81 个。

点击表格视图中的 按钮，可观察排在第一位的对接构象，继而可通过点击表格视图中的

和 按钮（或者 CTRL+Up/CTRL+Down），观察配体分子的每个 pose 同受体分子的结合模式。

### 1. 非键相互作用的直观显示与分析

在工具浏览器（Tools Explorers）中，展开 Receptor-Ligand Interactions | View Interactions，点击 Ligand Interactions。

在视图窗口中，受体原子与配体对接 poses 间的非键相互作用会通过不同颜色的虚线显示出来，且只有参与了同配体之间相互作用的残基才会显示。（图 4）

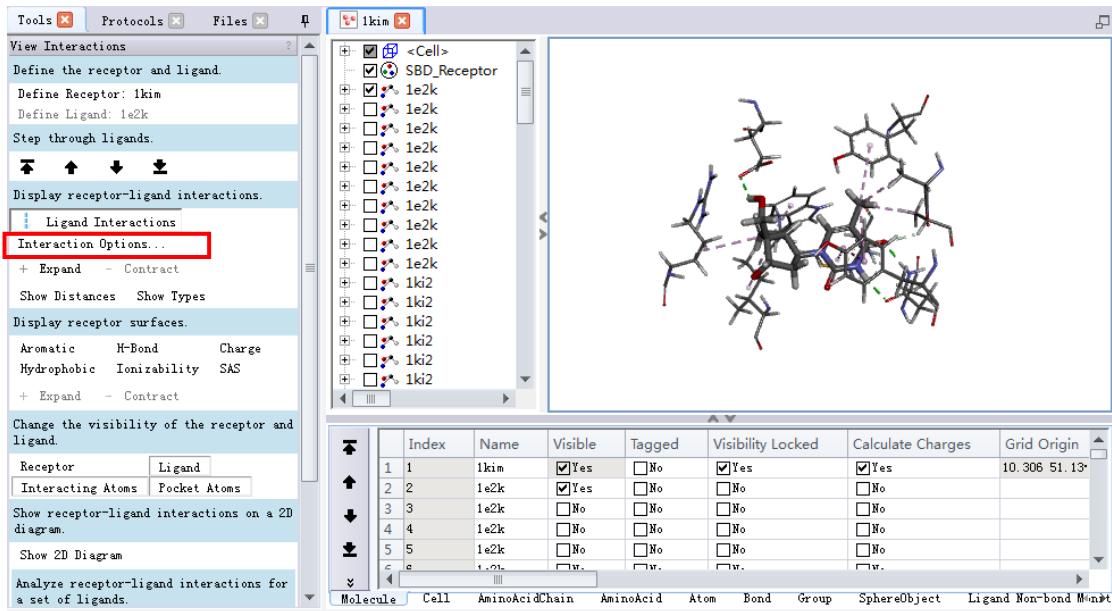


图 4 显示蛋白-配体之间非键相互作用

点击上述 View Interaction 工具面板下的 **Interaction Options...**，展开如下窗口（图 5）。该窗口中所列的非键相互作用类型即可以考虑的所有非键作用总类，其中黑色显示即在该蛋白和配体间存在的非键相互作用。

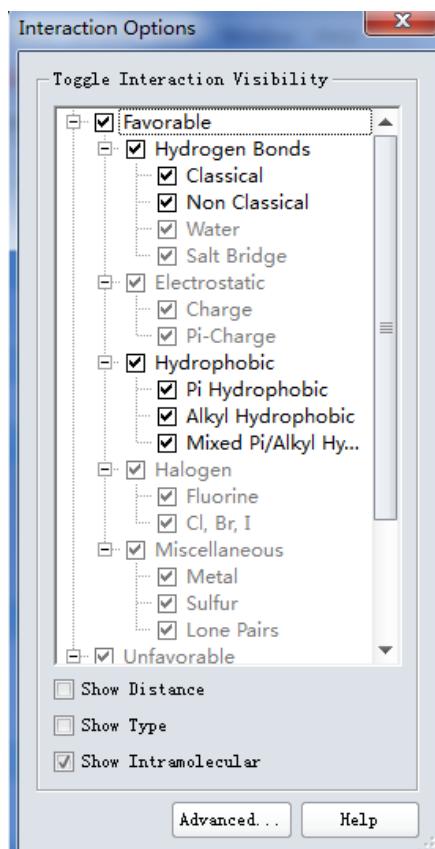


图 5

此外，在分子显示窗口任意选中某一虚线，在 DS 界面的左下方就会显示该非键作用类型及距离等相关信息。（图 6）

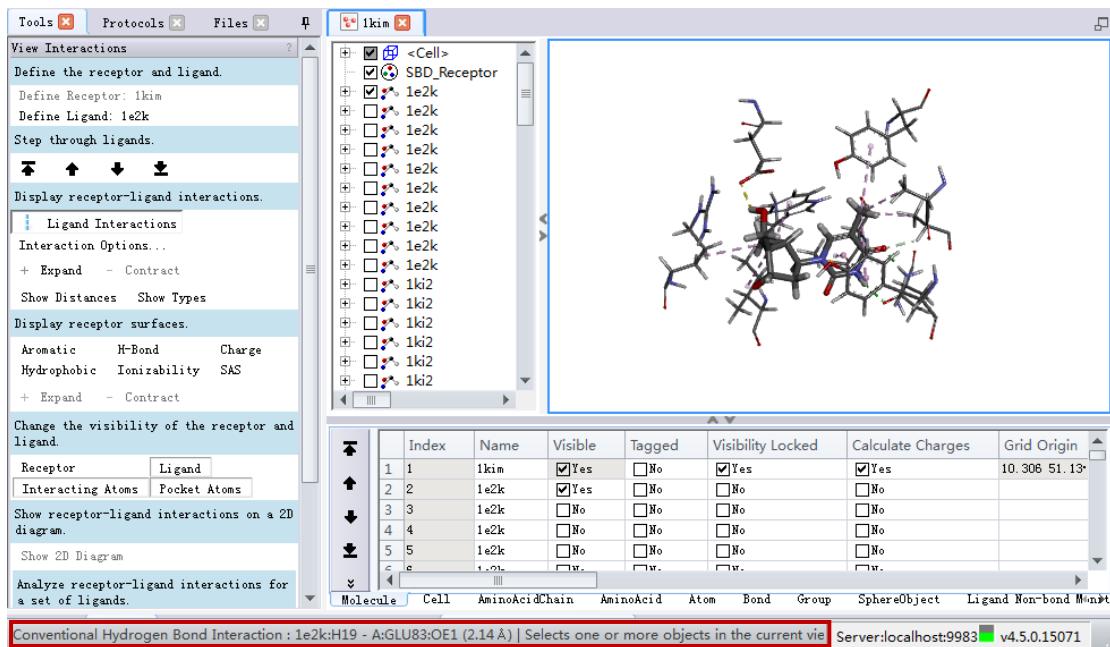


图 6

为了更好的观察受体分子与配体对接 pose 间的相互作用，可以对体系进行旋转以获得最佳的分析角度。

点击上述 View Interaction 工具面板下的 按钮可以观察不同的对接构象同蛋白之间的非键相互作用。

## 2. 生成配体-蛋白相互作用二维平面图

选中并显示要描述的配体（如第一个 1e2k），在工具浏览器（Tools Explorers）中，展开

**Receptor-Ligand Interactions |View Interactions**，点击 **Define Ligand**。

然后在该工具面板下，点击 **Show 2D Diagram**。

可以看到在一新窗口中打开配体-蛋白相互作用二维平面图，便于我们更直观的观察两者的相互作用及关键的氨基酸和基团。（图 7）

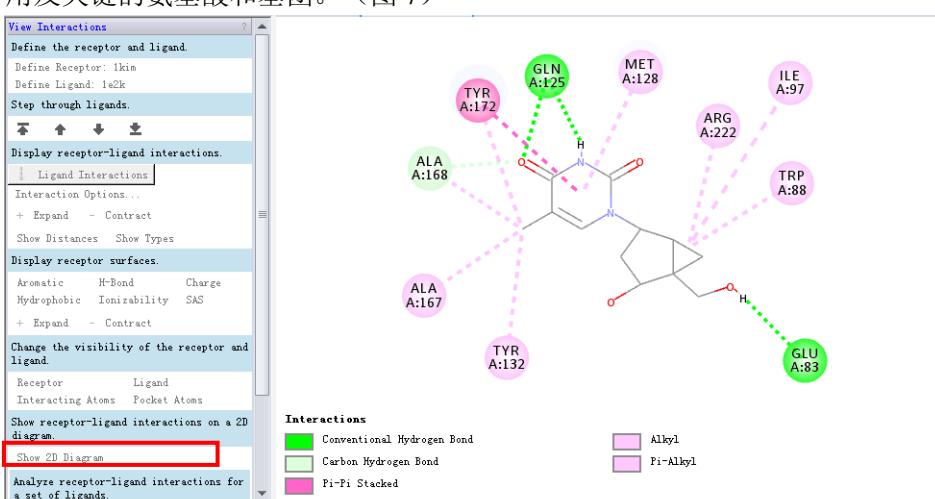


图 7 配体-蛋白相互作用二维平面图

得到的二维相互作用图中各种图形显示元素的详细定义可参考图 8:

Display	Description
	Interactions are shown as dashed lines between receptor residues and ligand atoms. If multiple ligand atoms are involved in an interaction, the line ends at the geometric center of those atoms.
	Residues involved in interactions are colored by interaction. See <a href="#">Nonbond interactions</a> for information about the interaction colors.
	Covalently bonded residues are represented by small discs.
	Neighboring residues are colored green.
	The solvent accessible surface of an interacting residue is represented by a halo around the residue. The diameter of the circle is proportional to the solvent accessible surface.
	The solvent accessible surface of an atom is represented by a blue halo around the atom. The diameter of the circle is proportional to the solvent accessible surface.

图 8 配体-蛋白相互作用二维图中各元素的定义

## Discovery Studio LigandFit 教程

### LigandFit – 基于几何形状匹配的快速分子对接技术

**目的:** 采用 LigandFit, 以一组配体及一个明确活性位点的蛋白质为实例, 示范分子对接及结果分析操作过程。

**所需功能和模块:** Discovery Studio Visualizer client, DS LigandFit, DS LigScore。

**所需数据文件:** 1kim.pdb 和 kinase\_ligands.sdf。

**所需时间:** 20 分钟

#### 介绍

基于结构的药物设计技术在药物研发中起着非常重要的作用。在药物分子产生药效反应的过程中, 药物分子与靶标相互结合, 首先就需要两个分子充分接近, 以合适的取向在特定的部位相互契合, 产生相互作用, 继而通过适当的构象调整, 得到一个稳定的复合物构象。

分子对接技术即基于结构的药物设计主要采用的手段, 该技术就是将配体分子置于受体分子活性位点的位置, 然后按照几何互补、能量互补以及化学环境互补的原则来实时评价配体与受体相互作用的好坏, 并找到两个分子之间最佳的结合模式。分子对接是从整体上考虑配体与受体结合的效果, 能比较好地避免其他方法中容易出现的局部作用较好而整体结合欠佳的情况。在药物设计中, 分子对接方法主要用来从小分子数据库中搜寻与受体生物大分子有较好亲和力的小分子, 并进行药理测试, 从而从中发现新的先导化合物。

Ligandfit 是应用精确的经典分子对接方法进行虚拟筛选的工具。它本身具备了受体分子活性位点的自动寻找和确认、构象柔性的多配体对接以及基于力场的相互作用打分评价的功能。LigScore 不仅包括 PLP、PMF 等在内的 8 种不同算法的打分函数, 还包括对所有打分函数评分结果的综合评价。

本教程将采用 LigandFit 将一组配体分子对接到胸苷激酶 (thymidine kinase) 中。

本教程包括:

- 准备分子对接体系, 执行分子对接计算
- 分析配体对接构象 (采用 2D 视图显示配体与受体的相互作用)

#### 准备分子对接体系, 执行分子对接计算

##### 1. 定义蛋白为受体分子

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| 1kim.pdb 文件。

该蛋白将在一个新的分子窗口中出现。

展开窗口左侧系统视图 (Hierarchy View) (可通过 CTRL+H 展开) (图 1), 在系统视图中点击选中 1kim。

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 点击 Define Receptor。

在系统视图中添加 SBD\_Receptor 一栏。

将前面选择的蛋白分子 1kim 定义为受体分子, 供下一步使用。

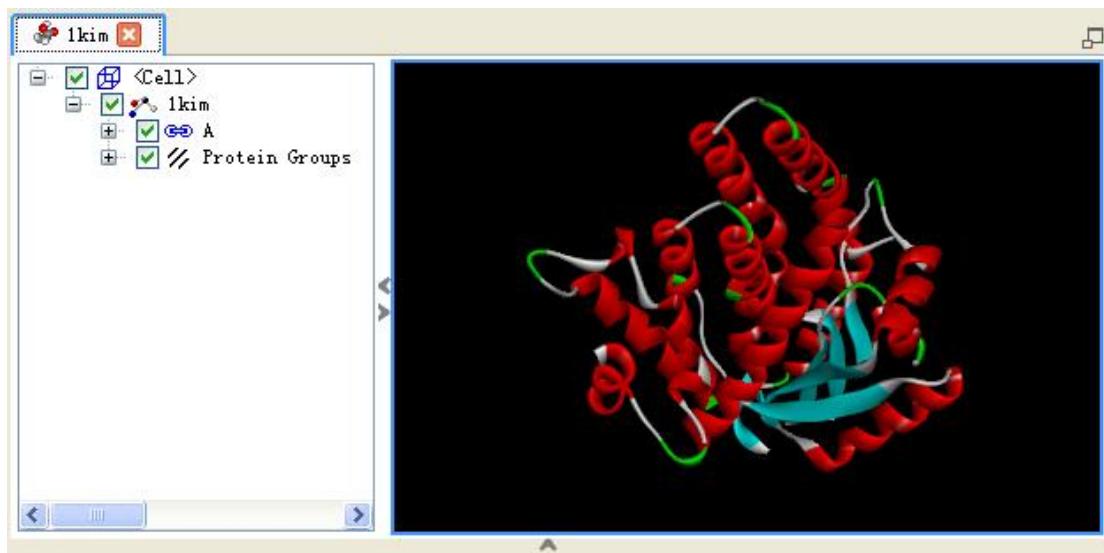


图 1 蛋白质三维结构示意图

## 2. 寻找受体中可能的结合区

如果晶体结构不包括 H 原子，也可在菜单栏中选择 Chemistry | Hydrogens | Add 加氢。

在工具浏览器（Tools Explorer）中，展开 Receptor-Ligand Interactions | Define and Edit Binding Site，在 Define Site 一栏下点击 From Receptor Cavities。

通过寻找受体中的空腔来寻找受体中可能的结合部位。

在系统视图中自动添加 9 个结合位点（Site1-9），即找到了 9 个可能的结合位点，最大的可能的结合位点则显示在图形视图中。（图 2）

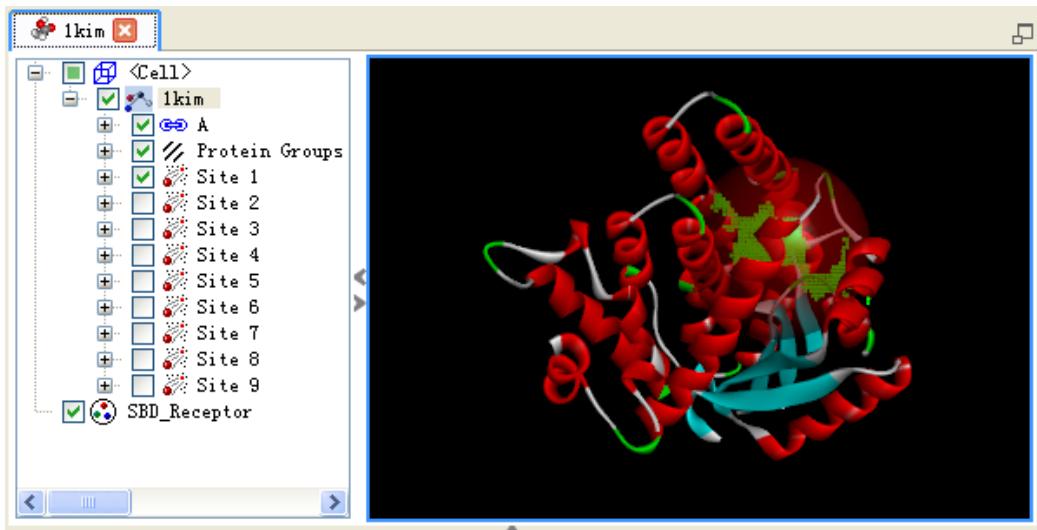


图 2 从受体中寻找结合位点

每一个找到的结合位点都包含了绿色的位点显示（Binding Site）和一个红色球体（SBD\_Site\_Sphere），该球体可通过在系统视图中展开每一个 Site，勾选或取消 SBD\_Site\_Sphere 来选择显示与否。

## 3. 打开配体文件

在文件浏览器（Files Explorer）中，找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| kinase\_ligands.sd 文件。

在新的分子窗口中打开一张共有 9 行的表格，每一行代表一个分子，共有 9 个配体分子。

点击表格左边的 按钮，在视图窗口中显示第一个分子（图 3）

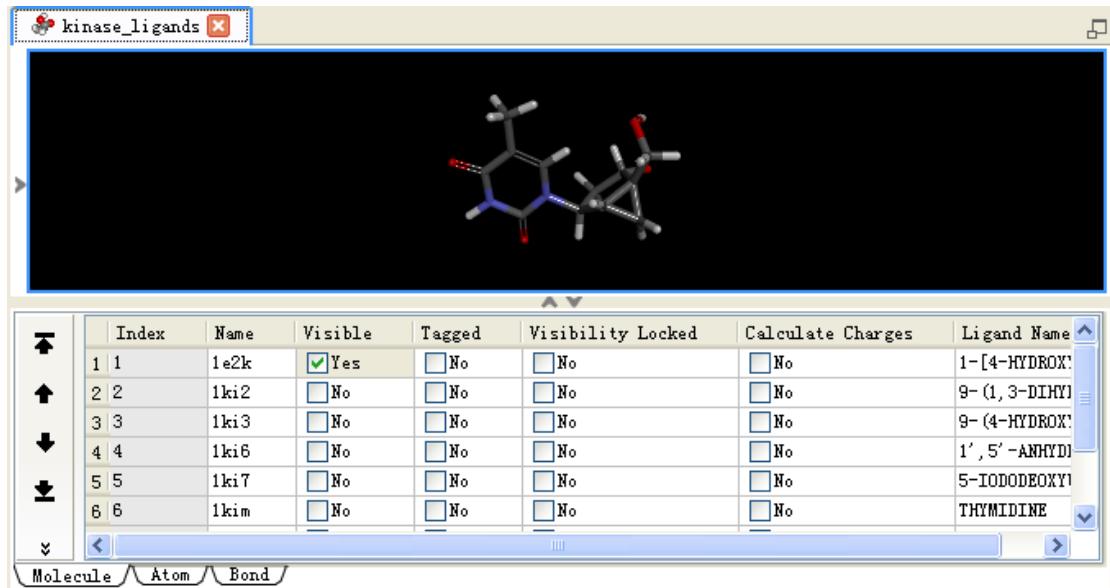


图 3 以表格形式浏览待对接配体

#### 4. 分子对接

在流程浏览器（Protocols Explorer）中，展开 Receptor-Ligand Interaction | Docking，点击打开 Dock Ligands (LigandFit) 流程。

流程对应参数在参数浏览器中打开。

在参数浏览器中，点击 Input Receptor 参数，从下拉列表中选择 1kim:1kim 设置受体蛋白。

点击 Input Binding Site 参数，从下拉列表中选择 Site1:1715 points,214.375Å^3, partition level 1，指定活性位点。

点击 Input Ligands 参数，从下拉列表中选择 kinase\_ligands:All，设置配体分子。

其他参数设置为缺省值。其中打分函数选取的是默认的 LigScore1、LigScore2、PLP1、PLP2、Jain、PMF 六种打分函数。（图 4）

点击流程工具栏（Protocols toolbar）中的 按钮运行作业，等待作业完成。

Dock Ligands (LigandFit) <span style="border: 1px solid red; padding: 2px;">X</span>	
Parameter Name	Parameter Value
Input Receptor	1kim:1kim
Input Binding Site	Site 1 : 1715 points, 214.375 Å^3, Partition level 1
Input Ligands	kinase_ligands:All
+ Input Control Ligands	
+ Energy Grid	Dreiding
Number of Monte Carlo Trials	"2 500 120,4 1200 300,6 1500 350,10 2000 500,25 3000 750"
+ Conformation Search	
+ Docking	Docking
Maximum Poses Retained	10
+ Pose Saving	
+ Interaction Filters	
+ Minimization Algorithm	Do not minimize
+ Scoring Functions	LigScore1,LigScore2,PLP1,PLP2,Jain,PMF
+ Parallel Processing	False

图 4 LigandFit 对接参数设置

#### 4. 浏览对接 poses

为清晰起见，从菜单栏中选择 **Window | Close All** 关闭所有窗口。

待作业完成后，双击作业浏览器(Jobs Explorer)中刚完成的分子对接作业，打开 **Report** 窗口，点击 **View Results**。

打开一个新的分子窗口 **1kim**，在系统视图中可以观察到该窗口中包含了所有 9 个配体分子的对接结果，共 90 个。

在窗口中点击鼠标右键，在 **Atom Display** 选取为 **Line**，**Protein Display** 选取 **Off**（图 5）。将受体蛋白分子的二级结构显示方式取消，将原子以线状方式显示。



图 5 鼠标右键快捷工具

展开窗口下方表格视图 (Table View)。

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interaction | Define and Edit Binding Site**，点击 **Show/Hide Residues Outside Sphere**。

将球外部的氨基酸残基隐藏。

在表格视图 (Table View) 中，点击 **Molecule** 选项，将第 1 行中位于 **Visibility Locked** 下的打勾取消。

在系统视图 (Hierarchy View) 中，将 **Cell | Site1** 前面的勾取消。

在表格视图 (Table View) 中，**Molecule** 选项下，将第 1 行中位于 **Visibility Locked** 下的打勾重新勾选。

点击表格视图中的 按钮，可观察排在第一位的对接构象，继而可通过点击表格视图中的 和 按钮（或者 CTRL+Up/CTRL+Down），观察配体分子的每个 pose 同受体分子的结合模式。

## 分析配体对接构象

### 1. 分析受体-配体间非键相互作用

在工具浏览器 (Tools Explorers) 中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

在视图窗口中，受体原子与配体对接 poses 间的非键相互作用会通过不同颜色的虚线显示出来。（图 6）

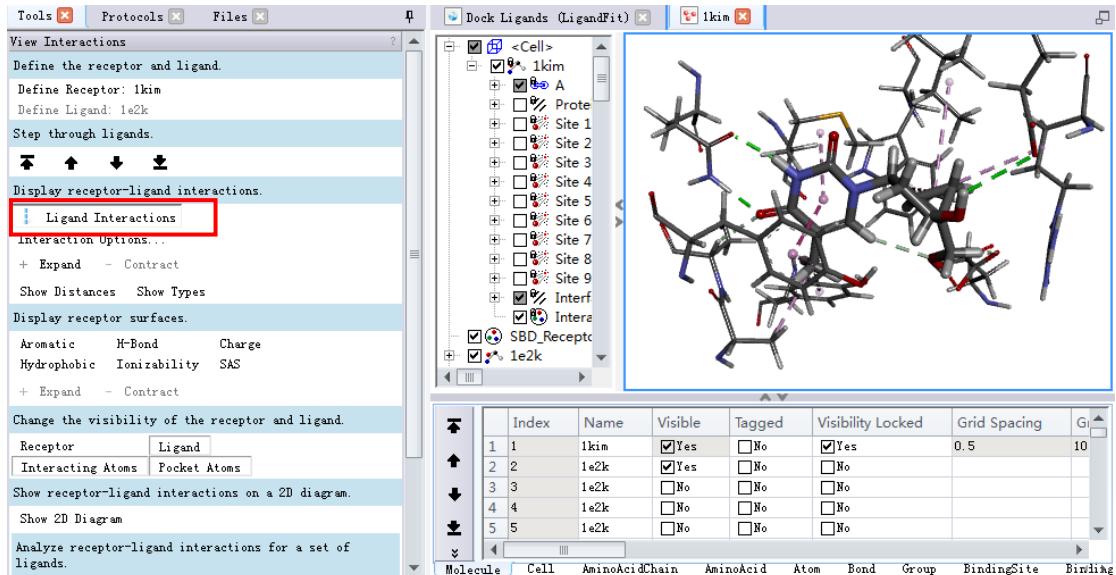


图 6 显示蛋白-配体之间非键相互作用

然后点击该工具面板下的 **Interaction Options**，展开如下窗口（图 7）。

该窗口中所列的非键相互作用类型即可以考虑的所有非键作用总类，其中黑色显示即在该蛋白和配体间存在的非键相互作用。

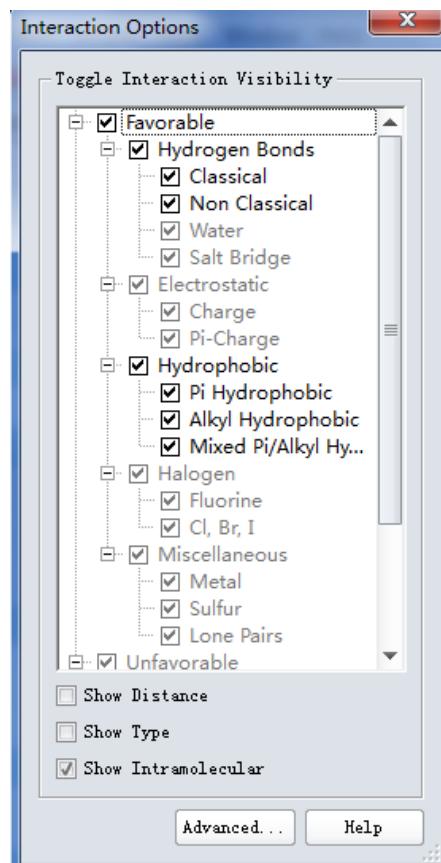


图 7

此外，在分子显示窗口任意选中某一虚线，在 DS 界面的左下方就会显示该非键作用类型及距离等相关信息。（图 8）

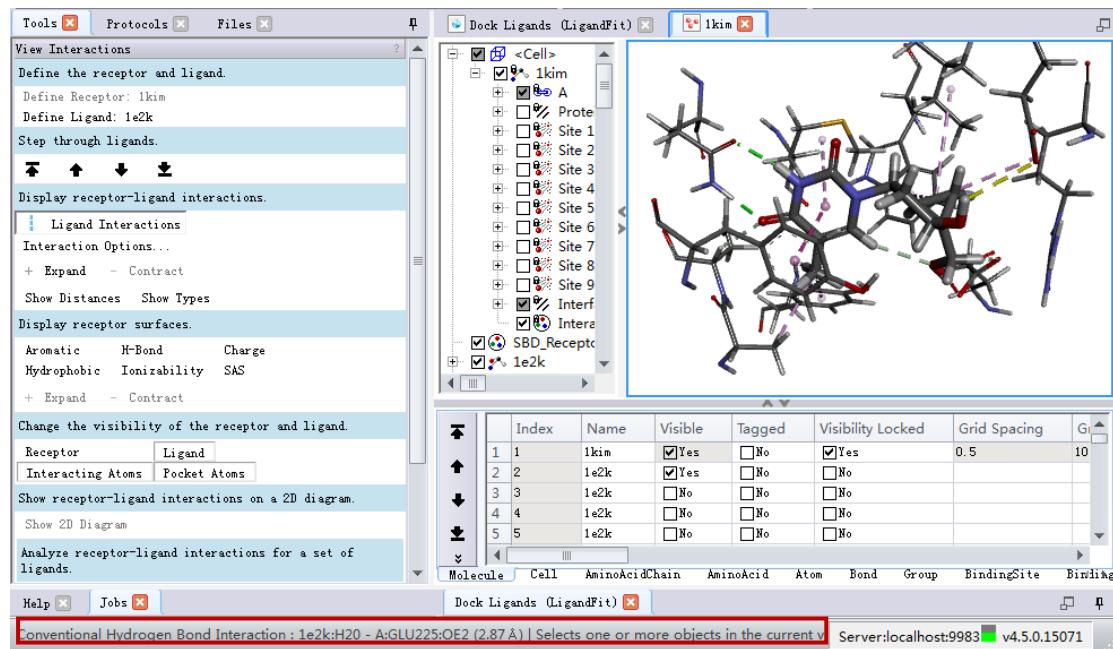


图 8

拖动表格视图中的滚动条至表格右端, 可以看到之前选取的多种打分函数对每个配体分子的每个对接构象的打分值以及 Dock\_Score 打分。

## 2.采用 2D 视图显示配体与受体的相互作用

在系统视图中点击选中要描述的配体分子 (第一个 1e2k)。

在工具浏览器 (Tools Explorers) 中, 展开 **Receptor-Ligand Interactions | View Interactions**, 点击 **Define Ligand**, 然后点击 **Show 2D Diagram**。

可以看到在一新窗口中打开配体-蛋白相互作用二维平面图, 便于我们更直观的观察两者的相互作用及关键的氨基酸残基和分子官能团。 (图 9)

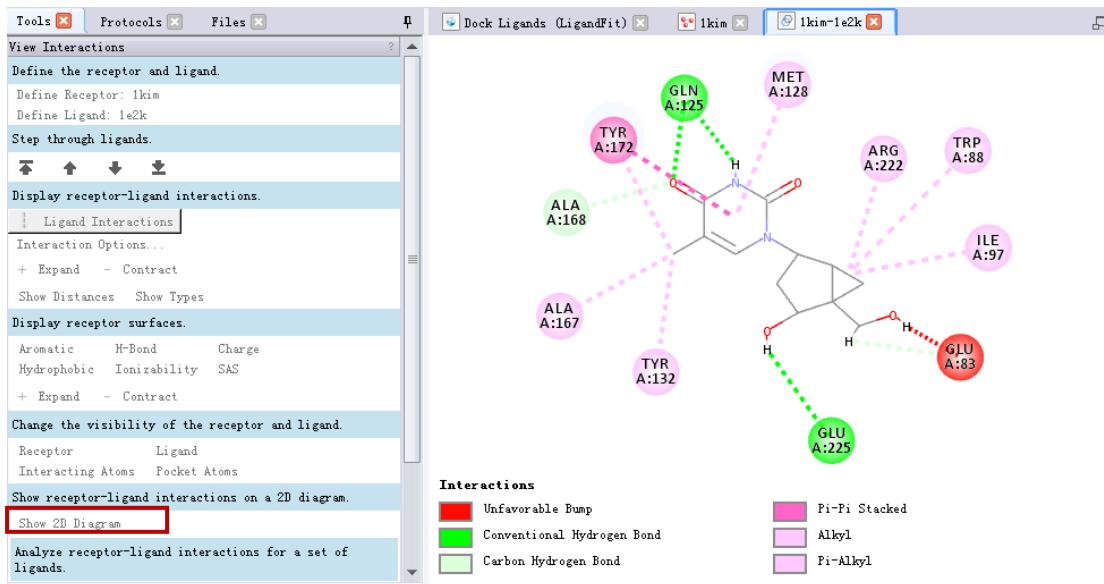


图 9 配体-蛋白相互作用二维平面图

## 3.采用一致性打分对多个打分函数进行一致性评价

回到 Report 界面, 点击 Results 一栏中的 **Docked Ligands** 链接。

在新的分子窗口 kinase\_ligands 中打开一个 90 行的表格, 代表了对接得到的 90 个配体构象。

在流程浏览器 (Protocols Explorer) 中, 展开 **Receptor-Ligand Interaction | Scoring And Analysis**, 点击打开 **Consensus Score** 流程。

流程对应参数在参数浏览器中打开。

在参数浏览器中, 点击 **Input Ligands** 参数, 从下拉列表中选择 **kinase\_ligands:All** 设置对接得到的各配体构象。

点击 **Input Properties** 右边的栅格, 打开 Input Properties 参数界面, 选取 **LigScore1\_Dreiding**、**LigScore2\_Dreiding**、**-PLP1**、**-PLP2**、**Jain**、**-PMF**、**DockScore** 七种打分函数。

展开 **Advanced** 一栏, 将 **Use Best Pose Only** 参数选取为 **False**, 即针对每个配体分子并不只考虑对接打分最高的那个对接构象。

其余参数采用默认设置, 其中 **Consensus Percentage** 为 **20**。 (图 10)

点击流程工具栏 (Protocols toolbar) 中的 按钮运行作业, 等待作业完成。

完成这个作业大约需要半分钟。

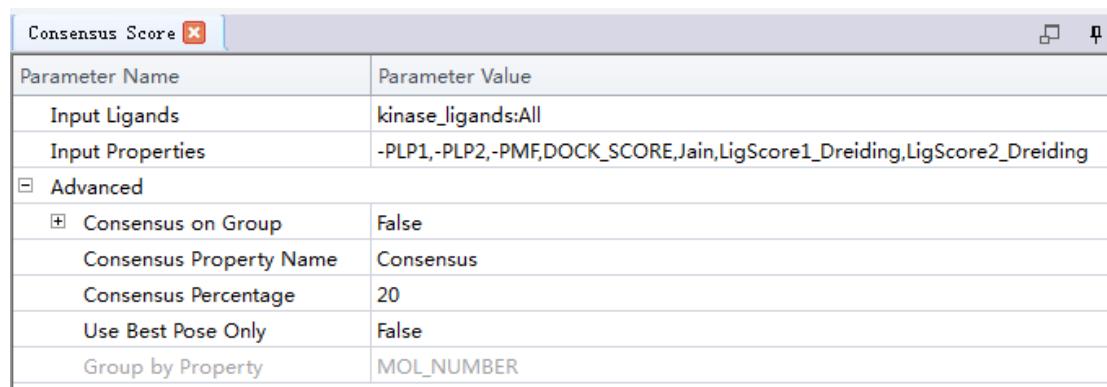


图 10 一致性打分参数设置界面

为清晰起见，从菜单栏中选择 **Window | Close All** 关闭所有窗口。

待作业完成后，双击作业浏览器中刚完成的一致性打分作业，打开 **Report** 窗口，点击 **View Results**。

在新的分子窗口 **kinase\_ligands** 中打开一张共 90 行的表格。

拖动表格下方的滚轮至表格的最右端。

可以看到添加了 **Consensus** 一列数值，即一致性打分值。（图 11）

双击该列表头两次，使该值从高到低排列，从而方便挑选一致性打分较高的对接构象。

	LIG_INTERNAL_ENERGY	Ligand Name	MOL_NUMBER	Consensus
1	i... 1.49	1-[4-HYDROXY-5-....	1	5
2	i... 7.278	1',5'-ANHYDRO-2... 4	4	5
3	i... 7.856	1',5'-ANHYDRO-2... 4	4	5
4	i... 6.443	1',5'-ANHYDRO-2... 4	4	5
5	i... -1.161	5-IODODEOXYURID...	5	5

图 11 一致性打分值

## Discovery Studio CDOCKER 教程

### CDOCKER - 精准的分子对接技术

**所需功能和模块:** Discovery Studio Client, DS CDOCKER.

**所需数据文件:** 1EQG.dsv, 1EQD-ibuprofen-conf.sd, 1EQG-ibuprofen.sd

**所需时间:** 15 分钟

#### 介绍

CDOCKER 是基于 CHARMm 力场的分子对接方法, 这种方法可以产生高精度的对接结果。在本教程中, 天然布洛芬配体分子对接回 COX-1 受体的结合位点中, 得到的对接构象和 X-ray 衍射得到的晶体结构中的配体天然构象进行比较。本教程包括:

- 准备对接体系
- 运行 CDOCKER
- CDOCKER 结果分析
- 计算对接结果的 RMSD 值

#### 准备对接体系

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | 1EQG.dsv。

在分子窗口中将打开一个带有活性位点的蛋白质三维结构 (图 1)。

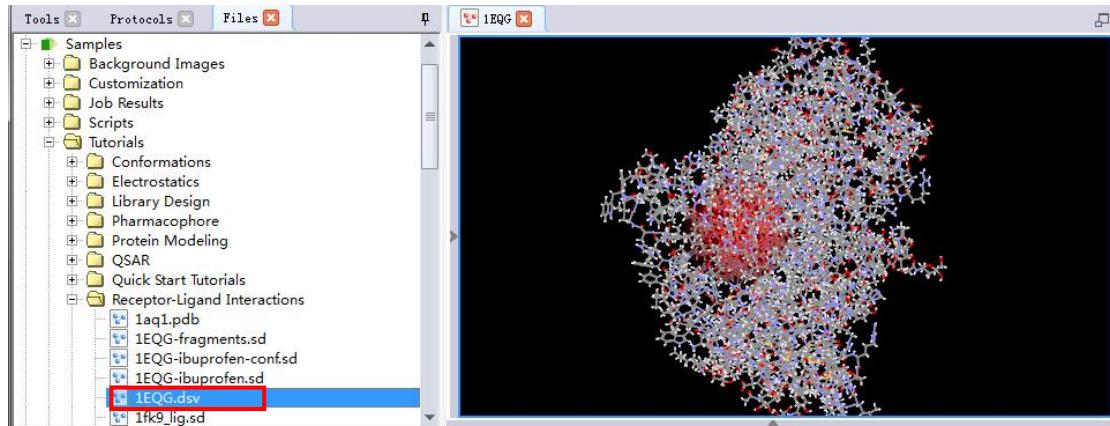


图 1 蛋白质三维结构示意图

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 依次点击 Show/Hide Residues Outside Sphere 和 Show/Hide Sphere。

展开菜单栏 View|Transform, 点击 Fit To Screen 将结合位点的氨基酸在窗口中居中显示(图 2)。

以上操作可以将结合位点外的残基以及球体隐藏，以便观察对接结果时更加便捷。

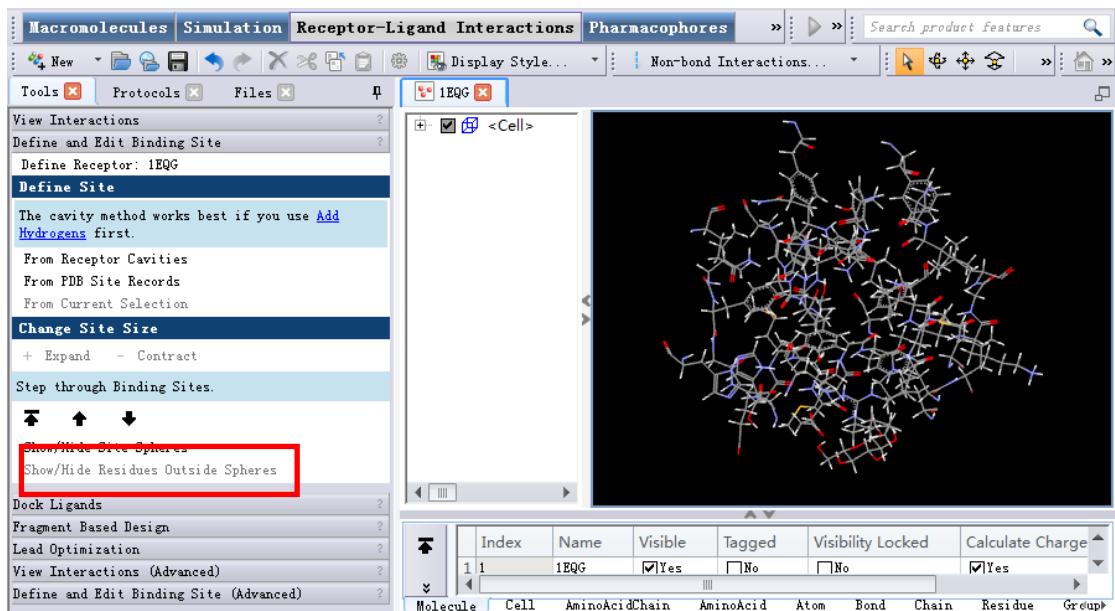


图 2 蛋白质活性位点氨基酸

找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | 1EQG-ibuprofen-conf.sd 文件。

将打开一个具有随机机构象的布洛芬分子（图 3）。

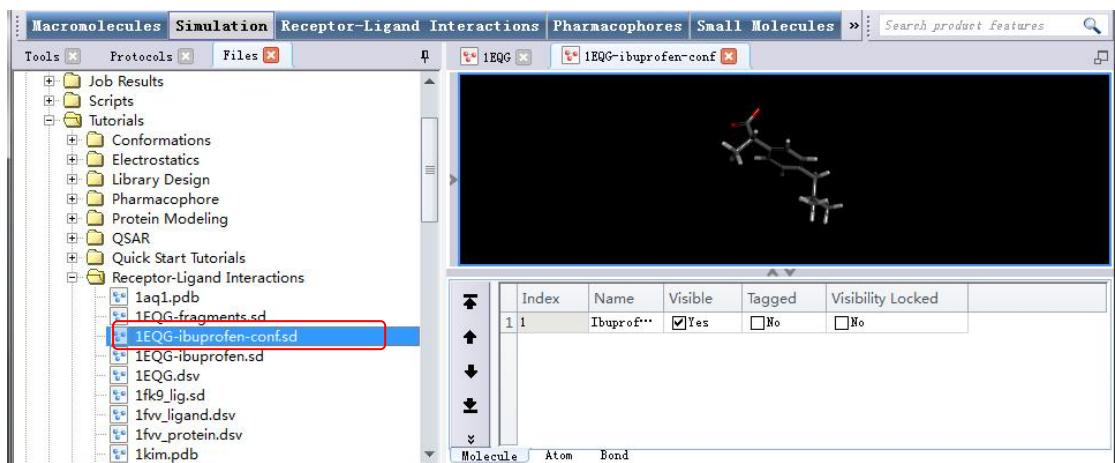


图 3 配体小分子结构

## 运行 CDKCKER

在工具浏览器（Tools Explorer）中，展开 Receptor-Ligand Interactions | Dock Ligands，点击 Dock Ligands (CDOCKER)，打开相应参数面板。

在参数面板中，将 Input Receptor 设置为 1EQG:1EQG。

参数 Input Ligands 设置为 1EQG-ibuprofen-conf:All。

点击 Input Site Sphere 参数，从下拉列表中选择该 sphere 的坐标及半径。

展开 Top Hits 参数，设置 Pose Cluster Radius 为 0.5。

将 RMSD 阈值设为 0.5 埃以确保对接构象尽可能具有多样性。

其余参数默认（图 4），点击 Run 运行。

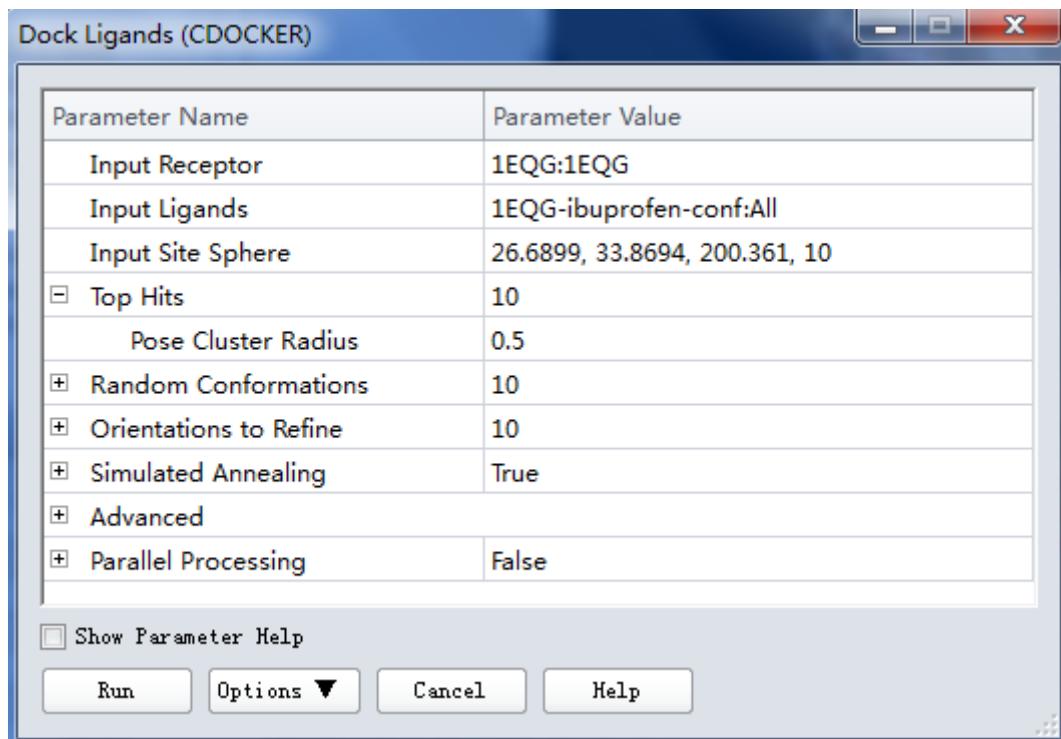


图 4 CDOCKER 参数设置

## CDOCKER 结果分析

作业完成后，在作业浏览器（Jobs Explorer）中，双击完成的分子对接任务，打开 Report 报告界面，点击 View Results，打开对接结果。

### 3. 非键相互作用的直观显示与分析

在工具浏览器（Tools Explorers）中，展开 Receptor-Ligand Interactions | View Interactions，点击 Ligand Interactions。

在视图窗口中，受体蛋白氨基酸残基与配体对接 poses 间的非键相互作用会通过不同颜色的虚线显示出来，且只有参与了同配体之间的相互作用的残基才会显示。（图 5）

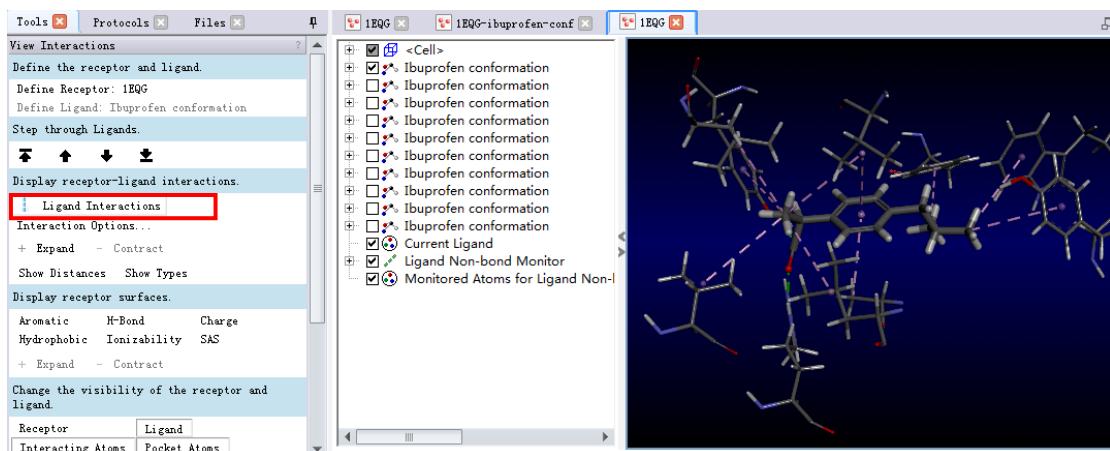


图 5 显示蛋白-配体之间非键相互作用

点击上述 View Interaction 工具面板下的 Interaction Options，展开如下窗口（图 6）。

该窗口中列出 DS 中所有非键相互作用类型，其中黑色显示表示该蛋白和配体间存在的非键相互作用。

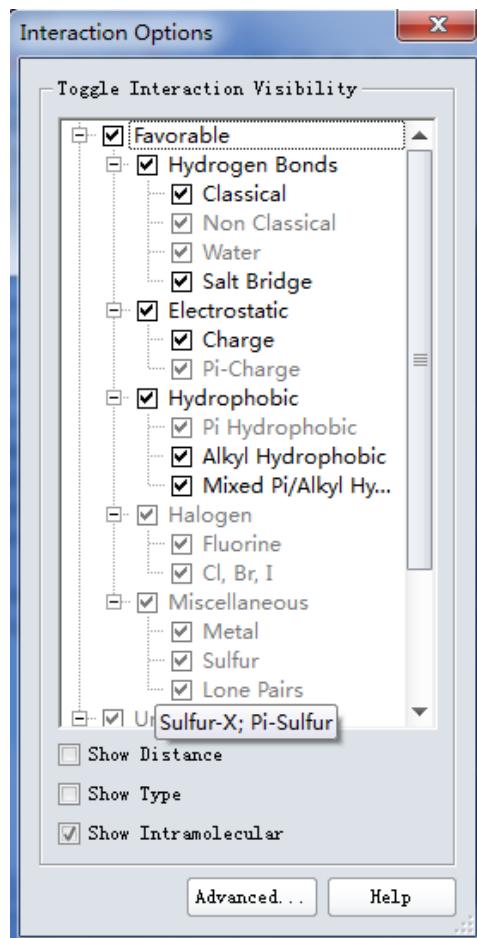


图 6

此外，在分子显示窗口任意选中某一虚线，在 DS 界面的左下方就会显示该非键相互作用类型及距离等相关信息。（图 7）

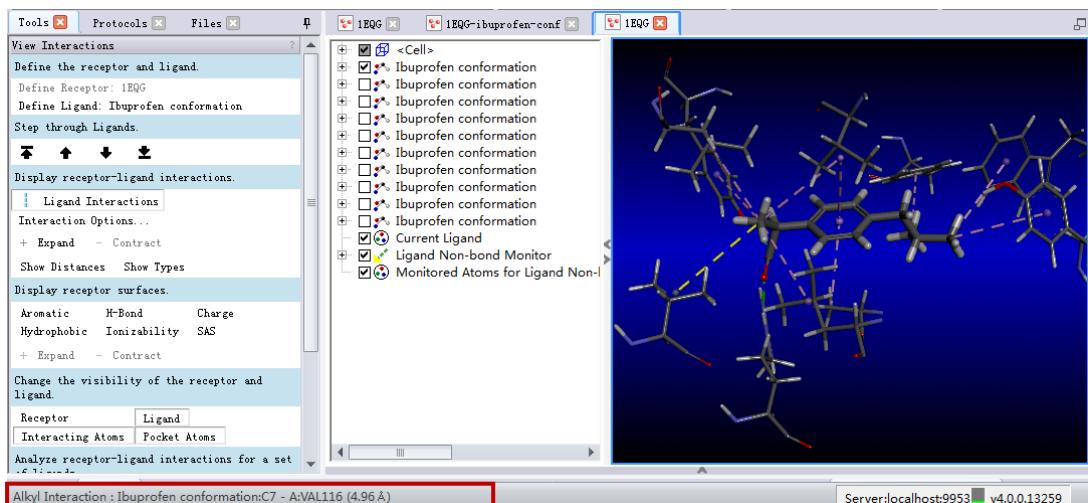


图 7

为了更好的观察受体分子与配体对接 pose 间的相互作用，可以对体系进行旋转以获得最佳的观赏角度。

Step through Ligands.

点击上述 View Interaction 工具面板下的 按钮可以观察不同的对接构象同蛋白之间的非键相互作用。

#### 4. 生成配体-蛋白相互作用二维平面图

展开 Receptor-Ligand Interactions |View Interactions，点击 Show 2D Diagram。

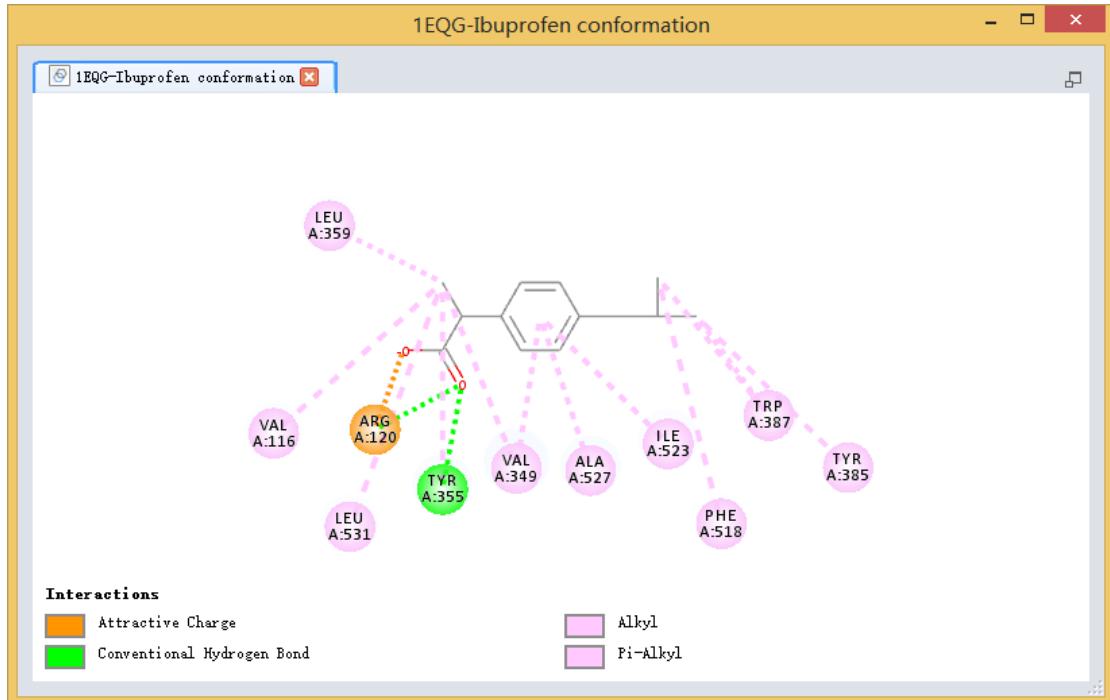


图 8 配体-蛋白相互作用二维平面图

可以看到在一新窗口中打开配体-蛋白相互作用二维平面图，便于我们更直观的观察两者的相互作用及关键的氨基酸和基团。（图 8）

#### 5. 对接配体和布洛芬天然晶体结构比对

在任务浏览器（Jobs Explorer）中，单击该任务条下（点开前面的+号）Docked Ligands 链接。

Protocol Name	Saved	Status	Details	Elapsed Time	Sta
Dock Ligands (CDOCKER)	<input checked="" type="checkbox"/> Yes	Success	10 poses: 1EQG-	0:03:36	

Report  
Docked Ligands  
Prepared Input Receptor  
View Results

对接配体将在一个新的分子窗口中打开。

按住 **CTRL+G**，显示分子图形窗口。

展开菜单栏 **File | Insert From**，点击 **File...**，选择 **Samples | Tutorials | Receptor-Ligand Interactions | 1EQG-ibuprofen.sd**。

在同一窗口中插入布洛芬的天然晶体结构 **1EQG-ibuprofen.sd**。

在表格视图中，设置最后一行的 **Ibuprofen** 分子的 **Visible** 和 **Visibility Locked** 为 **True**。

按住 **CTRL+Down**，观察每个配体文件和布洛芬天然晶体结构的比对情况。

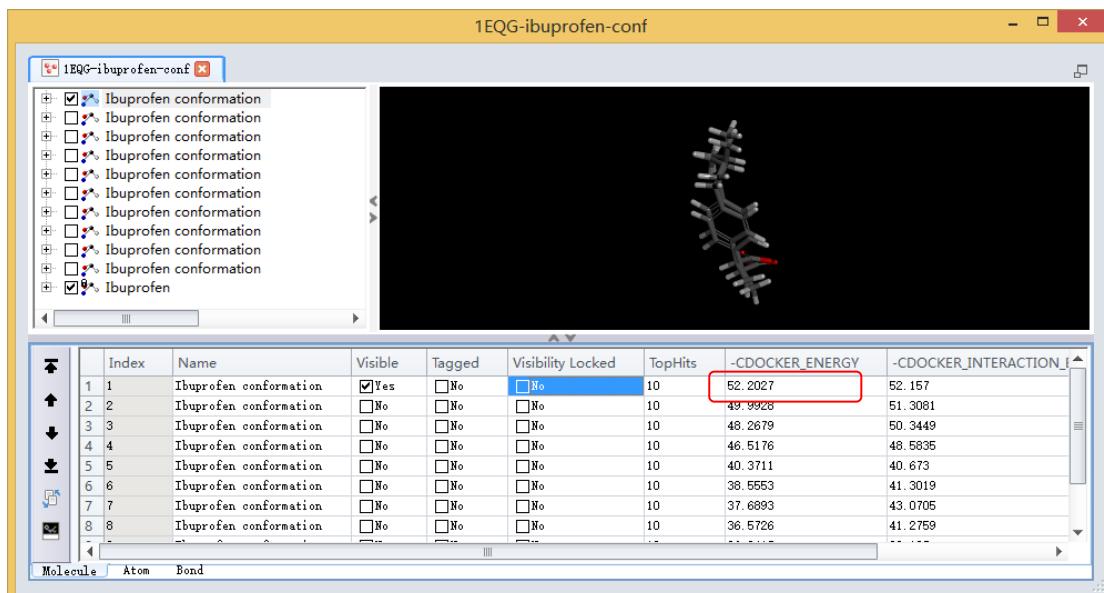


图 8 对接配体和布洛芬天然晶体结构比对

可以发现 CDOCKER 打分最高的位点，即-CDOCKER\_ENERGY 的值最高，和天然布洛芬晶体结构具有很好的叠合效果。

在表格视图中，选择最后一行的 Ibuprofen 分子，展开菜单栏 Structure|RMSD，点击 Set Reference。

在分子窗口中点击鼠标右键，选择 Show All，显示所有小分子结构。

展开菜单栏 Structure|RMSD，点击 Heavy Atoms。

打开一个新的窗口，显示所有 10 个对接构象同晶体构象之间的 RMSD 偏差。

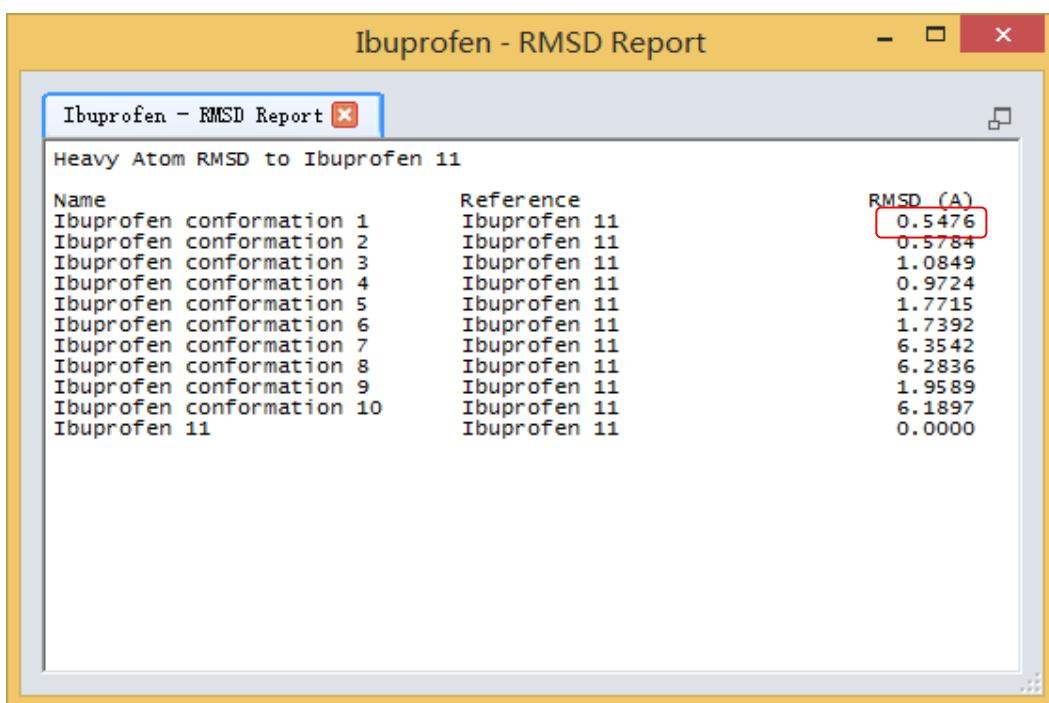


图 9

## Discovery Studio Flexible Docking 教程

### Flexible Docking – 全柔性受体-配体对接技术

**目的:** 受体和配体柔性均考虑的分子对接

**所需功能模块:** Discovery Studio Visualizer client, DS LibDock, DS Catalyst Conformation, DS CHARMM 及 DS Protein Refine

**所需数据文件:** 1s1x\_prot.dsv 和 1fk9\_lig.sd

**所需时间:** 25min

#### 介绍

柔性分子对接是指在对接过程中，蛋白的侧链和配体分子构象都可以自由变化，多用于精确考查分子间的结合模式，但由于计算量和计算时间的缘故，在实际应用中，一般只将活性位点处残基侧链定义为柔性，可以考虑其构象的改变。在 Flexible Docking 中，配体将按如下步骤对接到考虑活性部位氨基酸残基侧链柔性的受体中：

- 选择受体柔性氨基酸残基，使用 ChiFlex (CHARMM) 通过改变侧链构象计算出一组蛋白构象
- 使用 LibDock 将柔性配体对接到每一个受体构象的活性部位
- 使用 ChiRotor (CHARMM) 在刚性配体存在的情况下优化选定的蛋白侧链
- 使用 CDOCKER 优化最后的配体对接构象

在本教程中研究的两个蛋白分子为同一个 HIV-RT 受体，但由于所含的配体不同，两蛋白 PDB 编号不同，分别为 1s1x 和 1fk9。配体结构的差别会引起蛋白构象的改变，这也是柔性对接的理论基础。本教程将 1fk9 中结合的配体对接到 1s1x 中。

具体步骤包括：

- 准备对接的分子体系
- 执行分子对接计算
- 分析对接结果
- 计算对接结果的 RMSD 值

#### 准备对接的分子体系

##### 1. 蛋白和配体文件的准备

在文件浏览器 (Files Explorer) 中，找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| 1s1x\_prot.dsv 文件。

该蛋白将在一个新的三维窗口中打开。（图 1）

此蛋白已经过处理：已赋予 CHARMM 力场，所以在对接之前不需再进行力场的赋予；此外，结合位点也已识别，窗口中只显示该结合位点周围的 8 个残基及其相应的残基名，在对接过程中将会考虑这 8 个残基的侧链柔性，上述 8 个残基已被定义为一组并命名为 1s1x\_res8，系统视图中已显示。

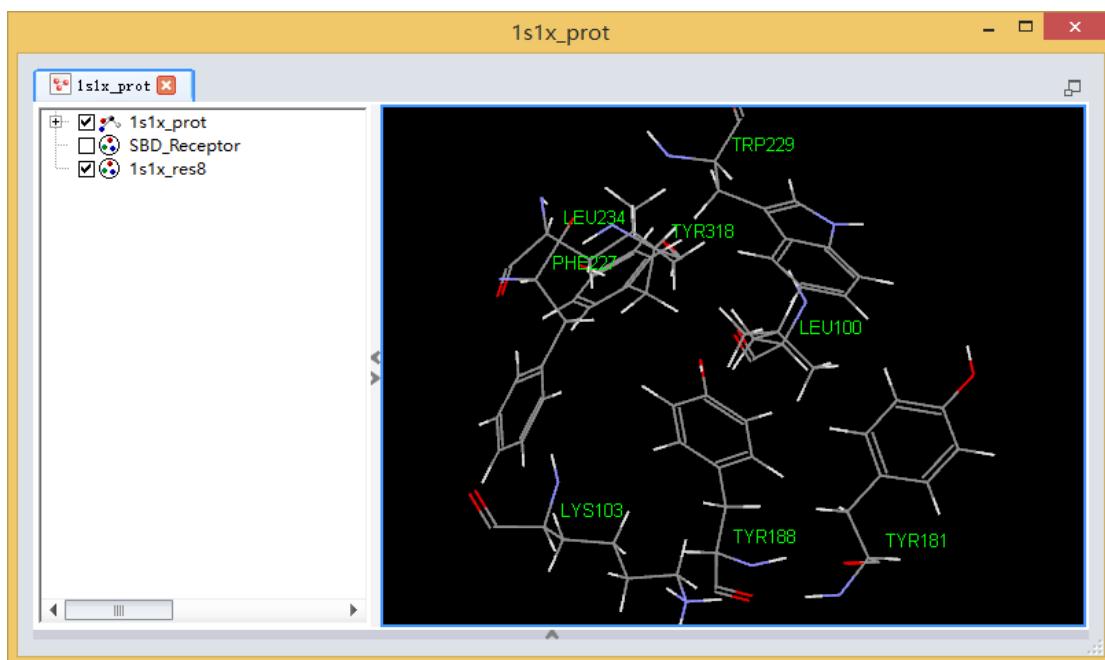


图 1 打开蛋白文件

在文件浏览器（Files Explorer）中，找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions|1fk9\_lig.sd 文件。

在一个新的窗口中打开配体分子，默认以表格视图形式显示。

按 **Ctrl+G**，展开分子视图窗口，显示配体分子结构（图 2）。

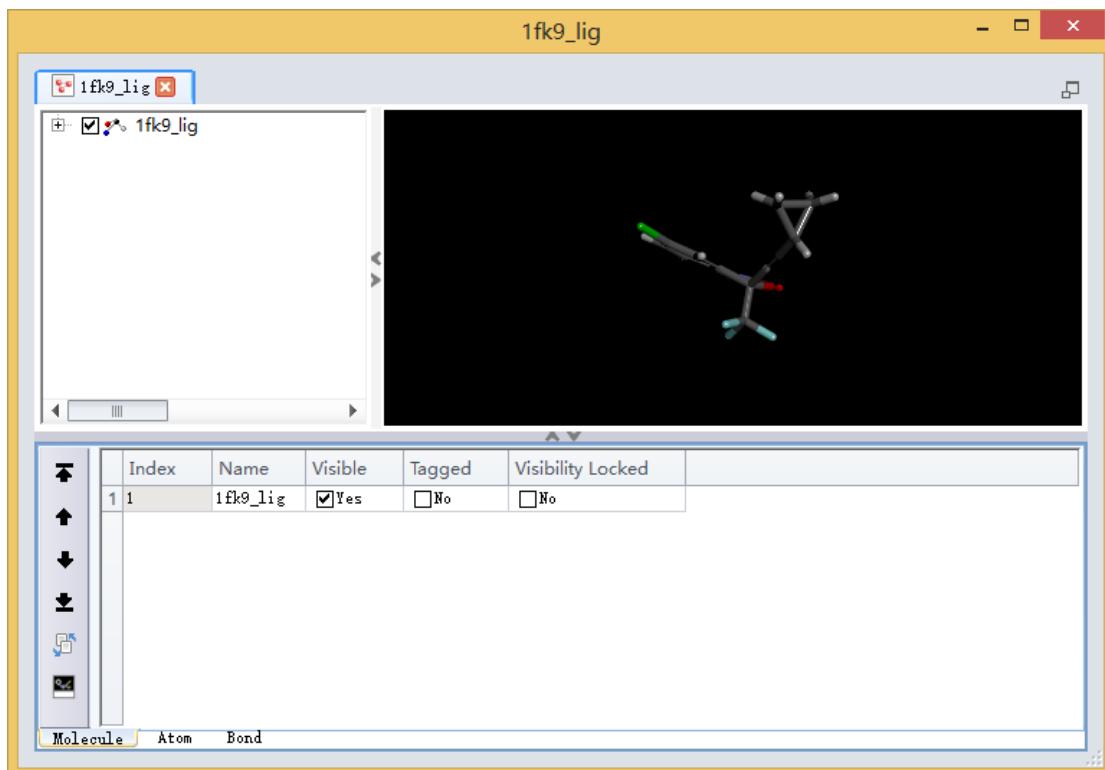


图 2 打开配体分子

## 执行分子对接计算

在流程浏览器 (Protocols Explorer) 中, 展开 **Receptor-Ligand Interactions | Docking**, 点击打开 **Flexible Docking** 柔性对接流程。

流程的对应参数项在参数浏览器中出现。

在参数浏览器中, 点击 **Input Typed Protein Molecule** 参数, 从下拉列表中选择 **1s1x\_prot:1s1x\_prot** 设置受体蛋白。

点击 **Input Ligands** 参数, 从下拉列表中选择 **1fk9\_lig:All**, 设置配体分子。

点击 **Input Site Sphere** 参数项, 从下拉列表中选择 **binding\_Sphere** 坐标, 指定活性位点。

点击 **Selected Residues** 参数项, 从下拉列表中选择 **1s1x\_res8**, 指定考虑侧链柔性的一组氨基酸残基。

该操作选择了蛋白活性位点处残基。这些残基事先已被选中并被定义为 **1s1x\_res8** 组。

点击 **Generate Protein Conformations** 参数项, 选择 **False**。

展开 **Generate Protein Conformations** 参数组, 点击 **Input File** 参数项, 载入之前已经产生的蛋白多构象。在文件浏览器 (Files Explorer) 中, 找到 **Samples| Tutorials| Receptor-Ligand Interactions| 1s1x\_prot-conformations.mol2** 文件, 载入即可。

本教程为了缩短运行时间, 蛋白多构象已事先产生好。

展开 **Generate Ligand Conformations** 参数组, 点击 **Conformation Method** 参数, 从下拉列表中选择 **BEST** 方法。

**注:** 虽然 **BEST** 方法不是构象生成方法参数缺省设置, 本例中使用该方法是因为对于含复杂环的配体它能更好的覆盖构象空间。

展开 **Docking** 参数组, 点击参数 **Max Hits to Save** 设置最多保存结果为 3。

如果电脑配置有多个核, 则可以设置并行设置来缩短运行时间。

点击 **Parallel Processing** 参数, 设置为 **True**。 (图 3)

点击流程工具栏 (Protocols toolbar) 中的 **Run**  按钮运行作业, 等待作业完成。

当作业完成时会出现作业完成对话框, 关闭作业完成对话框。

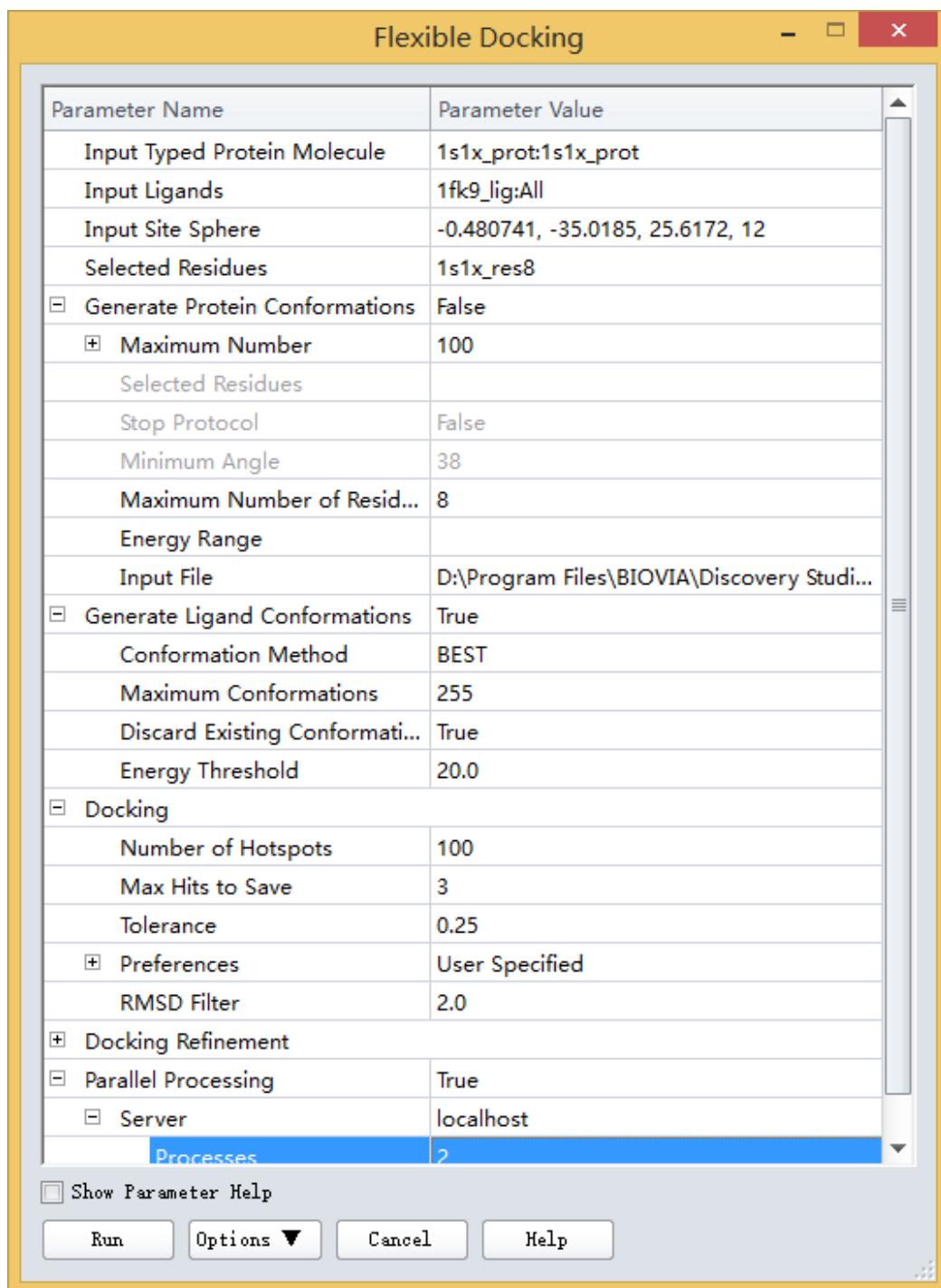


图 3 “Flexible Docking”参数设置

### 分析分子对接结果

在开始结果浏览之前，为了清晰起见，在菜单栏中选择 **Window | Close All** 关闭所有窗口。

在作业浏览器（Jobs Explorer）中，双击刚完成的作业。

在 Html 窗口中打开 Report.htm 文件。

在 Html 窗口中，点击 Results 部分的 **View Results**。

打开一个新的分子窗口，显示 1s1x\_prot.dsv 受体分子和 1fk9\_lig.sd 文件（配体的对接构象）。

在表格视图中，将第一个 1fk9\_lig 对接构象的 visible 勾选上，让其显示在分子视图窗口中。

(图 4)

然后点击表格中的 键和 键浏览对接得到的各配体构象。

观察 1s1x\_res8 组定义的氨基酸残基侧链构象将随着配体构象的变化而变化。

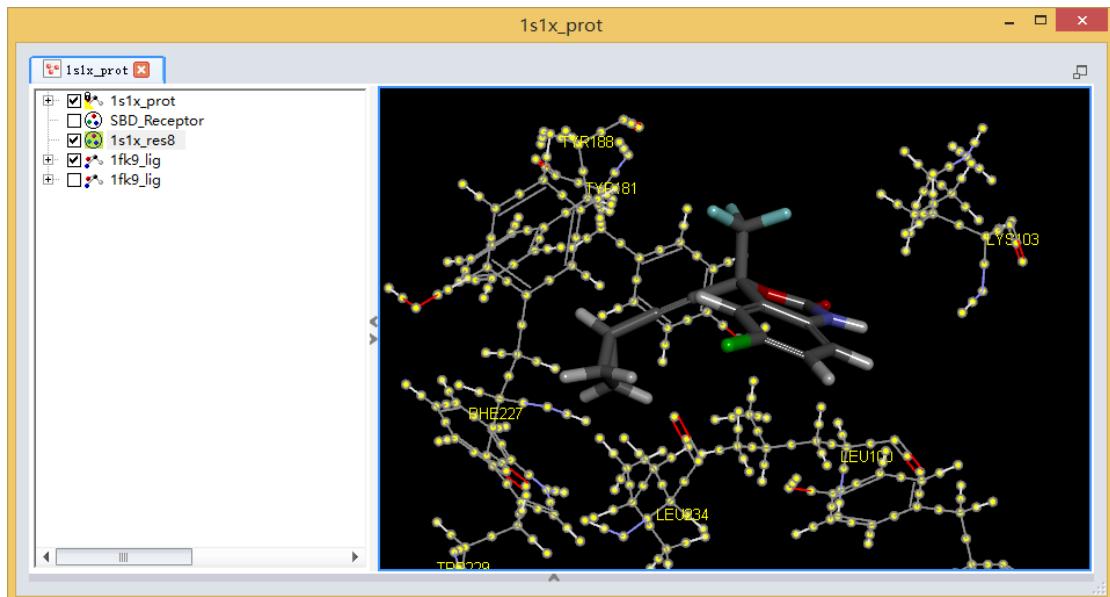


图 4 对接结果示意图

从表格中还可以查看每个对接构象的对接打分，包括 CDOCKER 的能量打分，也包括 LibDockScore 得分。

从菜单栏中选择 Window | Close All，关闭所有窗口。

## 计算对接构象的 RMSD 值

下面我们计算配体对接构象同晶体构象的 RMSD 值并分析其 CDOCKER 能量。

在任务浏览器 (Jobs Explorer) 中，单击该任务条下（点开前面的+号）Flexible Docking 链接。

对接配体将在一个新的分子窗口中打开。

按住 **CTRL+G**，显示分子图形窗口。

展开菜单栏 File | Insert From，点击 File...，选择 Samples | Tutorials | Receptor-Ligand Interactions | 1fk9\_lig.sd。

在同一窗口中插入 1fk9\_lig 的天然晶体结构。

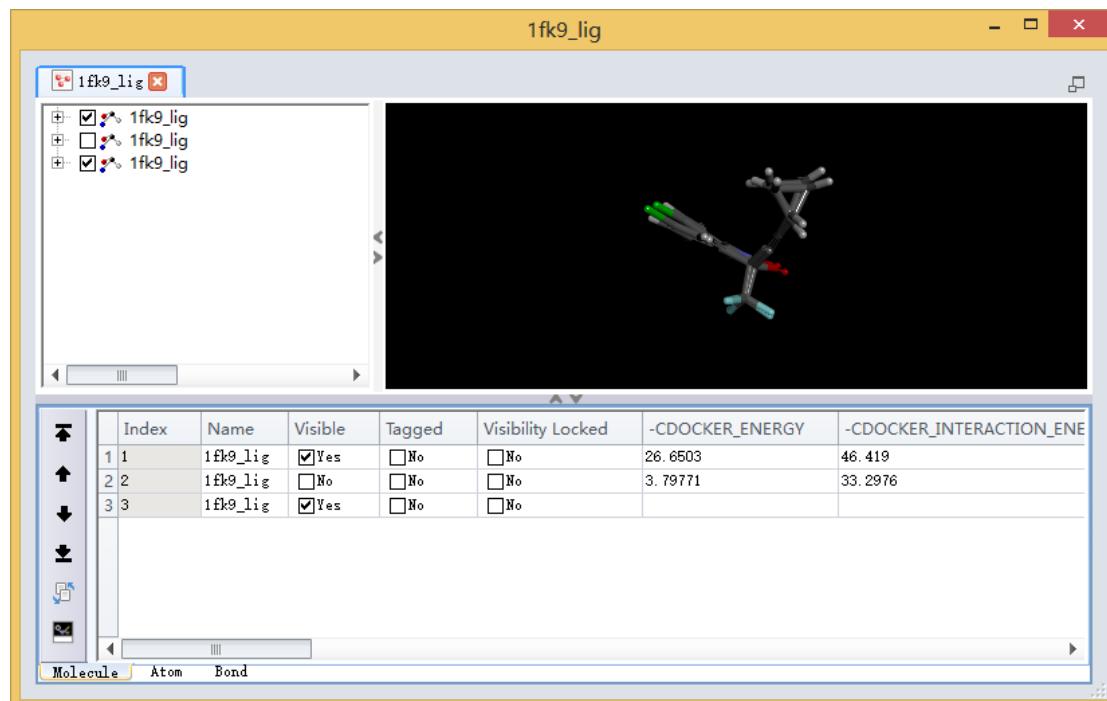


图 5 配体晶体结构与对接结果叠合

在表格视图中，选择最后一行的 1fk9\_lig 分子，展开菜单栏 Structure|RMSD，点击 Set Reference。

在分子窗口中点击鼠标右键，选择 Show All，显示所有小分子结构。

展开菜单栏 Structure|RMSD，点击 Heavy Atoms。

打开一个新的窗口，显示所有 2 个对接构象同晶体构象之间的 RMSD 偏差。

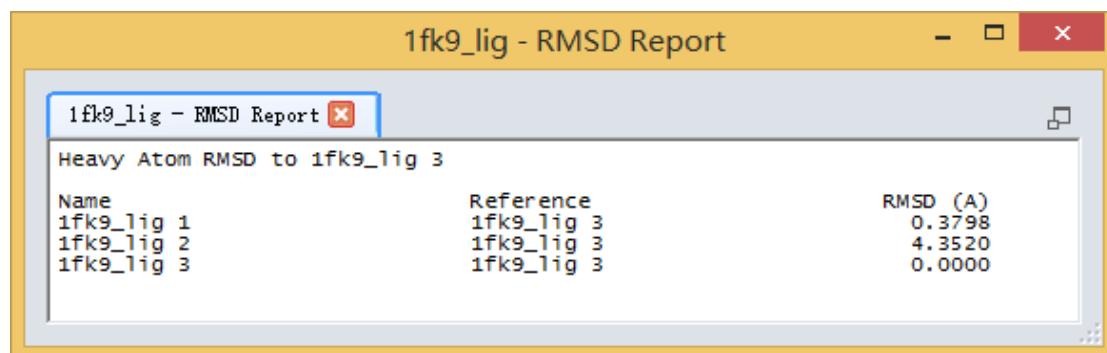


图 6

## Discovery Studio Analyze Ligand Poses 教程

**所需功能和模块:** Discovery Studio Client

**所需数据文件:** 3pwh\_dockedposes.csv

**所需时间:** 20 分钟

### 介绍

分子之间的非键相互作用是生命体系中分子识别的基础。在基于结构的药物设计过程中，识别和优化配体分子和受体分子间的相互作用是一个基本过程，因此，全面细致的分析非键相互作用非常关键。

本教程将采用热稳定的腺苷 A2A 受体和反相激动剂 ZM241385 的复合物，其中 A2A 受体是 G-蛋白偶联受体，是帕金森氏症的药物靶标，其拮抗剂 Preladenant 正处于临床阶段。本教程将分析对接至 A2A 受体的配体分子构象。

### 执行计算并分析结果

在文件浏览器（Files Explorer）中，双击打开 Samples\Tutorials\Receptor-Ligand Interactions\3pwh\_dockedposes.csv。

打开一个分子对接结果，包含 3pwh 蛋白分子和多个小分子的对接构象。（图 1）

说明：对接分子共有 50 个，具有一定多样性，其中 20 个为参照分子，带有 CHEMBL 前缀，30 个为诱饵分子，带有 ZINC 前缀。这 50 个分子基于 Dock Ligands（CDOCKER）模块对接至已经预处理过的腺苷 A2A 受体分子，最终得到 3pwh\_dockedposes.csv 文件中的对接结果。

本教程通过对该对接结果的分析来获取重要的非键相互作用信息，从而进一步优化虚筛流程。

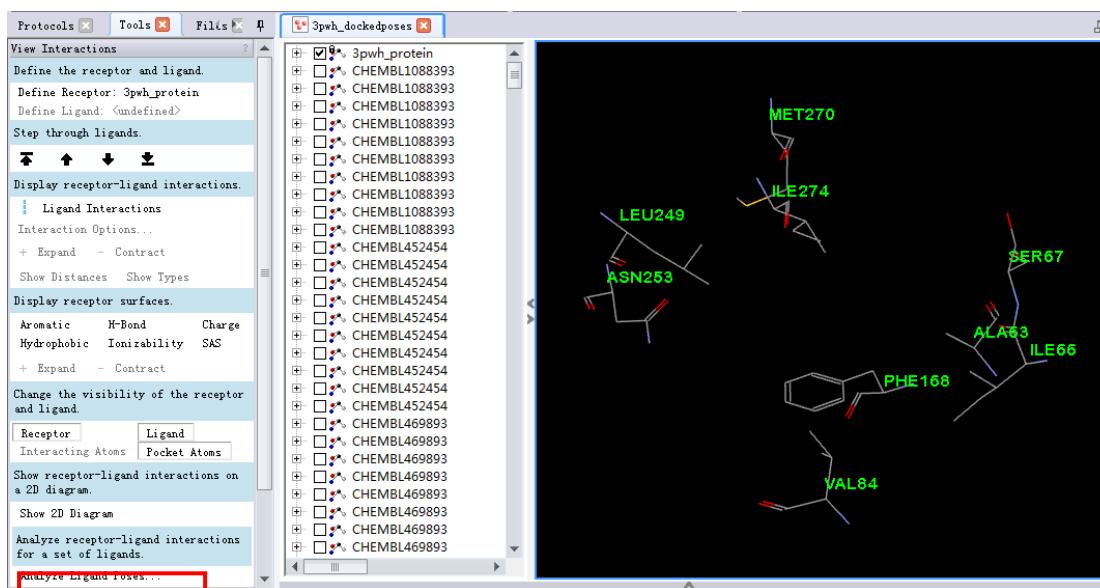


图 1 3pwh 对接结果

在工具浏览器（Tools Explorers）中，展开 Receptor-Ligand Interactions |View Interactions，点击 Analyze Ligand Poses。

相应的参数出现在参数浏览器中。

点击 **Input Ligands** 参数项，下拉列表中选择 **3pwh\_dockedposes:All**，选中 **3pwh\_dockedposes** 窗口中所有配体构象。

其余参数保留默认设置，点击 **Run** 运行任务。（图 2）

点击 **Background** 让任务后台运行，等待任务完成。

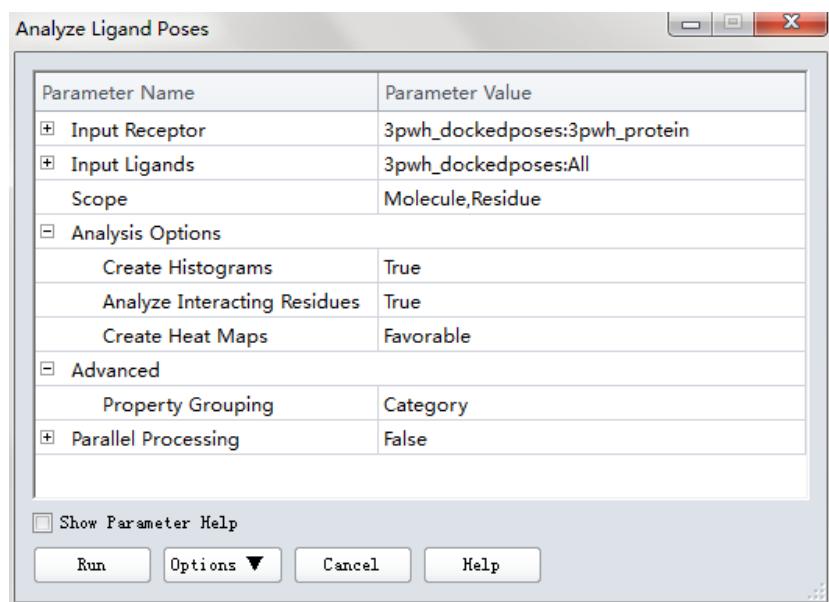


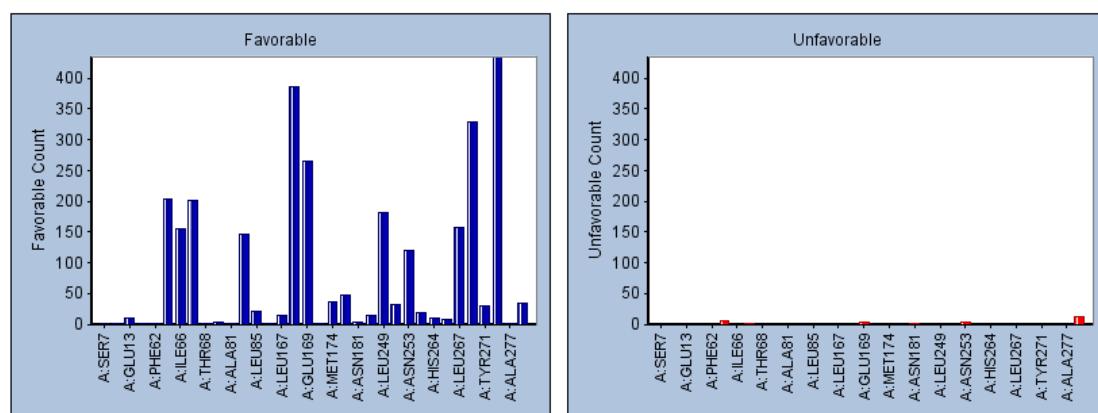
图 2 Analyze Ligand Poses 参数设置

在任务浏览器（Jobs Explorer）中，双击刚运行完的任务，打开 **Report** 界面。

在 **Report** 界面的 **Results** 一栏下，点击 **View Interaction Histogram (Residues)** 链接。

打开一个新的界面，该界面中包含了几幅柱状图。

### Residue Interaction Histograms



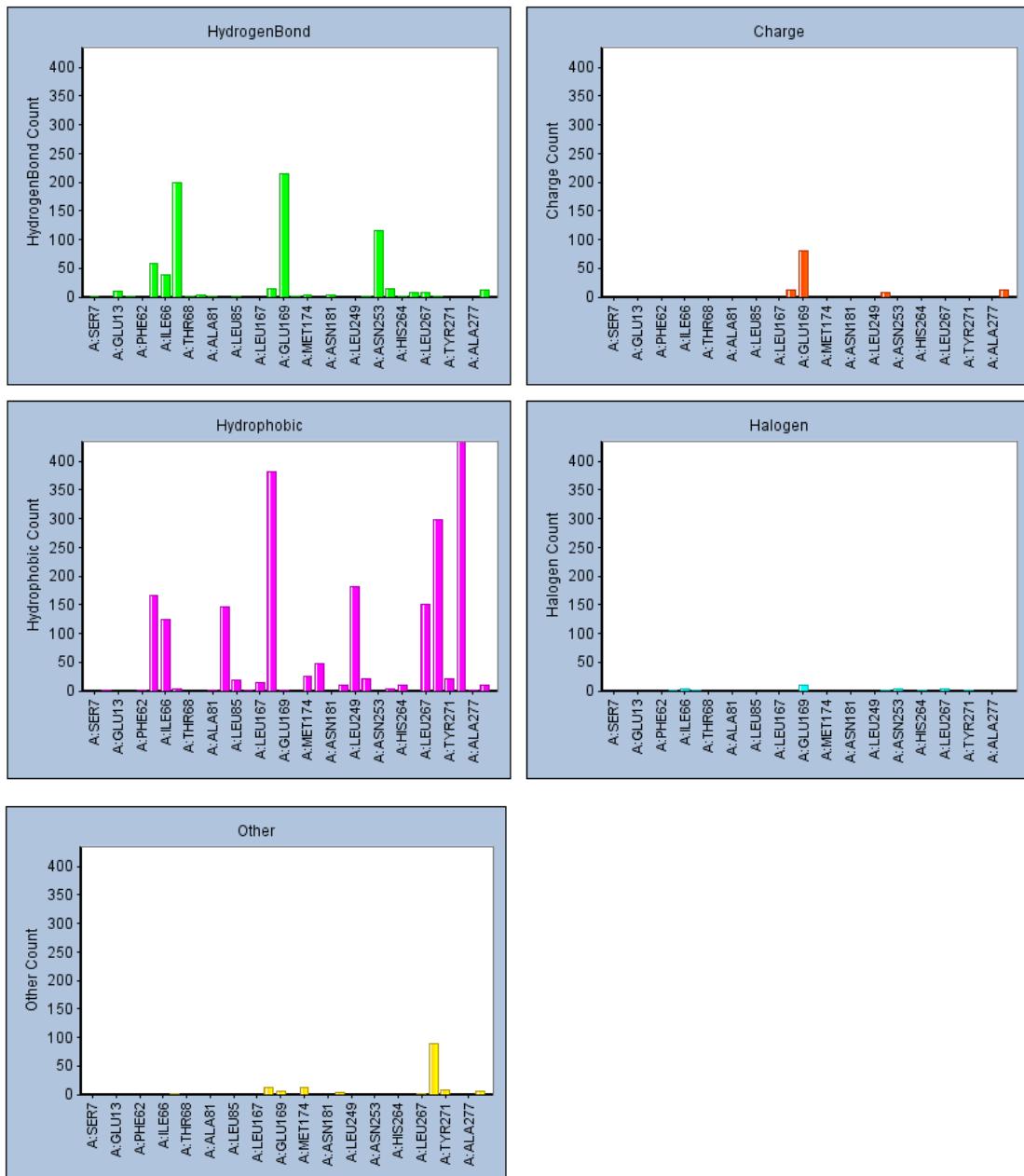


图 3 不同类型相互作用柱状图

以上柱状图总结了各对接构象同受体之间不同类型的相互作用力。对于每一种相互作用类型，从相应的柱状图可以看出受体蛋白中哪一个氨基酸残基同所有的对接构象形成的该类型作用数目最多。所有柱状图的横坐标都相同，便于进行不同类型作用的比较。

将鼠标放置在任一柱状之上，就会显示同该残基形成的该类型相互作用的配体对接构象数目。

点击 (通过在 DS 菜单栏下方点击鼠标右键勾选 Navigation 调出) 中的

Go Back 按钮回到 Report 界面。

点击 View Statistical Information 链接。

打开一个新的界面，该界面中包含了氨基酸残基的统计分析报告，该报告总结了相互作用数目及每一种作用类型所涉及的作用数目较多的氨基酸残基。

Statistical Residue Analysis							
Total Interaction Count							
Favorable	Unfavorable	HydrogenBond	Charge	Hydrophobic	Halogen	Other	
2867	27	713	109	2077	26	138	
Top 5 Residues with Favorable Interactions (5)							
Residue	Favorable	Unfavorable	HydrogenBond	Charge	Hydrophobic	Halogen	Other
A:ILE274	434	0	0	0	434	0	0
A:PHE168	387	0	15	12	381	0	13
A:MET270	328	0	2	0	297	0	89
A:GLU169	264	3	215	79	2	9	6
A:ALA63	203	5	58	0	165	2	0
Top 5 Residues with Unfavorable Interactions (5)							
Residue	Unfavorable	Favorable	HydrogenBond	Charge	Hydrophobic	Halogen	Other
A:HIS278	13	35	12	11	10	0	5
A:ALA63	5	203	58	0	165	2	0
A:GLU169	3	264	215	79	2	9	6
A:ASN253	3	119	115	0	0	4	0
A:SER67	2	202	199	0	4	1	1
Top 5 Residues with HydrogenBond Interactions (5)							
Residue	HydrogenBond	Unfavorable	Favorable	Charge	Hydrophobic	Halogen	Other
A:GLU169	215	3	264	79	2	9	6
A:SER67	199	2	202	0	4	1	1
A:ASN253	115	3	119	0	0	4	0
A:ALA63	58	5	203	0	165	2	0
A:ILE66	38	0	154	0	125	3	0
Top 5 Residues with Charge Interactions (4)							
Residue	Charge	Unfavorable	HydrogenBond	Favorable	Hydrophobic	Halogen	Other

图 4 统计分析报表

点击 中的 Go Back 按钮回到 Report 界面。

点击 View Heat Map (Favorable) 链接。

打开一个个分子窗口和图形窗口，分子窗口中包含了所有的共 500 个对接构象及相应的非键相互作用信息，图形窗口中含有一张相互作用热图（图 5）。

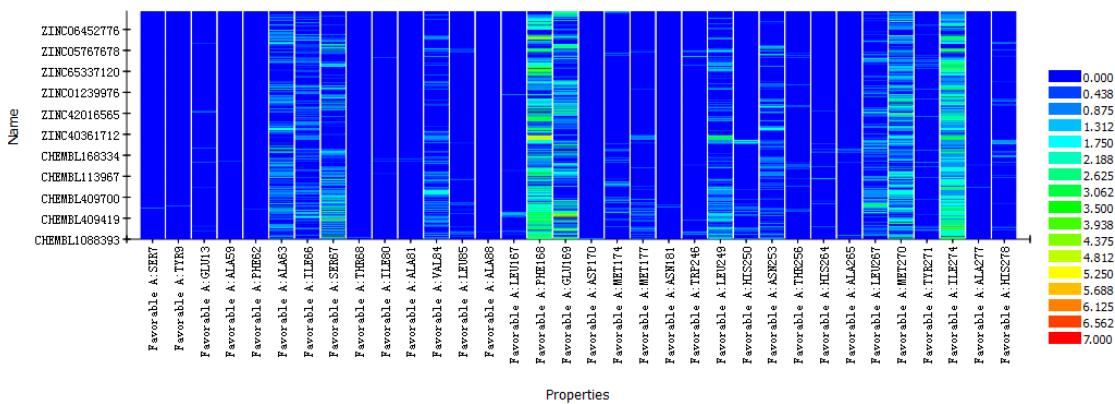


图 5 相互作用热图

从该图中可以观察到所有参与了同配体之间非键相互作用的残基信息，如 PHE168、ILE274。

在打开的分子窗口中，在表格视图中点击鼠标右键，选择 **Filter**。

在弹出的窗口中设置 **-CDOCKER\_INTERACTION\_ENERGY** 的 **Choice** 为 **Top (#)**, **Value** 为 **1**。

设置 **Apply filter to each group of identical values** 选项为 **Name**，点击 **OK**。

该分子窗口中分子会自动按上述设定条件进行过滤，最终每个对接分子只保留 -CDOCKER\_INTERACTION\_ENERGY 最高的一个对接构象，共 50 个对接构象。

注：-CDOCKER\_INTERACTION\_ENERGY 属性已通过事先的 ROC 曲线分析被证实能够较好的区分活性和非活性分子。

展开菜单栏 **Chart**，点击 **Heap Map**。

在弹出的窗口中设置 Y 轴为 **Name**，X 轴为 **Hydrophobic A:TYR9**，并 **SHIFT** 选中 **HydrogenBond A:HIS278**，点击 **OK**。

得到一张疏水、氢键相互作用热图（图 6）。

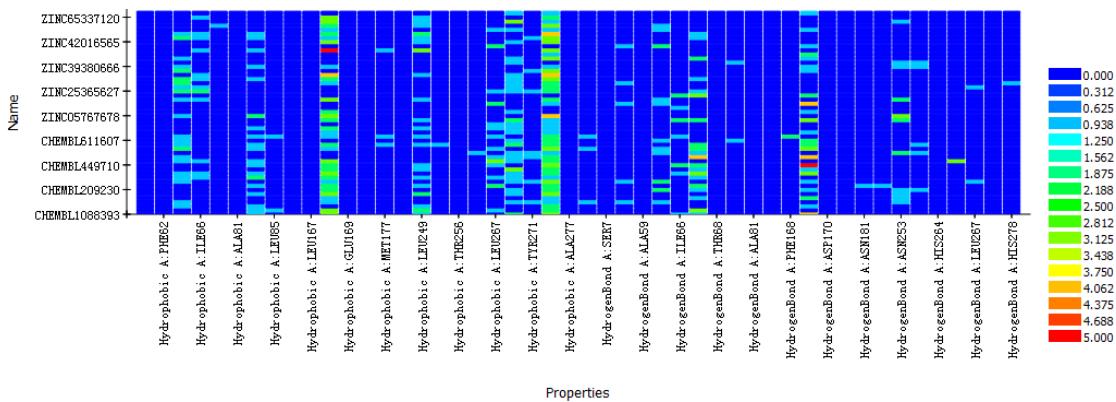


图 6 疏水、氢键作用热图

从该热图可以看出所有的参照分子都和 ILE274 形成了疏水作用，而大部分的诱饵分子也和该残基有疏水作用。而且很难轻易识别出另一种相互作用能够进一步区分参照分子和诱饵分子。然而通过接下来的操作，可以发现采用 ILE274 和另一种非键作用可以将虚筛结果提高。

切换回 filter 之后的分子窗口，选中所有 50 个对接构象，按 **CTRL+C** 进行拷贝。

打开一个新的分子窗口，按 **CTRL+V** 进行粘贴。

点击鼠标右键，选择 **Filter**。

在弹出的窗口中设置 **Hydrophobic A:ILE274** 的 **Choice** 为 **Greater than or equal**, **Value** 为 **1**;

设置 **HydrogenBond A:SER67** 的 **Choice** 为 **Greater than or equal**, **Value** 为 **1**。

点击 **OK**。

最终过滤得到 19 个对接构象，这 19 个对接构象满足以下条件：是每个配体分子对接结果中 -CDOCKER\_INTERACTION\_ENERGY 最高的对接构象，和 ILE274 可以形成疏水作用，和 SER67 可以形成氢键作用。

双击表头 **Control**。

其中有 13 个 control 分子，6 个 decoy 分子。虚筛结果有所提高，上述标准可以用来处理后续大规模未知活性分子的虚拟筛选。



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

基于片段药物设计专题:

**MCSS、Ludi、De Novo Evolution、Grow Scaffold、  
Replace Fragment**

## Discovery Studio MCSS 教程

### MCSS – 基于片段的分子对接技术

**目的:** 采用 MCSS, 以一组配体及一个明确活性位点的蛋白质为实例, 示范分子对接及结果分析操作过程。

**所需功能和模块:** Discovery Studio Visualizer client, DS MCSS

**所需数据文件:** 1EQG.csv 和 1EQG-fragments.sdf

**所需时间:** 15-20 分钟

### 介绍

目前, 分子对接技术已经成为基于结构的药物设计 (Structure-based drug design) 中一种常用的计算机辅助药物设计方法。但当对接的配体分子尺寸及柔性均很大, 采用分子对接方法很难准确的找到配体分子在靶标结构中正确的位置 (location)、取向 (orientation)、构象 (conformation)。故为了更加准确的确定配体分子中相应的片段 (基团) 在靶标结构中结合位点处的正确位置 (position), 可以采用基于片段的药物设计方法 (Fragment based drug design)。Discovery Studio 采用经典的 MCSS 算法实现 FBDD。

MCSS 作为一种片段对接的方法, 可以被用来分析靶标结构中配体结合位点的特性。采用此方法, 分子片段先被随机放置在结合位点中, 然后程序采用 CHARMm 对这些随机片段进行能量优化以找到最适的片段位置。片段采用独立的 MCSS\_Score 来打分和排序。

本教程中, 布洛芬片段被对接进入 COX-1 受体的结合位点中。对接得到的片段将与晶体结构中布洛芬的原始构象进行叠合比较。

该教程涵盖如下内容:

- 准备 MCSS
- 运行 MCSS
- 分析 MCSS 结果

### 准备 MCSS

在文件浏览器 (Files Explorer) 中, 打开 Samples | Tutorials | Receptor-Ligand Interactions | 1EQG.csv。

DS 将在一个新的 3D 窗口中打开该受体, 且受体活性中心的结合球显示在屏幕中心。(图 1)

在工具浏览器 (Tools Explorer) 中, 展开 Receptor-Ligand Interactions | Define and Edit Binding Site, 点击 Show/Hide Residues Outside Sphere, 然后点击 Show/Hide Sphere。

展开菜单栏 View | Transform, 点击 Fit To Screen 或者点击 View 工具栏中的 Fit To Screen

 以使结合位点显示在视窗正中。

这将隐藏结合位点球周围的氨基酸, 同时结合位点球本身也被隐藏。这样可以使结果的观看和分析更加容易。(图 2)

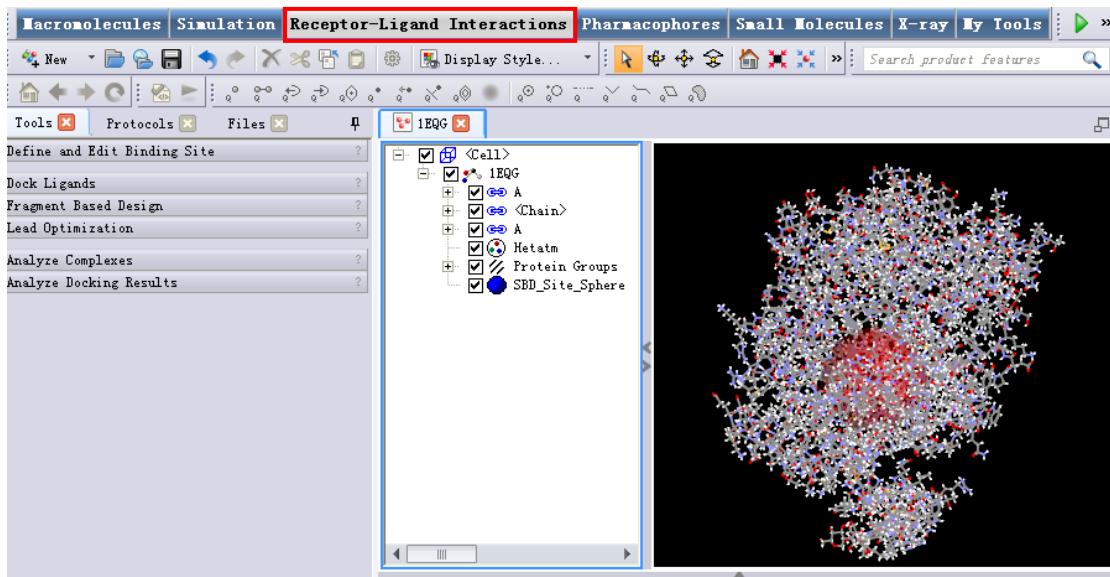


图 1 蛋白质三维结构示意图

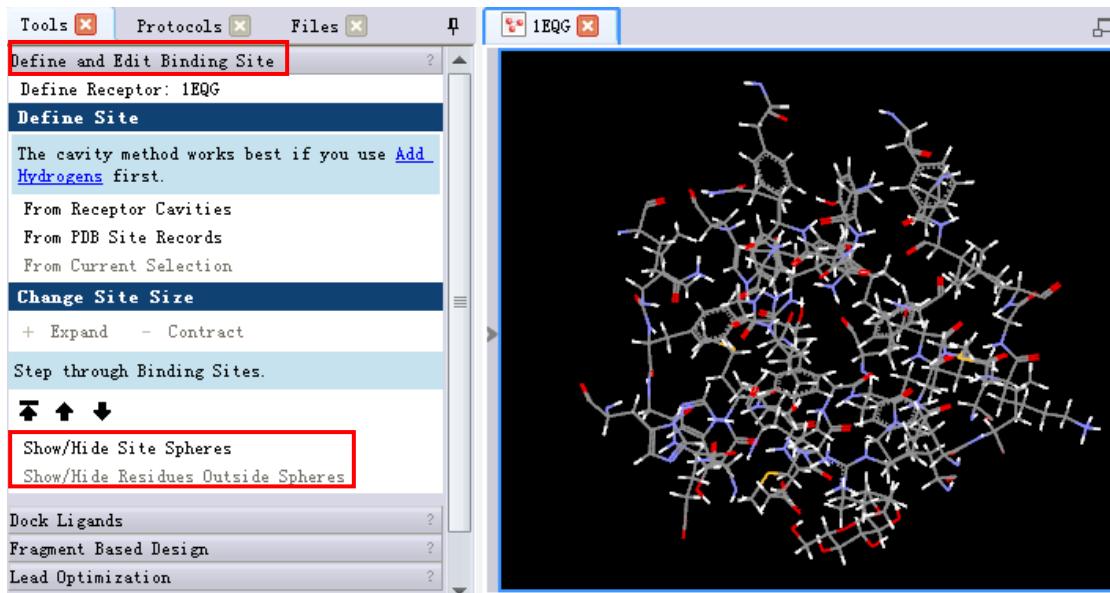


图 2 显示受体活性中心

在文件浏览器（Files Explorer）中，找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | **1EQG-fragments.sdf**。窗口中将显示两个片段：isobutyric acid 和 isopentane。（图 3）

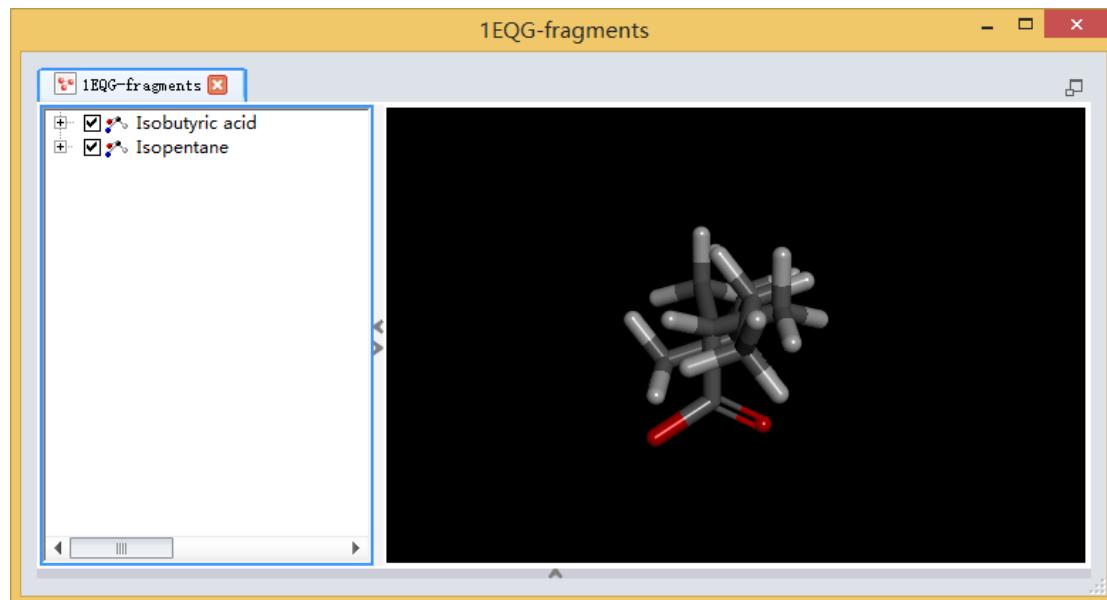


图 3 显示片段 isobutyric 和 isopentane

### 运行 MCSS

在工具浏览器 (Tools Explorer) 中, 展开 **Receptor-Ligand Interactions | Lead Optimization**, 点击 **Dock Fragments (MCSS)**。(图 4)

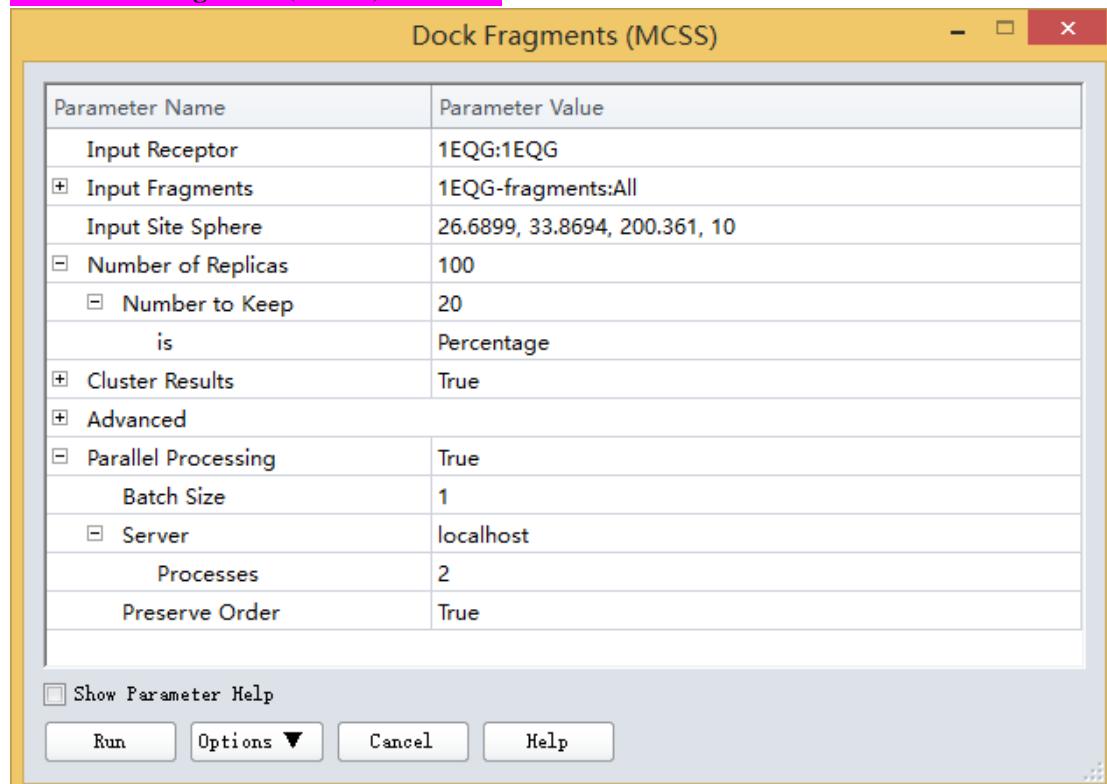


图 4 参数设置

在参数浏览器 (Parameters Explorer) 中, 点击 **Input Receptor** 设置为 **1EQG:1EQG**。点击 **Input Fragments** 参数, 选择 **1EQG-fragments>All**。这将 1EQG-fragments 窗口中所有的片段都对接入受体的活性位点中。

确保 **Input Site Sphere** 参数为“26.6899, 33.8694, 200.361, 10”。

展开 **Number of Replicas** 和 **Number to keep** 参数。将 **Number to Keep** 设置为 **20**, 同时将 **is** 参数设置为 **Percentage**。这将限制对接结果中所报告的 fragments 数目为打分排名在前 20%。可以通过设置并行计算将计算时间减少。如果计算机使用多处理器或多核, 将 **Parallel Processing** 参数设置为 **True**。

展开 **Parellel Processing** 参数, 并设置 **Batch Size** 为 **1**。通过设置 **Batch Size** 为 1, 每个处理器会单独处理 1 个片段。(图 4)

点击 **Run** 按钮, 等待计算完毕。完成该任务大约需要 3 分钟。

## 分析 MCSS 结果

从菜单栏中, 选择 **Window | Close All**, 关闭所有已打开的窗口。

在任务浏览器 (Job Explorer) 中, 双击 Dock Fragments(MCSS)任务, 打开 **Report** 界面。

在 Report 界面中, 在 **Results** 部分, 点击 **Top Scoring Fragments in Cluster** 链接。

打开一个新的窗口, 显示聚类分析之后的片段对接结果。

按 **CTRL+G** 打分子显示窗口。

在文件浏览器 (Files Explorer) 中, 选中 **Samples | Tutorials | Receptor-Ligand Interactions |1EQG- ibuprofen.sd**, 并将其拖拽至上述窗口中。

在表格视图 (Table View) 中, 选中最后一栏 (ibuprofen), 同时将 **Visibility** 设置为 **True**。

设定 **Visibility Locked** 为 **True**, 从而在视窗中显示并锁定该配体。

选中表格视图 (Data Table), 按住 **CTRL+Down** 浏览各对接的片段。(图 5)

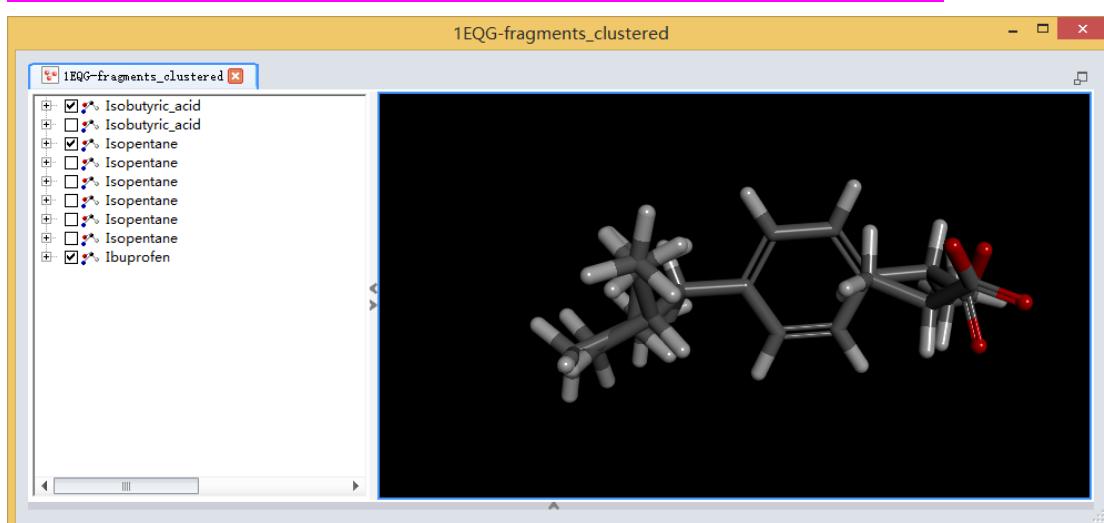


图 5 对接的片段与布洛芬原始构象进行比较

结果中, 对接的片段与布洛芬的晶体结构可以很好的叠合, 这表明 MCSS 可以准确找到疏水片段 (isopentane) 以及亲水片段 (isobutyric acid) 在受体结合位点中的位置。

## Discovery Studio LUDI 教程

### 使用 DS\_LUDI 进行全新药物设计

**所需功能和模块:** Discovery Studio client, DS\_Ludi, DS\_De novo evolution

**所需数据文件:** 1fvv\_protein.csv, 1fvv\_ligand.csv

**所需时间:** 10 分钟

#### 介绍

随着蛋白质结构解析技术的发展以及人类基因组计划的开展,对蛋白质晶体结构的研究发展迅速。怎样有效的分析这些靶点的信息以及设计出相匹配的分子是计算化学家面临的很大挑战。基于分子对接的数据库搜索方法一般是基于已有的数据库,分子结构是不发生变化的。从头药物设计则是一个全新的概念。

Ludi 是应用最为广泛的从头设计方法之一。使用 Ludi 来发现新的具有潜在活性的化合物,可节省研究者大量的时间。Ludi 强大的设计工具允许使用者在实验分析之前模拟筛选,并允许对已有的化合物进行改造。Ludi 易于操作,它包含有 drug-like 片段库,同时也允许用户将自己的分子片段加入到片段库中。

DS 中的 De Novo Protocols 使用的即为 Ludi 的算法,它帮助我们找到新的分子骨架或者修改已有的分子骨架来提高小分子的活性。De Novo 可以帮助我们进一步发展已经商业化的分子骨架数据库,并且通过对药物候选分子衍生物的打分来修饰已知的分子配体。这些 Protocol 包括 De Novo Receptor 可以用来寻找合适受体结合位点的潜在碎片,这种方法的优点是可以快速有效的对数据库中大量的碎片进行筛选; De Novo Link 可以将找到的活性片段连接或在已有骨架的基础上添加新的片段分子; De Novo Evolution 也称为 AutoLudi,主要用于进行 Me better 药物设计工作,不仅可以搜索活性片段分子并且可以将片段和骨架自动连接直接产生新的分子。

本教程中,用从头药物设计的方法设计细胞周期蛋白依赖性激酶 2 (CDK2) 抑制剂。

- 准备蛋白受体和骨架配体分子
- 运行 De Novo Receptor 及结果分析
- 运行 De Novo Link 及结果分析
- 运行 De Novo Evolution 及结果分析

#### 准备蛋白受体和骨架配体分子

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples| Tutorials| Receptor-Ligand Interactions| 1fvv\_protein.csv 文件。

DS 将在一个新的 3D 窗口中打开该受体,且受体活性中心的结合球显示在屏幕中心 (图 1)。

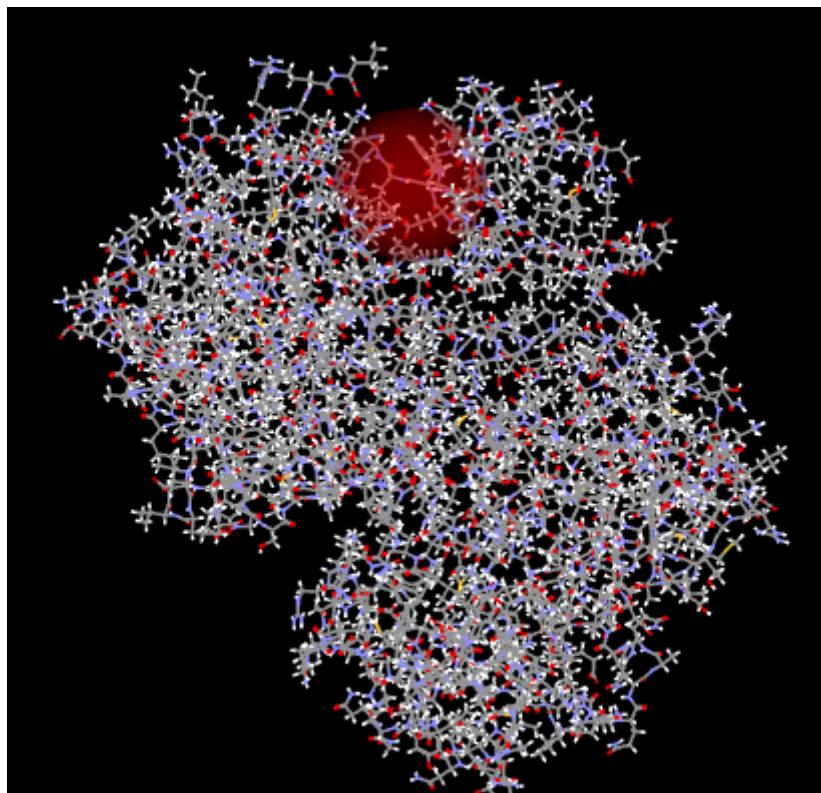


图 1 定义受体活性口袋

在工具浏览器(Tools Explorer)中, 展开 **Receptor-Ligand Interaction|Define and Edit Binding site** 工具组, 点击 **Show/Hide Residues Outside Spheres** 隐藏结合位点球周围的氨基酸, 再点击 **Show/Hide Sphere** 隐藏结合球。

展开菜单栏 **View | Transform**, 点击 **Fit To Screen** 或者点击 View 工具栏中的 **Fit To Screen**  
。

以使结合位点显示在视窗正中 (图 2), 便于结果的观看和分析。

在文件浏览器 (Files Explorer) 中, 找到并双击打开 **Samples | Tutorials | Receptro-Ligand Interactions | 1fvv\_ligand.dsv** 文件, 窗口中将显示配体分子结构 (图 3)。

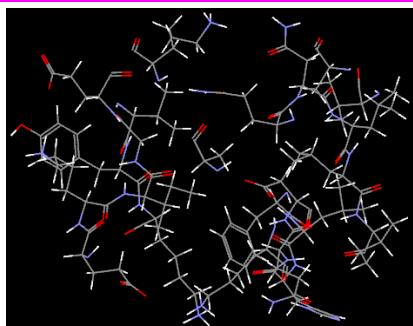


图 2 受体准备



图 3 配体准备

### 运行 De Novo Evolution 及结果分析

在工具浏览器 (Tools Explorer) 中, 展开 **Receptor-Ligand Interactions | Fragment Based Design**, 点击 **De Novo Evolution**, 打开参数面板。

确认 **Input Receptor** 设置为 **1fvv\_protein:1fvv\_protein**。

点击 Input Ligand Scaffold 参数，选择 1fvv\_ligand:1fvv\_ligand。

确保 Input Site Sphere 参数为 -9.42369, 207.623, 113.547, 8。

点击 Input Fragment Libraries 参数，浏览，选择 fragment\_link.inp（图 4）。

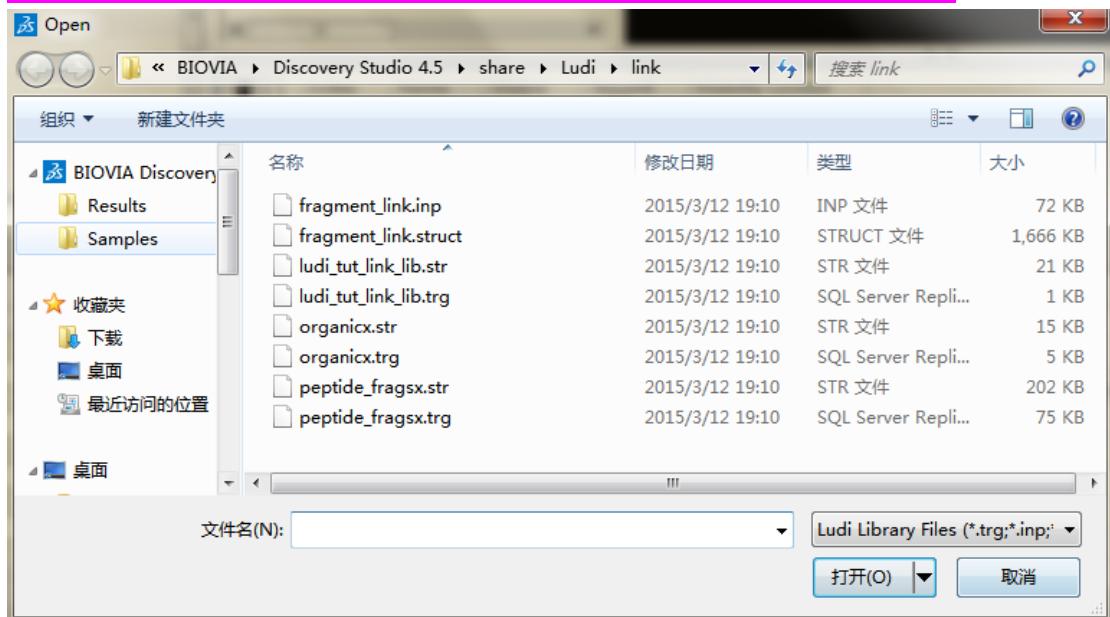


图 4

为了减少计算时间，本教程通过修改默认参数来限制 hits 数。

展开 Fitting 参数组，设置 Maximum RMSD 值为 0.3。

注意：本例中使用预先产生的片段库，该库可以通过 De Novo Library Generation protocol 产生。每个片段库包括两个文件，structure file 和 target file，同时产生，选其一即可。

其它参数均使用缺省值（图 5）。

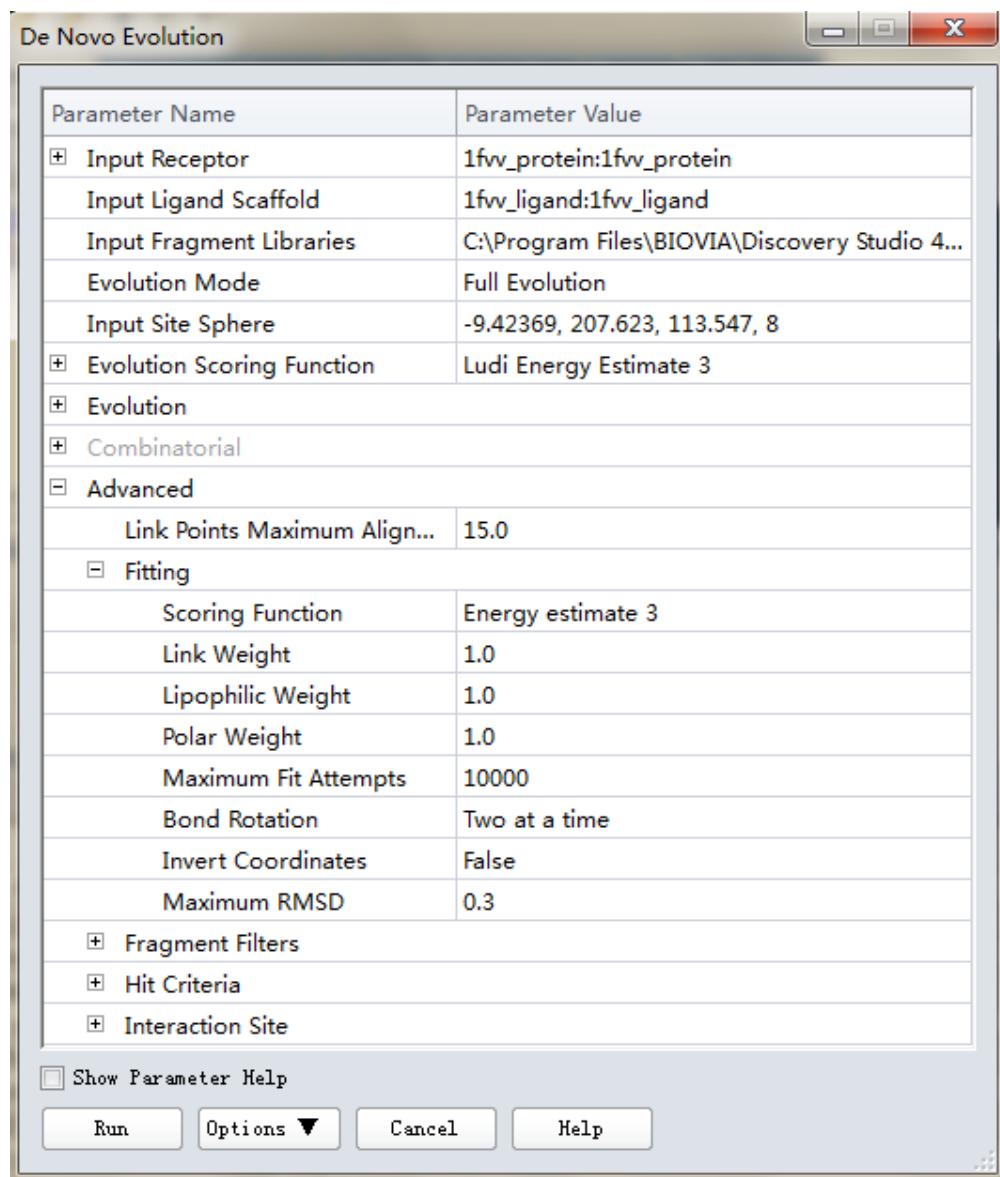


图 5 参数设置

点击 **Run** 按钮，等待计算完毕。

完成该作业大约需要 1 分钟。

待作业完成后，DS 会自动跳出一个新的以 1fvv\_protein 命名的分子窗口。

点击表格视图中 、 和 将生成的各个新的分子在受体结合位点中的可结合位置在图形窗口中显示出来。

在工具浏览器（Tools Explorer）中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

受体与配体分子间的非键相互作用会以虚线显示出来（图 6）。

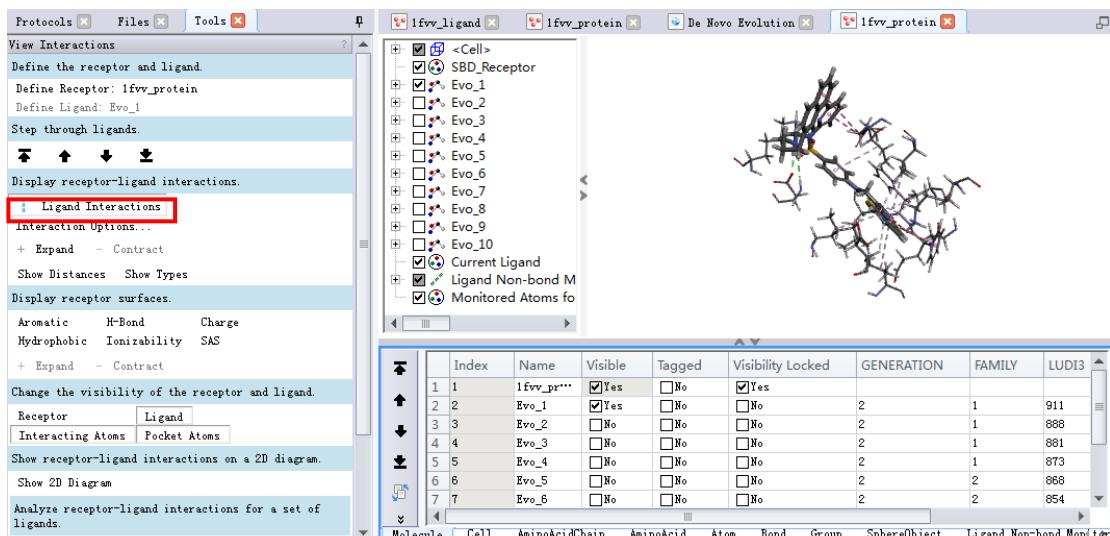


图 6 显示受体与配体间的非键作用

点击上述 **View Interaction** 工具面板下的 按钮可以观察生成的各个新的分子在受体结合位点中的结合位置，及其与受体间的非键作用。

我们会发现起始的分子骨架在不同位置上有基团取代。我们可以通过重叠起始的分子骨架观察取代基团所在位置。

将 1fvv\_ligand.csv 文件从文件浏览器（Files Explorer）中拖到当前激活的窗口中。并在 Data Table View 中将 1fvv\_ligand 的 Visibility Lock 设为 Yes。（图 7）

	Index	Name	Visible	Tagged	Visibility Locked
1	1	1fvv_protein	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input checked="" type="checkbox"/> Yes
2	2	Evo_1	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
3	3	Evo_2	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
4	4	Evo_3	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
5	5	Evo_4	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
6	6	Evo_5	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
7	7	Evo_6	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
8	8	Evo_7	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
9	9	Evo_8	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
10	10	Evo_9	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
11	11	Evo_10	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
12	12	1fvv_ligand	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input checked="" type="checkbox"/> Yes

图 7 Data Table

注意：在 table 里包含了 GENERATION 和 FAMILY 两列，分别代表了在运算过程中添加了多少代的取代基，以及在每个对接步骤中打分的排名情况。也可以通过在 Graphics View 中重叠结果文件中的 interaction map (在输出文件 Ludi.interactionmap.mol2) 来进一步观察取代基的位置。

在这个教程中，我们使用了 **Full Evolution mode**。在这种模式下，每一步的迭代过程只生长一个取代基。与之相比，**Quick mode** 每步迭代都有多个取代基连接在骨架上。

可以将 **Evolution Mode** 参数设置为 **Quick**，其它参数不变，重新跑一次 De Novo Evolution 运算，然后比较下结果（图 8）。

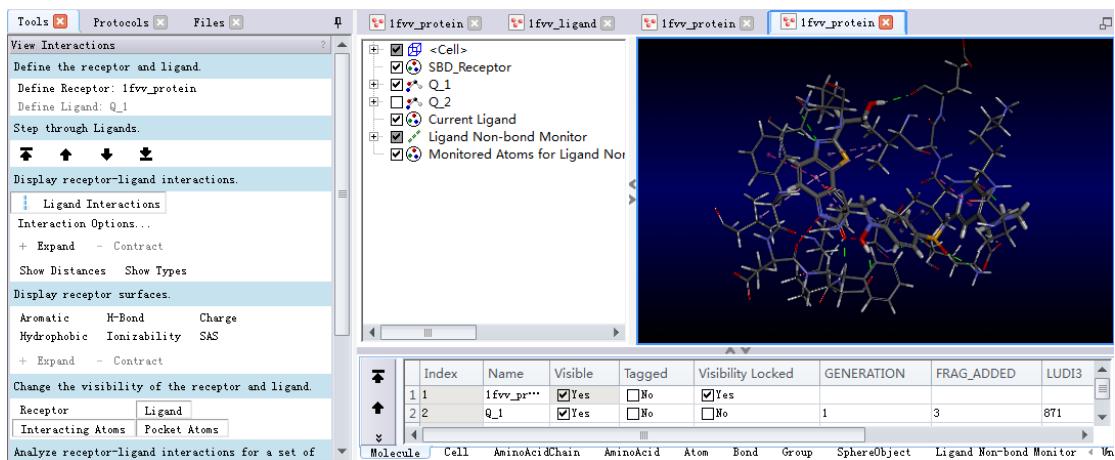


图 8 Quick mode 的计算结果

### 运行 De Novo Receptor 及结果分析

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interactions | Lead Optimization**，点击 **De Novo Receptor** 工具栏，打开参数面板。

确认 **Input Receptor** 设置为 **1fvv\_protein:1fvv\_protein**。

点击 **Input Fragment Libraries** 参数，浏览，选择 **fragment\_all.inp**。

注意：本例中使用预先产生的片段库，该库可以通过 **De Novo Library Generation protocol** 产生。每个片段库包括两个文件，structure file 和 target file，同时产生，选其一即可。

为了减少计算时间，本教程通过修改默认参数来限制 hits 数。

展开 **Hit Criteria** 参数组，设置 **Minimum Score** 为 **350**。

展开 **Fitting** 参数组，设置 **Maximum RMSD** 值为 **0.3**。

其它参数均使用缺省值（图 9）。

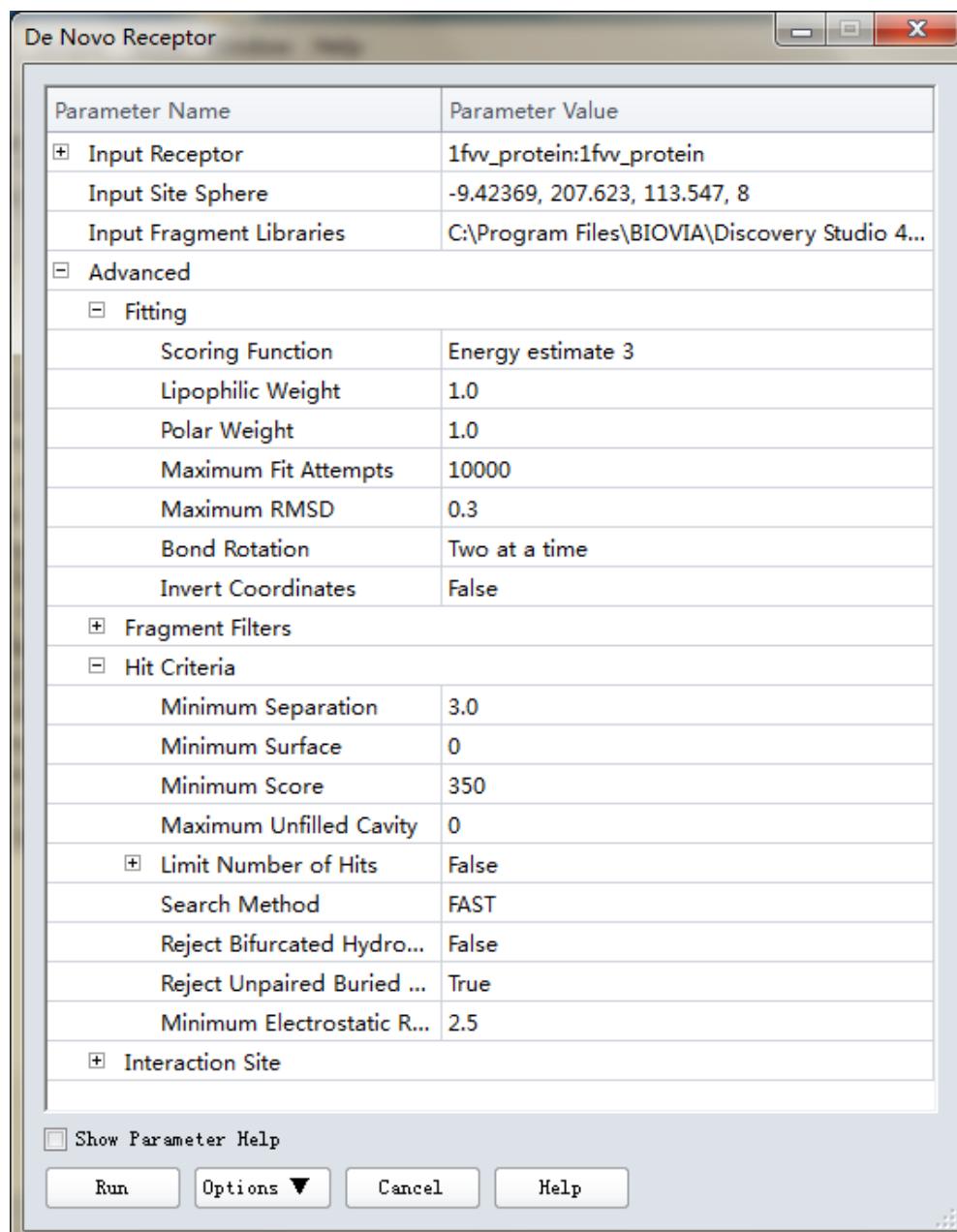


图 9 De Novo Receptor 参数设置

点击 Run 按钮，等待计算完毕。

完成该作业大约需要 1 分钟。

待作业完成后，DS 会自动跳出一个新的以 1fvv\_protein 命名的分子窗口。

点击表格视图中 、 和 将各片段在受体结合位点中的可结合位置在图形窗口中显示出来。

展开菜单栏 Files |Insert From...，点击 Files，找到 Samples| Tutorials | Receptro-Ligand Interactions | 1fvv\_ligand.dsv。

将天然配体显示在同一窗口中。

在表格视图（Data Table）中将 1fvv\_ligand 的 Visibility Lock 设为 Yes。（图 10）这样就会在视图窗口中锁定该天然配体。

	Index	Name	Visible	Tagged	Visibility Locked	Calculate Charge
1	1	1fvv_protein	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	2	TRP	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
3	3	A28	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
4	4	G17	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
5	5	G32	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
6	6	G33	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
7	7	H36	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
8	8	H76	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
9	9	AAO	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
10	10	AHO	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
11	11	E93	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
12	12	X13	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
13	13	X50	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
14	14	X84	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
15	15	X86	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
16	16	Y34	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No
17	17	1fvv_ligand	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No

图 10 Data Table

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

受体与配体分子片段间的非键相互作用会以虚线显示出来（图 11）。

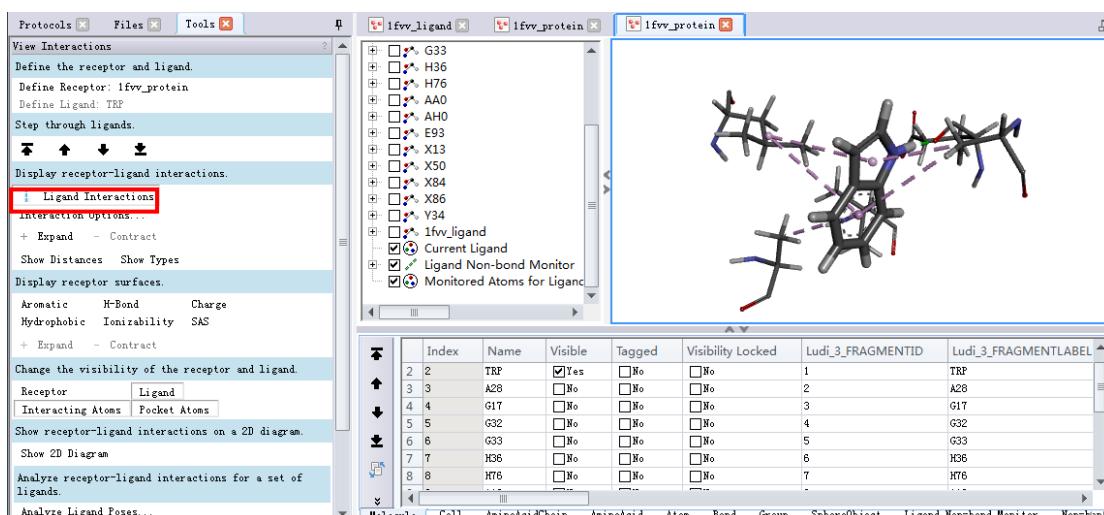


图 11 显示受体与配体间的非键作用

点击上述 **View Interaction** 工具面板下的 按钮可以观察不同的片段在受体结合位点中的可结合位置、同天然配体的叠合情况以及其与受体间的非键作用。

注意：也可在 **View Interaction** 工具面板下点击 **Interaction Options...**，来显示指定的非键作用类型。

此外，在表格里包含了各分子片段的 Ludi\_3\_RMSD 等碎片信息，以及每个碎片的打分情况。也可以通过在 Graphics View 中重叠结果文件中的 interaction map (输出文件中 Ludi.interactionmap.mol2 文件) 来进一步观察碎片的位置。

双击结果 Output 文件夹中的 Ludi.interactionmap.mol2 文件或者点击 Jobs Explorer 中该任务下的 **Interaction Map** 链接，将显示相互作用位点图。

这是分析受体活性口袋内的氨基酸残基的化学特征示意图，其中“蓝白线”代表此位置与受体匹配的相应配体基团应为氢键供体；“灰红线”则代表此位置的配体基团为氢键受体；灰色的点代表配体中的碳原子。（图 12）。

**注意：**如果找不到 Output 文件夹的路径，则可以在任务浏览器（Jobs Explorer）中，选中相应的任务，然后点击鼠标右键选择 Locate 即可找到相应 Output 文件夹的位置。

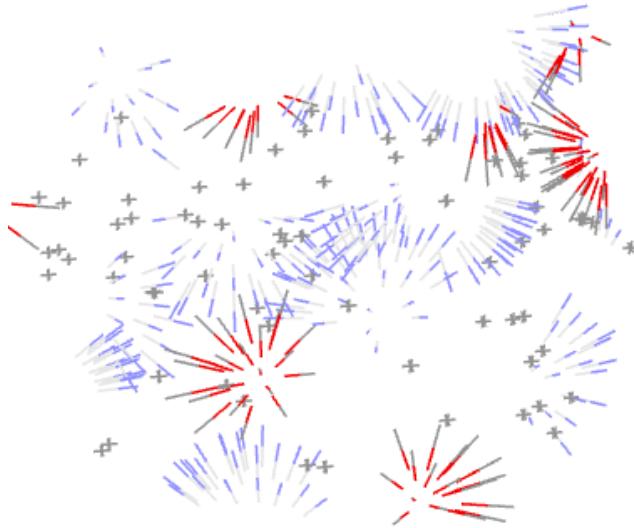


图 12 受体活性口袋内的氨基酸残基的化学特征

注意到有些对接片段和天然配体结构叠合的很好，而有些对接片段所在位置则没有天然配体，这就表明受体结合位点并没有被天然配体完全占据，而这也也可以被用来优化配体分子以加强同受体的结合。

### 运行 De Novo Link 及结果分析

在工具浏览器（Tools Explorer）中，展开 **Receptor-Ligand Interactions | Fragment Based Design**，点击 **De Novo Link**，打开参数面板。

确认 **Input Receptor** 设置为 **1fvv\_protein:1fvv\_protein**。

点击 **Input Ligand Scaffold** 参数，选择 **1fvv\_ligand:1fvv\_ligand**。

点击 **Input Fragment Libraries** 参数，浏览，选择 **fragment\_link.inp**。

**注意：**本例中使用预先产生的片段库，该库可以通过 De Novo Library Generation protocol 产生。每个片段库包括两个文件，structure file 和 target file，同时产生，选其一即可。

确保 **Link Points** 参数设置为 **Use all hydrogen atoms**。

所有 H 原子都是片段连接的潜在位置。

为了减少计算时间，本教程通过修改默认参数来限制 hits 数。

展开 **Advanced|Fitting** 参数组，设置 **Fitting Maximum RMSD** 值为 **0.3**。

其它参数均使用缺省值（图 13）。

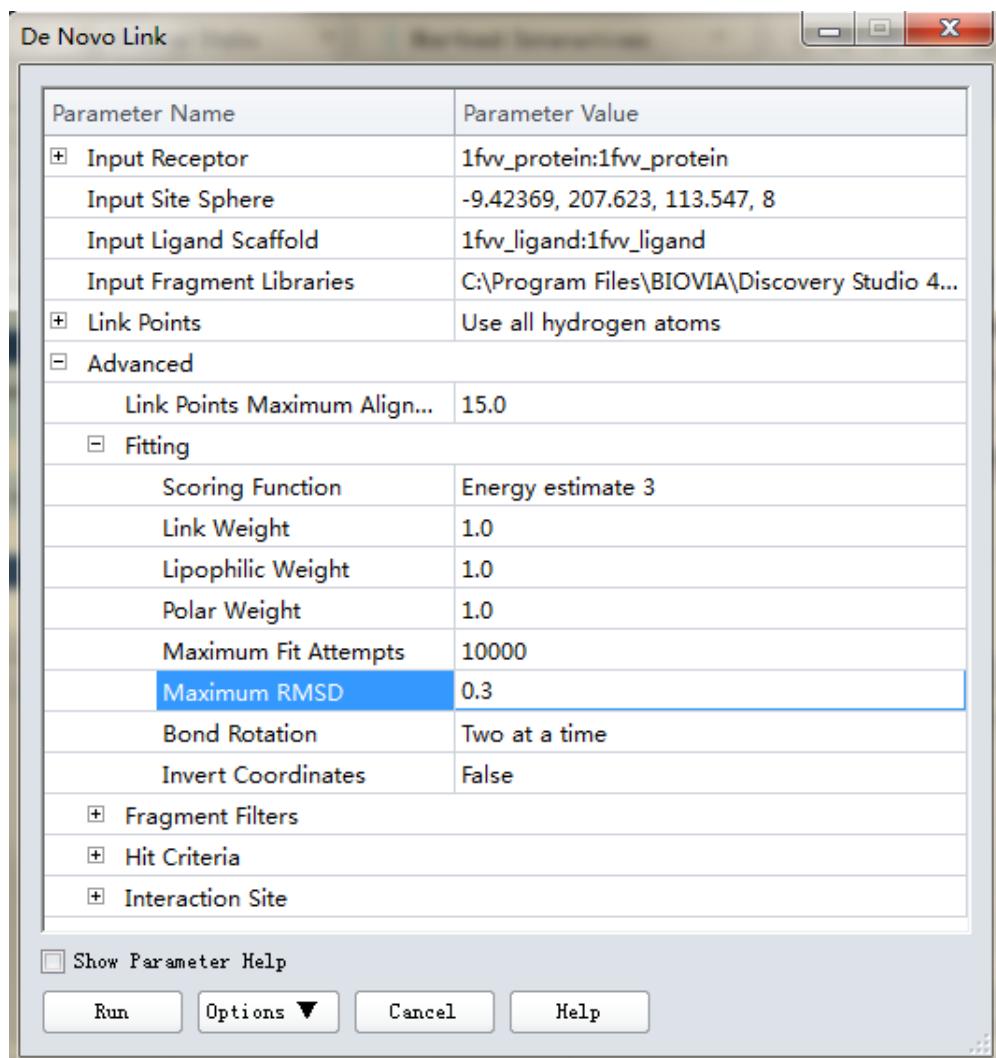


图 13 参数设置

点击 **Run** 按钮，等待计算完毕。

完成该作业大约需要 1 分钟。

待作业完成后，DS 会自动跳出一个新的以 1fvv\_protein\_complex 命名的分子窗口。

点击表格视图中 、 和 将各片段在受体结合位点中的可结合位置在图形窗口中显示出来。

在工具浏览器（Tools Explorer）中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

受体与配体分子片段间的非键相互作用会以虚线显示出来（图 14）。

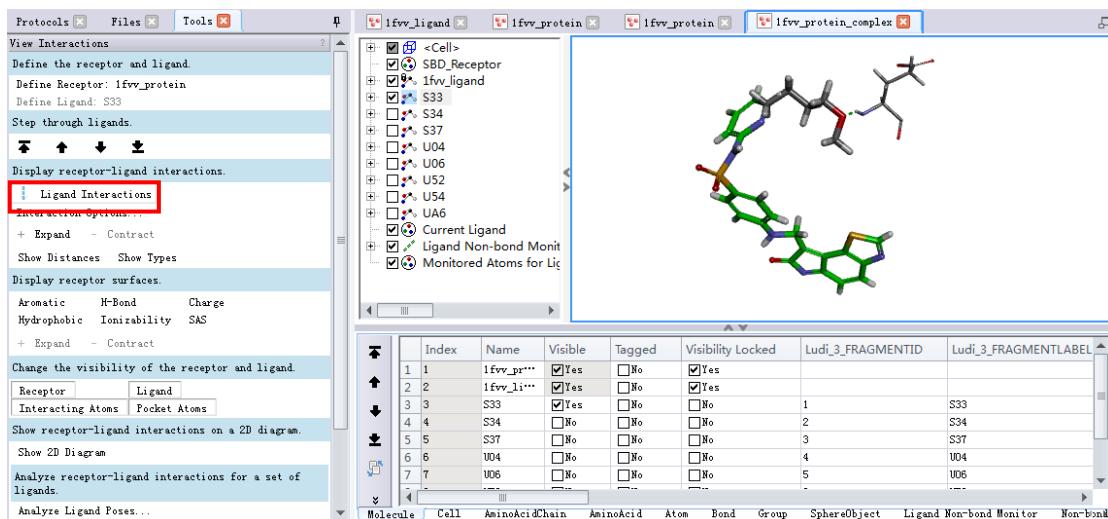


图 14 显示受体与片段间的非键作用

点击上述 View Interaction 工具面板下的 按钮可以观察不同的片段在受体结合位点中的可结合位置及其与受体间的非键作用。

双击结果 Output 文件夹中的 Ludi.interactionmap.mol2 文件或者点击 Jobs Explorer 中该任务下的 Interaction Map 链接，将显示分析化学特征的结果。

将该 interaction map 拷贝至 De Novo Link 结果窗口中，可以查看配体分子片段同 interaction map 的叠合情况，进而分析片段在受体结合位点中的位置。

图 15 为分析受体活性口袋内的氨基酸残基的化学特征示意图，其中“蓝白线”代表此位置与受体匹配的相应配体基团应为氢键供体；“灰红线”则代表此位置的配体基团为氢键受体；灰色的点代表配体中的碳原子；黄色的双行线代表片段的可连接位置。由于“Link Points”选择了程序默认的“Use all hydrogen atoms”，因此黄色线即为先导分子氢原子位置。

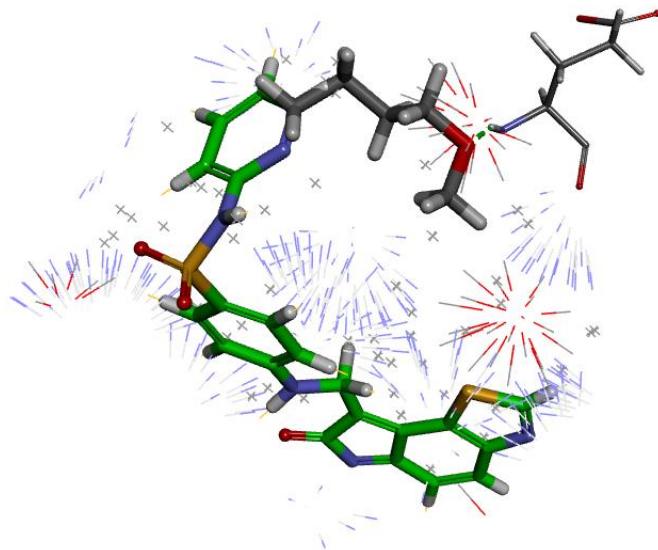


图 15 先导分子与化学特征的叠合

注：片段终端原子连接至骨架分子的 H 原子处，这也表明可以将片段分子和骨架分子进行共价连接，提高配体分子同受体的结合。

## Discovery Studio Grow Scaffold 教程

### Grow Scaffold – 活性位点先导化合物优化

**目的:** 采用 Grow Scaffold, 在蛋白活性位点内从骨架进行先导化合物优化

**所需功能和模块:** Discovery Studio Client, DS CHARMM Lite, and DS CATALYST CONFORMATION.

**所需数据文件:** 1kv2\_complex.csv, 2yis\_scaffold.sdf, diverse\_sulfides.sdf, sulfide\_halfreaction.rxn.

**所需时间:** 30 分钟

#### 介绍

先导化合物优化是一个复杂过程, 为了得到一个临床前候选药物, 通常需要对有前景化合物及骨架不断地进行化学结构优化, 以提高化合物的活性、选择性、生物利用度、药效及药代动力学性质, 并降低毒性。该过程通常是药物开发过程中的瓶颈。通过使用实用且有效的先导化合物优化软件, 根据可获取的化合物试剂的快速推荐出容易合成的候选化合物的结构。基于结构的先导化合物优化主要集中于蛋白靶点的活性位点的化合物结构的设计。Grow Scaffold 工具可以根据蛋白靶点活性位点的特点, 通过基于化学反应的原位生长 (reaction-based *in situ* enumeration) 方法来找出那些能够产生潜在化合物的试剂, 并对它们进行打分排序。

p38 $\alpha$  是一个典型的丝氨酸/苏氨酸蛋白激酶, 属于丝分裂原活化蛋白激酶 (mitogen activated protein kinase) 家族。它在内皮、免疫和炎性细胞中广泛表达, 在肿瘤坏死因子和白细胞介素-1 等促炎症细胞因子产生的调控中扮演着重要作用。实验已经证明, 选择性地抑制其中任何一个细胞因子都能够有效治疗各种炎症和免疫疾病, 如风湿性关节炎、炎症性肠病, 败血性休克, 和骨质疏松症。二芳基脲是 p38 $\alpha$  的一个先导化合物, Boehringer Ingelheim 首先对它开展了先导化合物优化, 随后 Pfizer 也开展了相关研究 (图 1)。

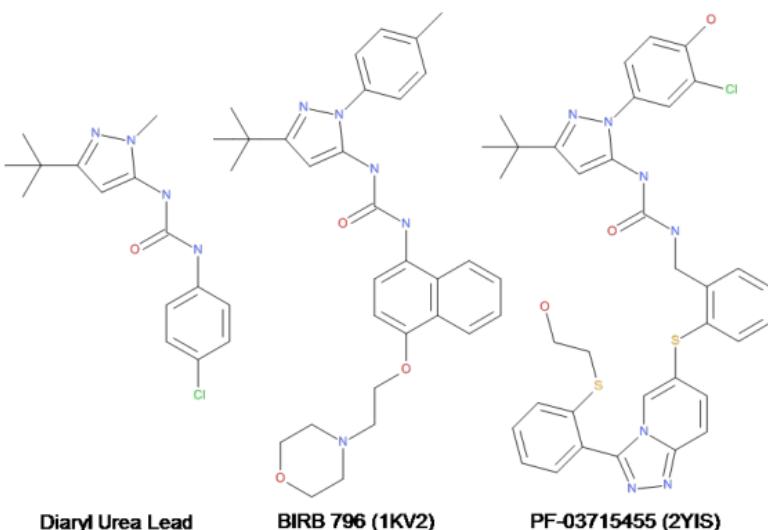


图 1 p38 $\alpha$  的先导化合物及优化过的先导化合物

在本教程中你将使用 Grow Scaffold 流程来重现 Boehringer Ingelheim 及 Pfizer 通过结构优化所发现的化合物。

本教程分为两部分：

- (一) 使用基于反应的原位生长方法产生优化过的先导化合物
- (二) 使用自定义反应来产生先导化合物

### (一) 使用基于反应的原位生长方法产生优化过的先导化合物

#### 1. 执行优化计算

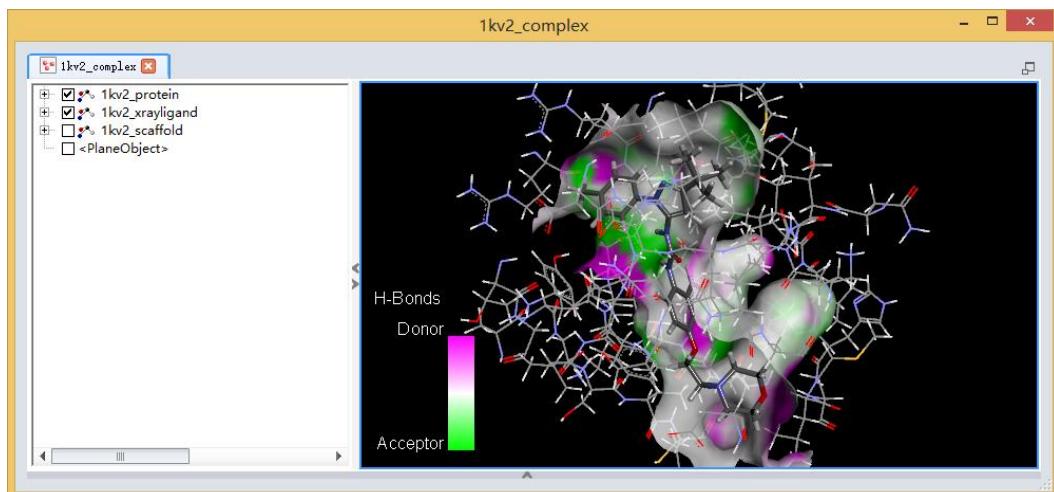


图 2 蛋白质三维结构示意图

复合物结构采用 PDB 号为 1KV2 的晶体结构，本教程中所采用的该结构文件已经从 PDB 库中下载并经过 Prepare Protein 模块处理。

在文件浏览器（File explorer）中找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | 1kv2\_complex.dsv。

该蛋白将在一个新的分子窗口中出现。与 1KV2 结合的配体(1kv2\_xrayligand)是 BIRB 796，它是第一报道的 p38 $\alpha$ II 型抑制剂。在本教程中,我们将在二芳基脲的萘环上引入一个吗啉乙氧基得到 BIRB 796。

展开窗口左侧系统视图（Hierarchy View）（图 2），勾掉 1kv2\_xrayligand 复选框并选择 1kv2\_scaffold. 选中脲基对位上萘基上的氢 (H31)。

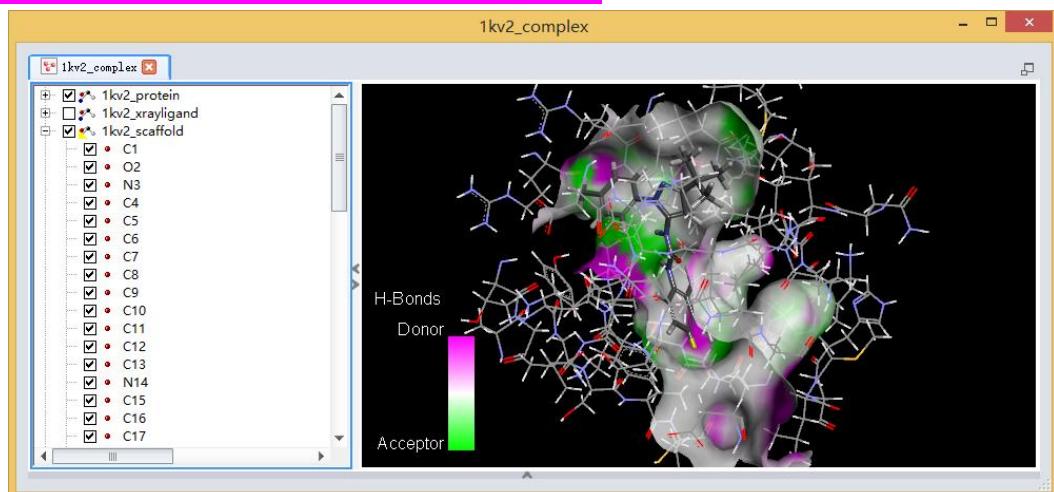


图 3 Scaffold 与受体分子

在工具浏览器（Tools Explorer）中，打开 **Receptor-Ligand Interactions | Lead Optimization** 工具栏，点击 **Grow Scaffold...** 打开 protocol 对话框。

也可以在流程浏览器（Protocol explorer）中的 **Recepotor-LiandInteractions|Fragment Based Design|Grow Scaffold** 打开 protocol。

在参数浏览器中，点击 **Input Receptor** 参数，从下拉列表中选择 **1kv2\_complex:1kv2\_protein**。

设置 **Input Ligand Scaffold** 参数为 **1kv2\_complex:1kv2\_scaffold**。

点击 **Scaffold Grow Points** 选项，在下拉菜单中选择 **Create New Group From Selection..**

点击 **Reactions** 参数，勾掉 **Amide Synthesis** 选项并选择 **Williamson Ether Synthesis** 选项。

点击 **Reagents** 参数，勾选 **Halides** 和 **Alcohols**。

其他参数为默认，点击 **Run** 按钮运行。可以点击 **Background** 让作业后台运行，等待作业结束。

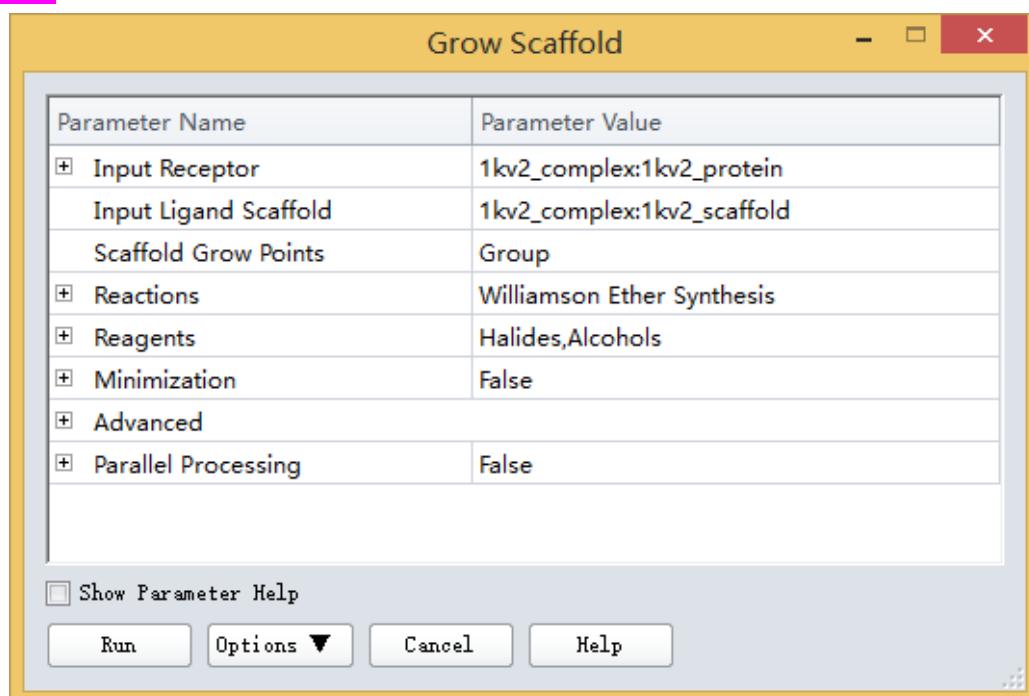


图 4 参数设置

## 2. 查看计算结果

点击作业浏览器（Job explorer）中刚完成的作业，点击 **View Results** 链接。

Protocol Name	Saved	Status	Details	Elapsed
Grow Scaffold	<input checked="" type="checkbox"/> Yes	Success	1kv2_scaffold: 460 ligands	0:01:48
	<a href="#">Report</a>			
	<a href="#">Enumerated Ligands</a>			
	<a href="#">View Results</a>			

图 5 查看计算结果

在打开的新的分子窗口 **1kv2\_protein** 中确保系统视图和表格视图打开(可以分别使用快捷键 Crtl+H 和 Crtl+T 进行开关)，在系统视图和表格视图中都可以观察到该窗口中包含先导化合物优化得到的 444 个分子。

点击表格视图中的 键和 键，观察所产生的配体分子同受体分子的结合模式。

## 3. 分析优化结果

## 非键相互作用的直观显示与分析

在工具浏览器 (Tools Explorers) 中, 展开 **Receptor-Ligand Interactions | View Interactions**, 点击 **Ligand Interactions**。

在视图窗口中, 受体原子与所选配体间的非键相互作用会通过不同颜色的虚线显示出来, 且只有参与了同配体之间的相互作用的残基才会显示。 (图 6)

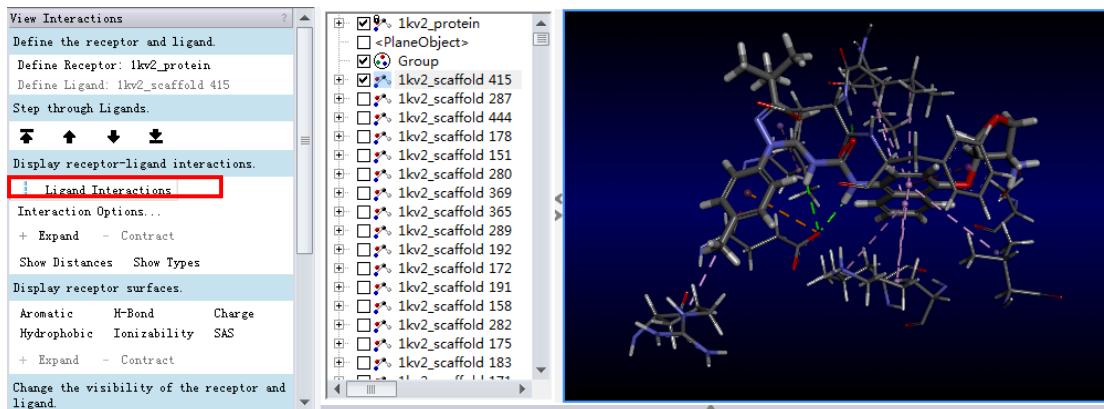


图 6 显示蛋白-配体之间非键相互作用

点击上述 **View Interaction** 工具面板下的 按钮可以观察优化后的分子同蛋白之间的各种非键相互作用。

为了更好的观察受体分子与配体对接 pose 间的相互作用, 可以对体系进行旋转以获得最佳的观赏角度。

化合物的 Pareto sorting 方法进行排序, 主要指标有:与蛋白形成的相互作用数目(最大化), 违法 lipiskk 五规则数目(最小化), 与受体发生碰撞数(最小化), 片段的新颖性(最大化)的数量. 数据列表中也显示了配体与蛋白质间的相互作用, 包括相互作用类型, 蛋白质残基, 作用距离等。

双击表格视图中 **Fragment Name R1** 的表头, 对该列按名称进行排序。

找到 **2-MORPHOLINOETHANOL** 并选中该行, 点击 (数据窗口左侧) 查看分子 (图 7)。

或者采用 filter 方法。

在表格视图中点击鼠标右键, 选择 **Filter...**

在弹出的窗口中, 设置“**Fragment Name R1**”的 **Choice** 为 **Contains**。

设置 **Value** 为 **MORPHO**, 点击 **OK**。

表格视图会自动按上述设置的限制条件过滤得到符合的分子。

然后选中 **2-MORPHOLINOETHANOL**。

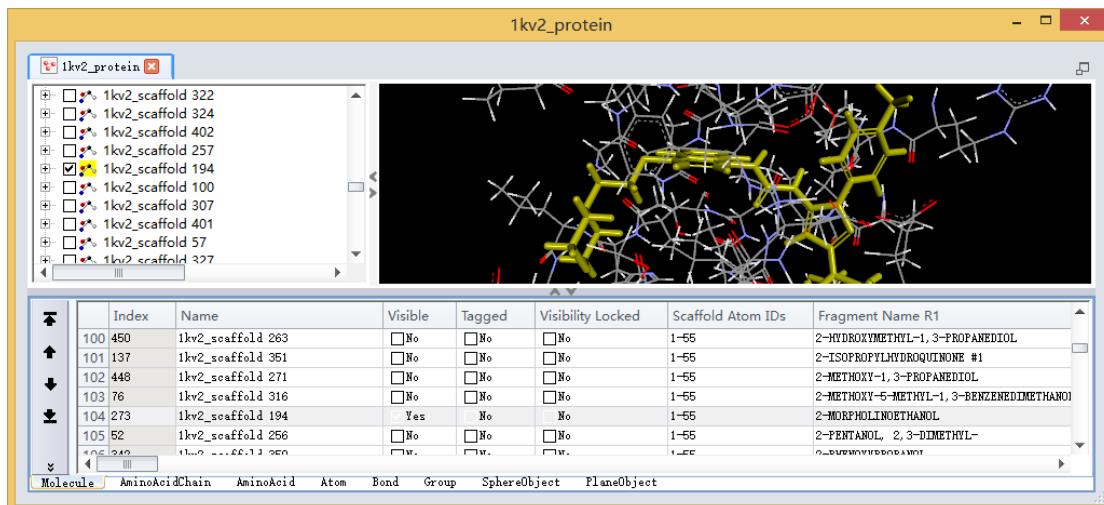


图 7 通过 grow scaffold 得到的 BIRB 796

## (二) 使用自定义反应产生先导化合物

尽管 Grow Scaffold 的所自带的反应类型已经能够概括药物化学家常用的化学反应，不过你还可以根据自己的需要，轻松地定制自己的反应。接下来我们将介绍如何使用非默认反应类型，来优化优化另一个 p38 $\alpha$  的抑制剂。

此外，由于 p38 $\alpha$  激酶 linker 区上的残基 MET109，具有重要作用，我们把它设为必须作用的残基，计算结果中只有与这个残基有相互作用的新分子才会被返回。

### 1. 执行计算

在文件浏览器 (File explorer) 中找到并双击打开，Samples | Tutorials | Receptor-Ligand Interactions | diverse\_sulfides.sd 文件。

再切换回刚才的 1kv2\_complex 窗口。

选择菜单栏中的 File | Insert From | File...，插入 Samples | Tutorials | Receptor-Ligand Interactions | 2yis\_scaffold.sd 文件。

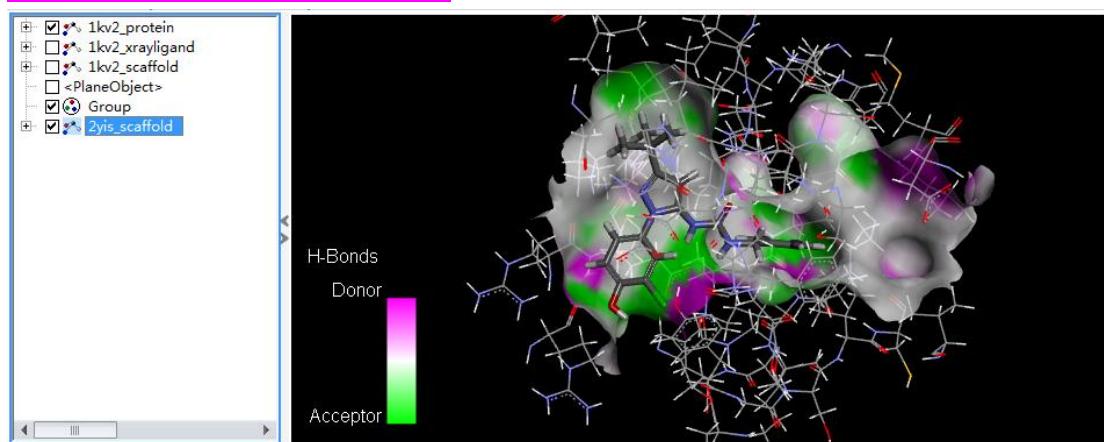


图 8 放入新的 scaffold (2yis\_scaffold)

在系统视图中，勾掉 1kv2\_scaffold 复选框，并选择 2yis\_scaffold 使它显示出来。(图 8)

在 2yis\_scaffold 配体中，选中受体表面开口方向上脲基邻位上的苯基上的氢 (H19)。

在工具浏览器 (Tool Explorer) 中，打开 Receptor-Ligand Interactions | Lead Optimization

工具栏，点击 **Grow Scaffold...** 打开流程对话框。

设置 **Input Receptor** 参数为 **1kv2\_complex:1kv2\_protein**。

展开 **Input Receptor** 参数组，将 **Interacting Residues** 参数设置为 **MET109**（该 group 已事先定义好，定义方法：展开 **1kv2\_protein**，展开 A 链，选择 Met109 残基，在分子窗口中右键**Group**，设置 group 名称为 MET109）。

设置 **Input Ligand Scaffold** 为 **1kv2\_complex:2yis\_scaffold**。

设置 **Scaffold Grow Points**，并选择 **Create New Group From Selection...**（刚才选中的 H19）。

展开 **Reactions** 参数组，先将 **Custom** 设置为 **Samples | Tutorials | Receptor-Ligand Interactions | sulfide\_halfreaction.rxn** 文件，然后再将 **Reactions** 参数设置为 **Custom: sulfide\_halfreaction**，并勾掉其他参数。（PS: **Custom: sulfide\_halfreaction** 只有设置了 custom 才会出现）

展开 **Reagent** 参数组，先将 **Custom** 参数设置为 **diverse\_sulfides:All**，然后在 **Reagents** 中勾选 **Custom: diverse\_sulfides** 选项。

将 **Minimization** 设为 **True**。

展开 **Advanced** 参数组，将 **Torsion Step Size** 设为 **10**，**Generate Reagents Conformations** 设为 **True**，并展开，将 **Maximum Conformations** 设为 **10**。

其他参数为默认（图 9），点击 **Run** 按钮运行作业，点击 **Background** 使作业在后台运行。

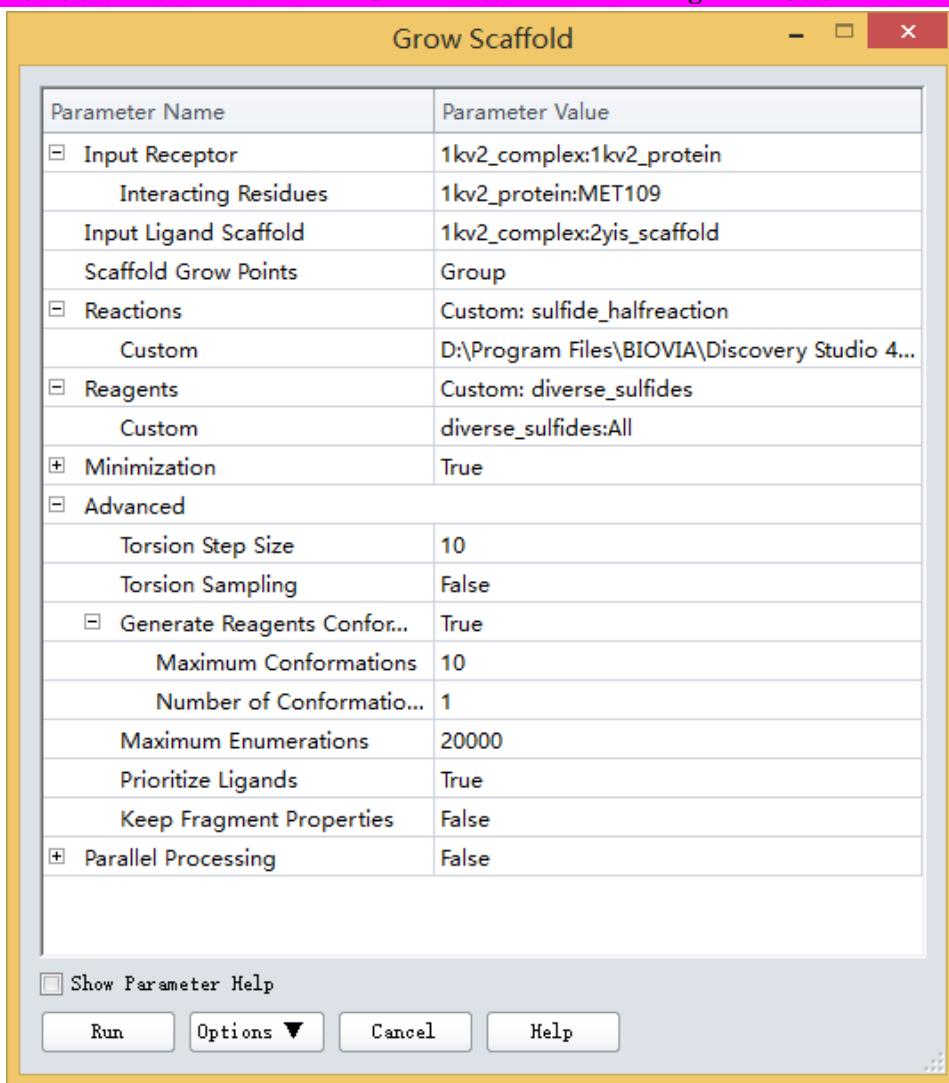


图 9 参数设置

该作业大概将运行 10 分钟，最终返回 14 个结果。

## 2. 查看结果

点击作业浏览器（Job explorer）中刚完成的作业，点击 **View Results** 链接。

在工具浏览器（Tools Explorers）中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

Step through Ligands.

点击上述 **View Interaction** 工具面板下的 按钮可以观察优化后的分子

同蛋白之间的各种非键相互作用。（图 10）

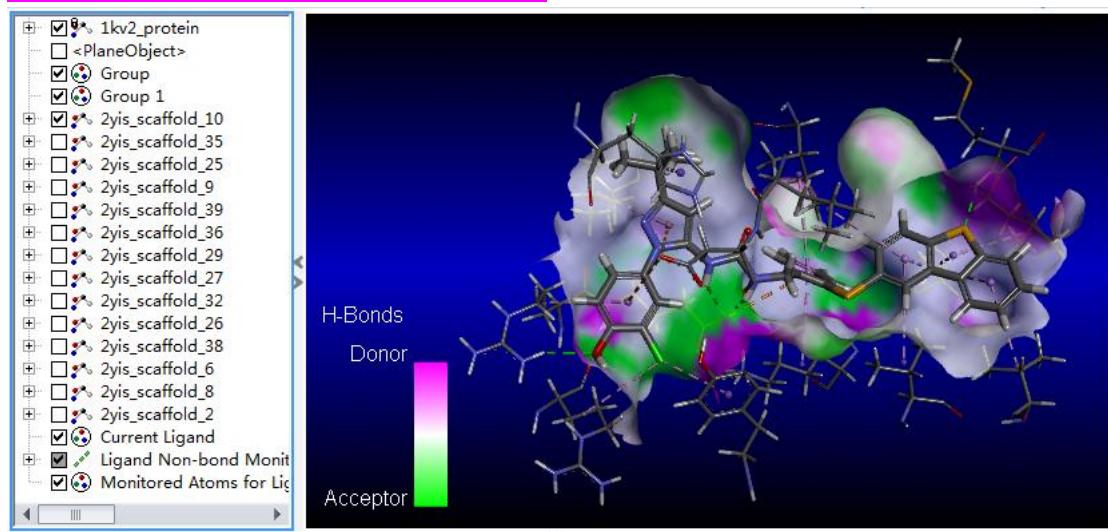


图 10 自定义反应先导化合物优化结果

尽管没有化合物与 PF-03715455 一样，因为 diverse\_sulfides.sd 文件中并没有包含形成 PF-03715455 所必需的子结构（hydroxyethylsulfanylphenyltriazolopyridinyl）。然而，你会发现一些有趣的新的取代，包括噻唑，benzoxazolinone，吲唑，苯并咪唑，和 imidazopyridine 子结构。

## Discovery Studio Replace Fragment 教程

**目的:** 采用 Grow Scaffold, 通过寻找适合的替代骨架进行骨架跃迁实现先导化合物优化

**所需功能和模块:** Discovery Studio Client, DS CHARMM Lite, and DS CATALYST CONFORMATION.

**所需数据文件:** 1o86\_ligand.sd, 1o86\_complex.dsv, ace\_inhibitors.sd, MyFragments-Cyclic-1R.smi.

**所需时间:** 30 分钟

### 介绍

骨架跃迁 (Scaffold Hopping) 是将一个配体的核心骨架替换为一个新的具有类似功能的基团来提高化合物的性质或者寻找具有类似功能的全新化合物。应用这种方法，在充分考虑活性的基础上，可以设计出突破专利保护，结构新颖，且可能改善药物代谢动力学性质的全新药物分子。

在本教程中，我们将使用 replace fragment 流程来对第三代血管紧张素肽转化酶 (ACE) 抑制剂赖诺普利进行骨架跃迁。ACE 催化非生物活性的十肽血管紧张素 I 转化为活性的八肽血管紧张素 II。血管紧张素 II 是一个强力的血管收缩剂，并能促进缓激肽（促进血管舒张）的降解。这种双重功效使得 ACE 在肾素血管紧张素系统 (RAS) 中发挥关键作用，并使得它成为治疗高血压，心脏衰竭，糖尿病肾病和 II 型糖尿病中一个非常有前景的靶点。

在本教程中,我们将使用 Replace Fragment 来寻找赖诺普利中的乙酰脯氨酸子结构的替代结构。主要任务如下:

- (一) 直接基于小分子产生替代片段
- (二) 使用蛋白活性位点优化产生的潜在的替代片段
- (三) 使用不同的性质来识别全新骨架

### (一) 直接基于小分子产生替代片段

在教程的第一部分，我们将使用 Replace Fragment 流程，只基于配体结构，来发现赖诺普利的乙酰脯氨酸子结构的替代结构。赖诺普利在 1983 年就被发现，而能够用于基于结构设计的 ACE 的晶体结构直到 2003 年才被公布，因此化合物的骨架跃迁经常是在没有蛋白结构下进行的。

#### 1. 执行骨架跃迁计算

在文件浏览器 (Files Explorer) 中，找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | **1o86\_ligand.sd** 文件。

显示小分子的三维结构。

在分子视图窗口 (按住 Ctrl+G 打开) 选中化合物的乙酰脯氨酸的子结构，或者在系统视图 (Hierarchy View, 按住 Ctrl+H 打开) 中，选中化合物的 **C15, O17, O22-C29** 原子 (图 1)。

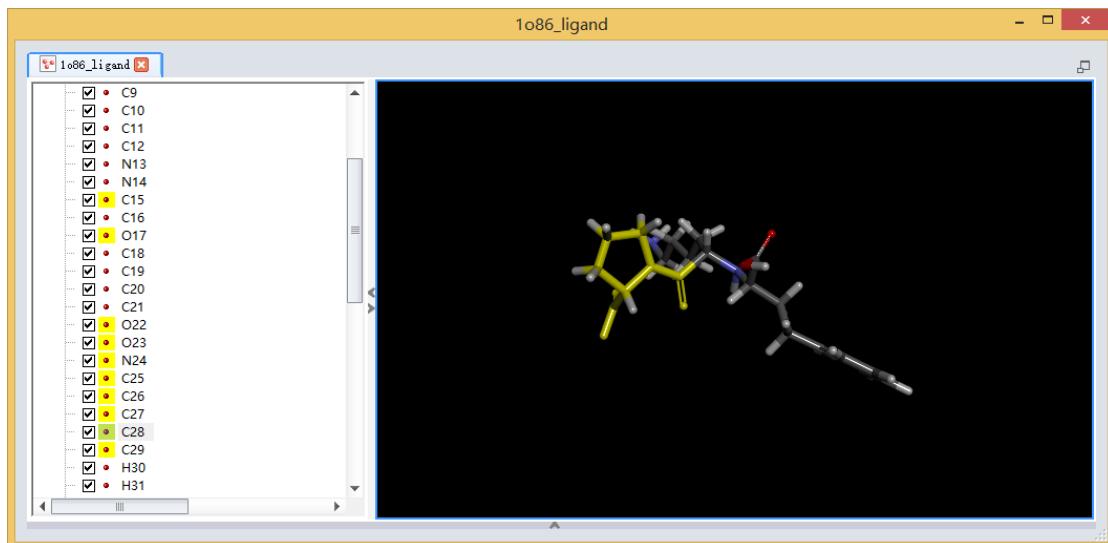


图 1 赖诺普利的三维结构，黄色加亮（选中状态）的为乙酰脯氨酸

在工具浏览器（Tools Explorer）中，展开 **Receptor-Ligand Interactions | Lead Optimization** 工具栏，单击 **Replace Fragment...** 工具，弹出流程窗口。

点击 **Input Ligand** 参数，设置为 **1o86\_ligand: 1o86\_ligand**。

将 **Fragment to Replace** 参数设置为 **Create New Group From Selection...**。

其他参数都默认，点击 **Run** 按钮，运行作业。

点击 **Background** 按钮使作业在后台运行。（图 2）

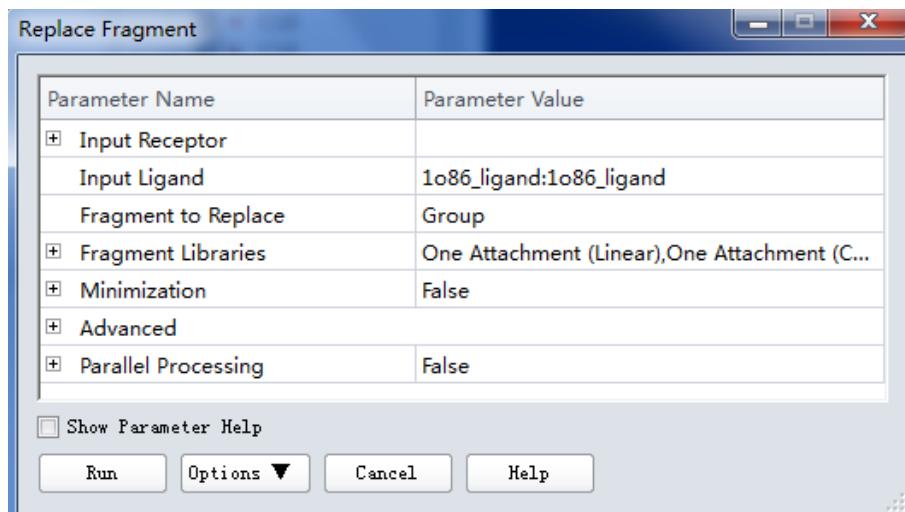


图 2 参数设置

作业运行的时间不到 3 分钟，最后得到 99 个化合物。

## 2. 查看计算结果

点击作业浏览器（Job explorer）中刚完成的作业，点击 **View Results** 链接。

Protocol Name	Saved	Status	Details	Elapsed Time
Replace Fragment	<input checked="" type="checkbox"/> Yes	Success	1o86_ligand: 99 ligands	0:01:30
	<a href="#">Report</a>			
	<a href="#">Novel Ligands</a>			
	<a href="#">View Results</a>			

图 3 查看计算结果

在表格视图中, 点击左侧的 按钮在图像窗口中查看第一化合物, 点击表格视图中的 键和 键进行浏览(图 4)。

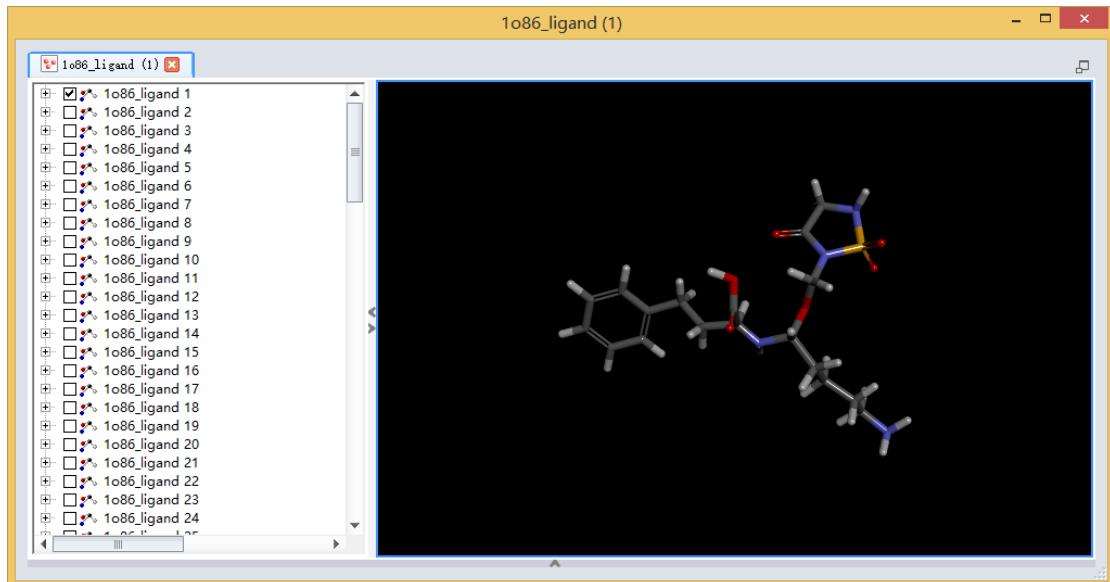


图 4 基于配体的骨架跃迁计算结果

由于我们只使用配体结构进行片段替换, 因此产生的新化合物只按照不符合 Lipinski 五规则数目 (Lipinski violations) 和片段的新颖性来进行排序。结果将给出替换片段与原始片段的相似性, 这有利于对分子进行排序, 关于相似性, 我们在教程的最后一部分还将进行讨论。

如果要考虑每一个匹配的片段, 那么所得到的新化合物在数量上对我们来说有可能会太多。目前 ACE 蛋白质的结构已经被解析出来了, 我们可以充分利用这些信息, 利用蛋白质的活性位点来对所得片段进行限制, 使得产生的的配体能够更好地符合结合位点的情况, 提高化合物的可结合性和活性。此外, 对于较大的化合物库, 你还可以通过计算化合物的 ADMET 及其他相关的性质来对得到化合物进一步过滤和进行排序。

## (二) 使用蛋白活性位点优化产生的潜在的替代片段

在这个部分, 我们将使用 ACE 的晶体结构来对所产生的可能替代片段进一步优化。我们这里使用的 ACE 与赖诺普利复合物的晶体结构的 pdb 号为 1O86, 该文件从 PDB 网站下载下来, 并且事先已经使用 Prepare Protein 流程进行优化。

### 1. 执行骨架跃迁计算

在文件浏览器 (Files Explorer) 中, 找到并双击打开 Samples | Tutorials | Receptor-Ligand Interactions | **1o86\_complex.csv** 文件。

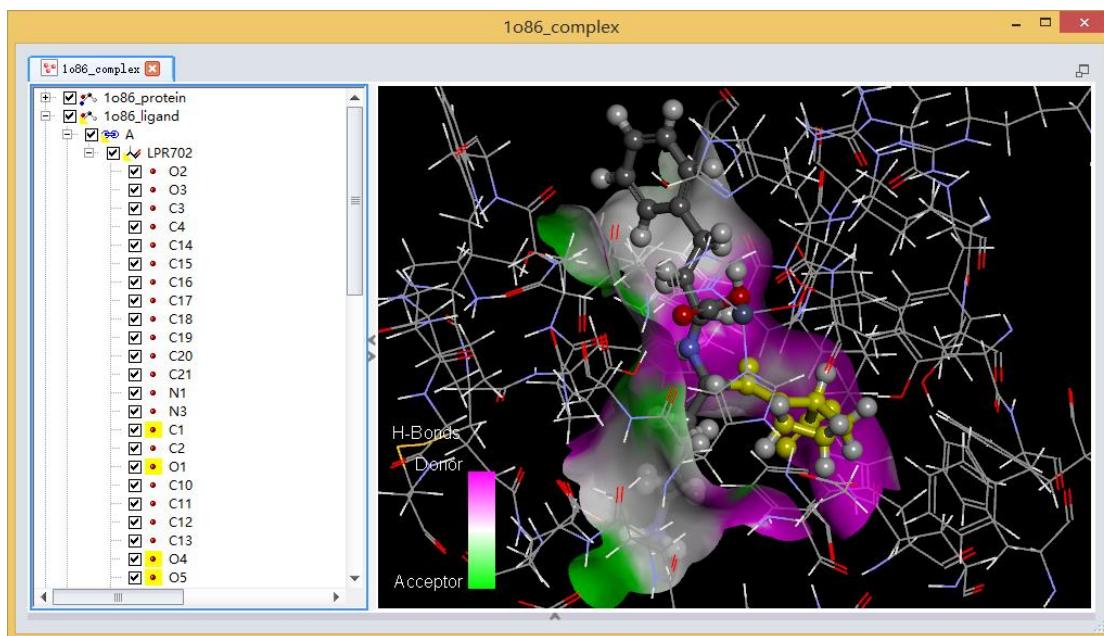


图 5 赖诺普利与 ACE 复合物的晶体结构，黄色高亮为被选中的乙酰脯氨酸

在分子窗口中选中配体的乙酰脯氨酸子结构，或在系统视图中选中配体的 C1, O1, O4, O5, N2, C5-C9 原子(图 5)。

在工具浏览器 (Tools Explorer) 中，展开 **Receptor-Ligand Interactions | Lead Optimization** 工具栏，点击 **Replace Fragment** 工具，打开流程对话框。

将 **Input Receptor** 设为 **1o86\_complex: 1o86\_protein**。

展开 **Input Receptor** 参数组，并将 **Interacting Residues** 设置为事先设定的 **interacting\_residues** 组。

点击 **Input Ligand** 参数，将其设为 **1o86\_complex: 1o86\_ligand..**

将 **Fragment to Replace** 设置为 **Create New Group From Selection...**。

其他参数都为默认参数，点击 **Run** 按钮运行作业。

点击 **Background** 按钮使作业后台运行。

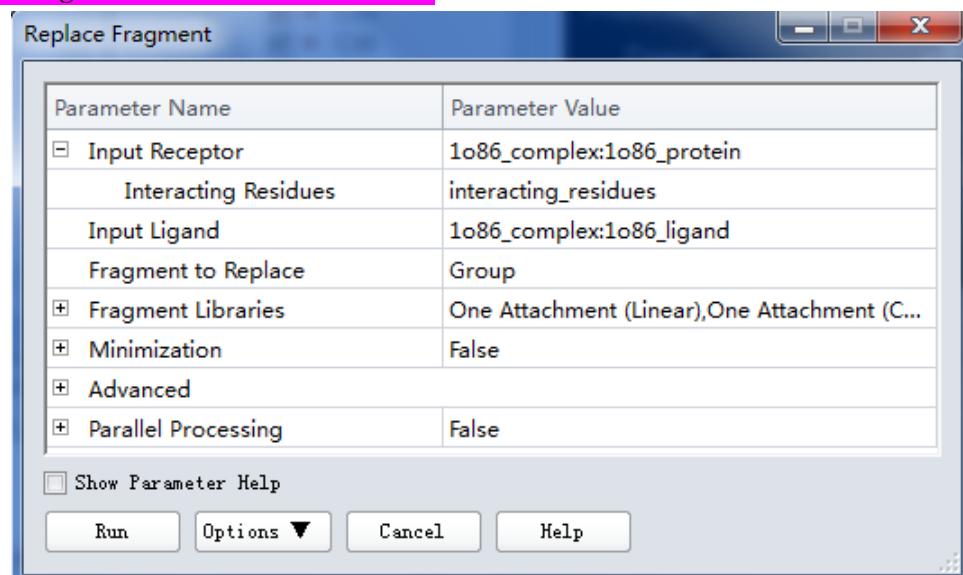


图 6 基于活性位点优化的骨架跃迁的参数设置

这个作业大致上要花 4 分钟，最终将产生 43 个配体。在这个例子中，活性位点的空间限制将用来对得到的配体进行优化。我们也可以规定产生的配体必须与受体中的残基形成相互作用来进行进一步优化。这里我们选用的残基有 GLN281, HIS353, LYS511, HIS513 和 TYR520，它们是赖诺普利的乙酰脯氨酸子结构与受体相互作用的残基。

## 2. 查看骨架跃迁优化结果

在作业浏览器（Jobs Explorer）中，展开新完成的作业，点击 **View Results** 链接，查看运算结果。（图 7）



图 7 查看计算结果

在表格视图中，点击左侧的 按钮在图像窗口中查看第一化合物，点击表格视图中的 键和 键进行浏览。

## 3. 骨架跃迁优化结果分析

### 非键相互作用的直观显示与分析

在工具浏览器（Tools Explorers）中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

在视图窗口中，受体原子与所产配体间的非键相互作用会通过不同颜色的虚线显示出来，且只有参与了同配体之间的相互作用的残基才会显示。（图 8）

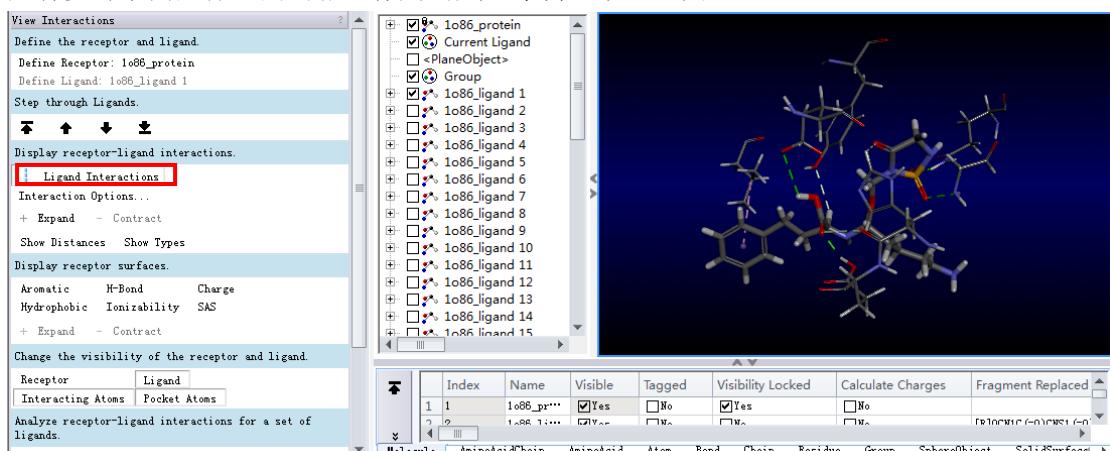


图 8 显示蛋白-配体之间非键相互作用

点击上述 View Interaction 工具面板下的 按钮可以观察优化后的分子同蛋白之间的各种非键相互作用。

为了更好的观察受体分子与配体的相互作用，可以对体系进行旋转以获得最佳的观赏角度。

在这个例子中，由于在片段替换过程中考虑了蛋白的活性位点，因此得到的配体所包含的将按照蛋白的相互作用数（最大），Lipinski 违法数（最小化），受体碰撞数（最小）和片段的新颖性（最大化）进行排序。所得到的配体都包含受体-配体相互作用性质，包括蛋白残基，相互作用类型及相互作用距离。所有的配体至少要与 interacting\_residues group 中的一个残基发生相互作用。

### (三) 使用不同的性质来识别全新骨架

在教程的最后一部分，我们通过设置参数来改变程序的运算结果。

#### 1. 执行骨架跃迁优化

在工作浏览器（Jobs Explorer）中找到第二部分的运算结果，展开，打开 report 链接。

在 report 页面里面展开 Parameters，然后点开 Protocol.pr\_xml 链接来打开 protocol（图 9）。该 protocol 包含之前运算 protocol 的所有参数。

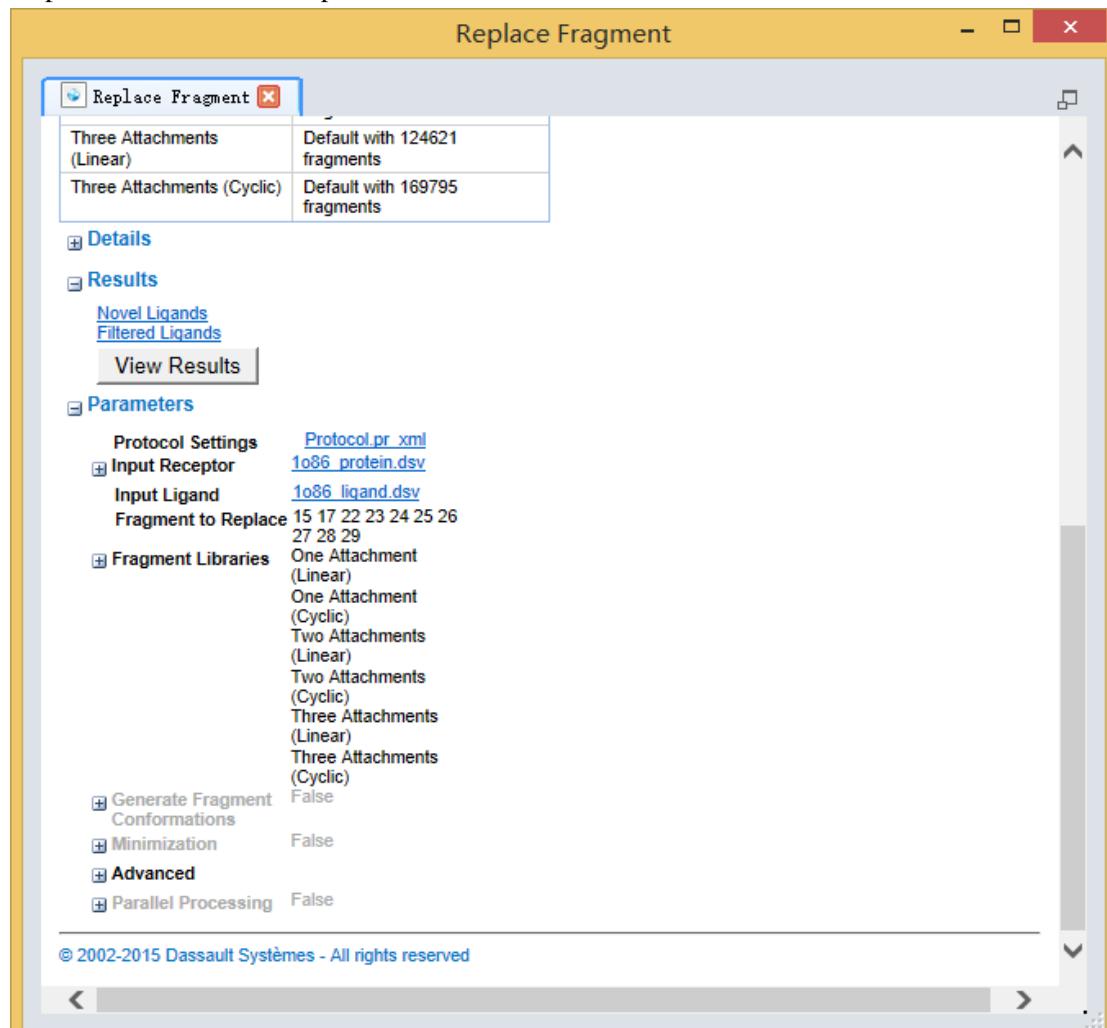


图 9 打开上次作业的 protocol 参数文件

在 protocol 中展开 Advanced | Fragment Search。

点击 Properties for Similarity 参数上的 [ ... ] 按钮，打开对话框。

移除右边窗口 (selected items) 的所有参数，然后添加 **FCFP\_4**。

将 **Maximum Distance** 参数改为 **0.7**。

其他参数都为默认参数 (图 10)，在流程浏览器窗口中点击鼠标右键，点击 **run** 运行作业 (或者按 **F5** 快捷键)，等待作业结束。

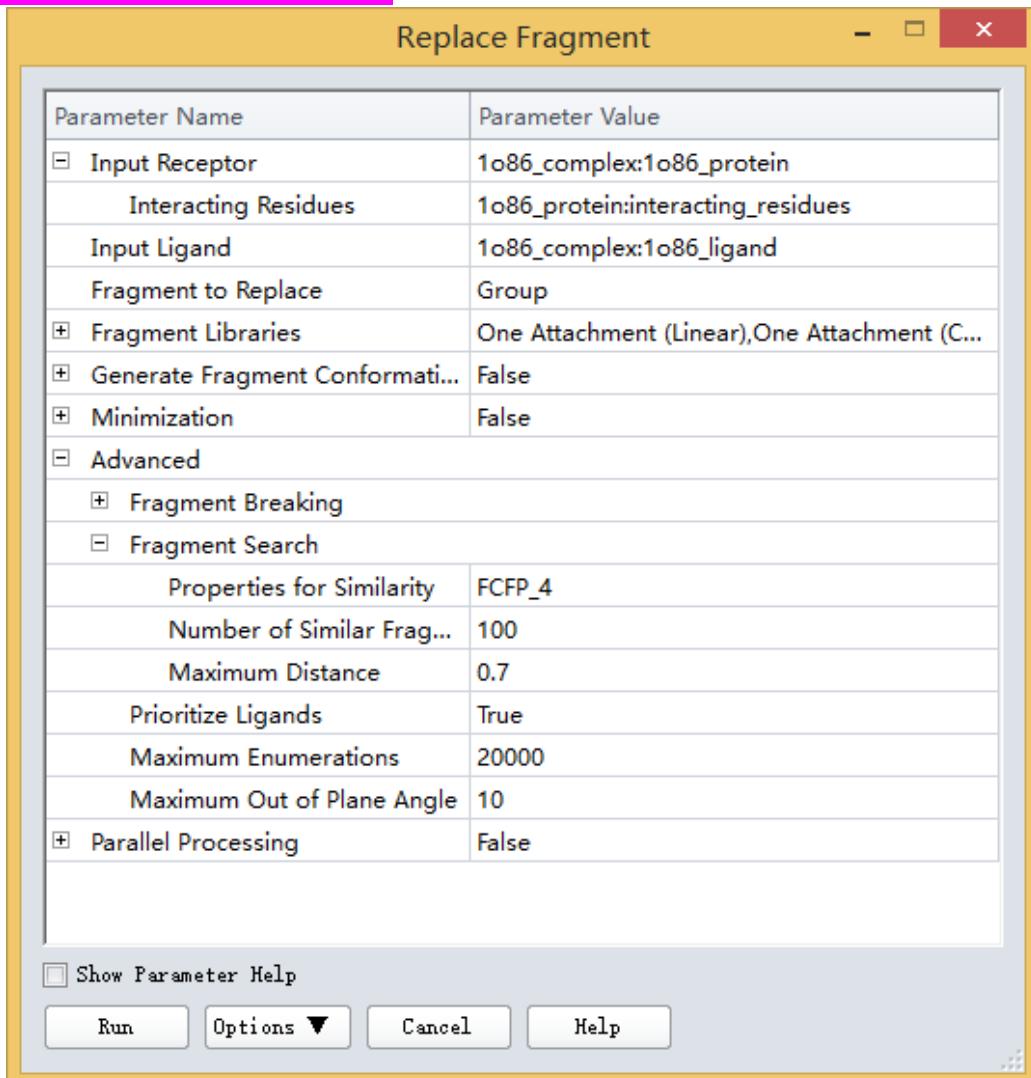


图 10 参数设置

这个作业运行不到 5 分钟，最后得到 75 个新分子。

需要注意的是，当使用分子指纹的时候，不同片段间的相似性将用 Tanimoto 系数来进行比较。两个很相似的片段它们间的 Tanimoto 系数可能会比较低，如使用 ECFP\_4 分子指纹时，苯环和嘧啶间的 Tanimoto 系数只有 0.333。因此，当只用分子指纹进行限时，为了得到较合理数目的结果，你需要适当地调高 Maximum Distance 参数。

## 2. 运算结果查看

在作业浏览器 (Jobs Explorer) 中，展开新完成的作业，点击 **View Results** 链接，查看运算结果。

在工具浏览器 (Tools Explorers) 中，展开 **Receptor-Ligand Interactions | View Interactions**，点击 **Ligand Interactions**。

在视图窗口中，受体原子与所产配体间的非键相互作用会通过不同颜色的虚线显示出来，且只有参与了同配体之间的相互作用的残基才会显示。（图 11）

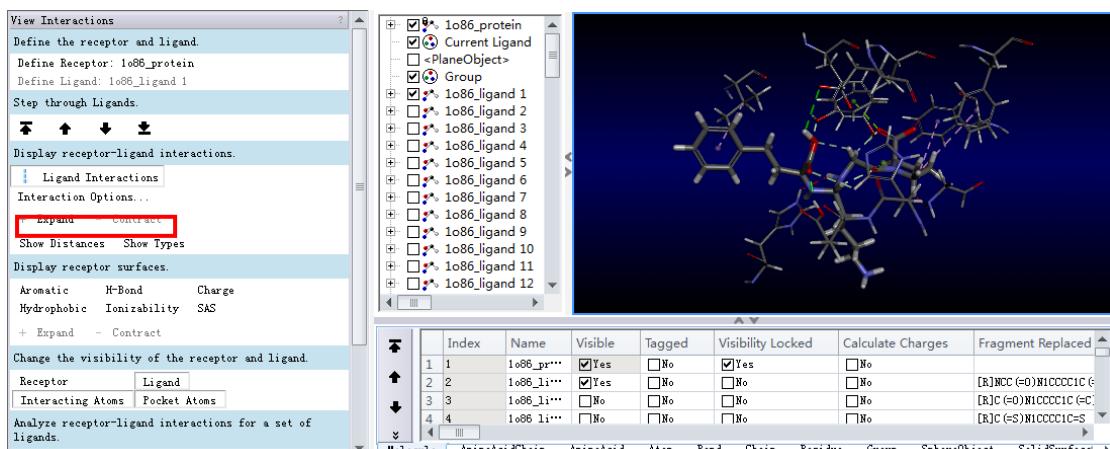


图 11 以分子指纹作为骨架新颖性评价标准时的运算结果

点击上述 View Interaction 工具面板下的 按钮可以观察优化后的分子同蛋白之间的各种非键相互作用。

为了更好的观察受体分子与配体的相互作用，可以对体系进行旋转以获得最佳的观赏角度。



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

药效团专题:  
**HipHop、HypoGen、SBP、CBP**

## 构建基于分子共同特征的药效团模型（HipHop）教程

**目的：**通过此教程，了解 Discovery Studio 中构建基于分子共同特征的药效团的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS Catalyst Conformation, DS Catalyst Hypothesis, DS Catalyst Score。

**所需数据文件：**5HT2c\_ligands.sd, sulfonamide5HT2c\_ligands.sd, sampledrugs20.sd。

**所需时间：**30 分钟

### 介绍

Common Feature Pharmacophore Generation protocol 用于发现一系列配体小分子所共有的化学特征，并基于这些共同特性结构的比对叠合自动生成药效团模型，用户可以使用共有的特征药效团去搜索化合物数据库来寻找可能的先导分子。基于分子共同特征的药效团模型同样也可以用于探索一系列具有相似活性但结构却不同或者结构柔性较大的分子的构效关系。

此教程以 6 个活性配体小分子所构成的训练集来构建基于分子共同特征的药效团模型，继而用于先导化合物的发现。

本教程包括以下步骤：

- 基于分子共同特征的药效团模型的构建（训练集）
- 基于分子共同特征的药效团模型的验证（测试集）
- 先导化合物的发现（数据库的筛选）
- 

### Common Feature Pharmacophore 的构建

#### 1. 训练集分子的准备

本教程采用一系列已知的 5-HT2c 配体（1, 2）来构建一个基于分子共同特征的药效团模型用于先导化合物的发现。5-HT2c 受体属于 GPCR 超家族。

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | Pharmacophore，双击打开 5HT2c\_ligands.sd。

在表格浏览器中可以看到一共有 6 行，代表了 6 个分子。这些分子的 Principal 和 MaxOmitFeat 属性都已事先定义。若无定义，则选择表格浏览器中剩下列的 heading，鼠标右键点击选择 Add Attributes，打开 Add Attributes 对话框，添加这两个属性。

Principal 属性定义了分子的活性水平：

数值	活性水平	描述内容
2	有活性	参考分子，分子中所有化学特征在构建药效团模型时都要考虑。
1	中等活性	定位药效团特征元素时需要考虑该化合物的构象空间。
0	非活性	该分子在定位药效团特征元素时不考虑，用于选择性地定义排除体积。

MaxOmitFeat 属性定义了每个分子中允许不与药效团模型匹配的特征元素的个数：

数值	描述内容
0	构建的药效团模型中所有特征元素都必需与化合物匹配上。
1	构建的药效团模型中允许有 1 个特征元素不与化合物匹配。
2	构建的药效团模型中所有特征元素都无需与化合物匹配。

本教程假设所有分子都是活性分子并且视其中某几个分子为参考分子。如果所选取的分子有着截然不同的药效特征类型和尺寸，那么 Principal 和 MaxOmitFeat 属性设置的不同对最终构建的药效团模型影响较大。

## 2. 药效团特征元素的选取

肉眼观察这几个活性化合物，选出可能的药效团特征元素，或者根据 Feature Mapping 工具来完成该步骤(如下)。

点击表格浏览器中的  按钮，在图形窗口中显示化合物 5HT2c\_ligand1。再在图形窗口右击选择 Show All，显示所有活性化合物的结构。

在工具浏览器（Tools Explorer）中，展开 **Pharmacophore | Edit and Cluster Features**，在工具面板中单击 **Feature Mapping** 打开 **Feature Mapping** 对话框，点击 **Features** 右边  按钮，打开 **Select Features** 对话框(图 1)，选择所要匹配的特征元素，此处设为默认值。

点击 **Run** 运行该功能，并等待计算完成。

该操作可以识别出所有已显示化合物中所有所选特征元素可能的位置。

在 **Edit and Cluster Features** 工具面板中单击 **Current Features** 查看训练集分子所表征的药效团特征元素的所有类型。

结果显示这六个化合物都包含有疏水中心 (Hydrophobe)、氢键供体 (Donor)、氢键受体 (Acceptor)、正电离子中心 (Ionizable Positive) 及芳香环中心 (Ring Aromatic) 这五种特征元素的类型。

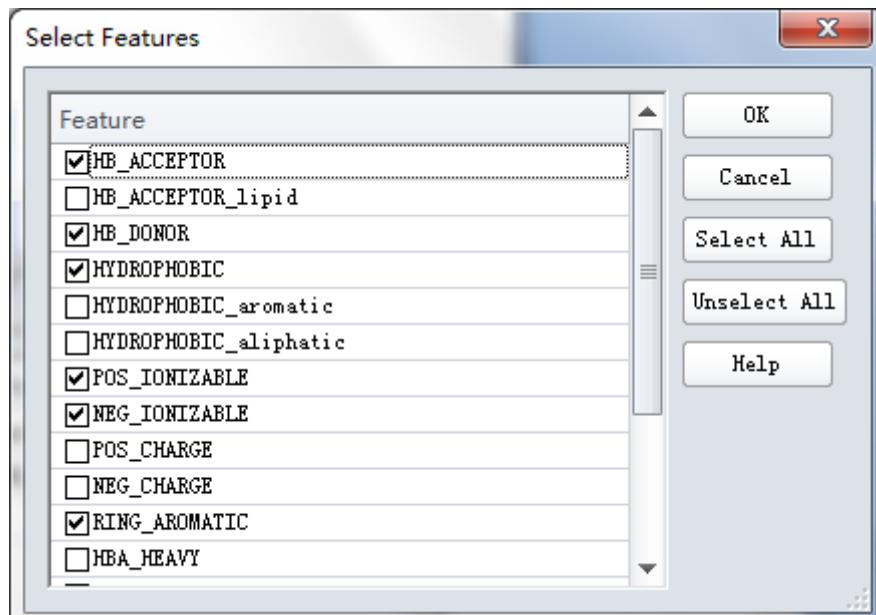


图 1 “Select Features”对话框

## 3. Common Feature Pharmacophore 的构建

在工具浏览器（Tools Explorer）中，展开 **Pharmacophores | Create Pharmacophores Automatically**，单击 **Common Feature Pharmacophore Generation**。

流程对应参数在参数浏览器中打开。

设置 **Input Ligands** 为 **5HT2c\_ligands>All**。

点击 **Feature** 右边  按钮，打开 **Select Features** 对话框。

勾选 POS\_IONIZABLE 和 RING\_AROMATIC 左边的复选框，点击 OK。（图 1）

在默认设置基础上添加上一个正电离子特征 POS\_IONIZABLE，一个芳香环特征 RING\_AROMATIC。

展开 Conformation Generation 参数组，点击 Conformation Generation 右边的栅格，下拉列表中选择 BEST，构象上限数 Maximum Conformation 设为 200，能量阈值 Energy Threshold 设置为 10，其他参数默认。（图 2）

对每个小分子产生最多 200 个构象用以表征小分子的构象空间。其中只有同最低能构象的能量差值在能量阈值 10kcal/mol 之内的构象被保留。

点击 Run 运行作业。

在弹出的窗口点击 Background。

等待作业完成。该作业大概需要 5 分钟。

从菜单中选择 Window | Close All，关闭所有窗口。

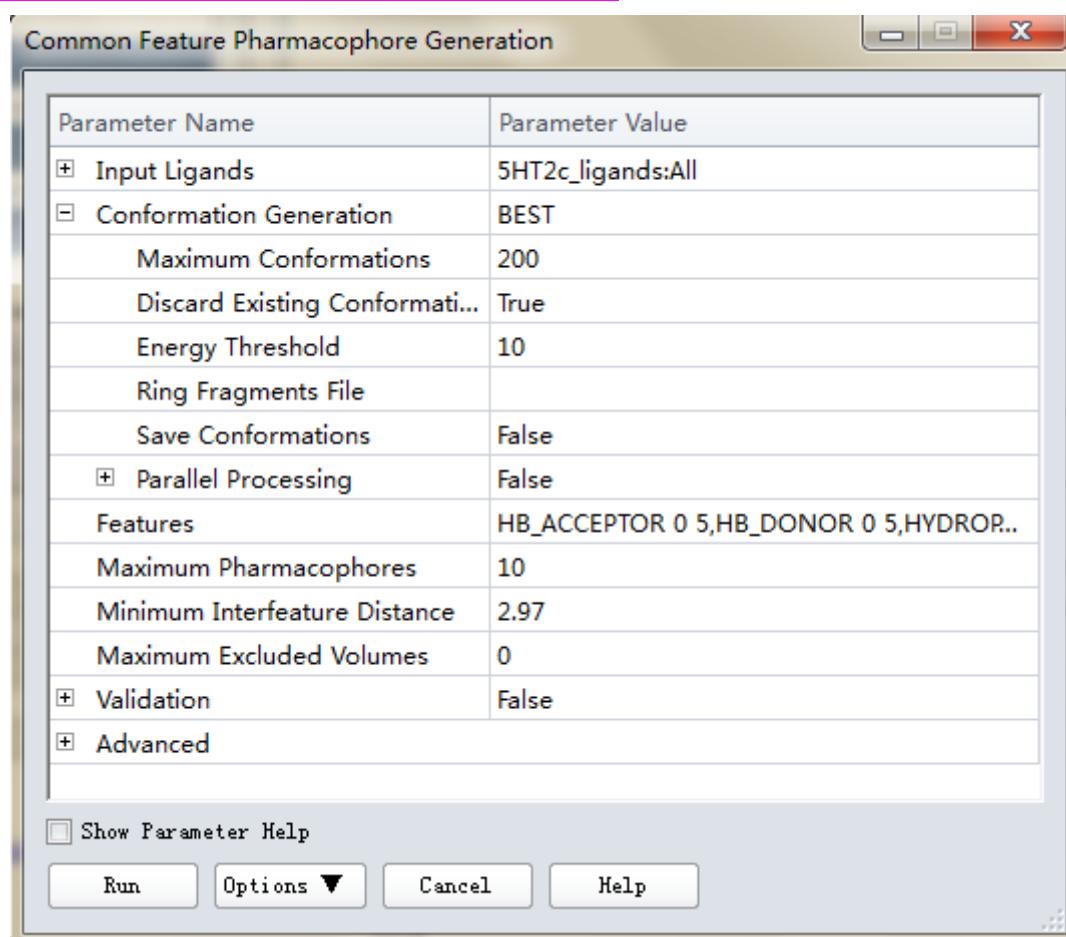


图 2 “Common Feature Pharmacophore Generation”流程参数设置

## 药效团结果分析

### 1. 查看结果

作业完成后，展开作业浏览器（Jobs Explorer）中该任务并点击 Report 链接，在 Html 窗口中打开 Report 页面。

从 Summary 一栏可知，此次运算一共产生了 10 个药效团。在 Details 一栏中我们可以了解到每个药效团更为详细的信息。Results 一栏罗列了各个结果的链接。Parameters 一栏显示了此次操作所采用的各参数。

这 10 个药效团模型是根据训练集分子与药效团模型的匹配度以及模型本身的稀有度来排名的。排名第一的模型未必是最好的模型，因此需要对这 10 个模型进行综合分析。

## 2. 查看药效团与训练集分子的匹配情况

在 Report 页面中展开 Results，点击 **View Aligned Ligands 01** 链接查看训练集分子同排名第一的药效团的匹配情况。

点击表格浏览器中的 键，查看排名第一的分子同药效团的匹配，继而点击 键和 键，可以查看到每个分子与该药效团的匹配情况。

在表格浏览器中，注意查看 5HT2c\_ligand4 和 5HT2c\_ligand5 的 FitValue 值。

回到 Report 页面，点击 **View Aligned Ligands 10** 链接查看训练集分子与排名第 10 的药效团的匹配情况。

在此次叠合中，5HT2c\_ligand5 的 FitValue 比 5HT2c\_ligand4 高，且匹配度要好。

分析 Common Feature Pharmacophore 的结果时需要考察所有药效团模型同训练集分子的匹配情况，了解每个模型是如何解释训练集分子的。

回到 Report 页面，展开 Details 一栏。（图 3）

该表格罗列了十个药效团模型的结果参数。查看该表格可以快速了解到十个药效团模型之间的不同。

Details

	Features	Rank	Direct Hit	Partial Hit	Max Fit
01	RRPHHA	105.092	111111	000000	6
02	RRPHHA	105.092	111111	000000	6
03	RRPHHA	104.778	111111	000000	6
04	RRPHHA	104.778	111111	000000	6
05	RRPHHA	104.646	111111	000000	6
06	RRPHHA	104.646	111111	000000	6
07	RRPHHA	103.983	111111	000000	6
08	RRPHHA	103.983	111111	000000	6
09	RRPHHA	103.983	111111	000000	6
10	RRPHHA	103.128	111111	000000	6

图 3 十个 Common Feature Pharmacophore 的结果参数

表格中每一行代表一个药效团。

每一栏的含义如下：（以模型 01 为例）

Features:RRPHHA 此药效团含有 2 个芳环中心，1 个正电离子中心，2 个疏水特征，1 个氢键受体特征；

Rank: 105.092 此药效团的打分值为 105.092；

DH: 111111 此药效团药效特征与 6 个小分子均可匹配；

PH: 000000 此药效团与 6 个小分子部分匹配的药效特征数目为 0；

Max Fit: 6 最大匹配值为 6，即 6 个药效特征均可匹配。

### 3. 查看药效团与测试集分子的匹配情况

最终决定选取哪一个药效团模型用于表征训练集分子以及用于寻找潜在的新型 5-HT2c 配体分子，是一个主观的过程。因此，可以基于对训练集化合物同关键化学特征间叠合情况的分析。同时也可以采用测试集化合物来验证确定。

注：同样可以通过设置上述 protocol 中的 Validation 参数来自动完成验证过程。产生的每一个药效团模型都会同诱饵分子数据库（含 actives 和 inactives）进行匹配并最终自动生成 ROC 曲线显示于 Report 页面。

本教程中所采用的测试集分子包含了负性分子，对模型作了一个简单的验证，从而确定最终的药效团模型用于识别潜在的新型小分子骨架。

从菜单中选择 Window | Close All，关闭所有窗口。

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | Pharmacophore，双击打开 sulfonamide5HT2c\_ligands.sd。在文件浏览器（Files Explorer）中，将 sampleddrugs20.sd 拖至同一窗口中。

第一个文件包含了 18 个已知 5-HT2c 活性的磺胺类分子[Garzya et al.,2007]。第二个文件包含了 20 个同 6 个训练集分子非常类似（基于 Lipinski 和 FCFP\_6 属性）的配体小分子，视这 20 个化合物为非活性化合物（理想情况应该是这些分子确实是负性对照分子，即确实不能与受体结合）。

在工具浏览器（Tools Explorer）中，展开 Pharmacophore | Search, Screen and Profile，点击 Ligand Profiler。

流程对应参数在参数浏览器中打开。

该流程可以将多个分子同多个药效团模型快速匹配。

设置 Input Ligands 为 sulfonamideSHT2c\_ligands:All。

点击 Input File Pharmacophores 右边 [...] 按钮，打开 Specify Queries 对话框。

点击 Add File Queries，找到之前运行 Common Feature Pharmacophore 构建流程所得到的 Output 文件。SHIFT-选取所有模型。

展开 Conformation Generation 参数组，点击 Conformation Generation 右边的栅格，下拉列表中选择 BEST，构象上限数 Maximum Conformation 设为 200，能量阈值 Energy Threshold 设为 10，点击 Save Conformations 右边的栅格，下拉列表中选择 True。

展开 Advanced 参数组，点击 Scale Fit Values 右边的栅格，下拉列表中选择 False。点击 Save Aligned Ligands 右边的栅格，下拉列表中选择 True。

点击 Options 选取 Save Protocol Settings 保存参数设置。

若下次运行该流程时想要采用原始参数的设置则可以点击 Options 选取 Restore Original Settings。

其余参数使用默认值。（图 4）

点击 Run 运行作业。

点击 Background 等待作业完成。

该作业大概需要 20min。

作业完成后，展开作业浏览器（Jobs Explorer）中相应的任务并且点击 View Results 链接，打开一张匹配热图，其中药效团模型 6 和 10 没有同任何测试集分子匹配上。

展开作业浏览器（Jobs Explorer）中相应的任务并且点击 5HT2c\_ligands\_01\_view.pl，在图形窗口中右击选取 Show All 查看所有同模型 01 匹配上测试集分子同模型 01 的匹配情况。

结果显示只有 7 个配体分子同药效团模型 01 相匹配，且 FitValue 值都偏低。造成这一现象的原因可能是 Maximum Omitted Features 参数设成了 0，即不允许任何一个药效团特征元素

不被匹配上。该参数设为**-1** 表示考虑所有药效团特征元素的子集。

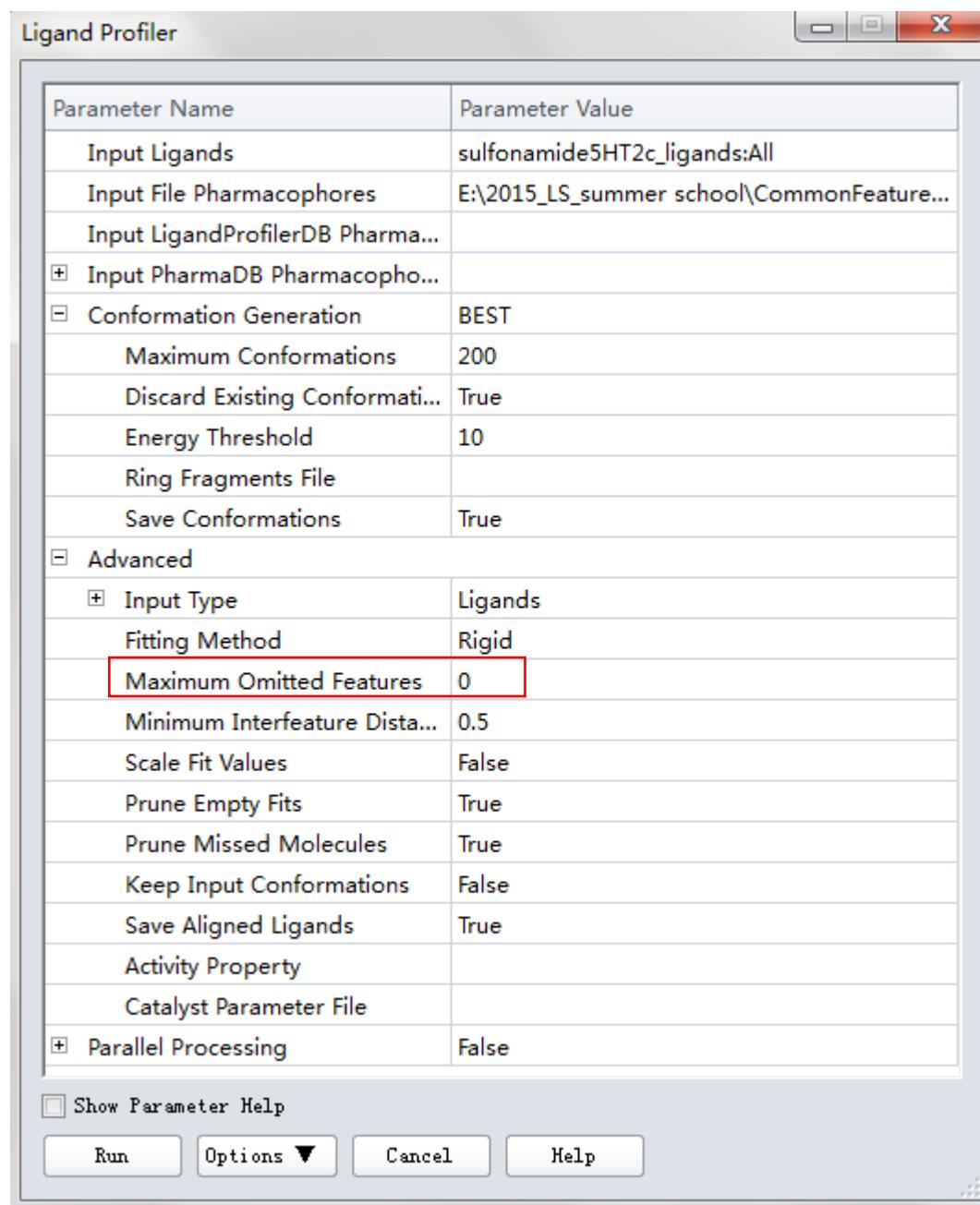


图 4. “Ligand Profiler”流程参数设置

再次点击 **Ligand Profiler**。

点击 **Input Ligands** 右侧栅格，下拉列表中选取 **Browse**，找到上一步运行得到的 **Output** 文件，选取 **sulfonamide5HT2c\_ligands\_BEST.sdf**。

将 **Conformation Generation** 改为 **NONE**。

**Conformation Generation** 设为 **NONE**，是因为上一步已经保存了所产生的构象（即 **Save Conformations** 设为 **TRUE**），不需要重复。

展开 **Advanced** 参数组，将 **Maximum Omitted Features** 右边的数值改为**-1** 以允许部分匹配。  
其余参数使用默认值。（图 5）

点击 **Run** 运行作业。

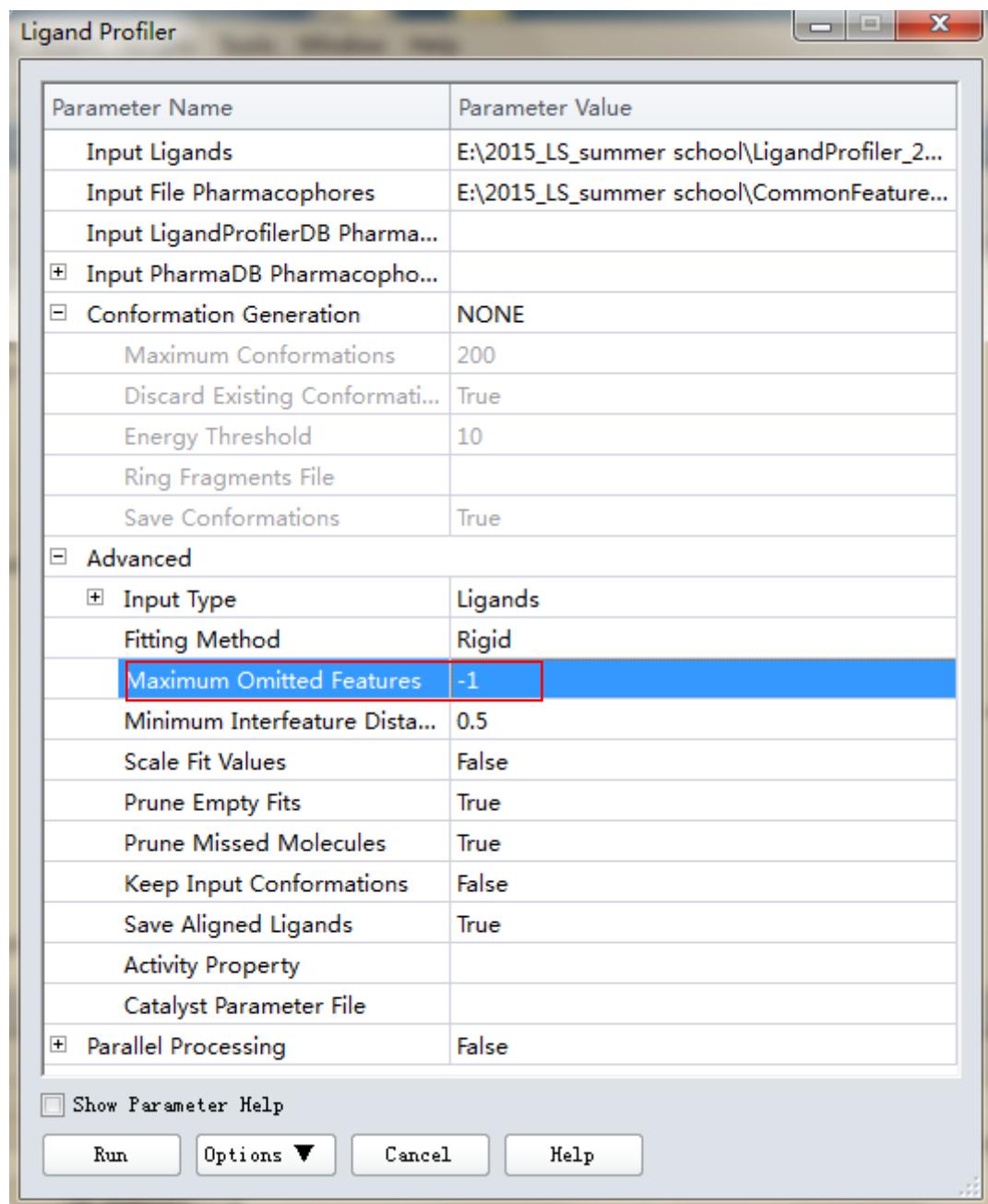


图 5. “Ligand Profiler”流程参数设置

点击 **Background** 等待作业完成。

该作业大概需要 5min。

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

展开 Results，点击 **View Results** 按钮。

自动生成一张表格和一个热图（图 6）。

在表格浏览器中，有 18 个活性磺胺类分子和 17 个非活性配体分子。

热图中横坐标第一列有几个区域显示为蓝色，表示同第一个药效团模型匹配的某些配体小分子 fit 值比较低。同样地，在药效团模型 5HT2c\_ligands\_04 一列有几个区域显示为橙色和红色，表示同 5HT2c\_ligands\_04 药效团模型匹配的某些配体小分子 fit 值比较高。

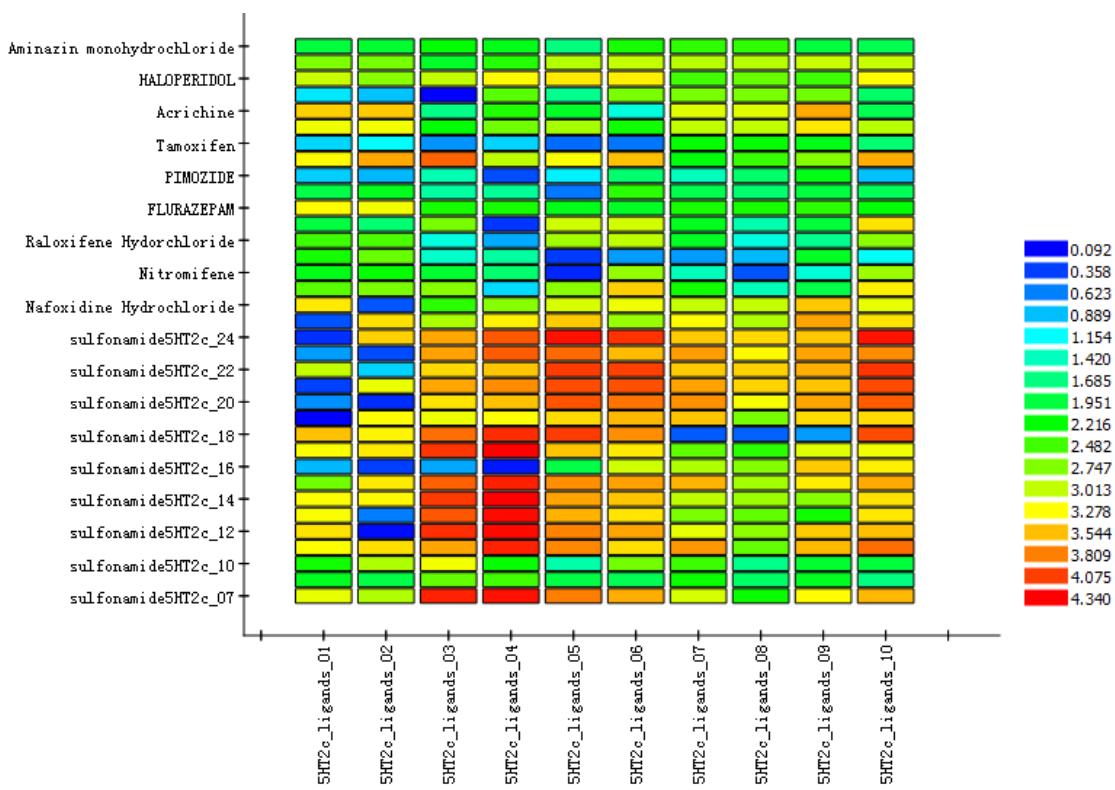


图 5. “Ligand Profiler”热图

展开 Results，点击 **5HT2c\_ligands\_04\_view.pl**。

在表格浏览器中，化合物按 FitValue 从高到低排列，前 14 个小分子都是活性化合物。

在 Hierarchy 窗口中，SHIFT-选取前 14 个分子，在 Graphics 窗口中右击选取 Show。

结果显示这 14 个分子的叠合情况有所改善，但所有的分子都没有同正电离子特征匹配。

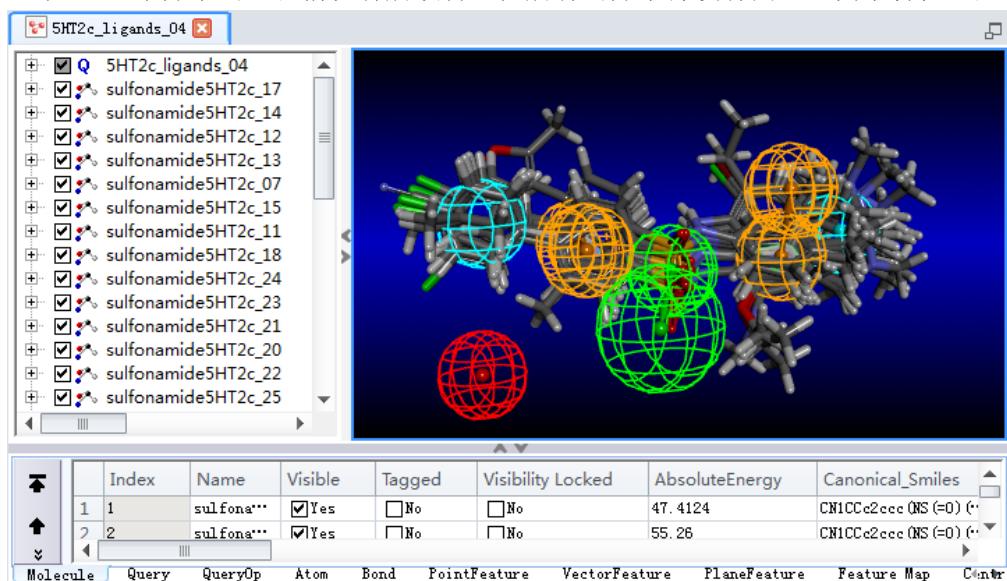


图 7

## 先导化合物的发现（数据库筛选）

采用上述得到的 3D 药效团模型对 MiniMaybridge 化合物数据库进行虚拟筛选，以期发现潜在的新型先导化合物。

## 1. 采用药效团进行数据库筛选

在工具浏览器（Tools Explorer）中，展开 **Pharmacophore | Search, Screen and Profile**，点击 **Search 3D Database**。

流程对应参数在参数浏览器中打开。

点击 **Input Database** 后面栅格，下拉菜单中选择 **MiniMaybridge[2000 mol]**。

此操作表示在 MiniMaybridge 数据库中进行搜索，该数据库中含有 2000 个小分子。

在 **Input Pharmacophore** 后选择 **5HT2c\_ligands\_04:5HT2c\_ligands\_04**。

此操作表示搜索可以与 5HT2c\_ligands\_04 药效团模型相匹配的小分子。

其余参数设为默认设置。（图 8）

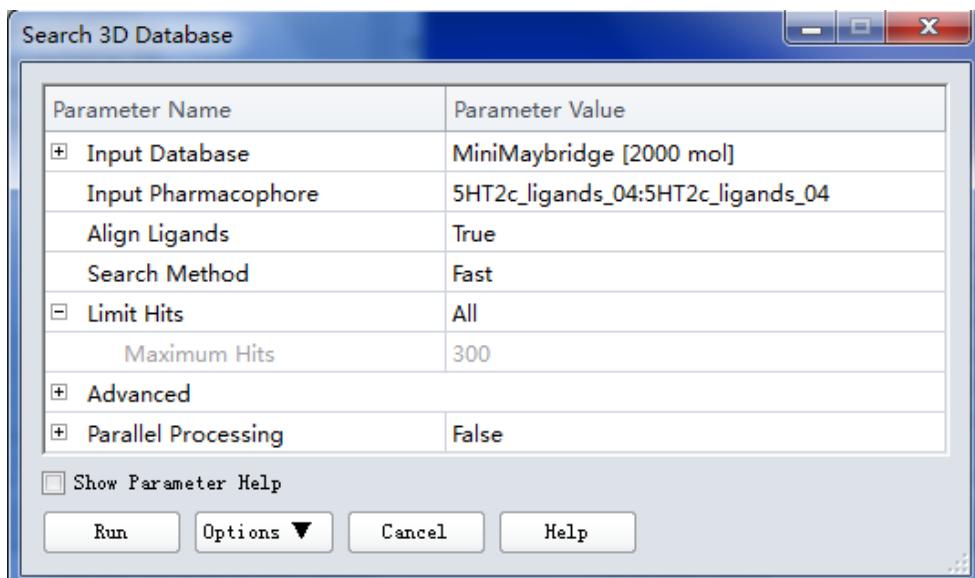


图 8. “Search 3D Database”参数设置

点击 **Run** 运行作业。

点击 **Background** 等待作业完成。

从菜单中选择 **Window | Close All**，关闭所有窗口。

## 2. 查看筛选结果

双击作业浏览器（Jobs Explorer）窗口中相应的行，打开 Report 页面点击 **ViewResults** 按钮可以查看小分子结构是否同原训练集分子及活性测试集分子有结构上的差别。

表格浏览器（Data Table）中的 **FitValue** 栏目中的数值可以评价小分子与药效团匹配的情况，FitValue 值越高，则小分子与药效团匹配的越好。

可以发现筛选到的这两个小分子从结构上来说，无论是同训练集分子还是测试集中活性分子相比，都有较大差异。

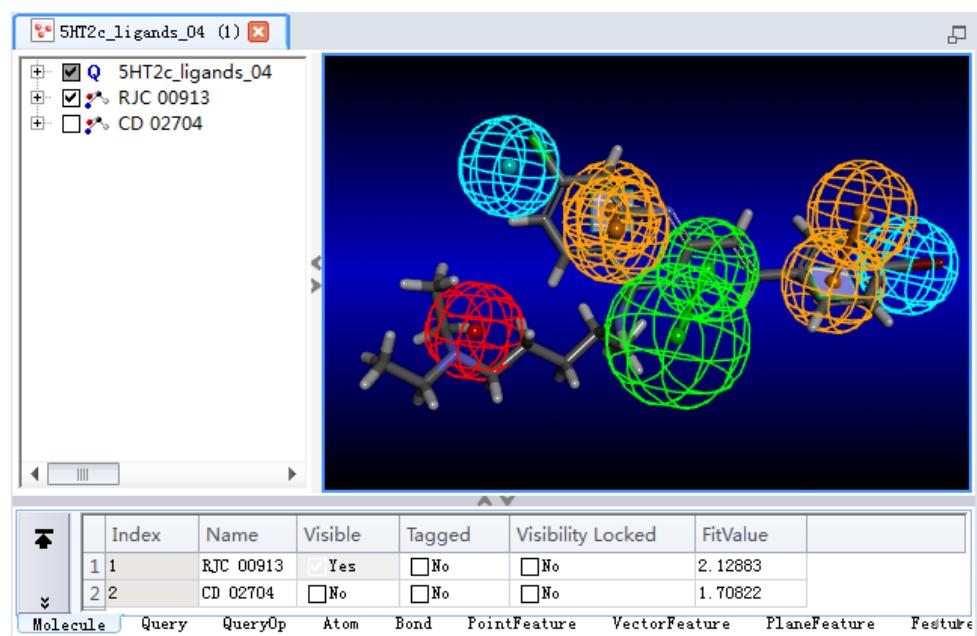


图 9

## 构建具有活性预测能力的药效团（Hypogen）教程

**目的：**通过此教程，了解 Discovery Studio 中构建具有活性预测能力的药效团的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS Catalyst Conformation, DS Catalyst Hypothesis, DS Catalyst Score。

**所需数据文件：**serm\_ligands.sd, test\_serm\_ligands.sd。

**所需时间：**30 分钟

### 介绍

3D Quantitative Structure-Activity Relationship (QSAR) Pharmacophore Generation protocol 可以基于一系列针对特定生物靶标具有明确活性数值的化合物构建出具有活性预测能力的药效团模型。该算法首先构建得到活性分子能共享而非活性分子则不能共享的初始药效团模型，然后再经过模拟退火进行模型的进一步优化。最终构建得到的模型可以预测化合物的活性，以及指导化合物的优化以提高其活性。

此教程中我们采用一系列已知活性值的训练集分子来构建一个 3D QSAR 药效团模型。

本教程包括以下步骤：

- 3D QSAR 药效团的构建(训练集)
- 药效团结果分析
  - ◆ 叠合&成本分析 (cost analysis)
  - ◆ 测试集
- 化合物的设计改造

### 3D QSAR 药效团的构建

#### 1. 训练集分子的准备

本教程采用一系列选择性的雌激素受体调节剂 (SERMs) [Broqi et al., 2009] 来构建一个 3D QSAR 药效团模型用于预测新型小分子的活性。

在文件浏览器 (Files Explorer) 中，展开 Samples | Tutorials | Pharmacophore，双击打开 serm\_ligands.sd。

在表格浏览器中可以看到一共有 24 行，代表了 24 个分子。这些分子的名字 (Name)、活性 (Activ，可以为 IC<sub>50</sub> 或 Ki 值，且活性跨度需跨越 4 个数量级) 和活性不确定度 (Uncert) 都已事先输入。Activ 值为 SERMs 对 MCF7 人体胸腺恶性肿瘤细胞增生的抑制活性 IC<sub>50</sub> ( $\mu\text{M}$ ) 值，Uncert 值为 1.5 (该值的设定取自于文献[Broqi et al., 2009])。

在表格浏览器中，点击 键和 键，在 3D Window 中观察各个分子的结构特征。

#### 2. 药效团特征元素的选取

在表格浏览器中，右击鼠标并选择 Color By Activity。

双击 Activ 一栏，使 24 个化合物按活性从高到低排列 (即活性值从低到高)。

根据药效团构建算法排名最靠前的 2 个化合物被定义为活性化合物 ( $\frac{MA \times Unc^{MA}}{Unc^A} - \frac{A}{Unc^A} > 0.0$ )，其所代表的两行显示为浅绿色。排名最靠后的 7 个化合物则被定义为非活性化合物。

( $\log(A) - \log(MA) > 3.5$ )，其所代表的行显示为浅粉色。观察前两个活性化合物，选出可能的药效团特征元素，或者根据 Feature Mapping tool 来完成该步骤(如下)。

点击表格浏览器中的 按钮，显示活性最高的化合物 (serm-40)。

在表格浏览器中勾选活性排名第二高化合物的 visible 复选框 (serm-34)。

在工具浏览器 (Tools Explorer) 中，展开 Pharmacophore | Edit and Cluster Features，单击 Feature Mapping，打开参数面板。

设置 Input Ligands 参数为 serm\_ligands:Visible。

在参数面板中点击 Features 右边 按钮，打开 Select Features 对话框，选择所要匹配的特征元素，此处设为默认值。

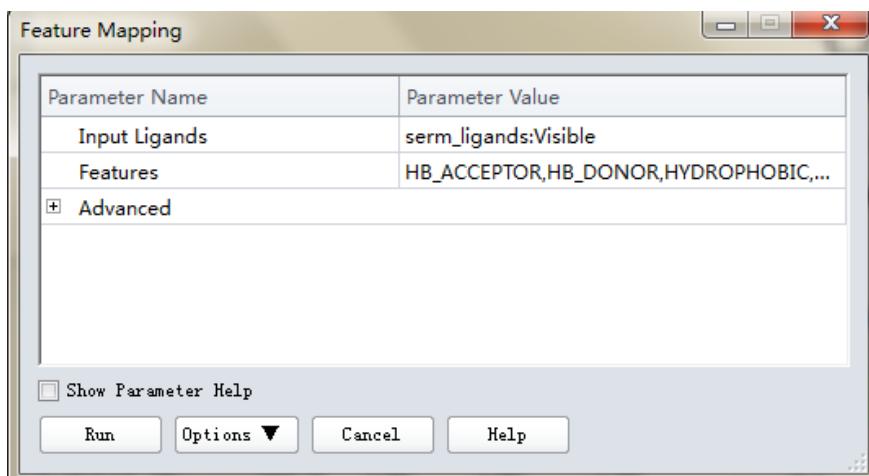


图 1 Feature Mapping 对话框

点击 Run 运行计算并等待计算完成。

该操作可以识别出所选特征元素在这两个化合物中所有可能的位置。

在 Edit and Cluster Features 工具面板中单击 Current Features: All Features 查看训练集分子所表征的药效团特征元素的所有类型。

结果显示这两个化合物都包含有疏水中心 (Hydrophobe)、氢键供体 (Donor)、氢键受体 (Acceptor)、正电离子中心 (Ionizable Positive) 及芳香环中心 (Ring Aromatic) 这五种特征元素的类型。在构建药效团模型时采用 Hydrophobic Aromatic 特征取代 Ring Aromatic 特征，因为前者限制条件少，只定义了一个位点，没有后者的平面和矢量限制。

一旦药效团模型构建好以后，通常需要依据模型对于测试集分子活性预测能力来评判药效团模型的优劣。在 3D QSAR Pharmacophore Generation 这一 protocol 中可以自动运行这一验证步骤。

在文件浏览器 (Files Explorer) 中，展开 Samples | Tutorials | Pharmacophore，双击打开 test\_serm\_ligands.sd。

### 3. 3D QSAR 药效团的构建

在工具浏览器 (Tools Explorer) 中，展开 Pharmacophores | Create Pharmacophores Automatically，单击 3D QSAR Pharmacophore Generation。

流程对应参数在参数浏览器中打开。

设置 Input Ligands 为 serm\_ligands:All。

展开 Validation 参数组，并设置为 True。

设置 Input Test Ligands 为 test\_serm\_ligands:All。

设置 Minimum Interfeature Distance 为 1.5。

展开 Conformation Generation 参数组，点击 Conformation Generation 右边的栅格，下拉

列表中选择 FAST，能量阈值 Energy Threshold 设置为 10，其他参数默认。（图 2）

对每个小分子产生最多 255 个构象用以表征小分子的构象空间。其中只有能量值在能量阈值 10kcal/mol 之内的构象被保留。

点击 Feature 右边  按钮，打开 Select Features 对话框。

勾选 POS\_IONIZABLE 和 HYDROPHOBIC\_aromatic 左边的复选框，点击 OK。（图 1）

在默认设置基础上添加上一个正电离子特征 POS\_IONIZABLE，一个芳香环特征 HYDROPHOBIC\_aromatic。

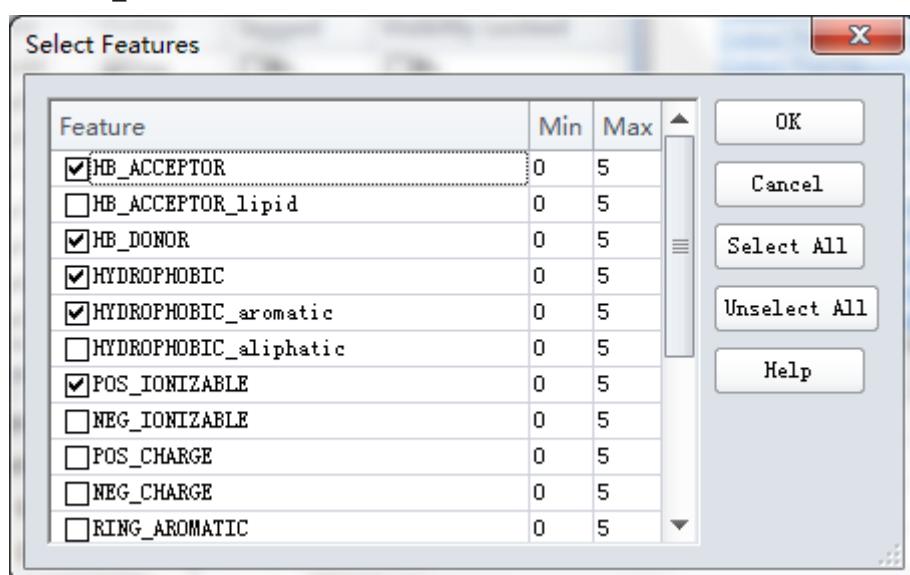


图 1 “Select Features”对话框

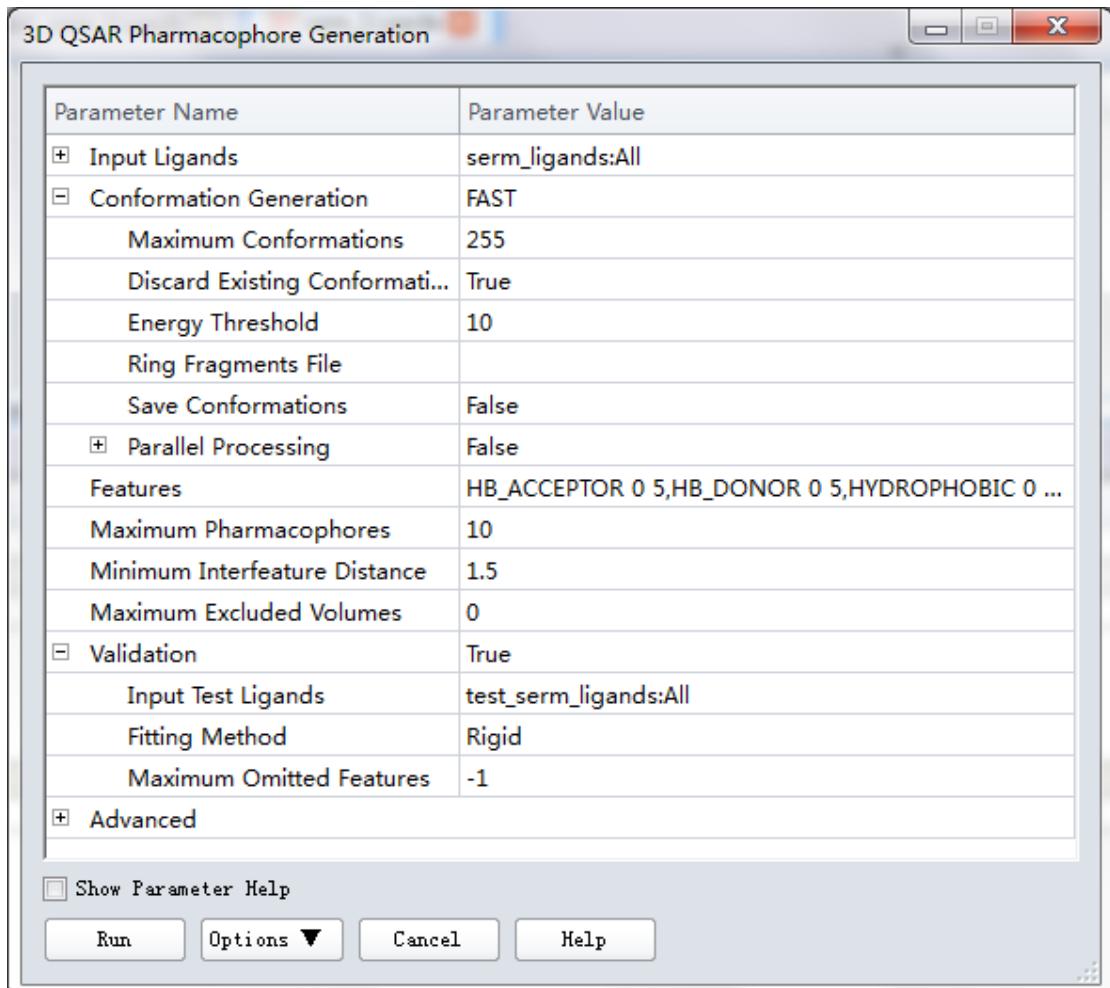


图 2 “3D QSAR Pharmacophore Generation”流程参数设置

点击 **Run** 运行作业。

点击 **Background** 等待作业完成。

该作业大概需要 10min。

## 药效团结果分析

选择一个最好的 3D QSAR 药效团模型不仅仅只是考虑药效团的排名(打分排名根据 cost 值的高低)。排名第一的药效团模型并不一定就是最好的模型。因此，除了分析模型的 cost 值，还需要考察训练集分子与模型的匹配情况从而考察模型是如何解释不同的小分子具有不同的活性值。此外最重要的，是用已知活性值的测试集分子来验证模型对训练集以外的分子活性值的预测能力。

### 1. 查看 Report 页面

作业完成后，双击作业浏览器 (Jobs Explorer) 中相应的行，打开 Report 页面。

从 Summary 一栏可知，此次运算一共产生了 10 个药效团，且包含测试集验证统计数据及测试集 LogEstimate vs LogActiv 的相关性曲线 (图 3)。在 Details 一栏中我们可以了解到每个药效团更为详细的信息。Results 一栏罗列了各结果的链接。Parameters 一栏显示了此次操作所采用的各参数。

利用上述测试集验证信息可以先快速的选取出预测活性能力较好的模型，此处为模型 5、8 最好。接着即可关注这些模型的 cost 值。

## ■ Summary

Generated 10 hypotheses

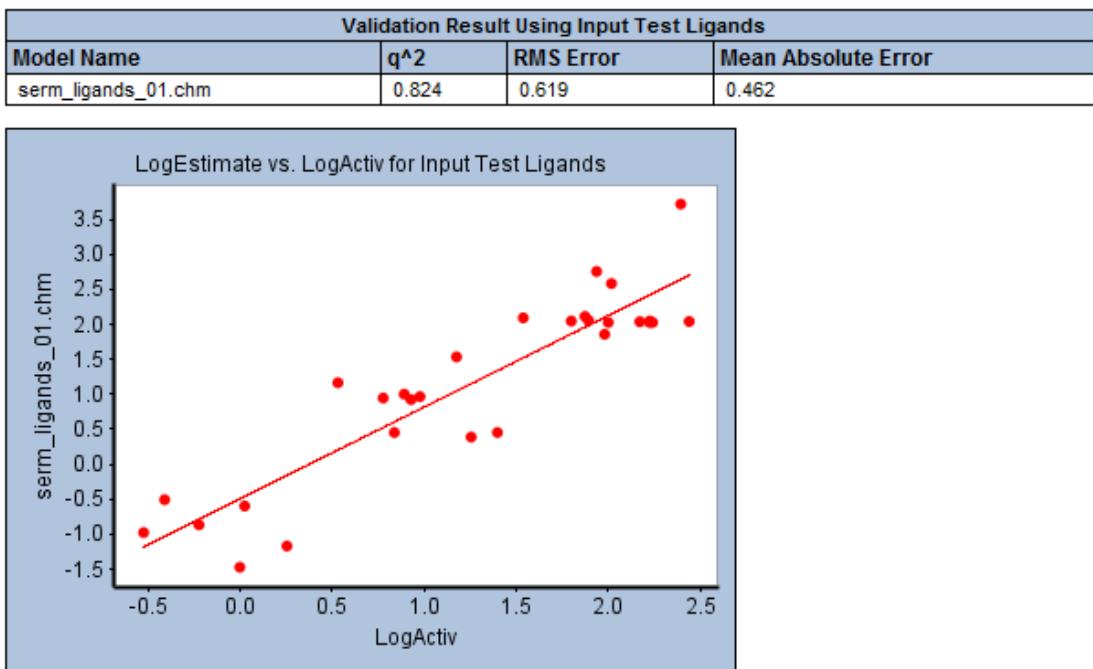


图 3 药效团模型的测试集验证统计数据及测试集 LogEstimate vs LogActiv 的相关性曲线

## 2. 查看药效团与训练集分子的匹配情况

查看训练集分子与所有模型的匹配情况相对比较耗时。该教程中用户可以简单查看训练集分子和药效团模型如何进行匹配。主要考察依据是相较非活性分子，活性分子同模型中匹配的特征元素应更多。也就是说，非活性分子应该缺失某一重要的特征元素，或特征元素存在但在空间上不能进行正确取向。因此，选取药效团模型的一个重要指标即基于训练集分子和药效团模型的匹配情况能够解释对应训练集分子的生物活性。

在 Report 页面中展开 Results，点击 **View Aligned Ligands 01** 链接查看训练集分子同排名第一的药效团的匹配情况。

在表格浏览器中，可以看到有一列名为 Estimate 的栏目，此列为使用该药效团对每个化合物的活性预测值。此列在制作 Common Feature 的药效团中是看不到的。

点击 键显示预测活性最高的化合物（serm-40）同该药效团匹配的情况。（图 4）

该化合物的实验活性值为  $0.02\mu\text{M}$ ，预测活性值则为  $0.023\mu\text{M}$ 。该化合物同药效团的所有特征元素都匹配的很好。

点击 键显示预测活性最低的化合物（serm-49）同该药效团匹配的情况。（图 5）

该化合物的实验活性值为  $235\mu\text{M}$ ，预测活性值则为  $408\mu\text{M}$ ，仍在同一数量级。该化合物只能同药效团中五个特征元素中的两个特征元素匹配上。

点击 键和 键，可以查看到每个分子与该药效团的匹配情况。

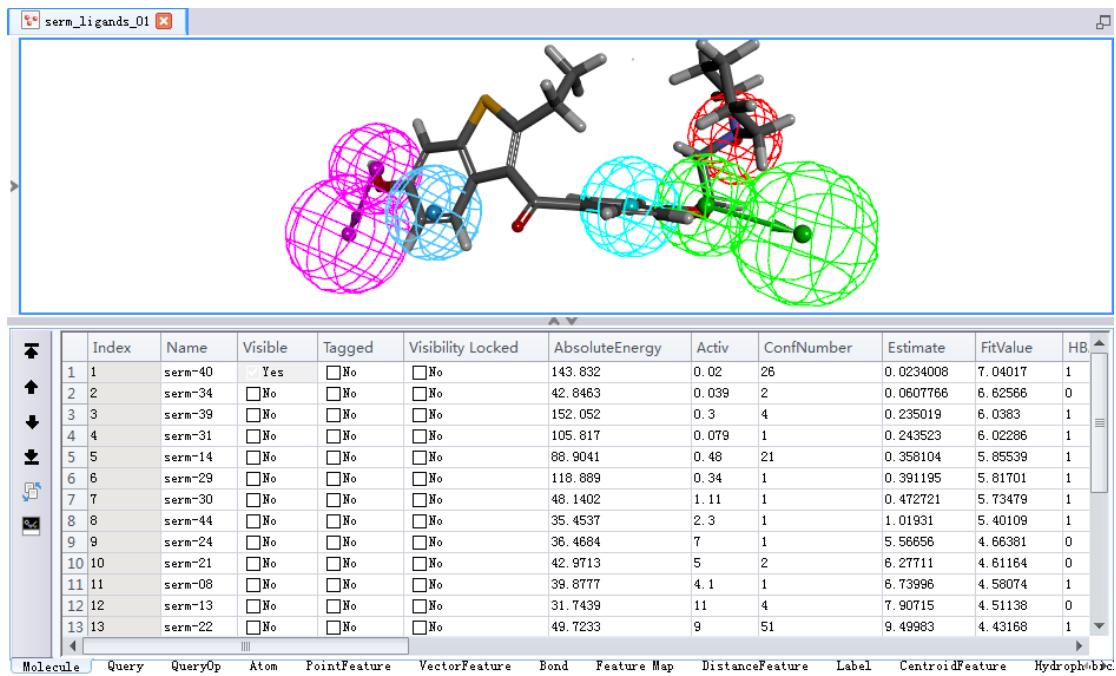


图 4 有预测能力的药效团及其与活性最高小分子匹配情况

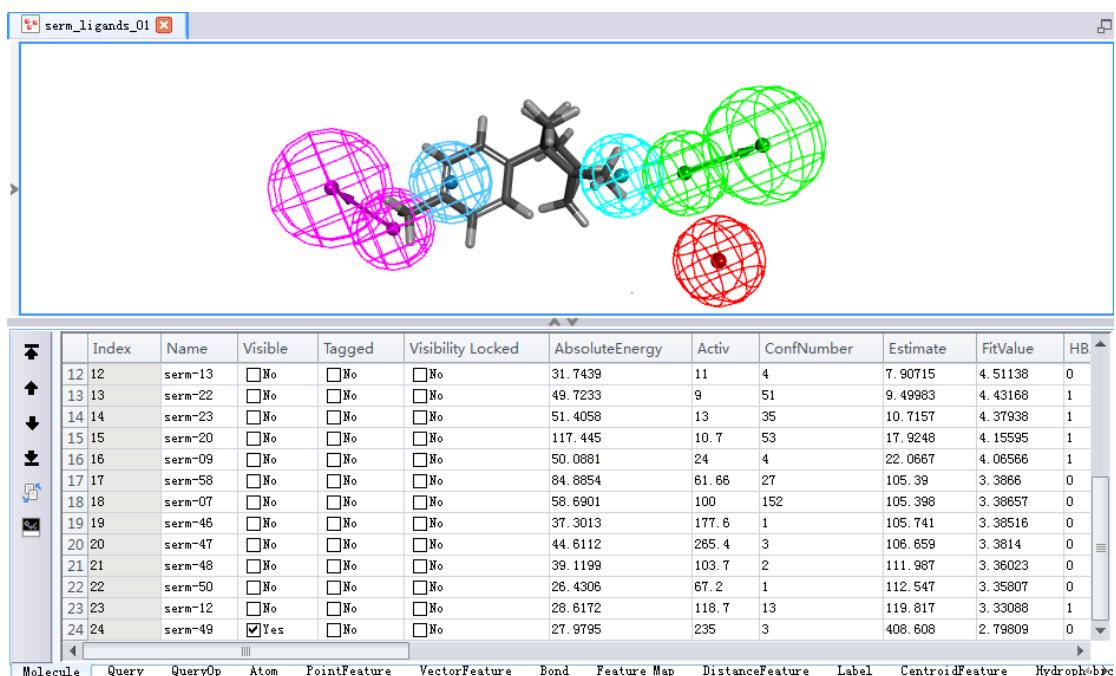


图 5 有预测能力的药效团及其与活性最低小分子匹配情况

接下去查看药效团模型的 log file。

展开作业浏览器 (Jobs Explorer) 中相应的行，点击 Detailed Analysis Report 打开图表形式的 cost 分析报告。（图 6）

该报告主要由三部分组成：

- 重要的操作参数及选取的特征元素；
- 十个药效团模型的 cost 分析摘要；

- 每个药效团模型同训练集分子匹配情况的图表表示。

$\Delta\text{cost}$  (Null cost — Total cost) 是评价一个药效团模型的重要指标，该值为 40-60，表明该模型的置信区间为 75%-90%，该值小于 40，则模型的置信区间降为 50% 以下。

对第三部分图表的简单分析表明第一个药效团模型对训练集化合物活性的预测能力最好，预测活性值和实验活性值之间的偏差最小。

第三部分最右边的表格总结了训练集中每个化合物同某一个药效团模型的匹配信息。表格最后一栏显示了每个化合物同药效团模型的各个特征元素的匹配信息，星号\*表示该特征元素未被匹配上。

一般评价药效团模型的通用规则是活性最高的两个化合物应该同模型中所有特征元素都匹配上。在该例中，只有模型 5、7、8 满足该条件。而且基于测试集验证数据，模型 5、8 也是其中两个验证数据较好的模型。因此接下去可以重点关注这两个模型进行后续研究。当然该规则并非必需条件，具体情况具体分析，例如活性最高的化合物虽然没有匹配上全部的特征元素但是能匹配上的特征元素匹配的更好，或者对于某些比较感兴趣的化合物相关性比较好等等。

### 1. Summary of Run Parameters:

HypoGen Parameters		Features Constraints		
		Name	Min	Max
Spacing	150	HBA	0	5
Variable Weight	No	HBD	0	5
Variable Tolerance	No	HYDROPHOBIC	0	5
		HYDROPHOBAromatic	0	5
		Poslonizable	0	5
		Total	1	5

### 2. Overall Results:

Pharmacophore Space:

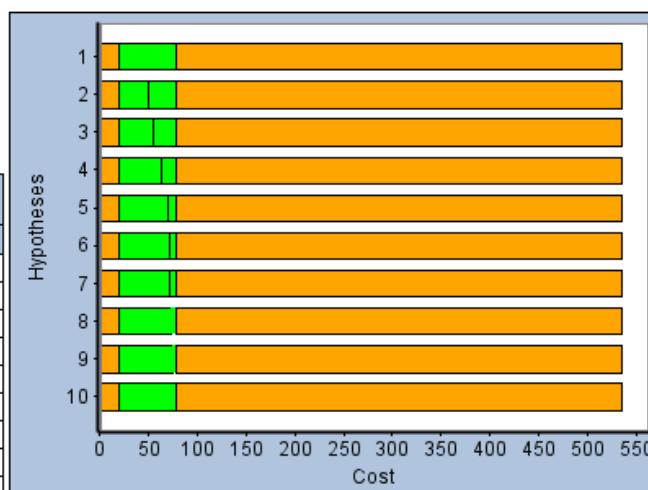
1.706844e+006

Best records in pass: 6.

Fixed Cost: 78.632

Null Cost: 613.375

Cost Analysis (Fixed/Null distance = 534.7 bits)	
Index	Null Cost Distance
1	516.16
2	485.13
3	481.17
4	472.29
5	465.60
6	463.58
7	462.88
8	462.12
9	460.39
10	457.16



**3. Hypotheses Results:****Hypothesis 1**

Results	
Maximum Fit	8.46646
Total Cost	97.2124
RMS	1.22694
Correlation	0.983641

Description	
Features	Weights
HBA	1.69329
HBD	1.69329
HYDROPHOBIC Aromatic	1.69329
HYDROPHOBIC	1.69329
Poslonizable	1.69329

Name	Fit	Est	Act	Err	Status	Mapping
serm-40	7.03	0.024	0.02	+1.2	active	[25 50 45 31 9]
serm-34	6.62	0.061	0.039	+1.6	active	[* 5 1 44 53]
serm-31	6.02	0.24	0.079	+3.1	moderately active	[7 18 1 14 *]
serm-39	6.07	0.22	0.3	-1.4	moderately active	[25 * 47 31 9]
serm-29	5.82	0.39	0.34	+1.2	moderately active	[4 27 22 1 *]
serm-14	5.91	0.32	0.48	-1.5	moderately active	[20 * 9 31 2]
serm-30	5.73	0.47	1.1	-2.3	moderately active	[7 18 1 21 *]
serm-44	5.40	1	2.3	-2.3	moderately active	[11 29 20 1 *]
serm-08	4.58	6.7	4.1	+1.6	moderately active	[32 30 29 11 *]
serm-21	4.61	6.3	5	+1.3	moderately active	[* 31 28 9 *]
serm-24	4.66	5.6	7	-1.3	moderately active	[* 15 18 9 *]
serm-22	4.44	9.3	9	+1	moderately active	[31 * 9 18 *]
serm-20	4.16	18	11	+1.7	moderately active	[18 * 44 23 *]
serm-13	4.54	7.4	11	-1.5	moderately active	[* 28 23 10 *]
serm-23	4.42	9.7	13	-1.3	moderately active	[29 * 12 28 *]
serm-09	4.08	21	24	-1.1	moderately active	[12 3 * 17 *]
serm-58	3.39	110	62	+1.7	moderately active	[* 26 32 *]
serm-50	3.36	110	67	+1.7	inactive	[* 36 11 **]
serm-07	3.39	110	100	+1.1	inactive	[* 12 58 *]
serm-48	3.36	110	100	+1.1	inactive	[* 7 34 *]
serm-12	3.33	120	120	+1	inactive	[* 35 * 22 *]
serm-46	3.39	110	180	-1.7	inactive	[* 6 27 *]
serm-49	2.83	380	230	+1.6	inactive	[* * 9 32 *]
serm-47	3.38	110	270	-2.5	inactive	[* * 6 16 *]

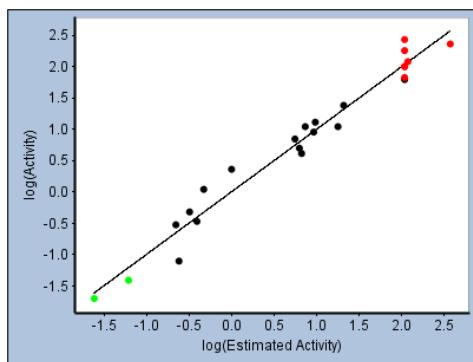


图 6 图表形式的 cost 分析报告

**基于药效团模型进行化合物的设计改造**

在工具浏览器（Tools Explorer）中，展开 **Pharmacophore | Search, Screen and Profile**，点击 **Ligand Profiler**。

流程对应参数在参数浏览器中打开。

该流程可以将多个分子同多个药效团模型快速匹配。

设置 **Input Ligands** 为 **test\_serm\_ligands:All**。

点击 **Input File Pharmacophores** 右下拉菜单，选择 **Browse** 选项，找到之前构建药效团模型所得到的 output 文件。

CTRL-选取药效团模型 5、8（或者 SHIFT-选取所有模型）。

展开 **Conformation Generation** 参数组，点击 **Conformation Generation** 右边的栅格，下拉列表中选择 **FAST**，能量阈值 **Energy Threshold** 设为 **10**。

展开 **Advanced** 参数组，将 **Maximum Omitted Features** 右边的数值改为 **-1** 以允许部分匹配。

点击 **Scale Fit Values** 右边的栅格，下拉列表中选择 **False**。

点击 **Save Aligned Ligands** 右边的栅格，下拉列表中选择 **True**。

其余参数使用默认值。（图 7）

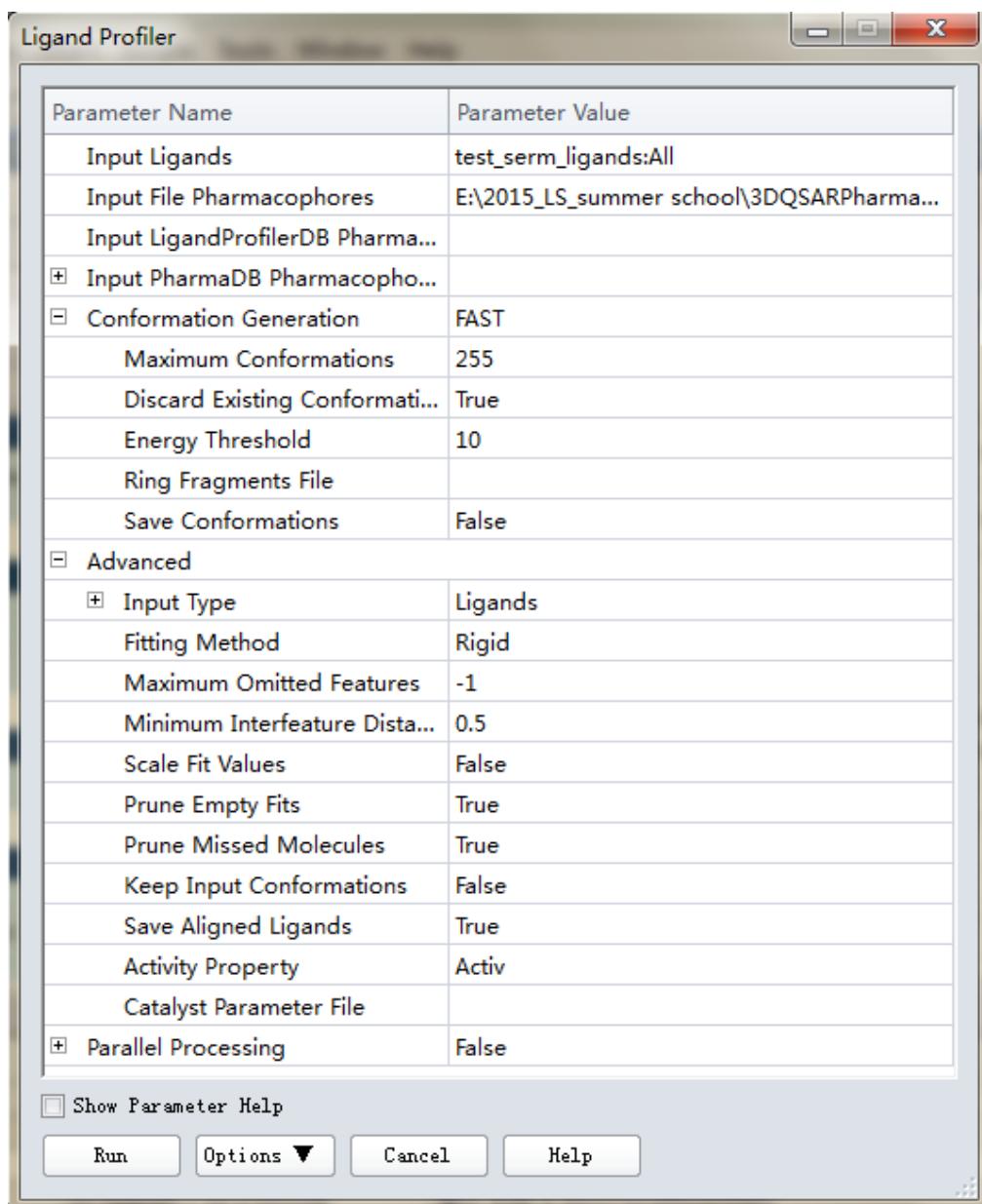


图 7 “Ligand Profiler”流程参数设置

点击 **Run** 运行作业。

点击 **Background** 等待作业完成。

作业完成后，展开作业浏览器（Jobs Explorer）中相应的行，点击 **serm\_ligands\_05\_view.pl**

查看测试集分子同模型 5 的匹配情况。

结果显示了 29 个测试集分子中每个分子同模型 5 匹配得到的最高 FitValue（最低 Estimate）的构象。

勾选 **testserm-38** 化合物的 **visible** 复选框，CTRL-选取化合物 **testserm-38** 和药效团模型 **serm\_ligands\_05**，复制粘贴到新的分子窗口。（图 8）

该化合物的实验活性值为  $63\mu\text{M}$ ，预测活性值为  $93\mu\text{M}$ 。该化合物只能同该模型五个特征元素中的两个特征元素匹配上。该匹配构象的能量值比较低，因此不足为虑。在有些情况下，该值会比较高，应具体分析。

对该化合物未能匹配的药效团特征元素，可以对该分子在结构上作两个小的改造。

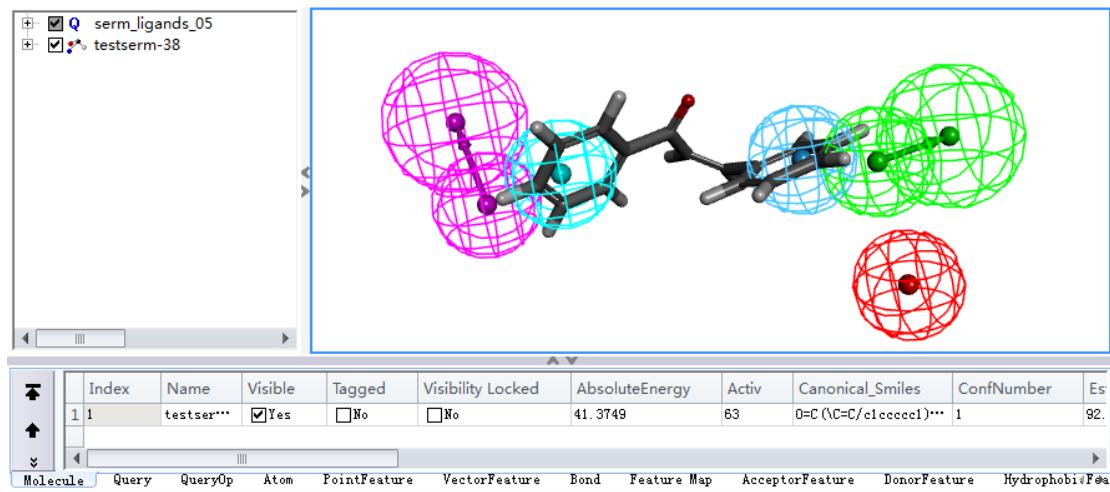


图 8

在系统视图 (Hierarchy View) 中，关闭 **serm\_ligands\_05** 药效团模型的复选框。

对 **testserm-38** 进行结构改造，在苯甲醛的苯环对位加一个酚羟基，在另一个苯环间位加一个酚羟基，对位加一个亚甲基胺。

修改之后的分子如图 9 所示：

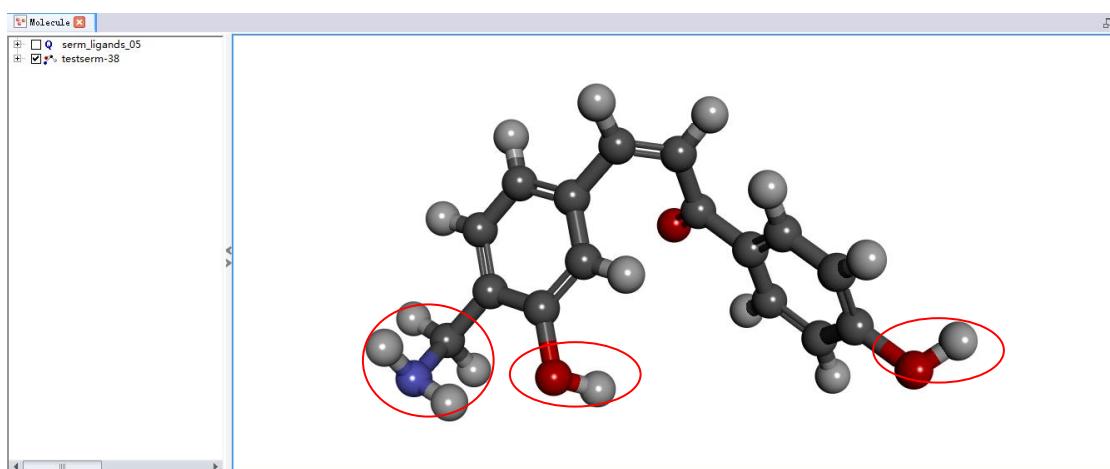


图 9 改造的 testserm-38 分子结构

在系统视图 (Hierarchy View) 中，重新选取 **serm\_ligands\_05** 药效团模型的复选框。

在工具浏览器 (Tools Explorer) 中，展开 **Pharmacophore | Search, Screen and Profile**，点击 **Ligand Pharmacophore Mapping**。

流程对应参数在参数浏览器中打开。

设置 **Input Ligands** 为当前窗口中改造过后的 **testserm-38** 分子。

设置 **Input Pharmacophore** 为当前窗口中的 **serm\_ligands\_05** 药效团模型。

点击 **Best Mapping Only** 右边的栅格，下拉列表中选择 **False**。设置 **Maximum Omitted Features** 为 -1。

展开 **Input Ligands/ Conformation Generation** 参数组，能量阈值 **Energy Threshold** 设为 10，其他参数均使用默认值。（图 9）

点击 **Run** 运行作业。

等待作业完成。该作业几秒钟即可完成。

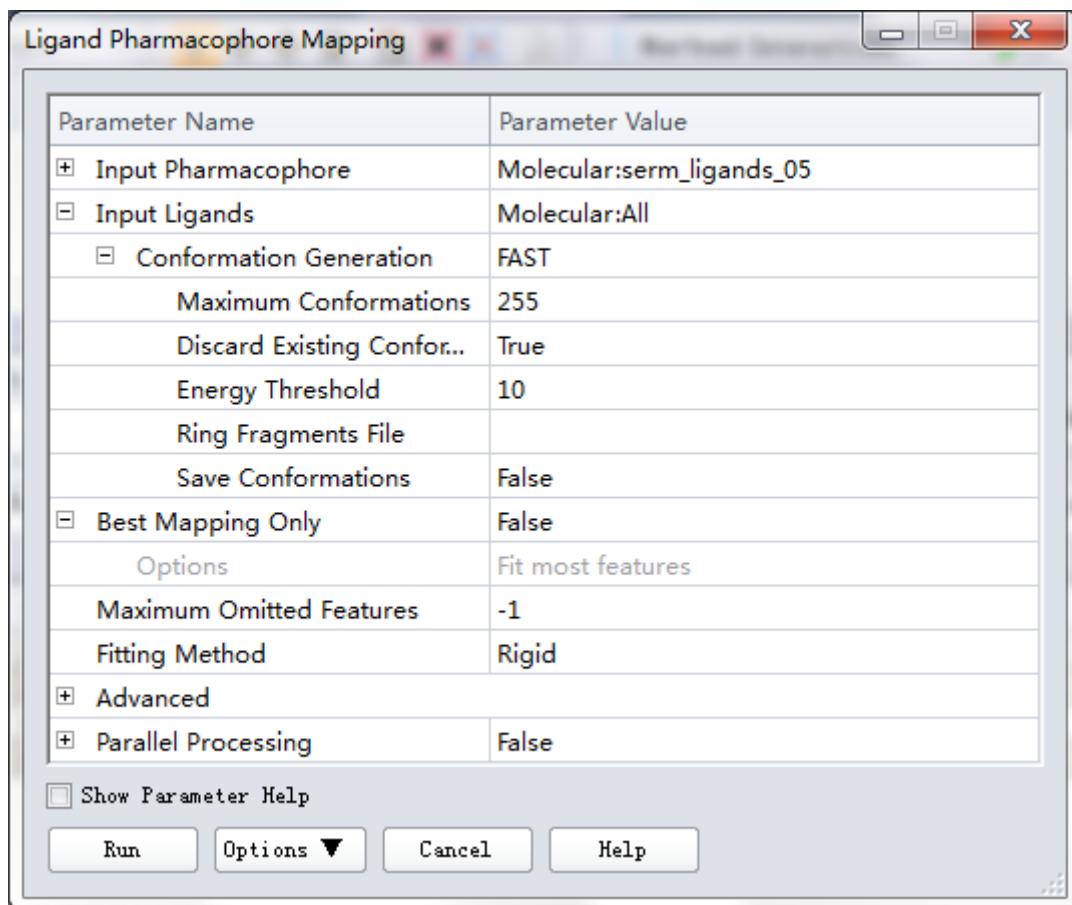


图 10 “Ligand Pharmacophore Mapping”参数设置

双击作业浏览器 (Jobs Explorer) 窗口中相应的行，打开 Report 页面点击 **ViewResults** 按钮。

在表格浏览器中点击 键，查看改造之后的 testserm-38 预测活性最高的匹配构象。

结果显示改造之后的化合物预测活性小于  $1\mu\text{M}$ 。药效团模型五个特征元素的四个都被匹配上 (图 10)，构象能量值仍保持较低的值。

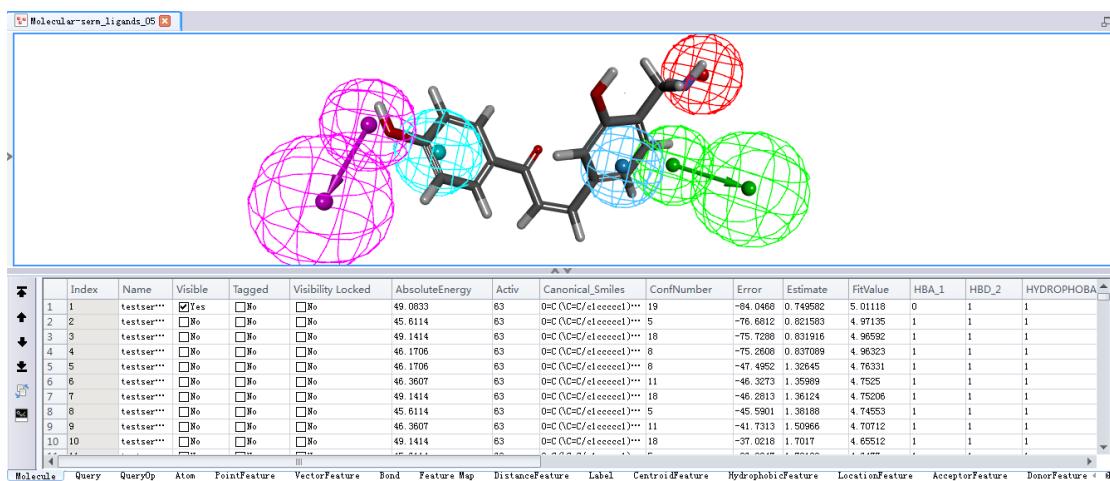


图 11 药效团模型同改造化合物的 ligand pharmacophore mapping 结果

在表格浏览器中点击  键，查看该化合物第五个匹配构象。

结果显示匹配情况有所改变。这一现象说明对不同的匹配情况进行研究有可能可以发现新的结合模式。

双击 **Pharmtype** 一栏两次，使匹配的特征元素个数由高到低排列，单击  按钮。

可以发现药效团模型五个特征元素都被匹配上，该匹配构象的预测活性值为  $3.20\mu\text{M}$ 。因此在有些情况下，化合物与特征元素匹配度更高比匹配的特征元素多但匹配度低要更好。

**提示：**如果运行某一个流程时经常使用相同的设置，但又不想在每次操作时对参数重新进行设定，可以在设定完参数以后，在参数设置浏览器中点击 **Options**，选取 **Save Protocol Settings** 以保存参数设置。

## 构建基于受体结构的药效团模型（SBP）教程

**目的：**通过此教程，了解 Discovery Studio 中构建基于受体三维结构药效团的操作方法及结果分析。

**所需功能和模块：**Discovery Studio client, DS CATALYST SBP, DS CATALYST SCORE, DS CATALYST CONFOMATION。

**所需数据文件：**1k22\_protein\_ligand.dsv, thrombin\_ligands.sd, kinase\_ligands.sd。

**所需时间：**30 分钟

### 介绍

来自学术界和制药企业的信息表明，Catalyst 是最好的基于小分子结构的药物设计工具。然而，科学家的愿望不仅仅是这些，大家也非常希望最大限度地利用已知的受体结构和药物—受体结合信息，有效地克服现有药物设计工具的缺点，更快地加速新药研发的过程。

Structure Based Pharmacophore (SBP) 利用已知或假设的蛋白活性位点从受体位点的特性直接得到相互作用位点图，并且将这个信息转化成适用于快速三维数据库检索的药效团模型。这些相互作用位点图还可以进行编辑、分类，以确保只有最重要的信息会被保留下来用于虚拟筛选。同时 SBP 中还可考虑受体活性位点处氨基酸残基的空间排布，使得通过虚拟筛选得到的分子不仅满足和受体结合位点化学特征上的互补，还可以在空间上很好地结合在活性位点空腔处。SBP 引入了那些使用受体结构信息的最佳方法，这些方法已经在高通量筛选的实验环境中所使用。

SBP 方法首先通过对活性部位的分析产生特征相互作用图 (feature interaction map)，特征相互作用图用于反映可能的配体与蛋白之间的合理的相互作用，接着通过特征相互作用图产生药效团模型，进一步利用药效团模型发现潜在的活性化合物。

本教程中使用  $\alpha$ -凝血酶作为例子，展示如何利用 SBP 方法建立药效团模型，开展数据库搜索，以及 Discovery Studio 如何提供了方便的结果分析工具帮助大家获得最好的结果。

本教程包括以下步骤：

- 构建基于蛋白结构的药效团模型
- 编辑药效团模型
- 配体库筛选
- 筛选结果的分析

### Structure-Based Pharmacophore 的构建

本教程基于  $\alpha$ -凝血酶蛋白结构 (PDB 号：1K22) 来构建药效团模。该蛋白结构从 PDB 库中下载得到并经 Prepare Protein 预先准备。

#### 1. 导入蛋白及配体结构

在文件浏览器 (Files Explorer) 中，展开 Samples | Tutorials | Pharmacophore，双击打开 **1k22\_protein\_ligand.dsv**。

$\alpha$ -凝血酶是丝氨酸蛋白酶，是胰凝乳蛋白酶家族中的一员。该蛋白共有三个结合口袋，口袋处残基分别用红色、蓝色和绿色表示。（图 1）

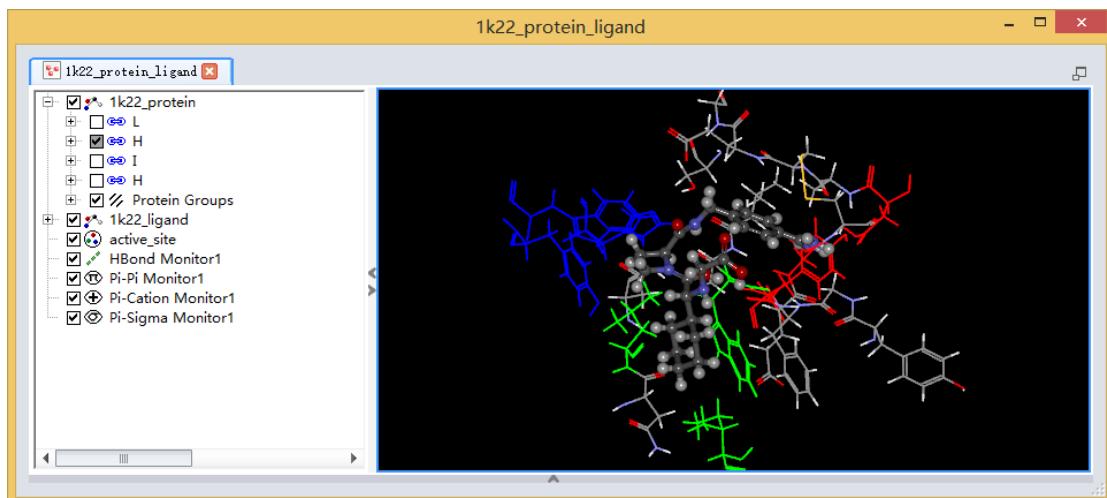


图 1 受体分子

## 2. 定义活性位点

在系统视图（Hierarchy View）中选取 **1k22\_protein**。

在工具浏览器（Tools Explorer）中，展开 **Receptor-Ligand Interactions | Define and Edit Binding Site**，在工具面板中单击 **Define Receptor**。

将前面选取的蛋白分子 **1k22\_protein** 定义为受体分子，供下一步使用。

在系统视图（Hierarchy View）中选取 **1k22\_ligand**。

在 **Define and Edit Binding Site** 工具面板中点击 **Define Site** 栏中的 **From Current Selection**。

在视图窗口（Graphic View）中自动生成了一个红色的球，该球将 **1k22** 配体包裹其中。同时在系统视图（Hierarchy View）中增加了 **SBD\_Site\_Sphere** 的选项。

如果蛋白结构中没有配体小分子的信息，则可以选择已知活性位点处的残基，或者用 **binding site** 工具自动搜索结合位点。

在视图窗口（Graphic View）中，选取红色的球体，或者系统视图（Hierarchy View）中选择 **SBD\_Site\_Sphere**，右击选取 **Attributes of SBD\_Site\_Sphere**。

打开 **SphereObjectAttributes** 对话框，将 **Radius** 设为 **9**。（图 2）

球体半径值可以调大以包含活性位点处的所有残基，但是这同样也会增加所要识别的相互作用的数量，其中可能包含一些非相关信息。

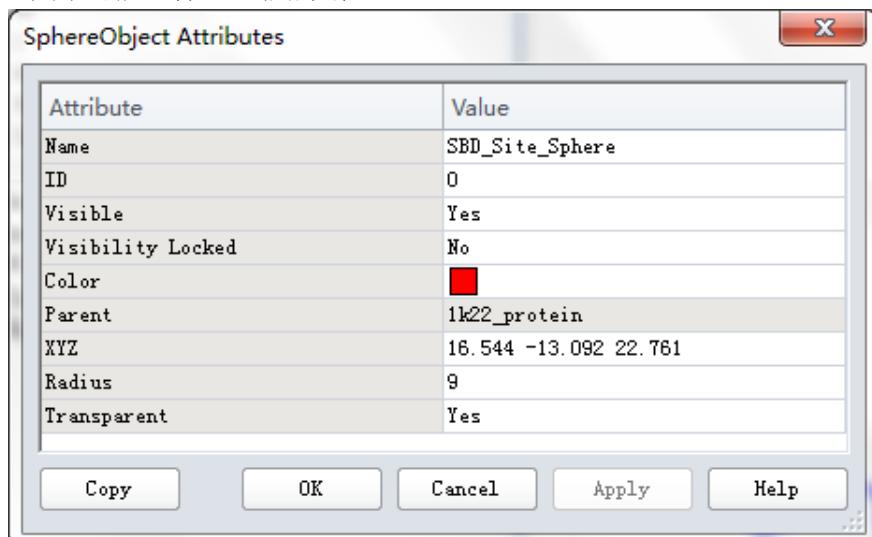


图 2 修改 Sphere 半径

### 3. 产生相互作用模型

在受体活性位点球体中产生相互作用模型。

在工具浏览器 (Tools Explorer) 中, 展开 **Pharmacophore | Edit and Cluster Pharmacophore Features**, 在工具面板中单击 Interaction Generation, 打开 Interaction Generation 对话框。

设置 Input Receptor 为 **1k22\_protein\_ligands:1k22\_protein**。

设置 Input Site Sphere 为已定义的半径为 9 埃的球体坐标。

其余参数使用默认值。 (图 3)

点击 Run 运行作业。

等待作业完成。

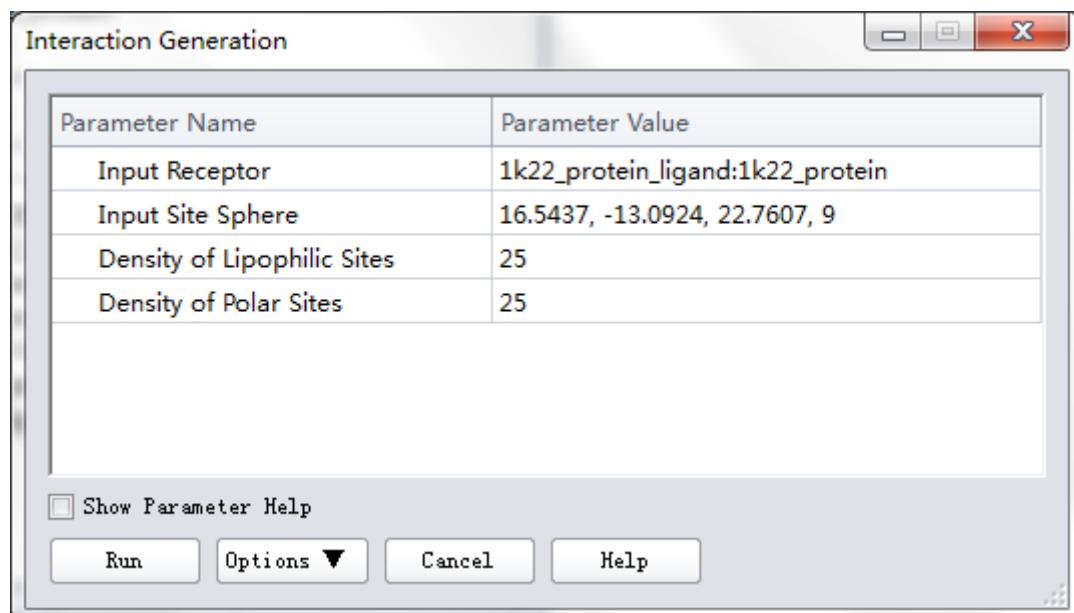


图 3“Interaction Generation”对话框

作业完成以后, 在视图窗口 (Graphic View) 中会自动插入产生的药效团特征。(图 4)

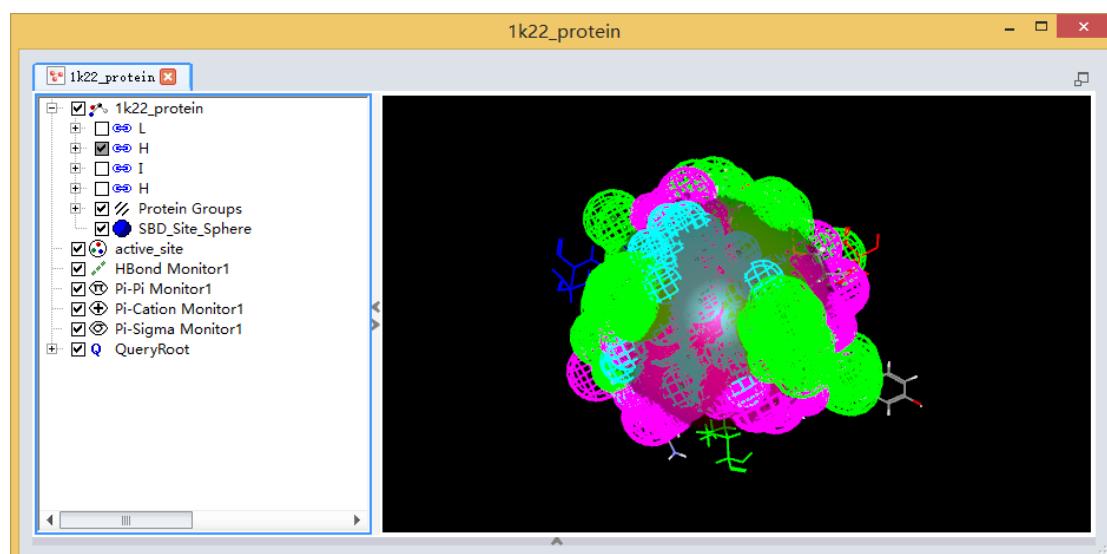


图 4 产生的药效团模型

## 基于结构药效团模型的编辑

初次构建相互作用模型时，20 个药效团特征元素是一个比较合理的数目。

### 1. 隐藏所有特征元素

在工具浏览器（Tools Explorer）中，展开 **Pharmacophore | Edit and Cluster Pharmacophore Features**，在工具面板中单击 **Show/Hide Location Spheres**，隐藏 feature 周围的球。

注意矢量化的特征元素可以表征同某一个特定的氨基酸残基间的相互作用。

在系统视图（Hierarchy View）中，关闭 **SBD\_Site\_Sphere**。

此时，视图窗口（Graphic View）中只有不同颜色标记的结合口袋，配体分子，分子间相互作用。

在 **Edit and Cluster Pharmacophore Features** 工具面板中，定义 **Current Feature** 为 **All Features**。

单击 **Show/Hide All Features**，隐藏所有特征元素。

### 2. 对 feature 进行聚类

首先定义 **Current Feature** 为 **Acceptor**，点击 **Edit and Cluster Pharmacophore Features** 工具面板中的 **Show/Hide Acceptor**，显示所有氢键受体特征。

点击 **Edit and Cluster Pharmacophore Features** 工具面板中的 **Cluster Current Features**，对 **Acceptor** 特征进行聚类分析。

自动出现一个新的窗口，包含聚类分析的树形图。（图 5）

将该 **Dendrogram** 窗口标签拖至视图窗口使得该窗口同 **1k22\_protein\_ligand** 分子窗口并列显示。

**注：**在该窗口中滚动鼠标滚轮，可以纵向调节图形大小，方便观察所有树形分支。移动该窗口中的 **slider** 线，显示不同的聚类中心。相应的该聚类中心会在分子窗口中高亮显示。该聚类是基于所有特征元素间的几何距离来定义的。自动聚类分析和手工选取相结合可以方便用户选取最佳特征元素的组合来表征蛋白活性位点。

移动窗口中的 **slider** 线，直到窗口右下角显示 **Number of cluster centers** 为 12。

在工具面板中点击 **Keep Only Cluster Centers**，关闭 **Dendrogram Window**。

这样所有的 **Acceptor feature** 将由 12 个 **Acceptor feature** 代表。

在系统视图（Hierarchy View）中，展开 **QueryRoot | Fit Fit\_1**。

CTRL-选取 **acceptor feature28, 46, 56, 77, 97**，单击工具面板中的 **Keep Only Current Feature Selections**，以保留这些 **Acceptor** 特征。

选取这些特定元素特征是因为这几个 **Acceptor** 特征对应了活性位点的特定氨基酸残基，而其他特征则要么离残基比较远要么取向偏离了可能的结合位点。

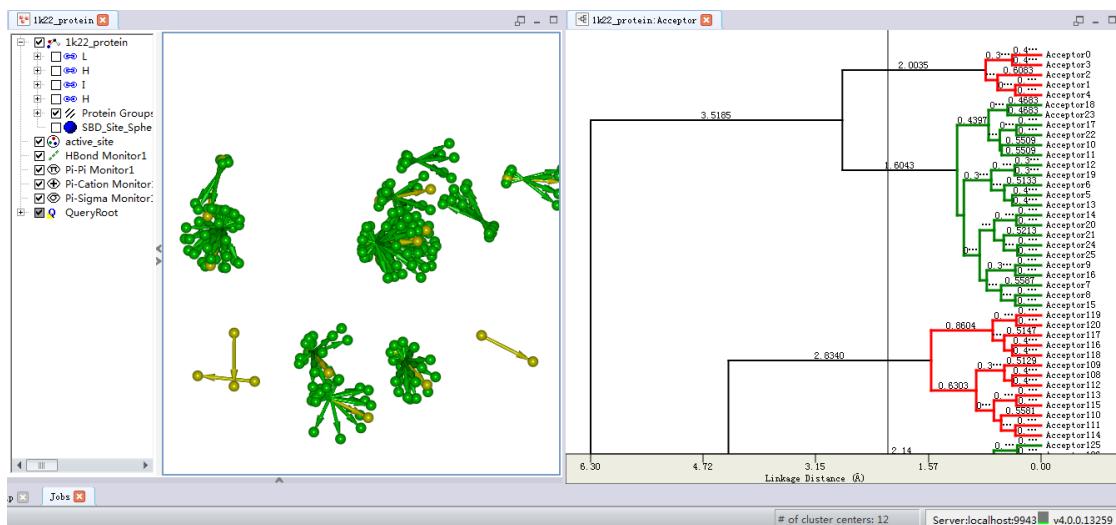


图 5Acceptor-Dendrogram 示意图

单击工具面板中的 **Show/Hide Acceptor**, 隐藏 Acceptor 特征。

设置工具面板中的 **Current Feature** 为 **Donor**, 点击 **Show/Hide Donor**, 显示所有氢键供体特征。

在系统视图 (Hierarchy View) 中, 展开 **QueryRoot | Fit Fit\_1**。

CTRL-选取 donor feature44, 63, 75, 76, 94, 101, 107, 141, 198, 207, 单击工具面板中的 **Keep Only Current Feature Selections**, 以保留这些 Donor 特征。

手动选取这些特定元素特征作为 cluster center 的代表同活性位点重要残基相互作用。这些特征的选取仅仅是依据基于蛋白结构信息得到的相互作用模型, 用户也可以依据已知的氢键相互作用来手工选取。

单击工具面板中的 **Show/Hide Donor**, 隐藏 Donor 特征。

设置工具面板中的 **Current Feature** 为 **Hydrophobe**, 点击 **Show/Hide Hydrophobe**, 显示所有疏水特征。

在系统视图 (Hierarchy View) 中, 展开 **QueryRoot | Fit Fit\_1**。

CTRL-选取 hydrophobe feature9, 48, 81, 单击工具面板中的 **Keep Only Current Feature Selections**, 以保留这些 Hydrophobe 特征。

设置工具面板中的 **Current Feature** 为 **All Features**, 点击 **Show/Hide All Features**, 显示所有保留特征元素。

蛋白活性位点处药效团模型共由 18 个特征元素表征, 特征元素的选取是基于观察到的配体和蛋白间相互作用, 以及从单一蛋白结构得到的作用。

**注:** 另一个用于编辑药效团模型的重要工具是 *Average Pharmacophore Features* 命令。选取同一类型的特征元素, 再用该命令基于这些特征元素得到一个平均特征用以代表整个类别。

在系统视图 (Hierarchy View) 中, CTRL-选取 acceptor feature28, 77, 97, donor feature63, 75, 141, 198, 从菜单栏中选取 **Edit | Group**。

命名为 S-interactions。

该步骤所定义的相互作用都表征的是 S-口袋处的残基。

### 3. 产生排除体积模型 (exclusion model)

在系统视图 (Hierarchy View) 中, 选择 **QueryRoot**, 然后在主菜单栏中选择 **Edit | Select**。

在 **Select** 对话框中, 点击 **Selection Mode** 右边的栅格, 下拉列表中选择 **Replace**, 点击 **Scope** 右边的栅格, 下拉列表中选择 **AminoAcid**。

选取 Radius, 设为 2.00, 点击 Apply 按钮。

点击 Selection Mode 右边的栅格, 下拉列表中选择 Intersection, 点击 Scope 右边的栅格, 下拉列表中选择 Atom。

选取 Property, 在最后一栏下拉菜单中选取 CA, 点击 Apply。 (图 6)

保留的特征元素周围 2 埃以内完整残基中的 CA 原子被选中。

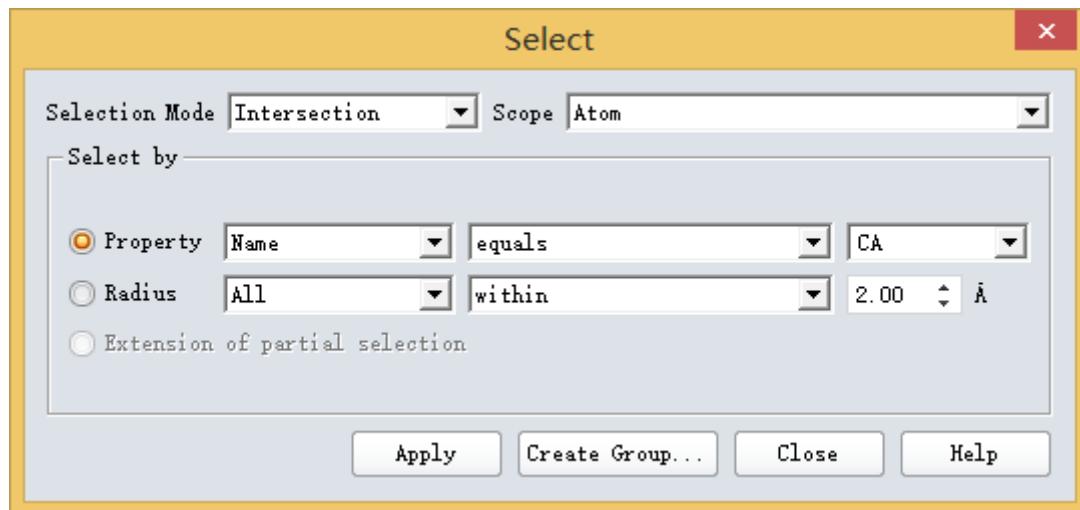


图 6“Select”对话框参数设置

从主菜单栏中选取 Structure | Query Features, 将 Feature 改设为 ExclusionSphere, 基于选取的原子产生排除体积。(图 7)

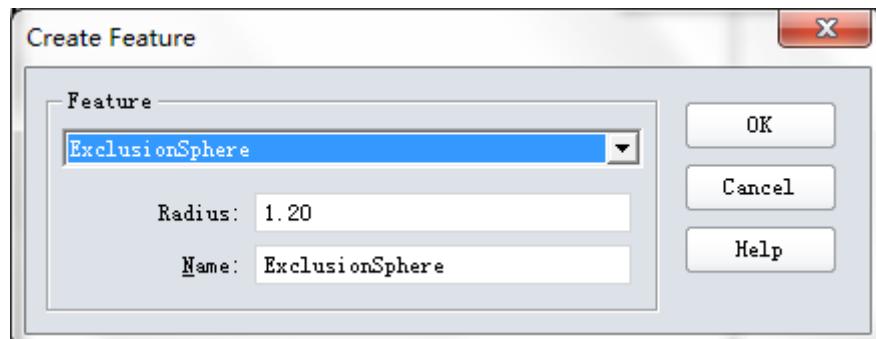


图 7“Create Feature”对话框参数设置

通过以上步骤, 最终得到了一个基于蛋白结构的药效团模型, 含有排除体积。(图 8)

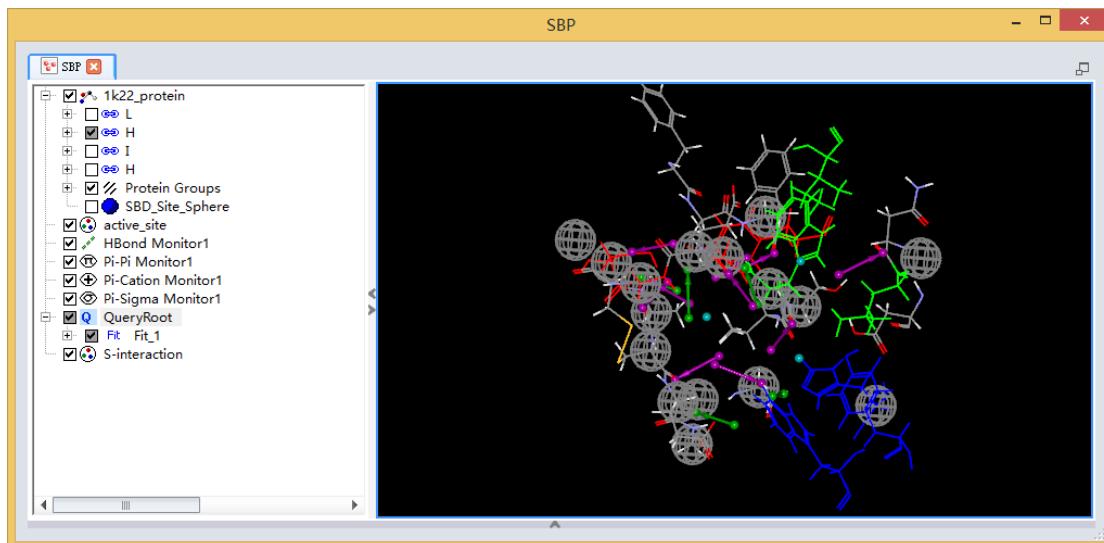


图 8 基于蛋白结构的药效团模型

## 配体库的筛选

在用该药效团模型筛选化合物库之前有必要对该模型进行验证，考察这 18 个药效团特征元素中那些特征元素比较关键。

本教程通过对一个包含了活性分子和非活性分子的测试集进行虚拟筛选来对该模型进行验证。

### 1. 采用药效团进行数据库筛选

在文件浏览器（File Explorer）中，展开 Samples | Tutorials | Pharmacophore，双击打开 thrombin\_ligands.sd。

读入 8 个有活性的凝血酶抑制剂，分子名都有\_ligand 的后缀。

接着展开 Samples | Tutorials | Receptor-Ligand Interaction，将 kinase\_ligands.sd 文件拖至 thrombin\_ligands 窗口中。

读入 9 个激酶抑制剂，这些分子都视为未知配体（理想情况下，这些分子应确保不能同凝血酶结合才可视为负性对照物）。

在系统视图（Hierarchy View）中，选中 Hydrophobe81 特征作为筛选过程中必须出现的特征元素。

在工具浏览器（Tool Explorer）中，展开 Pharmacophores | Search, Screen an Profile，点击 Screen Library，打开 Screen Library 对话框。

设置 Input Ligands 为 thrombin\_ligands:All。

确保 Input Pharmacophore 为 1k22\_protein\_ligand: QueryRoot。

接着指定筛选得到的配体分子必需匹配的特征元素个数，以及用于筛选的每个药效团模型中都必须包含的特征元素。

在 Screen Library 对话框中，展开 Input Pharmacophore 参数组。

点击 Required Features 右边的栅格，下拉列表中选择 True，并展开该参数组，设置 Required Features 为前面选中的 Hydrophobe81。

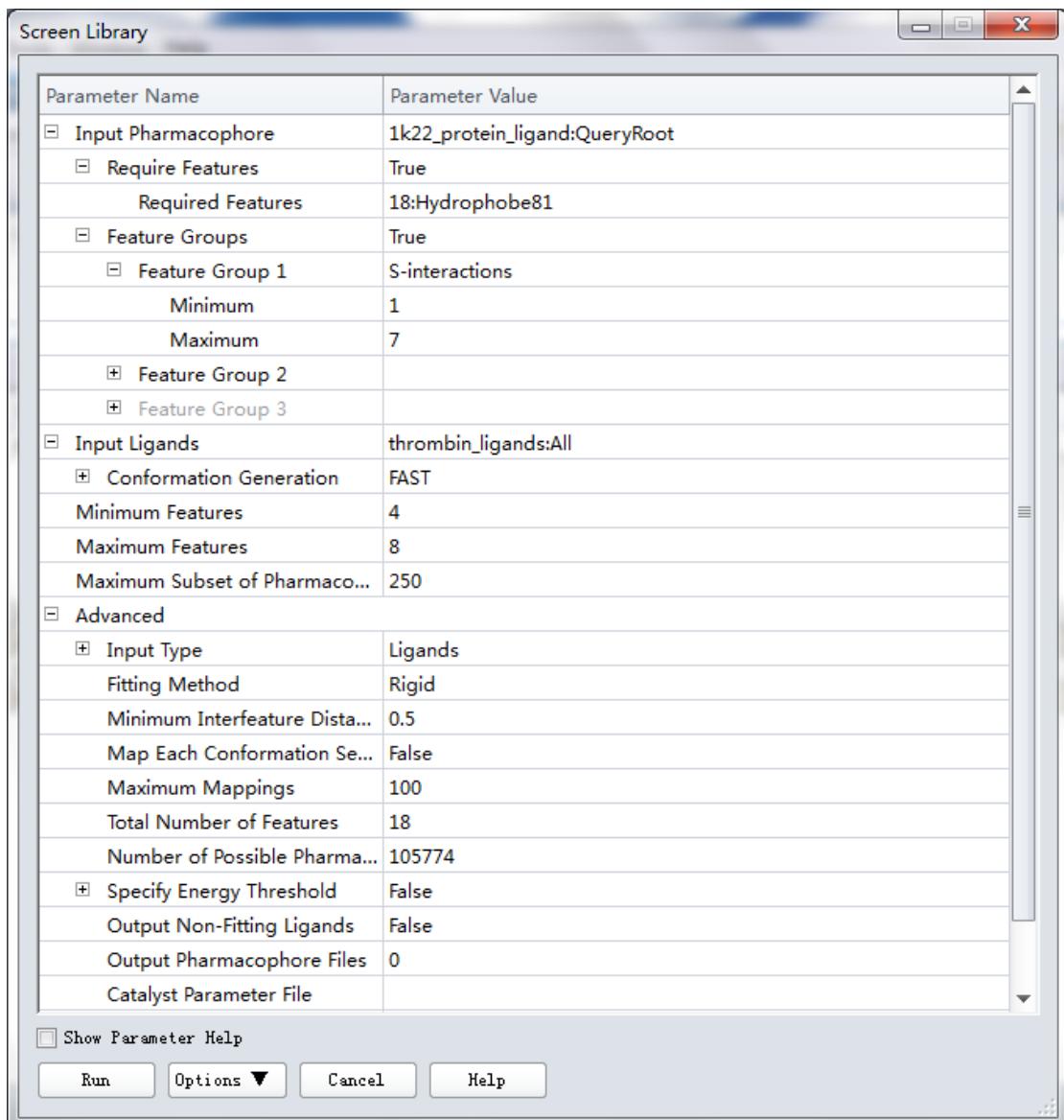


图 9“Screen Library”参数设置

点击 **Features Group** 右边的栅格，下拉列表中选择 **True**，并展开该参数组，设置 **Feature Group 1** 为前面定义的 **S-interactions**。

展开 **Feature Group 1** 参数组，将 **Maximum** 改为 **7**。

该值为该特征元素组中特征元素的上限个数。

将 **Maximum Features** 设为 **8**。

该值为输入药效团模型中需要考虑的特征元素的上限个数。

展开 **Input Ligands** 参数组，点击 **Conformation Generation** 右边的栅格，下拉列表中选择 **FAST**。

展开 **Advanced** 参数组，查看 **Number of Possible Pharmacophores**。

该值给出了药效团模型几种组合的可能性，是基于 **Total Number of Features**，**Minimum Features** 和 **Maximum Features** 这三个参数计算得来的，该值当前显示为 105774。

以上参数的设置表明最终筛选得到的配体小分子必须同编辑得到的 18 个特征元素中的 4 个特征元素相匹配，并且在这 4 个特征元素当中必须有一个是 **Hydrophobe81** 特征，同时还必

须有一个特征来自于特征元素组 S-interactions。

要想正确运行该流程还必须设置 Maximum Subset of Pharmacophores 参数以确保能够搜索到蛋白活性位点的所有药效团空间。

**将 Maximum Subset of Pharmacophores 改为 250。**

其余参数设为默认设置。(图 9)

点击 Run 运行作业。

等待作业完成。

## 2. 分析查看筛选结果

作业完成后，自动打开 Report 页面，点击 View All Results 链接，打开一个新的分子窗口。

该窗口中包含了一个 structure-based pharmacophore 和同所有类别的药效团匹配最好的命中分子。一共筛选得到 55 个分子，包含了 17 个凝血酶和激酶抑制剂中的 11 个配体分子。

在表格浏览器 (Data Table) 中，单击右键，选取 Filter Objects，打开 Specify Filter 对话框。

将 FitValue 的 Choice 选取为 Top (#)，Value 选取为 1。

勾选 Apply filter to each group of identical values 的复选框，选择 Name。(图 10)

针对表格中 Name 相同的分子只保留 FitValue 最高的一个。

可以看到 8 个凝血酶抑制剂中筛选出了 7 个，9 个激酶抑制剂中筛选出了 4 个。

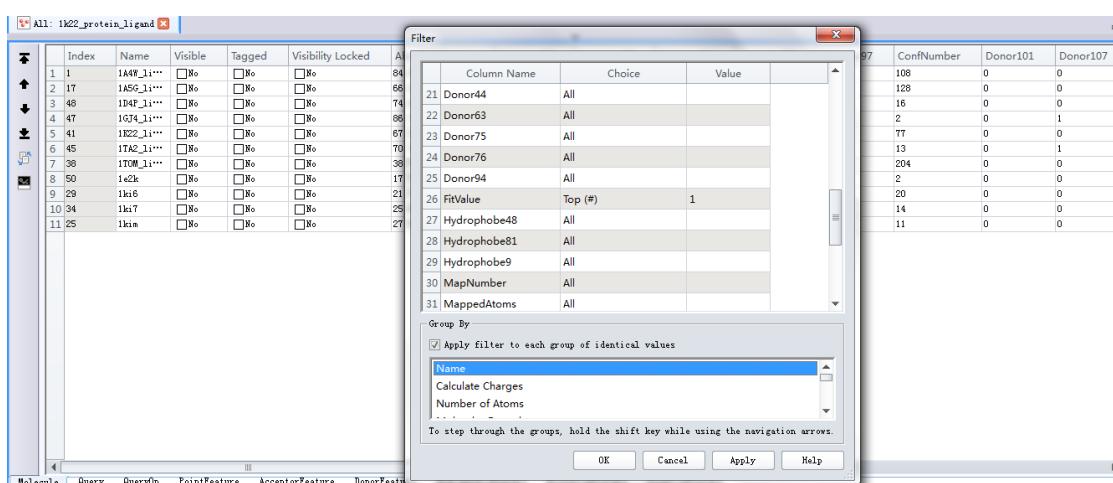


图 10“Specify Filter”参数设置

在表格浏览器 (Data Table) 中，单击右键，选取 Remove Filters，还原前步的设置。

双击 Name 一栏使得凝血酶抑制剂排在表格的前面，不选择表中任何一栏。

从主菜单栏中选取 Chart | Heat Map，跳出 Choose Plot Axes 对话框，设定热图的 X 轴和 Y 轴属性。

针对 Y 轴，选取 Name 属性。

针对 X 轴，选取 Acceptor28，接着 SHIFT-选取 Hydrophobe9。

将 ConfNumber 和 FitValue 属性的选取取消。(图 11)

即纵坐标为 Name，横坐标为所有共 18 个药效团特征元素。

坐标设定完以后，自动生成热图。(图 12)

从该图可以观察到 18 个药效团特征元素分别同每个分子的匹配情况。例如，Hydrophobe81 该特征同所有分子都相互匹配，正如前面所设定的。而 Donor94 则只同很少的分子相匹配，因此当筛选更大的分子库时可以考虑删除该特征。

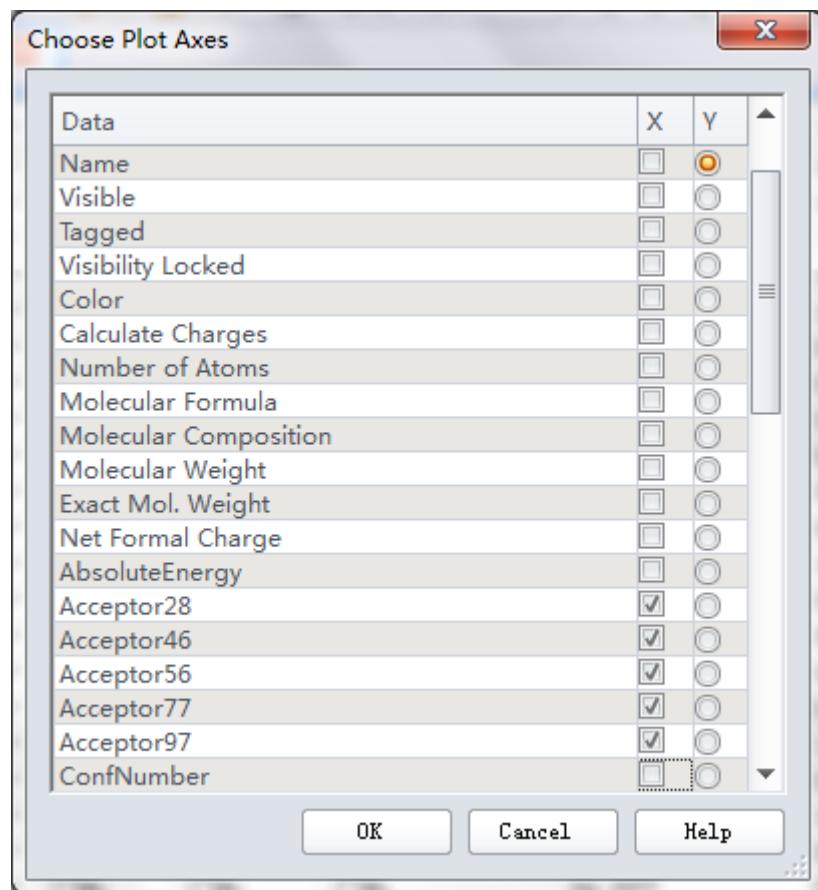


图 11 “Choose Plot Axes”参数设置

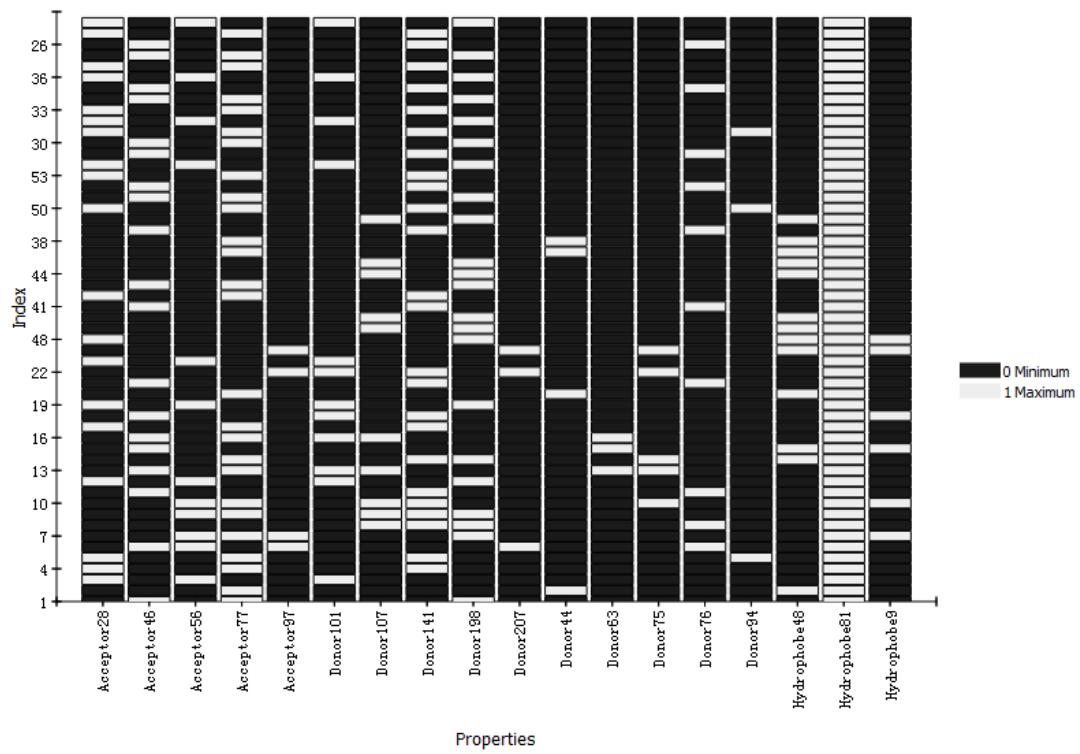


图 12 筛选结果热图

## 构建及验证基于受体-配体复合物药效团教程

**目的:** 通过此教程, 了解 Discovery Studio 中构建及验证基于受体-配体复合物药效团操作方法及结果分析。

**所需功能和模块:** Discovery Studio Client, DS CatalystConformation, DS CatalystHypothesis, DS Catalyst Score, DS Catalyst SBP。

**所需数据文件:** 2irz.pdb, active.sd 和 inactive.sd

**所需时间:** 50 分钟

### 介绍

随着 X-射线晶体衍射和核磁共振技术的进步, 大量的蛋白结构被解析, 尤其是如果受体和抑制剂复合物结构已知, 则从复合物结构中可以得到抑制剂中对活性贡献较大的基团及其空间分布。因此, 在基于活性配体构建药效团模型, 即 HipHop 和 HypoGen 方法以及基于受体的药效团模型 SBP 方法的基础上, 从 Discovery Studio 3.0 版本开始又增加了基于受体-配体晶体复合物构建药效团模型的功能 (Receptor-Ligand Pharmacophore Generation)。

研究分子间相互作用对于基于结构的药物设计非常重要。分子对接是常用的方法之一, 在传统的对接方法中, 对接的准确性往往要打折扣, 因为这些程序可以把化合物放在结合位点的任何位置。而相应的打分方程往往不能找到最有可能的结合位点。但是大部分情况下, 对一个给定的结合位点来说, 哪个相互作用对配体-受体相互作用起关键影响经常是已知的。对于这种情况, 就可以把以经验为主发现的结合位点和已知的结合模式考虑到对接过程中, 创建一个用于对接的药效团模型。这样就可以引导潜在的抑制剂结合到已知的、能量有利的相互作用上。

这个 demo 主要就是介绍如何根据受体-配体晶体复合物方便的构建药效团模型, 并在构建的过程中集成验证功能。以  $\beta$ -分泌酶以及其抑制剂 (Al-Nadaf A, et al. *Bioorganic & Medicinal Chemistry*, 2010, Volume 18, 3088-115.) 作为数据来源进行计算。

本教程包括以下步骤:

- 药效团模型的构建及验证
- 药效团结果分析
  - ◆ Selectivity Score
  - ◆ Validation
- 反向找靶

### 药效团模型的构建及验证

本教程采用  $\beta$ -分泌酶与其抑制剂(结构见图 1)晶体复合物构建药效团模型。

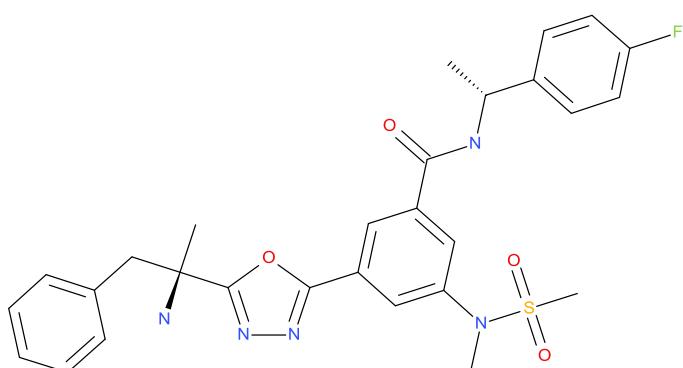


图 1 小分子抑制剂结构

### 1.蛋白的准备

在文件浏览器（Files Explorer）中，找到 2irz.pdb 文件，双击打开在分子窗口中显示(或者点击菜单栏 File| Open URL...，输入 PDB ID:2irz)，分别按下快捷键 Ctrl+H 和 Ctrl+T 打开描述此结构的相关树形窗口和表格窗口（图 2）

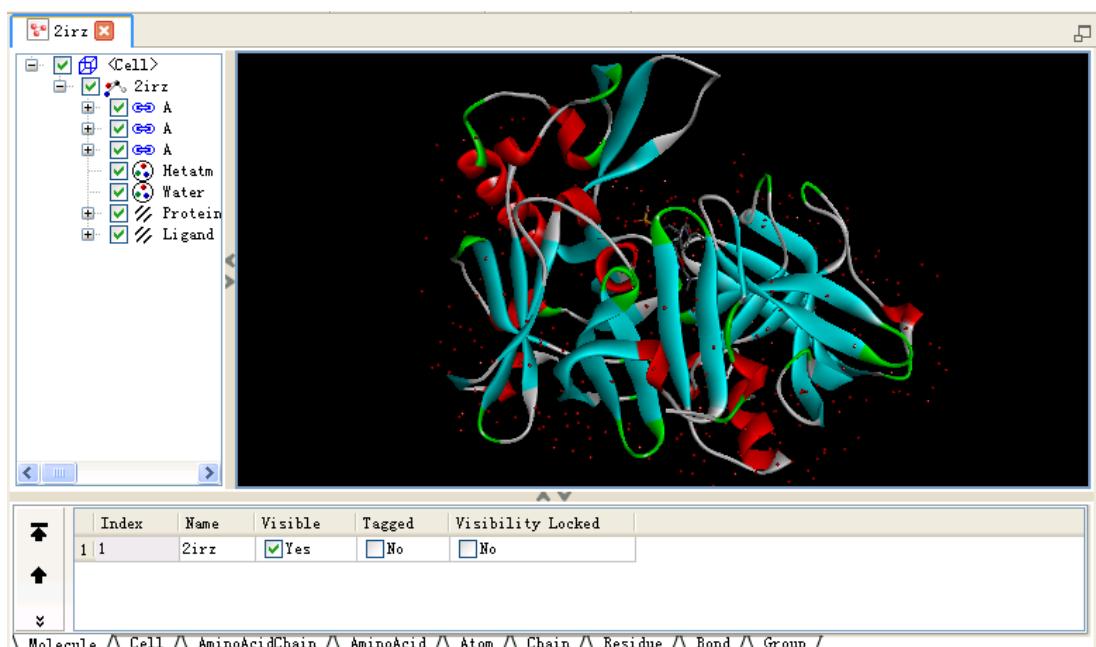


图 2 蛋白-配体晶体复合物 2irz 结构

在工具浏览器（Tools Explorer）中，展开 Macromolecules| Prepare Protein，在工具面板中单击 **Prepare Protein** 打开 Prepare Protein 对话框。

设置 **Input Protein** 为 2irz:2IRZ。

其余参数采用默认设置。

点击 **Run** 运行。

点击 **Background** 等待计算完成。

Prepare Protein	
Parameter Name	Parameter Value
Input Protein	2irz:2IRZ
Build Loops	True
Protonate	True
Advanced	

图 3 蛋白预处理

作业完成后，展开作业浏览器（Jobs Explorer）中该任务并点击 **View Results** 查看结果：

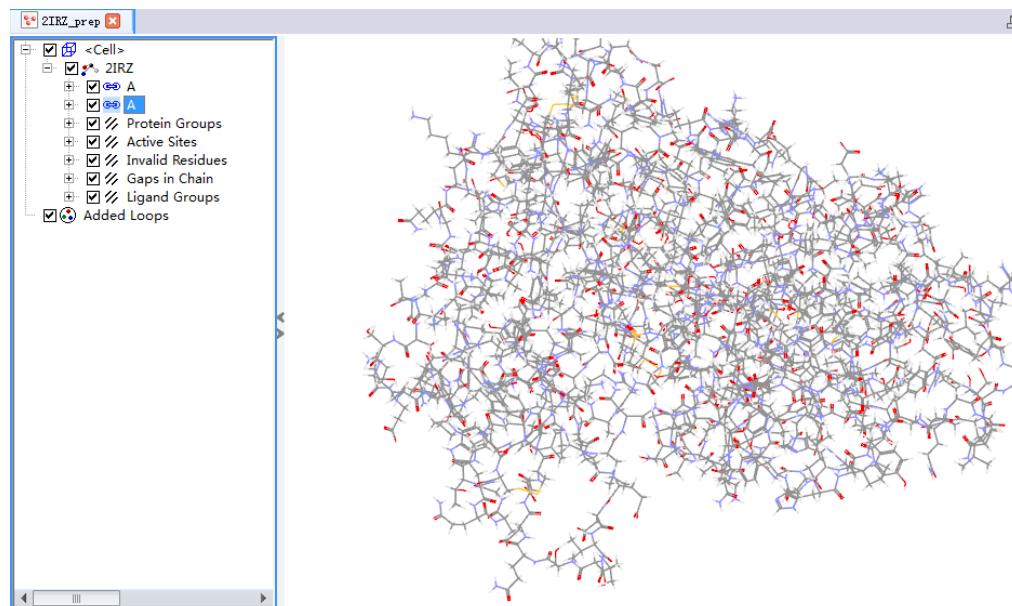


图 4 蛋白经过处理后结果

## 2.配体分子的准备（包括测试集）

点击菜单栏中 **File| New| Molecule Window** 打开一个新的分子窗口。

在 2IRZ\_prep 窗口中点击选中蛋白中第二个 A 链，即配体分子，单击鼠标右键选择 **Cut**，并 **paste** 到上述新建的分子窗口中。

点击菜单栏中的 **Structure| Crystal Cell| Remove Cell**，点击选中 2irz 并单击鼠标右键，选择最后一项 **Attribute of 2irz...**，出现下图对话框，将 Name 改为 Ligand

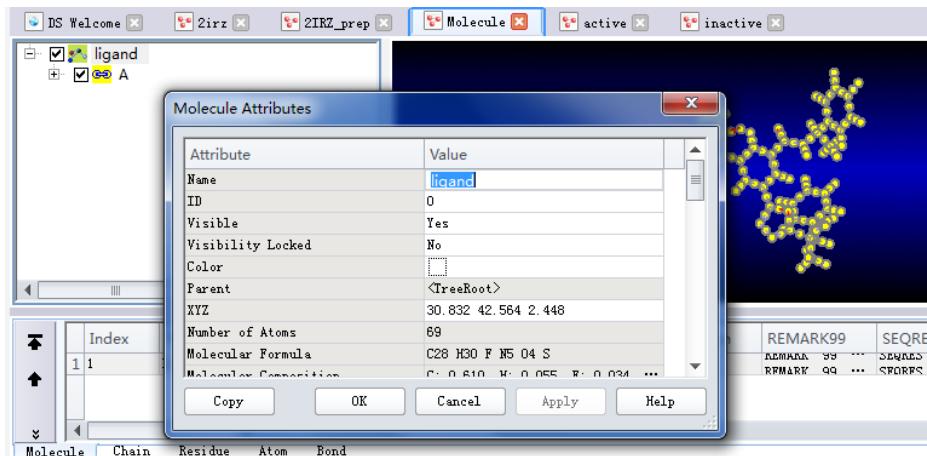


图 5 修改配体名称对话框

在文件浏览器（Files Explorer）中，找到 active.sd 以及 inactive.sd 两文件，并分别双击打开。

注：active.sd 中的 25 个分子为已知的  $\beta$ -分泌酶抑制剂，inactive.sd 中分子为从 zinc 库中下载的非活性分子。用所构建药效团模型命中活性分子的能力来判断该模型的准确性。

### 3.药效团模型的构建及验证

在工具浏览器（Tools Explorer）中，展开 **Pharmacophores | Create Pharmacophores Automatically**，单击 **Receptor-Ligand Pharmacophore Generation**。

流程对应参数在参数浏览器中打开。

设置 **Input Receptor** 为 **2IRZ\_prep:2IRZ**，**Input Ligand** 为 **Molecule:ligand**，

将 **Validation** 设为 **True**，并展开 **Validation**，设置 **Active Ligands** 为 **active>All**，**Inactive Ligands** 为 **inactive>All**。

其余参数采用默认设置。（图 6）

点击 **Run** 运行作业。

点击 **Background** 等待作业完成。

从菜单中选择 **Window | Close All**，关闭所有窗口。

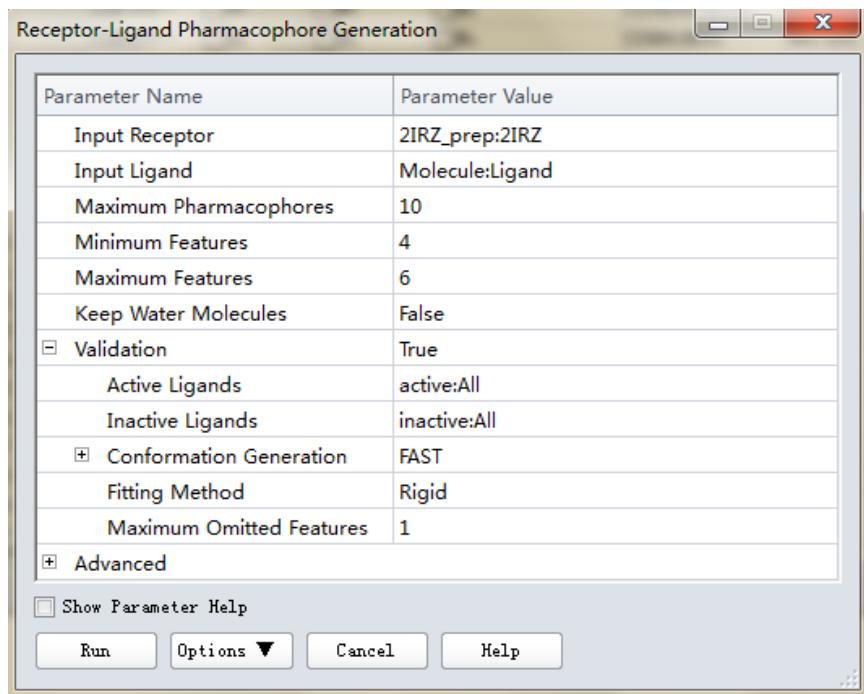


图 6“Receptor-Ligand Pharmacophore Generation”流程参数设置

## 药效团结果分析

该方法产生药效团模型过程中，加入了两种验证方式，包括流程参数设置时 Validation 操作和展开 Advanced 选项后的 Selectivity Scoring 设置。前者通过用已知的活性化合物与非活性化合物建立 decoy set 来验证模型区分活性化合物和非活性化合物的能力，后者判断药效团模型的选择性，分值越高说明选择性越高。

### 1. 查看 Report 页面

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

注：该计算流程为先枚举配体分子所有的化学特征，再保留其中与受体有相互作用的特征，再将其进行排列组合产生最后的药效团模型。

从 Summary 一栏可知，此次运算共产生了 10 个药效团模型。配体分子中共有 30 个化学特征，从 details 中可以看到此 30 个化学特征的详细描述。其中 10 个与受体相关即能够描述该抑制剂与受体的相互作用信息，为 AAAAADHHP。将其排列组合后保留前 10 个 Selectivity Score 最高的模型，具体特征组成和得分如下（图 7）：

Pharmacophore Summary			
Pharmacophore	Number of Features	Feature Set	Selectivity Score
Pharmacophore_01	6	AADHHP	12.732
Pharmacophore_02	6	AADHPR	12.732
Pharmacophore_03	6	AADHHP	12.732
Pharmacophore_04	6	AADHPR	12.732
Pharmacophore_05	6	AADHHP	12.732
Pharmacophore_06	6	AADHPR	12.732
Pharmacophore_07	6	AADHHP	12.732
Pharmacophore_08	6	AAADHP	12.732
Pharmacophore_09	6	AAADHP	12.732
Pharmacophore_10	6	AADHPR	12.732

图 7 程序运行得到的 10 个药效团模型特征组成和 Selectivity Score 值

注：选择性只是通过模型命中化合物库中分子数定义，并没有验证模型筛选能力，如果活性化合物信息了解较多，可以与该 demo 类似，通过用已知的活性化合物与非活性化合物建立 decoy set 来验证模型区分活性化合物和非活性化合物的能力。

该计算产生结果如下（图 8）：

Validation with Known Actives/Inactives								
Pharmacophore	Total Actives	Total Inactives	True Positives	True Negatives	False Positives	False Negatives	Sensitivity	Specificity
Pharmacophore_1	25	618	12	551	67	13	0.48000	0.89159
Pharmacophore_2	25	618	8	592	26	17	0.32000	0.95793
Pharmacophore_3	25	618	15	495	123	10	0.60000	0.80097
Pharmacophore_4	25	618	16	552	66	9	0.64000	0.89320
Pharmacophore_5	25	618	16	492	126	9	0.64000	0.79612
Pharmacophore_6	25	618	15	563	55	10	0.60000	0.91100
Pharmacophore_7	25	618	3	593	25	22	0.12000	0.95955
Pharmacophore_8	25	618	4	542	76	21	0.16000	0.87702
Pharmacophore_9	25	618	4	589	29	21	0.16000	0.95307
Pharmacophore_10	25	618	3	604	14	22	0.12000	0.97735

图 8 得到的 10 个药效团模型对活性化合物以及非活性化合物的命中结果表

其中的 **Sensitivity (SE)** 代表敏感度，**Specificity (SP)** 代表特异性，两值分别表示模型对活性分子以及非活性分子的识别能力。

两量化指标的具体定义如下：

敏感性 **SE(sensitivity) = TP (True Positives) /TP+FN (False Negatives)**

特异性 **SP(specificity) = TN (True Negative) /TN+FP (False Positives)**

两者得分越高，说明模型区分活性与非活性化合物能力越强。需要综合评估，具体可从 ROC 曲线判断，可以看到对于每个模型都具有相应的 ROC 曲线图，以 Pharmacophore\_04 为例（图 9）：

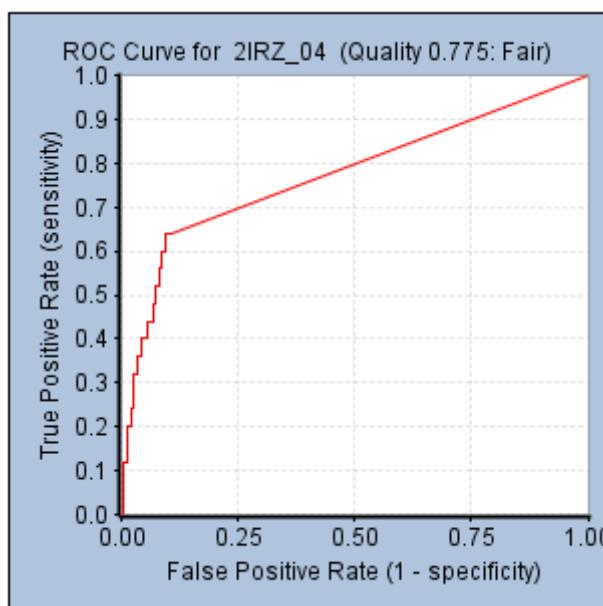


图 9 Pharmacophore\_04 ROC 曲线图

图中横坐标代表假阳性率，纵坐标代表真阳性率，用曲线下面积值代表最后的统计结果，图中最上方描述的 Accuracy0.775 即为曲线下面积，该值应大于 0.5，越大代表模型区分能力越强，后面的 Fair 代表 Pharmacophore\_04 模型（图 10）很好。

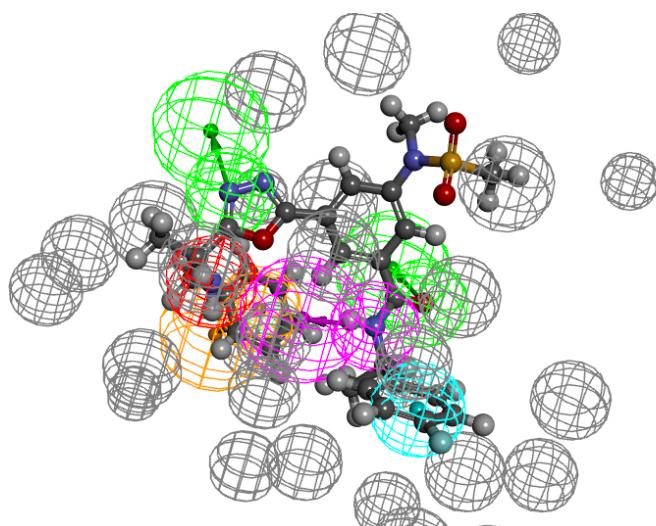


图 10 Pharmacophore\_04 与抑制剂小分子的叠合

## 反向找靶

### 1. 反向找靶

在文件浏览器（Files Explorer）中，找到 **1AJV.sd** 文件，双击打开在分子窗口中显示。

在工具浏览器（Tools Explorer）中，展开 **Pharmacophores | Search, Screen and Profile**，点击 **Ligand Profiler**。

流程对应参数在参数浏览器中打开。

**Input Ligands** 设置为 **1AJV:All**。

点击 **Input PharmaDB pharmacophores** 右边的按钮，在弹出的 **PharmaDB pharmacophores** 对话框中点击选择 **Shape**，点击 **OK**。

其中 **Shape** 表示添加形状束缚的搜索限制，本教程为节省计算时间，所以只选择了 **Shape**。在自行搜索的时候，可以选择 **ALL**，进行全库的搜索。

展开 **Input PharmaDB pharmacophores** 前面加号，**Model Selection** 设置为 **Most Selective**。

其中 **Most Selective** 便是只搜索每个蛋白的第一个药效团模型，本教程为了节省计算时间，在自行搜索时候，可以选择 **All Models** 进行全库搜索。

展开 **Advanced** 参数组，设置 **Save Aligned Ligands** 为 **True**。

其它参数默认，点击 **Run** 运行（图 11）。

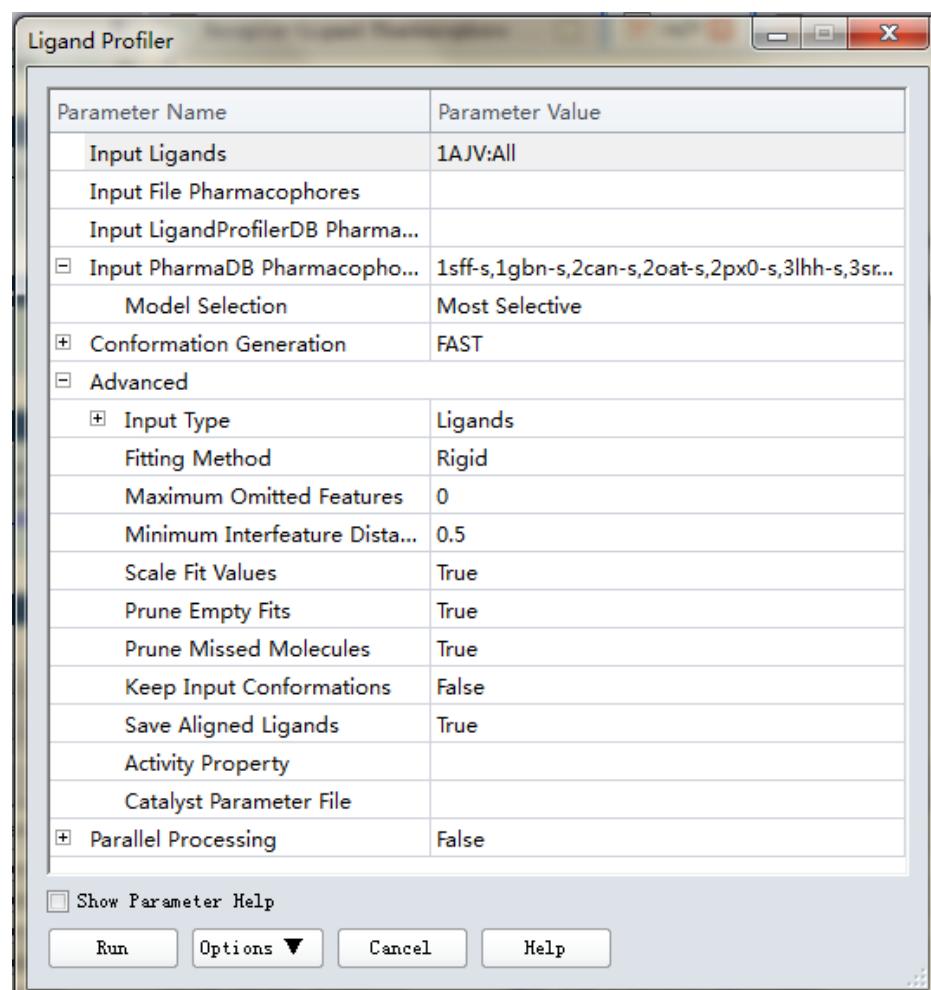


图 11

点击 **Background** 等待作业运行。

作业完成后，展开作业浏览器（Jobs Explorer）中该任务并点击 **Report** 链接，在 **Html** 窗口中打开 **Report** 页面。

## 2. 结果分析

在 Report 页面中点击 **Fits Organized by Molecule** 链接，查看 1AJV 与每个药效团模型匹配的 FitValue 值，所有药效团模型按 fitvalue 值由高到低排列。

FitValue 值越高，说明 1AJV 与相应药效团匹配的越好，药效团对应的靶标越有可能是 1AJV 作用的靶标。

Pharmacophores that fit each molecule		
Name	Pharmacophore	FitValue
1ajv	1fkh-01-s	0.913946
1ajv	1ajv-01-s	0.82808
1ajv	1hwr-01-s	0.771379
1ajv	1ajx-01-s	0.721309
1ajv	2qhz-01-s	0.623489
1ajv	2qnv-01-s	0.611275
1ajv	2zc9-01-s	0.599228
1ajv	2zdv-01-s	0.53702
1ajv	3si4-01-s	0.536889
1ajv	2cmf-01-s	0.535097
1ajv	4g50-01-s	0.533567
1ajv	2zf0-01-s	0.526377

图 12

右键返回，点击 **Pharmacophore properties** 链接，查看药效团模型对应的蛋白靶标的信息。

pharmacophore.Pharma-ID	pharmacophore/metadata/Gene-Name	pharmacophore/metadata/Uniprot-AC	pharmacophore/metadata/Kegg-Identifier	pharmacophore
1ajv	POL_HV1B1	P03366	KO_id not found	Enzymes
1ajx	POL_HV1B1	P03366	KO_id not found	Enzymes
1b3d	MMP3_HUMAN	P08254	K01394	Enzymes
1fkh	FKB1A_HUMAN	P62942	K09568	Enzymes
1hwr	POL_HV1H2	P04585	KO_id not found	Enzymes
1jla	POL_HV1H2	P04585	KO_id not found	Enzymes
1kqu	PA2GA_HUMAN	P14555	K01047	Enzymes
1m13	NR1I2_HUMAN	O75469	K08540	Transcription fact
1nlu	PICP_PSER	P42790	KO_id not found	Others
2a0c	CDK2_HUMAN	P24941	K02206	Enzymes
2cmf	ACES_TORCA	P04058	K01049	Enzymes
2q15	BACE1_HUMAN	P56817	K04521	Enzymes
2zdv	CDK2_HUMAN	P24941	K02206	Enzymes

图 13

点击 **View Results**，可以查看 1AJV 与各个药效团匹配的 FitValues 值匹配的热图。

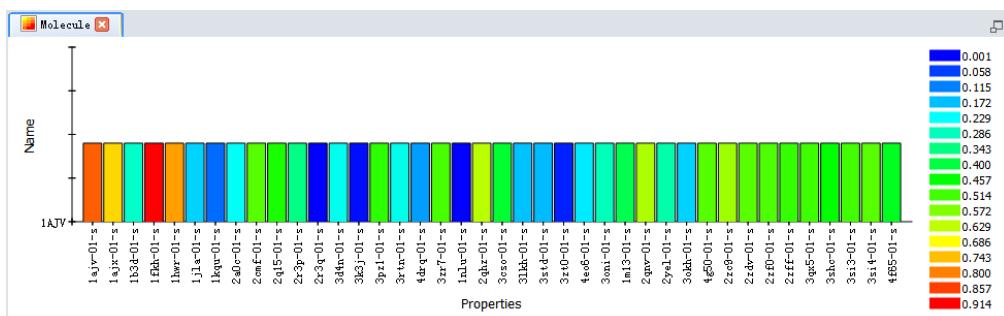


图 14

# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

**QSAR/SAR 专题:**

**2D-QSAR、3D-QSAR、MMP**

## 构建二维定量构效关系模型（2D-QSAR）教程

**目的：**通过此教程，了解并掌握 Discovery Studio 中通过 MLR 和 PLS 方法构建二维定量构效关系模型的操作过程。

**所需功能和模块：**Discovery Studio Client, DS QSAR。

**所需数据文件：**2D\_trainingset.sd, 2D\_test.sd。

**所需时间：**10 分钟

### 介绍

药物设计中，当受体的结构未知时，基于结构的药物设计方法（主要为分子对接）将无计可施。而QSAR方法是基于小分子配体的药物设计方法，目的是采用数理统计方法研究和揭示化合物活性与其分子结构或物理化学特征之间的定量变化规律。如果能够搜集到一系列结构类似物的生物活性数据，则可以采用QSAR（定量构效关系）的方法预测未知化合物的相关活性。

DS-QSAR 中包括以下几种产生模型的方式：

- ◆ 贝叶斯分类（Bayesian categorization）
- ◆ 遗传函数逼近（Genetic Function Approximation, GFA）
- ◆ 多重线性回归（Multiple Linear Regression, MLR）
- ◆ 偏最小二乘法（Partial Least Squares, PLS）
- ◆ 递归分区（Recursive Partitioning, RP）

本教程采用2D-QSAR的方法研究甘氨酸/N-甲基-D-天冬氨酸抗肿瘤抑制剂的二维定量构效关系。数据来源为如下的参考文献：2-Dimensional Quantitative Structure-Activity Relationship Modeling Study of Glycine/N-methyl-D-aspartate Antagonist Inhibition: Genetic Function Approximation Vis-à-vis Multiple Linear Regression Methods, Nitin S. Sapre, Nilanjana Pancholi, Swagata Gupta and Arun Sikarwar, *Acta Chim. Slov.* 2007, 54, 797–804。

参考文献中，作者采用GFA和MLR两种线性回归的方法构建构效关系模型。化合物的生物活性定义为LogIC<sub>50</sub>，建立构效关系时采用的理化参数为：Wiener指数（W），Randic接合系数（<sup>1</sup>X<sub>R</sub>），Balaban指数（J）。

本教程为了使用户对QSAR的方法流程有全面的了解，不采用文章中所使用的理化性质作为构建构效关系模型的参数（用户若有兴趣可以自己重复作者的QSAR分析），而采用的是DS中计算出的化合物2D性质。构建构效关系方程时，采用MLR/PLS的方法。

本教程包括以下内容：

- ◆ 构建化合物(包括训练集和测试集)的3D空间构象并进行能量优化
- ◆ 输入化合物相关的活性数据
- ◆ 计算训练集中化合物的相关性质
- ◆ 利用MLR构建QSAR模型
- ◆ 利用PLS构建QSAR模型
- ◆ 利用构建的QSAR模型预测测试集中化合物的生物活性

**注：**其中第1、2部分为数据准备部分，可选作。本教程已将前两步的数据和结果准备好。用户可以直接从第3步开始本教程的学习。

## (选作) 构建化合物的 3D 空间构象并进行能量优化

构建QSAR模型之前，一般需要搜集相关化合物的生物活性及其理化性质。本教程所使用的化合物活性数据均来自上述文献。

构建QSAR模型过程中，如果计算的化合物物理化性质不涉及化合物的3D结构，用户也可以只建立化合物的2D结构。建议如果化合物个数不是很多的话，还是使用分子的3D结构来计算相关的理化性质。

### 1. 对构建的初始构象进行能量优化

给初始结构做能量优化，首先需要给化合物赋力场参数。

在文件浏览器（Files Explorer）中，找到并双击打开**Molecule.sdf**文件。（图1）

该结构是直接从画图软件中拷贝的结构。

在工具浏览器（Tools Explorer）中，展开**Simulation | Change Forcefield**，点击**Forcefield**选择**MMFF**力场，再点击**Apply Forcefield**。

该步将MMFF力场赋给化合物。化合物会根据力场的定义自动加氢。

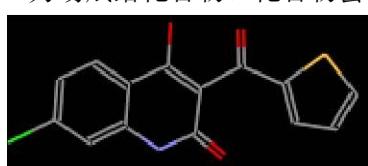


图1 优化前结构

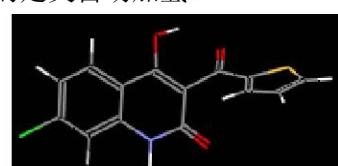


图2 优化后结构

在工具浏览器（Tools Explorer）中，展开**Simulation | Run Simulations**，点击**Minimization**打开Minimization对话框。

设置**Input Typed Molecule**为**Molecule:1**。

其余参数选择默认设置。（图3）

该步对小分子进行能量最小化，得到优化后的结构。（图2）

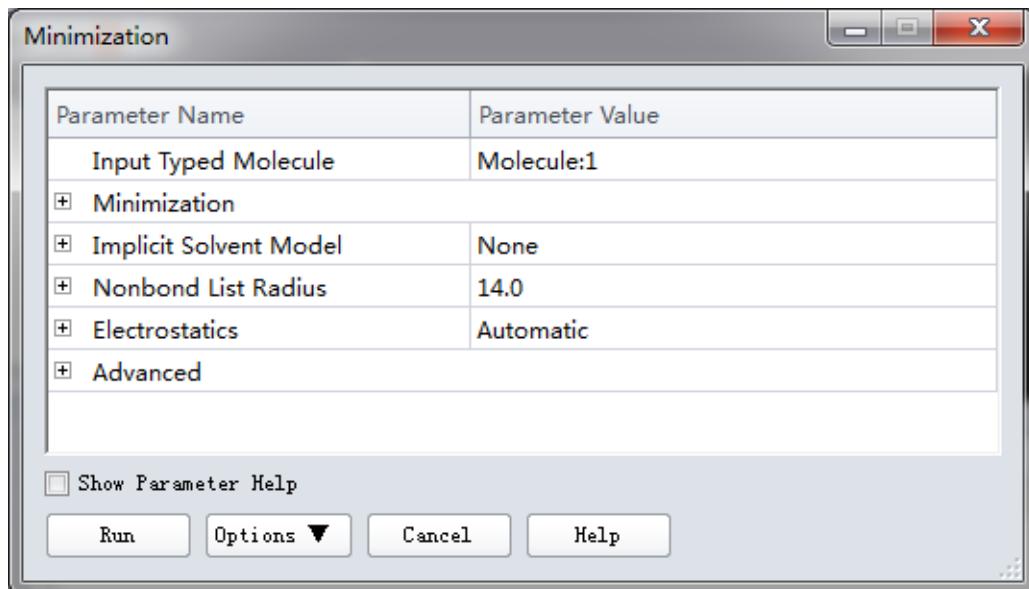


图3 “Minimization”对话框

依次对本次QSAR实验涵盖的所有化合物（包括训练集和测试集分子）进行能量优化，并将所有的结构放置于同一个3D窗口。也可以通过在工具浏览器（Tools Explorer）中，展开**Small Molecules | Minimize Ligands**，点击**Full Minimization**来批量对化合物进能量优化，如图4

所示。由于优化所有的结构需要消耗的时间比较多，本教程将使用已经提前优化好的结果作为后续研究的基础。提前优化好的训练集分子保存在2D\_training set.sd文件中。

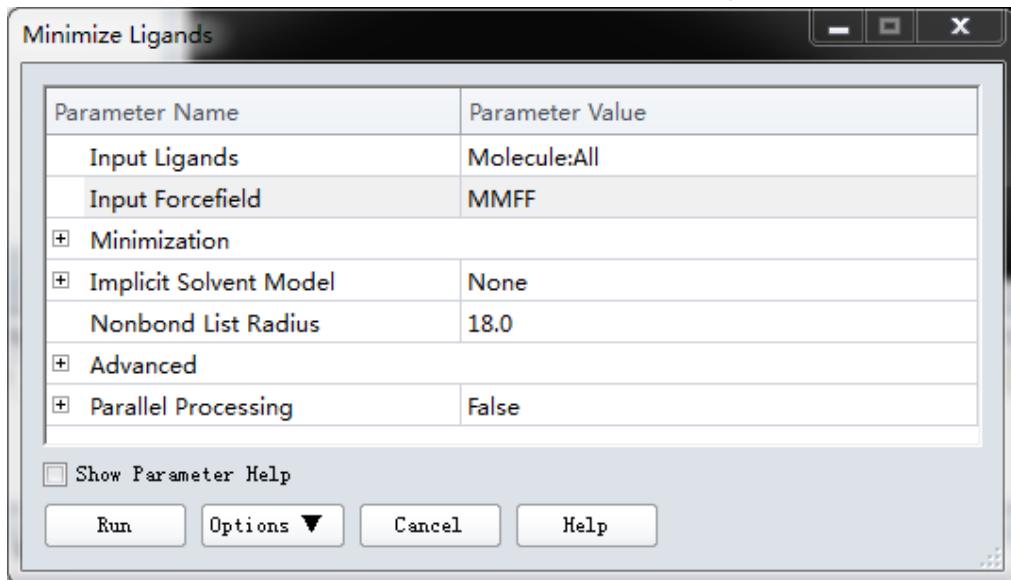


图4 “Full Minimization”对话框

### (选作) 输入相关化合物的数据

在文件浏览器（Files Explorer）中，找到并双击打开2D\_training set.sd文件。

DS将打开一个新的分子窗口，训练集中所有的分子都在一个表格浏览器（Tables Explorer）中显示。

	Index	Name	Visible	Tagged	Visibility Locked	Forcefield	ForcefieldBase	Cycle	CHARMin
1	1	1	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	54	20.6276
2	2	2	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	23.2743
3	3	3	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	23.6189
4	4	4	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	51.668
5	5	5	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	72	37.8266
6	6	6	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	168	52.9195
7	7	7	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	141	48.6857
8	8	8	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	155	43.9758
9	9	9	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	189	52.0223
10	10	10	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	71.8384
11	11	11	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	74.2041
12	12	12	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	74.8647
13	13	13	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	81.0612
14	14	14	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	181	73.9015
15	15	15	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	21.5682
16	16	16	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	21.5769
17	17	17	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	90.5354
18	18	18	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	MMFF	MMFF	200	80.0761

图5 2D\_training set.sd 列表窗口

下面需要添加化合物的活性数据。

在表格浏览器（Tables Explorer）中点击鼠标右键，选择Add Attribute。

弹出一个设置化合物新性质的对话框，将name设置为LogIC50，点击OK。



图6 AddAttribute 对话框

在表格浏览器（Tables Explorer）中会添加一列关于化合物活性数据LogIC50的数据。可根据文献报道的活性数据依次添加，已提前准备好，可直接进行下一步。

### 计算训练集中化合物的相关性质

由于本教程不使用文献报道的理化参数 ( $W$ ,  $^{1}X_R$ ,  $J$ ) 作为构建QSAR模型的参数，所以需要使用DS来计算训练集化合物的其他理化性质。

在工具浏览器（Tools Explorer）中，展开**Small Molecules | Calculate Molecular Properties**，点击**Calculate Molecular Properties**，打开Calculate Molecular Properties对话框。（图7）  
设置**Input Ligands**为**2D\_training set: All**。

点击**Molecular Properties**右边的 按钮，打开Molecular Properties对话框（图7），选择默认的**2D 性质：**

**ALogP, Molecular\_Weight, Num\_H\_Donors, Num\_H\_Acceptors, Num\_RotatableBonds, Num\_Rings, Num\_AromaticRings, Molecular\_FractionalPolarSurfaceArea.**

点击**Run** 运行该作业。

等待作业完成。

待作业完成以后，选取的2D性质计算结果自动添加于**2D\_training\_set**表单中。

在表格浏览器（Tables Explorer）中，点击鼠标右键，选择**Remove Attributes**，只保留**LogIC50**和用于计算的2D性质。（图8）

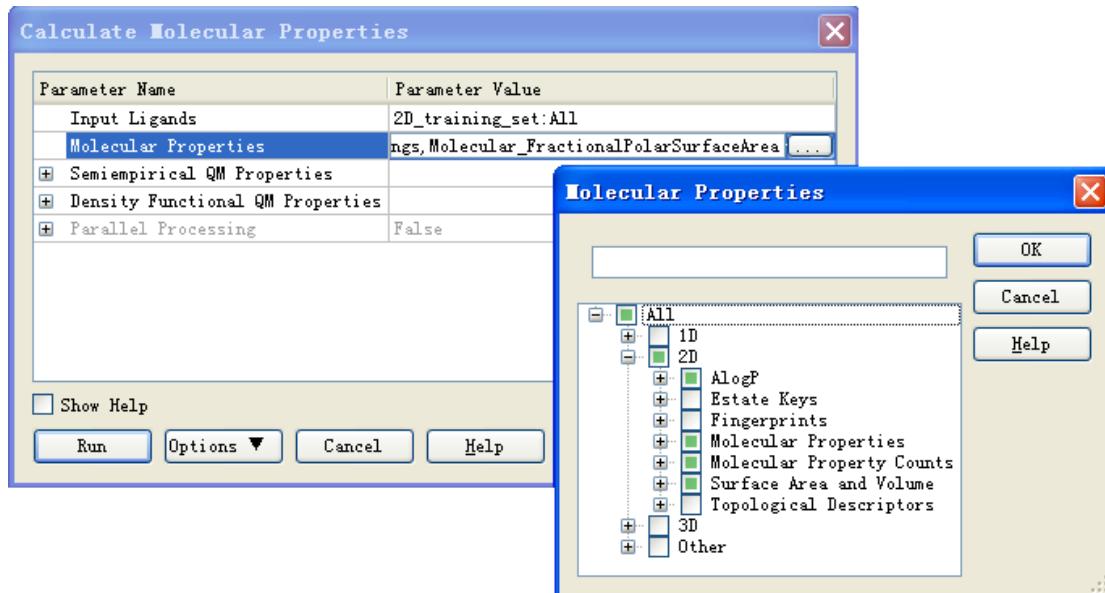


图7 “Calculate Molecular Properties”和“Molecular Properties”参数设置

	Index	Name	Visible	Tagged	Visibility Locked	logIC50	AlogP	Molecular_Weight	Num_RotatableBonds	Num_Rings
1	1	1	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	3.037	2.697	305.736	2	3
2	2	2	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	3.182	3.155	340.181	2	3
3	3	3	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.98	3.265	384.632	2	3
4	4	4	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.236	2.679	251.28	1	3
5	5	5	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.547	2.519	277.726	1	3
6	6	6	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.624	2.841	301.724	2	3
7	7	7	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.592	3.344	285.725	1	3
8	8	8	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.818	3.063	289.689	1	3
9	9	9	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.31	2.841	301.724	2	3
10	10	10	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.301	4.417	363.794	3	4
11	11	11	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	1.063	4.417	363.794	3	4
12	12	12	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	1.086	4.424	377.82	4	4
13	13	13	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.29	5.36	391.847	4	4
14	14	14	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.602	4.832	361.821	3	4
15	15	15	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.682	2.398	275.687	1	3
16	16	16	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.873	2.252	261.661	1	3
17	17	17	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.556	4.262	375.804	3	4
18	18	18	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.653	4.816	391.847	4	4
19	19	19	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.892	4.904	377.82	3	4
20	20	20	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.982	4.25	376.836	4	4
21	21	21	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.903	4.816	391.847	4	4

图8 计算后增加的2D性质

## 利用MLR建立QSAR模型

采用多元线性回归（MLR）方法建立回归方程。

在工具浏览器（Tools Explorer）中，展开**Small Molecules | Create QSAR Model**，点击**Create Multiple Linear Regression Model**。打开Create Multiple Linear Regression Model对话框。

设置**Input Ligands**为**2D\_training\_set:All**。

点击**Dependent Property**右边的栅格，下拉列表中选择**LogIC50**。

具体参数设置如图9。

注：Calculable Properties不选，在User Properties中选择已经计算好的性质。若事先没有计算过化合物的2D性质，则在Calculable Properties栏中选择需要计算的性质。

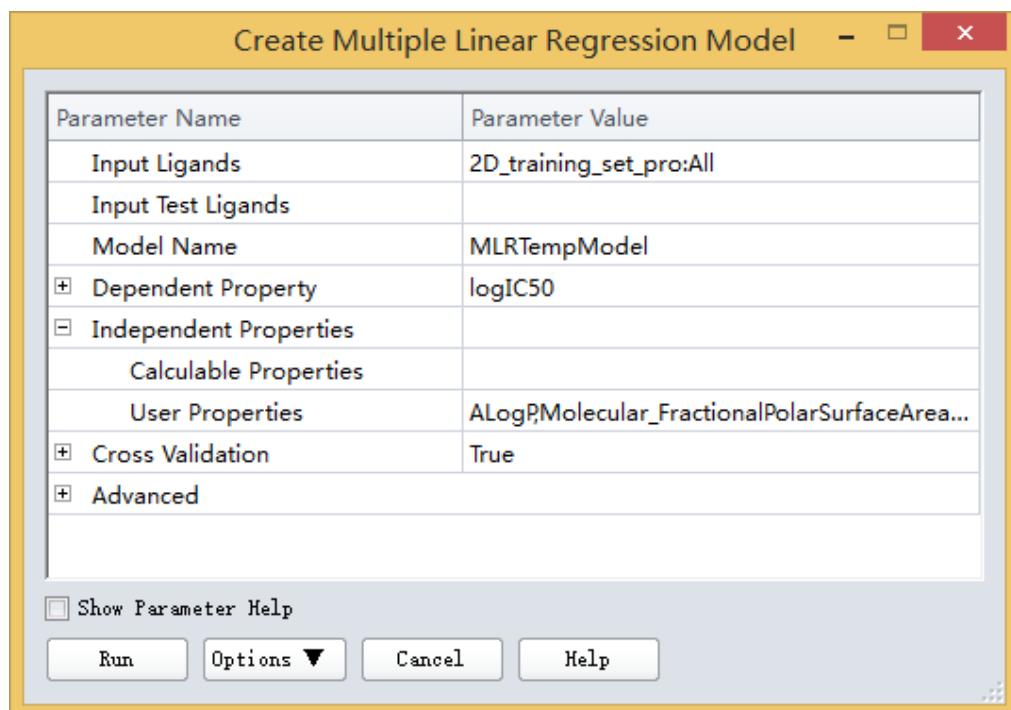


图9 “Create Multiple Linear Regression Model”参数设置

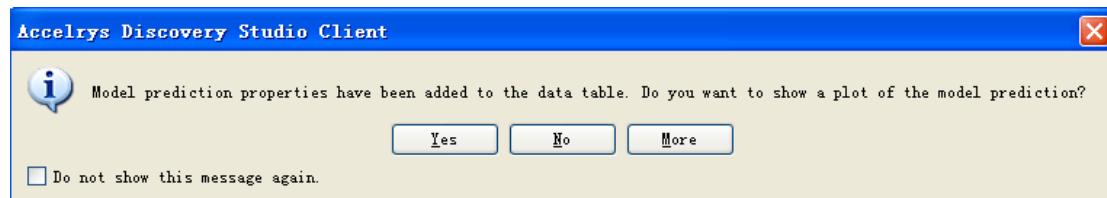
点击 **Run** 运行该作业。点击 **Background** 将作业设置为后台运行。

等待作业完成。

待作业完成以后，在任务浏览器（Jobs Explorer）中双击相应的行，打开Report结果文件。  
(图10)

点击 **View Results**。

打开一个新的分子窗口，包含了训练集分子的表单，同时弹出一个消息盒子（如下图）。



点击 **Yes**，打开一个 **plot** 窗口。（图 11）

横坐标为训练集分子的实验活性值，纵坐标为预测值， $R^2=0.654$ ，该值越接近 1，预测结果越准确。

该结果为MLR的线性回归方程。

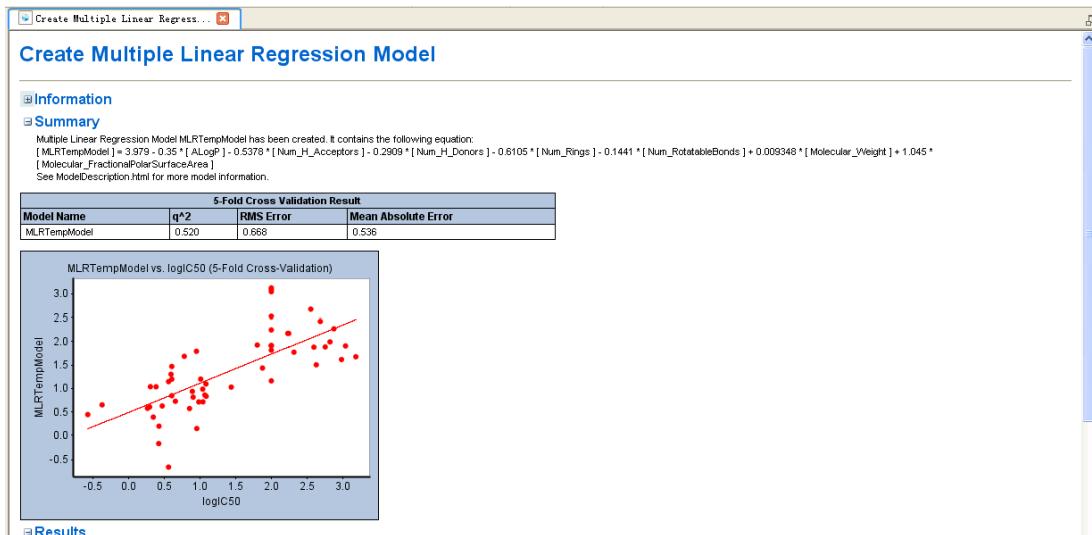


图10 结果文件

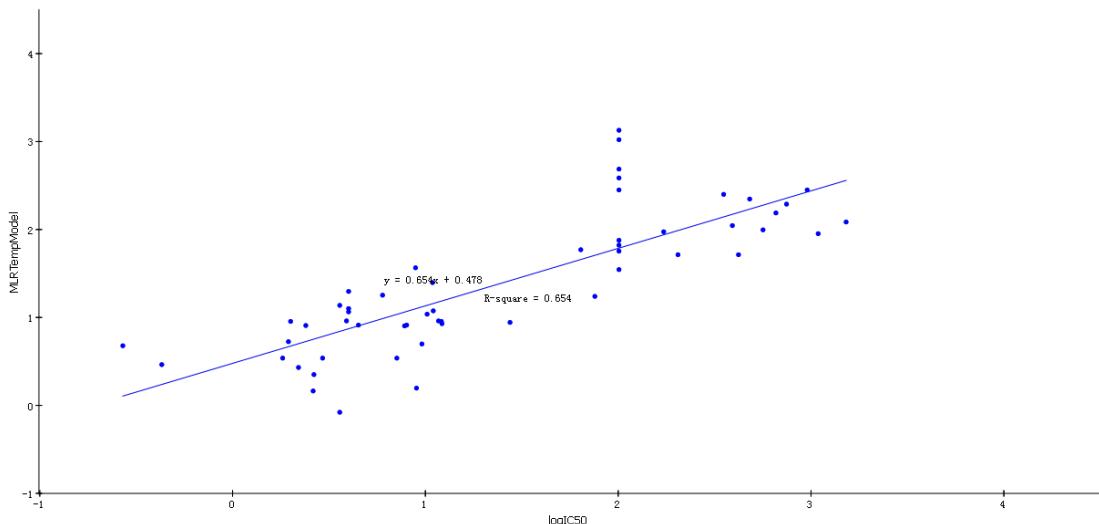


图11 训练集分子的“MLRTempModel vs. logIC50”plot

## 利用PLS建立QSAR模型

采用偏最小二乘（PLS）方法建立回归方程。

在工具浏览器（Tools Explorer）中，展开**Small Molecules | Create QSAR Model**，点击**Create Partial Least Squares Model**。打开Create Partial Least Squares Model对话框。

设置**Input Ligands**为**2D\_training\_set>All**。

点击**Dependent Property**右边的栅格，下拉列表中选择**LogIC50**。

具体参数设置如图12。

注：Calculable Properties不选，在User Properties中选择已经计算好的性质。若事先没有计算过化合物的2D性质，则在Calculable Properties栏中选择需要计算的性质。

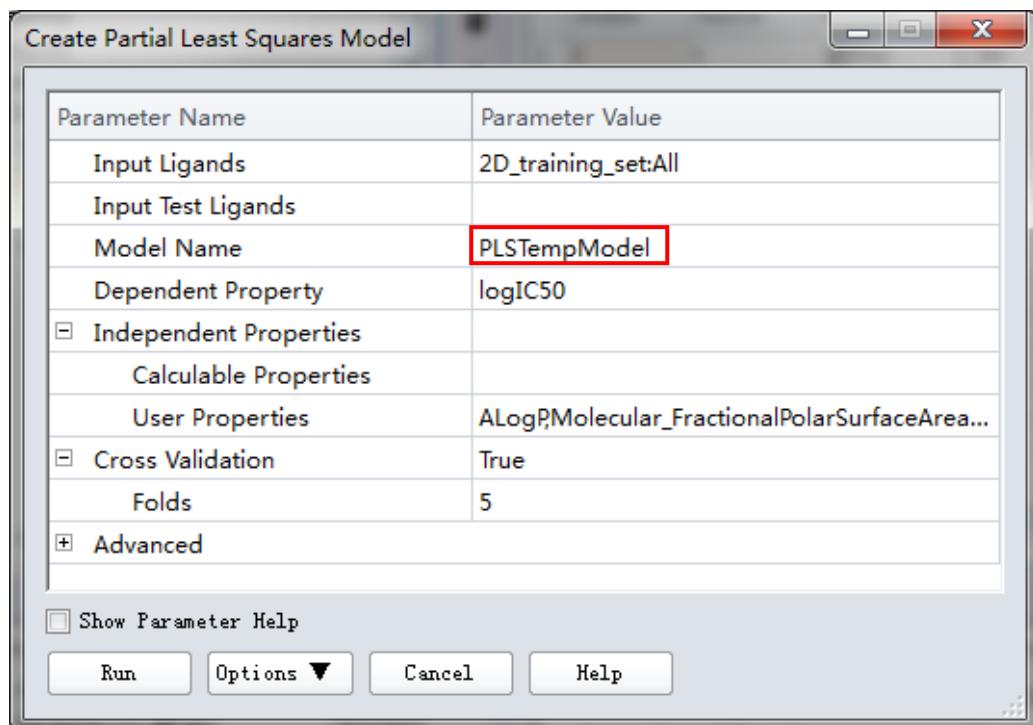
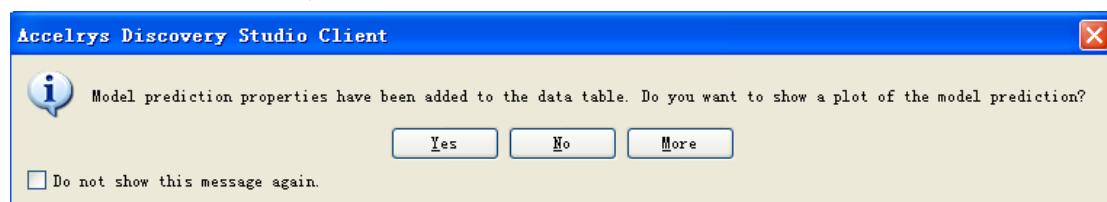


图12 “Create Partial Least Squares Model”参数设置

点击 **Run** 运行该作业。

等待作业完成。

待作业完成以后，自动弹出一个消息盒子（如下图）。



点击 **Yes**，打开一个 plot 窗口。（图 13）

横坐标为训练集分子的实验活性值，纵坐标为预测值， $R^2=0.640$ 。

该结果为PLS的线性回归方程。

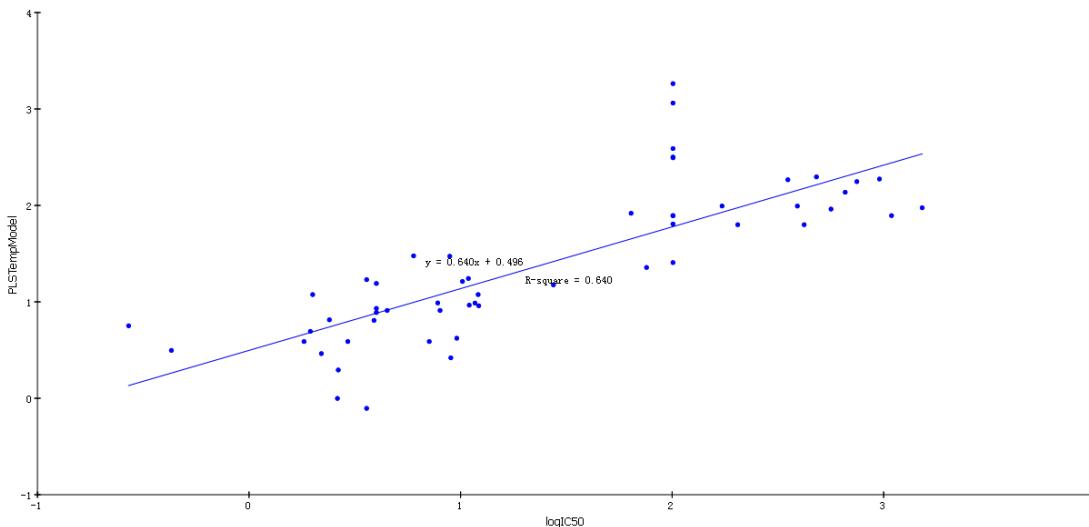


图13 训练集分子的“PLSTempModel vs. logIC50”plot

### 利用MLR\_QSAR/PLS\_QSAR模型预测测试集的生物活性

在文件浏览器（Files Explorer）中，找到并双击打开**2D\_test set.sdf**文件。

打开一个新的分子窗口，共有10个待测试的化合物显示在表格浏览器中，其实验活性值LogIC50值已人工输入。（图14）

	Index	Name	Visible	Tagged	Visibility Locked	logIC50	Forcefield	ForcefieldBase	Step-
1	1	T1	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.585	MMFF	MMFF	0.02
2	2	T2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2.004	MMFF	MMFF	0.02
3	3	T3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3.348	MMFF	MMFF	0.02
4	4	T4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.561	MMFF	MMFF	0.02
5	5	T5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.535	MMFF	MMFF	0.02
6	6	T6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	3.477	MMFF	MMFF	0.02
7	7	T7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.446	MMFF	MMFF	0.02
8	8	T8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.63	MMFF	MMFF	0.02
9	9	T9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	1.513	MMFF	MMFF	0.02
10	10	T10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2.913	MMFF	MMFF	0.02

图14 2D\_testset分子窗口

在工具浏览器（Tools Explorer）中，展开**Small Molecules | Calculate Molecular Properties**，点击**Calculate Molecular Properties**，打开Calculate Molecular Properties对话框。（图15）设置**Input Ligands**为**2D\_test: All**。

点击**Molecular Properties**右边的 按钮，打开Molecular Properties对话框，选择**Other**下的

**MLRTempModel**、**PLSTempModel**性质（图15）

点击**Run**运行该作业。

等待作业完成。

待作业完成以后，选取的2D性质计算结果自动添加于2D\_Test表单中。（图16）

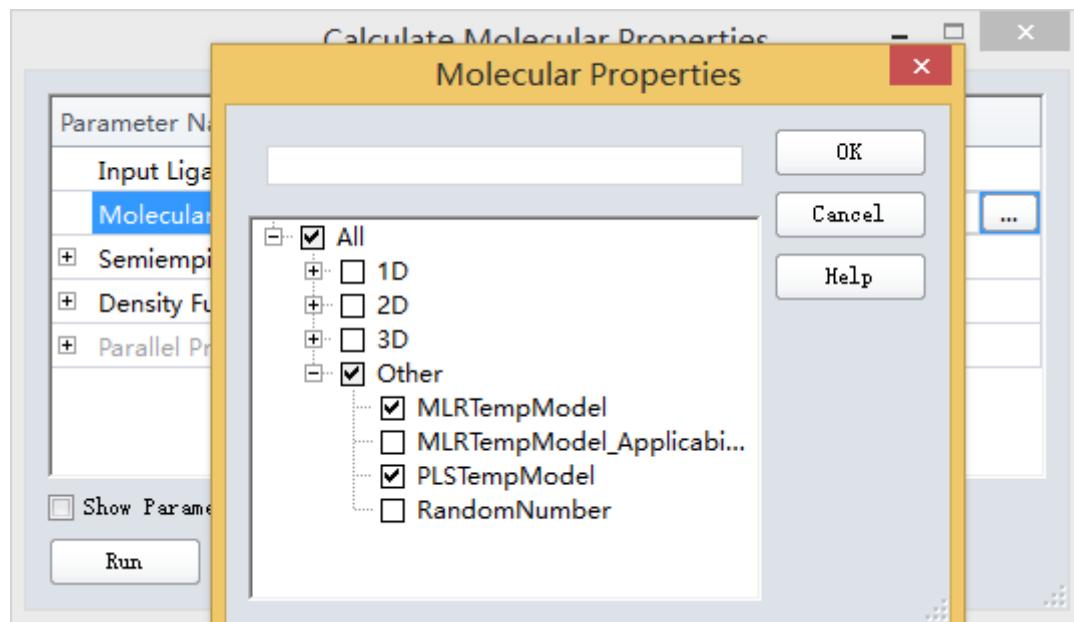


图15 “Calculate Molecular Properties”和“Molecular Properties”参数设置

2D_test_set (1)												
	Trial Energy	Cycle	CMAPs	Van der Waals Energy	RMS Gradient	Bond Energy	logIC50	W	J	XR	MLRTempModel	PLSTempModel
1	4	200	0	53.7734	0.35572	7.07898	1.585	1,930	1.***	11.9***	1.67748	1.70333
2		200	0	53.5223	0.37816	4.258	2.004	1,456	1.***	11.5***	1.36931	1.34119
3	4	177	0	44.1838	0.09386	4.08625	3.348	1,261	1.***	11.5***	1.43486	1.56276
4	3	95	0	44.6341	0.08979	6.08144	1.561	1,568	1.***	11.9***	1.08626	0.600864
5	5	200	0	55.4031	0.34544	5.31404	1.535	1,956	1.***	13.1***	1.33383	1.21391
6	7	200	0	31.1012	0.52048	2.59465	3.477	1,275	1.***	11.0***	1.33861	1.24705
7	8	200	0	72.9044	0.20282	8.48681	1.446	2,509	1.534	14.5***	0.813365	0.860851
8	8	200	0	68.6923	0.38472	6.81368	1.63	2,574	1.675	14.3***	0.575424	0.626252
9	2	200	0	54.6892	0.54361	5.65263	1.513	2,078	1.***	13.4***	0.509746	0.55534
10	8	101	0	32.4485	0.09112	2.96595	2.913	644	2.171	9.0754	2.27489	2.26629

图16 计算后增加的MLRTempModel性质

MLRTempModel 表示由 MLR-QSAR 回归方程预测得到的测试集化合物的 LogIC50 值。

PLSTempModel 表示由 PLS-QSAR 回归方程预测得到的测试集化合物的 LogIC50 值。

## 采用能量格点作为描述符构建 PLS 模型（3D-QSAR）教程

**目的：**通过此教程，了解并掌握 Discovery Studio 中构建三维定量构效关系模型的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS QSAR。

**所需数据文件：**trainingset.sd, testset.sd。

**所需时间：**20 分钟

### 介绍

药物设计即试图发现能够同生物大分子靶标在形状（steric）和电荷（静电势）上互补的小分子。

与 2D-QSAR 相比，3D-QSAR 方法更能间接反映配体小分子和蛋白大分子之间的非键相互作用特征，具有更加丰富的物理化学内涵，因此得到了迅速的发展和广泛的应用。

3D-QSAR 模型时基于小分子的立体（steric）和静电（electrostatic）场构建的回归模型，可以用于预测未知配体小分子的活性及观察受体-配体间有利和不利的相互作用。本教程利用能量格点作为描述符构建了一个偏最小二乘（PLS）模型。该能量格点是通过两种用于测量静电势和立体效应的探针计算得到的。

本教程包括以下步骤：

- ◆ 构建 3D-QSAR 模型
- ◆ 观察分析结果
- ◆ 基于 3D-QSAR 模型预测活性

### 3D QSAR 模型的构建

在构建 3D QSAR 模型之前，需要对训练集分子进行叠合，叠合好坏决定了最终模型的可信度。叠合方式一般有以下几种：

- ◆ 如果配体分子都来源于晶体结构且都与同一靶标相结合，则可以直接使用晶体结合构象
- ◆ 如果有配体小分子的药效团模型，则可将配体小分子匹配至药效团模型以实现对配体小分子的叠合
- ◆ 可直接基于配体小分子的公共骨架进行叠合
- ◆ 可将配体小分子公共骨架中的药效团特征元素加以提取，用于构建药效团模型，再通过配体药效团的匹配流程（Ligand Pharmacophore Mapping）进行小分子的叠合
- ◆ 可基于立体场和静电场通过场匹配的方式进行小分子的叠合（Structure | Superimpose | Molecular Overlay...）

#### 1. 3D-QSAR 的构建

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | QSAR，双击 trainingset.sd。

在分子窗口中打开一个表格，共 14 行，即 14 个训练集分子，该分子事先已进行叠合。

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | QSAR，双击 testset.sd。

同样在分子窗口中打开 8 个已叠合的测试集分子。

在工具浏览器（Tools Explorer）中，展开 Small Molecules | Create QSAR Model，点击 Create 3D QSAR，打开 Create 3D QSAR Model 对话框。

设置 Input Ligands 为 trainingset:All。

设置 **Input Test Ligands** 为 **testset:All**。

设置 **Activity Property** 为 **arNCTRlogRBA**。

在 **Model Name** 栏中输入 **GridBasedModel**。

其余参数保留默认设置。(图 1)

点击 **Run** 运行该作业。

等待作业完成。该作业需要运行 1min (2.4GHz 的 CPU)。

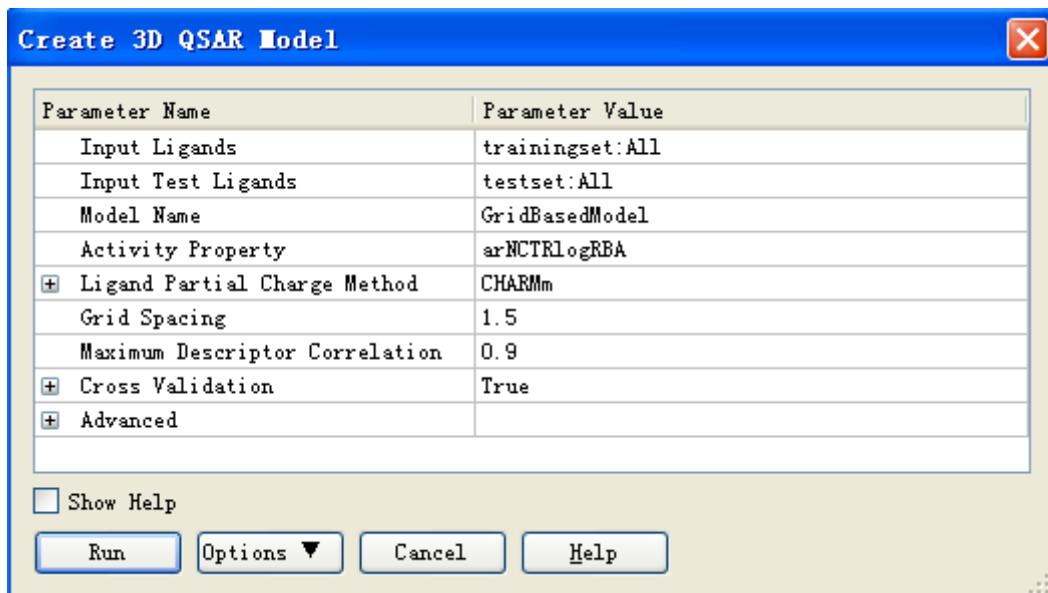


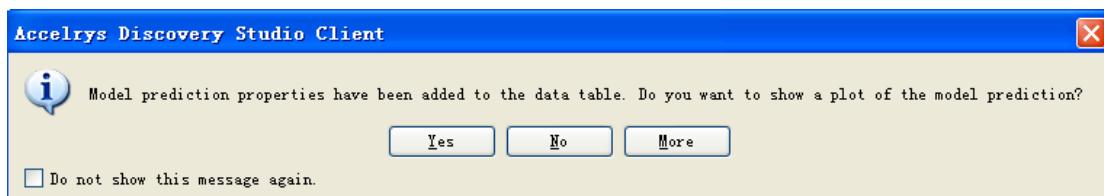
图 1 “Create 3D QSAR Model”流程参数设置

## 2. 观察结果

待作业完成以后，自动打开一个 Html 窗口，该窗口包含了一个 Report 报告文件。

在 Report 文件的 Results 部分，点击 **View Results**。

打开一个新的分子窗口，包含了训练集分子的表单，同时弹出一个消息盒子（如下图）。



点击 Yes，打开一个 plot 窗口。(图 2)

横坐标为训练集分子的实验活性值，纵坐标为预测值， $R^2=0.820$ ，该值越接近 1，预测结果越准确。

为了验证模型的预测能力，设置测试集进行外部验证。

在 Report 窗口中，点击 **View Test Results** 链接。

打开一个 plot 窗口。(图 3)

横坐标为测试集分子的实验活性值，纵坐标为预测值， $R^2=0.485$ ，表明该模型具有一定的可信度。

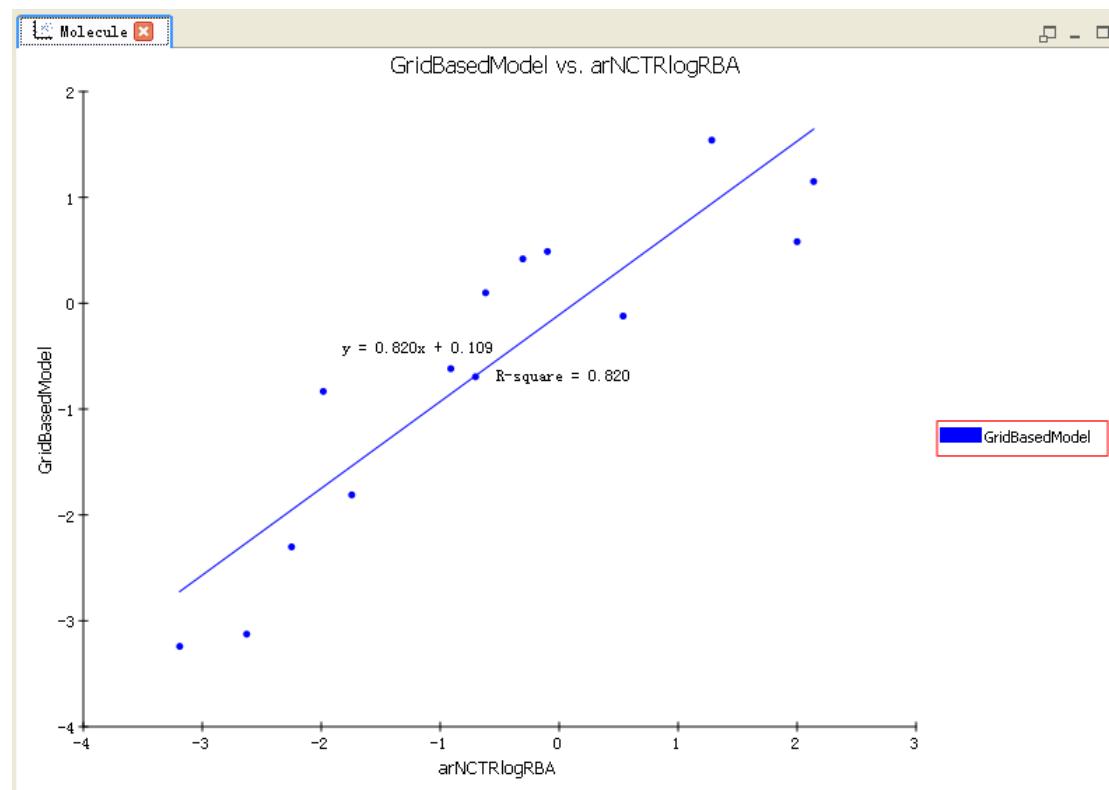


图 2 训练集分子的“GridBasedModel vs. arNCTRlogRBA”plot

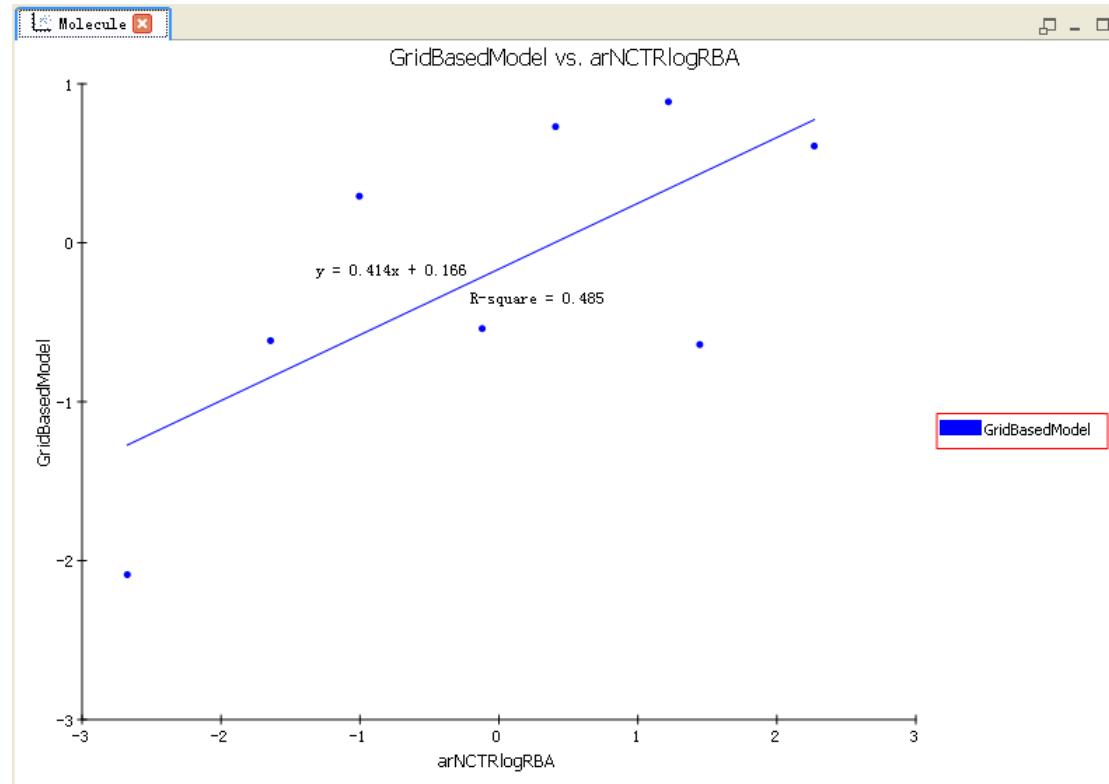


图 3 测试集分子的“GridBasedModel vs. arNCTRlogRBA”plot

点击 Results 部分的 Model Description 链接可以查看更多有关构建的 QSAR 模型的信息。该信息共包含以下五个部分：

- ◆ 回归统计参数
- ◆ 模型的相关系数和相应变量
- ◆ 训练集信息
- ◆ 排除变量信息
- ◆ 模型构建参数

3D QSAR 模型除了预测能力，其相应的 3D 格点图也非常有用。格点包含两部分内容：一部分表征静电势，另一部分是表征范德华作用。

在 Report 中，点击 **View Training Set Aligned to Model (EP)**。

打开一个分子窗口，视图窗口中显示了训练集分子匹配至 3D QSAR 模型的静电场系数等高图。(图 4)

红色区域表示该区域取代基的负电性越强越有利于提高化合物的活性；蓝色区域表示该区域取代基的正电性越强越有利于提高化合物的活性。

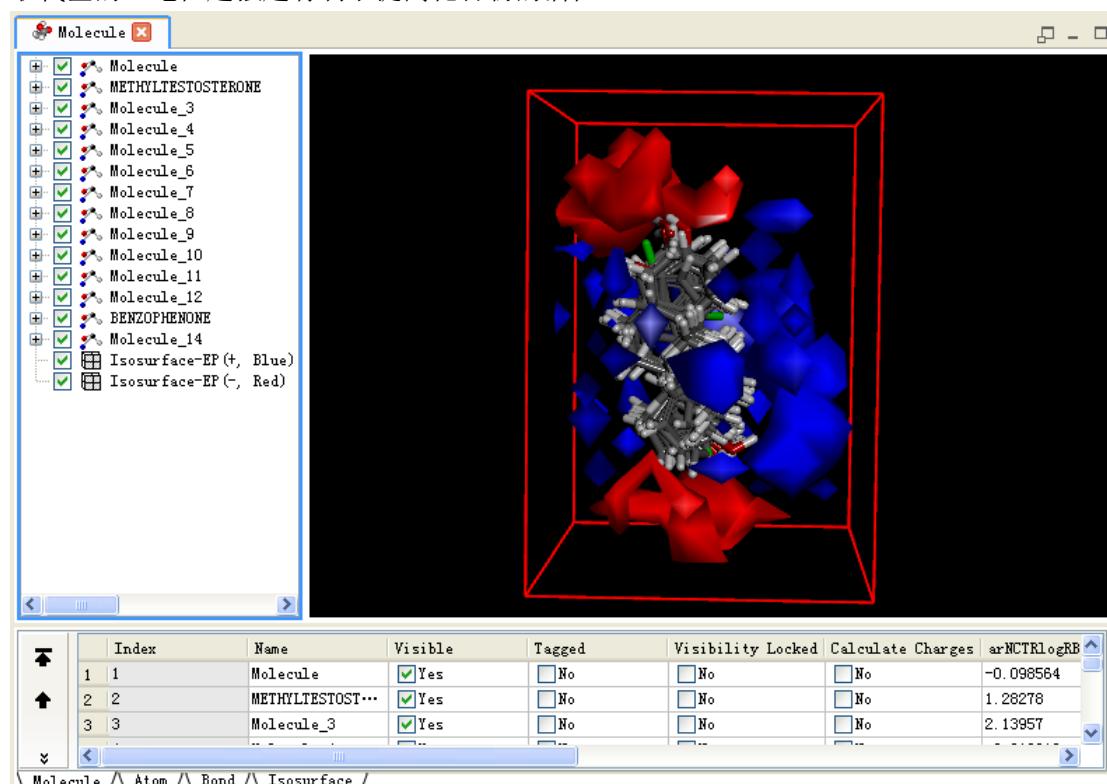


图 4 3D QSAR 模型的静电场系数等高图

在 Report 中，点击 **View Training Set Aligned to Model (VDW)**。

打开一个分子窗口，视图窗口中显示了训练集分子匹配至 3D QSAR 模型的立体场系数等高图。(图 4)

黄色区域表示该区域取代基体积的增大不利于提高化合物的活性；蓝色区域表示该区域取代基体积的增大有利于提高化合物的活性。

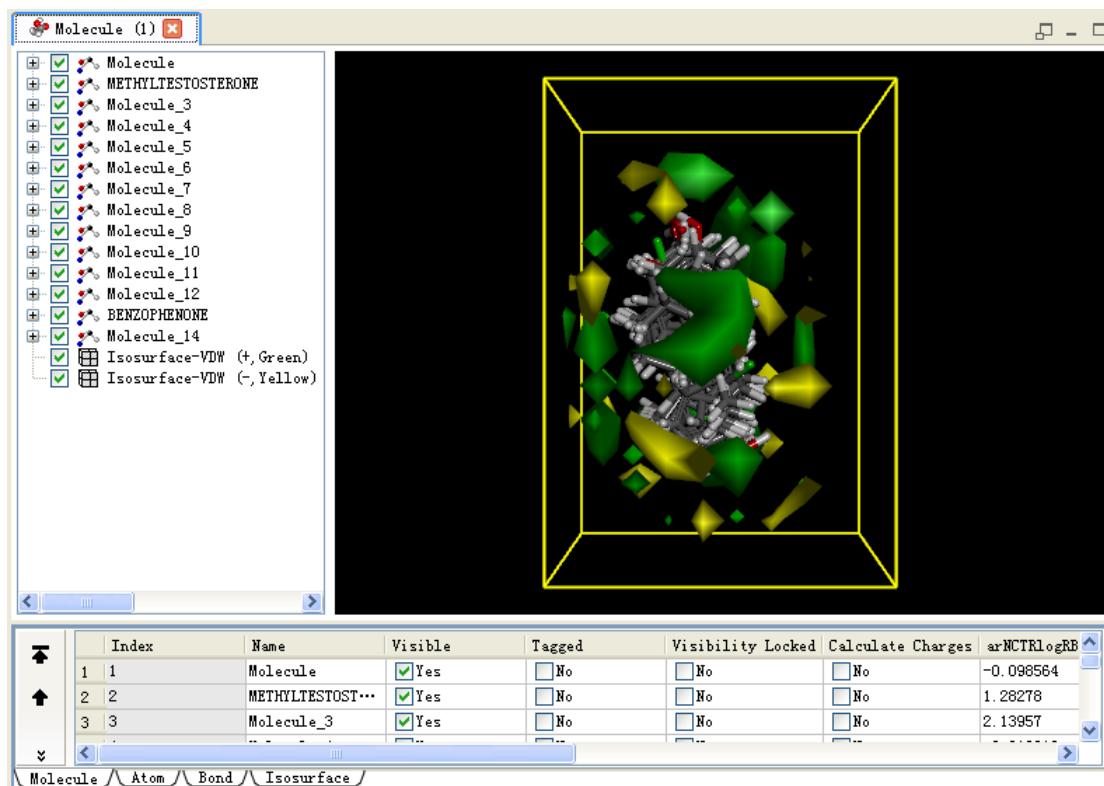


图 5 3D QSAR 模型的立体场系数等高图

## 基于 3D QSAR 模型预测化合物的活性

构建得到的 3D QSAR 模型可用于预测新化合物的活性。

关闭所有窗口。

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | QSAR，点击 testset.sd。

打开一个新的分子窗口。

在工具浏览器（Tools Explorer）中，展开 Small Molecules | Calculate Molecular Properties，

点击 Calculate Molecular Properties，打开 Calculate Molecular Properties 对话框。

设置 Input Ligands 为 testset>All。

点击 Molecular Properties 右边的 按钮，打开 Molecular Properties 对话框。

在对话框中，先选择再取消 All 左边复选框的选取，从而不选择任一选项。展开 Other，选取 GridBasedModel。（图 6）

点击 OK。

点击 Run 运行该作业。等待作业的完成。

该作业需要运行不到 1min 的时间（处理器：英特尔奔 4 2.4GHz，1GB 的内存）。

**注：**在运用模型预测活性之前，需要把新的小分子化合物叠合至先前的训练集分子。通常，该操作即将所有分子，包括训练集、测试集、新化合物，全部叠合在一起。

待作业完成后，弹出一个消息盒子（图 7），指示计算得到的分子性质已经添加到分子窗口中了。（图 8）

点击 OK。

图 8 中 GridBasedModel 列中的数值即基于 3D QSAR 模型预测得到的化合物的 arNCTRlogRBA 活性值。GridBasedModel\_Residual 一列即实验活性值同预测活性值之间的差值。

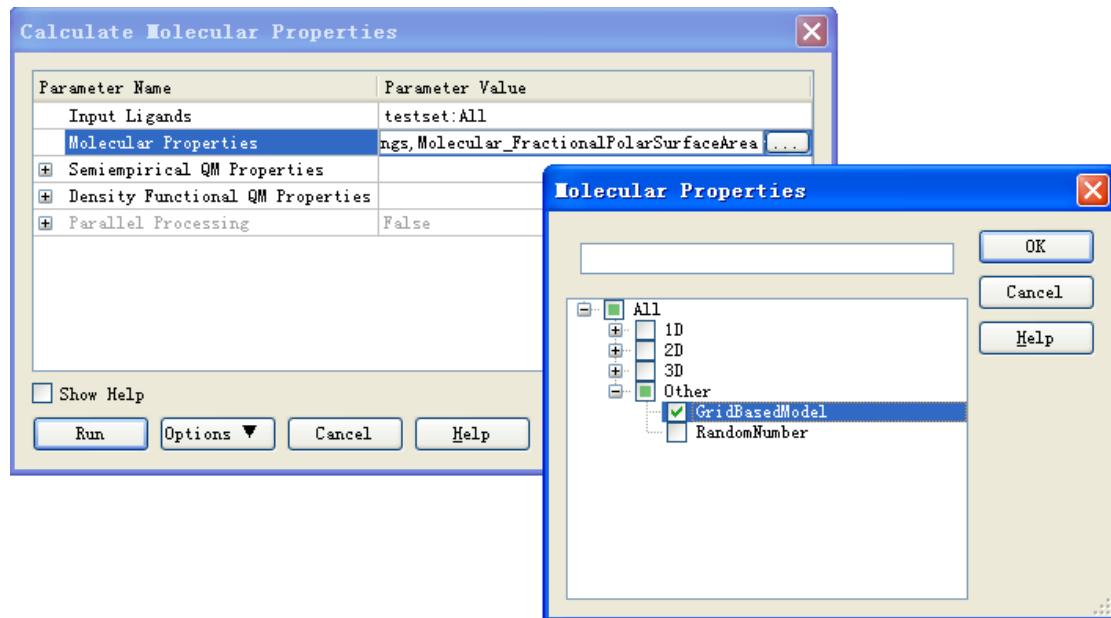


图 6 “Calculate Molecular Properties”参数设置



图 7 消息盒子

testset

	Index	Name	Visible	Tagged	Visibility Locked	arNCTRlogRBA	CAS	Structure_Category	KOWlogP	GridBasedTempModel
1	1	Molecule	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	-2.67447	611-99-4 DDT	2.19	-2.08816	
2	2	Molecule	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	-1.64344	1482-*** Steroid	1.21	-0.61604	
3	3	Molecule	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	-1.00446	481-30-1 Steroid	3.27	0.292866	
4	4	BETA-ES--	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	-0.120556	50-28-2 Steroid	3.94	-0.540296	
5	5	Molecule	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	0.407716	68-22-4 Steroid	2.99	0.730269	
6	6	D-(--)-W--	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	1.21973	797-63-7 Steroid	3.48	0.886806	
7	7	Molecule	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	1.4455	1225-*** steroid	5.12	-0.640568	
8	8	17ALPHA--	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	2.26941	3704-*** Steroid	3.69	0.608382	

图 8

## Discovery Studio MMPs 教程 1

### MMPs - Matched Molecular Pairs

所需功能和模块: Discovery Studio Client, DS QSAR.

所需数据文件: bindingdb\_cdk2.sd.

所需时间: 20 分钟

#### 介绍

构效关系 (SAR) 在药物化学中的应用是非常广泛的，主要分析化合物化学基团和生物效应之间的关系，然后对化合物分子进行化学修饰，从而提高化合物生化性能、选择性、药代动力学性质、生物利用度等。

MMPs (Matched Molecular Pairs) 分析是反向 SAR 技术，这种技术可以利用以前的实验数据快速获得常见性质，用来理解结构已知的化合物相关基团之间性质差异和从头设计结构新颖的化合物。MMPs 是一对分子，这对分子结构是不同的，但是仅仅一处是不同的。Activity cliffs 定义为：生物活性明显不同的 MMPs。

进行 SAR 研究时，生物活性对应生物学上相关化学空间的改变，化学空间可以比喻为一片地域的地形，其中 Activity cliffs 不连续的分布到这片地形中。Activity cliffs 对应着结构中很小的改变，可以比喻为一片地域中很小地形的改变，但可以引起活性或势能面的重要的变化。Activity cliffs 富含详细的 SAR 信息，但是用传统的 SAR 方法时，很难预测这种信息。

在本教程中，通过 CDK2 抑制剂的数据库的分析，使用 MMPs 方法来识别出 Activity cliffs。CDK2 是高度保守的丝氨酸-苏氨酸蛋白家族的一个成员，在真核细胞复制周期中起到调控的作用。CDK2 的缺失可以导致细胞增殖的中断，所以 CDK2 抑制剂可以给控制肿瘤细胞的生长提供一种有效的方法，是治疗肿瘤的有效的武器。

#### 使用 MMPs 识别 Activity cliffs

CDK2 配体集来自 BindingDB 数据库中 (<http://www.bindingdb.org/>)。CDK2 配体集有详细的 Ki 值，并且活性至少为 10mMol 以上。在本例中，Activity cliffs 被定义为活性差别至少 2 个数量级以上的 MMPs，以保证 Activity cliffs 的一员的活性保留在 1nMol 级别上。

打开文件 Samples | Tutorials | QSAR | bindingdb\_cdk2.sd。

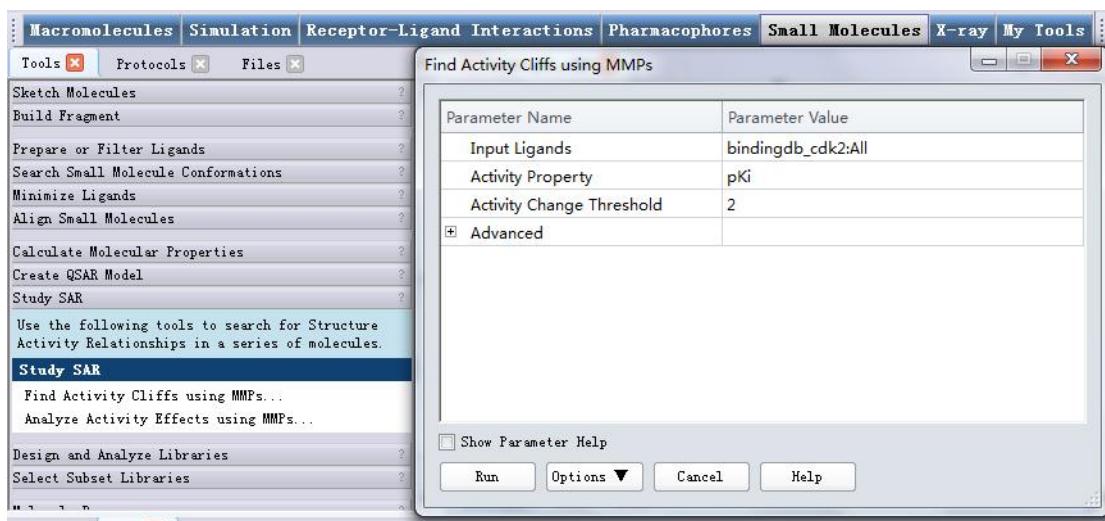
打开 Small Molecules | Study SAR 工具面板，点击打开 Find Activity Cliffs Using MMPs...。

Input Ligands 设置为 bindingdb\_cdk2:All。

Activity Property 设置为 pKi。

Activity Change Threshold 设置为 2。

点击 Run 运行。



作业运行时间 1min 左右，查看 Jobs Explorer 报告发现 27 对 Activity cliffs。Activity Change Threshold 用于测定是否 MMPs 是一对 Activity cliffs，如果活性变化是对数形式，活性变化的范围要求是两个数量级，那么阈值应该设置为 2。如果活性变化是线性形式，活性变化范围要求是两个数量级，那么阈值应该设置为 100。

在 Jobs Explorer 页面，点击 Summary 展开。

CDKs 配体集产生了 641 对 MMPs，仅仅 27 对是 Activity cliffs，剩余的 641 个被过滤掉。这 27 个活性 Cliffs 来自 20 个不同的核心结构。

在 Jobs Explorer 页面，点击 Results 展开，点击 View Results 按钮。

我们可以看到列表中包含 27 对 Activity cliffs，以及它们活性值的改变和每对 Activity cliffs 的核心结构。所有栏中，除了反应物和生成物栏，其它栏都可以隐藏。

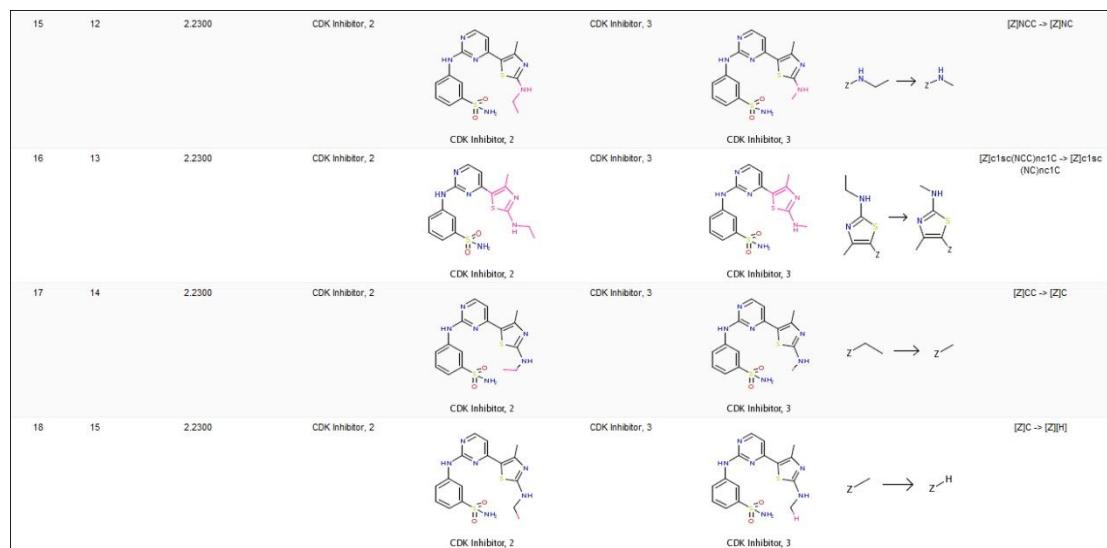
点击 Reactant Name 栏的右边的三角箭头，选择 Sort Ascending。

通过对反应物名称的升序排列，可以看到反应物一栏任何一个配体的所有转换，位于顶栏的 2,4-Diamino-5-keto-pyrimidine 23 被包含在 3 对 Activity cliffs 中。由于活性改变阈值总是正值，这就意味着在 3 对 Activity cliffs 中，2,4-Diamino-5-keto-pyrimidine 23 是活性低的伴侣。

Index	Core	Activity Change	Reactant Name ▾	Reactant	Product Name	Product	Reaction	Transformation
12	10	2.4200	2,4-Diamino-5-keto-pyrimidine 23		2,4-Diamino-5-keto-pyrimidine 29			[Z]C1CCNCC1 -> [Z] C1CCN(CC1)S(=O)(=O)C
13	11	2.0900	2,4-Diamino-5-keto-pyrimidine 23		2,4-Diamino-5-keto-pyrimidine 30			[Z][H] -> [Z]S(=O)(=O)CC
14	11	2.4200	2,4-Diamino-5-keto-pyrimidine 23		2,4-Diamino-5-keto-pyrimidine 29			[Z][H] -> [Z]S(=O)(=O)C

点击 Reactant Name 栏的右边的三角箭头，使三角向下，按名称降序排列，找到包括 CDK Inhibitor 2 和 CDK Inhibitor3 的四对 Activity cliffs。

这四对 Activity cliffs 本质上是相同的转换，MMPs 识别的原理关系到两个化合物化学改变的尺寸上的限制。因此，转换代表着取代基的尺寸由最小到最大的变换范围。



点击 Results 下的 Input Ligands with Activity Information 打开，双击 Num\_Cliffs 栏两次。  
此 SD 文件含有添加的性质的数据，从中我们可以看出每个配体被包含在几对 Activity cliffs 中，一对 Activity cliffs 中另一个化合物是什么和转换的部分是什么。  
观察 91 号化合物（Index 为 91），其中 Num\_Cliffs 为 3，活性的改变都为正值，都是作为 Activity cliffs 低活性的化合物。如果活性改变为负值，表明 Activity cliffs 中另一个化合物具有很好的改造潜能。

	Index	Name	Visible	Tagged	Visibility Locked	BindingDB monomerid
1	77	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,059
2	6	CDK Inh...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	81,430
3	5	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,061
4	24	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,046
5	67	CDK Inh...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	81,429
6	92	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,037
7	89	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,063
8	91	2,4-Dia...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	12,605
9	3	CDK Inh...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	81,433
10	17	2,4-Dia...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	12,611
11	32	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,050
12	76	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,069
13	1	CDK Inh...	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	81,438
14	2	CDK Inh...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	81,439
15	7	R547	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	12,621
16	9	pyrazol...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,337
17	30	2,4-Dia...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	12,612
18	51	CDK Inh...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	81,441
19	62	2-Anili...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	8,065
20	70	2,4-Dia...	<input type="checkbox"/> No	<input type="checkbox"/> No	<input type="checkbox"/> No	12,618

## Discovery Studio MMPs 教程 2

### 使用 MMPs 分析活性变化

所需功能和模块: Discovery Studio Client, DS QSAR.

所需数据文件: p450\_2d6\_chembl1614110.sd.

所需时间: 10 分钟

#### 介绍

药物化学家的大量时间可能花费到先导化合物优化的过程, 优化过程中需在活性、脱靶作用、毒性、药代动力学之间达到平衡, 为了寻找最大的机会使先导化合物进入后续的药物发现的渠道中。构效关系 (SAR), 是连接化学结构和分子性质的桥梁, 广泛的用于决定合成那个活性类似物。SAR 研究的焦点并不总是增加生物活性, 有时也可以应用现有的具有生物活性数据化合物来获得具有较好药代动力学性质的类似物。

在大型混合的数据库中, 通过识别结构倾向, 使用 MMPs 的方法能辅助优化非活性相关的性质。本教程中, 利用 Analyze Activity Effects using MMPs...工具, 通过识别所有可能的 MMPs 和对相同的转换进行聚类, 来识别数据库中的结构倾向。转换的频率表明了结构转换引起性质改变的可能性, 极值可能表明避开常规倾向的潜在途径。

本教程中, 包含各种配体的细胞色素 P450 2D6 (CYP2D6) 抑制剂数据库用来研究结构倾向, 指导分子设计, 辅助先导化合物优化过程。CYP2D6 是含血红素酶家族中的一员, 在药物代谢中起着重要的作用, 了解 CYP2D6 代谢可以提高药物的活性, 避免或最小化抑制代谢。

本教程中, 你能了解到在哪里的转化能增加或减少活性, 和哪里转换对活性没有影响。

### 使用 MMPs 分析活性变化

CYP2D6 配体可以在 ChEMBL 数据库下载 (<http://www.ebi.ac.uk/chembidb/>), 并且含有生物活性数据。配体通过去除盐离子和保存最大的片段进行处理, 去除重复的配体。

打开 Samples | Tutorials | QSAR | p450\_2d6\_chembl1614110.sd 文件

展开 Small Molecules | Study SAR 工具面板, 点击 Analyze Activity Effects Using MMPs... 打开

Input Ligands 设置为 bindingdb\_cdk2>All

Activity Property 设置为 pIC50.

Activity Change Threshold 设置为 2

点击 Run 运行

本次作业运行小于 1 分钟, 在作业浏览窗口报道产生 16093 个已经被识别的 MMPs。

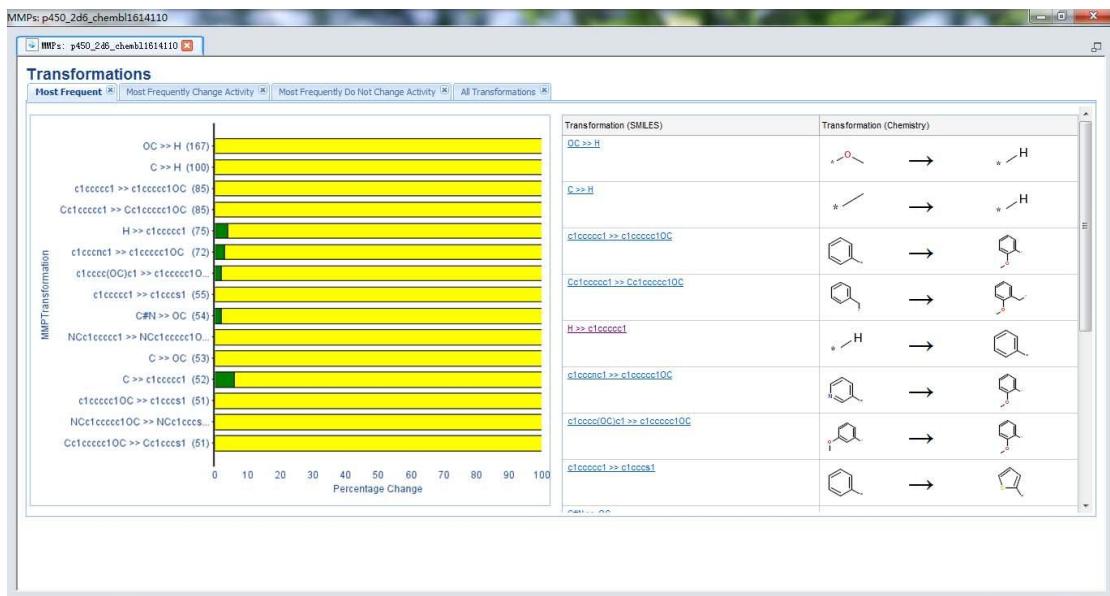
在作业浏览区展开 Summary

报道了 CYP2D6 数据库共产生 16093 个 MMPs, 并且含有 527 个不同的转换。

在作业浏览窗口展开 Results, 点击 View Results。

转换报告页面由四部分组成, Most Frequent 列表含有 15 个出现频率最高的 MMP 转换。

SMILES 表示转换虚拟的图像的表达。颜色表示转换频率导致活性的改变, 其中绿色表示增加, 红色表示减少, 黄色表示不变。



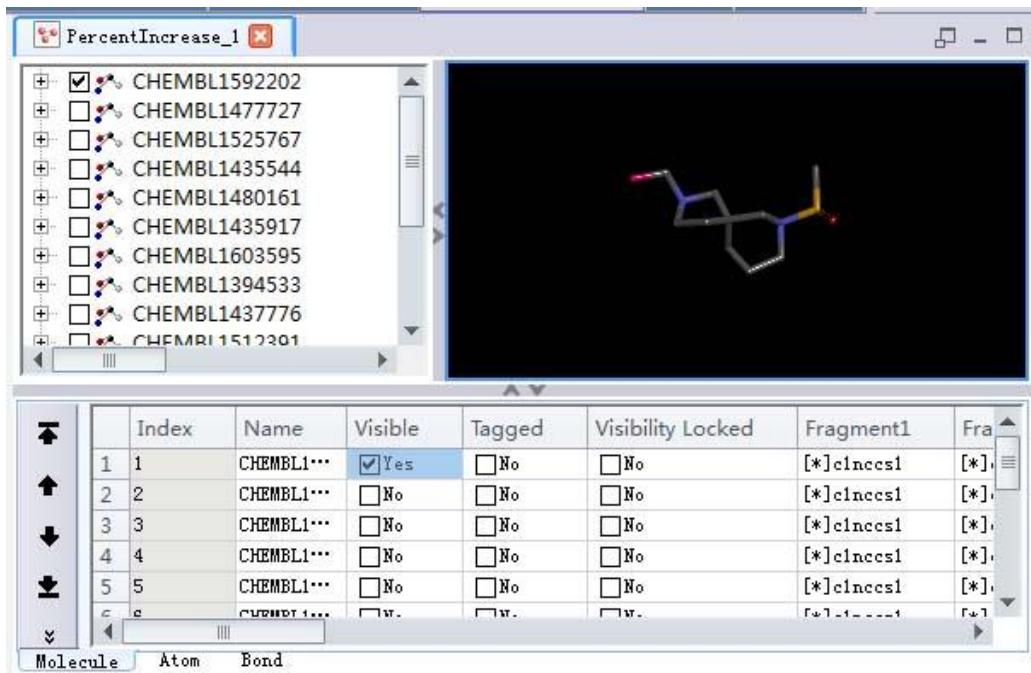
### 打开 Most Frequently Change Activity 窗口

这个窗口列举了 15 个频率最高的转换，这些转化导致的活性变化在两个数量级以上。

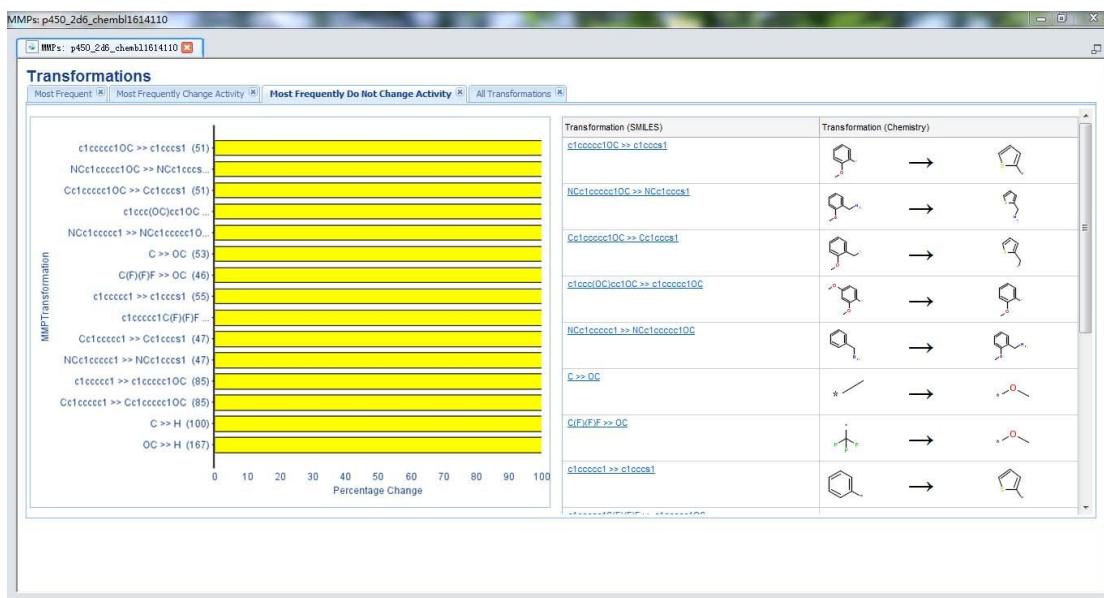


### 打开第 1 个转换 C(=O)c1cnccn1 >> S(=O)(=O)c1ccccc1, 打开一个数据窗口, 双击 MMPEffect 栏, 同时打开分子窗口。

暗橘色的原子表示转换的位置，所有的这些转换都包含甾类脂化合物。观察 Activity Change 一栏，有的化合物 MMP 效应减少，Activity Change 为 -2.7，有的化合物 MMP 效应增加，Activity Change 为 2.4，有的化合物 MMP 效应为中性，即 Activity Change 在 2 和 -2 之间。



回到 Transformations 总结栏，打开 Most Frequently Do Not Change Activity 一栏。此栏列举了 15 个转换频率最大的导致活性改变小于两个数量级。



### 打开 All Transformations 一栏

这个窗口总结了所有的已经识别的转换。注意到 527 个可能的转换中仅仅 300 个显示在窗口中。但显示所有可能的结果时，最大转换的数目可能减少。你可以通过以上的分析，通过频率和 MMP 效应（中性，增大或减少）来分析所有转换。



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

虚拟组合库设计与分析专题

## 构建及优化组合化合物库教程

**目的:** 通过此教程,了解 Discovery Studio 中构建组合化合物库并进行优化操作方法及结果分析。

**所需功能和模块:** Discovery Studio Client, DS Library Design。

**所需数据文件:** amines\_35.sd, acids\_45.sd 和 AmideFormation.rxn

**所需时间:** 10 分钟

### 介绍

组合化学为发现新的药物先导化合物提供了大量候选化合物,但它是一种高成本、低利用率的方法。虚拟组合库的出现和虚拟筛选的方法使组合化学扬长避短,成为一种更实用的药物研究方法。而在成功构建组合库后,将组合化学和合理药物设计结合起来,通过分子模拟和理论计算方法合理地设计虚拟化合物库,可以增加库中化合物的多样性和类药性,提高库的质量和筛选效率。因此,对于一个构建好的虚拟库,通常需要同时对多种性质进行优化,而通过帕累托优化则可以同时优化组合库的各类性质。

本教程主要介绍了如何通过同一个反应类型构建虚拟组合化合物库,并经帕累托优化使得到的子化合物库兼具多样性及类药性。

本教程包括以下步骤:

- 组合库的构建及优化
- 结果分析

### 组合库的构建及优化

本教程基于以下羧酸和氨缩合成酰胺的反应类型为基础构建虚拟组合化合物库:

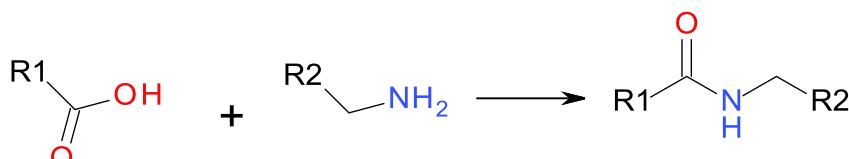


图 1 构建组合库参照的反应式

### 构建组合库操作及参数设置

#### 1.组合库构建参数设置

在文件浏览器（Files Explorer）中，展开 Samples| Tutorials| Library Design 文件夹，找到 amines\_35.sd 文件，双击打开在分子窗口中显示（默认为表格浏览器），在同一个文件夹中找到 acid\_45.sd，单击选中并拖拽进入同一窗口中。

展开分子视图（Graphical View）窗口（可通过 CTRL+G 展开）和系统视图（Hierarchy View）窗口（可通过 CTRL+H 展开）（图 2）

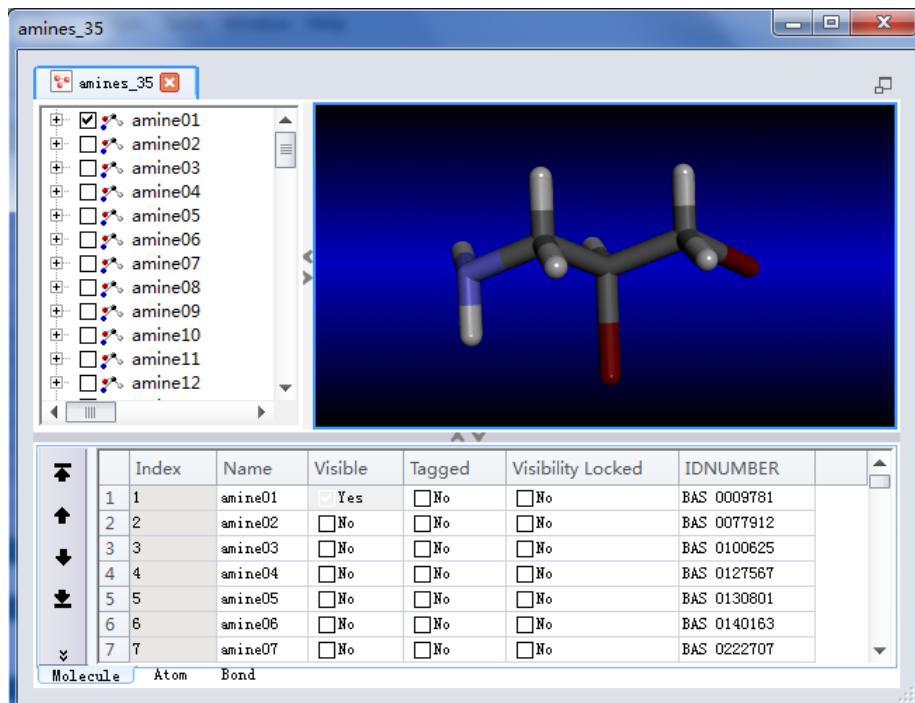


图 2 构建组合库需要的羧酸分子和含氨基分子

可通过表格视图 (Table View) 中的 按钮逐一查看各分子结构

在工具浏览器 (Tools Explorer) 中，展开 Small Molecules| Select Subset Libraries，点击 Optimize Combinatorial Library with Pareto Method。

在打开的参数设置面板中，点击 Input Library Type，在下拉列表中选择 Reaction。

点击 Input Reactants 一栏，在下拉列表中选择 amine\_35>All。

点击 Input Reaction File，浏览选取 Samples\Tutorials\Library Design\AmideFormation.rxn，指定依据的反应类型。

在 Library Dimensions 参数栏输入 7,8，指明每个片段库中分别用于后续组合库构建的片段个数，此处即由 7 个含氨基分子和 8 个羧基化合物反应得到含 56 个酰胺类化合物的化合物库。上述参数设置如下图 3 所示：

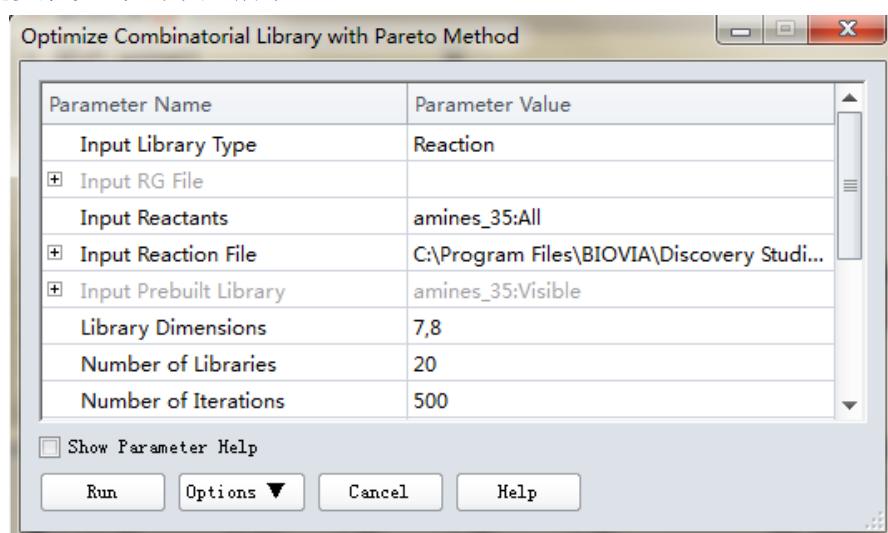


图 3 构建组合库部分参数设置（优化部分待续）

## 2. 定义帕累托优化参数设置

点击 Optimization Property 1 参数栏右边的浏览按钮，在出现的 Calculable Property 对话框（图 4）中输入 fcfp，在下拉列表中选择 FCFP\_12，点击 OK 载入，即从分子多样性的角度来优化库。

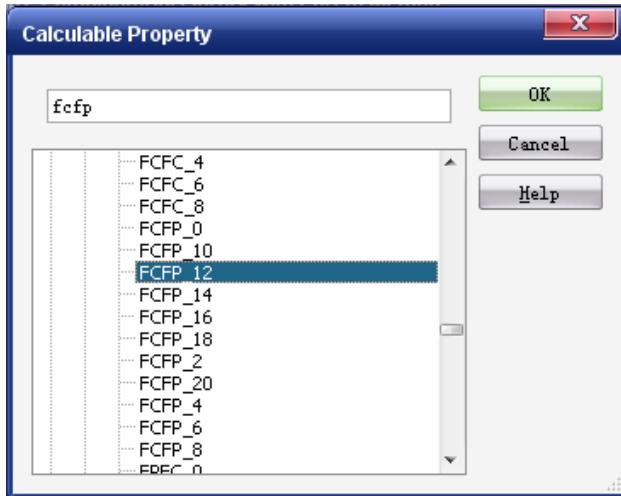


图 4 选择描述分子多样性参数 FCFP\_12 的选择

展开 Optimization Property 1 参数栏，点击 What to Optimize 参数，在下拉列表中选择 Unique Value Count，Goal 为默认的 Maximize。

通过优化得到此性质最大化的子化合物库，用来描述库的多样性

点击 Optimization Property 2 参数栏右边的浏览按钮，在出现的 Calculable Property 对话框（图 5）中输入 molecular\_weight，在下拉列表中选择此参数，点击 OK 载入

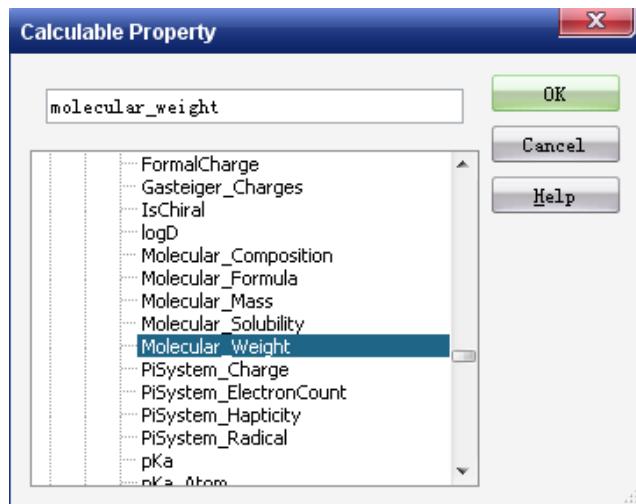


图 5 描述分子类药性参数分子量的选择

展开 Optimization Property 2 参数栏，点击 What to Optimize 参数，在下拉列表中选择 Range Penalty，点击 Range Minimum，输入 100，点击 Range Maximum 输入 500，Goal 选择 Minimize。

通过优化可得到分子量在 100-500 之间的分子，并且越小越好。

点击 Optimization Property 3 参数栏右边的浏览按钮，在出现的 Calculable Property 对话框（图 6）中输入 num\_h\_donors\_lipinski，在下拉列表中选择此参数，点击 OK 载入

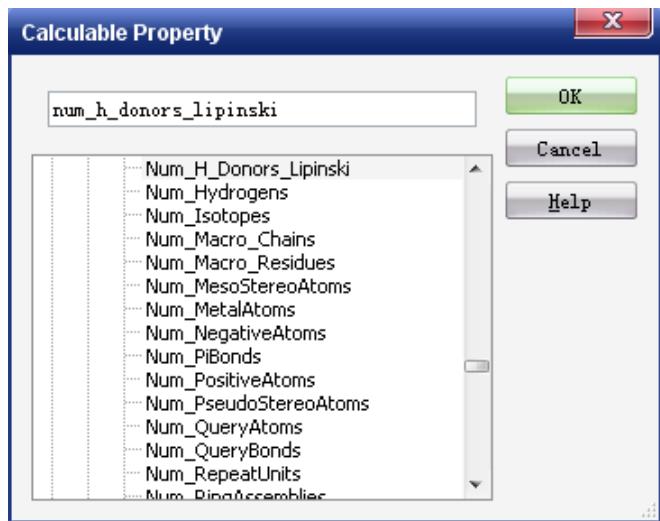


图 6 描述分子类药性参数氢键供体数目选择

展开 Optimization Property 3 参数栏, 点击 What to Optimize 参数, 在下拉列表中选择 Range Penalty, 点击 Range Minimum 输入 0, 点击 Range Maximum 输入 5, Goal 选择 Minimize。通过优化可得到氢键供体数目在 0 到 5 之间的分子, 且越少越好。

优化部分参数设置如下图 7 所示:

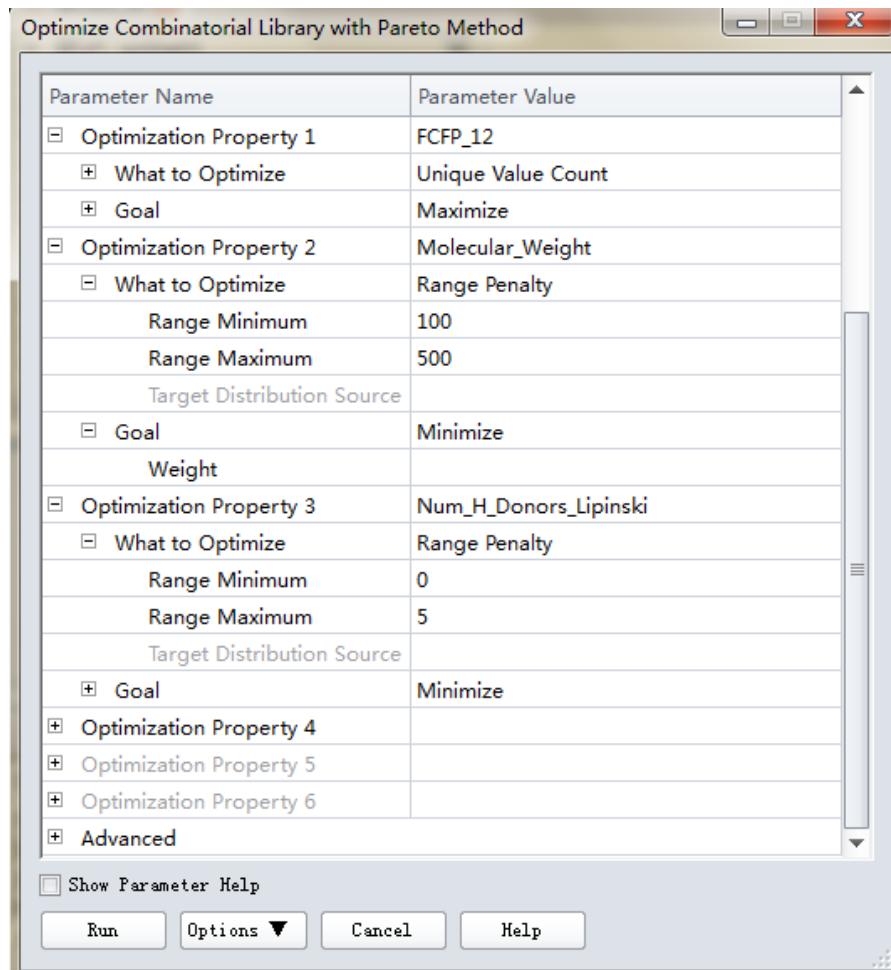


图 7 优化部分参数设置

点击 Run 运行作业。

等待作业完成，会自动弹出一个 Report 页面。

## 结果分析

### 1. 查看 Report 页面

从 Summary 一栏可知，该次运算产生了 20 个化合物库，每个库中包含 56 个分子，帕累托优化后结果和各个库的结果描述如下表（图 8）所示：

Pareto Subset Optimization								
LibraryID	CrowdingDistance	ParetoFront	FCFP_12_Uncount	Molecular_Weight_RangPenalty	Num_H_Donors_Lipinski_RangePenalty	R1	R2	
1	1e+099	1	1786	116.673984642857	1.14285714285714	11,12,16,17,20,22,34	42,45,46,56,60,65,72,77	
2	1e+099	1	1806	120.091960892857	0.982142857142857	11,12,16,17,20,22,34	42,45,46,56,65,66,67,72	
3	1e+099	1	1801	121.608399821429	0.928571428571428	11,12,16,17,20,22,34	45,46,56,59,65,66,69,72	
4	1e+099	1	1575	0	0	3,8,10,14,23,30,32	41,46,54,56,65,67,69,72	
5	1.46044937291182	1	1713	82.5957428571428	0.464285714285714	4,11,12,14,16,17,34	45,46,54,56,60,66,67,68	
6	1.27702189116704	1	1761	92.0853126785714	0.214285714285714	3,11,16,17,20,34,35	45,46,56,59,65,66,69,72	
7	1.24311921597887	1	1735	59.32207625	0.785714285714286	11,12,17,20,22,30,34	42,46,56,59,65,66,68,72	
8	1.22846867916971	1	1747	62.2007305357143	0.714285714285714	4,11,12,17,20,22,34	45,46,56,59,65,66,69,72	
9	1.07408692471086	1	1637	3.79305285714286	0	3,4,6,11,14,17,30	42,45,46,54,56,65,66,72	
10	1.0577751877791	1	1766	103.548130892857	0.321428571428571	11,12,16,17,20,28,34	45,46,53,54,56,62,66,72	
11	1.05501753265278	1	1588	3.59903196428571	0	3,4,8,11,14,17,32	47,56,59,65,66,67,68,69	
12	0.984145702996751	1	1678	13.63777	0	3,4,11,14,17,30,34	42,45,46,54,56,65,66,72	
13	0.973590267835944	1	1783	107.40329125	0.357142857142857	11,12,16,17,20,28,34	42,46,56,59,65,66,68,72	
14	0.931667922309399	1	1711	31.46880807142857	0	3,4,11,17,20,32,34	42,45,46,54,56,65,66,72	
15	0.910611847934342	1	1680	29.5620146428571	0.125	4,11,12,17,28,32,34	42,46,56,59,65,66,67,68	
16	0.89122780395702	1	1714	45.8818342857143	0.642857142857143	3,4,11,17,20,22,34	42,45,46,56,65,66,70,72	
17	0.875439760296529	1	1644	12.0255917857143	0	3,4,11,14,17,30,34	36,42,45,46,54,56,65,72	
18	0.829492700963843	1	1690	31.0722446428571	8.92857142857143e-002	4,11,12,14,17,28,34	45,46,56,59,65,66,67,69	
19	0.703798307050989	1	1797	119.682639464286	1.03571428571429	11,12,16,17,20,22,34	42,46,56,59,65,66,68,72	
20	0.6858425330877	1	1723	47.6319710714286	0.660714285714286	3,4,11,17,20,22,34	45,46,51,56,65,66,69,72	

图 8 构建的组合库经帕累托优化的结果

其中，ParetoFront 为 1 代表计算结果已收敛，否则，需要增加参数设置中的迭代步数（默认为 500），直到该值为 1 为止；

FCFP\_12\_Uncount 代表化合物多样性，该值越大代表多样性越好；Molecular\_Weight\_RangPenalty 代表化合物库中分子量平均值与设置范围的偏差，如该例中分子量设为 100 到 500，如果分子库平均分子量在两者之间，该值为 0，该值越大，说明离设定的范围越远；

同理，Num\_H\_Donors\_Lipinski\_RangePenalty 代表氢键供体数目平均值与设置范围的偏差，此值为 0 时说明满足所设定条件（该例中设定范围为 0 到 5 之间），该值越大说明离设定的范围越远；

最后两栏分别代表构建化合物库所用到的 7 个含有氨基的分子和 8 个含有羧基分子的编号；经以上分析可知，我们需要找到 FCFP\_12\_Uncount 值较大，Molecular\_Weight\_RangPenalty 和 Num\_H\_Donors\_Lipinski\_RangePenalty 两值较小的分子。

从 Summary 中，还可以看到 4 个二维点图，以第一副图为例（图 9）：

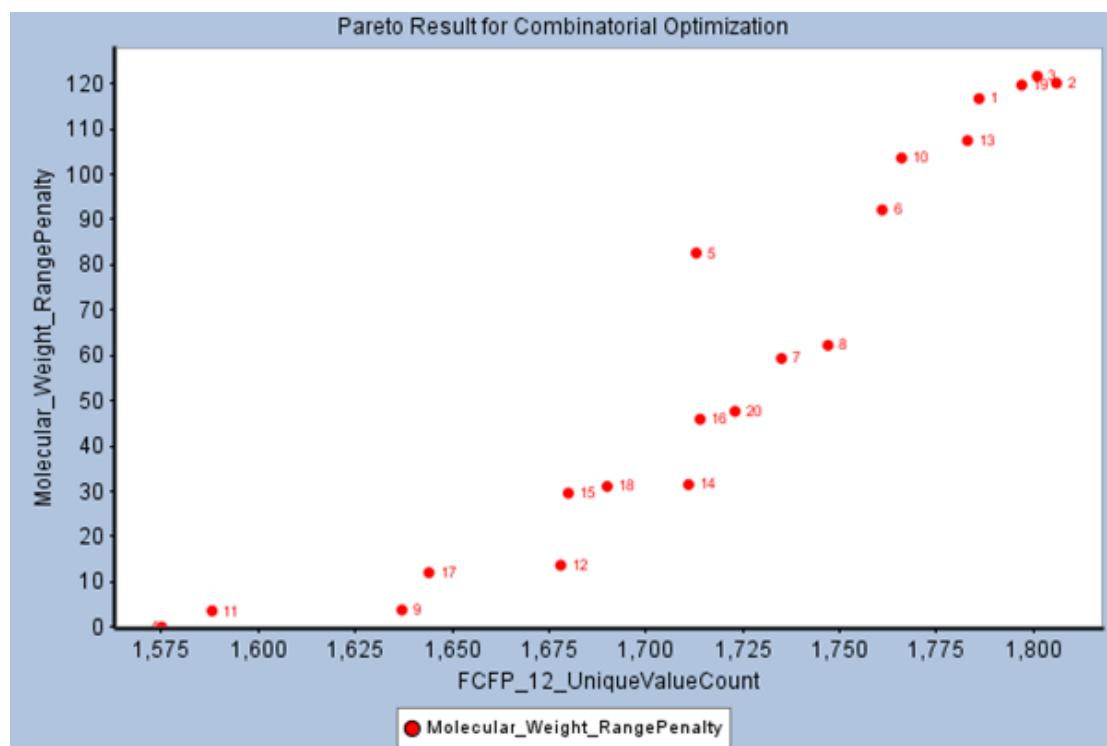


图 9 以 FCFP\_12\_UnciqeValueCount 和 Molecular\_Weight\_RangePenalty 值为纵坐标及横坐标所作二维点图

该图中靠近右上方分子库多样性较好但分子量偏高，左下方分子库分子量接近所设定范围但多样性偏差，如需两者兼顾可选中间化合物库进行研究。

点击 Results 中的 View Optimization Progress, 可得到三维显示点图（图 10），分析方法类似。

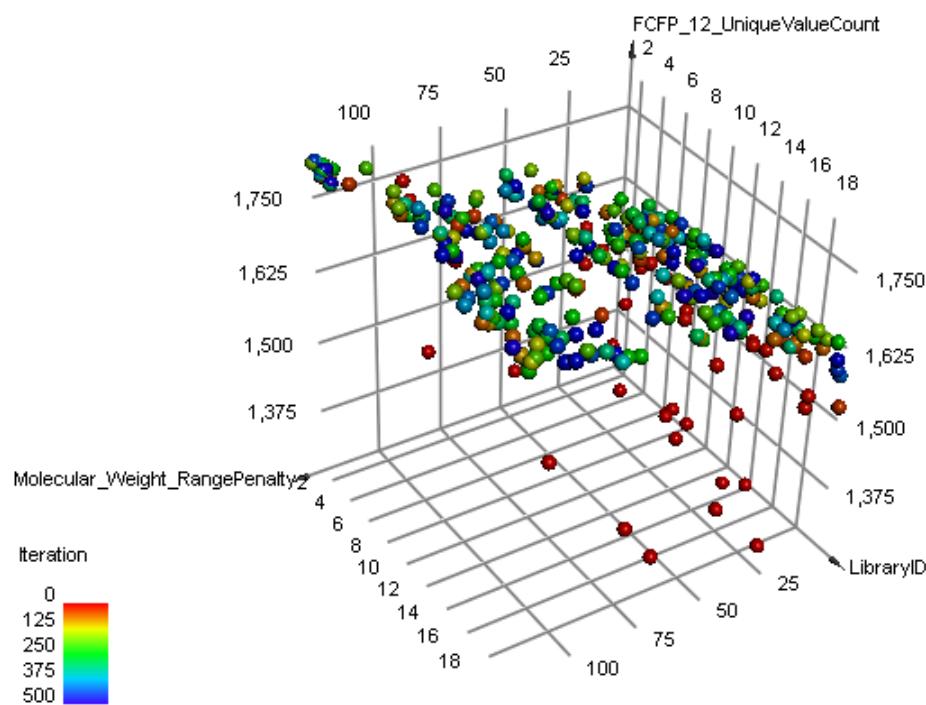


图 10 描述化合物库性质三维点图



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本： 4.5)**

化合物 ADMET 性质预测专题

**DS\_ADMET、 DS\_TOPKAT**

## 使用 Discovery Studio 进行化合物 ADMET 性质预测教程

**目的:** 采用 Discovery Studio, 以一组小分子化合物为实例, 计算其 ADMET 性质并对预测结果进行分析。

**所需功能和模块:** Discovery Studio Visualizer Client 和 DS ADMET。

**所需数据文件:** pk-test.sdf。

**所需时间:** 10 分钟

### 介绍

ADMET 性质是指分子在有机体内的吸收、分布、代谢、排泄和毒性等性质。如果在药物研发的早期阶段就能依据化合物的 ADMET 性质对先导化合物进行有针对性的选取和优化改造, 这对提高药物研发的成功率以及减少药物研发后期过程中由于 ADMET 性质问题所造成资金浪费问题, 是非常必要的。ADMET 描述符可以有助于及早排除 ADMET 性质不好的化合物从而避免后期耗资巨大的结构改造, 同时也可以评价结构优化的效果, 是否确实改善了 ADMET 属性, 从而避免合成所支出的过多资源。

本教程主要介绍了在 Discovery Studio 中如何进行化合物 ADMET 性质的预测, 以及预测结果的分析。

本教程包括:

- 运行 ADMET 性质计算流程
- 分析 ADMET 性质预测结果

Discovery Studio 中可以计算的 ADMET 性质包括:

- 25 摄氏度下水溶解度 (aqueous solubility)
- 血脑屏障通透性 (Blood brain barrier penetration, BBB)
- 细胞色素 P450 2D6 抑制性 (Cytochrome P4502D6 inhibition)
- 肝毒性 (hepatotoxicity)
- 人类肠道吸收性 (human intestinal absorption, HIA)
- 血浆蛋白结合率 (plasma protein binding)

### 运行 ADMET 性质计算流程

#### 1. 导入小分子化合物文件

在文件浏览器 (Files Explorer) 中, 展开 Samples | Tutorials | QSAR, 双击打开 pk-test.sdf 文件。

在表格浏览器中可以看到一共有 20 个丙酮酸盐激酶抑制剂。

#### 2. 选择计算性质, 运行计算流程

在工具浏览器 (Tools Explorer) 中, 展开 Small Molecules | Calculate Molecular Properties, 点击 ADMET Descriptors, 打开 ADMET Descriptors 对话框。

点击 Input Ligands 右边的栅格, 选择下拉列表中的 pk-test:All, 选择所有小分子化合物。

注: 如果 pk-test Molecule 窗口已经打开, 那么该处会自动设置。

ADMET Descriptors 参数选择默认设置, 即选取所有 ADMET 性质。(图 1)

此操作将计算所有小分子化合物的所有 ADMET 性质。

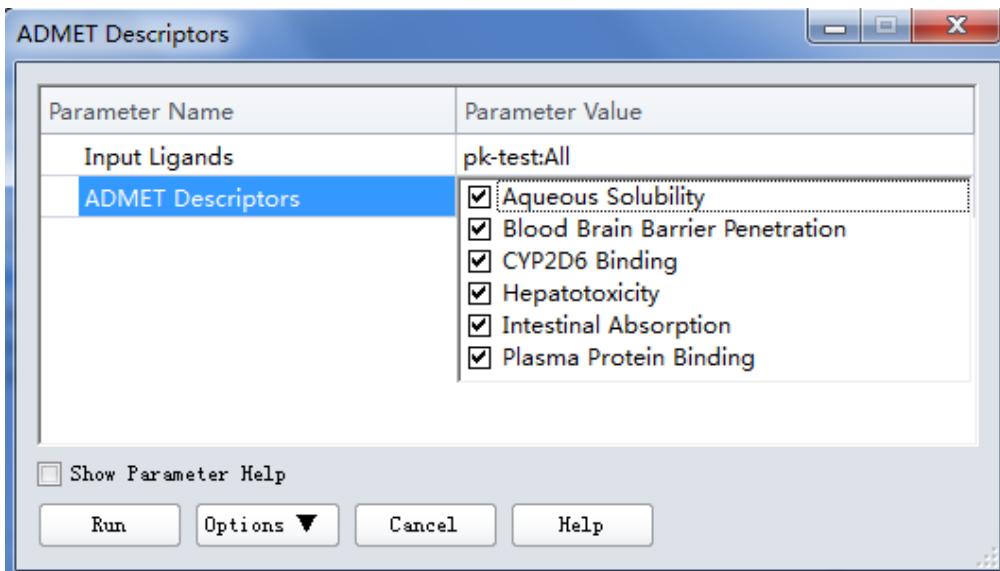


图 1“ADMET Descriptors”流程参数设置

点击 **Run** 运行作业。

等待作业完成。

注：该作业完成时间在 30s 以内（CPU 为 2.4GHz 的计算机）。

为清晰起见，从菜单栏中选择 **Window | Close All** 关闭不必要窗口。

### 分析预测结果

待作业完成后，自动跳出一个对话框（图 2）。

或者待作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。点击 **View Results** 按钮。同样会跳出该对话框。



图 2 ADMET plot 对话框

点击 **Yes**，打开 **ADMET plot** 窗口。（图 3）

该 ADMET plot 是一个 ADMET\_PSA\_2D vs ADMET\_AlogP98 的 2D 图。图中显示了两个系列的椭圆，分别表示了血脑屏障通透性（BBB）模型 95% 和 99% 的置信区间，以及人体肠道吸收性（HIA）模型 95% 和 99% 的置信区间。

注：如果在图中没有显示 4 个椭圆，则从菜单栏中选取 **Edit | Preferences**，打开 Preferences 对话框，展开 **Charts**，选取 **ADMET Plot**，检查 **Blood Brain Barrier** 和 **Absorption** 是否都已选中。（图 4）

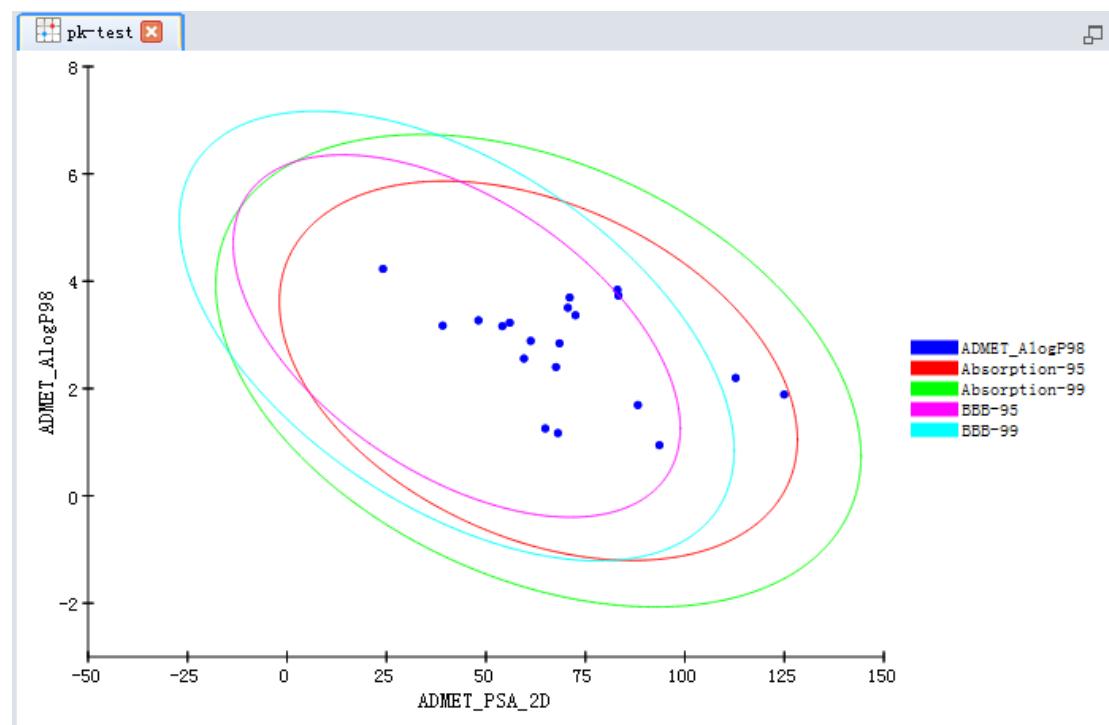


图 3 ADMET plot

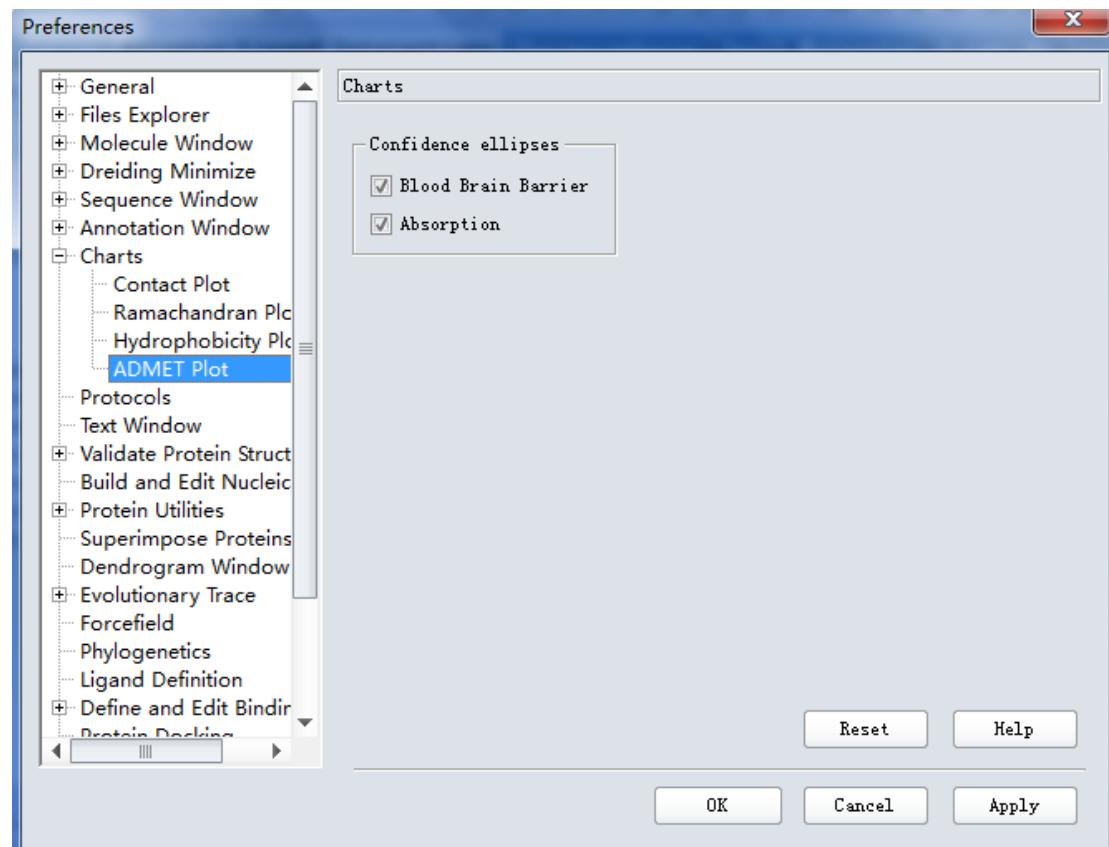


图 4 Preferences 设置面板

注：如果在 ADMET Plot 窗口中并没有如图 3 中所显示的图标，可以通过如下操作来获取：

在窗口中点击鼠标右键，选择 **Display Style**，在 **Axis and Font** 下勾选 **Show Legend**。（图 5）

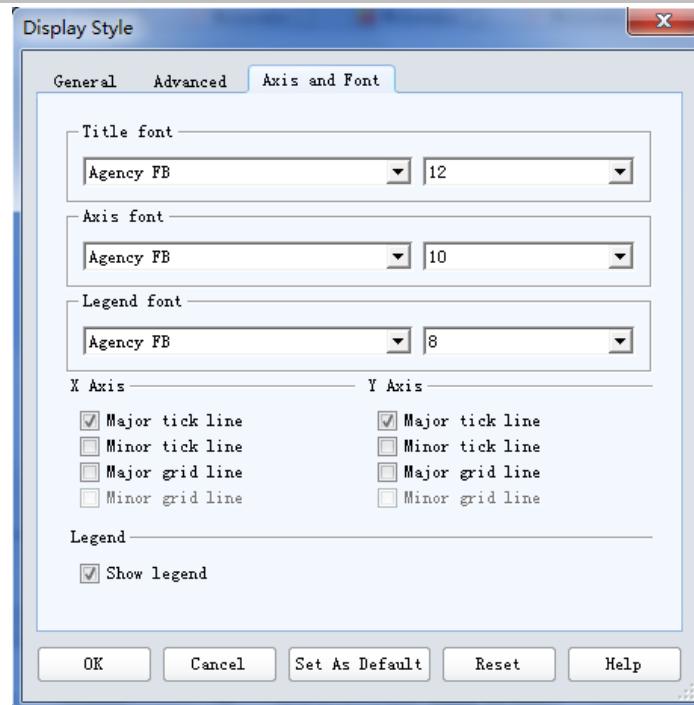


图 5

将 ADMET plot 标签拖至视图窗口，使得该窗口同 pk-test Molecule 窗口并列显示。（图 6）

点击蓝色椭圆（99% 血脑屏障的置信区间）外围的点。

与该点对应的分子将在 pk-test 分子窗口中突出显示。由于该分子处于 BBB 模型 99% 置信区间以外，对该分子的预测就视为不可靠的。

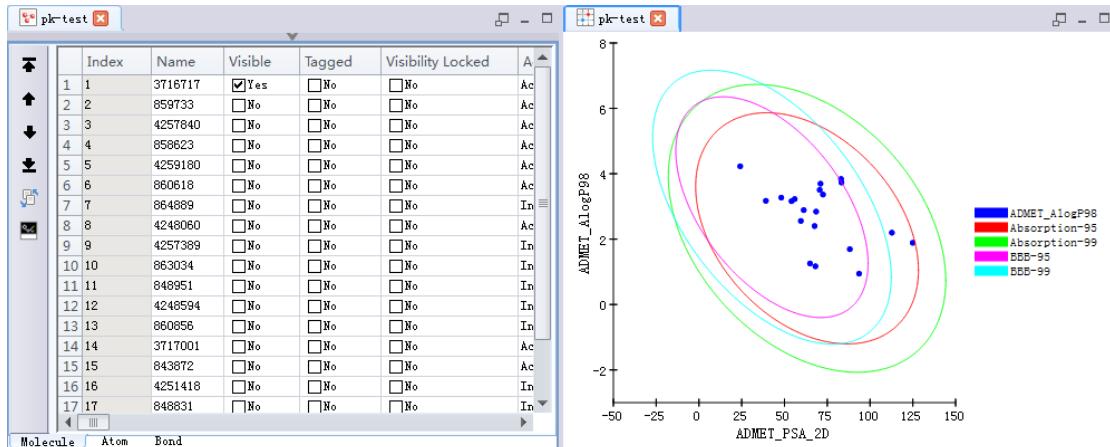


图 6 分子窗口同 ADMET plot 窗口并列显示

提示：按住 SHIFT 可以选中多个分子，或者使用 工具在 ADMET plot 外围画个圈以选中圈内的所有分子。同样在分子窗口选中某一分子，ADMET plot 中相应的点也会突出显示。

展开菜单栏 **Help**，点击 **Help Topics** 打开帮助文件，**Contents** 下点击选择 **Small Molecules tools | Theory-Small Molecules | Theory-ADMET Descriptors**，打开相关帮助文件，可以查看每一种预测的属性的详细信息。（图 7）

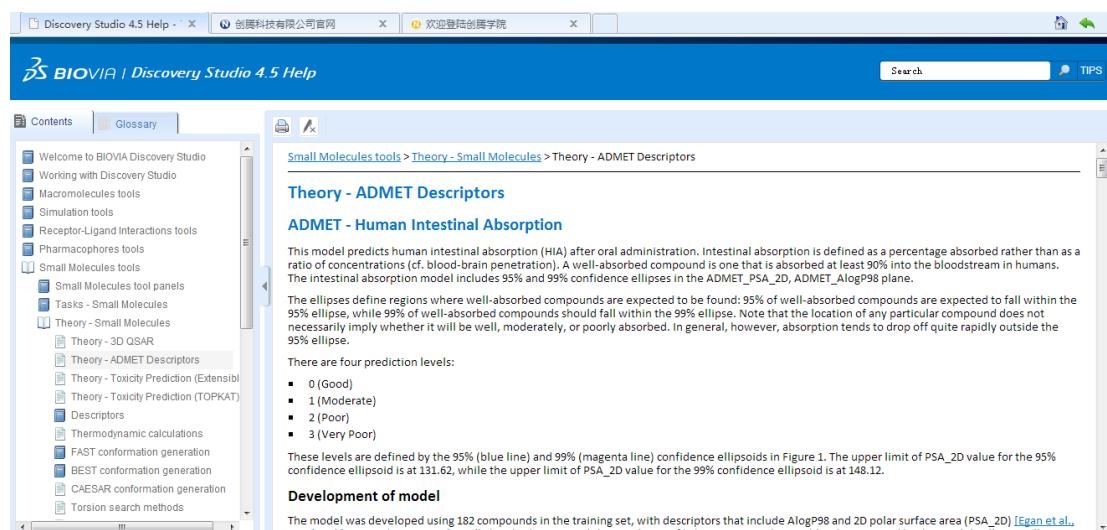


图 7 ADMET 帮助文件示意图

## 使用 Discovery Studio 进行化合物毒理性质预测（TOPKAT）教程

**目的：**采用 Discovery Studio，以一组小分子化合物为实例，示范毒理学性质预测及结果分析操作过程。

**所需功能和模块：**Discovery Studio Visualizer client 和 DS TOPKAT。

**所需数据文件：**1fvv\_ligand.dsv。

**所需时间：**10 分钟

### 介绍

毒理学性质预测流程 TOPKAT 基于分子的 2D 结构对化合物的毒性及环境效应进行快速且准确地计算预测并加以验证。该流程采用了一系列强大且经交叉验证过的定量结构毒性关系（QSTR）模型来评估各种毒性预测结果。

基于该流程，我们可以完成以下工作：

- 快速评估有机化合物一系列毒理学性质
- 检查子结构以及结构改造的结构毒性关系以及相关的作用机制
- 分子排名以用于实验测试或者进一步的开发
- 设计新的化合物

因此采用 TOPKAT 流程进行毒性预测在节省药物上市时间，减少动物实验，评估人体健康风险以及药物研发流程的战略性部署等方面都起了非常重要的作用。

本教程主要介绍了在 Discovery Studio 中如何进行小分子化合物毒理学性质的预测，以及预测结果的分析。

本教程包括：

- 运行 TOPKAT 流程
- 分析预测结果

Discovery Studio 中计算的毒理学性质包括：

- 潜在发育毒性（Developmental Toxicity Potential , DTP）
- 致突变性（Mutagenicity (Ames test)）
- 喙齿动物致癌性（Rodent Carcinogenicity）——包括 NTP 及 FDA 数据集
- 大鼠长期口服最低毒副反应水平（Rat Chronic Oral Lowest Observed Adverse EffectLevel, LOAEL）
- 皮肤致敏性（Skin Sensitization (GPMT) ）
- 皮肤刺激性（Skin Irritancy）
- 大鼠口服 LD50（Rat Oral LD50）
- 最大耐受剂量（Maximum Tolerated Dosage）
- 黑头呆鱼 LC50（Fathead Minnow LC50）
- 大型蚤 EC50（Daphnia magna EC50）
- VlogP
- 眼刺激性（Ocular Irritation）
- 吸入 LC50（Inhalational LC50）
- 好氧生物降解性能（Aerobic Biodegradability）
- 

### 运行 TOPKAT 毒性预测流程

#### 1. 导入小分子化合物文件

在文件浏览器（File Explor）中，展开 Samples | Receptor-Ligand Interaction，双击打开

**1fvv\_ligand.dsv** 文件，可以看到 1 个化合物显示在 Molecule Window 中。（图 1）

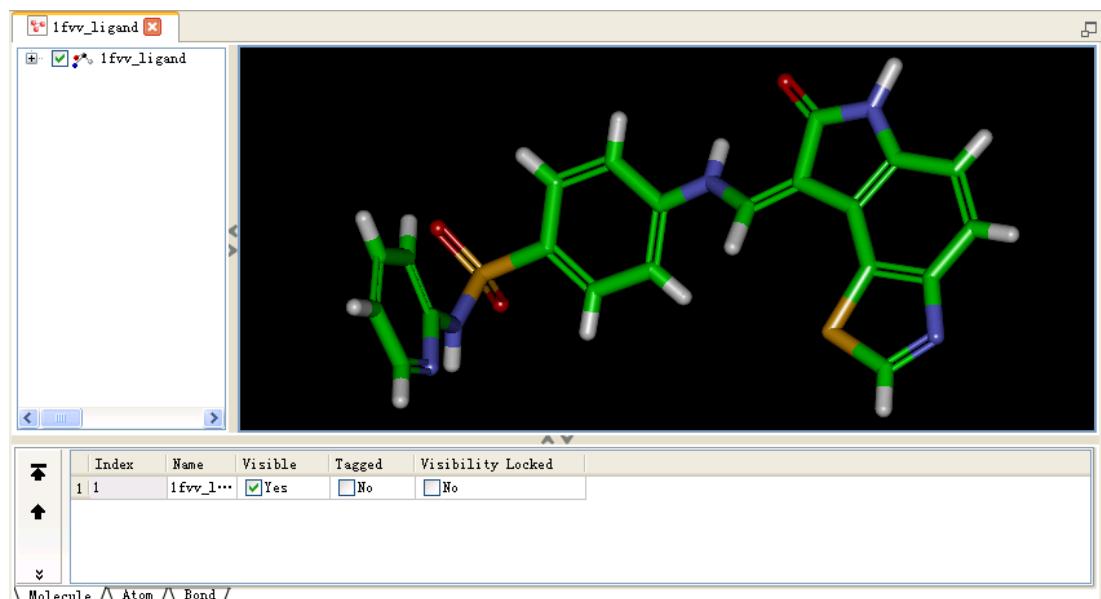


图 1 小分子化合物示意图

## 2. 选择计算性质，运行计算流程

在流程浏览器(Protocols Explorer)中，展开 Small Molecules | ADMET 文件夹，双击 Toxicity Prediction (TOPKAT)，打开参数浏览器。

点击 Input Ligands 右边的栅格，在下拉列表中选择 1fvv\_ligand:Visible，选择该小分子化合物。

点击 Models 右边的 ... 按钮，打开 Models 对话框，选择所需计算的毒理学模型。

在此教程中，我们选择 Ames Mutagenicity、Rat Oral LD50、Aerobic Biodegradability 以及 NTP Carcinogenicity Call (Male Rat) 模型为例。（图 2）

点击 Detailed Report 右边的栅格，在下拉列表中选择 True（图 3）。

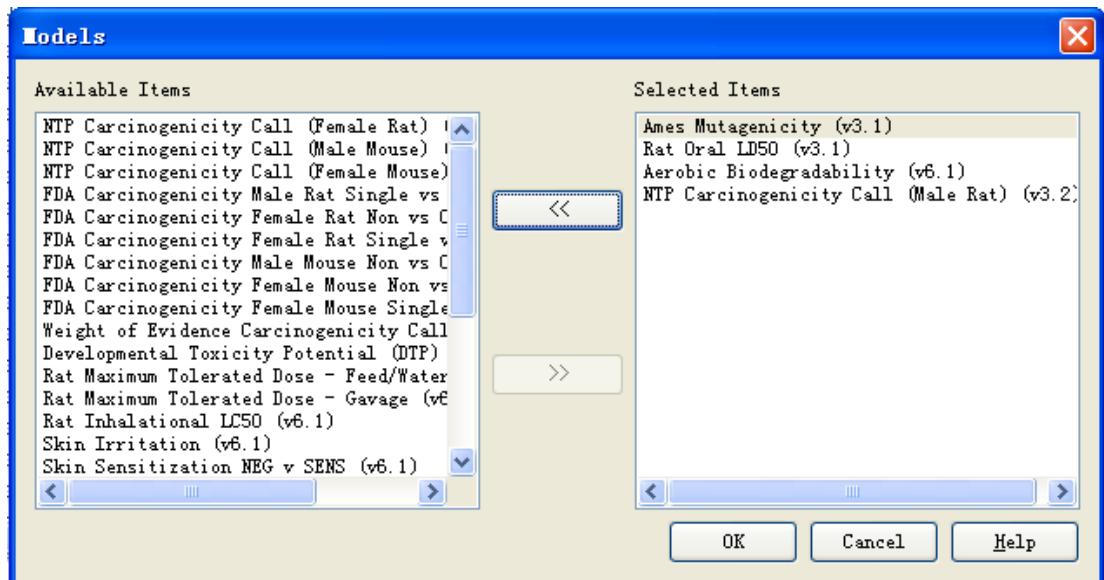


图 2 模型选择对话框

Toxicity Prediction (TOPKAT) ×	
Parameter Name	Parameter Value
Input Ligands	1fvv_ligand:Visible
Models	Ames Mutagenicity (v3.1), Rat Oral LD50 (v3.1), Aerobic Biodegradability (v6.1), NTP Carcinogenicity Call (Male Rat) (v3.2)
+ Submodels	
Input Models File	
Similarity Search	3
Descriptor Contribution	3
□ Detailed Report	True
Format	PDF

图 3 Toxicity Prediction (TOPKAT) 流程参数设置

点击流程工具栏 (Protocols toolbar) 中的 按钮运行作业，等待作业完成。

注：该作业在 2.4 GHz 电脑上完成时间不到 30 秒钟。

为清晰起见，从菜单栏中选择 Window | Close All 关闭不必要窗口。

### 3. 分析预测结果

待作业完成后，双击作业浏览器中刚完成的作业，打开 Report 页面。点击 Detailed Report (PDF) 链接。

该 PDF 结果文件中包含了 4 种毒理学模型的计算结果，包括模型名称、模型评价、预测数值等数据。其中 Page1（图 4）显示了该化合物所有预测的性质汇总及化合物的简单属性，Page2-5（图 5）分别显示的是各预测模型所提供的详细数据。

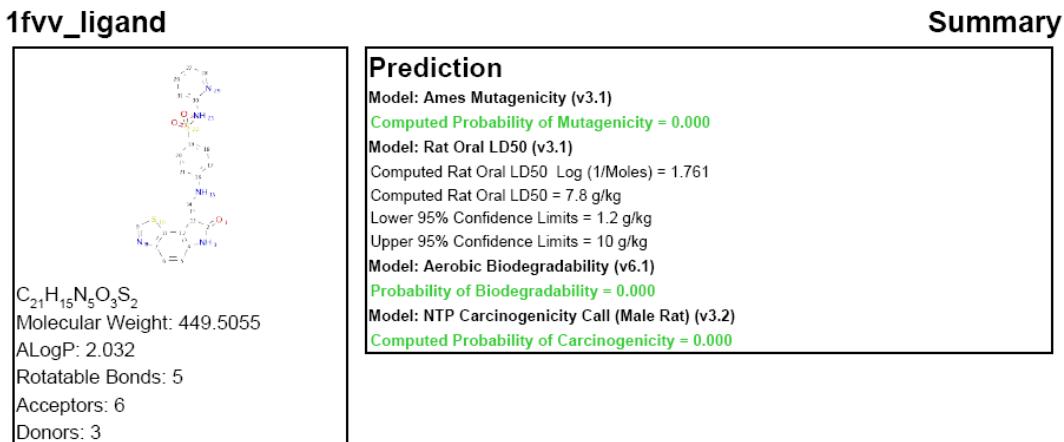


图 4 TOPKAT 流程计算结果示意图 (Page1)

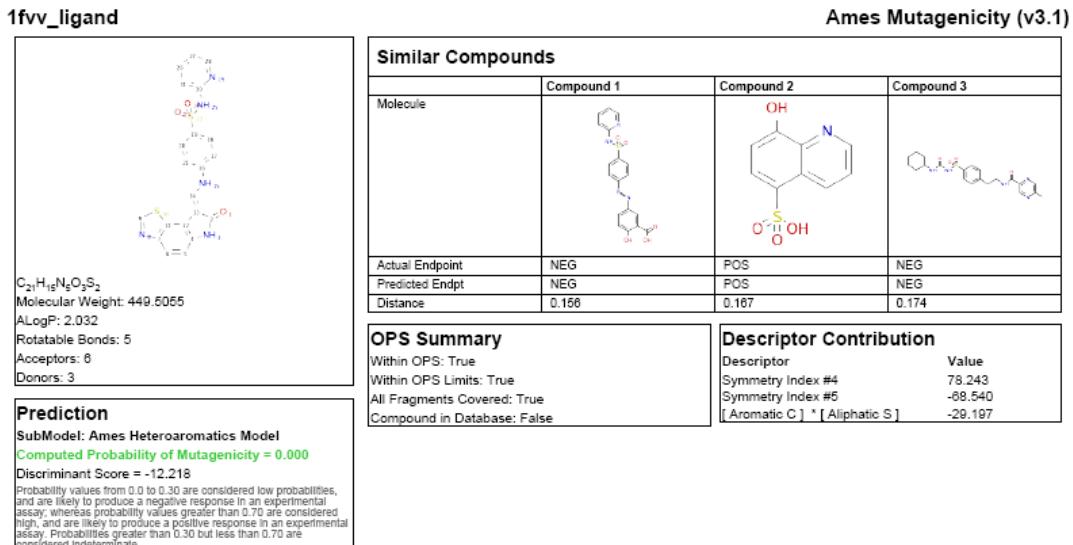


图 5 TOPKAT 流程计算结果示意图 (Page2)

其中 **Within OPS** 衡量该小分子化合物是否位于模型的最优预测空间，**Within OPS Limits** 衡量该小分子化合物是否位于模型最优预测空间界限范围之中。若两者均为 **True**，则所得结果可信；若前者为 **False**，后者为 **True**，则所得结果比较可信；若两者均为 **False**，则所得结果需考量。

**All Fragments Covered** 衡量该小分子化合物各片段是否都涵盖于模型中。

**Compound in Database** 衡量该小分子化合物是否是用来构建模型的化合物。

TOPKAT 结果中所得数值（如 Computed Probability of Mutagenicity）的分类范围及表示意义可在 Help 文件中查找。

展开菜单栏 Help，点击 Help Topics 打开帮助文件，Contents 下点击选择 Small Molecules tools | Theory-Small Molecules | Theory-Toxicity Prediction (TOPKAT)，打开相关帮助文件。(图 6)

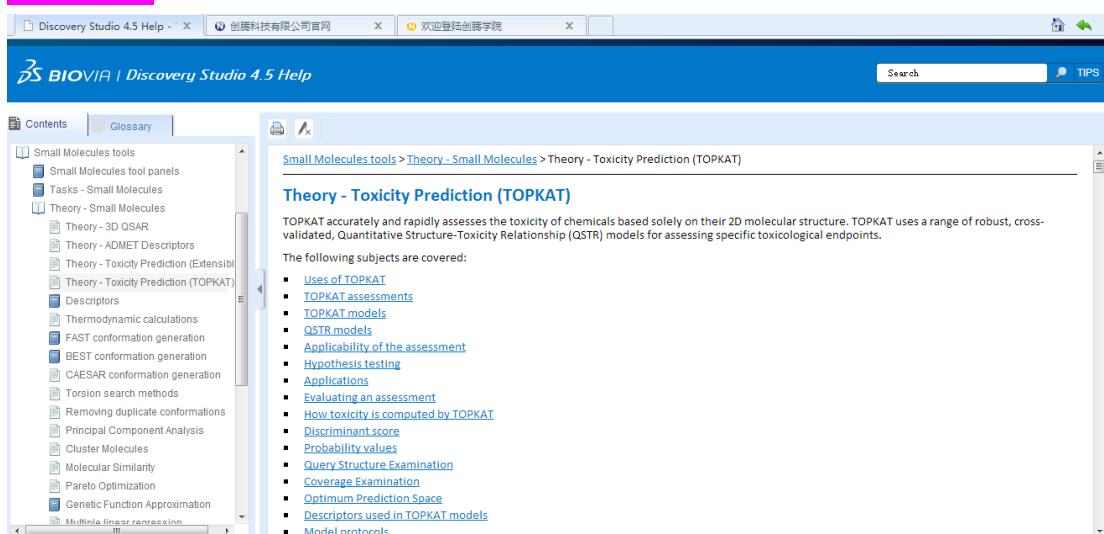


图 6 TOPKAT 帮助文件示意图

帮助文件中详细说明了 TOPKAT 结果中所得数值的分类范围及表示意义。

如此教程中计算 **Aerobic Biodegradability** 模型，所得结果 Probability of Biodegradability 是

一范围，为 0 到 1 数值，数值大于 0.7 则表示该小分子化合物可以被好氧微生物降解，数值小于 0.3 则表示该小分子化合物无法被好氧微生物降解，数值位于 0.3-0.7 之间则无法确定。（图 7）

#### Aerobic biodegradability

This is a single, self-contained module, consisting of four structurally based sub-models. It comprises a statistically significant and cross-validated QSTR model applicable to a specific class of chemicals, and the data from which these models were derived.

All data were determined according to a single study which reported the biodegradability of 894 compounds. This was assessed by the Japanese Ministry of International Trade and Industry (MITI) test protocol ([Loonen et al., 1999](#)). Molecular structure is the only input required to conduct an assessment of aerobic biodegradability.

For this module, the models compute the probability of a submitted structure as being capable of aerobic biodegradation (probability greater than 0.7) or incapable of being degraded aerobically (probability below 0.3). Probability values between 0.3 and 0.7 refer to an indeterminate region in which decisions should not be made except in special circumstances or under further analytical assessments.

This module is designed for operation with the TOPKAT interface, which (i) automatically determines whether the submitted structure belongs to the Optimum Prediction Space (OPS)\* of the model, and (ii) comprises QSTR similarity distance from chemicals with experimental biodegradability data in order to evaluate the reliability of the QSTR-based assessment.

图 7 帮助文件中计算数值的分类范围及表示意义



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

蛋白质序列分析专题

## 蛋白质序列的分析预测教程

**目的：**通过此教程，了解 Discovery Studio 中预测蛋白质二硫键的操作过程。

**所需功能和模块：**Discovery Studio Client, DS Sequence Analysis

**所需数据文件：**2CBA.csv

**所需时间：**1 小时

### 介绍

Predict Protein Sequence 功能可基于单一蛋白序列预测并计算如下性质：

- 基于蛋白序列预测蛋白翻译后修饰（PTM）位点：  
氧化位点、糖基化位点、水解位点、脱酰胺基位点、裂解位点、天冬氨酸异构化位点
- 基于蛋白序列预测抗体中保守氨基酸残基
- 基于蛋白序列预测抗原表位
- 基于蛋白序列识别半胱氨酸、二硫键中的半胱氨酸的识别（需提供蛋白结构）
- 基于蛋白序列计算生物物理学性质，包括分子量、等电点、净电荷、摩尔消光系数、包涵体表达发生几率；
- 基于蛋白序列预测蛋白亲疏水性、跨膜区。

### 蛋白质序列的分析预测

在文件浏览器（Files Explorer）中，找到 **1jfq.bsml** 文件，双击打开在分子窗口中显示。

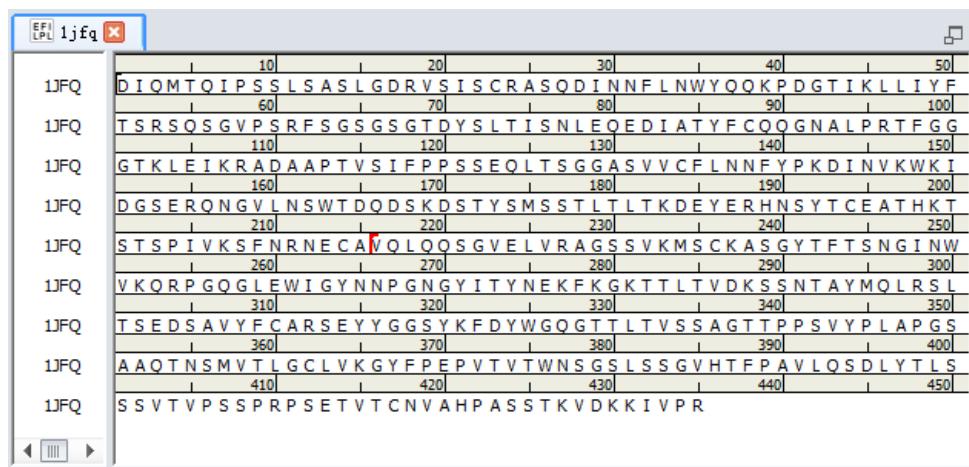


图 1

在工具浏览器（Tools Explorer）中，展开 **Macromolecules| Analyze Sequences**，点击 **Predict Sequence Properties**。

流程对应参数在参数浏览器中打开。

**Input Sequences** 设置为 **1jfg:All**。

如有相应蛋白结构，则可设置 **Input Protein Structures** 参数，若无法提供，则该参数空着即可。

设置 **Calculate Biophysical Properties** 参数为 **True**，用于计算蛋白的分子量、等电点、净电荷、摩尔消光系数、包涵体表达发生几率等生物物理学性质。

点击 Annotation Types 下拉菜单，可勾选 **Sequence motifs**，基于 PROSITE 预测蛋白翻译后修饰位点（PTM 位点）；也可勾选 **Conserved amino acids** 识别出抗体中的保守残基；也可

勾选 **Antigenic regions** 预测蛋白潜在的抗原线性表位。

点击 **Motifs** 参数右侧按钮，弹出对话框。(图 2)

可选择不同的翻译后修饰位点类型。

其余参数设置为默认参数，点击 **Run** 运行。(图 3)

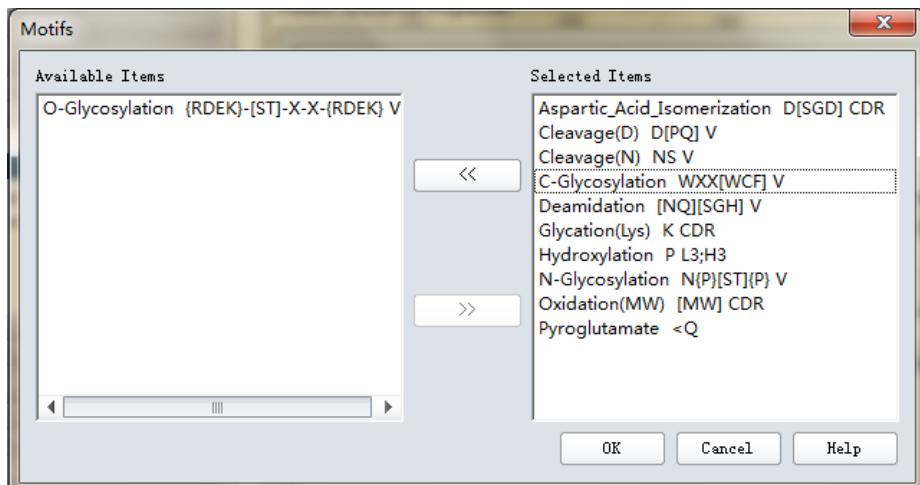


图 2

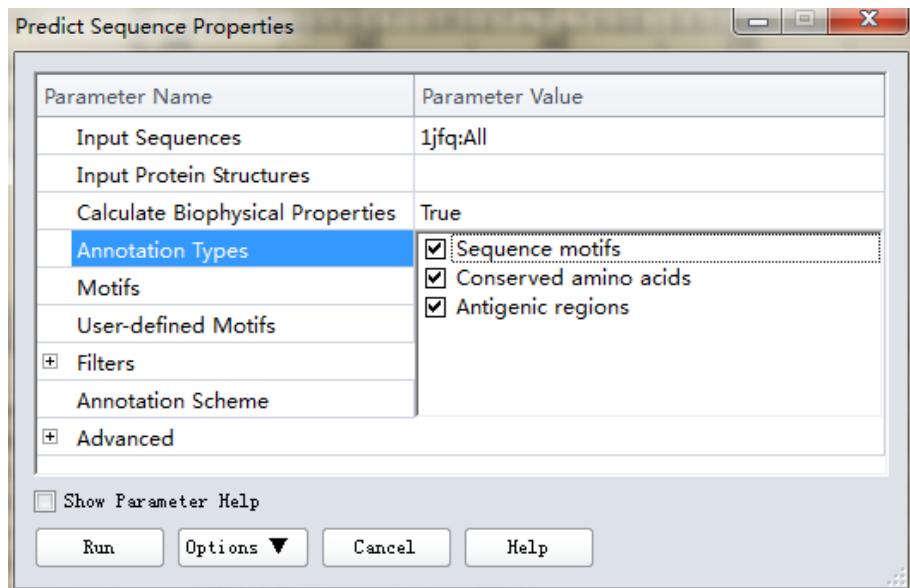


图 3

点击 **Background** 等待作业运行。

作业完成后，展开作业浏览器 (Jobs Explorer) 中该任务并点击 Report 链接，在 Html 窗口中打开 Report 页面。

## 结果分析

在 Report 页面中可以看到 Summary 一栏显示了所有基于蛋白序列预测计算得到的信息 (图 4)。

**Summary**

Sequence Properties											
Name	Sequence Features	Kyte-Doolittle Hydropathy	Free Energy Hydropathy	Mol Weight	Isoelectric Point	Charge	MolExtinctCoeff	MolExtinctCoeffSS	ExtinctCoeff	ExtinctCoeffSS	IEIB
1JFQ	<a href="#">36 Features</a>	<a href="#">Show Plot</a>	<a href="#">Show Plot</a>	47269	8.28	6.0	69790	70290	1.476	1.487	0.956

Legend	
Sequence Features	Link to open a data table in DS of information on the features found
Kyte-Doolittle Hydropathy	Link which opens data table and creates Kyte-Doolittle hydropathy plot in DS
Free Energy Hydropathy	Link which opens data table and creates Free Energy hydropathy plot in DS
Mol Weight	Molecular weight of the protein
Isoelectric Point	Isoelectric point calculated by EMBOSS pepstats
Charge	Net charge
MolExtinctCoeff	Molar extinction coefficient (reduced form) at A280
MolExtinctCoeffSS	Molar extinction coefficient (with disulfide bridges) at A280
ExtinctCoeff	Extinction coefficient 1mg/ml (reduced form) at A280
ExtinctCoeffSS	Extinction coefficient 1mg/ml (with disulfide bridges) at A280
IEIB	Improbability of expression in inclusion bodies

图 4

点击表格中 Sequence Features 栏的 36Features 链接，会打开一张表格。(图 5)

该表格中显示了该蛋白序列中预测出的 36 个 sequence feature 及每个 feature 所对应的序列。

	Feature Type	Start	End	Motif Name	Motif	Matched Residues	
14	Antigenic Region	340	351				
15	Antigenic Region	357	375				
16	Antigenic Region	381	409				
17	Antigenic Region	412	427				
18	Conserved Resi***	23	23	Conserved C:23			
19	Conserved Resi***	35	35	Conserved W:41			
20	Conserved Resi***	73	73	Conserved L:89			
21	Conserved Resi***	88	88	Conserved C:104			
22	Conserved Resi***	98	98	Conserved F:118			
23	Conserved Resi***	236	236	Conserved C:23			
24	Conserved Resi***	250	250	Conserved W:41			
25	Conserved Resi***	295	295	Conserved M:89			
26	Conserved Resi***	310	310	Conserved C:104			
27	Conserved Resi***	325	325	Conserved W:118			
28	Sequence Motif	55	56	Deamidation	[NQ][SGH]	QS	
29	Sequence Motif	90	91	Deamidation	[NQ][SGH]	QQ	
30	Sequence Motif	95	95	Hydroxylation	P	P	
31	Sequence Motif	220	221	Deamidation	[NQ][SGH]	QS	
32	Sequence Motif	246	247	Deamidation	[NQ][SGH]	NG	
33	Sequence Motif	257	258	Deamidation	[NQ][SGH]	QG	
34	Sequence Motif	269	270	Deamidation	[NQ][SGH]	NG	
35	Sequence Motif	321	321	Glycation(Lys)	K	K	
36	Sequence Motif	327	328	Deamidation	[NQ][SGH]	QG	

图 5

在 Report 页面中点击 View Results。

会打开两个窗口。(图 6)

其中一个窗口为序列窗口，不同 sequence feature 对应序列用不同颜色进行标注；另一窗口用色块标注出了所有预测的 sequence feature。这两个窗口可以交互式对应，如在下图右侧窗口中选中相应 sequence feature，则左侧窗口中相应序列即被选中。

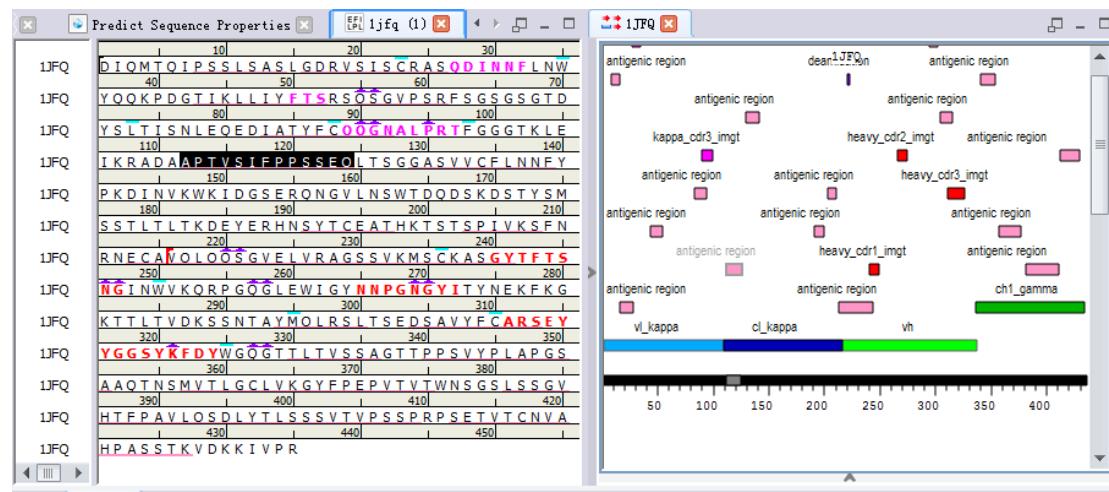


图 6



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

蛋白质结构预测专题:  
球蛋白建模、跨膜蛋白建模、抗体建模

## 基于蛋白序列构建同源蛋白模型（MODELER）教程

**目的：**通过此教程，了解 Discovery Studio 中基于蛋白序列构建同源蛋白模型的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS Sequence Analysis, DS MODELER, DS Protein Families, DS Protein Health。

**所需数据文件：**P41131.fasta。

**所需时间：**1 小时

### 介绍

蛋白质结构的解析对其功能的理解至关重要。然而，由于技术手段的限制，利用实验方法（主要为 X-ray, NMR）解析蛋白质结构投入大、周期长、风险大。对于某些膜蛋白，只利用现有技术条件，其结构甚至无法解析。另一方面，随着分子生物学技术的成熟及高通量测序技术的发展，越来越多的基因序列可以被轻松找到。这造成了现代蛋白质科学中一个奇怪的现象：蛋白质序列数据的累积量及积累速度远远超过蛋白质结构的解析。这种序列与结构间不平衡的现象极大地限制了我们对蛋白质功能及其相关作用机理的理解。因此，我们需要一种能够简单、快速且相对准确的技术来预测蛋白质的空间结构。

同源建模技术可以很好的解决上面的问题。该方法利用信息技术的手段，可以直接从蛋白的一级结构（氨基酸序列）预测蛋白质的高级结构（主要为三级结构）。根据最新一届国际建模大赛（CASP）的分类，目前主要的蛋白质建模方法包括两种：基于模板的建模（Template-based Modeling）和自由建模（Free Modeling）。前者又包括两种方法：同源建模法（Homology Modeling）和“穿线法”（Threading）。后者主要以从头计算法（ab initio）为主。所有的建模方法中，以同源建模法(Homology Modeling)使用最为广泛，预测结果的准确性最为可靠。

同源建模的理论基础为蛋白质三级结构的保守性远远超过一级序列的保守性。因此，人们可以通过使用一个或多个已知结构的蛋白（模板蛋白，template）来构建未知结构蛋白（目标蛋白，target）的空间结构。

Discovery Studio 为用户提供了一整套利用 Homology Modeling 方法自动预测蛋白质空间结构的工具。用户只需要提供蛋白质的氨基酸序列就可以轻松完成同源模型的构建及模型可信度评估的工作。DS 的 Homology Modeling 主要基于 MODELER 程序。目前 MODELER 已成为使用最为广泛，预测最为准确的同源建模工具之一。其主要的建模步骤包括：

1. 使用序列相似性搜索工具 BLAST 或 PSI-BLAST 搜寻目标序列的模板
2. 使用结构比对方法将模板进行比对、叠合
3. 使用序列比对方法将目标序列与模板结构的序列进行比对
4. 使用 MODELLER 产生目标序列的模型
5. 模型的评估

在识别目标序列的模板以及比对目标序列和模板结构的序列时，具体采用何种策略依赖于目标序列和模板序列间的同源性高低：

- ◆ 当序列同源性很高（一般大于 60%）时，BLAST 可以轻易识别出正确的模板，并且序列间的比对结果也很好。
- ◆ 当序列同源性不是很高，但仍在模糊区之上（一般介于 25% 和 60% 之间）时，BLAST 仍能够有效地识别出正确的模板，但是，简单的序列比对可能不能为同源建模产生最优的比对结果。在该情况下，序列比对结果可以通过利用无冗余序列库（non-redundant sequence database）中的同源序列创建 sequence profile 来改善。该

方法是最为常见的流程，该教程也会给出相应的例子来介绍该方法。

- 当序列同源性非常低（低于 25%）时，采用迭代搜索方法（PSI-BLAST）来搜寻模板，并且利用 sequence profile 来比对序列。

本教程中，以一个胞外淀粉酶的同源模建过程为例子，展示如何在 DS 中采用上面介绍的第二种方法来为该淀粉酶自动构建同源模型，并对所构建的模型进行评估，帮助大家获得 Homology Modeling 最直观的结果。

本教程包括以下步骤：

- 模板识别
- 目标序列和模板序列的比对
- 3D 模型的构建（MODELER）
- 3D 结构可靠性的评估

## 模板的识别

构建同源模型的第一步是基于序列相似性从已知的蛋白结构中识别出一个或多个模板蛋白。通常采用序列相似性搜索程序 BLAST 来完成该目的。进行 BLAST 搜索时，数据库可以使用 Protein Data Bank (PDB) 数据库也可以用 PDB\_nr95 (PDB 非冗余结构数据库)。为缩短搜索时间，通常使用 PDB\_nr95 数据库来寻找模板。

### 1. 载入序列

在文件浏览器 (Files Explorer) 中，展开 Samples | Tutorials | Protein Modeling 文件夹，双击打开 P41131.fasta 序列文件。

序列 P41131 在序列窗口中显示。(图 1)

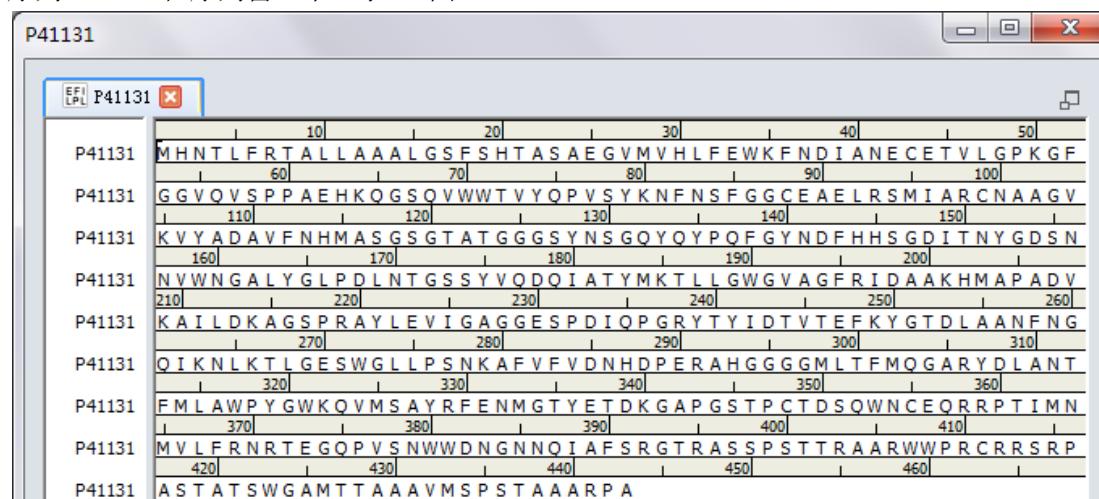


图 1 P41131 序列窗口

### 2. 搜寻模板

在工具浏览器 (Tools Explorer) 中，展开 Macromolecules | Search Sequences by Similarity，点击 BLAST Search (DS Server)，打开 BLAST Search 对话框。

在对话框中，点击 Input Sequence 参数右边的栅格，下拉列表中选择 P41131:P41131。

*Input Sequence* 中的文件名为 sequence window 的名字 (P41131) 与该窗口中的序列名称 (P41131) 的名字组合。

**点击 Input Database 参数右边的栅格，下拉列表中选择 PDB\_nr95。（图 2）**

该步骤指定了 BLAST 搜索的数据库为 PDB\_nr95 数据库，即序列同源性在 95% 时无冗余的信息。

**注：PDB\_nr95 序列数据库已经安装在 DS server 上。如果需要用 BLAST 搜索其它数据库，用户需要另外安装相应的数据库。**

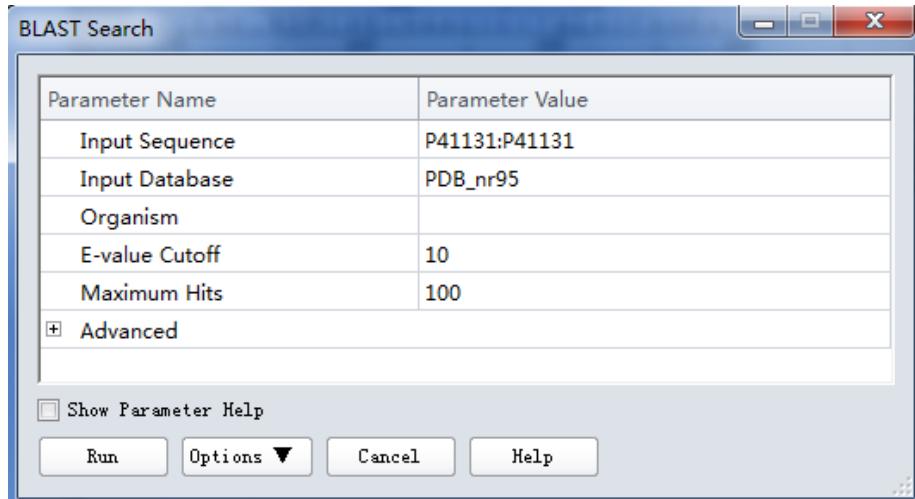


图 2 “BLAST Search”参数设置

**点击 Run 运行作业，等待作业完成。**

该作业大概需要一分钟的时间（奔腾 4，2.8GHz 的 CPU，2GB 的内存）。

待作业完成后，BLAST 搜索得到的序列会自动显示在 *BLAST Window* 中。

同时，弹出一个消息框“P41131 found 62 BLAST hits”。

**注：如果改动默认参数或使用不同的（或升级版）的 PDB\_nr95 数据库，BLAST 结果可能与本教程的结果不一致。**

**点击 OK 以关闭消息框。**

**在 P41131 - Blast 窗口，点击该窗口下的 Table View 标签。**

Table View 显示了命中的蛋白序列。每行表示一条命中的氨基酸序列。在 DS 中，灰色的 cells 是不能被编辑的。（图 4）

	Title/Description	Accession	Identity	Sequence Length	Alignment Length	Bit Score	E-value
1	ALPHA-AMYLASE...	1G94_A	48	448	381	317.39	4.44945e-103
2	Pancreatic al...	4GQR_A	44	496	425	295.049	8.68268e-94
3	ALPHA AMYLASE...	1HKO_A	44	496	425	291.582	1.83128e-92
4	ALPHA-AMYLASE...	1JAE_A	43	471	405	286.96	5.00318e-91
5	alpha-amylase...	3VM5_A	42	499	425	281.567	1.4827e-88
6	alpha-amylase...	3DCO_A	31	422	188	88.1965	5.70176e-19
7	Alpha-amylase...	1UAT_A	30	422	188	86.6557	1.80249e-18
8	CYCLODEXTRIN ...	1D3C_A	23	686	264	55.8398	1.80969e-08
9	CYCLODEXTRIN ...	1PAM_B	24	686	264	54.6842	4.28589e-08
10	Cyclomaltohex...	3BMW_A	24	683	266	54.299	5.555e-08
11	CYCLODEXTRIN ...	1CGT_A	22	684	231	46.595	1.39397e-05
12	4-ALPHA-GLUCA...	1LWJ_B	28	441	146	46.2098	1.67069e-05
13	Cyclodextrin ...	4JCM_A	25	669	171	45.4394	4.05439e-05
14	Cyclomaltohex...	4JCL_A	24	687	276	45.0542	5.1203e-05
15	Alpha-amylase...	3VMT_A	26	470	175	44.2838	6.52914e-05

图 4 BLAST 窗口的表格视图

**注：**命中的序列按照 E 值（序列无缝比对存在偶然性的可能性大小，表征了序列比对的可行度）进行排序。E 值最低的序列，结果最可靠，排在第一行。

点击 P41131 - Blast 窗口下的 Map View 标签。

Map View 将命中结果都显示在一张图中，每条线表示一条序列。每根横条根据 bit score 打分不同而配以不同的颜色（分数超过 400 为红色，最佳的命中结果）。本例中的目标序列 P41131 显示在窗口的最上方，为一条长度为 443 个氨基酸的直线。（图 5）



图 5 P41131 BLAST 窗口

用户可以将鼠标放置在某一个命中序列上，如下信息将会显示（如图 5 所示）：

- 序列数据库的描述
- 序列的编号
- 所属有机体的种类
- 目标序列中的起始氨基酸位置
- 数据库中命中序列的起始氨基酸位置
- 命中序列的长度
- 命中序列的分数
- 与目标序列的同源性

滑动鼠标的滚轮可以横向放大（缩小）Map View 中的结果

这时用户可以看到窗口顶端 target 的相应氨基酸。可能需要放大几次才能看见具体的氨基酸类型。（图 6）



图 6 P41131 BLAST 窗口

注: Map View 中命中序列的顺序并没有改变, 同 Table View 中的一致。

### 3. 模板的选取

本教程要构建蛋白序列为 P41131 的 3D 模型。为了达到这一目的, 我们需要挑选一个或多个合适的同源模板 (templates)。一个理想的 template 需要涵盖整个 target 的长度, 具有较高的序列等同性 (Sequence identity), 并且 E 值要够小 ( $< 1 \times 10^{-5}$ )。根据上述原则, 我们选用前 4 个命中序列作为 template。

一般而言, 若有多条模板与 target 都具有相似的同源性, 但模板之间相似性并不非常高, 那么通常使用多模板来构建同源模型。这些模板在核心区域一般都具有非常相似的结构, 而在一些 loop 区的长度和构象上可能有所不同。目标序列可能与其中某个模板在某 loop 区叠合的非常好, 而与另一个模板在另一 loop 区匹配的比较好。因此, 使用多模板可使建模过程中模型的每个部分都采用最合适的模板。

点击 P41131 - Blast 窗口。

点击 Map View 标签, 按住 SHIFT 键同时点击前五个命中序列 1G94\_A, 4GQR\_A, 1HX0\_A, 1JAE\_A, 3VM5\_A 以将其全部选中。

点击右键, 选择 Load Structure and Alignment。

这将在 DS 中打开两个新的窗口。一个为新的名为 1G94A 的分子窗口 (图 7), 包含了上述五个模板结构的 A 链以及结构中的水分子和配体分子。另一个为名为 P41131 (1) 的序列比对窗口 (图 8), 包含了目标序列 P41131 和五个模板结构的序列对比图。

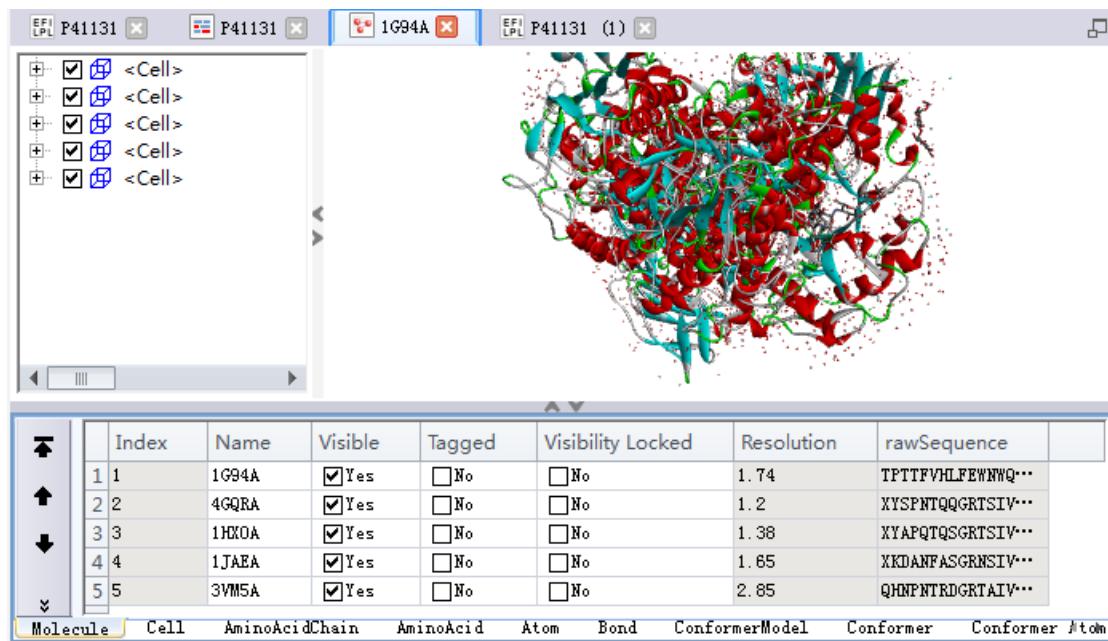


图 7 1G94 分子窗口

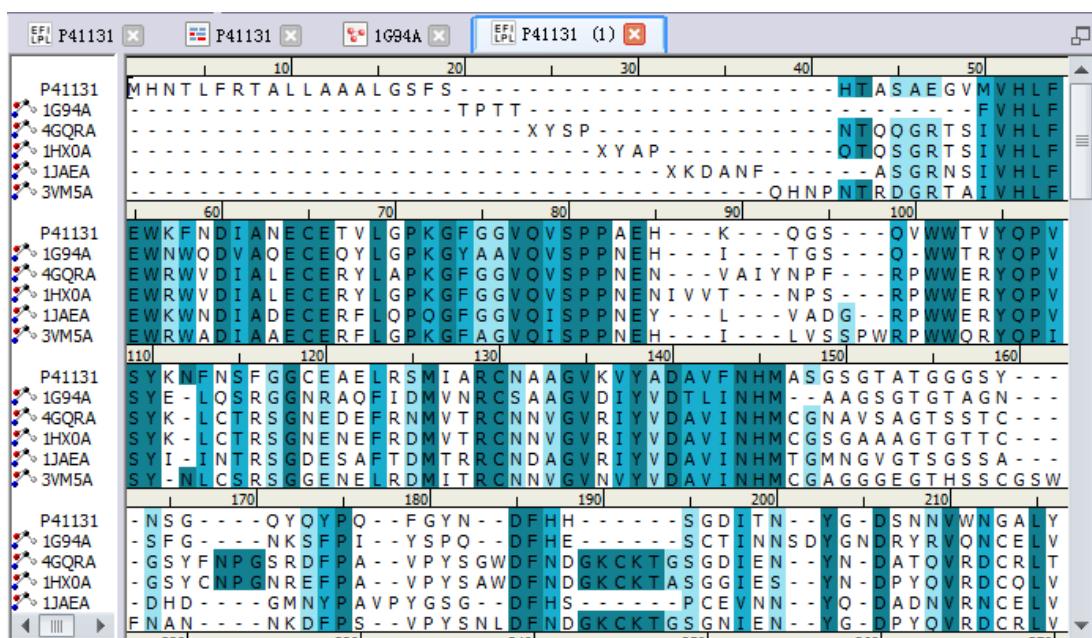


图 8 P41131(1) 序列比对窗口

在 Sequence Window 中，在菜单栏中点击选择 Sequence | Secondary structure | Visibility，选 PDB 和 Kabsch-Sander 选项，点击 OK。

在序列窗口中，五个模板序列的下方会出现根据 PDB 和 Kabsch-Sander 两种方法进行标识的相应序列的二级结构图。红色横条代表  $\alpha$  螺旋，蓝色箭头代表  $\beta$  折叠。

检查目标序列与五个模板序列间的比对结果。我们发现 C 端中约有 50 个氨基酸没有对比上。这样的比对结果并不适合于同源建模，需要对序列重新进行比对。下一步就展示如何将目标序列比对至模板序列上。

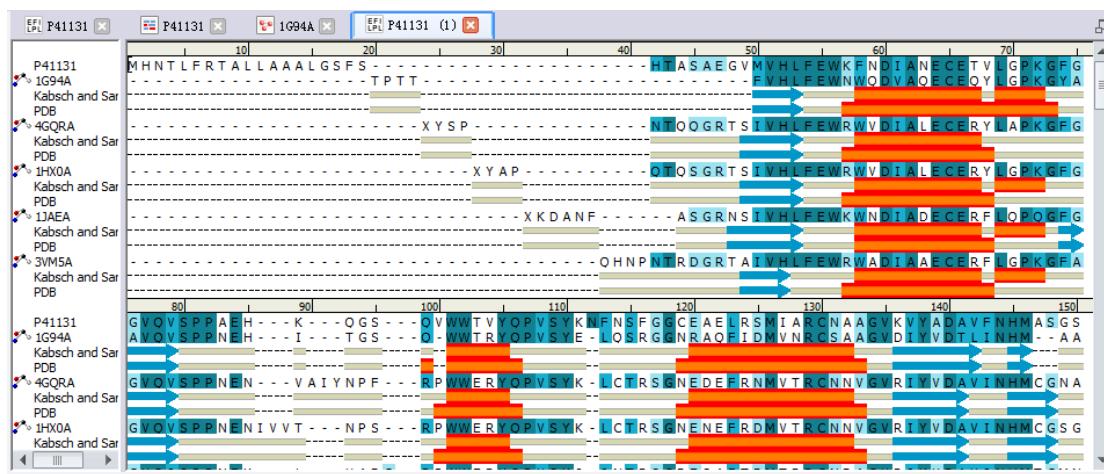


图 9 P41131(1) 序列比对窗口

### 将目标序列与模板序列进行重新比对

以下几种方法都可以用于将目标序列比对至模板序列：

- 通过多序列比对 (aligning multiple sequences) 将目标序列直接比对至模板序列
- 通过结构比对先将模板结构叠合，再通过 sequence 同 sequence profile 的比对将目标序列比对至比对好的模板序列。
- 通过结构比对先将模板结构叠合，再为目标序列生成一个 sequence profile，最后通过两个 sequence profile 间的比对将目标序列比对至比对好的模板序列。

接下来，我们将用 *Align Sequence to Templates* 工具将目标序列比对至所选的模板序列。你可以选择上述任一方法。在本例中，我们采用第二种方法，原因如下：

- 序列保守性和结构保守性通常会有细微的不同，这将会导致不同的结果。因此，对于同源模建，最佳方法是在模板结构同目标序列比对之前先基于结构的相似性将模板结构进行比对。

在工具浏览器 (Tools Explorer) 中，展开 Macromolecules | Create Homology Models，点击 Align Sequence to Templates，打开 Align Sequence to Templates 对话框。

点击 Input Model Sequence 右边的栅格，在下拉列表中选取 P41131-(1):P41131，点击 Input Template Structures 右边的栅格，选择 1G94A:All 确保 5 个模板结构都已选中。

其它参数都选择默认设置。（图 10）

点击 Run 运行作业，等待作业完成。

该作业大概需要五分钟的时间（奔腾 4，2.8GHz 的 CPU，2GB 的存储器）。

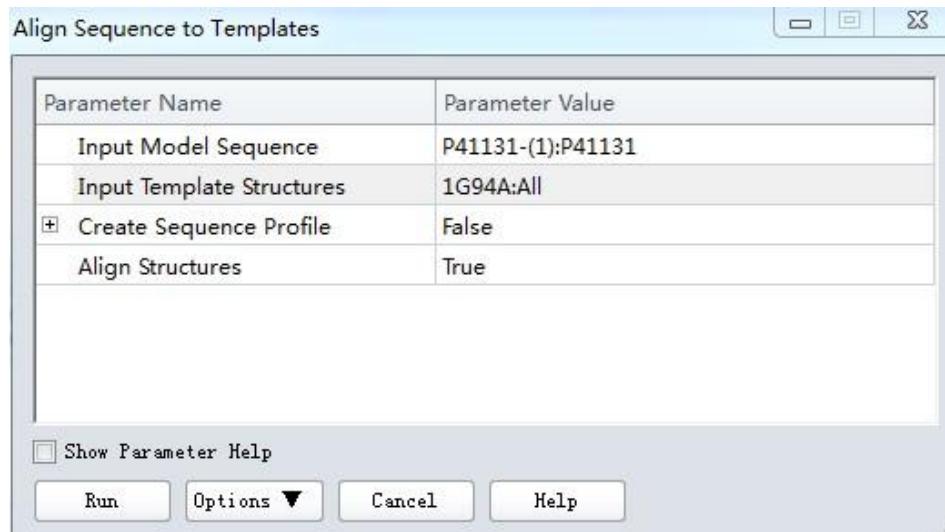


图 10 “Align Sequence to Templates”参数设置

待作业完成后，DS 自动打开两个新的窗口，一个是名为 P41131\_templates 的序列窗口，包含了目标序列同模板序列的比对结果（图 11），另一个是名为 1G94A(1) 的分子窗口，包含了五个叠合的模板结构（图 12）。在 report 界面中，Summary 栏里不仅显示了目标序列和模板序列之间的序列一致性（21.9%）和相似性（33.1%），还通过一张表格显示了五个模板结构两两主链之间的 RMSD 值及叠合的氨基酸残基数。

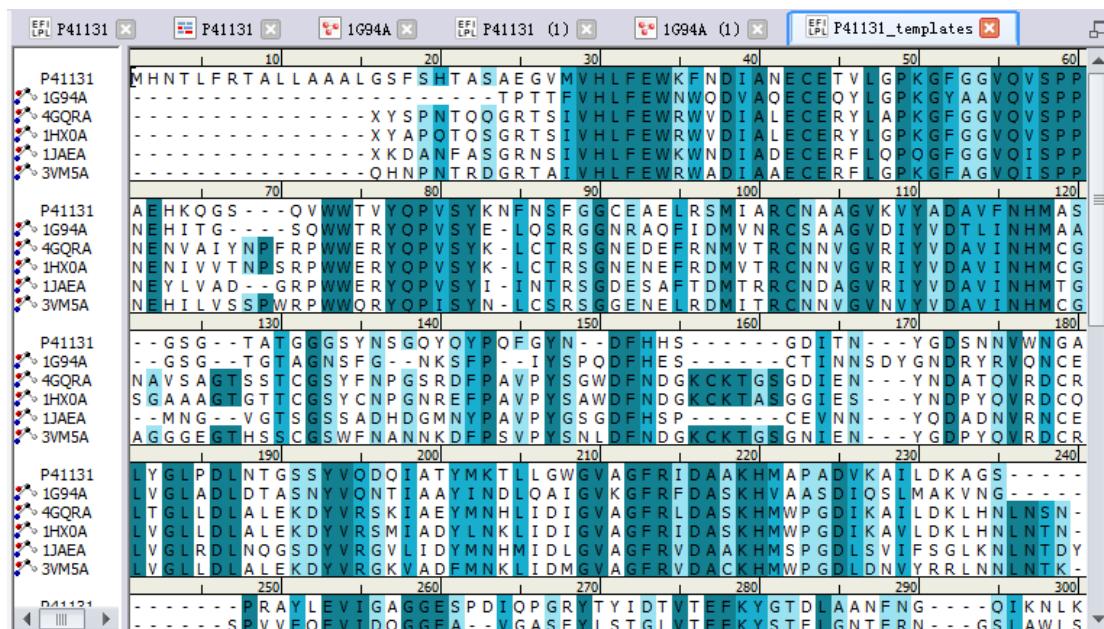


图 11 P41131\_templates 序列比对窗口

在上面的窗口中我们发现现在已经有更多的氨基酸被涂蓝，表示这些区域具有很强的序列一致性和相似性。

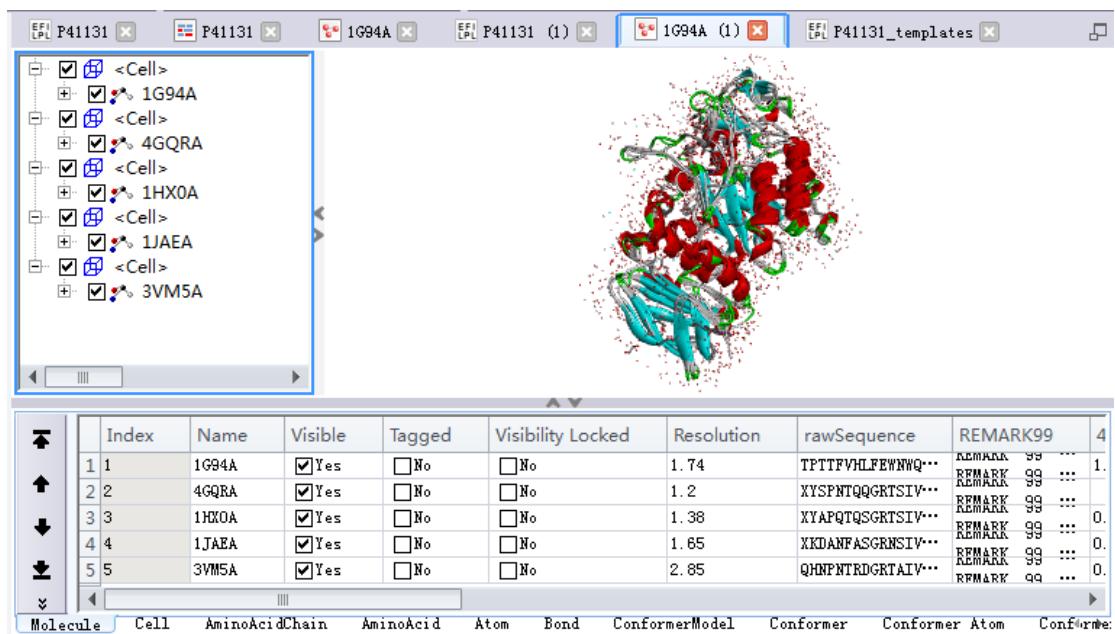


图 12 Superimposed 分子窗口

在上面的窗口中我们看到现在五个模板结构已经叠合在了一起。

### 使用 MODELER 构建目标序列的 3D 模型

在本小节中，我们将使用上小节所产生的比对结果以及四个模板结构来构建目标序列 P41131 的 3D 模型，在此基础上我们还可以将该模型同模板结构进行比较以及评估 MODELER 程序所产生的模型打分。

#### 1. 构建模型

点击名为 **1G94A(1)** 的分子窗口以将该窗口激活。

在视图窗口 (Graphics View) 中，点击鼠标右键，选取 **Show All**。

在工具浏览器 (Tools Explorer) 中，展开 **Macromolecules | Create Homology Models**，点击 **Build Homology Models**，打开 **Build Homology Models** 对话框。

点击 **Input Sequence Alignment** 右边的栅格，下拉列表中选取 **P41131\_templates:All**。

确保 **Input Templates Structures** 一栏中，**1G94A, 4GQRA, 1HXOA, 1JAEA, 3VM5A** 这五个模板蛋白都被选中。

点击 **Input Model Sequence** 右边的栅格，下拉列表中选取 **P41131**。

将 **Number of Models** 设为 **2**。

点击 **Optimization Level** 右边的栅格，下拉列表中选取 **Low**。（图 13）

注：将 *Optimization Level* 由默认值改为 **Low**，可以加快计算速度，但产生的模型的精度会下降。

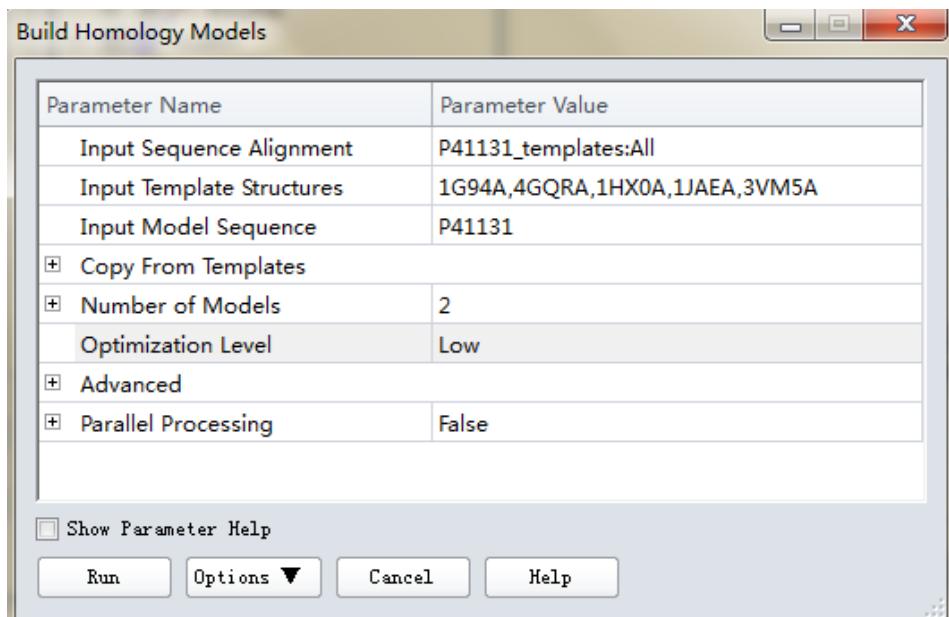


图 13 “Build Homology Models”参数设置

点击 Run 运行作业，等待作业完成。

该作业大概需要八到十分钟的时间（奔腾 4.2GHz 的 CPU，2GB 的存储器）。

待作业完成后，DS 自动打开两个新的窗口，一个是名为 P41131\_templates(1)的序列窗口，包含了两个模型序列同模板序列的比对结果（图 14），另一个是名为 P41131 的分子窗口，包含了五个叠合的模板结构和两个模型结构（图 15）。同时弹出一个消息框，显示了产生模型的个数及最佳模型相应的最低 PDF 总能量，点击该窗口中的 Report。

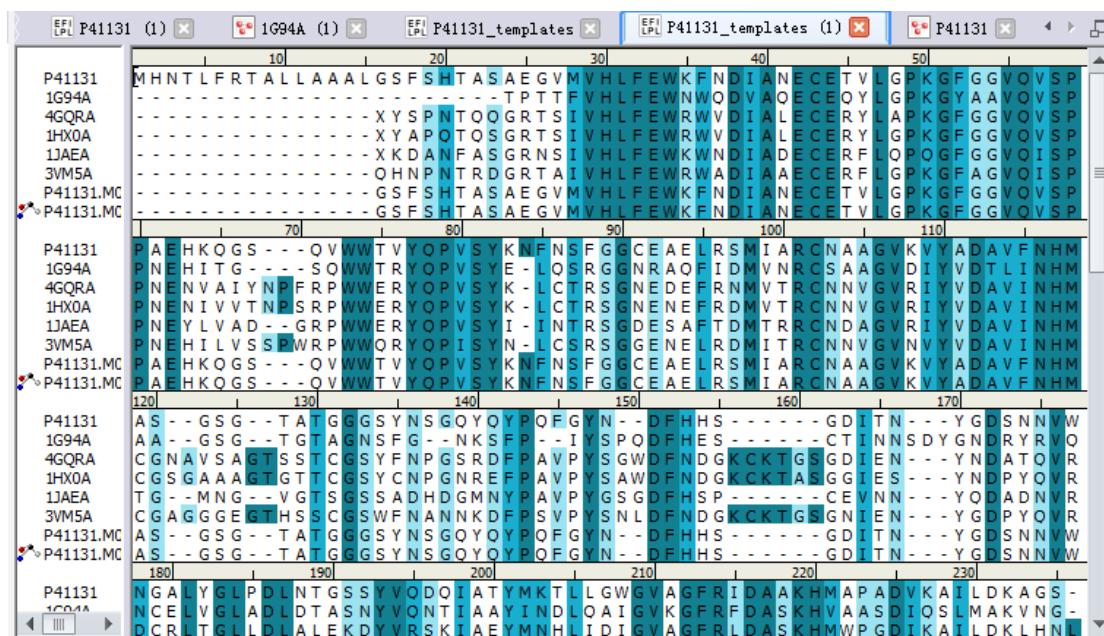


图 14 P41131\_templates(1) 序列比对窗口

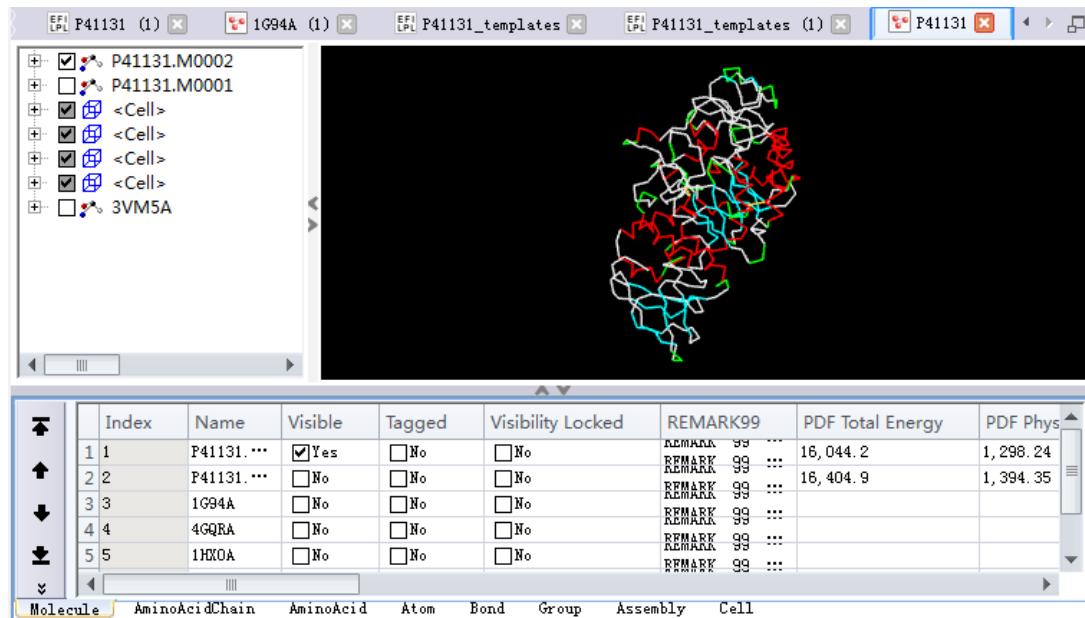


图 15 P41131 分子窗口

## 2. 根据 PDF 值或者 DOPE 值挑选最优模型

在打开的 Report 窗口（图 16）中，Summary 一栏综述了所构建模型的打分。模型的排名依据是 **PDF Total Energy**。建模过程中，DS MODELER 首先会提取模板（template）的几何特性，然后使用 PDF（probability density function）函数来定义蛋白结构中诸如键长、键角、二面角等几何特性。接着它会对 PDF 函数施加一定的约束条件，并以此来构建 target 的 3D 结构。所以 PDF 的函数值可以直接反应所构建模型的好坏。一般，模型的 PDF Total Energy 越低，表明该模型在同源约束条件下优化的越好；模型同限定的同源约束条件偏差越小，该模型的可信度越大。然而，如果构建的模型其 PDF Total Energy 相似，则可以利用基于原子统计势能的 **DOPE score** 作为衡量模型质量的依据。DOPE 是一个基于原子统计势能的程序，主要用于模型评估。它的分数可以认为是衡量同一分子不同构象可信度的标准，能够帮助选择预测结构的最优模型。分数越低，模型质量越可靠。

Model Scores			
Name	PDF Total Energy	PDF Physical Energy	DOPE Score
P41131.M0002	16044.2354	1298.2367995	-42544.765625
P41131.M0001	16404.9355	1394.3457499	-43791.621094

图 16 Report 窗口

注意: DOPE score 挑选的最优模型与 PDF Total Energy 挑选的最优模型不一致时, 可以使用其他的模型评估软件进行进一步的评估分析以选取较合理的初始模型。当没有其他模型评估软件可以使用时, 也可以粗略的选取 PDF Total Energy 最低的模型作为最合理的初始模型。本教程中选用 PDF Total Energy 最低的初始模型 P41131.M0002 作为最优模型。

点击 **P41131** 分子窗口以将该窗口激活。

在系统视图 (Hierarchy View) 中, 展开 **P41131.M0002 | Modeler Groups**。

改组下有两组数据, 分别是 *Identical* 和 *Non-Identical*。前者包含了同模板序列类型一致的氨基酸残基, 后者包含了同模板序列不一致的氨基酸残基。如果氨基酸残基没有同任何模板残基比对上, 则该残基单独分为一组, 命名为 *Insertion*, 该组包含了没有同模板残基比对上的保守残基。

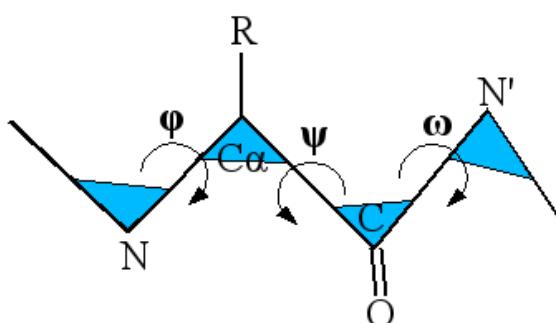
## 模型评估

模型构建完成后, 一般需要对其进行评估。DS 为用户提供了多种模型评估方式。主要包括: Ramachandran plot 和 Profile-3D。

在使用下列模型评估程序时, 关除 P41131.M0002 窗口外所有的窗口。如提示时候保存时, 选择否。

### 1. 使用 Ramachandran Plot 评估模型

Ramachandran plot 用于阐述蛋白质或肽立体结构中肽键内  $\alpha$  碳原子和羧基碳原子间的键的旋转度 (psi) 对  $\alpha$  碳原子和氮原子间的键的旋转度 (phi), 主要用来指明蛋白质或肽类中氨基酸的允许和不允许的构象 (如下图)。



通过对已知晶体结构的统计分析, 人们可确定氨基酸在 Ramachandran plot 中经常出现的区域。使用该图, 用户可以确定结构中每个氨基酸的构象是否正确。

从主菜单中, 选取 **Chart | Ramachandran Plot**, 显示 P41131.M0002 模型的拉氏图。 (图 17)

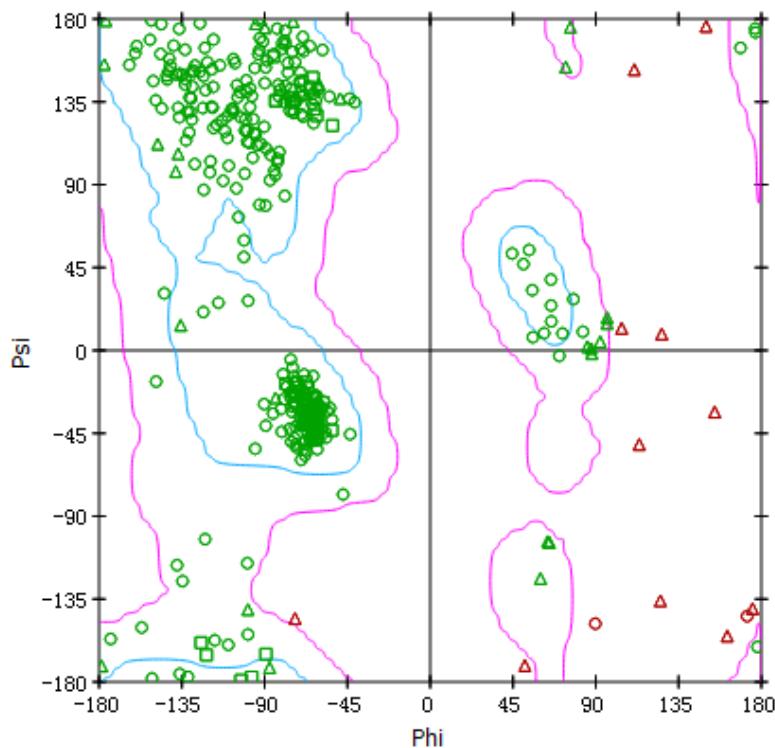


图 17 拉氏图

蓝色区域为“最适区”，该区域含有的氨基酸个数越多，结构越可信；紫色区域为“允许区”；其它区域的点（红色点）为 psi-phi 构象不合理的氨基酸，需要优化。

## 2. 使用 Profile-3D 评估模型

Profile-3D 是 UCLA 的 David Eisenberg 教授开发的一种基于“穿线”（threading）法的模型评估程序。该方法采用 3D-1D 的打分函数来检测所构建模型与自身氨基酸序列的匹配度关系。分数越高，说明同源模型的可信度越大。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Create Homology Models，点击 Verify Protein (Profile-3D) 打开 Verify Protein (Profile-3D) 对话框。  
点击 Input Protein Molecules 右边的栅格，选择 P41131:Visible。（图 18）  
点击 Run 运行作业，等待作业完成。

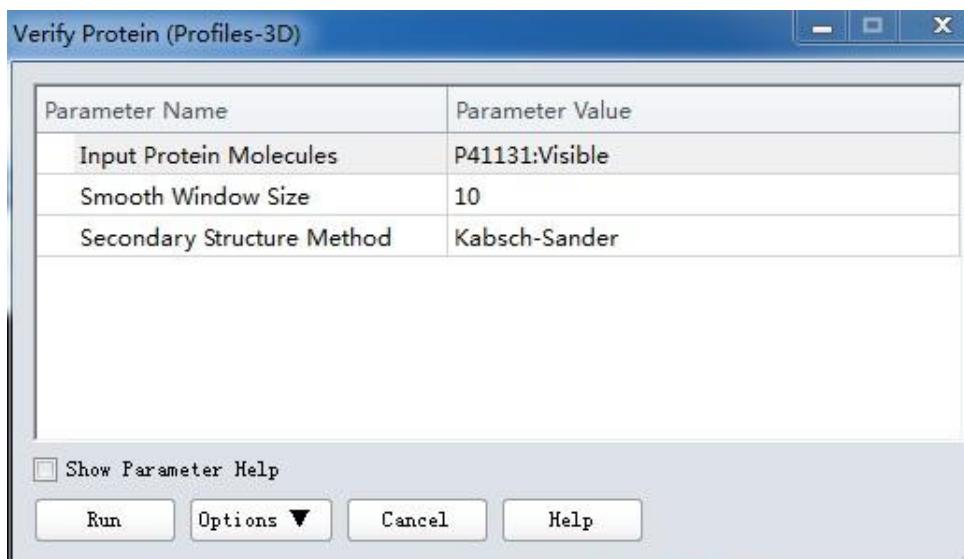


图 18 “Verify Protein (Profile-3D)”参数设置

该作业大概需要一分多钟的时间（奔腾 4, 2.8GHz 的 CPU, 2GB 的存储器）。待作业完成后，自动弹出一个显示了蛋白打分的消息框。（图 19）



图 19 消息框

点击 **OK**, 关闭该消息盒子。

点击 P41131 分子窗口，使得该窗口处于激活状态。

分子窗口中，Verify Protein 的结果打分会自动更新至表格浏览器（Data Table）中。

在 **Data Table View** 中，点击 **Molecule** 标签，滚动 **table** 至 **Verify Expected High Score**, **Verify Expected Low Score**, 以及 **Verify Score** 栏处。

如果模型的 **Verify Score** 高于 **Verify Expected High Score**，则模型的质量较高。**Verify Score** 越接近 **Verify Expected High Score**，模型的质量越好。

点击 **AminoAcid** 标签，滚动至 **table** 的尾列 **Verify Score**。

点击 **Verify Score** 一栏的表头以选取整列。

从主菜单中，选取 **Chart | Line Plot**。

这将打开一个新的图表，该图表显示了序列中每个氨基酸的打分。（图 20）

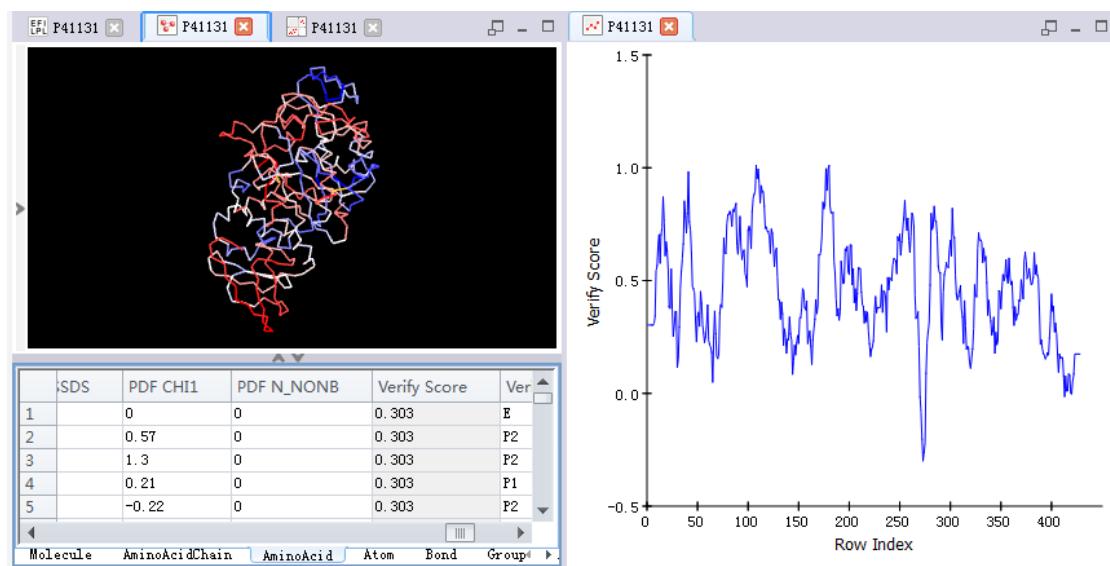


图 20 Profile-3D 图

下面我们来查看一下低分区的氨基酸。

将 **Line Plot** 窗口拖至分子窗口和序列窗口的并列位置，一同显示。

点击选中低峰周围区域，以选择 **Line Plot** 中对应的打分较低的氨基酸。

Profile-3D 图中选中的低分点（黄色）对应的氨基酸将在分子窗口和序列窗口高亮显示出来（黄色），以方便查看低分区域。

## 基于蛋白序列构建同源跨膜蛋白模型教程

**目的:** 通过此教程, 了解 Discovery Studio 中基于蛋白序列构建同源跨膜蛋白模型的操作方法及结果分析。

**所需功能和模块:** Discovery Studio Client, DS Sequence Analysis, DS MODELER, DS Protein Families。

**所需数据文件:** 2rh1.pdb, ADRB1\_HUMAN.fasta。

**所需时间:** 0.5 小时

### 介绍

膜蛋白在药物研发领域中是一类非常重要的靶标蛋白, 这是由于该类蛋白, 尤其是 GPCR (G 蛋白偶联受体), 是目前许多已知药物以及仍处于研发阶段的大量药物分子的目标靶点。GPCR 是一类七次跨膜受体蛋白, 它们广泛地参与细胞增殖、分化、迁移, 尤其是各类生理活动的调控。从近几十年药物发展的历史看, 全球 70% 所开发的药物都是针对 GPCR 的, 因此 GPCR 是令人瞩目的药物治疗靶点。

本教程即演示了如何使用 DS 中已有的一系列工具 (tools) 和流程 (protocols), 以 beta-2 肾上腺素受体 X-衍射晶体结构为模板, 来构建 beta-1 肾上腺素受体的同源模型并进行加膜处理。

本教程包括以下步骤:

- ◆ 输入 beta-1 肾上腺素受体的序列及 beta-2 肾上腺素受体模板结构
- ◆ 预测蛋白的跨膜区
- ◆ 将目标序列比对至模板序列
- ◆ 构建 3D 同源模型
- ◆ 添加隐性生物膜至构建得到的模型
- ◆ 调整生物膜的位置

### 打开输入文件

在文件浏览器 (Files Explorer) 中, 展开 Samples | Tutorials | Protein Modeling 文件夹, 双击打开 ADRB1\_HUMAN.fasta 序列文件。

这将在一个名称为 ADRB1\_HUMAN 新的序列窗口中打开 beta-1 肾上腺素受体的一级序列 (序列名为 ADRB1\_HUMAN)。

构建同源模型的第一步是识别合适的模板。如果同目标序列的序列一致性较高的模板存在, 通常可以通过简单的 BLAST 搜索得到该模板结构。对于膜蛋白而言, X-单晶衍射解析得到的晶体结构非常有限, 因此本教程不包含模板识别这一步骤 (该步骤细节可以参看 MODELLER 教程)。

本教程采用最近解析出来的 beta-2 肾上腺素受体晶体结构 (PDB 号: 2rh1) 作为模板。

在文件浏览器 (Files Explorer) 中, 展开 Samples | Tutorials | Protein Modeling 文件夹, 双击打开 2rh1.pdb 文件。

蛋白 2rh1.pdb 在分子窗口中显示。

注: 采用 BLAST 搜索 PDB\_nr95 数据库可以找到另一个可能的模板, 火鸡 beta-1 肾上腺素

受体。该晶体结构同目标序列的序列一致性高达 70.5%，是一个比较好的模板结构。然而，本教程采用 2rh1 作为模板是想突出 DS 的功能，因为在模板序列一致性如此高的情况下，软件功能在很大程度上是不太相关的。

建模的下一步：将目标序列与模板序列进行序列比对。首先需要将目标序列和模板序列置于同一个序列窗口当中。

**点击激活 ADRB1\_HUMAN 序列窗口。**

**在窗口中点击右键选择 Insert Sequence | From Windows...**

打开 Insert Sequence from Windows 对话框。

**选择 2rh1 点击 OK。**

将 2rh1 序列插入 ADRB1\_HUMAN 序列窗口当中。

**注：**此时两序列并未进行比对，状态栏（DS 窗口左下角）显示了两序列之间的序列一致性及相似性情况（分别为 7.0%，27.4%）。

## 预测蛋白的跨膜区

序列比对流程允许利用蛋白的二级结构信息来提高序列比对的精度。二级结构预测方法，如 DSC，是基于球蛋白的溶剂暴露模式来定义的。膜蛋白的表面大部分都暴露于生物膜内部的脂质环境中而非水环境。这就导致了氨基酸序列中疏水残基和亲水残基模式的不同，从而使得标准的二级结构预测方法来预测膜蛋白的二级结构不太可靠。本教程则采用专门用于预测膜蛋白螺旋区的 TransMem 方法。

**在工具浏览器（Tools Explorer）中，展开 Macromolecules | Analyze and Edit Transmembrane Proteins，点击 Predict Transmembrane Helices。**

在 ADRB1\_HUMAN 序列窗口中相应地添加了蛋白二级结构卡通图，红色横条表示 alpha-螺旋，蓝色箭头表示 beta-折叠。（图 1）

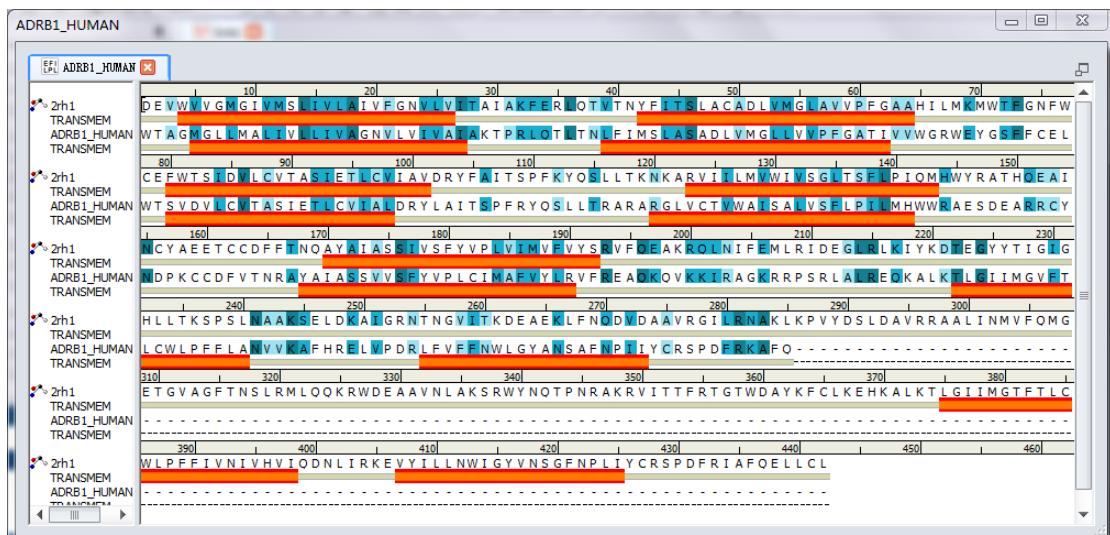


图 1

**注：**预测 2rh1 的跨膜螺旋时，有一段区域（ASN1002—TYR1161）是没有预测到螺旋结构的。该区域对应的是一个为方便结晶与  $\beta$ -2 肾上腺素受体融合的溶菌酶的序列。标准的二级结构预测方法可以预测出该区域的二级结构。如果将二级结构预测结果用于指导序列比对，如此大的序列插入会扰乱比对的结果。

## 将目标序列与模板序列进行比对

在工具浏览器（Tools Explorer）中，展开 **Macromolecules | Align Sequence and Structures**，点击 **Align Sequences...** 打开 Align Sequences 对话框。

确保 **Input Sequence Set** 设置为 **ADRB1\_HUMAN:All**。

点击 **Use Secondary Structures** 右边的下拉菜单，选择 **TRANSMEM**。（图 2）

点击 **Run** 运行作业，等待作业完成。

该作业大概需要 1.5min 的时间（奔腾 4，3GHz 的 CPU，1GB 的存储器）。

作业完成以后，**ADRB1\_HUMAN** 序列窗口会自动更新为序列比对之后的结果（图 3），同时弹出序列一致性和序列相似性（42.1%，53.3%）报告的对话框（图 3）。

点击 **OK** 关闭该对话框。

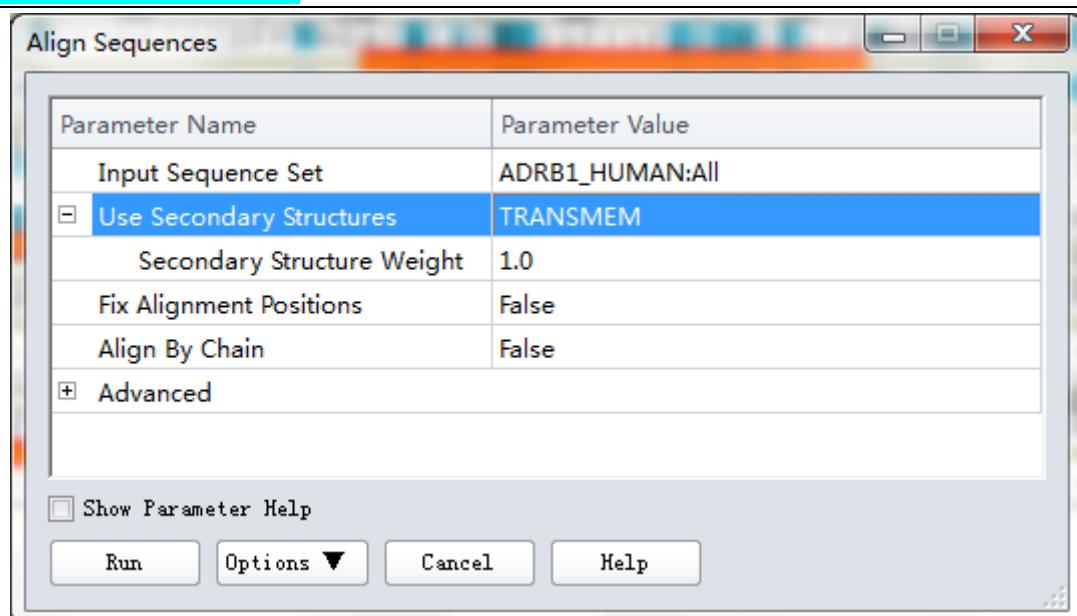


图 2“Align Sequences”参数设置



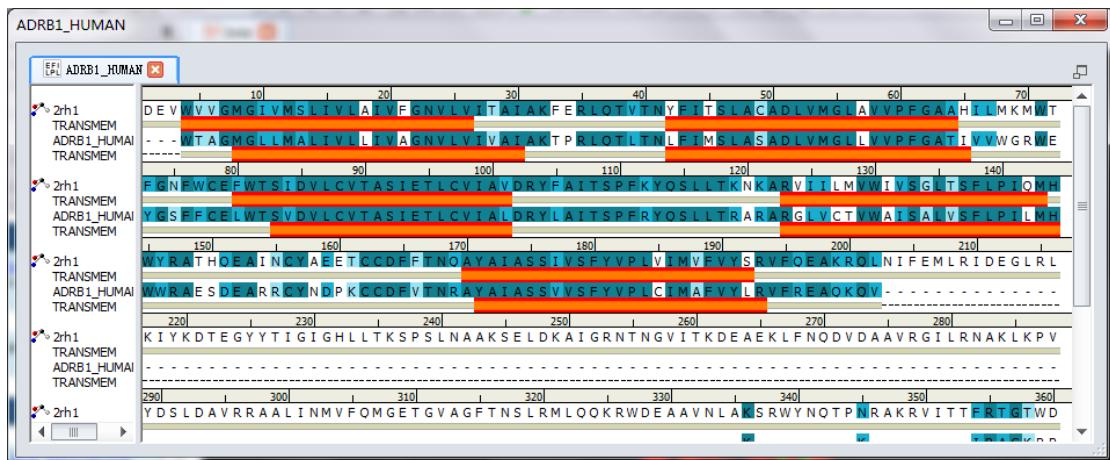


图 3

### 构建目标序列的 3D 同源模型

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Create Homology Models，点击 Build Homology Models，打开 Build Homology Models 对话框。  
点击 Input Sequence Alignment 右边的栅格，下拉列表中选取 ADRB1\_HUMAN:All。  
点击 Input Templates Structures 一栏，选择 2rh1。  
点击 Input Model Sequence 右边的栅格，下拉列表中选取 ADRB1\_HUMAN。  
展开 Copy From Templates 参数栏，点击 Ligands 参数栏，选择 2rh1:A:CLR412，  
2rh1:A:CLR413, 2rh1:A:CLR414, 2rh1:A:PLM415。

该步骤将棕榈酸和胆固醇分子从模板结构中直接复制到模型结构当中。这些分子有助于决定生物膜相对于模型结构的位置。

将 Number of Models 设为 1。

点击 Optimization Level 右边的栅格，下拉列表中选取 Low。（图 4）

注：将 Optimization Level 由默认值改为 Low，可以加快计算速度，但产生的模型的精度会下降。

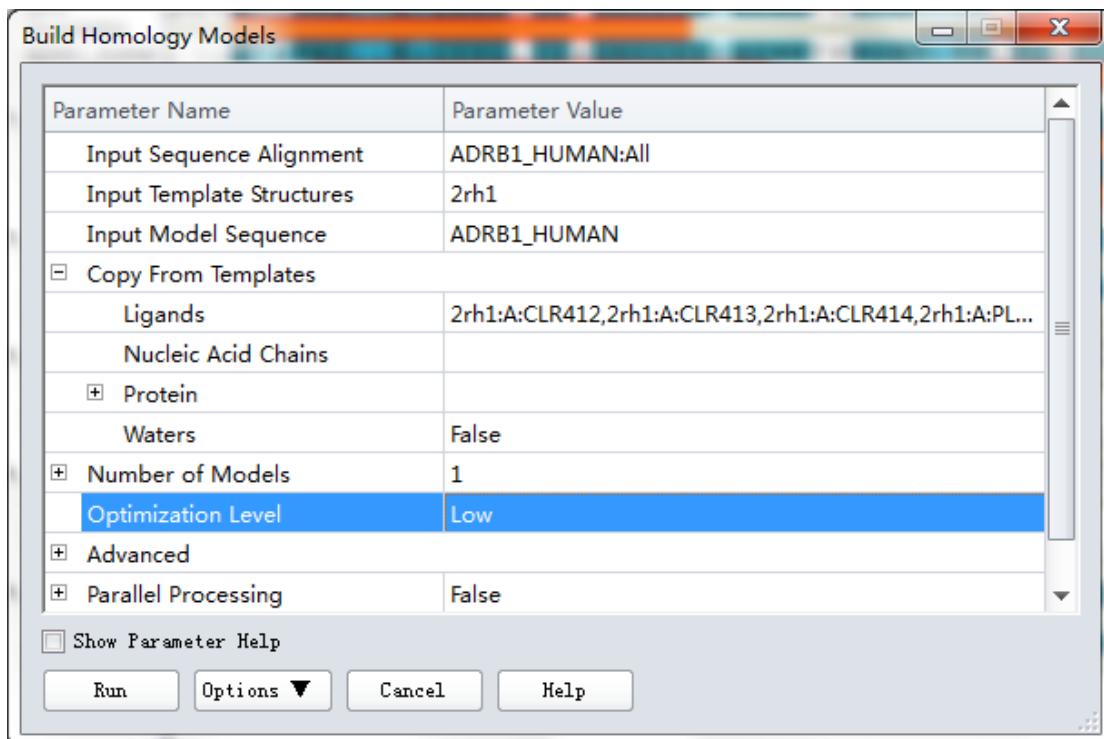


图 4“Build Homology Models”参数设置

点击 **Run** 运行该作业。

该作业大概需要 1min 的时间（奔腾 4，3GHz 的 CPU，1GB 的存储器）。

待作业完成以后，构建得到的模型在名为 ADRB1\_HUMAN 的分子窗口中打开（图 5），同时有一个报告 **PDF Total Energy** 的对话框弹出。

点击 **OK** 关闭该对话框。

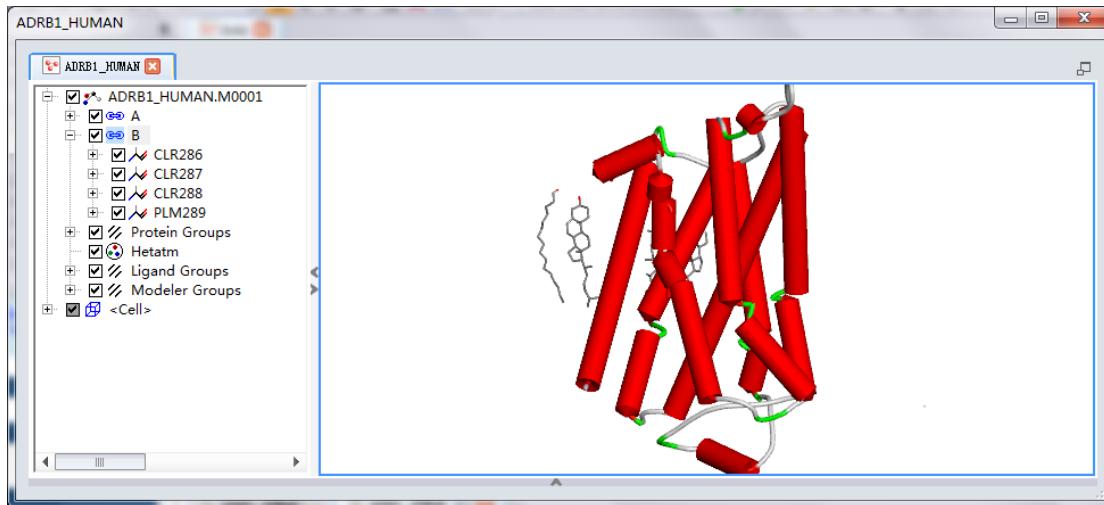


图 5

### 给模型添加隐性生物膜结构

DS 为用户提供了为蛋白质结构添加和控制隐性膜结构的工具。该膜结构的位置可以用在模拟（simulation）流程中以明确蛋白质在隐性溶剂模型中的溶剂化性质。

在菜单栏中，选择 **Edit | Preferences...** 以显示 Preferences 对话框。

在 Transmembrane Protein 页面中，确保 Membrane Thickness 的设置为 30 Å (图 6)。  
点击 OK 关闭 Preferences 对话框。

这里所说的 Membrane thickness 指的是膜的疏水区，不包括磷脂的极性头部所占据的区域。

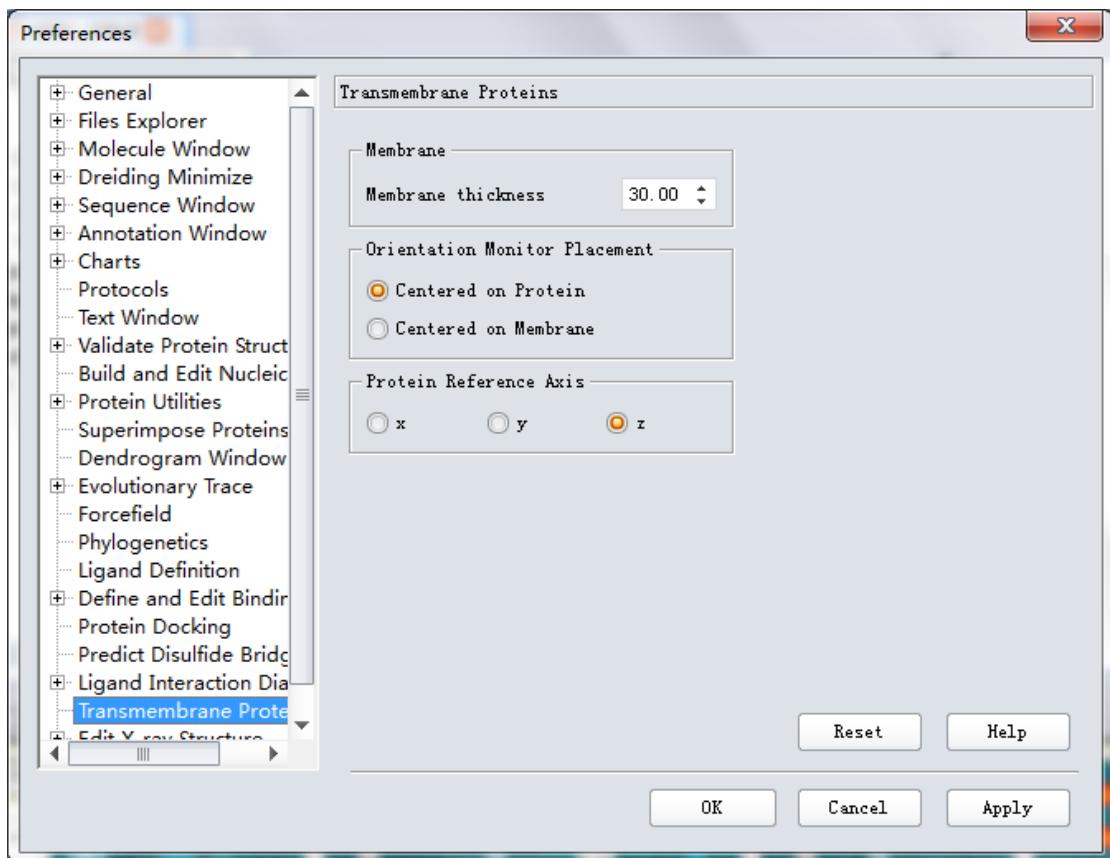


图 6

点击并激活 ADRB1\_HUMAN 分子窗口。

在工具浏览器(Tools Explorer)中，展开 Macromolecules | Analyze and Edit Transmembrane Proteins，在 Create and Edit Membrane 栏下点击 Add。

该步将为蛋白质结构添加一个膜结构。该膜在分子窗口的图形界面上以两个平行的平板表示 (图 7)。

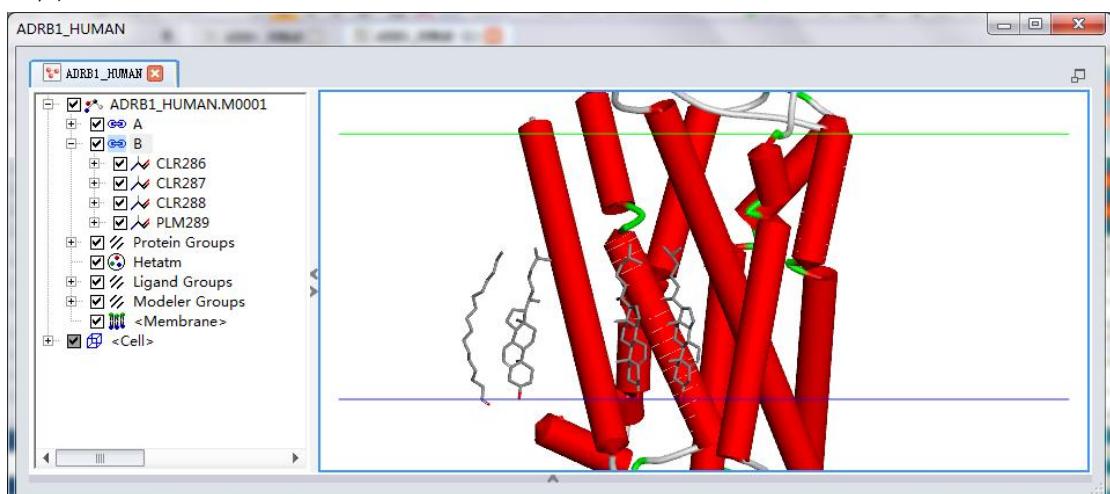


图 7

## 调整膜的位置

Add 命令只使用一个非常简单的溶剂模型来确定膜结构的位置。做这样的优化处理更多地是考虑到了速度而非精度。因此，通常都需要对膜的位置进行重新调整。

在系统（Hierarchy）视图中，展开 ADRB1\_HUMAN.M0001，点击选中 B 链以选中模型中的胆固醇和棕榈酸分子。

注意到，胆固醇和棕榈酸分子的极性区域都处于以两个平板所定义的膜结构的疏水区（图 8）。这表明初始构建的膜的位置需要调整。

使用 Add Membrane and Orient Molecule 流程可以自动完成上述膜结构位置的调整过程。该 protocol 使用广义伯恩（GB）溶剂模型对膜的位置进行全面优化。由于此步计算需要非常长的计算时间，所以本教程使用 **Analyze and Edit Transmembrane Proteins** 工具面板来调整膜结构的位置。

在系统（Hierarchy）视图中，展开 ADRB1\_HUMAN.M0001，点击选中 A 链。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Analyze and Edit Transmembrane Proteins，在 Create and Edit Membrane 栏下点击 Add Orientation Monitor。

这将在模型结构窗口中添加一个膜取向调节坐标（membrane orientation monitor），用以指示膜与蛋白质主轴间的移动（shift），倾斜（tilt）及旋转（rotation）角度（图 8）。

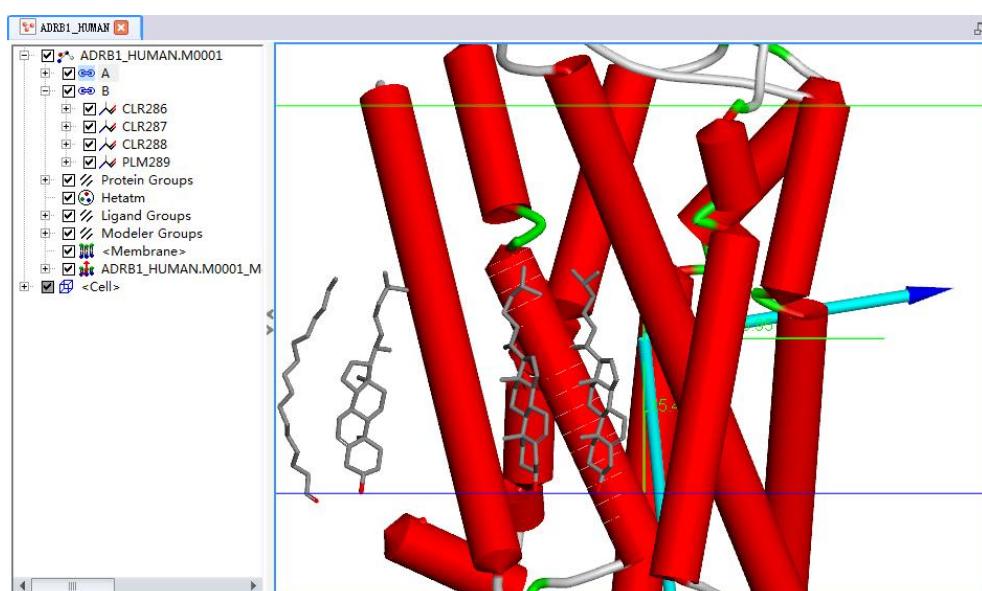


图 8

点击 Modify...

打开 Edit Membrane 对话框。

设定 Shift 为 4.8 Å，Tilt 为 15.5 度，Rotation 为 30.0 度。

点击 OK。

点击 3D 窗口中任一位置，从而不选中蛋白质。

调整后的膜结构中，胆固醇和棕榈酸分子的极性区都位于生物膜疏水区的外侧（图 9）。

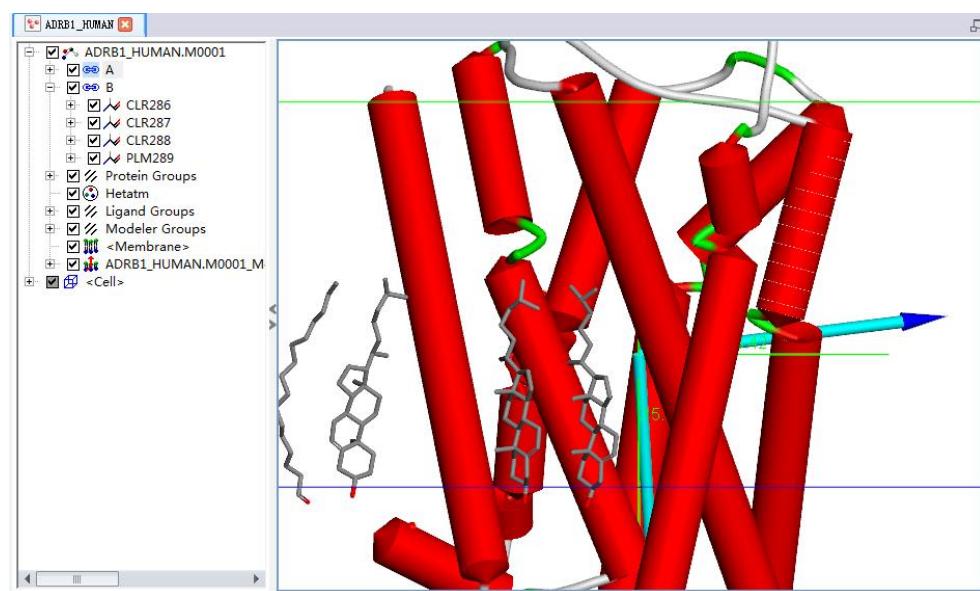


图 9

## 抗体 3D 结构的预测（MODEL ANTIBODIES）教程

**目的：**通过此教程，了解 Discovery Studio 中基于抗体序列构建抗体全长及 Framework 区模型的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS Sequence Analysis, DS MODELER, DS Protein Families, DS Validate Protein Structure。

**所需数据文件：**MA5\_H.bsml、MA5\_L.bsml

**所需时间：**1 小时

### 介绍

抗体分子是生物学和医学领域用途最为广泛的蛋白分子。以肿瘤特异性抗原或肿瘤相关性抗原、抗原独特型决定簇、细胞因子及其受体、激素及一些癌基因产物作为靶分子，利用传统的免疫方法或通过细胞工程、基因工程等技术制备的多克隆抗体、单克隆抗体、基因工程抗体广泛应用于疾病诊断、治疗及科学研究等领域，并以其毒副作用小、天然和高度特异性的疗效，创造出了巨大的社会效益和经济效益。

新版 Discovery Studio 在抗体设计领域进行了较大的更新，不仅可以方便用户构建杂合抗体 Framework 区的 3D 结构，而且还提供了构建抗体全长 3D 结构的工具。用户只需要提供抗体的氨基酸序列就可以轻松完成抗体同源模型的构建及模型可信度评估的工作。

本教程分别对上述两种抗体模型预测的方法（抗体 Framework 区 3D 结构和抗体全长 3D 结构的预测）进行介绍：

#### A. 抗体 Framework 区 3D 结构的预测，具体步骤如下：

- ◆ 载入序列
- ◆ Framework 区模板识别
- ◆ Framework 区模型的构建
- ◆ 模型可靠性的评估

#### B. 抗体全长 3D 结构的预测，即以免疫球蛋白 Immunoglobulin G 为模板分子，构建得到抗体的全长 3D 模型，具体步骤如下：

- ◆ 载入序列
- ◆ 抗体全长模型的构建

### A. 抗体 Framework 区 3D 结构的预测

#### 载入序列

在文件浏览器（Files Explorer）中，展开 Sample| Tutorials| Protein Modeling，双击打开 MA5\_H.bsml、MA5\_L.bsml 文件

上述两个序列依次分别为抗体 MA5 的重链（H）序列和轻链（L）序列。

#### Framework 区模板的识别

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Model Antibody，点击打开 Identify Framework Templates。

设置 Light Chain Sequence 为 MA5\_L:MA5\_L

设置 HeavyChain Sequence 为 MA5\_H:MA5\_H

其它参数默认，点击 Run 运行（图 1）

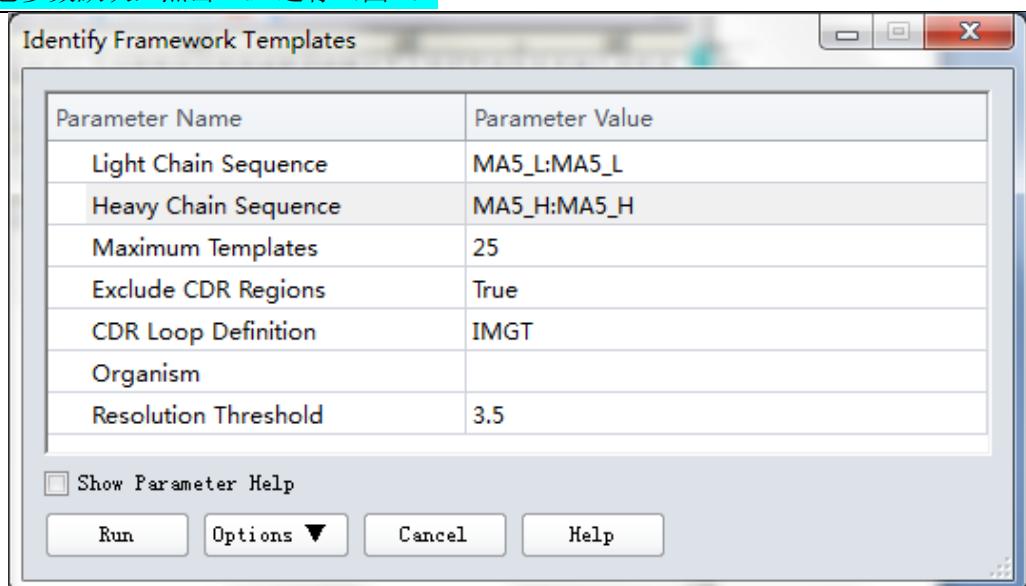


图 1

作业大概需要一分钟的时间。待作业完成后，点击 Report 页面，查看 Summary。报告的页面有四张表格，第一张表格是根据 Similarity 得到序列相似性最高的表面、轻链、重链模板，以下三张表格分别是搜索得到的所有的表面、轻链、重链模板。

抗体 Framework 区建模有两种方法：嵌合方式、单模板方式。嵌合方式能较全面的考虑轻链、重链模板的结构信息，同时根据表面模板决定轻链、重链的空间取向，但有时来自两条不同蛋白的轻链、重链模板与第三条表面模板叠合时，空间取向可能会不准确。无论嵌合方式或单模板方式建模，模型的质量都与搜索得到的表面模板或轻链模板、重链模板序列相似性有很大的关系，同时 CDR 区的序列相似性对 Framework 区结构质量也起很大的作用。

本教程采用单模板方式进行抗体 Framework 区建模，点击 Summary 下第二张表格 Overall Templates 中的 Identify (%) 两次，使其按照 Identify 降序排列，点击第二个 2XWT\_BA，载入结构（图 2）。

Identify Framework Templates									
Name	Similarity (%)	Identity (%)	Similarity (%) including CDRs)	Identity (%) including CDRs)	Resolution	VL_LAMBDA	VH	Organism	Redundant Hits
4M6N_LH	100.0	100.0	100.0	100.0	2	L:1-112	H:1-117	Homo sapiens	4M6N_LH_FV
2XWT_BA	94.4	89.4	84.8	76.5	1.9	B:1-110	A:1-119	Homo sapiens	
3G04_AB	93.3	86.6	83.6	75.9	2.55	A:1-111	B:1-121	Homo sapiens	
4BUH_AA	93.9	86.0	82.3	75.1	1.3	A:127-237	A:1-126	Homo sapiens	
4BUH_BB	93.9	86.0	84.4	77.1	1.3	B:120-230	B:1-120	Homo sapiens	
3MLW_LH	96.1	84.9	84.1	72.3	2.7	L:1-113	H:1-126	Homo sapiens	3MLW_MI_FV
4HQQ_LH	92.3	84.5	81.9	72.4	2.4	L:1-111	H:1-119	Homo sapiens	
4HPO_LH	91.7	84.0	79.1	69.9	1.694	L:1-111	H:1-126	Homo sapiens	
2B1H_LH	93.8	83.8	82.3	72.2	2	L:1-113	H:1-124	Homo sapiens	2B0S_LH_FV;2B1A_LH_FV
3NPS_CB	90.5	82.6	77.6	69.6	1.5	C:1-107	B:1-125	Homo sapiens	

图 2

## 构建目标序列的 3D 模型

### 1 构建模型

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Model Antibody，点击 Model Antibody Framework，打开 Model Antibody Framework 对话框。

点击 Light Chain Sequence 右边的栅格，下拉列表中选取 MA5\_L:MA5\_L。

点击 Light Chain Template 右边的栅格，下拉列表中选取 MA5\_H:MA5\_H。

点击 Interface Template 右边的栅格，下拉列表中选取 Overall Framework Hits:2XWT\_BA\_FV。

将 Number of Models 设为 1。

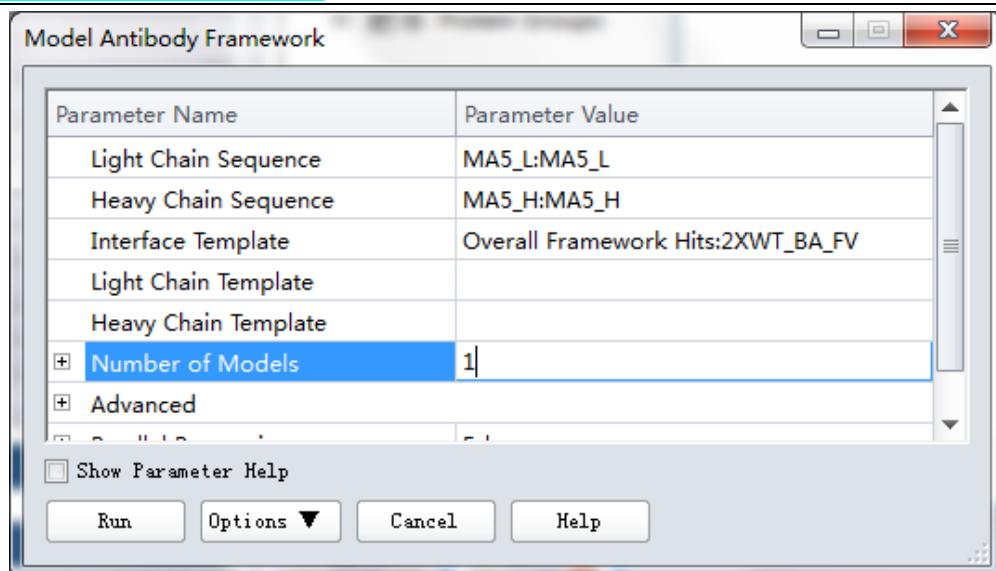
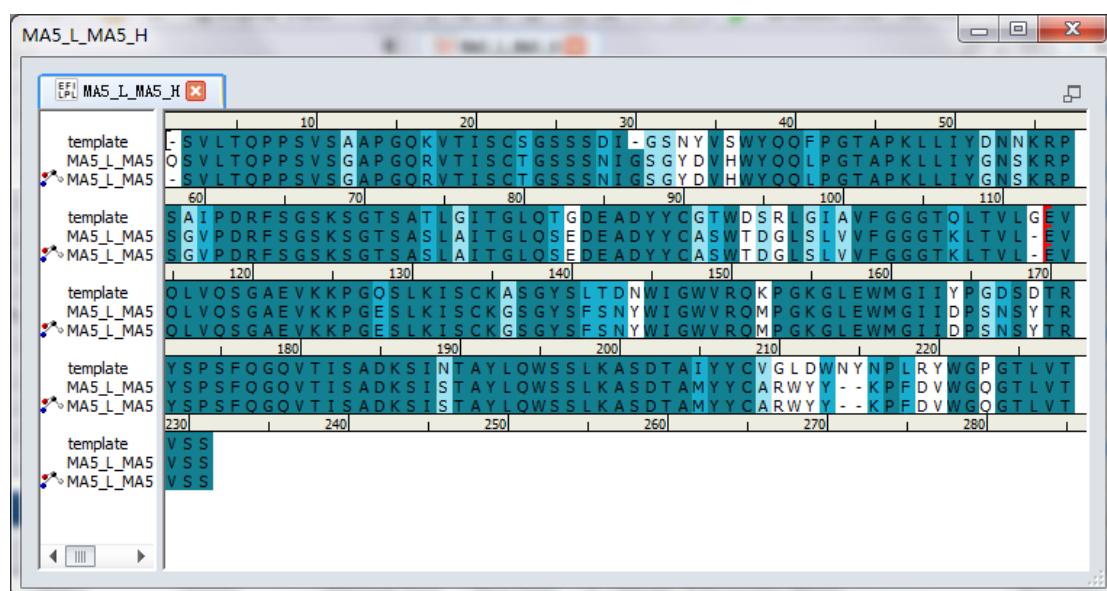


图 3

点击 Run 运行作业，等待作业完成（图 3）。

待作业完成后，点击 View Results 查看结果，一个是名为 MA5\_L\_MA5\_H 的序列窗口，另一个是名为 MA5\_L\_MA5\_H 的分子窗口，包含了表面模板和一个模型结构（图 4）。



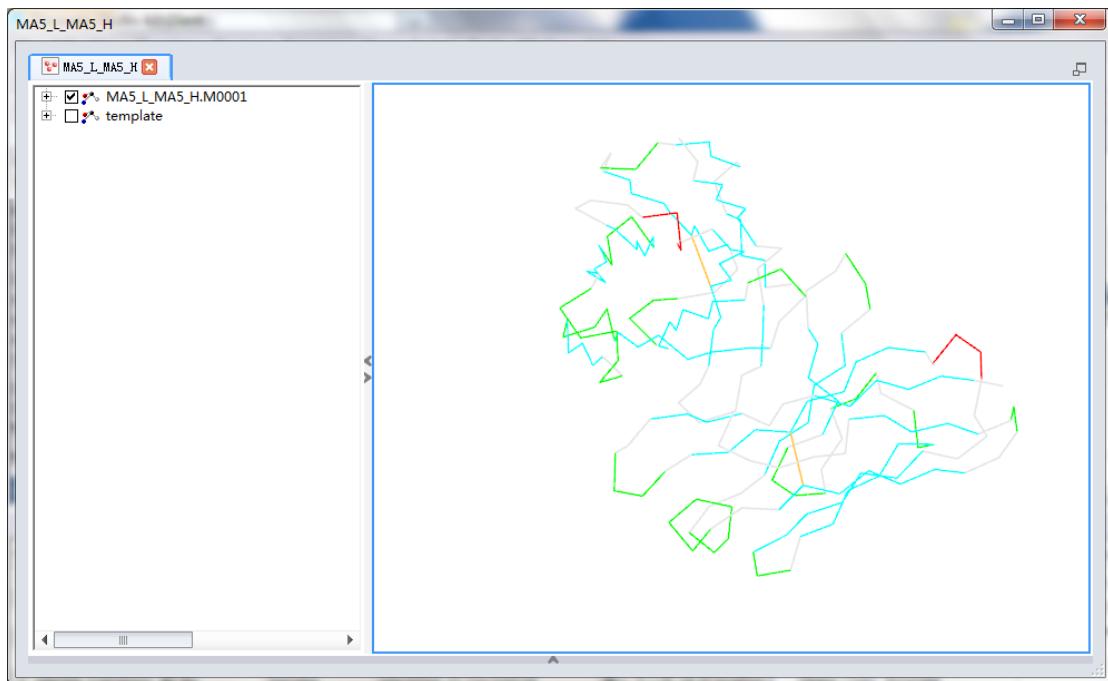


图 4

## 2 模型的 PDF 值或者 DOPE 值分析

建模过程中，DS MODELER 首先会提取模板（template）的几何特性，然后使用 PDF (probability density function) 函数来定义蛋白结构中诸如键长、键角、二面角等几何特性。接着它会对 PDF 函数施加一定的约束条件，并以此来构建 target 的 3D 结构。所以 PDF 的函数值可以直接反应所构建模型的好坏。一般，模型的 PDF Total Energy 越低，表明该模型在同源约束条件下优化的越好；模型同限定的同源约束条件偏差越小，该模型的可信度越大。然而，如果构建的模型其 PDF Total Energy 相似，则可以利用基于原子统计势能的 DOPE score 作为衡量模型质量的依据。DOPE 是一个基于原子统计势能的程序，主要用于模型评估。它的分数可以认为是衡量同一分子不同构象可信度的标准，能够帮助选择预测结构的最优模型。分数越低，模型质量越可靠。

在 MA5\_L\_MA5\_H 分子窗口的表格视图 (Table View) 中，点击 Molecule 标签栏。

可以查看每个模型 (本例中只有一个模型) 的 PDF Total Energy, PDF Physical Energy, DOPE Score。

在系统视图 (Hierarchy View) 中，展开 MA5\_L\_MA5\_H.M0001 | Modeler Groups。

在该组下可以看到，模型序列中同模板序列一致的 (Identical) 或者相似的 (Similarity) 以及没有比对上的 (No Match) 的氨基酸残基分别归于不同的组。

## CDR 区的优化

DS 可以基于序列识别出抗体中的高可变区，即互补决定区 (CDR)，并采用 BLAST 相似性搜索从抗体结构域以及 CDR 库中为每个 CDR 区搜索可能的 loop 模板，所搜寻到的模板将同相应的 CDR 区进行叠合，基于此最后构建优化得到抗体中的 6 个 CDR 结构。

点击激活 MA5\_L\_MA5\_H 分子窗口。

在工具浏览器 (Tools Explorer) 中，展开 Macromolecules | Model Antibody，点击 Model Antibody Loops，打开 Model Antibody Loops 对话框。

确保 Input Protein Molecules 为 MA5\_L\_MA5\_H:MA5\_L\_MA5\_H.M0001。

设置 Maximum Templates Per Loop 为 3。

将 Build Models 设为 True，并展开该组，将 Number of Models 设为 1。

点击 Optimization Level 右边的栅格，下拉列表中选取 Medium。（图 5）

点击 Run 运行任务。

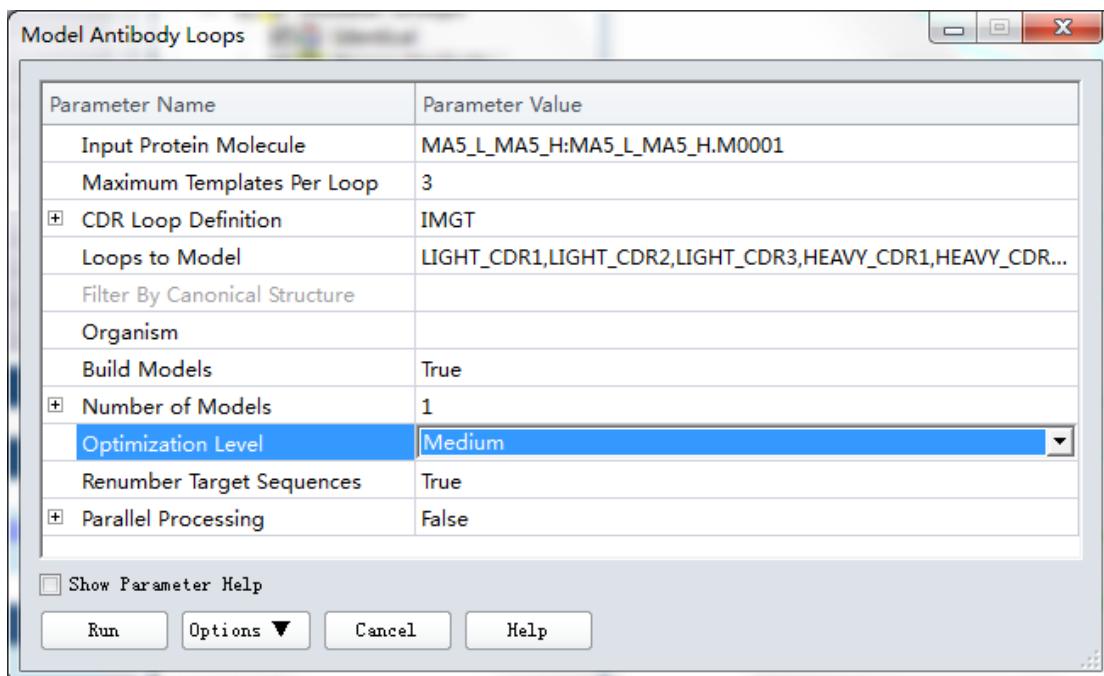


图 5

待作业完成后，查看作业结果，点击 View Results。

同时自动打开两个新的窗口，一个是名为 MA5\_L\_MA5\_H\_M0001 的序列窗口，包含了两个模型序列（loop 优化前和优化后）同 loop 模板序列的比对结果（图 6），另一个是名为 MA5\_L\_MA5\_H\_M0001 的分子窗口，包含了 loop 模板结构和两个模型结构的叠合结果（图 6）。

这两个窗口可以并排显示，并一一对应。

在分子窗口的系统视图中展开 MA5\_L\_MA5\_H\_M0001\_M.M0001|CDR Loops 文件夹。

可以看到模型中重链和轻链各三个 CDR 区。

点击选中 CDR Loops。

可以看到在分子窗口和序列窗口都分别由 loop 结构和序列被选中。

注：同一个模板结构可以用于构建多个 CDR Loop 结构，但是显示的时候则仍分开显示。

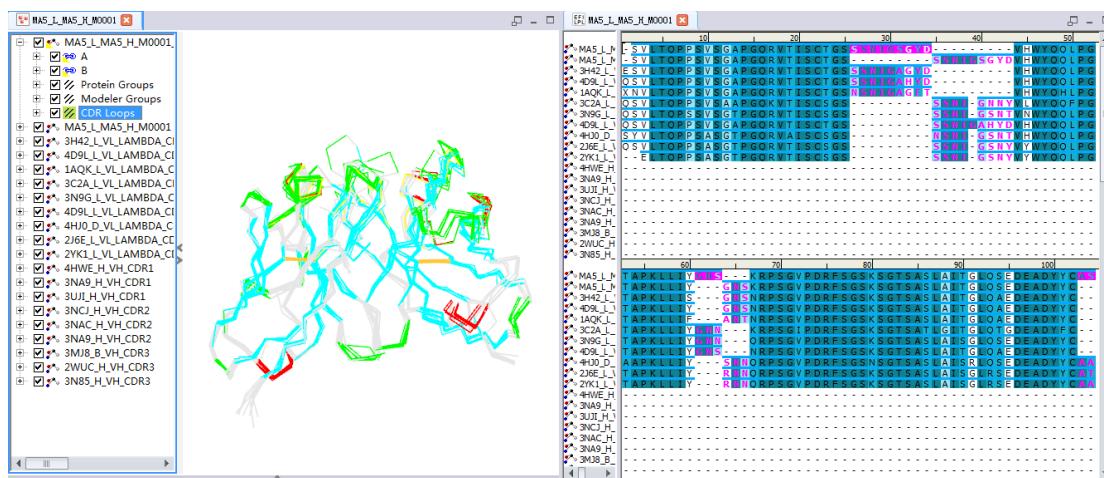


图 6

在分子窗口的表格视图 (Table View) 中点击 Molecule 标签栏。

滚动滑轮，可以查看到 loop 优化前后模型的 PDF Total Energy, PDF Physical Energy, DOPE Score, 以及每个模板结构的打分及同目标序列中相应 CDR 区序列的一致性及相似性等诸多信息。

## 模型评估

模型构建完成后，一般需要对其进行评估。DS 为用户提供了多种模型评估方式。主要包括：Ramachandran plot 和 Profile-3D。具体操作参照 Model\_Enzyme 教程。

## B. 抗体全长 3D 结构的预测

### 载入序列

在文件浏览器 (Files Explorer) 中，双击打开 LC.bmsl, HC.bmsl 两个序列文件。

上述两个序列依次分别为全长抗体的轻链 (L) 序列和重链 (H) 序列。

### 构建目标序列的 3D 模型

在工具浏览器 (Tools Explorer) 中，展开 Macromolecules | Model Antibody，点击 Model Full Length Antibody，打开 Model Full Length Antibody 对话框。

点击 Light Chain Sequence 右边的栅格，下拉列表中选取 LC:Light。

点击 Heavy Chain Sequence 右边的栅格，下拉列表中选取 HC:Heavy。

点击 Template Source 右边的栅格，下拉列表中选取 ligv。

点击 Copy Ligands 右边的栅格，下拉列表中选取 True。

将 Number of Models 设为 1。

点击 Optimization Level 右边的栅格，下拉列表中选取 Medium。（图 7）

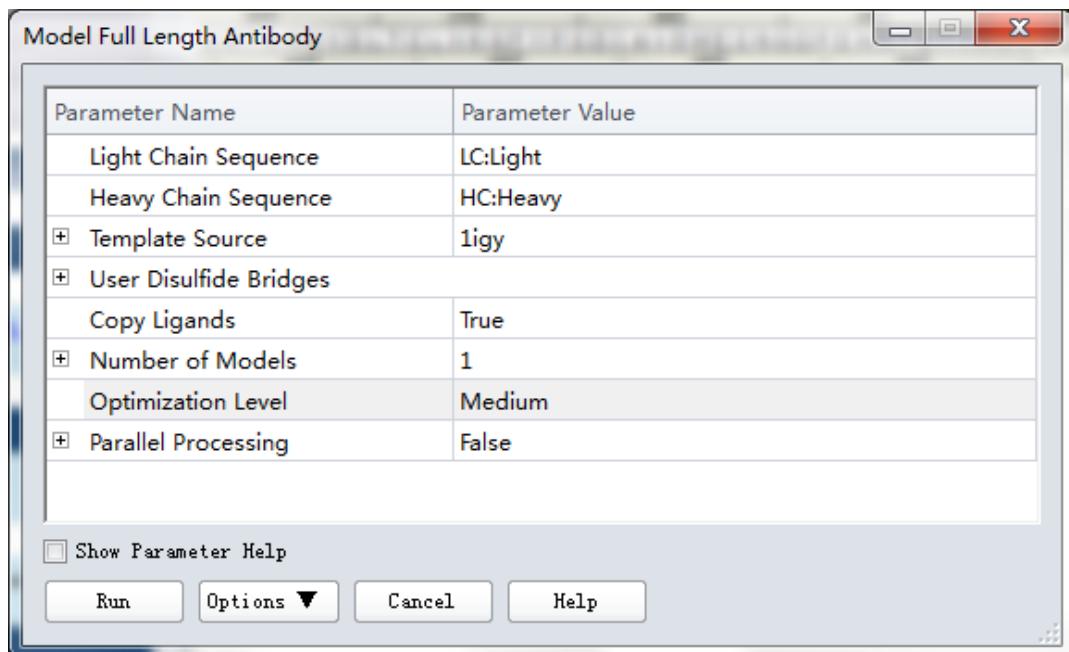


图 7

点击 **Run** 运行作业，等待作业完成。

待作业完成后，点击 **View Results** 查看结果。

同时自动打开两个新的窗口，一个是名为 **Light\_Heavy** 的序列窗口，包含了模型序列同模板序列的比对结果，另一个是名为 **Light\_Heavy** 的分子窗口，包含了模板结构和一个模型结构。这两个窗口可以并排显示，并一一对应。（图 8）

注：同样根据体系仔细检查序列比对结果，如有必要，可以手动进行细微调整，然后基于新的比对结果采用 **Build Homology Models** 流程重新进行模型的构建。

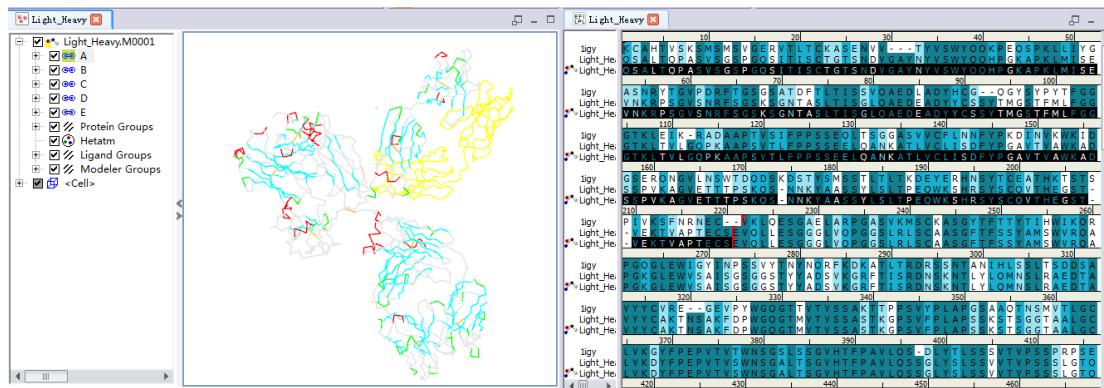


图 8

在 **Light\_Heavy** 分子窗口的系统视图（Hierarchy View）中，展开 **Protein Groups** 文件夹，  
点击 **Disulfide Residues** 可以显示抗体模型中形成二硫键的氨基酸残基。（图 9）

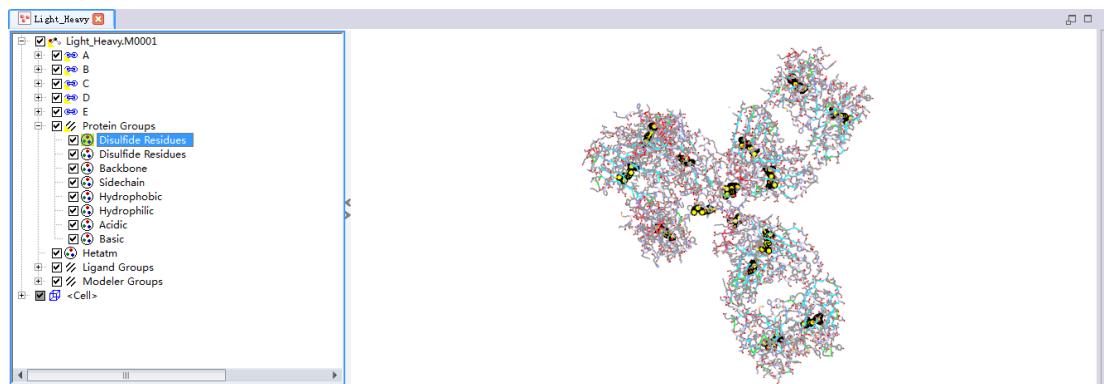


图 9

后续操作同抗体 Framework 区模型构建步骤类似，不再详述。

# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

蛋白对接专题

## 蛋白-蛋白对接（ZDOCK）教程

**目的：**通过此教程，了解 Discovery Studio 中蛋白序与蛋白质之间相互对接的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, Dock Proteins with ZDOCK, Refinement with RDOCK。

**所需数据文件：**2ptn.pdb, 2sta\_I.pdb。

**所需时间：**1 小时 30 分钟

### 介绍

蛋白对接技术是一种预测蛋白质相互识别以及相互作用的技术。在 DS 中，我们可以使用 ZDOCK 来实现蛋白质的对接计算。ZDOCK 是一种基于快速傅里叶转化相关性技术的刚性蛋白对接算法。算法中快速傅里叶转化相关性技术被用于搜索蛋白-蛋白系统的平动和转动空间。RDOCK 是一种基于 CHARMM 的能量优化过程，用于优化 ZDOCK 所寻找到的蛋白-蛋白复合物的结合构型，并使用能量打分函数给这些结合构型打分。

本教程使用 ZDOCK 来进行蛋白对接的实验，并分析对接的结果。从中选取的一些结合构型，利用 RDOCK 进行优化。本教程的蛋白对接实验中，我们选用牛  $\beta$ -胰岛素（PDB 号：2ptn）作为受体蛋白，另外胰岛素抑制剂 CMTI-I（PDB 号：2sta, I 链）将作为配体蛋白。该牛  $\beta$ -胰岛素及其抑制剂 CMTI-I 复合物的晶体结构是已知的（PDB 号为 1ppe）。

本教程涵盖如下内容：

- ZDOCK 运算的设定
- ZDOCK 结果的分析
- 对接构型的 RDOCK 优化
- 蛋白-蛋白结合界面氨基酸的 RMSD 分析

### 设定一次 ZDOCK 计算

#### 1. 在同一个 3D 窗口中打开受体和配体蛋白

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | Protein Modeling 文件夹，  
双击 2ptn.pdb 文件。

DS 将在一个新的 3D 窗口中打开该蛋白。

在同一文件夹中，将 2sta\_I.pdb 拖至上述同一分子窗口（2ptn 分子窗口）中。（图 1）

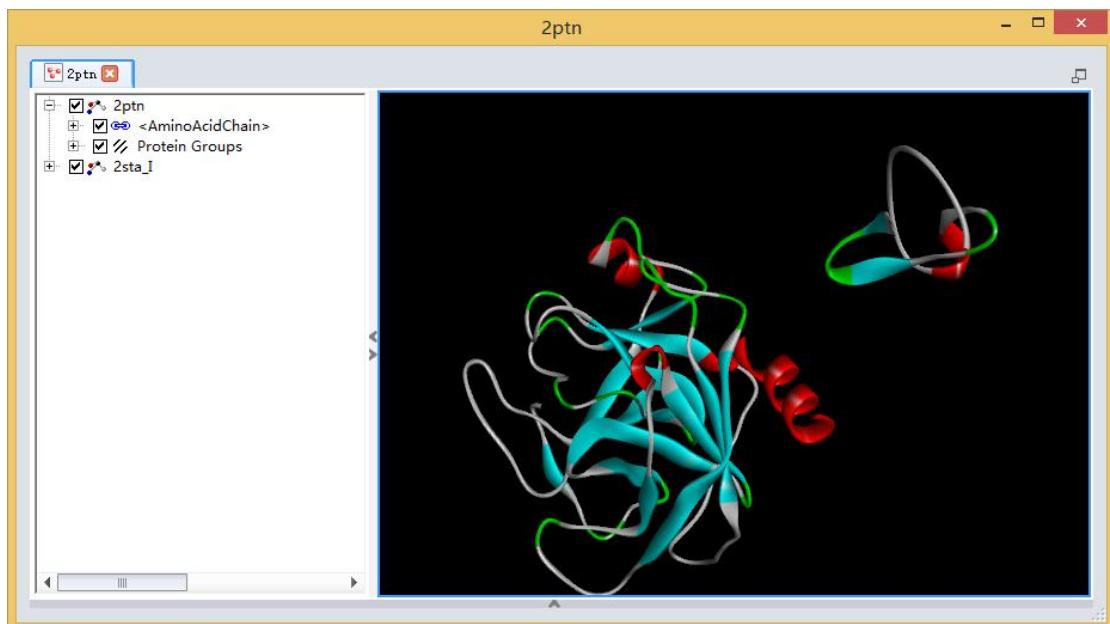


图 1 2ptn 分子窗口

## 2. 蛋白-蛋白对接 (ZDOCK)

在工具浏览器 (Tools Explorer) 中，展开 Macromolecules | Dock and Analyze Protein Complexes，点击 Dock Proteins (ZDOCK)，打开 Dock Proteins (ZDOCK) 对话框。(图 2)  
 设置 Input Receptor Protein 为 2ptn:2ptn。  
 设置 Input Ligand Protein 为 2ptn:2sta\_I。  
 点击 Angular Step Size 右边的栅格，下拉列表中选取 15。

注：ZDOCK 计算过程中可以采用两种欧拉角度进行结合构型的采样：6° 和 15°。采样角度为 6° 时，预测结果更为准确，因为它的最终样本数包括 54,000 个结合构型。尽管采样角度为 15° 的结果准确性有所下降（因为它的采样数只有 3600 个），但它的计算时间更短。

展开 Clustering 参数组。

点击 RMSD Cutoff 参数，将该值设置为 6.0。

点击 Interface Cutoff 参数，将该值设置为 9.0。

点击 Maximum Number of Clusters 参数，将该值设置为 60。

本教程中使用的配体抑制剂较小，对于这样小的体系，将 RMSD Cutoff 设置为 6.0 Å 的 cluster 半径并同时将 Interface Cutoff 设置为 9.0 Å，cluster 结果会更好。本教程中的 ZDCOK 运算，将不指定 blocked residues 也不指定 filtering binding site residues。

设置 ZRank 为 False。

ZRank 选项用于根据静电势、范德华、去溶剂化效应能来对 ZDOCK 初始预测的 pose 进行重新排名。本教程中，我们没有使用该选项。

注：ZDOCK 可以在采用 blocking 和（或）filtering 选项时进行计算：

如果有数据表明某些氨基酸残基不可能出现在蛋白-蛋白的作用界面，那么在受体蛋白中选中这些残基，然后设定 Receptor Blocked Residues 参数为 Create New Group from selection。

同样的，也可以通过设置 Ligand Blocked Residues 参数来指定配体蛋白中不可能出现在界面上的残基。

如果有数据表明，某些氨基酸一定会出现在蛋白质-蛋白质作用界面上，那么选中它们。然后展开 Filter Poses 参数组，设置 Receptor Binding Site Residues 参数和 Ligand Binding Site Residues 参数为 Create New Group from selection。

*Filter Poses* 功能也可以在 ZDOCK 运算完后单独运行，采用 Process Poses (ZDOCK) protocol 即可。

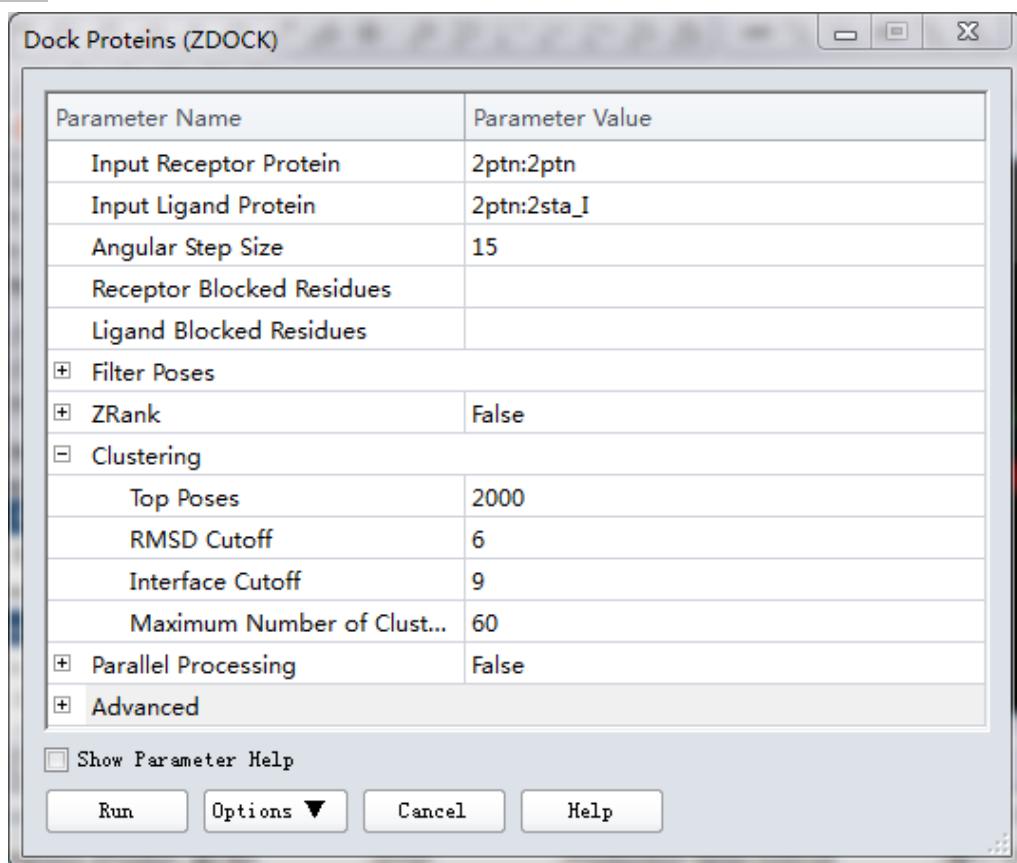


图 2 “Dock Proteins (ZDOCK)”参数设置

### 3. 运行 ZDOCK 并查看结果

点击 **Run** 运行作业，在对话框中观察作业运行状态。点击 **Background** 后台运行该作业。  
该作业大概需要 15 分钟的时间（奔 4 处理器，2Gb 的内存，2.8GHz 的显示器）。  
待作业完成以后，双击任务浏览器 (Jobs Explorer) 中相应的行，打开 Report.htm (图 3)。  
在 report 文件中分析 Summary 部分并点击 **View Results** 以打开一个包含对接 pose 的新窗口。(图 4)

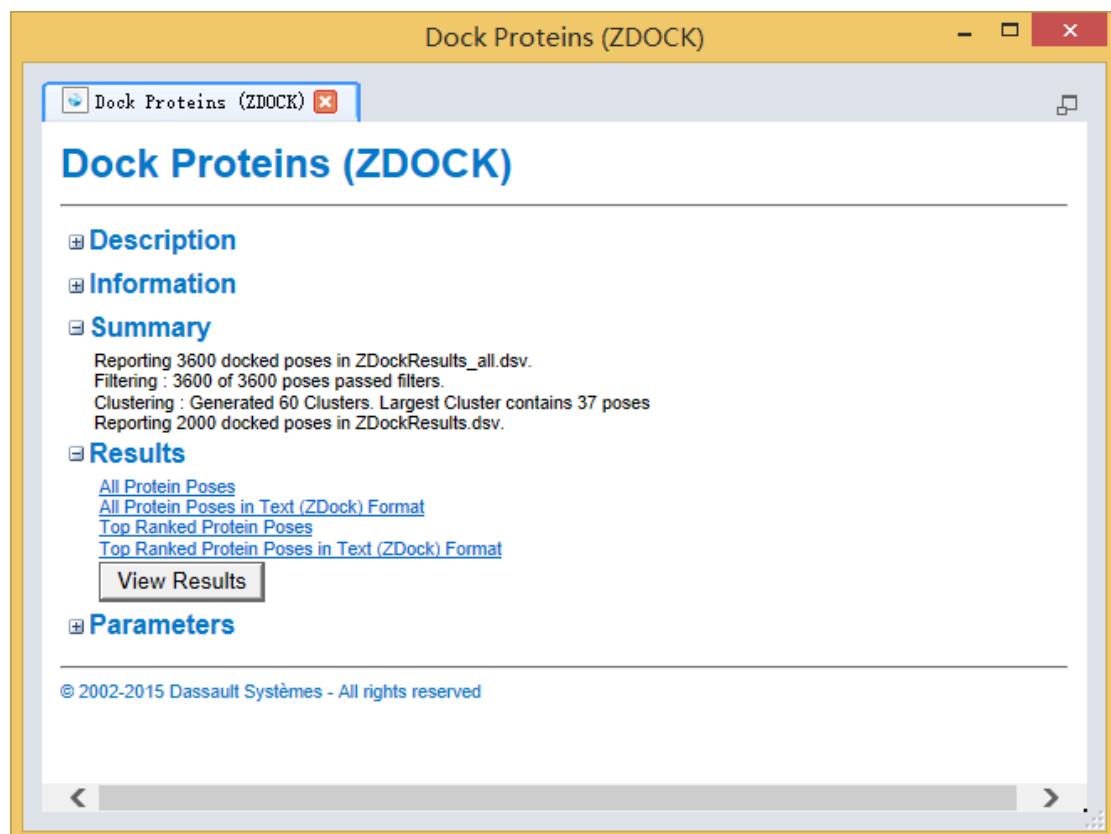


图 3 ZDOCK 运算的 Report 文件

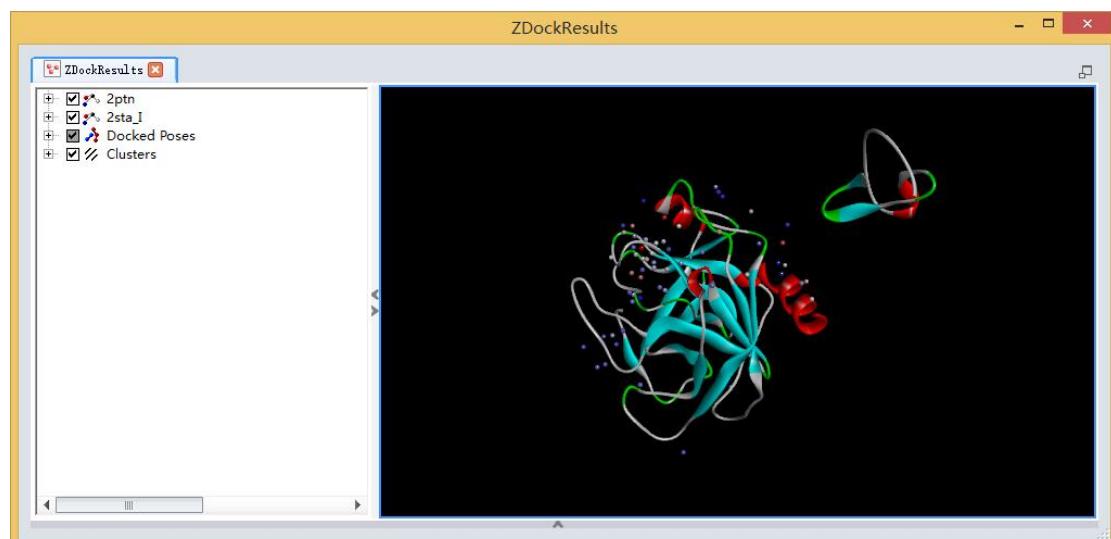


图 4 ZDockResults 窗口

## 分析 ZDOCK 结果

输入的受体蛋白和配体蛋白都显示在视图窗口当中。

如果两个蛋白分子并没有在视图窗口当中显示，则在工具栏中点击 Fit To Screen  按钮使两个蛋白分子居中显示。每一个聚类的中心 pose 都以点的形式显示在受体蛋白的周围。

### 1. 结合构型的聚类 (Cluster) 结果

在系统视图 (Hierarchy View) 中，展开 Docked Poses 和 Clusters。

这将打开几组 Clusters 和 poses。Cluster\_1 是最大的聚类(包含的 poses 最多), 后续的聚类所包含的 pose 逐渐减少。

在系统视图 (Hierarchy View) 中, 关闭 Docked Poses 和 Clusters 组。在表格浏览器 (Data Table) 中, 点击 **Protein Pose** 标签。

这将显示如下关于 ZDOCK 运算结果的信息:

- *ZDock Score* – 包含每个对接构型的 ZDOCK 分数 (PSC 打分函数), Pose 1 的 Zdock 分数最高 (最好)。
- *Cluster* – 指明每个对接构型所属的聚类组
- *ClusterSize* – 报告每个聚类含有的对接构型的个数。
- *Density* – 报告聚类过程中临近的对接构象数目。

关于聚类算法的详述, 可以参考 Discovery Studio Help 中的 [Clustering and analysis of docked protein poses](#)。

## 2. 结合构型作图分析

以下的步骤将解释如何使用不同的方法, 借助 plots 和 Dock and Analyze Protein Complexes 工具来帮助观看对接构型, 并选择一部分 ZDOCK 的对接构型进行进一步的 RDOCK 优化。

在工具浏览器 (Tools Explorer) 中, 展开 **Macromolecule | Dock and Analyze Protein Complexes**, 在 Browse Poses 一栏下设置 **Browse** 为 **Top Poses in Largest Clusters**。

这将显示 10 个最大的聚类中 100 个打分最好的 pose。

点击 **First** 显示第一个 pose 的配体分子。

这将在分子窗口的视图窗口中产生一个名为 Pose1\_Cluster2\_2sta\_I 的配体分子。

在 Calculae Ligand Interface RMSD 一栏下, 点击 **Calculate Binding Site RMSD**。

这将在标签为 ProteinPose 的表格中添加一栏名为 RMSD to Pose1\_Cluster2\_2sta\_I, 该栏显示了每个配体 pose 同定义的配体 pose 之间的 RMSD 值。该 RMSD 值是基于配体分子中结合位点的残基同定义的配体 pose 计算得到的。注意 Define Ligand 会自动设置为 first pose。

点击 **Select Binding Interface**。

该步选取了受体蛋白和配体蛋白结合界面处的氨基酸残基。

点击 **Lock Visibility** 保留配体分子的该 pose 并一直处于可见状态。

点击 **Next** 观察配体分子的下一个 pose, 是 cluster2 中的 pose3。

观察到该 pose 同第一个 pose 很相似。

点击 **Next** 两次显示 Pose7\_Cluster5\_2sta\_I。

观察到该 pose 对接至受体的位置很相似, 但是配体的对接界面却有较大的差距。

继续点击 **Next** 观察其他 pose, 如果想要保留某一个 pose 则点击 **Lock Visibility**。

在 Data Table 中, 点击 **ProteinPose**, 使之处于激活状态。

单击鼠标右键, 选取 **List Only Visible Objects**。

从主菜单栏中, 选择 **Chart| Point Plot**。

这将打开一个 Choose Plot Axes 的对话框。

在对话框中, 选择 **ZDock Score** 作为 X 轴, **Cluster** 作为 Y 轴。

点击 **OK**。

这将打开一个 2D 的点状图。从该图中可以非常清晰地看出 cluster2 包含了最多的打分较高的 pose。

从主菜单栏中, 选择 **Chart | 3D Point Plot**。

在 Choose Plot Axes 对话框中, 选择 **ZDock Score** 作为 X 轴, **Cluster** 作为 Y 轴, **Density**

作为 Z 轴, Color 轴选择 ZDock Score。点击 OK。

这将打开一个 3D 的点状图。

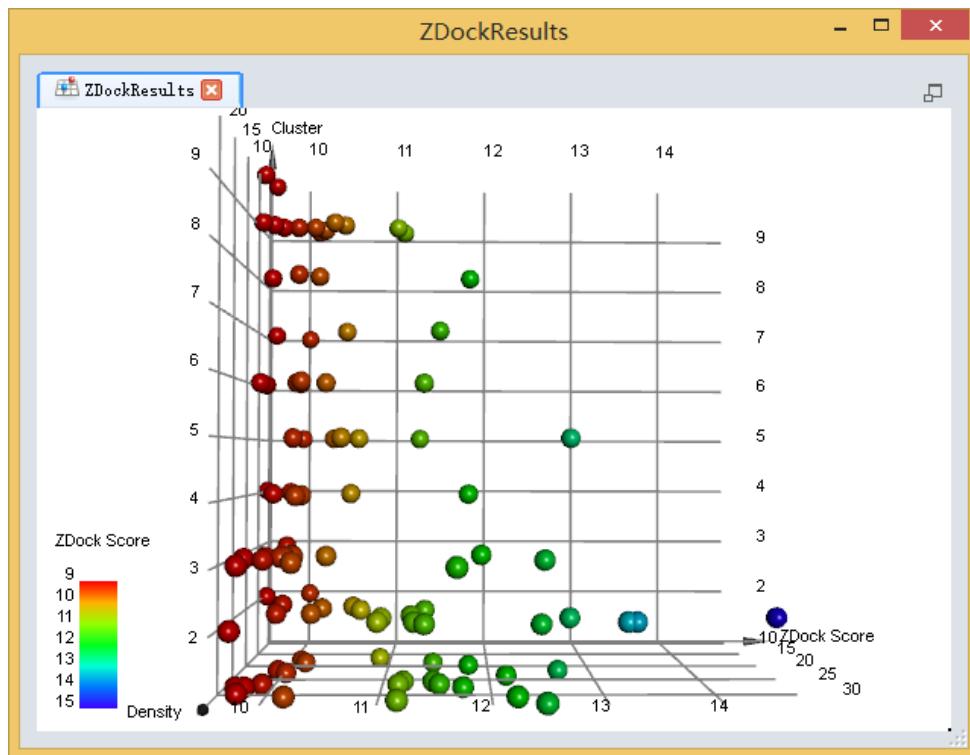


图 5 2D 和 3D 的 Point Plot 图形

点击 ZDockResult 点状图标签, 拖拽该窗口, 以使能够同时看到 ZDockResults 窗口。(图 6)

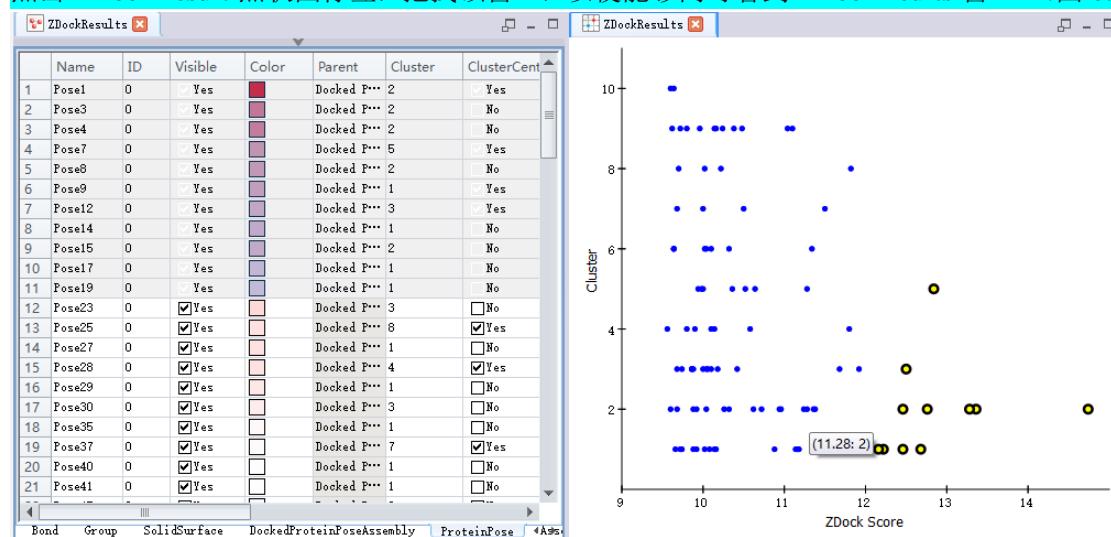


图 6 分子窗口和 Point Plot 图形窗口

在 View 工具栏中, 点击 Select 工具, 选中 ZDockResults 窗口中一些 ZDock 打分高于 12.0 的点。

该操作共选取了 11 个 pose, 5 个 pose 属于 cluster2, 4 个 pose 属于 cluster1, cluster3 和 5 各包含一个 pose。选中的对接构型在 3D 窗口中也会被选中。

## 利用 RDOCK 优化对接构型

### 1. RDOCK 优化

点击 ZDockResults 分子窗口，以使该窗口处于激活状态。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Dock and Analyze Protein Complexes，在 Refine Poses 一栏下，点击 Refine Docked Proteins(RDOCK)，打开 Refine Docked Proteins (RDOCK) 对话框。（图 7）

点击 Input Receptor Protein 右边的栅格，下拉列表中选择 ZDockResults:2ptn。

点击 Input Ligand Protein 右边的栅格，下拉列表中选择 ZDockResults:2sta\_I。

点击 Input Poses 右边的栅格，下拉列表中选择 Create New Group From Selection。

默认的会将该新产生的组命名为 Group，共包含之前选取的 11 个 pose。

**注：** Dielectric Constant 参数仍保留设置为 4.0。该值即运算 RDOCK 中计算 CHARMM 能量时所用的介电常数。

点击 Run 运行作业，在对话框中观察作业运行状态。

点击 Background 后台运行该作业。

该作业大概需要 5 分钟的时间（奔 4 处理器，2Gb 的内存，2.8GHz 的显示器）。作业结束时会出现一个提醒任务结束的对话框。

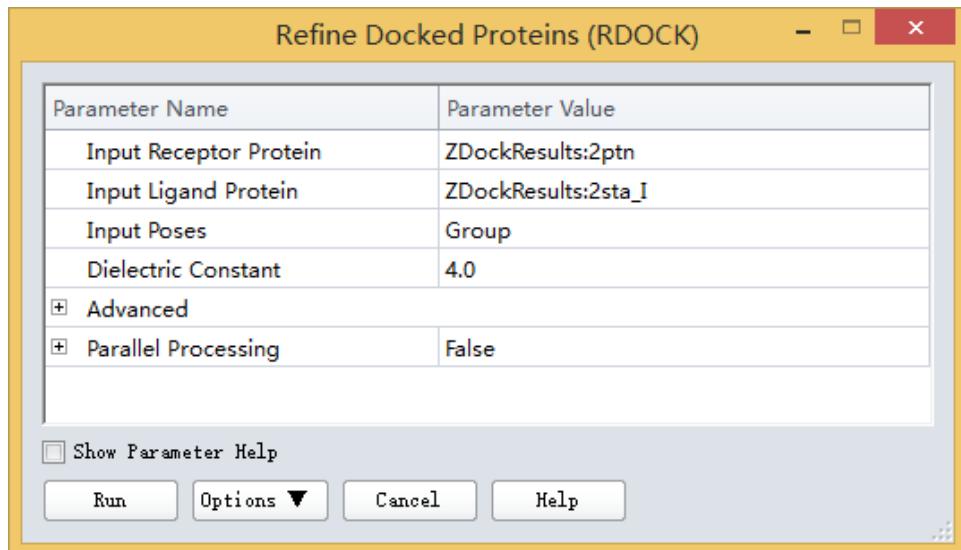


图 7 RDOCK 参数设置

### 2. 结果分析

待作业完成以后，双击任务浏览器（Jobs Explorer）中相应的行，打开 Report.htm。

在 report 文件中点击 View Results 打开一个新的分子窗口。（图 8）

经 RDOCK 优化的对接构型均出现在该窗口中，其中 top refined pose 在视图窗口中显示。

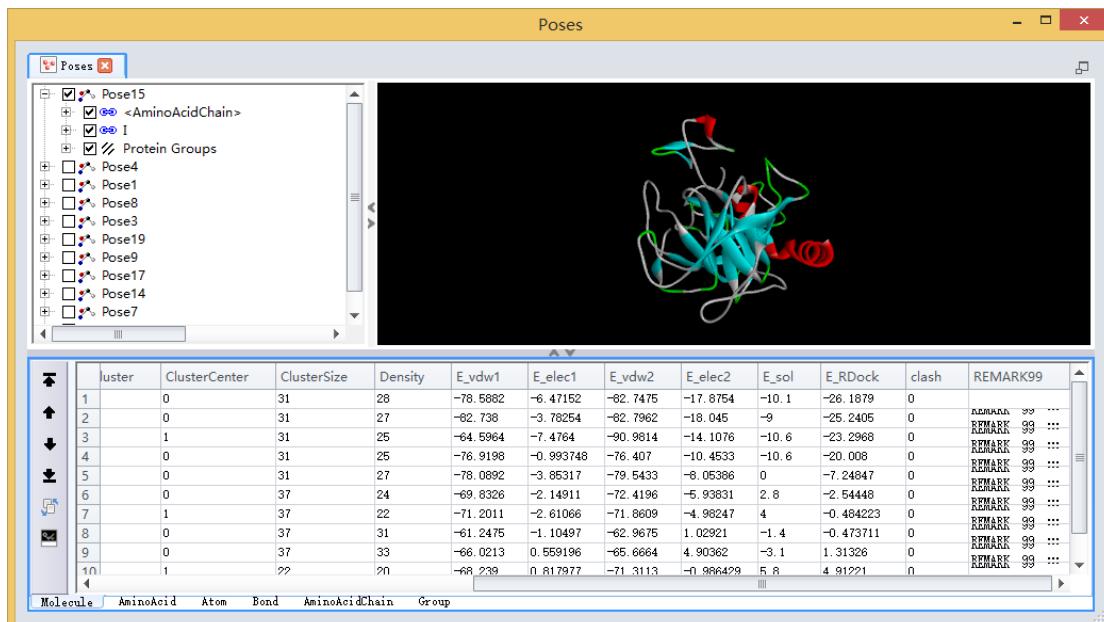


图 8 RDOCK 分子窗口

每一个优化好的 pose 其 RDOCK 所计算的各能量值都列在表格浏览器(Data Table View)中。

- $E_{elec1}$  and  $E_{elec2}$  – 经过第一、二轮 CHARMM 能量优化后蛋白质复合物的静电势能
- $E_{vdw1}$  and  $E_{vdw2}$  – 经过第一、二轮 CHARMM 能量优化后蛋白质复合物的范德华非键作用能
- $E_{sol}$  – 经 ACE 方法计算得到的蛋白质复合物的去溶剂化能
- $E_{RDock}$  – RDOCK 分数的定义为:  $E_{elec2} + \text{beta} \times E_{sol}$

表格中同样还包含了运算 ZDOCK 之后得到的各 pose 的性质，例如 ZDock Score 和聚类信息等。这些信息有利于结果的解释。

RDock 打分最好的 pose 属于 cluster2。这也证实了 cluster2 所包含的 pose 是对接最好的结果，有可能同真实的对接构象比较接近。

在 Table 中，点击 可以看到第一个 pose 的 3D 构象。分别点击 和 可以查看上一个和下一个 pose 的结构。

## 使用 RMSD 分析结合界面的氨基酸

本小节将把对接的构型同蛋白质复合物 PDB 1ppe 的晶体结构进行比较。

### 1 打开 PDB 1ppe

点击 Poses 分子窗口的标签，使该窗口处于激活状态。

从主菜单栏中，选取 File | Insert From | URL，打开 Insert From URL 对话框。

在 ID 文本框中输入 1ppe。确保地址栏设置为 Default PDB Structures。

点击 Open。

这将在 Poses 分子窗口的视图窗口中插入 1ppe 蛋白复合物结构。

在分子窗口中，单击鼠标右键，选取 Show All 显示所有 pose。

在系统视图 (Hierarchy View) 中，展开 Cell，然后展开 1PPE。

选中 Water, Delete。

这将删除 1ppe 中的水分子。

## 2. 将对接构型与 1ppe 叠合

在进行 RMSD 计算之前，首先应该将对接的 poses 和 1ppe 进行基于受体蛋白的叠合。

在视图窗口（Graphic View）中，单击鼠标右键，选取 Show Sequence。

这将打开一个 Sequence 窗口。（图 9）

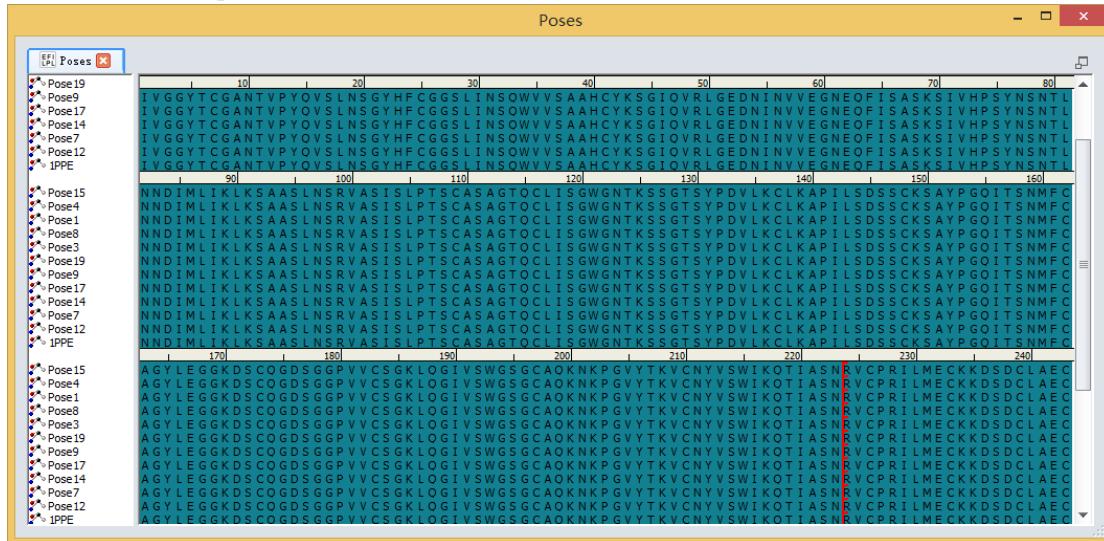


图 9 序列窗口

点击 Poses 分子窗口标签，激活该窗口。

在系统视图（Hierarchy）中，点击选中 1ppe 的 E 链。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Superimpose Proteins，在 Reference Protein 一栏下，选取 1PPE 中选中的链作为叠合时的参考分子。

在 Sequence Alignment 一栏中，设置 Sequence Alignment 为 Poses，并将弹出的窗口关闭。

在分子窗口中，确保 1ppe 的 E 链仍被选中。

在 Superimpose Proteins 一栏中点击 Superimpose。

该步将对对接 poses 的受体对接至 1ppe 的 E 链。

## 3. 计算结合界面氨基酸的 RMSD 值

在系统视图（Hierarchy）中，点击选中 1ppe 的 I 链。

在 Dock and Analyze Protein Complexes 工具面板下，点击 Define Ligand。

点击 Select Binding Interface。

选中处于蛋白-蛋白分子界面的氨基酸残基。

从菜单栏中，选择 Structure | RMSD | By Sequence Alignment....

这将打开 RMSD by Sequence Alignment 的对话框。

在该对话框中，点击 Reference molecule 右边的栅格，下拉列表中选择 1ppe。点击 Selected residues 左边的圆形按钮，关闭 Report at residue level 选项。点击 OK。（图 10）

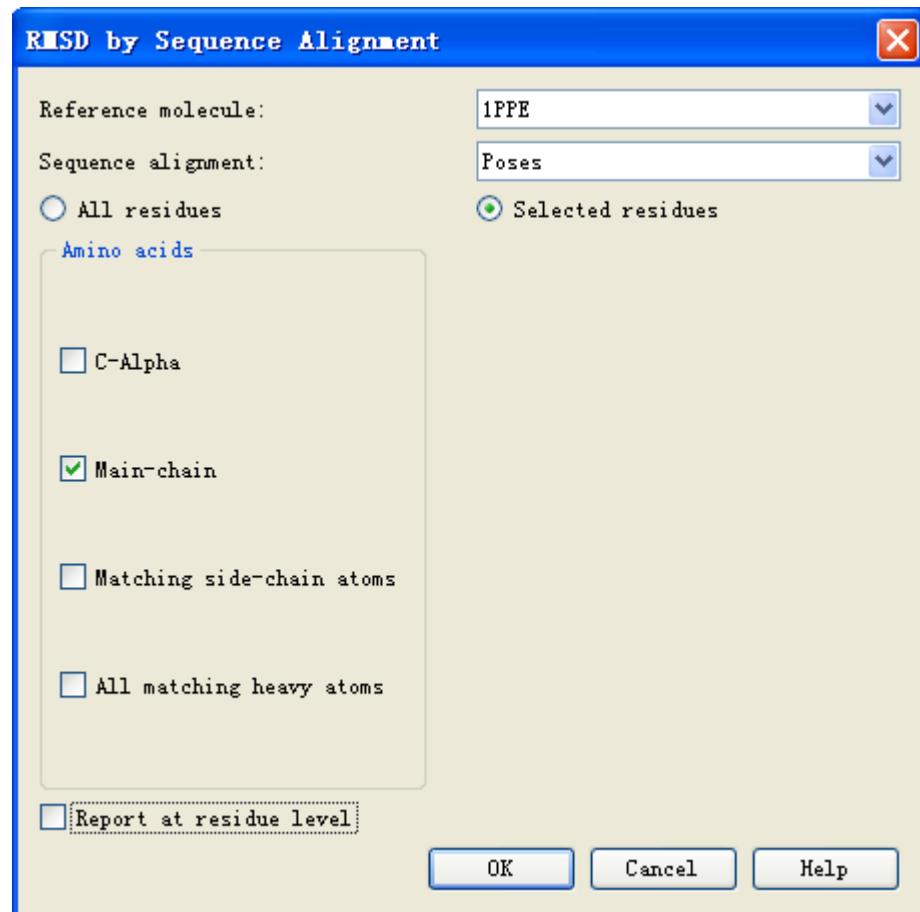


图 10 “RMSD by Sequence Alignment”参数设置

DS 将在一个新的 Html 窗口中打开一个基于选中氨基酸的 RMSD 报告。(图 11) 比较对接 poses 的 E\_RDock 打分和 RMSD 值,发现 RMSD 值最低的 poses 其 E\_RDock 打分也最低,都属于 cluster2 类别。

Molecule	Reference	Main-chain
Pose15	1PPE	0.707
Pose4	1PPE	1.945
Pose1	1PPE	1.147
Pose8	1PPE	1.767
Pose3	1PPE	2.184
Pose19	1PPE	6.337
Pose9	1PPE	6.950
Pose17	1PPE	6.007
Pose14	1PPE	5.824
Pose7	1PPE	6.868
Pose12	1PPE	6.803

图 11 RMSD 报告



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

分子力学/分子动力学专题

## Discovery Studio Molecular Dynamics 教程

### Molecular Dynamics – 分子动力学方法

**目的:** 采用 Discovery Studio, 以一组蛋白-配体复合物为实例, 示范分子动力学及结果分析操作过程。

**所需功能和模块:** Discovery Studio Visualizer client, CHARMM, Simulation protocols

**所需数据文件:** 1sem.pdb

**所需时间:** 1 小时

#### 介绍

分子动力学(Molecular Dynamics, MD)是分子模拟中最常用的方法之一。该方法基于分子力场, 能够动态的描述分子的运动状况, 继而描述生命的动态过程。分子动力学在生命科学领域中的应用非常广泛, 如蛋白质折叠的机理研究、酶催化反应的机理研究、与功能相关蛋白质的运动研究、生物大分子大范围构象变化的研究等等。近几十年来, 分子动力学方法已经成功的运用于大分子体系低能量构象的模建、X 射线晶体衍射以及 NMR 实验结果的处理。<sup>[1,2]</sup> 现在分子动力学方法已经成了理论生物学研究中必不可少的方法之一。<sup>[3,4]</sup>

分子动力学模拟主要包括以下几个步骤:

1. 模拟体系升温过程 (Heating Stage)
2. 模拟体系平衡过程 (Equilibrium Stage)
3. 模拟体系采样过程 (Production Stage)
4. 分析体系的目标性质 (Analysis)

然而, 由于一般的待模拟体系的初始结构或多或少的存在缺陷(比如, 初始结构不完整或存在不合理的结构区域), 所以我们往往需要对初始结构进行预处理才能进行分子动力学模拟。一个完整的分子动力学模拟过程应包括如下几个步骤:

1. 初始结构检查及预处理
2. 给模拟体系赋力场参数
3. 考虑溶剂效应
4. 初始结构能量最小化
5. 动力学模拟(包括升温、平衡、采样)
6. 结果分析

目前, 常用的分子动力学模拟软件都基于 Unix (Linux) 操作系统。实施每一个模拟的步骤都需要特定的命令来调用相关的程序对模拟体系进行处理。同时, 模拟者还需要熟知每个参数的意义并定义相关的参数值。最终分析过程也需要配合其他软件才能完成。这对初级分子动力学模拟者而言是非常困难的(模拟者不仅需要掌握分子动力学软件的使用命令, 还需要掌握操作系统相关的命令)。为了解决这样的问题, Discovery Studio 为用户提供了基于窗口的分子动力学模拟工具。使用 DS, 用户可以轻松的完成分子动力学的整个模拟过程(包括体系的准备、模拟过程的设置及结果的分析)。另外考虑到初级模拟者在整个模拟流程方面不熟悉, DS 专门设置了一个“Standard Dynamics Cascade”protocol。该 Protocol 将整个动力学模拟的步骤集合到一个参数设置界面当中, 方便初级模拟者的使用。

为了使大家对 MD 的操作流程有所了解, 本教程主要内容包括如下几个方面:

1. 初始结构的准备, 力场设置
2. 对结构进行溶剂化处理
3. 标准的分子动力学模拟流程

#### 4. 动力学模拟结果的轨迹分析

以下以一组蛋白-配体复合物为例进行部分束缚的动力学模拟，介绍 Discovery Studio 中分子动力学的基本操作方法：

- 准备蛋白-配体复合物
- 选择并定义原子组
- 添加溶剂
- 设置束缚部分
- 进行带束缚的分子动力学模拟
- 分析动力学模拟轨迹
- 计算蛋白-配体相互作用能量
- 

### 准备分子动力学体系，执行分子动力学计算

#### 1. 蛋白结构预处理

双击图标，启动 Discovery Studio，在文件浏览器（Files Explorer）双击打开 Smaples | Tutorials | Simulation | 1sem.pdb 文件，将蛋白-配体复合物导入 3D 窗口。（图 1）

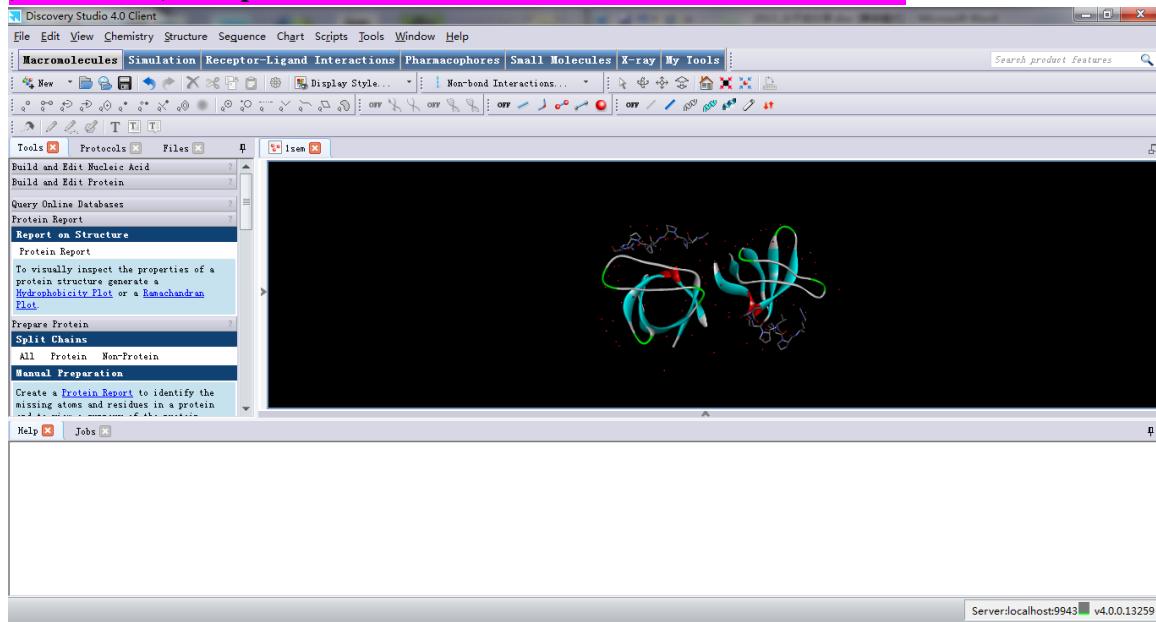
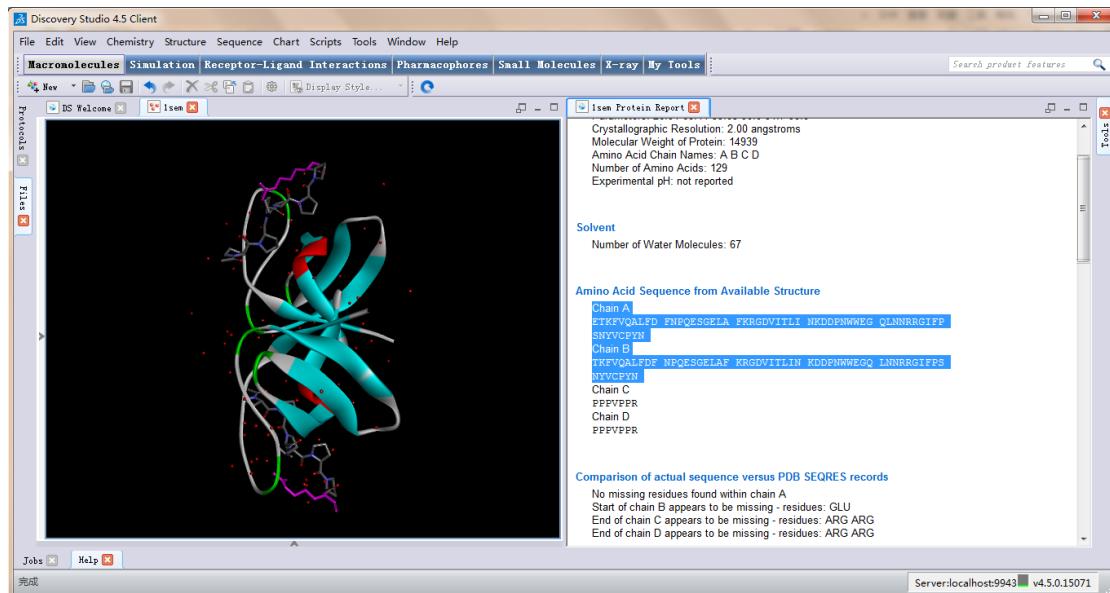


图 1

展开 Macromolecules | Protein Report 工具面板，单击 Protein Report。

Protein Report 是蛋白质结构的重要信息报告。在报告中不仅提供了结晶的分辨率、分子量、氨基酸链名称/数量、实验 pH 值等基本信息。还提供了缺失或不完整的氨基酸残基会以洋红色的颜色标注出来。为后续结构修正和优化提供参考。

可以看出蛋白质复合物有两个相同的亚基组成，下一步我们需要判断保留哪个亚基进行随后的动力学模拟。



点击 1sem 标签，确保其处于激活状态。然后点击标准工具栏（Standard toolbar）中 Display Style 的下拉箭头，选择右下角的显示方式（蛋白质以线状表示，小分子以球棍状表示）。  
(图 2)

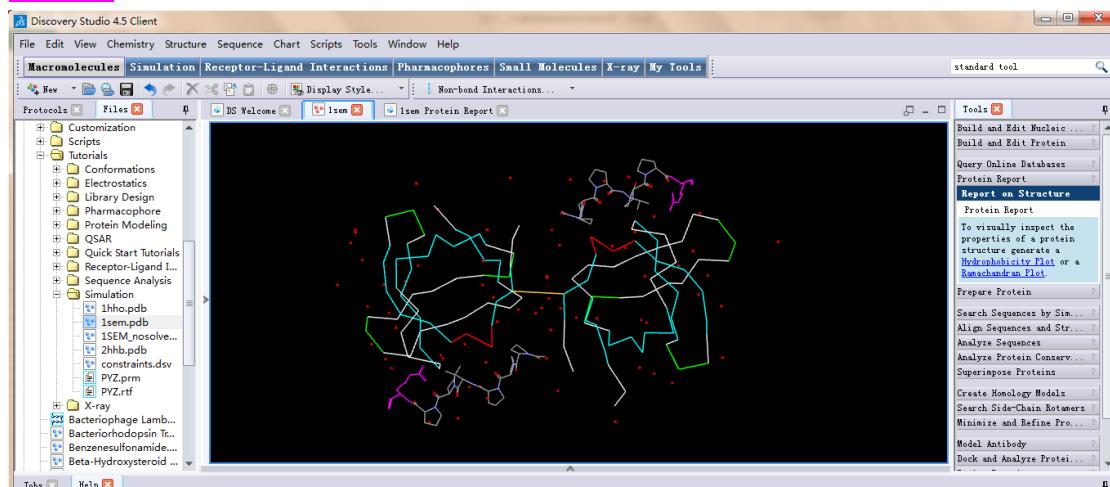


图 2

再次单击 Display Style，将 Color by 改为 Isotropic Displacement。单击 Colors...按钮，将颜色改为 Blue-White-Red。图 3

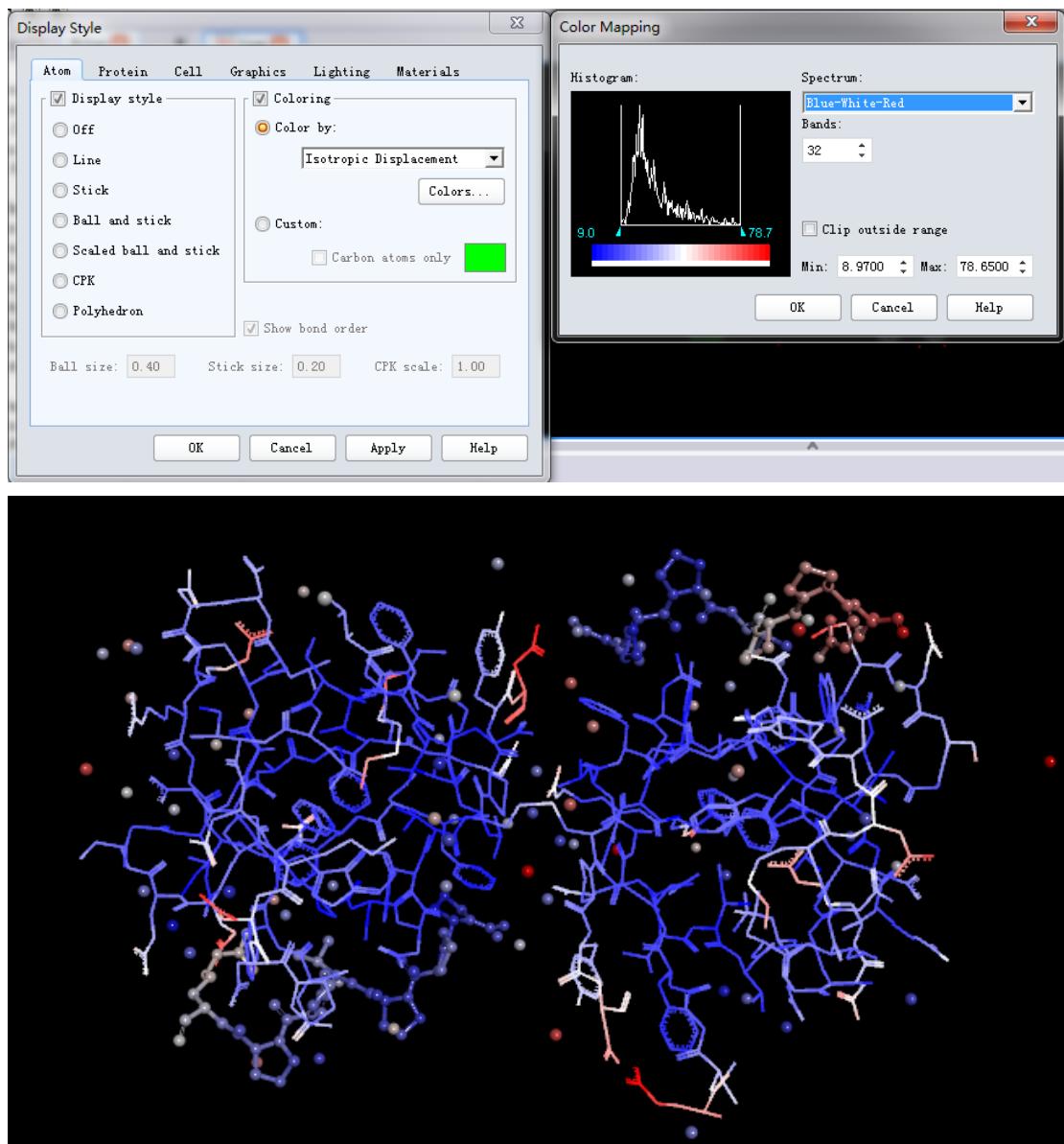


图 3

根据 Isotropic displacement（也被称为 B factor 或温度因素）的数值，我们能够清晰的识别出每个原子位置的不确定性。值越大（红色）表示原子的位置变化越大。值越小（蓝色）热振动越小，位置越精确。

我们将 C、D 链进行对比，显然 C 链的红色更多。而 D 链基本为蓝色。另外根据 Protein Report 显示两个亚基的氨基酸残基的缺失或不完整的情况均较好。因此综合考虑后，保留 B、D 链。快捷键 **CTRL+H**，打开层级窗口（Hierarchy View），然后选择所有 A、C 链，快捷键 **delete** 删除。

展开 Macromolecules | Prepare Protein 工具面板（Tool Panel），单击 **Prepare Protein....** 参数默认，点击 **run**。等待任务完成。运行大约半分钟（图 4）

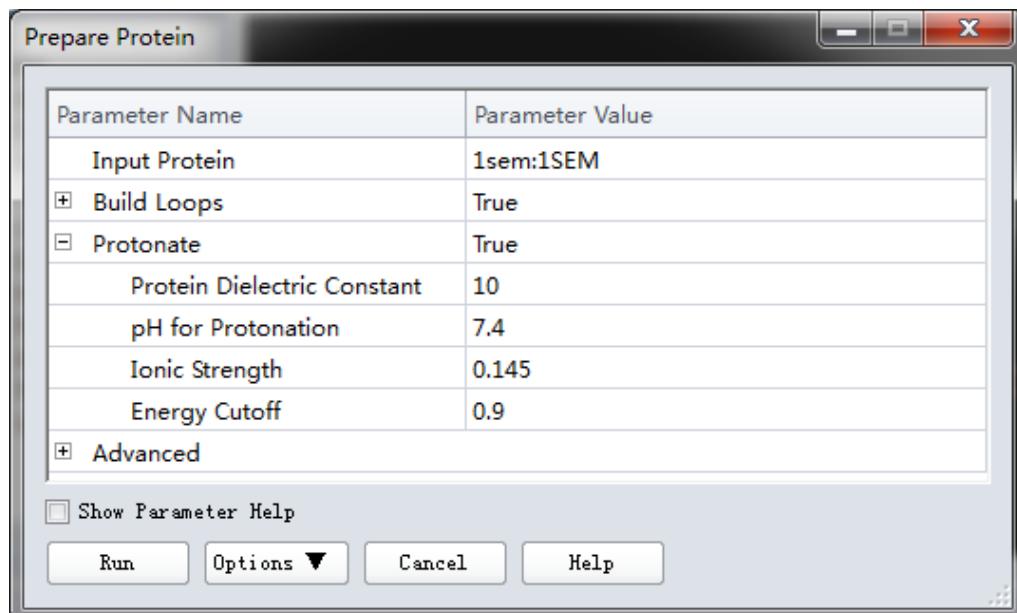


图 4

### 1. 溶剂化蛋白质

打开处理过的 1SEM\_Prep 文件，首先添加力场：打开 Simulation 工具面板，选择 change forcefield，将 Forcefield 改为 charmm36，然后再点击“Apply forcefield”。

从 Protocol 面板组中找到 Simulation | Run Simulation 文件夹，点击 Solvation 流程（图 5），该流程在参数浏览器（Parameters Explorer）中打开，Input Typed Molecular 输入 1SEM\_prep:1SEM，其它参数使用默认值，然后点击 Run。

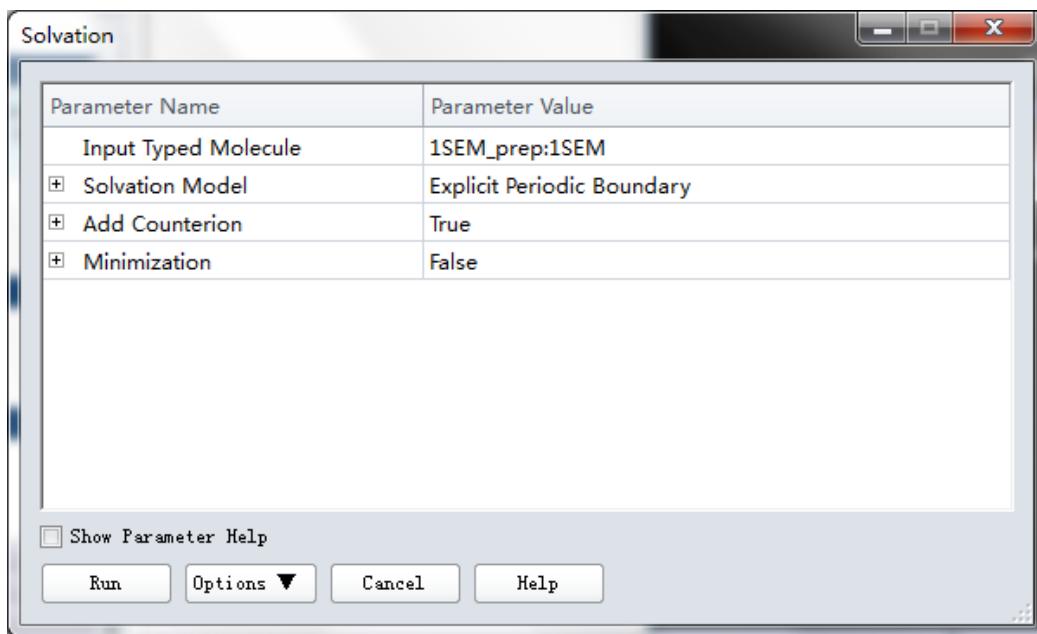


图 5 Solvation 参数设置

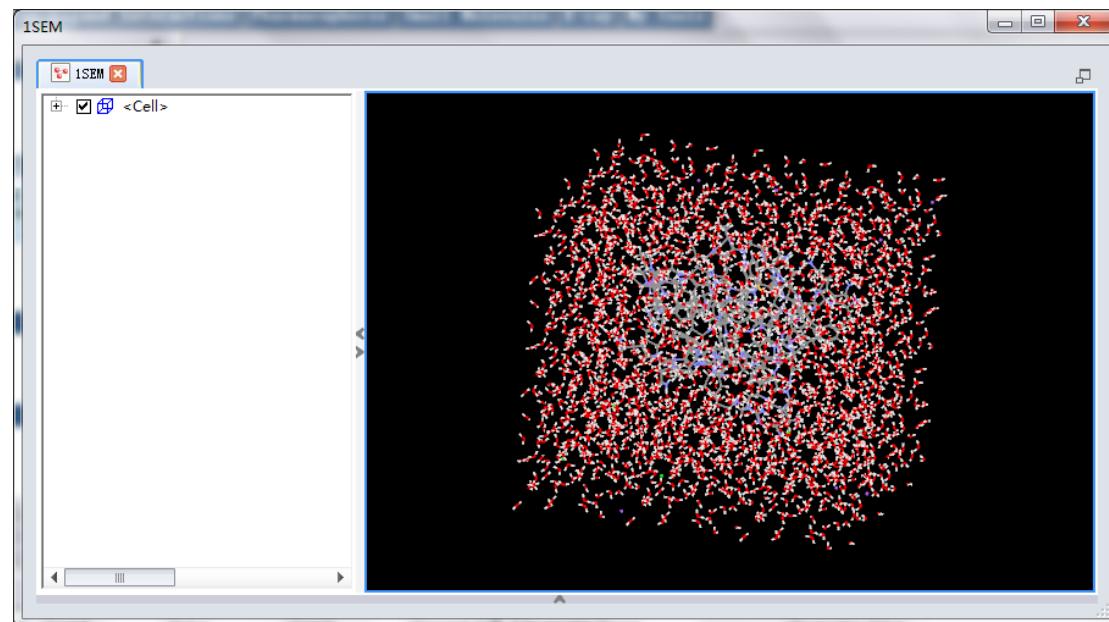


图 6 添加溶剂体系示意图

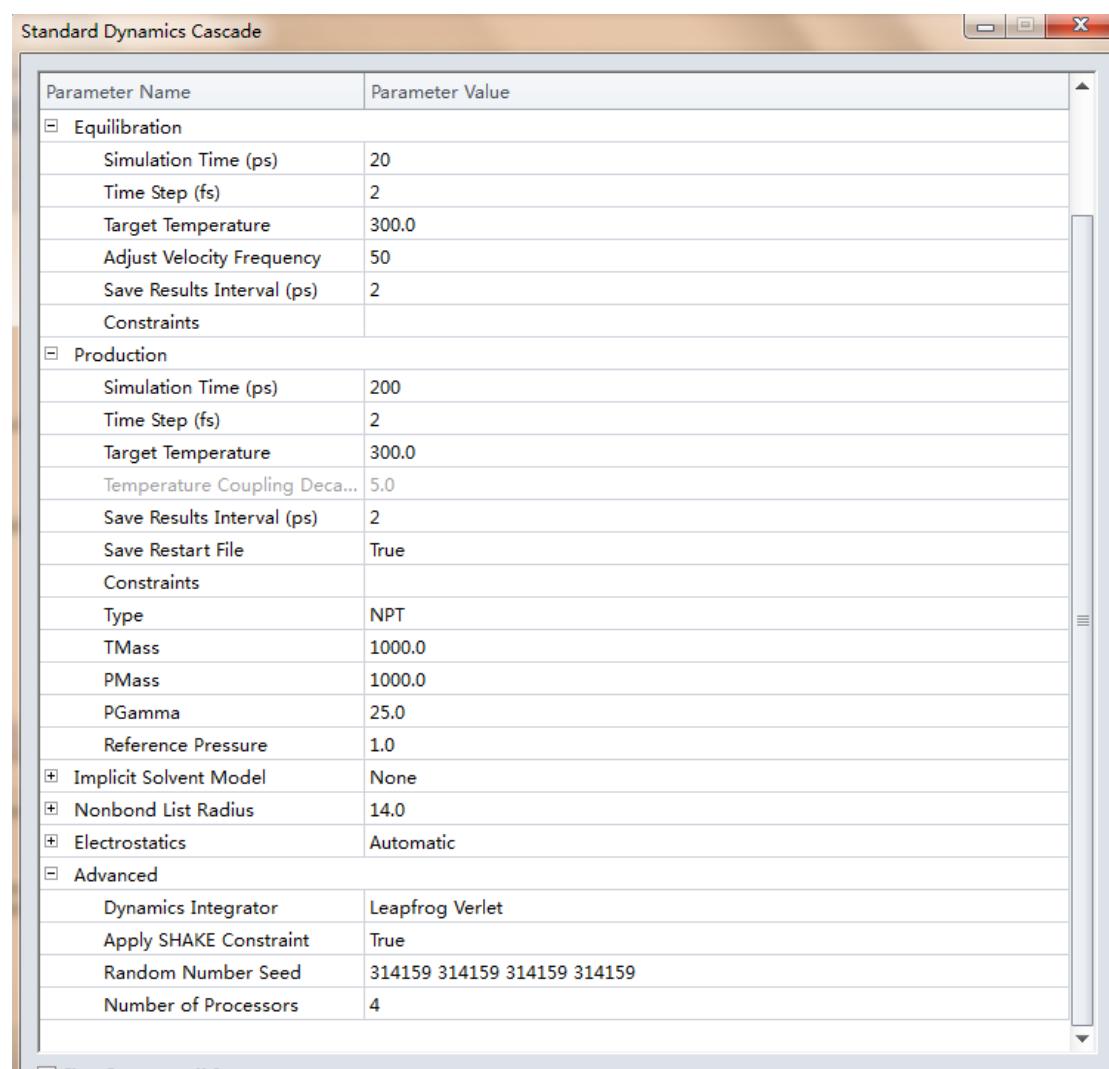


图 7 Standard Dynamics Cascade 流程参数设置

## 2. 打开动力学流程 Standard Dynamics Cascade，修改参数

确保添加了溶剂体系的 complex.csv 窗口为当前活跃窗口，文件流程浏览器（Protocol Explorer）中双击 Simulations 文件夹下 dynamics 文件夹下 Standard Dynamics Cascade 流程，该流程在参数浏览器（Parameters Explorer）中打开，“1SEM:1SEM”自动载入到 Input Typed Molecule 参数格。

Standard Dynamics Cascade 流程包括 5 个阶段：优化 I (Minimization)、优化 II (Minimization2)、升温 (Heating)、平衡 (Equilibration)、模拟采样 (Production)。

展开 Equilibration 阶段的参数组，设置 Simulation Time (ps) 为 20。

展开 Production 阶段参数组，设置 Simulation Time (ps) 为 200 (图 8)

展开 Advanced，设置 Number of Processors 为 8，其它参数保持默认，点击 Run。

8 核计算时，该作业需要 2 小时完成，作业的状态可以在“Jobs Explorer”中查看。当任务完成对话框打开，点击 OK 关闭。

## 3 查看结果，显示构象变化动画

在 Report 页面，点击 View Result 链接。如果您没有运行完上述长时间的动力学模拟，也可以打开 Samples | Tutorials | Simulation | 1SEM\_nosolvent.csv 文件

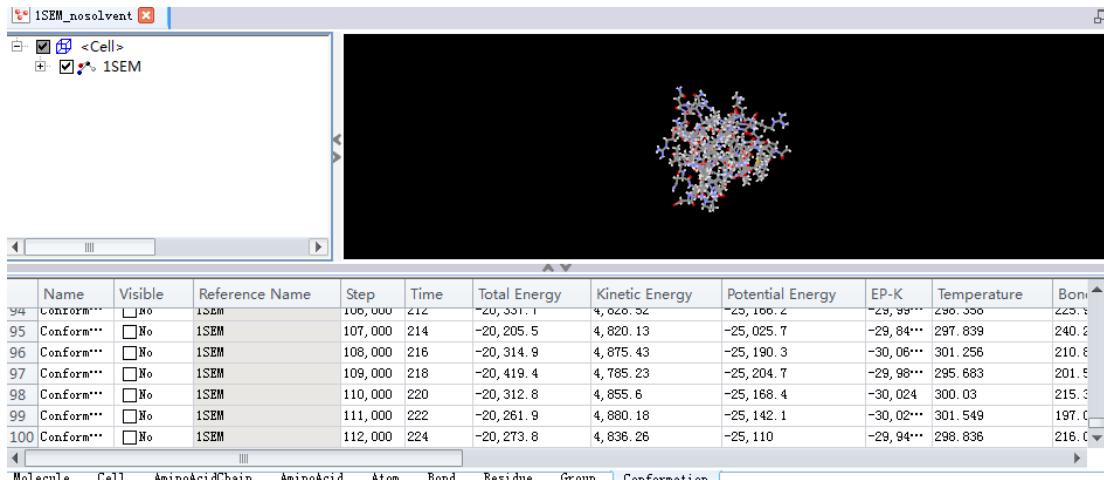


图 8

选择下拉菜单 Structure | Animation | Play 命令，则可以看到在 3D 窗口中的复合物开始运动，显示动力学模拟过程中的变化。动画循环往复播放。再次选择下拉菜单 Structure | Animation | Play 命令可以停止动画播放。

## 分析分子动力学计算结果

您将通过 Analyze Trajectory 方法分析分子动力学轨迹的一些结构信息。该方法能够计算一些几何性质，如：距离、角度、二面角、或者非键相互作用的数量。另外也能够计算 RMSD、RMSF。

### 1、定义非键相互作用监控

打开层级窗口 (Hierarchy View)，选择 D 链，然后点击标准工具栏的 Non-bond

Interactions... ，打开一个创建非键相互作用对话框。

将 Molecular Scope 改为 Intermolecular (between molecules)，点击 OK (图 14)

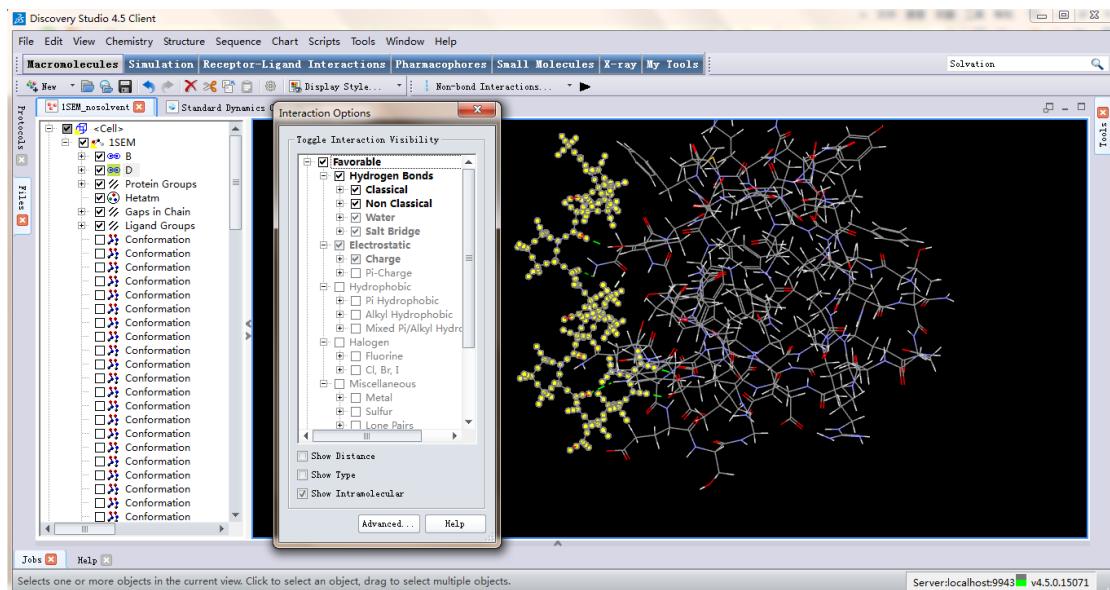


图 14

## 2. 建立 RMSD、RMSF，及编译力学轨迹监控数据

在流程浏览器(Protocol Explorer)中展开 Simulation | Analyze 文件夹, 双击 Analyze Trajectory 流程打开。

Input Molecule 输入 1SEM\_nosolvent:1SEM

Analysis Type 输入 RMSD、RMSF、Properties。

点击 RMSD Reference 参数, 下拉列表中选择 Specify a Molecule。

展开 RMSD Reference 参数组, Reference Molecule 输入 1SEM\_prep :1SEM, 确保 Scope 参数设置为 Molecule。设置 Atom Group 及 Atom Group to Fit 参数为 Backbone。

点击运行。

Parameter Name	Parameter Value
Input Molecule	1SEM_nosolvent:1SEM
Analysis Type	RMSD,RMSF,Properties
RMSD Reference	Specify a Molecule
Reference Frame	
Reference Molecule	1SEM_prep:1SEM
Scope	Molecule
Label	RMSD to conf
Atom Group	1SEM:Protein Groups:Backbone
Atom Group to Fit	1SEM:Protein Groups:Backbone
Save Fitted Trajectory	True
Fit Frames for RMSF	True
Properties	HBond Monitor 1
Protein Torsion Angles	Main Chain

图 10 Analyze Trajectory 流程参数设置

双击该作业行打开 Output 文件夹和 Report.htm 文件。点击打开 View Results。

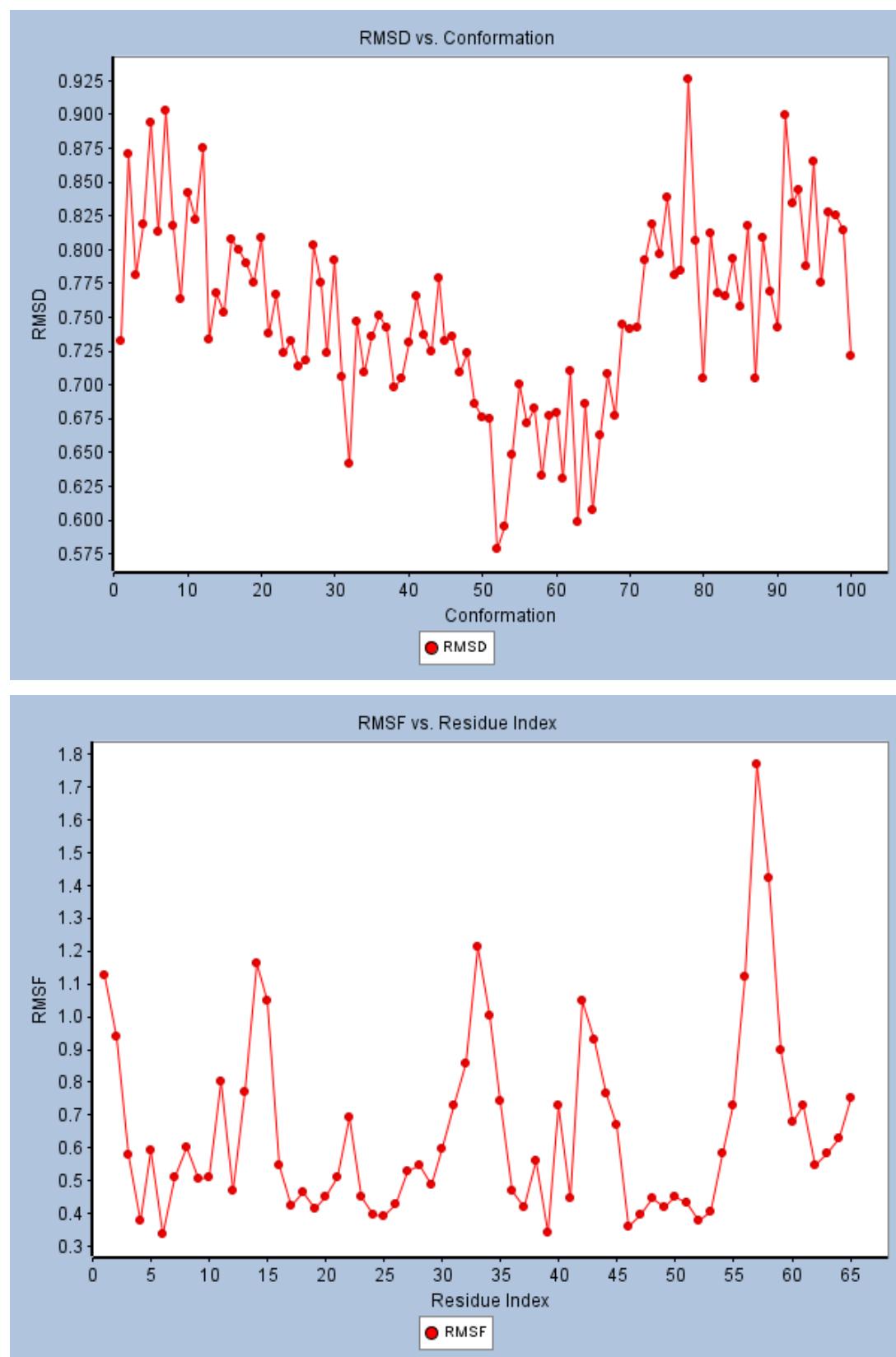


图 11: RMSD、RMSF 变化曲线

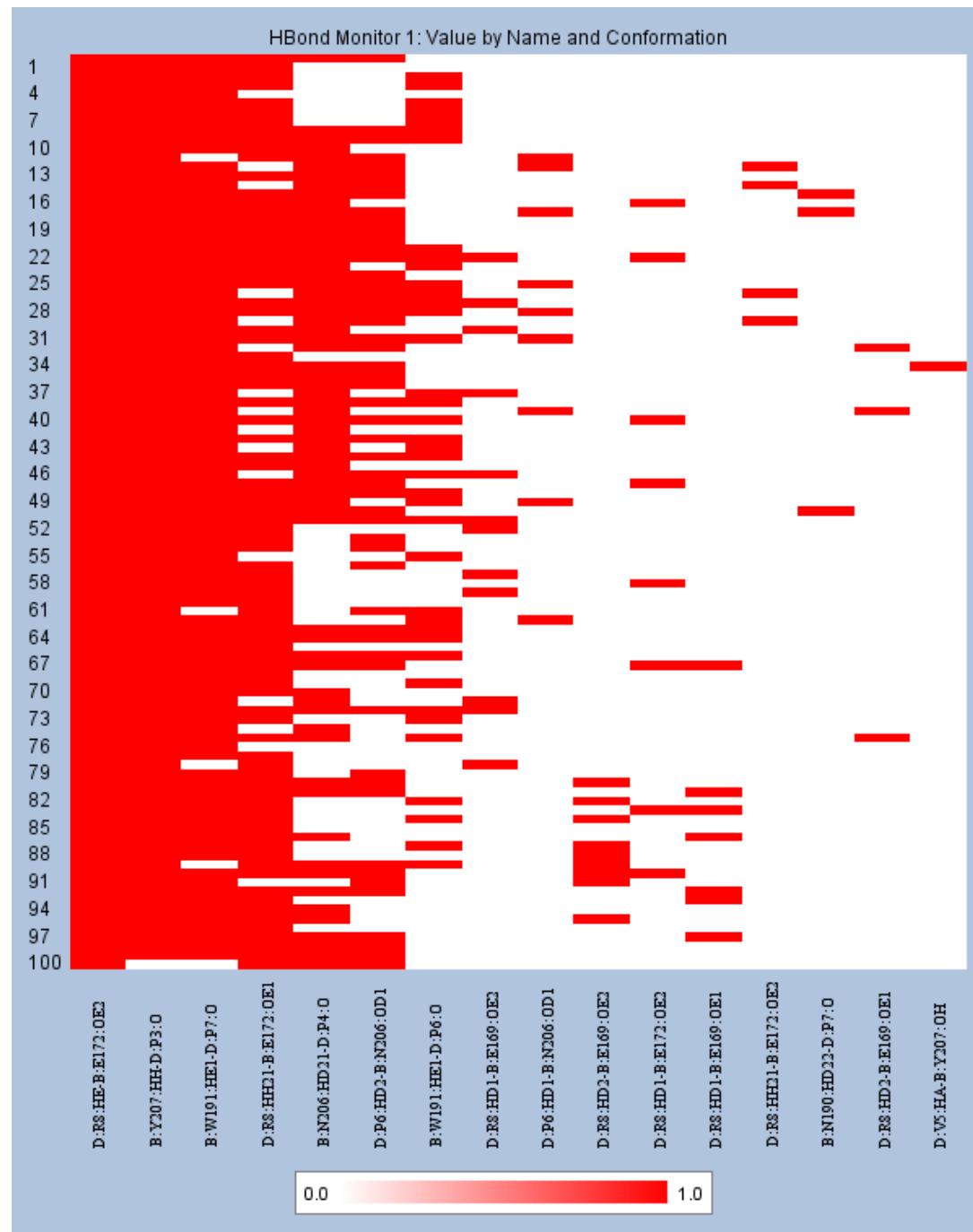


图 12 氢键热图示意图



# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

**分子力学/量子力学专题**

## QM/MM 结构优化

**目的：**采用 QM/MM 的方法对配体结构在活性位点中进行优化。

**所需功能和模块：**Discovery Studio Client, DS CHARMM。

**所需数据文件：**1UZE.dsv

**所需时间：**1 小时

### 介绍

本教程中，对一个配体分子将对接到 human testicular angiotensin I-converting enzyme 中，采用 QM/MM 的方法来优化对接所得到的结合模式。利用 CDOCKER 将配体分子对接到受体中，计算 RMSD 值，然后选择第一个 Pose 进行 QM/MM 优化。

#### 1. QM/MM 优化对接得到的配体位置

打开 1UZE.dsv，在窗口中选中 Zn 离子以及与 Zn 离子配位的配体，两个 HIS 和一个 GLU 氨基酸，点击鼠标右键，选中 Group... 选项，命名为 QMgroup。

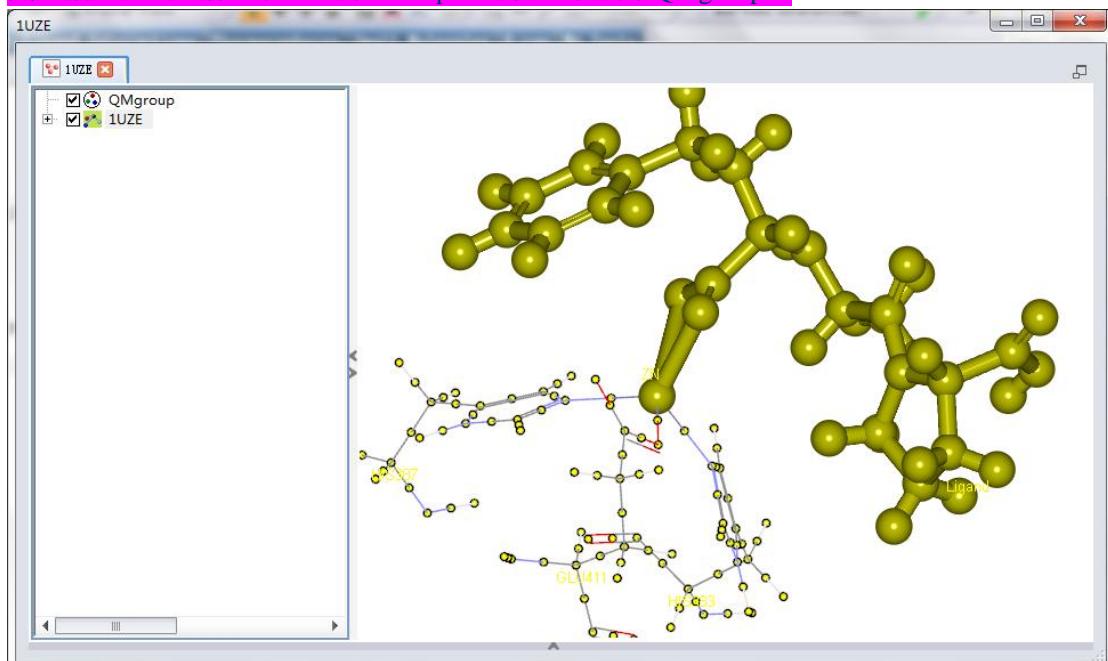


图 1

选中 1UZE 中非配体外的所有原子，然后点击 Tools | Simulation | Setup Constraints | Create Fixed Atom Constraint，

我们在接下来的 QM/MM 优化中只优化配体的空间结构，而保持金属蛋白的构型固定。在这里，我们默认晶体中金属蛋白的位置是准确的，而 CDOCKER 得到的配体位置还需要 refine，我们就沿用半柔性对接的思想，在 QM/MM 优化中，将金属蛋白的位置固定，而只优化配体的构型。金属蛋白上每个原子的位置不进行优化，但是在计算能量的时候还是包含了整个体系。（注意：QM 的区域不仅包含配体，还包括 Zn 离子，两个 HIS 和一个 GLU 氨基酸。）

可以看到金属蛋白上蓝色的小球代表了位置上被束缚的原子，在 1UZE 下面能看到定义的名字“Fixed:fix\_1”。

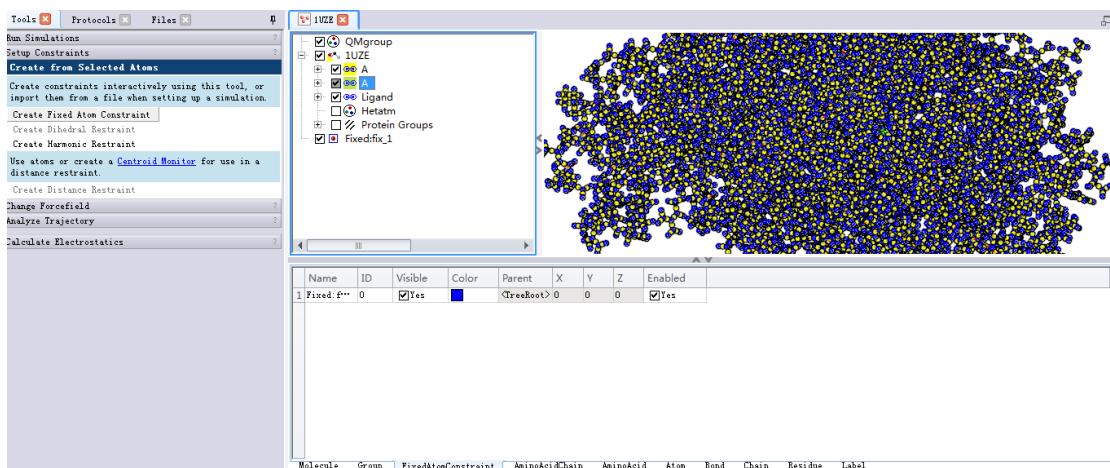


图 2

在 Protocols 浏览器中展开 Simulation 文件夹，双击 **Minimization (QM-MM)** 流程，流程对应参数在参数浏览器中打开。

在参数浏览器中，点击 **Input Typed Molecule** 参数，从下拉列表中选择“1UZE:1UZE”。

点击 **Quantum Atoms** 参数，从下拉列表中选择“QMgroup”，指定 QM 区域。

展开 **DMol3 Settings** 参数，点击 **Functional** 参数，从下拉列表中选择“BLYP”。(用户可以选择自己偏好的密度泛函)

展开 **Advanced** 下面的 **Minimization** 参数，点击 **Minimization Constraints** 参数，选中“**Fixed:fix\_1**”。

展开 **Advanced** 下面的 **DMol3 Settings** 参数，点击 **Charge** 参数，输入“0”(QM 区域的总电荷数)。我们可以仔细分析一下 QM 区域的总电荷，Zn 离子电荷为+2，配体为-1，GLU 氨基酸-1，所以总的电荷是+2-1-1=0。同时设置 **Expand Quantum Atoms** 参数，在下拉菜单中选中“**False**”。当我们设定了 QM 区域的 charge 之后，这一项必须选为“false”。(如果设为“true”的话，DS 会自动根据 QM 区域内每个原子的电荷来算出总的电荷数)。

选中 **Advanced** 下面的 **Number of Processors** 参数，输入“8”(需要进行并行计算的核的数目)。(这个需要根据具体机器来设定，由于 QM 计算比较耗时，我们推荐使用尽可能多的核来做并行计算。)

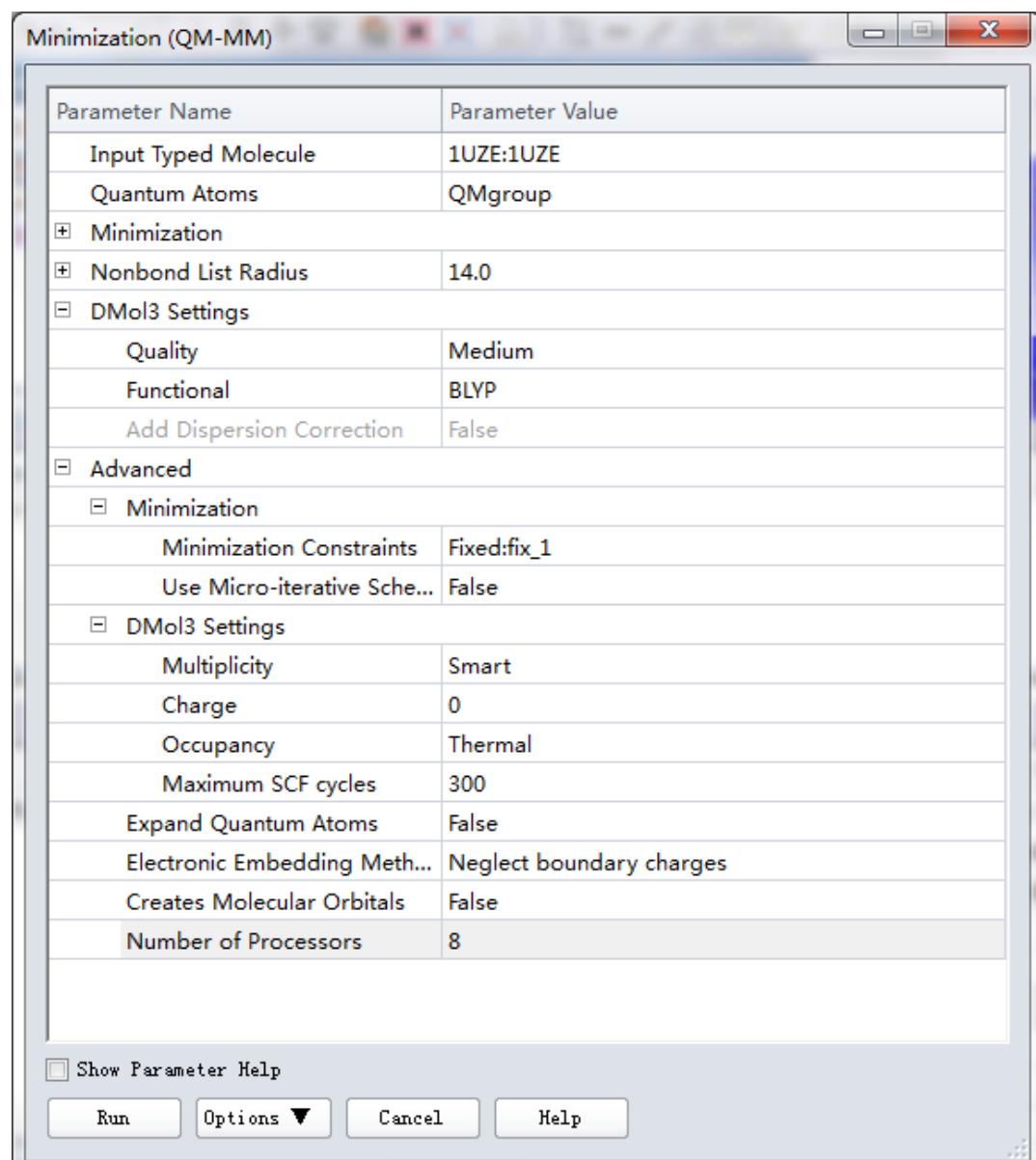


图 3

点击流程工具栏 (Protocols toolbar) 中的 按钮运行作业，等待作业完成。这个作业用 8 核并行计算，大约需要花一天的时间。

## 2. 分析 QM/MM 优化配体后的结果

在 Report.htm 的 Summary 中，我们可以分别看到 QM 区域，MM 区域的能量以及 QM 与 MM 区域之间的相互作用能，[点击 Output Files 中的 Output Molecules](#)，查看 QM/MM 优化后的配体结构。

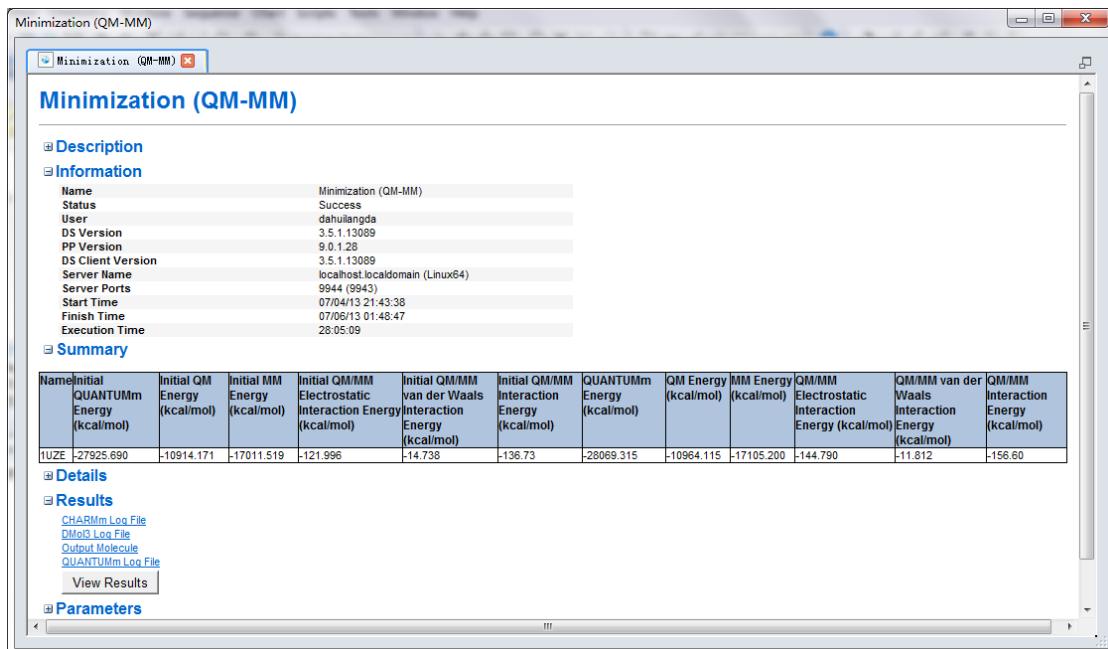


图 4

同时打开先前保存的 rmsd.csv 文件，在 1UZE.csv 中选中 ligand，然后将它复制到 rmsd.csv 窗口中，如下图所示，将优化后的四个配体分别一一复制到 rmsd.csv 窗口，并赋予名字 Pose1。

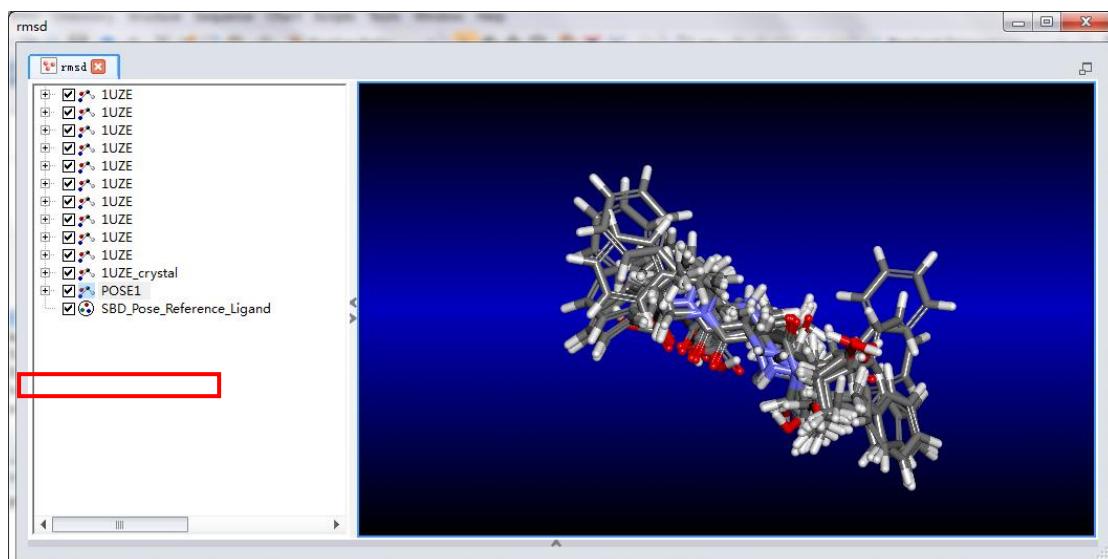


图 5

用之前介绍的计算配体关于晶体结构中配体位置的 RMSD 的方法，我们可以得到如下的 Report，

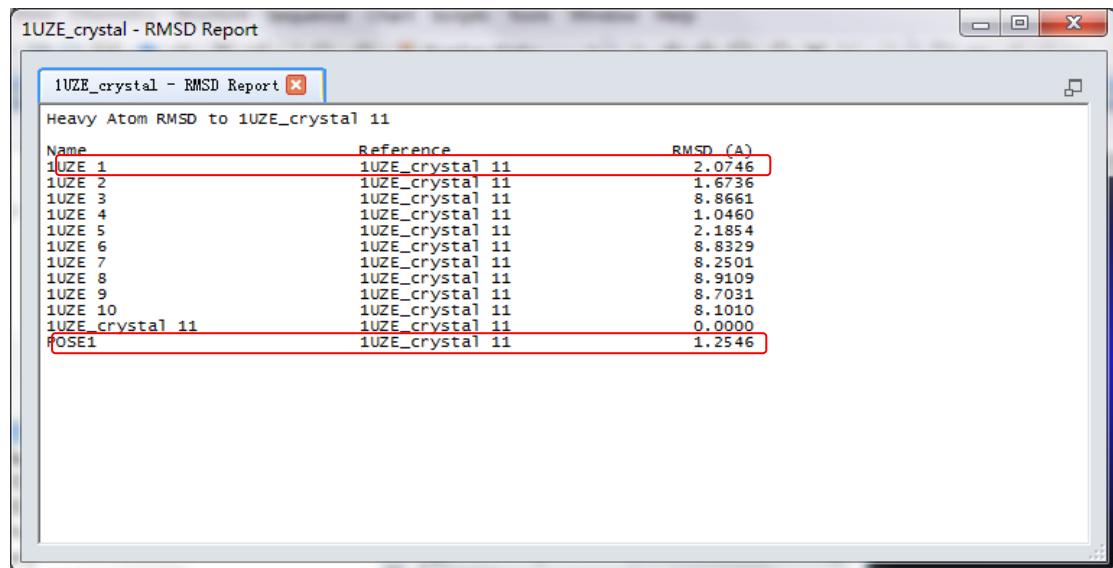


图 6

我们列出了 QM/MM 优化后配体构型关于晶体结构中配体位置的 RMSD（单位为埃）与 CDOCKER 的结果对比。在用 QM/MM 优化后，平均的 RMSD 由 2.07 埃降到了 1.25 埃。从这个案例看出，基于从头计算(ab initio)量子力学的 QM/MM 方法可以为我们提供更为精确的对接模型，为我们开展计算机辅助药物设计的方法提高到一个新的台阶。

# **Discovery Studio 上机操作 Tutorials**

## **(Discovery Studio 版本: 4.5)**

蛋白质设计专题:  
虚拟氨基酸突变、蛋白质二硫键的预测、  
蛋白质聚集效应的预测

## 虚拟氨基酸突变（Calculate Mutation Energy）教程

**目的：**通过此教程，了解 Discovery Studio 中虚拟氨基酸突变的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS CHARMM

**所需数据文件：**1aq1.pdb, 2sta\_I.pdb

**所需时间：**60 分钟

### 介绍

蛋白的氨基酸定点突变可以用于酶与抗体的设计，但是由于进行氨基酸选择时的盲目性而导致效率低下。虚拟氨基酸突变可以通过丙氨酸扫描和饱和突变确定最佳的氨基酸突变组合，从而为实验中的氨基酸定点突变提供指导。

本教程使用 Calculate Mutation Energy (Binding) 对一个蛋白-配体复合物进行基于相互作用力的虚拟氨基酸突变，确定了活性位点中的关键氨基酸，以及能提高亲和力的氨基酸突变目标。使用 Calculate Mutation Energy (Stability) 对一个蛋白进行基于热稳定性的虚拟氨基酸突变，预测了能提高蛋白热稳定性的突变目标，并利用 Predict Stabilizing Mutations 预测了最佳的氨基酸突变组合。

本教程涵盖如下内容：

- 虚拟氨基酸突变提高酶与底物的亲和力
- 虚拟氨基酸突变提高蛋白热稳定性
- 预测提高热稳定性的最佳氨基酸突变组合

### 虚拟氨基酸突变提高酶与底物的亲和力

在文件浏览器（Files Explorer）中，找到 Samples| Tutorials| Receptor Ligand Interaction 中的 **1aq1.pdb**，双击打开在分子窗口中显示。

分子窗口中展示出了一个带有配体的蛋白的结构（图 1）

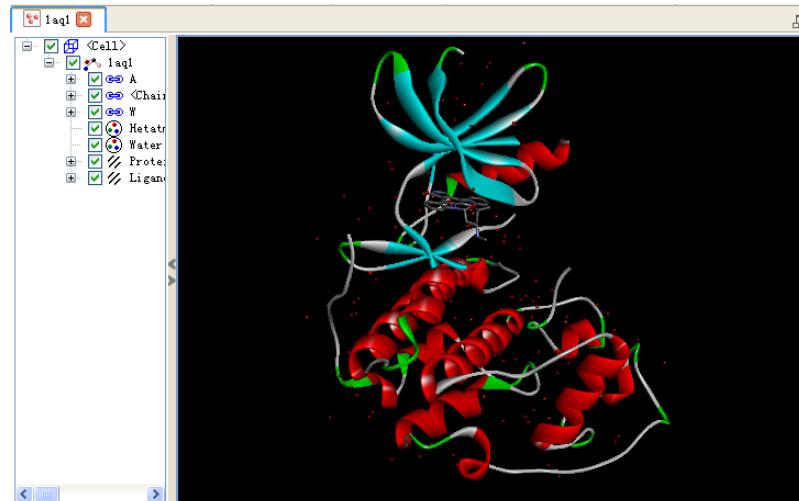


图 1

在 Hierarchy 窗口 (CTRL+H) 中选择 Water，点击键盘 Delete 以删去结晶结构中的结晶水。

在 Hierarchy 窗口 (CTRL+H) 中双击 1aq1 的<Chain>链，将配体重命名为 Ligand。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Prepare Protein，点击 Clean Protein 对蛋白的结构进行预处理。

然后在工具浏览器（Tools Explorer）中，展开 Simulation | Change Forcefield，点击 Apply Forcefield，将蛋白赋上 CHARMM 力场。

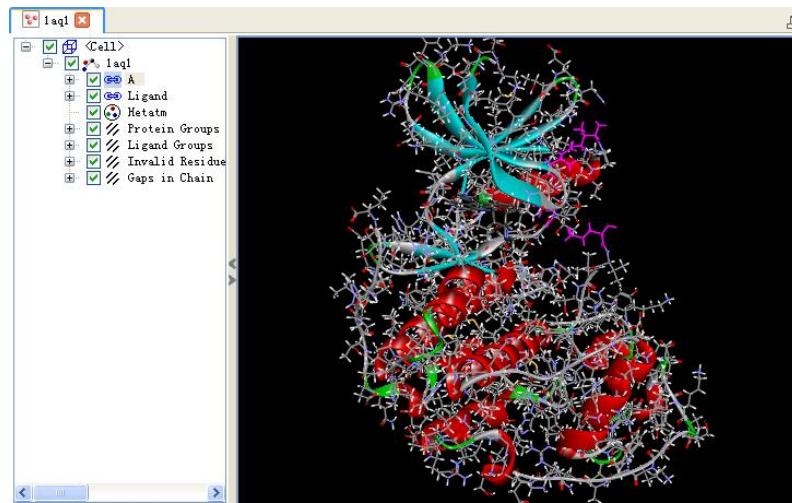


图 2

经过以上这些处理之后，在 Hierarchy 窗口中单击选中 Ligand，然后点击菜单栏 Edit | Select，在弹出的窗口中设置 Selection Mode 为 Replace(New Only)，Radius 为 3，然后点击 Apply。（图 3）

以选中配体周围 3 埃内的所有氨基酸作为后续突变点。

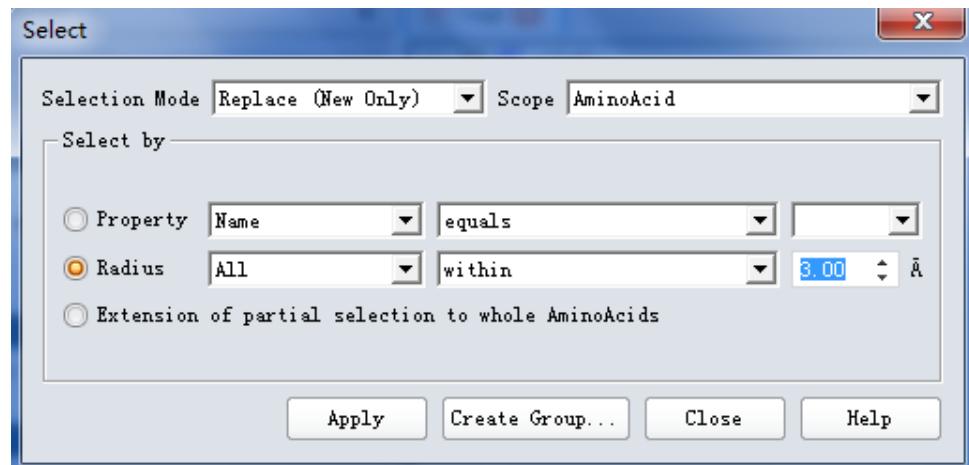


图 3

选中以后，点击鼠标右键，选择 Group，并在弹出的对话框中输入 mutation，这样我们就将配体 3 埃以内的所有氨基酸定义为一个名为 mutation 的组。

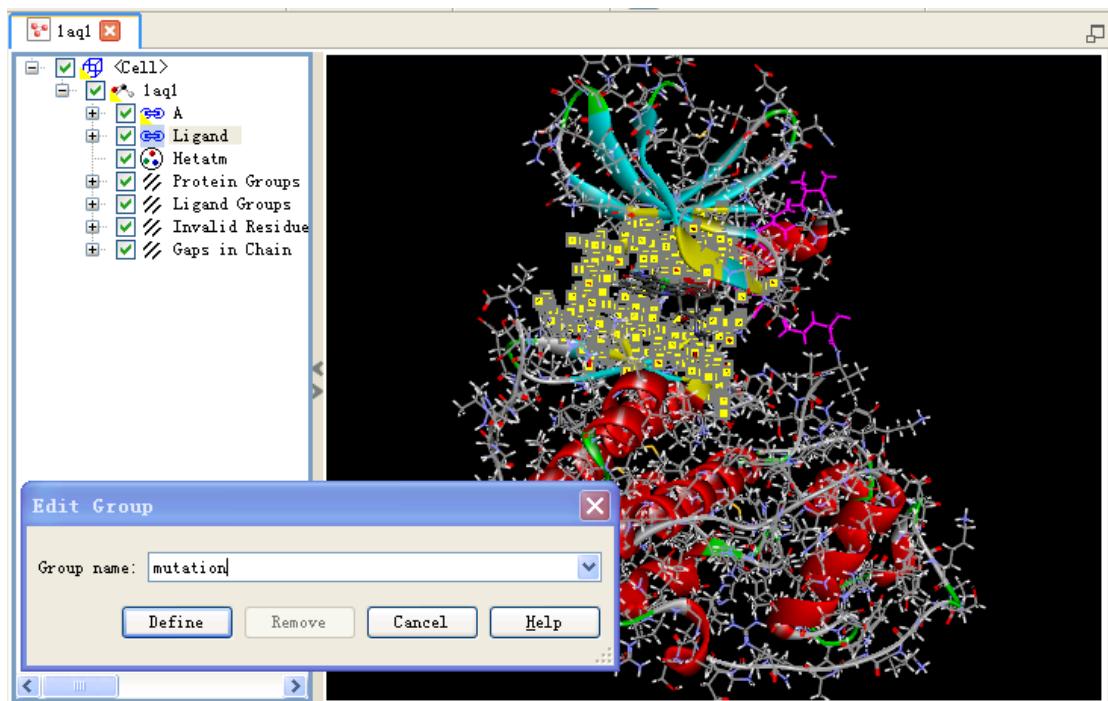


图 4

在工具浏览器（Tools Explorer）中，展开 Macromolecules| Design Protein，点击 Calculate Mutation Energy (Binding)。

流程对应参数在参数浏览器中打开。

设置 Input Typed Molecule 为 1aq1:1aq1。

设置 Ligand Chain 为 1aq1:Ligand\_2。

设置 Mutation Site 为 Single Mutations。

**Mutation Sites** 有四个模式可以选择：

Single Mutations：被选择的氨基酸残基进行单点突变。如果将 2 个氨基酸残基突变成 2 种不同的氨基酸类型，将会进行 4 次独立的突变。

Double Mutations：被选择的氨基酸残基进行两点同时突变。如果将 N 个氨基酸残基突变成 M 种不同的氨基酸类型，将会两两组合，进行  $[N!/2(N-2)!](M^2)$  次突变。

Triple Mutations：与 Double Mutations 类似，但进行三点同时突变。

All Selected Residues：将被选择的氨基酸残基同时突变。

本次教程将选择 Single Mutations 模式。

接着我们需要在 Mutations 参数栏设置氨基酸残基的突变类型。

点击 Mutations 参数栏右侧的 ，弹出如下图所示的对话框。我们这里将之前定义的 mutation 组中所有的氨基酸突变为 ALA。（图 5）

即将 mutation 组中氨基酸进行丙氨酸扫描。

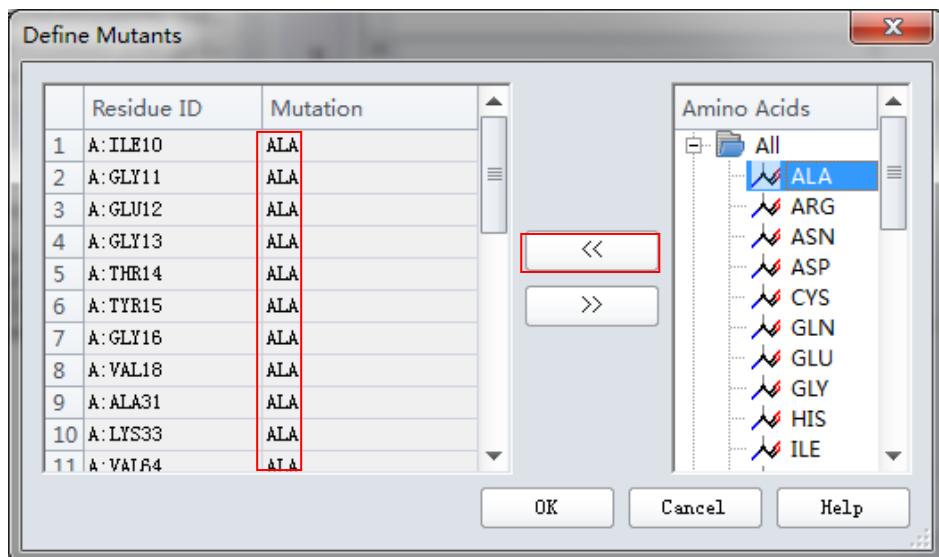


图 5

其余参数都采用默认设置。(图 6)

点击 Run 运行该 Protocol。

点击 Background 等待作业完成。

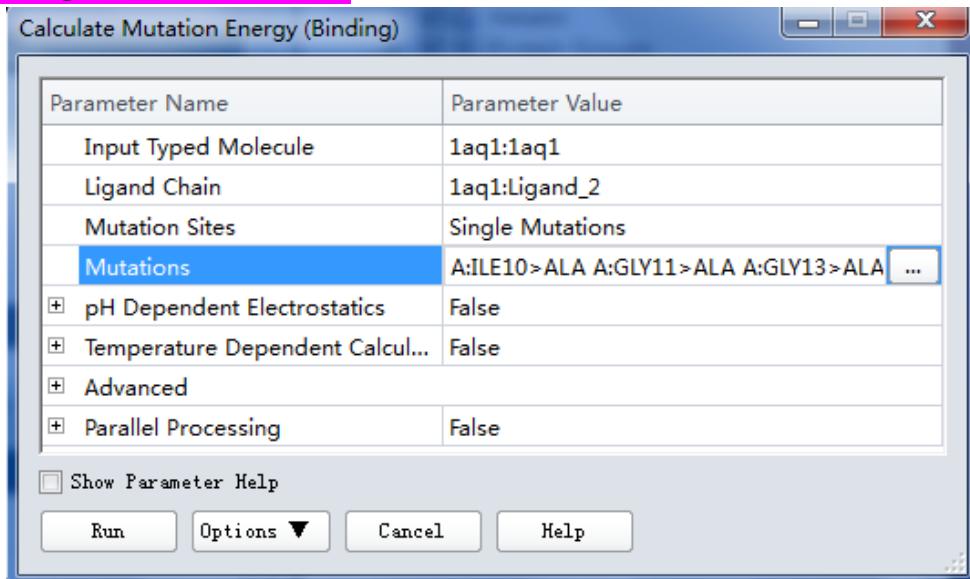


图 6

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

在 Report 页面的 **Summary** 栏，会出现一个表格(图 7)。该表格是按照突变能(Mutation Energy)从低到高对氨基酸突变的结果进行排序，只展示突变能最高的和最低的各 5 个氨基酸突变。对于每个氨基酸突变，表格的最右侧一列都有一个 Effect 评估，如果突变能在 -0.5 到 +0.5 之间，那么 Effect 为 neutral，即这种突变对于亲和力没有影响；如果突变能在 0.5 以上，那么 Effect 为 destabilizing，即这种突变会导致亲和力降低，相互作用关系减弱；如果突变能在 -0.5 以下，那么 Effect 为 stabilizing，即这种突变会导致亲和力上升，相互作用关系增强。

Index	Mutation	Mutation Energy (kcal/mol)	Effect
1	A:GLN131>ALA	-0.04	NEUTRAL
2	A:GLN85>ALA	-0.03	NEUTRAL
3	A:ALA144>ALA	-0.01	NEUTRAL
4	A:ALA31>ALA	0.00	NEUTRAL
5	A:HIS84>ALA	0.04	NEUTRAL
15	A:PHE80>ALA	0.82	DESTABILIZING
16	A:PHE82>ALA	0.96	DESTABILIZING
17	A:ILE10>ALA	1.01	DESTABILIZING
18	A:LEU134>ALA	1.89	DESTABILIZING
19	A:GLY11>ALA	6.28	DESTABILIZING

图 7

点击 Results 栏中 Mutation Energy Terms 链接。

打开一张含所有突变位点突变信息的表格。(图 8)

双击 Mutation Energy 表头，可按照突变能从低到高排列。

	Mutation	Mutation Energy ▲	Effect of Mutation	VDW Term	Electrostatic Term	Entropy Term	Non-polar Term
1	A:GLN131>ALA	-0.04	NEUTRAL	0.2	0	-0.17	0
2	A:GLN85>ALA	-0.03	NEUTRAL	-0.11	0.03	0.01	0
3	A:ALA144>ALA	-0.01	NEUTRAL	-0.05	0.03	0	0
4	A:ALA31>ALA	0	NEUTRAL	0	0	0	0
5	A:HIS84>ALA	0.04	NEUTRAL	0.06	0.03	0	0
6	A:ASN132>ALA	0.05	NEUTRAL	0.1	-0.01	0.01	0
7	A:GLU81>ALA	0.07	NEUTRAL	0.18	-0.05	0	0
8	A:ASP145>ALA	0.14	NEUTRAL	0.44	-0.13	-0.02	0
9	A:VAL64>ALA	0.33	NEUTRAL	0.6	0.04	0.01	0
10	A:GLY13>ALA	0.35	NEUTRAL	0.74	-0.02	-0.01	0
11	A:LYS33>ALA	0.37	NEUTRAL	1.01	-0.34	0.04	0
12	A:ASP86>ALA	0.39	NEUTRAL	1.44	-0.66	0	0
13	A:LEU83>ALA	0.51	DESTABILIZING	1.02	-0.01	0.01	0
14	A:VAL18>ALA	0.81	DESTABILIZING	1.55	0.08	-0.01	0
15	A:PHE80>ALA	0.82	DESTABILIZING	1.61	0.02	0.01	0
16	A:PHE82>ALA	0.96	DESTABILIZING	1.84	-0.01	0.06	0
17	A:ILE10>ALA	1.01	DESTABILIZING	1.86	0.13	0.02	0
18	A:LEU134>ALA	1.89	DESTABILIZING	3.79	0.02	-0.02	0
19	A:GLY11>ALA	6.28	DESTABILIZING	12.34	0.22	0	0

图 8

可以看到 Leu83、Val18、Phe80、Phe82、Ile10、Leu134 和 Gly11 氨基酸突变为 ALA 之后会导致亲和力降低，推测这些氨基酸是受体与配体相互作用的关键氨基酸。

回到 1aq1 的分子窗口，在 Hierarchy 窗口 (CTRL+H) 中选择 Leu83、Val18、Phe80、Phe82、Ile10、Leu134 和 Gly11 这 7 个氨基酸，点击鼠标右键选择 Group，命名为 keyresidues。(图 9)

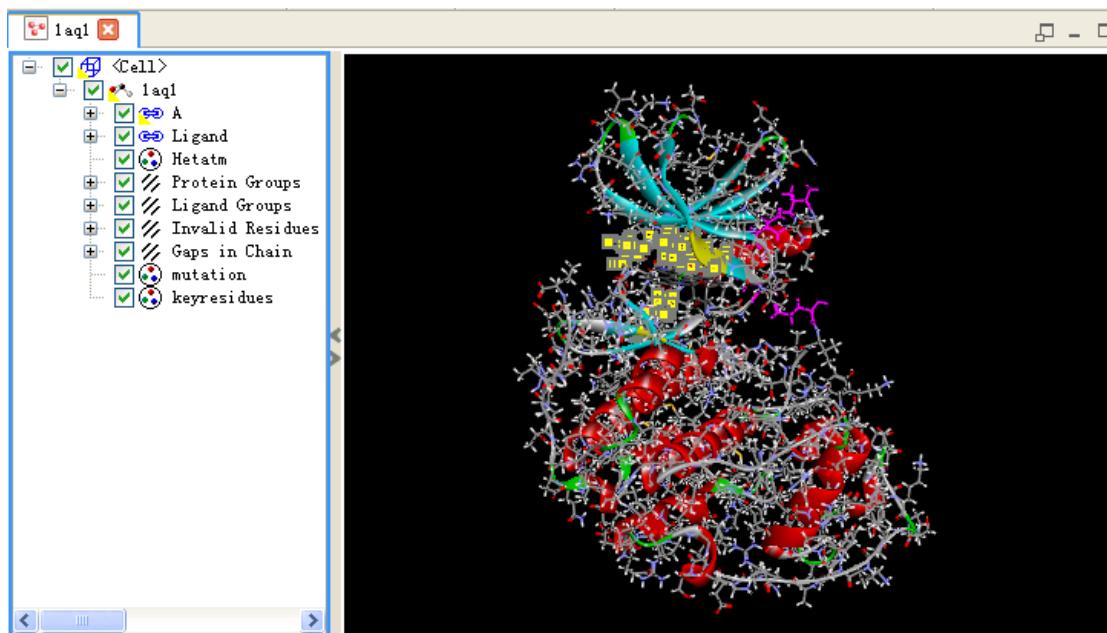


图 9

在工具浏览器（Tools Explorer）中，展开 Macromolecules| Design Protein，点击 Calculate Mutation Energy (Binding)。

流程对应参数在参数浏览器中打开。

设置 Input Typed Molecule 为 1aq1:1aq1。

设置 Ligand Chain 为 1aq1:Ligand\_2。

设置 Mutation Site 为 Single Mutations。

点击 Mutantions 参数栏右侧的 ..., 弹出如下图所示的对话框，将之前定义的 keyresidues 组中所有的氨基酸突变为其余 19 种标准氨基酸（右侧 ALL 中除了 ASPH、GLUH、LYSN、ARGN、HSC 之外的所有氨基酸）。（图 10）

即将 keyresidues 组中氨基酸进行饱和突变。

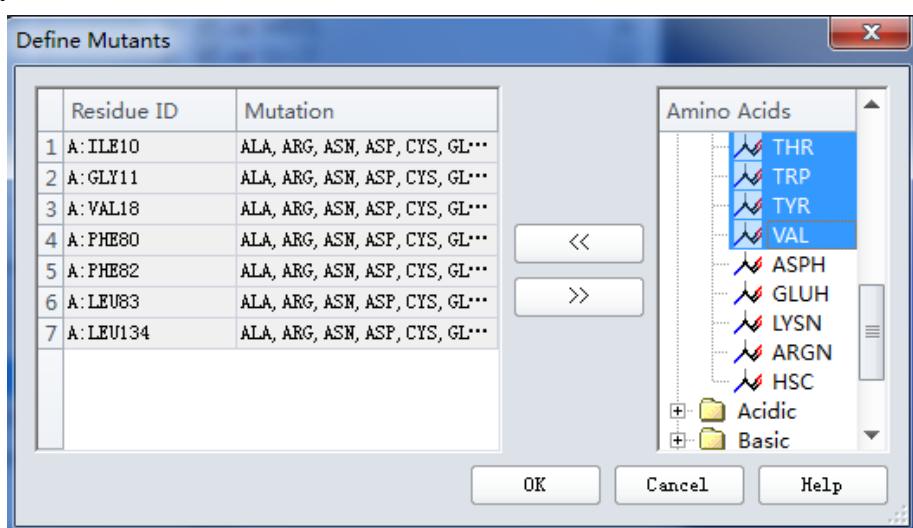


图 10

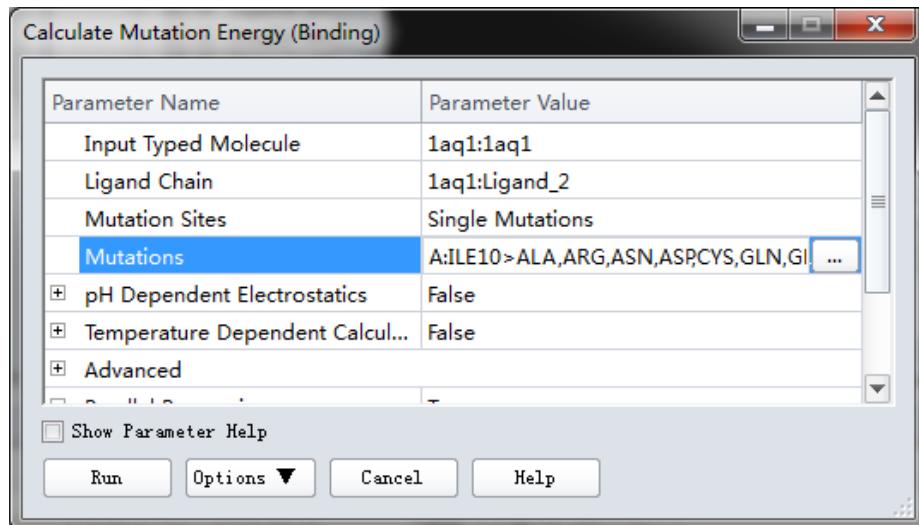


图 11

其余参数采用默认设置。(图 11)

点击 Run 运行该作业。

点击 Background 等待作业完成。

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

点击 Results 栏中 Mutation Energy Terms 链接。

打开一张含所有突变位点突变信息的表格。(图 12)

双击 Mutation Energy 表头，可按照突变能从低到高排列。

Mutation	Mutation Energy	Effect of Mutation	VDW Term	Electrostatic Term	Entropy Term	Non-polar Term
1 A:VAL18>TTR	-0.93	STABILIZING	-2.33	0.37	0.06	0
2 A:VAL18>HIS	-0.06	STABILIZING	-1.72	-0.01	0.01	0
3 A:VAL18>I12	-0.72	STABILIZING	-1.55	0.1	0.01	0
4 A:VAL18>ASN	-0.35	NEUTRAL	-1.02	0.28	0.02	0
5 A:ILE10>MET	-0.18	NEUTRAL	-0.52	0.07	0.06	0
6 A:PHE82>TTR	-0.18	NEUTRAL	-0.4	0.03	0.01	0
7 A:LEU83>TTR	-0.16	NEUTRAL	-0.35	0.03	0	0
8 A:ILE10>I12	-0.15	NEUTRAL	-0.37	0.05	0.01	0
9 A:LEU83>PHE	-0.12	NEUTRAL	-0.26	0.03	-0.01	0
10 A:LEU83>I12	-0.11	NEUTRAL	-0.27	0.05	0	0
11 A:GLY11>GLY	-0.06	NEUTRAL	-0.21	0.07	0.01	0
12 A:VAL18>LEU	-0.05	NEUTRAL	-0.19	0.06	0.01	0
13 A:PHE82>PHE	-0.05	NEUTRAL	-0.12	0.01	0	0
14 A:VAL18>THR	-0.05	NEUTRAL	-0.29	0.12	0.04	0
15 A:PHE80>TTR	-0.05	NEUTRAL	-0.13	0.03	0	0
16 A:VAL18>VAL	-0.04	NEUTRAL	-0.12	0.03	0	0
17 A:PHE80>TRP	-0.04	NEUTRAL	-0.13	0.09	-0.02	0
18 A:PHE80>ARG	-0.03	NEUTRAL	-0.3	0.2	0.03	0
19 A:PHE80>PHE	0	NEUTRAL	0	0	0	0
20 A:LEU83>LEU	0.01	NEUTRAL	0.01	0	0	0
21 A:VAL18>PHE	0.02	NEUTRAL	-0.23	0.23	0.03	0
22 A:PHE82>HIS	0.09	NEUTRAL	0.19	-0.04	0.02	0
23 A:PHE80>LEU	0.11	NEUTRAL	0.13	0.05	0.02	0
24 A:VAL18>ASN	0.14	NEUTRAL	-0.94	1.22	0	0
25 A:LEU83>VAL	0.14	NEUTRAL	0.23	0.04	0	0
26 A:ILE10>LEU	0.15	NEUTRAL	0.21	0.05	0.03	0

图 12

在结果中可以看到，当 Val18 变为 Tyr、His、Ile 三种氨基酸可以使得受体配体之间的亲和力有所提高。这些突变目标的预测，可以指导我们进行合理的氨基酸突变，从而可能提高酶的活力。

### 多点同时突变提高酶与底物的亲和力

回到 1aq1 的分子窗口，在 Hierarchy 窗口中选中 Ile10、Phe80、Phe82 这三个氨基酸。

在工具浏览器（Tools Explorer）中，展开 Macromolecules| Design Protein，点击 Calculate Mutation Energy (Binding)。

流程对应参数在参数浏览器中打开。

设置 Input Typed Molecule 为 1aq1:1aq1。

设置 Ligand Chain 为 1aq1:Ligand\_2。

设置 Mutation Site 为 Triple Mutations。

Mutation Sites 设为 Triple Mutations，最终的输出结果将是三个氨基酸突变的组合。

点击 Mutantions 参数栏右侧的 ，在弹出的对话框中选择如下突变方式：Ile10>MET、Phe80>Trp、Phe82>Tyr。

其余参数采用默认设置。(图 13)

点击 Run 运行该作业。

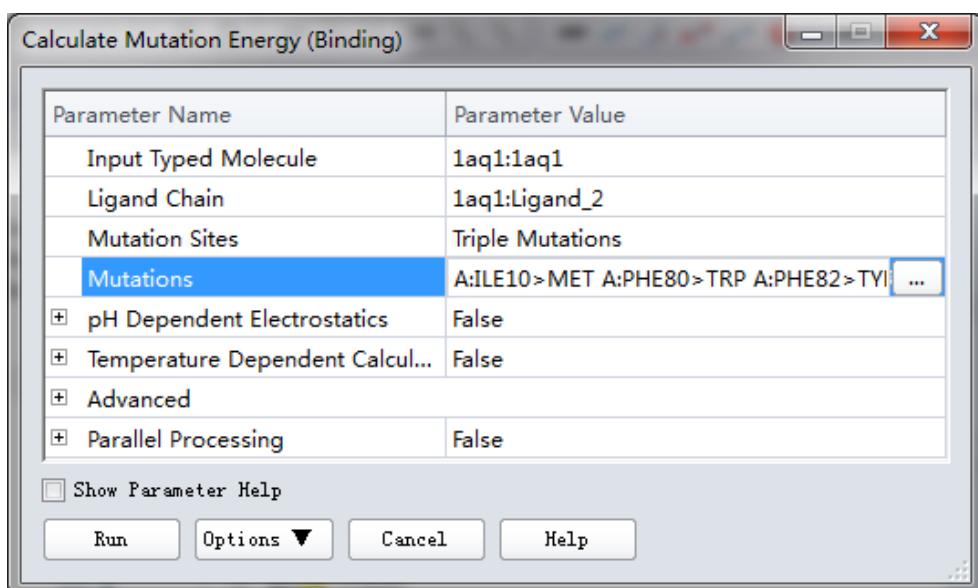


图 13

点击 Background 等待作业完成。

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

点击 Results 栏中 Mutation Energy Terms 链接。

打开一张三点突变信息的表格。(图 14)

双击 Mutation Energy 表头，可按照突变能从低到高排列。

	Mutation	Mutation Energy	Effect of Mutation	VDW Term	Electrostatic Term	Entropy Term	Non-polar Term
1	A:ILE10>ME...	-0.53	STABILIZING	-1.3	0.17	0.05	0

图 14

三点同时突变后，突变能为-0.53，亲和力增强。通过多点同时突变技术，我们能够考虑多点突变的协同效应对亲和力的影响。

### 虚拟氨基酸突变提高热稳定性

在文件浏览器(Files Explorer)中，找到 Samples| Tutorials| Protein Modeling 中的 2sta\_I.pdb，双击打开在分子窗口中显示。

分子窗口中展示出了一个蛋白多肽的结构。(图 15)

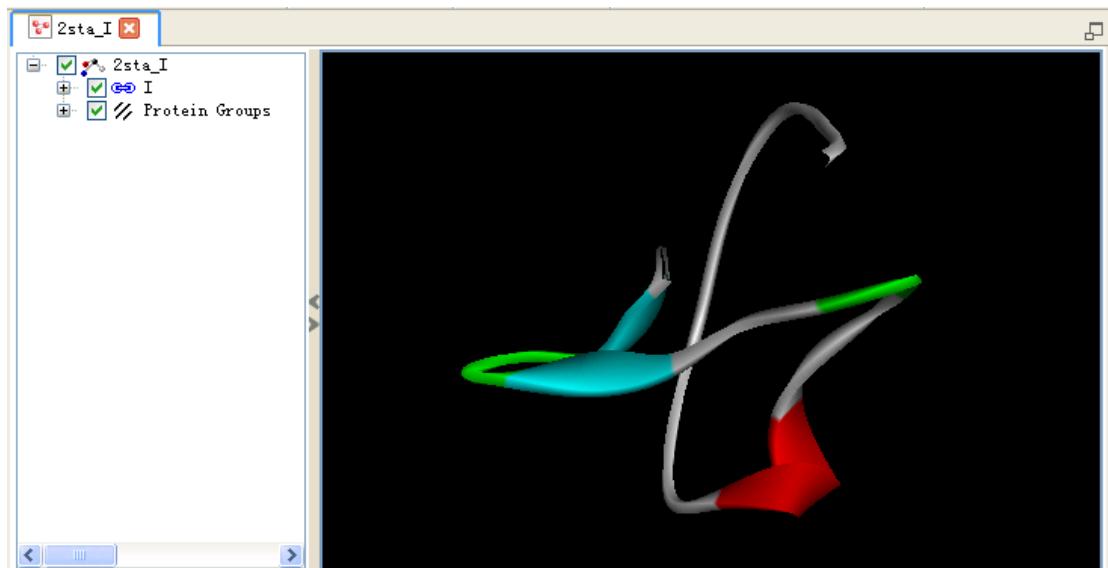


图 15

在工具浏览器（Tools Explorer）中，展开 Simulation | Change Forcefield，点击 Apply Forcefield，将蛋白赋上 CHARMm 力场。

在 Hierarchy 窗口中，选择 Pro504-Glu509 这 6 个氨基酸，点击鼠标右键选择 Group 将其命名为 mutation（图 16）。

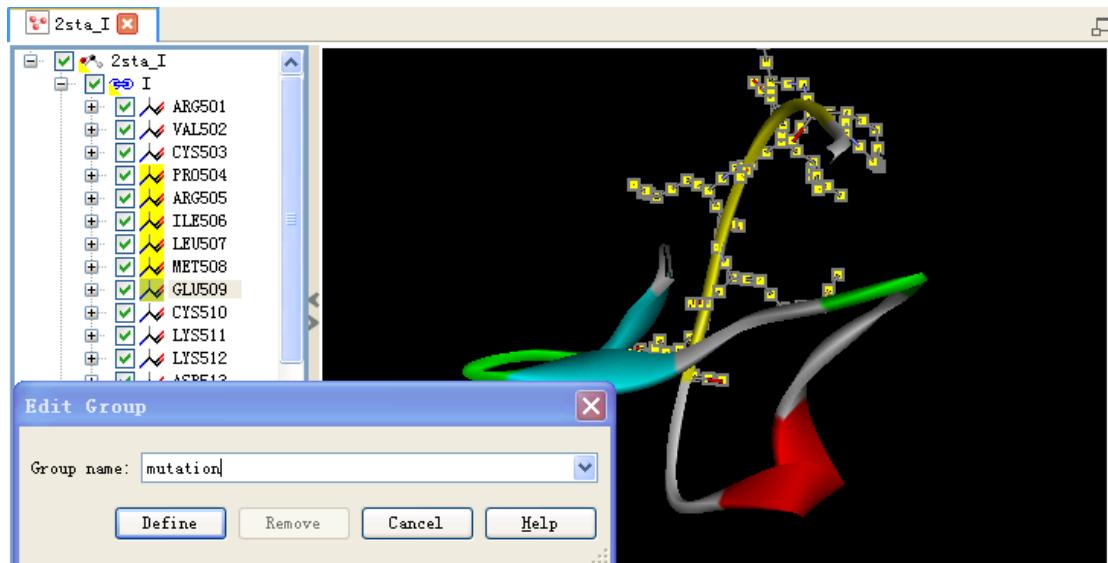


图 16

在工具浏览器（Tools Explorer）中，展开 Macromolecules| Design Protein，点击 Calculate Mutation Energy (Stability)。

流程对应参数在参数浏览器中打开。

设置 Input Typed Molecule 为 2sta\_I:2sta\_I。

设置 Mutation Site 为 Single Mutations。

点击 Mutantions 参数栏右侧的 ，在弹出的对话框中，将之前定义的 mutation 组中所有的氨基酸突变为其余 19 种标准氨基酸（右侧 ALL 中除了 ASPH、GLUH、LYSN、ARGN、

HSC 之外的所有氨基酸)。

其余参数采用默认设置。(图 17)

点击 Run 运行该作业。

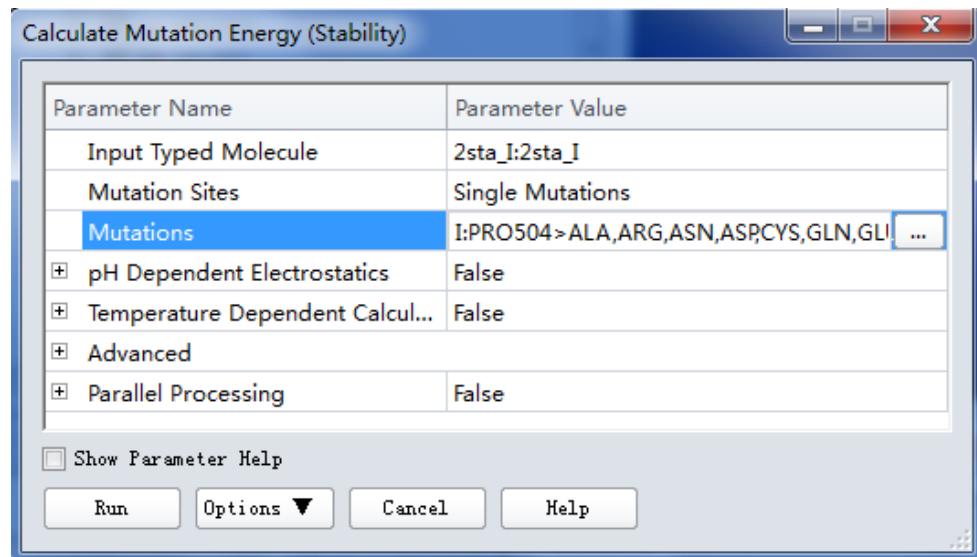


图 17

点击 Background 等待作业完成。

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

点击 Results 栏中 Mutation Energy Terms 链接。

打开一张含所有突变位点突变信息的表格。(图 18)

双击 Mutation Energy 表头，可按照突变能从低到高排列。

	Mutation	Mutation Energy	Effect of Mutation	VDW Term	Electrostatic Term	Entropy Term
1	I:GLU509>TRP	-1.4	STABILIZING	-2.4	-0.95	0.34
2	I:MET508>ILE	-1.2	STABILIZING	-0.83	-0.08	-0.94
3	I:GLU509>PRO	-0.86	STABILIZING	0.12	-1.28	-0.37
4	I:GLU509>CYS	-0.83	STABILIZING	1.59	-1.42	-1.14
5	I:GLU509>GLN	-0.83	STABILIZING	-0.55	-1.18	0.04
6	I:MET508>HIS	-0.81	STABILIZING	-0.61	-0.12	-0.56
7	I:ARG505>TRP	-0.73	STABILIZING	-1.48	0.04	-0.01
8	I:GLU509>LEU	-0.72	STABILIZING	-0.19	-1.22	-0.02
9	I:MET508>LEU	-0.69	STABILIZING	-0.05	-0.11	-0.77
10	I:ARG505>CYS	-0.68	STABILIZING	0.43	-0.16	-1.02
11	I:MET508>THR	-0.61	STABILIZING	0.36	-0.67	-0.57
12	I:MET508>TRP	-0.58	STABILIZING	0.37	0.23	-1.1
13	I:GLU509>TYR	-0.58	STABILIZING	-0.8	-1	0.4

图 18

从结果中可以看到，Glu509 突变为 TRP，突变能为-1.40，对应的 Effect 是 Stabilizing，即这种突变会使蛋白的热稳定性提高，而最后一行的 Ile506 突变为 Pro，突变能是 7.65，即突变之后整个蛋白的能量提高，对应的 Effect 是 destabilizing，也就是说这种突变会使蛋白的热稳定性降低。

## 预测提高热稳定性的最佳氨基酸突变组合

在工具浏览器 (Tools Explorer) 中, 展开 Macromolecules| Design Protein, 点击 Predict Stabilizing Mutations。

流程对应参数在参数浏览器中打开。

设置 Input Typed Molecule 为 2sta\_I:2sta\_I。

设置 Residues to Mutate 为选中的 Pro504-Glu509 这 6 个氨基酸。

点击 Mutante to 参数栏右侧的 , 在弹出的对话框中, 选择其余 19 种标准氨基酸。(图 19)

设置 Maximum Mutations 为 Triple Mutations。

其余参数采用默认设置。(图 20)

点击 Run 运行该作业。

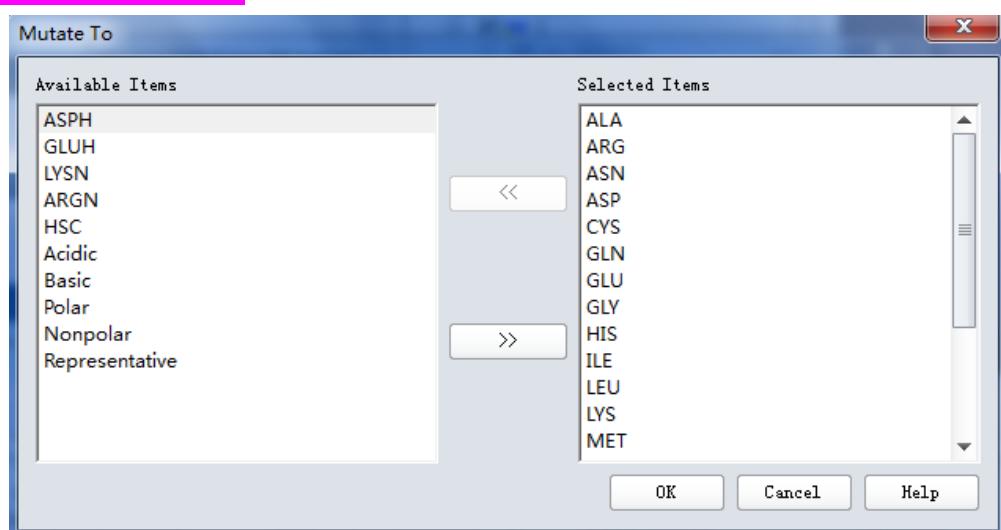


图 19

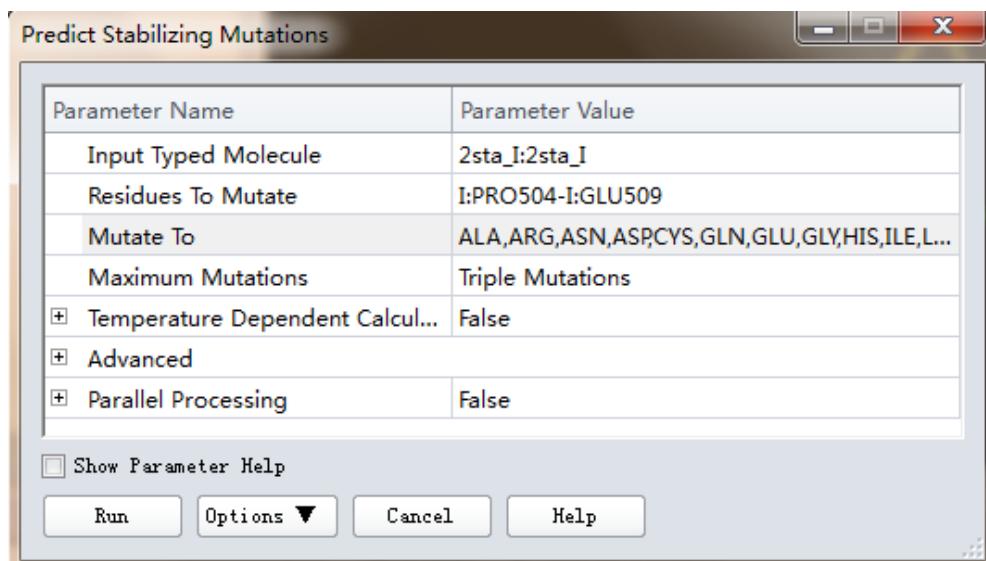


图 20

点击 Background 等待作业完成。

作业完成后，双击作业浏览器（Jobs Explorer）中相应的行，打开 Report 页面。

点击 Results 栏中 Mutation Energy Terms (Triple Mutations)链接。

打开一张含所有三点突变突变信息的表格。（图 21）

双击 Mutation Energy 表头，可按照突变能从低到高排列。

Mutation	Mutation Energy	Effect of Mutation	VDW Term	Electrostatic Term	Entropy Term	Non-polar Term
1 I-MET508 ILE,I-ARG505 CYS,I-GLU509 PRO	-3.52994	stabilizing	-1.98165	-1.53345	-2.21821	0
2 I-MET508 ILE,I-ARG505 CYS,I-GLU509 TRP	-3.43213	stabilizing	-3.1445	-1.19162	-1.58008	0
3 I-MET508 ILE,I-ARG505 TRP,I-GLU509 CYS	-3.39619	stabilizing	-4.8548	-1.00382	-0.58359	0
4 I-MET508 ILE,I-ARG505 TRP,I-GLU509 PRO	-3.32767	stabilizing	-3.3566	-1.36476	-1.20874	0
5 I-MET508 ILE,I-LEU507 CYS,I-GLU509 TRP	-3.26212	stabilizing	-2.1273	-1.43887	-1.84688	0
6 I-MET508 ILE,I-LEU507 CYS,I-GLU509 LEU	-3.21985	stabilizing	-2.12623	-1.34456	-1.65557	0
7 I-LEU507 PHE,I-ARG505 TRP,I-MET508 ILE	-3.21054	stabilizing	-5.3368	0.04702	-0.707065	0
8 I-MET508 ILE,I-ILE506 CYS,I-GLU509 PRO	-3.2012	stabilizing	-0.81598	-1.68308	-2.43962	0
9 I-MET508 ILE,I-ILE506 CYS,I-GLU509 VAL	-3.18714	stabilizing	-1.43941	-1.32976	-2.25232	0
10 I-LEU507 PHE,I-ARG505 THR,I-MET508 LEU	-3.18476	stabilizing	-5.4897	0.00214	-0.55123	0
11 I-MET508 ILE,I-LEU507 CYS,I-GLU509 LEU	-3.13636	stabilizing	-1.07877	-1.59215	-2.25113	0
12 I-MET508 ILE,I-ARG505 ILE,I-GLU509 PRO	-3.1193	stabilizing	-2.60026	-1.37556	-1.41425	0
13 I-LEU507 CYS,I-MET508 ARG,I-GLU509 PRO	-3.11381	stabilizing	-2.02991	-1.52983	-1.66818	0
14 I-MET508 ILE,I-LEU507 CYS,I-GLU509 GLN	-3.05332	stabilizing	-1.09553	-1.58301	-2.14444	0
15 I-MET508 ILE,I-ARG505 CYS,I-GLU509 CYS	-3.04578	stabilizing	0.21699	-1.5598	-2.96922	0
16 I-MET508 ILE,I-ARG505 TRP,I-GLU509 CYS	-3.03249	stabilizing	-1.5709	-1.3365	-1.97349	0
17 I-LEU507 CYS,I-MET508 LEU,I-GLU509 THR	-3.02901	stabilizing	-1.9598	-1.41662	-1.67601	0
18 I-MET508 ILE,I-ARG505 ILE,I-GLU509 TRP	-2.99458	stabilizing	-3.6767	-1.06262	-0.781156	0
19 I-MET508 ILE,I-ILE506 CYS,I-GLU509 TRP	-2.99167	stabilizing	-1.71541	-1.37582	-1.80757	0
20 I-MET508 ILE,I-ARG505 ILE,I-GLU509 VAL	-2.98585	stabilizing	-2.7917	-1.18658	-1.24565	0
21 I-MET508 ILE,I-ARG505 CYS,I-GLU509 THR	-2.98324	stabilizing	-0.97065	-1.98059	-1.88451	0
22 I-MET508 ILE,I-ARG505 TRP,I-GLU509 LEU	-2.97913	stabilizing	-3.28913	-1.12736	-0.963618	0
23 I-LEU507 PHE,I-ARG505 TRP,I-GLU509 GLN	-2.97477	stabilizing	-5.2506	-1.06551	0.2291	0
24 I-MET508 ILE,I-ILE506 CYS,I-GLU509 CYS	-2.93024	stabilizing	1.01999	-1.66883	-3.25761	0
25 I-LEU507 CYS,I-MET508 ARG,I-GLU509 ASN	-2.92928	stabilizing	-3.0228	-0.826366	-1.25588	0
26 I-MET508 ILE,I-ILE506 CYS,I-GLU509 LEU	-2.91637	stabilizing	-0.93343	-1.54988	-2.09341	0

图 21

从结果中可以看到，Met508 突变为 Ile，Arg505 突变为 Cys，并且 Glu509 突变为 Pro，这样的突变组合，突变能为-3.52994kcal/mol，远远低于之前的单点氨基酸突变。因此，这样的突变组合可以使蛋白的热稳定性提高的更多。基于这样的信息，可以指导设计高热稳定的酶和抗体。

另也可点击 Results 栏中 Mutation Energy Terms (Double Mutations) 和 Mutation Energy Terms (Single Mutations) 链接，查看双点突变和单点突变信息。

## 蛋白质二硫键的预测教程

**目的：**通过此教程，了解 Discovery Studio 中预测蛋白质二硫键的操作过程。

**所需功能和模块：**Discovery Studio Client

**所需数据文件：**2CBA.csv

**所需时间：**1 小时

### 介绍

蛋白质通过选择性的氨基酸突变形成二硫键，可能增强蛋白质稳定性。然而稳定的突变位点不仅仅由空间距离决定。稳定的二硫键之间除了有比较好的空间立体距离，还包括其它的一些因素，比如立体位阻、氨基酸埋藏深度等因素。

Predict Disulfide Bridges 工具尝试建立二硫键在一对氨基酸之间，保持它们之间的最好的几何空间距离，同时下列因素可能影响建立二硫键的稳定性，比如立体位阻、可溶剂表面积、氨基酸埋藏深度、体积改变、序列间隔、温度因子变化等。

### 蛋白质二硫键的预测

1、打开蛋白质结构文件，在菜单栏点击 Open，然后选择 2CBA.csv 打开。

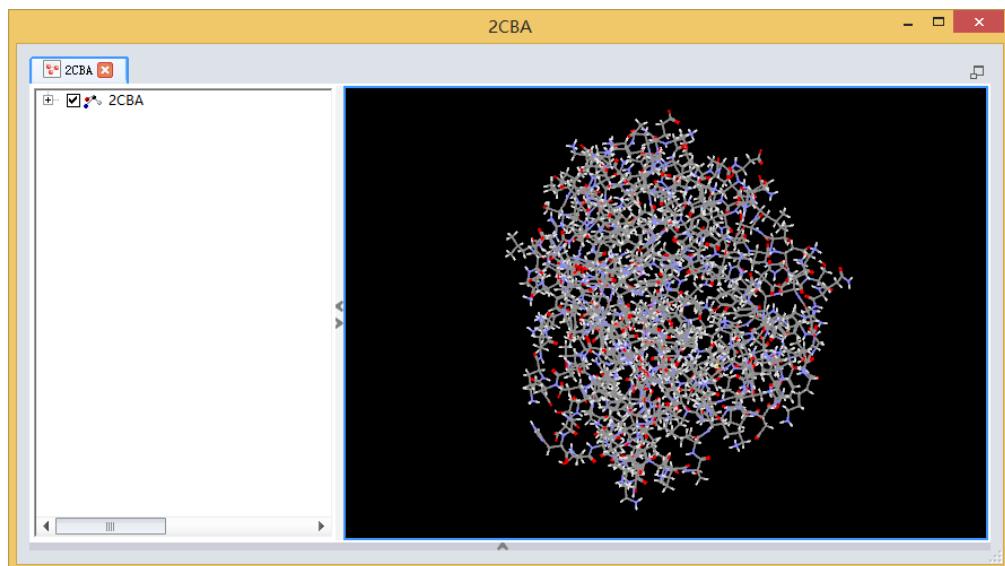


图 1

2、点击 Macromolecules| Design Protein| Predict Disulfide Bridges，进行二硫键的预测。然后点击 First 按钮查看第一个预测的二硫键的位置。

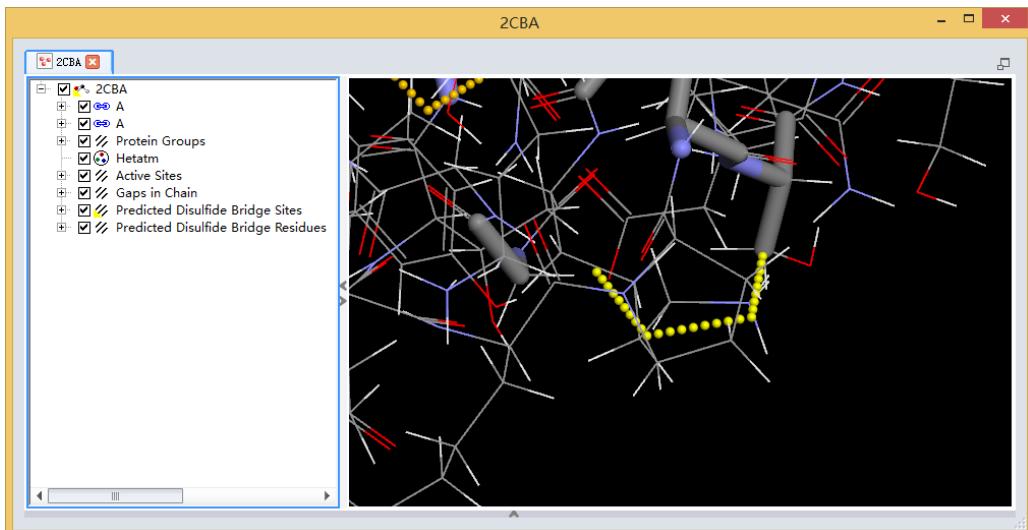


图 2

- 3、同时弹出一张二硫键预测报表，包含立体位阻、可溶剂表面积、氨基酸埋藏深度、体积改变、序列间隔、温度因子变化等信息。

Disulfide Prediction for 2CBA											
Disulfide Prediction for 2CBA											
Predicted disulfide bridges for molecule 2CBA											
Predicted Disulfide Bond	Original Residues	Score	Energy	Bad Contacts	Thermal Mobility	Sequence Separation	Residue Depth	Volume Change	Environment	Chi3	
CYS48:SG - CYS81:SG	A:SER48, A:GLY81	96.90	1.55	0	N/A	33	3.38	48.2	Sheet-Sheet	-83.5	
CYS38:SG - CYS258:SG	A:ALA38, A:ALA258	94.64	2.68	0	N/A	219	5.48	27.5	Coil-Sheet	136.1	
CYS153:SG - CYS219:SG	A:ALA153, A:SER219	93.78	3.11	0	N/A	66	3.77	22.5	Coil-Coil	-111.3	
CYS147:SG -	A:PHE147,	91.63	4.19	0	N/A	68	6.13	-109.1	Coil-Sheet	-	

图 3

- 4、将二硫键预测报表拉到最后，里面包含了每一项信息的最优区间、中等区间、最差区间，并分别以绿色、橙色、红色进行颜色标注。比如一个稳定的二硫键的位置应该符合以下条件：能量变化小于 5、过近接触为 0、温度因子变化大于 20、氨基酸间隔大于 25、氨基酸埋藏深度在 3 和 8 之间、体积变化在 25 至 50 之间、理想环境在 Coil-Coil 之间、Chi3 角度变化小于 20.

Color coding of factors which may influence stabilizing effect of a bridge			
Property	Green (Stabilizing)	Orange	Red (Destabilizing)
Energy	<= 5.0	<= 10.0	> 10.0
Bad Contacts	0	1	> 1
Thermal Mobility	>= 20.0	>= 5.0, < 20.0	< 5.0
Sequence Separation	>= 25 or Inter Chain	>= 5	< 5
Residue Depth	>= 3 and < 8	>= 2 or >= 8	< 2
Volume Change	>= -50.0 and < 25.0	>= -125 or >= 25.0	< -125.0
Environment	Coil-Coil	Other combinations	Sheet-Helix
Chi3 (minima -87, +97)	<= 20.0 difference from minimum	<= 30.0 difference from minimum	> 30 from minimum

图 4

5、构建二硫键的突变体，点击 Macromolecules| Design Protein| Create Disulfide Bridges

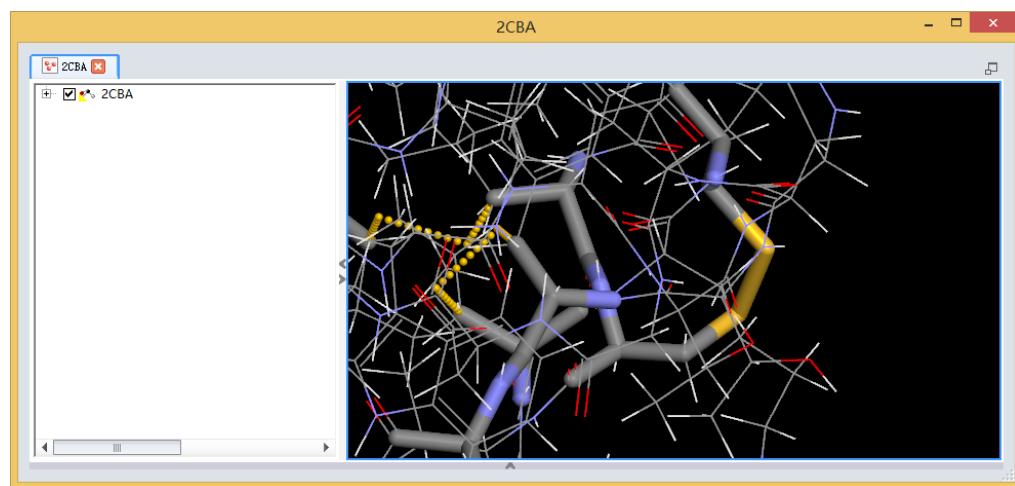


图 5

6、清除其它位置的预测信息，点击 Macromolecules| Design Protein| Clear Prediction，此时图形窗口中只含有打分最高的含有二硫键的蛋白质的结构。

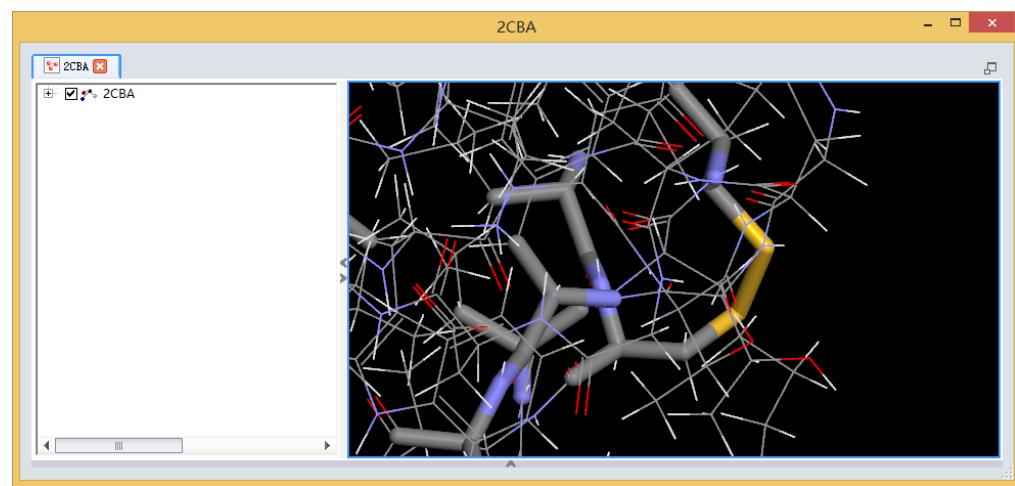


图 6

## 预测蛋白聚集位点（Protein Aggregation）教程

**目的：**通过此教程，了解 Discovery Studio 中预测蛋白聚集位点的操作方法及结果分析。

**所需功能和模块：**Discovery Studio Client, DS Protein Aggregation

**所需数据文件：**1hzh.pdb。

**所需时间：**20 分钟

### 介绍

抗体等具有治疗功能的蛋白，如果处于比较高的浓度下，会有发生聚集的趋势。这会导致抗体的活性下降，并引起免疫反应。抗体的聚集趋势计算是一种衡量蛋白表面氨基酸聚集倾向性的指标。具有比较高的聚集趋势得分的位点表明了该区域的氨基酸倾向于发生聚集。因此这些位点的预测，使得我们可以通过氨基酸定点突变的方法来改造蛋白，增强其稳定性。

本教程使用 Calculate Aggregation Scores 对一个全长 IgG1 抗体分子(PDB 号为 1h2h)进行蛋白聚集位点的预测计算，并分析预测的结果。

本教程涵盖如下内容：

- 聚集趋势得分的计算
- 分析蛋白聚集位点

### 聚集趋势得分计算

在文件浏览器（Files Explorer）中，展开 Samples | Tutorials | Protein Modeling 文件夹，双击 1hzh.pdb 文件。

DS 将在一个新的 3D 窗口中打开该蛋白。

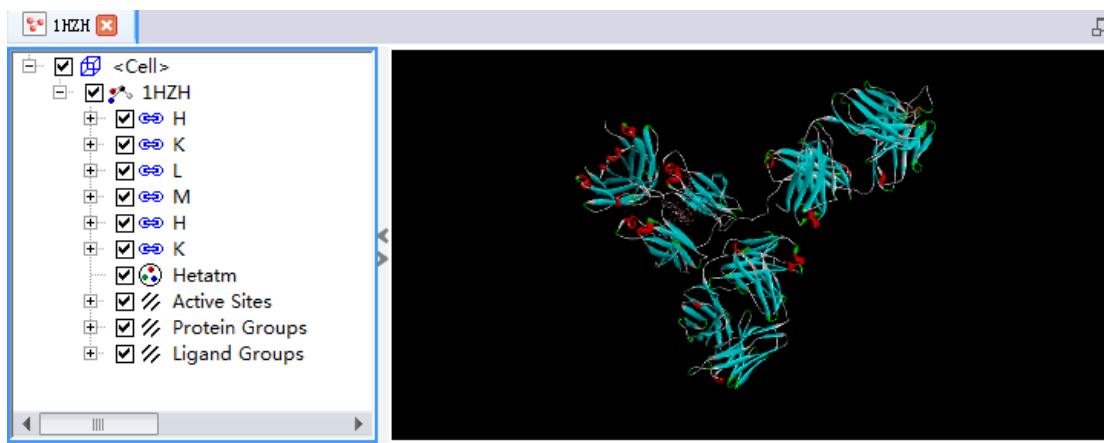


图 1 1hzh 分子窗口

**Ctrl+H** 打开 Hierarchy 窗口，然后选中 Water，点击 Delete 删除蛋白结构中的结晶水分子。  
在工具浏览器（Tools Explorer）中，展开 Simulation | Change Forcefield，将 Forcefield 设为 CHARMM Polar H，然后点击 Apply Forcefield，这将为蛋白赋上 CHARMM Polar H 力场。

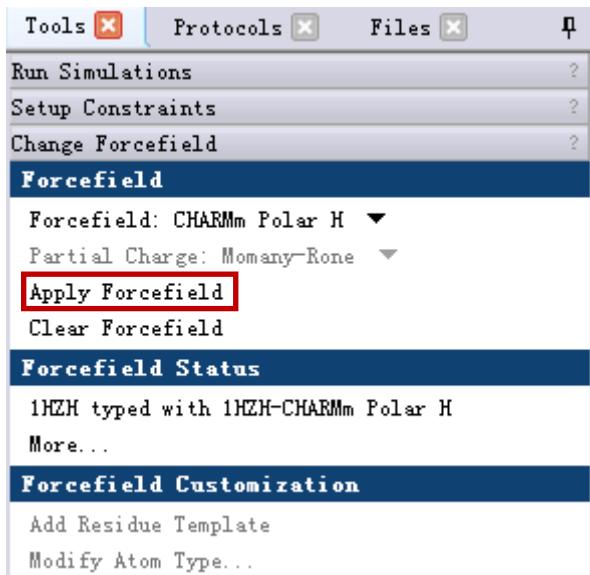


图 2 Apply Forcefield 设置界面

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Predict Protein Aggregation 工具面板，点击 Calculate Aggregation Scores。

在弹出的参数设置界面中，将 Input Typed Protein 设为 1hz:1HZH，将 Cutoff Radius 设为 5,7,10。

点击 Run 运行该任务。

该任务在奔腾 4, 2Gb 内存和 2.8GHz 的计算机平台上运行大约需要 3 分钟。

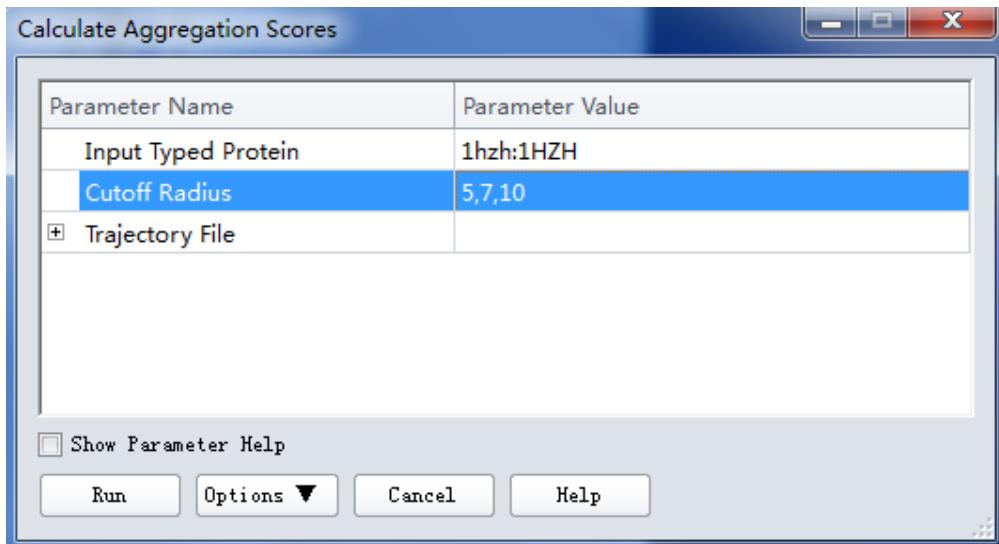


图 4 “Calculate Aggregation Scores”参数设置

计算完成之后，窗口中的蛋白结构将会自动更新，显示为带有计算完成的蛋白聚集趋势得分的结构。不同半径的蛋白聚集趋势得分将会添加到每个原子和每个氨基酸的属性中，分别以 **Aggr R5/7/10** 表示。聚集位点被定义为一组具有高蛋白聚集趋势得分，且暴露于蛋白表面的毗邻氨基酸的组合，并会在 Hierarchy 窗口中的 **Aggregation Site** 下创建相应的组。

计算完成后，程序将会自动创建一个溶剂表面，并依据最大的 Cutoff 半径下原子的蛋白聚集趋势得分来进行着色。具有高蛋白聚集趋势得分的原子将显示为红色，而具有低蛋白聚集

趋势得分的原子将显示为蓝色。此外，还会自动创建一张线图和一张点图。线图以氨基酸的序号为横坐标，以蛋白聚集趋势得分为纵坐标。点图以蛋白聚集位点的序号为横坐标，以蛋白聚集趋势得分为纵坐标。

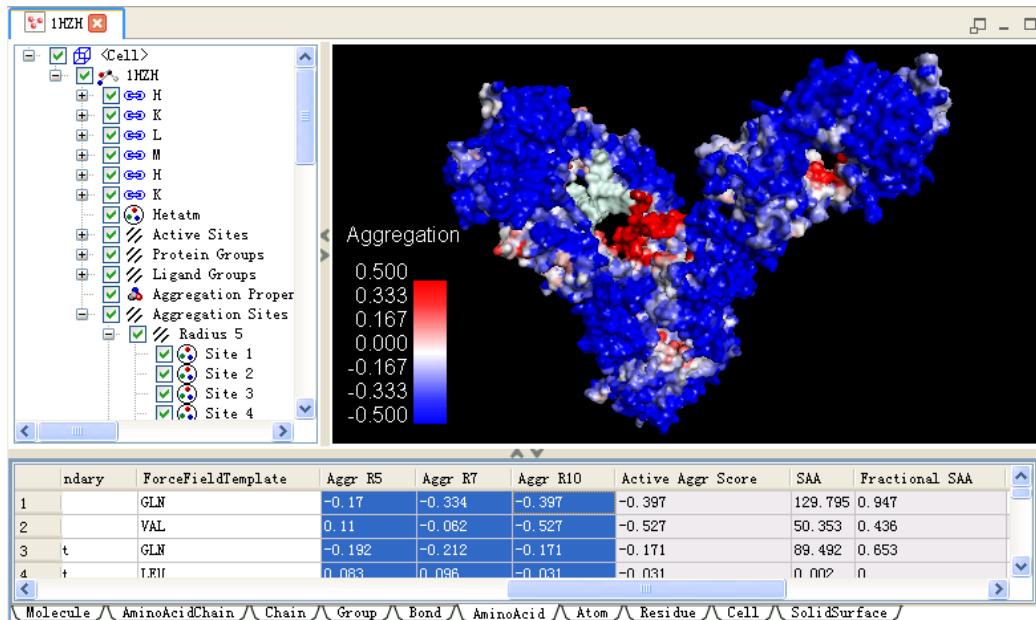


图 5 依据蛋白聚集趋势得分着色的溶剂表面

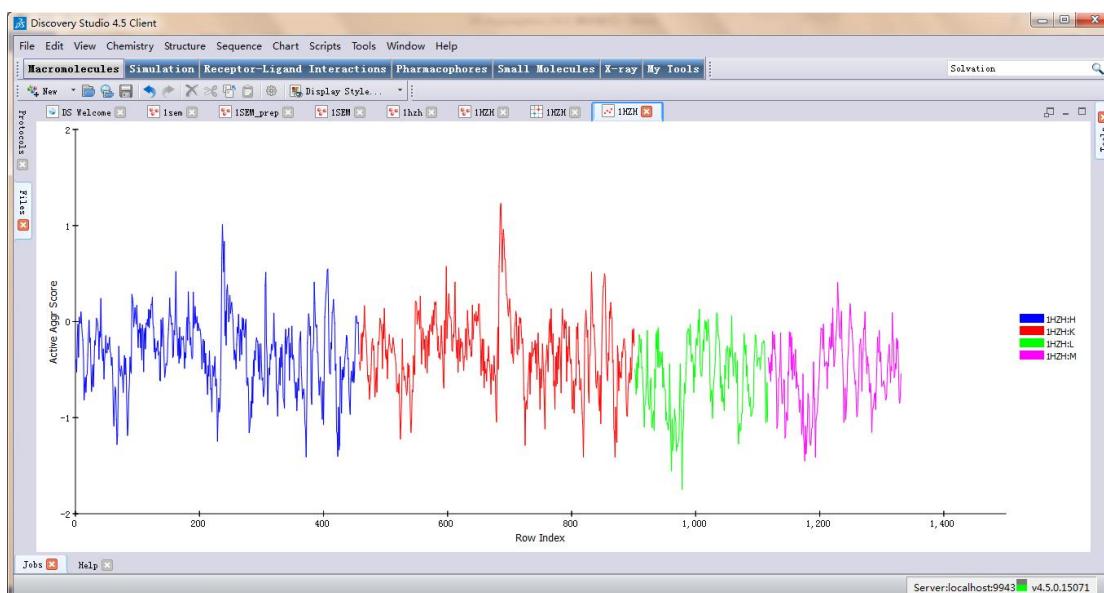


图 6 氨基酸序号 VS 蛋白聚集趋势得分的线图

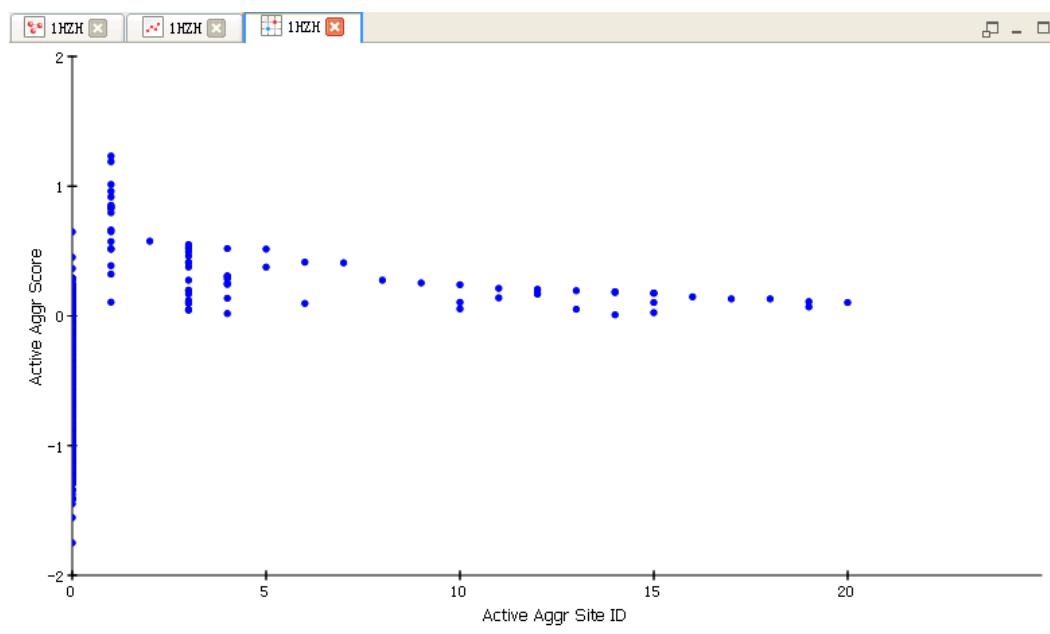


图 7 蛋白聚集位点序号 VS 蛋白聚集趋势得分点图

### 分析蛋白聚集位点

在 Graphic 窗口中展示了 Cutoff 为 10 的蛋白表面结构。旋转分子，并观察标为红色的蛋白表面区域，这些区域倾向于发生蛋白聚集。

在工具浏览器（Tools Explorer）中，展开 Macromolecules | Predict Protein Aggregation| View Aggregation Sites，将 View at Cutoff Radius 改为 5，观察蛋白表面颜色的变化。

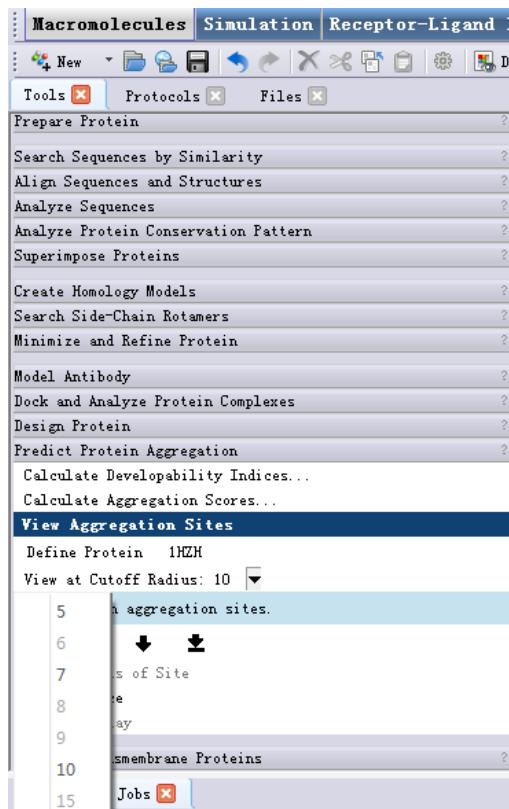
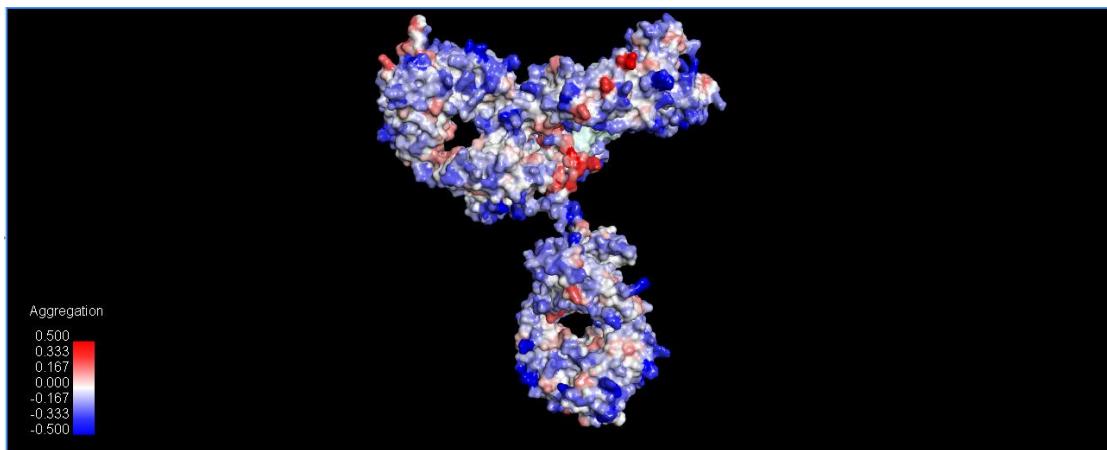


图 8 改变 Cutoff 半径



半径越小，单个聚集位点的面积越小，但是聚集位点的数目越多；半径越大，单个聚集位点的面积越大，而聚集位点的个数减少。

**Ctrl+H** 打开 Hierarchy 窗口，点开 Aggregation Site 以观察每个半径下的 Site group。

我们可以看到半径为 5 时，位点的数目要比半径为 10 时更多，而每个位点包含的氨基酸则更少。也可以通过观察之前得到的点图得到相同的结论。

在工具面板中，重新将 **View at Cutoff Radius** 半径设为 10 Å。点击 **First Site** 按钮 ，然后

点击 **Show Details of Site 1**。使用 **Fit To Screen** 按钮 ，然后通过旋转观察蛋白聚集位点。

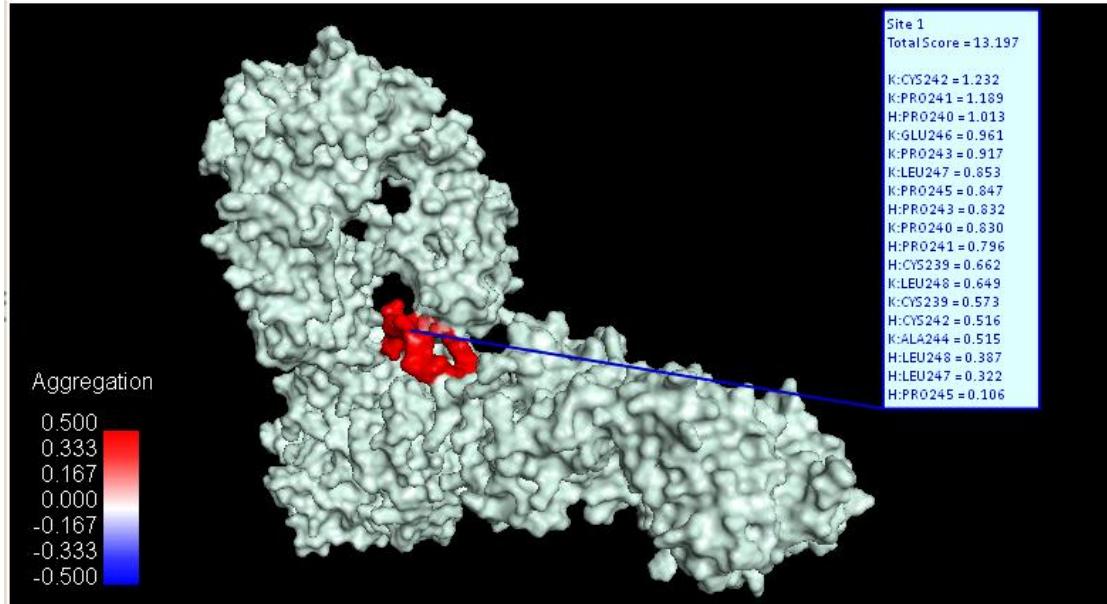


图 9 Show Details of Site 1 展示蛋白聚集位点

在图中 Site 1 被显示为红色，而蛋白的其他部分则显示为灰绿色。一个注释窗口被创建出来，指向 Site 1 的中心，并注明了组成 Site 1 的所有氨基酸以及每个氨基酸相应的蛋白聚集趋势得分。

在工具面板中，点击 **Hide Surface** 来隐藏蛋白表面的展示，并使得表面之下的氨基酸得以展示出来。

为了更好的展示氨基酸的细节，点击窗口上方 **Display Style** 右侧的箭头，并选择右下角的图示。

此时，蛋白中的氨基酸以线型进行展示，聚集位点的氨基酸呈现红色。

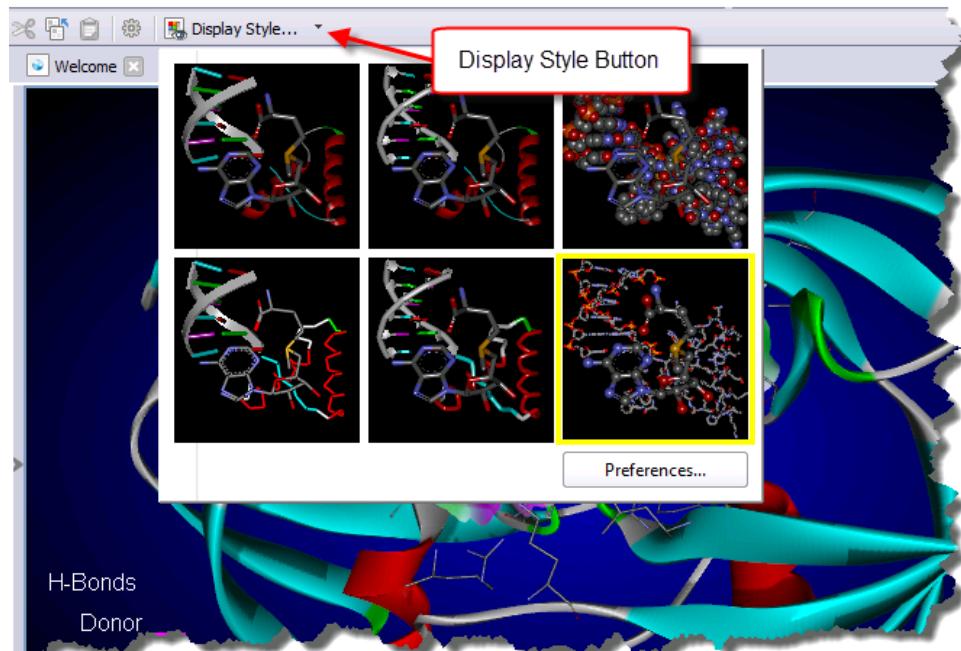
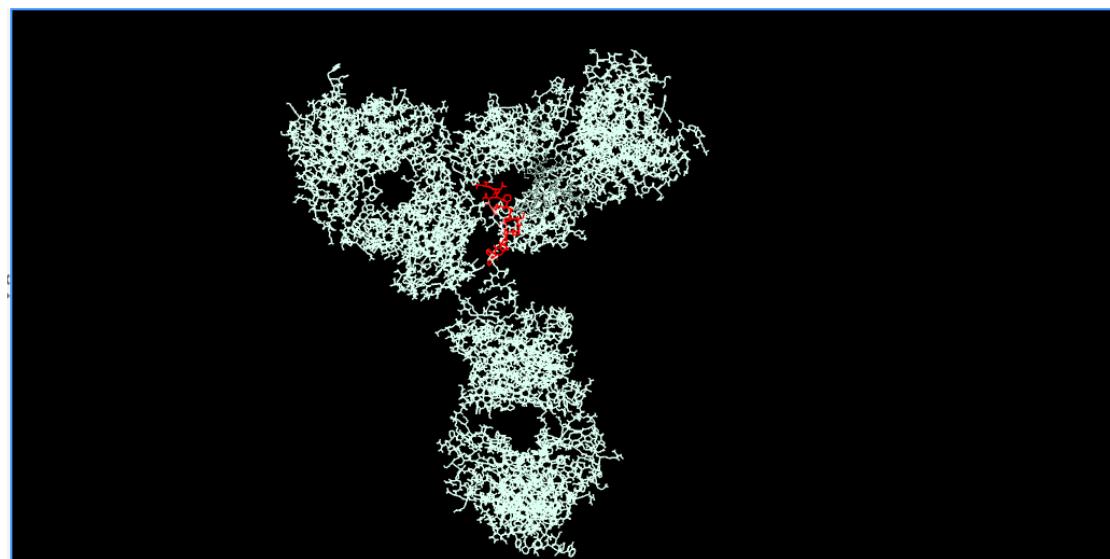


图 10 改变蛋白显示方法

根据同样的操作方法，通过点击 **Next** 可以查看 Site 2 所包含的蛋白聚集位点信息。通过 **Reset Display** 可以恢复到所有的蛋白聚集位点以红色展示，而其他位点以灰绿色展示的蛋白表面。



通过计算和分析蛋白聚集趋势得分和蛋白聚集位点，我们可以找到蛋白表面上聚集趋势得分比较高的氨基酸，并在实验中通过定点突变进行蛋白的改造。