

Cn3D 4.1 中文使用手册

Yang Xin (yx3240106@163.com) 翻译

生物软件网 (<http://www.bio-soft.net>)

整理提供

Welcome to the Cn3D 4.1 Tutorial!

目录

- [简介](#)
 - [Cn3D 的基本功能](#)
 - [下载和安装 Cn3D](#)
 - [文档约定](#)
 - [参考文献](#)
- [寻找特定的结构\(MMDB\)](#)
 - [根据 Entrez 文献研究](#)
 - [根据 Entrez 临近序列](#)
 - [通过 BLAST 搜索](#)
 - [通过已知的 PDB 标识](#)
- [在 Cn3D 中察看特定的序列结构](#)
 - [Cn3D 的基本控制操作](#)
 - [结构窗体的主菜单](#)
 - [样式面板](#)
 - [Cn3D 的序列阅读器](#)
- [结构比较\(VAST\)](#)
 - [在 Cn3D 中察看结构比较结果](#)
 - [Cn3D 的比对察看器](#)
 - [Cn3D 的比对模块](#)

- [导入和保存](#)
 - [导入序列和空间结构](#)
 - [可见的序列保守性](#)
 - [注释结构](#)
 - [骨架和界标](#)
 - [自定义样式](#)
 - [保存空间结构和图像](#)
 - [将 Cn3D 用于交互式图形](#)
 - [高级设置](#)
 - [动画控制](#)
 - [显示无序晶体](#)
 - [类型搜寻](#)
 - [导入用户 PDB 文件](#)
 - [比对编辑](#)
 - [使用编辑器](#)
 - [校正队列错误](#)
 - [将序列合并到多重序列中](#)
-

简介

这是 Cn3D 4.1 的使用手册。希望能够向初次使用或是曾经使用过 Cn3D 的用户提供一个关于本软件的基本特点的指导。新用户可能希望通过阅读这篇文档来学习如何使用 Cn3D,而有经验的用户则可以通过上面的目录和超连接直接跳转到自己感兴趣的章节。

本手册并不是对程序功能的详尽的介绍。在 Cn3D 的安装程序里包含有关于 Cn3D 的用户界面和详细功能介绍的帮助文档。— 见 [Cn3D_Commands.chm](#)。

Cn3D 的基本功能

Cn3D 是一个生物分子的三维结构、序列以及序列比对结果的可视化工具。Cn3D 可以将结构与序列的信息紧密的联系起来，这是它与其它软件的一个重要的区别：例如，一名科学家可以很快的从晶体结构中找出与导致已知疾病的突变相关的残基，或是保留同源序列家族的活性位点的残基。Cn3D 可以通过基于结构的序列比较来显示生物分子结构之间的比较，从而了解相关蛋白的那一个结构域在结构与序列上表现得更为保守。同时，可以自定义标签的特性，高品质的 [OpenGL](#) 的画质，还有多样的文件输出格式，都使得 Cn3D 成为文献注释的强大工具。Cn3D 的特色就是通过网络浏览器来作为 NCBI 的 [Entrez](#) 系统的一个辅助工具，但是它也可以作为一个独立的程序来使用。

在版本 4 当中，Cn3D 已经是一个完整的多序列比较编辑器了，除此之外，还包括一条已知序列和其他序列或是其他结构进行比较的算法。你可以新建一个比对结果或是评价一个已有的结果。Cn3D 可以被用来作为比较 [CDD project](#) 内容的基本的辅助工具。（保守结构域数据库）

下载和安装 Cn3D

Cn3D 可以应用于 [Windows](#)，[Macintosh](#)，和各种 [UNIX](#) 平台。这几页将说明如何下载和安装 Cn3D，并且如何 [配置网络浏览器](#) 来使用 Cn3D。

文档约定

Cn3D 的屏幕界面和序列窗体提供各种形式的示例；他们以极小的图片链接到大图。注意最大的图像是以 [PNG 格式](#) 存储的—这依靠所使用的浏览器，浏览这种格式的文件需要一个支持 PNG 的辅助程序。Cn3D 的 Windows 版可以用来创建这类图像，但是除了平台的用户界面和窗体变框外，图像基本上在任何平台上都是一样的。

当引用到用户界面当中的特定元素时（例如下拉菜单或弹出式菜单，按钮等等）本文将使用黑体字并且用冒号将需要进行的鼠标操作分开。例如，**Style:Edit Global Style:Details:Tube Radius** 意思是选择"Style" 下拉菜单（在结构窗体

中), 选择"Edit Global Style"项。一个分页的面板将出现; 选择"Details" 页, 然后设置"Tube Radius" 属性。

参考文献

为了增加手册的生动性和信息性, 我们使用当前文献中使用的有名的例子。大多数的例子是关注人的 PTEN 蛋白 ([Lee et al., 1999](#)), 一个在多种癌症中发现突变的肿瘤抑制因子(综述, 见 [Di Cristofano and Pandolfi, 2000](#))。

下面是手册中所引用的文献的详细列表(含有 Entrez 的连接)。

1. Di Cristofano A, Pandolfi PP. The multiple roles of PTEN in tumor suppression. Cell 2000 Feb 18; 100(4):387-90. ([Entrez](#))
2. Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, Dixon JE, Pandolfi P, Pavletich NP. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. Cell 1999 Oct 29; 99(3):323-34. ([Entrez](#))
3. Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, Bose S, Call KM, Tsou HC, Peacocke M, Eng C, Parsons R. Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat Genet. 1997 May; 16(1):64-7. ([Entrez](#))

Obtaining Structure Data for Cn3D

寻找特定的结构 (MMDB)

新用户的最大的问题之一是当他们开始使用 Cn3D 时怎样并且从哪里可以得到程序可以识别的格式的数据呢? Cn3D 有意的不去直接读取 PDB 格式的文件, 取而代之使用 NCBI 的 [MMDB 数据库](#)。简单的说, MMDB 从 [蛋白质数据库](#) 中获取数据信息, 然后分析每一个 PDB 文件, 进行深入的验证和纠错, 最后以一种更加

适合计算机的格式存储信息。参见 [MMDB 的主页](#) 可以了解到更多的关于这项工程的信息，并且可以从 MMDB 的结构摘要一页的选项和链接中得到更多的文件。

有很多种方法可以从 MMDB 中获取所要结构，包括直接的或是通过 NCBI 的 [Entrez](#) 服务中的链接。本章中描述了分子生物学家可能用到的通过不同的查询方式查找蛋白质结构一些典型方法。我们将以人的 PTEN 蛋白为例说明 ([Lee et al., 1999](#))。

Entrez 文献研究

为了使文献研究可以直接获得结构摘要，空间结构数据已经整合到 Entrez 当中。最简单的使用 Entrez 的方法是用关键词、作者等来查找空间结构注释。例如，到 [Entrez](#) 当中，在 "Search" 栏内选择 "Structure" 下拉菜单，然后在输入框中键入 "PTEN" 并且点击 "Go" 来执行查找。在这个例子当中，结果很少，1D5R 获得的唯一的查询结果，结果中含有到 MMDB 摘要页的链接。

空间结构相关链接也会出现在文献搜索结果中。在 "Search" 栏内选择 "PubMed" 下拉菜单，然后使用 "PTEN structure" 进行查询。在结果列表的后部有晶体结构文章：Lee et al., "Crystal structure of the PTEN tumor suppressor" ([Lee et al., 1999](#))。注意在左边有 "Structure" 相关链接—这个链接将会再次把我们带到 [1D5R 的 MMDB 摘要页](#)。

在这篇摘要中，选择 "Launch Viewer" 选项，然后点击 "View/Save Structure" 按钮来下载数据并且启动 Cn3D 显示结构信息—当然假设 Cn3D 已经作为一种辅助程序被恰当的 [安装和配置](#) 了。你也可以使用 "Save File" 选项来将下载的数据保存到磁盘中，这样你就可以手动的通过 **File:Open** 对话框将数据装载到 Cn3D 当中。之后将出现两个窗体：主窗体是 Cn3D 结构窗体，其中显示蛋白质的三维结构；另一个是序列窗体，其中显示蛋白质链的氨基酸序列。结果类似下面的样子：



如果是多链蛋白，将会有几条序列同时出现在序列窗体中。

根据 Entrez 临近序列

假设我们并不知道晶体的结构，并且与 PTEN 突变相关的疾病：例如，通过这个链接可以找到一篇关于 Cowden 氏病的文章：[Liaw et al., 1997.](#)。点击位于摘要右边的"Protein" 相关连结，然后跟随"O00633" 链接到达 [GenPept summary of PTEN](#)。在这一页中包含有大量的 PTEN 的文献和已知的突变信息。

这个例子中没有直接的三维结构的链接，因为用于测定晶体结构的蛋白与自然状态下蛋白不十分一样，这在 GenPept 的报告中有相应的描述。然而，我们可以在 GenBank 中找到与这个蛋白相似的序列列表，这些序列是已经知道空间结构的蛋白。从 GenPept 的摘要中，点击右上方的"BLink"。然后再点击结果页顶端的"3D Structures" 按钮，之后你将看到唯一的相关结构 1D5R, chain A。如果你点击结果线上的小蓝点，你将在 Cn3D 中看到比对的结果。

通过 BLAST 搜索

我们根据 PTEN 的序列通过 BLAST 搜索 PDB 数据库来直接得到空间结构。它的优点是可以使用任何序列进行查询，即使是新的或是个人所有的还没有提交到 GenBank 的序列。重要的一点是，我们可以评价比对结果，通过评分来判断查询序列和空间结构的序列相似性。

有许多的方法可以进行这样的查询，但是对于这个例子让我们从 NCBI 的 BLAST 服务开始。首先注意在 GenPept 摘要上部的 PTEN 的 accession 号是 "O00633" (第一个字母是大写字母 O，后面两位是零)。到 [BLAST 查询页面](#)，点击链接" Protein-protein BLAST (blastp) "（原文中是"Standard protein-protein BLAST [blastp]"）。然后在"Choose database " 项下选择 "pdb" 菜单（查找已知的空间结构），在输入框重键入"O00633" ，最后点击"BLAST!" 开始查询。查询序列将被送往 BLAST 队列，等待一段计算机下载的时间后，点击"Format!"按钮将出现结构。就在命中的图形摘要下边是找到的序列的列表。在顶部具有最高的 score 和 E-value 的序列，至少再写这篇文档时，就是现在熟悉的 PDB 入口 1D5R 。通过"pdb|1D5R|A"链接可到达这个结构的 [GenPept 摘要页](#)，在那里右边的结构链接最终带领我们到达 1D5R 的 [MMDB 摘要页](#) 。

这是一个价值不高的例子，因为找到的 PDB 结构就是用于 BLAST 查询的蛋白序列。但是这是一个非常强大的查找蛋白质空间结构的一般的方法，查找到的已有结构的蛋白质与目标蛋白质非常的相似，这样目标蛋白质的结构可以通过同

源性推断出来。参见这篇文档的[结构比较章节](#)我们可以学到如何在 Cn3D 的序列窗口中显示查询序列和其他蛋白序列的比对结果。

通过已知的 PDB 标识

如果一个蛋白质的 PDB 标识已知，那么就可以从[MMDB](#)中直接找到对应的空间结构。只需简单的在输入框内键入 4 个字符的 PDB 号，然后点击"Go"。

在关于结晶化和结构测定的期刊文章中如 PTEN ([Lee et al., 1999](#)), 我们可以找到作者提交结构数据到 PDB 中的标识号"1D5R"。把它输入 MMDB 的查询框就可以直接到达 1D5R 的[MMDB 摘要页](#)。

The Basics of Cn3D Controls

在 Cn3D 中察看特定的序列结构

这一章中将把[1D5R 的 PDB 结构](#) 作为例子；参见 [上一章](#) 介绍如何找到它的数据库。

Cn3D 的基本控制操作

- [结构窗体的主菜单](#)
- [样式面板](#)
- [自定义设置的一些例子](#)

因为并不是要对 Cn3D 用户界面进行全面的介绍，所以本章只提供对程序操作的总的说明，并且对不同科学方面数据进行可视化的图解说明。

• *结构窗体的主菜单*

File 菜单可以将数据输入 Cn3D，并且可以输出多种数据个适合图像类型；这些完全是无需解释的。仔细的说，**File:Save** 可以连同当前的视图，设置和其他用户注释一起保存分子的空间结构，以使所做的工作可以被保存。**File:Export PNG** 菜单项可将分子的空间结构图像内容保存为[PNG 格式](#)的图像，这种格式可以用于网页或是出版物(就象这篇手册！)。

View 菜单包含可以控制整个分子空间结构显示的菜单项。整个结构可以通过菜单项 **View:Zoom In** 和 **View:Zoom Out** 来放大或是缩小，并且可以通过菜单 **View:Restore** 使视图恢复到已保存数据的原始大小。菜单 **View:Reset** 可以使整个空间结构调整到适合窗体的大小。如果有多条序列的空间结构需要显示(例如一个 VAST 比对)，或者结构中包含多个亚基(例如一个 NMR 结构,常常有多个亚基一同被下载)，那么，每一个亚基将被分配一个单独的“框架” (“frame”)。菜单 **View:Frame** 下的各个条目可以控制哪一个框架在当前显示。**View:Animation** 将在[后续的章节](#)中讨论。

Show/Hide 菜单包含的操作可以实现特定的亚结构显示或隐藏。子菜单 **Show/Hide:Pick Structures** 对话框可以让你选择哪一个单一的结构，链和结构域显示的开关。**Show/Hide** 菜单下的其他项则与比对的视图有关。

Style 菜单下的选项可以影响空间结构中不同部分的形状和颜色。例如，默认的显示单个结构的方式是 **Style:Rendering Shortcuts:Worms** 加上 **Style:Coloring Shortcuts:Secondary Structure**，这种方式显示一个螺旋的骨架，没有侧链，用实心的记号—箭头和柱面—代表条带和螺旋。螺旋的颜色为绿色，条带为桔黄色，卷曲为蓝色。(这些二级结构元素的位置和范围是由 NCBI 确定的，这在 PDB 的原始文件中并没有纪录。)注意条带上的箭头 (螺旋柱面的方向是随机的)总是指向 N 端到 C 端的方向。了解其他的不同样式的最好办法就是试试看！

Cn3D 有意使颜色设置与绘制风格分开。菜单 **Style:Coloring Shortcuts** 下的选项的属性可以用来给结构的不同部分标上不同的颜色。使用这些选项，我们可以很容易的看到显示出来的 NCBI 确定的分子空间结构的组成 (结构域)，或者是来自 PDB 数据的晶体模型的温度等等。菜单 **Style:Rendering Shortcuts** 下的选项可以确定分子空间结构各个部分的形状，例如螺旋或是键球模型。

- *样式面板*

样式面板(**Style>Edit Global Style**) 包括对所有的绘制风格，颜色，标签，和其他 Cn3D 中可以自由选择的属性的控制。事实上，在菜单 **Style:Rendering Shortcuts** 和 **Style:Coloring Shortcuts** 中的所有选项都是样式面板各个不同选项之间组合的方便的快捷方式。

样式面板的 **Settings** 页下有 4 列可以设置各种结构元素的属性，在其左侧有相应的标签；这些设置应用与全局的整个空间结构。"On/Off" 控制元素是否被显示(也可以选择使用那种骨架模型)，"Rendering" 项控制用于着色的元素的几何形状，"Color Scheme"项控制元素的颜色。大多数的选项可以单独设置，可是螺旋风格除外，它只能用于实骨架样式。当"User Selection" 色彩设计被选中时，最右边的 "User Color" 列可以为结构元素设置颜色。整个背景色也是通过这种方式设置的。

样式面板的 **Label** 页可以控制骨架的标签，链的界标和离子。（事实上，所使用的字体是在结构窗体中的 **File:Set Fonts** 菜单下设置的。）

样式面板的 **Details** 页允许详细控制头视图的几何属性。例如，当选则以螺旋风格显示时，通过 **Details:Worm tube radius** 的设置就可以控制螺旋骨架的厚度。

Style:Annotate 面板将在[注释一章](#)单独讨论，但是简单的说，允许用户为选定的特定残基设置不同的风格，以使它们与空间结构中的其他区别开来。

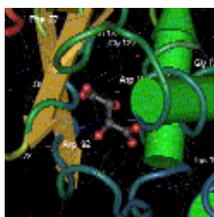
最后， **File:Preferences:Quality** 页可以控制 OpenGL 透视图的品质。事实上所有表面看起来圆滑的 OpenGL 对象都是有許多微小的多边形平面组成的。通过这个面板可以设置组成较大对象的小多边形面积的大小。弹出式菜单可以提供细节的设置，而右边的三个按钮则可以方便得将设置调整到预先设置的参数。图像外观平滑和显示速度之间是矛盾的：小多边形越多越小，对象就显得越平滑，但是这使得绘制时间变长。因此，在较老的计算机上为使 Cn3D 足够快的响应来达到实用的交互，较低的画质是必须的。相反，在很快的计算机上，或者要以 PNG 格式保存高质量的图片等对速度的要求不高时，高画质要更为可取。你可以选择正相图或透视图来保存。

- *自定义设置的一些例子*

图像的保存可以通过菜单 **File:Export PNG** 来完成。下面是 PDB 入口号为 1D5R 的结构，这里是一个活性位点内抑制剂的特写，用原色标记了球棍侧链，用温和色标记螺旋骨架，二级结构用不同颜色分开显示的三维对象，而且每隔 5 个残基标出了 PDB 序号。这些都可以在 **Style>Edit Global Style** 面板中进行设置。

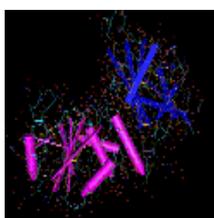
在图像右边是通过 **File:Save** 保存的相同空间结构的数据文件的链接。单击这个链接将会启动 Cn3D 并显示该结构，其初始图像与左边的静态图像文件一致。

举例来说，用户可以放大图像 (**View:Zoom Out** 或 **View:Reset**) 来察看为什么蛋白质表面比紧密填充的核心区域有较高的温度（晶体中有更多的运动）。这说明了 Cn3D 怎样创建和察看数字出版物中的交互式的图像；更多的主题请参见[注释一章](#)。



... [在此处单击就可以启动 Cn3D 并显示这张图像](#)

下面是相同蛋白的另一个视图，螺旋骨架根据二级结构用色彩区分，三维对象根据结构域用不同的色彩区分。还有（部分透明的）溶媒。



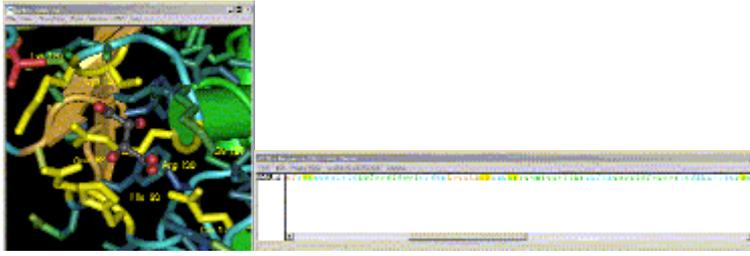
... [在此处单击就可以启动 Cn3D 并显示这张图像](#)

Cn3D 的序列阅读器

当一个单独的结构在 Cn3D 中被加载，序列阅读器就会显示结构中所有蛋白质和核酸链的序列。每个残基的颜色在结构窗体与序列窗体中都是相同的：序列中的每个字母代表结构中相应的残基，并且总是采用骨架中 **alpha** 碳的颜色（或对于核苷来说是磷的颜色），即使是结构窗体中侧链的颜色与骨架颜色不同。

更强大的一个应用是很容易将序列中的残基于结构中的原子一一对应。这就象是文本编辑器一样依靠高亮显示来实现。在序列窗体中按住并且拖动鼠标将会使字母反白显示，同时会使结构窗体中相应的原子以同样的反白颜色显示。反过来也是一样：双基结构窗体中的残基将会使它本身和序列窗体中相应的字母一起高亮显示。

因此，用户可以快速的从序列窗体中定位和高亮显示对蛋白质感兴趣的部分，比如活性位点相关的一系列残基。例如，下图中 **1D5R** 高亮显示了一些关键的催化和结合残基(参见[Lee et al., 1999](#)):



注意这有两个数字——序列位置和 PDB 分配号——当鼠标在序列上移动时在对应的残基上就会显示。这是因为结晶的蛋白质是被部分的截短的，其他高度混乱的残接已经从精制的结构中去掉；这些丢失的区域没有在结构中显示（例如，用空位字符或其他的图形设备标示），但是在 PDB 中是被计算在内的。参见[比对一章](#)。尽管如此，还是有巧妙的办法来处理！

Structure Alignments in Cn3D

结构比较 (VAST)

Cn3D 在单独结构的处理上的确优秀，但是在显示多个蛋白的结构比较方面更是出色。NCBI 创建并且维护了一个比对数据库，叫做[VAST](#)，可以方便的在 MMDDB 中查询所有核心区结构相似的蛋白质对。VAST 为每一对相关蛋白做两件事：计算保守区的最佳三维重叠，和构建基于空间结构相关的序列比对。

Cn3D 是 VAST 比对的基本的格式化工具。它将结构的显示与序列的显示结合起来，允许 VAST 的使用者察看由它创建的结构比较和序列比较结果，并且配色方案可以指示和加强保守区域。

让我们来看看 PTEN 的空间结构和它的 VAST 邻居吧。到[VAST 的主页](#)，在查询框中键入 "1D5R" 并且单击 "Get."。这将转到我们熟悉的 MMDDB 摘要页。其图像说明此蛋白质由一条单链 (A) 组成，NCBI 将它分解为两个结构域：N-端结构域 1，和 C-端结构域 2。点击图像顶部的标记 "Chain A" 的条形区将会显示基于整条链的 VAST 临近序列，如果点击标记 "1" or "2" 的彩色区域将会显示这些特定结构域的临近序列。

如点击 "1" 结构域可显示 N-端结构域的[临近序列列表](#)。关于 PTEN 结构的文献 ([Lee et al., 1999](#)) 强调了 PTEN 和双特异性磷酸酯酶(PDB 数据库入

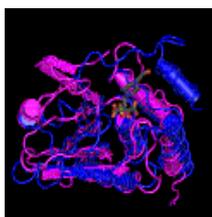
口"1VHR")之间的相似性。选择 1VHR 链 A 左边的选择框，然后点击"View 3D Structure" 在 Cn3D 中显示这个 VAST 比对结果。我们将在接下来的几节中使用这个例子。

在 Cn3D 中察看结构比较结果

在[前面的章节](#)中已经讲到,加载 1D5R 和 1VHR 的 VAST 比较结果。当 Cn3D 第一次启动时默认的视图是两个蛋白层次图并且使用骨架样式绘制的(alpha 碳由直棒链接), 核心区域紧密交迭的 alpha 碳突出显示。

Cn3D 4.1 结构比对中使用 VAST 算法, 在比对区域内默认的颜色是红和蓝, 相同的匹配残基为红色, 不同但可以匹配的残基为蓝色; 未匹配部分为灰色。注意 VAST 的工作方式, 对匹配区域和特定的二级结构域相联系—螺旋或条带, 而核心外面的环区长度和方向各异, 往往是非匹配序列。序列的显示中反映了结构的颜色, [下面](#)将详细讨论。

结构比对中初始的样式和颜色设置于使用 **Style:Rendering Shortcuts:Tubes** 和 **Style:Coloring Shortcuts:Sequence Conservation:Identity** 菜单选项等价。结构比对中的绘图设置与单一结构的调整方法是相同的。例如, 这是 Cn3D 中 PTEN 结构文献中 2A 的图像副本([Lee et al., 1999](#)), 他们有着相同的基本绘制样式合视角。图像清晰的展示了两个蛋白结构的紧密联系, 它们主要二级结构的方向和位置都很相似。但是活性位点周围的环区被抑制剂占据, 为了适应不同的底物, 他们的大小和位置都不同。



... [在此处单击就可以启动 Cn3D 并显示这张图像](#)

Cn3D 的比对察看器

当显示不止一个结构或已经比对的多条序列的结构时 Cn3D 的序列窗体同时可以作为比对察看器。

队列的显示非常的明了。每条序列占一行, 且主要的序列总是在最顶部。只有序列中用于比对的部分才显示出来, 尽管结构中可能存在多条链。并且, 序列

中的每个字母的颜色都是结构中相应的 **alpha** 碳(或磷)的颜色。高亮显示也一样在序列窗体和结构窗体中共享。

然而，**Cn3D** 中基于结构的比较和序列比较和通常的比较算法如 **BLAST** 或 **ClustalW** 在显示和根本的比对数据方面都有一些重要的不同。这将在下一节中介绍。

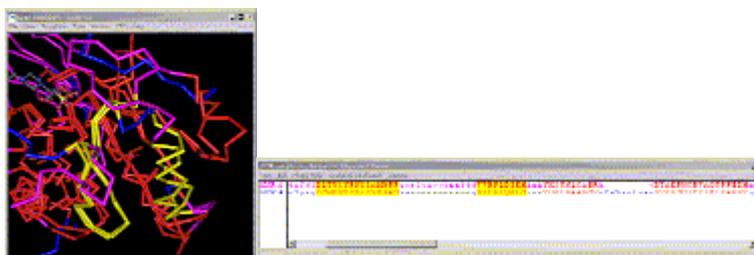
Cn3D 的比对模块

新的用户在 **Cn3D** 中察看一个 **VAST** 比对时的第一个问题可能是，"为什么一些字母要大写而另一些字母小写？并且这些有趣的 '~' 是什么意思？" 简而言之就是参与比对的残基以大写显示，未比对的用小写字母标示， '~' 代表未比对空位。但是，尤其对于后者，需要更多的解释。

动态规划比对算法比对残基是基于突变几率评分，与其它序列相比较使用空位来模拟生物进化过程中的插入和删除。一个插入的残基与一个空位匹配，并且相对应一个分值(空位罚分)。这些匹配的空位通常用 '~' 字母标出。也允许出现一个连续的空位覆盖两条序列的大的区域，即使当它们之间在进化上存在显著的不同。

在一个结构比对中(例如在 **VAST** 中)，一个残基和另一个残基匹配因为它们的 **alpha** 碳原子在空间上临近。因此，一个残基与空位匹配的概念是没有物理意义的，因为一个空间结构的 **alpha** 碳是不能和另一个结构中的空位空间相联系的。因此比对变得不连续了：匹配的区域—螺旋或条带在两个结构之间连续又紧密的交迭—由未匹配的区域分隔，这些构成了不能重叠的由各种不同长度不同方向的片断的交替。

如[前面](#)的讨论，以 **1D5R** 和 **1VHR** 之间的比较为例。



对于在屏幕中高亮显示的前亮块区域，第一块是与结构中的 **beta** 发夹相关，第二块与螺旋相关。就象可以从结构中看到的一样，他们在 **1D5R** 中被一段长的环分开，而在 **1VHR** 中是被一段短的环分开的。

这显示了未匹配的残基在匹配的区域之间，而保持匹配的残基正确的显示在另一条的上方需要不同长度的环区。这是靠字符'~'来填充较短的未匹配的环区-这只是显示出来的空位，并且这并不能说明也不能暗示任何比对实质。因此，1VHR的短环区由字符'~' 填补，这样 1D5R 的长的环区就有显示的空间了。

比对所默认的显示样式为 **Unaligned Justification:Split** 。这显示出一段短的非匹配序列将两段匹配序列联结起来的部分残基。这等价与使用 **Unaligned Justification:Center** 选项，它可以显示将分隔匹配序列的非匹配序列中间部分残基。这两种样式见下图所示。



... "split" 非匹配序列 (默认)

... vs. "centered" 非匹配序列

两种样式的不同是微小的，但是需要说明有一点是很重要的：显示非匹配序列对于用户是很方便的但是没有也不应该在非匹配的区域中干扰匹配序列的显示(例如使用小写字母)。这直接与空位动态规划法比对形成对照，动态规划法中，空位数、长度和位置都影响分值。上面的两种显示样式如果按照连续空位比对接是则分值是不同的。

另一个问题是当有多个些够需要显示时怎样创建多序列的比对，这其实是一系列的主/从序列对。这将在单独的部分讨论["intersect by master" 算法](#)。

Imports and Conservation in Cn3D

这个章节会举在 Cn3D 中一些比较高级的比对特点。

[导入序列和空间结构](#)，和[可见的序列保守性](#)。

导入序列和空间结构

Cn3D 是察看预先计算的序列比对的完美工具，象[上一章](#)所描写的从 VAST 中获取的序列。但是它也能导入用户选择的序列和结构，来演示当前主要结构的序列图，或比较新的结构。

我们从一个简单的例子开始。在[前面的章节](#)，我们举出当 1D5R 结构([Lee et al., 1999](#))第一次加载到 Cn3D，在序列窗体中残基的序数和在 PDB 文件中的不吻合，这是因为在精制的结构中残基的缺失。虽然不总是正确的，这种情况下 PDB 残基的序数是根据天然蛋白残基的真实序列序数。这可以通过 1D5R 与天然序列的比对得到很好的证实，这样可以显示出结构中缺失的残基。

因此，我们可以用[Entrez](#)去查找天然序列：在左边的“Search”菜单中选择“Protein”，在查询框输入"human PTEN"，然后点击"Go"。第一个条目就是 SwissProt 的 PTEN_HUMAN 入口。点击登记号"O00633"连接到 GenPept 报告。对于导入的文件，Cn3D 仅仅识别 FASTA 格式的文件，所以选择"Display"右边的"FASTA"，接着点击"Display"可以看到 FASTA 形式的序列，这个后面我们会用到。（**注意：**翻译这篇文章时，这样已经查不到 O00633 了，用户可以在查询框中输入"O00633"，然后点击"Go"。）

第一，在 Cn3D 中加载 1D5R 结构(可以参照[MMDB 说明](#))。在在序列窗口，选择 **Imports:Show Imports**，会出现（空的）输入窗口。

有两种方法可以将序列导入 Cn3D，从本地文件或网络下载。当使用本地文件，在你的电脑中剪切和粘贴 FASTA 格式的序列到一文本文件。然后选择 **Edit:Import Sequences:From FASTA File** 菜单项，在弹出的选择文件对话框中选择你刚才保存的文件。如果你没有防火墙，你可以让 Cn3D 直接从 NCBI 下载序列。这样需要选择 **Edit:Import Sequence:Network via GI/Accession** 菜单项，然后输入登记号"O00633"之后点击"Ok。"

注意，所有的残基是小写字母形式，这意味着没有进行比对。因此你需要通过一些算法去比较两条序列。Cn3D 有内建了一个基本的 BLAST 的界面，因此选择 **Algorithms:BLAST Single** 菜单，然后点击输入窗口中的任一条序列。这会利用 BLAST 默认的空位比对两条序列。

在输入窗口中滚动滑块就可以看到在 1D5R 中缺失的 N-端和大的 C-末端，这好象序列的 286 位到 308 位的残基被切除。同时需要注意 1D5R 的 PDB 的序列序数与天然序列的序数相匹配。

最后，为在序列窗口中进行多重序列比对，选择 **Alignments:Merge All** 菜单项。会将参与比较的序列对转到序列窗口，并且使用通常的比对用配色方案为空间结构标色。

导入结构也是同样的方法。举个例子，在[前面章节](#)中，我们举了重建 VAST 中 1D5R 和 1VHR 比对的例子。首先，转到 1VHR 的[MMDB 的摘要页](#)。对于这条序列，你可以自己下载 MMDB 文件（在 MMDB 页中选"save file"），或使用 Cn3D 下载。因此，再一次将 1D5R 导入 Cn3D，在输入窗口，选择 **Edit:Import Structure** 菜单。如你选"from file"，在谈出的对话框中选择你刚才保存的 MMDB 文件。如你选"via network"，输入 MMDB ID，号码是 4625。

无论你用那种导入方式，Cn3D 会问你想比较 1VHR 的那条链。如果，我们想比较 A 链，就选择它。Cn3D 会连接到 NCBI 的 VAST 数据库找到这两条序列的 VAST 比对。如果导入成功，你会看到和[前面章节](#)一致的比对结果。选择 **Alignments:Merge All** 将导入结果转移到比对窗口。

注意还没有可见的结构；Cn3D 仅仅显示当文件第一次打开时多序列比对中的结构。所以，选择 **File:Save** 菜单导出本地文件(无论什么问题，选择"Yes")，接着选择 **File:Load** 并且指向你刚才保存的文件。你应该可以看到这两个参与比对序列的结构，同时显示从 VAST 导入的比对结果。

如果没有 VAST 比对结果，或者你有防火墙或你由于一些原因不能连接到 NCBI，你只有自己手动比较结构和序列。选择 **Algorithms:BLAST Single** 然后点击序列中的一个启动一个 BLAST 比对。然后选择 **Alignments:Merge All** 将导入结果转移到连接窗口。保存并且重新打开文件，选择 **File:Realign Structures** 来计算基于比较残基的 alpha 碳原子的结构比对结果。

可见的序列保守性

Cn3D 可以基于比对结果中序列的保守性计算出各种各样的配色方案。序列中的保守残基因此可以通过为所显示的序列和结构应用保守的颜色而直接显示出来。

用红色指出高度保守的位置，蓝色是高度可变区。象在一个比对序列中的字母用不同的颜色标记一样，Cn3D 根据保守性使用各种各样的简单的标色算法。这些选项在 **Style:Coloring Shortcuts:Sequence Conservation** 子菜单中：

- **Identity** —— 统一序列红色，另外蓝色。
- **Variety** —— 从红到蓝的的等级依赖于组中不同残基的数目，与基团类型无关。

例如，四个一列四个残基，最保守的序列是单一的类型 (如: "AAAA")，然后是有二

种类型的("AAAV"), 之后是三种类型的("AAVW), 保守性最差是有四种类型的("AVIW")。所以使用四种颜色区分: 第一类 =红, 第二类 =青红, 第三类 =微红蓝, 第四类 = 蓝。每个空位是一种例外的类型。

- **Weighted Variety** —— 与上面相似, 只是"variety"用一个替代可能性矩阵(BLOSUM62)所加权。因此相似残基组(如"VVIL")会被认为比不同残基组("AAWK")有更大的保守性(颜色更加红), 尽管事实上"variety"是相同的。(在这个例子中权重值是3)

- **Information Content** 也利用一个 BLOSUM62 matrix 去计算保守性分值。用对数值观察在每组中残基期望的频率

- **Fit** 有一点不同, 它说明比较蓝的残基显示和组中其他的基团的保守性不能匹配。例如 一个组中的赖氨酸与缬氨酸)。

利用这些配色方案, 可以很快发现在结构和比对中哪些区域更保守(更红)。举个例子来说, 通常非常保守的残基都分布在活性位点周围, 通过这种方法可以很清楚的看到这一点。

Annotating and Saving Images from Cn3D

注释结构

这一章讲述了 Cn3D 的一些特点, 如可以用[标签](#)和[绘制样式](#)来注释结构、可以创建静态的图象文件、还可以将工作[保存](#)为 Cn3D 格式(ASN1)的数据文件, 以便在其他时候重新加载、或者在电子出版物中用作交互式图象。

骨架和界标

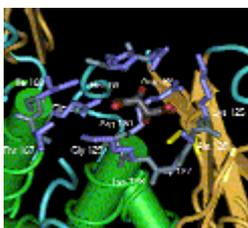
样式面板上的 **Labels** 页 (见 **Style>Edit Global Style**) 可以将原文标注自动的加到氨基酸或核苷酸链的不同位点上。Spacing(间距)表明注释之间的距离, 如每五个残基, 每二十个残基等等。Type(类型)说的是要用一个字母还是三个字母的残基缩写, Numering(编号方式)是让选择连续的数字(1---N), PDB 赋值的残基数目, 或是没有。如果选中 **Contrast with backbone**, 标注就会在背景中呈现可见的颜色, 否则便呈现残基主链的颜色。这里也可以控制链末端(Termini)和金属离子(Metal Ions)的标注。

当可以自动标注的时候，特别是对于那些较大的结构，就会有使用者想标注并突出一些个别的、他自己选择的残基，这些情况将在下一节中讲述。

自定义样式

假设 Cn3D 正被用来创建一个描绘某蛋白活性位点的图象。使用者可能想突出那些已知具有结合和催化功能的特殊氨基酸，那么用标注和对比绘图样式就能将这些残基与结构中其余部分明显的区分开。这种专门的注释是通过 **Style:Annotate** 对话框完成的。它允许使用者将标注和绘图设置（不同于 **global** 设备）用于在序列或结构窗口中已选择（高亮显示）的残基。

让我们再回到人 PTEN 晶体结构。Cn3D 可以被用来创建一个有点象 [Lee et al., 1999](#) 中的 2D 图形的图象以显示抑制剂周围侧链上的活性位点。



... 或[在此处单击就可以启动 Cn3D 并显示这张图像](#)

要创建一个专门注释式样可按下列程序进行：

1. 在序列窗口选择或在结构窗口中双击所要注释的残基。
2. 到 **Style:Annotate** 对话框，点击 **New**，将这种特征的名称和描述填入其中。点击 **Edit Style** 调出 **style** 面板，这时就只运用于所选择的残基。调整绘图和标注样式到所要的状态，可能与用于非特殊注释状态下的一切事物的整个体系（样式应用于全局而不是单一的注释）做对照，很快变化就会在结构窗口中显示出来。
3. 点击 **style** 面板中的“**Done**”然后点击“**ok**”显示注释。

上面的例子讲述了怎样使用 **Annotate** 对话框注释那些特殊的残基，同时还包含了发生接触反应的半胱氨酸的黄色的二次注释。

选择左边列表中的一个特征然后敲击 **Edit** 然后选择 **Edit Style** 调出某个注释的样式面板就可以改变那个注释的绘图和标注设置；确认设置成所要的状态后敲击“**Done**”就可以保存并显示新的设置。**Delete** 删除一个注释，**and Move** 会把注释设置该为当前高亮显示的残基的设置。选择右边列表中的个别注释然后敲击 **Turn Off** 可将它们隐藏，要再显示就在左边列表中选择并敲击 **Turn On**。

保存空间结构和图像

在 **Cn3D** 中有两种文件输出形式：静态图象和数据文件。这两种形式各有优缺点，正如下文所讨论的。这些特征在这篇文档中被用来创建了许多图象和交互式图形。

File:Export PNG 命令可将当前结构描绘区的内容保存至 **PNG 格式** 的图象文件中，这种图象文件会默认与原始图象相同的像素尺度，但是却具有更高的分辨率，比如用作印刷拷贝。这些文件可以被大部分的网络浏览器和文字处理器打开并显示，而且它们还具有很好的数据压缩和完全的色彩支持，是电子出版物的理想选择。

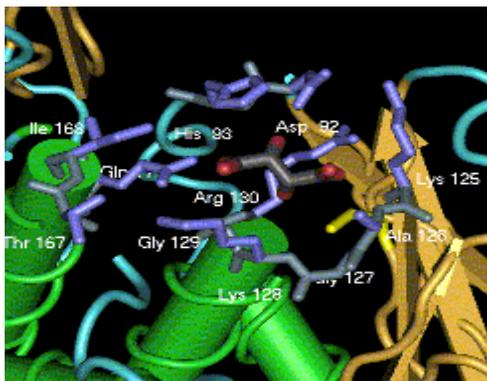
File:Save 命令可将当前视图保存至 **ASN1** 数据文件中，这种格式时 **Cn3D** 可以是别的数据格式。所得的文件不仅包含了原数据中所有的结构和序列信息，还有当前模型的定位（旋转/缩放），所有输入的附加序列，全局绘图和标注设置，还有前面所述的使用者自定义的注释。使用这种文件格式主要有下面两种目的：第一，将结果保存在文档中以便可以用文字处理器或其它编辑软件来保存并在后续环节中得到恢复。第二，使用者可以将以此种形式保存的 **Cn3D** 信息当作交互式图形使用，这一点可能是更重要的。下一节我们将以此为主题。

将 **Cn3D** 用于交互式图形

在本指南中多次出现了“交互式图形”的例子。交互式图形指的是那些与可下载数据相链接的图象，当将它们导入 **Cn3D** 时，可在 **Cn3D** 界面的交互式环境中复制。这样人们就可以准备一个静态图象用做数字出版物的插图，而在 **Cn3D** 中使用者还可以对该图形中蛋白的结构和序列作进一步的研究。当程序加载了这个数据，初始的视图将会与下边的静态图像相同，这样给了用户一个好的开端来仔细研究当前给出的信息。

这个很容易做到：在不改变结构的情况下，用 **File:Export PNG** 和 **File:Save** 命令各创建一个静态图象和一个数据文件，然后在某一可以将 **Cn3D** 与数据文件名作为命令行的上下文中将数据文件与图象相链接。

在网页中，可以单纯的用锚定链接机制将数据与图象连接起来。就拿上面的例子来说，点击图象将相同的图象导入到 **Cn3D** 中：



更加复杂的部分是让网络服务器发送正确的 mime (Multipurpose Internet Mail Extension protocol, 多用途的网际邮件扩充协议) 类型的数据以使用户的浏览器可以正确的启动 Cn3D 来下载数据。NCBI 目前使用 [Apache](#) 程序, 它是由 .htaccess 文件操作的, 使用的命令是:

```
AddType chemical/ncbi-asn1-binary .prt
```

这通知 Apache 发送以 .prt 结尾的 mime 类型 **chemical/ncbi-asn1-binary**, 这样就可以使网络浏览器启动 Cn3D。其他的网络服务器也可以提供同样的服务, 但可能基于不同的原理; 参考服务器的说明文档可以获得更多的信息。

Cn3D Advanced Topics

高级设置

这一章将讨论 Cn3D 软件中的各种高级设置

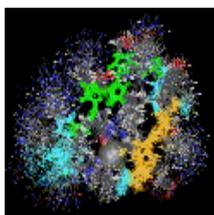
- [动画设置](#)
- [显示无序晶体](#)
- [导入用户的 PDB 文件](#)

动画控制

Cn3D 提供的动画设置可使用户既能人工得也能自动地浏览一整套模型的空间结构。这是一种简便的观察从一整套模型中单个分子的结构, 这些模型既可以来

自一大列蛋白中的几个,也可以是有多种模型的核磁共振图谱(NMR)或是一个混晶结构中的(“混晶”详见[下一部分](#))。

比如,下图所示的是 ZNF 的 MMDB 结构图.这是一个有 37 个亚基的锌指结构核磁共振图。图中每个锌离子都和两个半胱氨酸和两个组氨酸结合。当从 MMDB 下载多个结构时,为了看到 NMR 图中整个模型,请务必将“All Models”选项选定。



... 或[在此处单击](#)就可以启动 Cn3D 并显示这张图像

动画设置在 **View:Animation** 子菜单中。点击 **Play Frames** 浏览所有亚基; 动画播放的速度由 **Set Delay** 参数控制,参数值(百万分之一秒为单位)越小, 动画播放的速度越快。按下 **Stop** 键即可停止动画播放。

你也可以用 **View:Animation:Spin** 菜单项来自动旋转, 观察图像。

显示无序晶体

一些 X 衍射晶体图是通过对杂原子寻找多个匹配来处理杂基团的。这些构象异构体可以通过 Cn3D 来观察。当用动画形式运行时, 可以快速地同步显现所有这一结构中的杂基团。

比如,当尝试把 MMDB 入口号为 3LZT 的鸡溶菌酶的结构图加载到 Cn3D 时, 为了看到所有的构象异构体,在下载数据之前请把 MMDB 结构摘要页面中的” All Models”选项选中。

这一结构中蛋白表面的多种残基含有三套匹配体, 用户同样可以通过[上面](#)提到的动画控制功能来浏览这些结构。这可能是观察杂原子的最佳途径了—特别是使用 **Style:Rendering Shortcuts:Wire** 菜单项来观察侧链时。

PDB 格式并不区分有相互联系和没有相互关系的杂原子。Cn3D 则把所有的杂原子作为有相互关系的格式来处理。在某些情况下, 一套构象异构体是不够的, 就像 3LZT, 对于有些基团, 作者只提供了两套匹配体, 而其他的则为三套。Cn3D 将选择其中最为权威的一套进行填补。这就意味着如果要全部三套进行显示, 而有些基团只有两套时, 那么与之相配的例外一套将从其他地方借用过来补充这些基团。否则的话, 有些原子将会因为提供的数据不够而无法显示。

类型搜寻

Cn3D 4.0 提供了类型搜索引擎，以便你能按照特定的顺序去发现和锁定你要的基团。这些工作你可以通过在队列窗口和输入窗口中的 **View:Find Pattern** 子菜单来完成。只需简单地将类型规格([ProSite syntax](#) 中)输入到对话框中，在窗口中所有的非重复的匹配序列将会被显示。

导入用户 PDB 文件

某些用户可能想用 Cn3D 观测他们自己定制的 PDB 文件。Cn3D 不能直接导入 PDB 文件，而是利用 NCBI 网站提供的服务将其转化为 ASN1 格式数据。注意，有时对那些不太符合标准的 PDB 文件，从模型里或其他化学软件里输出不符合标准的文件，尤其是非自然残基存在时，处理效果就不能保证了。用不用这些数据就要看你自己的了。

通过[VAST search](#)网页，用户可以自由地将一个结构图上传到 NCBI 网页，在那里将其转化为 ASN1 数据。这一网页的功能实际上使用户可以同 MMDb 进行结构比较。数据结果和序列可以以 ASN1 数据的方式下载，并通过 Cn3D 察看。

Alignment Editing

使用编辑器

Cn3D 有一个功能齐全的比对编辑器，当你需要改变比对，比如[修改错误](#)，或是[增加新的序列](#)，都可以用到它。在编辑时有一些基本的规定需要遵守。你不能改变比对本身，而且[Cn3D 的比对模块](#)是固定的，也就是说，只允许没有间隔的比对区块。除此之外，用户可以利用以下所列的操作随意地更改比对。

比对编辑器既可以在比对窗口也可以在输入窗口中运行。在比对窗口中打开编辑器，用 **Edit:Enable Editor** 选项；编辑器在输入窗口中就会始终开着。

改变比对

比对编辑的基本操作就是简单地在比对区域中点住后拖动残基。这只会在一个区域内移动那一排的比对。这需要将 **Mouse Mode:Horizontal Drag** 菜单项选中，这一选项在编辑器运行时默认是选中的。

当对序列进行修改后，有一个 **Edit:Undo** 菜单项，可以撤销你可能的错误操作，你可以撤销从编辑器最近一次激活开始算的最多 50 次操作。

以下也会介绍改变比对区域数量和大小的方法。

区域指示行

当编辑器打开时，你会在比对的顶部发现还有一被标记行，用来指示比对区域的边界。标记“<”表示区域的左边界，“-”表示中间，“>”表示右边界，如果区域只有一个残基的宽度，则用“^”表示。

如果想将区域变小或变大，你可以简单得点住编辑标记“<”或“>”来拖动。当使区域变小的时候，残基字母会变成小写并添加到相邻关联区。而当区域变大的时候，你会看到更多的残基会被加到区域中来，因此字母会变成大写，有时序列区域图案会被着色。因为在序列区域中不允许存在间隔，可能就会对你区域大小产生限制，这取决于每一条序列中有多少可用非关联的残基在附近的非关联区。

区块操作

在编辑器中有四种区块操作方式：

- **Split block**:选择这一项后在序列区块中点击一个纵列，已有的区块会变切割成两个相邻的区块，而新的区块则从那一纵列开始。

- **Merge blocks**:点击这个选项然后点住鼠标在任何两个或更多的相邻关联区块间拖动。这些区块会融合成一个区块。

- **Create block**:选择这项后，点住鼠标在任何栏里的非关联残基之间拖动，如果这一比对里的每一条序列都由被选栏里的关联残基组成，那么一个新的区块就会由这些残基组成。

- **Delete block**:选择这一选项并点击比对区块的任何地方。被点击区块会被删除，而区块内的基团就会被添加到临近的非关联区。

行操作

用编辑器里 **Edit>Delete Row** 菜单选项你可以删除一行，当 **Mouse Mode:Move Row** 菜单项被选中时，你可以将行上移或下移，而且你也可以根据不同的标准来分类行，详见关于 **Edit:Sort Rows** 算法的内置帮助文件。

校正比对错误

.如果在多重比对中出现错误，根据错误的严重性，有连种基本方法可以校正。如果是轻微的错误，可以使用以上所说的手动操作解决。但有时在比对中的一条特定序列出现了非常严重的错误以致必须对这条序列重新排序。

为了使运算法则（就像 **BLAST**）运用到重排整条序列中，只有把序列转移到运算法则可用的输入窗口中。这一步是通过使用在比对窗口中 **Imports:Realign...** 菜单项来完成的。比如，当选中 **Imports:Realign Individual Rows** 菜单项，你可以点击任何的序列（除了主序列），被点击序列就会从多重比对转移到输入窗口中，并与主序列配对。

一旦转移到输入窗口后，就可以利用 **Cn3D** 中自带的各种算法来重排序列，或是和以往一样，手动地改变比对。

当序列在输入窗口中被排列完毕后，就可重新合并到多重比対中去，这一步将在下一部分介绍。

将序列合并到多重序列中

来自输入窗口的序列可作为新的行列加入到比对窗口中的多种序列。这一步叫做合并，可运用于[输入的序列](#)和从多种比対中删除的需要重排的序列（见[上文](#)）。

合并有时会很复杂，因为需要合并的成对比对必须和多重比対中的关联区块相容。也就是说，如果多重比対有一区块是从基团 **10-20**（相对于主序列而言），那么配对的需要合并的序列也必须有含从基团 **10-20** 的区块。

当一个新的序列加入进来时，为了保留区块结构，这些规定是必须的。在 **Cn3D** 的比対模型中，关联区块反映了比対的核心（通常和蛋白质的结构核心相对应），对于新加入到多重比対的序列，**Cn3D** 也运用了这种核心模型。如果，在主序列上的基团 **10-20** 呈现了一个 α 螺旋核心，那么就可以断定这一家族的任何成员都带有相应的 α 螺旋核心，因此比対的所有横列都有一个包涵这些基团的区块。

Cn3D 建议用户对于可合并的比対，把输入窗口中的任何配对序列区的背景设置成红色，以防序列被合并到多重比対中去。用户需在合并之前对染红的配对比对的区块结构做必要的修改。

同时你可以通过对多重比対的修改使配对的比対合并。尽管这种方法用的并不多，但是可能出现输入窗口中一族蛋白的核心模型确实存在错误，而需要修改多重比対的情况。这由用户自己决定，在设计上，**Cn3D** 不会自行做这样的决定去改变这些合并规则。

一旦有一个可合并的比对，而又它的配对比对没有着红的基团，那么选择 **Alignments:Merge Single** 菜单项，然后点击任何配对比对，该比对就会从输入窗口中删除而添加到比对窗口中的多重比对的最后，作为新的一行。你也可以用 **Alignments:Merge All** 菜单项合并输入窗口中所有的比对。
