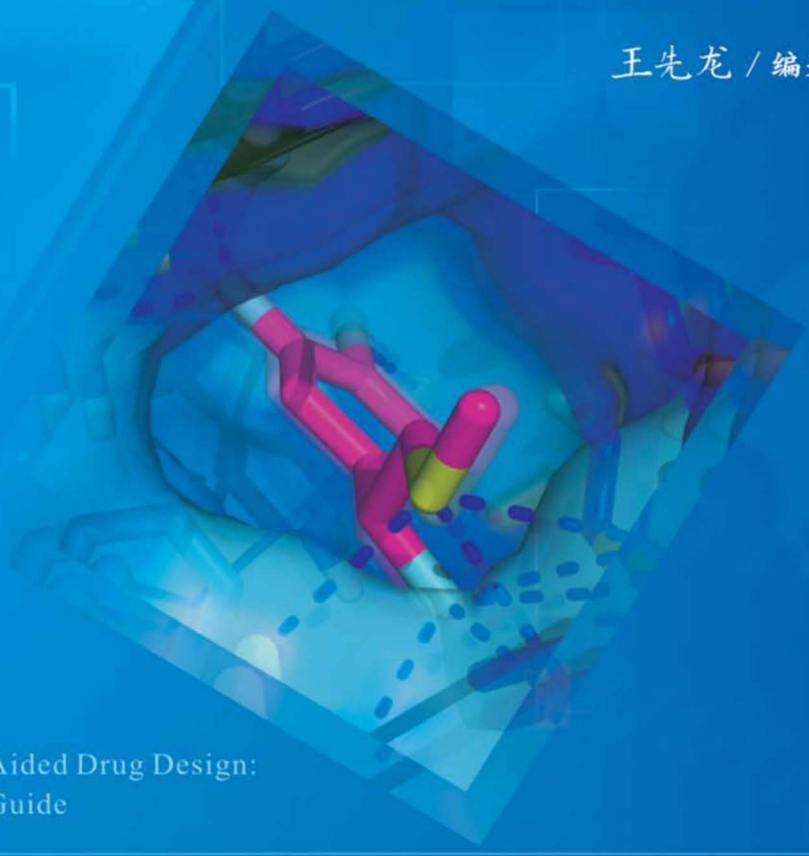
计算机辅助药物设计

实践指南

王先龙 / 编著



Computer-Aided Drug Design:
A Practice Guide

电子科技大学出版社

图书在版编目（CIP）数据

计算机辅助药物设计实践指南 / 王先龙编著.
—成都：电子科技大学出版社，2016. 6
ISBN 978-7-5647-3505-0

Ⅰ. ①计… Ⅱ. ①王… Ⅲ. ①药物－计算机辅助设计
－指南 Ⅳ. ①R914.2-39

中国版本图书馆 CIP 数据核字（2016）第 048608 号

计算机辅助药物设计实践指南

JISUANJI FUZHU YAOWU SHEJI SHIJIAN ZHINAN

王先龙　编著

出　　版：电子科技大学出版社（成都市一环路东一段 159 号电子信息产业大厦　邮编：610051）
策划编辑：高小红　李述娜
责任编辑：刘　愚
主　　页：www.uestcp.com.cn
电子邮箱：uestcp@uestcp.com.cn
发　　行：新华书店经销
印　　刷：四川煤田地质制图印刷厂
成品尺寸：175mm×230mm　　印张 21.25　　字数 415 千字
版　　次：2016 年 6 月第一版
印　　次：2016 年 6 月第一次印刷
书　　号：ISBN 978-7-5647-3505-0
定　　价：48.00 元

序

计算机辅助药物设计（Computer-Aided Drug Design, CADD）是以药靶互作理论，化学信息学，计算化学，组合数学，概率统计等学科为基础，采用高性能计算来设计，筛选和优化药物分子的过程，包含众多不同方法和应用侧面。它以理论科学为基础，弥补实验科学所面临的不足，但它并非取代理论科学或实验科学，而是作为两者之间的桥梁和重要补充。

计算机辅助药物设计的出现和广泛应用有其必要性和条件充分性。药物发现是一个漫长而高风险的过程，近些年的统计表明，一种全新药物的研发大致需要 12 年和 10 亿美元的投入。但目前的药物产出与研发投入并不成正比，一方面，研发投入大幅度提高，而获批上市的药物数量并没有增加。因此，市场的驱动力要求采用创新性的方法来提高研发产率。在技术上，20 世纪 40 年代出现的电子计算机以 Moore 定律高速发展，使得计算科学发展成为独立于理论和实验科学之外的第三支柱。在理论上，量子力学和经典牛顿力学的发展及其在分子模拟中的应用已相当成熟，形成计算化学这一学科，以 1998 年 Pople 和 Kohn，2013 年 Karplus, Levitt 和 Warshel 获得诺贝尔化学奖为标志性事件。在实验技术上，高通量实验技术提供海量数据促使计算辅助手段的发展和应用，形成各种组学，这以人类基因组工程和蛋白质组工程为代表。

计算机辅助药物设计的进化历史大致可分成四个阶段。（1）早期萌芽阶段。20 世纪 60 年代，美国化学会采用计算机来处理化学文摘数据，另外一些公司也构建了化合物结构性质的数据库。定量构效关系（QSAR）方法开始出现并应用在药物设计中。（2）快速发展期。20 世纪 70～80 年代，分子力学方法发展迅速，人们可以在较短时间内计算分子结构和性质，模拟蛋白质的分子动力学。美国 Gordon 会议出现了计算化学专题；Merck 公司科学家发表文章，声称“可以逐个原子地设计药物”。（3）发展低潮。20 世纪 90 年代，计算机辅助手段并没有如人们所期待那样带来革命性变化，大家对计算机辅助药物设计的认知回归理性：计算手段可辅助药物设计，并不能完全取代实验。（4）理性发展。21 世纪以来计算化学和化学信息学持续发展，与实验结合，普遍应用，成为各大制药公司每一新药研发中不可或缺的环节。整个新药研发是一循环上升的过程，计算辅助方法合理应用可大幅提高研发效率，节约成本。现在上市的新药在研发过程中没有计算机辅助药物设计的身影是件不可思议的事情。

由于计算机辅助药物设计的重要性，许多高校都开设了相关课程。由于计算机辅助药物设计既涉及计算机应用和编程知识，又要掌握许多物理，化学，生物等学科的理论知识，对大多数初学者来说，学习曲线坡度陡。特别是现在的研究生有着不同的学习背景，知识差异大，缺乏很多必要的知识准备。本书由作者在多年讲授“计算机辅助药物设计”和“计算机辅助药物设计综合实验”两门课程的基础上编写而成，希望为初学者提供较低的学习门槛。因此本书内容编排上强调实用，上手容易，通过许多具体的实例来讲解软件或技术的使用，而理论知识方面市面上已经有一些很好的教材或专著可选。但由于计算机辅助药物设计内容非常丰富，不可能面面俱到，本书只采撷了几点常用的技术来介绍，希望能帮助初学者叩开计算机辅助药物设计这一领域的大门。

本书适宜用作研究生，高年级本科生“计算机辅助药物设计”和其他一些相关课程的理论和实验课程教材，也可供青年学者自学使用。用作实验课程教材时，建议每章至少安排四学时。教师应在教学活动开始前，在一台服务器上安装好操作系统和相应的软件，并准备好相关的数据，学生在机房实习时，通过第一章介绍的方法访问服务器。Linux 操作系统方面，作者推荐使用 Gentoo，该分发版本具有极高的定制自由度，并能自动解决软件依赖关系问题。

受眼界和能力所限，本书不可避免地存在不少问题和缺陷，希望读者在使用过程中，将改进建议和意见反馈给作者。

王先龙

2015 年 12 月于成都

致　谢

本教材出版承蒙电子科技大学教务处，生命科学与技术学院等单位资助，并受到相关领导关心，特此致谢。

封面中文书名由黄体强先生题写，特此感谢。第 6 章部分内容由何彬硕士论文改编而成，特此说明。姜云峰，陶韵文等同学参与第 4 章部分内容早期版本的写作；张鹏，唐令利，朱小娟等同学参与了校对工作；许多选课同学在讲义使用过程中也提出了许多宝贵意见，在此一并感谢。

本着推广开源软件的精神，本书使用开源软件 LibreOffice 编辑和排版，并使用开源字体。中文标题使用文泉驿字体，中文内容使用 AR PL 明体，西文使用 Liberation 字体。

特别感谢电子科技大学出版社高小红和刘愚编辑，没有他们的大力支持和极大的耐心，本书无法付梓。

目　录

第 1 章
Linux 操作系统基础知识
工作环境入门

1.1 Linux 操作系统简介

由于大多数的读者可能仅熟悉 Windows 操作系统图形界面的使用，但对于计算机辅助药物发现和材料设计等科学计算任务来说，Linux 及其他多种类 Unix 的操作系统更方便，应用也更普遍，因此读者有必要了解 Linux 操作系统的基本架构和特性，掌握常用的命令。针对这一需求，本章初步介绍 Linux 操作系统的使用，内容以实用为出发点，面向零基础读者，不追求全面和系统性。在掌握了本章基本操作后，感兴趣的读者可参考其他资料进一步学习，甚至学习编写科学计算软件。

Linux 操作系统衍生于 Unix 操作系统，后者 1969 年诞生于美国 AT&T 公司。由于 Unix 受知识产权保护，不能自由使用与修改，因而在 20 世纪 80 年代和 90 年代初诞生了基于自由软件和开放源代码的 Linux 操作系统。1991 年 10 月，芬兰裔计算机科学家 Linus Torvalds 发布了第一版 Linux 内核（Kernel），标志着 Linux 的正式诞生。初期 Linux 操作系统是基于 Intel x86 架构，但发展至今，Linux 和其他多种类 Unix 操作系统在各种应用场合大显身手，包括大型计算机集群，个人桌面计算机和笔记本电脑，更以嵌入式系统形式存在于汽车，通信设备和智能家用电器中，如各种基于安卓（Android）和 Tizen 操作系统的智能手机，平板电脑，游戏机和电视机等。而在面向科学计算应用需求的超级计算机（Supercomputer）领域，Linux 更是占据绝对的领先地位。如 2015 年 6 月发布的第 45 个超级计算机 TOP500 榜单中的前 10 名全是使用 Linux 操作系统[1]。苹果公司也在 2000 年左右购买了一款商业 Unix 操作系统 NextStep，推出了基于 Unix 内核的 OS X 操作系统。由于 Linux，Unix，OS X 以及其他类 Unix 操作系统在架构和命令集上有很多相似性，掌握了 Linux 操作系统也有助于学习其他操作系统的使用。

微软公司的 Windows 系列操作系统在个人电脑市场应用的比例很大，其主要优势是使用简便，商业软件支持较好。相比较，Linux 操作系统的优点，特别是在科学计算应用中有很多，以下是一些典型优势。

- 免费。网络上有各种版本的 Linux 操作系统以及各种应用软件，都可以免费获得。完全可以取代商业操作系统和应用软件。
- 自由。大多数的 Linux 操作系统和应用软件都是基于 GPL 等开源代码共享协议，用户享有充分的修改和进一步开发的自由度。比如国内一些单位利用开源代码开发了国产的 Linux 操作系统和办公软件。

- 安全。Linux 是最为安全的操作系统之一，对各种病毒和恶意软件及网络攻击都具有很好的免疫能力和防范措施。特别是代码的透明性是杜绝后门软件的有力保证。
- 稳定。Linux 不需要经常性重启而可以保持系统的高效性。不会因使用时间的延长造成内存泄漏而变慢或死机。即使是硬盘在全满的情况下仍能稳定运行。
- 高效。可支持多用户多进程同时高效运行，充分利用 CPU 等硬件资源，高效率地完成逻辑和数值计算，并保持系统的稳定性。
- 多种选择。在解决各种需求中，Linux 系统通常存在多种解决方案，用户可根据使用习惯和特性需求选择最适合的方案。比如，图形化桌面系统，有功能丰富而视觉效果突出的 Gnome 和 KDE 系统，也有轻量级资源需求极低的 Xfce 和 Flubox 等系统。

过去，Linux 系统的安装被认为是个比较困难的过程，特别是解决各种硬件驱动模块上需要大量的用户干预。但现在，主流的 Linux 操作系统，如 Ubuntu 和 Fedora，安装过程几乎和安装 Windows 系统一样的方便。而桌面系统功能很全面，使用也非常方便，无须了解 Linux 命令的使用也能完成大部分的办公事务。然而，对于科学计算应用而言，只有掌握一定的 Linux 命令和脚本语言（Shell script）的使用，才能充分利用 Linux 系统的特点，把硬件和软件性能发挥到极致，更好地解决我们的科研问题。

古老 vs.现代

Strong representatives from each past era thrive today,such as programming in the thirty-year-old language known as Fortran, and even in the ancient script known as direct machine code. Some people might look on such relics as living fossils; others would point out that even a very old species might still be filling a particular ecological niche.

Alan Kay, Sci. Am. September 1984

Kay 这段二十多年前的评论说的是 Fortran 语言，由于其数值运算的高效性，至今仍在科学计算中占重要地位。这段话也适用 Unix 操作系统。历史在螺旋式上升，有淘汰，有保留。

1.2 Linux 操作系统结构

从用户角度来说，Linux 操作系统架构大致可分为如图 1-1 所示的 5 个层次：硬件平台，内核（Kernel），外壳（Shell），系统实用程序（System Utilities）和用户应用软件。前四层为操作系统的主要部分，支撑用户应用软件的运行。

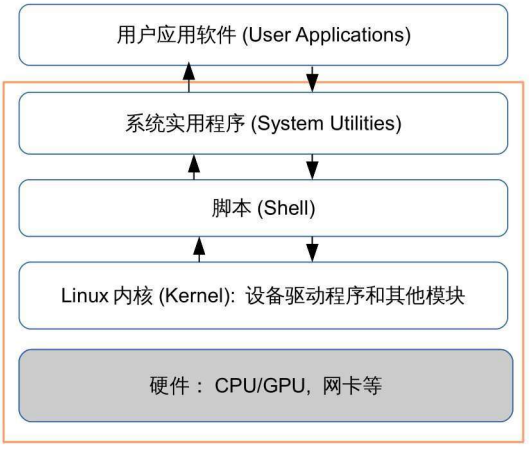


图 1-1 Linux 操作系统架构示意图

内核

内核是操作系统和物理硬件资源交互的接口层。它是操作系统最为主要的部分，开机启动后首先加载到内存中。它也是应用程序和支撑程序运行数据间的桥梁。内核负责任务管理，内存管理，磁盘管理等工作，主要功能是管理计算机资源，为其他需要使用资源的程序提供支撑。

外壳

外壳是用户与操作系统间的接口，为用户使用内核所提供的服务提供软件。它包括两部分：命令行（Command Line Interface，CLI）和图形界面（Graphic User Interface，GUI）。

命令行

这又通常称为狭义的 Shell，它接受和执行命令。常见的命令行系统有 Bash，Tcsh 等，在语法上稍有不同。使用命令行工作效率较高，而且在由于图形卡不工作或其他原因导致不能启动图形界面的情况下仍能使用。

图形界面

图形界面方便用户操作，用户无须记住命令。Linux 系统中广泛应用的图形界面是 X 窗口系统，又称 X11。在 X11 之上又有多种桌面系统，如 Gnome，KDE，Xfce 等。

系统应用程序

包括各种系统中断（System interrupt）和调用（System call）。实现系统在执行用户应用程序和系统进程两种状态间切换。

用户应用软件

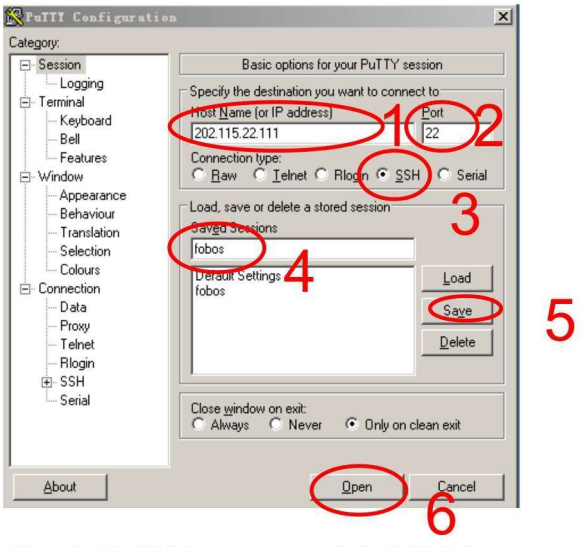
完成用户需求任务，如办公用的 OpenOffice 套件，图像处理软件 GIMP，高性能计算使用的分子动力学软件 Gromacs 等。内核为这些应用程序提供进程，支持其高效稳定运行。

1.3 Windows 系统与 Linux 系统通信方法

通常情形下，日常办公的电脑使用不同版本的 Windows 操作系统，而计算用的服务器运行着 Linux 系统。由于计算服务器噪声较大，对环境要求较高，一般会放在信息中心等专门的机房中。用户需要从远程登录服务器，办公电脑扮演着终端的角色。从用户终端到服务器的通信方式有多种，这里介绍两种方式：模拟终端和 VNC 虚拟 X 服务器。前者是命令行方式，后者是图形界面方式。

模拟终端

我们推荐使用开源免费的 PuTTY 软件。该软件可在 http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html 网页免费下载，文件大小仅几百千字节。该软件提供了 Telnet 和 ssh 两种服务器连接方式，两者都需要在服务器开启相应的服务（Service）。后者（ssh）使用基于 ssl 加密算法，连接更安全，在 Linux 服务器上应用普遍。ssh 连接需要服务器开启 sshd 服务（通常位于/etc/init.d/ sshd）。PuTTY 软件使用很简单，步骤如图 1-2 所示。



1 输入服务器 IP 地址或域名　　**4** 输入要保存为 Session 名称
2 确认端口号（ssh 默认使用 22）　　**5** 点击保存设置
3 确认使用 ssh 连接类型　　**6** 点击 Open 启动 Session

图 1-2 PuTTY 配置界面

双击 PuTTY 图标后，首先出现如图 1-2 所示的配置界面。在输入 IP 地址后（其他采用默认值），我们可以保存相应的配置：在标记 4 位置内输入要保存的名称，然后点击 Save（保存）即可。如果我们的配置信息已在标记 4 下方的列表中，双击相应的名称即可开启连接；也可先选择相应的名称，单击 Load（加载），再单击 Open 开启连接。

在首次登录服务器时，会出现如图 1-3 所示的有关 rsa2 指纹警告信息，选择“是”（Yes）即可进入服务器用户登录界面。

服务器用户登录过程见如下示例。首先是输入用户名，回车后输入密码。如登录成功会出现系统的欢迎信息，光标停在$符号之后，等待用户输入其他命令。如果用户名或密码错误，在等待几秒钟后会提示 Access denied 信息，再次输入密码。

```
login as: xwang
xwang@jordan's password:
Last login: Mon Sep  2 13:57:05 2013 from mercury.uestc.edu.cn
[xwang@jordan ~]$
```

图 1-3 PuTTY 登录新服务器出现的 rsa2 指纹警告信息

Notes

在 Linux 平台上，输入密码过程中不回显，输完后回车即可。在输入过程中可以使用回格（Backspace）键修改。

此外命令行提示符（上述示例中的“[xwang@jordan ~]$ ”部分）在不同系统中可能有所不同，具体形式取决于环境变量$PS1 的设置。用户可以修改。下文中会介绍如何修改系统环境变量。

在成功登录服务器后，我们就可以使用系统中任何非图形界面命令，除非另有权限限制。如果有多项作业同时进行，我们可以开启多个连接，称为 Sessions。

退出登录，使用 exit 命令。

```
[xwang@jordan ~]$ exit
```

VNC 虚拟 X 服务器

VNC 全称为 Virtual Network Computing（虚拟网络计算），它提供了一个方便的跨平台远程桌面共享操作通道。VNC 由两部分构成：客户端（Client）和服务器（Server）。采用 RFB 协议，客户端把用户鼠标，键盘操作等事件通过网络传输给服务器端，然后服务器把图形显示更新内容回传给客户端。因此，使用 VNC 方式登录服务器的先决条件是 Linux 服务器上已安装 VNC Server，并开启该服务。常见的 VNC 软件有RealVNC，TightVNC，TigerVNC 等，都分为客户端和服务器端。客户端一般又称为 VNC Viewer。下面的讲解以 TightVNC 为例，其他软件大同小异。TightVNC 可在 SourceForge 网站下载（http://sourceforge.net/projects/vnc-tight/）。

VNC Server 启动和关闭

VNC Server 可以通过运行 vncserver 命令开启，如下所示。在 VNC Server 启动后，会提示相应的配置信息。要特别注意第一行最后的数字:*n*（本例中 *n* =1，如果已有其他 Server 在运行，该数目会增加），该数字称为显示号（Display #），在后面 VNC Viewer 登录时要使用到，它对应该 Server 网络服务的端口号。实际的端口号为 5900＋*n*（本例中对应的端口号即为 5901）。

```
xwang@mercury ~ $ vncserver
New 'X' desktop is mercury:1
Creating default startup script /home/xwang/.vnc/xstartup
Starting applications specified in /home/xwang/.vnc/xstartup
Log file is /home/xwang/.vnc/mercury:1.log
```

在首次开启 VNC Server 后，我们还应该使用 vncpassword 命令设置登录密码。如下文所示。在该例中，一共输入 4 次两组密码。使用第一组密码服务器可以响应鼠标和键盘事件；而使用第二组密码只能观看显示内容，不能操作，该模式方便教学演示。此外，警告信息还表明了如果输入密码超过 8 位，实际只取前 8 位。

```
xwang@mercury ~ $ vncpasswd
Using password file /home/xwang/.vnc/passwd
Password:
Warning: password truncated to the length of 8.
Verify:
Would you like to enter a view-only password (y/n)? y
Password:
Warning: password truncated to the length of 8.
Verify:
```

VNC Viewer 使用

在 Windows 系统中，双击 VNC Viewer 应用程序图标，首先出现登录界面，如图 1-4 所示。

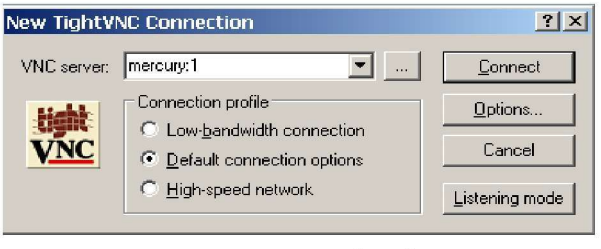


图 1-4 VNCViewer 登录窗口

在 VNC server 地址框中输入服务器的域名或 IP 地址＋:*n*。上文中机器名为 mercury，也可直接使用其 IP 地址（222.197.171.143）。冒号后的 *n* 是显示号（本例为 1）。然后点击 Connect。若连接成功，将进入图 1-5 所示的密码验证窗口。

图 1-5 VNC Viewer 连接过程中的密码验证窗口

输入密码后，即进入服务器的桌面系统。图 1-6 所示的是使用 Gnome 的桌面管理系统。在该界面下，用户可以使用鼠标和键盘进行互动操作，使用各种图形工具，方便进行分子结构查看，建模等图形类工作。

在 VNC 工作模式中，所有的计算工作和图形任务都在服务器端运行，客户端的 Viewer 只负责显示虚拟桌面上的图像和鼠标，键盘事件，因而对客户端的硬件要求和使用时的负载都非常低。我们甚至可以在智能手机，平板电脑等便携设备上管理服务器上的工作。唯一的约束条件是客户端和服务器端的网络通信保持通畅。

使用 VNC 的另一好处是在关闭 VNC Viewer 后（无论是主动关闭，或是因网络或终端断电等被动原因），服务器端的工作不受影响。重新连接后，即恢复到原先工作状态。而且一个 VNC Server 可以连接多个 Viewer，方便多人异地协同工作。